Visualization and Learning-based Classification Methods for the Analysis of Multi-frequency Polarimetric Interferometric Synthetic Aperture Radar Image Data

Sylvia Marlene Hochstuhl

Doctoral Thesis
Karlsruhe, 2025

Visualization and Learning-based Classification Methods for the Analysis of Multi-frequency Polarimetric Interferometric Synthetic Aperture Radar Image Data

Zur Erlangung des akademischen Grades einer

DOKTORIN DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

von der KIT-Fakultät für
Bauingenieur-, Geo- und Umweltwissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Sylvia Marlene Hochstuhl
aus Bonn

Tag der mündlichen Prüfung:	05.12.2024
Referent:	Prof. Dr.-Ing. Stefan Hinz Institut für Photogrammetrie und Fernerkundung Karlsruher Institut für Technologie
Korreferent:	Prof. Dr.-Ing. Michael Schmitt Institut für Raumfahrttechnik und Weltraumnutzung Universität der Bundeswehr München

Karlsruhe (2025)

Sylvia Marlene Hochstuhl

Visualization and Learning-based Classification Methods for the Analysis of Multi-frequency Polarimetric Interferometric Synthetic Aperture Radar Image Data

Doctoral Thesis
Date of examination: 05.12.2024
Referees:
Prof. Dr.-Ing. Stefan Hinz
Prof. Dr.-Ing. Michael Schmitt

Karlsruhe Institute of Technology
Department of Civil Engineering, Geo and Environmental Sciences
Institute of Photogrammetry and Remote Sensing
Kaiserstr. 12
76131 Karlsruhe

Abstract

Polarimetric Synthetic Aperture Radar (PolSAR) systems have established themselves as essential instruments in Earth observation because they can capture information-rich image data independent of weather and illumination conditions. Analyzing PolSAR data allows distinguishing different land cover and estimating geophysical parameters such as surface roughness and soil moisture. By additionally using SAR interferometry, complementary information is obtained from the phase difference of an image pair. Thus, the combination of SAR polarimetry and interferometry (PolInSAR) allows a comprehensive characterization of observed scenes. Due to the varying penetration depth of electromagnetic waves depending on their wavelength, multi-frequency PolInSAR data offer the potential for an even more detailed analysis. This dissertation aims to fully exploit the rich information content of multi-frequency PolInSAR data in the context of both manual, human-driven analysis and automatic analysis. For this purpose, new methods for data visualization and learning-based data classification are developed and evaluated.

To display the information of multi-frequency PolInSAR data through user-friendly visualizations, it is necessary to transform the multi-dimensional, complex-valued signals into three-dimensional, real-valued representations that optimally preserve the information content. In this work, the dimension reduction method Uniform Manifold Approximation and Projection (UMAP) is tailored and applied to PolInSAR data. Two approaches are proposed: Feature-UMAP (F-UMAP), in which a set of polarimetric and interferometric features is extracted and condensed using UMAP, and Wishart-UMAP (W-UMAP), which utilizes a statistical distribution model for PolInSAR data by integrating a Wishart-based distance metric into the UMAP algorithm. The results show that employing UMAP ensures good preservation of the data structure and the separability of land cover classes, outperforming comparable dimension reduction methods in this respect. The UMAP-based visualization proves to be a suitable tool for generating intuitively interpretable data representations. Visualizing the data structure using scatter plots facilitates the detection of patterns and clusters in complex data. In addition, easily interpretable color images can be generated to enable visual scene analysis, allowing analysts to incorporate contextual knowledge into the data analysis and identify potential applications. The resulting improved understanding of the data subsequently supports the targeted development of automatic analysis methods.

An essential application of automatic data analysis in remote sensing is land cover classification. Inspired by the success of Convolutional Neural Networks (CNNs), which has led to remarkable advances in the automatic interpretation of optical image data, this

thesis transfers CNN-based classification to multi-frequency PolInSAR data. To enable the use of established CNN models, different approaches are developed and compared, which transform the multi-layered information of the data into a real-valued representation that serves as model input. The tested representations are either based on directly using the PolInSAR coherency matrix or combining a wide range of previously extracted polarimetric and interferometric features. To identify the representation that is best suited as input for the CNN-based land cover classification, experiments with different U-Net based CNN models are performed on the custom-built open Pol-InSAR-Island benchmark dataset. The best classification performance is achieved by integrating polarimetric and interferometric features into the CNN-based classification. The classification results obtained in this way also outperform those obtained using traditional learning-based methods such as the Wishart classifier and the Random Forest. In order to reduce the computational and memory requirements of the proposed approach, different dimension reduction methods applied before or within the CNN classification are evaluated. The results show that applying a prior feature transformation with Principal Component Analysis (PCA) or feature selection via the minimum Redundancy Maximum Relevance (mRMR) method effectively reduces the dimension of the CNN input data with only minimal loss of classification performance. This reduction significantly shortens the training time of the CNNs. The findings regarding how to combine the multi-layered information of multi-frequency PolInSAR data with the ability of CNNs to recognize and classify complex image patterns lay the foundation for the future advancements of learning-based models for automatic PolInSAR image interpretation.

Kurzfassung

Polarimetrische SAR (Radar mit synthetischer Apertur) Systeme haben sich aufgrund ihrer Fähigkeit, informationsreiche Bilddaten unabhängig von Wetter- und Beleuchtungsbedingungen aufnehmen zu können, als wichtige Technologie in der Erdbeobachtung etabliert. Die Analyse polarimetrischer SAR (PolSAR) Daten ermöglicht die Unterscheidung verschiedener Landbedeckungen und die Schätzung geophysikalischer Parameter wie Oberflächenrauhigkeit und Bodenfeuchte. Durch den zusätzlichen Einsatz der SAR Interferometrie können darüber hinaus komplementäre Informationen aus der Phasendifferenz eines Bildpaares gewonnen werden. Die Kombination aus SAR Polarimetrie und Interferometrie (PolInSAR) erlaubt eine umfassende Charakterisierung beobachteter Szenen. Aufgrund der variierenden Eindringtiefe elektromagnetischer Wellen in Abhängigkeit von ihrer Wellenlänge bieten multifrequente PolInSAR-Daten Potenzial für eine noch detailliertere Analyse. Das Ziel dieser Dissertation ist es, den reichhaltigen Informationsgehalt multifrequenter PolInSAR Daten sowohl im Rahmen der manuellen, anwendergesteuerten Analyse als auch der automatischen Analyse vollständig auszuschöpfen. Um dies zu erreichen, werden neue Methoden für die Datenvisualisierung und die lernbasierte Datenklassifikation entwickelt und evaluiert.

Um die Information multifrequenter PolInSAR Daten durch einfach interpretierbare Visualisierung darzustellen, ist die Transformation der multidimensionalen komplexwertigen Signale in dreidimensionale, reellwertige Darstellungen erforderlich, die den Informationsgehalt optimal erhalten. In dieser Arbeit wird dazu das Dimensionsreduktionsverfahren Uniform Manifold Approximation and Projection (UMAP) auf PolInSAR Daten zugeschnitten und angewendet. Dazu werden zwei Ansätze vorgeschlagen: Feature-UMAP (F-UMAP), bei welchem eine Menge polarimetrischer und interferometrischer Merkmale extrahiert und mittels UMAP kondensiert wird und Wishart-UMAP (W-UMAP), bei welchem ein statistisches Verteilungsmodell für PolInSAR Daten genutzt wird, indem eine Wishart-basierte Distanzmetrik in den UMAP-Algorithmus integriert wird. Die Ergebnisse zeigen, dass der Einsatz von UMAP eine gute Erhaltung der Datenstruktur und der Trennbarkeit von Landbedeckungsklassen gewährleistet und in dieser hinsicht vergleichbare Dimensionsreduktionsmethoden übertrifft. Die UMAP-basierte Visualisierung erweist sich als geeignetes Werkzeug für die Generierung intuitiv interpretierbarer Darstellungen. Die Visualisierung der Datenstruktur durch Streudiagramme erleichtert die Erkennung von Mustern und Clustern in den komplexen Daten. Darüber hinaus können einfach interpretierbare Farbbilder erzeugt werden, die eine visuelle Szenenanalyse ermöglichen, und somit Analysten erlauben, Kontextwissen in die Datenanalyse einzubringen und Anwendungspotenziale

zu identifizieren. Das so erlangte verbesserte Datenverständnis unterstützt folgend die gezielte Entwicklung automatischer Analysemethoden.

Ein zentraler Anwendungsbereich der automatischen Datenanalyse in der Fernerkundung ist die Klassifikation der Landbedeckung. Inspiriert durch den Erfolg von Convolutional Neural Networks (CNNs), welche zu bemerkenswerten Fortschritten in der automatischen Interpretation optischer Bilddaten geführt haben, wird in dieser Arbeit die CNN-basierte Klassifikation auf multifrequente PolInSAR Daten übertragen. Um den Einsatz etablierte CNN Modelle zu ermöglichen, werden verschiedene Ansätze vorgestellt und verglichen, welche zunächst die vielschichtige Information der Daten in eine reellwertige Darstellung transformieren. Die getesteten Darstellungen basieren entweder auf der direkten Verwendung der PolInSAR Kohärenzmatrix oder der Kombination vielfältiger zuvor extrahierter polarimetrischer und interferometrischer Merkmale. Um die Darstellung zu identifizieren, die sich am besten als Input für die CNN-basierte Landbedeckungsklassifikation eignet, werden Experimente mit verschiedenen U-Net basierten CNN Modellen auf dem eigens erstellten offenen Pol-InSAR-Island Benchmark Datensatz durchgeführt. Die beste Klassifikationsleistung wird erzielt, indem polarimetrische und interferometrische Merkmale in die CNN-basierte Klassifikation integriert werden. Die auf diese Weise erzielten Klassifikationsergebnisse übertreffen dabei auch die Ergebnisse, die mittels traditioneller lernbasierter Verfahren wie dem Wishart Klassifikator und dem Random Forest erreicht werden. Um Rechen- und Speicherbedarf des vorgeschlagenen Ansatzes zu reduzieren, werden verschiedene Dimensionsreduktionsmethoden, die vor oder innerhalb der CNN Klassifikation angewendet werden, evaluiert. Die Ergebnisse zeigen, dass die Anwendung einer vorherigen Merkmalstransformation mit der Hauptkomponentenanalyse (PCA) oder einer Merkmalsauswahl durch die minimum Redundancy Maximum Relevance (mRMR) Methode die Dimension der CNN Eingangsdaten effektiv reduziert, während der Verlust der Klassifikationsleistung minimal bleibt. Dadurch verkürzt sich die Trainingszeit der CNNs erheblich. Die Erkenntnisse darüber, wie man die vielschichtige Information multifrequenter PolInSAR Daten mit den Fähigkeiten von CNNs komplexe Bildmuster zu erkennen und zu klassifizieren, kombiniert, legen die Basis für die zukünftige Weiterentwicklung lernbasierter Modelle für die automatische PolInSAR Bildinterpretation.

Acknowledgement

Many people accompanied and supported me during my Ph.D studies and contributed to the success of this thesis. I cannot thank every single person who played a part in positively shaping my research time, but I would like to mention the people who contributed in a particularly important way.

First and foremost, I would like to thank my Ph.D supervisor Stefan Hinz, who made my Ph.D project possible in the first place. He not only provided me with competent professional support but also inspired me with his positive attitude and enthusiasm. I very much appreciate that Stefan Hinz gave me a great deal of freedom in setting my thematic focus and conducting my research. I would also like to thank him for enabling me to attend many conferences that have assisted my professional and personal development. I would also like to thank my co-supervisor, Michael Schmitt, for his time and support during the final phase of my Ph.D.

A very sincere thanks go to my direct mentor and colleague Antje Thiele, who always strongly supported me throughout the entire time, and ensured that I never lost my motivation. As a group leader at the IOSB, she is a great role model, and over the last few years, I have been able to learn a lot from her, not only professionally but also personally.

I would also like to thank my colleague Horst Hammer, who has always been a great support to me. His constructive discussions have helped me progress professionally, and I very much appreciate that he has always made time for all my concerns.

I also thank my head of department at the IOSB, Karsten Schulz, who has always enabled me to harmonize my research work at the IPF with my research work at the IOSB. Many thanks also to the wonderful SAR team at the IOSB, who have always made a positive impact on my day-to-day work. I also thank my many lovely colleagues at the IPF for a wonderful time together in a great working atmosphere.

Niklas Pfeffer, whom I sincerely thank, also made an important contribution to my work. He worked closely with me as part of his Master's thesis. The discussions with him gave me valuable impulses, and I was able to build on the results and code implementation of his work.

Finally, I thank my friends and family, who have always accompanied and supported me. A very special thanks goes to my husband, Adam Hochstuhl, who supported me in every situation and motivated me constantly. His emotional support was invaluable, especially in the final phase. Thank you very much!

Contents

1 Introduction **1**
- 1.1 Problem Statement . . . 2
- 1.2 Outline and Main Contributions . . . 4

2 Basics on Polarimetric Interferometric Synthetic Aperture Radar Data **7**
- 2.1 SAR Principles . . . 7
  - 2.1.1 Side-Looking Airborne Radar . . . 7
  - 2.1.2 Synthetic Aperture Radar . . . 9
  - 2.1.3 SAR Image Formation . . . 11
- 2.2 SAR Polarimetry . . . 12
  - 2.2.1 Wave Polarimetry . . . 13
  - 2.2.2 Scattering Polarimetry . . . 15
  - 2.2.3 Polarimetric Speckle Filtering . . . 20
  - 2.2.4 Polarimetric Scattering Decomposition Theorems . . . 21
  - 2.2.5 Polarimetric Features . . . 27
- 2.3 Polarimetric SAR Interferometry . . . 29
  - 2.3.1 SAR Interferometry . . . 29
  - 2.3.2 Combining Polarimetry and Interferometry . . . 32

3 The Pol-InSAR-Island Benchmark Dataset **35**
- 3.1 Introduction . . . 35
- 3.2 PolInSAR Data . . . 37
- 3.3 Reference Data . . . 39
  - 3.3.1 Revised Biotope Type Map . . . 40
  - 3.3.2 Target Class Map . . . 41
  - 3.3.3 Pol-InSAR-Island Class Map . . . 46
  - 3.3.4 Training and Test Data Split . . . 48
- 3.4 Discussion . . . 48

4 Visualizing Multi-frequency PolInSAR Data **51**
- 4.1 Introduction . . . 51
- 4.2 Related Work . . . 52
  - 4.2.1 Visualizing Multi-dimensional Data . . . 54
  - 4.2.2 Dimension Reduction of Polarimetric Feature Space . . . 57
  - 4.2.3 Conclusion . . . 60
- 4.3 Method . . . 60

4.3.1 UMAP Algorithm . . . 60
4.3.2 Feature-UMAP . . . 63
4.3.3 Wishart-UMAP . . . 66
4.3.4 Visualization . . . 67
4.4 Experimental Setup . . . 68
4.4.1 Comparison Methods . . . 68
4.4.2 Hyperparameter Setting . . . 71
4.4.3 Evaluation . . . 71
4.5 Results . . . 74
4.5.1 2-dimensional Projections . . . 75
4.5.2 3-dimensional Projections . . . 81
4.6 Applications . . . 85
4.6.1 2-dimensional Projections . . . 85
4.6.2 3-dimensional Projections . . . 88
4.7 Discussion . . . 91

5 CNN-based PolInSAR Image Classification 95
5.1 Introduction . . . 95
5.2 Fundamentals . . . 98
5.2.1 Wishart Classifier . . . 98
5.2.2 Feature Selection . . . 99
5.2.3 Random Forest Classification . . . 101
5.2.4 Convolutional Neural Networks . . . 102
5.3 Related Work . . . 106
5.3.1 Input Representation . . . 106
5.3.2 CNN Architectures . . . 112
5.4 Methods . . . 116
5.4.1 Feature Stacking . . . 116
5.4.2 Automatic Feature Selection . . . 120
5.4.3 Feature Transformation . . . 123
5.4.4 CNN Architecture . . . 123
5.4.5 Cross-channel Feature Extraction . . . 128
5.4.6 Model Training . . . 130
5.5 Experimental Setup . . . 132
5.6 Results . . . 134
5.6.1 Baseline Results . . . 134
5.6.2 Single-frequency PolSAR Classification . . . 140
5.6.3 Multi-frequency PolInSAR Classification - Feature Stacking . . . 150
5.6.4 Multi-frequency PolInSAR Classification with Cross-channel Feature Extraction . . . 160
5.6.5 Multi-frequency PolInSAR Classification with Dimension Reduction 160
5.7 Discussion . . . 170

5.7.1 Data Representation 171
5.7.2 Handling High-dimensional Data 173
5.7.3 Comparison of Random Forest and CNN Classification 176

6 Conclusion and Outlook **179**
6.1 Conclusion 179
6.2 Outlook 181

Declaration of Supportive Resources **183**

Bibliography **185**

List of Figures **207**

List of Tables **209**

List of Abbreviations **211**

A Analysis of Learning Rate Schedule **215**

B Analysis of CNN Input Data Representations **217**

1 Introduction

Our planet is in a constant state of change, shaped by natural processes such as tidal movements and seasonal climate fluctuations, as well as human influences such as urbanization, deforestation, and agricultural expansion. Understanding and analyzing these dynamic changes is crucial in order to implement sustainable urban planning, for example, or to plan and evaluate environmental protection measures in a targeted manner. Advances in the development of various remote sensing technologies enable continuous monitoring of the state of the Earth's surface. One of the central aspects in this context is the automatic classification of land cover based on observation data, which allows, for example, the monitoring of deforestation, assessment of the condition of vulnerable habitats, or the condition of artificial infrastructure. The Synthetic Aperture Radar (SAR) system is a particularly powerful remote sensing technology. By sending and receiving microwave pulses, it can acquire high-resolution image data regardless of light conditions and cloud cover and has, therefore, established itself as an essential tool for Earth observation.

A further development of the SAR system is the Polarimetric SAR (PolSAR) system, which transmits and receives orthogonally polarized microwaves and thus detects not only the intensities and phases but also the polarization state of the backscattered waves. Since the polarization state changes depending on geophysical properties such as soil moisture and surface roughness, dielectric properties, and the shape and orientation of observed scatterers, corresponding qualitative and quantitative physical information can be derived from the measurement data [1].

If a scene is covered from at least two slightly different angles, potentially with a time offset, further valuable, complementary information can be obtained from the phase difference between the measurements. This imaging technique, referred to as Interferometric SAR (InSAR), enables the derivation of geophysical parameters such as surface topography or temporal variability of observed scatterers. Hence, the combination of polarimetry and interferometry, known as Polarimetric Interferometric SAR (PolInSAR) [2], results in a powerful observation space. This enables an even more extensive characterization of observed scatterers, allowing, for instance, the retrieval of vegetation height, biomass, or snow depth [3–5], and can improve land cover classification [6–8].

Current PolSAR systems (e.g., F-SAR or AIRSAR) enable simultaneous data acquisition in different frequency bands, further expanding the set of observable variables. Since electromagnetic waves penetrate media to different depths depending on their frequency, multi-frequency PolInSAR promises an even more differentiated distinction of land cover

classes or an improved estimation of geophysical parameters, such as vegetation density or tree canopy volume [9–11].

At present, multi-frequency PolInSAR data is recorded exclusively with airborne systems. However, due to the potential for diverse applications, the transfer to spaceborne systems is already planned for upcoming missions. Specifically, the soon-to-be-launched multi-frequency NASA-ISRO SAR (NISAR) system and the Radar Observatory System for Europe in L-band (Rose-L) promise continuous data acquisition on a global scale [12]. To use the information from the massive amount of data effectively for remote sensing applications, methods tailored to PolInSAR data are required that enable both human-driven analysis and automatic data analysis.

1.1 Problem Statement

The major challenge addressed in this thesis is to present the rich information of multi-frequency PolInSAR data in a form that can either be easily interpreted by humans or effectively analyzed by automated analysis techniques.

Human-driven Visual Data Analysis. Data visualization is pivotal in supporting human observers in understanding the relationship between multi-frequency PolInSAR data and possible target variables. It provides an intuitive tool for data interpretation and understanding that allows analysts to quickly identify patterns, anomalies, and relationships within the data that automated algorithms might miss. A profound understanding of the data is essential to develop automated analysis methods in a targeted manner and to better understand or validate their output. In addition, visualizations enable combined automated and manual analysis since analysts can combine their observations with background and domain knowledge and incorporate them into the data analysis.

Two types of visual representation are important for human-driven data analysis. The first is the representation of the data structure, represented by scatter plots. These plots depict similarities or differences in observed scattering processes and allow easy recognition of clusters and patterns in the data. This type of representation makes it possible, for example, to recognize which types of land cover can be separated on the basis of the data or how varying acquisition parameters affect the data. The second type of representation is a color composite of the observed scene, in which different colors symbolize different backscatter processes. This enables the direct recognition of spatial patterns, the identification of interesting areas, or the detection of unusual phenomena.

However, the spatial imagination of humans, as well as the color representation of standard displays (RGB), are limited to 3 dimensions. Both visualization approaches, therefore, face the challenge of representing the information of the multi-dimensional, complex-valued

multi-frequency PolInSAR data as completely as possible in only three dimensions. The data should be compressed in such a way that the information content is preserved to the best possible extent. A further aim is to guarantee the consistency and transferability of the visualization. The same backscatter properties should be displayed in the same color across the entire scene, and the color scheme should be transferable to new acquisitions to allow consistent analysis across multi-temporal or multi-location datasets.

Automatic Land Cover Classification. While human-driven visual analysis is particularly suitable for detailed analyses on a local scale, automatic analysis methods are essential for semantic data interpretation on a global scale. The pixel-wise land cover classification based on the acquired data is a widely required task. For the classification of PolSAR and PolInSAR image data, methods from the field of supervised machine learning, which use labeled sample data to learn the relationship between observation data and target classes, have proven successful and are well established. Typical classification approaches consist of the extraction of polarimetric (and interferometric) features, followed by the training of a learning model such as the Random Forest or the Support Vector Machine (SVM) [7, 13–19]. In this approach, the design and compilation of task-specific suitable features is of central importance for the quality of the classification result and represents a challenge.

Inspired by the tremendous success of Convolutional Neural Networks (CNNs) for the automatic interpretation of optical image data, its use in the SAR domain has also received increased attention in recent years. In particular, the ability of CNNs to learn extracting and classifying complex discriminative features makes it a suitable tool for land cover classification. The potential of CNNs, compared to traditional machine learning methods, to improve the classification of PolSAR data has already been demonstrated in several studies [20–25]. However, their application for (multi-frequency) PolInSAR analysis has been relatively limited, and further research effort is required to unlock the full potential arising from combining the highly informative image data with powerful feature extraction capabilities of CNNs.

An existing challenge in transferring CNNs to PolInSAR classification results from the fact that CNNs are originally designed to process RGB images, where three real values fully describe pixel information. The information that is contained in a single pixel of PolInSAR data, in contrast, is represented as a multi-dimensional complex-valued measurement. A significant amount of the polarimetric and interferometric information that can contribute to the classification task lies in the complex phase. Thus, to apply established CNN models for the classification of PolInSAR image data, the question of how the multi-layered information of the data can be represented in the real domain to fully exploit it within the classification process must be addressed.

1.2 Outline and Main Contributions

The structure of this thesis is outlined below. It describes how the aforementioned challenges are addressed and highlights the main contributions of the individual chapters.

Chapter 2 - Polarimetric Interferometric SAR. To provide the theoretical basis for the data analyzed in this thesis, principles and approaches for the acquisition, representation, and characterization of multi-frequency PolInSAR are described. In addition, the relevant features for land cover classification that can be derived from polarimetry and interferometry are discussed.

Chapter 3 - The Pol-InSAR-Island Benchmark Dataset. The success of learning-based analysis of optical image data has been strongly driven by the availability of large-scale benchmark datasets for model training and evaluation. In contrast, freely available labeled PolSAR data are rare and even non-existent for PolInSAR data. The lack of labeled data for training, testing, and comparing different approaches hinders the rapid development of machine learning algorithms for PolInSAR image analysis. The Pol-InSAR-Island benchmark dataset was therefore generated, which provides the data basis for the analyses in this thesis and is also made available to the research community to advance the development of learning-based methods. The main contribution thus lies in the publication of the first open multi-frequency PolInSAR dataset, labeled by land cover classes, that poses a challenging classification task and allows training and evaluation of learning-based models.

Chapter 4 - Visualizing Multi-frequency PolInSAR Data. To address the open challenge of generating consistent visualization that comprehensively represents the relevant information of multi-frequency PolInSAR data, the use of the dimension reduction method Uniform Manifold Approximation and Projection (UMAP) [26] is proposed in this chapter. While this technique is already successfully used for visualizing multi-dimensional data [27–30], its application for visualizing PolInSAR data has not yet been investigated. The analysis presented in this chapter fills this gap by tailoring UMAP to PolInSAR data and evaluating its potential for preserving and visualizing information compared to other dimension reduction and visualization methods.

Chapter 5 - CNN-based PolInSAR Image Classification. While CNNs has already proven to be a powerful model for the land cover classification of PolSAR images, integrating additional information from interferometry and multi-frequency into CNN-based classification has not yet been sufficiently studied. The analyses presented in this chapter aim to close this research gap. In particular, the question of how the multi-layered information of multi-frequency PolInSAR data can be represented to fully utilize it within a CNN-based

classification and achieve optimal results is addressed. For this purpose, the use of different data representations, based either directly on the measured signal or on physical interpretable features, as input for CNNs is proposed and compared. Further, the use of dimension reduction methods prior to CNN classification is proposed and evaluated. This approach avoids redundancies within the input data and thus enables more efficient model training.

Chapter 6 - Conclusion and Outlook. This chapter summarizes the findings of this work. In addition, an outlook is given for future work that can build on these insights.

2 Basics on Polarimetric Interferometric Synthetic Aperture Radar Data

The following chapter explains the basics of the PolInSAR technology. First, the operating principle of the SAR sensor is explained. This is followed by an explanation of wave and scattering polarimetry along with the information that can be derived from PolSAR measurements. Finally, the concept of SAR interferometry and its combination with SAR polarimetry is described.

2.1 SAR Principles

SAR is a technology that uses electromagnetic waves in the microwave range to generate high-resolution images of the Earth's surface. Since SAR systems are active sensor systems that autonomously transmit microwave pulses and receive the reflected signals, images can be captured independently of daylight. As microwaves can penetrate most atmospheric conditions, SAR systems are also independent of the weather and thus can provide reliable data even in cloudy conditions. These properties make the SAR technology a valuable tool for Earth observation. The SAR principle builds on the concept of the real aperture radar technology (also known as Side-Looking Airborne Radar (SLAR)), which is described below.

2.1.1 Side-Looking Airborne Radar

SLAR is an imaging radar system that is installed on a moving platform (e.g., aircraft or satellite) and alternately transmits electromagnetic pulses orthogonal to the flight direction and receives the backscattered signal [31, 32]. The antenna used to receive the signals can either be the same antenna that also transmits signals (monostatic radar) or an additional antenna (bistatic radar). The transmitted electromagnetic waves interact with the observed surface or the objects on it and are partially scattered back to the receiving antenna. The interaction causes the amplitude and phase of the signal to change depending on the properties of the detected areas, including surface roughness, object geometry, and dielectric conductivity, as well as the recording geometry. Depending on the frequency

Table 2.1.: Frequency bands used in radar remote sensing.

