OCM 2025

7th INTERNATIONAL CONFERENCE ON

Optical Characterization of Materials

MARCH 26th – 27th, 2025
KARLSRUHE | GERMANY

J. BEYERER | T. LÄNGLE | M. HEIZMANN (Eds.)

KIT Scientific Publishing

Jürgen Beyerer | Thomas Längle | Michael Heizmann (eds.)

OCM 2025

7th International Conference on Optical Characterization of Materials

March 26th – 27th, 2025
Karlsruhe | Germany

OCM 2025

7th International Conference on Optical Characterization of Materials

March 26th – 27th, 2025
Karlsruhe | Germany

edited by
Jürgen Beyerer | Thomas Längle | Michael Heizmann

KIT Scientific Publishing

Organizer

Fraunhofer Institute of Optronics,
System Technologies and Image Exploitation IOSB
c/o Karlsruhe Center for Material Signatures KCM
Fraunhoferstraße 1, 76131 Karlsruhe

The proceedings are also available as an online version
http://dx.doi.org/10.5445/KSP/1000178356

Impressum

Scientific Publishing

Karlsruher Institut für Technologie (KIT)
KIT Scientific Publishing
Straße am Forum 2
D-76131 Karlsruhe

KIT Scientific Publishing is a registered trademark
of Karlsruhe Institute of Technology.
Reprint using the book cover is not allowed.

www.bibliothek.kit.edu/ksp.php | E-Mail: info@ksp.kit.edu | Shop: www.ksp.kit.edu

Print on Demand 2025 – Printed on FSC-certified paper

ISSN 2510-7240
ISBN 978-3-7315-1408-4
DOI 10.5445/KSP1000178356

Preface

The state of the art in optical characterization of materials is advancing rapidly. New insights into the theoretical foundations of this research field have been gained and exciting practical developments have taken place, both driven by novel applications and innovative sensor technologies that are constantly emerging. The big success of the international conferences on Optical Characterization of Materials in 2013, 2015, 2017, 2019, 2021 and 2023 proves the necessity of a platform to present, discuss and evaluate the latest research results in this interdisciplinary domain. Due to that fact, the international conference on Optical Characterization of Materials (OCM) took place the seventh time in Karlsruhe, Germany from March 26-27, 2025. The aim of this conference was to bring together leading researchers in the domain of Characterization of Materials by spectral characteristics from UV (240 nm) to IR (14 µm), multispectral image analysis, X-ray methods, polarimetry, and microscopy. Typical application areas for these techniques cover the fields of, e.g., food industry, recycling of waste materials, detection of contaminated materials, mining, process industry, and raw materials.

The OCM 2025 was organized by the Karlsruhe Center for Spectral Signatures of Materials (KCM) in cooperation with the German Chapter of the Instrumentation & Measurement Society of IEEE. The Karlsruhe Center for Spectral Signatures of Materials is an association of institutes of the Karlsruhe Institute of Technology (KIT) and the business unit Advanced Sensing of the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB.

This year again the organizing committee has had the pleasure to evaluate a large amount of contributions. Based on the submissions, we selected 33 papers as posters and talks, a plenary lecture and several practical demonstrations. The present book is based on the conference and contains extended versions of the submitted abstracts.

The editors would like to thank all authors that have contributed to these proceedings as well as the reviewers, who have invested a

generous amount of their time to suggest possible improvements of the papers. The help of Lukas Dippon and Jürgen Hock in the preparation of this book is greatly appreciated. Last but not least, we thank the organizing committee of the conference, led by Britta Ost, for their effort in organizing this event. The excellent technical facilities and the friendly staff of the Fraunhofer IOSB greatly contributed to the success of the conference.

March 2025

Jürgen Beyerer
Thomas Längle
Michael Heizmann

General chairs

Jürgen Beyerer	Karlsruhe
Thomas Längle	Karlsruhe
Michael Heizmann	Karlsruhe

Program chair

Henning Schulte	Karlsruhe

Members

Jochen Aderhold	Braunschweig
Sebastian Bauer	Boston (USA)
Andrea Büttner	Freising
Alexander Ennen	Fürth
Alexander Feil	Aachen
Kathrin Greiff	Aachen
Gunnar Grün	Valley
Robin Gruna	Karlsruhe
Tino Hausotte	Erlangen
Andreas Herzog	Magdeburg
Thomas Hofmann	Würzburg
Olfa Kanoun	Chemnitz
Anna Kicherer	Siebeldingen
Andrea Krähmer	Berlin
Julius Krause	Basel (Switzerland)
Georg Maier	Karlsruhe
Heike Mempel	Freising
Félix Salazar	Madrid (Spain)
Heinar Schmidt	Kulmbach
Matthäus Speck	Waldheim
Bernhard Zagar	Linz (Austria)

Contents

Preface i

Advanced Imaging

Correspondence between Geometric Features and Parameters of a Single Voigt Profile 1
A. Kehrein, T. Camps, D. Bejtullahu, A. Lukyamuzi, and F. Sheikh

About the parameter determination of overlapping Voigt profiles from geometric features of the spectrum 11
A. Kehrein, T. Camps, D. Bejtullahu, A. Lukyamuzi, and F. Sheikh

Transfer learning for hyperspectral image classification 23
G. Ahmeti, H. Frank, and A. Zell

Assessment of lossless compression algorithms and their performance on near-infrared spectral images 33
A. Maghmoumi, H. Shekhar, T. Scherling, N. Kroell, A. Feil, and K. Greiff

Measurement of air-water gas exchange by fluorescence imaging . 43
D. Hofmann and B. Jähne

Hyperspectral data compression and its impact on spectral signatures of water bodies 53
J. Kuester, W. Gross, S. Schreiner, A. Michel, and M. Heizmann

Surface Parametrization

Material characterization by inverse rendering 65
J. Meyer, L. Dippon, and C. Kludt

Technique for the development of parametric directional reflectance models for macrostructured surfaces 77
A. Polywka and C.F. v. Carmer

Prediction of various material parameters using a combination of non-destructive sensors and artificial intelligence 91
F. Linscheid, M. Korkisch, L. Fischer, and M. Sause

Food and Agriculture

Leveraging grounded SAM and a weakly supervised CNN filtering for enhanced leaf segmentation in automated plant phenotyping 101
L. Fiedler, F. Braig, I. Howard, R. Gruna, and J. Beyerer

Non-destructive and inline capable characterization of grapes by mid-infrared spectroscopy 115
T. Tybussek, C. Wurm, and J. Xiang

Use of VIS-NIR spectroscopy to monitor mango ripening quality with ripening indices in professional ripening processes 125
S. Wittmann, S. Goisser, and H. Mempel

Optimization of plant experiments in CEA and vertical farming through standardization and modeling 137
K. Plat

Optimizing near-infrared spectroscopy for on-line grape must quality assessment: Addressing the impact of suspended solids ... 143
P. Gauweiler, X. Zheng, L. Cornehl, A. Kicherer, and R. Gruna

On the rapid aging of roasted coffee beans – A hyperspectral freshness analysis 155
P. Menz, L. Beland, J. Moritz, and A. Herzog

Classification of overall sensory acceptability of modified-air packaged (MAP) minced pork stored at different temperatures using fluorescence spectroscopy 165
J. Schlosser, S. Münch, D. Brüggemann, and H. Schmidt

Plastics Recycling

Optimizing illumination for mid-wave infrared hyperspectral imaging in black plastic recycling 175
M. Jernej, G. Jakob, M. Jaschik, and M. Kainz

Exploring near-infrared spectra of multi-material multi-layer packaging – Findings from the PET-stream 185
N. Kuhn, D. Treitler, G. Koining, M. Mager, J. Fischer, and A. Tischberger-Aldrian

Detection of bio-based additives in plastics using NIR data: Opportunity for bio-based markers 195
M. Vogelgesang, B. Trojanowski, S. Hanstein, W. Benner, and E. Ionescu

Robust model development for HSI-based characterization of post-consumer plastics 207
L. Roming, G. Maier, and P. Bäcker

MIR measurements combined with photon-up-conversion technology to measure and identify black polymers 219
W. Becker, K. Sachsenheimer, and M. Klemenz

Demolition Waste Recycling

Impact of demolition waste powders on the microstructure of cement mortars: A comparative analysis of concrete, ceramic and mixed wastes 231
A. Tokareva and D. Waldmann

PAH detection in road surface residue using various hyperspectral imaging sensors 243
P. Bäcker, G. Maier, T. Längle, and J. Beyerer

Increasing resource efficiency using the example of natural gypsum using hyperspectral imaging and machine learning methods 257
P. Hunhold, E. Linß, and S. Nowak

Process Optimization in Recycling

Automated data acquisition method for sensor-based real-time material flow characterization of recyclable waste streams using sensor fusion: A case study 267
F. Roth, A. Deliktas, K. Rattawa, A. Feil, and K. Greiff

Adaptive architectures for semantic segmentation in the field of sensor-based sorting systems 281
F. Kronenwett, R. Lehmann, H. Zheng, G. Maier, T. Längle, and W. Karl

Bulky waste classification from a distance: Challenges and first insights 295
F. Blum, P. Meyer, T. Lange, M. Trost, and T. Tiedemann

Towards continual learning with the artificial neural twin applied to recycling processes 305
R. Mendez, A. Maier, and J. Emmert

Supervised and unsupervised textile classification via near-infrared hyperspectral imaging and deep learning 319
M. Kainz, J. K. Krondorfer, M. Jaschik, M. Jernej, and H. Ganster

CNN-based copper reduction in shredded scrap for enhanced electric arc furnace steelmaking 329
G. Koinig, M. Neubauer, W. Martinelli, Y. Radmann, N. Kuhn, T. Fink, E. Rückert, and A. Tischberger-Aldrian

Correspondence between geometric features and parameters of a single Voigt profile

Achim Kehrein, Drilon Bejtullahu, Andrew Lukyamuzi, and Faraz Mehmood Sheikh

Rhine-Waal University of Applied Sciences,
Marie-Curie Str. 1, 47533 Kleve, Germany

Abstract A Voigt profile uses four parameters to model a spectral line: central location, intensity, shape parameter, and width parameter. The analysis of a line spectrum must determine the four parameters for each spectral line, so that the superposition of the Voigt profiles matches the measured spectrum sufficiently. For this indirect approach, general fitting algorithms may be used, which are computationally demanding and their success often depends on initial guesses of the parameters. In this paper we study the correspondences between geometric features of a Voigt profile and its parameters. The geometric features are the maximum point, the inflection points, and the area under the spectral line. The correspondences form the foundation for the design of Voigt-specific algorithms for analyzing line spectra. More efficient algorithms support sustainability and enable more real-time applications.

1 Introduction

A spectral line can be modeled by a Voigt profile using four parameters. We develop the parametric model in several steps. First, the centered Voigt profile

$$V(x;\sigma,\gamma) = \int_{-\infty}^{+\infty} G(x-z;\sigma)L(z;\gamma)\,dz \tag{1}$$

is the convolution of a centered Gaussian and a centered Lorentzian,

$$G(x;\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}} \quad \text{and} \quad L(x;\gamma) = \frac{\gamma}{\pi(x^2+\gamma^2)} \tag{2}$$

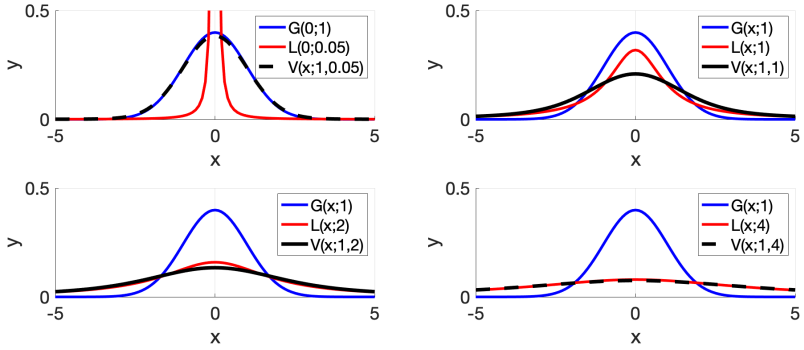


Figure 1: Some standardized Voigt profiles (black) $V(x;1,\rho)$ with the standardized Gaussian (blue), $\sigma = 1$, and the corresponding Lorentzians (red) with $\gamma = \sigma\rho$.

with width parameters $\sigma > 0$ and $\gamma > 0$. Kehrein and Lischtschenko [1] showed that the shape of a Voigt profile depends only on the ratio $\rho = \gamma/\sigma$, which is called the shape parameter. A standardized Voigt profile scales into general Voigt profiles via the width-scaling rule

$$V(x;1,\rho) = \sigma \cdot V(\sigma\, x;\sigma,\sigma\cdot\rho)\ . \tag{3}$$

Figure 1 shows some standardized Voigt profiles. For $\rho \to 0$, the Voigt profiles tend to the standardized Gaussian $G(x;1)$, while, for $\rho \to \infty$, the Voigt profiles $V(x;1,\rho)$ asymptotically approach the Lorentzians $L(x;\rho)$.

The parametric Voigt model of a spectral line is

$$I\sigma \cdot V\big(\sigma(x-\mu);\sigma,\sigma\cdot\rho\big) \tag{4}$$

with intensity I, central location μ, shape ρ, and width σ.

Figure 2 shows three stages of the parametric model. The red function is the standardized Voigt profile $r(x) = V(x;1;0.5)$. Its geometry depends on the shape parameter $\rho = 0.5$. The blue function, $b(x) = 2\cdot V(2x;2,1)$, is related to the red function by scaling with the width $\sigma = 2$. The green function, $g(x) = 1.5\cdot b(x) = 3\cdot V(2x;2,1)$, is the intensity multiplied version of the blue function and visualizes the effect of the parameter I. The simple horizontal shift belonging to the parameter μ is not shown.

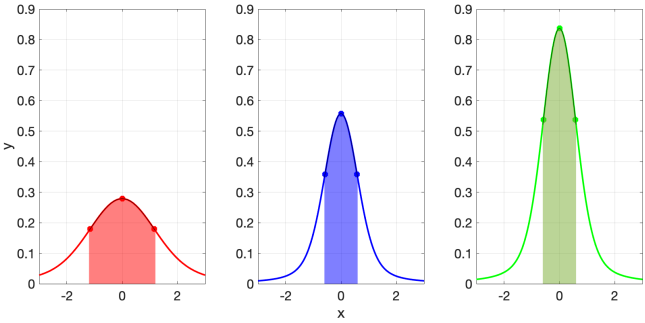


Figure 2: The red function is the standardized Voigt profile $r(x) = V(x; 1; 0.5)$, which has total area 1. Its width-scaled version, $\sigma = 2$, in blue, $b(x) = 2 \cdot V(2x; 2, 1)$, has still total area 1 because it is halved horizontally and doubled vertically. The intensity multiplied version, $I = 1.5$, in green, $g(x) = 3 \cdot V(2x; 2; 1)$ has area 1.5. The shaded area between the inflection points is always 56%.

The first derivatives of the three stages are related by $b'(x) = 4 \cdot r'(2x)$ and $g'(x) = 1.5 \cdot b'(x)$. They all have the same zero at $x = 0$. Hence, the location of the maximum does not change under width scaling and intensity multiplication. The second derivatives satisfy $b''(x) = 8 \cdot r''(2x)$ and $g''(x) = 1.5 \cdot b''(x)$. On the one hand, the horizontal distance between the inflection points and the maximum of the red curve is divided by $\sigma = 2$ for the blue function, but does not change further from the blue to the green function. On the other hand, the area between the inflection points does not change between the red and the blue functions: this area is compressed by the factor σ horizontally and stretched by σ vertically. Intensity multiplication multiplies the area , but does not change the ratio of the area between the inflection points and the total area. This ratio of the inter inflection area is invariant under width scaling and intensity multiplication; it only depends on the shape ρ. The horizontal distance between an inflection point and the maximum depends on ρ and σ.

We are going to show how the parameters I, μ, σ, and ρ can be computed from the following geometric features: maximum, inflection points, and the inter inflection area.

2 Asymptotic Case: Geometry of a Lorentzian

The centered Lorentzian has the maximum value $L(0;\gamma) = 1/(\pi\gamma)$ and the inflection points $(\pm\gamma/\sqrt{3}, 3/(4\pi\gamma))$. The antiderivative is

$$\int L(x;\gamma)\,dx = \frac{1}{\pi}\arctan\frac{x}{\gamma} + C \tag{5}$$

and the inter inflection area is constant,

$$\frac{2}{\pi}\arctan\left(\frac{1}{\sqrt{3}}\right) = \frac{1}{3}\,. \tag{6}$$

3 Geometric Features of a Standardized Voigt Profile

Figure 3 shows a standardized Voigt profile. The points are computed using the quick-and-dirty approach of [1]. But for computing

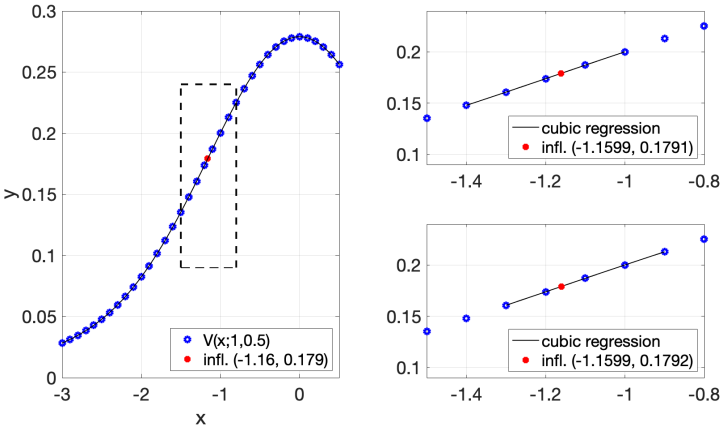


Figure 3: The standardized Voigt profile $V(x;1,0.5)$ (blue) and the locally defined cubic polynomials (black). Nicely, there are no overlaps visible between the polynomials. The two graphs on the right zoom into the dashed rectangle and show the overlapping polynomials that feature an inflection point (red) on the left or on the right of the central point of the polynomial domain respectively. The polynomial inflection points provide a consistent estimate of the Voigt inflection point.

the derivatives we use a new way: we fit a cubic polynomial to every five consecutive data points (x_i, y_i), $i = n-2, \ldots, n+2$ using least-squares. We get cubic polynomials with finite, overlapping domains

$$p_n(x) = a_n + b_n x + c_n x^2 + d_n x^3 \quad \text{for} \quad x_{n-2} \le x \le x_{n+2}\,, \tag{7}$$

which are the black curves in Figure 3. They match so well that we do not see differences in overlaps.

We use third-degree polynomials because that is the minimal degree, for which polynomials feature inflection points. Also, we use a least-squares fit instead of an interpolating polynomial since this is more appropriate for measured data: the fit reduces noise. As the fit uses only one degree of freedom, extrema and inflection points are not smoothened that much. The normal equations for the least-squares fit are stated and solved in Section 7.

We use the polynomials to estimate the Voigt profile between samples and to estimate the first and second derivatives: the polynomial p_n and its derivatives are restricted to the domain $\big((x_{n-1}+x_n)/2, (x_n + x_{n+1})/2\big)$ to avoid overlaps. For example, the inflection point of a Voigt profile is estimated by the inflection point at $x = -c_n/(3d_n)$ of the polynomial p_n with the nearest center to a sign change of the second derivative values $p_n''(x_n)$ and $p_{n\pm1}''(x_n)$. See Figure 3.

We also use the polynomials to estimate the area under the Voigt profile. By adding up the areas under p_n from $(x_{n-1}+x_n)/2$ to $(x_n + x_{n+1})/2$ and the boundary areas between inflection point and midpoint, we compute the area under the Voigt profile between the inflection points.

The computed results are documented in Table 1 together with the values of the asymptotic Lorentzian.

4 Geometric Feature versus Parameter Relationships

The maximum is at $x = \mu$ with value $I\sigma \cdot V(0; \sigma, \sigma\cdot\rho) = I \cdot V(0; 1, \rho)$, which we abbreviate to $I \cdot m(\rho)$. Figure 4 shows how the maximum value depends on the shape parameter. Although the Voigt maximum values approach the Lorentzian maximum values asymptotically for increasing shape parameter, the Lorentzian cannot be incorporated in a global model because of its vertical asymptote at $x = 0$. Instead we

Table 1: Geometric Properties of $\sigma \cdot V(\sigma \cdot x; \sigma, \sigma \cdot \rho) = V(x; 1, \rho)$ depending on ρ. $m(\rho)$ = maximum value, $i(\rho)$ = Voigt inflection points, L-IP = Lorentz inflection points, $a(\rho)$ = inter inflection area.

ρ	$m(\rho)$ in σ	$L(0;\rho)$	$i(\rho)$ in σ	L-IP in $(1/\sigma, \sigma)$	$a(\rho)$
0 (Gauss)	0.3989		$(\pm 1, 0.242)$		0.6826
0.5	0.2790	0.6366	$(\pm 1.16, 0.179)$	$(\pm 0.29, 0.6366)$	0.5620
1	0.2087	0.3183	$(\pm 1.34, 0.140)$	$(\pm 0.58, 0.2387)$	0.4902
2	0.1341	0.1593	$(\pm 1.74, 0.094)$	$(\pm 1.15, 0.12)$	0.4157
3	0.0970	0.1061	$(\pm 2.19, 0.070)$	$(\pm 1.73, 0.0796)$	0.3815
4	0.0753	0.0796	$(\pm 2.69, 0.055)$	$(\pm 2.31, 0.060)$	0.3644
5	0.0614	0.1593	$(\pm 3.20, 0.045)$	$(\pm 2.89, 0.048)$	0.3546
6	0.0517	0.0531	$(\pm 3.735, 0.0383)$	$(\pm 3.46, 0.040)$	0.3489
7	0.0446	0.0455	$(\pm 4.277, 0.0331)$	$(\pm 4.04, 0.034)$	0.3450
8	0.0392	0.0398	$(\pm 4.828, 0.0292)$	$(\pm 4.62, 0.030)$	0.3425
9	0.0349	0.0354	$(\pm 5.383, 0.0261)$	$(\pm 5.20, 0.027)$	0.3406
10	0.0315	0.0318	$(\pm 5.943, 0.0235)$	$(\pm 5.77, 0.024)$	0.3393
12	0.0263	0.0265	$(\pm 7.070, 0.0197)$	$(\pm 6.93, 0.020)$	0.3375
16	0.0198	0.0199	$(\pm 9.345, 0.0148)$	$(\pm 9.24, 0.015)$	0.3357
32	0.0099	0.0099	$(\pm 18.529, 0.0074)$	$(\pm 18.48, 0.007)$	0.3339

use the following partial fraction model with vertical asymptotes at negative horizontal positions,

$$m(\rho) \approx \frac{1.732658}{(\rho + 2.220241)^2} + \frac{0.341247}{\rho + 7.177915}\,, \tag{8}$$

which has the coefficient of determination $r^2 = 0.9999$.

Next, we study how the location of the inflection points depends on the shape parameter. We denote the horizontal distance measured in units of σ between an inflection point and the maximum by $i_x(\rho)$. The function value at an inflection point is $I\sigma \cdot V\big(\sigma \cdot i_x(\rho); \sigma, \sigma \cdot \rho\big) = I \cdot V\big(i_x(\rho); 1; \rho\big)$, which we abbreviate to $I \cdot i_y(\rho)$. The partial fraction model

$$i_y(\rho) \approx \frac{1.477978}{(\rho + 2.661468)^2} + \frac{0.249335}{\rho + 7.477085} \tag{9}$$

has the coefficient of determination $r^2 = 0.9999$. See Figure 4.

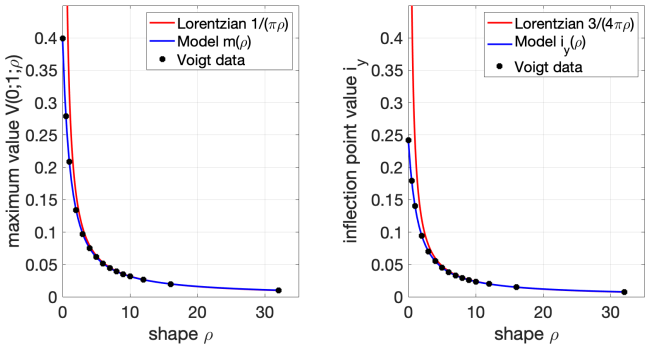


Figure 4: (Left) Maximum value of $V(x;1,\rho)$ versus shape ρ. The blue empirical model $m(\rho)$ is described in (8). (Right) The inflection point value versus the shape: $i_y(\rho)$. See Equation (9).

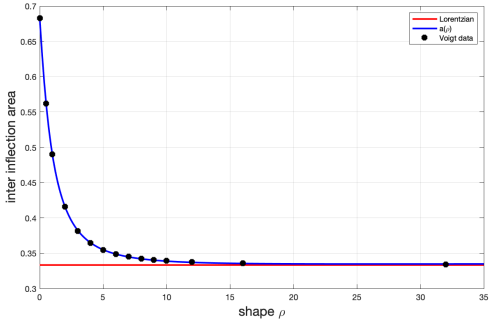


Figure 5: The inter-inflection area versus the shape parameter.

The inter-inflection area data is modeled by

$$a(\rho) \approx \frac{1.492210}{(\rho + 1.771379)^2} - \frac{0.056445}{\rho + 0.446671} + \frac{1}{3} \tag{10}$$

with the coefficient of determination $r^2 = 0.999929$. See Figure 5.

Figure 6 shows the invertible relationship between the quotient $q = m/i_y$ and the shape ρ. We have empirical relationships for the right-hand side of $q = m(\rho)/i_y(\rho)$. Until we have a sufficiently simple and

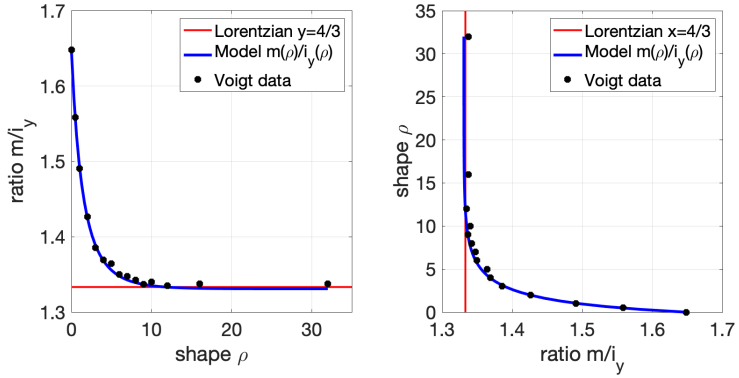


Figure 6: Left: The quotient data m/i_y and the ratio of the empirical functions $q(\rho) = m(\rho)/i_y(\rho)$. The visible deviations in the quotient data seems due to insufficient number of significant digits of m and i_y. Right: The inverse perspective, which is the direction in which we need the relationship.

accurate model of the inverse function $\rho = q^{-1}(m/i_y)$, we can apply Newton's method to $m/i_y = q(\rho)$ and find ρ.

5 Parameter Extraction Algorithm

The Voigt model of a spectral line uses four parameters: intensity $I > 0$, location μ, shape $\rho > 0$, and width $\sigma > 0$,

$$f(x) = I\sigma \cdot V\big(\sigma \cdot (x - \mu); \sigma, \sigma \cdot \rho\big) \; . \tag{11}$$

We determine the maximum, the inflection points, and the area between the inflection points. Then, we extract the parameters:

1. The maximum position is the estimate $\hat{\mu}$ of the central location.
2. The quotient of the maximum value and the inflection point value,

$$\frac{I\sigma \cdot V(0; \sigma; \sigma\rho)}{I\sigma \cdot V\big(\sigma \cdot i_x(\rho); \sigma, \sigma\rho\big)} = \frac{V(0; 1, \rho)}{V\big(i_x(\rho); 1; \rho\big)} = \frac{m(\rho)}{i_y(\rho)} \tag{12}$$

is indepedent of μ, σ, and I. It only depends on ρ. We determine the shape parameter $\hat{\rho} = q^{-1}(\hat{m}/\hat{i}_y)$ from this quotient.

3. The inter inflection area is $I \cdot a(\rho)$. We divide the measured area by the modeled area $a(\hat{\rho})$ to get the intensity $\hat{I}$.

4. Finally, the estimate $\hat{\sigma}$ is obtained by dividing the measured maximum value $I\sigma \cdot m(\rho)/\big(I \cdot m(\rho)\big) \approx \hat{m}/\big(\hat{I} \cdot m(\hat{\rho})\big)$.

A numerical example is presented on the poster at the conference.

6 Outlook

There are several ways to extend this work in future research.

For reliable results, we need to determine uncertainty bounds for our estimates. The functions are rather steep in relevant parts, and the use of empirical functions may contribute significantly to error propagation.

For an efficient calculation, we either need to find a sufficiently simple and accurate empirical function $q(\rho)$ (even better, its inverse $r(q)$) or we have to interpolate sufficiently many tabulated values.

So far, we have considered very densely sampled numerical data. Often the resolution of a spectrum is much smaller. Therefore, one should analyze how the proposed algorithm responds to lower sample densities. Is there a lower bound for the sample density similar to the Shannon-Nyquist Theorem of Signal Transmission? If so, what kind of aliasing will occur in the case of undersampling?

In our second contribution to this conference [2] we discuss how the overlap of two Voigt profiles shifts their geometric features and how we may use the above relationships to extract Voigt parameters from superimposed spectral lines.

7 Appendix: Locally-Defined Least-Squares Cubic Polynomials

We fit a least-squares cubic polynomial p_n to five consecutive data points (x_i, y_i), $i = n-2, n-1, \ldots, n+2$. Later computations use the function values, derivatives, and second derivatives of these locally defined cubic polynomials.

To simplify the calculation, we linearly transform the equidistant arguments via $k_i = (x_i - x_n)/\Delta x$. The transformed arguments are

$k = -2, -1, \ldots, +2$. We compute the best fitting cubic polynomial $\tilde{p}_n(k) = a + bk + ck^2 + dk^3$ or, equivalently,

$$p_n(x) = a + b\left(\frac{x - x_n}{\Delta x}\right) + c\left(\frac{x - x_n}{\Delta x}\right)^2 + d\left(\frac{x - x_n}{\Delta x}\right)^3 \tag{13}$$

for the original arguments.

The system of normal equations for (k, y_{n+k}), $k = -2, \ldots, +2$ is

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \\ -8 & -1 & 0 & 1 & 8 \end{pmatrix} \begin{pmatrix} 1 & -2 & 4 & -8 \\ 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \\ -8 & -1 & 0 & 1 & 8 \end{pmatrix} \begin{pmatrix} y_{n-2} \\ y_{n-1} \\ y_n \\ y_{n+1} \\ y_{n+2} \end{pmatrix}$$

with solution

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \frac{1}{10080} \begin{pmatrix} -864 & 3456 & 4896 & 3456 & -864 \\ 840 & -6720 & 0 & 6720 & -840 \\ 1440 & -720 & -1440 & -720 & 1440 \\ -840 & 1680 & 0 & -1680 & 840 \end{pmatrix} \begin{pmatrix} y_{n-2} \\ y_{n-1} \\ y_n \\ y_{n+1} \\ y_{n+2} \end{pmatrix}$$

References

1. A. Kehrein and O. Lischtschenko, "Quick-and-dirty computation of voigt profiles, classification of their shapes, and effective determination of the shape parameter," in *OCM 2023-Optical Characterization of Materials: Conference Proceedings*. KIT Scientific Publishing, 2023, pp. 169–178.
2. A. Kehrein, A. Lukyamuzi, D. Bejtullahu, and F.-M. Sheikh, "About the determination of parameters of overlapping voigt profiles from geometric features of the spectrum," in *OCM 2025-Optical Characterization of Materials: Conference Proceedings*. KIT Scientific Publishing, 2025.

About the parameter determination of overlapping Voigt profiles from geometric features of the spectrum

Achim Kehrein, Drilon Bejtullahu, Andrew Lukyamuzi, and Faraz Mehmood Sheikh

Rhine-Waal University of Applied Sciences,
Marie-Curie Str. 1, 47533 Kleve, Germany

Abstract A spectral line is often modeled as a Voigt profile depending on four parameters: intensity, central location, shape, and width. The analysis of a line spectrum has to determine these four parameters for each spectral line, so that the superimposed Voigt profiles match the measured spectrum. Nicely, the parameters of a single Voigt profile can be computed directly from its geometric features – the area under the curve, the maximum, and the inflection points – and, therefore, no iterative fitting methods are needed. The present contribution studies examples how the overlapping of two spectral lines affects the geometric features qualitatively and categorizes them into three cases. Moreover, we quantitatively determine the geometric features for the examples and present some preliminary heuristics for the Voigt parameter estimations. We believe that this contribution is a step towards more sustainable and faster algorithms for analyzing line spectra.

Keywords Voigt profile, shape parameter, overlapping spectral lines, line spectra analysis, spectroscopy

1 Introduction

A Voigt profile is a bell-shaped function, symmetric about its maximum, and has an inflection point on each side of the maximum. For examples, see the red and the blue functions in Figure 1. A Voigt profile

is defined as the convolution

$$V(x;\sigma,\gamma) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \cdot \frac{\gamma}{\pi(x^2+\gamma^2)}\, dz \tag{1}$$

of a Gaussian and a Lorentzian with width parameters $\sigma, \gamma > 0$. By the transformation rule

$$V\Big(x;1,\rho\Big) = \sigma\, V(\sigma\, x;\sigma,\sigma\rho) \tag{2}$$

each Voigt profile corresponds to a member of the standardized Voigt profile family $V(x;1,\rho)$, which provides the prototypes of Voigt profile shapes. The ratio $\rho = \gamma/\sigma > 0$ is the shape parameter.

To describe a spectral line with intensity $I > 0$ and its peak located at horizontal position μ we use the parametric model

$$I\sigma \cdot V\big(\sigma(x-\mu);\sigma,\sigma\cdot\rho\big)\ . \tag{3}$$

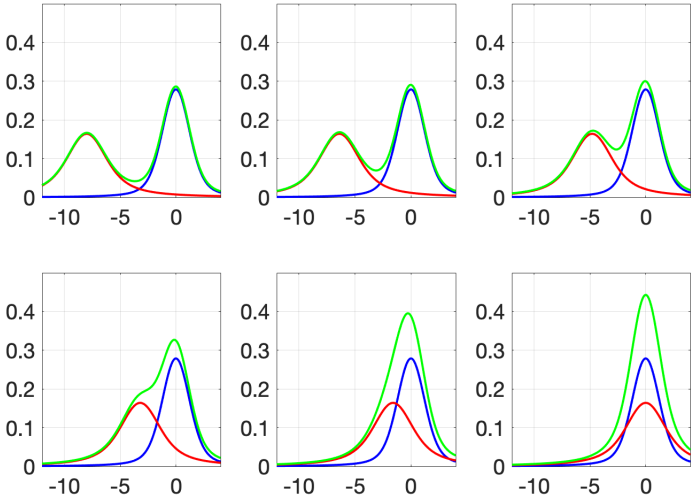


Figure 1: The superimposed spectra (green) of two Voigt profiles with various horizontal distances between them. The blue Voigt profile has shape $\rho_b = 0.5$, the red one, $\rho_r = 1.5$. Both have intensity $I = 1$ and width $\sigma = 1$.

The parameters ρ and σ are used to match the shape and the width of the spectral line. In [1] we showed how the geometric features—maximum, inflection points, and area—of this single spectral line determine the Voigt parameters. Now, we study geometric features of a spectrum with two overlapping spectral lines, which we visualize as red and blue respectively,

$$S(x) = \underbrace{I_r\sigma_r \cdot V(\sigma_r x; \sigma_r, \sigma_r\rho_r)}_{=V_r(x)} + \underbrace{I_b\sigma_b \cdot V\big(\sigma_b(x-\mu); \sigma_b, \sigma_b\rho_b\big)}_{=V_b(x)} . \qquad (4)$$

Figure 1 shows how the geometric features depend on the distance between the Voigt profiles. In particular, there are two clearly distinct cases: for large distances, the spectrum possesses a local minimum that separates the spectral lines; for small distances, we see only one peak in the spectrum. In the former case, the geometric features of the spectrum are shifted versions of geometric features of the individual spectral lines. The shifts increase with decreasing distance between the spectral lines. The latter case with one peak has two subcases. On the one hand, there can be four inflection points, three on one side and one on the other of the maximum. On the other hand, when the Voigt profiles are even closer together, there is just one inflection point on either side of the maximum. There cannot be an even number of inflection points on one side of the maximum since the horizontal asymptote requires a convex shape and the peak a concave shape. There must be an odd number of curvature changes.

The Voigt function values are computed with the numerical methods described in [2]. The derivatives and second derivatives are computed using the locally defined cubic polynomials described in [1]. These polynomials are least-squares fits to five consecutive data points.

2 Voigt Profiles Separated by a Minimum

First, we discuss a spectrum of two spectral lines that are separated by a local minimum. Figure 2 shows such spectrum together with its first and second derivatives for the individual spectral lines (red and blue) and for the superimposed spectrum (green). We use the local cubic polynomials that are the least-squares fit for five consecutive points to

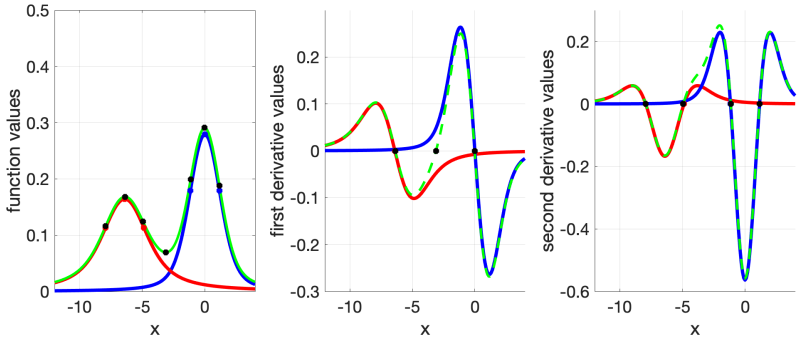


Figure 2: Function, first derivative, and second derivative of the superimposed spectrum (green) of two Voigt profiles (red, blue) separated by a minimum. Dots indicate inflection points and extrema; for better visibility the green function uses black dots.

estimate the inflection points, extrema, and inter-inflection areas. The results are documented in Table 1.

Table 1: The geometric features maximum, inflection point, and inter-inflection area of the Voigt profiles (red, blue) and their superposition (green) from Figure 2.

type	red	blue	green	perturbation
infl.	(-7.9313, 0.1132)	–	(-7.9243, 0.1162)	(+0.0070, 0.0030)
max.	(-6.4000, 0.1642)	–	(-6.3744, 0.1684)	(+0.0256, 0.0042)
infl.	(-4.8687, 0.1132)	–	(-4.9347, 0.1240)	(-0.0660, 0.0108)
min.	–	–	(-3.1153, 0.0695)	–
infl.	–	(-1.1599, 0.1792)	(-1.1404, 0.1994)	(+0.0195, 0.0202)
max.	–	(0.0000, 0.2790)	(-0.0201, 0.2908)	(-0.0201, 0.0118)
infl.	–	(+1.1599, 0.1792)	(+1.1553, 0.1883)	(-0.0046, 0.0091)
area	0.4417	–	0.4419	+0.0002
area	–	0.5515	0.5751	+0.0236

We observe the following effects of superposition. First of all, all function values of $S(x) = V_r(x) + V_b(x)$ are shifted upwards by the superposition of the positive summands.

The red inflection point on the left is perturbed to the right and up because its neighborhood is superimposed with a convex part of V_b. The sign change in the second derivative V_r'' from positive to nega-

tive is superimposed with a positive V_b'' and, hence, the zero moves to the right. The red maximum is perturbed to the right and up, since its neighborhood is superimposed with an increasing part of V_b. The sign change in the first derivative V_r' from positive to negative is superimposed with a positive V_b' and, hence, moves to the right. The red inflection point on the right is perturbed to the left and up because its neighborhood is also superimposed with a convex part of V_b. Now the change in sign of V_r'' is from negative to positive and the addition of $V_b'' > 0$ moves the inflection point to the left. See Figure 2.

Symmetrically, the inflection points of V_b are perturbed towards the maximum of V_b, and the maximum is perturbed towards the maximum of the overlapping V_r.

The perturbations at the inner inflection points are larger than at the outer inflection points. Moreover, the wider Voigt profile perturbs the narrower Voigt profile more than vice versa.

Now we apply the algorithm from [1] to the geometric features of the spectrum. The ratios of maximum value and inflection value for the red peak are

$$\frac{m_r}{i_{y,r}} = \frac{0.1642}{0.1132} = 0.1451 \,, \tag{5}$$

while the ratios for the corresponding green peak with the neighboring inflection points are

$$\frac{m_{g1}}{i_{y,g1}} = \frac{0.1684}{0.1162} = 1.4458 \quad \text{and} \quad \frac{m_{g1}}{i_{y,g2}} = \frac{0.1684}{0.1240} = 1.3358 \,. \tag{6}$$

For the blue peak and the corresponding green peak we get

$$\frac{m_r}{i_{y,r}} = 1.5569 \,, \quad \frac{m_{g2}}{i_{y,g3}} = 1.4584 \,, \text{ and } \quad \frac{m_{g2}}{i_{y,g4}} = 1.5443 \,. \tag{7}$$

As mentioned before, the inflection point farther from the other peak yields the better result. Using these ratios, we get the estimates $\hat{\rho}_r$ and $\hat{\rho}_b$, which are above the actual shape values. Then we obtain the inter-inflection area estimates $\hat{a}_r = a(\hat{\rho}_r)$ and $\hat{a}_b = a(\hat{\rho}_b)$. They produce the intensity estimates $\hat{I}_r = a_{g1}/\hat{a}_r$ and $\hat{I}_r = a_{g1}/\hat{a}_r$ and then $\hat{\sigma}_r = m_{g1}/\hat{I}_r$ and $\hat{\sigma}_b = m_{g1}/\hat{I}_b$.

For estimating the central location of the peak, we propose the use of the arithmetic mean between the spectral maximum and the mean of the neighboring inflection points. We know that the spectral maximum is biased towards the other peak. The mean of the inflection positions is biased away from the other peak. We obtain

$$\hat{\mu}_r = \frac{1}{2}\left(-6.3744 + \frac{-7.9243 - 4.9347}{2}\right) = -6.4020 \tag{8}$$

and $\hat{\mu}_b = -0.006325$.

Since the peaks are well separated, the estimates are quite accurate. They get worse as the peaks approach. An improved algorithm must use some proximity correction, which could be based on the relative height of the minimum with respect to the neighboring inflection points. The smaller relative height decreases to 1 with decreasing separation.

3 Non-Separated Voigt Profiles with Four Inflection Points

In this section, we consider two spectral lines that are not separated by a local minimum. However, the spectrum still has four points of inflection. See Figure 3. The geometric features are listed in Table 2.

Table 2: The geometric features maximum, inflection point, and inter-inflection area of the Voigt profiles (red, blue) and their superposition (green) from Figure 3.

type	red	blue	green	perturbation
infl.	(-4.5313, 0.1132)	–	(-4.4082, 0.1292)	(+0.1231, 0.0160)
max.	(-3.0000, 0.1642)	–	–	–
infl.	–	–	(-2.6796, 0.2007)	–
infl.	(-1.4867, 0.1132)	–	–	–
infl.	–	(-1.1599 , 0.1792)	(-1.1007, 0.2818)	(+0.0592, 0.1026)
max.	–	(0.0000, 0.2790)	(-0.1706, 0.3330)	(-0.1706, 0.0540)
infl.	–	(+1.1599, 0.1792)	(+1.1137, 0.2141)	(-0.0462, 0.0349)
area	0.1414	–	0.2906	+0.1492
area	–	0.5525	0.6616	+0.1091

The left red inflection point is perturbed right and up because its neighborhood is superimposed with a convex part of V_2. The sign

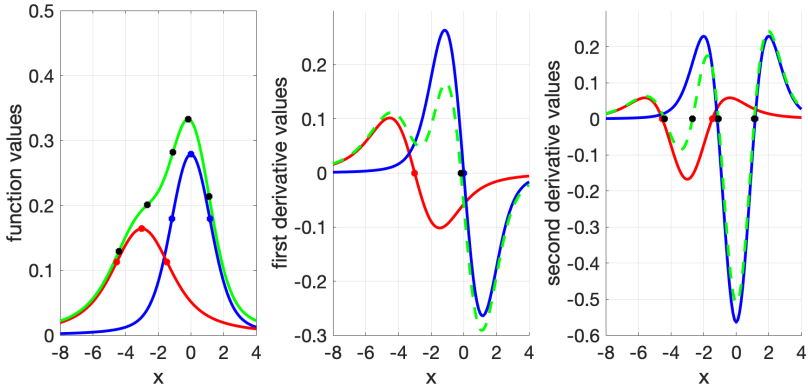


Figure 3: Function, first derivative, and second derivative of the superimposed spectrum (green) of two Voigt profiles (red, blue) not separated by a minimum. The black dots indicate the inflection points and the maximum.

change in the second derivative V_1'' from positive to negative is superimposed with a positive V_2'' and, hence, the zero moves to the right. Then, there is a discontinuous change. The red maximum has no perturbed counterpart in the total spectrum. The right red inflection point is significantly perturbed left and up because its neighborhood is also superimposed with a convex part of V_2. The sign change of V_1'' is from negative to positive and the addition of $V_2'' > 0$ moves the inflection point to the left. See Figure 3. The inflection points of the blue Voigt profile move to slightly perturbed inflection points in the green spectrum.

It is far from obvious how to extract the eight Voigt profile parameters from the geometric features. But there are five geometric features with two coordinates each. Thus, by studying, the x-coordinates of the inflection points, we obtain at least sufficiently many quantities. A shape-dependent model for the horizontal position of the inflection point of a single standardized Voigt profile is developed in Section 5.

Moreover, the example indicates, that the second point of inflection from the left may be a good estimate for the maximum of the left peak.

4 Two Voigt Profiles Become one Asymmetric Peak

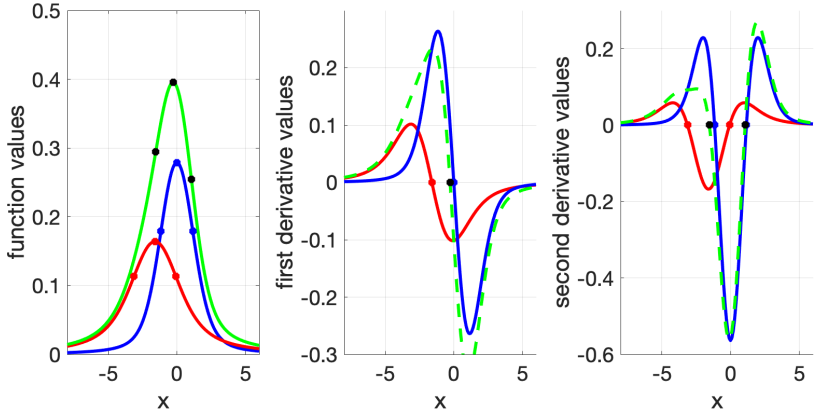


Figure 4: Function, first derivative, and second derivative of the superimposed spectrum (green) of two Voigt profiles (red, blue) not separated by a minimum and with only one curvature change on either flank.

In this case of closely overlapping Voigt profiles, the superimposed spectrum only features a maximum and two inflection points, one on either side of the maximum. Only the asymmetric location of the inflection points with respect to the maximum indicates the existence of the smaller second peak.

The asymmetry of the inflection points about the maximum reveals that the spectrum is a superposition of two peaks with a less intense peak to the left of the intenser one.

In this specific example, at least, the maximum and the inflection point to its right provides a good estimate of the shape,

$$\frac{m_g}{i_{y,g2}} = 1.5528 \tag{9}$$

which compares nicely with $\frac{m_b}{i_{y,b2}} = 1.5569$.

However, a simple count tells us that we have to extend our previous methods. We need to extract eight parameters, but maximum, inflection points, and the inter-inflection area provide only seven. This

Table 3: The geometric features maximum, inflection point, and inter-inflection area of the Voigt profiles (red, blue) and their superposition (green) from Figure 4. We dropped the column "perturbation" because the changes can hardly be considered as such with the exception of the horizontal location of the intenser peak.

type	red	blue	green
infl.	(-3.1313, 0.1132)	–	–
max.	(-1.600, 0.1642)	–	–
infl.	–	–	(-1.5416, 0.2945)
infl.	–	(-1.1599, 0.1792)	–
infl.	(-0.0687, 0.1132)	–	–
max.	–	–	(-0.2714, 0.3955)
max.	–	(0.0000, 0.2790)	–
infl.	–	–	(+1.0611, 0.2547)
infl.	–	(+1.1599, 0.1792)	–
area	0.4417	0.5542	0.8972

includes the horizontal coordinates of the inflection points. Then, we still have to find one more geometric feature in the asymmetric peak: In future research, we will consider estimates of the areas outside the inflection points.

5 Horizontal Distance between Inflection Point and Maximum

In this subsection, we return to studying a single standardized Voigt profile. We need to model the relationship between the horizontal distance between the symmetric inflection points and the maximum of a standardized Voigt profile $V(x;1,\rho)$. Figure 5 shows the computed numerical data and our empirical model. Considering the residuals of the data with respect to the asymptotic Lorentzian model, we obtain a partial fraction model as those in [1]. Thus, we obtain the model

$$i_x(\rho) \approx \frac{-1.850805}{(\rho + 0.581452)^2} + \frac{1.972509}{\rho + 0.304663} + \frac{\rho}{\sqrt{3}} \tag{10}$$

with coefficient of determination $r^2 = 0.9999982$.

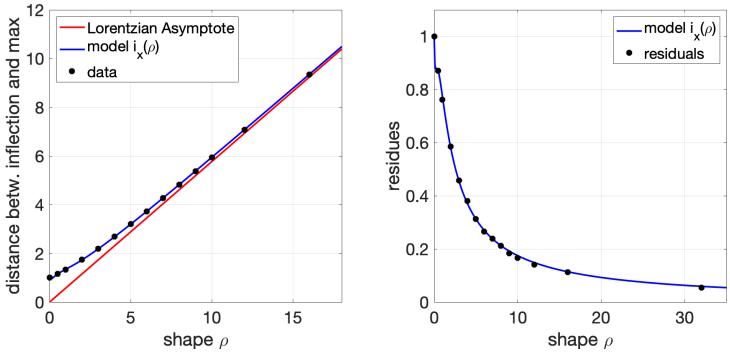


Figure 5: The horizontal distance between inflection point and maximum versus the shape parameter: $i_x(\rho)$ for standardized Voigt profiles $V(x;1,\rho)$

.

6 Outlook

There are several steps to take from here. Most importantly, the algorithms to estimate the Voigt parameters for all three cases must be made concrete. We have to study more examples. We have to vary σ and we must consider examples closer to the case boundaries as worst-case scenarios. For the separated case, we have to study if the spectral ratio $m_g/i_{y,g}$ is always less than the ratio of the corresponding Voigt profile. For the non-separated cases, we have to design adapted algorithms.

Moreover, we have to study how the accuracy depends on the density of the spectral data samples. Below some critical density, it will be impossible to model inflection points. But then, one has to question whether such sparse data should be modeled by Voigt profiles at all.

In particular, what minimal sample density is needed to resolve an asymmetric peak into its overlapping spectral lines depending on their proximity.

References

1. A. Kehrein, D. Bejtullahu, A. Lukyamuzi, and F.-M. Sheikh, "Correspondence between geometric features and parameters of a single voigt profile,"

in *OCM 2025-Optical Characterization of Materials: Conference Proceedings*. KIT Scientific Publishing, 2025.

2. A. Kehrein and O. Lischtschenko, "Quick-and-dirty computation of voigt profiles, classification of their shapes, and effective determination of the shape parameter," in *OCM 2023-Optical Characterization of Materials: Conference Proceedings*. KIT Scientific Publishing, 2023, pp. 169–178.

Transfer learning for hyperspectral image classification

Genc Ahmeti, Hannah Frank, and Andreas Zell

University of Tübingen, Computer Science Department , Cognitive Systems
Sand 1, 72076 Tübingen, Germany

Abstract Hyperspectral imaging (HSI) has revolutionized various fields such as remote sensing, agriculture, the food industry, and the recycling sector by providing detailed spectral information in addition to the spatial context. However, the high data dimensionality poses a significant challenge for hyperspectral image classification. This work explores the potential of supervised pretraining with transfer learning, implementing three fundamentally different approaches (classical, *HGen*, *DIAN*). An existing computer vision model for image classification (*ResNet-50*) as well as the specifically designed *DeepHS-Net* are expanded and incorporated into the transfer learning architectures. The pretraining is evaluated on the well-established and publicly available hyperspectral remote sensing scenes collection (HRSS), the *DeepHS Fruit* and the *DeepHS Debris* data set for the respective classification task, before and after fine-tuning as well as in a linear fashion in three different scenarios with varying degrees of class overlap. In addition, the performance on the basis of training variability and model confidence is assessed. Our work provides evidence that pretraining via transfer learning can increase the model performance and stabilize the training process in environments with heterogeneous hyperspectral data sets and tasks.

Keywords Transfer learning, pretraining, hyperspectral imaging, HSI classification, remote sensing, fruit ripeness

1 Introduction

Because objects of varying media emit and reflect electromagnetic waves of particular spectra, surface composition classification becomes possible. Current work supports the idea of enhancing hyperspectral image classification with already established deep learning methods [1,2] via supervised training to further improve model predictions. Still, the high dimensionality of hyperspectral data poses significant challenges for deep neural networks that usually require large labeled data sets for effective classification. Transfer learning, a technique where knowledge gained from one domain is applied to another domain or task, has already proven effective to mitigate these challenges for regular computer vision tasks. As HSI data comes from a variety of different domains, but typically in rather small individual data sets, finding common spectral features and transferring knowledge inbetween these data sets and tasks provides an alternative approach to stabilize the training of large classifier models and improve their overall performance. Currently, this approach promises higher accuracies than previously established pretraining and fine-tuning methods [2,3]. Additional benefits of incorporating transfer learning include a reduction of training time and costs by skipping most of the compute-heavy pretraining step while also increasing model stability during training. Additionally, smaller data sets can be used for model training without overfitting becoming very likely. However, applying already existing methods to the HSI classification environment is non-trivial and hardly investigated. With our work, we want to provide an insight into some techniques and how they can be properly adapted.

2 Experiments

2.1 Data Set

Our experiments are based on the three remote sensing data sets (i.e. Pavia Center, Pavia University and Indian Pines) from the publicly available HRSS collection [4], as well as the *DeepHS Fruit* [5, 6] and the *DeepHS Debris* data set [1]. These sets are accompanied by varying degrees of class overlap: Whereas both Pavia sets have an overlap of about 78%, all the other collections do not share any classes with each

other. However, the Pavia University set contains image patches that belong to meadows or trees such that a semantic relation can be established in plant-based structures when compared to the Indian Pines collection, even if there is no real class overlap. The semantic difference is much stronger for the other two data sets because their samples represent single objects in contrast to the landscape shots from the HRSS set. This gap is further enhanced due to the difference in spatial resolution, which is why we consider these sets as having low to no overlap.

2.2 Models

The classifier models consist of a backbone based on the typical networks that are utilized for classification in conventional computer vision tasks. For our purposes, we will use the *ResNet-50* [7] and the *DeepHS-Net* [5], which was specifically engineered to work with HSI data sets, and expand them with a final classification layer composed of fully-connected layers. On the one hand, the ResNet-50 has been proven to work exceptionally well in classical computer vision tasks. On the other hand, recent work shows that it can also be used as a backbone for HSI classification in heterogeneous environments with transfer learning (e.g. [2]) to achieve state-of-the-art accuracies, which is why we consider it as a valid option for the possible model backbones.

2.3 Transfer Learning Architectures

For the supervised pretraining of the model, we employ three transfer learning architectures: Regular pretraining with parameter transfer [8], HGen [2] and DIAN [9]. For the regular pretraining procedure, the classification network mainly consists of a model backbone that is pretrained on one task before getting fine-tuned on a different, but related task. For the HGen architecture, a novel spectral-spatial embedded feature extraction network design is incorporated into the regular approach: The spatial branch utilizes a conventional computer vision model pretrained on the ImageNet dataset as a spatial feature extractor. The spectral branch obtains spectral features by employing a separate CNN consisting of 1D convolutional layers with kernels of the size 7×1 and a subsampling stride size of 2. Both feature extractors

apply a fully-connected (FC) layer with an output size of 100 at the end. The generated feature vectors from each branch are concatenated and propagated through two FC layers and a dropout layer.

Lastly, the domain-invariant attention network (DIAN) design published by Ye et al. in 2022 introduces a different strategy for transfer learning between cross-scene hyperspectral images. Their core idea is to increase the training efficiency for data sets where the set of classes overlaps strongly by including dual-stream input processing: A single CNN is employed for each data stream to extract features from the source and target scenes at the same time. The goal is to map the heterogeneous features from both scenes into a low-dimensional subspace that they both share. To achieve cross-domain consistency, Ye et al. developed an attention network that takes each of the data streams as a single input and first propagates both through an autoencoder to receive encodings of the feature vectors of the same size. Then, for each stream, these decodings are used as weights, which are multiplied with the corresponding feature vector element-wise like in an attention block. Afterwards, a novel loss is used to direct the network into generating features that are similar to each other (in accordance with the L_2 distance) when they belong to the same class, otherwise they should be as dissimilar as possible. It is constructed by including all the possible L_2 distances for two feature vectors where they either belong to the same class or to a different one. Our experiments include a continuous extension of the class equality measure by computing similarity coefficients via the Goodness-of-Fit (GOF) measure [10] which is based on the Cauchy-Schwarz inequality.

Our networks were optimized with Adam with a weight decay of 10^{-6} and a learning rate of 0.01, which is steadily decreased via a stepwise learning rate scheduler employing a step size of 10 and $\gamma = 0.1$. We trained for 70 epochs with an effective batch size of 32. The training data was augmented via random flipping, random rotation and random cut, each of which occurring with a probability of 50%, in addition to random cropping with a probability of 10%. Finally, we apply checkpoint callback and early stopping based on the validation loss to reduce overfitting.

2.4 Evaluation

In accordance with the evaluation procedure proposed by Varga et al. [6], we measured the classification performance both before and after regular or linear fine-tuning. Without pretraining, the model either makes use of randomly initialized weights or comes preloaded with ImageNet weights. With pretraining, the model is pretrained only on the Pavia University data set with one of the three transfer learning architectures and then fine-tuned on Pavia City, Indian Pines, the Avocado subset of DeepHS Fruit and DeepHS Debris to simulate three scenarios with varying degrees of class overlap (high, middle, low/none). Following the recommendation of Varga et al. [6], only the last classification layer is trained at first, then the pretrained backbone is included for further fine-tuning. After the supervised learning, our models were evaluated on the test set with test time augmentations [11] at a probability of 50%. Using five different seeds each, we conducted experiments for all possible combinations of data sets and transfer learning architectures.

3 Results

The pretrained model was evaluated on the downstream classification task. Due to the number of combinations for the configuration, only a narrow selection can be seen in Figure 1. It includes a comparison of the classification performances depending on whether the model was pretrained or not and whether the fine-tuning was performed on the complete network or linearly with a frozen backbone. The model selection is comprised of a Baseline (ResNet-50 + CIE RGB Mapping), HGen-Net (ResNet-50 + HGen) and the DeepHS-Net.

In general, the pretraining did not always lead to a performance improvement. We make the following observations: Firstly, for the Pavia City data set, simple linear evaluation without pretraining already achieves high accuracies compared to the more computationally complex alternatives. In that case, linear evaluation always performs slightly worse than when training in a regular fashion, this difference at most being 3%.

Especially for the DeepHS-Net, not applying a pretraining routine leads to accuracies of more than 80% with the least complex training

			Pavia	Indian Pines	Debris	Fruit
Baseline	No Pretraining	Linear eval.	77.9 (± 3.7)	30.5 (± 2.0)	56.6 (± 12.2)	67.9 (± 6.3)
		Regular class.	81.2 (± 1.0)	27.4 (± 3.6)	67.6 (± 2.4)	66.1 (± 1.6)
	Pretraining	Linear eval.	59.7 ± (19.3)	52.7 (± 4.6)	51.0 (± 0.9)	63.4 (± 8.9)
		Finetuning	77.1 (± 3.6)	55.4 (± 2.8)	66.9 (± 3.7)	64.9 (± 5.8)
HGen-Net	No Pretraining	Linear eval.	80.3 (± 3.2)	21.5 (± 12.0)	49.3 (± 1.2)	40.3 (± 0.7)
		Regular class.	88.4 (± 1.2)	36.1 (± 1.5)	57.7 (± 7.5)	66.2 (± 5.4)
	Pretraining	Linear eval.	**88.6** (± 1.8)	49.2 (± 3.7)	51.8 (± 2.9)	69.8 (± 6.0)
		Finetuning	88.1 (± 3.3)	43.1 (± 4.4)	50.3 (± 5.9)	70.4 (± 1.7)
DeepHS-Net	No Pretraining	Linear eval.	82.7 (± 0.8)	50.7 (± 2.6)	55.3 (± 3.3)	67.0 (± 7.4)
		Regular class.	83.0 (± 1.0)	52.2 (± 0.7)	65.4 (± 1.5)	81.2 (± 9.2)
	Pretraining	Linear eval.	69.1 (± 11.2)	55.0 (± 5.4)	48.6 (± 7.2)	73.1 (± 4.2)
		Finetuning	83.3 (± 1.4)	**58.3** (± 2.2)	**67.9** (± 1.1)	**84.1** (± 3.6)

Figure 1: Classification accuracies [%] (mean and std) of the baseline, HGen and DeepHS model for each of the four data sets. Highest accuracies in **bold**

routine. However, it can be noted that pretraining is a viable option for models when the source domain is not related to the target domain: For Indian Pines, except for the DeepHS-Net model, pretraining almost doubles the accuracy, regardless of the fine-tuning method. For the Debris and Fruit set, fine-tuning or regular training is preferable over linear evaluation, possibly due to the target domain differing too much from the rest. Furthermore, the baseline performs better on the Debris data set without pretraining but with a regular training routine than the HGen-Net with pretrained weights and fine-tuning, indicating that the domain of the acquired knowledge differs too much from

the target, such that Transfer Learning negatively impacts the performance. Although we can note for the Fruit set that, especially for the HGen-Net, pretraining can give a substantial performance boost for linear fine-tuning. Additionally, for regular fine-tuning, a rise in accuracy of about 4% can be measured for both HGen-Net and DeepHS-Net when pretraining is included.

Lastly, in some of the cases where the performance was worse when applying pretraining, we can note that the standard deviation of the accuracy was reduced, indicating that pretraining can increase overall training stability.

4 Ablation Study

4.1 Model Confidence

Model confidences can be further categorized into situations where the model was either correct or wrong in its decision. Figure 2 displays them for all four data sets where the model prediction is wrong depending on whether the model was regularly trained or pretrained on the Pavia University data set. We can observe that pretraining leads to a decrease in confidence by more than 10% for Pavia Center, Debris and Fruit set, meaning that the model appropriately assigns choices that are wrong a lower likelihood. However, for Indian Pines the model becomes more confident in its wrong decisions. We believe that this may be the case because of the partial domain relatedness between urban and agricultural landscapes.

4.2 DIAN Loss

The original implementation of the DIAN loss [9] only considers overlapping classes that are identified to be either completely equal or entirely different. The coefficients for the L_2 distances of the source and target embeddings thus are either 1 or 0 depending on the class equality. For our approach, we want to consider a continuous extension for coefficients in the cases where the class overlap is low via the Goodness-of-Fit (GOF) similarity measure [10]. Thus, the DIAN architecture becomes viable for the Indian Pines, Debris and Fruit data set. Moreover, for the Pavia Center data set, these coefficients can either

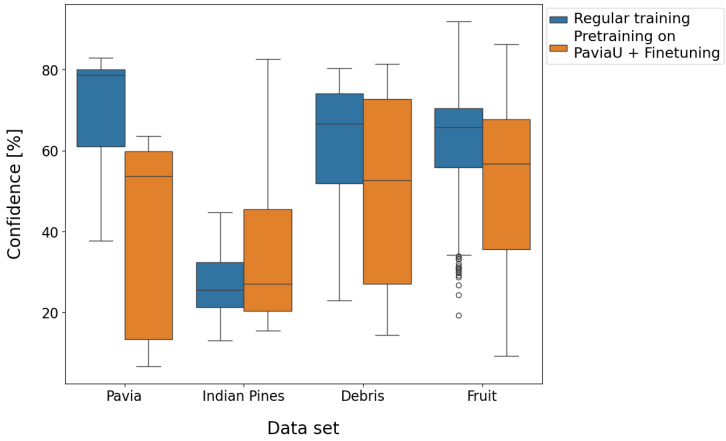


Figure 2: Model confidences depending on the supervised training routine for all four data sets where the model prediction is wrong. Each model employs a ResNet-50 backbone. For all TL architectures.

completely replace the ones assigned via class equality or simply be added for the cases where the original coefficients are 0. In Figure 3, we present the classification accuracies that can be achieved by implementing these coefficients. We can directly see that for the Indian Pines and Debris data set, the median accuracy is increased by about 5% in comparison to just using the cross entropy loss for both input streams. For the Pavia Center data set, leaving out the coefficients that were determined via class equality is tremendously detrimental to the model performance. Moreover, disregarding the original DIAN loss implementation seems to be more beneficial, likely due to the high inter- and intra-class covariances. However, adding in the similarity coefficients yields better accuracies which are increased by 4.3%, demonstrating that incorporating the embeddings of different classes enhances the training procedure.

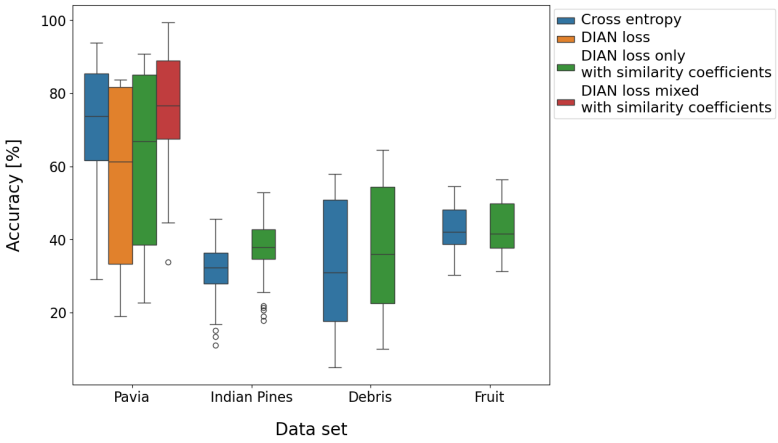


Figure 3: Classification accuracies depending on the choice of the loss function for the DIAN architecture on all four data sets and regular fine-tuning. HGen-Net is employed and pretrained on Pavia University.

5 Conclusion

In this work, supervised training with transfer learning was employed as pretraining for hyperspectral image classification of surface compositions. We show that it is possible to transfer the ideas of transfer learning from conventional computer vision environments to hyperspectral data. Further, we provide evidence that transfer learning can be applied in environments with heterogeneous hyperspectral data sets and tasks while also increasing the model performance via pretraining. Moreover, our work shows that the training process can be stabilized while allowing for the usage of larger classifier models without overfitting. For some special cases, we can even show that the overall classification accuracy can be improved. In addition, our experiments suggest that the likelihood of very confidently assigning wrong class labels is lessened when using specific transfer learning architectures. Lastly, our novel approach for the DIAN loss is promising in terms of enhancing dual-stream networks by including hyperspectral image similarities into the loss weighting, increasing the performance of DIAN-like models by about 5%.

References

1. H. Frank, K. Vetter, L. A. Varga, L. Wolff, and A. Zell, "Hyperspectral imaging for characterization of construction waste material in recycling applications," in *27th International Conference on Pattern Recognition, ICPR 2024 (accepted)*, 2024.
2. Y. Wang, M. Liu, Y. Yang, Z. Li, Q. Du, Y. Chen, F. Li, and H. Yang, "Heterogeneous few-shot learning for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.
3. B. Yang, S. Hu, Q. Guo, and D. Hong, "Multisource domain transfer learning based on spectral projections for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3730–3739, 2022.
4. M. Graña, M. A. Veganzons, and A. B., "Hyperspectral remote sensing scenes," *https://ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes*, 2011.
5. L. A. Varga, J. Makowski, and A. Zell, "Measuring the ripeness of fruit with hyperspectral imaging and deep learning," in *International Joint Conference on Neural Networks, IJCNN 2021*. IEEE, 2021, pp. 1–8.
6. L. A. Varga, H. Frank, and A. Zell, "Self-supervised pretraining for hyperspectral classification of fruit ripeness," in *6th International Conference on Optical Characterization of Materials, OCM 2023*. KIT Scientific Publishing, 2023, pp. 97–108.
7. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
8. F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
9. M. Ye, C. Wang, Z. Meng, F. Xiong, and Y. Qian, "Domain-invariant attention network for transfer learning between cross-scene hyperspectral images," *IET Computer Vision*, vol. 17, no. 7, pp. 739–749, 2023.
10. M. Agarla, S. Bianco, L. Celona, R. Schettini, and M. Tchobanou, "An analysis of spectral similarity measures," in *Color and Imaging Conference*, vol. 29. Society for Imaging Science and Technology, 2021, pp. 300–305.
11. A. G. Howard, "Some improvements on deep convolutional neural network based image classification," *arXiv preprint arXiv:1312.5402*, 2013.

Assessment of lossless compression algorithms and their performance on near-infrared spectral images

Abtin Maghmoumi[1], Harsh Shekhar[1], Tabea Scherling[1], Nils Kroell[2], Alexander Feil[1], and Kathrin Greiff[1]

[1] Department of Anthropogenic Material Cycles,
RWTH Aachen University,
Wuellnerstr. 2, D-52062 Aachen, Germany
[2] STADLER Anlagenbau GmbH,
Max-Planck-Str. 2, D-88361 Altshausen, Germany

Abstract Storing and streaming hyperspectral images (HSI) requires significant resources. One suitable strategy of reducing this resource utilizations is the use of compression algorithms. In this work, we benchmarked *Bzip2*, *Gzip*, *LZ4* and *lzf* algorithms on a dataset of HSIs of lightweight plastic packaging and achieved a lossless compression as high as 96.7% compared to the original filesize.

Keywords Near-infrared spectroscopy, hyperspectral image, lightweight plastic packaging

1 Introduction

In recent years, the analysis of infrared (IR) spectral data has become significant in waste management. Among the various IR techniques, near-infrared (NIR) spectroscopy has gained particular importance as a reliable solution for sorting different types of plastic [1]. Near-infrared hyperspectral image (HSI) includes more information of the spectral domain as the number of sub-bands they cover goes over 100. This means each pixel shows a spectrum, which enables a detailed spectral analysis and use of sophisticated classification models for inline

material characterization [2]. However, this higher dimension introduces a challenge when the recorded HSIs are transferred or stored. For example, an image with dimension of 1000 $\times$ 312 $\times$ 220 requires about 130 megabytes of storage space. With an acquisition rate of 300 frames per second (FPS), this will add up to 3.3 terabytes of data for a full day of recording, rendering the acquisition of raw spectral data in industrial processing scenarios infeasible. To tackle the storage challenge, Deore et al. developed a two-stage lossy compression algorithm that uses spectral decorrelation and discrete wavelet transformation (DWT) [3]. Beside lossy compression, lossless compression algorithms such as *Bzip2*, *7zip* and *Gzip* have shown promising results on satellite hyperspectral images [4]. One solution to improve compression performance is taking advantage of properties of near-infrared HSIs like replacing values of pixels in HSI which are not important for further analysis, i.e. spectra of background pixels in HSI of material being transported on a conveyor belt. Here, we benchmark four of the most common lossless compression algorithms *Bzip2*, *Gzip*, *LZ4* and *lzf* on a dataset of HSIs of lightweight plastic packaging post-consumer waste in two scenarios (with and without background suppression) to find an efficient compression pipeline for such HSIs.

2 Material and Methods

2.1 Dataset generation

The hyperspectral imaging camera used to record the dataset was an EVK HELIOS NIR G2–320 EVK Kerschhaggl GmbH (Raaba, Austria) with a wavelength range of 990 nm – 1678 nm with a spectral resolution of 3.1 nm/band. In total 170 HSIs were recorded, the details about polymer type and number of sample are in Table 1, each HSI has a spatial resolution of 312 and spectral resolution of 220, and was recorded with 300 frames per second. The G2-320 measures and records lines, so to create images we used the GenICam interface of the camera to read the lines from the camera's buffer and append them together to create the hyperspectral images (Fig 1).

Table 1: Dataset structure and number of samples per class

Packaging polymere	# of samples	filesize (GB)
Polyethylene terephthalate (PET)	95	8.896
Polypropylene (PP)	50	4.056
High Density Polyethylen (HDPE)	25	2.445

2.2 Compression algorithms

Here the following lossless compression algorithms have been investigated, *Gzip* [5], *Bzip2* [6], *lzf* [7] and *LZ4* [8]. As these methods are lossless, the reconstructed spectrum remains identical to the original before compression. However, as it is crucial for the compressed file to be decompressed within a reasonable time, the decompression time was also measured. The *Python* implementation of the mentioned algorithms were used on a computer equipped with an Intel® Core i9 14900 and 128 GB of RAM, for storage a 2 TB NVME SSD.

2.3 Background suppression

To improve compression performance, through a preprocessing step, the objects are detected using *scikit-image* [9], and their bounding box is calculated and all the pixels, which lie outside the bounding box are considered as background and were given the value of zero over all channels, the pipeline is illustrated in Fig 2. The time required for object detection was also taken into account as part of compression time.

2.4 Performance assessment

To assess their performance, each near-infrared HSI was compressed once with and once without background suppression using all the four algorithms. To compare the performance of benchmarked compression algorithms, the compression ratio (CR) (1) and compression speed for each HSI were calculated, and then the mean and standard deviation (std) over the whole dataset were derived.

$$\text{Compression Ratio} = \frac{\text{Original File Size} - \text{Compressed File Size}}{\text{Original File Size}} \quad (1)$$

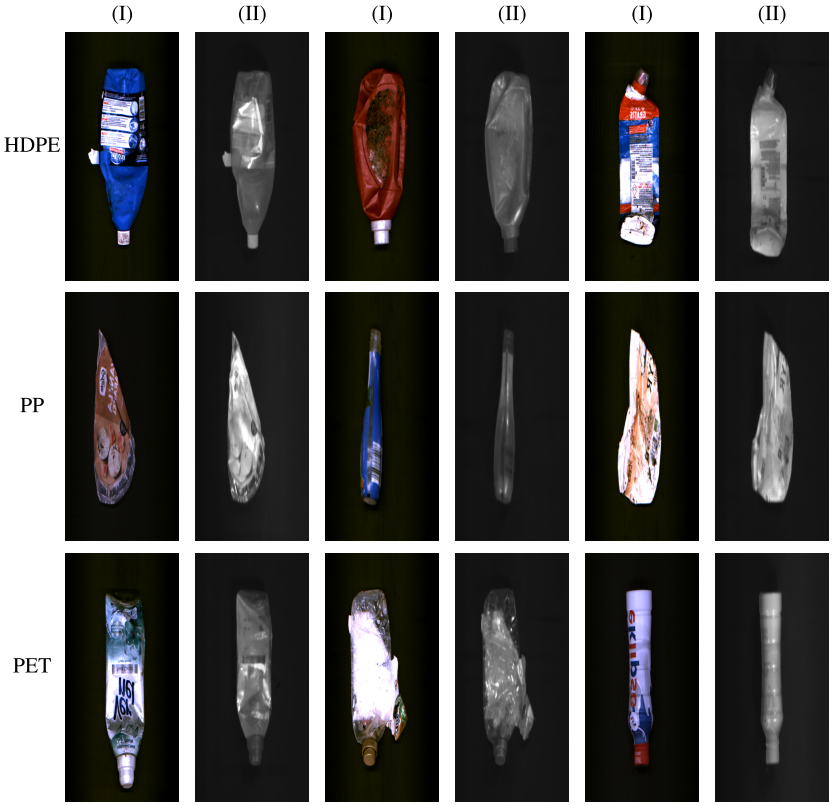


Figure 1: Sample images from dataset, (I) columns illustrate the RGB image and (II) columns visualize the HSI as a grayscale image (for each pixel an average was calculated)

3 Results and Discussion

3.1 Compression performance without background suppression

The achieved CRs for no background suppression scenario varies between 58.9 % and 90.1 % with *Bzip2* performing the best, while *LZ4* only achieved an average CR of 58.9 %. As seen in Table 2, *Bzip2* achieves the highest CR with *Gzip*, *lfz* and *LZ4* following. For both

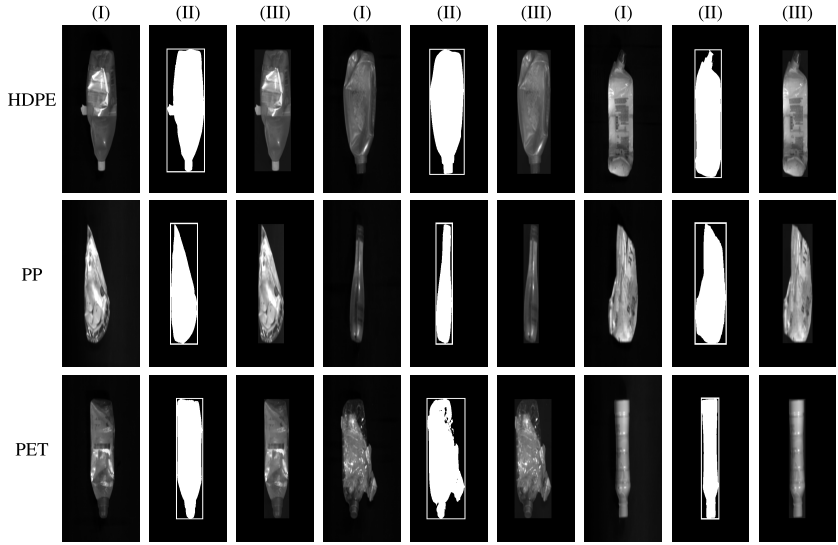


Figure 2: The pipeline for suppressing background pixels. (I) columns visualize the HSI as a grayscale image (for each pixel an average was calculated), (II) columns are binary images of the grayscale image with detected objects and their bounding boxes and (III) columns depict the grayscale images of the HSIs with suppressed background.

Gzip and *Bzip2* a compression level parameter can be defined: for *Gzip* it defines the search window and the intensity of Huffman tree optimization and for *Bzip2* it changes the block size used for compression. Varying the compression level parameter of both *Gzip* and *Bzip2* algorithms indicates that for *Bzip2* on average only 2.1 % higher CR are achieved, for *Gzip* on the other hand a 6.2% increase in CR is observed. Beside CR, the compression speed should also be evaluated. Though *Bzip2* achieves higher CR, it is much slower in comparison to other algorithms, which raises the concern for its real-time implementation.

3.2 Compression performance with background suppression

Suppresing the background in HSIs improves CR, on average the CR attained by algorithms were improved by 11.1 %, the maximum im-

Table 2: Compression results without background suppression (highest mean values are bolded)

		CR		Speed [MB/s]	
		mean	**std**	**mean**	**std**
Algorithm	**Compression level**				
Bzip2	**1**	88.0%	2.0%	15.3	1.8
	2	88.8%	1.8%	13.3	1.9
	3	89.2%	1.7%	12.7	2.1
	4	89.5%	1.6%	12.2	2.1
	5	89.7%	1.6%	11.5	2.1
	6	89.8%	1.5%	11.0	2.0
	7	90.0%	1.5%	10.5	2.0
	8	90.0%	1.4%	10.2	2.0
	9	**90.1%**	1.4%	9.7	1.9
Gzip	**1**	76.4%	2.6%	85.3	11.3
	2	77.7%	2.5%	75.0	9.6
	3	79.2%	2.6%	55.0	6.3
	4	80.9%	2.6%	56.2	7.1
	5	81.7%	2.6%	41.5	5.0
	6	82.4%	2.7%	27.6	2.7
	7	82.6%	2.7%	19.3	1.9
	8	82.6%	2.7%	10.0	1.4
	9	82.6%	2.7%	8.9	1.3
lzf	**-**	64.5%	3.3%	135.6	29.2
LZ4	**-**	58.9%	2.8%	**157.0**	38.9

provement in CR was for *LZ4* with a 30 % increase of CR. As for compression speed, both *Gzip* and *Bzip2* achieve more throughput, but *lfz* and *LZ4* slow down. Table 3 illustrates the results for compression with background suppression. The first observation is an evident increase in CR for all algorithms, with *Bzip2* achieving CR as high as 96.70 %, an explanation would be the decreased variance in HSI with background suppression and existence of repetitive zero values in the HSI, these enable the compression algorithms to achieve higher compression ratios. When it comes to compression speeds, for gzip and *Bzip2* a subtle boost in speed can be seen, but for *lzf* and *LZ4* a decrease is observed. Also, changing the compression level parameter for Bzip2 and gzip leads to minimal improvement of CR, so it is not necessary to sacrifice compression speed to achieve lower sized HSIs. Interestingly, *lzf* and *LZ4* achieve a CR close to 90 % while maintaining a fast compression speed.

Table 3: Compression results with background suppression (highest mean values are bolded)

		CR		Speed [MB/s]	
		mean	**std**	**mean**	**std**
Algorithm	**Compression level**				
Bzip2	**1**	95.6%	2.4%	32.6	11.0
	2	96.0%	2.2%	29.7	10.6
	3	96.2%	2.0%	27.7	10.3
	4	96.4%	1.9%	26.4	10.1
	5	96.5%	1.9%	25.4	9.8
	6	96.6%	1.8%	24.5	9.5
	7	96.6%	1.8%	22.5	9.3
	8	96.7%	1.8%	22.0	9.1
	9	**96.7%**	1.7%	21.0	9.2
Gzip	**1**	91.7%	3.9%	100.3	29.1
	2	92.0%	3.8%	100.4	26.9
	3	92.2%	3.7%	99.7	29.2
	4	92.7%	3.6%	84.6	18.4
	5	92.8%	3.6%	78.8	19.6
	6	92.8%	3.6%	72.3	19.4
	7	92.8%	3.6%	72.9	20.6
	8	92.8%	3.6%	66.6	21.2
	9	92.9%	3.6%	60.7	20.7
lzf	**-**	88.2%	5.5%	96.5	35.7
LZ4	**-**	89.1%	5.3%	**103.3**	32.0

3.3 Decompression performance

When compressing a file, it is also important to evaluate the decompression performance, specially how fast the original HSI can be reconstructed. In Table 4 the mean and standard deviation of decompression time and decompression speed are shown for the dataset in case of no background suppression. *LZ4*, *lzf* and *gzip* achieve much higher decompression speeds compared to *Bzip2*, it is also worth noting that different compression levels does not affect the decompression speed of gzip algorithm. The results for decompression of HSIs with suppressed background are presented in Table 5, showing a noticeable boost in the decompression speed. *Bzip2* gained the most boost, with an average 285.7 % increase. For *Gzip*, *lzf* and *LZ4* an average boost of about 30 % can be concluded.

Table 4: Decompression results for compressed HSIs without background suppression.

		Speed [MB/s]	
		mean	**std**
Algorithm	**Compression level**		
Bzip2	**1**	87.90	16.97
	2	98.81	6.44
	3	97.67	6.15
	4	92.82	10.56
	5	69.31	12.17
	6	76.56	17.86
	7	88.78	5.68
	8	67.42	13.01
	9	60.08	9.09
Gzip	**1**	279.63	13.23
	2	269.41	16.94
	3	281.21	19.54
	4	284.03	18.47
	5	253.21	26.06
	6	303.50	25.14
	7	297.71	24.76
	8	302.87	21.76
	9	290.21	22.52
lzf	**-**	**320.00**	23.33
LZ4	**-**	292.23	23.33

4 Conclusion

The dense 3D array structure of near-infrared hyperspectral images makes their storage particularly challenging. To tackle this, we have benchmarked four of the most common lossless compression algorithms on a dataset of HSIs of lightweight plastic packaging and shown that if the background pixels are suppressed before compression the achieved compression ratios of all four algorithms increases. The best algorithm was *Bzip2* with a CR as high a 96.7 % while maintaining an acceptable compression speed, *Gzip* also achieved CRs higher than 90 % with much faster compression speed, showing its feasibility for a real-time implementation. However, a real-time implementation depends highly on the equipment, in our case for recording with 300 FPS we would need a compression speed of 39.28 MB/s, which is still lower than the slowest settings of *Gzip*. In conclusion, two pipelines can be introduced depending on use case, for non-real-time applica-

Table 5: Decompression results for compressed HSIs with background suppression.

		Speed [MB/s]	
		mean	**std**
Algorithm	**Compression level**		
Bzip2	**1**	257.41	60.06
	2	254.68	56.12
	3	247.25	53.40
	4	247.33	55.24
	5	239.33	52.74
	6	237.85	56.33
	7	229.05	53.70
	8	222.04	52.44
	9	221.86	54.25
Gzip	**1**	440.73	51.74
	2	429.08	53.49
	3	423.07	48.67
	4	407.50	45.86
	5	366.34	67.48
	6	338.38	45.36
	7	411.94	47.61
	8	403.80	50.92
	9	404.44	42.64
lzf	**-**	**492.63**	59.37
LZ4	**-**	395.68	21.77

tions a pipeline of first suppressing background and then using *Bzip2* algorithm for compression is suggested but if real-time performance is deemed then *Gzip* should be utilized for compression at cost of higher storage requirement. For future work our results could be used for comparison with lossy algorithms and the trade-off between compression performance and classification results.

Acknowledgements

This work was funded the German Federal Ministry of Education and Research (BMBF) within the program ”Resource-efficient circular economy - plastic recycling technologies (KuRT)” under the project ReVise-UP (grant no. 033R390A). We thank Peter Bardenheuer for his assistance in recording the dataset used in this work. His contributions were essential in ensuring the data was collected effectively and accurately.

References

1. E. R. K. Neo, Z. Yeo, J. S. C. Low, V. Goodship, and K. Debattista, "A review on chemometric techniques with infrared, raman and laser-induced breakdown spectroscopy for sorting plastic waste in the recycling industry," *Resources, Conservation and Recycling*, vol. 180, p. 106217, 2022.
2. G. ElMasry and D.-W. Sun, "Chapter 1 - principles of hyperspectral imaging technology," in *Hyperspectral imaging for food quality analysis and control*, D.-W. Sun, Ed. London and Burlington MA: Academic Press an imprint of Elsevier, 2010, pp. 3–43. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123747532100012
3. G. dDeore, S. Rajaraman, R. Awate, and S. Bakare, "Spectral-spatial hyperspectral image compression based on measures of central tendency," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2015, pp. 2226–2232.
4. M. Grossberg, I. Gladkova, S. Gottipati, M. Rabinowitz, P. Alabi, T. George, and A. Pacheco, "A comparative study of lossless compression algorithms on multi-spectral imager data," in *DCC 2009*, J. A. Storer and M. W. Marcellin, Eds. Los Alamitos Calif.: IEEE Computer Society, 2009, p. 447.
5. J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 337–343, 1977.
6. J. Seward, "Bzip2 algorithm." [Online]. Available: http://www.bzip.org.
7. M. Lehmann, "Liblzf," 2008. [Online]. Available: http://oldhome.schmorp.de/marc/liblzf.html
8. "Lz4 - extremely fast compression," 2024. [Online]. Available: https://lz4.org/
9. S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "scikit-image: image processing in python," *PeerJ*, vol. 2, p. e453, 2014.

Measurement of air-water gas exchange by fluorescence imaging

Dennis Hofmann[1] and Bernd Jähne[1,2]

[1] Institute of Environmental Physics (IUP), Heidelberg University,
Im Neuenheimer Feld 229, 69120 Heidelberg, Germany
[2] Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University,
Berliner Straße 43, 69120 Heidelberg, Germany

Abstract A simplified fluorescence imaging technique using a fluorescent dye sprayed onto the water surface was employed to study gas transfer at the air-water interface in the Heidelberg Aeolotron. Experiments under varying wind conditions revealed the same turbulent surface structures as with erlier complex chemical techniques. The structures were analyzed using the structure tensor to quantify anisotropic regions. This approach enabled the estimation of gas transfer dynamics and offers a simple, non-invasive method for studying air-water gas exchange in any wind-wave facility and possibly also in the field.

Keywords Fluorescence imaging, orientation analysis, structure tensor

1 Introduction

The optical characterization of materials has become an indispensable tool in both fundamental research and applied sciences, offering a wide array of techniques to investigate material properties through phenomena such as refractive index, reflectance, absorption, fluorescence, and scattering. These methods have proven to be remarkably versatile, enabling insights into chemical composition, surface structures, and even three-dimensional shapes [1, 2]. However, the exploration of transport processes, particularly those at the interface between aqueous and

gaseous media, has received comparatively little attention within the field of optical material characterization.

Transport processes, such as the exchange of gases across the air-water interface, are of significant importance in diverse contexts, from environmental processes like ocean-atmosphere gas exchange and river reaeration to industrial applications in chemical engineering [3]. These processes are controlled by complex interactions within a thin sublayer, the mass boundary layer, less than one millimeter thick. The challenge lies in understanding and measuring these processes with sufficient resolution to capture their intricate dynamics.

Fluorescence imaging techniques have emerged as powerful tools for visualizing concentration fields and identifying transport mechanisms at such interfaces. In [4], Hofmann and Jähne introduced a fluorescence-based approach to study mass transfer processes in the Heidelberg Aeolotron. This method allowed for high-resolution measurements of gas transfer velocities and the identification of the underlying mechanisms by utilizing controlled experimental setups, including deionized water, a gastight airspace, and complex chemical systems. However, the complexity of this setup limited its broader applicability.

To address this limitation, a significant advancement was made by introducing a simplified fluorescence method that uses a fluorescent dye sprayed on the water surface. This new approach retains the ability to visualize and measure gas transfer processes while drastically reducing experimental complexity. By carefully controlling droplet size and velocity to ensure proper interaction with the water surface, the spray method enables similar measurements without requiring a pristine laboratory environment. This innovation allows the technique to be applied in diverse settings, including wind-wave facilities and even natural systems, provided a sufficiently powerful light source is available.

Beyond its application to gas transfer velocity measurements, the spray method opens new possibilities for analyzing surface structures and dynamics. Surface patterns, influenced by turbulence, surfactants, and wind-driven processes, are crucial for understanding transport mechanisms (see Figure 1). These structures can now be systematically characterized using advanced image processing techniques, such as the structure tensor. This mathematical tool is well-suited for analyz-

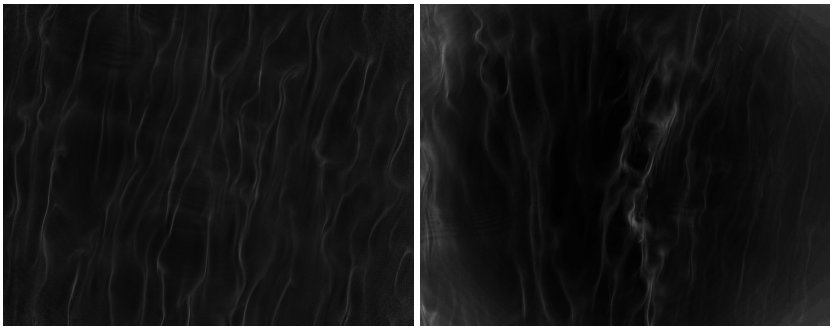


Figure 1: Comparison between surface structures captured using the classical fluorescence imaging method (left) and the fluorescence spray method (right) at wind speeds of 6.3 Hz.

ing the spatial and temporal orientation of patterns, providing insights into the coherence and directional features of surface dynamics.

In this contribution, first the basic parameters describing the transport across the air-water interface are outlined (Sect. 2). Then the measurements are described in Sect. 3. In Sect. 4 the turbulent surface structures on the water interface are analyzed using the structure tensor and it is shown how the speed of the gas transfer can be estimated from this analysis.

2 Air-Water Gas Transfer

Air-water gas exchange is governed by the interplay of molecular diffusion and turbulence within the water-side mass boundary layer, where turbulent transport diminishes near the interface. This layer is critical in controlling the transfer of gases across the air-water interface, as turbulence in the bulk water cannot penetrate the surface due to the significant density difference between air and water. Consequently, molecular diffusion dominates the transport within the thin boundary layer, where concentration gradients are highest. The transfer resistance in this layer, particularly on the water side for sparingly soluble gases like CO_2 and O_2, determines the overall efficiency of the exchange process.

The rate of gas exchange is characterized by the gas transfer velocity, which is defined as:

$$k = \frac{j}{\Delta c} \tag{1}$$

where j is the flux of gas across the interface, and is the concentration difference between the bulk water and the interface. This relationship directly ties k to measurable physical quantities and reflects the efficiency of mass transfer for a given driving force. Mathematically, k is inversely related to the total resistance to transport, with the water-side resistance typically dominating for low-solubility gases.

At the microscopic scale, the water-side mass boundary layer has a characteristic thickness, which depends on the diffusion coefficient D and the transfer velocity k via the relationship without the need to assume a specific transport model [3]

$$z_* = \frac{D}{k}. \tag{2}$$

This layer thickness reflects the distance over which molecular diffusion compensates for the concentration difference across the interface.

Because both the vertical speed of the transport is known and the thickness of the layer across which the transport takes place, the time constant can be expressed as

$$t_* = \frac{z_*}{k} = \frac{D}{k^2}. \tag{3}$$

This shows that k increases with the inverse square root of the renewal time t^*. The two equations express that only one the the three parameters k, z_* and t_* needs to measured. Then all three are known, because they are related to each other via the known diffusion coefficient D.

By spraying a dye onto the water surface, two of these parameters are directly accessible. The boundary layer thickness z_* can be inferred from the brightness of the fluorescence. By analyzing the temporal decay of fluorescence intensity patterns at the interface, the lifetime of turbulent surface structures can be estimated and thus t_*.

These structures represent the interface's exposure time to bulk water, effectively linking the observed fluorescence dynamics to. High-resolution imaging allows for the spatial and temporal characterization

of the surface turbulence, yielding robust estimates of the renewal time. Using the established relationship, the gas transfer velocity k can be derived, offering a non-invasive approach to quantifying gas exchange processes.

3 Measurements

These experiments were conducted under variable wind conditions, with the dye being sprayed for two different durations: once for 2 seconds and once for 5 seconds, at each wind speed. The fluorescence patterns were captured using seven high-speed cameras positioned underneath the water surface at a frame rate of 500 fps, enabling high-resolution imaging of the dynamic surface structures. This setup allowed for real-time visualization of turbulent surface features and their evolution in response to surface renewal events, providing insights into the interaction between the water surface and the bulk water below.

4 Analysis of Surface Structures

The analysis of the fluorescence image sequences from the spray experiments was performed using the structure tensor [5], a mathematical tool designed to capture the spatial organization of patterns within an image. The structure tensor allows for the identification of the primary orientation of the observed patterns and the quantification of the anisotropic areas, those exhibiting directional coherence, within the image, providing insights into the turbulent surface structures that develop on the water interface during gas exchange processes.

First, the dominant orientation of the surface structures was determined using the structure tensor (see Figure 2). The structure tensor enables the computation of the orientation of the surface patterns by analyzing the eigenvalues and eigenvectors associated with the tensor. Specifically, the orientation of the primary eigenvector, corresponding to the largest eigenvalue, indicates the direction in which the surface structures are most coherent. The analysis revealed that the dominant orientation stabilized at approximately 97°, tilted by 7° relative to the wind direction. This angle corresponds to the typical main orientation

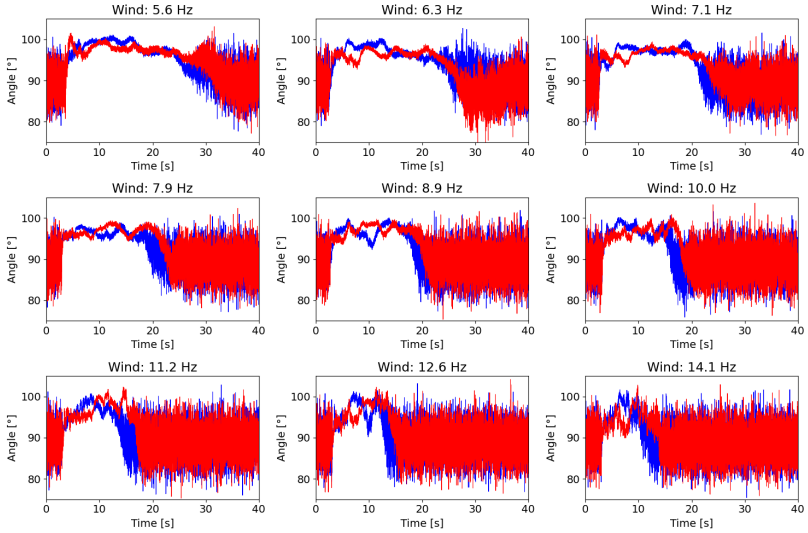


Figure 2: Temporal evolution of the dominant surface structure orientation at different wind speeds, with dye input durations of 2s (blue) and 5s (red).

of stable streaks at the water surface, driven by wind-induced turbulence. Furthermore, the time during which the main orientation remains stable was found to decrease with increasing wind speed, suggesting that higher wind speeds cause the turbulent structures to break up more quickly.

Next, the coverage of anisotropic regions on the water surface was analyzed by applying a coherence threshold to the structure tensor (see Figure 3). The threshold was set at a coherence value of 0.5, where regions with coherence values above this threshold were classified as anisotropic. These anisotropic regions exhibit clear directional structures, such as streaks, which are responsible for enhancing gas transfer. The coverage of these anisotropic regions increased with rising wind speeds, reaching a maximum of 25% at a wind frequency of 7.9 Hz. After this peak, the coverage began to decrease as wind speed continued to rise. Additionally, with higher wind speeds, the time it took for the anisotropic coverage to return to the baseline level was shorter. This

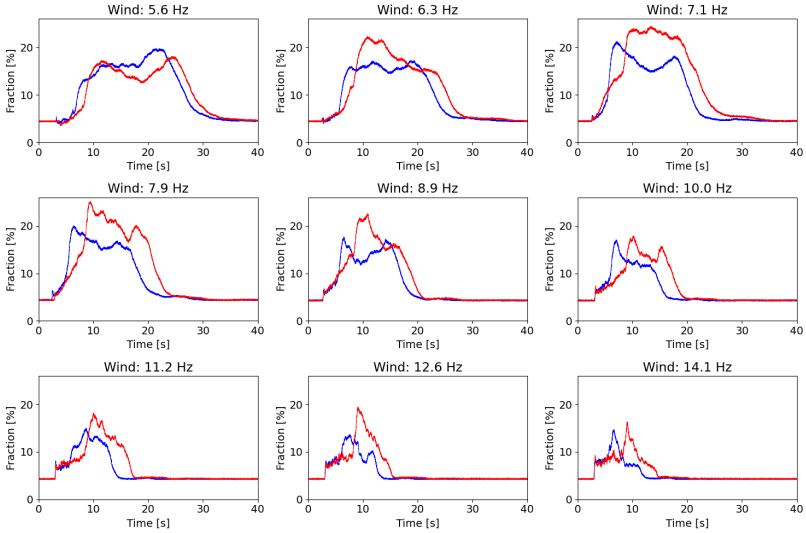


Figure 3: Time-dependent coverage of anisotropic regions across the image area, with dye input durations of 2s (blue) and 5s (red).

indicates that at higher wind speeds, the turbulent streaks dissipate more quickly, leading to a more dynamic surface renewal process that affects both the surface coverage and gas transfer dynamics.

The analysis can be further extended by utilizing the structure tensor, building on the previously discussed analysis of anisotropy. For this extension, we assume that all regions with a coherence value greater than 0.5 correspond to a surface structure. This value is used as a segmentation threshold, enabling the identification of coherent regions, which are then analyzed for both their intensity and geometric properties. The segmented structures are examined in terms of their average intensity and their size, defined by the pixel area. By quantifying these properties, histograms of the distribution of intensity and size can be generated.

The analysis of the segmented structures at different wind speeds reveals distinct patterns when examining the intensity of the structures over time (see Figure 4). As wind speed increases, the intensity of

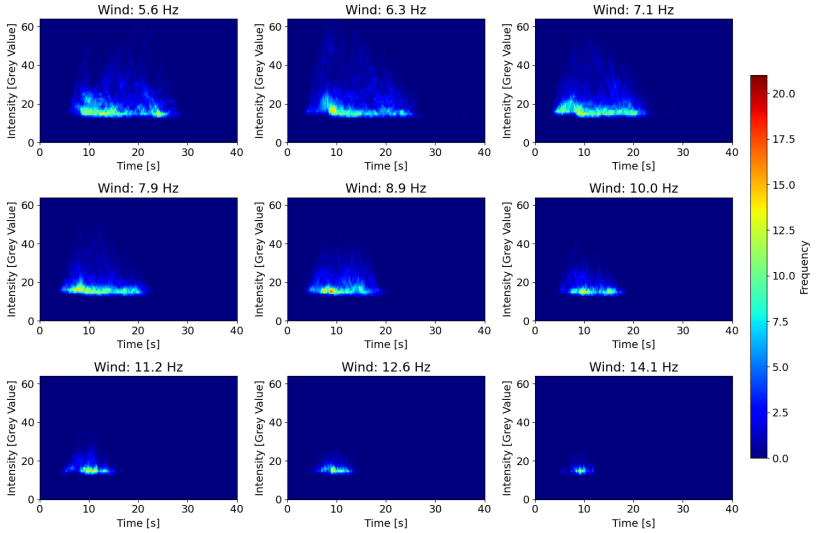


Figure 4: Temporal evolution of the average brightness of the surface streaks at different wind speeds, with dye input durations of 5s.

the segmented structures decreases. This behavior can be attributed to the more rapid transport of the dye into deeper water layers at higher wind speeds. As the turbulence intensifies with stronger winds, the dye is dispersed more quickly, preventing the formation of brighter, more defined structures at the surface. These findings suggest that the surface structures are less pronounced and more diffuse at higher wind speeds.

Additionally, the analysis of the size of the segmented structures provides further insights into the effect of wind speed on the surface renewal process (see Figure 5). With increasing wind speed, the size of the segmented structures appears to decrease. Smaller surface features form due to the faster transport of the dye into deeper regions, which limits the formation of larger structures. This observation aligns with the conclusion from the intensity analysis, where higher wind speeds result in more rapid dissipation of the turbulent streaks, leading to a reduction in the size of the coherent surface structures. These changes

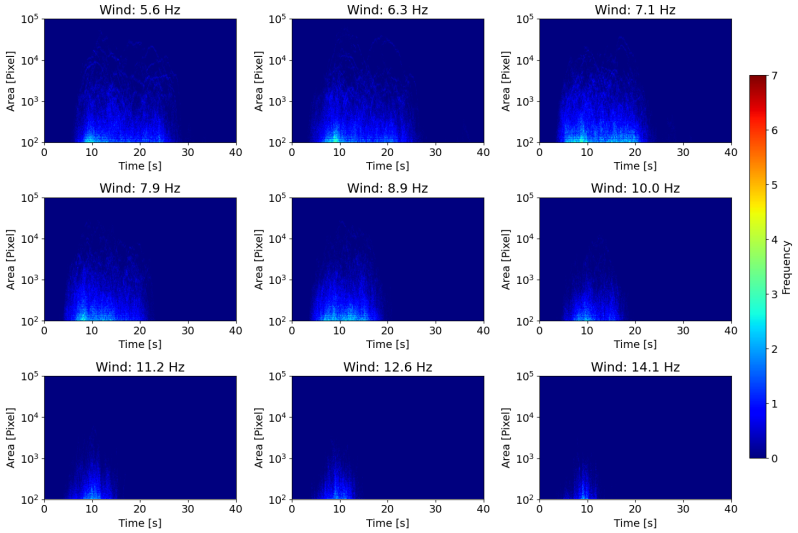


Figure 5: Time-dependent area of the segmented surface streaks, with dye input durations of 5s.

in both intensity and size highlight the dynamic nature of the surface renewal process and the influence of wind speed on the formation and dissipation of surface streaks.

In summary, the segmentation of surface structures using the structure tensor reveals that with increasing wind speed, not only does the intensity of the structures decrease, but their size also diminishes. This indicates that at higher wind speeds, the turbulence promotes a faster dissipation of surface structures, leading to smaller and less intense streaks. These observations are important for understanding how varying wind conditions can impact the formation of turbulent surface patterns and, consequently, the efficiency of gas transfer across the air-water interface.

5 Conclusions

The newly developed spray fluorescence method extends the application of fluorescence techniques beyond the controlled environment of the Aeolotron, enabling the detection of identical surface structures to those observed with classical fluorescence imaging. These include turbulent patterns, such as anisotropic streaks, which play a vital role in gas exchange processes. By combining advanced visualization with structure tensor analysis, the method quantifies key parameters like orientation, intensity, size, and temporal dynamics of these structures. At lower wind speeds, stable streaks form, maintaining consistent orientation and lasting longer, while increasing wind speeds intensify turbulence, leading to quicker dissipation and reduced stability, intensity, and size of the streaks. This versatile technique, offering valuable insights into surface dynamics and their impact on gas exchange, will be further validated through experiments at the Large Air-Sea Interaction Facility (LASIF) in Marseille in March 2025 and field measurements in the Baltic Sea in June 2025, paving the way for broader applications in natural environments.

References

1. J. Beyerer, F. P. León, and T. Längle, Eds., *OCM 2013 - Optical Characterization of Materials: Conference Proceedings*. KIT Scientific Publishing, 2013.
2. J. Beyerer, T. Längle, and M. Heizmann, Eds., *OCM 2023 - Optical Characterization of Materials: Conference Proceedings*. KIT Scientific Publishing, 2023.
3. B. Jähne, "Air-sea gas exchange," in *Encyclopedia of Ocean Sciences*, 3rd ed., J. K. Cochran, H. J. Bokuniewicz, and P. L. Yager, Eds. Academic Press, 2019, vol. 6, pp. 1–13.
4. D. Hofmann and B. Jähne, "Fluorescence imaging of concentration fields of dissolved gases at water interfaces," in *OCM 2023 - Optical Characterization of Materials: Conference Proceedings*, J. Beyerer, T. Längle, and M. Heizmann, Eds., 2023, pp. 129–138.
5. B. Jähne, *Digitale Bildverarbeitung und Bildgewinnung*, 7th ed. Berlin: Springer Vieweg, 2012.

Hyperspectral data compression and its impact on spectral signatures of water bodies

Jannick Kuester[1,2], Wolfgang Gross[1], Simon Schreiner[1], Andreas Michel[1], and Michael Heizmann[2]

[1] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Gutleuthausstrasse 1, 76275 Ettlingen, Germany
[2] Institute of Industrial Information Technology, Karlsruhe Institute of Technology, Hertzstraße 16, 76187 Karlsruhe, Germany

Abstract This paper analyzes the characteristics of water spectra and examines how varying training datasets, with different amounts of water spectra, impact the reconstruction accuracy of lossy hyperspectral data compression methods. The hypothesis is that the spectral signatures of water have lower intensity values than the spectral signatures of land and, therefore, have fewer significant features or that the features have a lower intensity. At lower intensity values, the reconstruction error caused by the lossy compression has a greater influence on the signal, as the relative error is more significant. Furthermore, the scarcity and less pronounced features in water spectra hinder the training of machine learning methods. To validate this hypothesis, lossy compression models are trained using three distinct training datasets: a combination of land and water spectra, exclusively land spectra, and exclusively water spectra. The experimental results substantiate the hypothesis that the lower intensity and weaker features of water spectra pose significant challenges for machine learning models. However, the reconstruction accuracy of the A1D-CAE demonstrates that combining land and water spectra enables a generalized restoration of critical features, thereby mitigating the limitations of models trained exclusively on land- or water-based datasets. Consequently, it is imperative to include a sufficient amount of water spectra when constructing training datasets.

Keywords Hyperspectral image (HSI) compression, spectral compression, feature extraction, autoencoder, spectral analysis

1 Introduction

The demand for hyperspectral data in remote sensing applications, such as disaster management, land cover classification, and object detection, has grown significantly. These applications benefit from the additional spectral information provided by hyperspectral sensors, which can capture material-specific features across hundreds of spectral channels. However, the large data volumes produced by these sensors pose challenges for data transmission and storage, particularly in satellite- and drone-based systems where bandwidth is limited. Lossy hyperspectral data compression offers a solution by enabling real-time transmission while balancing compression rate and reconstruction accuracy. This work focuses on investigating the reconstruction quality of lossy compressed water spectra. Previous studies, such as Kuester et al. (2023) [1], have shown that water spectra exhibit higher reconstruction errors compared to land spectra. This study hypothesizes that the lower intensity and consequently less pronounced features of water spectra make their reconstruction particularly challenging. To address this, lossy compression models are trained on three distinct datasets: land-only, water-only, and mixed land and water spectra. The evaluation is based on hyperspectral satellite data from the PRISMA mission, utilizing state-of-the-art machine learning methods including Adaptive 1D-Convolutional Autoencoder (A1D-CAE) [1], Nonlinear Principal Component Analysis (NLPCA) [2], Spectral Signals Compressor Network (SSCNet) [3], and 3D-Convolutional Autoencoder (3D-CAE) [4] for a compression rate of factor $c_R = 4$. These models are assessed to identify the optimal strategy for reconstructing critical features of water spectra while maintaining generalizability across diverse spectral categories.

2 Spectral analysis

In this chapter, the characteristics of the spectral signatures of water are examined. The analysis from Kuester et al. (2023) [1] reveals that higher reconstruction errors occur when reconstructing water spectra. It is hypothesized that the spectral signatures of water have lower intensity values compared to those of land and, therefore, exhibit less

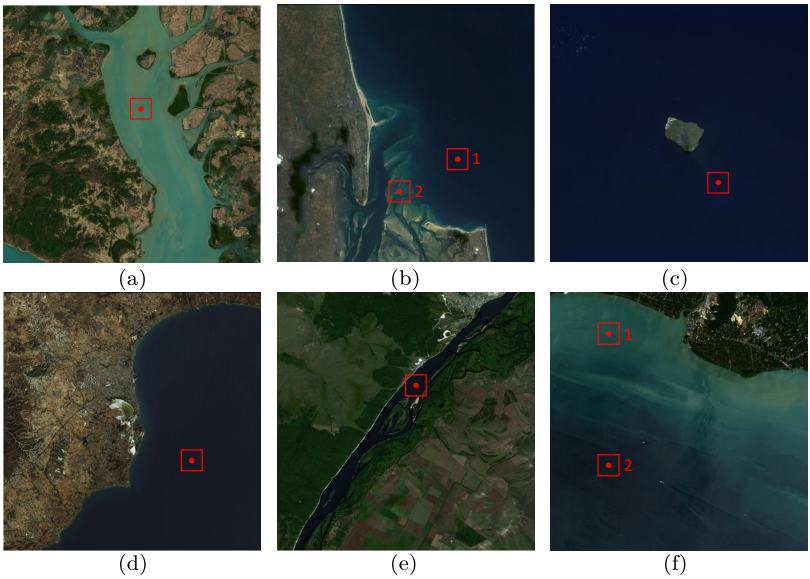


Figure 1: Datasets from the PRISMA mission in RGB composition with bands $R = 32$, $G = 21$, and $B = 10$. The test datasets were collected in Myanmar (a), Mozambique (b), Italy (c), Cyprus (d), Russia (e), and Malaysia (f), covering a wide range of spectral variability and, thus, a complex evaluation task.

significant or less pronounced features. For lower-intensity values, the reconstruction error introduced by lossy compression has a greater impact on the signal as the relative error becomes more significant. Furthermore, sparse and weakly pronounced features hinder the training of machine learning algorithms.

To analyze the properties of water's spectral signatures, six hyperspectral PRISMA datasets from different locations worldwide were selected, as shown in Figure 1. The visualization in Figure 1 illustrates that water pixels in the various datasets exhibit diverse shades of blue, making water a complex category. The red dot marks the pixels whose spectral signature are displayed in Figure 2, while the red square identifies the pixels used to calculate the metrics presented in Table 1.

The findings indicate that the spectral signatures of water in the recorded wavelength range of $420\,\text{nm} - 1010\,\text{nm}$ exhibit various char-

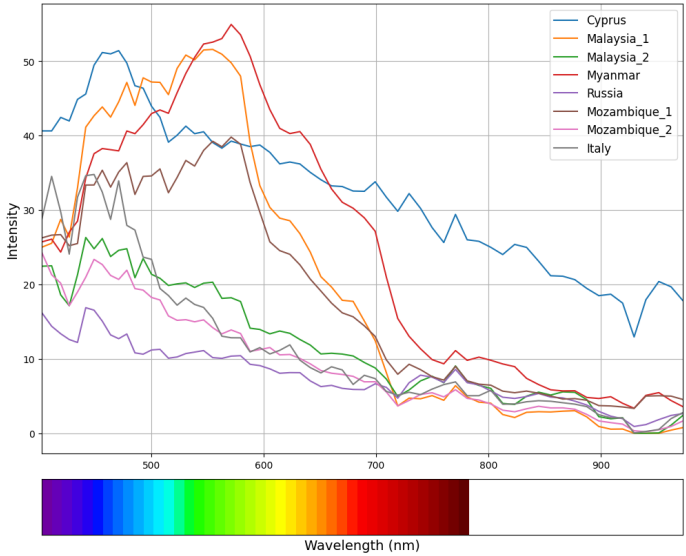


Figure 2: Comparison of water spectra from different locations, illustrating variability in intensity across the visible and near-infrared spectrum (420 nm – 1010 nm).

acteristic reflection properties, as shown in Figure 2. In the visible range (420 nm – 780 nm), reflectivity is generally low. Blue light (420 nm – 500 nm) is strongly scattered in the water, contributing to its characteristic blue appearance, while reflection in the green wavelengths (500 nm – 600 nm) slightly increases, particularly in water bodies containing suspended particles or chlorophyll. Red light (640 nm – 780 nm), however, is strongly absorbed, resulting in minimal reflection. In the near-infrared range (780 nm – 1010 nm), water absorption significantly increases, leading to very low reflectivity. Between 900 nm – 1010 nm, absorption is nearly absolute, making reflected light in this range almost undetectable.

In addition to visual analysis, metrics such as mean, standard deviation, maximum and minimum intensity values, area under the curve (AUC), and signal energy were calculated and compared for water and land; the results are shown in Table 1. Water consistently shows lower intensity values than land, with mean intensities in Myanmar and Malaysia at 27.15 and 22.89, compared to 33.05 and 66.09 for land.

Table 1: Statistical analysis for water and land spectral signatures across PRISMA datasets.

Dataset	Myanmar 2021_03_05						Mozambique 2021-08-20						Italy 2023-06-20					
Metric	mean	std_dev	max	min	auc	signal_energy	mean	std_dev	max	min	auc	signal_energy	mean	std_dev	max	min	auc	signal_energy
spectral signature water	27.15	16.85	54.57	3.34	1697.32	64369.04	18.10	12.57	36.49	1.27	1125.03	30603.58	12.55	9.91	33.33	1.92	773.43	16120.85
spectral signature land	33.05	21.00	72.94	3.40	2055.49	96626.89	36.80	11.75	65.42	16.00	2285.73	94030.16	43.82	10.34	66.88	30.21	2724.84	127742.15
Dataset	Cyprus 2020-08-18						Russia 2021-06-25						Malaysia 2021-03-02					
Metric	mean	std_dev	max	min	auc	signal_energy	mean	std_dev	max	min	auc	signal_energy	mean	std_dev	max	min	auc	signal_energy
spectral signature water	32.37	10.81	50.75	0.0	2031.02	73395.08	11.95	2.48	18.13	4.96	739.80	9394.26	22.89	19.04	52.19	0.00	1427.77	55868.78
spectral signature land	91.96	23.65	129.05	29.22	5736.12	568137.25	40.75	24.22	91.22	7.91	2530.43	141649.48	66.093	30.82	125.41	23.17	4107.27	335054.56

The amplitude range is also smaller for water, as seen in Russia, where maximum intensities are 52.19 for water and 125.41 for land. Similarly, the standard deviation, indicating feature variability, is consistently lower for water, e.g., 10.81 in Cyprus versus 23.65 for land. Signal energy and AUC, reflecting overall dynamics and magnitude, are significantly reduced for water; for instance, in Russia, signal energy is 9394.26 for water compared to 141649.48 for land, and the AUC in Cyprus is 2031.02 for water versus 5736.12 for land. These findings highlight water's weaker spectral signatures.

Overall, the results in Table 1 together with the visual analysis of the spectral signatures from Figure 2 demonstrate that the spectral signatures of water are not only less intense but also possess less pronounced features.

3 Compression Methods

This section describes the lossy compression methods applied in this study. Machine learning-based methods have become increasingly popular for hyperspectral data compression due to their superior performance compared to classical methods like JPEG and JPEG 2000 [5–7]. The models evaluated in this work include A1D-CAE, NLPCA, SSCNet, and 3D-CAE. All models were trained with a compression rate of $c_R = 4$ for 150 epochs using the AdamW optimizer with a learning rate of 0.001 and weight decay of 10^{-12}.

- **A1D-CAE:** The Adaptive 1D-Convolutional Autoencoder (A1D-CAE) processes individual spectral signatures, focusing on high reconstruction accuracy by excluding spatial features. The en-

coder compresses spectral data using 1D convolutional layers with strides and adaptive padding to accommodate varying spectral band numbers, ensuring compatibility with the compression rate (c_R). The decoder reconstructs the spectral signatures using transposed convolutional layers. The final layer scales the output to $[0, 1]$ using the Hardtanh activation function. Training optimizes performance using mean squared error (MSE) as the loss function.

- **NLPCA:** The Nonlinear Principal Component Analysis (NLPCA) model is a shallow autoencoder that compresses hyperspectral data using three linear hidden layers with sigmoid activations. This architecture ensures efficient compression while preserving and evenly distributing spectral information. The MSE loss function guides the training process, enabling the model to capture essential non-linear patterns in the data.

- **SSCNet:** The Spectral Signals Compressor Network (SSCNet) is a 2D-convolutional autoencoder designed for hyperspectral data compression and reconstruction. The encoder uses three pairs of 2D convolutional layers (Conv2d) with Parametric ReLU (PReLU) activations, followed by 2×2 max pooling for downsampling. The compressed representation is generated by flattening the feature map. The decoder mirrors the encoder with transposed convolutional layers (ConvTranspose2d) and PReLU activations for upsampling. A final sigmoid activation normalizes the reconstructed output to $[0, 1]$. Binary Cross-Entropy Loss measures reconstruction error during training. SSCNet processes data in tiles of $32 \times 32 \times 63$, integrating spectral and spatial features.

- **3D-CAE:** The 3D-Convolutional Autoencoder (3D-CAE) combines spectral and spatial compression using a 3D encoder-decoder structure. The encoder applies 3D convolutional layers with batch normalization (BatchNorm3d) and Leaky Rectified Linear Unit (LeakyReLU) activations, culminating in a 3D residual block (Res-block) for enhanced feature extraction. After quantization, the decoder reconstructs the data using 3D upsampling and convolutional layers with residual connections to maintain data fidelity. The MSE loss function aims to optimize reconstruc-

tion but depends on data and model characteristics. The 3D-CAE processes data in tiles of $32 \times 32 \times 63$, preserving spatial-temporal consistency and essential features.

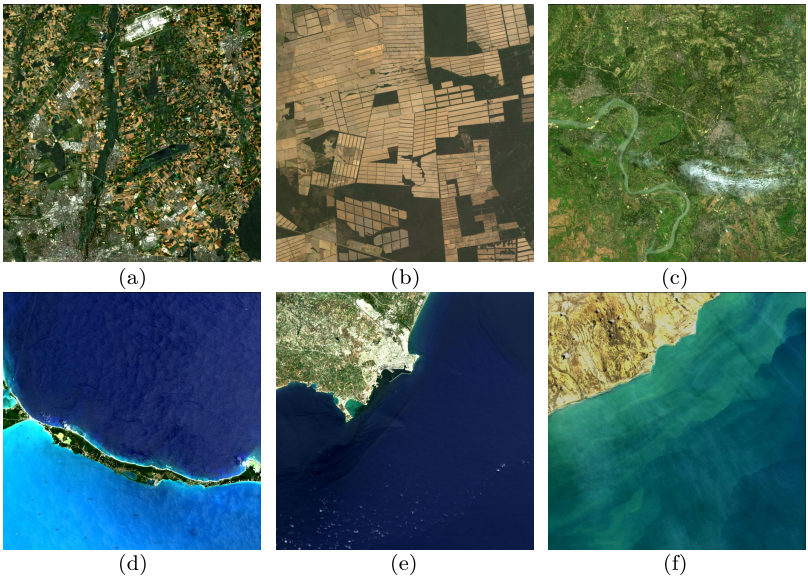


Figure 3: Selection of the training datasets from the PRISMA mission in RGB composition with bands $R = 32$, $G = 21$, and $B = 10$. The test datasets were collected in Germany (a), Argentina (b), India (c), Bahamas (d), Syria (e), and Venezuela (f), covering a wide range of spectral variability.

4 Experimental setup

To investigate the reconstruction quality for lossily compressed water spectra, we created three different training datasets. Hyperspectral data from different locations all over the world from the PRISMA mission were used to create the training datasets [8]. The sensor from the PRISMA mission covers the visible and near-infrared from $420\,\text{nm} - 1010\,\text{nm}$ with $b = 63$ bands and contains 1000×1000 pixels. The data are radiometrically corrected and georeferenced with a ground sampling distance of 30 m.

Three training datasets were created, incorporating varying proportions of water spectra from a total of 17 PRISMA datasets. Figure 3 shows a selection of the PRISMA training data. The composition of the first training dataset is predominantly vegetation, deserts, and urban signatures, with no water signatures. The second training dataset comprises 50 % water spectral signatures and 50 % land spectra. The third training dataset consists exclusively of water spectral signatures. As PRISMA datasets that exclusively contain water spectra or land signatures are rare, binary masks were created. The training datasets contain 8 million spectral signatures. For the training process, each of the three created training datasets is split into 80 % training and 20 % validation data with the test split function from scikit-learn [9]. The evaluation is performed on unknown test datasets that are spatially and temporally disjoint from the training datasets. To determine the information loss caused by the compression, we evaluate the reconstruction accuracy based on the peak-signal-to-noise ratio (PSNR) [10], the spectral angle (SA) [11], and the structural similarity index measure (SSIM) [10] metrics. The reconstruction accuracy is measured between the original and the reconstructed spectral signatures of each test dataset. Additionally, the value range of all hyperspectral training and test datasets is scaled to $[0, 1]$.

5 Results & Discussion

The results of the analysis on hyperspectral data compression using various machine learning methods demonstrate clear differences in reconstruction accuracy, as shown in Table 2, 3 and 4. The PRISMA datasets presented in Chapter 2 were employed as test datasets. Care was taken to ensure these datasets included a sufficient number of spectral signatures from both water and land.

Table 2 presents the reconstruction error for all compression methods trained on a land-based training dataset. Spectral Angle values below 3° indicate high reconstruction accuracy [2]. The A1D-CAE model achieved the best results, exhibiting high PSNR values and SSIM values close to 1. This suggests that spatial features are reconstructed with high quality. Additionally, the SA value for this model is the lowest compared to other models, further indicating superior reconstruction

Table 2: Evaluation of the reconstruction accuracy for the land-only training dataset based on the SA, PSNR, and the SSIM metric.

Training dataset	Myanmar 2021-03-05			Mozambique 2021-08-20			Italy 2023-06-20			Cyprus 2020-08-18			Russia 2021-06-25			Malaysia 2021-03-02		
land	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]
A1D-CAE	**1.45**	**45.00**	**0.99**	**3.51**	**48.68**	**0.99**	**8.76**	**34.44**	**0.94**	**1.18**	**53.72**	**0.99**	**1.33**	**51.96**	**0.99**	**2.52**	**50.22**	**0.99**
NLPCA	2.01	41.18	0.98	6.14	41.87	0.97	22.06	26.83	0.75	2.07	48.76	0.99	1.61	51.15	0.99	5.05	42.89	0.97
SSCNet	85.07	13.68	0.04	84.48	20.59	0.13	87.49	20.14	0.11	82.76	18.98	0.03	79.06	16.40	0.05	85.13	21.20	0.12
3D-CAE	11.78	26.05	0.72	67.24	19.94	0.21	67.09	18.66	0.11	27.99	27.15	0.62	17.79	26.91	0.67	75.18	19.52	0.06

Table 3: Evaluation of the reconstruction accuracy for the water-only training dataset based on the SA, PSNR, and the SSIM metric.

Training dataset	Myanmar 2021-03-05			Mozambique 2021-08-20			Italy 2023-06-20			Cyprus 2020-08-18			Russia 2021-06-25			Malaysia 2021-03-02		
water	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]
A1D-CAE	**1.47**	**44.38**	**0.98**	**2.28**	**49.62**	**0.99**	**2.80**	**46.34**	**0.98**	**1.72**	**49.25**	**0.99**	**1.84**	**47.13**	**0.99**	**1.62**	**50.87**	**0.99**
NLPCA	1.81	40.84	0.98	2.11	49.37	0.99	3.02	45.05	0.98	2.63	46.42	0.99	2.27	45.54	0.99	1.98	48.93	0.99
SSCNet	89.99	11.88	0.02	88.58	19.59	0.11	89.74	20.08	0.09	89.72	17.67	0.03	89.81	16.24	0.04	88.915	17.20	0.10
3D-CAE	9.71	26.77	0.76	10.49	36.00	0.90	7.76	36.71	0.93	8.02	34.97	0.89	14.16	28.08	0.76	10.15	34.01	0.90

Table 4: Evaluation of the reconstruction accuracy for the mixed training dataset based on the SA, PSNR, and SSIM metric.

Training dataset	Myanmar 2021-03-05			Mozambique 2021-08-20			Italy 2023-06-20			Cyprus 2020-08-18			Russia 2021-06-25			Malaysia 2021-03-02		
mixed	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]	SA [°]	PSNR [dB]	SSIM [ ]
A1D-CAE	**1.21**	**46.30**	**0.99**	**1.87**	**52.76**	**0.99**	3.92	43.08	**0.98**	**0.96**	**54.55**	**0.99**	**1.16**	**52.94**	**0.99**	**1.49**	**53.78**	**0.99**
NLPCA	1.54	44.09	0.98	2.44	51.15	0.99	**3.23**	**45.00**	0.97	1.31	51.08	0.99	1.64	49.62	0.99	2.08	51.03	0.99
SSCNet	89.98	12.63	0.02	88.88	17.59	0.08	89.77	20.00	0.06	89.68	16.98	0.03	89.75	14.40	0.06	89.04	18.20	0.09
3D-CAE	6.32	29.57	0.83	7.28	38.56	0.94	7.16	37.86	0.94	5.97	36.03	0.91	6.71	32.91	0.91	7.59	37.78	0.94

accuracy. The A1D-CAE also performed well on test datasets with a high proportion of water spectra, such as Cyprus and Malaysia, despite not being explicitly trained on such data. However, the reconstruction accuracy decreased for test datasets like Mozambique and Italy. This decline in accuracy was observed across other compression methods as well. Methods that prioritize spatial features, such as SSCNet and 3D-CAE, showed significantly reduced reconstruction accuracy.

Table 3 shows the reconstruction error for all compression methods trained on a water-based training dataset. As expected, the reconstruction error decreased for test datasets predominantly containing water spectra, such as Mozambique, Italy, and Malaysia. This effect was most pronounced for the Italy dataset, which contains the highest proportion of water spectra among all test datasets. Conversely, for test datasets with a higher proportion of land spectra, such as Cyprus and Rus-

sia, the reconstruction error increased. Nevertheless, the A1D-CAE achieved the highest reconstruction accuracy among all compression methods.

Table 4 presents the reconstruction error for all compression methods trained on a mixed training dataset. By adequately incorporating both water and land spectra, reconstruction accuracy was improved across all test datasets. This observation is particularly evident for the Mozambique, Italy, Cyprus, and Malaysia datasets. Improvements in reconstruction accuracy were observed for the A1D-CAE, NLPCA, and 3D-CAE methods. However, SSCNet exhibited a notably high reconstruction error, likely due to its exclusive focus on spatial features.

6 Conclusion & Outlook

The results, along with the analysis presented in Chapter 2, support the hypothesis that the lower intensity and less pronounced features of water spectra pose greater challenges for machine learning models. These characteristics lead to higher reconstruction errors, especially when models are trained solely on land-based datasets. Training datasets that include a mix of land and water spectra significantly improve reconstruction accuracy, as demonstrated by the A1D-CAE, which consistently outperformed other methods across diverse test datasets.

The results emphasize the importance of balanced training datasets that reflect the spectral diversity of the target application. Models trained on mixed datasets achieved generalized reconstruction capabilities, effectively mitigating the limitations observed in land-only or water-only datasets. Additionally, methods focusing on spectral fidelity, such as the A1D-CAE, proved superior to those prioritizing spatial features.

Future research should focus on expanding training datasets to include more diverse water bodies, such as turbid or chlorophyll-rich waters, to enhance model generalizability. Adaptive compression rates tailored to the spectral complexity of specific regions could further optimize reconstruction quality. Additionally, real-time implementation of these methods on satellite and drone platforms could enable efficient data processing under field conditions.

References

1. J. Kuester, W. Gross, S. Schreiner, W. Middelmann, and M. Heizmann, "Adaptive two-stage multisensor convolutional autoencoder model for lossy compression of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
2. G. Licciardi and J. Chanussot, "Spectral transformation based on nonlinear principal component analysis for dimensionality reduction of hyperspectral images," *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 375–390, 2018.
3. R. La Grassa, C. Re, G. Cremonese, and I. Gallo, "Hyperspectral data compression using fully convolutional autoencoder," *Remote Sensing*, vol. 14, no. 10, p. 2472, 2022.
4. Y. Chong, L. Chen, and S. Pan, "End-to-end joint spectral–spatial compression and reconstruction of hyperspectral images using a 3d convolutional autoencoder," *Journal of Electronic Imaging*, vol. 30, no. 4, pp. 041 403–041 403, 2021.
5. J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," *arXiv preprint arXiv:1611.01704*, 2016.
6. L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy image compression with compressive autoencoders," *arXiv preprint arXiv:1703.00395*, 2017.
7. Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Deep convolutional autoencoder-based lossy image compression," in *2018 Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 253–257.
8. R. Loizzo, R. Guarini, F. Longo, T. Scopa, R. Formaro, C. Facchinetti, and G. Varacalli, "Prisma: The italian hyperspectral mission," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 175–178.
9. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
10. S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, "scikit-image: image processing in Python," *PeerJ*, vol. 2, p. e453, 6 2014. [Online]. Available: https://doi.org/10.7717/peerj.453

11. F. A. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, "The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data," *Remote sensing of environment*, vol. 44, no. 2-3, pp. 145–163, 1993.

Material characterization by inverse rendering

Johannes Meyer, Lukas Dippon, and Christian Kludt

Fraunhofer Institute of
Optronics, System Technologies and Image Exploitation IOSB
Fraunhoferstr. 1
76131 Karlsruhe, Germany

Abstract The optical characterization of materials is crucial for understanding their fundamental properties and potential applications in various fields, including electronics, photonics, and materials science. In this contribution, we present a novel approach for material characterization through inverse rendering, focusing on the approximation of the bidirectional reflectance distribution function (BRDF). The BRDF is characterized by its dependence on various factors, including surface texture, material composition, and wavelength of light, making it a complex but vital tool for predicting how surfaces will appear under different lighting conditions. Using computational imaging and optimized illumination patterns, our method significantly enhances the efficiency of BRFD estimation compared to traditional mechanical scanning techniques. This advancement could be beneficial for practical applications in fields such as material identification and computer graphics, but also opens avenues for future research in automated visual inspection.

Keywords Inverse rendering, computational imaging, material characterization, BRDF measurement

1 Introduction

The optical characterization of materials is fundamental in both scientific research and industrial applications. It involves the study of how materials interact with electromagnetic radiation, particularly in the visible, ultraviolet, and infrared regions of the spectrum. This interac-

tion provides essential information on the structure, composition, and electronic properties of a material.

For instance, techniques such as UV-Vis spectroscopy allow researchers to determine the absorption spectra of materials, which can reveal information about electronic transitions and bandgap energies. This is particularly important for semiconductors and insulators, where understanding the bandgap can guide the design of devices such as solar cells, transistors, and photodetectors. Photoluminescence spectroscopy, however, can provide insights into the recombination processes of charge carriers and the presence of defects or impurities within a material, directly affecting its performance in optoelectronic applications.

Raman spectroscopy is another powerful tool that is used to characterize phonon modes in materials, offering insights into molecular vibrations, crystal structures, and phase transitions. This technique is crucial for materials development in fields such as nanotechnology and biomaterials, where understanding the behavior of materials at the nanoscale is essential.

Moreover, optical characterization is not limited to purely academic research; it plays a significant role in quality control and material selection in industry. By ensuring that materials meet specific optical criteria, manufacturers can optimize the performance of their products, ranging from consumer electronics to advanced optical components.

In summary, optical characterization serves as a bridge between fundamental science and practical application, enabling the advancement of technology through a deeper understanding of material properties [1].

A particularly important optical property of materials that also allows their discrimination is the so-called bidirectional reflectance distribution function (BRDF). The $BRDF(\theta_\mathrm{i}, \varphi_\mathrm{i}, \theta_\mathrm{e}, \varphi_\mathrm{e})$ is a fundamental concept in optics and surface science that characterizes how light is reflected off a surface. It provides a comprehensive description of the angular distribution of the reflected light as a function of both the incident angles $(\theta_\mathrm{i}, \varphi_\mathrm{i})$ and the reflected angles $(\theta_\mathrm{e}, \varphi_\mathrm{e})$. Mathematically, the BRDF is defined as the radiance reflected in a particular direction with respect to the incident irradiance from another direction [2, 3]. This function is crucial for understanding and modeling the optical properties of materials, particularly in fields such as remote sensing,

computer graphics, and radiative transfer. By analyzing the BRDF, one can assess how surfaces interact with light, which is essential for applications ranging from improving the accuracy of satellite-based Earth observations to enhancing the realism of virtual environments in rendering techniques, but also for discriminating materials. The BRDF is characterized by its dependence on various factors, including surface texture, material composition, and wavelength of light, making it a complex but vital tool for predicting how surfaces will appear under different lighting conditions. Notable models for BRDF include the Lambertian model, which describes diffuse reflection, and the Phong reflection model, which accounts for specular highlights. Understanding BRDF not only aids in material characterization, but also facilitates the development of advanced optical devices, for example, for automated visual inspection [4–13]. Unfortunately, completely capturing the BRDF with state-of-the-art techniques requires a time-consuming mechanical scanning of incident angles and of the observation angles [14].

To mitigate this drawback, we introduce a novel approach for approximating the BRDF of an object under test by means of computational imaging methodology. We employ a very high-dimensional light source (i.e., a computer display) to illuminate the inspected object with a certain pattern of different directions of incidence and observe the reflections resulting with a common camera. We employ a novel technique called inverse rendering to derive an approximated BRDF that best explains the obtained image [15]. To describe the BRDF, we will rely on so-called principled BRDFs, which are easy to interpret for humans and therefore allow the discrimination of different materials [16, 17]. Principled BRDFs use interpretable parameters like metallic, specular, roughness, etc. (c.f. Fig. 5) to describe the interaction of a surface with incident light. Instead of heuristically choosing the illumination pattern, we optimize it in a data-driven, end-to-end manner.

The remaining sections of this article are structured as follows. Section 2 describes how we use inverse rendering to approximate the BRDF. In Sec. 3, we explain how we optimized the employed illumination pattern. Section 4 shows and discusses the results of the experiments performed. Eventually, Sec. 5 summarizes our contribution and provides directions for future research.

2 Inverse Rendering

Inverse rendering is a sophisticated technique used in computer graphics and computer vision to deduce the underlying parameters of a virtual scene from a reference image, enabling the generation of realistic renderings [18]. The primary goal of inverse rendering is to estimate scene attributes such as lighting, material properties (i.e., BRDFs), and geometry, so that the synthesized image closely aligns with an observed reference image. This process typically involves formulating an optimization problem in which the objective is to minimize the difference between the rendered image produced by a virtual scene and the target reference image. A key ingredient in this optimization framework is automatic differentiation, which allows efficient computation of gradients with respect to the scene parameters [19]. Using automatic differentiation, the rendering process can be treated as a differentiable function, enabling the use of gradient-based optimization techniques to iteratively refine the parameters of the scene. This capability is particularly important in complex scenes, where traditional numerical methods may struggle to converge or require excessive computational resources. Automatic differentiation improves the speed and accuracy of the optimization process and facilitates the exploration of high-dimensional parameter spaces. By effectively combining rendering techniques with optimization methods and automatic differentiation, one can determine virtual scene parameters so that rendered images accurately reflect real-world lighting and material conditions.

For our approach, we employ the Mitsuba 3 raytracer [15], which supports inverse rendering due to its holistic differentiability. Our proposed optical setup for approximating the BRDF consists of a camera that observes the reflections of a computer screen via the surface of the inspected object, i.e., the object of which we want to approximate the BRDF (c.f. Fig. 1). Since the camera indirectly observes the screen over the surface of the object, the BRDF of the inspected object has a significant influence on the light transport between screen and camera. The camera image that results for a certain screen pattern (c.f. Sec. 3) is used as a reference image $\mathbf{g}_{\text{ref}}$.

We now replicate this optical setup in a virtual scene and model the BRDF of a virtual object using a principled BRDF whose parameter values are represented by the vector $\boldsymbol{\xi}$ and encode the pattern displayed

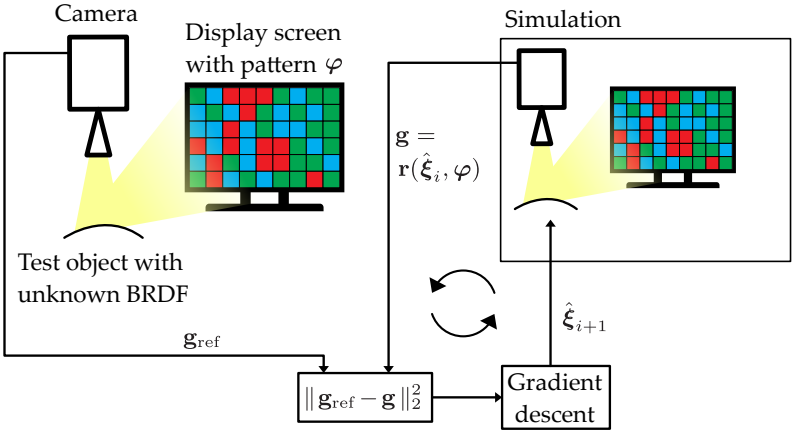


Figure 1: The concept of inverse rendering applied for BRDF approximation: (left) A camera captures the reference image $\mathbf{g}_{\text{ref}}$ of the inspected object under the illumination with pattern $\boldsymbol{\varphi}$; (right) by means of inverse rendering the approximation $\hat{\boldsymbol{\xi}}$ of the BRDF is iteratively optimized until the difference $\|\mathbf{g}_{\text{ref}} - \mathbf{g}\|_2^2$ between the reference image $\mathbf{g}_{\text{ref}}$ and the rendered image $\mathbf{g}$ is minimized.

by the screen with the vector $\boldsymbol{\varphi}$. We denote the rendering process by the function

$$\mathbf{g} = \mathbf{r}(\boldsymbol{\xi}, \boldsymbol{\varphi}) \,, \tag{1}$$

that yields the camera image $\mathbf{g}$ in vector notation resulting for the object BRDF parameterized by $\boldsymbol{\xi}$ and the displayed pattern $\boldsymbol{\varphi}$. To assess the similarity between the rendered image $\mathbf{g}$ and the reference image $\mathbf{g}_{\text{ref}}$ we employ the mean squared error calculated via

$$\ell(\boldsymbol{\xi}, \boldsymbol{\varphi}, \mathbf{g}_{\text{ref}}) = \|\mathbf{r}(\boldsymbol{\xi}, \boldsymbol{\varphi}) - \mathbf{g}_{\text{ref}}\|_2^2 = \|\mathbf{g} - \mathbf{g}_{\text{ref}}\|_2^2 \,, \tag{2}$$

which we call the loss function and which we want to minimize by choosing $\boldsymbol{\xi}$ adequately. Since the rendering function $\mathbf{r}(\,\cdot\,)$ is differentiable, we are able to compute the gradient $\nabla_{\boldsymbol{\xi}} \ell$ of ℓ with respect to the BRDF parameters $\boldsymbol{\xi}$ via automatic differentiation.

Eventually, this allows us to minimize ℓ by means of gradient descent in an iterative manner. Starting with an initial guess $\hat{\boldsymbol{\xi}}_1$ of the BRDF parameters, we can approach the sought minimum by performing steps

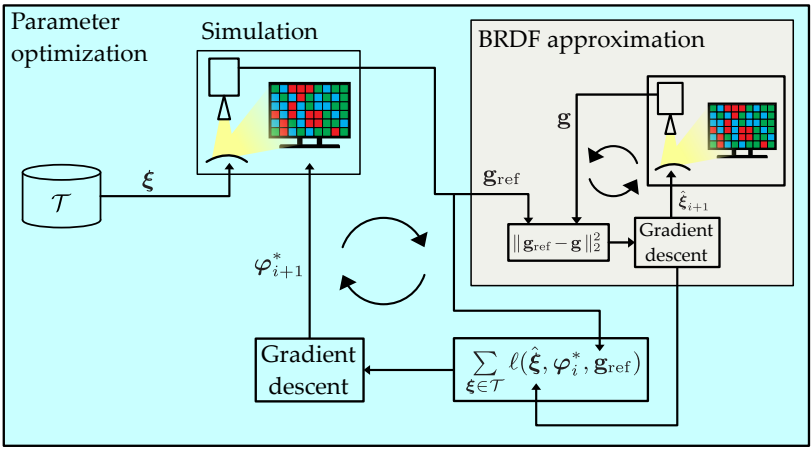


Figure 2: The parameter optimization: From training BRDFs $\xi \in \mathcal{T}$ the corresponding reference images $\mathbf{g}_{\text{ref}}$ are synthesized by using the current illumination pattern $\boldsymbol{\varphi}^*$ and fed into the BRDF approximation algorithm to obtain the reconstructed estimates $\hat{\xi}$. The resulting reconstruction errors are aggregated for all $\xi \in \mathcal{T}$ and then a gradient descent step for $\boldsymbol{\varphi}^*$ is performed to iteratively minimize this loss. The resulting illumination pattern is then used for the next iteration.

in the inverse direction of the gradient, i.e.,

$$\hat{\xi}_{i+1} = \hat{\xi}_i - \eta \nabla_{\xi} \ell(\hat{\xi}_i, \boldsymbol{\varphi}, \mathbf{g}_{\text{ref}}) \,, \tag{3}$$

with $\eta \in \mathbb{R}^+$ representing the so-called learning rate.

3 Illumination Pattern Optimization

Instead of employing a random pattern $\boldsymbol{\varphi}$ for the illuminating screen, we want to follow the approach of end-to-end optimization by also optimizing $\boldsymbol{\varphi}$ for the purpose of BRDF approximation.

To achieve this, we create a set $\mathcal{T}$ with $|\mathcal{T}| = N$ of training BRDFs. We follow the same approach as for the BRDF approximation described in the previous Section but now with the overarching goal of optimizing the screen pattern $\boldsymbol{\varphi}$. We now want to find an illumination pattern $\boldsymbol{\varphi}^*$ that minimizes the overall loss with respect to all training BRDFs,

i.e.,

$$\boldsymbol{\varphi}^* = \arg\min_{\boldsymbol{\varphi}} \ell_{\text{tot}}(\boldsymbol{\varphi}, \mathcal{T}) = \arg\min_{\boldsymbol{\varphi}} \sum_{\boldsymbol{\xi} \in \mathcal{T}} \ell(\hat{\boldsymbol{\xi}}, \boldsymbol{\varphi}, \mathbf{r}(\boldsymbol{\xi}, \boldsymbol{\varphi}))\,. \tag{4}$$

To find $\boldsymbol{\varphi}^*$, we again make use of the automated differentiability of Mitsuba to obtain the gradient $\nabla_{\boldsymbol{\varphi}} \ell_{\text{tot}}$ of the overall loss function ℓ_{tot} with respect to $\boldsymbol{\varphi}$ and iteratively perform gradient descent:

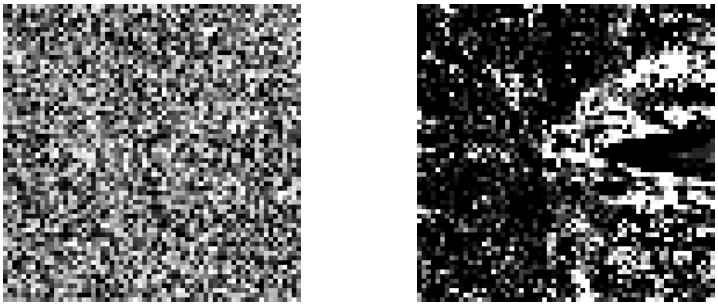
$$\boldsymbol{\varphi}^*_{i+1} = \boldsymbol{\varphi}^*_i - \tau \nabla_{\boldsymbol{\varphi}} \ell_{\text{tot}}(\boldsymbol{\varphi}^*_i, \mathcal{T})\,, \tag{5}$$

with $\tau \in \mathbb{R}^+$ denoting the respective learning rate. Please note that in order to evaluate the inner loss functions $\ell(\hat{\boldsymbol{\xi}}, \boldsymbol{\varphi}, \mathbf{r}(\boldsymbol{\xi}, \boldsymbol{\varphi}))$ for all $\boldsymbol{\xi} \in \mathcal{T}$, we rely on the optimization rule from (3) determined in the previous Section to compute the respective approximations $\hat{\boldsymbol{\xi}}$. So (5) can be imagined as two nested optimizations, the outer one optimizing $\boldsymbol{\varphi}^*$ and the inner one(s) optimizing the $\hat{\boldsymbol{\xi}}$.

4 Experiments

To evaluate our approach we employed Mitsuba to generate reference images corresponding to randomly chosen BRDFs and tried to reconstruct them with the optimization procedure from (3). Furthermore, we optimized the illumination pattern and then again tried to reconstruct the BRDFs for simulated reference images. For our experiments, we used a sphere as a test object and placed it beside a screen similar to the setup shown in Fig. 1. The BRDF has been parameterized as a principled BRDF with a focus on 8 parameters as listed in Fig. 5. Figure 3 shows the result of the pattern optimization approach described in Sec. 3. Clearly, there is more structure in the pattern and especially on the left half it seems to be less noisier. The cone-like black structure on the right half of the pattern corresponds to the screen region that is hidden in the shadows of the sphere. Hence, it cannot contribute to the camera image, consequently does not play a role in the BRDF estimation process and received values of zero during the optimization.

Figure 4 shows the qualitative results of the approximation of a slightly specular and greenly tinted BRDF. This already visually demonstrates the much faster convergence achieved by the optimized



(a) Random screen pattern (b) Optimized screen pattern

Figure 3: Optimization of the employed screen pattern: (a) the initial, random and unoptimized pattern; (b) the optimized pattern.

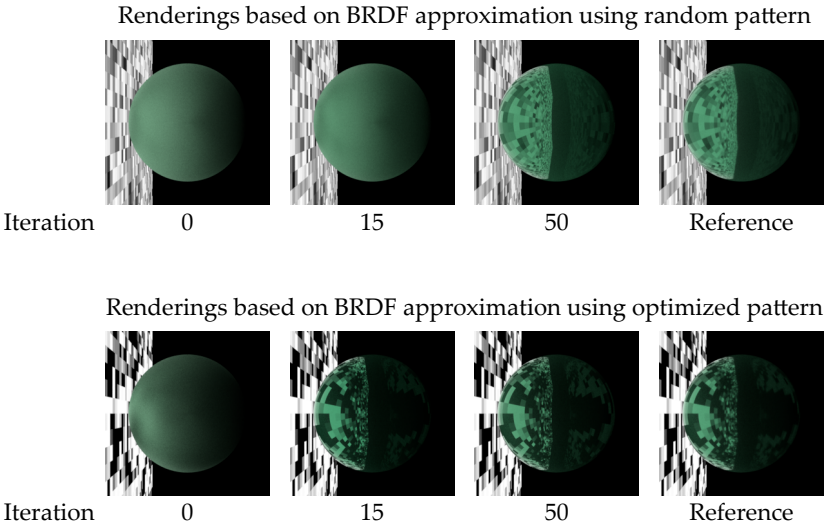


Figure 4: Qualitative results of the BRDF approximation: the random screen pattern (top row) leads to a notably slower convergence towards the target reference image (last column) compared to the optimized pattern (bottom row) for which the optimization already converged after 15 iterations.

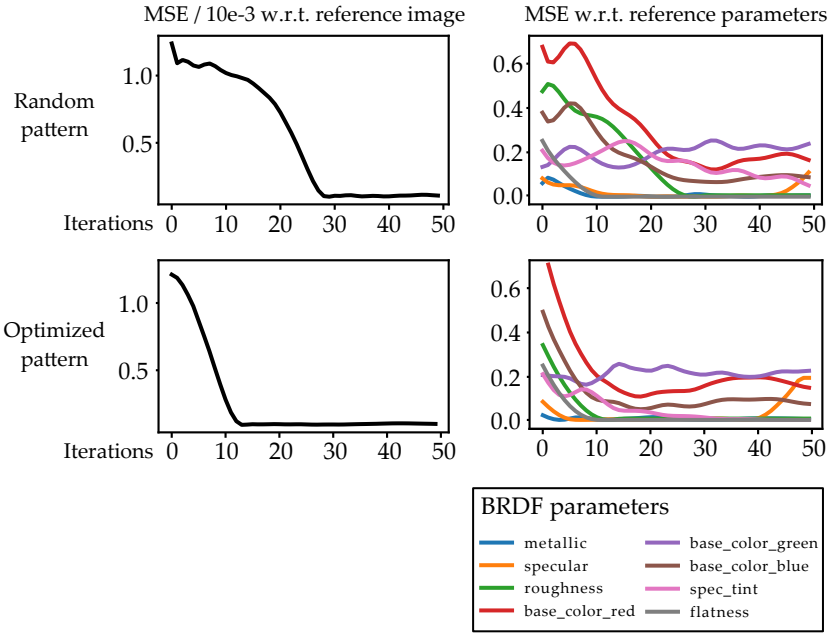


Figure 5: Quantitative evaluation of the mean squared error during the BRDF approximation: for the random screen pattern (top row) the pixel-wise error between the intermediate rendered images with respect to the reference image (left column) converges after about 30 iterations whereas convergence is reached after only 15 iterations with the optimized pattern (bottom row). A similar observation can be made for the errors of the intermediate BRDF approximations $\hat{\boldsymbol{\xi}}_i$ with respect to the ground truth BRDF $\boldsymbol{\xi}$ (right column).

screen pattern in contrast to the random pattern (which, however, also converged after more iterations). This observation is also quantitatively evident as visualized in Fig. 5. The approximation loss with respect to the individual BRDF parameters decreases more quickly for the optimized screen pattern compared to the random pattern. Interestingly, the error of the BRDF parameters does not completely disappear even though there is no further significant error in the image space between the image rendered with the approximated BRDF and the reference image. This could be explained by an effect similar to metamerism in

human color perception where multiple color spectra can lead to the same color impression due to the mapping from the high-dimensional spectral space to the only three-dimensional space of human color perception [2, 3]. In our case, this could mean that multiple principled BRDFs could lead to the same (RGB) color image.

5 Conclusion

In this contribution, we presented a novel approach for material characterization through inverse rendering, focusing on the approximation of the bidirectional reflectance distribution function. Using computational imaging and optimized illumination patterns, our method significantly enhances the efficiency of BRDF estimation compared to traditional mechanical scanning techniques. The experimental results demonstrated that our optimized illumination pattern facilitates faster convergence and more reliable approximations of the material properties. This advancement could be beneficial for practical applications in fields such as material identification and computer graphics, but also opens avenues for future research in automated visual inspection. Our findings underscore the potential of inverse rendering methodologies in improving material characterization processes, paving the way for the development of advanced optical devices for specific applications. Future work will explore further optimizations and the applicability of our approach to a wider range of materials and conditions. In addition, it is necessary to evaluate the approach by means of practical experiments outside of simulation.

References

1. M. Fox, *Optical properties of solids*, repr., with corr ed., ser. Oxford master series in physics Condensed matter physics. Oxford: Oxford Univ. Press, 2008, no. 3.
2. J. Beyerer, F. Puente León, C. Frese, and J. Meyer, *Automatische Sichtprüfung: Grundlagen, Methoden und Praxis der Bildgewinnung und Bildauswertung*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2024.
3. J. Beyerer, F. Puente León, and C. Frese, *Machine vision: automated visual*

inspection: theory, practice and applications. Berlin Heidelberg: Springer, 2016, oCLC: 910739334.

4. J. Meyer, W. Melchert, M. Hartrumpf, T. Längle, and J. Beyerer, "SNR-optimized image fusion for transparent object inspection," in *Unconventional Optical Imaging*, vol. 10677. SPIE, 2018, p. 106770A.
5. J. Meyer, T. Langle, and J. Beyerer, "Acquiring and processing light deflection maps for transparent object inspection," in *2016 2nd International Conference on Frontiers of Signal Processing (ICFSP)*. Warsaw, Poland: IEEE, Oct. 2016, pp. 104–109.
6. J. Meyer, T. Längle, and J. Beyerer, "General Cramér-von Mises, a Helpful Ally for Transparent Object Inspection Using Deflection Maps?" in *Image Analysis*, P. Sharma and F. M. Bianchi, Eds. Cham: Springer International Publishing, 2017, vol. 10269, pp. 526–537, series Title: Lecture Notes in Computer Science.
7. J. Meyer, T. Längle, and J. Beyerer, "About the Acquisition and Processing of Ray Deflection Histograms for Transparent Object Inspection," in *Irish Machine Vision & Image Processing Conference Proceedings*, 2016, pp. 9–16.
8. J. Meyer, "Light Field Methods for the Visual Inspection of Transparent Objects," Ph.D. dissertation, 2019, medium: PDF Publisher: KIT Scientific Publishing.
9. J. Meyer, "Visual inspection of transparent objects physical basics, existing methods and novel ideas," in *Proceedings of the 2014 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory*, vol. 20. KIT Scientific Publishing, 2015, pp. 37–47.
10. J. Meyer, "Overview on machine vision methods for finding defects in transparent objects," in *Proceedings of the 2015 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory*, vol. 24, 2015, pp. 103–112.
11. J. Meyer, T. Längle, and J. Beyerer, "About acquiring and processing light transport matrices for transparent object inspection," *tm - Technisches Messen*, vol. 83, no. 12, pp. 731–738, Dec. 2016.
12. J. Meyer, T. Längle, and J. Beyerer, "Acquisition and processing of light transport matrices for automated transparent object inspection," in *Forum Bildverarbeitung 2016. Hrsg.: M. Heizmann*. KIT Scientific Publishing, 2016, pp. 75–86.
13. J. Meyer, T. Langle, and J. Beyerer, "Towards light transport matrix processing for transparent object inspection," in *2017 Computing Conference*. London, UK: IEEE, Jul. 2017, pp. 244–248.

14. M. López Martínez and T. Hartmann, "Multispectral gonioreflectometer facility for directional reflectance measurements and its use on materials and paints," in *Target and Background Signatures IV*, K. U. Stein and R. Schleijpen, Eds. Berlin, Germany: SPIE, Oct. 2018, p. 32.
15. W. Jakob, S. Speierer, N. Roussel, M. Nimier-David, D. Vicini, T. Zeltner, B. Nicolet, M. Crespo, V. Leroy, and Z. Zhang, "Mitsuba 3 renderer," 2022.
16. S. McAuley, S. Hill, N. Hoffman, Y. Gotanda, B. Smits, B. Burley, and A. Martinez, "Practical physically-based shading in film and game production," in *ACM SIGGRAPH 2012 Courses*. Los Angeles California: ACM, Aug. 2012, pp. 1–7.
17. L. Fascione, J. Hanika, M. Fajardo, P. Christensen, B. Burley, and B. Green, "Path tracing in production - part 1: production renderers," in *ACM SIGGRAPH 2017 Courses*. Los Angeles California: ACM, Jul. 2017, pp. 1–39.
18. B. Nicolet, A. Jacobson, and W. Jakob, "Large Steps in Inverse Rendering of Geometry," *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, vol. 40, no. 6, Dec. 2021.
19. M. Bartholomew-Biggs, S. Brown, B. Christianson, and L. Dixon, "Automatic differentiation of algorithms," *Journal of Computational and Applied Mathematics*, vol. 124, no. 1-2, pp. 171–190, Dec. 2000.