Band	**Wavelength** [cm]	**Frequency** [GHz]	**Sample Application**
Ka	27 - 40	0.75 - 1.2	airport surveillance
Ku	12 - 18	1.7 - 2.5	satellite altimetry
X	8 - 12	2.5 - 4	urban monitoring
C	4 - 8	4 - 8	sea ice classification, agriculture, soil moisture
S	2 - 4	8 - 15	agriculture monitoring, vegetation mapping
L	1 - 2	15 - 30	geophysical monitoring, biomass estimation
P	0.3 - 1	30 - 100	geology, archaeology

band used by the radar, the point at which a signal is reflected varies. Longer wavelengths (i.e., lower frequencies) lead to greater penetration depths in media so that, for example, vegetation is penetrated. Table 2.1 lists frequency bands typically used in radar remote sensing along with samples of their applications according to [33].

The imaging geometry of a SLAR is shown in Figure 2.1. The area illuminated by the radar beams is referred to as a swath, and its ground range extent is denoted as swath width. The flight direction is called azimuth direction, and the line-of-sight (i.e., the beam direction) is referred to as range direction. The angle θ indicates the incidence angle.

An important parameter in imaging radar technology is the spatial resolution in range and in azimuth direction [32]. The range resolution δ_r specifies the minimum distance between two reflecting objects at which the radar can image them as separate objects. This resolution is achieved when the distance between the objects is large enough so that the reflected signals do not overlap. The range resolution depends on the duration τ of the transmitted pulse and the propagation speed c (speed of light) and can be described by:

$$\delta_r = \frac{c\tau}{2}. \tag{2.1}$$

An improvement in range resolution can be achieved by increasing the bandwidth B ($B = 1/\tau$) of the transmitted signal. However, as the signal energy is thereby distributed over a larger frequency range and thus the energy per frequency unit decreases, this leads to a higher signal-to-noise ratio. Pulse compression is an alternative method that achieves a higher range resolution while keeping the noise level low. A technique in which a long radar pulse is transmitted and then compressed during signal processing. This enables high energy transmission while maintaining a short, effective pulse duration.

The azimuth resolution δ_a specifies the minimum distance between two reflecting objects in the lateral direction (parallel to the direction of movement of the radar) at which the radar can image them as separate objects. For a SLAR, the azimuth resolution depends on the slant range distance r_0 between the sensor and the object and the radar beamwidth

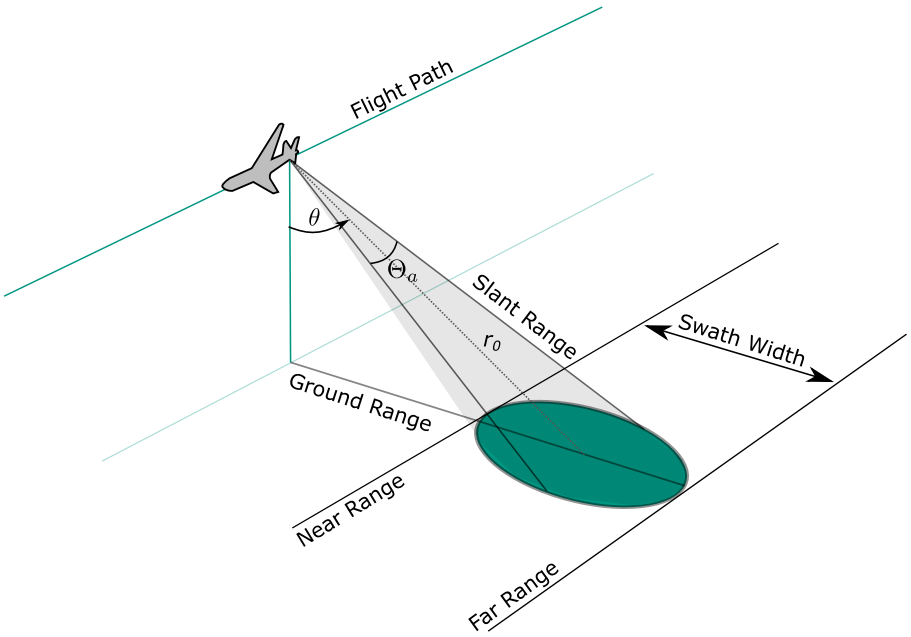


Figure 2.1.: Acquisition geometry of a SLAR. The incidence angle is denoted by θ, the azimuth beamwidth by Θ_a, and r_0 is the range distance between the antenna and the observed target. The graphic is adopted from [34].

Θ_a, which can be derived from the physical length of the antenna L and the wavelength λ of the signal [35]. The relationship can be described as:

$$\delta_a = \frac{\lambda}{L} \cdot r_0 = \Theta_a \cdot r_0. \tag{2.2}$$

Consequently, the azimuth resolution worsens with increasing range distance. An improvement in the azimuth resolution can be achieved by reducing the beamwidth. This requires either the use of high-frequency signals, which are, however, sensitive to atmospheric influences, or the use of longer antennas, which are, however, limited by the hardware. This results in only moderate to low azimuth resolutions being achievable with a SLAR, which represents the main disadvantage of this system.

2.1.2 Synthetic Aperture Radar

To counteract the issue of limited azimuth resolution, the SAR principle was developed, in which the movement of the radar carrier is used to create a virtual, long antenna. This is illustrated in Figure 2.2. As the radar carrier moves along a trajectory, the radar continuously transmits pulses and receives the reflected signals. Objects in the observed area are recorded from different angles throughout the course of the movement. By coherently combining the received radar echoes, the aperture of an antenna is synthesized, which is significantly longer than the physical antenna length. This allows the realization of a higher azimuth resolution.

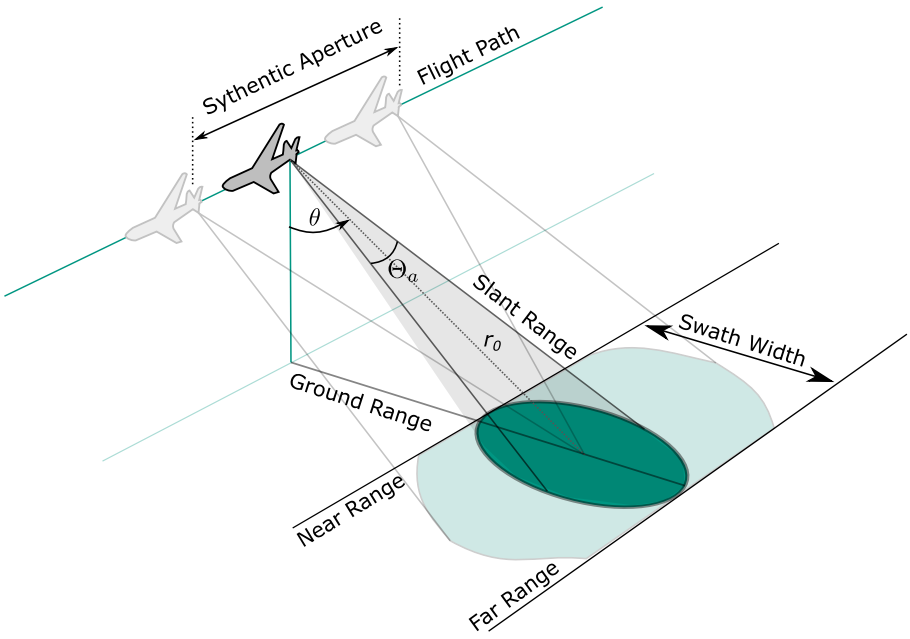


Figure 2.2.: SAR acquisition geometry.

The synthetic aperture length L_{sa} corresponds to the distance covered by the sensor during the transmission of the pulses (see Figure 2.2). It can be modeled as a function of the beamwidth Θ_a of the real aperture and the distance r_0:

$$L_{sa} = \Theta_a \cdot r_0 = \frac{\lambda}{L} \cdot r_0. \tag{2.3}$$

The virtual beamwidth Θ_{sa} is given by:

$$\Theta_{sa} = \frac{\lambda}{2L_{sa}}, \tag{2.4}$$

where the factor two accounts for the two-way measurement (transmit and receive path), which is necessary due to the displacement between the transmit and receive positions.

The azimuth resolution of the SAR system results in:

$$\delta_a = \Theta_{sa} \cdot r_0 = \frac{\lambda}{2L_{sa}} r_0 = \frac{\lambda}{2\left(\frac{\lambda}{L} \cdot r_0\right)} r_0 = \frac{L}{2}. \tag{2.5}$$

and, thus, is independent of the range distance.

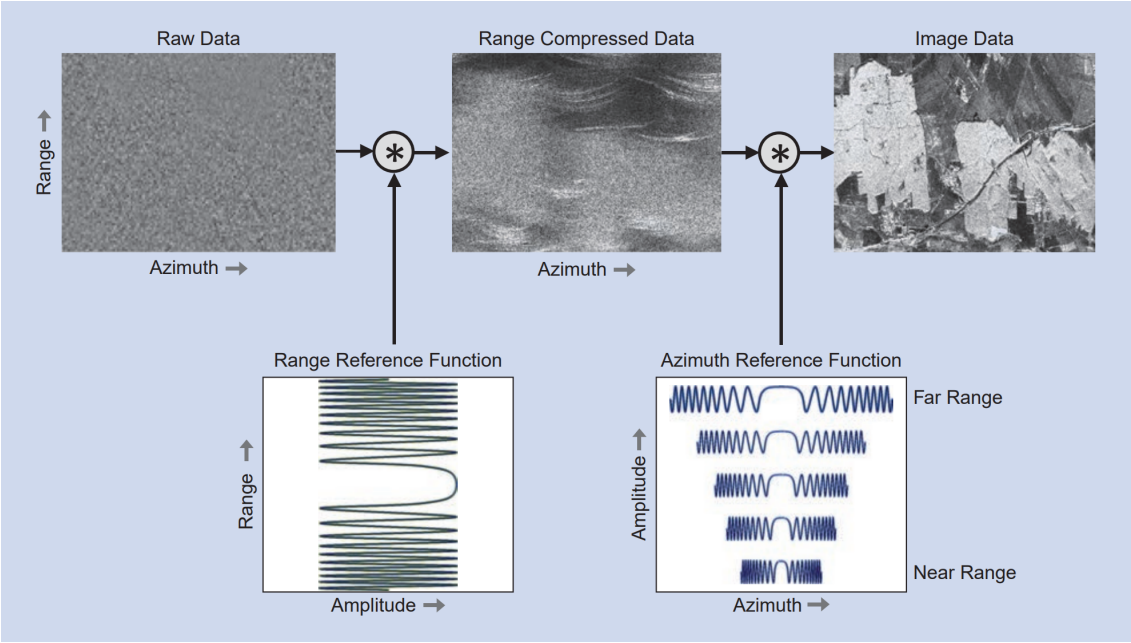


Figure 2.3.: Process of SAR image focusing, consisting of range and azimuth compression. The symbol ⊛ stands for the convolution operation. The figure is taken from [35].

2.1.3 SAR Image Formation

The received radar echoes, which are recorded and assembled during the overflight, are stored in a 2-dimensional matrix consisting of complex-valued elements. The first dimension of this matrix corresponds to the range, and the second to the azimuth direction. In raw format, however, human observers cannot yet interpret this data, as the backscatter signal of an object is distributed over range and azimuth direction. In order to display the observed scene, it is necessary to focus the data. As illustrated in Figure 2.3, the focusing process consists of two successive steps: range compression and azimuth compression [35].

Range compression. The signals transmitted by a SAR sensor are frequency-modulated signals known as chirps. The amplitude of the signal remains constant over the transmission period of a pulse, while the phase varies linearly over time. Within the range compression, the chirp signal is compressed to a short pulse in order to localize the total energy reflected from an object in one pixel along the range direction. For this purpose, each range line is convolved with the complex conjugate of the spectrum of the transmitted chirp.

Azimuth compression. Within the range compression, the received signal is also convolved with a reference function. In this case, however, the reference function is not derived from the chirp signal but corresponds to the complex conjugate of a physically modeled backscatter signal expected from a point target on the ground. This reference

function represents a signal with linear frequency modulation. The frequency varies linearly with time and has an azimuth-frequency rate that is inversely proportional to the slant range. This results in long chirps in the far range and a short chirp in the near range (see Figure 2.3).

In the resulting focused SAR image, each pixel is represented by a complex number. To display the SAR image, typically the intensity, defined as the square of the amplitude of the complex value, is used, which indicates the reflectivity of the corresponding point on the ground. Common post-processing steps include radiometric calibration, which ensures that the intensity indicates the radar cross section normalized to the respective area [32], and geocoding, in which pixels in the slant range are projected into ground range so that each pixel corresponds directly to its position on the ground.

Speckle Effect. A characteristic effect observed in SAR images is the so-called speckle effect [36, 37]. It occurs if several randomly distributed elemental scatterers are located within a resolution cell. This is reflected in the SAR image by strong differences in the backscattered coefficients of neighboring pixels. The effect occurs because the amplitude and phase of the backscattered signal result from the coherent sum of amplitudes and phases of the individual scattering processes. Consequently, the measured signal is not deterministic. Instead, the amplitudes follow a Rayleigh distribution, and the phases are uniformly distributed [37]. To improve the visual appearance of SAR images, various approaches have been developed to mitigate speckle. Comprehensive reviews can be found in [37–39]. Since speckle has a multiplicative character, it cannot be suppressed by increasing the transmitted power. A commonly used method to reduce speckle is a non-coherent spatial averaging over a small pixel neighborhood known as multilooking.

2.2 SAR Polarimetry

To understand the principle of PolSAR, it is necessary first to consider wave polarimetry, a concept used to describe the polarization state of electromagnetic waves. After outlining the fundamentals of wave polarimetry, the concepts of scattering polarimetry are presented.

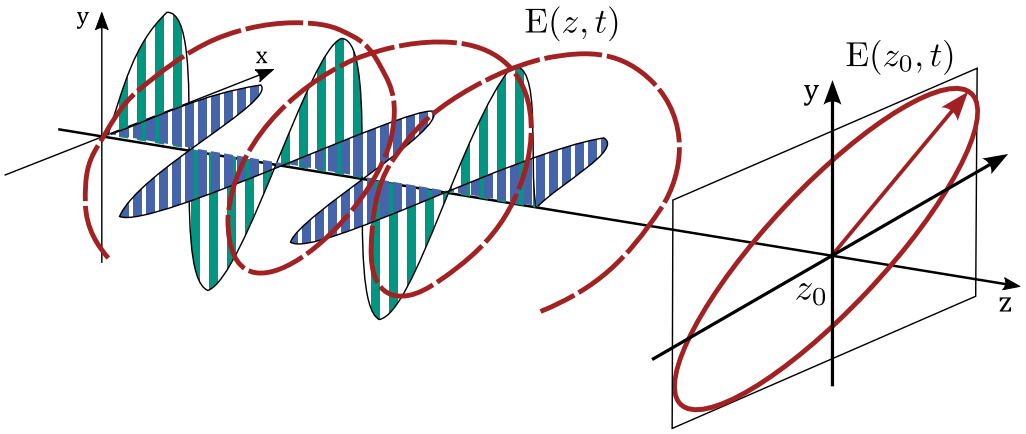


Figure 2.4.: The polarization ellipse describes the trajectory of the electric field at position z_0. The figure is redrawn from [41].

2.2.1 Wave Polarimetry

The real electric field of an electromagnetic plane wave propagating along the z-direction is always orthogonal to the propagation direction and can be described by [40]:

$$\mathrm{E}(z,t) = \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \begin{bmatrix} E_{0x}cos(\omega t - kz + \delta_x) \\ E_{0y}cos(\omega t - kz + \delta_y) \\ 0 \end{bmatrix}, \tag{2.6}$$

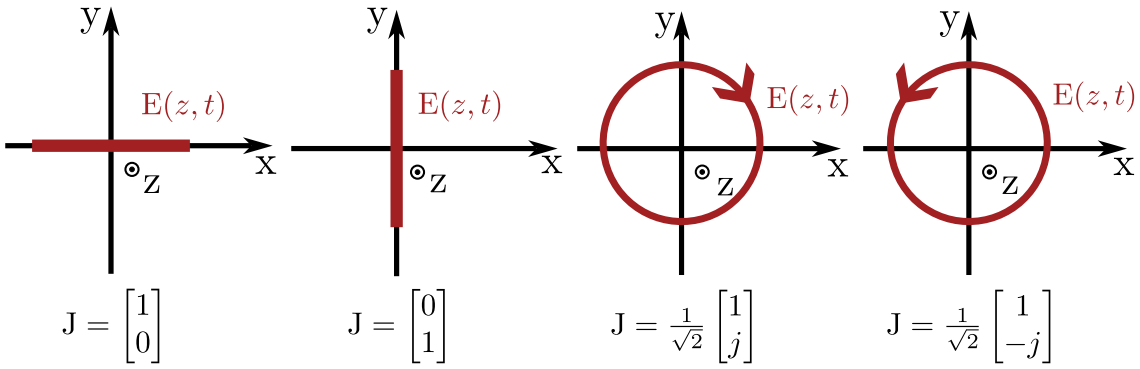
at a location z for a time t. Here, E_{0x} and E_{0y} indicate the amplitude of the wave in $x-$ and $y-$directions and δ_x, δ_y specify the corresponding phases of the wave components. The electric field rotates in the $x-y$ plane with a fixed angular frequency ω. The wave vector $\mathbf{k}_z$:

$$\mathbf{k}_z = \begin{bmatrix} 0 \\ 0 \\ k \end{bmatrix}, \tag{2.7}$$

indicates the propagation of the wave parallel to the z-direction.

The polarization state of an electromagnetic wave is completely determined by the ratio of the amplitudes E_{0x}/E_{0y} and the phase difference $\delta_y - \delta_x$. It can be illustrated by the polarization ellipse that is obtained by observing the electric field at a fixed location z_0 as shown in Figure 2.4.

Since amplitudes and phase differences of a polarized wave remain unchanged during propagation, the shape of the ellipse and, thus, the polarization state is independent of space and time. A space- and time-independent description of the polarization state is



(a) Horizontal polarization. **(b)** Vertical polarization. **(c)** Left circular polarization. **(d)** Right circular polarization.

Figure 2.5.: The polarization ellipse (red) is shown for special polarization cases in the H-V basis alongside their corresponding Jones vectors J. The figure is redrawn from [41].

given by the complex wave vector J, which is related to the space-time description of the real electric wave vector $\mathrm{E}(z,t)$ (given in Equation 2.6) at location $z = 0$ as follows:

$$\mathrm{E}(0,t) = \begin{bmatrix} E_{0x}cos(\omega t - \delta_x) \\ E_{0y}cos(\omega t - \delta_y) \end{bmatrix} = \mathrm{Re}\left(\begin{bmatrix} E_{0x}e^{j\delta_x} \\ E_{0y}e^{j\delta_y} \end{bmatrix} e^{j\omega t} \right) = \mathrm{Re}\left(\mathrm{J}e^{j\omega t} \right). \tag{2.8}$$

The complex-valued vector J is called the Jones vector and fully characterizes the polarization state of a monochromatic, uniform plane wave [42]. In Figure 2.5, some canonical polarization states are shown by means of the polarization ellipse and the corresponding Jones vector.

Linear Polarization. An electromagnetic wave is referred to as linearly polarized if the phase difference is $\delta_y - \delta_x = q\pi$ (where $q = 0, \pm 1, \pm 2, ...$). In this case, the electric field oscillates along one direction, and the polarization ellipse simplifies to a line. Special cases are the horizontal and vertical linear polarization shown in Figures 2.5a and 2.5b, in which the electric field oscillates along the x- or y-axis.

Circular Polarization. An electromagnetic wave is described as circularly polarized if the amplitudes of the x- and y-components are equal ($E_{0x} = E_{0y}$) and the phase difference equals $\delta_y - \delta_x = p\pi/2$ (where $p = \pm 1, \pm 3, \pm 5, ...$). This indicates that the field components are shifted by 90° relative to each other. In this case, the electric field moves along a circle (see Figures 2.5c and 2.5d). A distinction is made between left-circularly and right-circularly polarization, depending on whether the field rotates clockwise or anti-clockwise from the perspective of an observer looking towards the source.

2.2.2 Scattering Polarimetry

A polarimetric SAR system transmits pulses of two orthogonally polarized waves and receives the backscattered waves using orthogonally arranged antennas [40]. Generally, horizontally and vertically polarized waves (linear H-V basis) or left- and right-circularly polarized waves (circular basis) are used. The measured backscatter signal depends on the properties of the transmitted wave and the geophysical and geometrical characteristics of the observed scatterers. Thus, information about the scatterer itself can be derived by analyzing the relationship between the transmitted and backscattered electromagnetic waves. An important aspect in this respect is the analysis of the changes occurring in the polarization state.

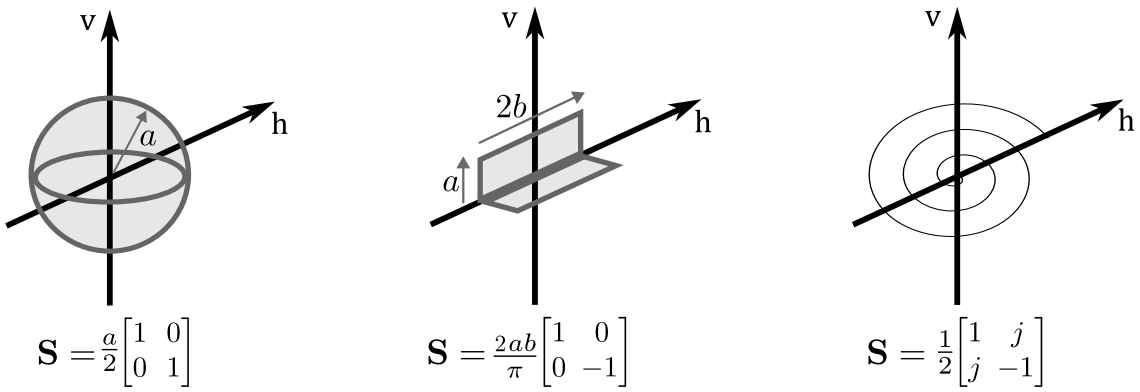
Scattering Matrix. The relationship between transmitted and received waves, which are described in the form of Jones vectors J^t and J^r, can be described by the complex-valued 2×2 scattering matrix $\mathbf{S}$:

$$\mathrm{J}^r = \frac{\mathrm{e}^{-jkr}}{r}\mathbf{S}\mathrm{J}^t = \frac{\mathrm{e}^{-jkr}}{r}\begin{bmatrix} s_{hh} & s_{hv} \\ s_{vh} & s_{vv} \end{bmatrix}\mathrm{J}^t. \tag{2.9}$$

The subscripts (h for horizontally polarized and v for vertically polarized) indicate the transmitted or received polarization, respectively. The factor e^{-jkr}, where k is the wavenumber and r, the distance between sensor and scatterer, indicates the phase shift resulting from the traveled distance. The coefficient $1/r$ is the attenuation of a spherical wave. The elements of the scattering matrix are referred to as complex scattering amplitudes. The complex scattering amplitudes s_{hh} and s_{vh} in the first column result from the transmission of a horizontally polarized wave and measurement of the backscattered signal in horizontal and vertical polarization, respectively, using two corresponding antennas. To measure the complex scattering amplitudes s_{hv}, s_{vv} (second column), a vertically polarized wave is transmitted and received in horizontal and vertical polarization.

As already mentioned, instead of the H-V basis, another orthogonal basis can equally be used to measure the scattering matrix. Changing the basis only changes the representation but not the information content of the measurement. By applying a basis transformation, the scattering matrix measured in one basis can be converted into the scattering matrix that would result from measurement in another basis.

If a monostatic configuration is used for data acquisition, for reciprocal scattering media, i.e., media that show direction-independent scattering behavior, the simplification $s_{hv} = s_{vh}$ applies. Further, in the analysis of polarimetric SAR data, the absolute phases of the complex scattering amplitudes are not of interest, and only the phase differences between the individual elements are considered. For a monostatic configuration, the polarimetric



(a) Sphere.

(b) Dihedral corner reflector.

(c) Right-handed helix.

Figure 2.6.: Scattering matrices that result from scattering processes on canonical bodies. The figure is redrawn from [40].

scattering information is thus defined by five independent parameters: three amplitude values $|s_{hh}|$, $|s_{hv}|$ and $|s_{vv}|$ and two relative phases $\varphi_{hv} - \varphi_{hh}$, $\varphi_{vv} - \varphi_{hh}$:

$$\mathbf{J}^r = \underbrace{\frac{\mathrm{e}^{-jkr}\mathrm{e}^{j\varphi_{hh}}}{r}}_{\text{Absolute phase}} \underbrace{\begin{bmatrix} |s_{hh}| & |s_{hv}|e^{j(\varphi_{hv}-\varphi_{hh})} \\ |s_{hv}|e^{j(\varphi_{hv}-\varphi_{hh})} & |s_{vv}|e^{j(\varphi_{vv}-\varphi_{hh})} \end{bmatrix}}_{\text{Relative scattering matrix}} \mathbf{J}^t. \tag{2.10}$$

As described in Section 2.2.1, a monochromatic, completely polarized electromagnetic wave can be fully described by its associated Jones vector J. Consequently, if the backscattered wave is also monochromatic and completely polarized, the scattering process can be fully described by the scattering matrix **S**. This requires a stable, monochromatic, and completely polarized incident wave as well as an anisotropic and stationary scatterer that maintains or changes the polarization state of the incident wave in a predictable way. Scatterers that fulfill these criteria are called deterministic targets. Models of scattering processes occurring on exemplary selected canonical bodies are shown in Figure 2.6.

Covariance and Coherency Matrix. In a PolSAR image, each pixel is described by its corresponding scattering matrix. The scattering matrix contains an absolute, non-measurable phase component that changes from pixel to pixel due to distance differences, topographical variations, different dielectric properties, and geometric orientation of the target objects. The change in absolute phase prohibits basic image processing steps such as spatial filtering, as the addition or averaging of scattering matrices with different absolute phases can lead to inconsistent or physically meaningless results. An alternative description of the scattering process is therefore presented below, in which the absolute phase is eliminated.