Effective directional reflectance modelling of fine-structured surfaces

Andreas Polywka and Carl F. v. Carmer

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Gutleuthausstr. 1, 76275 Ettlingen, Germany

Abstract An additive superposition approach is presented to model the effective directional surface reflectance of optically rough fine-textured planar surfaces. As examples, effective reflectance models for randomly droplet-occupied surfaces and for periodic two-dimensional sawtooth-like fine-structured surfaces are constructed and parameterised. Comparative ray-optical simulation results confirm the applicability of the approach to effective reflectance models of optically rough, finely structured surfaces.

Keywords Reflectance modelling, macrostructures, numerical simulation, BRDF

1 Introduction

The reflectance of plane technical surfaces is contigent on their surface properties and is described in its optical effect by the surface roughness, which is of the order of magnitude of the wavelength, $k_{opt} = \mathcal{O}(\lambda)$. Small geometric changes can also be randomly or regularly imprinted on the surface. These geometric fine-structures can possess a characteristic spatial dimension l, which is significantly larger than the wavelength of the radiation, but significantly smaller than the characteristic length of the plane surface, $\mathcal{O}(\lambda) \ll \mathcal{O}(l) \ll \mathcal{O}(L)$.

In order to perform numerical calculations of light propagation in large-scale scenes using ray-optical methods, it is necessary to consider the effect of such fine structures on large flat surfaces using suitable modelling approaches, as opposed to calculating the ray interaction

on fine structures with geometric resolution. To this end, a methodology for the directional reflectance modelling of macro-structures is hereby presented. This methodology employs the superposition of homogeneous models as the basis for a modelling and parameterisation procedure for effective, area-averaged reflectance modelling of heterogeneous macro-structured surfaces.The reflectance modelling method is illustrated using two concrete examples: droplet-covered surfaces and technically two-dimensionally fine-structured surfaces.

Related Work. Semi-analytical reflectance models can elegantly and efficiently describe complex optical properties of homogeneous surfaces, e.g. [1–3]. Various methods are used for the development and parameterisation of BRDF models. Many analytical models describe large-scale material phenomena with the geometry of a microfacet-based BRDF and depict the optical appearance of rough surfaces [4,5]. Shi et al. [6] use machine learning methods to form the BRDF of periodically arranged structures. Magnin et al. [7] have developed a ray tracing algorithm to calculate the reflectance of pyramidal structures in the micrometre range (approx. 100 µm). Gartley et al. [8] use special BRDF simulations to determine the reflectance of contaminated surfaces, but do not address the special optical properties of the contaminations. None of the methods can be used generically to develop an analytical BRDF model for geometric fine-structures.

2 Test scenarios and configuration of numerical ray-optical simulations

2.1 Simulation of plane carrier surface

The reflectance is determined using ray-optical methods (numerical simulation) in the simulation software COMSOL Multiphysics®. The main application Comsol Multiphysics Version 6.2, as well as the *Ray Optics* module and a CAD module are used for this investigation. For ray tracing, Comsol Ray Optics uses 4 Intel Xeon Gold 6248 processors with 80 cores on a Windows Server 2019 operating system. The available RAM is 512 GB.

The simulations are always carried out with unpolarised, perfectly

parallel and homogeneous radiation. The azimuth is set to 0°. The BRDF (Bidirectional Reflectance Distribution Function) is determined with at least 20,000 rays and for 6 to 12 angles of incidence. The wavelength is 1550 nm, which was chosen due to its technical relevance in practice. For example, LIDAR [9] or autonomous driving systems [10] often use this wavelength. For the simulation data determined in this way, the corresponding angles of incidence are calculated and converted into an angle-dependent intensity representation. This is done using a Python script.

Like most real surfaces, the analysed surface is not optically smooth. The plane carrier surface is therefore associated with a reflectance function to model the roughness effect. The surface reflectance for simulating optical roughness can be described, for example, by a Gaussian function (1) [11–14].

$$R_{surface}(\vartheta) = \frac{I}{\sqrt{2\pi}\cdot\sigma}\cdot e^{-\frac{1}{2}\left(\frac{\vartheta-\vartheta_i}{\sigma}\right)^2} \quad . \tag{1}$$

I is the sum of all ray intensities and corresponds to the area of the Gaussian function. The standard deviation of the Gaussian function σ has a direct physical relationship to the geometric roughness [11–14]. ϑ_i is the angle of incidence from which the angle of specular reflection is derived. Using BRDF measurement data, the scattering width of the carrier material is determined as $\sigma = 3°$.

2.2 Simulation of individual droplets on carrier surface

Knowledge of the refractive index and the contact angle is required for the simulation of an (idealised) water droplet in the COMSOL simulation environment. The refractive index of water has been extensively researched [15, 16], so that further material investigations are not necessary for this study. The contact angle of a water droplet is formed by the interaction of the surface tension of the fluid, the surface energy of the carrier material and the surrounding medium [17], as well as the roughness of the carrier material [18, 19]. As a rule, the formation of the contact angle is very complex and the contact angle cannot always be reliably calculated using these parameters, but must be measured physically.

In the experimental measurements, the Young-Laplace model is used to determine the drop contour on the carrier surface in order to determine the resulting contact angle. The measurement is carried out for water droplets with different radii. The result shows an almost constant contact angle of 72° for radii from 1.75 mm to 3 mm. The water droplet is modelled in Comsol Multiphysics using the same Young-Laplace model and with a constant contact angle (72°) and radius (3 mm).

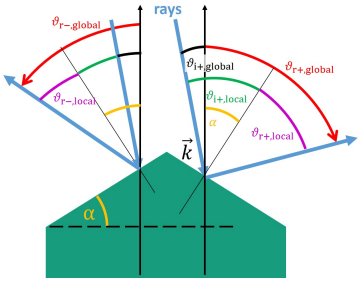
For later verification of the area-averaged reflectance model, ensembles of 1, 10, and 15 drops are randomly distributed on the substrate and irradiated over their entire area. The reflectance data obtained from the ensemble simulations is used to evaluate the accuracy of the model (see section 3.2).

2.3 Simulation of fine-structure elements on carrier surface

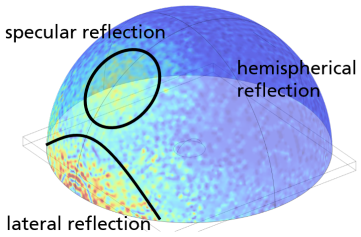
For the second application example, a generic fine-structure is selected as a single element, which can be described as a periodic two-dimensional grid structure with a sawtooth-like cross-sectional profile (see figure 1(a)). The 'sawtooth' flanks are directly adjacent to each other, there are no horizontal surface strips of the unaffected carrier surface in between. The structure angle α is used as a variable geometric parameter of the fine-structure. The structure angle is defined as the angle between the horizontal plane and the flanks. Consequently, a structure angle of 0° results in a structure-free plane surface, whose reflectance is determined solely by the optical roughness and the refractive index of the substrate surface. The flank length is chosen to be constant at 2 cm, resulting in further geometric parameters such as the height of the fine-structure and the grating constant, depending on α.

3 Effective reflectance model: modelling, plausibility check and discussion

The effective reflectance model of a plane surface demonstrated in section 3.4 will use a superposition approach for modelling the directional surface reflectance by optical roughness of the surface (section 3.1) as well as by structural roughness elements imprinted on the surface (sections 3.2 and 3.3).



(a) Definitoric sketch of fine-structure



(b) Reflectance regimes of individual droplets

Figure 1: Definitoric sketches of individual structural elements of a plane surface: droplet-covered surface, and sawtooth-like fine-structured surface. (a): Definition sketch of sawtooth profile geometry and parameter definitions of fine-structure. (b): Reflectance regimes of a single droplet on a plane, moderately rough carrier surface, shown on a hemispherical reflected radiance distribution resulting from exclusive irradiation of the droplet itself.

3.1 Reflectance modelling of plane carrier surface without fine-structures

In this study, the reflectance of the rough carrier surface is reproduced using the Generalised-Harvey-Shack (GHS) model [13, 14] as the most suitable model. This is a semi-analytical physical reflectance model for optically smooth to moderately rough surfaces, which requires the geometric surface roughness and the complex refractive index of the substrate material as input information. The material parameters for parameterising the GHS model were determined from a reflectance measurement of the real substrate object using the Fraunhofer-IOSB gonioreflectometer.

The complete determination of the (effective) complex refractive index requires a transmission measurement, which cannot be carried out with opaque materials. However, a successful transmission measurement is absolutely essential for determining the absorption coefficient. Instead, the complex refractive index and surface roughness can be determined from the inverse solution of the GHS model with a single BRDF measurement of the homogeneous carrier surface without

further material measurements by fitting the model parameters. The fitting requires an assumption on the statistical height distribution of the surface (PSD power spectrum density). Polished surfaces have statistical height distributions that can be described with an inverse power law [14]. In many cases, the local height distribution of a rough surface is normally distributed [12–14].

The Python script by Johansen [20] was modified for this purpose; in particular, a new PSD function that resembles a Lorentz or Cauchy distribution has proven successful. The parameterisation of the inverse GHS model using the modified script provides a refractive index of the carrier with $n_{carrier} = 1.63 + 0.94j$ at 1550 nm. The roughness is obtained as $\sigma = 0.724$ µm.

The theoretical reflectance (perpendicular incidence of radiation) is 16%, while 13% is measured. The calculated value is close to the measured value, which is why the refractive index determined is considered plausible. The ratio σ/λ is high at 0.5, which indicates an optically moderately rough surface. The narrow width of the reflectance speaks against this. Moderately rough surfaces can reach a half-width of 20° and more in the reflectance [14,21]. If the parameters are used in the GHS model with a Gaussian PSD, a half-width of 10°is measured, which is due to the different PSD and makes the result plausible. Reference samples made of aluminium with clearly defined roughness (Ra $\approx$ 1µm) support the result. The reference sample shows a significantly higher scattering.

3.2 Reflectance modelling of single droplets

Figure 1(b) shows the simulation result for a single drop. The hemispherical distribution of the reflected radiance shows that a reflectance model of a drop has to take three optical effects into account.

Specular reflection. Specular reflection follows the law of reflection and is characterised here by $\vartheta_i \approx \vartheta_r$. It is largely independent of the angle of incidence. In the zenith angle range around $\vartheta_i \approx 0^\circ$ (zenith) there is an underestimation of approx. 26 % at the maximum.

Hemispherical reflectance. The hemispherical reflectance is caused by the interaction at the water-air interface. Due to the droplet shape, the propagation is not directional, but spreads radially in space. According to Fresnel's formulae and the refractive index of water [15], the reflectance is low and approx. 2 % for angles of incidence up to 45°. It can, therefore, be modelled in simplified form as isotropic reflectance.

Lateral reflectance. The lateral reflectance spreads parallel to the carrier surface. It only starts to become relevant from an angle of incidence of 27.5° (zenith) and increases rapidly in intensity.

The sum of the reflectances of the three reflectance regimes forms the reflectance R_{drop} of the individual droplet, which can be described by an empirical model as follows (for detail cf. [22])

$$R_{drop} = \underbrace{0.0094 \cdot e^{\frac{-0.5 \cdot (\vartheta - \vartheta_i)^2}{18.8^2}}}_{specular\ regime} + \underbrace{0.02 \cdot e^{\frac{-(\vartheta - 120° + \vartheta_i)}{12.5}}}_{lateral\ regime} + \underbrace{0.003}_{hemisph.\ regime} \tag{2}$$

3.3 Reflectance modelling of fine-structured surfaces

Reflectance modelling makes it necessary to distinguish between the 'global' angle of incidence $\vartheta_{i,global}$, defined as the angle of incidence with respect to a featureless surface, and the 'local' angle of incidence at the fine-structure. The latter is of decisive importance for the application of Fresnel's formulae. The (global) reflectance angle is the last interaction of the radiation with the fine-structure. Figure 1(a) shows the most important parameters.

The figure shows the interaction of the rays with a structural element. The ray bundles are split by the triangular elements. The position of the flanks must be taken into account. Due to the laws of geometrical optics, the angles of the fine-structure can be subtracted or added to the angle of incidence. Each partial ray bundle has its own reflectance angle, which is calculated with $\vartheta_{r+,global} = \vartheta_{i,global} + 2 \cdot \alpha$ or with $\vartheta_{r-,global} = \vartheta_{i,global} - 2 \cdot \alpha$. The reflected intensity is calculated by the 'local' angle of incidence using Fresnel's formulae $R_{Fresnel}(\vartheta_{i,local})$.

As long as $\vartheta_{i,local} = \vartheta_{i,global} + \alpha < 90°$ is fulfilled, the rays bundles are split. With perpendicular incidence, rays are split in a ratio of 1:1;

the ratio shifts with increasing angle of incidence $\vartheta_{i,global}$. The splitting ratio is in accordance with the conservation of energy:

$$E_- = 0.5 \frac{l_{flank} \cdot \left(\sin(90^\circ - \alpha) + \sin(\vartheta_{i,global}) \right)}{l_{flank} \cdot sin(90^\circ - \alpha)} \quad \text{and} \tag{3}$$

$$E_+ = 0.5 \frac{l_{flank} \cdot \left(\sin(90^\circ - \alpha) - \sin(\vartheta_{i,global}) \right)}{l_{flank} \cdot sin(90^\circ - \alpha)} \quad , \tag{4}$$

where $E_- + E_+ = 1$.

The reflectance model for the first, direct interaction with the fine-structure can thus be represented as follows

$$R_+ = R_0(\vartheta_{r+,global}) \cdot R_{Fresnel}(\vartheta_{i,global} + \alpha) \cdot E_+ \quad \text{and} \tag{5}$$

$$R_- = R_0(\vartheta_{r-,global}) \cdot R_{Fresnel}(|\vartheta_{i,global} - \alpha|) \cdot E_- \quad , \tag{6}$$

where R_0 denotes the reflectance function of the structure-free surface according to equation (1).

The first-order reflectance model is extended by a second reflection, which takes place when $\vartheta_{r+,global} = \vartheta_{i,global} + 2\,\alpha \geq 90^\circ$ is fulfilled. By further applying the angle theorems, the global reflectance angle $\vartheta_{2r+,global} = 180^\circ - \vartheta_{i,global} - 4 \cdot \alpha$ is calculated. To simplify matters, it is assumed that no further splitting of the ray bundles takes place and *all* rays interact once more with the fine-structure. The reflectance function R_+ is extended by an additional Fresnel term. The reflectance function for the second-order reflection at the fine-structure is

$$\begin{aligned} R_{2+} =& R_0(\vartheta_{r+,global}) \cdot R_{Fresnel}(\vartheta_{i,global} + \alpha) \cdot E_+ \\ & \cdot R_{Fresnel}(\vartheta_{2r+,global} + \alpha) \quad . \end{aligned} \tag{7}$$

The total reflectance R_{fs} of the fine-structure then results as the sum of the parts (5), (6) and (7) to

$$\begin{aligned} R_{fs} =& R_+ + R_- + R_{2+} \\ =& R_0(\vartheta_{r+,global}) \cdot R_{Fresnel}(\vartheta_{i,global} + \alpha) \cdot E_+ \\ & \cdot \left(1 + R_{Fresnel} \left(\vartheta_{2r+,global} + \alpha \right) \right) \\ & + R_0(\vartheta_{r-,global}) \cdot R_{Fresnel} \left(\left| \vartheta_{i,global} - \alpha \right| \right) \cdot E_- \quad . \end{aligned} \tag{8}$$

3.4 Effective reflectance model for plane surfaces with additional fine-structures

The effective reflectance model R_{eff} of a plane carrier surface with imprinted or impressed structural elements superimposes the reflectance R_{free} of the unaffected optically rough surface and the reflectance R_{strc} due to fine-structures. The total area of a surface is spit into an unaffected part and a part with structural elements, $A = A_{free} + A_{strc}$; the following therefore applies to its fill factors: $A_{free}/A + A_{strc}/A = FF_{free} + FF_{strc} = 1$. Thus, the effective reflectance model is formulated as

$$R_{eff} = R_{free}\,(1 - FF_{strc}) + R_{strc}\;FF_{strc} \quad , \tag{9}$$

where a parameterised GHS model is used as the reflectance model R_{free} of the unaffected surface (see section 3.1). The reflectance model R_{strc} is obtained from the radiation interaction on a structure object of the surface. Interactions between individual structural elements can also be taken into account, so that a structure object can be thought of as an ensemble of a few individual elements. Equations (2) and (8) show examples of structure reflectance models.

Application 1: Effective reflectance model for droplet-covered surfaces. The effective reflectance model (9) is for a droplet-covered surface (see section 2.2):

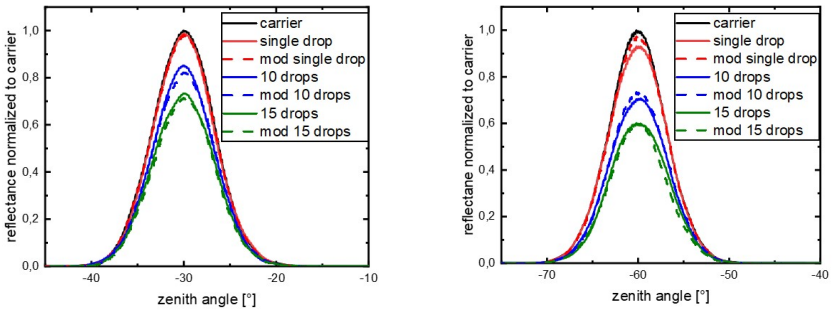
$$R_{ndrop} = R_{carrier} \cdot (1 - FF_{ndrop}) + R_{drop} \cdot FF_{ndrop} \quad , \tag{10}$$

where is the fill factor of the drop coverage is

$$FF_{ndrop} = \left(\frac{0.5}{\cos(\vartheta_i)} + 0.5\right) \cdot \frac{n_{drop} \cdot A_{singledrop}}{A_{carrier}} \quad . \tag{11}$$

Depending on the angle of incidence ϑ_i, a drop casts a shadow on the carrier surface, the size of which is approximated by the term $0.5/\cos(\vartheta_i)$. Equation (2) is used as the reflectance model of a drop.

Figure 2 shows the results of the model as an example in comparison to simulations of a single drop (red), and of ensembles of 10 drops (blue) and 15 drops (green) for two representative angles of incidence of 30° and 60°. The result of the reflectance model is shown as



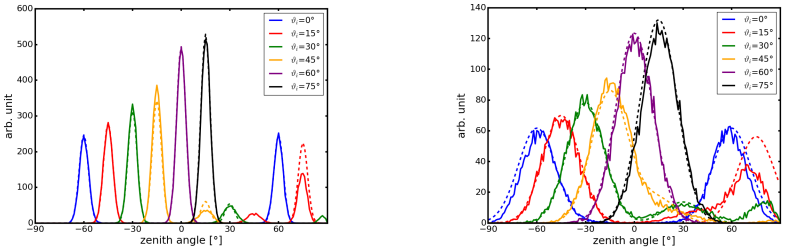
(a) $\vartheta_i = 30°$ **(b)** $\vartheta_i = 60°$

Figure 2: Comparison of the effective reflectance model (10) for a droplet-loaded surface (dashed line) with results of ray-optical simulations for ensembles with 1, 10, and 15 droplets (solid lines, red, blue, green) for incidence angles ϑ_i of 30° and 60°.

a dashed line, the numerically determined area-averaged reflectances as full lines. The results are normalised to the maximum of the carrier reflectance, whereby the normalised reflectance of the carrier surface alone is also indicated by a black full line.

The graphs show good agreement between the simulated reflectance and the reflectance model for different angles of incidence. The deviations are always below 4 %. In further investigations, the contact angle and the radius of the individual droplets were varied, assuming a Gaussian size distribution of droplets. The simulation results of the droplet ensembles also showed good agreement with the model in these cases.

The area-averaged reflectance model for drop–covered surfaces is linear and based on the reflectance of a single isolated drop. The investigations show that a surface–to–surface distance of the order of the drop radius is necessary to keep the influence of the interaction on the specular reflectance low. However, the interaction in the lateral reflectance regime cannot be completely avoided. The overall good results confirm the effective reflectance model (10) and thus the modelling approach (9).



(a) optically moderately rough, $\sigma_s = 3°$ **(b)** optically rough, $\sigma_s = 11°$

Figure 3: Comparison of the effective reflectance model (12) for a two-dimensionally fine-structured surface with a structure angle $\alpha = 30°$ (dashed line) with results of ray-optical simulations for angles of incidence $\vartheta_i \in [0°, 15°, \ldots, 75°]$ (solid lines) for roughnesses σ_s of 3° and 11°.

Application 2: Effective reflectance model for two-dimensional triangular fine-textured surfaces. The effective reflectance model R_{eff} (9) for a plane surface with a two-dimensional periodic fine-structure with a triangular cross-section is formulated using the carrier reflectance model R_{surf} from (1) and the reflectance model R_{fs} from (8) for the structural elements as

$$R_{eff} = R_{surf}\left(1 - FF_{fs}\right) + R_{fs}\, FF_{fs} \quad , \tag{12}$$

where FF_{fs} is the fill factor of the carrier surface provided with imprinted fine-structures.

For the plausibility check of the reflectance model, $FF_{fs} = 1.0$ was selected as the most unfavourable case with complete coverage of the carrier surface with fine-structures. The effective reflectance model (12) then merges into the structure reflectance model (7). The parameter study includes the structure angle α and the roughness σ_s (see equation (1)). Figures 3(a) and 3(b) show the results for roughness angles σ_s of 3° and 11° at a structure angle of 30°. For this structure angle, we see generally good agreement for all angles of incidence from 0° to 75° for both an optically moderately rough and an optically rough surface. Significant deviations only occur for $\vartheta_{i,global} = 15°$.

Comparisons of the effective reflectance model with simulation re-

sults at other structural angles α show generally good agreement up to $\alpha = 45°$ and confirm the model. For structural angles above $\alpha = 60°$, the model is incomplete and further interactions must be taken into account. These include additional interactions with the fine-structure and the further splitting of the beam's ray bundles.

4 Conclusions

Effective reflectance models of plane fine-structured surfaces, equation (9), enable a more efficient simulation of radiation propagation in large-scale scenarios. The superposition of reflectance effects on unaffected surface parts and surface parts with additional fine-structures leads to a simple additive modelling concept. A parameterised Generalised-Harvey-Shack model is used for the reflectance of the unaffected area components (section 3.1). The reflectance of a single imprinted or impressed structural element is modelled on a physical basis, as demonstrated in two use cases. A randomly drop-covered carrier surface is modelled by an empirical reflectance model (2), a two-dimensional triangular periodic fine-structure by a simplified analytical model (8). The resulting effective reflectance models for both sample surfaces, (10) and (12), are then checked for plausibility using comparative ray-optical simulation results. The good agreement between the results demonstrates that the modelling approach is also suitable for efficient use in complex large-scale propagation computations.

Acknowledgements

The financial support of the Federal Ministry of Defence under contracts no. T/K2AL/PA000/K1110 and no. E/E71S/M0587/EF032 is gratefully acknowledged.

References

1. T. Tongbuasirilai, J. Unger, J. Kronander, and M. Kurt, "Compact and intuitive data-driven BRDF models," *The Visual Computer*, vol. 36, no. 4, pp. 855–872, may 2019.

2. D. Guarnera, G. Guarnera, A. Ghosh, C. Denk, and M. Glencross, "BRDF representation and acquisition," *Computer Graphics Forum*, vol. 35, no. 2, pp. 625–650, may 2016.
3. S. Rusinkiewicz, "Reflectance models," in *Computer Vision: A Reference Guide*, 2nd ed., K. Ikeuchi, Ed. Cham: Springer International Publishing, 2021, pp. 1063–1066.
4. M. Ashikmin, S. Premože, and P. Shirley, "A microfacet-based BRDF generator," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques - SIGGRAPH '00*, ser. SIGGRAPH '00. ACM Press, 2000, pp. 65–74.
5. M. Oren and S. K. Nayar, "Generalization of Lambert's reflectance model," in *Proceedings of the 21st annual conference on Computer graphics and interactive techniques - SIGGRAPH '94*, ser. SIGGRAPH '94. ACM Press, 1994, pp. 239–246.
6. W. Shi, J. Dorsey, and H. Rushmeier, "Learning-based inverse bi-scale material fitting from tabular BRDFs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 4, pp. 1810–1823, Apr. 2022.
7. V. Magnin, J. Harari, M. Halbwax, S. Bastide, D. Cherfi, and J.-P. Vilcot, "Angle-dependent ray tracing simulations of reflections on pyramidal textures for silicon solar cells," *Solar Energy*, vol. 110, pp. 378 – 385, 2014. [Online]. Available: https://hal.science/hal-01403918
8. M. G. Gartley, J. R. Schott, and S. D. Brown, "Micro-scale modeling of contaminant effects on surface optical properties," in *Imaging Spectrometry XIII*, S. S. Shen and P. E. Lewis, Eds. SPIE, aug 2008.
9. S. Junttila, M. Holopainen, M. Vastaranta, P. Lyytikäinen-Saarenmaa, H. Kaartinen, J. Hyyppä, and H. Hyyppä, "The potential of dual-wavelength terrestrial lidar in early detection of Ips typographus (L.) infestation – leaf water content as a proxy," *Remote Sensing of Environment*, vol. 231, p. 111264, Sep. 2019.
10. J. P. Bos, A. M. Kurup, D. J. Chopp, and Z. D. Jeffries, "The Michigan Tech autonomous winter driving data set: year two," in *Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure 2021*, M. C. Dudzik, T. J. Axenson, and S. M. Jameson, Eds. SPIE, Apr. 2021, p. 8.
11. S. O. Rice, "Reflection of electromagnetic waves from slightly rough surfaces," *Communications on Pure and Applied Mathematics*, vol. 4, no. 2-3, pp. 351–378, aug 1951.
12. P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*. Pergamon Press, 1963.

13. A. Krywonos, J. E. Harvey, and N. Choi, "Linear systems formulation of scattering theory for rough surfaces with arbitrary incident and scattering angles," *Journal of the Optical Society of America A*, vol. 28, no. 6, p. 1121, may 2011.

14. J. E. Harvey, *Understanding Surface Scatter Phenomena: A Linear Systems Formulation*. SPIE, jul 2019.

15. G. M. Hale and M. R. Querry, "Optical constants of water in the 200-nm to 200-µm wavelength region," *Applied Optics*, vol. 12, no. 3, p. 555, mar 1973.

16. D. Segelstein, "The complex refractive index of water," Master's thesis, University of Missouri–Kansas City, 1981.

17. T. Young, "An essay on the cohesion of fluids," *Philosophical Transactions of the Royal Society of London*, vol. 95, no. 0, pp. 65–87, jan 1805.

18. R. N. Wenzel, "Resistance of solid surfaces to wetting by water," *Industrial & Engineering Chemistry*, vol. 28, no. 8, pp. 988–994, aug 1936.

19. A. B. D. Cassie and S. Baxter, "Wettability of porous surfaces," *Transactions of the Faraday Society*, vol. 40, p. 546, 1944.

20. V. E. Johansen, "Preparing the generalized Harvey–Shack rough surface scattering method for use with the discrete ordinates method," *Journal of the Optical Society of America A*, vol. 32, no. 2, p. 186, jan 2015.

21. K. A. O'Donnell and E. R. Mendez, "Experimental study of scattering from characterized random surfaces," *Journal of the Optical Society of America A*, vol. 4, no. 7, p. 1194, jul 1987.

22. A. Polywka, "Reflektanz von Schiffsoberflächen: Modell für tropfenbedeckte Flächen," Fraunhofer–IOSB, Ettlingen, Tech. Rep. 23-21402-31000Z-0XX, 2023.

Prediction of various material parameters using a combination of non-destructive sensors and artificial intelligence

Florian F. Linscheid, Marco Korkisch, Leon Fischer, and Markus G. R. Sause

Universität Augsburg, Institute for Materials Resource Management
Am Technologiezentrum 8, 86159 Augsburg

Abstract Nondestructive testing (NDT) methods offer a promising alternative for material quality control, although most current solutions are focused on specific parameters or need sample preparation. To address this challenge, we propose combining adapted NDT methods with artificial intelligence (AI). This approach leverages AI to analyze raw NDT data and predict material properties. The results demonstrate that NDT models, combined with AI-based models, achieve average relative errors of $5-20$ %, depending on the parameter. These findings pave the way for integrating AI-enhanced NDT techniques into automated production environments, reducing testing time and costs while enabling batch-wide material characterization.

Keywords NDT, AI, material characterization, quality control

1 Introduction

Material properties are critical in determining the design, manufacturing, and performance of products across modern production processes [1]. Understanding material behavior under varying influences is essential for selecting appropriate materials. Material characterization also underpins quality control by verifying that produced or purchased materials meet the specifications required for downstream production stages. The importance of this step is reflected in the extensive number of test standards to evaluate specific material parameters.

Traditional testing methods, while accurate and reliable in the past decades, come with significant limitations [1, 2]. These methods are often time-consuming, costly and labor-intensive, requiring specialized equipment and trained personnel. As industries transition into the era of Industry 4.0, integrating such methods into automated production environments becomes increasingly challenging.

Nondestructive testing (NDT) methods provide a compelling alternative, although most current solutions are focused on specific parameters, need sample preparation [3–5] or are more focused on classification rather than prediction of material parameters [6, 7]. Although traditionally employed for flaw detection and structural health monitoring [8], NDT methods can also provide data related to material properties, such as electrical conductivity from eddy current measurements [9]. Their key advantage is compatibility with automated processes. However, extracting material parameters directly from NDT data is complex as the relationship between NDT measurements and material characteristics involves intricate, nonlinear dependencies that cannot be easily derived analytically.

To address this challenge, we propose an integration of adapted NDT methods with artificial intelligence (AI). This approach leverages AI to analyze raw NDT data and predict material properties. In this study, we validate this concept by demonstrating how selected NDT methods, paired with custom feature extraction and machine learning models, can accurately predict material parameters.

This approach has significant potential for practical applications where a fast and cost-effective in-line quality control is held back by time- and labor-intensive lab tests. By tailoring NDT methods and AI models to specific use cases, it becomes possible to design custom measurement systems for quality control.

2 Materials and Methods

2.1 Materials

This study involved a total of 65 materials, spanning six distinct material classes: **Stones/Minerals** including granite, fire clay, graphite, cement and slate, **glass** (borosilicate, soda-lime, quartz and glass ceramic), **Wood and natural materials** including wood from beech, oak,

spruce, pine and poplar, as well as cork and composite wood products (HDF, MDF, OSB, particle board), **ceramics** in the form of Al_2O_3 and ferrite, **polymers** from the two main groups of thermoplastics (ABS, PA-6, PE-HD, PE-HMW, PE-UHMW, PETG, PMMA, PP, PS, PTFE, PVC) and thermosets (epoxy, acrylic and PU resins) and **metals** with copper, brass, different types of aluminum, steel and stainless steel.

In many cases, multiple variants of a material were sourced, either in different forms or from different suppliers. This approach aimed to include both minor and major variations in material properties to ensure comprehensive coverage.

All materials were processed into standardized sample types, part of which were subsequently used for characterization and others for evaluation with the nondestructive testing (NDT) methods. The samples for NDT testing were around 30×30 mm with thicknesses ranging from 3 to 18 mm.

2.2 Materials characterization

The classical lab methods were used to perform a wide characterization of all materials, which provide the ground truth as input data for the AI models' training process. Tests adhered to relevant standards, with at least three samples per material, and results averaged and stored with standard deviations. The following material parameters were chosen:

- Thermal Conductivity/Diffusivity: Measured at room temperature using a Linseis HFM 300 or LFA 1000.
- Emissivity: Determined using a calibrated custom setup utilizing two infrared thermometers
- Specific Heat Capacity: Assessed with a DSC 214 Polyma (NETZSCH) in temperature-modulated mode near room temperature.
- Density: Helium pycnometer (Micrometrics AccuPyc II).
- Sound Velocity: Longitudinal and transverse velocities measured via ultrasonic phase spectroscopy (UPS) using an Advantest R3754A network analyzer and Olympus transducers.

- Elastic Parameters: Elastic modulus and Poisson's ratio obtained through tensile or compression tests with a ZwickRoell universal testing machine and GOM Aramis 12M system.
- Electrical Conductivity: Measured with a bench power supply and precision voltmeters.

2.3 NDT methods

To obtain qualitative measurement data, a series of different sensor systems, based on established NDT methods, was used. Although these are not directly capable of determining quantitative material parameters, their measurement data are in correlation with certain material characteristics. Therefore, the following methods were selected:

- Active Thermography: Heat decay and distribution were analyzed using a thermal camera after laser excitation (450 nm) to infer thermal properties such as conductivity, capacity, and emissivity.
- VIS/NIR Spectroscopy: Polymer identification and spectral analysis used a SparkFun Triad AS7265x and a Spectral Engines NIRone Sensor S ($1.55 - 1.95\ \mu m$).
- Radar: An Infineon 60 GHz radar kit provided data on material interaction with radar waves, sensitive to dielectric properties.
- Acoustics: Samples excited by a vibrating ceramic plate emitted impact sounds recorded via a MEMS microphone with a sample rate of 60 kHz, linking frequency spectra to mechanical properties.
- Eddy Current: Custom sensors with excitation/sensing coils recorded responses across frequencies to analyze conductivity and magnetic properties.

2.4 Experimental setup and procedure

The experimental setup integrates all NDT sensors into a single system and allows sample thickness compensation, as well as position variability by tilting and raising the sensor head in relation to the sample. The

measurement procedure begins with spectroscopy measurements, followed by radar and thermography, where the 450 nm laser is heating the sample, while the thermal camera records video of the temperature changes. Simultaneously, eddy current data are collected. Finally acoustic measurements are performed by vibrating the ceramic plate and recording the sound of the sample. The process is repeated for 50 iterations per sample, with 65 materials tested, using three samples each. A centering mechanism resets the sample's position after each iteration, completing a cycle in approximately 30 seconds.

2.5 Feature extraction and model training

Each NDT method underwent a tailored feature extraction. For spectroscopy, data smoothing and normalization yielded key spectral features, while eddy current measurements produced amplitude and phase features from Bode plots. Radar data was summarized using channel statistics, and the acoustic signals were analyzed for time-domain characteristics using the Python software package *tsfresh* [10]. Thermography data yielded key features from the temporal evolution of the heated zone. The combined dataset was cleaned via variance thresholds and PCA.

The target vector consisted of the respective material parameters determined by the lab measurements. To stabilize the algorithm and avoid discrete values, the target values were perturbed using a normal distribution with a standard deviation of 5 % around the true value. Furthermore, the target values were scaled for training. Most values were normalized to a standard deviation of 1 and a mean of 0. Metrics varying across several orders of magnitude (e.g., Young's modulus and conductivities) were transformed into a normal distribution using a Box-Cox power transformation.

A neural network with two hidden layers (512 and 128 neurons) was selected as the machine learning algorithm. A dropout of 0.5 was added, and an Adam optimizer with a learning rate of 0.001 was trained, using the mean squared error (MSE) as loss function. The training was done in Python with the help of the libraries scikit-learn and tensorflow.

3 Results

3.1 Material parameters

The material parameter measurements provided a comprehensive dataset across six material classes. The material parameters for all materials were determined as described in chapter 2.2. In general, the measured values aligned well with datasheet and literature values, where available. The calculated standard deviation for each parameter and material was typically around 5 % or less. However, higher variance was observed in certain cases, particularly for natural materials.

3.2 Predicted material parameters

Figure 1 displays the machine learning model's predictions for all 65 materials, divided into the eight material parameters. True values are on the x-axis, and predictions on the y-axis, with blue markers representing validation data (used in training) and orange markers showing six entirely new materials: two PE, a PP, Al 1050, brass and beech. The dashed line indicates the ideal agreement between actual and predicted values. The mean absolute percentage error (MAPE) ranges from 13 % for pressure wave speed to 37 % for Young's Modulus, excluding thermal diffusivity and electrical conductivity, which are harder to predict due to their wide ranges and uneven spacing. The training was also repeated for each measurement method separately to be able to estimate the individual contributions to the model's predictive capabilities (Table 1). The results show that parameters like the pressure wave speed (c_{speed}) and specific heat capacity (c_{p}) are generally easier to predict, while thermal diffusivity (α) and electrical conductivity (σ) perform quite poor.

The results show that even individual measurement methods can help perform some parameter prediction but their combination drastically improves the results to a mean value of 22 % in the test dataset. These results show to be a promising approach to characterize a broad range of materials. In real world applications though it is unlikely that material parameters have to be predicted from this broad range. This is why the polymer class was picked to train a more specialized machine learning model. The results are shown in Figure 2. Due to the

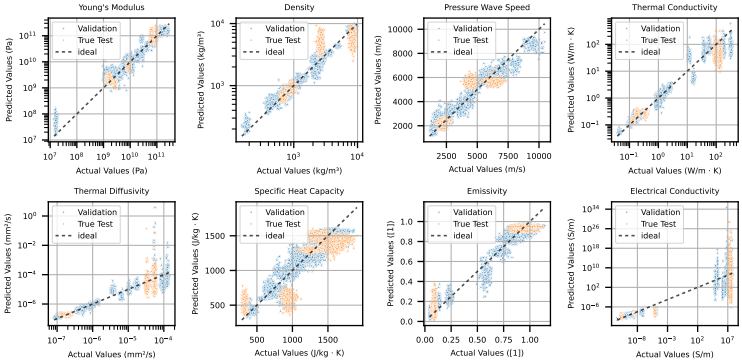


Figure 1: Predicted material parameters for a broad range of materials ranging from cork to Al_2O_3. Points in orange are materials that the machine learning model was never trained on.

poor performance, σ and α were left out of training. As test material, two PEs from different brands and one PP sample were chosen. The results increase significantly now, as shown in the last row of Table 1. In many cases now the percentage errors lie in the vicinity of standard deviations that were achieved with the lab tests and are thus inherent to the materials themselves.

4 Discussion

The results demonstrate that predicting material parameters based on NDT sensor data is a viable approach, yielding promising results. While standard material characterization achieved standard deviations of around 5 % (and up to 10 % in certain cases), the predicted parameters showed relative errors averaging between 5 % and 20 % for the polymers. Although this indicates that predictions do not yet match the accuracy of standard testing methods, it highlights the feasibility of using NDT data for parameter estimation.

The analysis of model quality revealed several key influences. Narrowing the range of training data significantly improved predictions for materials within the trained range. The choice of features also played a critical role in prediction quality. Although models using features

Table 1: Comparison of the prediction results of different machine learning models. Shown are the mean deviations from the true values in percent for Young's modulus (E), density (ρ), pressure wave speed (c_{speed}), thermal conductivity (κ), thermal diffusivity (α), specific heat capacity (c_{p}), emissivity (ε), electrical conductivity (σ) and the mean over all columns. The first six rows show results using individual measurement methods, while the last two rows present combined methods for all materials and a polymer subset.

Methods	Materials	E	ρ	c_{speed}	κ	α	c_{p}	ε	σ	Mean
NIR	all	86.3	38.9	17.3	59.4	55	28.4	77.6	10^{15}	**44.0**
VIS	all	36.5	58.9	28.1	166.5	96.1	33.9	93.7	10^{14}	**88.2**
Radar	all	61.4	39.2	17.2	50.7	42.1	25.3	94.6	10^{5}	**41.2**
Thermo	all	63.4	32.65	12.8	43.7	113.6	32.7	80.1	10^{10}	**37.9**
Acoustic	all	112.7	36.1	18.9	84.4	61.9	31.5	210.4	10^{8}	**70.6**
EC	all	141	47.6	31.2	42.9	30.8	30.6	35.5	540	**47.0**
Combined	all	37.1	24.1	12.8	35.1	226.1	22.2	22.1	10^{15}	**21.9**
Combined	poly.	14.5	9.1	5.6	23.1	-	6.4	4.1	-	**10.5**

derived from a single NDT method were already viable to predict specific parameters, they showed inferior overall performance compared to models incorporating features from all methods. Combining multiple NDT methods increases the number of relevant features and captures more direct connections to physical properties, improving overall prediction accuracy. For instance, while the best average error of a single method (thermography features) was 37.9 %, combining all methods reduced the error rate to 21.9 %.

Prediction quality also varied across material parameters. Properties such as pressure wave speed, specific heat capacity, and density exhibited relatively low error rates, whereas Young's modulus, electrical conductivity, and thermal diffusivity proved harder to predict. One potential reason for this discrepancy is the distribution of actual values. Parameters like density and specific heat capacity often have a more uniform distribution, whereas others, such as modulus and conductivity, exhibit greater variability, even on a logarithmic scale.

Additionally, the physical interdependence of certain parameters, such as thermal diffusivity (a function of density, thermal conductivity, and specific heat capacity), may introduce unwanted complexity into the model. This interdependence warrants further investigation to mitigate its impact on prediction accuracy.

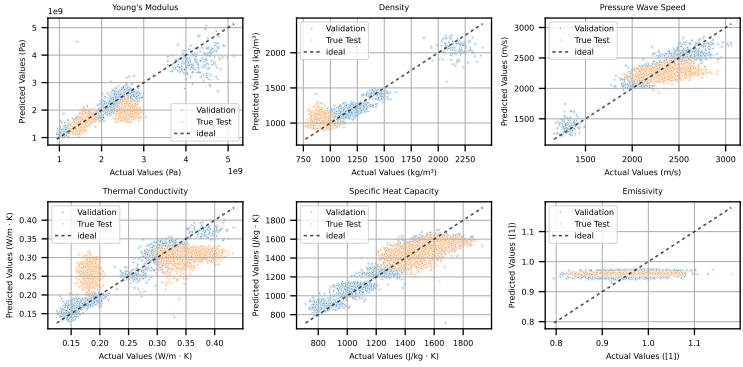


Figure 2: Predicted material parameters for the material class of polymers.

5 Conclusion

This study demonstrated that NDT methods, combined with AI-based evaluation, can predict material parameters with reasonable accuracy, achieving average relative errors of $10 - 20$ %, depending on the parameter. These findings pave the way for integrating AI-enhanced NDT techniques into automated quality control systems, reducing testing time and costs while enabling batch-wide material characterization. While the models performed well within the range of training data, their predictive accuracy decreased for materials outside this range, highlighting the need for tailored models for specific industries. In the next stages of this research, we aim to focus on more specific applications, particularly targeting the characterization of recycled polymers, a growing area of industrial importance. To further enhance the prediction accuracy and broaden the scope of measurable parameters, we will explore the integration of additional nondestructive testing methods, such as Raman spectroscopy and terahertz analysis.

Acknowledgements

We would like to express our gratitude to the Bavarian Ministry of Economic Affairs, Regional Development and Energy, as well as Bayern

Innovativ, for their funding of the project “sensAI” at the University of Augsburg through the Validierungsförderung program.

References

1. J. O. Cross, R. L. Opila, I. W. Boyd, and E. N. Kaufmann, “Materials characterization and the evolution of materials,” *MRS Bulletin*, vol. 40, no. 12, pp. 1019–1034, Dec. 2015.
2. S. V. Kalinin, M. Ziatdinov, M. Ahmadi, A. Ghosh, K. Roccapriore, Y. Liu, and R. K. Vasudevan, “Designing Workflows for Materials Characterization,” Feb. 2023, arXiv:2302.04397 [cond-mat].
3. M. Scheller, “Real-time terahertz material characterization by numerical three-dimensional optimization,” *Optics Express*, vol. 19, no. 11, p. 10647, May 2011.
4. J. Huang, C. Tian, J. Ren, and Z. Bian, “Study on Impact Acoustic—Visual Sensor-Based Sorting of ELV Plastic Materials,” *Sensors*, vol. 17, no. 6, p. 1325, Jun. 2017.
5. L. Lauck, C. Zimmer, and K. Szielasko, “Hoch ortsauflösende, bildgebende mikromagnetische Materialcharakterisierung mit magnetooptischer Sensorik,” in *DGZfP-Jahrestagung 2022*, Kassel, 2022.
6. J. Weis and A. Santra, “Material Classification using 60-GHz Radar and Deep Convolutional Neural Network,” in *2019 International Radar Conference (RADAR)*. TOULON, France: IEEE, Sep. 2019, pp. 1–6.
7. H.-S. Yeo, G. Flamich, P. Schrempf, D. Harris-Birtill, and A. Quigley, “RadarCat: Radar Categorization for Input & Interaction,” in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. Tokyo Japan: ACM, Oct. 2016, pp. 833–841.
8. S. K. Dwivedi, M. Vishwakarma, and P. A. Soni, “Advances and Researches on Non Destructive Testing: A Review,” *Materials Today: Proceedings*, vol. 5, no. 2, Part 1, pp. 3690–3698, Jan. 2018.
9. N. Bowler and Y. Huang, “Electrical conductivity measurement of metal plates using broadband eddy-current and four-point methods,” *Measurement Science and Technology*, vol. 16, no. 11, p. 2193, Sep. 2005.
10. M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, “Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package),” *Neurocomputing*, vol. 307, pp. 72–77, Sep. 2018.

Leveraging grounded SAM and a weakly supervised CNN filtering for enhanced leaf segmentation in automated plant phenotyping

Leo Fiedler[1], Felix Braig[1], Ian Howard[1], Robin Gruna[2], and Jürgen Beyerer[23]

[1] Carl Zeiss AG
Hermann-von-Helmholtz-Platz 6, 76344 Eggenstein-Leopoldshafen
[2] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Fraunhoferstraße 1, 76131 Karlsruhe, Germany
[3] Karlsruhe Institute of Technology (KIT), Vision and Fusion Laboratory (IES), Haid-und-Neu-Str. 7, 76131 Karlsruhe, Germany

Abstract This study presents a novel framework for leaf segmentation, integrating Grounded SAM with a lightweight CNN to enhance automated plant phenotyping. By leveraging prompt-derived bounding boxes, Grounded SAM provides initial segmentation, which is refined by a CNN trained on minimal data to isolate leaf segments. The method was validated on potato and CVPPP datasets, demonstrating improved recall and precision over the baseline. This approach reduces reliance on extensive training data, offering a scalable solution for diverse plant phenotyping applications.

Keywords Phenotyping, automated segmentation, CVPP , segment anything

1 Introduction

Phenotyping is the characterization of plant traits that result from genomic and environmental influences; these traits collectively form the phenome [1]. Accurate phenotyping is crucial in crop breeding for selecting desirable traits to enhance crop yield, improve disease resistance, or increase adaptability to environmental stresses. It also plays a

vital role in monitoring crop health and development [2]. Traditionally, phenotyping has relied on visual observation combined with destructive analysis methods such as picking and weighing. To address the inefficiencies of traditional phenotyping, methods based solely on image analysis are becoming increasingly popular [1]. However, image-based phenotyping faces challenges due to the complexity of segmenting images to measure changes in leaf morphology over a plant's lifetime, which is influenced by the dynamic nature of leaf appearance throughout growth [3], compounded by the diversity of plant species and leaf geometries that complicate the development of universally applicable segmentation methods for leaf identification across various species and growth stages [3]. In recent years, multiple computer vision (cv) automated approaches have been developed to tackle this problem. Nevertheless, these models often require extensive training data to achieve high accuracy [4].

This paper extends the approach of Williams et al. [5] by utilizing Grounded Sam (GSAM) [6] (explained in detail below) to segment individual leaves based on prompt-derived bounding boxes rather than segmenting the entire image. We then employ a lightweight CNN to filter out only the leaf segments, using only few images per plant type and growth stage for training. This approach transforms the annotation process from a segmentation task into a classification task, making it faster and easier, and is schematically illustrated in Figure 1. The methods are tested and validated on the Potato Leaf Dataset (PLD) (Solanum tuberosum) created by [5], and to demonstrate transferability, they are also applied to the Computer Vi- sion Problems in Plant Phenotyping (CVPPP) datasets [7] with the plants Arabidopsis (Arabidopsis thaliana) and tobacco (Nicotiana tabacum). We applied the methods to 936 images and find that our methods outperforms the baseline on boths datasets with an Average Recall at an Intersection over Union (IoU) threshold of 0.75 (AR_{75}) of 63 and an Average Precision at an IoU threshold of 0.75 (AP_{75}) of 64.

1.1 Related work

Recently, the segment anything (SAM) framework [8] was introduced as a groundbreaking foundational model for image segmentation. The most notable feature of SAM is its impressive zero-shot segmentation

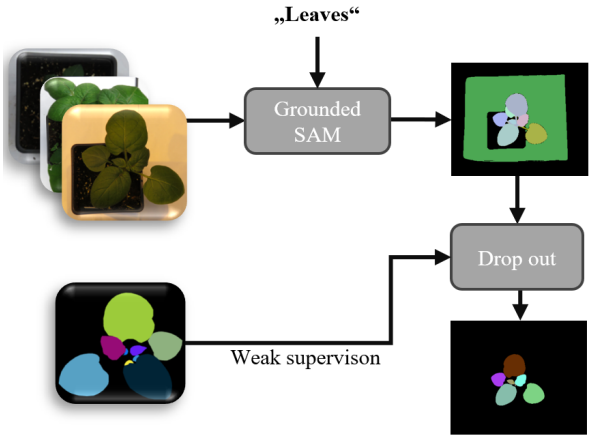


Figure 1: Overview of the process depicts the initial segmentation using GSAM, followed by a CNN-based dropout refinement. Plant and reference images adapted from [5]

performance on previously unseen datasets and tasks [9]. This capability is facilitated by various prompts, such as points and boxes. SAM has demonstrated remarkable versatility and effectiveness in handling a wide array of segmentation tasks [10] and has been successfully implemented in various fields. Including but not limited to: medical segmentation [11] [10], material science [12], remote sensing [13] [14].

Williams et al. [5] proposed a method where a plant image is initially segmented using SAM, followed by the application of rule-based geometric and color filters to isolate leaf segments. These filters include green color detection, exclusion of the entire plant, circle fitting to exclude non-leaf objects, and IOU exclusion with a set threshold. Duplicate masks are subsequently removed. The authors validated their method using a dataset of potato plant images, consisting of top-down views across four growth stages, with threshold parameters manually optimized for this dataset. Zhang et al. [15] employed SAM for semantic fruit segmentation, distinguishing fruits from other segments based on segment area, and filtered overlapping masks by their IoU. Balasundaram et al. [16] used the prompt ”plant” with SAM to segment plants and estimate their health via their Normalized Difference

Vegetation Index values. Additionally, [17] utilized GSAM to segment brown spots on coffee leaves.

2 Material and Methods

Using GSAM, plant images are segmented based on prompt-derived bounding boxes, resulting in a collection of annotation masks. GSAM integrates the Grounding DINO model as an open-set object detector with SAM to enable precise segmentation based on arbitrary text prompts. Initially, Grounding DINO [18] generates bounding boxes for objects or regions within an image using textual information as a condition. These bounding boxes serve as prompts for SAM, which then produces detailed mask annotations, allowing for accurate segmentation of specified regions within diverse visual tasks [6]. Since the segmentation goal of this study is leaf detection, the prompt we used is ”leaf.”

To eliminate incorrectly segmented leaves, we employ a binary decision process for each mask, utilizing a CNN classifier to determine whether it represents the fully visible portion of a leaf. This classifier is trained only on a few representative instances from each dataset.

All masks are resized to match the input dimensions of the trained CNN. The regions segmented by GSAM are then filled with the original pixels from the input image, while the rest of the image is filled with black pixels (0,0,0) in the RGB color space.

Training

The resulting images, created as described above, are then annotated as either *leaf* or *no leaf* and used to fine-tune a pre-trained CNN.

To enhance the robustness of the model, the training data will be augmented through various techniques such as rotation, mirroring, brightness adjustment, and cropping.

In this case, the ResNet10t [19] network is chosen for its ability to generalize well with minimal data. Its lightweight structure enables it to efficiently learn important features and patterns from the training data, even when the dataset is relatively small.

2.1 Experiments

Datasets

For the experiments, two datasets were utilized:

The PLD: Created by [5], comprised of 126 top-down images of potato plants (*Solanum tuberosum*) at various growth stages. The RGB images are of resolution 4608 x 3456 pixels, with reference data provided in a JavaScript Object Notation (JSON) file that characterizes each leaf as a geometric shape.

The CVPPP dataset [7]: This dataset contains images of Arabidopsis (*Arabidopsis thaliana*) and tobacco (*Nicotiana tabacum*) plants, which are frequently employed in plant phenotyping research and were created by [20]. The reference data is presented as a multicolored PNG mask. The CVPPP dataset is organized into four subsets: A1 (128 images), A2 (31 images), A3 (27 images), and A4 (624 images). A1, A2, A4 contain images of Arabidopsis (Arabidopsis thaliana) plants, while A3 consists of images of tobacco (Nicotiana tabacum) plants and P1 of potato (Solanum tuberosum) plants.

Representative examples of the dataset can be seen in Figure 2. It is worth noting that the quality of annotations in the CVPPP dataset is higher, as the PLD used polygons for annotation, which can lead to errors along the edges of the plants.

Evaluation Metrics

Evaluating the accuracy of multi-object segmentation presents challenges. This study employs the evaluation method from the Common Objects in Context (COCO) dataset [21] to ensure robust and standardized assessment. The process involves comparing the IOU of each predicted mask with reference masks, acknowledging that pixel-wise annotations are inherently imperfect and subject to human error [22]. For each segment predicted as a leaf, the IoU between the predicted and reference masks is calculated. An annotation is deemed a true positive (TP) if the IoU surpasses a set threshold. When multiple predicted masks have an IoU greater than zero with the same reference segment, the mask with the highest IoU is designated as the TP, while others are classified as false positive (FP). Reference masks that do not sufficiently overlap with any predicted masks are recorded as false negative (FP).

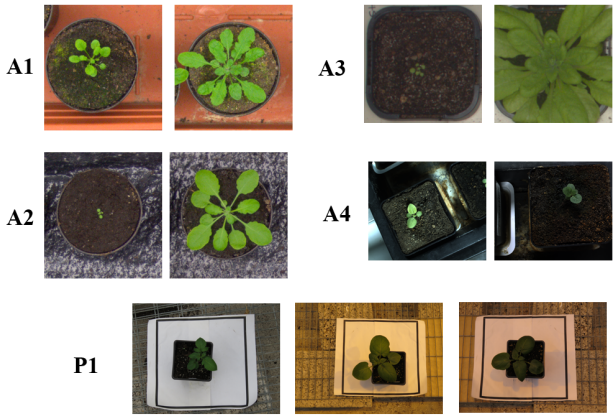


Figure 2: Example images from the datasets used in this study: The CVPPP dataset (A1-A4) and the PLD [5] dataset.

The threshold for determining whether an IOU is considered a match is set at 0.75 for AR_{75} and AP_{75} metrics. For generalized AR and AP the metrics are calculated at each IoU threshold in the set $T \in [0.5 : 0.05 : 0.95]$, then the average AR and AP are calculated across these varying thresholds.

Experimental Setup

First, a baseline was established by applying the original Leaf only SAM (LOS) method [5] to both datasets. Due to the high resolution of the potato dataset images (4608 x 3456 pixels), the resolution was reduced to 10% to facilitate processing. To ensure this reduction did not adversely affect baseline performance, the original code was applied to a subset (10 images) of the dataset at full resolution, and metrics were calculated, providing benchmark scores comparable to [5]. Some deviations are expected due to the undisclosed data split. These metrics were then compared to those from the same 10 reduced-resolution images, revealing a slight degradation in AR and AP, as shown in Table 1. This is attributed to lower SAM performance and higher IoU penalties at reduced resolution. For processing speed, comparisons between

methods were conducted on these reduced-resolution images, with a slight improvement in absolute performance anticipated at full resolution. Nonetheless, comparisons at reduced resolution remain fair.

Table 1: Comparison of the benchmark LOS model AR_{75}, AP_{75}, AR, and AP values that we obtain for the 10 random images in full resolution, and reduced resolution. Also, the values from the original paper that slightly deviate, likely due to different validation images.

Condition	**AR_{75}**	**AP_{75}**	**AR**	**AP**
Williams et al. [5]	63	60	59	56
Full Resolution	68	56	63	53
Reduced Resolution	51	49	50	49

Next, we tested the proposed method against the baseline. To train the CNN classification seven images from each dataset were used to train a model.

The prompt utilized to identify the bounding boxes is ”leaf,” with thresholds of 0.1 for both text and box. This low threshold is selected due to the dropout nature of the process, where an excess of segmentations is beneficial, as they can be removed but not added.

3 Results and discussion

Results

The performance of the proposed method was evaluated on two datasets: the CVPPP dataset and the PLD. Exemplary results are shown in Figure 3, and the complete performance results are summarized in Tables 2 and 3. In Table 2 the performance metrics for the data set A4 alone are presented, alongside the results for all data we used from the CVPPP (A1-A4). A4 was chosen because it encompasses all growth stages of a plant, making it an exemplary case for the challenges of automated phenotyping. The metrics are defined as above. In Table 3, the results are compared on only the PLD, on which the parameters for the baseline model were manually tuned.

For the CVPPP dataset, the baseline method (LOS) [5] achieved an AR_{75} of 46, AP_{75} of 25, AR of 42, and AP of 23. Our proposed method

Table 2: Results of the proposed methods in comparison to only SAM, the baseline and only GSAM on the CVPPP [7] dataset. Shown are the individual results for A4 as well as the averaged results from A1-A4.

	A4				CVPPP			
Method	**AR_{75}**	**AP_{75}**	**AR**	**AP**	**AR_{75}**	**AP_{75}**	**AR**	**AP**
Base SAM	64	11	56	10	68	20	62	19
Baseline LOS [5]	25	9	22	8	46	25	42	23
Baseline GSAM	66	47	58	42	70	44	64	41
Ours (GSAM + CNN)	64	59	56	52	66	58	60	53

outperformed the baseline. It achieved an AR_{75} of 66, AP_{75} of 58, AR of 60, and AP of 53 for the averaged results from subsets A1 to A4.

Table 3: Results of the proposed methods in comparison to only SAM, the baseline and only GSAM on the PLD [5] dataset.

Model	**AR_{75}**	**AP_{75}**	**AR**	**AP**
Base SAM	58	10	54	9
Baseline LOS [5]	58	62	53	57
Baseline GSAM	63	48	59	46
Ours (GSAM + CNN)	59	69	55	65

The images in Figure 3 show that the method reliably removed incorrect masks. However, some false positive (FP) classifications are also visible, such as the submask in the example from PLD(a). Additionally, it is evident that, compared to the reference, the leaves in the image from A1 are recognized without stems which leads to a smaller IoU. This smaller IoU can contribute to a poorer performance on the AR_{75} metric.

3.1 Discussion

The proposed prompt and CNN-based method exhibits improvements over the baseline established by [5]. The segmentation provided by GSAM surpasses the baseline in terms of average recall (AR) and average precision (AP) while our CNN dropout further enhances precision by eliminating incorrect masks. Additionally, our approach permits the use of low thresholds for text and bounding box detection, as erroneous

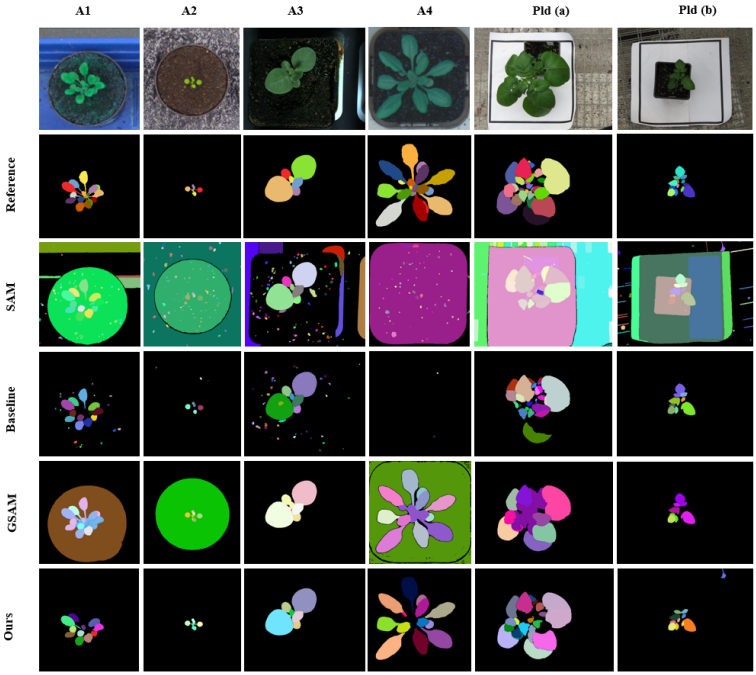


Figure 3: Segmentation results across different datasets and methods. A1-A4 are taken from the CVPPP dataset, Pld a and b from the PLD. For each example, the complete segmentation results from SAM, LOS [5], GSAM("Leaf"), and our proposed approach are shown.

segments are subsequently filtered out.

The annotation process for the CNN-based method is particularly swift, as it only requires binary classification rather than manual instance segmentation and mask creation, significantly reducing the time and effort needed for annotation. For instance, the data used to train the networks for the experiments were labeled in under 5 minutes.

The baseline method's performance on the A4 dataset was likely poor because the green filtering misclassified many leaves of the slightly differently colored Arabidopsis. Similarly, in the A1 dataset, the green filter likely caused moss-covered areas on the soil to be recognized as false positives (FP).

Due to the dropout pipeline of the method we note that the recall of this approach can only be as good as that of the initial segmentation given by the GSAM model. To determine where this upper bound (placed by GSAM) of the approach lies, we also calculated the metrics for the unmodified GSAM mask on both datasets. We find for the dataset A4 the AR_{75} for the total output of SAM was 66, so the with the AR_{75} of 64 from our approach, 97% of the maximum possible recall was reached while increasing the AP_{75} by 32%.

4 Conclusions and Outlook

The primary aim of this study was to create a robust framework for automated leaf segmentation with minimal dependence on training data or manual tuning, beneficial for plant phenotyping applications. Specifically, an automated procedure resilient to variations in plant species and growth stages was targeted. The challenges posed by diverse plant morphologies and leaf geometries can be mitigated by using GSAM for initial segmentation, followed by a sparsely trained lightweight CNN to isolate segments corresponding to plant leaves. For further improvement, the segmentation provided by GSAM should be carefully considered and optimized to ensure high absolute recall for leaves, with different prompts for Grounded DINO [18] tailored to specific plant phenotypes. Moreover, various parameters of SAM should be explored, along with the potential use of SAM2 [23] in the segmentation step. Additionally, future work could involve testing on more complex datasets, such as those featuring leaves from different perspectives or under varying environmental conditions, to enhance the robustness and applicability of the method.

References

1. D. Chen, K. Neumann, S. Friedel, B. Kilian, M. Chen, T. Altmann, and C. Klukas, "Dissecting the Phenotypic Components of Crop Plant Growth and Drought Responses Based on High-Throughput Image Analysis ," *The Plant Cell*, vol. 26, no. 12, pp. 4636–4655, 12 2014. [Online]. Available: https://doi.org/10.1105/tpc.114.129601

2. F. A. van Eeuwijk, D. Bustos-Korts, E. J. Millet, M. P. Boer, W. Kruijer, A. Thompson, M. Malosetti, H. Iwata, R. Quiroz, C. Kuppe, O. Muller, K. N. Blazakis, K. Yu, F. Tardieu, and S. C. Chapman, "Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding," *Plant Science*, vol. 282, pp. 23–39, 2019, the 4th International Plant Phenotyping Symposium. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168945217311548

3. G. R. S and V. Vasanthi, "Crop growth monitoring and leaf area index estimation using wireless sensor network and cnn," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2021, pp. 1031–1036.

4. J. Li, M. Xu, L. Xiang, D. Chen, W. Zhuang, X. Yin, and Z. Li, "Foundation models in smart agriculture: Basics, opportunities, and challenges," *Computers and Electronics in Agriculture*, vol. 222, p. 109032, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016816992400423X

5. D. Williams, F. MacFarlane, and A. Britten, "Leaf only sam: A segment anything pipeline for zero-shot automated leaf segmentation," 2023.

6. T. Ren, S. Liu, A. Zeng, J. Lin, K. Li, H. Cao, J. Chen, X. Huang, Y. Chen, F. Yan, Z. Zeng, H. Zhang, F. Li, J. Yang, H. Li, Q. Jiang, and L. Zhang, "Grounded sam: Assembling open-world models for diverse visual tasks," 2024.

7. M. Minervini, A. Fischbach, H. Scharr, and S. A. Tsaftaris, "Finely-grained annotated datasets for image-based plant phenotyping," *Pattern Recognition Letters*, vol. 81, pp. 80–89, Oct. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865515003645

8. A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollar, and R. Girshick, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 4015–4026.

9. P. Shi, J. Qiu, S. M. D. Abaxi, H. Wei, F. P.-W. Lo, and W. Yuan, "Generalist vision foundation models for medical imaging: A case study of segment anything model on zero-shot medical segmentation," *Diagnostics*, vol. 13, no. 11, 2023. [Online]. Available: https://www.mdpi.com/2075-4418/13/11/1947

10. Y. Huang, X. Yang, L. Liu, H. Zhou, A. Chang, X. Zhou, R. Chen, J. Yu, J. Chen, C. Chen, S. Liu, H. Chi, X. Hu, K. Yue, L. Li, V. Grau, D.-P. Fan, F. Dong, and D. Ni, "Segment anything model for medical images?"

Medical Image Analysis, vol. 92, p. 103061, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841523003213

11. J. Wu, W. Ji, Y. Liu, H. Fu, M. Xu, Y. Xu, and Y. Jin, "Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation," *arXiv e-prints*, p. arXiv:2304.12620, Apr. 2023.
12. K. Nichols, M. Hauwiller, N. Propes, S. Wu, S. Hernandez, and M. Kautzky, "Segment anything model for grain characterization in hard drive design," 2024. [Online]. Available: https://arxiv.org/abs/2408.12732
13. L. P. Osco, Q. Wu, E. L. de Lemos, W. N. Gonçalves, A. P. M. Ramos, J. Li, and J. Marcato, "The segment anything model (sam) for remote sensing applications: From zero to one shot," *International Journal of Applied Earth Observation and Geoinformation*, vol. 124, p. 103540, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569843223003643
14. D. Wang, J. Zhang, B. Du, M. Xu, L. Liu, D. Tao, and L. Zhang, "Samrs: Scaling-up remote sensing segmentation dataset with segment anything model," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36. Curran Associates, Inc., 2023, pp. 8815–8827. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/1be3843e534ee06d3a70c7f62b983b31-Paper-Datasets_and_Benchmarks.pdf
15. W. Zhang, L. M. Dang, L. Q. Nguyen, N. Alam, N. D. Bui, H. Y. Park, and H. Moon, "Adapting the segment anything model for plant recognition and automated phenotypic parameter measurement," *Horticulturae*, vol. 10, no. 4, 2024. [Online]. Available: https://www.mdpi.com/2311-7524/10/4/398
16. A. Balasundaram, A. Sharma, S. Kumaravelan, A. Shaik, and M. S. Kavitha, "An improved normalized difference vegetation index (ndvi) estimation using grounded dino and segment anything model for plant health classification," *IEEE Access*, vol. 12, pp. 75 907–75 919, 2024.
17. F. Mumuni and A. Mumuni, "Segment anything model for automated image data annotation: empirical studies using text prompts from grounding dino," 2024. [Online]. Available: https://arxiv.org/abs/2406.19057
18. S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu *et al.*, "Grounding dino: Marrying dino with grounded pre-training for open-set object detection," *arXiv preprint arXiv:2303.05499*, 2023.
19. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

20. H. Scharr, M. Minervini, A. Fischbach, and S. A. Tsaftaris, "Annotated image datasets of rosette plants," in *European conference on computer vision. Zürich, Suisse*, 2014, pp. 6–12.

21. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 740–755.

22. L. Zhang, R. Tanno, M.-C. Xu, C. Jin, J. Jacob, O. Cicarrelli, F. Barkhof, and D. Alexander, "Disentangling human error from ground truth in segmentation of medical images," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 15 750–15 762. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/b5d17ed2b502da15aa727af0d51508d6-Paper.pdf

23. N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer, "Sam 2: Segment anything in images and videos," *arXiv preprint arXiv:2408.00714*, 2024. [Online]. Available: https://arxiv.org/abs/2408.00714

Non-destructive and inline capable characterization of grapes by mid-infrared spectroscopy

Thorsten Tybussek[1], Christoph Wurm[1], and Jiaxin Xiang[1,2]

[1] Fraunhofer Institute of Process Engineering and Packaging IVV, 85354 Freising, Germany
[2] Technical University of Munich, Chair of Analytical Food Chemistry 80333 Munich, Germany

Abstract The microbial spoilage of grapes during ripening and after harvest poses a serious threat to quality. The suitability of mid-infrared (MIR) spectroscopy for the quantification of mold growth/load in grape products was assessed using grape juice inoculated with three different fungal species. Data from reference analytics and MIR spectra were combined in different PLS regression models resulting in a R^2 of 0.774 (NRMSE = 13.4%) for the prediction of fungal biomass and a R^2 of 0.652 (NRMSE = 17.9%) for the determination of the ergosterol content based on infra-red spectra. These results imply that MIR spectroscopy may be suitable for a rough assessment of mold contamination in grape juice. However, its application in monitoring fermentation processes as non-destructive and inline capable method appears significantly more meaningful.

Keywords Infrared spectroscopy, MIR, grape, mold, process control, ergosterol.

1 Introduction

Biological and environmental factors, such as grape variety and ripening stage, as well as climate and soil, significantly impact the chemical composition of wine grapes [1]. Additionally, mold spoilage presents a substantial concern during ripening and before processing. While

certain mold species, such as *Botrytis cinerea*, can be advantageous in specific contexts, infestation with filamentous fungi, such as *Aspergillus* spec. and *Penicillium* spec., can result in off-flavors and the potential formation of mycotoxins [2, 3]. Mid-infrared (MIR) spectroscopy is already used for the determination of the most substantial quality parameters in grape products, including sugar content, pH and total acidity [4–7]. The potential for detection of mold and mold varieties by MIR spectroscopy in grapes and grape products could be also demonstrated [8, 9]. However, the fungal infection of grapes or food in general is still predominantly detected by visual inspection or quantification of chemical marker substances. The most commonly used fungal biomarker is ergosterol, a hormone located in the cell wall of molds and yeasts. Due to the general specificity, this molecule is a good marker for fungal infestation and is already used for the determination of fungal biomass in grapes or fruit juices [10–13]. Nevertheless, would a rapid and non-destructive determination of ergosterol and therefore fungal biomass by mid-infrared spectroscopy allow a quick, labor-efficient and inline capable analysis of quality and safety of grapes. The scope of this work was to investigate the suitability of mid-infrared spectroscopy for the determination of fungal biomass in grape juice by applying a well-defined experimental setup to obtain spectral data as well as solid ground truth data. This dataset was subsequently used for regression analysis to predict the fungal biomass in inoculated grape juice based on mid-infrared spectroscopy.

2 Material & Methods

2.1 Preparation of Mold Spore Suspensions

Aspergillus carbonarius (DSM 1982), *Botrytis cinerea* (DSM 1282) and *Penicillium expansum* (DSM 877) were obtained from the DSMZ-German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). Spores of *A. carbonarius* were harvested with sterile sea sand after 10 days of incubation at 30 °C on yeast extract glucose agar plates supplemented with chloramphenicol. Spores were resuspended with Ringer's solution and ultrasonic bath treatment for 2 minutes. The supernatant was filtered using a fritted glass funnel, and the entire spore harvesting process was repeated two more times. The spore suspen-

sion was centrifuged and thrice washed with Ringer's solution. The resulting spore suspension was stored in a sterile glass bottle with glass beads at 4 °C. The spores of *B. cinerea* and *P. expansum* were harvested after 14 days of growth at 24 °C on potato dextrose agar plates. The conidia were released by the addition of Ringer's solution and gentle scraping with a sterile fork. The spore suspension was filtered using a fritted glass funnel and transferred to a sterile glass bottle with glass beads for storage at 4 °C.

2.2 Grape Juice Inoculation

Commercial pasteurized red grape juice was obtained from a local supermarket and originated from the same production batch. Aliquots of grape juice were aseptically inoculated with the previous prepared mold suspensions to achieve an initial spore count of approximately of 10×10^5 CFU/ml. The inoculated juice was transferred in 6 mL aliquots to individual sterile 24 mL glass vials and the lid was sealed with a 0.45 µm filter membrane to avoid further microbial contamination during storage. The prepared inoculated grape juice samples, along with the corresponding controls, were stored for up to 10 days in climate chambers at a controlled temperature of 10 °C, 16 °C or 20 °C and a relative humidity of 80%. During storage the inoculated samples were analyzed regularly. For each sampling three whole 24 mL glass vial were taken and analyzed individually.

2.3 Fungal Biomass Quantification

Fungal (dry) biomass was determined by sterile filtration of the grape juice samples and subsequent determination of the dry matter of the filter residue. Grape juice samples were filtered through a 0.45 µm filter membrane (Merck KGaA, Germany) using a Buchner funnel with slight vacuum pressure. The resulting filtrate was subsequently used for further analysis (FTIR, pH and brix). The filter containing fungal residue was rinsed with deionized water to remove remaining grape juice, and dried at 72 °C overnight. Fungal dry biomass corresponded to the weighed difference (0.1 mg precision) between the initially empty filter and the biomass loaded filter after drying.

2.4 Fourier-Transform Infrared Spectroscopy (FTIR)

Infrared spectra of the filtered grape juice samples were acquired using an ALPHA II FT-IR spectrometer equipped with a Platinum-ATR sampling module and OPUS 8.2.28 software (Bruker Optics GmbH & Co. KG, Germany). The filtered grape juice sample was applied evenly on the ATR sensor so that its entire surface was covered. Spectra were acquired between 400 cm^{-1} and 4,000 cm^{-1} with a resolution of 4 cm^{-1} and with 24 scans, each.

2.5 pH and Brix

pH and brix were determined from the recovered grape juice filtrate of the biomass determination step (see section 2.3). The pH was measured with a SevenCompact pH meter S220 (METTLER TOLEDO, Switzerland), while brix measurement was performed with a Brix-Acidity Meter Master Kit (ATAGO, Japan).

2.6 Ergosterol Content

The ergosterol content in the samples was measured using High-performance liquid chromatography (HPLC) based on the dried filters of section 2.3. The filter was placed in a reaction tube and 10 mL of a 10 g/mL methanolic potassium hydroxide solution (Th. Geyer GmbH Co. KG., Germany), as well as choleocalciferol (Merck KGaA, Germany) in n-hexane as internal standard was added. The mixture was blended with an ART-MICCRA D-8 blender for 1 minute. After additional shaking at 400 rpm for 30 minutes, the mixture was incubated in an oven at 80 °C for 1 hour. After cooling down, 1 ml of deionized water and 2 ml of n-hexane were added to the suspension and the tubes were shaken at 200 rpm for 2 minutes. The samples were allowed to rest until phase separation occurred. The upper phase was then transferred to an HPLC vial, and the solvent was completely evaporated using a gentle stream of nitrogen. The solid residue was dissolved in 1 ml of the mobile phase for HPLC, consisting of 80% methanol and 20% acetonitrile. A recovery sample consisting of ergosterol (Sigma-Aldrich, USA) was additionally prepared and processed alongside the biomass samples in parallel. HPLC analysis was performed using an Ultimate 3000 (Ther-

moScientific, USA) system equipped with a Nucleosil 120-3 C18 column (Macherey-Nagel, Germany) and a variable wavelength detector. Separation of peaks was achieved with an isocratic mobile phase consisting of 80% methanol and 20% acetonitrile (isocratic) at a flow rate of 1 mL/min and 20 °C. Ergosterol detection was conducted at 254 nm, whereas the wavelengths 210 nm, 290 nm and 363 nm were additionally used to verify the presence of analyte and internal standard, and the absence of interfering substances. HPLC operation, data acquisition and subsequent processing were done with Chromeleon 7 (ThermoScientific, USA).

3 Results & Discussion

Ergosterol content, brix and pH of the inoculated grape juice samples changed over a storage time of 7 days at 20 °C (Table 1). Significant differences in these parameters were also visible within the different mold species. The grape juice samples inoculated with *Aspergillus carbonarius* and *Penicillium expansum* showed the largest changes in brix value and pH, whereas inoculation with *Botrytis cinerea* resulted in slight alterations. Interestingly, the direction of these significant changes was dependent on the mold species. The juice samples inoculated with *A. carbonarius* showed a significant lower pH and brix value compared to *P. expansum* inoculated samples. As expected, the fungal biomass and ergosterol content increased in all species during storage. Species-dependent conversion factors for ergosterol content to biomass are published in literature [11] and were also observable in this study. While samples inoculated with *B. cinerea* displayed only slight non-significant changes in brix and pH over time, the significant increase in fungal biomass during storage demonstrated sufficient mold growth under the experimental conditions.

The reference analysis results were confirmed by the obtained FTIR spectra. The highest relevant infrared absorption of inoculated grape juice samples could be detected between 900 and 1500 cm^{-1} (Figure 1). Due to the fact that the water absorption is almost absent in this region, absorption effects can be mainly linked to dissolved substances [4]. The largest differences in the spectra were observed in the region between 1000 to 1100 cm^{-1}, which is typical for sugar absorption. The grape

Table 1: Measured chemical and physical parameters of the inoculated grape juice over the incubation period. Displayed is the mean across all mold species at day 0 (initial values) and for each species separately after 7 days of storage at 20 °C and 80% relative humidity. The displayed values correspond to the mean ± standard deviation. Different letters indicate significant different groups within rows ($p<0.05$, ANOVA and Tukey's range test).

Parameter	Initial Values	*A. carbonarius*	*P. expansum*	*B. cinerea*
Brix	16.3 ± 0.1^{a}	15.2 ± 0.3^{b}	14.9 ± 0.1^{c}	16.5 ± 0.1^{a}
pH	3.28 ± 0.02^{a}	2.39 ± 0.02^{b}	3.08 ± 0.05^{c}	3.26 ± 0.01^{a}
Biomass [mg]	9.00 ± 8.0^{a}	61.6 ± 3.9^{b}	43.8 ± 0.8^{c}	30.2 ± 1.9^{c}
Ergosterol [mg/L]	0.09 ± 0.1^{a}	46.0 ± 6.2^{b}	53.4 ± 4.9^{b}	11.5 ± 2.8^{c}

juices inoculated with *A. carbonarius* and *P. expansum* showed the highest decrease in absorption in this region indicating sugar catabolism. Additionally, a slight difference in the absorption patterns can be found between *A. carbonarius* and *P. expansum*. The first had its maximum of absorption at around 1064 cm^{-1}, while the samples inoculated with *P. expansum* exhibited a maximum at around 1042 cm^{-1}. These differences are likely caused by a different sugar composition. Glucose has its maximum absorption at around 1037 cm^{-1}, while fructose demonstrates at 1064 cm^{-1} its typical maximum [14]. This indicates that *P. expansum* favorably consumed fructose in contrary to *A. carbonarius*.

The obtained MIR spectroscopic data were used without further processing to predict the amount of mold in the juice samples during storage. Therefore, fungal biomass and ergosterol content were independently modeled using partial least squares (PLS) regression. Both target values were separately calculated using four factors and random sampling, where 66% of the data was used for training and 34% for testing purposes, with five repetitions each. Comparisons of the predicted target values with their respective ground truth revealed differences in both models, even though the applied steps during model building were the same (Figure 2). The PLS regression model for fungal biomass showed a R^2 of 0.774 and a NRMSE of 13.4%, whereas the PLS regression model for ergosterol content revealed a R^2 of 0.652 and a NRMSE of 17.9%. This result indicated a higher correlation between the spectral data and the fungal biomass compared to the ergosterol

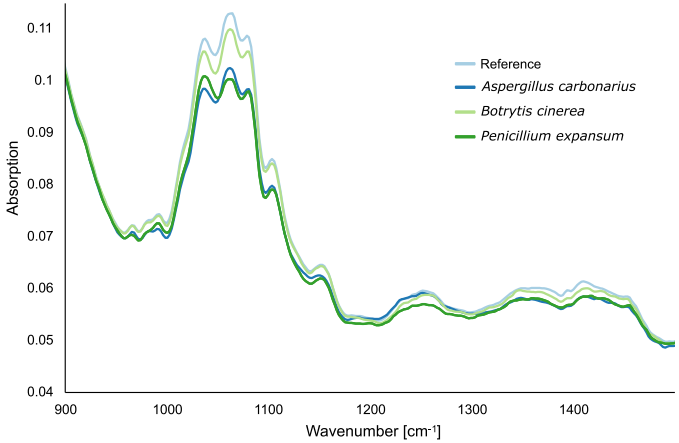


Figure 1: Excerpt of the measured mean raw infrared spectra of the inoculated grape juice samples and non-inoculated reference grape juice after 7 days of storage at 20 °C and 80 % relative humidity.

content. Regarding the quantitative estimation of mold spoilage in grape juice, ergosterol represents therefore not the ideal ground truth for reliable prediction levels. A certain factor contributing to this effect is the varying ergosterol content in fungal cell walls across different fungal species. This circumstance must be taken into account when converting ergosterol content to fungal biomass [11].

4 Conclusions & Outlook

The largest observed differences in the MIR spectra could be contributed to the individual sugar content and composition, which is correlated with mold growth. Thus, the sugar content is the most significant factor in estimating mold growth in the experiment. Since sugar content, as well as sugar composition and other constituents highly fluctuate in grapes by nature, an application of the calibration on other grape juices or even other fruit varieties is very unlikely. The two separate PLS regression models showed distinct results caused by the different target values. The PLS regression model for ergosterol content was less precise compared to the model for biomass prediction. In the-

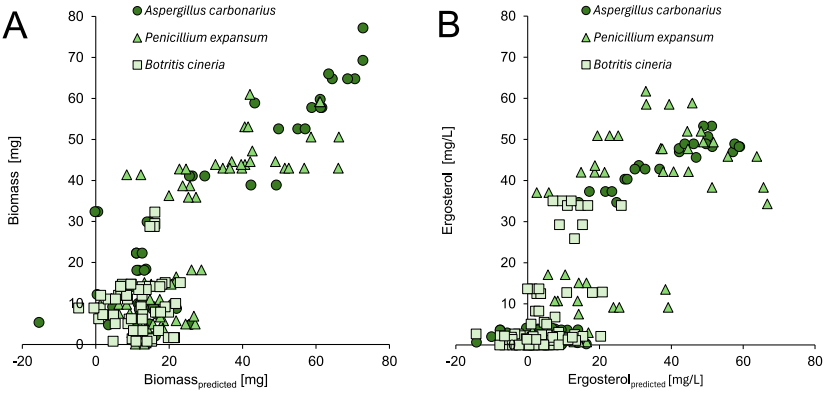


Figure 2: Partial least squares (PLS) regression models for (A) fungal biomass with R^2 of 0.774 (NRMSE = 13.4%) and (B) ergosterol content with R^2 of 0.652 (NRMSE = 17.9%) in the inoculated grape juice samples based on MIR spectra. The data shown consists of samples incubated at 10 °C, 16 °C or 20 °C and collected at 0, 2, 4, 7 and 10 days after start of the experiment. The obtained data were analyzed in Orange version 3.38.1. NRMSE was calculated as RMSE divided by X_{max}–X_{min}.

ory, it is possible to calculate the fungal biomass from ergosterol content using species-specific conversion factors [11,15] probably resulting in an increased precision of the PLS regression model for ergosterol. However, in practical application, this approach might not be effective because the present mold species are often not identified, or conversion factors are subject to necessary adjustments based on various intrinsic and extrinsic factors [16, 17]. Additionally, ergosterol is also synthesized by common yeasts such as *Saccharomyces cerevisiae*, which are widely present on grapes and other food [18]. The herein described experiments were performed under controlled and reproducible conditions, which are typical for fermentation processes. An application of MIR spectroscopy in e.g. the process control of fermentation processes seems to be very promising. The therein well-defined conditions allow the application as calibration or continuous measurements for further use in novelty or abnormality detection to ensure process stability.

Acknowledgments

This work was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program.

References

1. S. Blotevogel, E. Schreck, C. Laplanche, P. Besson, N. Saurin, S. Audry, J. Viers, and P. Oliva, "Soil chemistry and meteorological conditions influence the elemental profiles of west european wines," *Food Chemistry*, vol. 298, p. 125033, 2019.
2. J. E. Welke, "Fungal and mycotoxin problems in grape juice and wine industries," *Current Opinion in Food Science*, vol. 29, pp. 7–13, 2019.
3. S. Rousseaux, C. F. Diguta, F. Radoï-Matei, H. Alexandre, and M. Guilloux-Bénatier, "Non-botrytis grape-rotting fungi responsible for earthy and moldy off-flavors and mycotoxins," *Food Microbiology*, vol. 38, pp. 104–121, 2014.
4. M. Swanepoel, M. Du Toit, and H. H. Nieuwoudt, "Optimisation of the quantification of total soluble solids, ph and titratable acidity in south african grape must using fourier transform mid-infrared spectroscopy." *South African Journal of Enology and Viticulture*, vol. 28, no. 2, pp. 140–149, 2007.
5. R. Croce, C. Malegori, P. Oliveri, I. Medici, A. Cavaglioni, and C. Rossi, "Prediction of quality parameters in straw wine by means of ft-ir spectroscopy combined with multivariate data processing," *Food Chemistry*, vol. 305, p. 125512, 2020.
6. E. Mendes and N. Duarte, "Mid-infrared spectroscopy as a valuable tool to tackle food analysis: A literature review on coffee, dairies, honey, olive oil and wine," *Foods*, vol. 10, no. 2, 2021.
7. V. Thanasi, S. Catarino, and J. Ricardo-da Silva, "Fourier transform infrared spectroscopy in monitoring the wine production," *Ciência e Técnica Vitivinícola*, vol. 37, 2022.
8. A. Versari, G. P. Parpinello, A. U. Mattioli, and S. Galassi, "Determination of grape quality at harvest using fourier-transform mid-infrared spectroscopy and multivariate analysis," *American Journal of Enology and Viticulture*, vol. 59, no. 3, p. 317, 2008.

9. L. M. Schmidtke, L. J. Schwarz, C. Schueuermann, and C. C. Steel, "Discrimination of aspergillus spp., botrytis cinerea and penicillium expansum in grape berries by atr-ftir spectroscopy," *American Journal of Enology and Viticulture*, vol. 70, no. 1, p. 68, 2019.
10. M. Strydom, A. Kirschbaum, and A. Tromp, "Ergosterol concentration of several different saccharomyces cerevisiae yeast strains," *South African Journal of Enology & Viticulture*, vol. 3, no. 1, 1982.
11. J. U. Porep, R. Walter, A. Kortekamp, and R. Carle, "Ergosterol as an objective indicator for grape rot and fungal biomass in grapes," *Food Control*, vol. 37, pp. 77–84, 2014.
12. C. C. Steel, L. J. Schwarz, Y. Qiu, C. Schueuermann, J. W. Blackman, A. C. Clark, and L. M. Schmidtke, "Thresholds for botrytis bunch rot contamination of chardonnay grapes based on the measurement of the fungal sterol, ergosterol," *Australian Journal of Grape and Wine Research*, vol. 26, no. 1, pp. 79–89, 2020.
13. F. M. Vella, R. Calandrelli, A. Del Barone, M. Guida, and B. Laratta, "Rapid evaluation of ergosterol to detect yeast contamination in fruit juices," *European Food Research and Technology*, vol. 249, no. 2, pp. 465–472, 2023.
14. G. Sinnaeve, P. Dardenne, R. Agneessens, M. Lateur, and A. Hallet, "Quantitative analysis of raw apple juices using near infrared, fourier-transform near infrared and fourier-transform infrared instruments: A comparison of their analytical performances," *Journal of Near Infrared Spectroscopy*, vol. 5, no. 1, pp. 1–17, 1997.
15. H. Montgomery, C. Monreal, J. Young, and K. Seifert, "Determinination of soil fungal biomass from soil ergosterol analyses," *Soil Biology and Biochemistry*, vol. 32, no. 8, pp. 1207–1217, 2000.
16. M. O. Gessner and E. Chauvet, "Ergosterol-to-biomass conversion factors for aquatic hyphomycetes," *Applied and Environmental Microbiology*, vol. 59, no. 2, pp. 502–507, 1993.
17. J.-Y. Charcosset and E. Chauvet, "Effect of culture conditions on ergosterol as an indicator of biomass in the aquatic hyphomycetes," *Applied and Environmental Microbiology*, vol. 67, no. 5, pp. 2051–2055, 2001.
18. T. Jordá and S. Puig, "Regulation of ergosterol biosynthesis in saccharomyces cerevisiae," *Genes*, vol. 11, no. 7, 2020.

Use of VIS-NIR spectroscopy to monitor mango ripening quality with ripening indices in professional ripening processes

Sabine Wittmann[1], Simon Goisser[2], and Heike Mempel[1]

[1] Hochschule Weihenstephan-Triesdorf, Institute of Horticulture, Freising, Germany
[2] Fruchtimport vanWylick GmbH, Cologne, Germany

Abstract This study investigated the ripening process of Kent mangoes using various quality parameters and non-destructive spectroscopy techniques in professional ripening facilities. The research tracked changes in firmness, total soluble solids (TSS), titratable acidity, and color values over 8 days of ripening, revealing significant transformations during the climacteric phase with firmness decreasing from 99 to 35 N and TSS increasing from 13.6 to 16 °Brix. Two ripening indices, the Ripening Index (RPI) and Internal Quality Index (IQI), were applied and analyzed to quantify mango maturation in professional ripening processes. The results demonstrated minimal changes in the first three days ripening, followed by substantial quality shifts. Visible and Near-Infrared (VIS-NIR) spectroscopy models were developed to predict quality parameters, with the most successful correlations observed for color values like chroma (r^2 = 0.83) and hue (r^2 = 0.94). The study highlights the potential of non-destructive spectroscopic techniques to calculate ripening indices for monitoring mango ripeness while revealing challenges in developing robust prediction models across all quality parameters.

Keywords Mango ripening, VIS-NIR, ripening index, temperature sum, quality parameters

1 Introduction

Mangoes (Mangifera indica L.) are among the most popular tropical fruits, with increasing demand in European markets. As climacteric fruits, mangoes undergo a genetically regulated ripening process characterized by biochemical changes that influence acidity, soluble solids, and physical attributes such as firmness, weight, and color [1,2]. These transformations enhance aromatic compounds and nutritional qualities, which are critical for consumer acceptance and palatability. In Germany, the controlled ripening of mangoes in specialized facilities has been practiced for about 10 years, often involving a warming phase and ethylene treatments. Before this, mangoes imported by ship were marketed directly without controlled ripening. Defining the point of perfect mango ripeness remains challenging. Despite the importance of ripening control, current postharvest protocols lack standardization, and mango maturity assessments are typically subjective, based on small sample sizes and destructive methods such as visual appearance, firmness, or total soluble solids (TSS). Due to mango quality and ripening process variations, these measurements often cause resource inefficiencies and food loss [3]. Advances in non-destructive technologies offer promising alternatives for monitoring fruit quality throughout the supply chain. Several reports indicate visible and near-infrared spectroscopy as practical tools for predicting mango qualities like TSS, firmness, dry matter (DM), and flesh color. However, they rarely address controlled ripening practices commonly used in Germany. Another study used a portable VIS-NIR spectrometer to determine maturity parameters (soluble solids: r^2_{CV} = 0.87, $RMSE_{CV}$ = 1.39 %; dry matter: r^2_{CV} = 0.84, $RMSE_{CV}$ = 8.81 g kg^{-1}) [4]. A interval partial least square regression (iPLSR) was applied to improve mango firmness prediction (r^2_P = 0.75, $RMSE_P$ = 5.92 $Hz^2g^2/^3$) in portable NIR spectroscopy [3]. Nevertheless, models for predicting mango quality using NIR or VIS-NIR need to achieve high robustness. Since modeling a single parameter involves similar variability as a destructive reference measurement, it is challenging to make concrete statements about the ripeness of mangoes in the ripening room. It is important to combine relevant parameters into an index to obtain more reliable results. A Ripening Index (RPI) was successfully used to monitor postharvest ripening of mangoes, using firmness, titratable acidity, and TSS [5]. Another ripening

index, the Internal Quality Index (IQI), was further developed, integrating TSS, firmness, and color attributes of mangoes [6]. The report also used VIS-NIR to predict both RPI and IQI with good results. However, such models have not been adapted to typical industrial ripening conditions, where mangoes are ripened over 3–5 days at increasing temperatures up to 22 °C in specialized ripening rooms. The cumulative temperature sum provides insights into optimal mango storage conditions. However, its use has not been adopted as a parameter for defining ripening times in ripening facilities. This study evaluates whether RPI and IQI can serve as valuable ripening parameters for professional mango ripening. Furthermore, it seeks to determine if VIS-NIR models can be used for non-destructive measurement of ripening parameters, as described in the literature. Additionally, the study will use the temperature sum to assess its potential as a parameter for determining ripening times.

2 Material and methods

2.1 Sample material and storage conditions

In this study, Kent mangoes were used for data collection across five separate ripening cycles. The fruits were transported to Europe by refrigerated container ship for about three weeks. In the artificial ripening process, the fruits are typically warmed for 24 hours at 16 ± 1 °C / 61 ± 9 % rH and exposed to ethylene gas for 24 hours to initiate climacteric ripening. To counteract the climacteric phase, the fruits are then cooled down to about 12 ± 1 °C for 2-3 days still in the ripening room. This ensures that the fruits reach their ready-to-eat stage another 2-3 days later, on their arrival at the store or end consumer. Three of these cycles were conducted in rooms of a professional mango ripening facility (Axesso Ripening room, Interko Heat Exchangers and Ripening Solutions, Moerkapelle, Netherlands). After placing the mangos in the ripening rooms, four boxes were defined for further sampling, and a logger for temperature and humidity (Testo 184 H1, Testo SE & Co. KGaA, Titisee-Neustadt, Germany) was placed inside. The other two were carried out at constant room temperatures of 20 ± 1 °C and 54 ± 5 % RH.

2.2 Acquisition of reference data

Each day, measurements were taken on ten mangoes (ripening room) and five mangoes (room temperature) at three defined measurement points. The temperature sum was calculated by aggregating the average hourly temperature over the storage period for each ripening process. Color values were measured at three defined and marked points on each fruit using a colorimeter (PCE-CSM 2, PCE Instruments, Meschede, Germany). Firmness was assessed using a penetrometer (Fruit Texture Analyzer, GUESS Manufacturing Ltd., Cape Town, South Africa). Afterward, the skin was carefully removed from two measurement sites, and the spectrum, color values, and firmness were recorded again. A cylindrical cutter (approx. 25 mm diameter, 20 mm length) was used to take samples from each measurement point. One sample was dried in an oven at 30 °C for 24 hours to determine dry matter content (DM in %), with weight measured before and after drying. The second sample was analyzed for total soluble solids (TSS) and titratable acidity (TA). It was homogenized at 24,000 rpm using an Ultra-Turrax (IKA Ultra-Turrax T25, Janke & Kunkel), followed by centrifugation at 10,000 rpm for 5 minutes at 4 °C using a Heraeus Megafuge X1R Centrifuge (Thermo Scientific). The supernatant was used to determine TA with an automatic titration system (HI-902, Hanna® Instruments, Germany) using 0.25-N sodium hydroxide (NaOH). The remaining supernatant was analyzed for TSS using a handheld refractometer (Hi 96822, Hanna Instruments, Germany). Two indices, a ripening index (RPI) and an internal quality index (IQI), were calculated using Eqs. (1) and (2), where F is firmness (Newtons), TA is titratable acidity (%), TSS is total soluble solids (%), and L^*, h_{ab}, and C_{ab} are the color attributes of the flesh.

$$RPI = \ln(100 \cdot F \cdot TA \cdot TSS^{-1}) \tag{1}$$

$$IQI = \ln(100 \cdot F \cdot L^* \cdot h_{ab} \cdot TSS^{-1} \cdot C_{ab}^{-1}) \tag{2}$$

2.3 Recording of NIR spectra

Spectra of fruit with and without skin was recorded at pre-marked locations in contact with the fruit using the F-750 Produce Quality Meter (Firmware v.1.2.0 build 7041, Felix Instruments, Portland, USA). The

device records spectra for the wavelength range from 477–1059 nm. Spectral data from the F-750's SD-Card was transferred to a computer for subsequent analysis.

2.4 Statistical and chemometrical analysis

Spectra was loaded into the Model Builder Software (version 1.3.0.177, Felix Instruments, Portland, USA). The software uses non-linear iterative partial least squares (NIPALS) regression and a leave-one-out method for cross-validation. Spectra in second derivative form were used to build prediction and cross-validation models. The data processing of the reference values and the temperature and humidity loggers was carried out in Excel (Version 2021, Microsoft, Redmond).

3 Results and discussion

During the first three days, when the mango fruits were in the ripening rooms (Days 1-3), firmness decreased on average across all ripening variants from 99 to 93 N (Table 1). Total soluble solids (TSS) increased

Table 1: Mean and standard deviation of firmness, total soluble solids (TSS), titratable acidity (TA), dry mass (DM) and flesh/peel color parameters of mangoes at different days of storage.

Day of storage	Firmness (N)	TSS (°brix)	TA (%)	DM (%)
1	103.3 ± 12.4	13.3 ± 0.5	0.7 ± 0.0	14.6 ± 1.3
2	94.0 ± 8.2	14.1 ± 0.8	0.6 ± 0.1	14.5 ± 0.4
3	93.1 ± 5.6	14.0 ± 0.5	0.6 ± 0.0	13.9 ± 0.9
4	72.9 ± 12.0	15.0 ± 0.3	0.6 ± 0.0	14.8 ± 0.7
5	58.1 ± 9.1	14.8 ± 0.7	0.5 ± 0.1	14.4 ± 0.7
6	nA	nA	nA	nA
7	41.3 ± 9.7	15.9 ± 0.1	0.4 ± 0.0	13.8 ± 1.4
8	27.9 ± 4.9	15.9 ± 0.0	0.6 ± 0.0	nA

Day of storage	Hue (*)		Luminosity (L*)		Chroma (c*)	
	peel	flesh	peel	flesh	peel	flesh
1	97.4 ± 6.8	81.8 ± 2.7	47.4 ± 0.8	75.1 ± 2.0	31.3 ± 0.9	55.0 ± 4.9
2	96.5 ± 7.1	81.3 ± 0.9	48.5 ± 1.9	74.7 ± 1.6	32.3 ± 0.7	55.3 ± 0.5
3	98.2 ± 2.1	82.7 ± 1.3	47.2 ± 1.4	74.1 ± 1.3	32.0 ± 0.7	52.0 ± 2.7
4	89.9 ± 7.8	80.3 ± 0.9	46.8 ± 2.7	73.3 ± 0.6	33.9 ± 3.2	59.7 ± 2.3
5	94.1 ± 8.8	81.8 ± 1.6	50.2 ± 0.9	74.1 ± 2.1	36.0 ± 0.7	53.2 ± 4.4
6	nA	nA	nA	nA	nA	nA
7	100.2 ± 1.9	80.8 ± 0.1	48.5 ± 0.5	69.9 ± 1.5	33.9 ± 1.1	60.2 ± 1.0
8	89.2 ± 0.3	80.8 ± 0.0	52.7 ± 1.2	71.7 ± 0.9	38.0 ± 0.5	57.0 ± 0.0

slightly from 13.6 to 14.1 °Brix. Titratable acidity showed high variance, reducing from 0.7 to 0.6 on average across all variants. The color values for the peel and flesh remained stable for h (hue) and L (lightness), with 96-94 (peel), 81-82 (flesh), 47-47 (peel), and 75-74 (flesh), respectively. Color values such as c (chroma) for the flesh showed a minimal increase (31 to 32), while they decreased for the peel (57 to 52). The ripening process can explain the observed trends. In the initial phase (Phase: Initiation), the climacteric phase was initiated by warm temperatures and the addition of ethylene gas, leading to minimal changes in the parameters. These slight trends generally align with known quality changes in mango ripening literature [7]. In the second phase (Climacteric, Days 4-6), corresponding to the climacteric peak, more significant changes in firmness and TSS were observed. Firmness continuously decreased from 93 N (Day 3) to 35 N (Day 7), while TSS increased from 14 to 16 °Brix during the same period. This reduction aligns with several reports of softening of the texture due to enzymatic activity typical for the climacteric phase [1]. However, the extent of reduction depends on temperature conditions [8]. Titratable acidity (TA) also decreased, with high variability, from 0.6 % (Day 3) to 0.5 % (Day 7). The color values for h, L, and c remained practically constant with minimal changes. The assignment of ripening phases can also be confirmed through the changes in the calculated ripening indices RPI and IQI, as shown in Figure 1. A clear trend can be observed throughout the ripening process in both cases. However, both indices change minimally during the first phase (ripening room, Days 1-3). This initial phase, with an RPI of 6.1 (Day 1) to 5.9 (Day 3) and an IQI of 11.2 (Day 1) to 11.1 (Day 3), corresponds to unripe mangoes (RPI of 6 indicates unripe mangoes) [6, 9]. The minimal change in RPI during the first three days can be attributed to the time delay after applying heat and ethylene to initiate the ripening process. Both ripening indices significantly decrease once the mangoes are cooled in the ripening room and the climacteric phase begins. The RPI decreases from 5.8 (Day 3) to 4.3 (Day 7). This range between RPI 6-4 has been defined as intermediate-ripe mangoes and is associated with the best quality [6]. An RPI below 4 corresponds to an over-ripe mango. This value is not reached, indicating that the mangoes in commerce would have reached a ready-to-eat status (Days 7-8). The data clearly shows that the RPI can adequately and logically represent the changes and ripening of mango fruits. However,

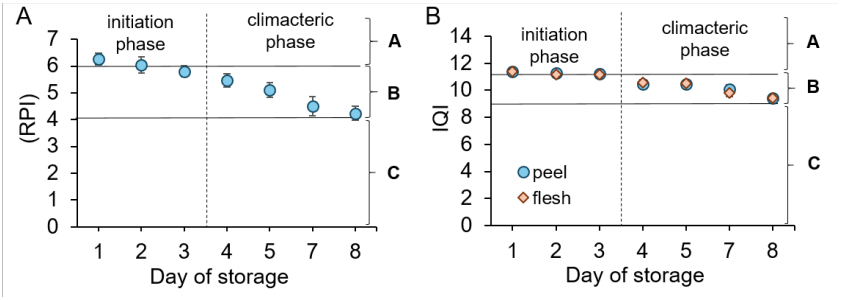


Figure 1: RPI and IQI of mangoes at different days of storage. Mangos ripening stages are separated by horizontal lines and marked as unripe (A), intermedium-ripe (B) and over-ripe (C).

the changes during the first three days of ripening are minimal, with no significant differences in RPI or IQI. It is easier to determine daily quality changes as soon as the climacteric phase begins. The results for predicting quality parameters during mango ripening are listed in Table 2. Models were developed for the VIS (477-729 nm) and VIS-NIR (729-1059 nm) ranges. Additionally, models were created based on NIR measurements on the skin and flesh. The correlations with the NIR spectrum on the skin were weak for all quality parameters, with firmness $r^2 = 0.32$ and TSS $r^2 = 0.41$. The correlations in the VIS range on the flesh showed higher accuracy, particularly for color values with chroma (c*, $r^2 = 0.83$), hue (h*, $r^2 = 0.94$), and lightness (L*, r^2=0.67), and were also suitable for firmness ($r^2 = 0.56$) and TSS ($r^2 = 0.45$). The VIS-NIR range improved the TSS correlation to $r^2 = 0.58$ as well as dry matter from $r^2 = 0.32$ (VIS) to $r^2 = 0.55$ (NIR). However, no sufficient correlations could be established for TA with any method. In contrast to the results in this experiment, other studies have established good models for the prediction of TSS in mangoes, with model accuracies ranging from $r^2 = 0.86$ [10] to $r^2 = 0.98$ [11]. The feasibility of predicting mango quality parameters was also demonstrated through the skin with validation models reaching $r^2 = 0.72$ for pulp firmness and $r^2 = 0.50$ for TA, which could not be proven in this study [12]. Another study achieved prediction models for the h° value through the mango peel with a maximum RMSE of 3.16°, which is significantly more accu-

rate than the results of this experiment [13]. However, the portable NIR spectrometers used in these cited studies apply wavelength ranges up to 1600 and 2300 nm and thus deviate significantly from the wavelength range of the F-750 device. A better prediction models with the F-750 was achieved in terms of dry matter in a study [14] than the results of this experiment ($r^2 = 0.84$). However, several mango varieties were used for the modeling, and different data pre-processing methods were carried out before the correlation. Based on our results, non-destructive measurement of the ripeness index (RPI) using VIS-NIR showed some limitations, as no model could be generated for TA. The possibility of capturing all necessary quality parameters for calculating IQI in one spectral measurement could present a significant time-saving advantage for assessing mango ripeness. While the destructive measurement of the spectrum on the flesh reduces some advantages of this method, the overall potential for non-destructive assessment remains promising. The firmness of the mangoes decreased linearly in relation to the

Table 2: Results of the NIR models for the prediction of fruit quality parameters in mango samples

	With Skin (477-729 nm)					Without Skin (477-729 nm)				
Parameter	n	Value range	PC	r^2_{CV}	RMSECV	n	Value Range	PC	r^2_{CV}	RMSECV
Firmness (kg)	160	2,7 - 14,8	6	0.32	2.63	532	0,3 - 14,8	6	0.56	2.42
Acidity (g/L)	270	2,7 - 12,4	3	0.01	1.97	270	2,7 - 12,4	3	0.16	1.77
Brix (%)	270	9,9 - 18,9	6	0.14	1.53	270	9,9 - 18,9	6	0.45	1.23
C*	540	15,1 - 51,1	5	0.29	4.61	540	19,4 - 81,5	5	0.83	3.93
h°	540	20,8 - 117,7	6	0.29	20.23	540	73,6 - 99,2	6	0.94	0.95
L*	540	29,6 - 60,6	5	0.17	4.88	540	59,1 - 83,9	5	0.67	2.21
DM (%)	195	9,3 - 19,4	5	0.11	1.73	195	9,3 - 19,4	5	0.32	1.49

	With Skin (729-1059 nm)					Without Skin (729-1059 nm)				
Parameter	n	Value range	PC	r^2_{CV}	RMSECV	n	Value Range	PC	r^2_{CV}	RMSECV
Firmness (kg)	160	2,7 - 14,8	6	0.3	2.68	532	0,3 - 14,8	6	0.45	2.7
Acidity (g/L)	270	2,7 - 12,4	3	0.01	1.94	270	2,7 - 12,4	3	0.09	1.85
Brix (%)	270	9,9 - 18,9	6	0.41	1.27	270	9,9 - 18,9	6	0.58	1.06
C*	540	15,1 - 51,1	5	0.12	5.16	540	19,4 - 81,5	5	0.34	7.65
h°	540	20,8 - 117,7	6	0.06	23.41	540	73,6 - 99,2	6	0.31	3.25
L*	540	29,6 - 60,6	5	0.09	5.15	540	59,1 - 83,9	5	0.46	2.84
DM (%)	195	9,3 - 19,4	5	0.42	1.39	195	9,3 - 19,4	5	0.55	1.21

temperature sum, consistently across all ripening variants (Figure 2). This demonstrates the dependency of firmness on temperature. Literature also describes this relationship, indicating that firmness decreases rapidly at higher storage temperatures. The trend for TSS shows a weaker correlation with temperature, with $r^2 = 0.57$. All other quality

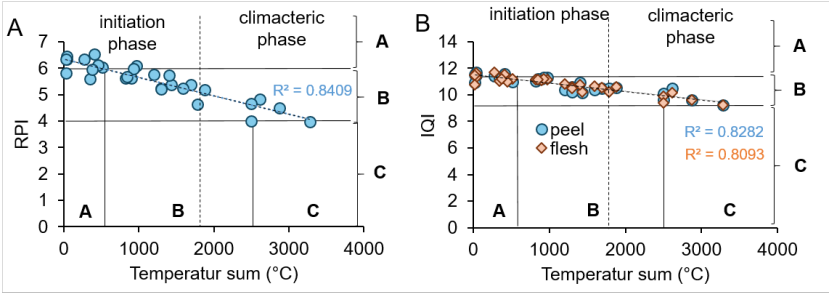


Figure 2: Mean of RPI and IQI of mangoes and the temperature sum. Mango's ripening stages are separated by horizontal lines and marked as unripe (A), intermedium-ripe (B) and over-ripe (C).

parameters exhibit weak correlations. Correspondingly, high correlations were found between the ripening indices RPI and IQI and the temperature sum (RPI, $r^2 = 0.84$ and IQI, $r^2 = 0.82$). Thus, an unripe mango, with an RPI of 6 during the ripening room phase (Phase 1), corresponds to a temperature sum of 0-970 °C. An intermediate-ripe mango (RPI 6-4), corresponding to Phase 2 in our study (cooling, delivery), falls within a temperature sum range of 970-2500 °C. The changes observed in the first three days of ripening indicate minimal variation in firmness and TSS. While this makes these indices challenging for quality determination during early ripening stages, it also suggests that this phase of minimal changes could be considered in future models or ripening concepts.

4 Summary

This study demonstrates that RPI and IQI are effective parameters for assessing mango ripening, with high correlations to the temperature sum. VIS measurements on the flesh show strong correlations for color values and moderate success for firmness and TSS. However, VIS-NIR models are not feasible for non-destructive measurement of TA. The temperature sum effectively characterizes the ripening stages, with explicit linear dependency observed for firmness. These findings suggest that while VIS-NIR technology enhances quality assessment, early-

stage ripening indices are challenging, emphasizing further refinement. Overall, the study highlights the potential of these methods to improve professional mango ripening practices.

Acknowledgement

The authors want to thank Fruchtimport vanWylick GmbH for providing mango fruit and a commercial ripening room for exotic fruit during the course of this study. We thank furthermore Lukas Süßmeier for the data collection. The work was supported by the QS-Wissenschaftsfond Obst, Gemüse, Kartoffeln.

References

1. B. Liu, Q. Xin, M. Zhang, J. Chen, Q. Lu, X. Zhou, X. Li, W. Zhang, W. Feng, H. Pei, and J. Sun, "Research progress on mango post-harvest ripening physiology and the regulatory technologies," *Foods*, vol. 12, no. 1, p. 173, 2022.
2. T. D. Le, T. Viet Nguyen, N. V. Muoi, H. T. Toan, N. M. Lan, and T. N. Pham, "Supply chain management of mango (mangifera indica l.) fruit: A review with a focus on product quality during postharvest," *Frontiers in Sustainable Food Systems*, vol. 5, p. 799431, 2022.
3. P. Mishra, E. Woltering, and N. El Harchioui, "Improved prediction of 'kent' mango firmness during ripening by near-infrared spectroscopy supported by interval partial least square regression," *Infrared Physics I& Technology*, vol. 110, p. 103459, 2020.
4. J. P. dos Santos Neto, M. W. D. de Assis, I. P. Casagrande, L. C. C. Júnior, and G. H. de Almeida Teixeira, "Determination of 'palmer' mango maturity indices using portable near infrared (vis-nir) spectrometer," *Postharvest Biology and Technology*, vol. 130, pp. 75–80, 2017.
5. A. L. V'asquez-Caicedo, P. Sruamsiri, R. Carle, and S. Neidhart, "Accumulation of all-trans-ß-carotene and its 9-cis and 13-cis stereoisomers during postharvest ripening of nine thai mango cultivars," *Journal of Agricultural and Food Chemistry*, vol. 53, no. 12, pp. 4827–4835, 2005.
6. V. Cortés, C. Ortiz, N. Aleixos, J. Blasco, S. Cubero, and P. Talens, "A new internal quality index for mango and its prediction by external visible and near-infrared reflection spectroscopy," *Postharvest Biology and Technology*, vol. 118, pp. 148–158, 2016.

7. S. Dea, J. K. Brecht, M. C. do Nascimento Nunes, and E. A. Baldwin, "Optimal ripeness stage for processing 'kent' mangoes into fresh-cut slices," *HortTechnology*, vol. 23, no. 1, pp. 12–23, 2013.

8. M. K. Assefa, B. A. Demessie, E. A. Shimelis, P. Verboven, M. Hertoga, and B. Nicolai, "Kinetic modeling of quality change in ethiopian kent mango stored under different temperature," in *Advances of Science and Technology: 7th EAI International Conference, ICAST 2019*. Springer International Publishing, 2020, pp. 71–81.

9. N. Vélez-Rivera, J. Blasco, J. Chanona-Pérez, G. Calderón-Domínguez, M. de Jesús Perea-Flores, I. Arzate-Vázquez, S. Cubero, and R. Farrera-Rebollo, "Computer vision system applied to classification of "manila" mangoes during ripening process," *Food and Bioprocess Technology*, vol. 7, pp. 1183–1194, 2014.

10. D. G. Abdullah Al-Sanabani, M. I. Solihin, L. P. Pui, W. Astuti, C. K. Ang, and L. W. Hong, "Development of non-destructive mango assessment using handheld spectroscopy and machine learning regression," *Journal of Physics: Conference Series*, vol. 1367, no. 1, p. 012030, 2019.

11. M. S. Amirul, R. Endut, C. B. M. Rashidi, S. A. Aljunid, N. Ali, M. H. Laili, A. R. Laili, and M. N. M. Ismail, "Estimation of harumanis (mangifera indica l.) sweetness using near-infrared (nir) spectroscopy," *IOP Conference Series: Materials Science and Engineering*, vol. 767, no. 1, p. 012070, 2020.

12. E. J. N. Marques, S. T. de Freitas, M. F. Pimentel, and C. Pasquini, "Rapid and non-destructive determination of quality parameters in the 'tommy atkins' mango using a novel handheld near infrared spectrometer," *Food Chemistry*, vol. 197, pp. 1207–1214, 2016.

13. T. Nordey, J. Joas, F. Davrieux, M. Chillet, and M. Léchaudel, "Robust nirs models for non-destructive prediction of mango internal quality," *Scientia Horticulturae*, vol. 216, pp. 51–57, 2017.

14. N. T. Anderson, K. B. Walsh, P. P. Subedi, and C. H. Hayes, "Achieving robustness across season, location and cultivar for a nirs model for intact mango fruit dry matter content," *Postharvest Biology and Technology*, vol. 168, p. 111202, 2020.

Optimization of plant experiments in CEA and vertical farming through standardization and modeling

Kay Plat

greenhub solutions GmbH,
Koburgerstr. 10, 04277 Leipzig

Abstract Vertical Farming and Controlled Environment Agriculture (CEA) enable the precise control of environmental factors such as light, temperature, and nutrients, allowing the study of complex growth processes. To ensure the comparability and reproducibility of experimental results, standardized methods and innovative modeling approaches are required. This paper highlights the challenges of standardization in CEA experiments, the integration of image and sensor data for analyzing growth parameters, and the use of resource-efficient technologies for optimizing agricultural systems. The presented approaches contribute to the development of reproducible and sustainable cultivation systems.

Keywords Controlled environment ggriculture (CEA), plant growth, standardization, modeling, resource efficiency, sensor data, image analysis

1 Introduction

The potential of Vertical Farming and Controlled Environment Agriculture (CEA) is transforming plant research. Through the precise control of environmental factors such as light, temperature, and nutrients, entirely new scientific questions can be explored. However, to obtain reliable and comparable results, standardized methods are essential. greenhub actively works on developing such standards and creating solutions that can be directly applied in practice. Standardization is

not only a tool for science but also a way to make technologies and processes more accessible for broader applications. The integration of optical and hyperspectral techniques significantly expands the possibilities for monitoring plant growth processes and offers new approaches for early stress detection. [1].

2 Challenges and Solutions

Diversity of Conditions: The challenges posed by varying growth conditions should not be underestimated. To ensure that research results are consistent and reproducible, uniform methods must be used [2]. Harmonizing growth parameters between different research institutions can also promote data exchange and facilitate global collaboration.

Technology as the Key: The technologies developed by greenhub enable precise monitoring and adjustment of growth conditions in real time. This ensures stabilized experimental conditions and reliable results [3]. Advances in sensor technology, automation, and vertical farming allow maximizing yield per unit of area and optimizing resource efficiency [4].

3 Modeling of Plant Growth and Its Application

Understanding and Prediction: Growth models are essential tools for understanding how plants respond to various environmental factors. These models are not only useful in research but are also applied in agriculture to optimize yields and improve resource efficiency [5]. Sensitivity analyses provide valuable insights into how environmental changes impact biochemical processes and help model and optimize these processes accordingly [6]. Image data is playing an increasingly important role in plant monitoring. Even non-visible changes within the plant can serve as early warning indicators, making camera systems indispensable in future plant research projects [7].

Implementation in Research and Industry : The models developed in research can find broad applications in agriculture and industry. They help farmers and companies refine their cultivation strategies and operate more sustainably. Image analysis techniques also allow for

non-destructive measurements of growth parameters such as weight or leaf size and enable the early detection of deficiencies [8]. Model predictive control (MPC) further enhances the dynamic regulation of environmental parameters, maximizing resource efficiency [9]. One major challenge in this field is the lack of valid growth data, which still requires considerable effort to generate. greenhub aims to provide an automated environment with the GreenResearcher, ensuring that these models and technologies are tailored to real-world requirements. Additionally, inline production measurement and plant condition detection using optical and sensor-based methods will be enabled. This will not only accelerate research but also significantly advance the development of new approaches.

4 Practical Application

Adaptation of Cultivation Systems: Findings from standardized experiments and model predictions should help optimize cultivation systems, particularly regarding light and nutrient management. Image data enables the detection of changes in plants before they become visible to the human eye, allowing for earlier interventions than before [10].

Resource Efficiency : These optimized systems contribute to more sustainable agricultural practices that offer both economic and ecological benefits. Spectral analysis can help make resource use more efficient in indoor production and experimental setups. Particularly in the use of light, fertilizers, and plant protection products, these analyses enable precise determination of requirements, reducing overall consumption [11]. At the same time, such approaches help reduce the environmental impact of modern agricultural systems.

5 Conclusion and Outlook

The Interaction of Standards and Technology: Standardized methods, combined with growth modeling and state-of-the-art technology, are key to advancing plant research. There is still significant untapped potential in the field of image analysis. In many cases, standardized procedures are still lacking. These approaches would not only provide

Figure 1: greenhub GreenResearcher, Vertical Farming System

scientific benefits but also offer concrete advantages for practical applications [7]. Greater integration of automation and artificial intelligence could further improve the efficiency and scalability of such systems in the future.

greenhub's Role: greenhub is playing a key role in driving these developments by creating solutions that advance research while also meeting the needs of the agricultural sector. The novelty of greenhub's approach lies in the combination of cutting-edge technologies such as inline measurements and automated detection of plant conditions using optical and sensor-based methods. This not only improves analysis quality but also significantly facilitates integration into real production environments and optimization processes. The goal is to bridge the gap between research and practice and to make innovations directly applicable.

References

1. S. Gorji and et al., "Applications of optical sensing and imaging spectroscopy in indoor farming: A systematic review," *Journal of Agricultural*

Research, 2024.

2. T. Kozai, *Smart Plant Factory: The Next Generation Indoor Vertical Farms*. Springer, 2018.
3. D. Touliatos, I. C. Dodd, and M. R. McAinsh, "Vertical farming increases lettuce yield per unit area compared to conventional horizontal hydroponics," *Food and Energy Security*, 2016.
4. S. H. Van Delden and et al., "Feeding the world: The future of vertical farming," *Trends in Plant Science*, 2021.
5. T. Li and et al., "Crop modeling for controlled environment agriculture," *Frontiers in Plant Science*, 2020.
6. S. Streif, R. Findeisen, S. Waldherr, and F. Allgöwer, "Steady state sensitivity analysis of biochemical reaction networks: A brief review and new methods," *Journal of Theoretical Biology*, 2009.
7. D. Singh and et al., "Imaging techniques for phenotyping of plant traits," *Plant Methods*, 2021.
8. M. Rahman and et al., "Application of ai in precision agriculture," *Sustainability*, 2021.
9. T. Padmanabha and et al., "Model predictive control for environmental optimization in controlled agriculture," *Environmental Control Journal*, 2020.
10. H. Poorter and et al., "A meta-analysis of plant responses to light intensity," *Plant Physiology*, 2019.
11. Y. Zhang and et al., "Spectral reflectance as a tool for resource efficiency in agriculture," *Agronomy*, 2021.

Optimizing near-infrared spectroscopy for on-line grape must quality assessment: Addressing the impact of suspended solids

Pascal Gauweiler[1], Xiaorong Zheng[2], Lucie Cornehl[2], Anna Kicherer[2], and Robin Gruna[1]

[1] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Fraunhoferstr. 1, 76131 Karlsruhe, Germany
[2] Julius Kühn-Institut (JKI), Federal Research Center for Cultivated Plants Institute for Grapevine Breeding, Geilweilerhof, 76833 Siebeldingen, Germany

Abstract What distinguishes good wine from premium wine is the quality of the berry material. The sugar and acid content are the key parameters in this case. To measure these, a time-consuming and costly labor analysis is usually required during delivery process to the cellar. Near-infrared (NIR) spectroscopy offers a faster alternative, but is also used in laboratories due to its size and integrability. The development of the miniaturization of NIR-spectrometer opens up new possibilities for integration. To explore this, such a spectrometer and a pump with particle filter were integrated into a harvester to measure the quality parameters online during harvesting. The filter size of the pump that delivers the grape must to the NIR-spectrometer for analysis is a critical parameter as suspended solids in the must affect accuracy. This study examined the filtration requirements to reduce this problem using 38 samples from four grape varieties, filtered in 9 steps from 1 mm to 40 µm and centrifuged as the final step considered as free of suspended solids resulting in 359 spectra. Chemometric analysis via PLSR showed that, for sugar prediction, a finer filter is needed, whereas a proper preprocessing effectively eliminates the influence of solids on acids.

Keywords NIR-Spectroscopy, agriculture, food quality, chemometrics, vine, grape must

1 Introduction

Sugar and acid content are key factors in assessing grape must quality, especially in wine cooperatives, where these metrics form the basis for payment. In the current process, the grape delivery is being analyzed in laboratories and is therefore time-consuming and costly. A reliable method for determining must constituents during the harvest would eliminate the need for laboratory analyses, facilitate cellar management, and enable early decisions to be made if quality is not up to standard. Near-infrared (NIR) spectroscopy has emerged as an efficient method to detect various ingredients in biological samples [1] [2].

In a previous study [3], a miniaturized NIR spectrometer was calibrated with samples from four varieties of vine. The transmission spectra of the respective must were recorded and a model using PLSR (patial least-square regression) was build. The idea is to integrate that spectrometer in a harvester to get the quality parameters such as glucose, fructose, malic acid and tartaric acid online during harvest. With that information, the cooperative or the winery cellar can skip the expensive and time-consuming laboratory analysis. During practical application, the filter device did not work well. The must contained suspended solids, which led to a decrease in intensity and thus to a poorer estimate of the spectral analysis.

In this study, the previous NIR spectrometer with its flow cell was used to examine the influence of suspended solids to determine the minimum requirement of the filter device for still good sugar and acid estimation. 38 samples of four grape varieties, two red and two white, were used, pressed and then filtered in 9 steps, started at 1 mm to 40 µm. The samples were centrifuged in the 10th step and considered free of suspended solids. Spectra were acquired during each step totaling 359 samples. The assumption is that the smaller the particle size in the sample, the better the performance of the model since the data become better as the intensity increases and the signal-to-noise ratio (SNR) improves.

2 Materials and Methods

2.1 Sample Collection

Two red ('Dornfelder' and 'Pinot Noir') and two white ('Chardonnay' and 'Riesling') wine grape varieties of *Vitis vinifera* (L.) were sampled weekly from onset of véraison (16th August 2023) up to harvest (27th September 2023) from four commercial vineyards located at the "Wollmesheimer Mütterle" site, near Landau in southern Rhineland-Palatinate, Germany (49°10'41.8"N 8°05'36.7"E). Each variety had four sampling time points. At each time point, 200 berries were randomly collected from each vineyard. At harvest, four additional samples were taken for each variety, except 'Pinot Noir', directly from harvesters.

The berry samples were divided into two subsamples of 100 berries each. One subsample was crushed using a blender (BL 6280, Grundig, Germany), while juice was extracted from the other using a manual juicer. Each subsample was further split into three parts: one was centrifuged at 20,340 × g for 10 minutes (Sigma 6K 15, Sigma Laborzentrifugen GmbH, Germany), another was used for the determination of the reference value and the third was subjected to filtering tests. For filtering, a series of filters with pore sizes ranging from 1,000 µm to 40 µm (pluriStrainer® Set, pluriSelect Life Science, Germany) were applied to non-centrifuged samples. To determine reference values for glucose, fructose, malic acid and tartaric acid, juice from a subsample was transferred to a 2 mL reaction tube, centrifuged at 12,100 × g for 5 minutes (Minispin, Eppendorf, Hamburg, Germany) and diluted 1:3 with double-distilled water. The samples were centrifuged again at 12,100 × g for 5 minutes before analysis.

A HPLC system (Agilent 1290 Infinity II, Agilent Technologies, Santa Clara, USA) equipped with a multisampler (G7167B), binary pump (G7120A), column oven (G7116B), diode array detector (DAD, G7117B), and refractive index detector (RID, G1362A) was used for analysis. A Rezex ROA-Organic Acid H+ ion exclusion column (300 × 7.8 mm, 8 µm) with a guard column (Carbo-H+, Phenomenex Inc., Torrance, USA) was maintained at 75 °C for separation. Samples (5 µL) were injected and analyzed in a 16.5-minute run using 0.4 mM sulfuric acid as the mobile phase at a flow rate of 0.6 mL/min. Malic and tartaric acids were quantified at 210.4 nm, while glucose and fructose were detected

with the RID set at 50 °C. A dilution series (1.5 – 90 g/L for glucose and fructose; 0.15 – 9 g/L for malic and tartaric acid) served as an external standard. Chromatograms were processed using Agilent OpenLab ChemStation software (Agilent Technologies, Santa Clara, USA). The distribution of the various compounds is shown in Tab. 1.

Table 1: Distribution of the sample content in g/L.

Content	Mean	Std. dev.	Min	Max
Glucose	73.88	26.97	9.9	115.13
Fructose	75.33	30.48	7.33	120.47
Malic Acid	7.89	7.77	1.88	27.43
Tartaric Acid	8.45	2.77	5.18	17.02

2.2 Spectral Data Collection

The samples were analyzed using the NIRONE Sensor S 1.4 spectrometer by SpectralEngines, Steinbach, Germany. It operates within a wavelength range of 1,100 nm to 1,350 nm, offering a resolution between 12 nm and 16 nm, and delivering a SNR of 15,000. The spectrometer uses a single-element InGaAs detector combined with a Fabry-Pérot interferometer for optical filtering.

A HL-2000-HP halogen lamp (Ocean Optics, Florida, USA) was used for illumination, which covers a broad spectral range from 360 nm to 2,400 nm. The samples were placed in a 1/4" flow cell (Avantes, Apeldoorn, Netherlands) equipped with an extension hose and a tap at the end to stabilize the sample during measurement. The optical path with a length of 5 mm is aligned orthogonally to the flow cell, which corresponds to the diameter of the flow cell. The samples are therefore measured in transmission. To connect the light source, flow cell and spectrometer, a step-index fiber optic cable (Thorlabs, New Jersey, USA) with a diameter of 365 µm and a numerical aperture of 0.22 was used (Fig. 1). This configuration ensured optimal utilization of the spectrometer's dynamic range while preventing signal overloading.

Before starting a new measurement series, a white and a dark reference was recorded. The measurement of the white reference μ_{white} was realized by using distilled water with a temperature of 20 °C . For measuring the dark reference μ_{dark}, the empty flow cell was recorded

with the light turned off. The spectra μ_i are balanced by the formular $\mu_{i,\text{bal}} = (\mu_i - \mu_{\text{dark}})/(\mu_{\text{white}} - \mu_{\text{dark}})$.

Each sample was filtered step by step from coarse to fine and measured in each case resulting in 342 spectra. 17 of the centrifuged samples were measured totaling 359 spectra. After every measurement, the flow cell was cleaned with distilled water. Due to hydrogen bonding, the temperature of the sample has an impact on spectral data [4]. Therefore, the sample temperature was kept constant at 20 °C. The mean spectra of each filter size is shown in Fig. 2.

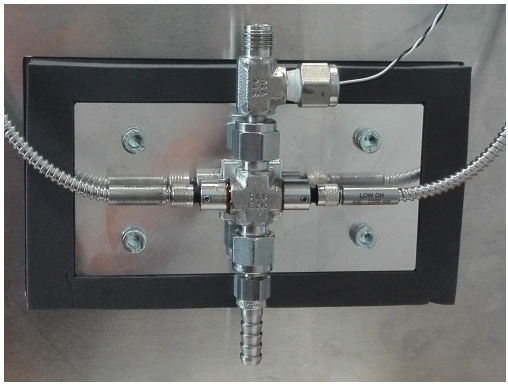


Figure 1: Flow cell with an optical path of 5 mm. The light shines through a fibre with a diameter of 365 µm orthogonal to the vertical flow. The must is inserted vertically.

2.3 Data treatment

To ensure clean data, an outlier removal step was performed beforehand. For each filter size, a principal component analysis (PCA) was applied. Data points that deviate from the mean of the PCA space by more than 1.96 standard deviations across the first four components were classified as outliers, capturing 95 % of the data within this range. Each filter size was analyzed independently, resulting in a different number of samples per filter size (see Tab. 2 and Fig. 3).

For the modeling process, all data were split according to their filter size. Data with one filter size were used for test, the remaining nine

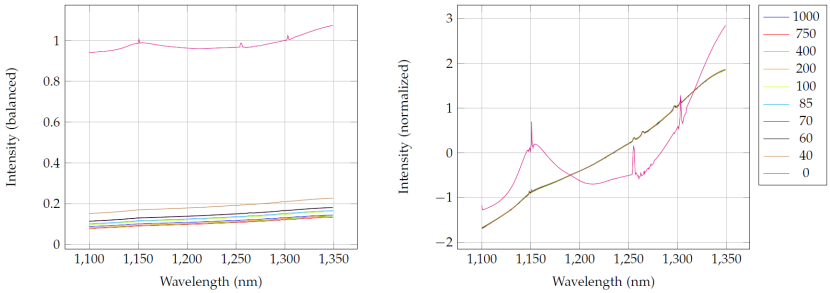


Figure 2: Raw (left) and normalized (right) average raw spectra, separated by filter size. The spectrum is subjected to artifacts due to the sensor used. Except for the centrifuged samples, the data are very similar so there is a large overlap in this graph. The filter size is correlated with the intensity.

were used for training and parameter optimization. The quality of the training data is considered to be consistent, as the data remains 88 % the same every run.

Table 2: Distribution of the samples for each filter size before and after outlier removal.

Filter size (µm)	centr.	40	60	70	85	100	200	400	750	1,000
Sample number	17(16)	38(34)	38(33)	38(33)	38(34)	38(31)	38(33)	38(32)	38(32)	38(33)

2.4 Preprocessing

To reduce effects such as scattering, noise and variation in intensities, all spectra ran through a preprocessing step to earn a proper estimation. The combination of normalization and smoothing with an optional derivation is recommended [5]. Normalization helps to reduce the impact of varying intensities caused by the sample's own absorption characteristics. There are multiple approaches to normalization, with two of the most widely used being standard normal variate (SNV) and multiplicative scatter correction (MSC) [5]. Both methods are common preprocessing techniques, and since their outcomes are generally comparable, the choice between them is not particularly critical. SNV works by a straightforward two-step process: subtracting the mean

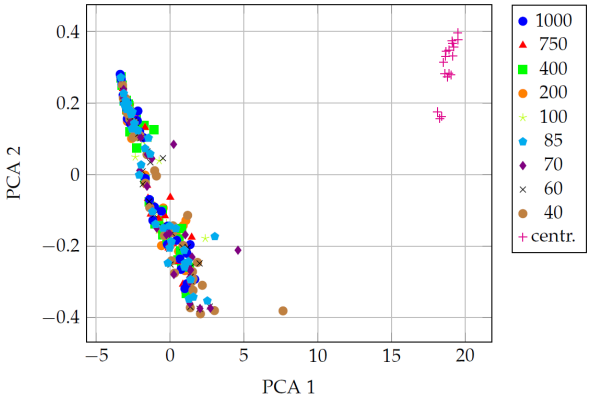


Figure 3: PCA Plot of the first two components grouped by filter size in µm after utlier removal. No clusters formed, except for the centrifuged samples.

value from the spectrum and then dividing the result by the standard deviation.

For noise reduction, filter algorithms like Savitzky-Golay (SavGol) are frequently applied [5] [6]. This method involves fitting a polynomial to a defined window of the spectra and replacing the center value with the result of the fit. As the algorithm moves through each point in the spectra, it reduces noise by filtering out high-frequency components that the polynomial cannot capture. A key benefit of this approach is that spectral peaks remain in their original positions and derivatives can be computed during the fitting process, which is useful for correcting baselines caused by scatter effects.

2.5 Modeling

Partial Least Squares Regression (PLSR) is the most widely used regression approach in chemometrics [7]. It is used as a method for dimensionality reduction that optimizes the covariance between the predictor and response variables. This technique enables the prediction of one dataset from another using a linear model. The key parameter of the performance of the model is the number of components. With a low number, the model tends to generalize poorly, while a high number

leads to overfitting.

With the parameters of SavGol filter and the number of components of PLSR, the question is which parameter settings works best for the data. Optuna, a framework for hyperparameter optimization [8], was used to determine the optimal parameters. The Tree-structured Parzen Estimator sampler [9] was chosen due to its advantages for hierarchically dependent parameters and its suitability for relatively small datasets ranging from 100 to 1,000 samples. The metric to measure the performance of the model is the root-mean-squared-error which was calculated as RMSE $= (\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2)^{1/2}$. During the parameter search, cross-validation was performed whereas the training data were split into five folds for every parameter set chosen by TPE algorithm. The data within each fold were randomly selected, ensuring that all five folds cover the entire training dataset. One fold was defined as the evaluation set, while the other four served as the calibration set. The model's performance was rated based on this division, and the process was repeated with each fold taking a turn as the evaluation set. After completing all iterations, the performance metrics from each fold were averaged to receive the final performance value of the respective parameter set. This approach provided a more representative evaluation by using all training data and minimizing the influence of favorable splits. The best performing parameter set was considered the final model.

The parameters that needed to be optimized included the window size, polynomial order, and derivative of the Savitzky-Golay filter, as well as the number of components of the PLSR model. Since SNV has no parameters, it did not require any optimization. The entire modeling process and parameter optimization was carried out using Python 3.9, with the use of SciKit-Learn 1.0.2 and Optuna 3.6.1.

3 Results and discussion

For each compound and filtration stage, an individual model was developed and optimized according to the previous described method. The resulting RMSE for each compound is shown in Fig. 4 divided into train and test. The RMSE of the training data of all four compounds is nearly independent of the filter size, which is consistent with the

Table 3: RMSE (g/L) of train and test data for different filter sizes (µm).

Filter	Train:				Test:			
	Glucose	Fructose	Malic Acid	Tartaric Acid	Glucose	Fructose	Malic Acid	Tartaric Acid
1,000	8.5	6.76	3.39	1.09	16.21	28.2	7.47	1.53
750	7.31	4.16	3.32	1.21	7.98	10.07	3.31	1.25
400	7.15	7.3	3.29	1.31	12.49	12.28	3.83	1.34
200	7.96	7.65	3.41	0.86	7.32	7.52	3.62	1.38
100	7.15	5.82	3.29	1.25	10.01	9.27	3.46	1.38
85	8.13	6.81	3.3	1.3	7.81	9.43	3.25	1.26
70	7.31	5.98	3.29	1.32	7.42	7.97	3.48	1.32
60	7.73	7.4	3.41	1.14	8.71	8.68	3.66	1.07
40	8.0	7.0	2.28	1.23	6.35	6.8	2.74	1.24
centr.	6.17	6.38	3.15	0.85	200.47	260.92	67.34	7.35

assumption made. For both sugars, the RMSE showed a decreasing trend with the particle size decreases; however, a typical converging behavior was not observed. The uneven trend of the sugar values can be explained by one sample that was not detected at the outlier detection stage and could not be well described by the model. The sample number is relatively small whereby RMSE metric is sensitive to outliers. Deleting this particular sample would lead to a smoother trend. The impact of this sample is not visible in the acid results.

The curves for sugars and acids are similar in terms of unevenness. Therefore, the information of each sugar and acid is also distributed in similar ways in the spectra. The RMSE at 1,000 µm is the highest value of all compounds and decreases rapidly with the next filtration stage. This could be the limit where the signal starts to drown in noise. Due to the particle size, the intensity of the spectra is decreased (see Fig. 2), resulting in a worse SNR. In general, the acid concentration is lower than the sugar (see Tab. 1) and should therefore be more affected by noise, but the performance at the different stages remains the same. The RMSE of the training set of tartaric acid by 200 µm filter size performs surprisingly well, which is usually an indicator of overfitting. However, the RMSE of the test set is integrated into the trend as expected.

The RMSE of the centrifuged samples are not included in Fig. 4 since

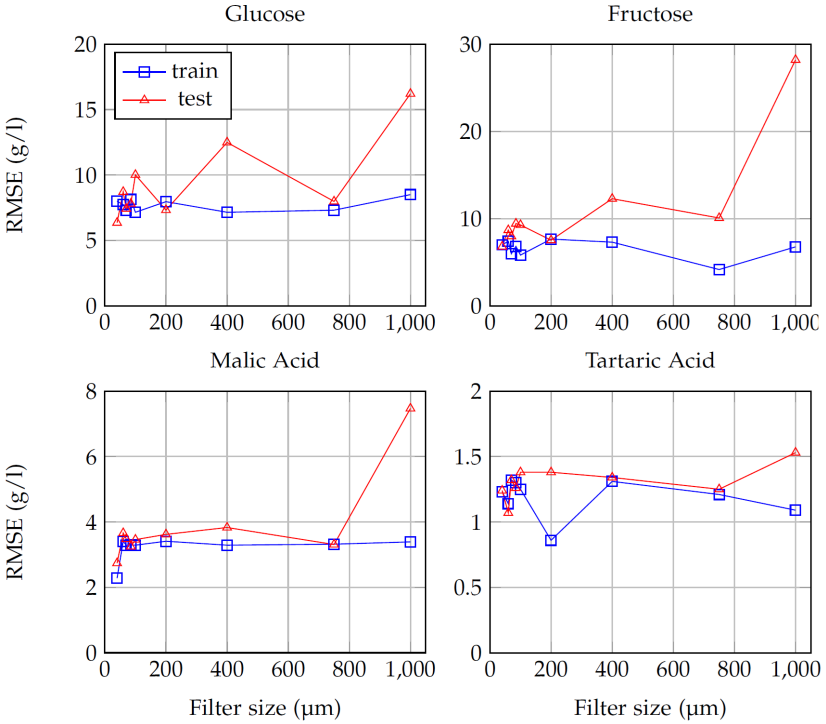


Figure 4: RMSE (g/L) of filter size 1000, 750, 400, 200, 100, 85, 70, 60 and 40 µm of glucose, fructose, malic acid and tartaric acid.

the prediction was too poor (Tab. 3). Fig. 2 and Fig. 3 show that the centrifuged spectra are very different from the rest. It is therefore not surprising if the prediction does not work as the model has never seen a centrifuged sample during training. Hence, the performance of the training set is as expected.

All contents performed the best at the finest filter size of 40 µm which is explainable as the SNR is the best in this area.

4 Conclusion

The 38 samples were filtrated by 10 filter steps starting from 1,000 µm to 40 µm and as the final step centrifuged, which is considered free of suspended solids. The resulting 359 spectral data were analyzed for each filter size case via PLS-Regression and basic preprocessing. For modeling, all data were used except those of one filter size. The optimal model parameters were determined using an optimization framework.

The resulting data show that for sugar, the use of a finer filter leads to better prediction accuracies. This can be attributed to the improved SNR as the signals intensity increases with each finer filter level. The prediction of acids is almost constant along the filter sizes, which is an indicator that proper preprocessing effectively eliminates the influence of solids in this case. Both contents are important parameters in practice, so the choice of filter size is based on the sugar performance which provides sufficient results at 100 µm. A filter size of 750 µm is recommended as the upper limit for the available data, as the signal is drowned out by noise with coarser filters.

References

1. V. Cortés, J. Blasco, N. Aleixos, S. Cubero, and P. Talens, "Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review," *Trends in Food Science & Technology*, vol. 85, pp. 138–148, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924224418304825
2. B. Vincent and P. Dardenne, *Application of NIR in Agriculture*. Singapore: Springer Singapore, 2021, pp. 331–345. [Online]. Available: https://doi.org/10.1007/978-981-15-8648-4_14
3. L. Cornehl, J. Krause, X. Zheng, P. Gauweiler, F. Schwander, R. Töpfer, R. Gruna, and A. Kicherer, "Determination of sugars and acids in grape must using miniaturized near-infrared spectroscopy," *Sensors*, vol. 23, no. 11, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/11/5287
4. V. H. Segtnan, v. Šašić, T. Isaksson, and Y. Ozaki, "Studies on the structure of water using two-dimensional near-infrared correlation spectroscopy and principal component analysis," *Analytical Chemistry*, vol. 73, no. 13, pp. 3153–3161, 2001, pMID: 11467567. [Online]. Available: https://doi.org/10.1021/ac010102n

5. Å. Rinnan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165993609001629
6. R. W. Schafer, "What is a savitzky-golay filter? [lecture notes]," *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 111–117, 2011.
7. S. Wold, M. Sjöström, and L. Eriksson, "Pls-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001, pLS Methods. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169743901001551
8. T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," *CoRR*, vol. abs/1907.10902, 2019. [Online]. Available: http://arxiv.org/abs/1907.10902
9. S. Watanabe, "Tree-structured parzen estimator: Understanding its algorithm components and their roles for better empirical performance," 2023. [Online]. Available: https://arxiv.org/abs/2304.11127

On the rapid aging of roasted coffee beans – A hyperspectral freshness analysis

Patrick Menz[1], Linda Beland[1,2], Joshua Moritz[1], and Andreas Herzog[1]

[1] Fraunhofer Institute of Factory Operation and Automation IFF,
Sandtorstr. 22, 39106 Magdeburg, Germany
[2] Otto-von-Guericke University Magdeburg,
Universitätspl. 2, 39106 Magdeburg, Germany

Abstract This work investigates the aging process of roasted Arabica and Robusta coffee beans using hyperspectral measurement technology. Measurements spanning the visual near-infrared to short-wave infrared ranges were conducted almost daily for four weeks. The spectral data analysis highlights significant changes in the 1900-2005 nm range, closely linked to chemical composition changes in coffee beans. Storage conditions impact aging, with the coffee canister preserving freshness better than refrigeration. The study identifies 1930 nm as a critical wavelength, emphasizing the importance of understanding spectral signatures in coffee quality assessment. The study's dataset provides a valuable resource for further research, offering insights into the spectral evolution of roasted coffee beans and its implications for quality control and flavor enhancement.

Keywords Hyperspectral analysis, coffee aging, coffee freshness, rapid coffee aging, NIR, hyperspectral imaging

1 Introduction

From predicting plant health to quality control in the food industry, spectral sensors combined with AI offer a wide range of potential applications. In the global food industry, coffee is an important issue. Spectral measurement technology is already being used to distinguish between different green coffee beans [1]. The storage of green coffee,

as well as the storage after roasting, plays an important role in the flavor [2,3]. In particular, important degassing processes take place after roasting which only perfect the flavor experience after a certain time [4] and define the quality of the coffee by its freshness [5]. This is about 14 days after roasting. Other studies in this area have looked at much longer timescales and different spectral ranges [4].

In this work, we aim to significantly reduce the time between spectral measurements and also want to measure in the ranges of visual near-infrared and short-wave infrared ranges. Measurements are therefore taken almost daily to monitor the degassing effect, and thus the aging after roasting. In a proof-of-concept study, we used hyperspectral measurement technology to investigate how quickly the spectrum changes during the degassing process of freshly roasted Arabica and Robusta coffee beans from different varieties under different storage conditions. For this purpose, spectra were acquired using a dual hyperspectral camera system in the range of 400-2500 nm under constant laboratory conditions. The measurements were taken 14 days after roasting and were carried out almost daily over a period of 4 weeks. Various analyzes of the spectral data were performed to determine how well the aging of the coffee beans was reflected in the spectra. In this study, the aging process is analyzed in terms of changes in the ingredients of the coffee bean that are visible in the reflectance spectrum.

2 Experimental design

2.1 Coffee samples and storage conditions

The most popular varieties of coffee are Arabica and Robusta. Arabica beans from Costa Rica and Robusta beans from Ecuador were selected for this experiment. After roasting, the beans were kept in an airtight package with a CO2 valve for 14 days. The package was then opened and prepared for the experiment. Petri dishes, each containing 30 g of coffee beans, were used as sample carriers. As coffee beans are stored in different ways in households, four common storage methods were selected to study the aging process under different conditions. These are:

(A) Coffee canister with CO2 valve in the laboratory at 22°C

(B) Petri dish with closed lid in the laboratory at 22°C

(C) Petri dish with closed lid in the refrigerator at 2°C

(D) Petri dish under constant active air flow in the laboratory at 22°C,

whereby the constant air flow was achieved by a fan directly oriented to the beans.

2.2 Measurement setup

The measurement trials were completed in a dark laboratory under constant environmental conditions. The hyperspectral measurement system used was the push-broom camera system from Norsk Elektro Optikk (NEO), consisting of Hyspex VNIR-1600 and Hyspex SWIR-320m-e, covering the spectral range from 400 nm to 2500 nm. A translation stage was used to continuously move the samples under the push-broom camera system to produce a 2D spectral image. Two 250 W line-producing halogen light sources were used to illuminate the samples. In addition, a graduated reflectance standard with 90%, 50%, 25%, 10% and 5% reflectance was used for subsequent calibration of the spectra.

Due to the different storage conditions, it was ensured that the coffee samples had acclimatized to the constant ambient temperature before each measurement. The camera system and lighting were also switched on at least 30 minutes before the start of the measurements to allow them to stabilize. Once constant conditions were achieved, a measurement was taken for each storage condition of the coffee beans, including the reflectance standard and the Arabica and Robusta coffees (see Figure 1).

2.3 Processing

After taking the spectral data, the data was at first processed using the software Hyspex Rad from NEO. The dark current was subtracted from the images to determine the true radiance. The images were then further processed using the in-house MATLAB based HawkSpex®Flow pipeline. To convert the radiance data into comparable reflectance data, a calibration was performed using the reflectance standard recorded in each measurement. Due to the different graded reflections of the

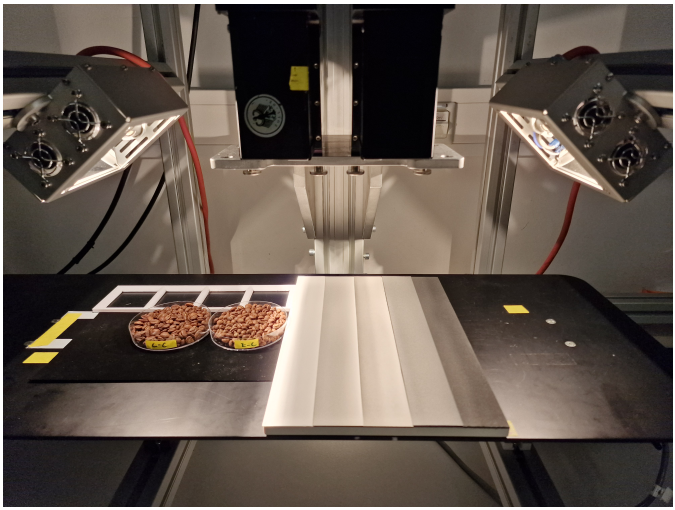


Figure 1: Laboratory measurement setup including two camera systems, graduated reflectance standard, halogen light source and coffee beans in Petri dishes.

standard, a multi-point calibration was used instead of the usual two-point calibration [6]. This is based on a least squares fit of a quadratic function and more accurately represents the true reflectance. From the labeled data (see Figure 2), 2000 samples per coffee type were randomly selected. In order to minimize noise, these samples were filtered with a third-order 1-D median filter.

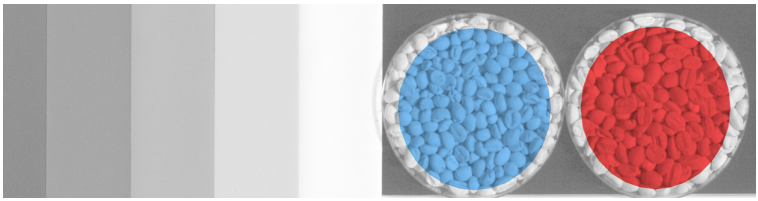


Figure 2: One spectral band of processed and labeled spectral measurement, where left is Robusta from Ecuador and right Arabica from Costa Rica.

2.4 Timeline

The measurement trial began on the day all the samples were filled into the respective storage containers, even before they were exposed to the different storage conditions. From this point, the samples were measured almost daily over a period of three weeks. No measurements were taken on days 2, 8, 9, 11, 15 and 16. In addition, a further measurement was taken on day 35 to record the deviation after approximately one month, as has been done in other studies over longer periods [4].

3 Experimental results

3.1 General observations

In order to analyze in general whether and how changes in the spectrum are related to the age of the coffee beans, this sub-section concentrates exclusively on storage condition (B). First, it is analyzed whether there are any differences at all in the spectra at the wavelengths chosen for this work. This is done by analyzing the reflectances over all the measurement days for both camera systems.

Due to the differences in height and orientation of the coffee beans within the Petri dish, it is useful to normalize the data before a valid statement can be made. An L1 normalization was used for this purpose (see Figure 3).

It turns out that the data from the VNIR camera system does not show any significant changes regarding the aging. Therefore, further analysis of the VNIR data is not carried out. The data in the SWIR wavelength range, on the other hand, show significant differences, particularly in the range from 1900-2005 nm. A trend can be observed that shows the aging process day by day with decreasing normalized reflectance. It can be seen that the changes within the first few days are big and become smaller over time. The strongest change in absorption is at 1930 nm, one of the known absorption bands of coffee, besides 1430 nm, 1730 nm, 1760 nm and 2250 nm [7]. In particular, the characteristic wavelengths of caffeine are at 1940-1960 nm [8].

The absorption bands of the second overtone of C=O and the first overtone of the C=O and O-H combination band are located in the wavelength range from 1900-2000 nm, where a superposition of the

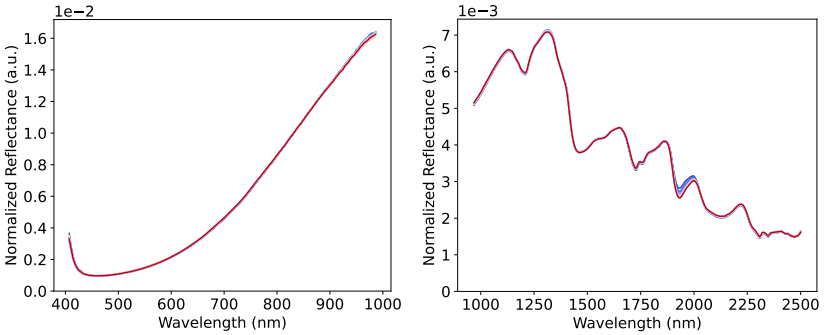


Figure 3: Left VNIR and right SWIR mean normalized spectra of all measurements from Arabica coffee beans. Color gradient of blue (day 1) to red (day 35) is used to differentiate between days.

main chemical components of coffee beans occurs [9]. These major components are carbohydrates, water, lipids, proteins (including amino acids), chlorogenic acid and caffeine, with carbohydrates being the most abundant and caffeine the least abundant. It is possible that the aging process has only changed the water content of the coffee beans. It is also possible that a combination of all the major components of the coffee bean have changed.

For a detailed analysis of the aging process, only the first and last day of measurement are shown in Figure 4. The standard deviation is also taken into account in order to visualize the variation of the recorded pixels of different coffee beans.

In the non-normalized state, the overlap of the standard deviation is large. This can be attributed to the fact that coffee beans were also measured in the lower layers due to the Petri dish. After applying L1 normalization, the standard deviation drops sharply and shows only a small overlap in the relevant wavelength range. Even from a purely visual observation, it is clear that a machine calibration model could work very well to distinguish between the first and last day.

So far, the focus has been on Arabica coffee beans, but Robusta shows similar behavior (see Figure 5).

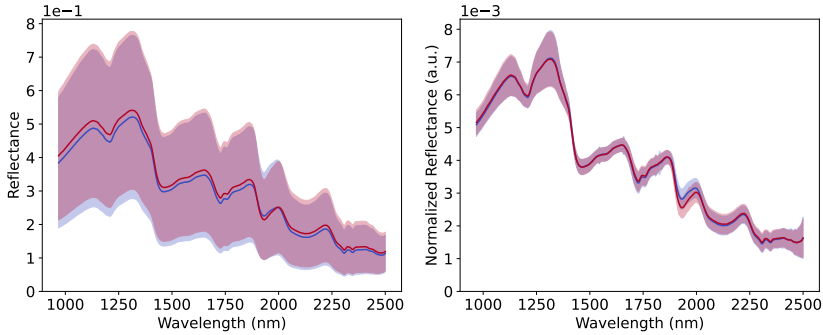


Figure 4: Left non-normalized and right normalized mean spectra (opaque lines) with standard deviation (transparent area) of measurement taken day 1 (blue) and day 35 (red) from Arabica coffee beans.

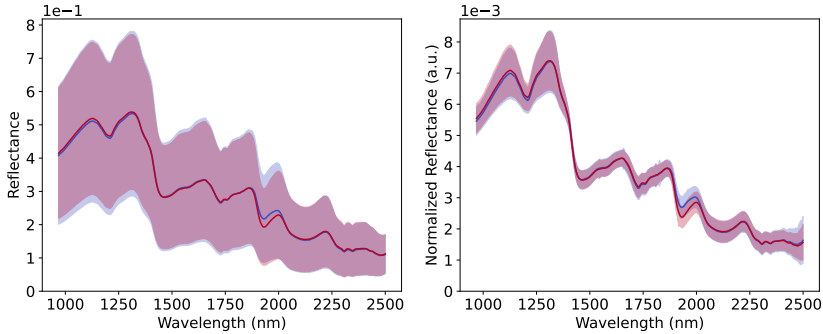


Figure 5: Left non-normalized and right normalized mean spectra (opaque lines) with standard deviation (transparent area) of measurement taken day 1 (blue) and day 35 (red) from Robusta coffee beans.

3.2 Storage condition observations

Since it has already been shown for storage condition (B) that the rapid aging process of the coffee beans is visible in the spectrum for both varieties, it is now investigated whether this process is dependent on the storage conditions described in section 2.1. For this purpose, storage condition (B) is taken as a reference and the relative difference to other storage conditions is analyzed. Figure 6 shows the difference spectrum

for Arabica as a relative difference.