To construct this descriptor, the scattering matrix $\mathbf{S}$ is first vectorized so that the scattering vector $\mathbf{k}$ is obtained:

$$\mathbf{k} = \frac{1}{2}\text{trace}(\mathbf{S}B). \tag{2.11}$$

Here, B is a set of 2×2 complex basis matrices that form an orthonormal set under a hermitian inner product. In SAR polarimetry usually the lexicographic ordering basis:

$$B_l = \left\{ 2\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, 2\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, 2\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, 2\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \tag{2.12}$$

or the Pauli basis:

$$B_p = \left\{ \sqrt{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sqrt{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \sqrt{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sqrt{2}\begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix} \right\} \tag{2.13}$$

is used [40]. This results in the scattering vectors $\mathbf{k}_l$ and $\mathbf{k}_p$:

$$\mathbf{k}_l = \begin{bmatrix} s_{hh} \\ s_{hv} \\ s_{vh} \\ s_{vv} \end{bmatrix}, \quad \mathbf{k}_p = \frac{1}{\sqrt{2}}\begin{bmatrix} s_{hh} + s_{vv} \\ s_{hh} - s_{vv} \\ s_{hv} + s_{vh} \\ j(s_{hv} - s_{vh}) \end{bmatrix}, \tag{2.14}$$

which simplifies in the monostatic case (assuming reciprocity property) to:

$$\mathbf{k}_l = \begin{bmatrix} s_{hh} \\ \sqrt{2}s_{hv} \\ s_{vv} \end{bmatrix}, \quad \mathbf{k}_p = \frac{1}{\sqrt{2}}\begin{bmatrix} s_{hh} + s_{vv} \\ s_{hh} - s_{vv} \\ 2s_{hv} \end{bmatrix}. \tag{2.15}$$

The information contained in the two vector representations is equivalent. The choice of the basis depends only on the final purpose. The lexicographic representation is suitable for system-related analyses or statistical data analysis. The Pauli representation, on the other hand, is more suitable for relating the measurement signals to physical processes and interpreting them. This can be seen from the matrix decomposition using the Pauli basis:

$$\mathbf{S} = \begin{bmatrix} a+b & c-jd \\ c+jd & a-b \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + c\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix}. \tag{2.16}$$

The scattering matrix is decomposed into matrices that correspond to the scattering process induced by canonical bodies (some of them illustrated in Figure 2.6). The complex

coefficients a, b, c, and d, which represent the elements of the scattering vector $\mathbf{k}_p$, reflect the contribution of different scattering mechanisms:

- $a = (s_{hh} + s_{vv})/\sqrt{2}$ corresponds to single scattering from a sphere or plane surface, commonly denoted as surface or odd-bounce scattering.
- $b = (s_{hh} - s_{vv})/\sqrt{2}$ corresponds to dihedral scattering, commonly denoted as double-bounce scattering.
- $c = (s_{hv} + s_{vh})/\sqrt{2}$ corresponds to dihedral scattering occurring with a relative orientation of $-\pi/4$ along the line of sight.
- $d = j(s_{hv} + s_{vh})/\sqrt{2}$ corresponds to helix scattering, in which the backscattered wave changes its polarization state from linear to circular polarization.

To eliminate the absolute phase, which is contained both in the scattering matrix $\mathbf{S}$ and in the scattering vectors $\mathbf{k}_l$ and $\mathbf{k}_p$, the outer product of the scattering vector is formed with its conjugate complex. This results in a matrix that contains only relative phases. Assuming reciprocity and a monostatic configuration, the resulting matrices are:

$$\mathbf{C} = \mathbf{k}_l \mathbf{k}_l^{T*} = \begin{bmatrix} |s_{hh}|^2 & \sqrt{2} s_{hh} s_{hv}^* & s_{hh} s_{vv}^* \\ \sqrt{2} s_{hv} s_{hh}^* & |s_{hv}|^2 & \sqrt{2} s_{hv} s_{vv}^* \\ s_{vv} s_{hh}^* & \sqrt{2} s_{vv} s_{hv}^* & |s_{vv}|^2 \end{bmatrix} \tag{2.17}$$

and

$$\mathbf{T} = \mathbf{k}_p \mathbf{k}_p^{T*} = \begin{bmatrix} |s_{hh} + s_{vv}|^2 & (s_{hh} + s_{vv})(s_{hh} - s_{vv})^* & 2(s_{hh} + s_{vv}) s_{hv}^* \\ (s_{hh} - s_{vv})(s_{hh} + s_{vv})^* & |s_{hh} - s_{vv}|^2 & 2(s_{hh} - s_{vv}) s_{hv}^* \\ 2 s_{hv} (s_{hh} + s_{vv})^* & 2 s_{hv} (s_{hh} - s_{vv})^* & 4 |s_{hv}|^2 \end{bmatrix}. \tag{2.18}$$

The symbol * denotes the complex conjugation, and ^{T} indicates the transposition operation. The matrices $\mathbf{C}$ and $\mathbf{T}$ are complex, hermitian matrices with rank one. The rows and columns of $\mathbf{C}$ and $\mathbf{T}$ are therefore linearly dependent, and the number of independent parameters is still five, analogous to the scattering matrix $\mathbf{S}$. Notably, the information content of $\mathbf{C}$ and $\mathbf{T}$ is identical, and the two representations can be converted into each other using a basis transformation. The advantage of the descriptors $\mathbf{C}$ and $\mathbf{T}$ compared to the scattering matrix $\mathbf{S}$ is their independence from the absolute phase, which enables different measurements (i.e., different pixels) to be added together without causing inconsistencies or loss of information. The information in $\mathbf{C}$ or $\mathbf{T}$ remains equivalent to the information described by $\mathbf{S}$ and thus only enables the complete description of scattering processes caused by deterministic scatterers.

However, real-world scenes are usually not composed exclusively of deterministic scatterers but contain so-called distributed scatterers. Distributed scatterers are composed of several

randomly distributed deterministic scatterers that fall into a resolution cell and whose scattering processes overlap. This leads, besides the speckle effect described earlier, to the result that the backscattered wave is not a fully polarized wave but is partially polarized, i.e., the polarization state varies in space and time [1]. The backscattered wave can, therefore, no longer be described using the introduced descriptors. To describe the polarization state of partially polarized waves, a second-order statistical formalism must be applied instead. This usually involves estimating the covariance matrix $\langle \mathbf{C} \rangle$ or the coherency matrix $\langle \mathbf{T} \rangle$, which are defined as:

$$\langle \mathbf{C} \rangle = E\{\mathbf{k}_l \mathbf{k}_l^{T*}\}, \tag{2.19}$$

$$\langle \mathbf{T} \rangle = E\{\mathbf{k}_p \mathbf{k}_p^{T*}\}. \tag{2.20}$$

Here, $E\{\cdot\}$ represents the expectation value. The maximum likelihood estimator of the covariance and coherency matrix corresponds to the incoherent spatial averaging (multilook) [43]. The estimated covariance and coherency matrices are called sample covariance and sample coherency matrices. However, for convenience, they are only referred to as covariance and coherency matrices in the course of this thesis. The matrices are, by definition, hermitian positive semi-definite (i.e., the diagonal elements are non-negative, and all eigenvalues are real and non-negative) and are generally of full rank. The number of independent parameters of the matrices is, in the most general case, nine, resulting from the three real diagonal elements and the magnitude and phase of each of the three off-diagonal complex cross-correlations. Covariance or coherency matrices allow the complete description of scattering processes on distributed scatterers.

Depending on symmetry properties of observed scatterers, the number of independent parameters can be reduced. Important examples are reflection and rotational symmetry. For scatterers that are reflection-symmetric with respect to the line-of-sight, the covariance and coherency matrix have five independent parameters and are simplified to:

$$\mathbf{C} = \begin{bmatrix} c_{11} & 0 & c_{13} \\ 0 & c_{22} & 0 \\ c_{31} & 0 & c_{33} \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & 0 \\ t_{21} & t_{22} & 0 \\ 0 & 0 & t_{33} \end{bmatrix}. \tag{2.21}$$

If an observed surface is additionally rotationally symmetric, the so-called azimuthal symmetry results, in which the number of independent parameters is reduced to two. The simplified coherency matrix results in:

$$\mathbf{T} = 2 \begin{bmatrix} t_{11} & 0 & 0 \\ 0 & t_{22} & 0 \\ 0 & 0 & t_{33} \end{bmatrix} \quad \text{with} \quad t_{22} = t_{33}. \tag{2.22}$$

2.2.3 Polarimetric Speckle Filtering

Estimating the covariance or coherency matrix that describes the scattering process of distributed scatterers can be achieved by performing a spatial averaging of neighboring pixels. However, this method has the disadvantage of reducing the spatial resolution. To counteract this effect, a number of adaptive filtering methods have been developed that preserve the resolution during multilooking while aiming to maintain the image quality as well as possible by avoiding the smearing of edges. A review of established filters for PolSAR data is given in [44]. In this thesis, the Refined Lee filter [45] is used for estimating the coherency matrix, whose functionality is outlined below.

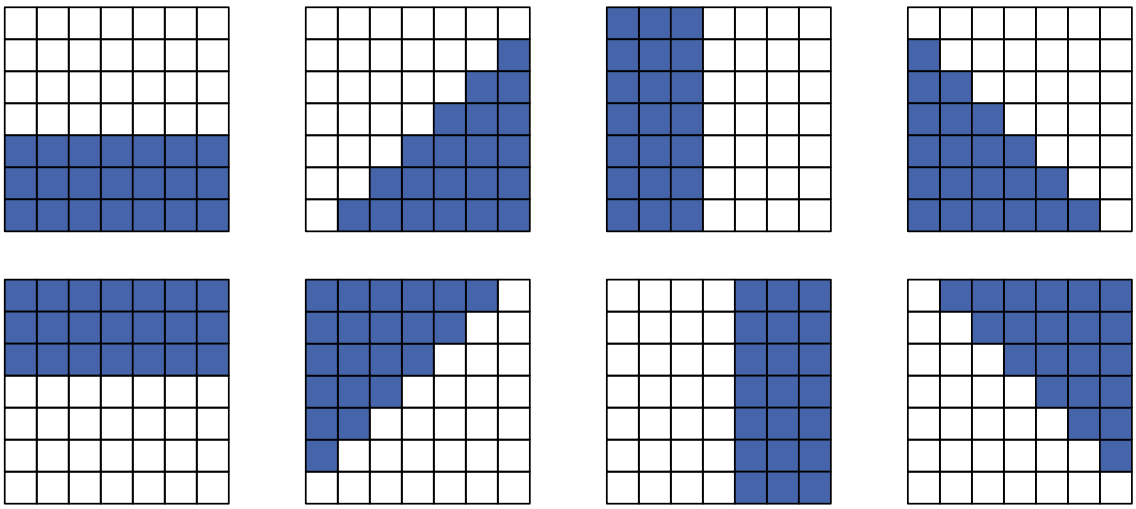
Refined Lee Filter. The Refined Lee filter, originally proposed in [46] for speckle reduction of single-polarized SAR image data, is based on local statistics. Its adaptation to PolSAR data, whereby special care has to be taken to preserve the polarimetric properties of the data, is described in [45]. The filtering process essentially consists of three steps: first, the selection of an edge-aligned filter window. Second, the calculation of local image statistics. Third, the computation of the filtered covariance/coherency matrix based on the selected window and local statistics. These steps are explained below using the example of estimating the coherency matrix $\langle \mathbf{T} \rangle$ for a monostatic configuration. The process can be carried out analogously for the estimation of the covariance matrix $\langle \mathbf{C} \rangle$.

The use of non-squared edge-aligned filter windows is intended to preserve the sharpness of edges in the image. The possible edge-aligned filter windows that determine which pixels are considered during local filtering are shown in Figure 2.7. The selection of a suitable filter is based on the dominant edge direction of the span image within a squared pixel neighborhood. The span specifies the measured total scattered power and is defined by:

$$y = |s_{hh}|^2 + 2|s_{hv}|^2 + |s_{vv}|^2. \tag{2.23}$$

The edge direction is calculated using simple edge filter masks [46]. For each edge direction, two possible edge-aligned windows qualify, the appropriate one of which is selected so that the mean value of the included pixels is closest to the value of the central pixel. In the second step, filtering weights are calculated based on the concept of local statistical filters [47]. Specifically, a linear minimum mean-square filter, assuming a multiplicative noise model ($y = v_n x$), is applied:

$$\hat{x} = \overline{y} + b(y - \overline{y}). \tag{2.24}$$



(a) 0°. **(b)** 45°. **(c)** 90°. **(d)** 135°.

Figure 2.7.: Edge-aligned windows of the Refined Lee filter, corresponding to edges with orientations of 0°, 45°, 90°, or 135° relative to the horizontal image axis.

Where $\hat{x}$ is the filtered pixel value of the span image, y is the measured span value, and $\overline{y}$ denotes its local mean. The weighting factor b is calculated using:

$$b = \frac{\mathrm{var}(x)}{\mathrm{var}(y)}, \quad \mathrm{var}(x) = \frac{\mathrm{var}(y) - \overline{y}^2\sigma_v^2}{(1+\sigma_v^2)}. \tag{2.25}$$

Here, $\mathrm{var}(x)$ indicates the variance of the total scattered power without speckle noise, and $\mathrm{var}(y)$ is the local variance of the measured span value. The parameter σ_v^2 denotes the variance of the assumed noise v_n, which can be estimated by the standard-deviation-to-mean ratio in homogeneous areas. The weighting factor b approaches one for inhomogeneous areas (i.e., where $\mathrm{var}(x)$ is high), so that $\hat{x} \approx y$ results. In contrast, for homogeneous areas (i.e., where $\mathrm{var}(x) \approx 0$) b approaches zero and $\hat{x} \approx \overline{y}$ results. For the derivation of Equation 2.24 and Equation 2.25 see [46].

In the last step, the filtered coherency matrix $\langle \mathbf{T} \rangle$ is computed for every pixel using the neighboring pixel within the selected edge-aligned window and the corresponding filter weight b by:

$$\langle \mathbf{T} \rangle = \overline{\mathbf{T}} + b(\mathbf{T} - \overline{\mathbf{T}}), \tag{2.26}$$

where $\overline{\mathbf{T}}$ represent the local mean of $\mathbf{T}$ computed within the edge-aligned window.

2.2.4 Polarimetric Scattering Decomposition Theorems

The aim of the polarimetric target decomposition is to represent the measured backscatter signal as a weighted sum of elementary scattering mechanisms. This decomposition

facilitates the physical interpretation of PolSAR data. Depending on whether the target decomposition is performed using the scattering matrix $\mathbf{S}$ or the coherency matrix $\langle \mathbf{T} \rangle$, a distinction is made between coherent and incoherent target decomposition methods. In the following, the target decomposition methods that are employed in this thesis are described.

Coherent Target Decomposition

Coherent target decomposition can be used to describe scattering processes that arise at deterministic scatterers. For this purpose, the scattering matrix $\mathbf{S}$ is represented as a combination of scattering matrices describing elementary scattering mechanisms:

$$\mathbf{S} = \sum_{i=1}^{k} c_i \mathbf{S}_i. \tag{2.27}$$

Here $\mathbf{S}_i$ is the scattering matrix, describing an elementary scattering mechanism, and the coefficient c_i represents the contribution of $\mathbf{S}_i$ to the measured scattering matrix $\mathbf{S}$. A commonly used coherent target decomposition is the Pauli decomposition, which has already been described in Section 2.2.2 (Equation 2.16).

Further coherent target decomposition methods are the Krogager decomposition [48] and the Cameron decomposition [49]. These are not used in this thesis and are therefore not described in detail here. The interested reader is referred to the indicated references.

Incoherent Target Decomposition

The incoherent target decomposition aims to represent scattering processes occurring at distributed scatterers as a weighted sum of second-order descriptors that describe simple scattering processes. This facilitates a better understanding of complex scattering processes, which is difficult to achieve by directly considering the covariance or coherency matrix. The general definition of the decomposition of the covariance or coherency matrix is:

$$\langle \mathbf{C} \rangle = \sum_{i=1}^{k} p_i \mathbf{C}_i, \quad \langle \mathbf{T} \rangle = \sum_{i=1}^{k} q_i \mathbf{T}_i. \tag{2.28}$$

Methods of incoherent decomposition can be divided into two approaches: eigenvector and eigenvalue-based decomposition and model-based decomposition. A detailed survey of established decomposition methods is provided in [50]. In the following, only decomposition methods that are used in this work are described in detail (exemplified by the decomposition of the coherency matrix $\langle \mathbf{T} \rangle$).

Eigenvalue Decomposition. The eigenvector and eigenvalue-based target decomposition and the subsequent extraction of physically interpretable polarimetric features were introduced by Cloude et al. in [51, 52].

According to the eigendecomposition theorem, the hermitian positive semi-definite coherency matrix $\langle \mathbf{T} \rangle$ can be decomposed into:

$$\langle \mathbf{T} \rangle = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{-1}. \tag{2.29}$$

Here, $\mathbf{\Lambda}$ is a diagonal matrix composed of the eigenvalues of $\langle \mathbf{T} \rangle$:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \quad \text{where } \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0. \tag{2.30}$$

The unitary matrix $\mathbf{U}$ is formed by the corresponding eigenvectors:

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}. \tag{2.31}$$

The eigenvectors can be parameterized by:

$$\mathbf{u}_i = \begin{bmatrix} \cos(\alpha_i) \\ \sin(\alpha_i)\cos(\beta_i)e^{j\delta_i} \\ \sin(\alpha_i)\cos(\beta_i)e^{j\gamma_i} \end{bmatrix}. \tag{2.32}$$

Based on this, the coherency matrix $\langle \mathbf{T} \rangle$, which is generally of rank three, can be represented as a combination of three coherency matrices with rank one:

$$\langle \mathbf{T} \rangle = \mathbf{T}_1 + \mathbf{T_2} + \mathbf{T_3} = \lambda_1 \cdot (\mathbf{u}_1 \cdot \mathbf{u}_1^{*T}) + \lambda_2 \cdot (\mathbf{u}_2 \cdot \mathbf{u}_2^{*T}) + \lambda_3 \cdot (\mathbf{u}_3 \cdot \mathbf{u}_3^{*T}). \tag{2.33}$$

Here, the matrices $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ describe deterministic scattering contributions.

The eigenvalues and eigenvectors can be used to derive polarimetric features that can be interpreted physically. The first feature is the entropy H, which is derived from the eigenvalues:

$$H = -\sum_{i=1}^{3} p_i \log_3(p_i) \quad \text{with} \quad p_i = \frac{\lambda_i}{\sum_{j=1}^{3} \lambda_j}, \tag{2.34}$$

where p_i indicates the relative importance of an eigenvalue in relation to the total backscattering power and is often referred to as pseudo probability. The values of entropy H range from 0 to 1 and are considered a measure of the randomness of the observed scattering process. An entropy of $H = 0$ indicates that only one of the eigenvalues is non-zero and thus corresponds to a deterministic scattering process without depolarization. An entropy

of $H = 1$ results when all eigenvalues are equal, which corresponds to a random scattering process that completely depolarizes the incident wave.

The second feature is the anisotropy A, which describes the relationship of secondary scattering processes (related to the second and third eigenvalue):

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3}. \tag{2.35}$$

This feature, which has a value range from 0 to 1, can only be interpreted in combination with the entropy H. For $0.7 < H < 1$, a high anisotropy indicates that only one dominant secondary scattering mechanism is present.

The third polarimetric feature of the eigenvalue decomposition is the mean alpha angle α, defined as:

$$\alpha = \sum_{i=1}^{3} p_i \alpha_i \quad \text{with} \quad \alpha_i = \arccos(|u_{1i}|). \tag{2.36}$$

The angle α can be directly related to occurring scattering mechanisms and is interpreted as follows: an angle with $0^\circ \leq \alpha \leq 30^\circ$ indicates surface scattering, $40^\circ \leq \alpha \leq 50^\circ$ indicates volume scattering, and $60^\circ \leq \alpha \leq 90^\circ$ indicates double-bounce scattering.

Freeman Durden Decomposition. The Freeman Durden decomposition, proposed in [53], is one of the first target decomposition methods based on the physical modeling of the scattering process. Assuming reflection symmetry of the scattering process, the measured coherency matrix is modeled as the sum of three elementary scattering processes: First, surface scattering, modeled as Bragg scattering; second, double-bounce scattering, modeled as scattering caused by dihedral targets composed of two orthogonal surfaces with different dielectric properties; third, volume scattering, modeled as a volume of randomly oriented dipole scattering elements. The decomposition model results in:

$$\langle \mathbf{T} \rangle = \underbrace{f_s \begin{bmatrix} 1 & b^* & 0 \\ b & |b|^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{surface scattering}} + \underbrace{f_d \begin{bmatrix} |a|^2 & a & 0 \\ a^* & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{double-bounce scattering}} + \underbrace{\frac{f_v}{4} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{volume scattering}}. \tag{2.37}$$

The parameterization of the surface scattering component is modeled using the Bragg scattering coefficients R_s and R_p (parallel and orthogonal to the incidence plane), which

depend on the incidence angle and the permittivity of the surface under consideration. The relationship between the parameterization and the model coefficients is:

$$b = (R_s + R_p)/(R_s - R_p) \quad f_s = |R_s - R_p|^2. \tag{2.38}$$

The parameters f_d and a, which describe the double-bounce component, result from a generalized scattering model of a corner reflector. With the horizontal and vertical backscattering coefficients R_{th} and R_{tv}, caused by the vertical surface, and horizontal and vertical backscattering coefficients R_{gh} and R_{gv}, caused by the horizontal surface, the parameters are represented as:

$$f_d = \frac{1}{2}|R_{th}R_{gh} + R_{tv}R_{gv}e^{j\Phi}|^2 \quad \text{and} \tag{2.39}$$

$$a = \frac{R_{th}R_{gh} - R_{tv}R_{gv}e^{j\Phi}}{R_{th}R_{gh} + R_{tv}R_{gv}e^{j\Phi}}, \tag{2.40}$$

where Φ accounts for occurring phase shifts.

The volume scattering process is modeled by a volume of randomly oriented very thin cylinder-like scatterers, whereby the orientation angle is assumed to be uniformly distributed. A detailed derivation of the resulting coherency matrix model can be found in [53].

The equation system, given in Equation 2.37, consists of five unknowns (f_s, f_d, f_v, a, b) but only four equations and is therefore under-determined. In order to solve the equation system, a further assumption must be made. For this purpose, $a = -1$ or $b = 1$ is assigned depending on the sign of $\mathrm{Re}(\langle s_{hh}s_{vv}^* \rangle)$, which indicates whether surface scattering or double-bounce scattering is dominant.

Finally, scattering powers of surface, double-bounce, and volume scattering are described by the trace of the corresponding coherency matrix:

$$P_s = f_s(1 + |b|^2), \quad P_d = f_d(1 + |a|^2), \quad P_v = f_v. \tag{2.41}$$

Yamaguchi Decomposition. The Freeman Durden decomposition is based on the assumption that the described scattering process exhibits reflection symmetry. This assumption applies to a majority of scattering processes that occur at natural distributed scatterers, such as forest or meadow surfaces. However, this symmetry property cannot be assumed for more complex structures. For this reason, in [54], the Freeman Durden decomposition is extended to the four-component Yamaguchi decomposition, which takes into account the non-reflection symmetry cases. The modeling of surface and double-bounce scattering is analogous to the Freeman Durden decomposition. The model for volume scattering,

however, is modified. While for Freeman Durden decomposition, the orientation of randomly oriented dipoles is assumed to be uniformly distributed, Yamaguchi decomposition takes into account that the orientation may have a predominant direction. The log-ratio $r_{vv/hh} = 10\log(|s_{vv}|^2/|s_{hh}|^2)$ is used to determine whether an observed volume preferably contains horizontally oriented or vertically oriented dipoles or whether there is no predominant direction. For the individual cases, the following coherency matrix models result:

$$\text{if} \quad r_{vv/hh} < -2\text{dB} \qquad \mathbf{T}_V = \frac{1}{30}\begin{bmatrix} 15 & 5 & 0 \\ 5 & 7 & 0 \\ 0 & 0 & 8 \end{bmatrix} \tag{2.42}$$

$$\text{if} \quad -2\text{dB} \leq r_{vv/hh} < 2\text{dB} \qquad \mathbf{T}_V = \frac{1}{4}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.43}$$

$$\text{if} \quad r_{vv/hh} \geq 2\text{dB} \qquad \mathbf{T}_V = \frac{1}{30}\begin{bmatrix} 15 & -5 & 0 \\ -5 & 7 & 0 \\ 0 & 0 & 8 \end{bmatrix} \tag{2.44}$$

In addition to surface, double-bounce, and volume scattering, the Yamaguchi decomposition adds a fourth component to the model, which describes a left- or right-handed helix scattering process. This is described by:

$$\mathbf{T}_h = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & \pm j \\ 0 & \pm j & 1 \end{bmatrix}. \tag{2.45}$$

The choice of sign shows whether a left- or right-handed helix scattering process is described.

Overall, the decomposition model thus results in:

$$\langle \mathbf{T} \rangle = \underbrace{f_s \begin{bmatrix} 1 & b^* & 0 \\ b & |b|^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{surface scattering}} + \underbrace{f_d \begin{bmatrix} |a|^2 & a & 0 \\ a^* & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{double-bounce scattering}} + \underbrace{\frac{f_v}{4} \begin{bmatrix} t_{v,11} & t_{v,12} & 0 \\ t_{v,21} & t_{v,22} & 0 \\ 0 & 0 & t_{v,33} \end{bmatrix}}_{\text{volume scattering}} + \underbrace{\frac{f_c}{2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & \pm j \\ 0 & \mp j & 1 \end{bmatrix}}_{\text{helix scattering}}. \tag{2.46}$$

Van Zyl Decomposition. A disadvantage of Freeman Durden and Yamaguchi decomposition is that the resulting matrices, describing surface and double-bounce scattering, may exhibit negative eigenvalues, i.e., represent negative power. In this case, they do not

represent a physically possible scattering process. Negative eigenvalues can occur because, in both approaches, surface and double-bounce scattering is determined after subtraction of the estimated volume contribution. If the volume scattering component is overestimated, negative eigenvalues may occur.

To prevent the decomposition components from having negative eigenvalues, the non-negative eigenvalue decomposition is proposed in [55]. The measured covariance matrix $\langle \mathbf{C} \rangle$ is modeled by:

$$\langle \mathbf{C} \rangle = \alpha \langle \mathbf{C}_{model} \rangle + \mathbf{C}_{remainder}. \tag{2.47}$$

Here, $\langle \mathbf{C}_{model} \rangle$ represents the covariance matrix based on a physical model, and $\mathbf{C}_{remainder}$ contains the remaining components of the measured matrix that is not covered by the model matrix. The parameter α is chosen to be maximal under the condition that all three eigenvalues of $\mathbf{C}_{remainder}$ are greater than or equal to zero.

2.2.5 Polarimetric Features

In the following, further polarimetric features used in this thesis are presented, each characterizing a specific aspect of polarimetric information.

Total Scattered Power. An important parameter that characterizes the scattering process at an observed scatterer is the total scattered power, which is referred to as *span* and results from

$$span = |s_{hh}|^2 + 2|s_{hv}|^2 + |s_{vv}|^2. \tag{2.48}$$

This parameter is independent of the polarization basis used for the measurement and representation.

Polarimetric Coherence. The polarimetric coherence is a measure of the correlation between SAR images of different polarizations [1]. Based on the coherency matrix $\langle \mathbf{T} \rangle$, it is determined as the amplitude of the complex correlation coefficient $\rho_{t_{ij}}$:

$$\rho_{t_{ij}} = \frac{\langle t_{ij} \rangle}{\sqrt{\langle t_{ii} \rangle} \sqrt{\langle t_{jj} \rangle}}. \tag{2.49}$$

Where $\langle t_{ij} \rangle$ is the complex-valued off-diagonal element of the coherency matrix $\langle \mathbf{T} \rangle$ in row i and column j. An equivalent calculation can be performed using the covariance matrix $\langle \mathbf{C} \rangle$, which is then denoted as $\rho_{c_{ij}}$.

Conformity Coefficient. The conformity coefficient μ is proposed in [56] as indicator for bare surface scattering. It is defined as:

$$\mu = 2\frac{\mathrm{Re}(s_{hh}s_{vv}) - |s_{hv}|^2}{|s_{hh}|^2 + |s_{vv}|^2 + 2|s_{hv}|^2}. \tag{2.50}$$

A conformity coefficient value of $\mu = 1$ corresponds to surface scattering, whereas a value of $\mu = -1$ indicates double-bounce scattering.

Optical Polarimetry Features. In [57], Praks et al. introduce polarimetric features that provide physical interpretation inspired by features used in optical polarimetry. These are namely scattering predominance $\hat{H}_p$, scattering diversity $\hat{H}_d$, degree of purity d_p and the depolarization index d_{idx}.