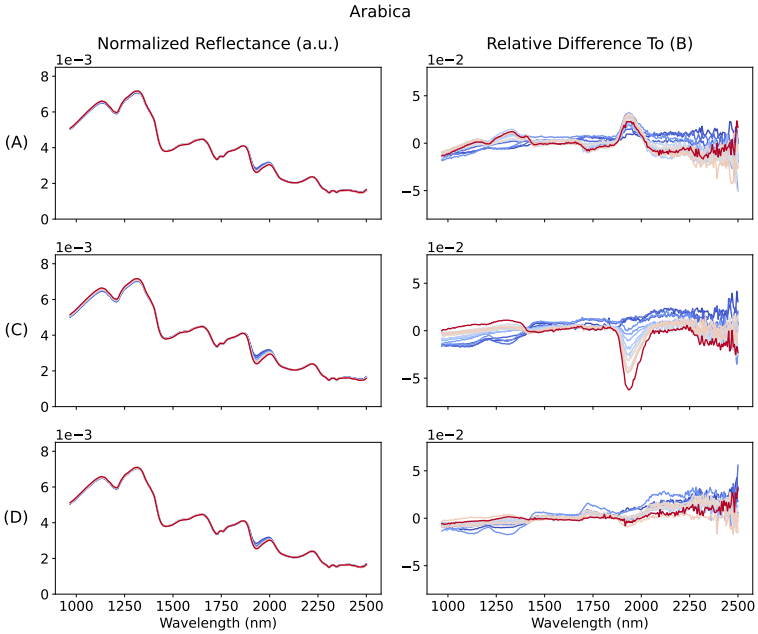


Figure 6: Left normalized mean spectra and right relative difference mean spectra to storage condition (B) from all other storage conditions in this work, including all measurement days. Color gradient of blue (day 1) to red (day 35) is used to differentiate between days.

It can be seen that storage in the coffee canister (A) has a positive effect on the freshness of the coffee beans. The changes in reflectance in the 1900-2005 nm wavelength range are significantly lower, which is confirmed by the positive difference. On the other hand, storage in the refrigerator (C) seems to accelerate the decrease in reflectance. Active air flow with the lid of the Petri dish open (D) shows no significant change in the difference. They have therefore spectrally aged almost identically to those in storage condition (B). Fluctuations from 2100 nm seem to be noise.

The same can be observed for Robusta. The relative differences are slightly different to those for Arabica, but show the same tendency.

4 Conclusion and further work

The results show a change in the spectrum over time, regardless of the type of coffee beans. This change is more prominent in the first few days and decreases after approximately two weeks, depending on the storage conditions. In this study, the coffee canister (A) seems to preserve freshness best. On the other hand, storage in the refrigerator (C) seems to accelerate the aging process. There was no significant difference between storage with active air flow (D) and without (B). It can therefore be concluded that the aeration did not lead to a faster changes in the main components of the coffee beans.

It was possible to identify 1930 nm as a significant wavelength, which is in a range where the major compounds of coffee beans overlap. A low-cost sensor that specifically monitors the aging process of the roasted coffee beans presented here is conceivable in a similar way to [10]. For this, only the wavelength range presented here needs to be covered and another to ensure standardization. It has already been shown in [11] that the degassing process depends on the roasting of the coffee. Therefore, future work should investigate whether the results can be reproduced for other coffees with different roasting conditions. Furthermore, a machine learning model that estimates the actual aging is conceivable. As the changes may not have a negative impact, laboratory analysis of the ingredients and tasting by coffee sommeliers would be useful.

References

1. P. Menz, A. Backhaus, and U. Seiffert, "Transfer learning for transferring machine-learning based models among hyperspectral sensors," in *ESANN 2019 proceedings*, 2019, pp. 589–594.
2. J. Błaszkiewicz, E. Nowakowska-Bogdan, K. Barabosz, R. Kulesza, E. Dresler, P. Woszczyński, Ł. Biłos, D. B. Matuszek, and K. Szkutnik, "Effect of green and roasted coffee storage conditions on selected characteristic quality parameters," *Scientific reports*, vol. 13, no. 1, p. 6447, 2023.
3. J. Barrera-López, A. F. González-Barrios, L. F. Vélez, L. F. Tarquino, H. López, and M. Hernandez-Carrión, "Evaluation of roasting and storage conditions as a strategy to improve the sensory characteristics and shelf

life of coffee," *Food science and technology international*, vol. 30, no. 3, pp. 207–217, 2024.

4. M. C. Lázaro, E. J. Ferreira, J. A. Gomes Neto, and E. C. Ferreira, "Characterization and predictive modeling potential of aging time of roasted coffee using infrared spectroscopy," *Analytical methods : advancing methods and applications*, vol. 14, no. 36, pp. 3486–3492, 2022.
5. C. Yeretzian, I. Blank, and Y. Wyser, "Protecting the flavors—freshness as a key to quality," in *The craft and science of coffee*, B. Folmer, Ed. Saint Louis: Elsevier Science, 2017, pp. 329–353.
6. J. Burger and P. Geladi, "Hyperspectral nir image regression part i: calibration and correction," *Journal of Chemometrics*, vol. 19, no. 5-7, pp. 355–363, 2005.
7. N. Caporaso, M. B. Whitworth, S. Grebby, and I. D. Fisk, "Non-destructive analysis of sucrose, caffeine and trigonelline on single green coffee beans by hyperspectral imaging," *Food research international (Ottawa, Ont.)*, vol. 106, pp. 193–203, 2018.
8. T. Li, X. Yu, W. Zu, Y. Cai, and W. Hong, "Review: Application and research progress of near infrared spectroscopy for multidimensional assessment of coffee beans," *Journal of Near Infrared Spectroscopy*, vol. 32, no. 6, pp. 179–191, 2024.
9. J. S. Ribeiro, M. M. C. Ferreira, and T. J. G. Salva, "Chemometric models for the quantitative descriptive sensory analysis of arabica coffee beverages using near infrared spectroscopy," *Talanta*, vol. 83, no. 5, pp. 1352–1358, 2011.
10. P. Menz, V. Vaquet, B. Hammer, and U. Seiffert, "From hyperspectral to multispectral sensing – from simulation to reality: A comprehensive approach for calibration model transfer," in *ESANN 2022 proceedings*, 2022, pp. 469–474.
11. X. Wang and L.-T. Lim, "Effect of roasting conditions on carbon dioxide degassing behavior in coffee," *Food Research International*, vol. 61, pp. 144–151, 2014.

Classification of overall sensory acceptability of modified-air packaged (MAP) minced pork stored at different temperatures using fluorescence spectroscopy

Johannes Schlosser[1], Siegfried Münch[2], Dagmar Brüggemann[2], and Heinar Schmidt[1]

[1] University of Bayreuth, Bioanalytical Sciences and Food Analysis, Fritz-Hornschuch-Str. 9, 95326, Kulmbach
[2] Max Rubner-Institute, Federal Research Institute of Nutrition and Food, Department of Safety and Quality of Meat, E.-C.-Baumann-Str. 20, 95326 Kulmbach

Abstract This study investigates whether fluorescence spectra can predict the overall sensory impression of minced MAP meat. Minced pork was packaged under 70% O_2 and 30% CO_2 and stored at constant 2, 10 and 14°C. Fluorescence measurements and sensory assessment were regularly conducted up to 17 days at 2°C storage and with a shorter periode for 10 and 14°C. A trained sensory panel graded the samples, individual grades were merged into an overall acceptance score from 1 to 9. Samples with a score >5 were considered inacceptable for consumption. Fluorescence spectra were measured on opened packages with a portable device at random positions. Averaged spectra per sample of 5 batches were used to calibrate partial least squares discriminant analysis models. About 100 samples were used for calibration and 60 independent samples for validation. Cross-validated two-class models had averaged accuracies of 91% (89%), sensitivities of 90% (87%), and specificities of 92% (91%) for "acceptable" (validation results). Hence, fluorescence spectra measured with a handheld fluorescence device have potential to classify MAP minced pork according to the overall sensory impression into "acceptable" and "not acceptable".

Keywords Sensory assessment, fluorescence spectra, spoilage, minced pork, MAP

1 Introduction

Minced meat is a perishable food, which has to be stored at 2°C to maintain its shelf life. [1] Microbial parameters have to be monitored to ensure good hygienic conditions. Among these, the determination of total viable counts (TVC) is important to assess the hygienic status of minced meat. [2] However, there are a number of drawbacks, which are limiting the application: The sampling is destructive. The cultivation-based method is labour and cost-intensive and results are only available after several days. In addition, some bacteria are known, which are viable, but not cultivable and, hence, invisible for TVC [3].

Various spectroscopic rapid methods have been investigated to determine the TVC in meat including mid infrared, Fourier transform infrared (FTIR), VIS/near-infrared, multispectral image acquisition, fluorescence and Raman spectroscopy. All these methods showed correlations between the spectra and TVC when samples were stored during the experiment. Only few of the publications focused on minced meat or modified atmosphere packaging. Grimmler et al. reported correlations with an R^2cv of 0.83 and root mean squared error of cross validation (RMSECV) of 0.69 log(cfu/g) for aerobically stored minced pork and beef using fluorescence spectroscopy [4]. Fengou et al. investigated minced pork under aerobically conditions with FTIR, MSI and VIS. For the different methods, they reported root mean squared errors of prediction (RMSEP) from 0.915 to 1.173 and R^2 values from 0.727 to 0.834 [5]. Arias et al. obtained an R^2val of 0.694 in the prediction and an RMSEP of 1.037 log(cfu/g) using near-infrared spectroscopy with MAP pork fillets and measurement on the opened MAP trays [6]. Given prediction errors of 0.7 to 1.1 log units, the predictability of TVC is not very precise with $\pm$ 1.4 to 2.3 log(cfu/g) at a 95% confidence level which may hinder application of TVC as a threshold parameter to evaluate fitness for consumption. On the other hand, sensory analysis is well suited to describe the edibility of meat. Ramírez et al. investigated the sensory analysis of pork with an electronic nose. They found correlations between the results of their device and the sensory assessment [7]. However, sensory analyses are also destructive, cost and labour-intensive as they require a trained panel. Nevertheless, they are meaningful with regard to the perception of the consumer. A rapid method to determine the sensory properties of minced meat could be

interesting for different stakeholders. Therefore, the aim of this study was to investigate the feasibility to determine the overall acceptability of minced MAP pork using a handheld fluorescence device.

2 Materials and Methods

Minced meat was produced batch wise from pork (shoulder) and packaged in trays under modified atmosphere with 70% O_2 and 30% CO_2. The batches were stored at constant temperature until the samples were spoiled, in total three independent experiments for 2, 10 and 14°C, each. In general, samples were measured on day 1, 3, 4, 6, 8, 10, 13, 15 and 17 for 2°C storage and day 1, 2, 3, 6, 8 and 10 for experiments with a storage temperature of 10 or 14°C, see Table 1. One package was used for sensory analysis and three packages for fluorescence analysis for each measurement day. As there was no sensory analysis for the first 10°C experiment, the fluorescence spectra were not included in this study. In experiment C, measuring day 4 was skipped and in experiment H, the measurements were cancelled after day 6 due to spoilage of the samples (see Table 1).

Table 1: Overview of experiments, storage temperatures and number of samples.

Experiment	Storage temperature	Measurement days	Number of samples	
			Sensory analysis	Fluorescence
A	2°C	9	9	26
B	2°C	9	9	27
C	2°C	8	8	24
D	10°C	6	6	18
E	10°C	6	6	18
F	14°C	6	6	17
G	14°C	6	6	18
H	14°C	4	4	12

For the sensory analysis, a trained sensory panel with 4 - 9 panellists graded the sample according to appearance and smell per sampling day. The individual assessments were discussed and merged into a

consent report including an overall acceptability with scores ranging from 1 to 9, with 1 being impeccable and 9 totally spoiled. Samples with a score >5 were considered as inacceptable for consumption. Intermediate marks were reported in the consent report rounded to 0.5.

In parallel, fluorescence spectra were collected with a portable fluorescence device (BFD-100, FreshDetect GmbH, Germany), which excited fluorescence at 405 nm and recorded the spectra from 460 to 900 nm. Samples were opened and measured at 11 randomly-selected positions per package. The spectra were manually checked for outliers and these were discarded if required. The remaining spectra were averaged to one mean value spectrum per sample using MS Excel. In doing so, less than 1% of the spectra was discarded and 2 samples were completely lost (experiment A, day 13 and F, day 8).

For data analysis, sensory and fluorescence data were combined, resulting in 160 pairs of overall acceptance scores and fluorescence spectra. Partial least-squares discriminant analyses (PLSDA) were calculated using Solo Toolbox 9.1 (Eigenvector Research Inc., Manson, WA, USA). Fluorescence spectra were pre-processed with Savitzky-Golay smoothing (order 0, filter width 7), baseline correction (automatic weighted least squares, order 2), standard normal variate (SNV) normalization and mean centring (MC). In a first step, the samples were subdivided into two classes according to their sensory overall acceptability scores: Class 1 with scores ≤ 5 were considered as "acceptable", while class 2 with scores >5 were considered as "not acceptable". In a second step, the samples were divided into three classes and an intermediate class was introduced to account for uncertainties at the threshold. In doing so, class 1 ("acceptable") contained samples with a score of 4 or lower, class 2 ("still acceptable") included samples with scores higher than 4 and lower than 6 and class 3 ("not acceptable") included samples with overall acceptability scores of 6 or higher. In both cases, 3 PLSDA models with different combinations of calibration and validation experiments were calculated each. Model 1 was built on the experiments B, C, E, G, H and was validated with experiments A, D and F. Model 2 included data of the experiments A, C, D, F, H and was validated experiments B, E, G. Model 3 comprised experiments A, B, E, F, G and was validated with experiments C, D and H.

3 Results and Discussion

For the two-class PLSDA models, good classification results were obtained. The number of correctly predicted samples varied between the three models, see confusion matrices, Figure 1. To compare the performance of the models, accuracy, specificity and sensitivity were calculated for each model for calibration, cross-validation and validation according to equations 1 - 3, respectively. Class 1 (acceptable) was conventionally considered as "negative" and class 2 (not acceptable) was considered as "positive".

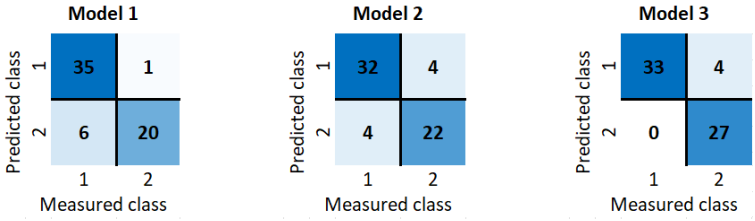


Figure 1: Confusion matrices of the predictions of the PLSDA models 1 - 3 with two classes; class 1 = "acceptable"; class 2 = "not acceptable"; validation samples from independent experiments.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{2}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

Table 2 shows that the accuracies of the three models were in the same range and only slightly decreased on average from cross-validated 91.3% in the calibration to 89.7% in the validation, varying between 87% and 93%. Larger discrepancies were observed for sensitivity and specificity, which might be due to the small number of samples in the split data sets. The analysis of misclassified samples per

Table 2: Performance metrics for PLSDA models 1 - 3 with 2 classes.

	Parameter	Model 1	Model 2	Model 3
Calibration	Accuracy	0.94	0.94	0.93
	Specificity	0.93	0.95	0.97
	Sensitivity	0.95	0.91	0.87
Cross-validation	Accuracy	0.91	0.93	0.90
	Specificity	0.90	0.94	0.91
	Sensitivity	0.92	0.91	0.87
Validation	Accuracy	0.89	0.87	0.93
	Specificity	0.85	0.89	1.00
	Sensitivity	0.95	0.85	0.81

overall acceptability score (see Table 3) showed, that all 3 models correctly classified samples with scores <4 "as acceptable" and rated all samples with scores >6 as "not acceptable" without exception. Therefore, a third class was introduced to cover the region of uncertainty between 4.5 and 5.5. All samples in this range were classified as "still acceptable", while other sample remained in their previous class.

Table 3: Overview of misclassified samples in predictions of the 2 class-PLSDA models.

Acceptability score	Number of samples					
	Model 1		Model 2		Model 3	
	Total	Misclassified	Total	Misclassified	Total	Misclassified
1 - 3	24	0	27	0	21	0
3.5	3	0	0	0	3	0
4	6	0	6	4	3	0
4.5	3	3	3	0	0	0
5	5	3	0	0	6	0
5.5	3	1	0	0	0	0
6	0	0	9	4	9	4
6.5 - 7.5	0	0	0	0	0	0
8 - 9	18	0	17	0	12	0

The classification results of the PLSDA models with three classes are shown in Figure 2. Again, the number of correctly predicted samples varied slightly between the three models. Accuracy, specificity and sensitivity are summarised in Table 4 to compare the performance of the models. The accuracy decreased on average from cross-validated 91.3% of the two-class model to 81.3% of the three-class models. In the validation, the accuracy decreased on average to 76.3%. Against expectations, the performance of the three-class PLSDA models decreased compared to the PLSDA models with two classes. Obviously, the introduction of the additional class increases the likelihood of errors rather than reducing it or the thresholds were not well chosen.

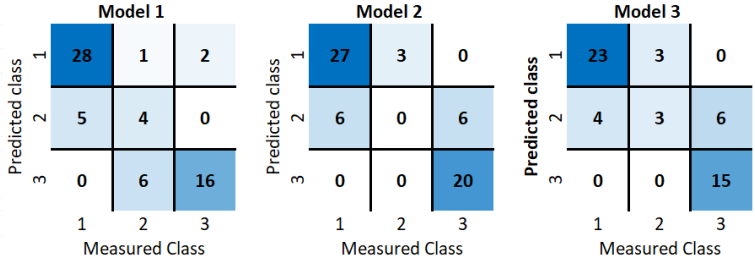


Figure 2: Confusion matrices of the predictions of the PLSDA models 1 - 3 with three classes; class 1 = "acceptable"; class 2 = "still acceptable"; class 3 = "not acceptable". The samples used for validation originated from independent experiments, which were not included in the calibration data.

Table 5 displays the misclassifications for the PLSDA models with 3 classes. Mostly one-class misclassifications occurred in the range from 3.5 to 6. The acceptability scores could not be strictly differentiated according to the spectra between "acceptable" and "still acceptable" as well as between "still acceptable" and "not acceptable". On the other hand, the three class PLSDA models had only 2 two-class misclassifications out of 178 predictions (1.1%) It is worth mentioning that this misclassification occurred with highly spoilt samples in a range where the two-class model performed correctly. This may indicate that two-class discrimination is feasible, but no further refinement. This is underlined by the finding that partial least-squares regression (PLSR) did not yield high correlations between the fluorescence spectra and the overall ac-

Table 4: Overview of misclassified samples in predictions of the PLSDA models with 3 classes.

	Parameter	Model 1	Model 2	Model 3
Calibration	Accuracy	0.81	0.81	0.81
	Specificity	0.80	0.80	0.81
	Sensitivity	0.96	0.94	0.94
Cross-validation	Accuracy	0.82	0.80	0.82
	Specificity	0.84	0.78	0.82
	Sensitivity	0.96	0.94	0.94
Validation	Accuracy	0.77	0.76	0.76
	Specificity	0.85	0.82	0.85
	Sensitivity	0.90	0.90	0.89

ceptability scores with coefficients of determiantion (calibration and cross-validation) of $R^2 = 0.71$ (data not shown).

Table 5: Overview of the incorrectly classified samples in prediction for the PLSDA models with 3 classes.

Acceptability score	Number of samples								
	Model 1			Model 2			Model 3		
		Misclassified			Misclassified			Misclassified	
	Total	One-Class	Two-Class	Total	One-Class	Two-Class	Total	One-Class	Two-Class
1 - 3	24	0	0	27	0	0	21	0	0
3.5	3	3	0	0	0	0	3	3	0
4	6	2	0	6	6	0	3	1	0
4.5	3	2	0	3	3	0	0	0	0
5	5	2	0	0	0	0	6	3	0
5.5	3	3	0	0	0	0	0	0	0
6	0	0	0	9	6	0	9	6	0
6.5 - 7.5	0	0	0	0	0	0	0	0	0
8 - 9	18	0	2	17	0	0	12	0	0

4 Conclusion

The study has shown that fluorescence spectra recorded with a portable device have potential to calculate (or predict) the overall sensory acceptability of minced pork which was packed under high oxygen modified atmosphere with 70% O_2 and 30% CO_2 and which was stored at 2, 10 or 14°C with a joint model for all three temperature regimes. PLSDA models with 2 classes were able to distinguish correctly between "acceptable" and "not acceptable" samples. The accuracy of the predictions in a validation with independent samples was 87 - 93%. The introduction of an intermediate class to account for uncertainties in the range between "acceptable" and "not acceptable" did not improve the classification results. Au contraire, the total number of misclassifications increased and occurred in the area between neighbouring classes.

Further investigations are required to confirm the results with a larger data basis and samples from different production origins to demonstrate suitability for industrial and commercial applications. To this end, fluorescence measurements should also be recorded through the packaging material so that the overall sensory acceptability can be determined non-destructively with the packed meat samples.

Acknowledgements

The project was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany. The Federal Office for Agriculture and Food (BLE) provides coordinating support for artificial intelligence (AI) in agriculture as funding organization, grant numbers 28DK126C20 and 28DK126E20. We would like to thank Manfred Behrschmidt, Joef Haida and Marco Zäh for preparing the meat samples, the trained panellist for the sensory assessment and Kathrin Fiedler for spectra recording and assistance in data processing.

References

1. E. Parliament and Council, "Regulation (ec) no 853/2004 of the european parliament and of the council of 29 april 2004 laying down specific hygiene rules for on the hygiene of foodstuffs," *Official Journal of the European Union*, vol. L139, pp. 55–205, 2004.
2. W. Wei, F. Zhang, F. Fu, S. Sang, and Z. Qiao, "Rapid detection of total viable count in intact beef dishes based on nir hyperspectral hybrid model," *Sensors*, vol. 23, no. 23, p. 9584, 2023.
3. N. K. A. El-Aziz, Y. H. Tartor, A. A. El-Aziz Gharib, and A. M. Ammar, "Propidium monoazide quantitative real-time polymerase chain reaction for enumeration of some viable but nonculturable foodborne bacteria in meat and meat products," *Foodborne pathogens and disease*, vol. 15, no. 4, pp. 226–234, 2018.
4. C. Grimmler, T. Kuhfuß, M. Heiden, and H. Schmidt, "Non-invasive assessment of the bioburden of minced pork using a hand-held fluorescence device," *tm-Technisches Messen*, vol. 85, no. 3, pp. 177–183, 2018.
5. L.-C. Fengou, E. Spyrelli, A. Lianou, P. Tsakanikas, E. Z. Panagou, and G.-J. E. Nychas, "Estimation of minced pork microbiological spoilage through fourier transform infrared and visible spectroscopy and multispectral vision technology," *Foods*, vol. 8, no. 7, p. 238, 2019.
6. E. Arias, V. Sierra, N. Prado, P. González, G. Fiorentini, J. Díaz, and M. Oliván, "Development of a portable near-infrared spectroscopy tool for detecting freshness of commercial packaged pork," *Foods*, vol. 11, no. 23, p. 3808, 2022.
7. H. L. Ramírez, A. Soriano, S. Gómez, J. U. Iranzo, and A. I. Briones, "Evaluation of the food sniffer electronic nose for assessing the shelf life of fresh pork meat compared to physicochemical measurements of meat quality," *European Food Research and Technology*, vol. 244, pp. 1047–1055, 2018.

Optimizing illumination for mid-wave infrared hyperspectral imaging in black plastic recycling

Maria Jernej, Gerhard Jakob, Malte Jaschik, and Maria Kainz

JOANNEUM RESEARCH Forschungsgesellschaft mbH, Institute DIGITAL, Steyrergasse 17, 8010 Graz, Austria

Abstract Sorting black plastics is challenging due to their strong absorption in the visible and short-wave infrared ranges induced by added carbon black particles as colorant. Conventional sorting methods struggle due to too low reflectivities. Mid-wave Infrared (MWIR) hyperspectral imaging offers a possible solution, providing sufficient signal level for classification. In this work, the illumination of the *Specim FX50* MWIR hyperspectral camera was optimized by the developmentof a line-focused illumination system using quartz tube heaters and elliptical reflectors. The new setup improved signal-to-noise ratios, eliminated temporal drifts, and allowed successful classification of black plastics (HDPE, ABS and PP). However, sensitivity to ambient temperature variations highlights the need for a temperature stabilized environment to achieve reproducible results and enable reliable operation in industry.

Keywords Hyperspectral imaging, mid-wave infrared, black plastics, focused illumination

1 Introduction

The global demand for plastics has created an urgent need to find efficient ways to recycle them, especially since most are derived from non-renewable resources. Among the many challenges associated with plastics recycling, sorting black plastics remains particularly difficult.

Conventional sorting techniques based on the visible or shortwave infrared (SWIR) portion of the electromagnetic spectrum struggle to separate black plastics. Added carbon black particles absorb strongly in this spectral region, resulting in a signal-to-noise ratio (SNR) that is too low for successful material classification. One potential solution to this challenge is to shift the analysis of the samples to the mid-wave infrared (MWIR) region. This region provides sufficient signal response even for black samples. [1]

Despite its potential, MWIR hyperspectral imaging still faces challenges such as sensitivity to environmental temperature variations, drift, and poor signal due to inappropriate illumination. As a result, it is rarely used for industrial applications. Limitations in data reproducibility and signal quality were also observed during initial testing of the *Specim Fx50* at the Imaging Laboratory of Joanneum Research. Since the original illumination, two conventional patio heaters, was suspected to be the main cause of the poor data quality, it was decided to completely revise the illumination setup. The previous setup was replaced by two quartz tubes whose light is focused onto the scan line of the camera by custom-made elliptical reflectors. In addition, the temporal stability and warm-up times of the hyperspectral imaging system were evaluated. The sensitivity of the system to environmental variations was investigated. With the final setup, colored and black plastic samples were acquired and analyzed to demonstrate the capabilities of the optimized system.

2 Material and Methods

2.1 Hyperspectral Cameras

The hyperspectral imaging system at the Joanneum Research Imaging Laboratory is equipped with several pushbroom hyperspectral cameras, covering a total spectral range from 400 nm to 5300 nm. The sample transport is realized by a variable speed conveyor belt. A rotary encoder at the start of the conveyor acts as an external trigger source to match the frame rate to the speed of the conveyor, ensuring adequate resolution in the y-direction.

This work focuses mainly on improving the quality of data generated by a *Specim Fx50* hyperspectral camera. The camera operates in

the MWIR with 308 spectral bands ranging from 2.7 μm to 5.3 μm. The camera's focal plane array is cryogenically cooled to reduce dark current and features temperature stabilized optics.

To demonstrate the superiority of the redesigned system over typical plastic sorting setups in the challenging task of classifying black plastic samples, comparison images of the samples were acquired with a hyperspectral SWIR camera (combination of *PhotonFocus MV3-D640I-M01-144-G2-12* area sensor and *Specim IMSPECTOR N17E* spectrograph) covering the spectral range from 900 nm to 1700 nm.

2.2 Plastic Samples

The use case considered a colored and a black sample of high density polyethylene (HDPE), acrylonitrile butadiene styrene (ABS), and polypropylene (PP), respectively. The colored HDPE and ABS samples were taken from the *Plasticprop* essentials sample kit. [2] The remaining samples were collected from domestic packaging waste, following the recycling code labeling on the objects. It should be noted that these samples, mainly corks and lids, may contain additives.

2.3 Acquisition and Calibration

For the acquisition of hyperspectral data, an exposure time of 1.5 ms is chosen. The sample transport speed is set to 0.03 ms^{-1} as this is fast enough to prevent melting of the samples but still guarantees a high signal response.

To bring the acquired images into a meaningful context and compensate for effects such as non-uniform illumination or varying sensor sensitivity, calibration needs to be done. Since spectral calibration, where the spectral pixels are mapped to their corresponding wavelengths, was already completed by the manufacturer of the camera, only radiometric calibration is still missing. In this work, the measured intensities are calibrated against a reference standard of high, (almost) constant reflectance (relative radiometric calibration). The true reflectance can be calculated by normalizing the dark current corrected signal of an

image against the dark current corrected reference intensity:

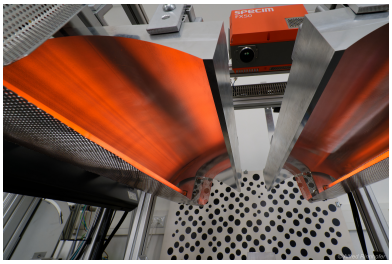
$$\text{reflectance} = \rho_w \cdot \frac{\text{measurement} - \text{dark signal}}{\text{white reference} - \text{dark signal}} \tag{1}$$

ρ_w is the (unknown) reflectance of the aluminum reference target.

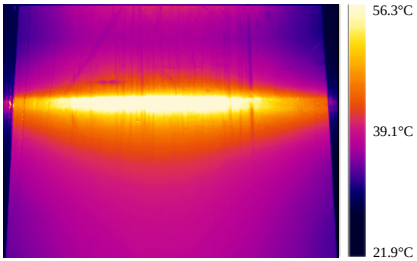
3 Line-Focused Illumination

The illumination of a hyperspectral imaging setup should have a continuous, preferably uniform spectrum along the spectral range of the camera. Especially for low reflective samples, a high illumination intensity is important to obtain an acceptable SNR at profitable sample transport speeds. In addition, temporal stability is essential to produce comparable results. Although the initial illumination, a conventional unfocused patio heater, was suspected to be the cause of poor data quality, again a thermal emitter was chosen for reasons of cost and simplicity, and because of the broad, continuous emission spectrum optimal for pushbroom hyperspectral imaging. A small radius quartz tube heater with a relatively low surface temperature of 800 K to 1000 K was chosen to have peak emission in the camera's spectral range.

To overcome the limitations of the previous system, the radiation of the thermal emitter should be focused on the scan line of the camera in order to achieve higher signal levels at lower power consumption and to avoid unnecessary warming of the environment. For this purpose, reflectors with an elliptical cross section have been designed. By geometrical considerations of the reflected rays, the optimal properties of the elliptical shells were evaluated, which are a compromise between the maximum ratio of reflected rays and the fulfillment of the size requirements. Aluminum was chosen as the reflector material because of its good reflectivity in the MWIR and its good machinability. The reflectors were milled from a single block of aluminum and then the surface was cleaned. Then the quartz tubes and reflectors were assembled and protective cages were installed. Finally, the two light sources were mounted on opposite sides of the camera's scan line and aligned using a thermal camera, see figure 1. To obtain uniform illumination of uneven samples, the illumination should be slightly defocused from the conveyor belt. [3]



(a) Final setup of the MWIR hyperspectral imaging system: *Specim Fx50* and focused lamps.



(b) Thermal image of the focus line of one of the light sources across the conveyor belt.

Figure 1: Final MWIR hyperspectral imaging setup and the corresponding focused illumination line.

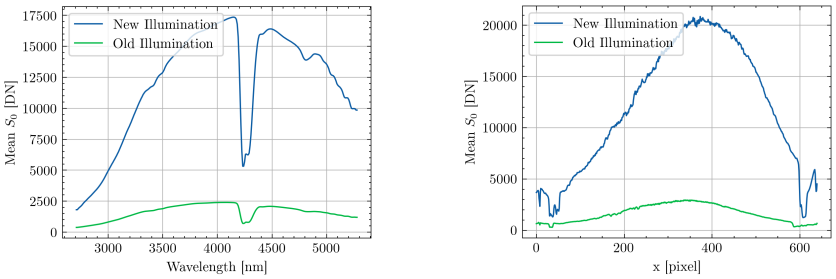
4 Results and Discussion

4.1 Improvements in Intensity and Signal-to-Noise Ratio

To evaluate the improvements in intensity with the new illumination, reference images of a diffusely reflective aluminum target, representing the emission characteristics of the light source itself, were captured. One of the initial light sources and one of the new, focused lamps were successively used to illuminate the sample.

As shown in figure 2, the dark current corrected signal increased significantly along both the spatial and spectral axes, indicating that the illumination intensity on the sample was effectively altered. Since both the old and the new illumination are thermal emitters, the spectral intensity distributions are a consequence of Planck's blackbody radiation law and thus are relatively similar. The reduced intensities at low and high wavelengths are mainly a consequence of the detector's wavelength dependent spectral response. The decrease in intensity along the spatial axis for both illumination scenarios is due to the limited width of the lamps. The sudden drop in spatial intensity can be explained by the sharp end of the aluminum target. For the new illumination, the intensity distribution is not perfectly centered, but this is compensated by the inverted distribution of the opposing lamp.

The overall SNR is summarized by averaging the dark current corrected signal along the sensor and dividing by the median of the pix-



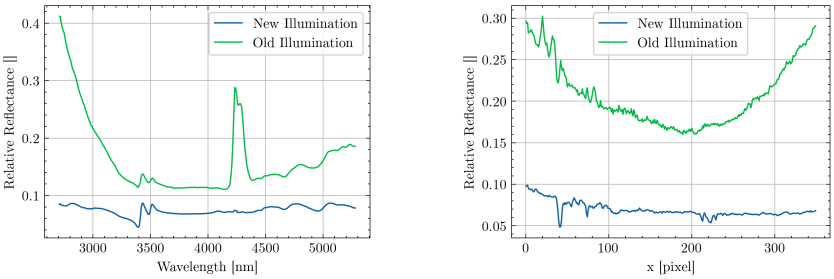
(a) spectral distribution (b) spatial distribution

Figure 2: Increase of signal from the initial unfocused illumination to the optimized illumination. Mean spectral and spatial response of the aluminum target.

elwise sensor noise. Through the new illumination, the overall SNR measured on aluminum reference target could be improved almost four-fold from 178 to 715.

4.2 Drift at the Edges of the Sensor

The data acquired with the old illumination often showed a bowl shape along both the spectral and spatial axis after radiometric calibration. Two examples, the calibrated mean reflectance spectrum of an HDPE sample and the mean intensity along the conveyor belt of the imaging system, are shown in figure 3 for both illumination setups. The increase in intensity towards low and high wavelengths or pixel numbers (thus the edges of the area sensor) appears only for the data acquired with the old illumination setup. In the average HDPE spectrum, the CO_2 absorption band is also very prominent and inverted, indicating a significant intensity drift between the acquisition of the white reference and the sample. In contrast, the curves acquired with the new illumination setup show no evidence of drift towards the edges of the detector. The fact that the two curves also do not match in the center region, even though the same reference target was used for calibration, again supports the theory that the old setup was not stable over time and that there was a strong drift between sample and reference acquisition.



(a) Average spectrum of a HDPE sample

(b) Mean intensity along the conveyor belt

Figure 3: Spectral and spatial intensity drift towards the border of the photo detector in the calibrated data acquired with the old illumination setup (green curve). The data acquired using the optimized illumination setup (blue curve) after calibration behaves as expected.

4.3 Stability

The signal increase during the heat-up process of the quartz tube was monitored, and it was found that it must be turned on for 5 min before recording hyperspectral data to ensure stable continuous emission. Otherwise, the heating curve of the thermal emitter would cause a strong drift in the signal. By monitoring the internal temperatures of the camera and the average signal, it was also observed that the *Specim Fx50* camera causes a signal drift as long as its uncooled internal components are still getting warmer at the beginning of operation. It was found that the camera needs to run at least for 2 h in advance to reach a stable state. In addition, care must be taken to prevent the optics of the camera from overheating, as this would lead to a rapid increase in the measured signal. It is therefore necessary to ensure sufficient ventilation of the optical heat sink. Experiments with an air conditioner have shown that the influence of the change in ambient temperature on the signal is severe. Cooling down the camera environment leads to a rapid downward shift of the spectrum in an order far above the noise level. Hence, the camera environment must be temperature stabilized to make the generated data truly comparable.

4.4 Use Case: Classification of Black Plastics

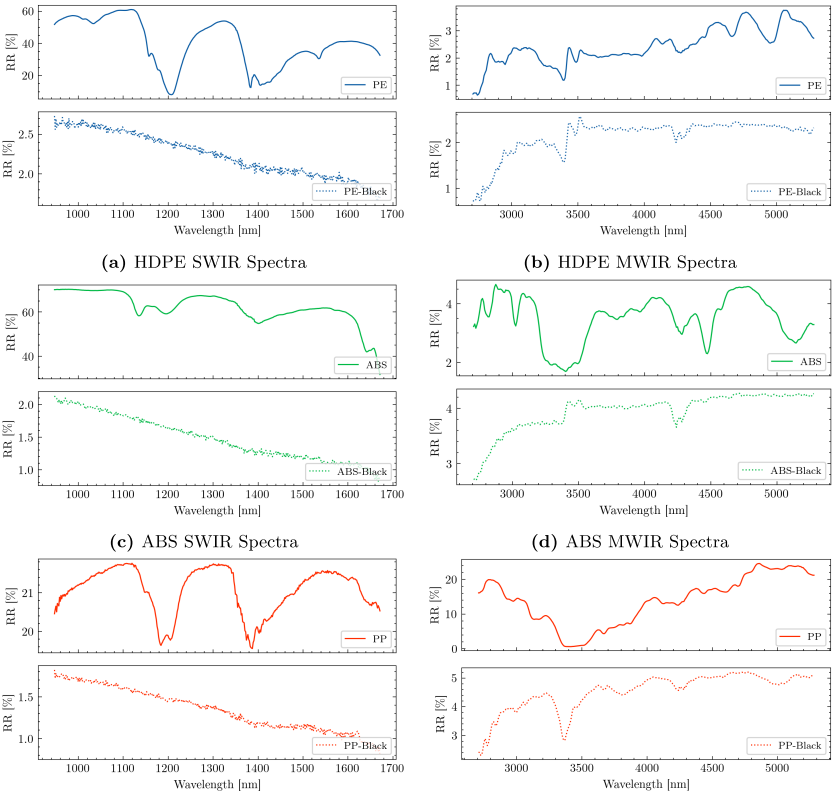
To demonstrate the capabilities of the redesigned imaging system, images of color and black plastic samples have been acquired and will be compared to images acquired with the SWIR camera set-up. The average spectra of all plastic samples are shown in figure 4. While all colored samples have spectral features captured by both cameras, the spectral response of the black samples is only present in the MWIR range. Although the SWIR range is generally more suitable for spectral characterization of plastics, it cannot be used to classify black plastics.

Since both HDPE and PP belong to the group of polyolefins, they share most of the spectral characteristics and thus have similar spectral responses in both spectral ranges. Similar to Lima Ribeiro, Fuchs, Lorenz, et al., characteristic spectral features of both colored and black HDPE at 3430 nm and 3512 nm could be resolved, while for ABS the nitrile stretching absorption peak at 4470 nm is only visible in the spectrum of the colored sample [4]. Overall, the spectrum of black ABS looks more like that of HDPE, suggesting that the sample may has been mislabeled or has a PE content. The four closely spaced peaks in the region of 3390 nm to 3525 nm of PP [5], resulting in a broad absorption band due to the poor spectral resolution of 35 nm of the camera, are less pronounced for the black sample.

The preliminary results show that MWIR hyperspectral imaging can overcome the limitations of lower wavelength methods and reveal the spectral characteristics of black plastics. However, as with all spectroscopic techniques, the effectiveness of MWIR hyperspectral imaging is highly task and material dependent. Variations in chemical composition, surface structure, and additives can significantly affect a material's spectral signature and, consequently, classification accuracy.

5 Conclusion and Outlook

A new illumination for the *Specim Fx50* MWIR hyperspectral camera has been developed and tested. By replacing the previous illumination, the intensity on the sample and the SNR could be significantly altered. Unexplained drifts at the detector edges and temporal variations were eliminated. Most of the signal drift can be avoided by allowing the illumination and camera to reach their stable state. The



(a) HDPE SWIR Spectra

(b) HDPE MWIR Spectra

(c) ABS SWIR Spectra

(d) ABS MWIR Spectra

(e) PP SWIR Spectra

(f) PP MWIR Spectra

Figure 4: Mean absorption spectra of black and colored HDPE, ABS and PP samples in the SWIR and MWIR regime. For the colored samples (solid line), a characteristic spectral response is visible in both spectral regions. In contrast, there is no spectral response for the black samples (dashed line) in the SWIR region because of the strong absorption of the black additive.

camera's operating temperature is still dependent on ambient temperature fluctuations. A temperature-stabilized environment is necessary to obtain reproducible results in future experiments. The results of the optimized MWIR hyperspectral imaging system for the classification of colored and black HDPE, ABS and PP plastics were promising. Com-

pared to SWIR reference images, most of the spectral features could also be revealed for the black samples. However, its effectiveness depends on the specific task and material, as variations in composition, surface structure, and additives significantly affect spectral signatures and classification accuracy.

Acknowledgements

Funded by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK).

This work was carried out as part of the master's thesis 'Optimizing Illumination Concepts and Calibration Procedures for Mid-wave Infrared Hyperspectral Imaging'. [6]

References

1. O. Rozenstein, E. Puckrin, and J. Adamowski, "Development of a new approach based on midwave infrared spectroscopy for post-consumer black plastic waste sorting in the recycling industry," *Waste Management*, vol. 68, pp. 38–44, Oct. 2017.
2. Sytyte Oy, "Plasticprop essentials sample kit." [Online]. Available: https://www.plasticprop.com/
3. P. Barnabé, G. Dislaire, S. Leroy, and E. Pirard, "Design and calibration of a two-camera (visible to near-infrared and short-wave infrared) hyperspectral acquisition system for the characterization of metallic alloys from the recycling industry," *Journal of Electronic Imaging*, vol. 24, no. 6, p. 061115, Nov. 2015.
4. A. de Lima Ribeiro, M. C. Fuchs, S. Lorenz, C. Röder, J. Heitmann, and R. Gloaguen, "Multi-sensor characterization for an improved identification of polymers in WEEE recycling," *Waste Management*, vol. 178, pp. 239–256, Apr. 2024.
5. C. Signoret, A.-S. Caro-Bretelle, J.-M. Lopez-Cuesta, P. Ienny, and D. Perrin, "MIR spectral characterization of plastic to enable discrimination in an industrial recycling context: II. specific case of polyolefins," *Waste Management*, vol. 98, pp. 160–172, Oct. 2019.
6. M. Jernej, "Optimizing illumination concepts and calibration procedures for mid-wave infrared hyperspectral imaging (in submission)," Master's thesis, Graz University of Technology, 2025.

Exploring near-infrared spectra of multi-material multi-layer packaging – Findings from the PET-stream

Nikolai Kuhn[1], David Treitler[1], Gerald Koinig[1], Moritz Mager[2], Jörg Fischer[2], and Alexia Tischberger-Aldrian[1]

[1] Montanuniversitaet Leoben, Chair of Waste Processing Technology and Waste Management,
Franz Josef-Str. 18, 8700 Leoben, Austria
[2] Johannes Kepler University Linz, Institute for Polymeric Materials and Testing,
Altenberger Str. 69, 4040 Linz, Austria

Abstract Multilayer, multi-material packaging impedes the mechanical recycling of plastic packaging. We characterized 30 trays made of PET or PET/PE with an IR-Microscope and examined the NIR-spectra of more than 300 trays recorded with a hyperspectral camera. We found that PET dominates the layer structure with layer thickness percentages of more than 80 %. Furthermore, the PE-peaks are detectable in PET/PE trays and thus indicate its separability via sensor-based sorting. Moreover, the orientation of the analyzed particle in relation to the sensor influences the NIR spectrum only minor. Especially the quality of the spectrum is influenced by the material thickness and its surface properties

Keywords Sensor-based sorting, material characterization, recycling

1 Introduction

During mechanical recycling, plastics are reprocessed into secondary raw materials in the form of flakes, pellets, and granules. Recycling typically includes the steps of collection, separation, sorting, washing,

and compounding [1]. To avoid inferior properties of the secondary raw material, the reprocessing requires a high purity of the respective target material and hence the removal of unwanted materials [2]. Through concatenating several liberation and separation steps, heterogeneous material streams such as post-consumer packaging waste can be processed into material concentrates ready for conversion [3]. However, the more complex the material composition of a particle, the more complex its recycling becomes [4]. For example, post-consumer plastic waste is dominated by food packaging which in turn places specific demands on protecting the packaged goods. Hence, packaging designs must combine different materials in different layers, increasing their complexity. Each layer fulfills specific functions, such as an extended shelf life, protection against mechanical stress, or simplified packaging and transportation [5]. As the share, order, and types of the polymers used greatly vary, these complex packaging are hard to separate through direct separation techniques such as swim-sink separation, sorting centrifuges, or zigzag classifiers. Alternatively, indirect separation techniques such as sensor-based sorting can be applied by using the chemical composition as a characteristic property [6]. For sorting plastics, near-infrared spectroscopy is widely used [7] [8]. However, even here the combinational complexity remains a challenge. [9] showed that it is possible to differentiate between packaging made of mono-material and multi-material on a sensor-based sorting system using near-infrared spectroscopy. They defined an overarching level of abstraction above the actual material composition based on overlying characteristic spectral information for the desired overarching class. Nevertheless, it remains unclear to what extent the spectral information and thus the classification is influenced by the sample presentation, share/thickness of the materials used and additional features such as sealing films. In this article, we investigate this conundrum using PET trays from lightweight packaging waste collections. These trays can either consist of the mono-material PET or of a multi-layer combination of PET and LDPE and possibly another barrier material such as EVOH. In addition, such packaging is usually sealed PET/PE or PET films [10].

In this article, we examine the following research questions:

1. What are typical designs of packaging trays made of PET and

PET/PE regarding used materials, layer thickness, and shares?

2. Are there spectral differences based on the side of the PET/PE tray facing the sensor?
   a) Sealing film: top vs. bottom
   b) Tray: top vs. bottom
   c) Sealing film & tray: top vs. bottom)
3. Are there spectral differences between a PET/PE sealing film and a PET/PE tray?
4. Are there spectral differences between PET/PE (tray and sealing film) and PET?
5. Are additional barrier layers (e.g., EVOH) recognizable in the analyzed spectrum?

2 Material and Methods

The used material stems from a sampling campaign of lightweight packaging waste of the separate collection in Austria conducted in 2023 by [11]. Amongst others, PET trays were separated and analyzed using a near-infrared handheld device (microPHAZIR, Thermo Fisher Scientific Inc., USA). The total amount of PET-trays amounts to 302, of which 212 (70 %) were mono-material PET and 90 (30 %) were multi-material PET/PE trays. For a better understanding of the sample properties, 33 packages were selected, differing in brand, size, and application. These were then examined using an IR-microscope (Spotlight 400, PerkinElmer Inc., USA) regarding layer thickness and material type.

Next, the samples were recorded on a sensor-based sorting machine equipped with a near-infrared hyperspectral camera (EVK-Helios G2 NIR 1, EVK DI Kerschhaggl GmbH, Austria). As the objects can pass the sensor in a wide variety of positions, the same samples were recorded in different orientations to the sensor. For this purpose, the samples were recorded both with the sealing film open and closed. In addition, images were taken from both angles, i.e., from the bottom or top side facing the sensor, resulting in a total of six different perspectives of the sample (cf. Fig. 1). All consequent data handling was

performed using Matlab 2022b (The MathWorks, Natick, USA). The spatial pixels of the sample were then manually selected from the hyperspectral image, avoiding overexposed and underexposed areas as well as those with decorations such as labels. The pixel values were then derived, smoothened, normalized, and combined into a mean spectrum. This procedure removes baseline effects from the spectra and overlapping bands are highlighted making it easier to distinguish the spectra from each other [12]. Figure 3 depicts normalized and derived spectra from the respective areas together with spectra obtained from PET-trays.

3 Results and Discussion

The results of the examined packaging trays are presented in Figure 1 and the results of the obtained near-infrared spectra are in Figure 3.

1) What are typical designs of packaging trays made of PET and PET/PE regarding used materials, layer thickness, and shares?

The 22 PET packaging trays analyzed had an average thickness of 341 µm, with a minimum of 85 µm and a maximum of 594 µm and a standard deviation of 123 %. The 11 multi-material multilayer trays analyzed had an average thickness of 297 µm, a minimum of 223 µm, a maximum of 437 µm, and a standard deviation of 73 %. A quarter of them consisted of an additional EVOH barrier embedded between LDPE layers. Furthermore, the top layer, which touches the packaged goods and to which the sealing film is attached, is composed of LDPE while the bottom layer is composed of PET. This is because the sealing films, which are often also made of PET/PE, are easier to seal with the same material, whereas the PET layer is the top layer for better printability (source). Lastly, PET accounted for at least 80 % of the total thickness of all multimaterial, multilayer trays. 2 shows the cross-section of three typical structures (PET, PET/LDPE, PET/LDPE/EVOH/LDPE).

2) Are there spectral differences based on the side of the PET/PE tray facing the sensor?

Depending on the area examined, the near-infrared spectra collected differ only slightly in their course or quality.

a) Sealing film, top vs. bottom

Layer 1			Layer 2			Layer 3			Layer 4			
Material	Thickness [µm]	Share of total thickness [%]	Material	Thickness [µm]	Share of total thickness [%]	Material	Thickness [µm]	Share of total thickness [%]	Material	Thickness [µm]	Share of total thickness [%]	Total Thickness [µm]
LDPE	42	18	PET	187	82	-	-	-	-	-	-	229
LDPE	19	7	EVOH	8	3	LDPE	19	7	PET	243	84	289
LDPE	41	16	PET	219	84	-	-	-	-	-	-	260
LDPE	35	15	PET	204	85	-	-	-	-	-	-	239
LDPE	42	16	PET	220	84	-	-	-	-	-	-	262
LDPE	14	4	EVOH	5	1	LDPE	18	5	PET	313	89	350
LDPE	42	10	PET	395	90	-	-	-	-	-	-	437
LDPE	39	15	PET	217	85	-	-	-	-	-	-	256
LDPE	17	6	EVOH	6	2	LDPE	20	7	PET	246	85	289
LDPE	38	9	PET	394	91	-	-	-	-	-	-	432
LDPE	40	18	PET	183	82	-	-	-	-	-	-	223

Figure 1: Composition of the examined PET/PE trays in terms of used material, layer thickness, and their respective share of the total thickness. Layer 1 equals the top layer touching the packaged good while layer 2 / 4 is the bottom outside layer.

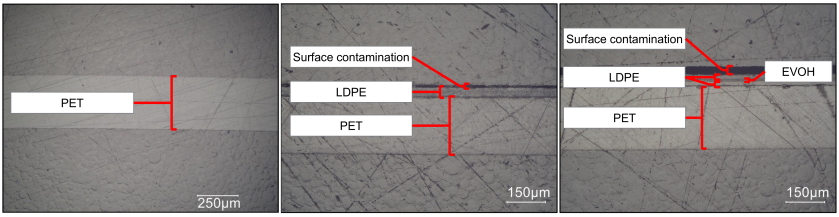


Figure 2: Cross-section of three typical structures of the examined trays (from left to right: PET, PET/LDPE, PET/LDPE/EVOH/LDPE) obtained with an IR-Microscope. The top layer touches the packaged goods.

There are barely visible spectral differences between the NIR-image of the sealing film captured from its top or bottom. However, it can be seen that the proportion of sinusoidal noise is higher on the front of the sealing film than in the other images. This is due to interferences caused by the interaction of radiation with the low layer thickness [13]. Another cause might be an additional treatment of the sealing film to make it more reflective, as a higher gloss is more appealing to customers.

b) Tray top vs. bottom

Again, there are barely visible spectral differences between the NIR-image of the tray captured from its top or bottom. However, when looking at the tray from the top, the PE peak at 1220 nm is somewhat

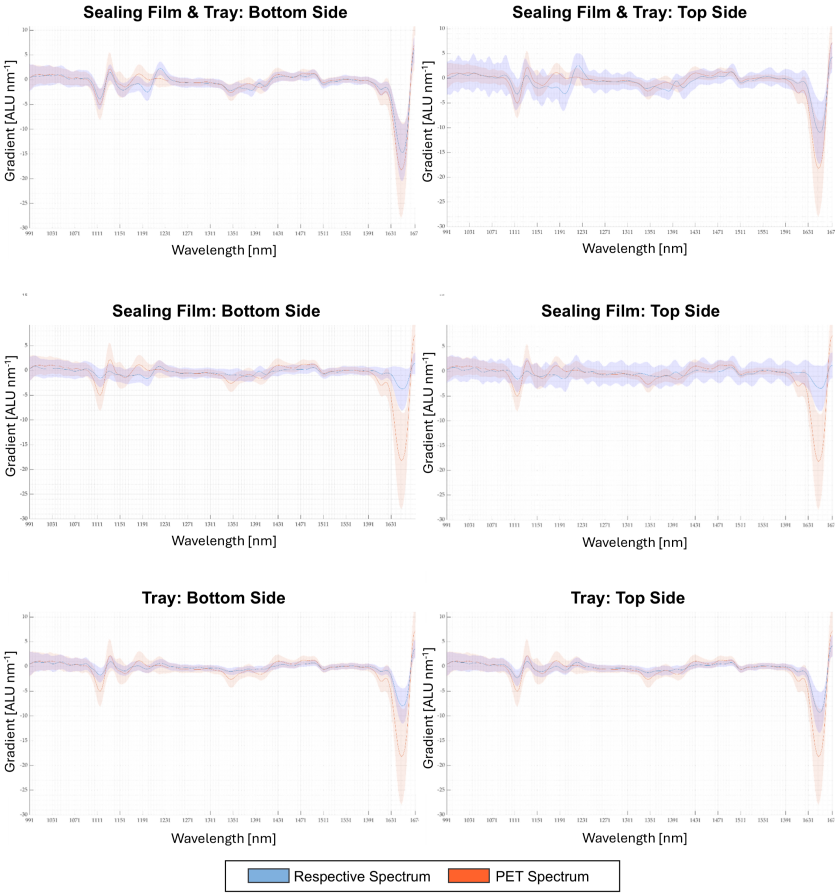


Figure 3: Near-infrared spectra of PET/PE trays presented to the sensor in different orientations.

more pronounced than when looking at it from its bottom. One reason for this could be found when viewing the images with an IR microscope (cf. 2: It is noticeable that the PE layer is often on the inside of the tray, i.e. when viewed from the top.

c) Sealing film & tray, top vs. bottom

When looking at the entire packaging, i.e. the sealing film and tray

together, from above or below, only very slight differences are recognizable. Here too, a slightly more pronounced peak is recognizable when viewed from below, i.e. the side of the tray facing the sensor. As described above, this could be because the PET layer is often located on the outside of the tray.

It is also noticeable here that the sinusoidal noise is more pronounced on the side of the reflective sealing film.

3) Are there spectral differences between a PET/PE sealing film and a PET/PE tray?

When comparing the tray spectra with the sealing film spectra, the aforementioned tendency of films to form sinusoidal noise due to interference becomes apparent again. This phenomenon generally poses a challenge in the sensor-based sorting of thin packaging such as films (cf. [14]). It can also be seen that the PET peak at 1650 nm is more pronounced for the trays than for the sealing film. This is probably due to the generally higher share (min. 80 %) of PET in the PET/PE trays.

4) Are there spectral differences between PET/PE (tray and sealing film) and PET?

In general, when examining the spectra, it can be seen that the PET/PE spectra from tray or sealing film differ from the mono-material PET spectra within the examined spectral range. The peaks characteristic of PET always dominate the PE peaks in the PET/PE spectrum. One reason for this could be the higher share of PET in the tray and the sealing film. This assumption is supported by the fact that the ester peaks characteristic of PET (especially in the wavelength range of 1120 nm (fourth band) and 1650 nm (third band), cf. [15] are particularly pronounced in the tray images, especially those with the tray-side facing the sensor. When comparing the PET spectra with the PET/PE spectra, it is noticeable that the PET/PE spectra generally show a slightly less characteristic course. Furthermore, a shift of the PE peak from 1180 nm to 1220 nm PE peak can be observed, which could be caused by the additive effects of the two materials.

5) Are additional barrier layers (e.g., EVOH) recognizable in the analyzed spectrum?

Although not plotted as a separate category, the spectra were also analyzed for their content of other barrier layers, in particular EVOH. Although, in principle identifiable by its OH groups in the near-infrared range [15], EVOH is not visible in the analyzed spectra. This is proba-

bly due to the very low concentrations, which are between 1-3 % of the total thickness of the analyzed trays.

4 Conclusion

Due to its complex composition, multi-material, multi-layer packaging poses a challenge for the recycling industry. To generate secondary raw materials with a high substitution potential, it would make sense to separate the current PET material stream from the separate collection of plastic packaging waste into a PET fraction and PET/PE fraction. However, this requires interdisciplinary cooperation between waste processors, plastics technology, and recycling technology. This article aimed to increase the understanding of PET/PE trays at the end of their life cycle through a systematic investigation.

For this purpose, multi-material, multi-layer trays from the separate collection of plastic packaging were analyzed. From a sample of 302 PET or PET/PE trays, 33 trays were analyzed using an IR microscope concerning their layer structure. It was found that almost exclusively two- or four-layer structures were used. EVOH could be detected as an additional barrier layer in some samples at a layer thickness of 1-3 %. It was also found that PET dominates the layer structure with layer thickness percentages of more than 80 %.

In the next step, various aspects of sensor-based sorting were analyzed. The aim was to identify possible factors influencing the near-infrared spectra. It was found that the orientation of the analyzed particle in relation to the sensor has observable but only very slight influences on the NIR spectrum. Furthermore, the material thickness and surface properties have a greater influence, particularly regarding the quality of the spectrum. In addition, spectral differences between the PET spectrum and the PET/PE spectrum are clearly recognizable. The PET/PE spectrum is more similar to the spectrum of PET than that of PE. This is probably due to the generally higher share (min. 80 %) of PET in the PET/PE trays. Finally, it was determined that EVOH is not visible in the analyzed spectral range due to its low proportion (1-3 %) in the trays examined.

One interesting aspect of future investigations could be to determine the layer thickness percentage at which certain plastics are detectable

in a multi-material multilayer packaging system during sensor-based sorting. In the future, this could result in a better sorting of plastic waste and subsequently increase its recycling rate.

Acknowledgements

This work was funded by the Austrian Research Promotion Agency (FFG) within the program "Energy of the Future. circular economy 2021 (KP)" under project *circPLAST-mr* (grant no. 4238670).

References

1. K. Ragaert, L. Delva, and K. Van Geem, "Mechanical and chemical recycling of solid plastic waste," *Waste Manag.*, vol. 69, pp. 24–58, 2017.
2. I. Antonopoulos, G. Faraca, and D. Tonini, "Recycling of post-consumer plastic packaging waste in the EU: Recovery rates, material flows, and barriers," *Waste Manag.*, vol. 126, pp. 694–705, 2021.
3. A. Feil and T. Pretz, *Plastic Waste and Recycling: Environmental Impact, Societal Issues, Prevention, and Solutions*. Elsevier, 2020.
4. C. Moyaert, Y. Fishel, L. Van Nueten, O. Cencic, H. Rechberger, P. Billen, and P. Nimmegeers, "Using recyclable materials does not necessarily lead to recyclable products: A statistical entropy-based recyclability assessment of deli packaging," *ACS Sustain. Chem. Eng.*, vol. 10, no. 36, pp. 11 719–11 725, 2022.
5. B. Morris, *The Science and Technology of Packaging*. Elsevier, 2017.
6. G. Koinig, B. Rutrecht, K. Friedrich, C. Barretta, and D. Vollprecht, "Latent recycling potential of multilayer films in austrian waste management," *Polymers (Basel)*, vol. 14, no. 8, p. 1553, Apr. 2022.
7. B. Bilitewkski and G. Härdtle, *Abfallwirtschaft*. Springer, 2013.
8. M. Kranert, *Einführung in die Kreislaufwirtschaft*. Springer Fachmedien: Wiesbaden, 2017.
9. G. Koinig, N. Kuhn, T. Fink, E. Grath, and A. Tischberger-Aldrian, "Inline classification of polymer films using machine learning methods," *Waste Manag.*, vol. 174, pp. 290–299, 2024.

10. M. Seier, V.-M. Archodoulaki, T. Koch, B. Duscher, and M. Gahleitner, "Polyethylene terephthalate based multilayer food packaging: Deterioration effects during mechanical recycling," *Food Packag. Shelf Life*, vol. 33, p. 100890, 2022.
11. N. Kuhn, G. Konig, and A. Tischberger-Aldrian, "Comprehensive characterisation of lightweight packaging waste to improve recycling," in *Eurowaste, Symposium Proceedings Sardinia*, Cagliary (Italy), October 2023.
12. K. W., *Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik: Ein Lehrbuch*. Weinheim: Wiley-VCH, 2006.
13. E. Jeszenszky, L. Kocsányi, and P. Richter, "Eliminating the interference pattern in near-infrared spectra used for identification of thin plastic foils," *Applied Spectroscopy*, 2004.
14. G. Koinig, N. Kuhn, C. Barretta, K. Friedrich, and D. Vollprecht, "Evaluation of improvements in the separation of monolayer and multilayer films via measurements in transflection and application of machine learning approaches," *Polymers (Basel)*, vol. 14, no. 19, p. 3926, Sep. 2022.
15. J. Workman, *The Handbook of Organic Compounds*. Elsevier, 2020.

Detection of bio-based additives in plastics using NIR data: Opportunity for bio-based markers

Malte Vogelgesang[1], Bettina Trojanowski[1], Stefan Hanstein[1], Andreas Züge[2], Wladislaw Benner[1], Emanuel Ionescu[1]

[1] Fraunhofer IWKS, Research Institution for Materials Recycling and Resource Strategies, Brentanostr. 2a, 63755 Alzenau
[2] SKZ - German Plastics Center, Friedrich-Bergius-Ring 22, 97076 Würzburg

Abstract In a research project on bio-based materials, characteristic near-infrared (NIR) signatures of various inorganic and organic fillers, including wood-based fibers, were identified in polymers. These signatures are strong enough for sensor-based sorting, enabling the removal of these materials from unfilled streams or directing them into specific recycling pathways. This led to the idea of using bio-based materials as markers, offering a sustainable alternative to markers made from rare earth elements like neodymium. The study aimed to develop bio-based markers detectable with conventional NIR sorting equipment. The substances were extracted from apple pomace and cocoa shells, including hemicellulose, theobromine, DEAE glycan, catechin, and others. After initial testing, depolymerized hemicellulose emerged as the most promising candidate for large-scale tests. Bottles and plates containing varying concentrations of markers were produced and cut into flakes for sorting tests. Support vector machines achieved 99.8 % accuracy of classification for inorganic markers and 95.0 % for organic markers. Sorting experiments indicate that bio-based markers can sustainably differentiate polymer materials using existing sorting equipment, but are prone to error from organic dust contamination. Additional research is needed to reduce the required concentrations for robust sorting, making it a viable market product.

Keywords Near infrared spectroscopy, polymer compounds, tracer-based sorting

1 Introduction

Plastic production is rapidly increasing and still almost exclusively relying on fossil resources [1]. When carbon-based polymers disintegrate or are incinerated, the carbon is released, such as into the atmosphere in the form of CO_2. Keeping the materials in loops through reuse or recycling reduces the need for additional resources. Another approach is the substitution of fossil feedstock with bio-based resources, e. g., through bio-based additives in the form of fillers, which are added to conventional base polymers. In the recent past, fiber-filled polymers have received a backlash in media, as to not being recyclable and therefore lowering the quality of recyclates in the established material streams [2]. In a research project on regionally-sourced bio-based materials, we found characteristic NIR signatures of different inorganic and organic fillers, including wood-based fibers, in the tested polymers. These signatures were strong enough for sensor-based sorting, which might offer a chance to remove the materials from unfilled material streams or to channel them into their own respective recycling pathways. This led to the idea of using bio-based materials as markers for individual material or product categories. Currently available markers for polymers are often made from rare earth elements [3], such as neodymium, which are critical for high-tech applications. Bio-based materials may offer a less critical alternative. The goal of the study was to develop bio-based markers which are detectable with conventional NIR sorting equipment.

2 Previous Work

In the research project "reGIOcycle", new plastic compound formulations were developed with a focus on bio-based fibers in polyolefins. The mechanical properties were improved through the application of specialized coupling agents. To investigate the behavior of these new materials within the waste streams of conventional plastic waste in commercial sorting plants, experiments were conducted at the pilot plant at Fraunhofer IWKS (Figure 1).

Samples of 1 kg of stepped plates were sorted three times on the sorting machine. The results showed that with current state of the art

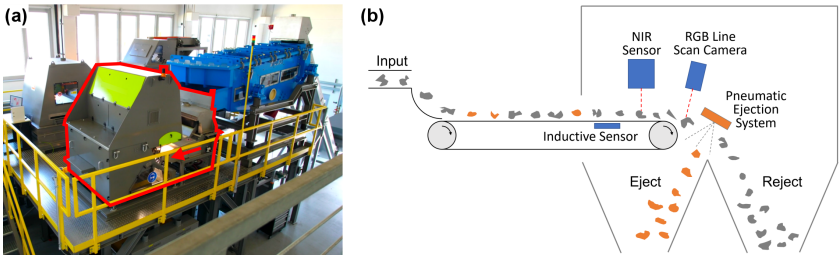


Figure 1: a) Pilot scale sorting system at Fraunhofer IWKS, sensor-based sorting unit marked in red. b) Principle drawing of sensor-based sorting.

NIR sorting technology, commercially available filled materials (wood-based fibres and also inorganic fillers) end up in the respective fractions of their base polymers polyethylene (PE) (Table 1).

Table 1: Sorting results of unfilled and filled HDPE (high-density polyethylene) compounds with commercial sorting parameters for PE.

Compound	Eject Rate
HDPE	100%
HDPE + Wood-based Fibres (20%)	99.6%
HDPE + Inorganic Filler (20%)	100%

When training the software of the sorting machine towards detecting fillers, it was possible to separate these materials reliably from a binary mixture with unfilled materials (Table 2).

Table 2: Sorting results of filled materials from unfilled materials in binary mixture of identical sample size.

Compound	Purity	Yield	F1-Score
HDPE + Organic Filler	98.3%	99.5%	0.99
HDPE + Inorganic Filler	98.3%	100.0%	0.99

With the promising results from this study, a project called ”bIR-mark” was started to specifically develop and test bio-based marker materials for plastics in commercial NIR sorting systems.

3 Material and Methods

3.1 Samples

To find possible candidates for bio-based markers, various substances were extracted at the extraction pilot plant at Fraunhofer IWKS [4–6]. The feedstock was apple pomace, the residue after producing apple juice, and cocoa shells. This was done in a way to keep it usable as animal feed and for soil applications, to not compete with food production. Catechin extracts were provided by Martin Bauer GmbH & Co. KG. The challenge in producing bio-based markers was to create or isolate molecules with a strong NIR signature. A strong and characteristic NIR signature is of great importance because many of the overtone vibrations coincide, leading to an overlap of the marker vibrations with, e. g., PE vibrations. The most prominent absorption bands are caused by the overtone and combination bands of the molecular fundamental vibrations of the functional groups C–H, N–H, O–H, and S–H. PE has very strong absorption bands due to the multitude of CH_2 chains. Natural substances differ from PE in that, besides CH_n vibrations, they exhibit characteristic vibrations of functional groups like C–O (alcohols, ethers, acids, esters), N–H (amines, amides), NO_2 (nitro groups), C=C (aromatics, alkenes), and C=O (aldehydes, ketones, acids, esters). Another limitation was the measurement range of the NIR detector in the sorting machine, which detects in a range of 1300–1900 nm. This range includes the first and second overtone vibrations of the molecular vibrations of CH, CH_2, CH_3, CONHR, Ar–OH, $CONH_2$, RNH_2, ROH, and H_2O groups.

The tested substances included variants of hemicellulose (HEC) in differing molecular masses, theobromine (TB), DEAE glycan (HEC-D), hemicellulose stearate (HEC-SA), 4-tert-butylbenzoic acid (HTBBA) and catechin (CAT). These substances were then tested for their NIR detectability in the pilot scale sorting plant at Fraunhofer IWKS, their stability under extrusion-relevant temperatures, and their compounding properties. Starting with pure samples of the substances (Figure 2), the most promising candidates were chosen for larger scale tests. To compare the new developments with state-of-the-art technology, a commercially available inorganic marker (IM) by Tailorlux GmbH was included in the tests. The substances were incorporated into HDPE

compounds by project partners at SKZ. These were then processed via injection moulding into coin-sized chips and later to palm-sized discs (diameter of 7 cm) for a larger area, leading to more NIR spectra. Concentration series were conducted to find the lower limits of reliable detection.

In a final step, hemicellulose with 9 kDa (HEC9) and IM were incorporated in HDPE, LDPE (low-density polyethylene) and PP (polypropylene) matrix materials by SKZ. Sample parts were produced using common manufacturing processes. Bottles and plates with a wall thickness of 500 µm were produced using blow molding and plate pressing. The samples were manually cut into flakes of about 2 cm diameter for sorting tests. The influence of thermal and mechanical stress on the marker compounds during processing and mechanical recycling was examined. No significant changes in optical appearance or in thermal and rheological properties were observed. Sample plates were also subjected to thermal and UV aging.

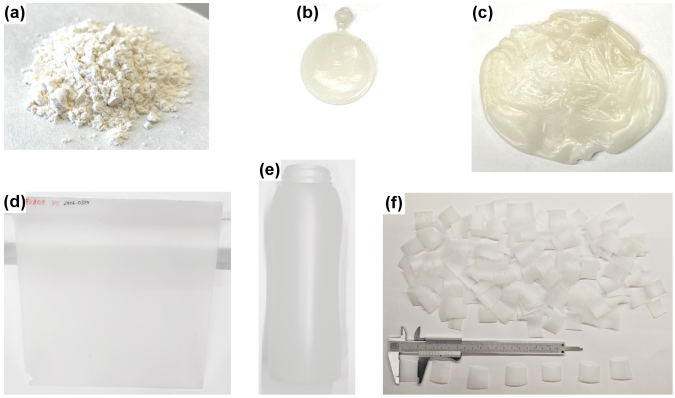


Figure 2: Sample shapes: a) Pure marker substance, b) Chip, c) Disc, d) Plate, e) Bottle, f) Flakes (from Bottles)

3.2 Machine Learning-based NIR Detection

The developed bio-based markers were recorded in pure form using the NIR sensor of a Varisort Compact multi-sensor sorting machine by

Sesotec GmbH with a spectral range of 1300 - 1900 nm at the pilot sorting plant of Fraunhofer IWKS. From the comparison of the spectra with the reference spectra of polyethylene, promising candidates for compounding could be identified. A manual analysis of the spectra from HDPE compounds was not able to identify characteristic influence of the low marker concentrations (1 % and below) to facilitate detection. Due to the resolution of the NIR sensor with pixels of about 5 x 5 mm, only a very limited number of independent spectra could be recorded per chip sample, inhibiting robust training of the classification algorithms. Therefore, larger sample discs were produced and NIR spectra recorded of front and back. Black and white balancing was performed and erroneous spectra manually removed. Spectra of marked and unmarked samples were visually compared. To reduce the influence of individual training approaches on the classification results, machine learning with identical settings were applied to all marker compounds. For their high versatility, artificial neural networks in the form of a multilayer perceptron (MLP) were chosen, as well as upport Vector Machines (SVM) for their good performance in weak signals, such as in edge pixels [7].

In Python, pairings of marked and unmarked HDPE samples were created. Classes were balanced and split into 70/30 train/test-split and the spectra were smoothened, differentiated and normalized. After training with MLP and SVM, the classification accuracy was evaluated using the unseen test data, as proportion of correctly assigned spectra relative to the total number of spectra (Formula 1).

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

3.3 Sorting Experiments

The flakes from plates and bottles were then used to test the marked and unmarked materials for their sortability, which is a necessary criterion for technical recyclability. They were again recorded with the NIR sensor, black and white corrected, and erroneous spectra removed.

Material types were combined for each sorting scenario and classes balanced. Normalization and mean centering were performed for each spectrum and per wavelength band across all spectra. A partial least squares (PLS) algorithm was trained over the entire spectral range. For sorting operation, noise filtering was applied and a pixel-based error correction was integrated by assessing 3 direct neighbors.

In the sorting trials, various scenarios were investigated. The suitability of the markers in HDPE was practically tested by sorting marked from unmarked HDPE flakes, for both the bio-based HEC9 and the inorganic marker in different concentrations. The influence of a color masterbatch on IM detection was assessed, as well as IM in two types of LDPE (a: FA3227, b: FA7220). To investigate sorting of HDPE and LDPE, HDPE with markers was sorted from unmarked LDPE, and compared with pure HDPE and LDPE. For each of the sorting scenarios, the sorting plant was set to eject the target material (t), and the samples were processed individually by material type. To statistically validate the results, each trial was repeated three times. The evaluation was then carried out based on weight of the eject and pass fractions. In the case of the aged samples, due to the small sample size, particle count was used.

From the weights/particle counts of the eject fraction (e) and the pass fraction (p), Eject Ratio (ER), i. e., the yield (Formula 2), was calculated for each run of the respective material. The arithmetic mean of the three runs was then calculated to determine the average yield of the target material in this sorting scenario:

$$\mathrm{yield_t} = \frac{\mathrm{m_{e,\,t}}}{\mathrm{m_{e,\,t}} + \mathrm{m_{p,\,t}}} = \mathrm{ER_t} \tag{2}$$

Assuming an equal mass distribution of the materials in each sorting scenario, we subsequently calculated the theoretical purity of the target material for illustration purposes from the ERs (Formulae 3, 4):

$$\mathrm{purity}_t = \frac{\mathrm{m_{e,}}_t}{\sum_i \mathrm{m_{e,}}_i} \tag{3}$$

$$\text{For equal masses of all materials:} \quad \mathrm{purity}_t = \frac{\mathrm{ER}_t}{\sum_i \mathrm{ER}_i} \tag{4}$$

Sorting exhibits a trade-off between purity and yield. Since their individual manifestation is a matter of process adjustment and not the capability of the marker, the F1-score was calculated (Formula 5) to combine both indicators [8]. An F1 of 1 would indicate perfect sorting, 0.5 with two equally-sized material quantities would indicate a random distribution without meaningful sorting, whereas 0 would indicate a purity and/or yield of 0.

$$F1 = 2 \times \frac{\text{Purity}_t \times \text{Yield}_t}{\text{Purity}_t + \text{Yield}_t} \tag{5}$$

4 Results and Discussion

4.1 Machine Learning-based NIR Detection

The spectral signatures of CAT (blue) and HEC with 30 kDa (black) deviate significantly from the average HDPE spectrum (red), which may indicate good recognizability (Figure 3a).

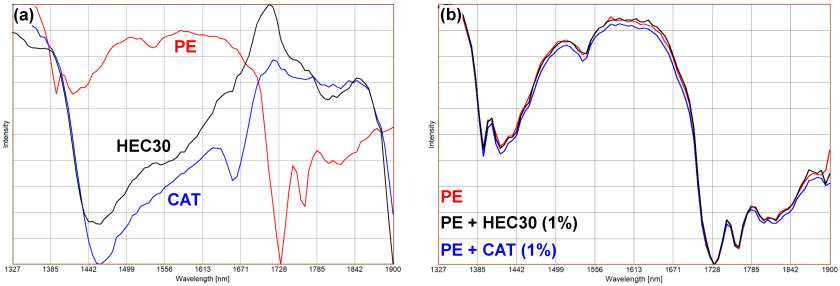


Figure 3: a) Average spectra of pure catechin (CAT, blue) and hemicellulose 30 kDa (HEC30, black) compared to PE (red), b) Average spectra of 1 % marker in HDPE chips.

During manual analysis however, there were hardly any differences visible between the spectra of pure HDPE and HDPE with marker concentrations of 1 % and below (Figure 3b). Surprisingly, this was also the case for the commercially available IM. Due to the low spatial resolution of the NIR sensor relative to the sample size and mostly only

single samples per substance/concentration, the samples were passed through the sensor several times, even though this would not yield fully independent measurements.

Here, the machine learning came into play. The MLP showed error-free distinction of the marked samples from the pure HDPE samples in almost all cases, which resulted in limited informational value. Accordingly, the results of the SVM were used as the significant decision criterion for the suitability of the markers (Table 3).

Table 3: Results of SVM on test dataset for the accuracy of distinguishing compounds with markers from samples without.

Marker	Concentration [wt%]	Sample	Accuracy
CAT	1%	Chip	84.8%
CAT	10%	Chip	99.7%
HEC 30 kDa	0.1%	Chip	76.8%
HEC 30 kDa	1%	Chip	87.1%
HTBBA	1%	Chip	70.1%
HEC 16 kDa	0.5%	Disc	94.6%
HEC-D	0.5%	Disc	78.9%
HEC-SA	0.5%	Disc	88.8%
IM	0.5%	Disc	99.9%
TB	0.5%	Disc	91.5%

The results show very good recognizability of CAT with a concentration of 10 wt% and of depolymerized HEC16 with 0.5 wt%. The commercial marker showed the best results, even in low concentrations of 0.5 wt%. Since low quantities of markers are pursued, depolymerized hemicellulose was chosen for further trials. Due to better results of lower molecular weights, the substance was further depolymerized.

4.2 Sorting Experiments

The results of SVM with the flake spectra was promising (Table 4). In the sorting test for HEC9 however, it was found that with each pass, the proportion of incorrectly ejected unmarked HDPE increased. Subsequently, the sample spectra were re-recorded and retrained, with results already significantly worse in the software compared to the initially recorded spectra. The latest sorting results are presented below.

This decrease could be attributed to the contamination of the samples with fine dust, as inevitably occurs in waste sorting facilities. The organic deposits on the initially pure HDPE reference samples exhibit a spectrum comparable to the samples with the hemicellulose marker, causing them to be incorrectly ejected as well. Only at a higher concentration of 1 % was an improvement in distinguishability achieved.

The inorganic marker shows robust results for HDPE and PP that increase with concentration. The sortability was hardly affected by the color masterbatch. For both LDPE types, the sorting accuracies are relatively low. One possible explanation is that the additives used in the LDPE types react in similar wavelength ranges as the marker. This could obscure the actual effect of the marker, which may explain why Type 1, which is UV-stabilized and designed for high stress, seems to be less suitable for combination with the marker.

It is evident that a certain distinction is generally possible even with unmarked HDPE and LDPE samples. The combination with markers showed practically no influence on the result. Since in practice there is a far greater variety of formulations than could be represented in the project, overlaps in additives could complicate NIR-based differentiation between LDPE and HDPE. Oven storage showed no clear change and, in particular, no deterioration in sortability (not shown). Starting from 14 days of UV weathering, there seems to be a slight deterioration in results, although statistical fluctuations cannot be ruled out.

5 Conclusions and Outlook

The detection of bio-based marker substances using NIR was investigated on a pilot scale sorting plant. Initial tests with small samples lacked robust detection due to sparse spectral data, leading to the use of machine learning methods like Support Vector Machines. Hemicellulose emerged as a promising candidate and, along with an inorganic marker, was incorporated into larger samples for sorting experiments. While initial results were encouraging, contamination with fine dust adversely affected the detection of hemicellulose in HDPE and PP samples, requiring higher concentrations (1 %) for better distinguishability. The inorganic marker did not show this issue and performed well even at low concentrations (0.1–0.2 %). Aging tests indicated that

Table 4: SVM and Sorting Results for HEC9 and IM in HDPE, LDPE and PP

Polymer	Marker	Accuracy (SVM)	Purity	Yield	F1-Score
HDPE + Marker, HDPE					
HDPE	HEC9 (0.1%)	95,0%	50.8%	69.2%	0.59
HDPE	HEC9 (0.2%)	95,8%	48.9%	67.0%	0.57
HDPE	HEC9 (1.0%)	98,1%	59.5%	85.3%	0.70
HDPE	IM (0.1%)	99,8%	88.5%	75.4%	0.81
HDPE	IM (0.2%)	100%	89.5%	83.7%	0.87
HDPE	IM (0.5%)	100%	97.0%	88.4%	0.93
HDPE	IM (0.2%) + Color (0.75%)	100%	85.1%	82.9%	0.84
LDPE + IM, LDPE					
LDPEa	IM (0.1%)	88,6%	50.5%	71.4%	0.59
LDPEa	IM (0.2%)	99,7%	53.3%	90.9%	0.67
LDPEb	IM (0.1%)	99,3%	56.6%	86.3%	0.68
LDPEb	IM (0.2%)	99,8%	59.3%	88.5%	0.71
HDPE (+ Marker), LDPEa					
HDPE	Unmarked	99,8%	95.8%	81.6%	0.88
HDPE	HEC9 (0.2%)	99,9%	90.4%	83.7%	0.87
HDPE	IM (0.2%)	99,9%	93.0%	84.6%	0.89
PP + Marker, PP					
PP	HEC9 (0.1%)	75,9%	49.8%	75.3%	0.60
PP	IM (0.2%)	99,4%	90.8%	83.0%	0.87

both markers maintained their sortability without significant deterioration. Tested HDPE and LDPE types were possible to be sorted without need for markers. Further research needs have been identified. The development of the hemicellulose marker should continue, exploring modifications that could enhance its NIR signature. To effectively mark different materials, it's essential to identify additional marker systems and explore combinations, including the possibility of combining organic and inorganic markers.

Acknowledgements

The study of reGIOcycle was supported by funds of the German Federal Ministry of Education and Research (BMBF) under the FONA initiative and Stadt Land Plus (FKZ: 033L218). bIRmark was supported

by funds provided by the IGF program of the German Federal Ministry for Economic Affairs and Climate Action (BMWK) (FKZ: 01IF22106N). Production of hemicelluloses with adjusted size by the team of Marius Wolf, Fraunhofer IWKS, is gratefully acknowledged.

References

1. R. Geyer, J. R. Jambeck, and K. L. Law, "Production, use, and fate of all plastics ever made," *Science advances*, vol. 3, no. 7, p. e1700782, 2017.
2. X. Zhao, K. Copenhaver, L. Wang, M. Korey, D. J. Gardner, K. Li, M. E. Lamm, V. Kishore, S. Bhagia, M. Tajvidi, H. Tekinalp, O. Oyedeji, S. Wasti, E. Webb, A. J. Ragauskas, H. Zhu, W. H. Peter, and S. Ozcan, "Recycling of natural fiber composites: Challenges and opportunities," *Resources, Conservation and Recycling*, vol. 177, p. 105962, 2022.
3. C. Olscher, A. Jandric, C. Zafiu, and F. Part, "Evaluation of marker materials and spectroscopic methods for tracer-based sorting of plastic wastes," *Polymers*, vol. 14, no. 15, 2022.
4. A. Möller and S. Hanstein, "Process for obtaining glycan extracts from apple pomace producing functionalized glycans, and functionalized glycan that can be produced by the process and its use," German Patent WO2 022 013 161A1, 20.01.2022.
5. M. Wolf, F. Berger, S. Hanstein, A. Weidenkaff, H.-U. Endreß, A. M. Oestreich, M. Ebrahimi, and P. Czermak, "Hot-water hemicellulose extraction from fruit processing residues," *ACS omega*, vol. 7, no. 16, pp. 13 436–13 447, 2022.
6. M. Wolf, S. Hanstein, O. Schmitz, P. Czermak, and M. Ebrahimi, "Depolymerization of hemicelluloses utilizing hydrothermal and acid catalyzed processes proceed by ultrafiltration as fractionation media," *Carbohydrate Polymer Technologies and Applications*, vol. 6, p. 100355, 2023.
7. X. Chen and A. Feil, "Detection and classification of heterogeneous materials as well as small particles using nir-spectroscopy by validation of algorithms," in *4th International Conference on Optical Characterization of Materials*, 2019.
8. N. Kroell, T. Dietl, A. Maghmoumi, X. Chen, B. Küppers, A. Feil, and K. Greiff, "Assessment of sensor-based sorting performance for lightweight packaging waste through sensor-based material flow monitoring: Concept and preliminary results," in *Proceedings of 9th Sensor-Based Sorting & Control Conference*, 2022.

Robust model development for HSI-based characterization of post-consumer plastics

Lukas Roming, Georg Maier, and Paul Bäcker

Fraunhofer Institute of
Optronics, System Technologies and Image Exploitation IOSB
Fraunhoferstr. 1, 76131 Karlsruhe, Germany

Abstract The need for effective recycling solutions has led to the investigation of advanced technologies for material characterization. This study explores chemometric methods including artificial neural networks for characterizing post-consumer plastics using near-infrared hyperspectral data. We analyze different classification methods in terms of overall performance, their ability to generalize onto another dataset as well as outlier detection. The investigated methods are Linear Discriminant Analysis (LDA), Artificial Neural Networks (ANN), Spectral Angle Mapper (SAM), and Partial Least Squares Discriminant Analysis (PLS-DA). We show that the performance on the main dataset is similar for all methods, with an accuracy of approximately 98%, whereas the models varied in terms of generalization and outlier detection. SAM performed best in terms of outlier detection, outscoring other methods by 8 percentage points.

Keywords hyperspectral imaging, robust model development, post-consumer plastics, near-infrared

1 Introduction

With an increasing demand for sustainable plastic products, high-quality recyclates are needed, which require a pure separation of lightweight packaging (LWP) waste into its material components. Key technologies that offer a great potential for material separation are sensor-based sorters, employing near-infrared hyperspectral imaging

(NIR-HSI), which allows to capture detailed spatial and spectral information. This information can be analyzed using chemometric methods to distinguish different types of polymers. Sorting purity depends on the sensor, the separation unit, and data analysis methods. These methods are usually based on statistical models, which may fail, when the statistics of the input data slightly changes e.g. under varying conditions in the sorting plant. We adress this issue by comparing chemometric methods in terms of their adaptability to a different measurement scenario and the detection of outliers in the form of foreign material and background (defined by the conveyor belt).

2 Related works

Sorting post-consumer plastics using NIR-HSI has been used for many years and is an established method in research and industry [1]. Most common methods for spectral analysis are Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM) [1] [2]. Based on these methods, various polymers have been successfully classified. Investigated polymers include PET, HDPE, LDPE, PP, PVC, PS, ABS, PA, HIBS, and more [1] [2]. Recent studies suggest further potential in improved analysis methods.

Mehrubeoglu et al. [3] investigated 13 plastic samples of 11 different polymer types. Four ratios of spectral responses were utilized as features. A sample is classified by the shortest Euclidean distance to labeled feature cluster centers. Some miss-classified pixel existed in the final semantic segmentation result.

In a study by Zheng et al. [4] ABS, PS, PP, PE, PET, and PVC were explored using NIR-HSI and a fisher discrimination model, which is closely related to the LDA. The model showed an accuracy of 84.9% in cross-validation and 100% for 21 unknown cleaned samples.

Wu et al. [5] analyzed 298 samples of WEEE plastics, specifically PP, PS, and ABS. Model robustness has been put to the test by training with virgin plastics and testing with WEEE plastics. Among the tested methods, PCA-LDA performed best in accuracy and robustness (99.7%), while SAM was also accurate (99.3%), but sensitive to the choice of data preprocessing. PLS-DA was less effective (91.9%) despite the fact,

that it is one of the most widely used methods.

Xia et al. [6] used a CNN for classifying polymers reaching 98.00% accuracy, outperforming common methods like SIMCA (69.98%) and PLS-DA (57.00%). The analysis was based on 159 clear and pure samples, including 84 black plastics of 7 different polymers. In their outlook the authors stated that future work will involve actual daily life samples to gain a more robust model.

Most studies utilize small datasets of up to 150 samples, resulting in limited variance in the sample collection and the risk of performance drop, when being applied in a sorting facility. Further, datasets often consist of clean, virgin material. As stated in [5], training samples from a variety of sources would be necessary to prevent misclassification in the circumstances of changes of coating and contamination on the surface.

To meet the aforementioned gaps, this study aims at

1. a comparative analysis on chemometric methods based on an extensive dataset of untreated post-consumer plastics,
2. a cross-dataset analysis between laboratory and industry conditions,
3. and evaluating the models ability to detect outliers

to find the best model with a low sensitivity to changes in the samples appearance and the measurement setup. Both circumstances can change during the lifetime of the classification system in a sorting facility.

3 Material and methods

In this section the recorded datasets are described in detail. After that, the measurement setups will be presented, before the methods for classifying the data are introduced.

3.1 Datasets

Two datasets have been created for the cross-dataset analysis using a hyperspectral camera. Both datasets include samples of Polypropylene (PP), Polyethylene terephthalate (PET), Polyethylene (PE), and

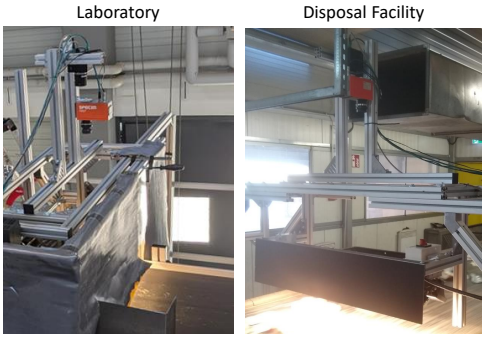


Figure 1: Measurement setups used to record RGB and NIR-HSI images. On the right, the setup is placed on top of a conveyor belt at the disposal facility in Iserlohn, Germany. A similar setup has been used to record the data in the laboratory as depicted on the left.

Polystyrene (PS). For each hyperspectral image there is also an RGB image, which data is used only for annotation of the samples and visualization in this study. A few samples of the datasets can be seen in Figure 2. By presorting the samples according to their material category, dataset annotation has been simplified. After recording, segmentation was applied to extract foreground spectra.

The **laboratory dataset** was captured at Fraunhofer IOSB in Karlsruhe, Germany. The dataset includes 102, 119, 196, and 131 samples of PP, PET, PE, and PS respectively. A single sample represents one post-consumer plastic packaging. For all 548 samples, multiple NIR spectra were recorded, resulting in a total number of more than 133,000 spectra.

The **disposal facility dataset** has been recorded at the Lobbe RSW GmbH in Iserlohn, representing images from a realistic sorting scenario. The disposal facility dataset includes 381, 94, 152, and 290 samples of PP, PET, PE, and PS respectively. In total, this results in 917 samples, of which more than 85,000 NIR spectra were taken.

3.2 Measurement setup

The measurement setups are shown in Figure 1. The line-scan camera FX17e from SPECIM was used to capture the hyperspectral images, fea-

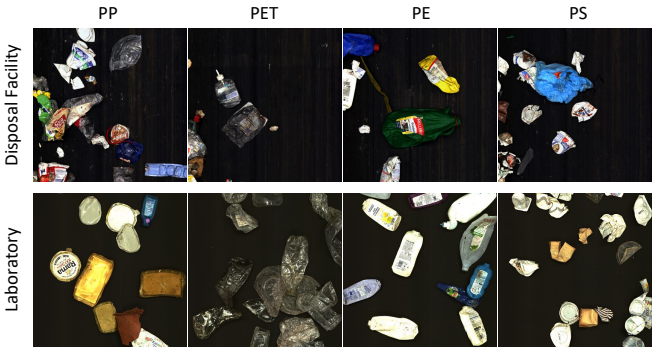


Figure 2: A few samples from the dataset measured in the disposal facility (first row of images) and from the lab (second row of images). The pictures shown here are from the RGB camera.

turing 640 spatial pixels and 224 channels covering wavelengths from 900 nm to 1700 nm. To ensure proper illumination in the NIR range, 12 halogen lamps, each with a power rating of 120 W, were employed. For both setups, we utilized the same illumination and camera type (same producer and product name). The recordings differ in the background, with a cleaner conveyor belt used in the lab and a higher recording speed in the disposal facility. The conveyor belt operated at approximately 1 m/s in the disposal facility, compared to 0.2 m/s in the lab.

We perform white-balancing on all spectra following the method used in [5]. Averaged spectra are depicted in Figure 3. It can be seen that offset and scale of spectra varies when comparing the two datasets. The key spectral features defined by the position of peaks and valleys are preserved.

3.3 Methods

The algorithms investigated in this study are common chemometric methods, namely PLS-DA, LDA, and SAM as well as artificial neural networks. The goal is to predict the class (PP, PET, PE, or PS) for an unknown sample.

All methods can be used with different preprocessing steps and parameters, which will be described for each method briefly.

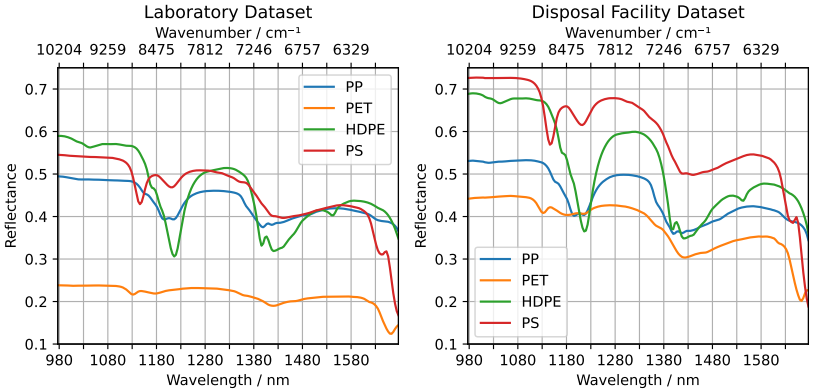


Figure 3: Averaged spectra for different polymer types. The data from the laboratory is shown on the left and from the disposal facility on the right. Given reflectance values are relative reflectance in the range from 0 to 1.

PLS-DA and LDA

PLS-DA and LDA are closely related since both are linear classifier. The methods are commonly used in combination with the preprocessing methods standard normal variate correction (SNV) and differentiation of the spectra (often approximated by Savitzky-Golay filter). Detailed explanations of PLS-DA and LDA can be found in [7].

SAM

Spectral angle mapper (SAM) is less famous than PLS-DA and LDA, but an effective technique for classifying NIR spectra. It is based on the angle between the sample spectrum to precalculated prototypes, whereby the prototypes are mean spectra of each category.

For given spectra a and b, the angle of inclination can be calculated using the formula

$$\theta_{\mathrm{SAM}} = \cos^{-1}\left(\frac{a \cdot b}{\|a\|\,\|b\|}\right),$$

with $a \cdot b$ being the dot product [8]. In this study, SAM worked best

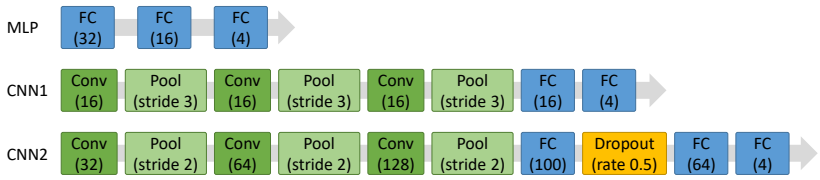


Figure 4: Model architectures of investigated artificial neural networks with a multilayer perceptron (MLP) at the top. In the middle and bottom are convolutional neural networks, whereby CNN2 is adapted from [6].

in combination with the second derivative of the spectral data using Savitzky–Golay filter.

Artificial neural network

A multilayer perceptron and two convolutional neural networks (CNN1 and CNN2) are also investigated in this study and belong to the more recently discovered methods for spectral analysis. The multilayer perceptron consists of fully connected layers only, whereas the CNNs are built using convolutional, pooling, and fully connected layers. The architecture of CNN2 corresponds to the one proposed by Xia et al. [6], since it showed very promising results in the underlying study. Figure 4 presents all investigated architectures.

4 Results

The results are presented by first comparing the models accuracies on the laboratory test data. Then, the accuracies on the data from the sorting facility are shown, before we compare the capability of detecting outliers for each model.

4.1 Results on laboratory dataset

For evaluation, all models have been trained on 85% of the dataset recorded in the laboratory. Then, the models were tested on the remaining 15% of the laboratory dataset (from now on referred as lab

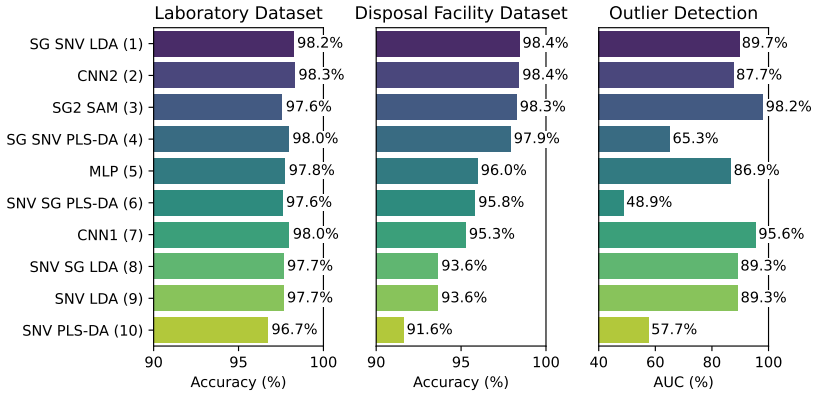


Figure 5: Accuracies and AUC for all methods on unseen data. LDA, PLS-DA, and SAM were combined with preprocessing methods applied in the order stated in the models name. Outlier detection performance is evaluated using the area under ROC curve (see section 4.3).

data) and on 100% of the dataset from the disposal facility (from now on referred as fab data).

The previously described models are used together with different preprocessing methods. We evaluated 6 different variants of LDA and PLS-DA, one SAM, one MLP, and two CNNs. In Figure 5 each method is listed together with the preprocessing steps being Savitzky–Golay filter estimating first derivative (SG), Savitzky–Golay filter estimating second derivative (SG2), and/or standard normal variate (SNV). The order of names corresponds to the order of application. For this analysis better results were achieved by performing SNV after Savitzky–Golay filtering.

Accuracies of the models are presented on unseen data only. The models are sorted according to its accuracy on fab data. On the left side of Figure 5 it can be seen that the accuracies of all models on lab data are about the same, except for model (10), though it is with 96.7% only 1.5 percentage points below the best (CNN2) with 98.3%.

4.2 Results on sorting facility dataset

On fab data there is a wider range of accuracies as shown in the middle of Figure 5. Savitzky–Golay filter, SNV and LDA reached together with CNN2 the highest accuracy of 98.4%. Models (3) and (4) are with 98.3% and 97.9% respectively not significantly lower. Model (10) has the worst performance in this comparison on both datasets. Therefore, we conclude that PLS-DA should only be used here in combination with Savitzky–Golay filter or a similar preprocessing method. Interestingly, there is no indication that LDA or PLS-DA outperforms the others in general, instead, it depends on the preprocessing steps, if the model performs well or not compared to the others. For LDA and PLS-DA the best results were achieved by first applying Savitzky–Golay filer and then SNV.

The artificial neural networks CNN2 (2), MLP (5), and CNN1 (7) reached very high accuracies with 97.8% up to 98.3% on lab data, being among the top models. But their accuracies dropped by about 2 percentage points on fab data, except for CNN2, which achieved the highest accuracy together with model (1).

4.3 Outlier detection performance

For sensor-based sorting of post-consumer LWP waste, it is not only necessary to predict the polymer category, but also, to identify foreign material. This is due to the presence of all kind of materials in post-consumer LWP, including items that do not belong to the LWP fraction. For an automated sorting scenario, it is also important to have a background class, which indicates that there is no object underneath the sensor and no actuators have to be activated. For this reason, we set up another spectral dataset of foreign objects (paper, wood, metal, and glass) and background (conveyor belt without samples).

We took the models that were trained on lab data and extended it by applying a threshold to the score values. If the score of every class is below this threshold, the model outputs -1 instead of the class index of the polymer (which would be 0 for PP, 1 for PET, 2 for PE, and 3 for PS). For the LDA, MLP, and CNNs we used the posterior probabilities as scores. For the PLS-DA we used the regression outputs and similarity scores for SAM. The models class prediction can easily be derived from

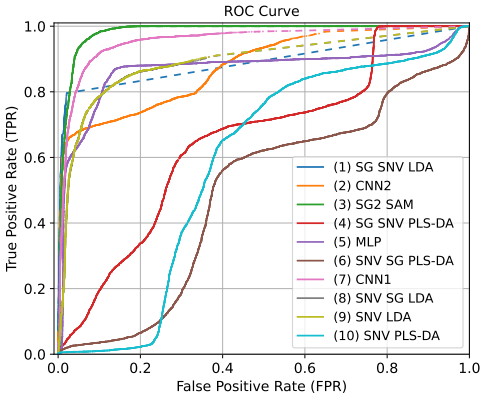


Figure 6: Receiver Operating Characteristic (ROC) curves for all methods on dataset from disposal facility. Dotted lines connect neighboring data points and indicate that there are no actual data points at these positions.

these scores by finding the class index with the highest score.

For evaluating the anomaly detection, we used 50% of normal polymer spectra, 25% foreign material and 25% of background captured in the disposal facility. Both foreign material and background are considered as anomaly, whereby the models were not trained on these samples. Instead, they detect the anomalies based on the threshold as described before. The models are expected to output -1 if there is an anomaly and a number greater or equal to 0 otherwise.

For this binary decision, we calculated Receiver Operating Characteristic (ROC) curves as shown in Figure 6. As the threshold of a model is increased, it will predict more of the spectra as anomaly resulting in a better true positive rate (TPR). However, this also increases the false positive rate (FPR). A perfect model should detect all anomalies (TPR=1), while there are no false positives (FPR=0). SAM (3) gets closest to this point in Figure 6, followed by CNN1 (7), which can be seen as second best. Model (1) being the best model in terms of accuracy on fab data, is performing very well until a TPR of 0.8. For a TPR of 0.8, the threshold for model (1) was set to $1 - 2.22 \cdot 10^{-16}$. Further increasing the threshold is numerically barely possible without resulting in a

threshold of 1, which would classify all spectra as anomaly resulting in the point at the top right corner. This is why the curve is printed in dotted line from $TPR = 0.8$ up to 1 for model (1). This numerical issue is highlighting the fact that model (1) is way to certain about some of the foreign materials belonging to one of the trained polymers.

A common way to evaluate the ROC curve using a single number is the area under ROC curve (AUC), which is added to Figure 5 on the right hand side. The AUC underline the conclusions made based on Figure 6. Model (3) and (7) are among the best two with 98.2% and 95.6%, respectively. All other models show an AUC below 90%.

5 Conclusions and outlook

We have shown that LDA, CNN2, SAM, and PLS-DA being trained on the laboratory dataset were able to reach top accuracies of approximately 98% on the disposal facility dataset. For that, the best performing LDA and PLS-DA were achieved by preprocessing the spectra with Savitzky–Golay filter and then with SNV. By introducing foreign material and background spectra, the models were tested on their capability of outlier detection. Based on thresholding predicted class scores, the outliers are identified without any further training. Out of the best 4 models of the previously mentioned results, only SAM was able to reach a very good AUC score of 98.2%. In contrast to that, the other 3 models showed an AUC below 90%. In conclusion, SAM demonstrated the best overall results. To enhance outlier detection capabilities of the other models, exploring alternative methods beyond thresholding of classifiers scores could yield improvements. This presents an opportunity for future research in this area.

Acknowledgement

We would like to thank Lobbe RSW GmbH for the opportunity and assistance in recording a dataset at their waste disposal facility in Iserlohn, Germany. Moreover, we thank Steffen Rüger and his team from the Fraunhofer Institute for Integrated Circuits (IIS) for their support in dataset annotation. Funding was provided by the Federal Ministry of Education and Research (BMBF) under the funding reference 033KI201.

References

1. J. Yang, Y.-P. Xu, P. Chen, J.-Y. Li, D. Liu, and X.-L. Chu, "Combining spectroscopy and machine learning for rapid identification of plastic waste: Recent developments and future prospects," *Journal of Cleaner Production*, vol. 431, p. 139771, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095965262303929X
2. O. Tamin, E. Moung, J. A. Dargham, F. Yahya, and S. Omatu, "A review of hyperspectral imaging-based plastic waste detection state-of-the-arts," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 3, pp. 3407–3419, 2023. [Online]. Available: https://ijece.iaescore.com/index.php/IJECE/article/view/27804
3. M. Mehrubeoglu, A. V. Sickle, and J. Turner, "Detection and identification of plastics using SWIR hyperspectral imaging," in *Imaging Spectrometry XXIV: Applications, Sensors, and Processing*, E. J. Ientilucci and P. Mouroulis, Eds., vol. 11504, International Society for Optics and Photonics. SPIE, 2020, p. 115040G. [Online]. Available: https://doi.org/10.1117/12.2570040
4. Y. Zheng, J. Bai, J. Xu, X. Li, and Y. Zhang, "A discrimination model in waste plastics sorting using nir hyperspectral imaging system," *Waste Management*, vol. 72, pp. 87–98, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0956053X17307602
5. X. Wu, J. Li, L. Yao, and Z. Xu, "Auto-sorting commonly recovered plastics from waste household appliances and electronics using near-infrared spectroscopy," *Journal of Cleaner Production*, vol. 246, p. 118732, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652619336029
6. J. Xia, Y. Huang, Q. Li, Y. Xiong, and S. Min, "Convolutional neural network with near-infrared spectroscopy for plastic discrimination," *Environmental Chemistry Letters*, vol. 19, no. 5, pp. 3547–3555, 2021.
7. M. Bevilacqua and F. Marini, "Local classification: Locally weighted–partial least squares-discriminant analysis (lw–pls-da)," *Analytica Chimica Acta*, vol. 838, pp. 20–30, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003267014007557
8. L. Madan Kumar, B. Pavan, P. Kalyan, N. S. Paul, R. Prakruth, and T. Chinnu, "Design of an embedded based control system for efficient sorting of waste plastics using near infrared spectroscopy," in *2014 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2014, pp. 1–6.

MIR measurements combined with photon-up-conversion technology to measure and identify black polymers

Wolfgang Becker, Kerstin Sachsenheimer, and Melanie Klemenz

Fraunhofer ICT,
Joseph-von-Fraunhofer-Straße 7, 76327 Pfinztal

Abstract Identifying black polymers, which typically contain 0.5 to 3 mass percent of soot or black masterbatch, remains a considerable challenge in recycling sorting processes. Near-infrared spectroscopy (NIRS) is well-suited for the rapid and reliable identification of non-black polymers, making it a preferred choice for industrial sorting. However, due to the complete absorption of light in the NIR spectrum by black polymers, NIRS proves ineffective for such materials. Mid-infrared (MIR) spectroscopy offers a promising alternative for detecting black polymers. While conventional Fourier-transform infrared (FTIR) systems in the MIR range provide high accuracy, they lack the speed required for large-scale sorting operations. This paper introduces an innovative spectrometer system utilizing photon up-conversion, enabling the high-speed detection of black polymers and presenting a viable solution for industrial applications. Experimental results highlight the system's potential for efficient sorting of black polymers, addressing a critical gap in recycling technology.

Keywords Identification of black polymers, middle infared, photon up conversion, black plastic sorting

1 Introduction

Black polymers are extensively used in various consumer products, including electronics, automotive parts, and IT equipment. Their characteristic color results from adding approximately 0.5 to 3 mass percent of

carbon black or black masterbatch during production. Recycling these materials offers significant environmental and economic benefits, yet current methods for identifying black polymers are still under development. For industrial applications, these methods need to be efficient, reliable, and easy to implement.

NIRS (Near-Infrared Spectroscopy) is currently a preferred technique for identifying non-black plastics, especially common household waste, due to its accuracy and efficiency. However, NIRS is ineffective for black polymers because the carbon black content absorbs nearly all light in the NIR spectrum, making material identification impossible. By contrast, MIR (Mid-Infrared) spectroscopy has shown potential for characterizing black polymers [1–4]. This approach can partially overcome the absorption challenges by utilizing different wavelengths, although it is also impacted by the composition and structure of the material.

Black polymers differ from typical household plastics, as they are primarily used in technical applications requiring specific properties. These properties are often achieved by incorporating various additives and fillers, which complicates spectral identification. Even among polymers of the same type, spectra can vary widely due to differences in additive composition, such as flame retardants, reinforcing fibers, or levels of carbon black. Additionally, surface treatments or coatings, including layers for color, protection, or electromagnetic compatibility (EMC), can further interfere with spectral analysis and may need to be removed before analysis.

Interest in identifying black polymers has led to the exploration of several IR techniques, such as specular or diffuse reflection [1,3,5], Attenuated Total Reflectance (ATR) [5,6], IR transmission [5,7] and emission spectroscopy [8]. Reflectance measurements have been studied in multiple research projects, showing promise for accurately analyzing soot-filled polymers [9–11]. In recent developments, researchers have explored alternative spectral techniques to further enhance polymer identification. One approach is hyperspectral imaging, which combines spatial and spectral data, allowing for detailed chemical mapping and identification of complex polymer blends [12]. But up to now hyperspectral techniques are using NIR spektroscopy and FTIR is not realized up to now. Additionally, fluorescence spectroscopy is being investigated for detecting certain polymers with unique fluorescent

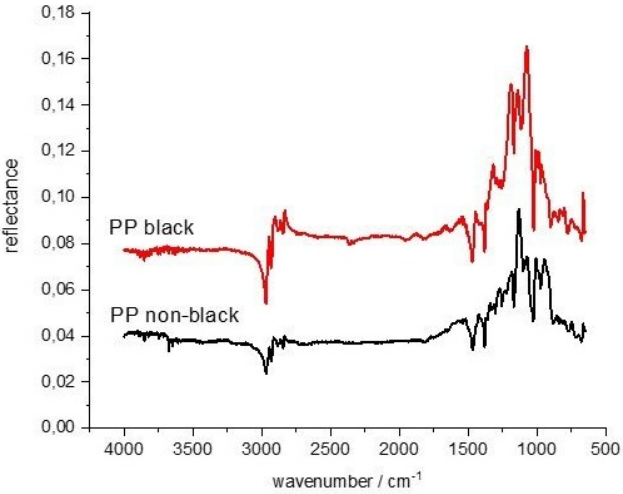


Figure 1: MIR spectra of a black and a non-black PP polymer sample

markers, providing another method for polymer sorting and identification [13]. An example of a polypropylene polymer is shown in Figure 1. One significant limitation of MIR spectroscopy, however, is that it works effectively only with non-conductive polymers. If a polymer is electrically conductive, MIR light is fully absorbed, resulting in unusable spectra without detectable signals.

The importance of recycling black polymers will increase in the future, as these materials are commonly found in devices such as smartphones and other electronic equipment containing valuable elements, including rare earth metals. Rare earth elements, critical to modern technologies, are used in components like magnets, batteries, and displays. Efficient recycling of these materials is essential to meet growing demand, minimize environmental impact, and promote sustainability. Improved recycling methods for black polymers could significantly contribute to the conservation of these valuable resources, supporting efforts toward sustainable development and resource management in the face of increasing demand for technology.

A summary or overview of the current issues in plastic identification within sorting facilities and measurement techniques for polymer

detection and classification is provided in [14]. This also includes a discussion of the future challenges in this area.

1.1 Principles of identification of polymer samples

The primary method for identifying industrial plastics is near-infrared (NIR) spectroscopy. Detector systems based on indium gallium arsenide (InGaAs) are particularly well-suited for this purpose, as they offer both the sensitivity and the speed necessary for sorting applications. These dispersive spectrometer systems are designed without moving parts, which makes them resistant to mechanical vibrations and simplifies their protection against dust. Advanced systems can capture measurement rates of up to tens of thousands of spectra per second, enabling economical operation in sorting facilities.

The NIR wavelength range is from 780 nm to 2500 nm , which corresponds to a wavenumber range of approximately 12800 cm^{-1} to 4000 cm^{-1}. Due to the different detectors required, the NIR range is often divided into two sections. Silicon-based sensors are primarily used for the short near-infrared (SNIR) range, covering wavelengths from 780 nm to about 1050 nm , while InGaAs sensors are predominantly applied to the "classical" NIR range from 1050 , nm to 2500 nm . The NIR range is characterized by overtone and combination bands associated with C-H, N-H, O-H, and C-O functional groups. Identification of plastics relies heavily on the stretching vibration modes of CH, CH_2, and CH_3 groups, with the second overtone found around 1.1 µm to 1.25 µm, the first overtone around 1.65 µm to 1.7 µm, and combination bands within these ranges.

One significant limitation of plastic identification with NIR spectroscopy is its inability to detect black plastics. The soot within these materials absorbs all NIR light, resulting in no detectable signal and thus no material differentiation.

In the mid-infrared (MIR) spectral region, spanning wavelengths from 2.5 µm to approximately 16 µm (or a wavenumber range from 4000 cm^{-1} to around 600 cm^{-1}), plastics display additional vibrational modes — such as deformation, rocking, and twisting modes — arising from their molecular structure. Besides the C-H groups, molecular structures like O-H, N-H, and C-O also contribute to the spectral characteristics through fundamental vibrations. These distinct molecular

groups and their specific vibrational modes produce a unique spectrum for each polymer, which allows for accurate identification. This spectral region, often known as the "fingerprint region," makes MIR spectroscopy a leading analytical technique for the analysis and characterization of polymers.

An added benefit of the MIR spectral range is that it enables the identification of black polymers. By measuring reflectance spectra, it is possible to distinguish black polymers, as illustrated in Figure 1, which compares the reflectance spectra of a black polypropylene (PP) sample with those of a non-black sample. The spectra reveal the primary spectral features, and the filler material appears to have minimal impact on the polymers' optical properties. Thus, MIR spectroscopy proves to be an effective technique for identifying black polymers.

A limitation in the MIR range, however, is that no measurable signal can be obtained from electrically conductive polymers. This occurs due to the presence of a free electron system within the polymer material, which interferes with signal detection in this spectral region.

1.2 Non-linear photon up conversion

In the photon up conversion process, infrared (IR) photons are converted into higher-energy photons in the near-infrared (NIR) range, specifically within the silicon spectral range, through sum-frequency generation. In this mechanism, a mid-infrared (MIR) photon with energy

$$\omega_{\mathrm{MIR}} + \omega_{\mathrm{LS}} = \omega_{\mathrm{SF}} \tag{1}$$

This process has long been recognized as an effective method for high conversion of signals [15–17]. By employing periodically poled non-linear crystals (PPLN) made of lithium niobate oxide, and through proposals to conduct the high conversion directly within the laser resonator to amplify the laser power, conversion efficiencies exceeding 10% can be achieved. However, beyond energy conservation as described by eq.1, phase matching of the light waves within the crystal is essential. This requirement can be formulated in the k-space with k-vectors and in the one dimensional case it corresponds to eq. 2

$$\frac{1}{\Lambda_{PP}} + \frac{n_{LS}}{\lambda_{LS}} + \frac{n_{MIR}}{\lambda_{MIR}} = \frac{n_{SF}}{\lambda_{SF}} \tag{2}$$

Λ_{PP} is the polarization period of the conversion crystal and the different λ_i are the wavelengths of the interacting waves. The n_i are the refractive indexes of the media at the respective wavelength. The refractive index of the crystal is temperature dependent. By choosing a temperature and the polarization period of the crystal the wavelength range of the converter is determined. More details and applications of photon up conversion can be found in [18].

2 Measurement setup

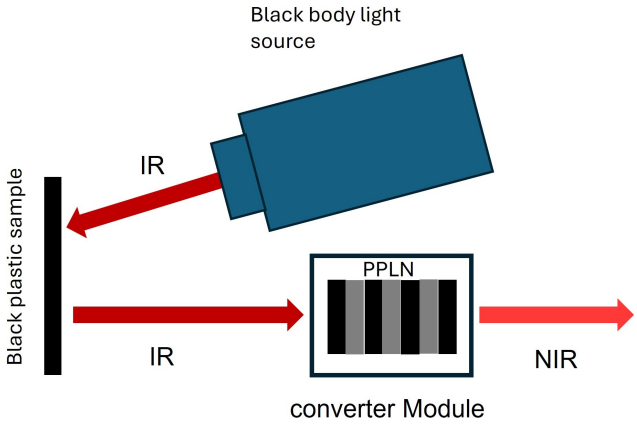


Figure 2: measurement setup with the Litran converter and a black plastic sample. A black body was used as a IR light source with temperatur of 900°C. The Litran system was constructed and built by the Fraunhofer Institut für Physikalische Messtechnik, Fh-IPM, Freiburg (Germany)

The measurement setup utilized for analyzing the plastic samples is depicted in Fig.2. An infrared (IR) light source, specifically a black body radiator operating at 900°C, was employed. The samples consisted of 5 mm thick plastic plates, which were non-electrically conductive. This characteristic is crucial, as electrically conductive black

plastic samples absorb all radiation in the MIR spectral range, rendering them unsuitable for characterization.

The black body source irradiated the sample, and the reflected light was directed through the converter module containing the PPLN crystal. The resulting NIR signal was detected using a single-photon avalanche diode (SPAD) area sensor. This sensor was developed as part of the Fraunhofer Gesellschaft's LITRAN research project, conducted by the Fraunhofer Institut für integrierte Schaltungen, (Fh-IIS), Erlangen (Germany).

The measurement duration for each sample ranged between 0.3 and 0.5 seconds. For reference, IR measurements were performed using an uncoated aluminum mirror. Reflectance values were determined according to the corresponding equation.

$$R = \frac{I_S - I_N}{I_R - I_N} \tag{3}$$

- I_S is the sample intensity spectrum, representing the measured intensity from the sample under investigation.
- I_R is the intensity of the reference, typically measured using a standard material like an uncoated aluminum mirror.
- I_N is the wavelength-dependent noise and the dark current spectrum, which includes background signals and detector noise contributions.

3 Results

The poling period of the PPLN crystal was Λ_{PP} = 21.34 µm and the temperature of the PPLN crystal was 50.02 °C for all measurements. The measured NIR spectra were then assigned to wavelengths in the MIR. A black polypropylene (PP) sample was analyzed using the photon up-conversion system, which is schematically depicted in Fig.2 and shown in a photograph in Fig.5.

The poling period of Λ_{PP} = 21.34 µm allowed for spectral measurements within the the wavelength rande of 3.2 µm to 4.15 µm wavelength range. A single measurement was conducted without averaging.

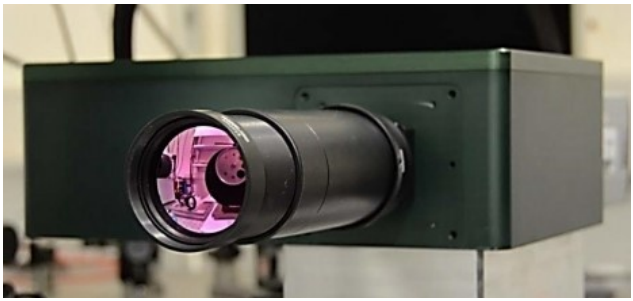


Figure 3: Photo of the Litran system which was constructed and built by the Fraunhofer Institut für Physikalische Messtechnik, Fh-IPM, Freiburg (Germany)

For comparative purposes and to assess the LITRAN system's spectral quality, the plastic samples were also examined using a commercial FTIR spectrometer. The resulting spectra are presented in Figure .

The FTIR spectra were acquired at a resolution of $4\,\mathrm{cm}^{-1}$, with specific reflection peaks corresponding to the functional groups in the PP polymer. Upon comparison, it is evident that the LITRAN system exhibits a lower spectral resolution than the FTIR measurements. Although the reflection peaks in the LITRAN spectra are broader, they remain discernible and can be utilized for identifying and distinguishing various types of plastics. In the following section the measurements with the photon up-conversion system are referred to as LITRAN (from Light Transformation) and spectra measured with the Fourier spectrometer will be assigned FTIR. This will simplify the presentation.

Functional groups of the PP polymer are assigned to the corresponding reflection peaks. By comparing the two spectra one can see that the spectral resolution of the LITRAN system is lower compared to the FTIR measurements. The reflection peaks are wider but present and can be used for the recognition and differentiation of other types of plastic.

4 Conclusion

MIR spectroscopy has proven to be an effective technique for the spectral characterization of black plastic components and their classifica-

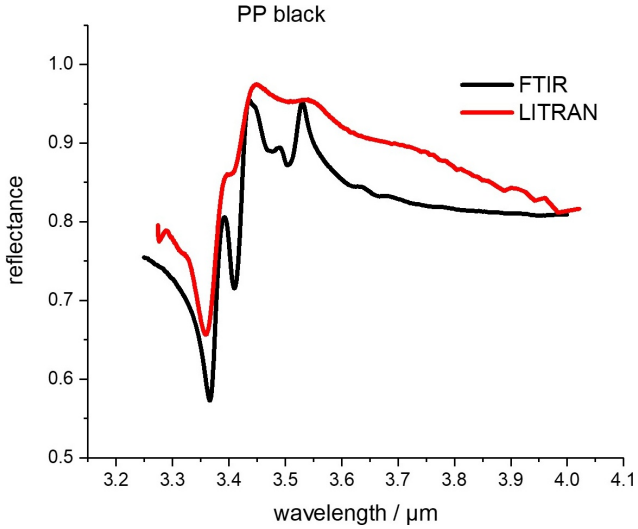


Figure 4: MIR spectrum of PP black sample measured with LITRAN system and with FTIR spectrometer. Spectra are given in wavelength.

tion into corresponding polymer types. Using the LITRAN system, reflectivity measurements in the spectral range between 3.2 µm and approximately 3.6 µm enable the identification of pure plastic grades as well as blends, which can even be traced to specific manufacturers. For re-granulates, the proportions of base materials can theoretically be quantified through appropriate calibration models. These capabilities are essential for ensuring precise sorting and producing high-quality recycled black plastic materials.

The LITRAN system supports the economical sorting of black plastics, as it can achieve measurement rates of several thousand measurements per second, enabling sorting systems with adequate mass throughput. For engineering plastics, the identification of potential polymer additives and mixtures is also critical. Many engineering plastics, for instance, contain flame retardants, some of which are restricted and must be removed prior to re-granulation. The LITRAN system's spectral range can be customized to detect such substances.

Additionally, the LITRAN system offers the potential to enhance ex-

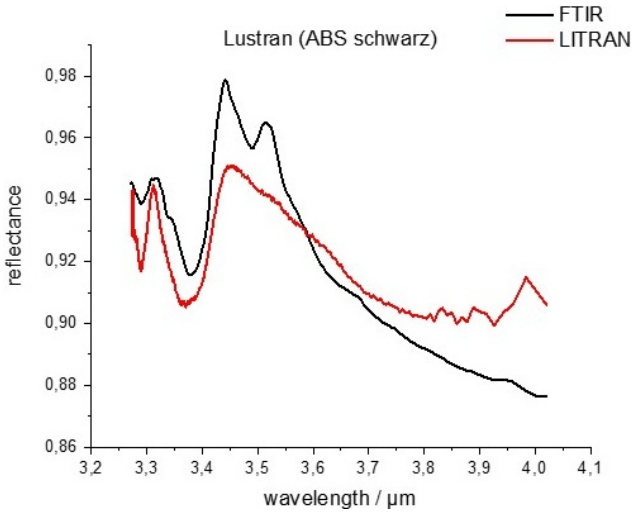


Figure 5: MIR spectrum of ABS (brand name Lustran) black sample measured with LITRAN system and with FTIR spectrometer. Spectra are given in wavelength

isting NIR-based sorting systems by extending them to the MIR range. In the MIR range, absorption bands are 10 to 100 times stronger than in the NIR range, allowing the system to sort standard non-black household plastics effectively. However, achieving this would require measurement rates of several tens of thousands to hundreds of thousands per second. The optical components and measurement setup of the LITRAN system are, in principle, capable of broadening the application of photon upconversion to additional industrial processes.

References

1. Becker, W. and Eisenreich, N., "Identification of black polymers in the MID-infrared spectral region with an FTIR-spectrometer," *CARE Innovation*, pp. 88–95, 1998.
2. *Becker, W. Using Spectroscopic Methods for Identification of Technical Plastics from E & E and Elv Sector. In Proceedings of the Going Green—Care Innovation Conference, Vienna, Austria, 13–16 November 2006*, 2006.

3. Becker, W., *Prozessanalytik: Strategien und Fallbeispiele aus der industriellen Praxis*, 1st ed. Weinheim: Wiley-VCH, 2006.
4. N. Eisenreich and T. Rohe, "Infrared spectroscopy in analysis of plastics recycling," in *Encyclopedia of Analytical Chemistry*, R. A. Meyers, Ed. Wiley, 2007.
5. W. W. Hart, P. C. Painter, J. L. Koenig, and M. M. Coleman, "A fourier transform infrared method of studying elastomers filled with carbon black," *Applied Spectroscopy*, vol. 31, no. 3, pp. 220–224, 1977.
6. R. T. Graf, J. L. Koenig, and H. Ishida, "Comparison of ft-ir transmission, specular reflectance, and attenuated total reflectance spectra of polymers," pp. 385–395, 1987.
7. E. F. Devlin, "The fourier transform infrared spectrum of cured, black–reinforced sbr," *Journal of Polymer Science: Polymer Letters Edition*, vol. 19, no. 4, pp. 189–192, 1981.
8. Chalmers, John M. and Mackenzie, Moray W. and Poole, Neil, "Some observations on FTIR emission spectroscopy of black solid samples," *Microchimica acta*, vol. 95, no. 1-6, pp. 249–253, 1988.
9. M. Claybourn, P. Colombel, and J. Chalmers, "Characterization of carbon-filled polymers by specular reflectance," *Applied Spectroscopy*, vol. 45, no. 2, pp. 279–286, 1991.
10. O. Rozenstein, E. Puckrin, and J. Adamowski, "Development of a new approach based on midwave infrared spectroscopy for post-consumer black plastic waste sorting in the recycling industry," *Waste management (New York, N.Y.)*, vol. 68, pp. 38–44, 2017.
11. K. Nogo, K. Ikejima, W. Qi, N. Kawashima, T. Kitazaki, S. Adachi, K. Wada, A. Nishiyama, and I. Ishimaru, "Identification of black microplastics using long-wavelength infrared hyperspectral imaging with imaging-type two-dimensional fourier spectroscopy," *Analytical methods : advancing methods and applications*, vol. 13, no. 5, pp. 647–659, 2021.
12. A. de Lima Ribeiro, M. C. Fuchs, S. Lorenz, C. Röder, J. Heitmann, and R. Gloaguen, "Multi-sensor characterization for an improved identification of polymers in weee recycling," *Waste management (New York, N.Y.)*, vol. 178, pp. 239–256, 2024.
13. C. Olscher, A. Jandric, C. Zafiu, and F. Part, "Evaluation of marker materials and spectroscopic methods for tracer-based sorting of plastic wastes," *Polymers*, vol. 14, no. 15, 2022.
14. T. Werner, I. Taha, and D. Aschenbrenner, "An overview of polymer identification techniques in recycling plants with focus on current and future challenges," *Procedia CIRP*, vol. 120, pp. 1381–1386, 2023.

15. T. R. Gurski, H. W. Epps, and S. P. Maran, "Upconversion of broadband infrared spectra," *Applied optics*, vol. 17, no. 8, pp. 1238–1242, 1978.

16. Q. Hu, J. Seidelin Dam, C. Pedersen, and P. Tidemand-Lichtenberg, "High-resolution mid-ir spectrometer based on frequency upconversion," *Optics letters*, vol. 37, no. 24, pp. 5232–5234, 2012.

17. P. Tidemand-Lichtenberg and C. Pedersen, "Broadband long-wavelength upconversion in ultra-short nonlinear crystals," *Optics letters*, vol. 49, no. 9, pp. 2233–2236, 2024.

18. D. K. Singh, *Modern Techniques of Spectroscopy*. [S.l.]: Springer Singapore, 2021.

Impact of demolition waste powders on the microstructure of cement mortars: A comparative analysis of concrete, ceramic and mixed wastes

Anna Tokareva[1] and Daniele Waldmann[2]

[1] University of Luxembourg, Faculty of Science, Technology and Medicine, Department of Engineering
Avenue de la Fonte 6, L-4364 Esch-sur-Alzette
[2] Technische Universitat Darmstadt, Institut fur Massivbau
Franziska-Braun-Straße 3, 64287 Darmstadt

Abstract The study evaluates the potential of various Construction and Demolition Waste (CDW) powders as Supplementary Cementitious Materials (SCMs) in cement mortars. Four CDW types—concrete screening fines (CS), mixed fines (MS), terracotta (RT), and porcelain (RP)—were used to replace 20% of the cement. Techniques including laser granulometry, XRF, XRD, and SEM analyzed particle size, composition, and microstructure of the materials. Ceramic powders RT and RP showed pozzolanic properties, though RP's surfactant contamination caused a porous structure. The properties of CS and MS were not significantly different. The findings help to understand the importance of separating CDW to use them in resource-efficient and sustainable cement production.

Keywords Construction and demolition waste, supplementary cementitious materials, microstructure, concrete, ceramics

1 Introduction

The rapid growth of the global population has led to a rising demand for the construction of new buildings, infrastructure and roads. Additionally, the demolition of outdated structures to make way for mod-

ern, more efficient buildings is increasingly common. This trend results in the depletion of natural resources and a substantial increase in construction and demolition waste (CDW). Despite the vast quantities generated, about 35% of world CDW is landfilled, contributing to landscape degradation, occupying significant land areas, and causing environmental pollution [1], [2]. However, this waste stream contains valuable recyclable materials that are often underutilized. For instance, concrete waste can be processed to separate the cement matrix from aggregates, followed by thermal treatment of the matrix to dehydrate and reactivate the hydrated cementitious binder [3], [4], [5], [6]. This reprocessed material can partially replace cement clinker in concrete production, thereby reducing the associated carbon dioxide emissions generated during clinker manufacture. Similarly, finely ground ceramic waste has potential as a partial substitute for clinker, acting as a pozzolanic supplementary cementitious material (SCM) [7], [8], [9], further reducing the environmental impact of cement production. However, this requires separation of construction and demolition waste (CDW), which is often a difficult task.

This study compares the results of two previous works [10], [11] and aims to evaluate the necessity of separating demolition waste by comparing the effects of concrete, ceramic, and mixed wastes on the microstructure of the binder matrix in mortar samples. It gives a comparative analysis of the properties of various types of CDW, including concrete screening fines (CS) composed of a separated cement matrix, mixed screening fines (MS) containing 60% of cement matrix with ceramic and glass fines inclusion in amount of approximately 40%, terracotta powder (RT) and porcelain powder (RP).

2 Materials and Methods

2.1 Materials preparation

Concrete-containing materials (CS and MS) were collected from Tradecowall, a Belgian inert waste processing company. The ceramic materials (RT and RP) were sourced from various sites in Luxembourg and France. Following sampling, all materials were dried at 105°C to a constant mass, then crushed in a jaw crusher and milled in an impact mill equipped with a 2 mm mesh bottom sieve. The CS and MS

powders were subsequently sieved through a 125 µm mesh, and the resulting 0/125 µm fraction was calcined in a muffle furnace at 500°C for 2 hours. This process reactivated the hydrated cement phases without CO_2 emissions, as the temperature remained below the calcite decomposition threshold. Basically, the RT and RP powders did not require thermal activation, however, to unsure their pozzolanic activity, they were sieved on a finer sieve of 63 µm mesh.

The processed waste powders were incorporated as supplementary cementitious materials (SCMs) in cement mortar specimens, with 20% of the cement replaced by the respective waste powders. Mortar bars measuring 40×40×160 mm were prepared according to the ISO 679:2009 standard. The water-to-binder (w/b) and sand-to-binder (s/b) ratios for all mixtures were set at 0.5 and 3, respectively.

Given the pozzolanic nature of ceramic powders, which can be considered as calcined clay with admixtures, the microstructure of ternary cement blends, known as Limestone Calcined Clay Cement (LC3), was also analyzed. These LC3 blends comprised 70% Portland cement, 20% ceramic powder, and 10% limestone filler. LC3 mortar bars followed the same w/b and s/b ratios of 0.5 and 3, respectively.

In addition to the waste powders, the study utilized CEM I 52.5 R Portland cement provided by the French manufacturer Vicat, CEN Standard Sand in accordance with EN 196-1 and tap water for the preparation of all mortar specimens.

2.2 Experimental Methods

The characterization of the resulting powders was conducted using a HELOS & RODOS laser granulometer to determine particle size distribution, X-ray fluorescence (XRF) to assess chemical composition, and X-ray diffraction (XRD) to analyze mineralogical composition. Mineralogical analysis was performed with a D2 PHASER X-ray diffractometer under the following conditions: Cu-Kα radiation at 10 mA and 30 kV, with a rotation angle from 5° to 70° in 2θ, a 0.02° step size, and a step time of 0.6 s. Data were processed using Profex software.

The microstructure of cement mortars containing 20% waste powders was examined after 90 days of curing in a moisture chamber at 20 °C using scanning electron microscopy (SEM). For SEM analysis, mortar specimen debris was coated with a 4 nm layer of gold and examined

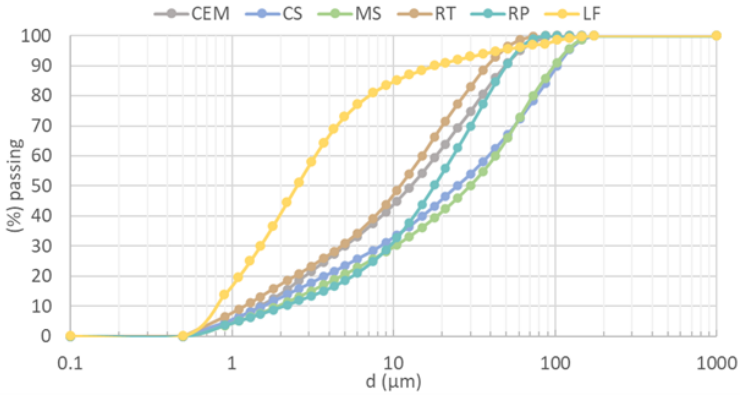


Figure 1: Particle size distribution of the powders used in the study.

with a JEOL JSM-6010LA field emission scanning electron microscope.

3 Results and Discussion

3.1 Particle Size Distribution

The results of laser granulometry shown on the Figure 1 revealed similar particle size distributions for both concrete-containing wastes, CS and MS. However, despite undergoing the same preparation procedure, the ceramic powders, RT and RP, exhibited significantly different granulometric compositions. The coarser particle size distribution of the porcelain powder (RP) can be attributed to its higher hardness compared to terracotta (RT), which makes it more difficult to grind. Overall, due to the finer sieve used during processing, the ceramic powders have a finer particle structure than the cement-containing powders.

3.2 Chemical Composition

XRF analysis (Table 1) revealed that the predominant oxide in the composition of MS, RT, and RP was SiO_2, with concentrations of 41.94%, 50.64%, and 64.38%, respectively, whereas the major oxide in CS was

CaO, at 33.29%. Both ceramic powders contained sufficient amounts of silicon, aluminum, and iron oxides to meet the 70% requirement for pozzolanic SCMs after ASTMC618 standard. However, the presence of ceramic and glass impurities in MS did not contribute enough pozzolanic oxides to the mixed waste screening powder. Consequently, the sum of pozzolanic oxides (SiO_2, Al_2O_3, and Fe_2O_3) was 76.72% for RT, 85.74% for RP, and only 55.55% for MS. It can be noted that the oxide composition of MS is more similar to the composition of CS than to that of ceramic wastes.

Table 1: Oxide composition of the investigated materials, mass-%.

	SiO_2	Al_2O_3	Fe_2O_3	CaO	MgO	TiO_2	MnO	Na_2O	K_2O	P_2O_5	SO_3	LOI*
CEM	16.07	3.91	3.58	66.72	1.45	0.37	0.08	0.26	1.18	0.39	3.88	2.10
CS	28.42	5.90	4.12	33.29	1.63	0.45	0.12	0.35	1.25	0.17	0.80	23.50
MS	41.94	8.10	5.51	21.24	1.11	0.77	0.11	0.47	2.15	0.44	2.56	15.60
RT	50.64	17.66	8.42	10.32	1.51	1.08	0.19	0.39	3.37	0.23	3.28	2.90
RP	64.38	19.14	2.22	6.74	0.45	1.36	0.04	0.64	3.03	0.08	0.21	1.70
LF	8.54	1.96	1.06	47.68	1.81	-	-	0.12	0.37	-	0.1	38.36

*Loss on ignition at 950 °C

3.3 Mineralogical Composition

XRD patterns given on the Figure 2 showed that the mineralogical composition of all powders was primarily dominated by quartz. The cement-containing wastes also exhibited significant amounts of calcite, originating from both the aggregates in the concrete and carbonation of the cement, as well as feldspars. In contrast, the composition of the ceramic wastes differed more markedly from each other. RT contained high amounts of feldspars, while RP exhibited a notable presence of mullite, which was absent in terracotta. This distinction is attributed to the different firing temperatures used in the production of these two types of ceramics.

Notably, the diffraction patterns of both ceramic powders displayed a pronounced amorphous halo (Figure 3), indicative of a high glass phase content, which is crucial for pozzolanic activity. The cement-containing materials also exhibited a glass phase, but their amorphous halo is less pronounced.

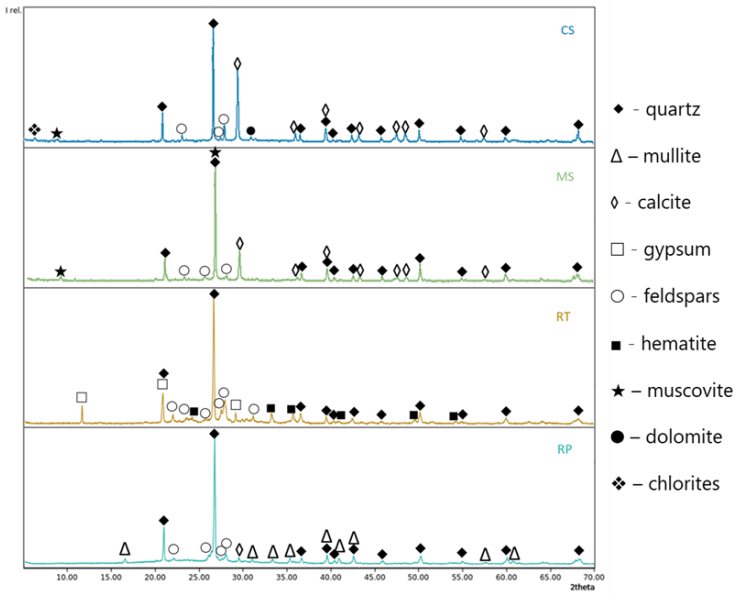


Figure 2: XRD patterns of the investigated waste powders.

3.4 Microstructure

Figure 4a shows SEM images at 200x magnification, revealing a dense matrix structure in mortar samples containing 20% CS, MS, and RT. In contrast, the samples with RP exhibited a more porous structure. At higher magnification (1000x; Figure 4b), the RT-containing mortars demonstrated a notably dense microstructure, while those with CS and MS displayed a moderately loose microstructure. The RP-containing samples exhibited a highly porous microstructure with numerous voids.

The compressive strength test, considered the most critical evaluation of cement quality, aligns with the SEM microstructural analysis. According to previous studies [10], [11], the mortars containing CS and MS, which exhibited a relatively dense microstructure, achieved compressive strengths after 90 days of curing that were only 10% and 12% lower, respectively, than that of the reference mortar without SCMs.

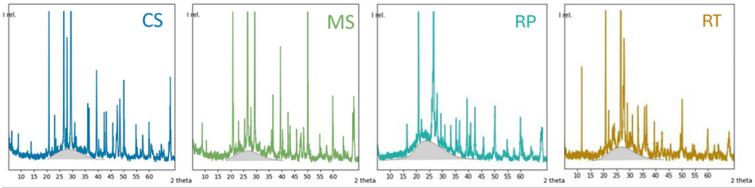


Figure 3: Amorphous halo on the XRD patterns.

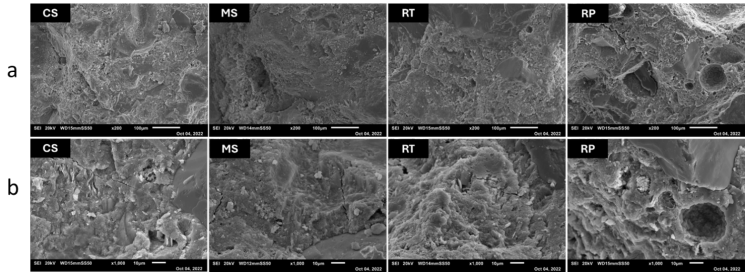


Figure 4: SEM pictures of mortars containing 20% SCMs: a – magnification 200x, b – magnification 1000x.

These results suggest that CS and MS are acceptable for use in structures not requiring high-strength concrete. The RT mortars, exhibiting the densest microstructure, demonstrated the best performance, with compressive strength 3% higher than the reference sample. In contrast, RP mortars, characterized by a porous structure with numerous voids, showed the poorest performance, with compressive strength 37% lower than the reference. These results suggest that RT can be used in the production of high-strength concrete, while untreated RP powder is unsuitable as an SCM.

The dense microstructure and good mechanical properties observed in RT-containing mortars indicate favorable pozzolanic activity of the material. However, the porous structure of RP-containing mortars is attributed to contamination of the sanitary porcelain with surfactants during its lifespan, as confirmed by foaming tests. Calcination of RP powders at 500°C eliminated the contaminants and significantly improved the mechanical properties of mortars containing 20% RP, re-

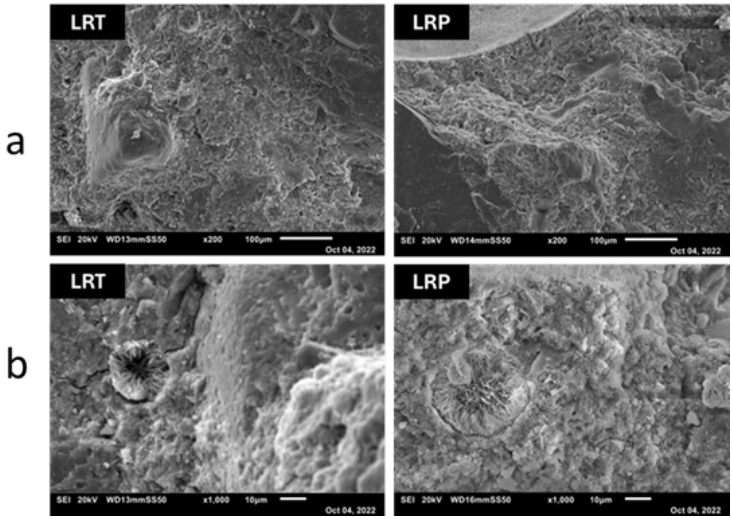


Figure 5: SEM pictures of LC3 mortars with ceramic waste powders: a – magnification 200x, b – magnification 1000x.

ducing porosity and enhancing compressive strength.

The pozzolanic properties of ceramic powders suggest their potential application in the production of highly sustainable LC3 cement. SEM images of samples containing 20% ceramic powders and 10% limestone filler (Figure 5) revealed pore filling with secondary ettringite. This phenomenon has a dual effect on the quality of the hardened binder. On one hand, pore filling compacts the structure, increasing its density and improving mechanical properties. On the other hand, the growth of secondary ettringite can induce internal stresses, potentially leading to cracking, which negatively impacts the material's durability. A more detailed investigation of this phenomenon is needed to assess its implications for long-term performance.

4 Conclusion

This study demonstrated that the properties of pure concrete waste (CS) and mixed concrete-ceramic waste (MS), as well as the microstruc-

ture of cement mortars containing 20% of these materials, are comparable. These findings suggest that ceramic and glass inclusions have minimal impact on the performance of concrete SCMs. Consequently, the separation of concrete waste from ceramic materials may not be necessary for producing cement intended for low-strength structures.

In contrast, only pure ceramic waste appears suitable for the production of high-strength cement with a dense microstructure. It is essential to consider the conditions under which the original ceramics were used, particularly for materials like sanitary porcelain, which may require washing or calcination to remove surfactants.

Further studies are needed to assess the life cycle impact of the investigated SCMs and evaluate the economic feasibility of separating and purifying ceramic waste for cement production. Additionally, more detailed research on the microstructural development of LC3 based on ceramic waste is required, particularly to understand the effect of secondary ettringite growth in pores on the properties of hardening cement.

Acknowledgements

We sincerely thank Mr. Marc Seil, Mr. Ed Weyer, Mr. Gilbert Klein and Mr. Mehdi Saeidi from the Solid Constructions Laboratory at the University of Luxembourg for their technical support. We are also very grateful to Mr. Robin Chapelle and Mr. Nicolas-Alexandre Eugene from Tradecowall company (Belgium), to Mr. Patrick Copus from Polygone Sàrl (Luxembourg) and to Mr. Cedric Langard from Xardel Demolition (Luxembourg) for providing the researched materials, Dr. Claude Simon and Mr. Sven Jung from Cimalux company (Luxembourg) for providing cement and technical assistance in laser granulometry analysis, Dr. Neven Ukrainchyk from the Technical University of Darmstadt for technical assistance with XRF and XRD analyses, and Mrs. Zornitza Tosheva from the Department of Physics and Materials Science at the University of Luxembourg for the training of the SEM use.

References

1. I. Papamichael, I. Voukkali, P. Loizia, and A. A Zorpas, "Construction and demolition waste framework of circular economy: A mini review," *Waste Management & Research: The Journal for a Sustainable Circular Economy*, vol. 41, no. 12, pp. 1728–1740, 2023.
2. K. Kabirifar, M. Mojtahedi, C. Wang, and V. Tam, "Construction and demolition waste management contributing factors coupled with reduce, reuse, and recycle strategies for effective waste management: A review," *Journal of Cleaner Production*, vol. 263, no. 121265, August 2020.
3. L. Xu, J. Wang, K. Li, S. Lin, M. Li, T. Hao, Z. Ling, D. Xiang, and T. Wang, "A systematic review of factors affecting properties of thermal-activated recycled cement," *Resources, Conservation and Recycling*, vol. 185, no. 106432, October 2022.
4. A. Carrico, J. Bogas, S. Hu, S. Real, and M. Costa Pereira, "Novel separation process for obtaining recycled cement and high-quality recycled sand from waste hardened concrete," *Journal of Cleaner Production*, vol. 309, no. 127375, August 2021.
5. Y. Sui, C. Ou, S. Liu, J. Zhang, and Q. Tian, "Study on properties of waste concrete powder by thermal treatment and application in mortar," *Applied Sciences*, vol. 10(3), no. 998, August 2020.
6. A. Carrico, J. Bogas, and M. Guedes, "Thermoactivated cementitious materials – a review," *Construction and Building Materials*, vol. 250, pp. 1–13, July 2020.
7. E. Vejmelkova, D. Konakova, M. Cachova, M. Zaleska, P. Svora, M. Keppert, P. Rovnanikova, and R. Cerny, "High-strength concrete based on ternary binder with high pozzolan content," *Structural Concrete*, vol. 19, no. 5, pp. 1258–1267, March 2018.
8. L. Li, W. Liu, Q. You, M. Chen, and Q. Zeng, "Waste ceramic powder as a pozzolanic supplementary filler of cement for developing sustainable building materials," *Journal of Cleaner Production*, vol. 259, no. 120853, June 2020.
9. J. Shao, J. Gao, Y. Zhao, and X. Chen, "Study on the pozzolanic reaction of clay brick powder in blended cement pastes," *Construction and Building Materials*, vol. 213, pp. 209–215, August 2019.
10. A. Tokareva, S. Kaassamani, and D. Waldmann, "Fine demolition wastes as supplementary cementitious materials for co2 reduced cement production," *Construction and Building Materials*, vol. 392, no. 131991, August 2023.

11. A. Tokareva, S. Kaassamani, and D. Waldmann, "Using ceramic demolition wastes for co2-reduced cement production," *Construction and Building Materials*, vol. 426, no. 135980, May 2024.

PAH detection in road surface residue using various hyperspectral imaging sensors

Paul Bäcker, Georg Maier, Thomas Längle, and Jürgen Beyerer

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Fraunhoferstr. 1, 76131 Karlsruhe, Germany

Abstract Road destruction sites still yield tar-contaminated rubble, a material that was used until the mid 1980s in many European countries and has been shown to be carcinogenic and mutagenic. Usually, this tar-contaminated road rubble is mixed with bitumen and minerals, which are also present in road construction and are valuable for recycling. Therefore, a challenge lies in distinguishing tar from bitumen and being able to sort them for further treatment and recycling. In this work, we investigate whether it is possible to distinguish tar- from bitumen-based road surface material with hyperspectral imaging in the short- and mid-wave infrared spectrum. The findings reveal that SWIR imaging is less effective, while MWIR imaging provides more promising results by distinguishing between different sources of samples based on their chemical signatures. However, generalizability across various sources poses a significant challenge due to the variability in sample composition, which is heavily influenced by the presence of different additives. Notably, the study demonstrates the efficacy of MWIR hyperspectral imaging in detecting polycyclic aromatic hydrocarbons (PAHs) in controlled sand samples, suggesting potential for identifying PAH contamination in road debris. The results highlight several challenges and emphasize the necessity for a more diverse dataset to enhance model robustness and accuracy.

Keywords Hyperspectral, MWIR, PAH

1 Introduction

Until the 1980s, tar was primarily used as a binder for road surface construction before being outlawed in most European states [1]. The material was outlawed, as it contains Polycyclic Aromatic Hydrocarbons (PAHs). They are a class of organic compounds with multiple aromatic rings, such as Naphthalene, Biphenyl, Fluorene and Anthracene. These compunds have been identified as carcinogenic, mutagenic, and genotoxic and can contaminate groundwater [2]. In addition, the use of recycled materials containing tar as the foundation of new road surfaces has been restricted. Therefore, potentially tar-containing construction and demolition waste from roads must be decontaminated, which is currently highly energy consuming.

This problem is amplified by the lack of separation of the different materials contained in the road rubble. In addition to tar, these are usually bitumen, which replaced tar as binder material, and minerals, which make up the largest part of the road-surface-mixture and account for approximately 95% [3]. Both bitumen and minerals would be valuable for recycling. However, in many cases, these bitumen and mineral fractions are mixed with tar-containing road rubble. In these cases, the entire batch is considered tar-contaminated, as mixing is prohibited in most countries [4]. This tar-containing rubble is land-filled or recycled using thermal methods that destroy the chemical structures of the embedded minerals, rendering it unusable for further recycling [5,6]. Separating tar from other road materials before decontamination would also improve the efficiency of decontamination and reduce transportation costs.

1.1 State of the art and research

Currently, a lot of uncontaminated road rubble is declared tar-containing, as this is in some cases cheaper for the demolition crews than carrying out the extensive testing which is currently mandated [4]. This extensive testing includes taking point samples in a certain raster and having them analyzed in a laboratory. The methods applied here are usually High Performance Liquid Chromatography, Gas Chromatography, or mass spectrometry [7]. Due to the high testing efforts, some material is falsely labeled tar-contaminated.

To get a rough overview of possible PAH concentration, solvent-based paints can be sprayed onto the rubble, which react with the PAHs creating a fluorescent effect that is visually observable. However, this method is not sufficient for official classification, as this detection method only works for a subset of possible PAHs. For large-scale dense testing, an optical solution would be much preferred, allowing for higher throughput.

Optical methods that use intrisic fluorescence have been shown to be capable of detecting PAH concentration in soil [8, 9]. However, the methods require high-intensity excitation and a slow scanning process. Although some methods are inline capable, the throughput of these systems is still orders of magnitude lower than required.

Some studies have used spectroscopy in the ultraviolet to visual wavelength regions to detect PAH in gas streams [10]. Although these methods are inline-capable, they require vaporization of the samples, which renders them unusable for the target recycling applications.

Hyperspectral short-wave infrared (SWIR) sensors have also been investigated in this context. It has been shown that the information allows distinguishing between different additives in bitumen-based binders [11]. Furthermore, [12] were able to detect PAH in soil using near- to short-wave infrared. However, accuracy is rather limited and detection is likely related to the interrelationship between organic carbon and clay, which is not present on the road surface.

1.2 Contribution

In this work, a preliminary study is presented that examines the differences in spectral reflectivity of tar, bitumen, and mineral in both the short- and mid-wavelength infrared spectrum with the goal of identifying methods to distinguish the different materials. Hence, we consider the problem as a classification task. Our work forms the basis for the development of a sensor-based sorting approach [13], which allows materials to be separated before further treatment, thus reducing transport and processing costs. Therefore, we consider sensors that are suitable for this field of application. More precisely, we use a hyperspectral SWIR camera operating in the wavelength range from $1000nm$ to $2500nm$ as well as a hyperspectral MWIR camera that is sensitive from $2700nm$ to $5300nm$.

We investigate which wavelengths are particularly suitable for finding relevant features for distinguishing tar and bitumen. Related work suggests that the regions between 6250nm and 14286nm as well as around 3300nm [14] are relevant. The MWIR sensor covers the range around 3300nm, the SWIR sensor might, however, be able to detect some overtones of these spectral bands.

2 Materials and Methods

In the following, we provide information on the samples considered in this study, as well as on the system used for data acquisition. We further present initial insights obtained via exploratory analysis of the acquired spectral data and present the final processing pipeline developed and used in the remainder.

2.1 Description of samples

The dataset for this study was obtained using 151 bitumen, 118 mineral and 148 tar samples with a grain size of approximately 16mm to 32mm. The samples were acquired from 3 different sites (mainly road destruction sites), which is a total of 12 sources. The samples were fixed to sample carriers, see 1. This allows for reproducible measurements and the fusion of different sensors. The sample carriers are fixed to the linear stages of the cameras using dowel pins.

Figure 1: Example for samples on sample carriers. Left: mineral samples, middle: bitumen samples, right: tar samples.

2.2 Description of Acquisition System

For our analysis, both a hyperspectral camera in the SWIR (HySpex SWIR-640) and in the MWIR range (Specim FX50) were used. SWIR imaging systems are widely used in current sensor-based sorting facilities to separate materials such as different polymers or to detect impurities in different consumer goods. Therefore, it would be ideal if these sensors and facilities could be used with small modification for sorting road debris. However, most known absorption bands of PAH are well outside the spectral range of SWIR. Therefore, detection relies on the detection of high-order spectral overtones, which are challenging to detect. It is assumed that these bands may be easier to detect in the MWIR. Therefore, we also use a MWIR sensor.

2.3 Processing Pipeline

For this study, several step analysis was performed. As a basis, a classification pipeline has been created, incorporating an Autoencoder and a kfolds-classification. This pipeline is applied to both sensors and different separations for train and test samples. Further, the influence of the Autoencoder on the linear separability of the dataset is shown.

Overview

2 shows the data processing pipeline that incorporates an Autoencoder step. In our work, train/test splits are done sample-wise, but classification is performed pixel-wise. The full pipeline is shown, however we evaluated multiple versions with different or no implementation for some of the steps.

The samples are first captured using a sensor (either SWIR, MWIR or both sensors combined). The result is the captured image data (1). This data is then preprocessed. A SVN normalization is applied, dead pixels are interpolated, and some parts of the spectrum are masked (CO_2 and noisy parts of spectrum). From the resulting preprocessed data (2), the mineral samples are split off to train an autoencoder. The Autoencoder uses an MLP structure. The tar and bitumen samples are used for classification. They are split into train- and test samples using different methods and following a cross-validation approach (3).

If feature selection is applied, it is fitted using the training set (4) and then applied to both train- and test samples (5). The Autoencoder is then applied to the training samples. The 20% of the pixel with the lowest reconstruction error is considered an embedded mineral and is removed (6). In addition, the 20% quantile of the reconstruction error is learned as a threshold. The Autoencoder is now applied to the test samples. All pixels with a lower reconstruction error than the threshold are removed (8). An artificial neural network is trained as a pixelwise classifier using (6). The classifier is applied to (7) and compared to the ground-truth data for evaluation. The following sections go into further detail on the different pipeline steps.

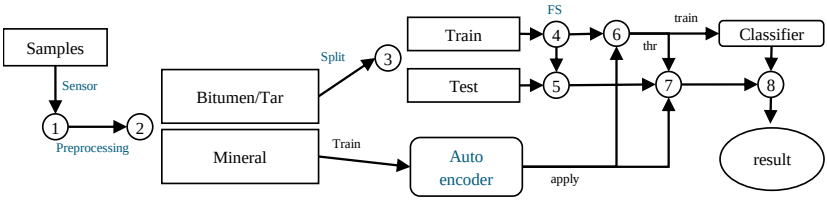


Figure 2: Data Pipeline for classification. The marked steps are altered during evaluation. *FS* denotes Feature Selection, which is trained on the training samples and then applied to all samples. All steps that have been varied for analysis are marked in blue.

Preprocessing

Preprocessing serves the purpose of removing irrelevant or misleading information from the measurements. The preprocessing in this work includes object segmentation, dead pixel correction by spectral interpolation, removal of CO_2 band as well as noisy edge-wavelengths and Standard Normal Variate (SNV) transformation to remove influence of mean-value (on a per-pixel base)

Object segmentation was achieved by training a simple classifier on background vs. foreground spectra. The resulting masks were then filtered and used as individual object masks. A dilation morphological filter was applied to the masks to remove edge pixels, which often suffer from mixed spectra, shadows, or steep surface angles. The CO_2band was ignored for the MWIR data.

Autoencoder for Removal of Embedded minerals

Another unwanted factor are the minerals embedded in the tar and bitumen samples. There are two possible negative effects of the embedded minerals. First, learning differences between tar and bitumen is more difficult if the training data set also contains spectra that are actually of class mineral but are wrongly labeled as tar or bitumen. Secondly, it is possible that there is a correlation between the type of the embedded mineral and the actual class. However, there is no indication that this link would generalize to all tar and bitumen samples.