The starting point for calculating these features is the power-normalized coherency matrix $\mathbf{N}$:

$$\mathbf{N} = \frac{\langle \mathbf{T} \rangle}{\mathrm{trace}(\langle \mathbf{T} \rangle)}. \tag{2.51}$$

An important characteristic of this matrix is that its eigenvalues are equal to the pseudo probabilities computed in the context of eigenvalue-based target decomposition (see Equation 2.34). The proposed features are based on the Frobenius norm of $\mathbf{N}$, which is defined as the square root of the sum of all squared matrix elements n_{ij} with the following relationships:

$$\|\mathbf{N}\|_F = \sqrt{\sum_{i=1}^{3}\sum_{j=1}^{3}|n_{ij}|^2} = \sqrt{\mathrm{trace}(\mathbf{N}^{*T}\mathbf{N})} = \sqrt{\lambda_{N,1}^2 + \lambda_{N,2}^2 + \lambda_{N,3}^2}, \tag{2.52}$$

where $\lambda_{N,i}$ are the eigenvalues of $\mathbf{N}$.

The square of the Frobenius $\|\mathbf{N}\|_F^2$ is called scattering predominance $\hat{H}_p$ and can be interpreted as a measure for the amount of power that a single scattering mechanism contributes to the total scattered power. If only one scattering mechanism is present, $\hat{H}_p$ becomes 1. For three equally occurring scattering mechanisms, $\hat{H}_p$ is $1/3$. The scaled version of scattering predominance is called scattering diversity $\hat{H}_d$ and is given by:

$$\hat{H}_d = \frac{3}{2}(1 - \|\mathbf{N}\|_F^2). \tag{2.53}$$

A further feature derived from optical polarimetry [58] is the degree of purity d_p defined as:

$$d_p = 2\sqrt{\|\mathbf{N}\|_F^2 - 0.25}. \tag{2.54}$$

If the transmitted fully polarized wave is completely depolarized within the scattering process, then the degree of purity is $d_p = 0$. If no depolarization occurs, as is the case for a deterministic point target, the result is $d_p = \sqrt{3}$. When scaling the degree of purity to a value ranging from 0 to 1, it is called the depolarization index d_{idx}.

2.3 Polarimetric SAR Interferometry

SAR interferometry is based on the comparison of the phases of two SAR images that depict the same scene from slightly different positions. The resulting information can be used to realize a number of applications, including the highly accurate measurement of ground deformation, topography mapping, and monitoring of geophysical processes such as volcanic activities or glacier movement [35].

This section first describes the operating principle of SAR interferometry [59] and subsequently presents the combination of SAR polarimetry and interferometry [2].

2.3.1 SAR Interferometry

Interferometric SAR image pairs can be acquired by along-track measurement, where the receiving antennas are laterally displaced, or by across-track measurement, where antennas are locally separated along the flight direction. Along-track interferometry can be used to measure line-of-sight velocities, while across-track interferometry is primarily used for topographic measurements. Since only across-track interferometry is used in this work, this concept is outlined in the following.

For across-track interferometry, either a single-pass configuration, in which only one antenna transmits a signal and two laterally displaced antennas receive the backscattered signal, or a repeat-pass configuration, in which each antenna transmits and receives its own echoes, can be used. Figure 2.8 shows the recording principle of an interferometric across-track measurement. It can be seen that the displacement of the recording antennas, which is referred to as the perpendicular baseline $B_\perp$, leads to a difference in the range distance Δr, which depends on the surface height Δh. Assuming a large range distance r_0 and short baseline $B_\perp$, the following relationship applies [35]:

$$\Delta r \approx \frac{B_\perp}{r_0 \sin(\theta)} \cdot \Delta h. \tag{2.55}$$

The range difference Δr is related to the measured phase difference $\Delta\varphi$ as follows:

$$\Delta\varphi = m\frac{2\pi}{\lambda}\Delta r, \tag{2.56}$$

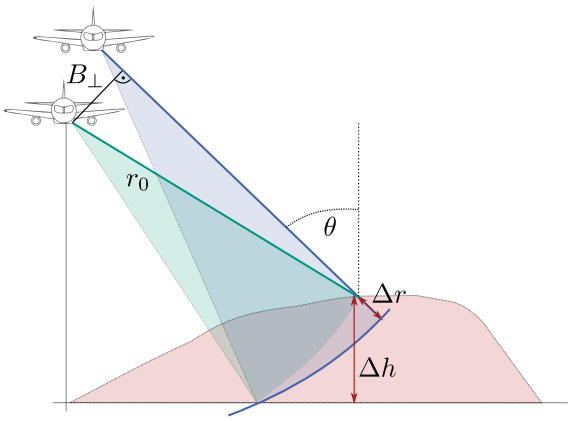


Figure 2.8.: Configuration of an across-track SAR interferometry measurement. The figure illustrates that the range difference Δr changes due to a change in the surface height Δh. The figure is redrawn from [35].

where m is one for a single-pass and two for a repeat-pass configuration.

To determine the phase difference of an interferometric image pair, both images must first be registered to each other so that the same pixel coordinates represent the same section of the scene. This step is also referred to as co-registration. To form the so-called complex interferogram, the average hermitian product of two complex scalar signals s_1, corresponding to the first image, and s_2, corresponding to the second image, is calculated [2]:

$$\mathbf{I} = \left\langle \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \begin{bmatrix} s_1^* & s_2^* \end{bmatrix} \right\rangle = \begin{bmatrix} \langle s_1 s_1^* \rangle & \langle s_1 s_2^* \rangle \\ \langle s_2 s_1^* \rangle & \langle s_2 s_2^* \rangle \end{bmatrix}. \tag{2.57}$$

The resulting phase difference $\Delta\varphi$ which is also referred to as the interferometric phase Φ, is derived as:

$$\Phi = \arg(\langle s_1 s_2^* \rangle) = \arctan \frac{\mathrm{Im}(\langle s_1 s_2^* \rangle)}{\mathrm{Re}(\langle s_1 s_2^* \rangle)} + 2\pi q \quad (q = 0, \pm 1, \pm 2, \ldots), \tag{2.58}$$

and is ambiguous by integer multiples of 2π. The interferometric phase Φ shows a systematic trend along the range direction that can be considered as the phase resulting from the recording of an ideally flat Earth. This phase component is usually extracted from the measured interferometric phase during pre-processing, which is referred to as flat Earth removal. The remaining phase component primarily results from the topography of the imaged scene.

An important measure of phase accuracy is the magnitude of the complex interferometric coherence $|\gamma|$, which describes the degree of correlation between the interferometric SAR image pair. The interferometric coherence is defined by the normalized complex cross correlation:

$$\gamma = \frac{\langle s_1 s_2^* \rangle}{\sqrt{\langle s_1 s_1^* \rangle} \cdot \sqrt{\langle s_2 s_2^* \rangle}}. \tag{2.59}$$

The interferometric coherence can be modeled as a combination of different contributions:

$$\gamma = \gamma_{SNR} \cdot \gamma_{baseline} \cdot \gamma_{temporal}. \tag{2.60}$$

The first factor γ_{SNR} describes the loss of coherence due to receiver noise. The second factor $\gamma_{baseline}$ corresponds to decorrelation caused by different look angles used for capturing the two images. The third factor $\gamma_{temporal}$ denotes temporal decorrelation that occurs due to changes in the scattering behavior of the observed area between the two time points of acquisition.

The sensitivity of the interferometric phase to small height differences Δh can be described by [59]:

$$\frac{\Delta\varphi}{\Delta h} = \frac{2m\pi B_\perp}{\lambda R_0 \sin(\theta)}. \tag{2.61}$$

Instead of the height sensitivity, the height of ambiguity is often used, which describes the height that leads to a phase change of 2π:

$$h_{\text{amb}} = \frac{\lambda m r_0 \sin(\theta)}{B_\perp}. \tag{2.62}$$

The height measurement is thus ambiguous by multiples of h_{amb}, which is related to the ambiguity of the interferometric phase (Equation 2.58). These ambiguities can be resolved using phase unwrapping methods.

Phase unwrapping is an ill-posed inverse problem that requires assumptions to be made or additional information to be included in order to solve it unambiguously. In general, the phase continuity assumption [60] is assumed, which states that the absolute phase difference between adjacent pixels in the 2-dimensional interferogram does not exceed π. If this condition applies to all positions in the interferogram, phase unwrapping can be solved by simple path integration. However, since this is usually not the case in real data due to phase noise and phase aliasing, correct phase unwrapping is a challenging task that is being actively researched. A detailed overview of existing phase unwrapping methods can be found in [61]. Phase unwrapping methods can be broadly divided into path-following methods, which aim to find the most reasonable integration path based on external information, and optimization-based methods, which formulate phase unwrapping as a global optimization problem. Path-following methods can be further categorized into branch-cut algorithms and quality-guided methods. The most prominent approach of the first category is Goldstein's branch cut algorithm [62], which identifies and places cuts along discontinuities in the phase data based on the distribution of residuals indicating phase jumps. This divides the phase path into consistent segments, which are unwrapped separately. In contrast, quality-guided path-following methods [63] use quality metrics

(for example, interferometric coherence) to first unwrap the phase in regions with high reliability and then gradually progress to regions of lower reliability.

In order to convert the phase into height information, various approaches were developed and optimized [64]. However, height retrieval is not the focus of this thesis, and only interferometric phases are used. The interested reader is therefore referred to the referenced literature at this point.

2.3.2 Combining Polarimetry and Interferometry

In order to transfer the expression of the interferometric phase and the interferometric coherence derived from Equation 2.57 to polarimetric images, the measured signals are represented using the scattering vectors $\mathbf{k}_{p,1}$, $\mathbf{k}_{p,2}$ instead of the complex scalar signals s_1, s_2. This results in the 6×6 hermitian positive semidefinite matrix $\langle \mathbf{T_6} \rangle$:

$$\langle \mathbf{T_6} \rangle = \left\langle \begin{bmatrix} \mathbf{k}_{p,1} \\ \mathbf{k}_{p,2} \end{bmatrix} \begin{bmatrix} \mathbf{k}_{p,1}^* & \mathbf{k}_{p,2}^* \end{bmatrix} \right\rangle = \begin{bmatrix} \langle \mathbf{T}_{11} \rangle & \langle \mathbf{\Omega}_{12} \rangle \\ \langle \mathbf{\Omega}_{12} \rangle^{*T} & \langle \mathbf{T}_{22} \rangle \end{bmatrix}. \tag{2.63}$$

The matrices $\langle \mathbf{T}_{11} \rangle$ and $\langle \mathbf{T}_{22} \rangle$ represent the polarimetric coherency matrices corresponding to the first and the second image. The complex matrix $\langle \mathbf{\Omega}_{12} \rangle$ contains the interferometric relations of the different polarimetric channels. Since in general $\mathbf{k}_{p,1} \neq \mathbf{k}_{p,2}$, $\langle \mathbf{\Omega}_{12} \rangle$ is not hermitian positive semidefinite.

An important PolInSAR observable is the interferometric coherence, already introduced in Equation 2.59 for single-polarized interferometric SAR image pairs. For polarimetric data, it is expressed as:

$$\gamma = \frac{\mathbf{w}_1^{*T} \langle \mathbf{\Omega_{12}} \rangle \mathrm{w}_2}{\sqrt{\mathbf{w}_1^{*T} \langle \mathbf{T}_{11} \rangle \mathrm{w}_1} \sqrt{\mathbf{w}_2^{*T} \langle \mathbf{T}_{22} \rangle \mathrm{w}_2}}. \tag{2.64}$$

The vectors $\mathbf{w}_1$, $\mathbf{w}_2$, are normalized complex projection vectors, which can be interpreted as two scattering mechanisms [2]. Since the choice of projection vectors influences the interferometric coherence, the coherence can be optimized by selecting suitable projection vectors. Various optimization methods have been developed for this purpose [2, 65, 66]. The most general approach to identify projection vectors that maximize interferometric coherence is proposed in [2], which consists of maximizing the complex Lagrangian L:

$$L = \mathbf{w}_1^{*T} \langle \mathbf{\Omega}_{12} \rangle \mathbf{w}_2 + l_1 (\mathbf{w}_1^{*T} \langle \mathbf{T}_{11} \rangle \mathbf{w}_1 - C_1) + l_2 (\mathbf{w}_2^{*T} \langle \mathbf{T}_{22} \rangle \mathbf{w}_2 - C_2). \tag{2.65}$$

Here l_1 and l_2 are Lagrange multipliers. A solution can be found by setting partial derivatives to zero:

$$\frac{\delta L}{\delta \mathbf{w}_1^{*T}} = \langle \mathbf{\Omega}_{12} \rangle \mathbf{w}_2 + l_1 \langle \mathbf{T}_{11} \rangle \mathbf{w}_1 = 0 \tag{2.66}$$

$$\frac{\delta L}{\delta \mathbf{w}_2^{*T}} = \langle \mathbf{\Omega}_{12} \rangle^{*T} \mathbf{w}_1 + l_2^* \langle \mathbf{T}_{22} \rangle \mathbf{w}_2 = 0. \tag{2.67}$$

This can be formulated as two eigenvalue problems with common eigenvalues $u = l_1 l_2^*$:

$$\langle \mathbf{T}_{22} \rangle^{-1} \langle \mathbf{\Omega}_{12} \rangle^{*T} \langle \mathbf{T}_{11} \rangle^{-1} \langle \mathbf{\Omega}_{12} \rangle \mathbf{w}_2 = u \mathbf{w}_2 \tag{2.68}$$

$$\langle \mathbf{T}_{11} \rangle^{-1} \langle \mathbf{\Omega}_{12} \rangle \langle \mathbf{T}_{22} \rangle^{-1} \langle \mathbf{\Omega}_{12} \rangle^{*T} \mathbf{w}_1 = u \mathbf{w}_1. \tag{2.69}$$

Based on this formulation, the optimal scattering mechanisms $\mathbf{w}_{opt,1}$ and $\mathbf{w}_{opt,1}$ are obtained as the corresponding eigenvectors. They can be inserted into Equation 2.64 to compute the optimal interferometric coherence vector. The corresponding optimal interferometric phase vector is given by:

$$\Phi_{opt} = \arg(\mathbf{w}_{opt,1}^{*T} \langle \mathbf{\Omega_{12}} \rangle \mathbf{w}_{opt,2}). \tag{2.70}$$

While in this thesis, only optimal interferometric coherences and phases are used directly, the polarimetric interferometric coherence can also be used to derive information about the vertical structure of volume scatterers [40]. These allow several applications, including the estimation of forest height [67], biomass [3], or snow depth [4].

3 The Pol-InSAR-Island Benchmark Dataset

The content of this chapter is published in:

[68] Sylvia Hochstuhl, Niklas Pfeffer, Antje Thiele, Stefan Hinz, Joel Amao-Oliva, Rolf Scheiber, Andreas Reigber, and Holger Dirks. “Pol-InSAR-Island - A benchmark dataset for multi-frequency Pol-InSAR data land cover classification”. In: *ISPRS Open Journal of Photogrammetry and Remote Sensing* 10 (2023), p. 100047.

The adopted text passages are marked with a blue line. Omitted or inserted text passages are marked in square brackets [].

3.1 Introduction

In the current development of effective methods for land cover classification of PolSAR images, attention is mainly focused on methods from the field of machine learning, especially on its subfield deep learning as summarized in [69] and [25]. This trend is primarily motivated by the tremendous success of deep learning methods for the analysis of optical image data. While this success has been strongly driven by public benchmark datasets such as ImageNet [70] and CityScale [71], which contain a large number of labeled training and test data, comparable datasets for the PolSAR domain are rare. Due to the inherent differences in imaging techniques, the use of models pre-trained on optical data is inappropriate for the analysis of PolSAR data. Therefore, models must be trained from scratch, which requires domain-specific datasets. The lack of highly demanded large and representative expert-labeled benchmark datasets for the SAR community is highlighted by Zhu et al. in [69] in their analysis of the current state of deep learning-based SAR image analysis.

Only a few open PolSAR datasets labeled by land cover classes exist that are frequently used in the respective literature to evaluate approaches for land cover classification. These are:

- The Flevoland dataset, which contains a PolSAR image (1024 × 1024 pixels) acquired by the NASA/JPL AIRSAR system and ground truth data, in the form of 14 crop type classes. The ground truth was collected within the 1991 MAC-Europe campaign [72].
- The Oberpfaffenhofen E-SAR dataset[1], which contains a PolSAR image (1300 × 1200 pixels) labeled with three land cover classes (*Built-up Area*, *Wood Land* and *Open Area*).
- The PolSF dataset [73], which contains PolSAR images (1024 × 900 to 2784 × 5056 pixels) from various spaceborne systems (Gaofen-3, RADARSAT-2, RISAR, ALOS2) and an airborne system (AIRSAR) over San Francisco alongside with different versions of land cover labels.
- The AIR-PolSAR-Seg [74] dataset, which contains a PolSAR image (9082 × 9805 pixels) acquired by the Gaofen-3 system and labeled by six land cover classes (*Natural*, *Water*, *Land Use*, *Industrial*, *Housing* and *Others*).

The enumerated datasets, however, have several limitations that make them inadequate for evaluating machine learning classifiers. An inherent limitation of the Flevoland, Oberpfaffenhofen, and PolSF datasets is the absence of predefined training and test areas. This lack of standardization poses a significant challenge when attempting to compare research results across various studies. Another shortcoming is the lack of complexity of the provided classification task. For instance, the Oberpfaffenhofen dataset employs merely three easily distinguishable classes. The Flevoland dataset, on the other hand, consists of 14 distinct and inherently self-resembling classes. However, they are arranged rectangularly, which greatly simplifies the image segmentation. As a result, very high classification performance can already be achieved with simple classifiers for these datasets, making it not feasible to compare more sophisticated classifiers that are needed for tackling more challenging real-world tasks. The recently presented AIR-PolSAR-Seg [74] dataset represents an attempt to provide an open labeled large-scale PolSAR benchmark dataset with high scene complexity and thus a more challenging land cover classification task. However, a major drawback of this dataset is that only amplitude images are available, and consequently, a significant part of the polarimetric information cannot be included in the classification.

Despite their shortcomings, these datasets have made an important contribution to the development of learning-based classification of PolSAR data. However, no freely available labeled dataset can be found for the analysis of PolInSAR data, which also explains the limited research activity in this area.

To fill this gap [and to create the database for this thesis], the Pol-InSAR-Island dataset was generated, which provides multi-frequency PolInSAR data along with a challenging

[1] At the time of publishing this work, unfortunately, the dataset is no longer available through the commonly referenced link: `https://earth.esa.int/web/polsarpro/data-sources/sample-datasets`.

classification task consisting of 12 land cover classes. The difficulty of the classification task is mainly due to the focus on natural land cover classes, which are similar in appearance and continuously merge into each other. The dataset is intended to enable the training and fair comparison of different models for the classification of single- or multi-frequency PolSAR and PolInSAR data. All in all, the Pol-InSAR-Island dataset stands out from existing benchmark datasets due to the following advantages:

1. **Open provision of labeled multi-frequency PolInSAR data in high resolution**: Pol-InSAR-Island is the first open benchmark dataset for the classification of multi-frequency PolInSAR data.
2. **Challenging classification task due to high scene complexity**: The dataset comprises 12 different natural land cover classes characterized by high intra-class variance and inter-class similarity. Another challenge arises from the imbalance in the number of samples per class that is typical for real-world data.
3. **Controlled training and test setting**: In total, the dataset contains more than 5 million labeled pixels, which are equally divided into spatially disjoint training and test sets. When partitioning the dataset, care was taken to ensure similar class strengths between the training and test sets.

The following chapter is divided into three sections. First, in Section 3.2, the PolInSAR data, their properties, and their representation form are described. The process of class labeling based on reference data is outlined in Section 3.3. Finally, in Section 3.4, the advantages and limitations of the presented dataset are discussed.

3.2 PolInSAR Data

The PolInSAR data of the Pol-InSAR-Island dataset were acquired over the German Wadden Sea in April 2022 on behalf of the Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency (NLWKN). The measurement campaign was conducted with the airborne F-SAR system developed at the German Aerospace Center (DLR) [75]. The captured area covers the islands Norderney, Baltrum, and Langeoog, including the surrounding tidal flat area. The Pol-InSAR-Island dataset does not contain all the SAR data acquired during the campaign but only the data segments covering the island of Baltrum. The reason for this is the high time effort required for quality-assured labeling of the land cover. The decision to initially focus on the island of Baltrum is due to the fact that a high diversity of land cover classes is available within a comparatively small area. As shown in Figure 3.1, the selected data sections, each covering an area of 2.5 km $\times$ 3.6 km, originate from two overlapping flight paths, which were acquired in opposite flight directions. The incidence angle interval used to capture these sections ranges from 26 $^\circ$ to 58 $^\circ$. As described in [75], the F-SAR system allows the acquisition

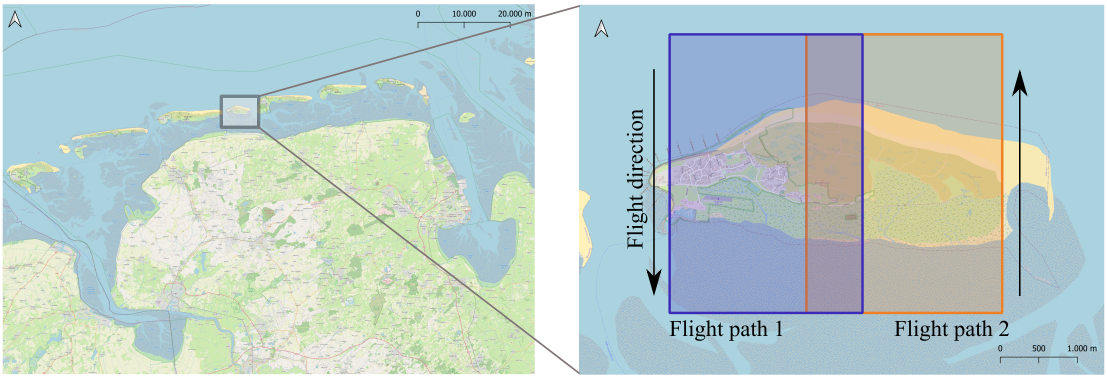


Figure 3.1.: Geographic location of the F-SAR measurement campaign 2022 and selected areas for the Pol-InSAR-Island dataset. The flight paths, shown in the right figure, were each captured twice using a vertical baseline of 40 m, a horizontal baseline of 0 m, and a time offset of 12 minutes.

of fully polarimetric SAR data ($s_{hh}, s_{hv}, s_{vh}, s_{vv}$) in five frequency bands: P, L, S, C and X. The data considered for the Pol-InSAR-Island dataset were acquired simultaneously in S- and L-band, corresponding to frequencies of 3.25 GHz and 1.325 GHz, respectively. The resulting images have a resolution of 0.65 m (S) and 1.29 m (L) in range and 0.5 m (S) and 0.6 m (L) in azimuth. The interferometric analysis is enabled by imaging in a repeat-pass configuration, where the area was imaged twice with a time offset of 12 minutes, a nominal vertical baseline of 40 m, and a horizontal baseline of 0 m. Further information about the concepts of data acquisition and processing, which have been developed within the GeoWAM project [76], can be found in [77]. Figure 3.2 shows the multi-frequency PolInSAR imagery included in the Pol-InSAR-Island dataset in the form of color-coded images based on Pauli decomposition and interferometric coherences. Based on co-registered single-look complex interferometric image pairs, the matrix desciptor $\mathbf{T}_6$ is calculated for each pixel: [...]

$$\mathbf{k}_i = \frac{1}{\sqrt{2}} \begin{bmatrix} s_{hh} + s_{vv} & s_{vv} - s_{hh} & 2s_{hv} \end{bmatrix}^T \tag{3.1}$$

$$\mathbf{T}_6 = \begin{bmatrix} \mathbf{k}_1 \\ \mathbf{k}_2 \end{bmatrix} \begin{bmatrix} \mathbf{k}_1^{*T} & \mathbf{k}_2^{*T} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{T}_{22} \end{bmatrix}. \tag{3.2}$$

Here, * denotes the complex conjugation, and ^{T} stands for the matrix transposition. [...] A postprocessing step is performed to remove the flat-earth phase from the components in $\mathbf{\Omega}_{12}$ and $\mathbf{\Omega}_{21}$. Subsequently, the resulting representation is projected from the slant range to the ground range geometry on a 1 m $\times$ 1 m grid. In the Pol-InSAR-Island dataset, the geocoded image data are provided in S- and L-band. In addition, the dataset contains auxiliary files that map the incidence angle as well as the vertical wavenumber.

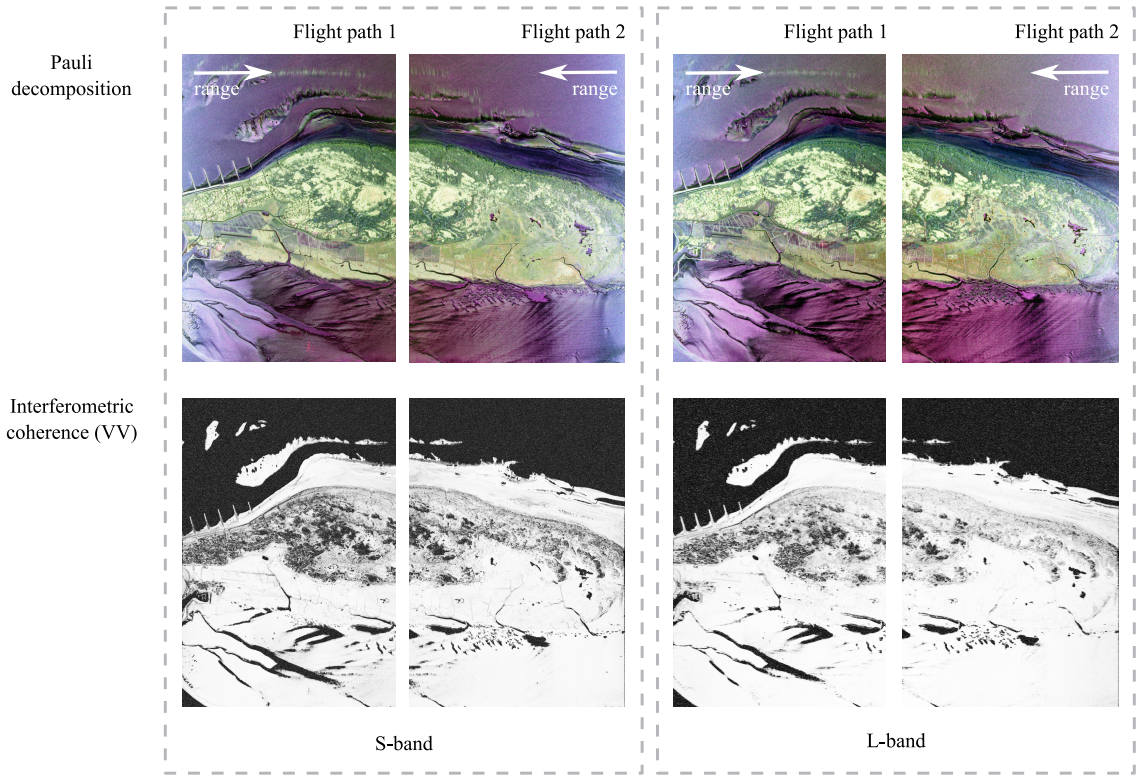


Figure 3.2.: Geocoded PolInSAR image data of the Pol-InSAR-Island dataset. The image data of both frequency bands are visualized by polarimetric composites (Pauli decomposition) and interferometric coherence in VV polarization.