We therefore chose to remove the embedded mineral from the tar and bitumen using an Autoencoder approach. We chose this approach instead of a classification approach, because a classification approach would require additional tar and bitumen samples. The Autoencoder is a MLP model that tries to reproduce input spectra. In our case, the model is trained on mineral spectra and is therefore capable of reproducing mineral spectra accurately, while it struggles to reproduce tar or bitumen spectra. The reconstruction error for each spectrum is compared to a threshold to detect the embedded minerals. For a visual confirmation of the overall validity of the approach, the result of the Autoencoder is merged onto a high-resolution SWIR image. An example is provided in 3.

To visualize the effect in feature space, we performed a partial least squares (PLS) transformation on the raw input data for tar and bitumen, the filtered data (without minerals) and for cross-checking a dataset with only the spectra identified as minerals. 4 displays the results as kernel density plots. It is apparent that the linear separability of the classes is improved by removing the embedded minerals.

Feature Selection

A main motivation for feature selection is an improvement in inference time, as fewer features have to be processed. Since the underlying chemical properties of the different materials usually only affect a small part of the spectrum (for example, some resonance frequencies of different bonds), the approach in many cases also helps remove irrelevant information and can improve robustness and accuracy.

To detect relevant features in the tar and bitumen samples, we per-

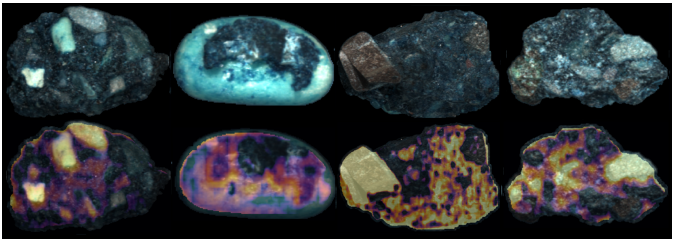


Figure 3: Visual result of Autoencoder filter. Detected minerals are marked in yellow.

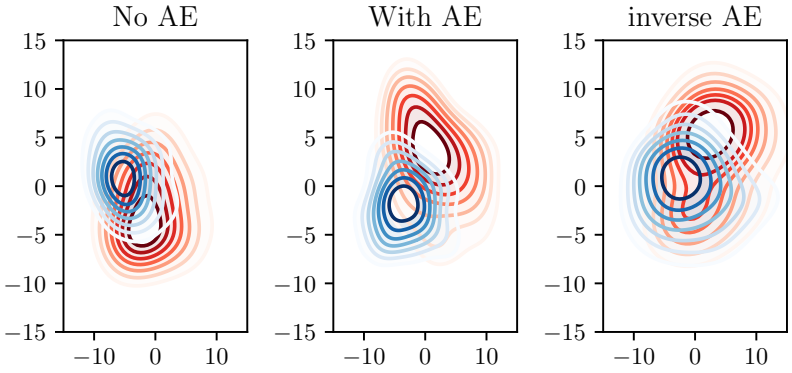


Figure 4: Change of linear separability when applying Autoencoder on the MWIR data.

form a synergy interval PLS (SiPLS) [14] with 3 bands of 10 adjacent features on a kfold cross validation. 5 shows that the features around 3000nm and 4250nm are consistently selected.

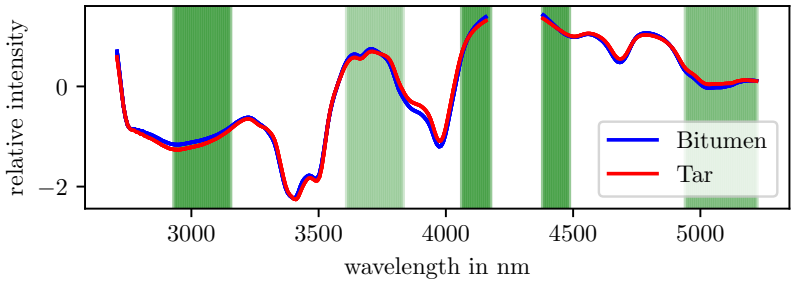


Figure 5: Feature selection bands. Reference spectra are means over all tar and bitumen samples.

3 Results and discussion

We have applied different versions of the pipeline presented in 2.3. The different variables that were tested are sensor choice, preprocessing, influence of embedded minerals / Autoencoder, influence of train/test split variation and influence of performing feature selection

The sensor choice refers to the selection of sensors used. In this study, a SWIR and a MWIR sensor were considered. In addition, the performance of the combined sensors was tested.

Two different methods for cross validation and train/test split were evaluated. Firstly, the separations of train and test samples for each class is performed without considering the source (road destruction site) of the materials by randomized kfold cross validation. Secondly, the samples are split along the sources, so that each training set contains two sources from each class and the test set the other. The latter separation is in the following referred to as *source aware*.

The influence of performing a SiPLS feature selection was tested. The feature selection was trained on the training samples and applied to both train and test samples before classification.

1 shows the accuracy both in terms of the percentage of correctly classified pixels as well as percentage of correctly classified objects. The object classification is achieved by making a majority decision over all pixels that belong to a given object.

Table 1: Classification performance for different settings.

Sensor	Split	FS	AE	Ob Acc	Px Acc
SWIR	Random	No	Yes	63.05	58.27
SWIR	Source aware	No	Yes	44.57	48.53
Both	Random	No	No	87.81	81.05
Both	Random	No	Yes	**89.53**	79.14
MWIR	Random	No	No	88.44	**81.89**
MWIR	Random	No	Yes	88.06	80.42
MWIR	Source aware	No	No	51.92	50.65
MWIR	Source aware	No	Yes	54.38	52.73
MWIR	Source aware	Yes	Yes	70.80	60.69
MWIR	Random	Yes	Yes	83.29	76.11

The results show that the SWIR sensor is not suitable for distinguishing tar and bitumen-based road rubble. Using both sensors does not improve accuracy compared to using MWIR alone.

However, one major caveat to the results is that they only generalize within different sources. When samples from all sources are in the training set, the accuracy is much higher than when a model trained on a set of sources is applied to other sources. This is a strong indication that differentiation is mostly not based on tar or bitumen specific properties but rather on different additives in the samples. Further, although the samples of the tar class have been confirmed to be PAH containing, the PAH distribution within the class is unknown. Therefore it could be possible, that individual samples within this class do not contain any PAH. We were able to show that feature selection helps to significantly increase accuracy when testing on novel sources. This may be because tar and bitumen are most easy to distinguish in these wavelength regions, due to the differences in chemical properties. Either PAH itself can be detected, or some other proxy linked to tar or bitumen. Another cause may be changes in environment conditions during the dataset capture process. However, the background test in **??** has shown that the environment influence does not carry that much

information.

3.1 Remaining Challenges

It is not yet clear how well the results generalize to the distinction between tar and bitumen. The current models appear to mostly learn additive-specific features, as indicated by the far worse performance when testing on samples from unknown sources. To see whether a tar/bitumen specific separation is possible, a much larger and more diverse dataset would be required.

In a second step, it could be determined whether the MWIR data allows for the detection of PAH content in tar and bitumen or is sensitive to other chemical differences between tar and bitumen.

4 Conclusion

In this study, we investigated the separability of tar and bitumen from road debris using hyperspectral imaging techniques in the short- and mid-wave infrared spectrum. Motivated by the need to improve recycling processes and reduce costs associated with handling potentially hazardous materials, we utilized hyperspectral SWIR and MWIR cameras to classify samples of tar, bitumen, and minerals obtained from road destruction sites.

Our findings indicate that, while SWIR imaging did not prove effective, MWIR imaging enabled a distinction between different sources of samples on the basis of their chemical signatures. We have also shown that masking and ignoring the embedded minerals inside the tar and bitumen samples improves the linear separability of tar and bitumen. However, generalizability across different sources remains a challenge, primarily due to variability in sample composition influenced by various additives. Notably, our mid-wave hyperspectral imaging approach successfully detected PAHs in controlled sand samples, demonstrating its potential to identify PAH contamination in road debris.

The study's results underscore several challenges and highlight the importance of expanding the dataset to improve model robustness and accuracy. Future work should focus on enriching the dataset by collecting a more diverse range of samples to enhance the model's ability

to generalize across different types of tar and bitumen. Another developement avenue would be moving away from a classification approach towards a regression method, directly estimating the PAH content for each pixel. This would allow detecting PAH contaminated bitumen samples (for example as a result of oil spills) and allow distinctions between different levels of PAH contamination Finally, exploring the impact of different additives on the MWIR spectra could help in improving the confidence in the results.

These steps are essential to advance the capability to sort and recycle road debris effectively, which can lead to significant environmental and economic benefits.

Acknowledgment

This work was supported by the Fraunhofer Internal Programs under Grant No. PREPARE 40-02829. Training samples were kindly provided by Zwisler GmbH.

References

1. Wiley, *Polycyclic aromatic hydrocarbons (PAH) [MAK Value Documentation, 2012]*. John Wiley & Sons, Ltd, 2013.
2. P. B. Farmer, R. Singh *et al.*, "Molecular epidemiology studies of carcinogenic environmental pollutants: Effects of polycyclic aromatic hydrocarbons (PAHs) in environmental pollution on exogenous and oxidative DNA damage," *Mutation Research/Reviews in Mutation Research*, vol. 544, no. 2, pp. 397–402, 2003.
3. T. S. Arnold, "What's in your asphalt?" vol. 81, no. 2, 2017.
4. J. Booth. (2019) Managing reclaimed asphalt. Association of Directors of Environment, Economy, Planning and Transport (ADEPT). [Online]. Available: https://adeptnet.org.uk/documents/managing-reclaimed-asphalt-2019
5. V. D. Zementwerke, Ed., *Zement-Taschenbuch 2002*, 50th ed. Düsseldorf: Verlag Bau u. Technik, 2002.
6. S. Breuer and F. R. Schilling, "Anisotropic thermal transport properties of quartz: from $-120\,^{\circ}$c through the α–β phase transition," *European Journal of Mineralogy*, vol. 33, no. 1, pp. 23–38, 2021.

7. Y. Zhang, L. Yuan, S. He, H. Tao, W. Xie, X. Zhang, X. Ren, T. Jiang, L. Li, and Z. Zhu, "Contemporary research progress on the detection of polycyclic aromatic hydrocarbons," vol. 19, no. 5, 2022.
8. F. Qazi, E. Shahsavari, S. Prawer, A. S. Ball, and S. Tomljenovic-Hanic, "Detection and identification of polyaromatic hydrocarbons (PAHs) contamination in soil using intrinsic fluorescence," vol. 272, p. 116010, 2021.
9. V. T. Tran, J. S. Seok, I. Yoon, and H. Ju, "Ultraviolet fluorescence-based quantitative detection of polycyclic aromatic hydrocarbons," vol. 15, no. 4, pp. 356–361, 2021.
10. T. Weide, V. Guschin, W. Becker, S. Koelle, S. Maier, and S. Seidelt, "Analysis of pure tar substances (polycyclic aromatic hydrocarbons) in the gas stream using ultraviolet visible (UV-vis) spectroscopy and multivariate curve resolution (MCR)," vol. 69, no. 1, pp. 143–153, 2015.
11. B. Jahangiri, K. Barri, A. H. Alavi, and W. G. Buttlar, "A molecular sensing method integrated with support vector machines to characterize asphalt mixtures," vol. 179, p. 109528, 2021.
12. R. N. Okparanma and A. M. Mouazen, "Visible and near-infrared spectroscopy analysis of a polycyclic aromatic hydrocarbon in soils," vol. 2013, no. 1, p. 160360.
13. G. Maier, R. Gruna, T. Längle, and J. Beyerer, "A survey of the state of the art in sensor-based sorting technology and research," *IEEE Access*, vol. 12, pp. 6473–6493, 2024.
14. M. Li, Y. Feng *et al.*, "Quantitative analysis of polycyclic aromatic hydrocarbons in soil by infrared spectroscopy combined with hybrid variable selection strategy and partial least squares," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 257, p. 119771, Aug. 2021.

Approach to increase resource efficiency in natural gypsum using hyperspectral imaging and machine learning methods

Patrick Hunhold, Elske Linß, and Saskia Nowak

Materials Research and Testing Institute at Bauhaus-Universität Weimar
Coudraystraße 9, 99423 Weimar, Germany

Abstract The sustainable use of limited resources has become a critical priority in modern economic practices. The planned cessation of coal-fired power generation by 2038 and the subsequent discontinuation of flue gas desulfurization (FGD) gypsum will deprive the gypsum industry of this significant resource. This deficit cannot be effectively mitigated through recycling alone. Expanding the extraction of natural gypsum and anhydrite poses substantial environmental concerns. This study explores the classification of underutilized gypsum-bearing rocks (mixed rocks), a byproduct of natural gypsum mining for the application as a replacement. Hyperspectral imaging (HSI) is utilized to capture spatial and spectral information, which is then analyzed using machine learning techniques. Achieving recognition rates above 99 %, the approach demonstrates great potential for identifying mineral phases within mixed rocks, highlighting its viability for enhancing resource efficiency.

Keywords Hyperspectral imaging, machine learning, SWIR, natural resources, resource efficiency, gypsum, anhydrite

1 Introduction

1.1 Problem and Objective

The construction industry as one of the major CO_2 producers plays a central role in sustainability within the broader societal context, particularly in the efficient use of resources. Gypsum is a significant raw

material, with its demand expected to rise to approximately 10.7 million tons per annum (10.7 Mt/a) in Germany by 2035. The current demand of around 10.0 Mt/a (as of 2020) is entirely met by domestic sources, divided into three categories: natural gypsum and anhydrite, flue gas desulfurization (FGD) gypsum (by-product of coal-fired power generation), and recycled gypsum. With the planned coal phase-out by 2038, a substantial portion of FGD gypsum will be eliminated, necessitating an increased reliance on natural gypsum and recycled gypsum. Since the share of recycled gypsum is projected to grow only to around 1.0 Mt/a, the extraction of natural gypsum will become increasingly significant. However, this development is controversial due to the potential impact on sensitive ecosystems, such as the biodiversity-rich South Harz region [1–6].

A posible potential is given by the so-called transitional or mixed rock types, which have been economically neglected due to their variable content of crystal water and other properties. However, given the increasing demand, the efficient utilization of these resources is gaining greater attention [2, 4].

The following section presents a hyperspectral imaging technique that, in combination with machine learning methods, enables rapid and non-destructive determination of gypsum and anhydrite fractions in transitional and mixed rocks. This approach facilitates a goal-driven management of material flows and a more efficient use of limited resources.

1.2 Related Work

The ability to distinguish between gypsum and anhydrite based on the spectral information in the near-infrared (NIR) range has been well-documented. Differences can be identified particularly in certain spectral bands. Cloutis et al. [7] describe that the dominant bands in the visible-near-infrared (VNIR) region, influenced by water (H_2O), are located around 1400 nm to 1500 nm and 1900 nm to 2000 nm, which is why these regions are referred to as water bands. Additional bands specific to the near-infrared spectrum (NIRS) of gypsum can be found at approximately 1000 nm, 1200 nm, and 1750 nm. During the conversion of gypsum to bassanite, Harrison [8] observed that the characteristic bands of gypsum shift towards shorter wavelengths. In contrast,

it is generally assumed that anhydrite lacks these water bands. However, experiments by Cloutis et al. [7] indicate that weak expressions of these bands can occur, suggesting that anhydrite samples may be partially hydrated or in a transitional state towards bassanite or gypsum.

Müller et al. [9] report that optical sorting based on NIRS in free-fall experiments involving a mixture of 62.6 wt% anhydrite and 37.6 wt% gypsum resulted in an increase of the anhydrite content to 98.0 wt% and in the gypsum content to 88.2 wt% in the seperated fractions, leading to significant improvements in material quality.

2 Materials and Methods

2.1 Materials

For the investigation of the classification potential of mixed/transitional rocks using hyperspectral imaging, three samples within the grain size range 8/16 (8 mm to 16 mm), provided by CASEA GmbH from the Ellrich mine, are provided. These samples were pre-classified by experts into the categories of anhydrite, gypsum, and mixed rocks as depicted in Figure 1. Visual differentiation of these categories proves to be challenging, as all samples exhibit a color variation ranging from white to gray. While anhydrite samples tend to show a more whitish hue, gypsum samples appear slightly grayer. However, this visual distinction is not considered a definitive differentiating feature.

Figure 1: Material from left to right: gypsum, mixed rocks and anhydrite.

2.2 Experimental Setup

The experimental setup as shown in Figure 2 comprises a hyperspectral line-scan camera (HySpex SWIR-384 [10]) with a sensitivity in the short-wavelength infrared (SWIR) range from 1000 nm to 2500 nm and two halogen lamps (emission spectrum from 400 nm to 2500 nm). The lamps are positioned at approximately 45° to the sample to create a line illumination. A moveable sample stage allows line-by-line scanning. The setup is complemented by a white reference standard measuring 200 mm by 50 mm, which exhibits a nearly constant reflectance of 95 %.

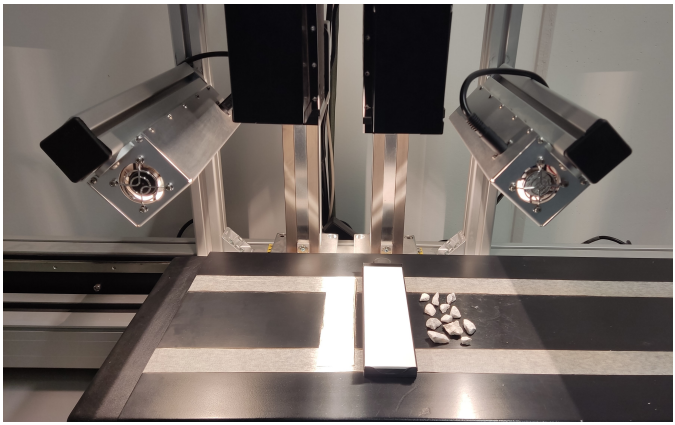


Figure 2: Experimental setup for hyperspectral image acquisition.

2.3 Data Acquisition

Before the generation of the hyperspectral image data begins, the preclassified material samples provided by CASEA GmbH are divided into two groups as shown in Figure 3. The first group consists of 20 randomly selected particles per material, with a high degree of variance in shape and color. This group is used on the one hand to check the purity of the material by determining the crystal water content using thermogravimetric analysis (TGA) and on the other hand the images of these samples are later used to investigate the classification of gyp-

sum and anhydrite in the mixed rocks. It should be noted that each particle in this group was individually labeled so that the results of the destructive TGA can later be traced back to the corresponding image data. The second group from the material provided forms the data basis for the training and evaluation of various classification models. For this purpose, around 500 particles each were extracted from gypsum, anhydrite and mixed rocks.

The TGA is carried out in several steps. First, each particle was ground to an approximate grinding fineness of 50 µm. The initial mass of each ground particle was then recorded. The samples were then dried at a maximum temperature of 40°C for 24 h to remove any free moisture. After drying, the particle mass (m_0) was measured again. To determine the water of crystallization, the particles were heated to a temperature of 350°C for 5 h to release the water bound in the crystal lattice. Finally, the mass (m_1) is measured again. The content of crystal water ($CW_{\%}$) was calculated based on the differences in the measured masses (Δm_{H_2O}) using equation 1 and equation 2.

$$\Delta m_{H_2O} = m_1 - m_0 \tag{1}$$

$$CW_{\%} = \frac{\Delta m_{H_2O}}{m_0} \cdot 100\ \% \tag{2}$$

2.4 Preprocessing

The preprocessing of the data is carried out in several steps. The first step involves correcting the image data, which includes identifying defective and overexposed pixels that manifest as distinctive outliers within the image data. The correction of individual pixel spectra is achieved by calculating an average pixel spectrum from the surrounding pixels, which is then used to replace the data of the defective pixel. The next preprocessing step involves normalizing the image data using the recorded white reference standard. This correction is applied individually to each pixel within one line of the image. Finally, the particles

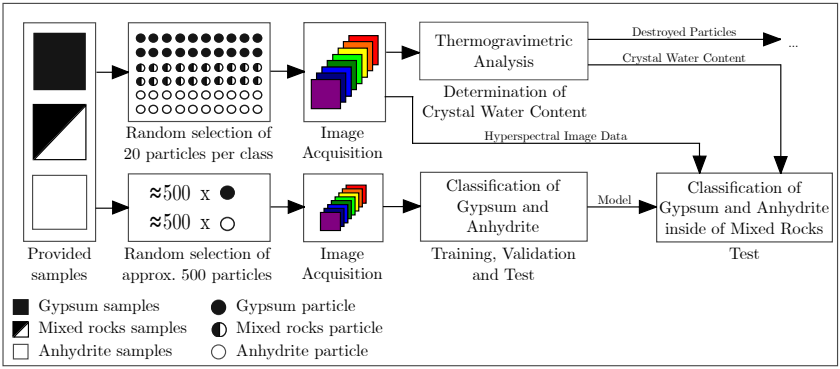


Figure 3: Process flow regarding data creation, training and testing.

in the acquired hyperspectral data are segmented. Due to the high contrast between the particles and the sample stage, a simple thresholding method can be employed.

2.5 Feature Extraction

For feature extraction in hyperspectral data, both spatial and spectral features can be utilized. However, since the focus of this study is on the spatial distribution of mineral phases within the mixed rocks, only spectral features are considered. These features are then combined with dimensionality reduction techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

The extracted feature focuses on the pixel spectra of the recorded particles within a selected Region of Interest (ROI) in the center of each image. The ROI size is set to 25 px by 25 px (ignoring background pixels). Representative examples of gypsum and anhydrite and their mean spectra are shown in Figure 4.

2.6 Classification

Before starting the classification process, the dataset is split into training and test datasets in an 80/20 ratio. The split is performed using

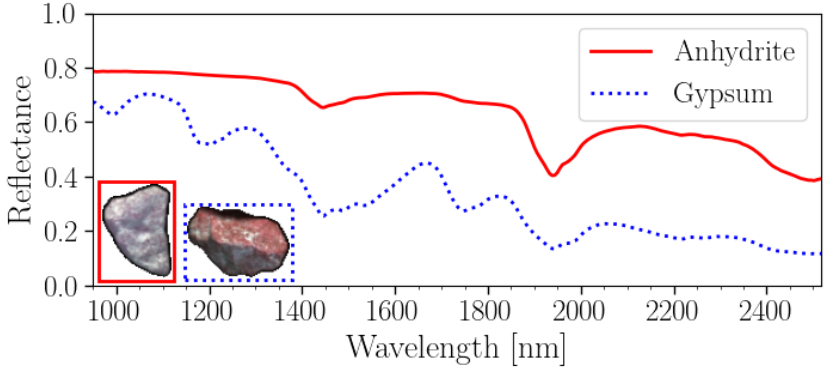


Figure 4: Mean reflectance spectra of gypsum and anhydrite including visual examples as false RGB images in the SWIR range.

stratified cross validation over five iterations to ensure that the class proportions in the dataset are preserved.

For the classification task, various Machine Learning (ML) algorithms are employed, including Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). It is important to note that the MLP requires a validation dataset during training. For this purpose, approximately 10 % of the training dataset is randomly selected and used as the validation set.

After training and evaluating the ability to distinguish between gypsum and anhydrite, an additional test is conducted. In this step, the trained machine learning models are applied to the previously described mixed rocks, and their data are classified on a pixel-by-pixel basis. For validation, the previously determined contents of crystal water are used as reference.

3 Results

3.1 Distinguishing Gypsum and Anhydrite

Figure 5 presents the mean accuracy and standard deviation for different ML algorithms with and without dimensionality reduction meth-

ods. In general, distinguishing gypsum and anhydrite based on their reflectance spectra in the SWIR range proves to be an effective approach. Achieving accuracies of 99 %, all ML algorithms show a high reliabibilty in this classification task. Using PCA and LDA as dimension reduction methods slightly enhance the overall accuracy and reduce the standard deviation as expected, but not in a decisive degree.

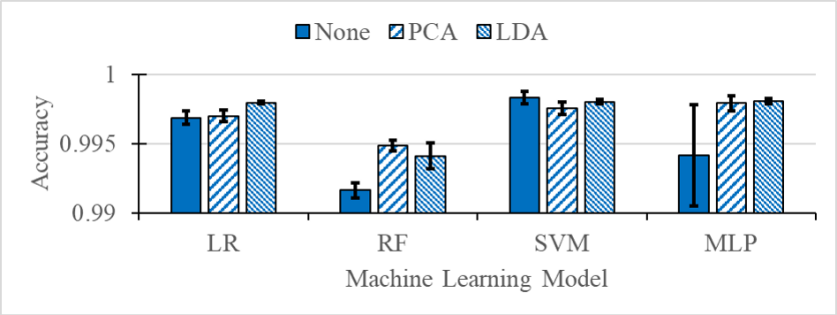


Figure 5: Mean accuracy and standard deviation of all classifiers with and without dimensionality reduction methods.

3.2 Detection of Gypsum and Anhydrite in Mixed Rocks

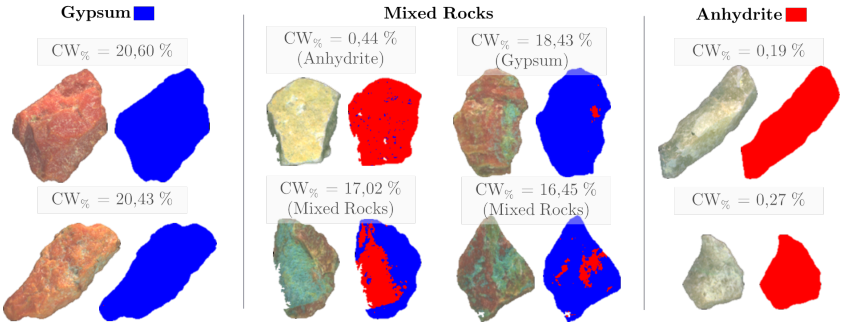


Figure 6: Pixelwise classification of gypsum and anhydrite in pure material and mixed rocks including determined content of crystal water.

Based on the successful classification of gypsum and anhydrite, the next task focuses on detecting these minerals in mixed rocks. As shown in Figure 6, the results from Sect. 3.1 were effectively reproduced using an SVM classifier with LDA. As it seems, it is the best method according to high accuracy and low standard deviation, applied to previously unknown samples, which had been used earlier to determine the content of crystal water (see Sect. 2.3). Pure gypsum and anhydrite samples were correctly classified, with all pixels assigned to their true labels. When applied to mixed rocks, a clear distinction between gypsum and anhydrite was observed within the samples. Additionally, the dataset of mixed rocks included samples that could be reliably classified as gypsum or anhydrite based on their determined content of crystal water.

4 Conclusion and Future Work

In conclusion, the integration of HSI and ML for distinguishing between gypsum and anhydrite demonstrates promising results, particularly for pure materials. For mixed rocks, this approach offers potential to enhance resource utilization, as pure gypsum and anhydrite often remain hidden within unused piles of mixed material. However, several aspects require further investigation. The current experiments were conducted under controlled conditions with a limited number of samples from a single region. It is crucial to evaluate whether the high accuracy can be maintained under more challenging conditions, such as varying levels of dust and humidity. Additionally, regional variations and the presence of impurities from other minerals may impact the classification accuracy of gypsum and anhydrite in mixed rocks. Future work will address these challenges to further optimize the efficient utilization of these resources.

Acknowledgements

The research and development project “Entwicklung von Leichtgipsen aus Schaumgips unter Nutzung von Ersatzbaustoffen - Leichtgips” in which this work was a part of, is funded by the Thuringian Ministry of Economics, Science and Digital Society. Special thanks go to the

funding body, whose financial support made the research on the abovementioned topic possible in the first place.

References

1. Bundesverband der Gipsindustrie e. V., "GIPS-Datenbuch," May 2013.
2. S. Reyer and K. Fohlert, "Untersuchung zur Rohstoffsicherung der Rohstoffart Gips/Anhydrit in Nordthüringen," e.t.a. Sachverständigenbüro Reyer, Erfurt, Tech. Rep., Dec. 2017.
3. Bundesamt für Justiz, "KVBG - Gesetz zur Reduzierung und zur Beendigung der Kohleverstromung*," p. 46, Aug. 2020.
4. H. Alwast, "Gutachten: "Umweltverträgliche Alternativen zum Abbau von Naturgips"," Bund für Umwelt und Naturschutz Deutschland e.V. – BUND, Tech. Rep., Sep. 2020.
5. BUND, GRÜNE LIGA, NABU, Verband deutscher Karst- und Höhlenforscher, and Architects for Future, "Abbau von Naturgips in Deutschland: Verbände legen Positionspapier vor," https://www.bund.net/service/presse/pressemitteilungen/detail/news/abbau-von-naturgips-in-deutschland-verbaende-legen-positionspapier-vor/, Oct. 2021.
6. J. Demmich, "Einsatz von REA-Gips und Recyclinggips – ein Beitrag zur Ressourceneffizienz," in *Wasser, Energie und Umwelt*, M. Porth and H. Schüttrumpf, Eds. Wiesbaden: Springer Fachmedien Wiesbaden, 2022, pp. 169–176.
7. E. Cloutis, F. Hawthorne, S. Mertzman, K. Krenn, M. Craig, D. Marcino, M. Methot, J. Strong, J. Mustard, and D. Blaney, "Detection and discrimination of sulfate minerals using reflectance spectroscopy," *Icarus*, vol. 184, no. 1, pp. 121–157, Sep. 2006.
8. T. N. Harrison, "Experimental VNIR reflectance spectroscopy of gypsum dehydration: Investigating the gypsum to bassanite transition," *American Mineralogist*, vol. 97, no. 4, pp. 598–609, Apr. 2012.
9. S. Müller, A. Müller, I. Döring, and U. Palzer, "Sensorbasierte Sortierung mineralischer Abfälle und Rohstoffe - Aufbereitungstechnik," https://www.at-minerals.com/de/artikel/at_Sensorbasierte_Sortierung_mineralischer_Abfaelle_und_Rohstoffe-3648506.html, May 2021.
10. HySpex, "HySpex SWIR-384," https://www.hyspex.com/hyspex-products/hyspex-classic/hyspex-swir-384/.

Automated data acquisition method for sensor-based real-time material flow characterization of recyclable waste streams using sensor fusion: A case study

Fabian Roth, Aydin Deliktas, Kaushik Rattawa, Alexander Feil, and Kathrin Greiff

RWTH Aachen University, Department of Anthropogenic Material Cycles, Wüllnerstraße 2, 52062 Aachen

Abstract In recent years, the development of real-time applications has become increasingly popular in the field of sensor-based systems and recyclable waste streams. One promising method is sensor-based real-time characterization (SBRTC) involving object detection or instance segmentation models as well as specific datasets containing recyclable materials to assess the quality of material flows. Building such models requires image data for training, testing and validation. This process is labor-intensive and prone to error, mainly when conducted manually. Here, we explore two approaches for the acquisition on conveyor belts: In approach I, a rotary encoder (a) and pre-defined time intervals (b) are compared to reduce acquisition gaps and redundancies, thereby improving data quality. The data acquisition was possible with a mean relative acquisition error (MRAE) of about 0% (a) and up to 66% (b). Approach II demonstrates the technical feasibility of object tracking which allows the counting of particles in a real-time video stream by leveraging Kalman Filter (KF), K-Nearest Neighbor (KNN), and Hungarian Maximum Matching (HMM). An accuracy of 99% ± 1% could be achieved. Therefore, this work contributes to novel data acquisition methods for high resolution RGB area images of SBRTC applications to effectively address the challenges of noisy and biased real-world datasets making it easier to perform data splitting.

Keywords Sensor-based real-time characterization, data acquisition, data quality, object tracking, computer vision

1 Introduction

In 2020, the European Union generated a total of 8.6 million tons of non-hazardous, non-ferrous metal waste, of which Germany accounted for 16.6%, ranking as the largest producer [1]. One key factor in reducing the environmentally and economically expensive consumption of primary non-ferrous raw materials is the transition towards a circular economy (CE). This aims at substituting more and more primary raw materials with secondary raw materials [2]. Therefore, reducing, reusing, and recycling has gained increasing importance, and within recycling, the characterization of recyclable waste streams has taken on a key role [3]. For this reason, gaining information about material characteristics is necessary for SBRTC of material flows. In order to leverage SBRTC, extracting frames from real-world video sequences is required to create novel datasets for inline analytics [4].
However, collecting and capturing images is both labor-intensive and time-consuming, particularly when conducted manually [5,6], as data collection, acquisition, and preparation can account for 45% to 90% of the total pre-processing time [7]. Moreover, a significant challenge in extracting frames from video sequences using an RGB area camera can be the issue of spatial and temporal correlation. That is, extracting two consecutive frames can lead to redundancies, such as the repeated occurrence of objects [8]. According to [6], a model's perceived performance can be artificially enhanced positively. This effect can be attributed to the fact that objects extracted twice can be part of both training and validation sets. Once a model is trained, it already knows the object. Moreover, duplicates only in training data can also falsely influence the model performance concerning generalization capabilities, as the model memorizes objects which are 'very similar', 'near duplicate', or 'exact duplicate' [9]. That is, data sparsity and redundancy can negatively impact the model's performance in its generalization capability by skewing the model and introducing a bias. This bias is also known as data leakage [9,10].
An automated data acquisition method for SBRTC of recyclable waste streams using sensor fusion offers the possibility of overcoming these limitations, i.e., avoiding acquisition gaps and redundancies.

Prior research has already demonstrated the suitability of rotary encoders and various sensor types. While in the automotive and manufacturing industries, rotary encoders are used to measure the movement of conveyor belts accurately [11], an optical encoder was used by [12] to start the acquisition of a line scan camera once the conveyor belt reached a stable velocity for the reconstruction of a 3D material transportation status. Another study demonstrated the calculation of the object length on a conveyor belt as it passes the scanline of a LiDAR camera [13]. Furthermore, the combination of RGB and magnetic sensors allows the determination of the rotational movements while utilizing an endoscopic capsule robot [14]. To deal with overlapping images, another study demonstrated a static acquisition method for scrap, where five frames per second were saved [15]. Moreover, rotary encoders allow high reliability and accuracy by converting a rotary motion into a digital or analog signal comprising several pulses per revolution [11]. However, using rotary encoders with RGB area cameras for data acquisition, i.e., preventing duplicate particle occurrences in the dataset by synchronizing camera acquisition framerate and conveyor belt speed, seems largely unexplored in literature. That is, sensor-based systems using optical sensors in combination with a rotary encoder and object tracking represents a relatively unsophisticated technology, thus offering untapped potential to improve dataset creation. For this reason, this study aims at improving the data quality of a data acquisition process by exploring methods to address spatial redundancies in consecutive frames and accurately determine ground truth data (i.e., true particle values) of real-time video sequences. To this end, we address the following research questions:

- RQ1: What is the impact of data acquisition without rotary encoder and how can spatial redundancies be prevented using a rotary encoder in order to improve data quality?
- RQ2: To which extent is it possible to determine the ground truth data of video sequences in real-time?

2 Materials and Method

In this study, two approaches are investigated: Approach I compares two methods to overcome the issue of spatial and temporal correlation. For approach II, object tracking is leveraged to determine the dataset's ground truth.

2.1 Acquisition principles

The experimental setup for data acquisition comprises several key components, as shown in Fig. 1: An industrial RGB camera (GV-77Q5WP-C-HQ) (a) with a resolution of 4500 by 4500 pixels, a pixel size of 3.2 µm and a color depth of 12-bit to extract color features from a 2D image with a framerate of 42 frames per second (FPS) at full resolution. This results in 0.089 mm per pixel with a recording area of 370.3 mm by 370.3 mm, which is illuminated by a diffuse illumination chamber (Planistar 60-60-Sled-3-VAD-19w-O, 5700K) (b). Images were taken by triggering a rotary sensor (MSK320) (c), which has a maximum angle resolution of 0.006 °. Additionally, a rotary encoder (MR320) (d) is mounted at the motor shaft, which sends up to 2000 impulses per revolution on a single signal line(SIKO Global 2024). Afterwards, an Arduino Nano v3 (e) receives data, which is transmitted to an industrial workstation (f). Processing the received data in (e) requires the circumference of a circle, which includes the radius of the motor shaft (60 mm) and the thickness of the belt (1.8 mm). Following, the total distance moved by the rotary encoder is compared with the total frame length of the recording area (y-axis). Once the rotary encoder has been moved to 370.3 mm, the total distance is set to zero, and the current image is saved. Otherwise, the distance travelled by each impulse is accumulated until the total distance is reached. Consequently, the recording software requests a periodic update from the Arduino Nano v3, repeating the acquisition process until the conveyor belt speed reaches zero.

2.2 Experiment I: Spatial and temporal correlation (RQ1)

Extracting frames from a video sequence requires addressing the challenge of spatial and temporal correlation. According to [16], spatial and temporal correlation is fundamental to analyzing object movement in

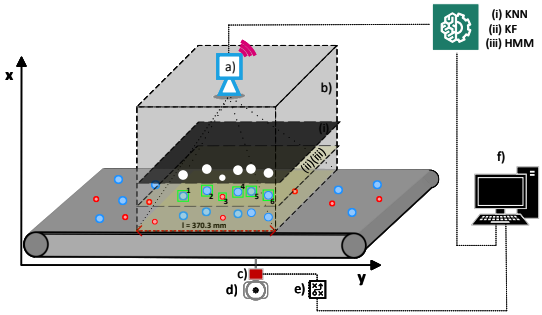


Figure 1: General structure of the experimental setup: (a) rgb camera, (b) illumination chamber, (c) rotary sensor, (d) rotary encoder, (e) Arduino Nano v3, f) industrial workstation and (i) K-Nearest Neighbor (KNN) background subtraction, (ii) Kalman Filter (KF), (iii) Hungarian Maximum Matching (HMM).

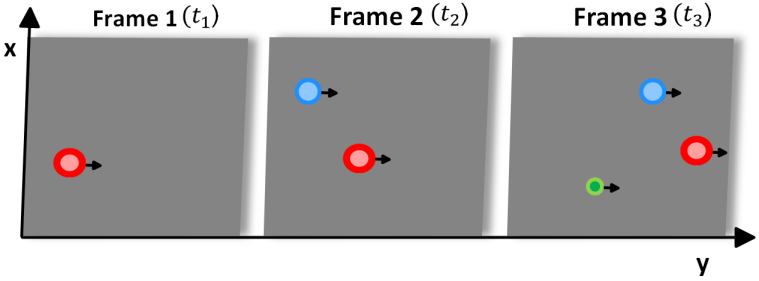


Figure 2: Consecutive frames over time (t_n) of a video sequence with particles in $t_1 =' red'$, $t_2 =' blue'$, $t_3 =' green'$.

a video sequence as objects change their positions dynamically. In this context, the spatial correlation is used to describe the location of objects within a single frame. In contrast, the temporal correlation describes the relationship between consecutive frames in a video sequence. As shown in Fig. 2, the position of the objects 'red', 'green, and 'blue' changes over time across consecutive frames.

As discussed in Section 1, one method for extracting frames from a video sequence is to select the sixth frame (static), while another

approach involves using an encoder (dynamic). For this reason, three speed levels were initialized at 39 FPS, i.e. $0.1\frac{m}{s}$, $0.15\frac{m}{s}$ and $0.24\frac{m}{s}$ (see Table 1), whereby the speed interval was set in incremental steps. The time interval t_{woe} for the static method is a fixed time interval that remains constant at different speed levels. It was calculated using an r.p.m. counter (PCE-DT 66) and the frame length of 370.3 mm. Each n-frame (f_{woe}) was captured, e.g. a frame was extracted after 3.7 seconds at $0.1\frac{m}{s}$, which equates to one image being captured every 144.42 frames. Using an encoder involves the dynamic time interval t_{we}, which varies at different speed levels or intervals. That is, every frame (f_{we}) is extracted depending on the conveyor belt's speed.

Speed Level	**Speed Interval**	**Time Interval**		**Frame Interval**	
$[\frac{m}{s}]$	$[\frac{m}{s}]$	t_{woe}	t_{we}	f_{woe}	f_{we}
0.10	0.10	3.70 s	3.70 s	144.42	144.42
0.10	0.15	3.70 s	2.47 s	144.42	96.28
0.10	0.24	3.70 s	1.54 s	144.42	60.17
0.15	0.10	2.47 s	3.70 s	96.28	144.42
0.15	0.15	2.47 s	2.47 s	96.28	96.28
0.15	0.24	2.47 s	1.54 s	96.28	60.17
0.24	0.10	1.54 s	3.70 s	60.17	144.42
0.24	0.15	1.54 s	2.47 s	60.17	96.28
0.24	0.24	1.54 s	1.54 s	60.17	60.17

Table 1: Speed levels and intervals at 39 FPS with the corresponding time interval (t_{woe};t_{we}) and frame interval (f_{woe};f_{we}).

To evaluate the acquisition performance, 50 particles (n_p) were recorded at pre-defined speed and time intervals. Afterwards, the total number of particles (c_p) within the recorded images was manually counted to calculate the mean relative acquisition error (MRAE), which can be denoted as Eq. (1):

$$\mathrm{MRAE}(n,c) = \frac{1}{N}\sum_{p=1}^{N}\frac{|n_p - c_p|}{|n_p|} \tag{1}$$

Where N is the number of the speed intervals.

2.3 Experiment II: Object tracking

In order to acquire ground truth particles at different speed levels (see Table 2) from a monolayer input stream, object tracking was utilized by leveraging three distinct algorithms, including KNN, KF and HMM (see Fig. 1) : (1) KNN is a pixel-level background subtraction method that determines the background by analysing the nearest neighbours in a pixel's "short-term-long-term" history [17]. In other words, the number of frames (or pixel values) used to model the background is determined by the history of the model (i). The model's sensitivity to changes is controlled by the Dist2Threshold (ii). Both parameters were set to 60 (i) and depending on the speed level: 200, 300, 320, and 340, respectively (ii). (2) KF is a statistical method for estimating the future position of an object based on its previous and current position, including both process noise covariance (PNC) (i) and measurement noise covariance (MNC) (ii). While (i) quantifies the degree of uncertainty associated with the behaviour and movements of the tracked object, (ii) represents the uncertainty inherent in the observations, capturing the inaccuracies and noise present in the measurement process [18]. In enhancing KF's responsiveness to changes at higher speed levels, (i) a scaling factor of 0.9 was set using a 4x4 matrix. Furthermore, a 2x2 matrix with a scaling factor of 0.5 and 0.3, respectively (ii), was defined to prevent missing changes in fast-moving objects. (3) In order to overcome the assignment problem in consecutive frames, HMM leverages inputs from (1) and (2) by calculating a cost matrix with a maximum distance of 70, thus assigning unique identifiers to each of the objects [19].

3 Results and Discussion

3.1 Experiment I: Spatial and temporal correlation (RQ1)

The data acquisition process could achieve a mean relative acquisition error of 27.3% at 0.1 $\frac{m}{s}$, 24.7% at 0.15 $\frac{m}{s}$, and 66% at 0.24 $\frac{m}{s}$. Furthermore, the relative acquisition error at 0.1 $\frac{m}{s}$, at 0.15 $\frac{m}{s}$ and 0.24 $\frac{m}{s}$ is zero, as the acquisition rate relates to the conveyor belt speed. An increase in conveyor belt speed from 0.1 $\frac{m}{s}$ to 0.15 $\frac{m}{s}$ and 0.24 $\frac{m}{s}$ with the same acquisition rate results in a relative acquisition error of 34% (48%).

Speed level	**Speed interval**	**KF**		**KNN**		**HMM**
$[\frac{m}{s}]$	$[\frac{m}{s}]$	PNC	MNC	History	Dist2 Threshold	max Distance
0.10	0.10	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.5$	60	320	70
0.10	0.15	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.5$	60	300	70
0.10	0.24	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.5$	60	320	70
0.15	0.10	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	340	70
0.15	0.15	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	340	70
0.15	0.24	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	340	70
0.24	0.10	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	200	70
0.24	0.15	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	200	70
0.24	0.24	$I(4 \times 4) \cdot 0.9$	$I(2 \times 2) \cdot 0.3$	60	200	70

Table 2: Object tracking parameters for (i) KF, (ii) KNN and (iii) HMM based on speed levels and intervals.

Starting with a speed level of 0.15 $\frac{m}{s}$ while decreasing (increasing) the conveyor belt speed to 0.1 $\frac{m}{s}$ (0.24 $\frac{m}{s}$) leads to a relative acquisition error of 38% (36%). The relative acquisition error at the speed level of 0.24 $\frac{m}{s}$ is 64% when the speed is decreased to 0.15 $\frac{m}{s}$ and 134% when it is decreased further to 0.1 $\frac{m}{s}$. As shown in Figure 3a, the number of particles counted at a speed level of 0.24 $\frac{m}{s}$, with a speed interval of 0.1 $\frac{m}{s}$, is 117. That is, the acquisition rate depends on the conveyor belt speed, which remains constant at this specific speed level and interval. Furthermore, an increase in the conveyor belt speed at a constant acquisition rate, for example, from 0.1 to 0.24 $\frac{m}{s}$, results in 48% of the particles being missed, whereas a decrease in the conveyor belt speed from 0.24 to 0.1 $\frac{m}{s}$ results in more than double the number of particles. Consequently, this demonstrates that pre-defined time intervals probably do not enhance model performance, even when employing the best deep-learning models. In contrast, the use of a rotary encoder couples the temporal component to the conveyor belt speed, resulting in a relative acquisition error of zero for each speed level. That is, the spatial correlation shows neither redundant nor missing particles, whereas edge cases cannot be avoided, as there are 1 at 0.1 $\frac{m}{s}$ and 1 at 0.24 $\frac{m}{s}$. Nevertheless, this enables precise control of the data acquisition

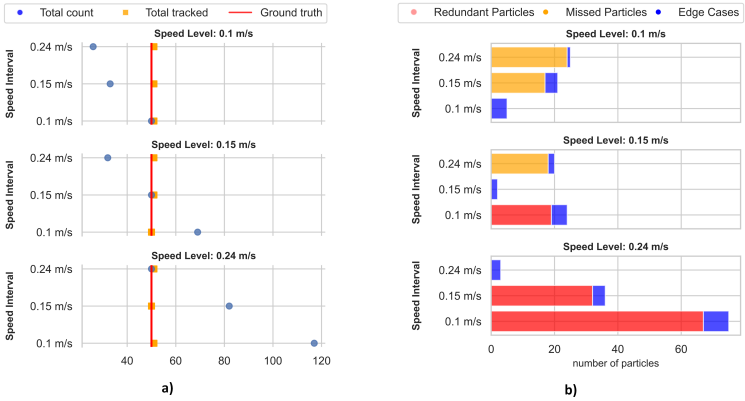


Figure 3: Acquisition without encoder using time intervals (a) Manual counted vs. tracked particles in consecutive frames and (b) redundant and missed particles with edge cases at different speed levels: $0.1\frac{m}{s}$, $0.15\frac{m}{s}$ and $0.24\frac{m}{s}$.

system through automated fine-tuning of the acquisition rate based on the conveyor belt speed. This ensures data integrity of datasets used for deep-learning models, thereby helping to create datasets that are free from bias and noise.

3.2 Experiment II: Object tracking (RQ2)

Additionally, object tracking provides the capability to count particles in consecutive frames within the acquisition system, independently of whether it is controlled by a rotary encoder. Consequently, the combination of KF, KNN, HMM and a rotary encoder ensures the real-time creation of datasets with ground truth information, achieving a count accuracy of almost 100%. As shown in Fig 3.a, the system can track particles with a deviation of ± 1 particles (=total tracked) of the total ground truth. This dual capability serves to reduce labor time and enhance the validation of the data acquired by using knowledge of the ground truth to address class imbalance in classification problems. It also enables targeted strategies to effectively balance different classes. However, increasing the conveyor belt speed above $0.24\ \frac{m}{s}$ requires enhancing the responsiveness of the KF to changes by increasing the PNC.

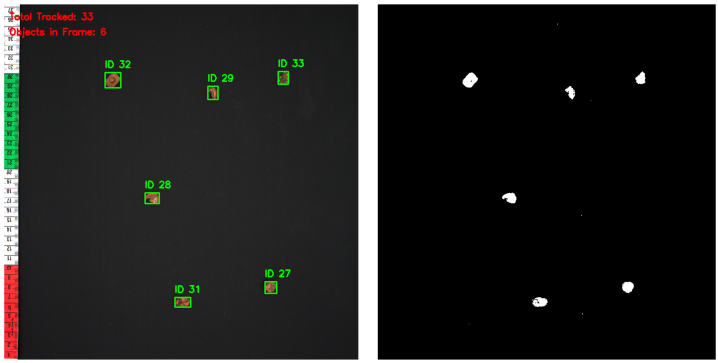


Figure 4: a) object tracking: unique identifiers for each particle by b) extracting particles from background using background subtraction.

Consequently, a reduction in MNC is essential to prevent the system from missing rapid changes. As shown in Fig. 4, background subtraction allows particles to be separated from the background. In this context, a lower Dist2Threshold provides the capability to track fast-moving objects, while the history remains resilient to higher conveyor belt speeds (see Table 2).

4 Conclusion and Outlook

In this study, an automated data acquisition method for dataset creation was investigated by leveraging two approaches. The first approach allows for addressing the issue of temporal and spatial correlation to deal with data sparsity and redundancy by avoiding noisy and biased data. The mean relative acquisition error was observed to be 27.3% at 0.1 $\frac{m}{s}$, 24.7% at 0.15 $\frac{m}{s}$, and 66% at 0.24 $\frac{m}{s}$ when extracting frames using a predefined time interval. In contrast, employing a rotary encoder achieved an impressive accuracy of 100%. On the other hand, object tracking can track particles in real-time to determine the dataset's ground truth with an accuracy of 99% ± 1% to allow for enhanced data validation by leveraging ground truth knowledge to tackle class imbalance in classification problems, ultimately leading to more

reliable and robust deep-learning models. It remains to be seen how performance changes at higher speed levels. Furthermore, a data acquisition system that considers dynamic temporal correlation has the potential to enhance the data quality of high resolution RGB area image datasets, thus allowing for higher generalization capabilities as a trained model does not memorize duplicated objects (i.e., skewing the model). Consequently, the capacity for more effective material assessment is enabled for sensor-based real-time characterization while reducing labor time and biased training, testing, and validation data during pre-processing. Further research is recommended to explore the effects of higher conveyor belt speeds, which is necessary for scaling up to technical or industrial scale aimed at acquiring novel datasets from various recyclable waste streams. Additionally, decoupling of the conveyor belt speed from a rotary encoder could enable the design of an RGB-based speed measurement system for flexible integration in industrial applications.

Acknowledgements

We would like to thank REMONDIS SmartRec GmbH for providing the non-ferrous metal sample material and technical assistance.

References

1. Eurostat, “Generation of waste by waste category, hazardousness and nace rev. 2 activity.” [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/env_wasgen__custom_12053946/default/table?lang=de
2. European Commission, “Eur-lex - 52020dc0098 - en - eur-lex,” 2020. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1583933814386&uri=COM:2020:98:FIN
3. N. Kroell, X. Chen, K. Greiff, and A. Feil, “Optical sensors and machine learning algorithms in sensor-based material flow characterization for mechanical recycling processes: A systematic literature review,” *Waste management (New York, N.Y.)*, vol. 149, pp. 259–290, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0956053X22002628

4. G. Lanza, B. Haefner, L. Schild, D. Berger, N. Eschner, R. Wagner, and M. Zaiß, "In-line measurement technology and quality control," in *Metrology*, ser. Precision Manufacturing, W. Gao, Ed. Singapore: Springer, 2019, pp. 399–433.
5. M. Elasri, O. Elharrouss, S. Al-Maadeed, and H. Tairi, "Image generation: A review," *Neural Processing Letters*, vol. 54, no. 5, pp. 4609–4646, 2022.
6. K. Li, M. I. Fathan, K. Patel, T. Zhang, C. Zhong, A. Bansal, A. Rastogi, J. S. Wang, and G. Wang, "Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations," *PloS one*, vol. 16, no. 8, p. e0255809, 2021.
7. S. E. Whang, Y. Roh, H. Song, and J.-G. Lee, "Data collection and quality challenges in deep learning: a data-centric ai perspective," *The VLDB Journal*, vol. 32, no. 4, pp. 791–813, 2023.
8. H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "A review of video object detection: Datasets, metrics and methods," *Applied Sciences*, vol. 10, no. 21, p. 7834, 2020.
9. B. Barz and J. Denzler, "Do we train on test data? purging cifar of near-duplicates," *Journal of Imaging*, vol. 6, no. 6, 2020.
10. S. Shahinfar, P. Meek, and G. Falzon, ""how many images do i need?" understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring," *Ecological Informatics*, vol. 57, p. 101085, 2020.
11. F. Paredes, C. Herrojo, and F. Martín, "Position sensors for industrial applications based on electromagnetic encoders," *Sensors*, vol. 21, no. 8, 2021.
12. Z.-L. Li, Q. Wu, F. Zeng, and Z.-S. Yue, "3d reconstruction of material transportation status on belt conveyors," in *2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. IEEE, 2015, pp. 168–172.
13. A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezzani, Eds., *Pattern Recognition. ICPR International Workshops and Challenges*, ser. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021.
14. M. Turan, J. Shabbir, H. Araujo, E. Konukoglu, and M. Sitti, "A deep learning based fusion of rgb camera information and magnetic localization information for endoscopic capsule robots," *International journal of intelligent robotics and applications*, vol. 1, no. 4, pp. 442–450, 2017.
15. M. Schäfer, U. Faltings, and B. Glaser, "Does - a multimodal dataset for supervised and unsupervised analysis of steel scrap," *Scientific data*, vol. 10, no. 1, p. 780, 2023.

16. J. Zhang and Y. Peng, "Video captioning with object-aware spatio-temporal correlation and aggregation," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 2020.

17. Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865505003521

18. M. Khodarahmi and V. Maihami, "A review on kalman filter models," *Archives of Computational Methods in Engineering*, vol. 30, no. 1, pp. 727–747, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s11831-022-09815-7

19. A. Frank, "On kuhn's hungarian method—a tribute from hungary," *Naval Research Logistics (NRL)*, vol. 52, no. 1, pp. 2–5, 2005.

Adaptive architectures for semantic segmentation in the field of sensor-based sorting systems

Felix Kronenwett[1], Roman Lehmann[2], Haoxiang Zheng[1], Georg Maier[1], Thomas Längle[1], and Wolfgang Karl[2]

[1] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Fraunhoferstr. 1, 76131 Karlsruhe, Germany
[2] Research Group Computer Architecture and Parallel Processing (CAPP), Institute of Computer Engineering (ITEC), Karlsruhe Institute of Technology (KIT), Haid-und-Neu-Str. 7, 76131 Karlsruhe, Germany

Abstract In recent years, the demand for efficient and accurate sorting solutions across various industries has surged due to the need for enhanced material recovery and sustainability. Sensor-based sorting systems have emerged as pivotal technologies. They employ visual inspection to ensure the precise classification and sorting of bulk materials. Several challenges hinder the potential of deep learning models in industrial systems for image data analysis in complex sorting tasks. Due to different hardware, traditional static deep learning models often fail to handle the dynamic requirements of varying material throughput and execution times, leading to inefficient sorting accuracy. This paper investigates the integration of adaptive architectures for semantic segmentation. These architectures dynamically adjust their computation pathways based on input complexity, optimizing performance and resource utilization. Implementing an architecture with early exit mechanisms improved accuracy, enabling sorting decisions regardless of hardware limitations. Experimental validation using real-world data from sorting plants demonstrates adaptive models' practical applicability and benefits for sensor-based sorting systems.

Keywords Adaptive architectures, semantic segmentation, sensor-based sorting, early-exit mechanisms

1 Introduction

Sensor-based sorting is a real-time application in the field of machine vision. Corresponding systems provide solutions for separating single particles from a material stream and are used in various industries such as recycling, food processing, and waste management [1].

The functional principle of a common kind of sensor-based sorting systems is as follows. Bulk material particles are transported on a conveyor belt. After being discharged from the belt, they pass through an inspection line and are detected by imaging sensors. The recorded data is analyzed to make sorting decisions for the physical separation of different material classes. Individually activated, fast-switching pneumatic valves, arranged orthogonally to the main transport direction of the conveyor belt, are used to eject individual objects. Assuming a constant speed of the bulk material due to conveyor belt transport, a fixed delay time between detection and ejection can be established, resulting in real-time requirements to initiate physical separation promptly [2].

1.1 Motivation

A crucial subsystem of sensor-based sorting systems is the visual inspection unit. For sorting tasks based on color and shape, color sensors are used. In this case, the aim of data evaluation is semantic segmentation, typically by assigning each pixel to a specific material class. However, in demanding sorting tasks with complex materials with similar optical properties, this is challenging.

Particularly in the area of image analysis, significant progress has been made thanks to innovative processes from the field of deep learning [3, 4]. Yet, sensor-based sorting systems face various challenges that require different aspects of adaptability from the neural networks employed that can be summarized as:

- Varying material throughput and load
- Varying hardware and execution times
- Wasteful computations on simple inputs

A particular challenge for the use of deep learning models in sensor-based sorting systems is the long execution time [5]. Generally, the

timing constraints depend on the distance between the sensor and the ejection unit. In addition, the model inference on a graphics processing unit (GPU) is subject to time jitter. The deployment of a deep learning model is therefore difficult, as not only the time constraints can change, but also the execution time due to the different hardware used. In standard neural networks, the execution time remains relatively constant due to static inference, where the compiotational graph and parameters remain fixed after training.

This leads to the concept of adaptive networks, which can adapt their structure or parameters to the input during inference. An adaptive network is often realized using early-exit architectures, where intermediate exits are placed within the neural network. These intermediate results can be utilized at different points during inference, and increasing computation time yields more accurate results. When used in sensor-based sorting, adaptive networks have the advantage that sorting decisions can be made regardless of hardware properties. The time between inspection and ejection can be optimally utilized by making results more accurate through more complex, therefore time-consuming evaluation, but still always reliably delivering timely results.

1.2 Contribution

This paper explores the use of adaptive architectures for semantic segmentation of sensor data, emphasizing their potential to enhance the use of deep learning models in sensor-based sorting systems. By using adaptive networks that adapt their structure depending on the situation, these architectures promise improved accuracy, flexibility, and robustness. They offer a significant advantage by optimizing the time between material inspection and rejection, ensuring timely and precise sorting decisions regardless of hardware limitations.

1.3 Related Work

A deep neural network (DNN) can be formulated as a function $y = f(x) = f_L(f_{L-1}(\dots f_1(x))$, where x is the input data, L is the number of layers in the DNN and f_i is the differentiable operator at layer i. The output of a layer i can be denoted as $h_i = f_i(h_{i-1})$, where h_{i-1} is the output of the previous layer $i-1$.

To convert a DNN to an early-exit architecture, an early-exit branch $c_b(h_b) = y_b$ is placed at every selected branch location $b \in B \subseteq \{1, \ldots, L\}$. c_b can be an intermediate classifier or regressor, producing the early result y_b. We call the original network $f(x)$ the backbone.

Early works for early exits were introduced in classification tasks [6–8]. Later, the idea was further developed and improved. Depending on the goal required, the quality of the decision, the speed of the decision, or both were improved [9]. The field of application has also been extended to various other areas. For example, NLP [10–12] and image segmentation [13–17] have been investigated using this technique.

For image segmentation, MLSNet [14] implements an early exit mechanism by incorporating a hierarchically connected decoder with ResNet [18] as the backbone and multiple classifiers at different layers. Experimental results in the complex Cityscapes [19] and BDD100K [20] datasets, two datasets based on various street scene images with pixel-level annotations, demonstrate that MLSNet surpassed the baseline models in terms of both accuracy and inference speed.

On the other hand, ADP-C (Anytime Dense Prediction with Confidence) [15] uses HRNet [21] as its backbone. When confidence is high enough, further computation for pixels is minimized. Once a pixel reaches sufficient confidence, its prediction is retained for subsequent exits without additional processing. ADP-C shows better performance on the Cityscapes dataset compared to the baseline.

The above works show that the networks were tested on well-known benchmark datasets (Cityscapes, BDD100K, etc.). We want to extend these ideas to real-world sorting system scenarios. In this work, we show how we adjust the network architecture to fulfill the given task and test the networks on real data.

2 Methods and Datasets

2.1 Datasets

Collecting real, labeled data from a sensor-based sorting system is expensive and challenging, and no large open datasets exist. Therefore, the datasets used for training and validation in our study are generated synthetically. Firstly, we record real data for the materials to be sorted in the lab. Individual instances are then algorithmically cut out from

actual recordings using basic image processing algorithms. Instances are randomly selected, rotated, and arranged on a background. Using this synthesis method, the labels required for training can be generated automatically. The synthesized images contain complex situations, i.e., highly occupied images with numerous object overlaps [5]. For the material types to be sorted, we select three groups, and each group consists of two kinds of material: raisins and coffee beens, bricks and sand-lime bricks, and peanuts and hibiscus, as shown in Figure 1. We use real data captured from the device for testing, and the labels are manually generated.

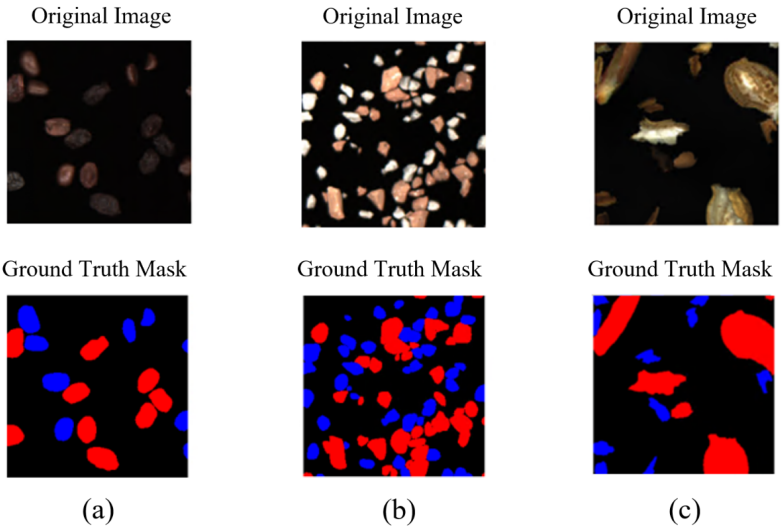


Figure 1: Synthetically generated training dataset: (a) raisins and coffee beens, (b) bricks and sand-lime bricks, (c) peanuts and hibiscus.

2.2 Adaptive Architectures

The difference between our datasets and the datasets of other domains in the literature is their complexity. Our data sets usually consist of very few classes, which means that the neural network supposed to

make the decisions does not have to be as complex as in the literature. We adapt the work from [14] by reducing the complexity. They use an encoder-decoder structure for segmentation. We replaced the backbone (encoder), originally consisting of a ResNet-18 [18], with a much narrower backbone. Now, the backbone consists of only a few layers. We further specify that the final architecture should have three exits. Every exit can be seen as a decoder to fulfill the encoder-decoder structure.

In Figure 2, the backbone is marked with a blue box, and each exit's level is highlighted differently. The feature maps of each level are computed sequentially, starting from level 1. The decoders consist of convolution layers with kernel size 3×3 followed by batch normalization, a ReLU-activation, and max-pooling. For upsampling in the decoder, bilinear interpolation is used, concatenated to the previous activations. The stride indicates the feature map dimension ratio to the final output resolution. The final classification is done by a 1 convolution to produce the multiscale segmentation outputs.

We use FLOP (Floating-Point Operation) as a measurement of the complexity of a neural network. FLOPs indicate how many floating-point operations are needed to calculate the result of the neural network to a given input. The advantage of FLOPs compared to a time measurement is the independence of the underlying hardware. Our architecture consists of two early exits (1 + 2) and the final exit (3) at the End. The first exit needs $0.038G$ FLOPs to calculate the desired output. The second exit needs $0.096G$ FLOPs, and the final (third) exit needs $0.18G$ FLOPs. In Table 1 you can see the difference in FLOPs of our model compared to [14].

Table 1: Comparison of FLOPs between the established and adapted model architecture in gigaflops.

Exit	New Model (FLOPs in G)	Old Model (FLOPs in G)
Exit 1	0.038	3.4
Exit 2	0.096	4.7
Exit 3	0.180	6.5

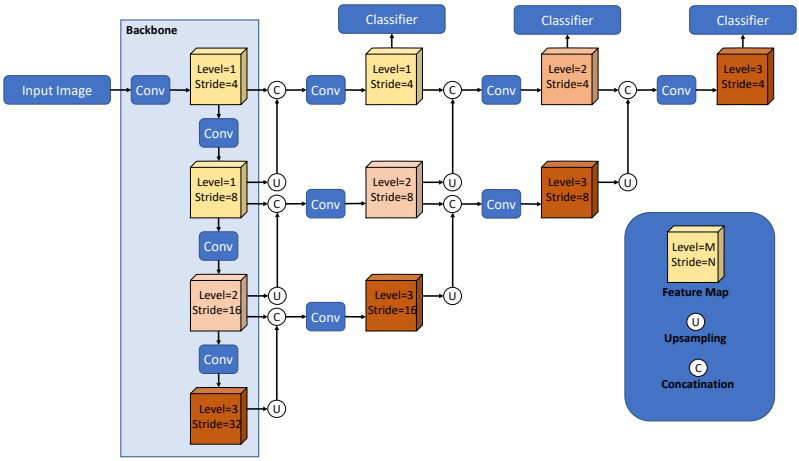


Figure 2: Our adapted MLSNet with a smaller backbone. „Conv“consists of a convolution operator with a kernel size of 3×3, followed by a batch normalization and a ReLU.

3 Experimental Results

In order to better understand the results of our own model, we first train and evaluate three established static architectures for semantic segmentation on the three datasets. This is followed by training and evaluation with the customized, own model. We trained all models on an NVIDIA Tesla T4 GPU with a batch size of 16. We used the ADAM optimizer [22] with a learning rate of 10^{-3} and a weight loss of 10^{-3} for further improvement. The PyTorch Lightning Framework was used to streamline training and monitoring. In particular, we use the mean intersection over union (mIoU) as a metric for our evaluation's semantic segmentation accuracy. With the number of classes n and $p_{x,y}$ representing the number of pixels of class x, that are predicted to class y, we can define mIoU as

$$mIoU = \frac{1}{n} \sum_{i=1}^{n} \frac{p_{i,i}}{\sum_{j=1}^{n} p_{i,j} + \sum_{j=1}^{n} p_{j,i} - p_{i,i}}. \tag{1}$$

The achieved accuracy is analyzed at various exits. The correlation between Accuracy and FLOPs, as a measure of possible interference speed, is specifically considered.

3.1 Baseline Semantic Segmentation Models

Three established semantic segmentation models were trained on the three described datasets to determine achievable segmentation accuracies. Especially regarding the fact that synthetic training data is used, but testing is done with real, manually labelled data The results of the mIoU metric with FCN32s [23], U-Net [24], and SegNet [25] models are shown in Table 2. U-Net is the largest model, with 45 GFLOPs, and it achieves the highest mIoU on the test data for all three datasets. The other two models each performed slightly worse. The results on the datasets also differ; the peanut/hibiscus dataset achieves very high mIoUs, over 90% in some cases. The brick/sand-lime brick data set seems to be the most difficult. This may be due to a high degree of similarity between the two classes, a high difference between training and actual data, or synthetic errors in generating synthetic data.

Table 2: Class-level mIoU (%) of baseline models (numbers in parentheses indicate FLOPs) on various datasets.

Model	Datasets		
	Raisins and Coffee	Bricks and Sand-Lime	Peanuts and Hibiscus
FCN32s (19G)	87.2	73.3	91.2
U-Net (**45G**)	**90.1**	**85.4**	**95.1**
SegNet (11G)	88.0	81.3	93.0

3.2 Early Exit Model

In the following, our own modified MLSNet was trained. All three early exists were trained simultaneously, and the model was then analyzed based on its performance on the test data. Particular attention is paid to the discrepancy between the different outputs. The results can be found in Table 3. FLOPs also provide the number of required computing operations and thus an estimate of the runtime.

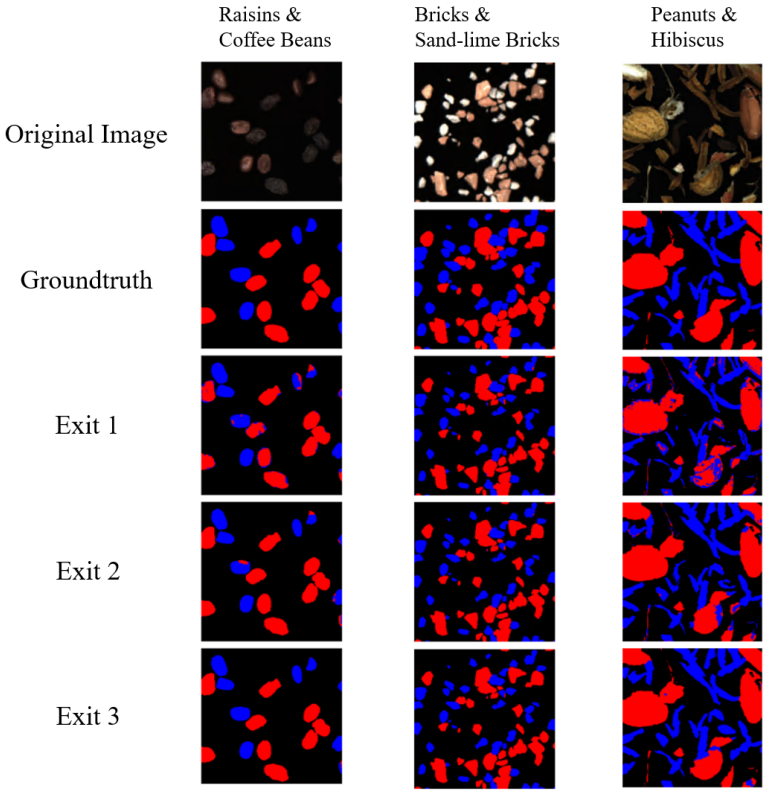


Figure 3: Comparison of the visual quality of the results of the different exits of our model for the three datasets.

Firstly, we look at the performance achieved at different points in time at the various outputs of the models. As expected, the Exit 3 output has the highest segmentation accuracy on all datasets. The achieved mIoUs are slightly lower than the performance of the baseline models, but this can be explained by the significantly smaller size of the architecture due to the lower depth and frequency of the convolutional layers. While the Exit 3 output has 0.18 GFLOPs, the U-Net has 45 GFLOPs.

In addition, increased segmentation performance can be seen with increasing output. The difference between the first and last exit is in the range of Δ7%. However, the FLOPs of the two exits have already more than quadrupled (factor 4.7).

The visual observation of the sorting results at the three exits in Figure 3 shows the improvement in segmentation performance. However, it can be argued that sorting would already be possible with the result at Exit 1 output in a sensor-based sorting scenario.

Table 3: Exit-level mIoU (%) of the i-th exit classifier (numbers in parentheses indicate FLOPs) of our Model on various datasets.

Model	Datasets		
	Raisins and Coffee	Bricks and Sand-Lime	Peanuts and Hibiscus
Exit 1 (0.038G)	80.6	76.8	78.0
Exit 2 (0.096G)	85.1	80.0	91.5
Exit 3 (0.18G)	**87.6**	**82.9**	**93.5**

3.3 Training Strategy Evaluation

For a better understanding of the model architecture, the existing model was split into three individual models, each of which was reduced to just one of the three outputs. The three models were then trained separately to enable a comparison between the separate and end-to-end training of the mixed model. The results in Table 4 show that the combined model with all exits together performs better than the results obtained at the same exit of the individual model. The previous exits can therefore benefit from the later layers and exits during training and enable better generalization. This behavior fits the known phenomenon called *Knowledge Distillation* [26]. As a result, the model is not only more flexible in use but also performs better than individual models of the same size that have been specifically customized.

4 Conclusion and Future Work

This paper shows how early-exiting neural networks as adaptive architecture can perform on sensor-based sorting systems. Such techniques

Table 4: Exit-level mIoU (%) of the i-th exit classifier using separate training strategies (numbers in parentheses indicate FLOPs).

Model	Datasets		
	Raisins and Coffee	Bricks and Sand-Lime	Peanuts and Hibiscus
Exit 1 (0.038G)	72.3	71.4	77.9
Exit 2 (0.096G)	84.5	79.4	90.9
Exit 3 (0.18G)	87.3	80.8	93.0

offer advantages in accuracy and the ability to adjust resource needs. Furthermore, with later exits, earlier exits can profit in training such that these exits can outperform networks with similar FLOPs without early exits. Due to this, training a big network with an early exit has to be considered, even if the big network can't be used in production.

In the future, we are eager to investigate more neural architectures and datasets. We also want to prove this technique in production with a real sorting system.

References

1. Y. Zhao and J. Li, "Sensor-Based Technologies in Effective Solid Waste Sorting: Successful Applications, Sensor Combination, and Future Directions," *Environmental Science & Technology*, Nov. 2022, publisher: American Chemical Society.
2. G. Maier, R. Gruna, T. Längle, and J. Beyerer, "A survey of the state of the art in sensor-based sorting technology and research," *IEEE Access*, vol. 12, pp. 6473–6493, 2024.
3. F. Kronenwett, G. Maier, N. Leiss, R. Gruna, V. Thome, and T. Längle, "Sensorgestützte Charakterisierung von Bauschutt bei hohen Belegungsdichten mittels Deep Learning," in *Vorträge-Konferenzband Zur 16. Recy & DepoTech-Konferenz*. Leoben, Österreich: Montanuniversität Leoben, Lehrstuhl für Abfallverwertungstechnik und Abfallwirtschaft, 2022.
4. F. Kronenwett, G. Maier, N. Leiss, R. Gruna, V. Thome, and T. Längle, "Sensor-based characterization of construction and demolition waste at high occupancy densities using synthetic training data and deep learning," *Waste Management & Research*, p. 0734242X241231410, 2024.
5. F. Kronenwett, G. Maier, and T. Längle, "Potential of deep learning methods for image processing in sensor-based sorting: data generation, train-

ing strategies and model architectures," *10th Sensor-Based Sorting & Control 2024*, p. 45, 2024.

6. E. Park, D. Kim, S. Kim, Y.-D. Kim, G. Kim, S. Yoon, and S. Yoo, "Big/little deep neural network for ultra low power inference," in *2015 International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*, 2015, pp. 124–132.
7. S. Teerapittayanon, B. McDanel, and H. T. Kung, "Branchynet: Fast inference via early exiting from deep neural networks," *CoRR*, vol. abs/1709.01686, 2017.
8. Y. Kaya, S. Hong, and T. Dumitras, "Shallow-deep networks: Understanding and mitigating network overthinking," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 09–15 Jun 2019, pp. 3301–3310.
9. G. Huang, D. Chen, T. Li, F. Wu, L. van der Maaten, and K. Q. Weinberger, "Multi-scale dense convolutional networks for efficient prediction," *CoRR*, vol. abs/1703.09844, 2017.
10. J. Xin, R. Tang, J. Lee, Y. Yu, and J. Lin, "Deebert: Dynamic early exiting for accelerating BERT inference," *CoRR*, vol. abs/2004.12993, 2020.
11. W. Liu, P. Zhou, Z. Zhao, Z. Wang, H. Deng, and Q. Ju, "Fastbert: a self-distilling BERT with adaptive inference time," *CoRR*, vol. abs/2004.02178, 2020.
12. J. Wang, K. Chen, G. Chen, L. Shou, and J. McAuley, "Skipbert: Efficient inference with shallow layer skipping," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 7287–7301.
13. X. Li, Z. Liu, P. Luo, C. Change Loy, and X. Tang, "Not all pixels are equal: Difficulty-aware semantic segmentation via deep layer cascade," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
14. S. Yokoo, S. Iizuka, and K. Fukui, "Mlsnet: Resource-efficient adaptive inference with multi-level segmentation networks," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 1510–1514.
15. Z. Liu, T. Darrell, and E. Shelhamer, "Confidence adaptive anytime pixel-level recognition," *CoRR*, vol. abs/2104.00749, 2021.
16. C.-H. Kung and C.-R. Lee, "Add: A fine-grained dynamic inference architecture for semantic image segmentation," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 4792–4799.

17. A. Kouris, S. I. Venieris, S. Laskaridis, and N. Lane, "Multi-exit semantic segmentation networks," in *Computer Vision – ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Cham: Springer Nature Switzerland, 2022, pp. 330–349.

18. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

19. M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

20. F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, T. Darrell *et al.*, "Bdd100k: A diverse driving video database with scalable annotation tooling," *arXiv preprint arXiv:1805.04687*, vol. 2, no. 5, p. 6, 2018.

21. J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep high-resolution representation learning for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3349–3364, 2021.

22. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

23. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

24. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

25. V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

26. G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015.

Bulky waste classification from a distance: Challenges and first insights

Fridolin Blum, Philipp Meyer, Timo Lange, Matthis Trost, and Tim Tiedemann

Hamburg University of Applied Sciences
Berliner Tor 5, 20099 Hamburg

Abstract Research on autonomous waste detection is primarily focused on conveyor belt systems. Large objects are typically shredded to fit within a conveyor belt system. This work investigates material detection in bulky waste before it is processed by shredders, as sorting large objects before shredding has the potential to improve the recycling process. Multispectral cameras are employed to capture high dynamic range images across the ultraviolet, visible, near-infrared, and shortwave infrared spectra. Deep learning techniques are applied for pixel classification and patch segmentation. We evaluate our approach on a small laboratory dataset consisting of 17 images. The results demonstrate that the multispectral imaging approach outperforms RGB-only imaging, achieving a 10% higher accuracy. Furthermore, the study demonstrates that spectral and spatial convolutions enhance the performance of material detection.

Keywords Bulky waste, material classification, multispectral imaging, deep learning

1 Introduction

Recycling is a key factor for climate protection and resource conservation [1]. Sorting discarded materials (e.g., plastics, wood, metal) is mandatory for recycling waste. Continuous development is taking place for conveyor belt-based sorting technologies, supported by initial robotic systems [2]. A wide variety of sensors are used in short distances and under nearly constant environmental conditions.

In contrast, bulky waste (Germany: approximately 2.5 million tons, with a recycling quota of 56% [3]) is still sorted using manually operated heavy machinery or requires shredding to be processed by conveyor belt systems [4]. Effective automated processes for sorting bulky waste must be developed to increase material recovery for recycling.

To detect materials in a heap of bulky waste, sensors need to capture objects from a distance under overlapping and varying environmental conditions (e.g., changing lighting conditions, dust clouds). In previous work [5], we proposed an approach where an autonomous crane either picks detectable objects or grabs regions with homogeneous materials for automated bulky waste sorting.

This work focuses on material classification for autonomous bulky waste sorting, presenting the following three contributions: (1) A small-scale laboratory material classification dataset[1], (2) A comparison of multispectral and RGB data for material classification[2], (3) An evaluation of the spectral and spatial multispectral image feature space[2].

The remainder of this work is organized as follows: Section 2 presents the background and reviews related work. In Section 3, we describe the used data and how the processing pipeline is built to classify materials, including tested neural network architectures. Section 4 presents evaluation scenarios and their results. We finish with a final conclusion with discussion and future work in Section 5.

2 Background and Related Work

For bulky waste sorting, distinguishing different materials is essential for effective separation. Various techniques are employed on this task. Near-infrared spectroscopy is used to differentiate materials from household waste [6] and to identify plastics in bulky waste [7]. Although both works demonstrate that materials are distinguishable by their spectral reflectance, they are impractical for classifying materials in a pile of coarse bulky waste. Scanning individual objects separately would be too time-intensive for high-throughput sorting systems.

In applications where waste is sorted on conveyor belts, hyperspectral and multispectral imaging techniques are utilized. For example,

[1] Dataset: `https://zenodo.org/records/14616139`
[2] Code: `https://github.com/autosys-lab/msi_material_classification_ocm`

Serranti et al. [8] employ a hyperspectral near-infrared imaging system to separate PE and PP plastics from construction waste, while Casao et al. [9] use a combination of RGB and near- to short-wave-infrared hyperspectral cameras to identify objects in waste streams.

Another approach for waste detection utilizes RGB images in combination with deep learning algorithms [10]. These systems offer the advantage of low-cost RGB imaging systems and pre-trained networks. However, RGB imaging requires material detection based solely on the objects color and shape, whereas multispectral and hyperspectral imaging exploit the differing reflectance properties of materials across various spectra. Material could sometimes be identified based on the object (e.g., wooden pallet), but often objects can be made of various materials. Moreover, objects in waste and recycling contexts are often dirty, fragmented, or deformed, making them difficult to recognize.

Due to inconsistent lighting, contamination with dust and dirt, the deformation of objects in recycling facilities, and the high dimensionality of multispectral imaging systems, deep learning algorithms are applied for material classification instead of classical computer vision approaches. Neural networks have been applied successfully in multispectral remote sensing image classification [11, 12]. The U-Net architecture is designed to classify all pixels in a given image patch [13], taking spatial features into account without the need of generating patches for every pixel.

For material classification this work adapts and combines the U-Net [13], SpectrumNet [12] architectures, and convolutions on the spectral feature space [11].

3 Methods and Materials

In this section, the dataset and data acquisition are described, followed by a description of the applied neural networks.

3.1 Dataset and Data Acquisition

For the acquisition of the multispectral images, we slightly adapted the setup from [5]. Using different cameras for the ultraviolet, visual, near-infrared, and short-wave infrared spectra, bandpass filters

are employed with filter wheels to record small partial spectra, which are registered and then merged into a multispectral image. In contrast to [5], only the first spectrum (190−1100 nm) is used with the ultraviolet camera. In addition, an extra RGB camera is used and the infrared cut filter was removed from the camera for visual/near-infrared camera. The complete multispectral image consists of 14 spectra from 190−1700 nm. To balance different lighting conditions, high dynamic range images are taken with varying exposure times for each channel and combined into a single image.

(a) The multispectral image consists of 14 channels spanning the visible, shortwave infrared, and ultraviolet spectra.

(b) Pixel level label mask with the following color coding: plastic, paper, wood, metal, textiles, minerals

Figure 1: Multispectral image captured in a laboratory setting with small bulky objects (a) and the corresponding label mask (b).

The dataset (see Fig. 1) comprises 17 laboratory multispectral scenes, with 14 training and 3 validation scenes. Lighting is applied from various directions, and every object in the test dataset appears one to five times in the training dataset. In our laboratory setup, the multispectral scenes are captured from a distance of approximately two meters.

Six material classes are included in the dataset: metal, wood, plastic, minerals, textiles, and paper. Deep learning networks are tested for the classification of single-pixel vectors and the segmentation of multispectral scene patches. Due to the fact that the material properties are unknown for many objects (e.g., small fragments, heavy contamina-

tion), not all pixels in an image are labeled. For pixel-vector classification, labeled pixel-vectors are extracted from the multispectral scenes. For segmentation, the multispectral scenes and label masks are divided into patches, while only patches containing at least one labeled pixel are considered for training and validation. To obtain a representative validation dataset, the images with the most labeled objects are chosen for validation. The chosen validation data results in a 7M/1M train-validation split for pixel classification and a 32k/4k train-validation split for patch segmentation.

3.2 Network Architectures

To evaluate the importance of spectral and spatial features, various network architectures are implemented and tested. Spectral-Conv is used as a pixel classifier to evaluate convolutions on the spectral features. Spectrum-UNet is implemented as a patch segmentation network to assess spatial convolutions, and 3DSpectrum-UNet is employed to test a combination of spatial and spectral convolutions.

Table 1: Network architecture for pixel classifier with spectral convolutions

layer name	output size	filter-size/ stride
input	14x1	
Conv	14x128	3/1
Max-Pool	7x128	2/2
Conv	7x128	3/1
Max-Pool	4x128	2/2
Conv	4x256	3/1
Global-Max-Pool	256	
Fully Connected	6	

Spectral-Conv: The network (Table 1) classifies each pixel vector separately. Features are extracted by three convolutional layers and two pooling layers. The classifier consists of a global pooling layer and a fully connected layer. All convolutional layers are ReLU activated, while the fully connected prediction layer uses softmax.

For the **Spectrum-UNet** the SpectrumNet [12] and UNet [13] architectures are combined. SpectrumNet was designed for multispectral remote sensing image classification. It uses Spectrum Modules for feature extraction, each consisting of one bottleneck layer with 1x1 filters, followed by two separate layers: one with 1x1 filters and one with 3x3 depthwise filters, both using the bottleneck layer as input.

UNet [13] is a neural network architecture originally designed for semantic segmentation of medical images without the need to classify

each pixel separately. UNet-like architectures typically consist of two paths: one contracting and one expansive path. The contracting path downsamples the input to learn high-level features, while the expansive path upsamples the outputs to increase the resolution. Between these two branches, skip connections are employed to localize high-level features in higher-resolution layers.

Table 2: Detailed architecture of the Spectrum-UNet expansive path, where (x, y, z) represents a spectrum module with x bottleneck filters, y 1x1 convolutions, and z 3x3 depthwise convolutions. A modified SpectrumNet [12] (see Section 3.2) is used as contraction path.

expansive path			
layer name	input	output size	filters /stride
upsampling1	spectral8	6x6x512	
spectral9	upsampling1	6x6x512	(64, 384, 128)/1
spectral10	spectral9, spectral7	6x6x512	(64, 384, 128)/1
upsamping2	spectral10	12x12x512	
spectral11	upsamping2	12x12x256	(32, 192, 64)/1
upsampling3	spectral11, spectral3	20x20x512	
spectral12	upsampling3	20x20x96	(24, 74, 24)/1
spectral13	spectral12	18x18x96	(24, 74, 24)/1
conv_pred	spectral13	16x16x6	3x3/1

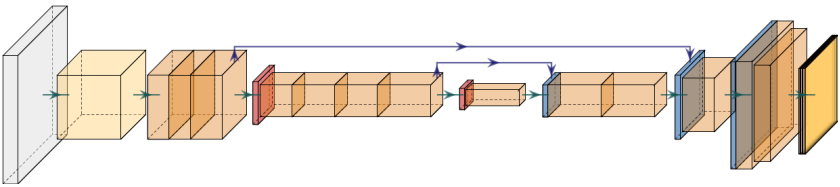


Figure 2: Spectrum-UNet architecture: Patch segmentation with a combination of SpectrumNet [12] and UNet [13]. The network performs semantic segmentation on multispectral patches (gray). The UNet-like network consists of Spectrum modules (orange), pooling layers (red), upsampling layers (blue), and convolutional layers (yellow).

In comparison to the original SpectrumNet, 32 filters are used in the first bottleneck layer instead of 16. This configuration performed

better in previous experiments and could be explained by the number of input channels, as this work uses 14, while the original work used only 10 [12]. In comparison to the original UNet the first horizontal skip connection is removed, as it contributed to overfitting in previous experiments. For a detailed architecture, see Fig. 2 and Table 2.

To extract even more from the spectral features the **3D-Spectrum-UNet** incorporates 3D convolutions (see Fig. 3). In the last downsampling block, three 3D convolutional layers are applied, each with 128 1x1x7 filters, followed by a 1x1x5 max-pooling layer with a stride of 1x1x5. The max-pooling layer is essential; otherwise, the 3D convolutions would expand the feature space too much.

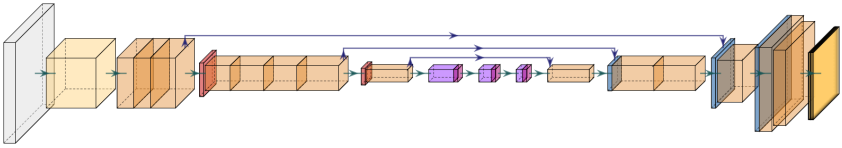


Figure 3: 3D-SpectrumUNet uses the Spectrum UNet architecture in combination with 3D convolutions in the lowest branch (purple), followed by an additional spectral module to merge 3D and 2D features.

4 Experiments and Results

Experimental setups and results are discussed in this section. In all experiments, the Adam optimizer is employed with a learning rate of 0.001. Each training session runs for a minimum of 40 epochs, after which early stopping is applied, and training is halted if the validation accuracy does not improve for ten consecutive epochs. Each training is conducted ten times, and the average accuracy, along with the standard deviation, is reported in the results.

1. RGB vs MSI: This experiment is conducted to evaluate whether the complex multispectral setup is superior to a simpler and more cost-effective RGB setup. For both setups, a Spectrum U-Net is trained for patch segmentation, with both networks being identical except for the input channels: one uses all multispectral channels, while the other utilizes only the RGB channels.

While the RGB-based network, with an accuracy of 51.5% and a standard deviation of 1.4%, still learns and performs better than random classification, the inclusion of multispectral features results in a performance gain of approximately 10%, with an accuracy of 62.1% and a standard deviation of 2.8%. The experiment shows that multispectral imaging is advantageous over RGB for material classification.

2. Learning Spectral and Spatial Features: Since multispectral imaging is advantageous over RGB imaging for material segmentation (see Section 4: RGB vs. MSI), an experiment is conducted to determine how spectral features should be learned.

Two different network architectures are tested for pixel classification: a multilayer perceptron (five fully connected layers: 512-128-64-32-6) and the Spectral-Conv network (see Section 3.2). Two additional networks are evaluated for patch segmentation: the Spectrum-UNet and the 3D-Spectrum-UNet (see Section 3.2). The Spectrum-UNet applies only spatial convolutions, while the 3D-Spectrum-UNet applies both spatial and spectral convolutions.

Table 3: The results show that patch segmentation outperforms pixel classification, and spectral convolutions enhance pixel classification performance.

	spectral-conv	spatial-conv	mean_acc	std_dev
MLP	no	no	54.9	1.5
Spectral Conv	yes	no	56.9	1.6
SpectrumUNet	no	yes	62.1	2.8
3DSpectrumUNet	yes	yes	**63.9**	2.8

The results (Table 3) show that both spectral and spatial convolutions are beneficial for material classification. The impact of applying convolutions on the spectral feature space to enhance spatial convolution-based networks is unclear, as the observed accuracy improvement is smaller than the standard deviation.

3. Leave-One-Out Cross Validation: To test how much the results depend on the chosen train-validation split, the Spectrum-UNet and 3D-Spectrum-UNet models are tested using leave-one-out cross-validation. The 3DSpectrumUNet achieves an mean accuracy of 60.2% and SpectrumUNet achieves 59.9%, but according to the standard deviation of 12%, the difference is not significant and thus their performance can be regarded as equal. The two best and worst performing

validation scenes are the same for both networks. Some objects occur only once in the dataset and generalization to new material surfaces seem to be challenging with our small dataset.

5 Conclusion and Outlook

This work investigates the classification of bulky waste before it is shredded. Therefore, conveyor-belt systems cannot be used, and the classification entirely depends on optical sensors. To gain enough samples from our 17-image dataset to train deep learning models, the images are divided into single pixel vectors (8 million) or patches of 24x24 pixels (36 thousand).

It is shown that multispectral imaging outperforms RGB imaging for bulky waste classification. We also demonstrate that both spatial and spectral features contribute to the classification accuracy. But further experiments have to be made to test if a combination of both is superior to only spatial convolutions. By applying convolutions on the spectral and spatial feature spaces, an overall accuracy of about 63% is achieved with only 14 training images.

In a real-world scenario, we anticipate lower accuracy with the same number of labeled images due to more variable environmental conditions. However, with a sufficiently representative dataset, classifying recurring materials should be achievable. We are currently collecting a significantly larger dataset in laboratory and field settings to check our results with additional data and under more realistic conditions.

Acknowledgements

This work was funded by the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) within the SmartRecycling-Up project (grant no. 67KI21013).

References

1. M. A. T. Alsheyab, "Recycling of construction and demolition waste and its impact on climate change and sustainable development," *Int. J. of Environmental Science and Technology*, vol. 19, no. 3, pp. 2129–2138, Apr. 2021.

2. T. Kiyokawa, J. Takamatsu, and S. Koyanaka, "Challenges for Future Robotic Sorters of Mixed Industrial Waste: A Survey," *IEEE Trans. on Automation Science and Engineering*, vol. 21, no. 1, pp. 1023–1040, Jan 2024.
3. Federal Statistical Office of Germany (Destatis), "Abfallbilanz 2022," Online; accessed 13-Nov-2024, 2024.
4. M. Hyvärinen, M. Ronkanen, and T. Kärki, "Sorting efficiency in mechanical sorting of construction and demolition waste," *Waste Management & Research*, vol. 38, no. 7, pp. 812–816, 2020.
5. T. Lange, A. Babu, P. Meyer, M. Keppner, T. Tiedemann, M. Wittmaier, S. Wolff, and T. Vögele, "First lessons learned of an artificial intelligence robotic system for autonomous coarse waste recycling using multispectral imaging-based method," in *Int. Symp. on Waste Mgmt, Resource Recovery and Sustainable Landfilling 2023*, ser. Proceedings SARDINIA 2023, 2024.
6. Z. Erickson, N. Luskey, S. Chernova, and C. C. Kemp, "Classification of household materials via spectroscopy," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 700–707, 2019.
7. H. Masoumi, S. M. Safavi, and Z. Khani, "Identification and classification of plastic resins using near infrared reflectance," *Int. J. Mech. Ind. Eng*, vol. 6, pp. 213–220, 2012.
8. S. Serranti, A. Gargiulo, and G. Bonifazi, "Classification of polyolefins from building and construction waste using nir hyperspectral imaging system," *Resources, Conservation and Recycling*, vol. 61, pp. 52–58, 2012.
9. S. Casao, F. Peña, A. Sabater, R. Castillón, D. Suárez, E. Montijano, and A. C. Murillo, "Spectralwaste dataset: Multimodal data for waste sorting automation," *arXiv preprint arXiv:2403.18033*, 2024.
10. D. Bashkirova, M. Abdelfattah, Z. Zhu, J. Akl, F. Alladkani, P. Hu, V. Ablavsky, B. Calli, S. A. Bargal, and K. Saenko, "Zerowaste dataset: Towards deformable object segmentation in cluttered scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2022, pp. 21 147–21 157.
11. Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, no. 1, 2017.
12. J. J. Senecal, J. W. Sheppard, and J. A. Shaw, "Efficient convolutional neural networks for multi-spectral image classification," in *2019 International Joint Conference on Neural Networks*, 2019, pp. 1–8.
13. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241.

Towards continual learning with the artificial neural twin applied to recycling processes

Ronald Mendez[1], Andreas Maier[1,2], and Johannes Emmert[1]

[1] Fraunhofer IIS, Fraunhofer Institute for Integrated Circuits IIS, Division Development Center X-Ray Technology,
Flugplatzstr. 75, 90768 Fürth
[2] Pattern Recognition Lab, Friedich-Alexander Universität Erlangen-Nürnberg,
Schlossplatz 4, 91058 Erlangen

Abstract With the increasing use of AI models in industrial processes, comes the need for more training data. Further, material packaging industry is an ever evolving sector, which makes the task of the AI models used by post-consumer package sorting facilities equally dynamic, rapidly outdating the training data, and requiring the generation of new expensive datasets. We propose to apply continual learning in combination with the Artificial Neural Twin (ANT), to continually train models without generating new data manually. We initially train with a small dataset, then, apply Orthogonal Weight Modification with training stimuli from quality control measurements collected by the ANT, and poof through experiments that this can replace the expensive process of dataset generation.

Keywords Continual learning, artificial neural twin, process optimization, backpropagation, orthogonal weight modification

1 Introduction

With rise in use of Artificial Intelligence (AI) models in material characterization, comes the need for datasets suitable for training such models. The publicly available data generally consist only on RGB images, that are not useful for material characterization, where AI models need

to be trained to perform the task of sorting different types of materials (i.e. sorting post consumer packaging, in terms of metallic, non-metallic, or in terms of plastic class: PET, HDPE, LDPE, PP, and PS). Generation of this data is expensive, and generally kept private afterwards [1].

Once trained, the model performs accurately at a specific set of tasks. Extending this set for new tasks (i.e. and extended set of material classes) means generating new datasets that involve both old and new tasks which translates into increased expenses, or, generating datasets only for the new tasks which despite less, is still expensive. This is particularly true for recycling plants, where one of the key operations is material classification, and AI models are expected to perform this task effectively. The recycling plant has no way to predict the new materials that reach the market and thereafter their sorting processes, compromising the accuracy of their models [2].

Re-training AI models on data generated only for the new task enables classification of new materials, but its ability to perform old tasks diminishes or in some cases can be completely overwritten, which is known as catastrophic forgetting [3], which can be avoided using Continual Learning (CL) approaches. A novel decentral process optimization approach, the Artificial Neural Twin (ANT) can be used to continuously collect data (i.e. sensor readings, Quality Control QC) from the industrial process, and propagate it to all the AI models associated to that process [4], which targets the dataset generation problem. Then, we will select the CL approaches applicable for the mentioned data, and test the performance of this training stimuli in combination with CL in a virtual sorting facility.

1.1 Continual Learning

Continual learning is an approach that in the last years has been increasingly used for AI models to become adaptable to changing conditions. It summarizes techniques for AI models to accumulate knowledge about new tasks in an offline or online fashion, without losing the previously acquired [5].

In offline CL, models are trained and tested outside the operation loop with a dataset that include old and new tasks. With the accumulation of new tasks this can become memory expensive, therefore

approaches like the work presented by J. Liu et. al. [6], use generative models that learn in an online fashion, to generate replay data representing old knowledge so that the model can learn from it offline.

Online CL, aim to resemble the human learning process [7]. Throughout life, our brains can acquire, refine, and transfer knowledge, which allows us to interact better with our environment. Even though, the source of learning information might still be in our surroundings, we rarely revisit it once we've learned from it, which resembles a learning process from a continuous stream of data [3]. In online CL, training data is received in small batches that are accessible only once. Therefore, the learning process must be efficient, so that models can extract sufficient knowledge from the online data stream, to perform new tasks effectively [8].

On the model's side, to avoid catastrophic forgetting while training, several CL approaches were developed, following two main strings: regularization, and parameter isolation [6].

Regularization based approaches intent to dictate specific rules to modify the model parameters, although this term traditionally has been used to describe the influence of a penalization parameter introduced in the loss function used for model training. It makes reference to any modification to the training algorithm, made to influence the parameter variations. Example techniques are Batch Normalization, Dropout, Label smoothing, Weight minimization norm, as S. Gonzalez et. al. [9] summarized.

Parameter insulation approaches propose model-architecture based strategies that generally block either neurons or sub-trees in a neural network structure. Knowledge for previous tasks is stored in these blocked sections, avoiding catastrophic forgetting while the free sections learn the new tasks. However, since each task has a corresponding network section, the model's capacity is limited.

1.2 Artificial Neural Twin (ANT)

Inside the industrial production field, given the complex nature of the tasks, the dynamicity of processes, and the diversity of the associated datasets; the use of AI models in combination with CL techniques to optimize such processes has proven to be effective [10]. Before applying AI-powered optimization, the process needs to be digitalized,

which involves the insertion of sensors that feed information to a network of process models that continuously learn how to better monitor and control machinery [11]. The Artificial Neural Twin (ANT) is developed as an industrial digitalization interface, with which, operators can create distributed systems of communicating and self-optimizing process machines.

The ANT uses AI models and decentral data fusion to predict future machine states using Model Predictive Control (MPC) principles [12], then applies the concept of backpropagation (Generally used to train artificial neural networks [13]) on the physically distributed machines to optimize the production process. Backpropagation allows for determining local loss gradients at each digitalized process machine (ANT node) [4]. Afterwards, the ANT nodes update their parameters through gradient descent [14]. This means that loss gradients are the stimuli for the process optimization. While this treats the problem of process optimization, we suggest that the ANT can similarly generate training stimuli for CL.

2 Methods

As mentioned in the previous section the ANT allows for the creation of self-organized distributed systems that apply differentiable data fusion to estimate the state of each section of the system (ANT-node), and through backpropagation and gradient descent optimize parameters of the mentioned nodes.

2.1 ANT data propagation

Exactly as in an Artificial Neural Network (ANN), where neurons are placed in their corresponding layers and connected to each other according to the network structure; the ANT interprets the physical process steps as nodes that are placed in their respective process stage, and links them together accordingly.

During the training process of an ANN, input data is required to infer an output, which is used to calculate loss gradients, that are backpropagated through the network, and through gradient descent used to optimize the network parameters while minimizing the loss func-

tion. The ANT operates similarly, exchanging information between its nodes to infer their corresponding states, then, loss gradients are calculated in terms of the state estimates and backpropagated trough the physical network to optimize ANT-node parameters.

The inference period of the ANT is similar to a decentral data fusion scenario, here the ANT-nodes exchange information about their state as shown in Figure 1, perform data fusion among the gathered information, their parameter settings and previous state estimates, to update their state estimate which is thereafter communicated to the network again. An ANT-node state includes every information relevant for the operation of a process section (i.e. machine parameters, material flow data, maintenance data, etc.).

The data fusion problem conducted during inference period on each ANT node can be expressed as the inference of a maximum a posteriori (MAP) process state estimate, x^*_{MAP}. In a simplified form this MAP estimate can be expressed as a nonlinear least squares problem,

$$\begin{aligned} \boldsymbol{\mu}_{\text{post}} &= \mathbf{x}^*_{\text{MAP}} = \\ &= \arg\min_{\mathbf{x}^*} \left\| \begin{bmatrix} \mathbf{L}_{\text{pr}}\mathbf{M}_{\text{pr}}\mathbf{x}^* \\ \mathbf{L}_{\text{proc}}\mathbf{f}_{\text{proc}}(\mathbf{x}^*) \\ \mathbf{L}_{\text{pred}}\mathbf{f}_{\text{pred}}(\mathbf{x}^*) \\ \mathbf{L}_{\text{p}}\mathbf{M}_{\text{p}}\mathbf{x}^* \\ \mathbf{L}_{\text{s}}\mathbf{M}_{\text{s}}\mathbf{x}^* \\ \mathbf{L}_{\text{c1}}\mathbf{M}_{\text{c1}}\mathbf{x}^* \\ \cdots \end{bmatrix} - \begin{bmatrix} \mathbf{L}_{\text{pr}}\boldsymbol{\mu}_{\text{x},t} \\ \mathbf{L}_{\text{proc}}\mathbf{0} \\ \mathbf{L}_{\text{pred}}\mathbf{0} \\ \mathbf{L}_{\text{p}}\mathbf{b}_{\text{p}} \\ \mathbf{L}_{\text{s}}\mathbf{b}_{\text{s}} \\ \mathbf{L}_{\text{c1}}\mathbf{b}_{\text{c1}} \\ \cdots \end{bmatrix} \right\|^2_2 ,, \end{aligned} \tag{1}$$

where $\mathbf{L}_\square$ are the whitening operators with respect to the covariances $\Gamma_\square = (\mathbf{L}^{\mathbf{T}}{}_\square \mathbf{L}_\square)^{-1}$, $\mathbf{M}_\square$ are observation matrices used to project concatenated state vectors into current state vectors, $\mathbf{f}_\square$ correspond to the information packages present in a node (i.e.: $\mathbf{f}_{proc}(x^*)$ corresponds to state estimates of process model, $\mathbf{f}_{pred}(x^*)$ to the prediction model; here information from the linear observation models, sensors, parameters, and communications can be included as well). A full description of the data fusion process inside the ANT can be found in J. Emmert et. al. work. [4]

During the backpropagation period, the ANT calculates gradients in an arbitrary loss function and propagates them to each node as Figure

1 depicts; then, inside the nodes they are used to tune machine parameters. The optimization step through backpropagation operates along the data fusion graph and thus backpropagates through the data fusion inference in Equation 1, which can be expressed in a differentiable manner.

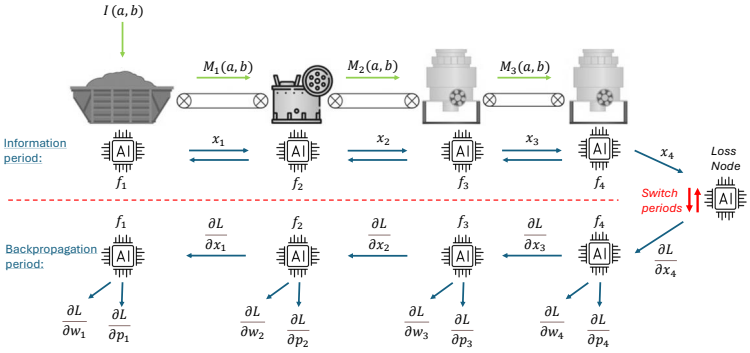


Figure 1: ANT info and backpropagation periods, where f_i is the model of process stage i set inside the ANT-node, with $i \in \{1, 2, \ldots, n\}$, $I(a,b)$ is the input material flow, $M_i(a,b)$ is the output material flow, and M_n is the output of process. $\frac{\partial L}{\partial x_i}$ is the gradient with respect to the process-stage state x_i, $\frac{\partial L}{\partial p_i}$ is the gradient with respect to the process-stage parameters p_i, and $\frac{\partial L}{\partial w_i}$ is de gradient with respect to the model parameters w_i

To control the alternation between the previously described two periods, the ANT uses a loss node. This node receives information from others, as well as quality control (QC) measurements from the user. This means that at any stage of the process, production recordings can be performed and introduced into the digitalized process through the loss node. There, loss functions are evaluated (i.e parameter optimization loss, quality control loss) and gradients are calculated and backpropagated.

2.2 CL approach selection

As shown in Figure 2, for each strategy several approaches have been developed, where the most success was registered in the approaches that combine several strategies [20]. Therefore, we sided towards one

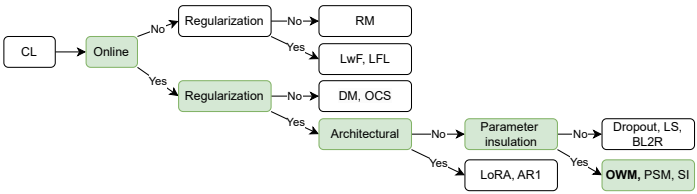


Figure 2: Binary tree for CL approach classification in terms of the type of strategy used. Where: Rehearsal Methos (RM) [6], Learning Without forgetting (LwF) [15], Less forgetting learning (LFL) [16], Diffusion model (DM) [17], Online Coreset Selection (OCS) [18], Low Rank Approximation (LoRA) [19], Architectural regularization (AR1) [20], Label Smoothing (LS) [9], Bayesian L2 regularization (BL2R) [21], Orthogonal Weight Modification (OWM) [22], Parameter soft masking (PSM) [23], Synaptic Intelligence (SI) [24].

of the online approaches with regularization (lower side of the tree). Considering that architectural approaches require the inclusion of additional model parameters where knowledge for every new task can be saved, which leads to memory expensive models with limited number of newly learnable tasks, we decided not to use an architectural approach. Parameter insulation approaches do not overcome the later limitation of architectural approaches, but they do not expand the memory requirements of the model, since their target is to use the parameters already included in the model, to learn new tasks.

Inside this subset, we consider OWM algorithm has a data efficient way to preserve old knowledge. During the model's training, OWM creates a subspace using generated loss gradients. These are then used to determine a projector matrix $P = I - A(A^T A + aI)^{-1} A^T$ where I is the identity matrix, a is a small constant and A is the matrix formed by the loss gradients generated during model's training. This matrix projects all incoming CL loss gradients into a subspace orthogonal to the original training loss gradients, thus preserving the knowledge.

Once the projector P is calculated, the weight update during CL is performed as follows $\Delta W = kP\Delta W^{BP}$. Where, k is the learning rate, and ΔW^{BP} is the gradient generated for model training.

This approach assumes that most models are significantly underdetermined, which translates on parameters available for learning new tasks. Every new task will be learned in a subsection of these pa-

rameters, which results in a limited number of tasks that a model can potentially learn with this approach. Despite its limitations, this approach encloses, a clever manner of preserving old knowledge, gradient based learning which avoids storing input data, and a higher use of full model's capacity.

2.3 Experimental setup

As mentioned in the previous section, we chose OWM as CL algorithm. For this, access to a continuous source of data was necessary, for which we simulated a section of a bulk material recycling process. To do that, we used the physics engine of Unity game development framework.

This controlled environment provides us with a material stream that emulates the complexity of a real material sorting task. Inside the simulated process, we begin with a machine ("siever") that sorts materials by size (Small, Medium, Large), on each outlet, we placed conveyor belts that emulate the delay between sorting machines, and after each of them we placed magnetic sorters, which sort materials as metallic and non-metallic (Flow diagram of this process in J. Emmert et. al. [4]).

For the process, seven different types of articles were simulated (Metallic: cans and caps. Paper: cups, balls and newspapers. Plastic: caps, and bottles), which travel inside the described environment.

The siever model combines, linear functions that describe the probability of an object exiting on one of the outlets, with a convolution of past input time series and an impulse response that models the mixing behavior inside the sieving drum. The magnetic sorters use a small ANN as process model to link output and input mass flows, and the flow sensors use a polynomial function i.e. $y(x) = \sum_{i=1}^{n} x_i * w_i$, where x_i represent the number of articles of type i registered by a flow sensor in a time step, and w_i is the corresponding mass of the article i.

The initial task of this model is set to learn the masses of all other articles w_i where $i \in \{2, 3, .., 7\}$, except for the mass of metallic cans, which were not present during this training period. Here the projector P is learned as well. Then, we embed the model $y(x)$ and its projector into the ANT, and include the metallic cans in the controlled environment. During CL period, we feed the ANT system with mass flow quality control measurements, which are used to calculate loss gradients that are later backpropagated to the flow sensor where they are

used by our CL algorithm to learn the mass of the metallic cans, allocating this knowledge in the previously unused model parameter w_1.

In this simple example every step of the process can be directly seen and evaluated. Afterwards we used the same setup for the ANN process model of the magnetic sorters, where we initially trained on very few samples that do not include cans material mass flows, resulting in a low fidelity model predicting inaccurate mass flows. During CL this model is expected to self improve on similar quality control measurements.

3 Discussion

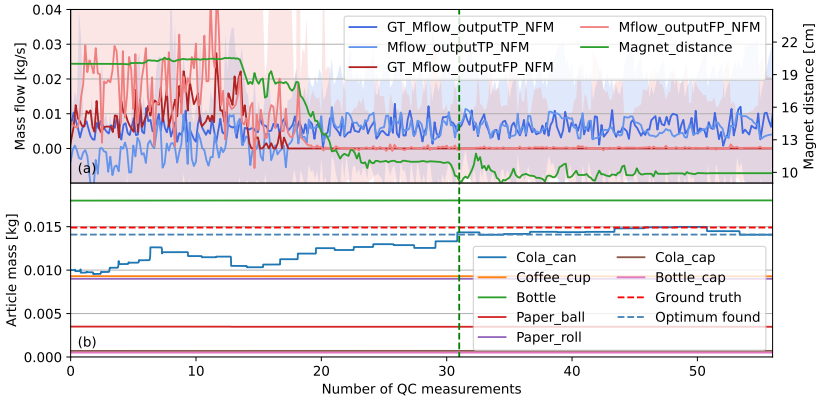


Figure 3: (a) Material mass flows registered by the sensor during CL while recording QC measurements (1 per min.). Non-metallic material flow exiting on the non-metallic outlet (Mflow_outputTP_NFM), metallic material exiting the non-metallic outlet (Mflow_outputFP_NFM), their corresponding ground truth (GT), and the machine parameter (Magnet_distance), (b) sensor model parameters representing article masses.

As mentioned in the first experiment we apply OWM algorithm and gradients from QC measurements, to a mass flow sensor. Figure 3 (a) depicts how after 31 QC measurements (Vertical dashed green line) the ANT process optimization converges to a magnet distance of about 10cm, at this point, the sensor CL has converged as well, registering

mass flows similar to GT, with respective uncertainty values (shaded areas in corresponding colors). Figure 3 (b) shows how convergence is reached as the mass of a cola can is learned, converging to 0.0140kg (dashed blue line) compared to the GT value 0.0149kg (dashed red line). Knowledge preservation granted by OWM approach is visible in the other parameters of the model, which remained constant (Straight horizontal lines in Figure 3 (b)).

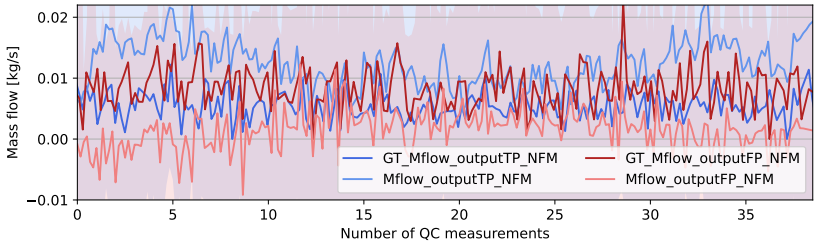


Figure 4: Material mass flows registered in magnetic sorter during CL while recording QC measurements (1 per min.). Non-metallic material flow exiting on the non-metallic outlet (Mflow_outputTP_NFM), metallic material exiting the non-metallic outlet (Mflow_outputFP_NFM) and their corresponding ground truth (GT).

In second experiment, CL is applied to magnetic sorter model, while sensors use fully converged static models. Figure 4 shows how OWM influences the model and thereafter its mass flow predictions as the number QC measurements increases, unfortunately the model did not converge , probably due to an unsuitable ANN architecture for OWM. Nevertheless, this and the previous experiment are proof of the ANT's capability to provide an environment for the implementation of CL algorithms, and stimuli for AI models to learn new tasks through continual learning. Thereafter we positively consider that large AI models can learn new tasks through CL in combination with the ANT; factors to consider are: larger initial training datasets, a dynamic learning rate, different CL algorithms among others, which will be evaluated in future research.

References

1. N. Basedow, K. Hadasch, M. Dawoud, C. Colloseus, I. Taha, and D. Aschenbrenner, "Open data sources for post-consumer plastic sorting: What we have and what we still need," *Procedia CIRP*, vol. 122, pp. 1042–1047, 2024, 31st CIRP Conference on Life Cycle Engineering. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212827124001847
2. K. Özkan, S. Ergin, Şahin Işık, and İdil Işıklı, "A new classification scheme of plastic wastes based upon recycling labels," *Waste Management*, vol. 35, pp. 29–35, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0956053X14004528
3. G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural Networks*, vol. 113, pp. 54–71, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608019300231
4. J. Emmert, R. Mendez, H. M. Dastjerdi, C. Syben, and A. Maier, "The artificial neural twin — process optimization and continual learning in distributed process chains," *Neural Networks*, vol. 180, p. 106647, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608024005719
5. L. Wang, X. Zhang, H. Su, and J. Zhu, "A comprehensive survey of continual learning: Theory, method and application," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 8, pp. 5362–5383, 2024.
6. J. Liu, W. Li, X. Yue, S. Zhang, C. Chen, and Z. Wang, "Continual offline reinforcement learning via diffusion-based dual generative replay," 2024. [Online]. Available: https://arxiv.org/abs/2404.10662
7. A. Chrysakis and M.-F. Moens, "Online continual learning from imbalanced data," in *Proceedings of the 37th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, H. D. III and A. Singh, Eds., vol. 119. PMLR, 13–18 Jul 2020, pp. 1952–1961. [Online]. Available: https://proceedings.mlr.press/v119/chrysakis20a.html
8. Z. Mai, R. Li, J. Jeong, D. Quispe, H. Kim, and S. Sanner, "Online continual learning in image classification: An empirical survey," *Neurocomputing*, vol. 469, pp. 28–51, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231221014995
9. S. Gonzalez and R. Miikkulainen, "Effective regularization through

loss-function metalearning," 2021. [Online]. Available: https://arxiv.org/abs/2010.00788

10. J. Chen, J. He, F. Chen, Z. Lv, J. Tang, W. Li, Z. Liu, H. H. Yang, and G. Han, "Towards general industrial intelligence: A survey on iiot-enhanced continual large models," 2024. [Online]. Available: https://arxiv.org/abs/2409.01207
11. D. T. Matt, G. Pedrini, A. Bonfanti, and G. Orzes, "Industrial digitalization. a systematic literature review and research agenda," *European Management Journal*, vol. 41, no. 1, pp. 47–78, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263237322000019
12. H. Scheu and W. Marquardt, "Sensitivity-based coordination in distributed model predictive control," *Journal of Process Control*, vol. 21, no. 5, pp. 715–728, 2011, special Issue on Hierarchical and Distributed Model Predictive Control. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959152411000163
13. A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 86–101, 2019, special Issue: Deep Learning in Medical Physics. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S093938891830120X
14. S. ichi Amari, "Backpropagation and stochastic gradient descent method," *Neurocomputing*, vol. 5, no. 4, pp. 185–196, 1993. [Online]. Available: https://www.sciencedirect.com/science/article/pii/092523129390006O
15. Z. Li and D. Hoiem, "Learning without forgetting," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 614–629. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-46493-0_37
16. H. Jung, J. Ju, M. Jung, and J. Kim, "Less-forgetting learning in deep neural networks," 2016. [Online]. Available: https://arxiv.org/abs/1607.00122
17. R. Gao and W. Liu, "DDGR: Continual learning with deep diffusion-based generative replay," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., vol. 202. PMLR, 23–29 Jul 2023, pp. 10 744–10 763. [Online]. Available: https://proceedings.mlr.press/v202/gao23e.html
18. J. Yoon, D. Madaan, E. Yang, and S. J. Hwang, "Online coreset selection for rehearsal-based continual learning," 2022. [Online]. Available: https://arxiv.org/abs/2106.01085

19. E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," 2021. [Online]. Available: https://arxiv.org/abs/2106.09685

20. D. Maltoni and V. Lomonaco, "Continuous learning in single-incremental-task scenarios," 2019. [Online]. Available: https://arxiv.org/abs/1806.08568

21. J. L. Ticknor, "A bayesian regularized artificial neural network for stock market forecasting," *Expert Systems with Applications*, vol. 40, no. 14, pp. 5501–5506, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417413002509

22. B. C. Guanxiong Zeng, Yang Chen and S. Yu, "Continual learning of context-dependent processing in neural networks," *Nature Machine Intelligence*, pp. 2522–5839, 2019. [Online]. Available: https://doi.org/10.1038/s42256-019-0080-x

23. T. Konishi, M. Kurokawa, C. Ono, Z. Ke, G. Kim, and B. Liu, "Parameter-level soft-masking for continual learning," 2023. [Online]. Available: https://arxiv.org/abs/2306.14775

24. F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," 2017. [Online]. Available: https://arxiv.org/abs/1703.04200

Supervised and unsupervised textile classification via near-infrared hyperspectral imaging and deep learning

Maria Kainz[1], Johannes K. Krondorfer[2], Malte Jaschik[1], Maria Jernej[1], and Harald Ganster[1]

[1] JOANNEUM RESEARCH Forschungsgesellschaft mbH, DIGITAL, Steyrergasse 17, 8010 Graz, Austria
[2] Graz University of Technology, Institute of Experimental Physics Petersgasse 16, 8010 Graz, Austria

Abstract Recycling textile fibers is critical to reducing the environmental impact of the textile industry. Hyperspectral near-infrared (NIR) imaging combined with advanced deep learning algorithms offers a promising solution for efficient fiber classification and sorting. In this study, we investigate supervised and unsupervised deep learning models and test their generalization capabilities on different textile structures. We show that optimized convolutional neural networks (CNNs) and autoencoder networks achieve robust generalization under varying conditions. These results highlight the potential of hyperspectral imaging and deep learning to advance sustainable textile recycling through accurate and robust classification.

Keywords Textile recycling, textile fibres, hyperspectral imaging, near infrared spectroscopy, deep learning, one-dimensional convolutional neural network, autoencoder, classification

1 Introduction

The textile industry is a major contributor to global pollution, mainly due to its resource-intensive production processes and massive waste generation. A significant amount of textiles are discarded each year, even though they could be recycled. To counteract the negative effects,

it is essential to improve textile recycling rates, as the European Union's guidelines require 100% of textile waste to be recycled by 2025. [1,2]

The hyperspectral near-infrared (NIR) imaging technology offers a potential solution for efficient recycling strategies. By analyzing the spectral signatures of different fibers – whether natural, artificial, or synthetic – this technology can accurately classify textiles based on their chemical composition. When combined with advanced classification algorithms, it enables textile recognition and efficient sorting of various textile types. [2–4] While standard machine learning approaches have been successfully applied to hyperspectral data [5] in the past, deep learning (DL) methods typically outperform these standard methods, due to the high versatility inherent to DL models. Especially, DL models based on convolutional neural networks (CNNs), have shown strong performance in previous work, demonstrating their capacity for accurate classification and generalization. [6–8]

In this study, we investigate the application of DL algorithms for textile classification. We explore two key use cases. First, we focus on supervised classification, where unseen textile samples are classified into one of 12 pre-defined categories – including pure textile fibers such as cotton, polyester, silk, wool, viscose, and nylon, cotton-polyester blends in various mixing ratios as well as cotton-elastane and linen-polyester-viscose blends. Second, we examine unsupervised classification, where only data from a single textile category are available for training. In this scenario, an autoencoder network is employed to identify normal and anomalous data based on reconstruction performance, detecting textiles that deviate from the class norm. This method proves useful in situations where not all potential textile classes are available for training, a common challenge in real-world applications.

To ensure realistic performance evaluation, we use test sets that differ from the training sets in aspects such as color, weave and yarn thickness, providing a more robust assessment of the model's out-of-sample performance and generalization. Our results demonstrate that optimized convolutional neural networks (CNNs) and autoencoder networks achieve robust generalization under varying conditions, showcasing the potential of combining hyperspectral imaging with deep learning for automated textile sorting.

2 Textile data, Acquisition and Preprocessing

Textile categorization: First, we describe the textile data recorded in this study and explain how they are categorized and summarized in Table 1. Primarily, the textiles are categorized by their *fiber type*, which represents the basic raw material of a textile. Fibers are divided into natural, artificial and synthetic fibers. Natural fibers either come from plants, such as cotton or linen, or from animals, such as wool or silk. Artificial fibers are derived from materials of natural origin. However, the production of yarn takes place through chemical transformation (e.g. viscose from cellulose). Unlike the previous ones, synthetic fibers (e.g. polyester) are produced by chemical synthesis. Combinations of different fiber types are called blended fibers. Here, we aim to classify textiles according to their fiber type, to facilitate the recycling of fibers. In addition to the fiber type, the *structure type* or *production type*, provides information about the processing of the textile fibers into textile end products. Fibers are spun into yarns, which are then processed into woven or knitted fabrics, which can differ in yarn properties, thread density, fabric type and other manufacturing techniques. There are also non-woven fabrics that are made directly from fibers and not from yarn. Fiber type and structure type together result in the *textile type* (*fiber type* + *structure type* = *textile type*). Textile samples within a textile type, i.e. a row in Table 1, differ purely in their coloring. This is also illustrated in Figure 1 for different types of polyester.

Hyperspectral imaging: The textile samples were analyzed on a conveyor belt with a pushbroom hyperspectral camera (*ImSpector N17E* spectograph [9] and a *PhotonFocus* camera [10]). This setup was used to examine the spectrum between 990 nm and 1700 nm (400 channels).

Figure 1: Comparison of polyester structure types (Panne Velvet, Satin, Fleece).

Table 1: Specification of textile data and assignment of objects to different data sets. The objects are differentiated by their fiber- and structure-type. Each line corresponds to one textile-type, that can entail different colors. The first column shows abbreviations used throughout the manuscript. D1 - D6 indicate the assignment of textiles to different data sets, that are studied in this paper. D1 is used for training, validation, and testing of the supervised classification model. D2 and D3 are used to evaluate the generalization and out-of-sample performance of the supervised model, differing in color and structure, respectively. Training of the unsupervised classification is performed on D4, and the performance on pure textiles and blends is tested via D5 and D6, respectively.

	Fiber		Structure		Sup. Cl.			Unsup. Cl.		
Textile	Type	%	Type	Nr.	D1	D2	D3	D4	D5	D6
C1	Cotton (plant-based)	100	Flag Fabric	1	45	15		60		
C2	Cotton	100	Twill	2			29			29
C3	Cotton	100	Cord	3			22			22
C4	Cotton	100	Canvas	4			32			32
C5	Cotton	100	Cretonne	5			37			37
P1	Polyester (synthetic)	100	Panne Velvet	1	37	12			49	
P2	Polyester	100	Satin	2			22		22	
P3	Polyester	100	Fleece	3			47		47	
S1	Silk (animal-based)	100	-	1	1	1			2	
S2	Silk	100	-	2			1		1	
L1	Linen (plant-based)	100	-	1	3	1			4	
N1	Nylon (synthetic)	100	Organza	1	11	4			15	
W1	Wool (animal-based)	100	Walkloden	1	41	14			55	
V1	Viscose (artificial)	100	Voile	1	28	9			37	
VLP1	Viscose/Linen/Polyester	40/30/30	-	1	25	8			33	
CP1 9:1	Cotton/Polyester	90/10	Frottee	1	19	6			25	
CP1 8:2	Cotton/Polyester	80/20	Sweatshirt	1	24	8			32	
CP1 7:3	Cotton/Polyester	70/30	Sweatshirt	1	38	12			50	
CE1	Cotton/Elastane	95/5	Jersey	1	62	21			83	
CE2	Cotton/Elastane	95/5	French Terry	2			48		48	

Dark textiles: It was found that black and dark-grey samples of textile type C5 and dark-blue samples of textile type CP1 9:1 absorb almost all radiation, thus providing no information about the fiber type in the recorded reflection spectrum. These samples are excluded from the dataset. This phenomenon is also mentioned in Ref. [2], where it is attributed to the complete absorption of radiation by carbon pigments. For all other dark-colored samples, this phenomenon was not observed.

Preprocessing: To mitigate the influence of the light source's spectral signature and the background reflection pattern, the measured intensities are calibrated by subtracting the dark reference and normalizing with respect to the white reference. Furthermore, we apply Stan-

dard Normal Variate (SNV) to remove baseline drift, followed by mean smoothing (i.e. averaging 5x5 pixel into one spectrum) to reduce noise, and apply a first derivative Savitzky-Golay-Filter to emphasize subtle spectral features, ensuring improved data quality for subsequent analysis. The combination of SNV+Mean+SG showed the best performance during hyperparameter optimization for both investigated scenarios.

3 Supervised Classification of Hyperspectral Textile Data

Datasets: For the supervised classification task, three datasets were prepared to evaluate the model's performance and generalization capabilities, as summarized in Table 1. We perform stratified splitting of D1 (1300 spectra per class) into training (60%), validation (20%), and test (20%) subsets. D2 consists of spectral data from textile samples exhibiting visible color variations compared to D1. And D3 is composed of spectral data from textiles differing in structural properties such as yarn thickness, thread density, and weave pattern. This dataset design enables a robust evaluation of model generalization under different conditions of variability.

Model: The final model employs a one-dimensional convolutional neural network optimized for hyperspectral data. The model is obtained by variation of different hyperparameters, such as number of layers, activation functions, normalization, and preprocessing. The final architecture employs the (SNV+Mean+SG) preprocessing discussed in Section 2, and an input layer size of 400, corresponding to the number of channels in the hyperspectral dataset. This is followed by two 1d-convolutional layers with ReLU activation function and a kernel size of 5. The number of filters is set to 20 and 32 respectively. Then a fully connected dense layer is used, with a batch normalization layer and a dropout layer with rate 0.5, followed by a ReLU activation and a 12 class softmax output layer.

Training: The network is trained using the categorical cross-entropy loss function and optimized with the Adam optimizer [11] with initial learning rate (LR) of 0.001, and a mini-batch size of 128. A LR-scheduler is used to reduce the LR by a factor of 0.2 when the valida-

tion loss failed to improve for five consecutive epochs. Early stopping with a patience of seven epochs is employed to prevent overfitting.

Results: To assess the model's performance, we compute pixel-based accuracy (i.e. the fraction of correctly classified pixels across all samples) and object-based accuracy (determined by majority voting of classified pixels assigned to an object), and display the corresponding confusion matrices in Figure 2. On the test set of D1 all samples are classified correctly. Also on D2 and D3 exceptional performance can be observed, only with slight misclassification rates for some cotton blends and structurally different cotton fibers for a pixel-based analysis. With respect to an object-based analysis, however, perfect classification is observed across all data sets, showing the exceptional generalization and out-of-sample performance of this comparably simple model.

4 Unsupervised Classification Using Autoencoders

Datasets: For the unsupervised classification task, datasets are structured to evaluate the autoencoder's ability to filter target textiles. As test case, we choose cotton as target textile. We train the model on D4 (1300 spectra) of Table 1, with a stratified split into training (60%), validation (20%), and test (20%) subsets. We test the detection capabilities in D5, consisting of non-cotton textiles and different cotton blends, and D6, comprised of cotton textiles, differing in structural properties, to assess robustness.

Model: The final autoencoder model, designed for target textile detection in hyperspectral data, is optimized through systematic hyperparameter variations, including the number of layers, activation functions, and units per layer. It employs preprocessing steps (SNV+Mean+SG) discussed in Section 2. The input and output layers are fixed at size 400, i.e. the number of spectral channels. The encoder compresses the input into a 20-dimensional latent space using two fully connected layers with 100 units and ReLU activations. The decoder mirrors the encoder, consisting of two layers with 100 units, reconstructing the input via a linear activation in the final layer.

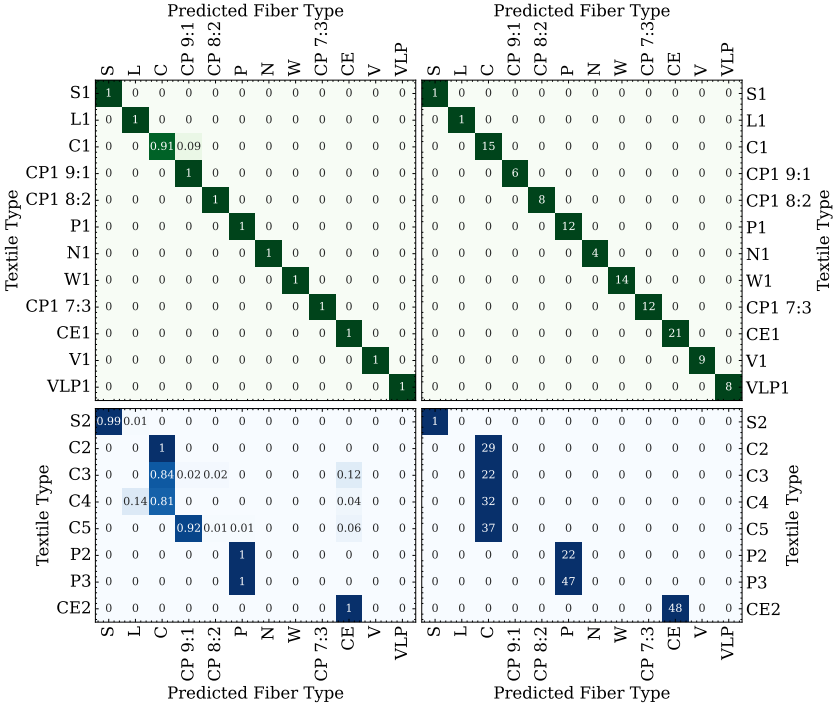


Figure 2: Confusion matrices for supervised classification of pixel-based (left) and object-based (right) analysis on D2 (top) and D3 (bottom). On the test set of D1 all data are correctly classified. Also on D2 and D3 exceptional performance can be observed, with slight misclassification rates for cotton blends or structurally different cotton fibers. With respect to an object-based accuracy, however, all textiles are classified correctly, despite the different structure and coloring.

Training: The autoencoder is trained to minimize the mean squared error (MSE) between input and reconstructed data. The Adam optimizer [11] is used for training with an initial LR of 0.001, and a mini-batch size of 16. A LR-scheduler is used to reduce the LR by a factor of 0.5 if the MSE does not improve over five consecutive epochs. Early stopping, monitoring the MSE of the validation set, is implemented with a patience of seven epochs to prevent overfitting.

Results: To assess the model's performance, we compute the reconstruction error (RE) on the different data sets. The classification is based on the observed RE value and the 95% quantile of the RE distribution over D4. Samples with RE lower (higher) than this threshold value are classified as cotton (non-cotton). The results are summarized in Figure 3, where the distribution of the RE over D4, D5 and D6 is shown, and Table 2, where pixel- and object-based accuracies of the textiles are listed. The autoencoder effectively identifies non-cotton textiles based on reconstruction error, with significant separation observed for synthetic fibers (e.g., polyester and nylon). Cotton-polyester blends exhibited intermediate RE values, making them more challenging to classify correctly. Structurally different cotton types from D6 show a wider RE distribution than on the training set and only moderate generalization is observed, leading to reduced accuracy.

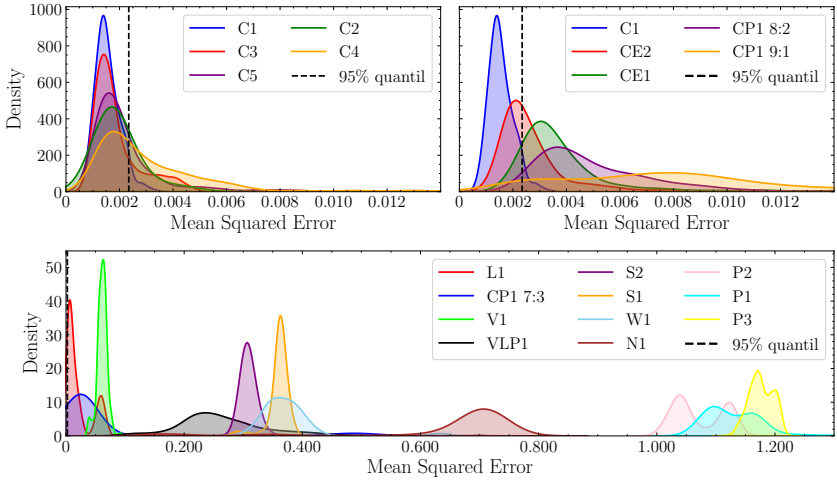


Figure 3: Distribution of reconstruction error (RE) for different textiles. The decision threshold is based on the 95% quantile of the RE distribution over D4. Pure textiles from D5 can be well distinguished and yield 100% recognition (bottom), whereas cotton blends are much harder to distinguish and only reduced recognition can be observed (upper left). The RE distribution of different cotton types from D6 is much wider than on the training set and only moderate generalization is observed, leading to reduced accuracy. The corresponding pixel- and object-based accuracies are given in Table 2.

Table 2: Pixel- and object-based accuracies (in %) of unsupervised textile classification.

	P1	S1	L1	N1	W1	V1	CP1 9:1	CP1 8:2	CP1 7:3	CE1	VLP1	P2	P3	S2	CE2	C2	C3	C4	C5
px	100	100	100	100	100	100	90	98	100	90	100	100	100	100	41	76	56	87	77
obj	100	100	100	100	100	100	92	100	100	100	100	100	100	100	42	86	69	86	83

5 Conclusion and outlook

In this study, we investigated supervised and unsupervised deep learning models and their generalization capabilities. For this purpose, we tested the performance of the developed models on data sets that differ from the training sets in aspects such as color and textile structure.

The supervised classification approach showed exceptional performance for hyperspectral textile data, achieving high in-sample and out-of-sample accuracy. However, its success depends on the availability of a representative and diverse dataset for different fiber types, which can be challenging, due to limited availability of all relevant fibers.

In contrast, the unsupervised approach using autoencoders showed lower performance but is less restrictive in terms of dataset requirements, as only one fiber type is needed for training. This makes it a promising alternative for scenarios where comprehensive labeled datasets are not available. Due to reduced performance on blends, this technique is better suited for pre-sorting of different fibers.

Overall, the results highlight that even relatively simple models, when suitably optimized and combined with preprocessing, can achieve robust classification and generalization performance. This provides a solid foundation for further advancements in automated textile analysis and recycling applications.

Acknowledgements

Funded by the Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation, and Technology (BMK).

References

1. European Commission and Joint Research Centre and S. Donatello, J. Danneck, C. Löw, D. Watson, A. Konstantas, S. Trzepacz, R. Liu, A. Köhler and G. Faraca, *Circular economy perspectives in the EU textile sector – Final report*. Publications Office, 2021.
2. A. Tischberger-Aldrian, H. Stipanovic, N. Kuhn, T. Bäck, D. Schwartz, and G. Koinig, "Automated textile sorting—status quo, challenges and perspectives," *Österreichische Wasser- und Abfallwirtschaft*, vol. 76, 11 2023.
3. J. Huang, H. He, R. Lv, G. Zhang, Z. Zhou, and X. Wang, "Non-destructive detection and classification of textile fibres based on hyperspectral imaging and 1d-cnn," *Analytica Chimica Acta*, vol. 1224, p. 340238, 2022.
4. J. Rodgers and K. Beck, "Nir characterization and measurement of the cotton content of dyed blend fabrics," *Textile Research Journal*, vol. 79, 04 2009.
5. A. Plaza, *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, pp. S110–S122, 2009, imaging Spectroscopy Special Issue. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425709000807
6. W. Du, J. Zheng, W. Li, Z. Liu, H. Wang, and X. Han, "Efficient recognition and automatic sorting technology of waste textiles based on online near infrared spectroscopy and convolutional neural network," *Resources, Conservation and Recycling*, vol. 180, p. 106157, 05 2022.
7. R. Grewal, S. Singh Kasana, and G. Kasana, "Machine learning and deep learning techniques for spectral spatial classification of hyperspectral images: A comprehensive survey," *Electronics*, vol. 12, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/3/488
8. M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 279–317, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271619302187
9. Specim. (n.d.) Imspector n17e. Accessed: 2025-01-08. [Online]. Available: https://www.specim.com/products/imspector-n17e
10. Photonfocus. (n.d.) Mv3-d640i-m01-144-g2 camera. Accessed: 2025-01-08. [Online]. Available: https://www.photonfocus.com/de/produkte/kamerafinder/kamera/mv3-d640i-m01-144-g2
11. D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 12 2014.

CNN-based copper reduction in shredded scrap for enhanced electric arc furnace steelmaking

Gerald Koinig[1], Melanie Neubauer[2], Walter Martinelli[3], Yves Radmann[3], Nikolai Kuhn[1], Thomas Fink[1], Elmar Rückert[2], and Alexia Tischberger-Aldrian[1]

[1] Montanuniversitaet Leoben, AVAW, Leoben, Austria,
[2] Montanuniversitaet Leoben, CPS, Leoben, Austria,
[3] Scholz Austria GmbH, Vienna, Austria

Abstract This study explores the detection of copper-containing particles in shredder fractions using Convolutional Neural Networks (CNNs). Twenty CNN architectures were evaluated for prediction accuracy and inference latency. The best-performing models were then evaluated in inline trials.

Keywords Metal recycling, CNN, computer vision

1 Introduction

The United Nations report on metal recycling defines the end-of-life recycling rates (EOL-RR) for 60 metallic and metalloid elements [1]. It highlights two key points: common metals like iron, copper, and zinc have EOL-RRs slightly above 50%, and copper, in particular, requires greater recycling efforts. Its inclusion in the EU's Critical Raw Material Act reflects growing demand driven by its essential role in the automotive and energy sectors [2] [3] [4]. Besides securing the EU's critical raw materials, metal recycling requires less energy than virgin production, e.g., up to 88% for copper. [5] [6].

Additionally, copper's presence in the feedstock for steel scrap recycling diminishes the mechanical properties of the recycled steel, thus making it only suitable for the least taxing applications [7] [8].

The metallurgical properties of copper make this further problematic as it cannot be metallurgically removed once in the melt as it neither

moves into the slag phase nor into the gaseous phase during scrap metal recycling in electronic arc furnaces [8] [9] [10].

Currently, copper removal from metal recycling feedstock is often done manually, which is physically and mentally demanding, making it difficult for recyclers to maintain a workforce. An alternative method uses X-ray fluorescence (XRF), but its high cost impedes widespread application. This article proposes a third method for detecting copper-containing particles in post-shredder fractions. It introduces an inline sorting technique using low-cost RGB sensors, offering reasonable accuracy. The approach evaluates various publicly available CNN architectures and one-stage detectors, which have become more applicable and mature for sorting tasks in waste management.

2 Materials and Method

2.1 Materials

As the presented sorting solution aims at replacing the manual sorting step, the samples were collected between shredding and manual sorting. This ensures that the training samples taken mimic the material the classification system would encounter in-situ as closely as possible. The copper fraction in particular is highly heterogeneous, incorporating wires, pipes, copper meatballs, and electric motors in different shapes and sizes. Thus, a high number of particles was necessary to ensure that the training material is a good representation of the material that is encountered after the shredding step. Before any recordings of the particles were performed, the material was split into a training set and a test set. This was done to prevent any data leakage and thus avoid overfitting by ensuring that the models could be tested on unseen particles. Figure 1 shows a collection of samples.

2.2 Applied Machine Learning Methods

Convolutional neural networks (CNN) were used for classification. In total, 20 different CNN architectures were trained using the training image dataset. In addition to these, one-stage detectors were evaluated for their performance. For this, all available sizes of YOLOv8 [11]

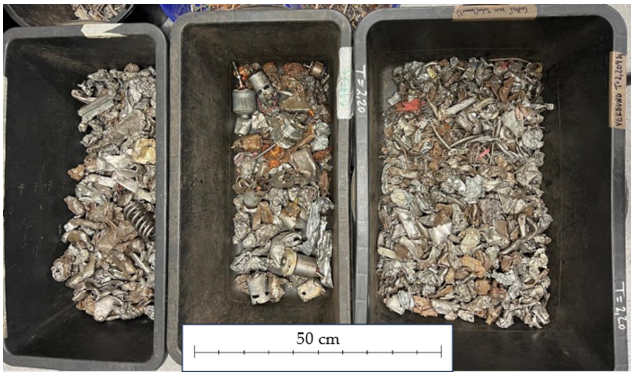


Figure 1: Depiction of Sample Material

and YOLOv11 [12] were used and compared for their mean average precision and their inference latency on a test set.

2.3 Sensor-Based Sorting Stand

The CNN stages of this work were performed on a sensor-based chute sorting aggregate, mirroring industrial facilities in optical-based sorting facilities. This setup and its working principle are shown in Figure 2. This sorting aggregate employs an AViiVA® SC2 CL Camera Link® Color Linescan Camera for image acquisition. The connection to the PC performing the classification was done with a Matrox Zebra Radient eV-CL frame grabber. The connection to the frame grabber and the Color Linescan Camera was established via the GigeCam interface over Ethernet.

2.4 Digital Waste Research Lab

For testing the one-stage detectors in this study, a conveyor belt was acquired and integrated into the existing setup at the Digital Waste Research Lab (DWRL). The belt is 80 cm wide and 500 cm long. A camera bridge for mounting industrial cameras (Basler ACE Pro) and lighting was installed on the conveyor. Figure 3 illustrates the technical facilities used in this work.

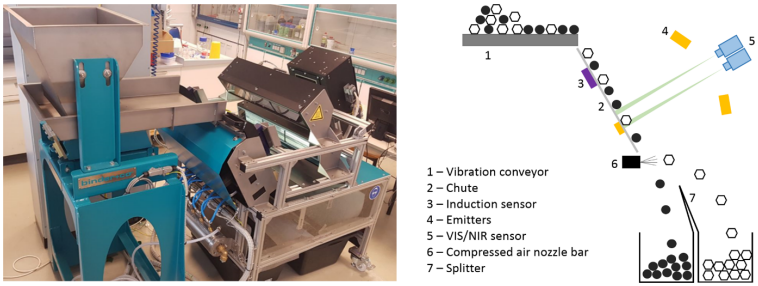


Figure 2: Depiction of Sensor Based Sorting Stand

Figure 3: Depiction of Digital Waste Research Lab

2.5 CNNs

CNN Model Creation, Validation, and Comparison

Each of the 20 CNN architectures was trained on 2,200 images, with 30% reserved for validation. Validation was performed on a separate set of unseen images to evaluate performance and prevent overfitting. Performance on unseen data was further tested by classifying 600 images in an offline phase, recording accuracy, and inference latency. To assess resilience to varying lighting during inline classification, 50 additional unseen particles were tested in live trials, where they were classified in situ as they were dropped onto the sorting chute.

Classification Process

As CNNs are image classification methods and thus incapable of object detection, the segmentation of individual objects in an image must be performed before classification. The image processing and classification steps shown in Figure 4 are described next: First, each recorded frame[A] was preprocessed to increase the difference between background and objects [B]. From this, a binary mask was created [C]. Blob analysis was then performed on that mask, yielding the bounding boxes of each object in the given frame. These bounding boxes were used to crop the individual objects from the recorded frame. The classification was then performed on these individual cropped objects and the resulting confidence and inference results were recorded and displayed [D].

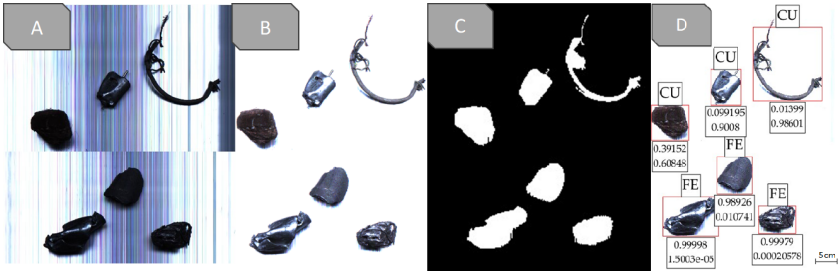


Figure 4: Depiction of CNN Image Processing and Classification

2.6 One-stage detectors

Training of one-stage detectors

All sizes of YOLOv8 and YOLOv11 were initially trained on a training set of images recorded on the DWRL conveyor belt. After training, each variant was evaluated by inferring testing images with varying occupation densities of 5%, 10%, 25%, and 40%. For each density, 5 runs with 10 testing images each were classified, and the mean, maximum, and minimum $mAP^{50\text{-}95}_{val}$ were recorded. Based on the results of this initial evaluation, the most accurate and fastest network was selected for further training and evaluation. This further training constituted training

for up to 10,000 epochs, with a learning rate of 0.001 and a classification error weight of 7.5. These networks were then further tested on images recorded at the DWRL.

Semi-Automatic Labelling Process

To streamline the labour-intensive process of creating and annotating training images, a semi-automated workflow using a custom-built application was employed. Figure 5 illustrates the workflow for generating the training images.

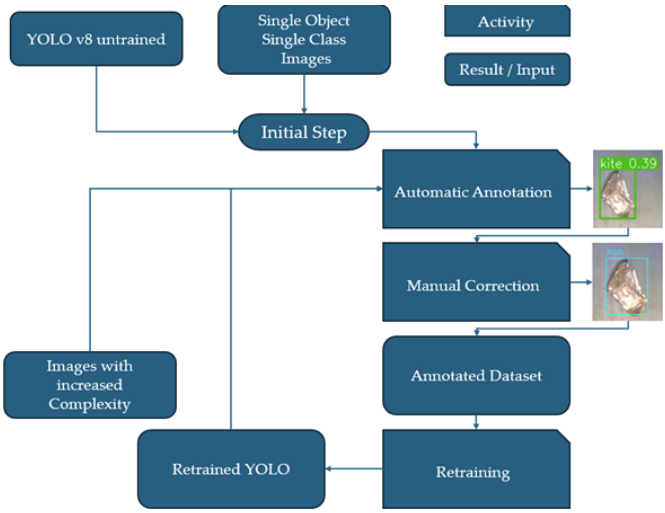


Figure 5: Workflow of semi-automatic labelling process

Initially, the one-stage detector annotated simple images containing a single particle. At first, the classifier could only create bounding boxes but misclassified the objects. These annotated images were manually corrected and used to train the classifier. This process was repeated with increasingly complex images until the model's predictions were sufficiently accurate for testing on the unseen dataset.

3 Results

3.1 CNNs

The recorded metrics allow the evaluation of the tested network architectures, with inline testing accuracies being particularly important as they best reflect real-world applications. The results of the most promising architectures are discussed below, with a full list of recorded metrics available in Table 1. The primary criterion for selecting a suitable network is the desired accuracy for the sorting task, which dictates the required architectural complexity and hardware. DenseNet-201, offering the highest accuracy, is the best choice for precise predictions. However, this comes with increased inference latency, taking twice as long as its less accurate tested counterparts. If some accuracy compromises are acceptable or if the classification must occur on limited hardware, AlexNet is a viable option. However, AlexNet was hindered by low confidence in its copper predictions. GoogleNet and MobileNetV2, with similar accuracy (90%), lower inference latency than AlexNet, and higher confidence in copper predictions, should also be considered when developing a sorting model.

Table 1: Results of best performing CNN architectures

Name	Training Accuracy [%]	Offline Test Accuracy[%]	Inline Test Accuracy[%]	Inference Latency[s]
Dense Net	100	94	98	0.0575
DarkNet-53	100	93.5	96	0.0984
VGG-16	100	93.16	96	0.1870
Inception-v3	100	89	94	0.0626
AlexNet	98	80.5	92	0.0275
NasNet-Large	100	85.6	92	0.6521
GoogleNet	98	90.8	90	0.0225
MobileNet-v2	100	82.5	90	0.0265
ResNet-101	100	80.2	90	0.0224

3.2 One Stage Detectors

The one-stage detectors of the YOLO family showed highly divergent performance. Firstly, the inference time increases substantially with increasing size of the architecture itself. This is not always proportional to the increase in classification accuracy as shown in Figure 6. The most accurate architecture was YOLOv8m, so a YOLO network of medium

size. The fastest network was YOLOv8n, which was the smallest architecture of the YOLO detectors evaluated. After further training, the YOLOv8m and YOLOv8n showed high $mAP_{val}^{50\text{-}95}$ on images from the testing set. Figure 6 shows the resulting $mAP_{val}^{50\text{-}95}$ for each tested YOLO architecture after the initial training run and the resulting $mAP_{val}^{50\text{-}95}$ for YOLOv8n and YOLOv8m after extensive training.

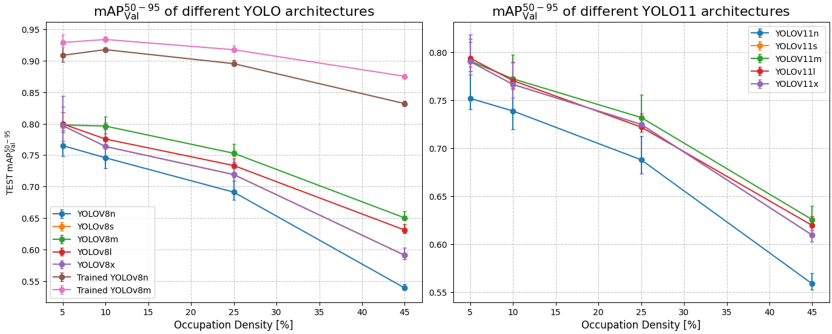


Figure 6: Resulting $mAP_{val}^{50\text{-}95}$ for each tested YOLO architecture depending on occupation density

Figure 7 shows a classified image after classifying the objects with YOLOv8m. It can be seen that the model handles overlapping objects reasonably well and is capable of handling relatively high occupation densities. Unfortunately, it can also be seen that some objects, e.g. a piece of wire, has not been recognised by the model while another object, namely a piece of copper pipe, was misclassified. In this area, the models still need further training. Specifically, objects that are only partly visible usually result in a rather low prediction confidence score and are thus excluded from the prediction.

Acknowledgements

This work was created as part of the research project “KIRAMET” which is funded by the Federal Ministry Republic of Austria Climate Action, Environment, Energy, Mobility, Innovation and Technology (FFG No.: FO 999899661) and would like to extend further gratitude to

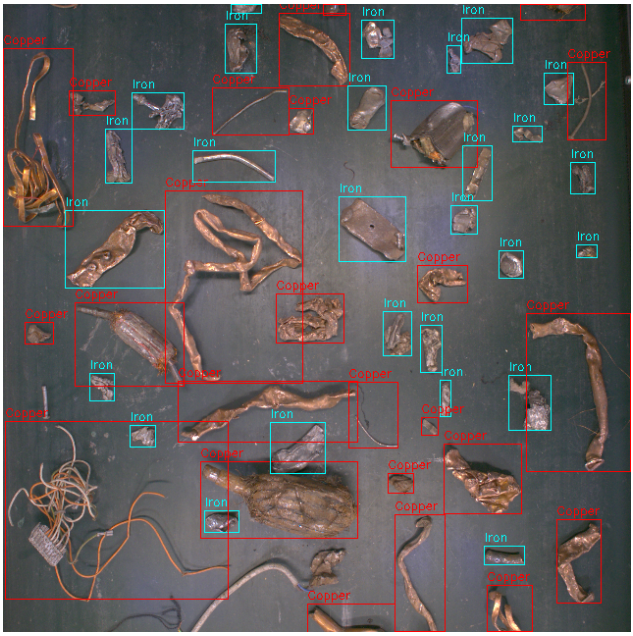


Figure 7: Classified Image of Copper Containing Scrap Metal with YOLOv8m

Scholz Recycling GmbH for kindly supplying the essential specimens for the examinations discussed in this manuscript and their ongoing support in the presented research.

References

1. Council of European Union, "Fifth list 2023 of critical raw materials for the eu," 2023, https://single-market-economy.ec.europa.eu/sectors/raw-materials/areas-specific-interest/critical-raw-materials_en#fifth-list-2023-of-critical-raw-materials-for-the-eu, Accessed on: 18/12/2024.
2. A. Elshkaki, T. Graedel, L. Ciacci, and B. K. Reck, "Copper demand, supply, and associated energy use to 2050," *Global Environmental Change*, 2016.

3. UN Environment Programm, “Geo scenarios framework,” 2004, accessed on: 18/12/2024. [Online]. Available: https://www.unep.org/resources/report/geo-scenarios-framework
4. G. L. van Acker, “Metals for clean energy: Pathways to solving europe’s raw materials challenge.” [Online]. Available: https://eurometaux.eu/metalscleanenergy
5. UN Environment Programm, “Environmental risks and challenges of anthropogenic metal flows and cycles,” 2013, accessed on: 18/12/2024. [Online]. Available: https://www.resourcepanel.org/reports/environmental-risks-and-challenges-anthropogenic-metals’-flows-and-cycles
6. F.-W. Wellmer and C. Hagelüken, “The feedback control cycle of mineral supply, increase of raw material efficiency, and sustainable development,” *Minerals*, vol. 5.
7. T. Watari, K. Nansai, and K. Nakajima, “Major metals demand, supply, and environmental impacts to 2100: A critical review,” *Resources, Conservation and Recycling*, vol. 164, 2021.
8. K. E. Daehn, A. C. Serrenho, and J. Allwood, “Finding the most efficient way to remove residual copper from steel scrap,” *Metallurgical and Materials Transactions B*, vol. 50, no. 3, 2019.
9. K. Nakajima, O. Takeda, T. Miki, K. Matsubae, S. Nakamura, and T. Nagasaka, “Thermodynamic analysis of contamination by alloying elements in aluminum recycling,” *Environmental Science and Technology*, 2010.
10. T. Hiraki, O. Takeda, K. Nakajima, K. Matsubae, S. Nakamura, and T. Nagasaka, “Thermodynamic criteria for the removal of impurities from end-of-life magnesium alloys by evaporation and flux treatment,” *Science and Technology of Advanced Materials*, vol. 12, 2011.
11. G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics yolov8,” 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
12. G. Jocher and J. Qiu, “Ultralytics yolo11,” 2024. [Online]. Available: https://github.com/ultralytics/ultralytics

International Conference on Optical Characterization of Materials

Each material has its own specific spectral signature independent if it is food, plastics, or minerals. New trends and developments in material characterization have been discussed as well as latest highlights to identify spectral footprints and their realizations in industry.

Conference topics

- **Agriculture**
- **Food Inspection**
- **Recycling and Environment**
- **Spectral Applications**
- **Spectral Data Processing**
- **Spectral Sensors**

The International Conference on Optical Characterization of Materials (OCM-2025) was organized by the Karlsruhe Center for Spectral Signatures of Materials (KCM) in cooperation with the German Chapter of the Instrumentation & Measurement Society of IEEE.

KCM is an association of institutes of the Karlsruhe Institute of Technology (KIT) and the business unit Advanced Sensing of the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (Fraunhofer IOSB).

ISBN 978-3-7315-1408-4

ISSN 2510-7240
ISBN 978-3-7315-1408-4

Printed on FSC-certified paper