3.3 Reference Data

The island of Baltrum is 6.5 km^2 in size and thus the smallest of the seven East Frisian Islands of Germany. The island faces the North Sea to the north, while to the south, there are tidal flats between the island and the mainland. These tidal flats form a unique ecosystem that is regularly flooded and drained again. The island itself is home to many different types of biotopes typical of North Sea islands. These include salt marshes, bogs, and dune valleys. In addition, part of the island is built up with settlements. Man-made coastal protection structures can be found in the northwestern coastal area. This diverse land cover results in the target classes of the demanding Pol-InSAR-Island dataset. To realize fine-grained labeling of the land cover with a high degree of semantic detail, an existing biotope type map is used, published by the Lower Saxon Wadden Sea National Park Authority (NLPV) and provided via the geodata portal Marine Dateninfrastruktur Niedersachsen (`https://mdi.niedersachsen.de/`). This biotope type map was generated in 2013 as part of the Trilateral Monitoring and Assessment Program (TMAP), which aims to provide a substantial basis for policy-making on the protection of the Wadden Sea ecosystem. In the map, the island of Baltrum is divided into 40 biotope types. The biotope type map is visualized in Figure 3.3. Since the biotope type map was generated several years before the F-SAR measurement campaign, significant changes in the extent of the

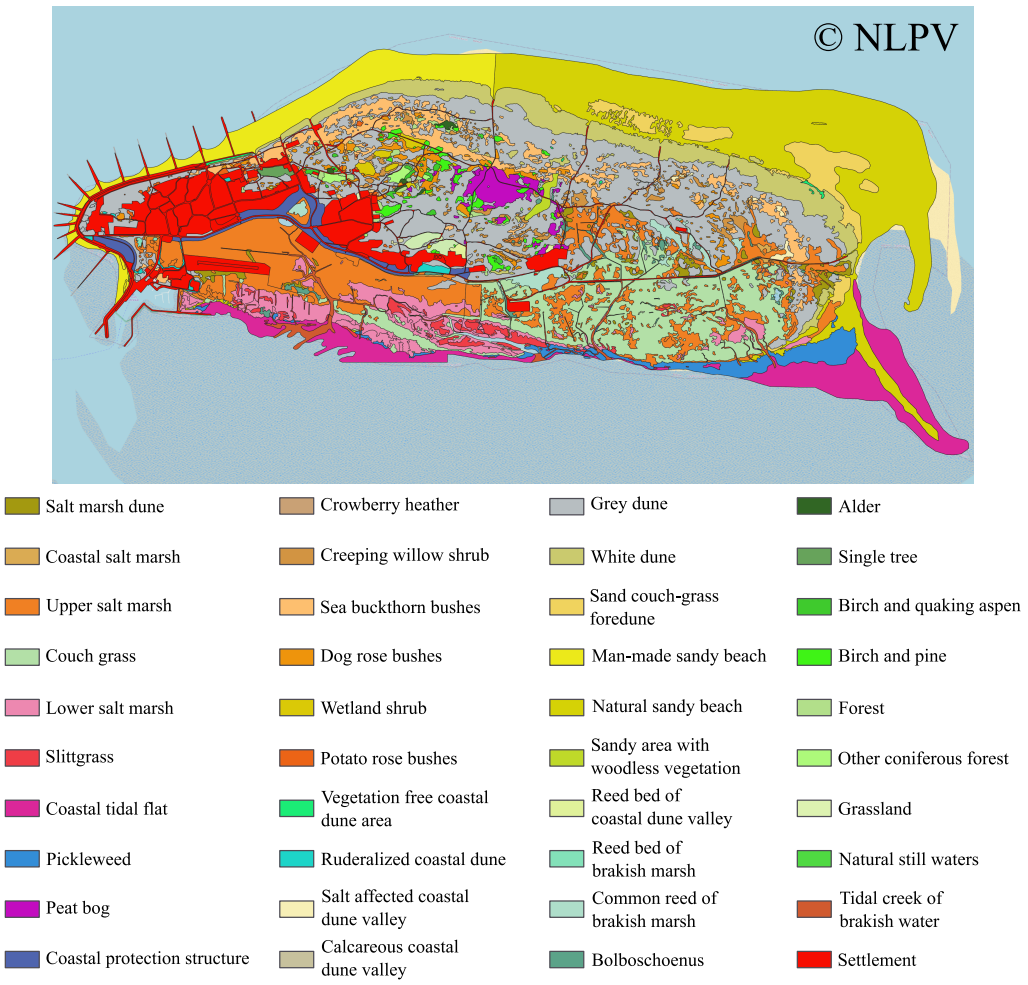


Figure 3.3.: Biotope type map (2013) of the island of Baltrum provided by the NLPV.

different biotopes are to be expected, making the map not sufficient for accurate labeling. Therefore, the map is revised within a semi-automatic process, which is schematically illustrated in Figure 3.4.

3.3.1 Revised Biotope Type Map

The biotope type map used as starting point is given as vector file, where each polygon contains exactly one biotope type attribute. The first step of the revision buffering & filtering aims at masking out areas where the class assignment of the outdated biotope type map is highly unlikely to match the land cover at the time of the SAR acquisition. It is assumed that this is particularly the case at transition areas of adjacent biotope boundaries. Moreover, the boundaries between two biotopes are usually fuzzy by nature, so that an unambiguous class assignment is not possible. Therefore, all polygons of the biotope type map are adjusted with a negative buffer operation to shrink their extent and to create an empty space without an assigned biotope type between two different biotope types. Subsequently, biotope types that collectively cover less than 1000 m^2 over the complete

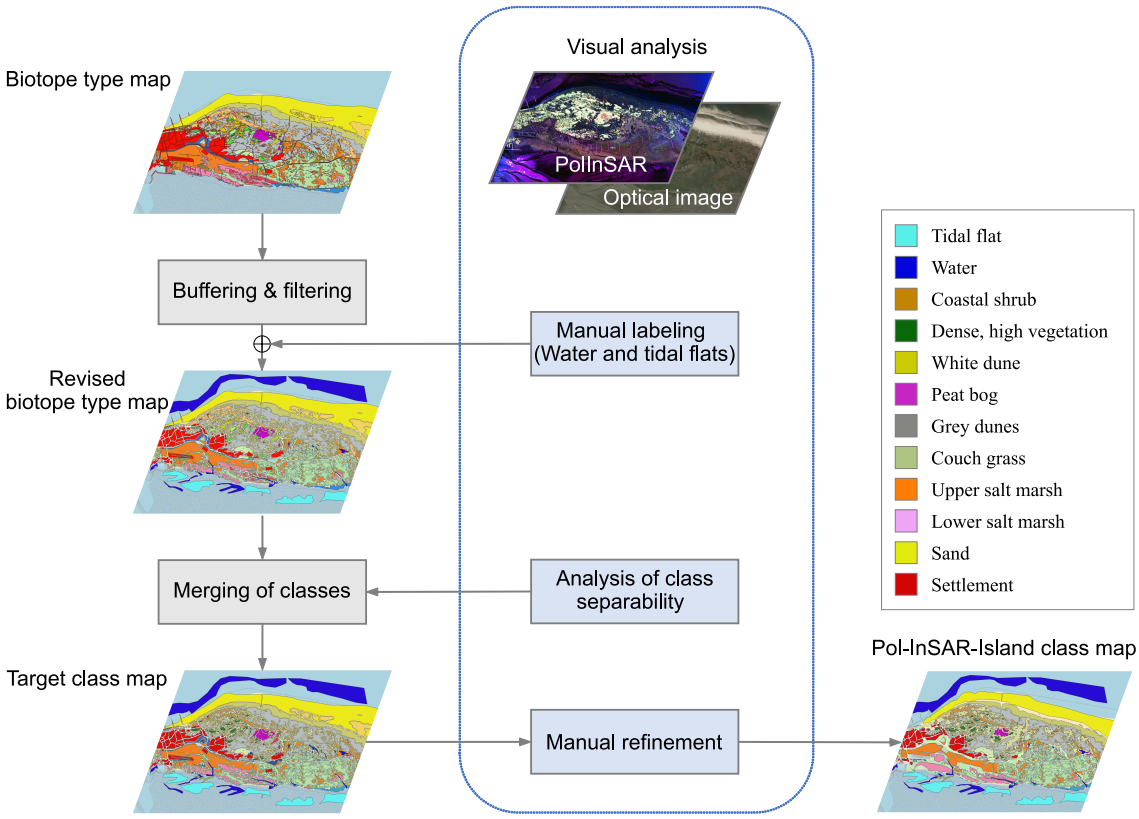


Figure 3.4.: Workflow of the semi-automatic process for data labeling.

study area are excluded from the dataset. Furthermore, areas that are assigned to the biotope type *Coastal tidal flat*, *Pickleweed*, or *Tidal creek of brackish water* are masked out. The reason for this approach is that the extent of exposed tidal flats is primarily determined by the current tidal phase, i.e., the current water level. Thus, the water level at the time of the SAR data acquisition can only be identified through an interpretation of the PolInSAR data itself. Therefore, a visual analysis of SAR backscatter amplitudes and interferometric coherences is performed to manually identify tidal flat areas and water areas that are added to the map. The resulting map is called the revised biotope type map.

3.3.2 Target Class Map

In the following processing step merging of classes, the original semantically fine-grained biotope types are grouped and assigned to generalized land cover classes, which form the final target classes of the benchmark dataset. The requirement for these target classes is that class separation using the multi-frequency PolInSAR data is possible but not trivial, so the dataset provides a challenging classification task. To determine an appropriate grouping of biotope types, a multi-stage visual class separability analysis is performed. This analysis is based on a visualization of the data structure of multi-frequency PolInSAR features. For this purpose, a 54-dimensional feature vector is generated for each pixel by stacking

Table 3.1.: List of considered polarimetric and interferometric features.

Expression	Feature description
$\langle t_{ii} \rangle$	Diagonal elements of $\langle \mathbf{T}_{11} \rangle$ (in dB)
$\lvert\langle t_{ij} \rangle\rvert$	Magnitude of off-diagonal elements of $\langle \mathbf{T}_{11} \rangle$
$\arg(\langle t_{ij} \rangle)$	Phase of off-diagonal elements of $\langle \mathbf{T}_{11} \rangle$
H, A, α	Entropy, anisotropy and mean alpha angle derived by eigenvalue decomposition [52]
p_1, p_2, p_3	Normalized eigenvalues of $\langle \mathbf{T}_{11} \rangle$
$\underline{\lambda} = \sum_{i=1}^{3} p_i \lambda_i$	Weighted mean of eigenvalues
$P_{s,Y}, P_{d,Y}, P_{v,Y}$	Surface scattering, double-bounce scattering and volume scattering power by Yamaguchi decomposition [54]
μ	Conformity coefficient defined in [56]
$\hat{H}_p$ $\hat{H}_d$ d_p d_{idx}	Scattering predominance, scattering diversity, degree of purity, and depolarization index [57]
$\lvert\gamma_{xx}\rvert$	Magnitude of interferometric coherence in VV, HV and HH polarization

polarimetric and interferometric features listed in Table 3.1, that are extracted from S- and L-band data. The polarimetric features are computed based on the coherency matrix $\langle \mathbf{T}_{11} \rangle$ of the reference scene [estimated through spatial averaging within a 7×7 window]. To visualize the high-dimensional feature representation of the data, an embedding in a 2-dimensional Euclidean space is computed using the dimension reduction method UMAP [26]. [A more in-depth exploration of this visualization method is carried out in Chapter 4 of this thesis.] The UMAP algorithm aims at preserving local and global neighborhood relations of the data, which in the present application describe the similarity of the selected PolInSAR features. A resulting embedding of the multi-frequency PolInSAR data is shown in Figure 3.5. Here, data points are colored according to their membership in one of six generic classes *Tidal flat*, *Water*, *Dunes landscape*, *Salt marshes*, *Sand* and *Settlement*. This class division was defined manually by grouping the specific biotope types based on their semantic meaning. The clustering observed in the scatterplot suggests a good separability of the six generic classes. While the classes *Tidal flat*, *Water*, *Sand* and *Settlement* represent only one biotope type each, the classes *Dunes landscape* and *Salt marshes* are composed of many different biotope types, whose separability is further investigated. For this purpose, two additional embeddings are computed, in which only the pixels of the classes *Salt marshes* and *Dunes landscape* are included. The resulting

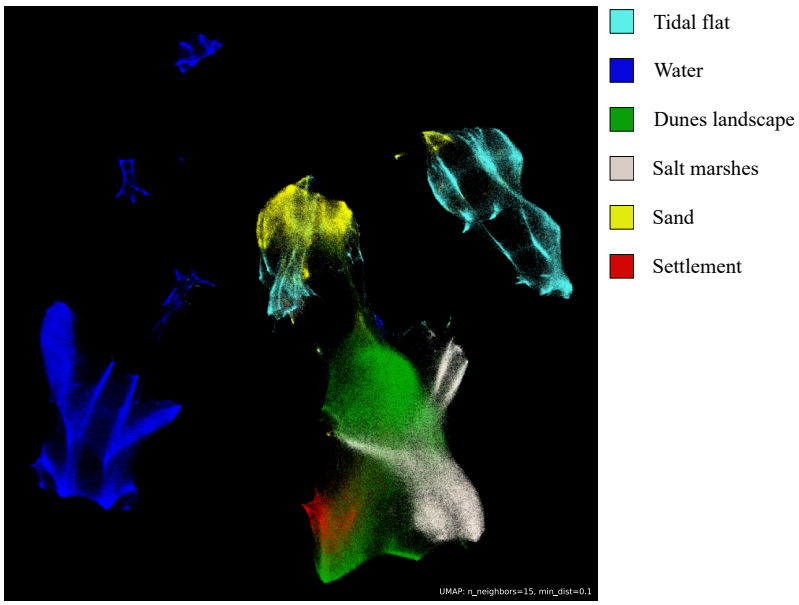


Figure 3.5.: UMAP projection of PolInSAR features. Projected data points are colored according to six generic land cover classes.

embeddings are shown in Figure 3.6. In the scatterplot, data points are colored based on their class as specified by the revised biotope type map. Note that reapplying the nondeterministic UMAP algorithm to the corresponding data subsets results in slight variations in the arrangement of data points in the 2-dimensional feature space. From the scatterplot corresponding to the class *Salt marshes*, a separability of the biotope types *Couch grass* (light green) and *Upper salt marsh* (orange) can be derived. Another cluster is formed by the data points of the biotope types *Slit grass* and *Upper salt marsh*, which overlap strongly and are therefore combined into one target class *Upper salt marsh*. The generic class *Dunes landscape* is composed of 14 biotope types whose similarities with respect to their corresponding PolInSAR features are shown in the lower part of Figure 3.6. The three biotope types *White dune* (yellow), *Peat bog* (violet), and *Grey dune* (grey) stand out as distinct clusters. The remaining biotope types have similar PolInSAR features and, therefore, partially overlap in the 2-dimensional feature space. However, a separation between biotope types that can be assigned to the class *Dense, high vegetation* and biotope types that can be assigned to the class *Coastal shrub* is possible. The presented analysis results in 12 target classes for the Pol-InSAR-Island dataset: *Tidal flat*, *Water*, *Coastal shrub*, *Dense and high vegetation*, *White dune*, *Peat bog*, *Grey dune*, *Couch grass*, *Upper salt marsh*, *Upper salt marsh*, *Sand* and *Settlement*. In Figure 3.7, the feature representation of the entire dataset is shown, with the data points colored based on these 12 target classes. This figure gives a first impression regarding the degree of difficulty of class separation. A brief description of each target class is given below.

- *Tidal flats* (*TF*) are sandy, muddy areas that are repeatedly flooded and drained. They lie between the island and the mainland, and tideways are present during low tide

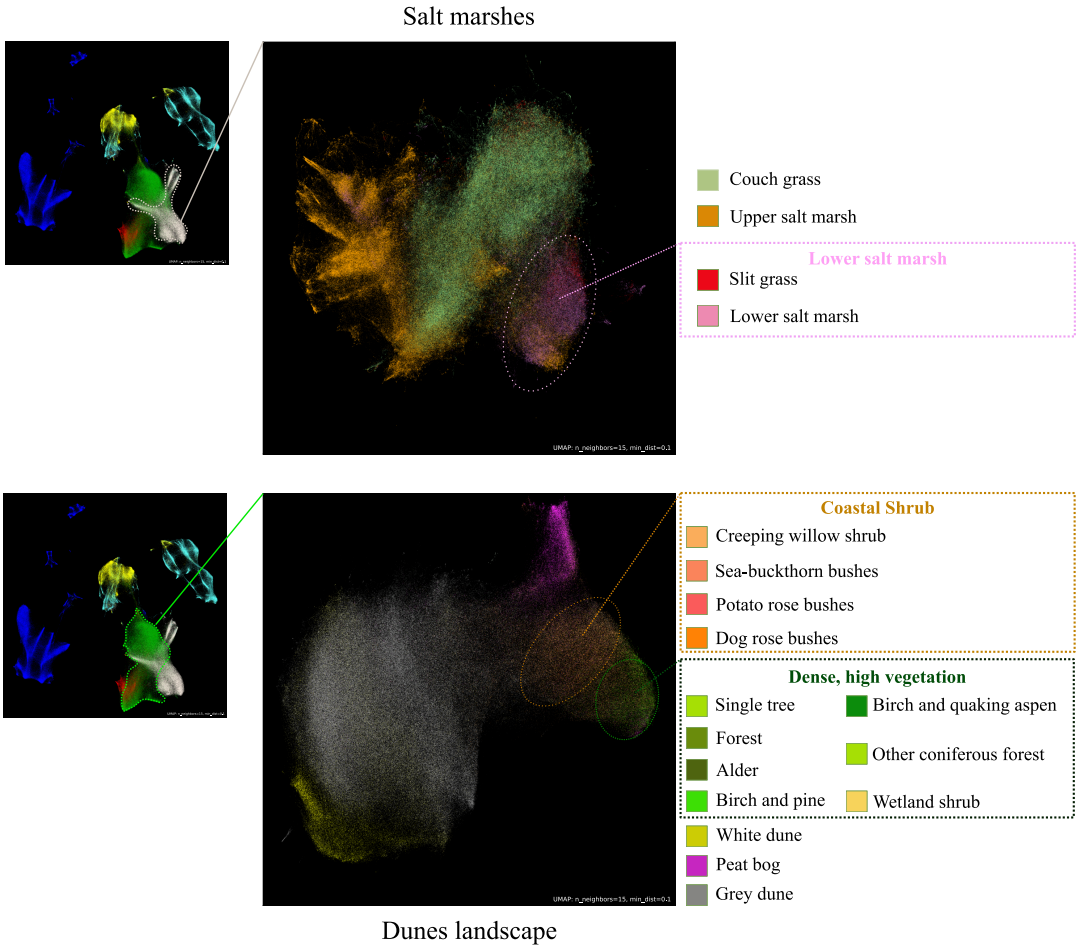


Figure 3.6.: Separability analysis of different biotope types, which belong to the generic class *Salt marshes* (top) and the generic class *Dunes landscape* (bottom), respectively.

periods. As they are regularly flooded, they are characterized by high moisture and little vegetation.

- *Water* (*W*) represents the North Sea, tideways running through the tidal flats, as well as waterways and small lakes located on the island.
- *Coastal shrub* (*CS*) represents low-growing shrubbery, such as sea-buckthorn, creeping willow, and potato rose bushes.
- *Dense, High vegetation* (*HV*) [referred to only as *High vegetation* in the remainder of the thesis] contains all the biotope types covering forested areas, as well as dense and high shrubbery.
- *White dune* (*WD*) represents areas with multiple meter-high dunes located next to the beach. They are covered by a more or less open low-growing vegetation cover of beach grass.

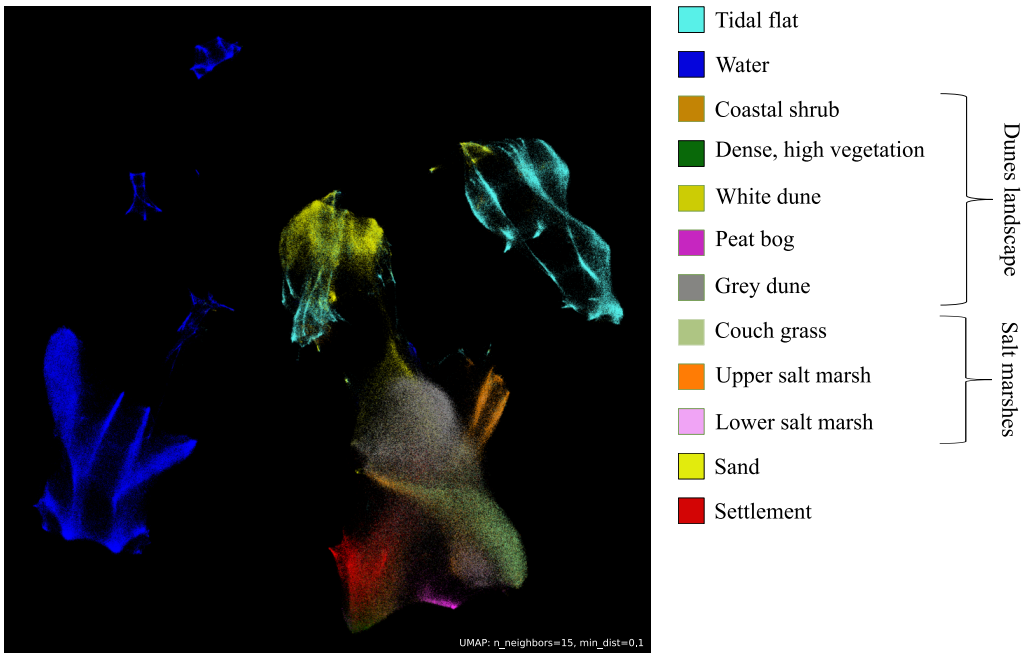
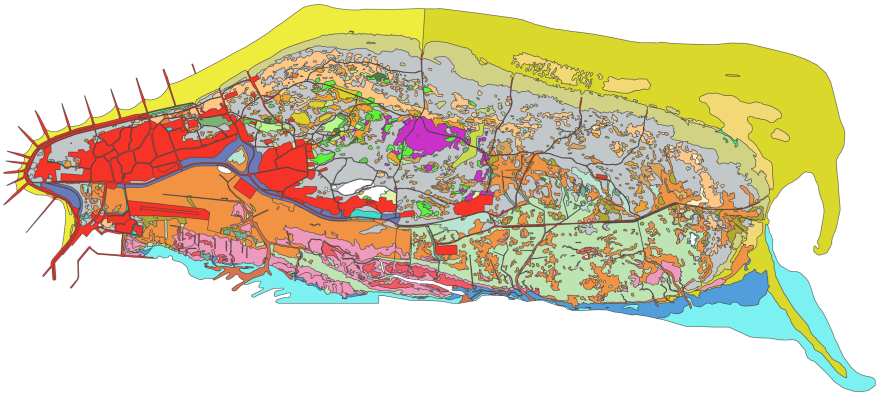
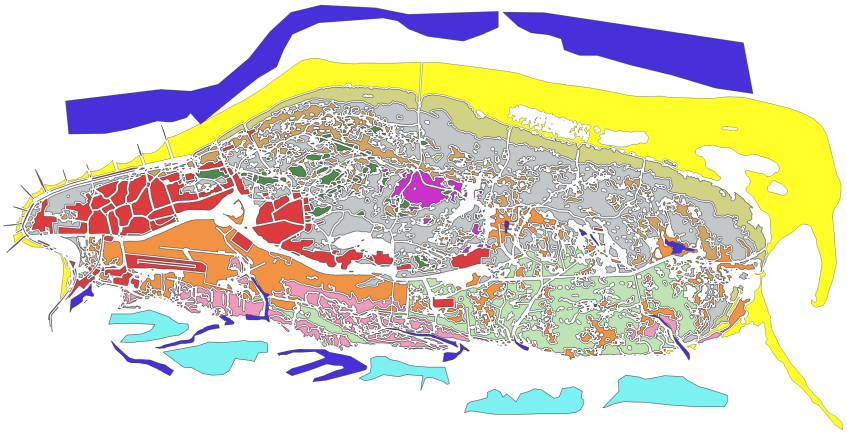


Figure 3.7.: UMAP projection of PolInSAR features. Projected data points are colored according to the 12 target classes of the Pol-InSAR-Island dataset.

- *Peat bog* (*PB*) is the rarest class in the dataset and occurs as one large area in the center of the island Baltrum. It is a moist to swampy area covered by vegetation such as floodplains or creeping willow bushes.
- *Grey dune* (*GD*) represents a dune landscape, which is covered with fireweed, thistles, or rubus.
- *Couch grass* (*CG*) identifies fallow land of a formerly pastured upper salt marsh with couch grass dominance.
- *Upper salt marsh* (*US*) extends to about 70 cm above the high tide line and is flooded only 25-70 times per year. It is home to numerous plant species, including salt marsh red fescue, milkweed, beach mugwort, and beach quackgrass.
- *Lower salt marsh* (*LS*) extends to about 30 cm above the high tide line and is flooded 100-300 times per year. Prevalent species are andelgrass, beach aster, beach lilac, beach trident, and purslane wedgewort.
- *Sand* (*S*) covers the beach areas, as well as sandy areas without vegetation near dunes. *Sand* is mostly found in the western and northern areas of the island, facing the seaside.
- *Settlement* (*SE*) represents the major settlement area in the west of the island as well as isolated buildings and settlements. No individual houses are labeled, but rather the whole area of the settlement, containing adjacent trees and surrounding meadows.



(a) Original biotope type map (40 classes).



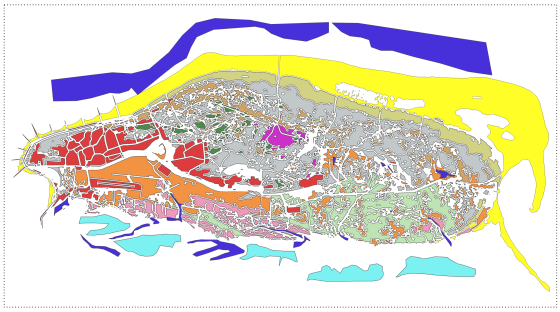
(b) Target class map (12 classes).

Figure 3.8.: Land cover maps for the island Baltrum, where (a) shows the original biotope type map and (b) shows the target class map, which results after manual labeling, buffering & filtering, and merging of classes. Corresponding legends can be found in Figure 3.3 and Figure 3.7.

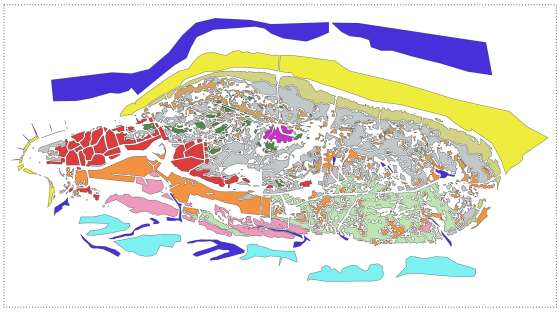
The resulting target class map, juxtaposed with the original biotope type map, is shown in Figure 3.8.

3.3.3 Pol-InSAR-Island Class Map

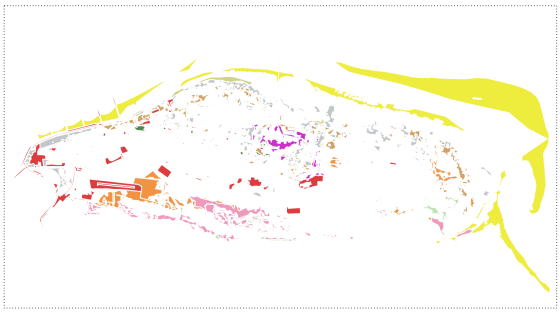
Finally, the target class map is manually refined to optimally align the class labels based on the outdated biotope type map (2013) to reflect the current state of the more recent F-SAR acquisition (2022). The manual refinement is realized by visual interpretation of PolInSAR feature images and recent optical data. The optical data consulted for this purpose were acquired in an airborne survey one day after the F-SAR acquisition and have a ground resolution of 1 cm. Performing the highly time-consuming manual revision and validation



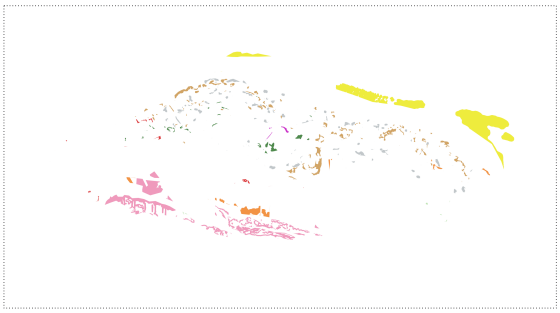
(a) Target class map.



(b) Pol-InSAR-Island class map.



(c) Removed/changed areas of the target class map.



(d) Added/changed areas of the Pol-InSAR-Island class map.

(e) Legend.

Figure 3.9.: Changes made to the target class map during manual refinement to obtain the final Pol-InSAR-Island class map. (a) shows the target class map, which serves as a starting point for the manual refinement. (b) shows the resulting Pol-InSAR-Island class map. (c) highlights areas that are removed from the map or assigned to a different class. These areas are colored according to their labels in the target class map. (d) highlights areas that are added to the map or assigned to a different class. These areas are colored according to their labels in the Pol-InSAR-Island class map.

of the labels is indispensable due to the large time difference between the generation of the biotope type map and the F-SAR acquisition. Figure 3.9 shows the changes made to the target class map in this step to obtain the final, quality-assured Pol-InSAR-Island class map. For this purpose, both class maps are shown side by side in Figures 3.9a and 3.9b. In addition, Figures 3.9c and 3.9d highlight areas that have either been removed from the map, added, or replaced with another class. As expected, areas of the class *Sand*, which are highly exposed to sea disturbances, have undergone the most significant changes. A further notable aspect is the removal of several areas labeled as *Settlement* (red) in the target class map. This is due to the fact that in the Pol-InSAR-Island class map, only densely developed settlement areas were included, and open areas with very sparse built-up areas were not taken into account. In the area of *Salt marshes*, large-scale changes occur locally, as the boundary between *Upper* and *Lower salt marsh* has changed over the years. For the remaining classes, primarily small-scale changes are evident. For example, the extent of *Coastal shrub* on the *Grey dune* has changed to a certain degree, and the area of the class

Peat bog visible in the PolInSAR data has decreased in size due to increased vegetation cover. After the manual revision and validation of the labels, the final Pol-InSAR-Island class map is provided as part of the open benchmark dataset.

3.3.4 Training and Test Data Split

The Pol-InSAR-Island dataset is divided into two spatially disjoint subsets for training and testing and thus for a fair comparison of different classification approaches. A chessboard grid is used for the partitioning, which alternately defines rectangular training and test regions with an extent of 512 m × 512 m each. This ensures that classes that are spatially concentrated on the island are adequately represented in both the training and the test set. In addition, this method of data splitting has the advantage that training and test samples are taken from the entire incidence angle interval ranging from 26 ° to 58 °. This is relevant because the incidence angle strongly influences the polarimetric backscatter signal, especially for surface scatterers such as *Sand*, *Water* and *Tidal flat* [78]. [A disadvantage of the splitting method, however, is that the training and test areas are directly adjacent to each other and thus partially resemble each other. Nonetheless, this method offers the best trade-off between sensible class coverage and heterogeneity of the training data on the one hand and correlation of training and test data on the other hand.] It should be noted that the choice of spatially overlapping flight paths (see Figure 3.1) results in an area that is doubly covered in the dataset. However, since the intersecting area was imaged from opposite flight directions, the SAR signatures vary between the two image sections. For this reason, both image sections of the intersecting area are retained in the dataset. In total, the dataset yields 2,677,248 labeled pixels for training and 2,773,559 for testing, distributed over the 12 target classes as listed in Table 3.2. As is typical for real-world data, the samples are not evenly distributed across the classes. The class imbalance complicates the task of land cover classification and needs to be addressed in the development of classifiers.

Table 3.2.: Percentage of labeled pixels per class contained in the training and the test set.

	TF	W	CS	HV	WD	PB	GD	CG	US	LS	S	SE
Training	7.09	17.98	5.29	2.04	6.72	0.87	17.78	8.17	8.01	5.85	14.92	5.28
Test	8.07	17.31	3.65	1.43	6.44	1.15	19.88	7.56	9.51	4.75	14.62	5.63

3.4 Discussion

The presented Pol-InSAR-Island dataset is the first open benchmark dataset for land cover classification of multi-frequency PolInSAR data, which [serves as a database in this thesis

and is further] intended to advance the development of machine learning approaches in this field. The novelty lies in the public availability of such information-rich, high-resolution data products, as well as the corresponding well-designed class labels, which pose a challenging land cover classification task. [...] The Pol-InSAR-Island benchmark dataset provides reliable training and test data that aims to encourage and enable researchers to develop and compare different machine learning approaches. The fixed definition of the training and test regions given in the dataset [...] ensures direct comparability of different approaches.

A limitation of the Pol-InSAR-Island dataset is the fragmentary labeling. In general, labeled areas of different target classes are not directly adjacent to each other but are separated by an unlabeled area. By employing machine learning methods, there is a risk of inaccurately detecting class boundaries and facing challenges in evaluating classifiers in this regard. Seamless labeling would, therefore, be desirable but is difficult or impossible to achieve with the available reference data. A further issue is the limited size of the dataset. The Pol-InSAR-Island dataset contains image data covering an area of about 15 km^2 with a ground resolution of 1 m $\times$ 1 m and more than five million labeled pixels in total. This exceeds the size of most existing datasets for PolSAR data classification but does not yet reach the size of datasets used in computer vision to train large deep learning models. In the future, it is conceivable to extend the dataset by incorporating additional data acquired during the F-SAR measurement campaign over the neighboring islands of Norderney and Langeoog. However, this requires a considerable time effort due to the manual improvement and validation of automatically generated labels, which is indispensable because of the time difference between the reference biotope type map and the SAR data acquisition. [...] Expanding the dataset with image data acquired from other SAR sensors (e.g., TerraSAR-X, RADARSAT-2, or Sentinel-1) also presents challenges. One major issue is the difficulty in directly matching the fine-scaled labels to lower spatial resolution data. Additionally, reliable adjustment and validation of labels become infeasible without additional (optical) reference data that have to be captured near the time of the SAR data acquisition. The limitation to a single sensor and a single location represents a weak point of the Pol-InSAR-Island dataset. As a result, classifiers trained on this dataset are unlikely to be transferable to other sensors and test sites without further adaptation. Thus, the main goal of providing this dataset [and its use in this thesis] is not to train universally applicable classifiers for practical use but rather to enable the development and comparison of novel multi-frequency PolInSAR data classification methods in a controlled training and test environment.

Convenient access to the data is provided via KITopenData: https://doi.org/10.35097/1700.

4 Visualizing Multi-frequency PolInSAR Data

Even if not quoted verbatim, the following preliminary publications are referenced, as they address the methodological concepts of UMAP-based visualization for PolInSAR data.

The Feature-UMAP method based on single-frequency PolInSAR data is introduced in:

[79] Sylvia Schmitz, Uwe Weidner, Horst Hammer, and Antje Thiele. "Evaluating Uniform Manifold Approximation and Projection for dimension reduction and visualization of PolInSAR features". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. V-1-2021. 2021, pp. 39–46.

The extension of the Feature-UMAP method to multi-frequency PolInSAR data is presented in:

[80] Sylvia Schmitz, Horst Hammer, and Antje Thiele. "Multi-frequency PolInSAR data are advantageous for land cover classification – A visual and quantitative analysis". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. V-1-2022. 2022, pp. 49–56.

The concept of the Wishart-UMAP method is published in:

[81] Sylvia Hochstuhl, Horst Hammer, Antje Thiele, and Stefan Hinz. "Wishart-UMAP-Translating pixel similarity of PolSAR images to real-valued feature descriptors". In: *2023 IEEE International Geoscience and Remote Sensing Symposium*. 2023, pp. 5966–5969.

4.1 Introduction

The unique strength of multi-frequency PolInSAR data lies in its rich information content, enabling the derivation of geophysical parameters, vegetation structures, and terrain characteristics of an observed scene. Despite the advancements in automated data analysis techniques for land cover classification and parameter retrieval, there remains an essential

need for human-driven visual interpretation of PolInSAR data. Human analysts play a vital role as they bring their domain and contextual knowledge to the analysis and are able to detect complex, large-scale patterns and anomalies that automated algorithms might miss.

Although effective visualization is essential to enable human-driven analysis, which supports automated analyses, existing visualization techniques do not adequately represent the comprehensive information contained in complex-valued multi-dimensional PolInSAR data. Commonly, visualization is achieved by creating color composites based on surface, double-bounce, and volume scattering derived from target decomposition techniques. Although these methods offer some insights, they are inherently limited in the number of features they can display simultaneously, leading to an incomplete representation of the full capabilities of the data.

In response to this challenge, this thesis proposes the use of UMAP [26], a dimension reduction technique that has shown promise for the visualization of multi-dimensional data in various application domains, including mass spectrometry imaging [27], single-cell analysis [28], and remote sensing [29, 30]. The key advantages that render UMAP a suitable candidate for PolInSAR data visualization include its capability to preserve local and global data structures and its scalability to large datasets due to computational efficiency. Additionally, its ability to incorporate new data into an existing embedding allows consistent analysis across multi-temporal or multi-location datasets.

This work has a dual objective: firstly, to evaluate the effectiveness of UMAP in preserving the intrinsic data structure and class separability of PolInSAR data compared to other dimension reduction techniques; secondly, to explore two distinct approaches for applying UMAP to PolInSAR visualization: Feature-UMAP (F-UMAP) and Wishart-UMAP (W-UMAP). The F-UMAP approach creates visualizations based on a wide range of expert-designed features incorporating domain-specific knowledge. In contrast, W-UMAP directly explores statistical similarities of complex-valued coherency matrices, aiming to capture a more holistic view of the data without requiring prior feature selection and extraction.

The primary aim of using UMAP-based visualization is to create comprehensive and easy-to-interpret representations that reflect the plentiful information present in PolInSAR data. This will enhance the visual analysis capabilities of PolInSAR data, thus unlocking new possibilities for its application and interpretation.

4.2 Related Work

To visualize Pol(In)SAR data in a human-interpretable manner, the data are commonly represented by RGB color composites. Commonly, the red, green and blue intensity channels are used to reflect elementary scattering mechanisms. For this purpose, the contributions of

surface, double-bounce, and volume scattering are calculated using target decomposition methods, including Pauli decomposition [1], eigenvalue decomposition [52] or Yamaguchi decomposition [54]. As shown in [82–84], additionally considering interferometric coherences improves target decomposition, which in turn allows a more detailed visual differentiation of land cover classes. Alternatively, visualization can also employ different types of features. For example, the three eigenvalues from the coherency matrix can be used to highlight scattering diversity within a resolution cell as proposed in [85]. Considering multi-frequency data, backscattering amplitudes measured across various frequency bands can be combined to visualize the frequency-dependent variations in backscattering [86]. In addition to visualizing polarimetric information, colorization based on feature selection is also proposed for the human-centered analysis of multi-temporal SAR data. To allow visual analysis of changes caused by seasonal climate variation, Amitrano et al. [87] propose to combine the cross-normalized amplitude of a reference image (dry season) and a test image with its interferometric coherence image. Similarly, [88] generates a color composite from two cross-normalized amplitude images (assigned to blue and green channels) and the difference between the two images (assigned to the red channel), which is used in the context of flood monitoring.

A major disadvantage of visualizing multidimensional SAR data using feature selection is that the information contained in unselected features is completely ignored. Thus, if the scattering mechanism contributions are exclusively used, any information about the temporal stability of scatterers and the topography of the observed scene contained in the interferometric measurements, as well as information resulting from the complementary multi-frequency measurement, is missed. In view of these considerations, dimension reduction by feature transformation may be more effective than feature selection. It allows for the preservation of relevant information from all available features without fully discarding several potentially valuable features, which can lead to richer and more informative visualizations. However, in the literature, the visualization of PolSAR or PolInSAR data using feature transformation for dimension reduction has not yet been addressed. Nevertheless, there are related studies in neighboring fields that offer valuable perspectives. These related works can be divided into two categories:

1. Research studies that focus on visualizing multi-dimensional image data but are not related to the SAR domain and,
2. Research studies that apply dimension reduction by feature transformation on PolSAR data, but not aiming at visualization.

Studies of both categories are presented and discussed below.

4.2.1 Visualizing Multi-dimensional Data

The challenge of effectively presenting multi-dimensional data in a way that is comprehensible to human users for visual pattern recognition and data interpretation is common across various application areas. These include, for instance, multi-dimensional remote sensing data analysis [89–103], bioinformatics [28, 104–111], medical data evaluation [112], time series analysis [113] or text categorization [114, 115].

Typically, the solution involves the application of dimension reduction methods, which aim to reduce the number of features under consideration while preserving as much information as possible. Reducing the feature dimension can be realized either by selecting a discriminative subset of features (feature selection) or by projecting the features into a lower-dimensional space (feature transformation). In the following, only dimension reduction methods based on the latter approach are considered. Numerous methods following this approach have been developed, which can generally be categorized into linear and non-linear methods, as well as global and local methods. Linear dimension reduction methods assume that the data lies on a linear subspace of the original high-dimensional space. These methods transform the data using linear combinations of the original features. Non-linear dimension reduction methods, on the other hand, can capture more complex structures in the data by applying non-linear transformations. Global methods consider the entire dataset to find a lower-dimensional representation. They aim to preserve the overall structure and relationships between all data points. Local methods focus on preserving the structure of small neighborhoods within the data. They prioritize maintaining the relationships between nearby points, making them more effective in capturing local patterns and clusters. An overview of some existing methods, categorized by linear-global, linear-local, non-linear-global, and non-linear-local, is presented in Table 4.1. Another important characterization of dimension reduction algorithms indicated in Table 4.1 is their ability to handle out-of-sample data, which refers to the ability to project new, unseen data points into the reduced-dimensional space without recomputing the entire transformation.

Several surveys of dimension reduction methods exist in the literature, providing detailed comparisons and evaluations [142–145]. Among the most extensive works is the survey conducted by Espadoto et al. [145]. The authors evaluated a broad range of dimension reduction methods across various datasets using metrics that measure neighborhood preservation, class separability, and point-pairwise distances. In this comparison Projection by Clustering (PBC) [114], Interactive Document Maps (IDMAP) [115], t-distributed Stochastic Neighbor Embedding (t-SNE) [139] and UMAP [26] acvhieve the best results.

Visualizing Data Structure. The methods t-SNE and UMAP are frequently utilized to visualize the data structure of high-dimensional datasets within 2-dimensional scatter plots, which are suitable for straightforward visual cluster analysis. These methods have been widely adopted in single-cell analysis [28, 104–107, 109–111] and for the evaluation

Table 4.1.: Dimension reduction methods categorized by linearity and locality and out-of-sample capabilities.

	Method	Out-of-sample	Reference
Linear			
global	Principal Component Analysis (PCA)	✓	[116]
	Factor analysis (FA)	✓	[117]
	Independent Component Analysis (ICA)	✓	[118]
	Nonnegative Matrix Factorization (NMF)	✓	[119]
	Locality Preserving Projection (LPP)	✓	[120]
local	Large-Margin Nearest Neighbor (LMNN)	✗	[121]
	Linear Local Tangent Space Alignment (LLTSA)	✗	[122]
Non-linear			
global	Sammon Mapping	✗	[123]
	Multi Dimensional Scaling (MDS)	✗	[124]
	Self-Organizing-Maps (SOM)	✓	[125]
	Kernel PCA (KPCA)	✓	[126]
	Curvilinear Component Analysis (CCA)	✗	[127]
	Generalized Discriminant Analysis (GDA)	✗	[128]
	ISOMAP	✗	[129]
	Gaussian Process Latent Variable Model (GPLVM)	✗	[130]
	Random Projection (RP)	✓	[131]
	Diffusion Maps	✗	[132]
	Variational Autoencoder (VAE)	✓	[133]
local	Locally Linear Embedding (LLE)	✗	[134]
	Manifold Charting	✓	[135]
	Lapalcian Eigenmaps (LE)	✗	[136]
	Local Tangent Space Alignment (LTSA)	✗	[137]
	Maximally Collapsing Metric Learning (MCML)	✗	[138]
	Projection by Clustering (PBC)	✗	[114]
	Interactive Document Maps (IDMAP)	✗	[115]
	t-distributed Stochastic Neighbor Embedding (t-SNE)	✗	[139]
	Least Square Projection (LSP)	✗	[140]
	LargeVis	✗	[141]
	Uniform Manifold Approximation and Projection (UMAP)	✓	[26]

of features extracted by deep learning methods [146, 147]. Their application extends to remote sensing, where they have been used to visualize spatiotemporal hydrological data [89], support image labeling for machine learning training [91], and interpret spectral signatures of hyperspectral image data [90]. Studies comparing the performance of t-SNE and UMAP [26, 28, 110, 148, 149] show that both perform equally well at visualizing meaningful clusters in the data, however, UMAP preserves global neighborhood structures more effectively, is more computationally efficient, offers better reproducibility and can handle out-of-sample data.

Visualizing Image Data. In addition to the visualization of the data structure in a 2-dimensional space using scatter plots, dimension reduction methods are used to project pixels of multi-dimensional image data into a 3-dimensional space to create color composites. Notably, the approach by Smets et al. [27], which successfully employs UMAP to create color composites of mass spectrometry images, closely aligns with the methodology proposed in this thesis. Furthermore, in the remote sensing field, specifically within the SAR domain, Convolutional Autoencoder (CAE) have been utilized to reduce the temporal dimensions of SAR image time series to generate color composites. This approach aids in visualizing dynamic environmental changes, such as glacier movements and variations in snow and vegetation cover [150–152].

While the use of dimension reduction for visualizing SAR data is relatively unexplored, it has been extensively applied in the hyperspectral imaging domain. Early works apply the Principal Component Analysis (PCA) to map numerous feature bands to the RGB, HSV (Hue, Saturation, Value), or CIEL*a*b [153] color space [92–96]. To address the limitations of the linear PCA transformation and capture the non-linear relationships inherent in hyperspectral data, more sophisticated non-linear dimension reduction methods have been employed, including ISOMAP [97], Diffusion maps [98], Self-Organizing-Mapss (SOMs) [99, 100], Multi Dimensional Scaling (MDS) [101], Stochastic Proximity Embedding [102], and Locality Preserving Projection (LPP) [103].

Despite their effectiveness, these methods face a common challenge: they inherently map data points to an unbounded 3-dimensional Euclidean space, leading to a sparse distribution of points when linearly scaled to fit a bounded color space, thus degrading visualization quality. To mitigate this, various strategies have been developed. Cui et al. [154] implemented a two-step approach, initially using PCA for data projection, followed by convex optimization to fine-tune color distances within the HSV space. Additional innovative strategies aim to balance the preservation of pairwise distances with enhanced visual contrast [101]. Moreover, Liao et al. introduced a sophisticated framework that merges hyperspectral images with high-resolution color images through manifold alignment, enabling the transfer of natural color and spatial information to produce color images that are both natural-looking and easier to interpret [103, 155].

Summary. Dimension reduction techniques have proven highly effective in visualizing complex data structures across various disciplines, yet their potential in PolInSAR data visualization remains largely unexplored. UMAP, in particular, has demonstrated its effectiveness in analyzing data structures via 2-dimensional scatter plots in areas such as bioinformatics and hyperspectral imaging. This success highlights its ability to enable the visual analysis of complex data and is thus assumed to be well suited for visualizing multi-frequency PolInSAR data. Moreover, the generation of color composites through dimension reduction by feature transformation, a well-studied field in hyperspectral imaging, has not yet been explored in the PolInSAR or PolSAR domain. Given that dimension reduction enables the creation of interpretable color composites in hyperspectral imaging, adapting these methods for multi-frequency PolInSAR data could unlock new insights. It can help to deepen the understanding of the interplay of numerous features resulting from the combination of frequencies, polarimetry, and interferometry and its influence on possible target tasks, such as land cover classification.

The identified research gaps, combined with insights from reviewed studies, highlight a significant opportunity to apply and customize UMAP for multi-frequency PolInSAR data to visualize underlying structures and generate color composites for straightforward scene interpretation.

4.2.2 Dimension Reduction of Polarimetric Feature Space

Methodologically related to this work are studies that apply dimension reduction techniques to PolSAR data to enhance automatic classification methods with respect to computational demand and accuracy. Although the primary focus of these studies is not on visualization, the methods offer interesting approaches for generating low-dimensional data representations that can be transferred to the visualization of PolInSAR data. In the following different dimension reduction approaches applied to PolSAR data are presented.

Pixel-based Dimension Reduction. One of the pioneering works in the field of dimension reduction for the efficient automatic analysis of PolSAR data was conducted by Ainsworth et al. [156]. They proposed a sub-optimal extension of ISOMAP [129] to cost-effectively represent inherently high-dimensional PolSAR datasets in low-dimensional space while retaining most of the information contained in the data. In [157], Tu et al. proposed a novel framework for PolSAR land cover classification using Lapalcian Eigenmaps (LE) to project a 42-dimensional feature vector into a 3-dimensional feature space. The results based on SVM and K-Nearest-Neighbor (KNN) classifications demonstrate that using the resulting embedding achieves superior class separability compared to applying PCA or using the original feature vector. Shi et al. [158] enhanced this approach by introducing supervised graph embedding based on LPP (linear approximation of LE). The method allows the

incorporation of label information of training data and handles the projection of unlabeled test data. Their results demonstrate a clear advantage in classification performance over the unsupervised dimension reduction methods PCA, Kernel PCA (KPCA), and LPP. In [159, 160], linear graph embedding methods are combined with the concept of sparse dictionary learning to incorporate local and global data structures into dimension reduction. While these methods outperforme PCA, a significant improvement in the results compared to LPP is not proven. While not primarily focused on dimension reduction as its main goal, t-SNE has been employed in several studies in the context of PolSAR data analysis. For instance, Quan et al. [161] employed dimension reduction by t-SNE to facilitate the generation of superpixels. He et al. [162] leveraged t-SNE to represent polarimetric information that is subsequently combined with spatial features, improving land cover classification accuracy. Zhang et al. [147] employed t-SNE to visually evaluate and compare the class separability of features derived from deep learning models versus raw polarimetric features. While t-SNE serves as a key tool for visualization and aids in various analytical tasks within these studies, its direct assessment as a dimension reduction method for PolSAR data has not been a central area of investigation. A new approach for the reduction of polarimetric features was proposed by Guo et al. [163]. The approach is based on stacked sparse autoencoders [164] that are applied to reduce the dimension of a large feature set (256 features) extracted from multi-temporal PolSAR data. Classification results using an SVM or CNN for crop type identification shows significant improvements by using stacked sparse autoencoders over traditional dimension reduction methods such as PCA and Locally Linear Embedding (LLE).

Spatially-Aware Dimension Reduction. Tao et al. [165] developed a dimension reduction technique that combines Independent Component Analysis (ICA) with tensor decomposition, leveraging both spatial features and polarimetric features to enhance classification accuracy and computational efficiency. This approach surpasses traditional pixel-based methods based on PCA, ICA, and LE. In subsequent studies this concept is further extended by applying tensor-based enhancements to various existing dimension reduction techniques. For instance, tensorial extensions of LLE [166], Local Discriminant Embedding (LDE) [167], and several linear dimension reduction techniques, including LPP, PCA, LDE, and Linear Discriminant Analysis (LDA) [168], show improvements in land cover classification results. Moreover, Liu et al. [169] proposed to combine multi-frequency, polarimetric, and spatial features using tensor-extensions of PCA and LDA. This method notably outperformed pixel-based approaches, particularly with tensor-based LDA.

Statistic-Aware Dimension Reduction. The previously discussed methods typically involve the extraction of an exhaustive feature set to represent PolSAR data and aim to preserve the neighborhood structure based on these features. The drawbacks of this approach are the high computational effort required to compute and store the exhaustive feature

set and potential biases introduced by manual feature selection. As described in Section 2.2, the polarimetric scattering process is fully described by the polarimetric coherency matrix. Therefore, generic approaches that quantify pixel similarities based directly on the coherency matrices instead of calculating the similarity of feature vectors provide a promising alternative. Matrix distance measures based on the complex Wishart distribution have already demonstrated their value in PolSAR classification [170–174], indicating its potential for efficient and unbiased dimension reduction. Anfinsen et al. [171] defined two distance measures, the Bartlett distance and the Symmetric Revised Wishart Distance (SRWD), that are derived from the complex Wishart distribution of the PolSAR coherency matrix and are suitable metrics to quantify pairwise similarities. Using one of these metrics enables the direct application of neighbor graph-based dimension reduction methods on PolSAR coherency matrices. Integrating the SRWD into the LE method has been shown to provide better initial segmentation for Wishart classifiers compared to traditional entropy and alpha feature-based clustering [175].

Further adaptations using affinity graphs based on the SRWD have been made, linking them with polarimetric features to explore and represent the intrinsic structure of PolSAR data. Ren et al. [176] combined Wishart-based and feature-based graph representations using subspace analysis on a Grassmann manifold [177]. The combined graph is then projected into a low-dimensional space using Laplacian regularized low-rank subspace clustering [178]. Additionally, Huang et al. [179] proposed a co-regularized graph embedding method to construct and embed a joined graph representation. Upon reviewing the outcomes of these studies, it is evident that the discriminative power of the low-dimensional representation based solely on Wishart distance is comparable to those based on extracted polarimetric features. Notably, the proposed combinations did not significantly enhance classification results, underscoring that the Wishart-based approach alone is adequate to represent discriminative information.

Summary. The review of dimension reduction techniques applied to PolSAR data reveals that non-linear dimension reduction methods can effectively represent polarimetric information in low-dimensional spaces while maintaining class separability. These findings encourage the transfer of the concept based on feature extraction followed by dimension reduction using non-linear feature transformation to multi-frequency PolInSAR data. Additionally, the results of statistic-aware dimension reduction methods highlight the feasibility of deriving meaningful low-dimensional representations directly from PolSAR coherency matrices using statistical measures such as the SRWD. This approach avoids extensive feature extraction and minimizes bias. In this thesis, the concept is transferred to PolInSAR data and is tested in the context of data visualization.

4.2.3 Conclusion

Common visualization techniques for PolInSAR data typically involve selecting three polarimetric features to create color composites, which only partially convey the data's full potential. A more comprehensive representation can be achieved using dimension reduction methods similar to those used in hyperspectral image visualization, as highlighted in existing literature. Yet, specific discussions on applying these methods to PolInSAR data are scarce. This thesis aims to fill this research gap by leveraging UMAP, a dimension reduction technique demonstrated effective across various domains, for PolInSAR data visualization. Drawing insights from PolSAR classification literature, this study will investigate two novel UMAP adaptations for PolInSAR data: first, projecting a broad spectrum of polarimetric and interferometric features into a low-dimensional space and second, integrating the SRWD with UMAP for direct data application. These approaches are intended to improve the visual interpretation of PolInSAR data by human observers, enhancing the ease of data evaluation and supporting decision-making in its practical applications.

4.3 Method

In this work, the UMAP algorithm is applied and investigated with regard to its ability to find low-dimensional real-valued representations of multi-frequency PolInSAR data that enable visual data exploration. Two approaches, which are referred to as F-UMAP and W-UMAP, are developed to adapt and apply UMAP to a PolInSAR dataset. The general workflow, which includes the initialization of the high-dimensional representation of the PolInSAR dataset, the UMAP-based projection of the dataset into a low-dimensional space and the subsequent data visualization, is shown in Figure 4.1. In the following, after reviewing the fundamentals of the UMAP algorithm, F-UMAP and W-UMAP are described in detail.

4.3.1 UMAP Algorithm

For clarity, several notations are introduced. Let $\mathbf{x}_i = [x_i^1\ x_i^2 ... x_i^n]$, represent an n-dimensional feature vector, which describe the i^{th} sample of a dataset. The corresponding n-dimensional dataset comprising N samples is denoted as $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$. Considering the Pol-InSAR-Island dataset, $\mathbf{X}$ signifies a set of PolInSAR image pixels, with $\mathbf{x}_i$ being an n-dimensional representation of the i^{th} pixel. The projection rule, which maps n-dimensional samples into an embedding space with d dimensions (where $d < n$), is denoted as f_d. The expression $\mathbf{Z}_d = f_d(\mathbf{X})$ describes the projection of the entire dataset $\mathbf{X}$ into the d-dimensional dataset $\mathbf{Z}$, and $\mathbf{z}_d = f_d(\mathbf{x})$ represents the projection of an individual sample.

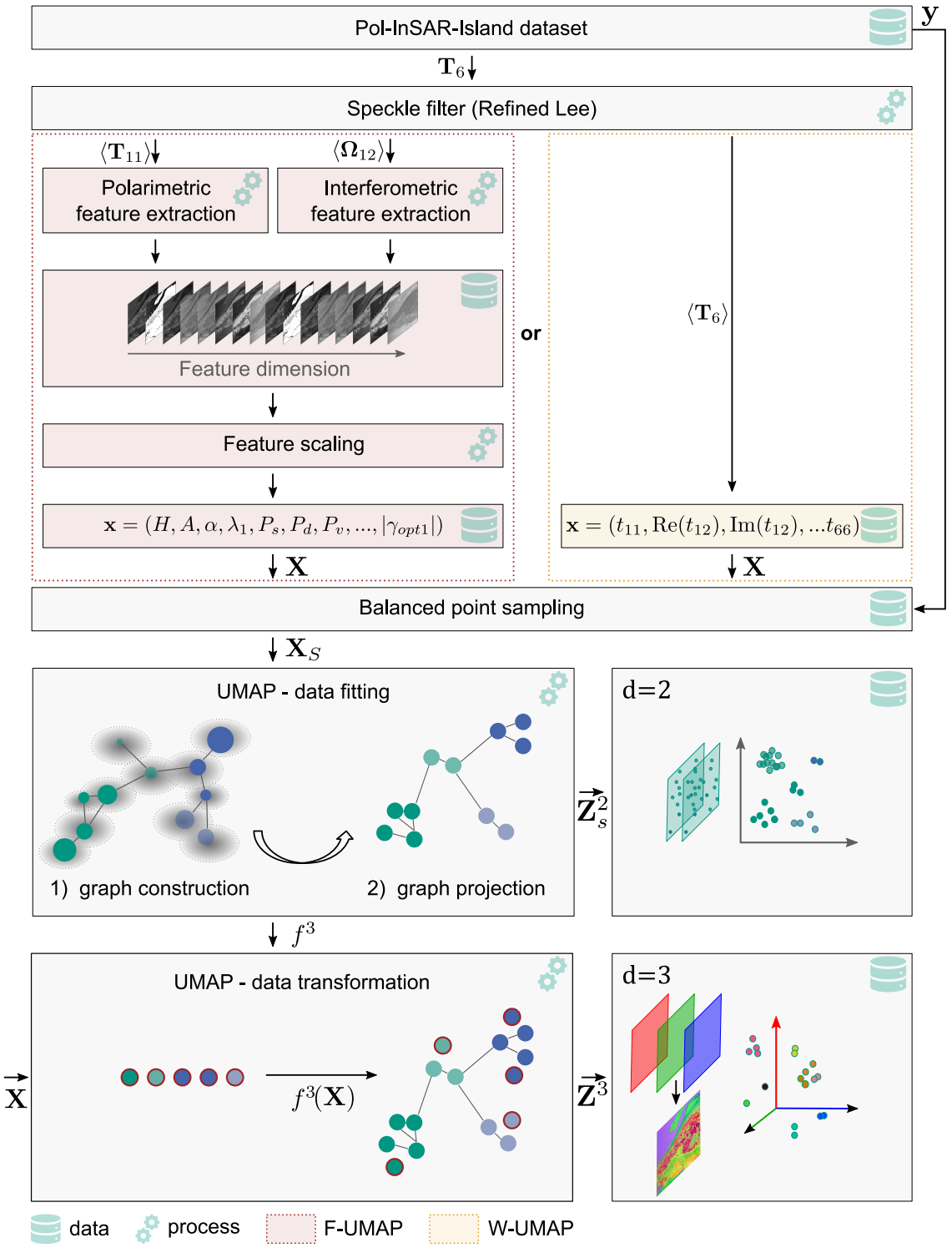


Figure 4.1.: Workflow of UMAP-based visualization of PolInSAR data. The F-UMAP approach is highlighted in red. The W-UMAP approach is highlighted in yellow.

The objective of the UMAP algorithm is to identify an embedding $\mathbf{Z}$ for a dataset $\mathbf{X}$, ensuring that samples close in $\mathbf{X}$ remain close in $\mathbf{Z}$. The UMAP algorithm first constructs a topological representation, specifically a fuzzy simplicial complex, of the dataset $\mathbf{X}$. Second, the low-dimensional representation $\mathbf{Z}$ is optimized to ensure that its topological representation retains maximal structural similarity to that of the high-dimensional dataset $\mathbf{X}$.

For a comprehensive description with mathematical underpinnings on the construction of the topological representation of a dataset and its suitability in approximating the topology of a dataset, refer to the original UMAP article [26]. Basically, a topological representation can be considered as an undirected weighted neighbor graph, where nodes represent individual data points ($\mathbf{x}_i$ or $\mathbf{z}_i$), and the weights on the edges denote the probabilities of connections between data points. The UMAP algorithm is divided into two principal phases: first, the construction of the graph G that represents the n-dimensional dataset $\mathbf{X}$ and second, the computation of a low-dimensional layout of that graph. The computational steps involved in these phases are outlined subsequently.

1. **Graph construction**: To construct the undirected weighted neighbor graph G, initially, a directed weighted neighbor graph $\bar{G}$ is computed. For each data point $\mathbf{x}_i$, a set of k nearest neighbor points $\{\mathbf{x}_{i_1}, ..., \mathbf{x}_{i_k}\}$ under a distance measure d is determined. Subsequently, edge weights $w_{\bar{G}}((\mathbf{x}_i, \mathbf{x}_{i_j}))$ (with $1 \leq j \leq k$) between each point and its neighbors are computed according to:

$$w_{\bar{G}}((\mathbf{x}_i, \mathbf{x}_{i_j})) = \exp\left(\frac{-\max(0, \mathrm{d}(\mathbf{x}_i, \mathbf{x}_{i_j}) - \rho_i)}{\sigma_i}\right). \tag{4.1}$$

   The parameter ρ_i is defined as the distance from $\mathbf{x}_i$ to its nearest neighbor, which guarantees the local connectivity of the graph. Specifically, it ensures that $\mathbf{x}_i$ is connected to at least one other point through an edge with a weight of one (refer to [26] for the derivation). The scaling factor σ_i is selected such that:

$$\sum_{j=1}^{k} \exp\left(\frac{-\max(0, \mathrm{d}(\mathbf{x}_i, \mathbf{x}_{i_j}) - \rho_i)}{\sigma_i}\right) = \log_2(k). \tag{4.2}$$

   This scaling process allocates a local notion of distance to each point based on its distance to the k^{th} nearest neighbor. The scaling factor is small for data points located in a dense region and large for those in a sparse region. For comprehensive reasoning and derivation regarding employing local distance metrics to approximate the topology of real-world datasets, the reader is directed to [26].

   To derive the undirected weighted graph G, the edge weights of the directed graph $\bar{G}$ between two points are combined to:

$$w_G((\mathbf{x}_i, \mathbf{x}_j)) = \left(w_{\bar{G}}((\mathbf{x}_i, \mathbf{x}_j)) + w_{\bar{G}}((\mathbf{x}_j, \mathbf{x}_i))\right) - w_{\bar{G}}((\mathbf{x}_i, \mathbf{x}_j)) \cdot w_{\bar{G}}((\mathbf{x}_j, \mathbf{x}_i)). \tag{4.3}$$

The resulting weight indicates the probability of the existence of at least one edge between the points $\mathbf{x}_i$ and $\mathbf{x}_j$.

2. **Graph layout**: The objective of the second phase is to find a low-dimensional representation $\mathbf{Z}$ of the data that closely matches the topological structure of $\mathbf{X}$. For this purpose, $\mathbf{Z}$ is first initialized using spectral embedding [136]. Analogous to the construction of G for the dataset $\mathbf{X}$, a graph Q is calculated for $\mathbf{Z}$. Here, the (symmetric) edge weight between the points $\mathbf{z_i}$ and $\mathbf{z_j}$ is calculated by:

$$w_Q((\mathbf{z}_i, \mathbf{z}_j)) = (1 + a||\mathbf{z}_i - \mathbf{z}_j||^{2b})^{-1}. \tag{4.4}$$

The parameters a and b represent hyperparameters that can either be defined directly by the user or indirectly via a hyperparameter that is easier to interpret and defines the minimum distance between points in the low-dimensional space.

Within an optimization process, $\mathbf{Z}$ is adjusted so that the graph representations G and Q are as similar as possible. The cross-entropy cost function is employed to assess the discrepancy between the representations:

$$c = \sum_{i \neq j} w_G((\mathbf{x}_i, \mathbf{x}_j)) \log \frac{w_G((\mathbf{x}_i, \mathbf{x}_j))}{w_Q((\mathbf{z}_i, \mathbf{z}_j))} + \left(1 - w_G((\mathbf{x}_i, \mathbf{x}_j)) \log \frac{(1 - w_G((\mathbf{x}_i, \mathbf{x}_j)))}{(1 - w_Q((\mathbf{z}_i, \mathbf{z}_j)))}\right). \tag{4.5}$$

The first term of the cost function can be interpreted as an attractive force applied on neighboring points, while the second term acts as a repulsive force on non-neighbor points to pull them apart in the low-dimensional space.

4.3.2 Feature-UMAP

The main idea of F-UMAP (shown in the yellow framed part in Figure 4.1) is to initially represent multi-frequency PolInSAR data by a combination of polarimetric and interferometric features, which are subsequently projected into a low-dimensional feature space using UMAP. The starting point is the Pol-InSAR-Island dataset presented in Chapter 3, containing the PolInSAR data where each pixel is described by the matrix descriptor $\mathbf{T}_{6i}$ and the corresponding class label y_i. The first processing step applied to the PolInSAR data is speckle filtering, which generates ensemble averages to estimate the coherency matrix $\langle \mathbf{T}_6 \rangle$ that allow subsequent polarimetric feature extraction and reduce speckle noise. For this purpose, the Refined Lee filter [46] described in Section 2.2.3 is employed, using a filter window size of 5×5 pixels. The choice of the comparatively small window aims to preserve spatial details. Subsequently, the resulting data is radiometric calibrated by normalizing the data to a reference area perpendicular to the slant range direction (known as γ^0 calibration) using the local incidence θ_i and and assuming a flat terrain.

To extract relevant information at pixel level, the subsequent processing step involves polarimetric feature extraction based on the 3×3 polarimetric coherency matrix $\langle \mathbf{T}_{11} \rangle$, and interferometric feature extraction based on $\langle \mathbf{\Omega}_{12} \rangle$. A detailed summary of all extracted features is presented in Table 4.2.

In total, 42 features (36 polarimetric and 6 interferometric features) have been selected, each described in detail in Section 2.2.4, Section 2.2.5, and Section 2.3.2. The polarimetric features are further categorized into features that are directly derived from $\langle \mathbf{T}_{11} \rangle$ ($\langle \mathbf{T}_{11} \rangle$-descriptors), features based on model-based and eigenvalue-based incoherent target decomposition methods, and features drawn from concepts of optical polarimetry. While additional polarimetric features exist, this diverse set of features is assumed to capture the essential polarimetric information comprehensively.

The interferometric features are derived from the complex interferometric coherences $\gamma_{opt,i}$ ($i = 1,2,3$) (see Equation 2.64) using optimized projection vectors [2]. The features considered include the magnitude $|\gamma_{opt,i}|$, which reflects the effects of temporal decorrelation, volume decorrelation, and system noise, and the estimated unwrapped phase $\Psi_{opt,i}$, which reveals information about the topography of the observed scene. For phase unwrapping, the quality-guided path-following method proposed in [180] is utilized. This method offers an effective balance between computational efficiency and reliability [61, 180]. In the phase unwrapping process, the interferometric coherence $|\gamma_{opt,i}|$ serves as the quality metric, with areas of low coherence ($|\gamma_{opt,i}| < p_{25\%}(|\gamma_{opt,i}|)$) being masked out.

Within the step feature scaling, features characterized by a high dynamic range ($|\langle t_{ij} \rangle|$, λ_i, $\underline{\lambda}$, P_s, P_d, P_v) undergo logarithmic scaling. This adjustment expands the data in the lower value range, where most observations are typically concentrated, and compresses the upper range, which contains fewer observations. Further, the value ranges of the various extracted features vary significantly. To prevent a bias toward features with higher absolute values, which could occur, for example, during the nearest neighbor calculation in the graph construction phase of the UMAP algorithm, feature scaling is necessary. Since some features contain outliers with extremely low or high values, all features are first clipped at their 1^{st} and 99^{th} percentiles before being linearly scaled to the range [0, 1]. After feature extraction and scaling, each pixel is described by a 42-dimensional feature vector $\mathbf{x}_i$ for single-frequency data. To represent multi-frequency information, the feature vector $\mathbf{x}_i$ is constructed by concatenating features from two frequencies (S-band and L-band), resulting in 84 dimensions. The dataset $\mathbf{X}$ then consists of these feature vectors $\mathbf{x}_i$ (with $1 \leq i \leq N$).

Point subsampling is performed to limit the computational effort. Considering the class labels $\mathbf{y}$, a predefined number of points is selected so that all classes are represented approximately equally. The selected subset, denoted as $\mathbf{X}_s$, is used as the input for deriving mapping functions f^2, f^3, which project the data into either a 2-dimensional or 3-dimensional Euclidean space. This step is denoted as UMAP - data fitting in Figure 4.1.

Table 4.2.: Polarimetric and interferometric features, extracted to build a high-dimensional feature representation of PolInSAR data.

		Expression	Description
Polarimetry	$\langle \mathbf{T}_{11} \rangle$-descriptors	$\langle t_{ii} \rangle$	Diagonal elements of $\langle \mathbf{T}_{11} \rangle$
		$\lvert\langle t_{ij} \rangle\rvert$	Magnitude of off-diagonal elements of $\langle \mathbf{T}_{11} \rangle$
		$\arg(\langle t_{ij} \rangle)$	Phase of off-diagonal elements of $\langle \mathbf{T}_{11} \rangle$
		$10\log(r_{hh/vv})$	Co-polarized differential reflectivity
		$10\log(r_{hv/hh})$	Cross-polarized differential reflectivity
		$10\log(r_{hv/vv})$	Cross-polarized differential reflectivity
		$\lvert\rho_{12}\rvert, \lvert\rho_{13}\rvert, \lvert\rho_{23}\rvert$	Polarimetric coherences
		μ	Conformity coefficient [56]
	Incoherent decomposition	λ_i	Eigenvalues of $\langle \mathbf{T}_{11} \rangle$
		p_i	Normalized eigenvalues (pseudo probabilities)
		$\underline{\lambda}$	Weighted mean of eigenvalues
		H, A, α	Entropy, anisotropy and mean alpha angle derived from eigenvalue decomposition of $\langle \mathbf{T}_{11} \rangle$ [52]
		$P_{s,VZ}, P_{d,VZ}, P_{v,VZ}$	Odd-bounce, double-bounce and volume scattering power derived from Van Zyl decomposition [55]
		$P_{s,Y}, P_{d,Y}, P_{v,Y}$	Odd-bounce, double-bounce and volume scattering power derived from Yamaguchi decomposition [54]
	Optical polarimetry	$\hat{H}_p$	Scattering predominance [57]
		$\hat{H}_d$	Scattering diversity [57]
		d_p	Degree of purity [57]
		d_{idx}	Depolarization index [57]
Interferometry		$\lvert\gamma_{opt,i}\rvert$	Magnitude of optimal interferometric coherences [2]
		$\Psi_{opt,i}$	Optimal interferometric phases (unwrapped) [2]

The F-UMAP approach uses the Euclidean distance $\mathrm{d}(\mathbf{x}_i, \mathbf{x}_j) = ||\mathbf{x}_i - \mathbf{x}_j||$ as the distance metric in the UMAP algorithm within the graph construction phase. The outputs of the UMAP transformation process are twofold: firstly, the 2-dimensional and 3-dimensional data representations $\mathbf{Z}_s^2$, $\mathbf{Z}_s^3$, which are readily visualizable in scatter plots, and secondly, the optimized projection functions f^2 and f^3. While the mapping function f^2 is not used any further, the mapping function f^3 is utilized in the subsequent UMAP - data transformation step to project unseen data points, e.g., all remaining points of the dataset, into the 3-dimensional space. The resulting representation $\mathbf{Z}^3$ of the entire dataset serves as a foundation to generate color composites, which will be described in more detail in Section 4.3.4.

4.3.3 Wishart-UMAP

The W-UMAP approach, highlighted in yellow in Figure 4.1, differs from F-UMAP in two significant aspects: the construction of the input $\mathbf{X}$, representing the PolInSAR data, and the choice of distance metric used in the UMAP algorithm to construct the neighbor graph. Instead of representing PolInSAR data by combining extracted polarimetric and interferometric features, the coherency matrix is used directly to represent a data point $\mathbf{x}_i = \langle \mathbf{T}_6 \rangle_i$. This approach eliminates the necessity of selecting, calculating, and storing features.

While the F-UMAP approach employs the Euclidean distance to measure the distance between two feature vectors, the W-UMAP approach uses a metric that directly quantifies the distance between coherency matrices $\langle \mathbf{T}_6 \rangle_i$ and $\langle \mathbf{T}_6 \rangle_j$. A suitable metric for this application is the SRWD which is proposed in [171]. The metric builds upon assumptions about the statistical distribution of the data. The elements of the scattering matrix are assumed to follow a multivariate complex Gaussian distribution with mean zero so that the distribution of the scattering vector $\mathbf{k}_p$ results in [181]:

$$p(\mathbf{k}_p) = \frac{1}{\pi^q |\mathbf{\Sigma}|} \exp\left(-\mathbf{k}_p^{*T} \mathbf{\Sigma}^{-1} \mathbf{k}_p\right), \tag{4.6}$$

where $\mathbf{\Sigma} = E\{\mathbf{k}_p \mathbf{k}_p^{*T}\}$ is the complex covariance matrix, q correspond to the dimension of the scattering vector $\mathbf{k}_p$ (which is 3 in the monostatic case) and the operator $|\cdot|$ denotes the determinant. From this, it can be derived that the multilooked coherency matrix $\langle \mathbf{T} \rangle$, computed by averaging n single-look samples:

$$\langle \mathbf{T} \rangle = \frac{1}{n} \sum_{i=1}^{n} \mathbf{k}_p \mathbf{k}_p^{*T}, \tag{4.7}$$

follows a complex Wishart distribution [181, 182]:

$$p(\langle \mathbf{T} \rangle) = \frac{n^{nq} |\langle \mathbf{T} \rangle|^{n-q} \exp\left(-n\text{trace}(\mathbf{\Sigma}^{-1}\langle \mathbf{T} \rangle)\right)}{K(n,q)|\mathbf{\Sigma}|^n}. \tag{4.8}$$

Here $K(n,q)$ is a constant defined by:

$$K(n,q) = \pi^{q(q-1)/2} \prod_{k=1}^{q} \Gamma(n-k+1), \tag{4.9}$$

where $\Gamma(\cdot)$ is the standard gamma function. This applies also to the PolInSAR coherency matrix $\langle \mathbf{T}_6 \rangle$.

Based on this assumption, the SRWD is derived from a generalized likelihood ratio test and is defined as:

$$d_{SRW}(\langle \mathbf{T} \rangle_i, \langle \mathbf{T} \rangle_j) = \frac{1}{2}\text{trace}(\langle \mathbf{T} \rangle_i \langle \mathbf{T} \rangle_j^{-1} + \langle \mathbf{T} \rangle_j \langle \mathbf{T} \rangle_i^{-1}) - u. \tag{4.10}$$

Here, u represents the dimension of the matrix $\langle \mathbf{T} \rangle$, which is three for PolSAR data and six for PolInSAR data. For an in-depth derivation of this distance measure, the reader is referred to [171]. The selection of SRWD is justified because it incorporates all elements of the matrix, thereby utilizing complete information for distance calculation. Furthermore, SRWD fulfills all metric requirements, including generalized non-negativity, identity of discernibles, symmetry, and subadditivity [171].

To employ the SRWD for comparing multi-frequency data, the sum of the SRWD of the coherency matrices for each frequency band is used. The subsequent processing steps, e.g., data fitting and data transformation, are analogous to those of the F-UMAP approach.

4.3.4 Visualization

In the processing of the Pol-InSAR-Island dataset using the proposed approaches, two primary outputs are generated: a 2-dimensional representation $\mathbf{Z}_s^2$, including a label-balanced sample set of data points, and a 3-dimensional representation $\mathbf{Z}^3$, containing all data points of the dataset. The 2-dimensional representation is used to visualize the structure and neighborhood relations within the PolInSAR data in an easily understandable scatter plot. It should be noted that the scatter plot axes do not have a direct physical interpretation. Instead, the relative positioning of the scatter points is meaningful, as it indicates the similarity in the backscattering properties of the represented data points. Thus, the scatter plot aids in identifying patterns, clusters, and outliers that would not be easily detected in the original high-dimensional dataset. Additionally, coloring the data points based on the given labels $\mathbf{y}$ allows for visual analysis of class separability.

To generate color images based on the 3-dimensional representation of multi-frequency PolInSAR data, the following steps are employed to translate the position of a point in the feature space into a color. Initially, the data values are clipped to the 1^{st} and 99^{th} percentiles to mitigate the influence of outliers. This step prevents inefficient use of the color space resulting in low contrast, which can occur when most points are concentrated within a narrow value range. Subsequently, the position of each data point is translated into a color by linearly scaling the three feature axes to a range from 0 to 255 based on the global minimum and maximum values across all feature dimensions. The scaled values are then assigned to red, green, and blue intensities, respectively. The resulting color composite displays regions with similar scattering behaviors in similar colors, offering a visually intuitive way to analyze the spatial distribution of scattering characteristics.

4.4 Experimental Setup

The Pol-InSAR-Island dataset serves as the foundation for evaluating the presented approaches. The class labels within the dataset are used exclusively for selecting a balanced subset of 100,000 points and for evaluation purposes. Notably, points labeled as *Settlement* are entirely excluded. This exclusion is due to the fact that regions labeled as *Settlement* comprise various land cover types (including man-made objects, asphalt, meadows, and vegetation), making this reference unsuitable for pixel-based evaluation.

4.4.1 Comparison Methods

In the evaluation, the proposed UMAP-based methods are compared to three other dimension reduction techniques: PCA, LE and t-SNE. PCA, a well-established linear dimension reduction method that preserves the global data structure, is selected due to its popularity and ease of implementation. LE and t-SNE are nonlinear methods focusing on local neighborhood preservation. LE is included in this comparison due to its proven effectiveness in reducing dimensions of PolSAR observables in prior studies [157, 171]. t-SNE is chosen because it is regarded as a state-of-the-art method for visualizing high-dimensional datasets [145, 149] and its successful application in PolSAR data analysis as described in Section 4.2. While PCA can only be used for dimension reduction based on the feature representation of the PolInSAR data, LE and t-SNE can additionally be applied directly to the coherency matrix representation analogous to the proposed W-UMAP method. The functional principles of the comparison methods are briefly outlined below.

PCA. PCA is a statistical technique that transforms a dataset consisting of possibly correlated features into a set of linearly uncorrelated features called principal components. The

main idea behind PCA is to identify the directions (principal components) along which the variance of the data is maximized. The data is transformed into a new feature space where the first feature axis (corresponding to the first principal component) captures the greatest variance, and each subsequent component, orthogonal to the previous ones, captures the maximum remaining variance.

The computational steps of PCA include:

1. **Data centering**: The original data are centered to have mean zero:

$$\mathbf{x}_c = \mathbf{x} - \mathbf{x}_\mu, \tag{4.11}$$

   where $\mathbf{x}_\mu$ is the mean feature vector over all samples.

2. **Covariance matrix computation**: Calculate the covariance matrix to represent how features are related to each other. The covariance matrix $\mathbf{C}$ of dataset $\mathbf{X}_c$ with N data points is given by:

$$\mathbf{C} = \frac{1}{N-1}\mathbf{X}_c^\top \mathbf{X}_c. \tag{4.12}$$

3. **Eigen decomposition**: Decompose the covariance matrix into its eigenvalues and eigenvectors by solving:

$$\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i. \tag{4.13}$$

   The eigenvectors $\mathbf{v}_i$ determine the orientations of the new feature axes, and the eigenvalues λ_i correspond to the amount of variance that the new axes explain. The eigenvectors are sorted by corresponding eigenvalues in descending order.

4. **Projection**: Project the original data into the space spanned by the eigenvectors:

$$\mathbf{Z} = \mathbf{X_c}\mathbf{V}, \tag{4.14}$$

   where $\mathbf{V}$ is the matrix whose columns are the ordered eigenvectors.

A low-dimensional representation with d dimensions is obtained by taking the first d columns of $\mathbf{Z}$.

Laplacian Eigenmaps. LE is a nonlinear dimension reduction technique used to capture the local structure of high-dimensional data based on nearest-neighbor connectivity. It builds on the assumption that high-dimensional data lie on a low-dimensional manifold embedded within the high-dimensional space.

The computational steps of LE include:

1. **Graph construction**: Construct a graph G where each node represents a data point. Edges between nodes are formed if they are among the k nearest neighbors of each other. Assign edge weights w_{ij} that reflect the closeness of the points using a Gaussian kernel similarity function:

$$w_{ij} = \exp\left(-\frac{\mathrm{d}(\mathbf{x}_i, \mathbf{x}_j)^2}{2\sigma^2}\right), \tag{4.15}$$

   where $\mathrm{d}(\mathbf{x}_i, \mathbf{x}_j)$ is the distance between the data points $\mathbf{x}_i$ and $\mathbf{x}_j$.

2. **Compute graph Laplacian**: Compute the graph Laplacian matrix $\mathbf{L}$, which is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{W}$ is the weight matrix and $\mathbf{D}$ is a diagonal matrix $d_{ii} = \sum_j w_{ij}$.

3. **Eigen decomposition:** Solve the generalized eigenvalue problem:

$$\mathbf{L}\mathbf{v} = \lambda \mathbf{D}\mathbf{v}. \tag{4.16}$$

   Sort the eigenvalue-eigenvector pairs in ascending order.

The eigenvectors corresponding to the smallest non-zero eigenvalues provide the embedding coordinates of the data points in the low-dimensional space.

t-SNE. t-SNE is a dimension reduction technique which works similar to the UMAP algorithm. It is also based on constructing a weighted neighbor graph and finding a low-dimensional data embedding with a similar graph representation. However, there are some differences concerning the derivation of edge weights and the optimization objective:

1. **Calculating edge weights**: While UMAP uses a uniform distribution over a fuzzy simplicial set to model neighborhood relationships, t-SNE uses a joint probability distribution based on Gaussian distribution in high dimensions resulting in the definition of edge weights:

$$w_{\bar{G}}((\mathbf{x}_i, \mathbf{x}_{i_j})) = \frac{\exp(-\mathrm{d}(\mathbf{x}_i, \mathbf{x}_{i_j})^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-\mathrm{d}(\mathbf{x}_i, \mathbf{x}_k)^2/2\sigma_i^2)} \tag{4.17}$$

$$w_G((\mathbf{x}_i, \mathbf{x}_j)) = \frac{w_{\bar{G}}((\mathbf{x}_i, \mathbf{x}_j)) + w_{\bar{G}}((\mathbf{x}_j, \mathbf{x}_i))}{2N} \tag{4.18}$$

   where σ_i indicates the variance of the Gaussian centered on the datapoint $\mathbf{x}_i$. For the definition of the edge weights of Q, the Student t-distribution is used, resulting in:

$$w_Q((\mathbf{z}_i, \mathbf{z}_j)) = \frac{(1 + \|\mathbf{z}_i - \mathbf{z}_j\|^2)^{-1}}{\sum_{k\neq l}(1 + \|\mathbf{z}_k - \mathbf{z}_l\|^2)^{-1}}. \tag{4.19}$$

2. **Optimization objective**: UMAP minimizes the cross-entropy between the representation of the data in high-dimensional and low-dimensional spaces. In contrast, t-SNE minimizes the Kullback-Leibler divergence:

$$c = \sum_{i \neq j} w_G((\mathbf{x}_i, \mathbf{x}_j)) \log \frac{w_G((\mathbf{x}_i, \mathbf{x}_j))}{w_Q((\mathbf{z}_i, \mathbf{z}_j))}. \tag{4.20}$$

4.4.2 Hyperparameter Setting

Various hyperparameters must be defined for the implementation of the proposed workflow and the comparison methods. An important hyperparameter in LE and UMAP is the number of nearest neighbors k, which is used for constructing the graph representations of the high-dimensional data. In this study, $k = 50$ is set for both methods. This choice ensures a balance between the size of the considered neighborhood and the computational effort. Analogous to the number of nearest neighbor points in UMAP and LE, a suitable value for the so-called *perplexity* must be set for the t-SNE algorithm. The *perplexity* acts as a smooth measure of the effective number of neighbors, which is considered for each point when constructing the neighbor graph. However, it is important to note that *perplexity* does not directly set the number of neighbors. In the original paper [139], an interval between five and 50 is suggested. In this work, the value is set to 30. To enhance computational efficiency, the Barnes-Hut implementation of t-SNE is used for experiments.

The remaining hyperparameters of all methods are kept at their default values, provided by the employed implementations. For PCA, LE, and t-SNE, the scikit-learn implementations [183] are used. For UMAP, the Python implementation provided by the authors of [26] is utilized [184].

4.4.3 Evaluation

The various dimension reduction methods are employed to project the sample set $\mathbf{X}_s$ of the Pol-InSAR-Island dataset, consisting either of 84-dimensional feature vectors or of coherency matrices, into a 2-dimensional Euclidean space. The resulting representations $\mathbf{Z}_s^2$ are evaluated regarding their effectiveness in capturing the structure and informational content of PolInSAR data. Besides qualitative visual assessments through scatter plots, four distinct quality metrics are employed for quantitative analysis:

- Trustworthiness m_t [185],
- Continuity m_c [185],
- KNN classification accuracy m_{KNN}, and
- Distance Consistency m_{dc} [186].

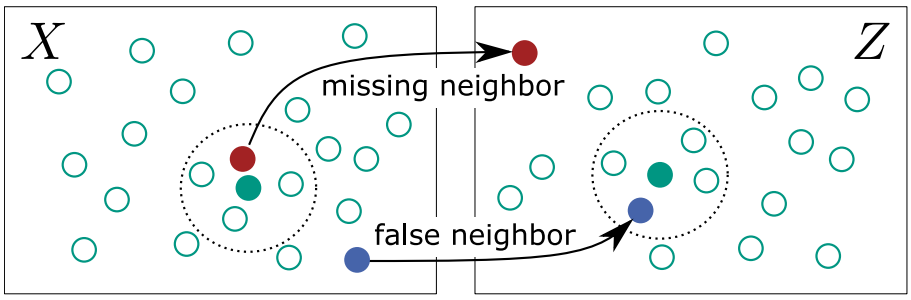


Figure 4.2.: Two projection error types related to local neighborhood preservation: False neighbor (blue) and missing neighbor (red). The figure is redrawn from [143].

Trustworthiness and Continuity. These metrics evaluate changes in local neighborhood relationships when transitioning from the original n-dimensional data space to the low-dimensional embedding space. Generally, two types of errors may occur: false and missing neighbors, as illustrated in Figure 4.2. Trustworthiness evaluates if neighboring points in the embedding space also share proximity in the original n-dimensional space, thus focusing on false neighbors. This measure indicates the degree of confidence one should have that observed patterns, such as clusters, in the embedding space actually exist in the original n-dimensional data. The Trustworthiness m_t is expressed as:

$$m_t = 1 - \frac{2}{NK(2N - 3K - 1)} \sum_{i=1}^{N} \sum_{j \in U_i^{(K)}} (r(i,j) - K), \tag{4.21}$$

where $U_i^{(K)}$ represents a set of data points such that $\mathbf{z}_j$ is among the K nearest neighbors of $\mathbf{z}_i$ in the embedding space, but the corresponding $\mathbf{x}_j$ is not among the K nearest neighbors of $\mathbf{x}_i$ in the original space; $r(i,j)$ denotes the rank of $\mathbf{x}_j$ among the nearest neighbors of $\mathbf{x}_i$. Thus, the metric penalizes false neighbors in the embedding space relative to their distances in the original data space.

Conversely, Continuity assesses the preservation of original neighborhood relationships. It is analogous to Trustworthiness but examines the change in neighborhood relationships from the opposite direction. Missed neighbors are penalized according to their distances in the embedding space. Continuity is defined as:

$$m_c = 1 - \frac{2}{NK(2N - 3K - 1)} \sum_{i=1}^{N} \sum_{j \in V_i^{(K)}} (\hat{r}(i,j) - K), \tag{4.22}$$

where $V_i^{(K)}$ includes data points such that $\mathbf{x}_j$ is among the K nearest neighbors of $\mathbf{x}_i$ in the original space, but the corresponding $\mathbf{z}_j$ is not among the K nearest neighbors of $\mathbf{z}_i$ in the embedding space; $\hat{r}(i,j)$ signifies the rank of $\mathbf{z}_j$ among the nearest neighbors of $\mathbf{z}_i$.

In this study, both metrics are computed for various neighborhood sizes ($K = 7, 15, 30, 50, 100, 300$) across five sets consisting of 3,000 sample points each. Sampling is employed to

manage computational demands, and a 5-fold evaluation on different samples ensures a robust estimation.

KNN Classification Accuracy. The KNN method is a supervised classification algorithm that predicts the class label of a test data point by analyzing the class labels of its nearest neighbors from a training set. It is commonly used to assess the quality of dimension reduction methods. By evaluating the accuracy of the KNN classifier trained and tested on the low-dimensional data representation, it can be assessed how well the dimension reduction method preserves the underlying structure and discriminative information. If the accuracy remains high after dimension reduction, it indicates that the low-dimensional representation retains sufficient information for separating different classes.

For the KNN-based evaluation, 5-fold cross-validation is applied. For this purpose, the 2-dimensional dataset $\mathbf{Z}_s^2$, alongside its corresponding labels $\mathbf{y}_s$ derived from the Pol-InSAR-Island dataset, is initially divided into five folds using stratified sampling. Five training and test runs are carried out, utilizing four folds for training and one for testing in each run. The accuracy obtained on each test set is quantified using the mean Intersection-over-Union (mIoU). This metric measures the overlap between predicted and reference classes as follows:

$$m_{KNN} = \frac{1}{l} \sum_{c \in C} \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c + \mathrm{FN}_c}. \tag{4.23}$$

Here, C is a class set containing l classes (with $l = 11$ in the conducted experiments), TP_c (True Positives) represents the number of correctly predicted points belonging to class c, FP_c (False Positives) represents the number of points incorrectly predicted as class c, and FN_c (False Negatives) represents the number of points of class c that were missed. The mIoU ranges from 0 to 1, where 1 signifies a perfect agreement between the predicted and actual class labels.

Distance Consistency. Distance Consistency evaluates the visual quality of 2-dimensional scatter plots by quantifying the perceived visual separation of clusters. This is measured by assessing the degree of overlap among data points from different classes within the rendered scatter plot image. The class overlap within a small image region is quantified using the entropy of the class probability density:

$$H = -\sum_{c \in C} \frac{n_c}{n} \log_2 \left(\frac{n_c}{n} \right), \tag{4.24}$$

where n_c is the number of data points belonging to class c, and n is the total number of data points within the considered image region. For regions containing only points from the same class, the entropy H is zero. The maximum value of entropy H, which is $\log_2(l)$, occurs when all l classes are equally represented in a region.

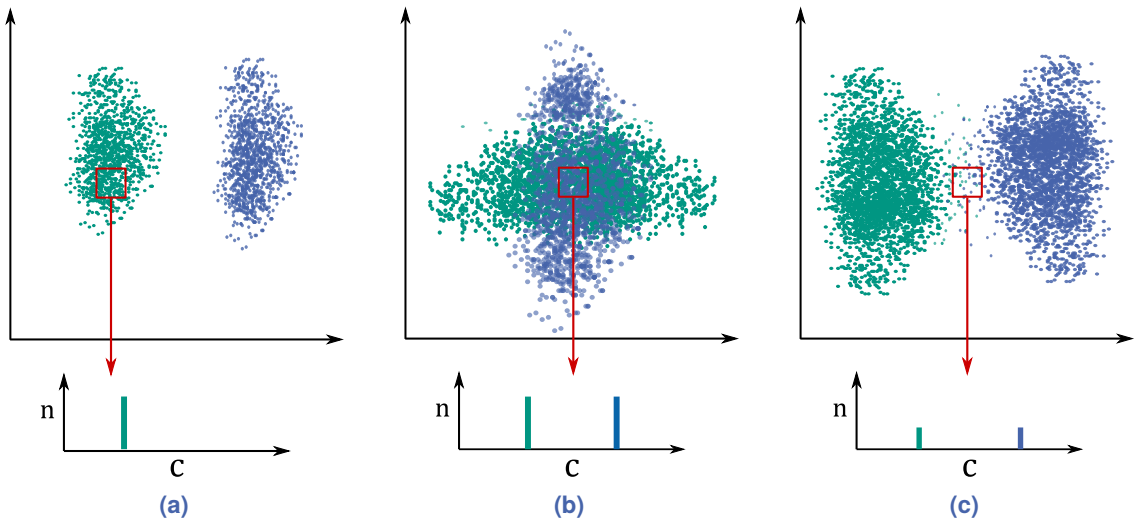


Figure 4.3.: Basic concept of Distance Consistency, redrawn from [186]. The top row shows exemplary data distributions in a 2-dimensional space, and the bottom row depicts the corresponding histograms that count class occurrences in a small image region (red rectangle). Example (a) shows clear class separation resulting in low entropy value and thus high m_{dc}. Example (b) shows overlapping classes resulting in high entropy and thus low m_{dc}. In (c), overlapping classes result in high entropy. However, due to the sparse point density, the indicated image region contributes only weakly to m_{dc}.

Distance Consistency is then calculated by a weighted sum of entropy H across the image, which is divided into non-overlapping rectangular regions denoted r_i. The weight corresponds to the proportion of data points within each region r_i, emphasizing class consistency in areas with higher data concentration. The formula for Distance Consistency is given by:

$$m_{dc} = 1 - \frac{1}{t} \sum p_{r_i} H_{r_i}, \tag{4.25}$$

where t is a normalization factor to scale m_{dc} to a value range of [0, 1]. The concept of Distance Consistency is illustrated in Figure 4.3 to provide a clearer understanding.

In this study, Distance Consistency is calculated for different region sizes (50×50, 30×30, 15×15 and 10×10 pixel). This approach ensures a comprehensive assessment of visual quality at different levels of detail.

4.5 Results

The subsequent section is structured into two main parts. The first part explores the projection of multi-frequency PolInSAR data into a 2-dimensional feature space, comparing UMAP to PCA, LE and t-SNE via the introduced quality metrics. The second part focuses on

projecting multi-frequency PolInSAR data into a 3-dimensional feature space using UMAP to generate color composites that are presented and discussed.

4.5.1 2-dimensional Projections

Qualitative Evaluation. The projection results based on the feature representation of multi-frequency PolInSAR data obtained using the different dimension reduction methods are depicted as scatter plots in Figure 4.4. Each scatter point represents a pixel of the PolInSAR data and is color-coded according to its respective class. Note that class information is solely used for visualization purposes and does not influence the computation of the projections. An arrangement of points according to their class is apparent across all plots, supporting the hypothesis that points representing the same land cover class yield similar backscatter signals. The scatter plots exhibit variation in the shape and separability of class clusters. In the PCA embedding, class clusters appear as diffuse blobs with partially overlapping fuzzy borders. Intra-class structures are not evident. In the LE embedding, data points are concentrated in very narrow areas, complicating the visual analysis of data structure and class separability. In contrast, the t-SNE embedding distributes points throughout the space, forming distinct clusters that facilitate inter- and intra-class separation. Notably, classes such as *High vegetation* (dark green), *Water* (dark blue), *Tidal flat* (light blue), and *Sand* (yellow) are segmented into neighboring subclusters. Similar discernible class clusters and intra-class subgroupings are evident in the UMAP embedding. Compared to the t-SNE embedding, the clusters in the UMAP embedding are more compact, which enhances visual perception. The packing density of points in the embedding space depends on the attractive force applied within the graph layout optimization. In the utilized UMAP implementation, this behavior can be easily adjusted by modifying the hyperparameter min_dist (or a and b, see Equation 4.4). In contrast, influencing the layout of t-SNE embeddings is more challenging as there is no direct control parameter. The known phenomena of the expanded cluster in t-SNE embeddings for large datasets (already observed in [104]) can only be countered by more sophisticated methods, such as those proposed in [105].

Figure 4.5 shows the projection results of LE, t-SNE and UMAP based on SRWD of coherency matrices. Compared to Feature-based projections (Figure 4.4), the biggest improvements are observed for the LE method. However, for cluster identification and analysis, the t-SNE and UMAP projections remain superior for visualization. In comparison to Feature-based t-SNE and UMAP embeddings, notable differences include improved separation of the *Peat bog* class cluster (depicted in purple) and the fragmentation or elongation of clusters associated with the classes *Upper salt marsh* (orange), *Lower salt marsh* (rose), and *Coach grass* (light green). The underlying reason for the latter observation will be explored further in Section 4.5.2.

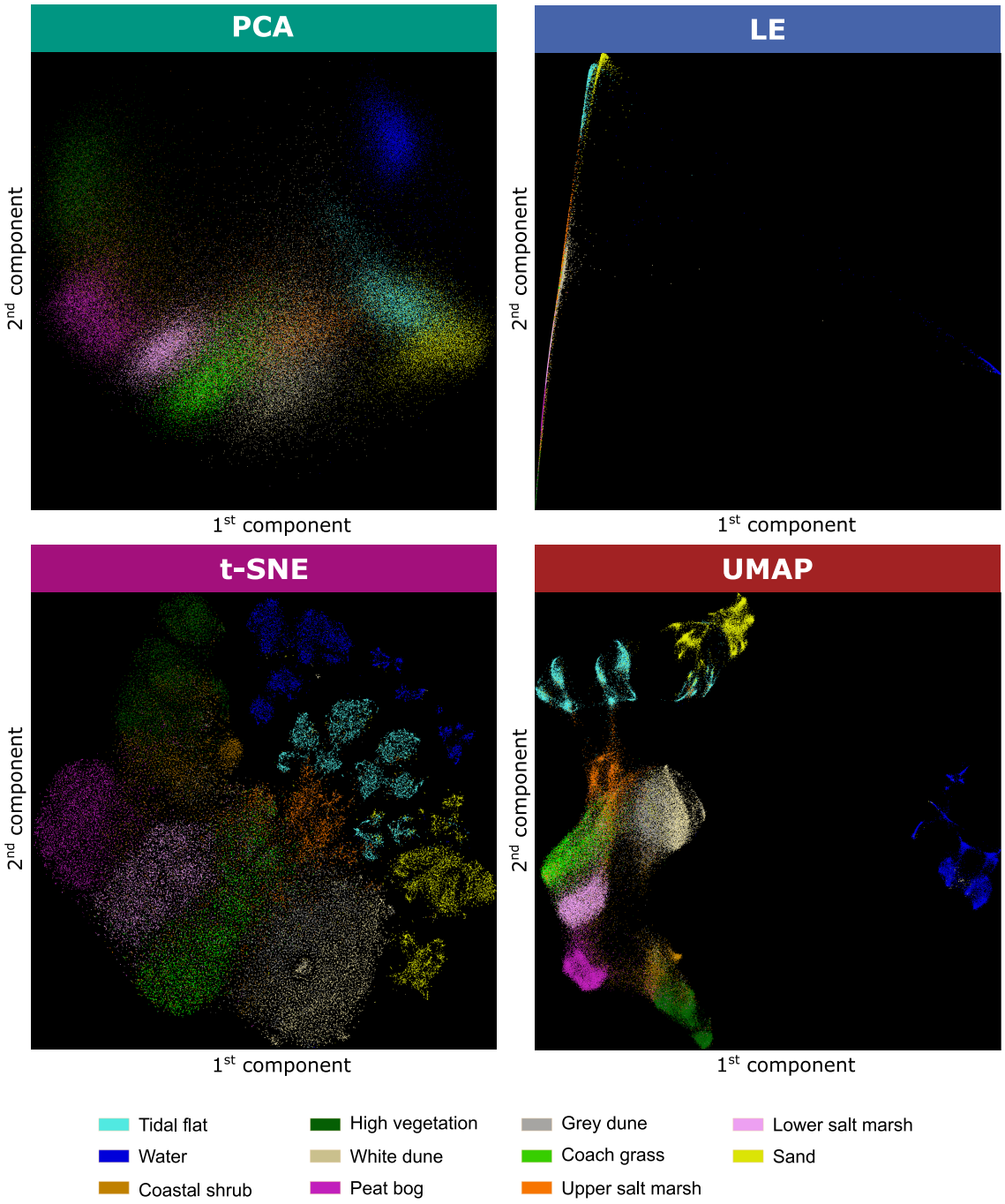


Figure 4.4.: 2-dimensional embeddings of multi-frequency PolInSAR features based on Euclidean distance using PCA, LE, t-SNE and UMAP for dimension reduction. The points are colored according to their class labels.

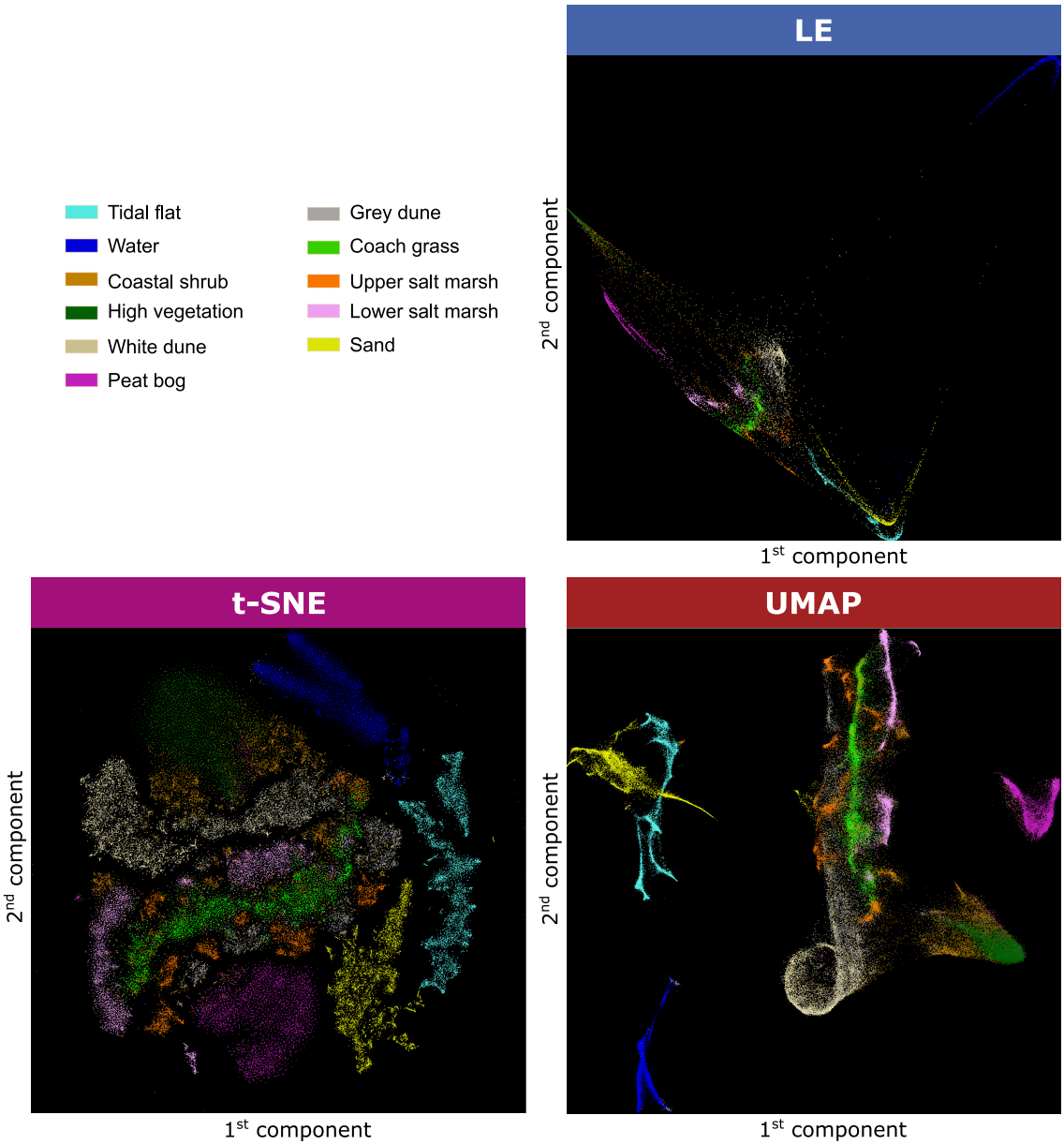


Figure 4.5.: 2-dimensional embeddings of multi-frequency PolInSAR coherency matrices based on the SRWD using LE, t-SNE and UMAP for dimension reduction. The points are colored according to their class labels.

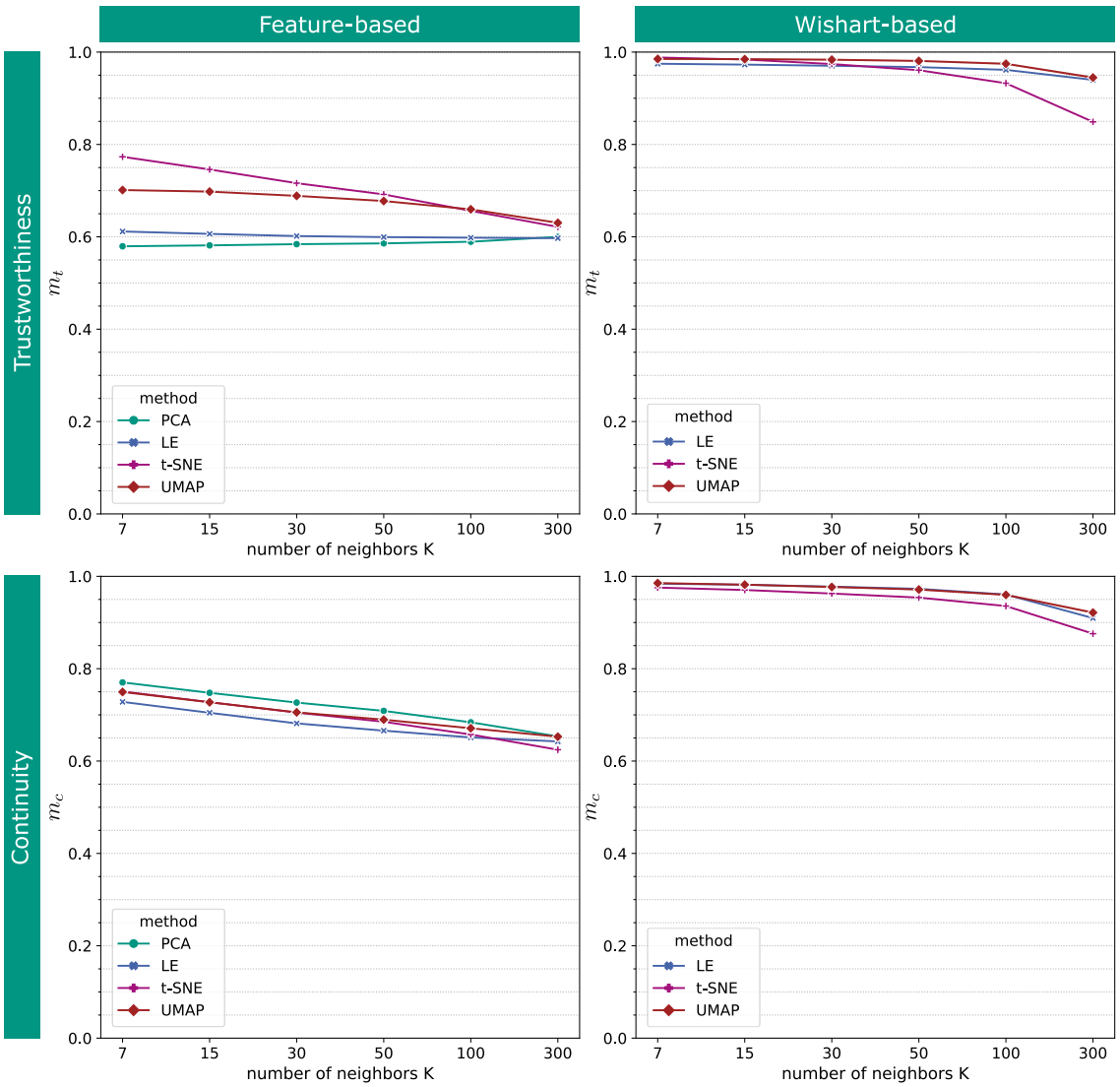


Figure 4.6.: Results for Trustworthiness m_t (top row) and Continuity m_c (bottom row) of Feature-based (left column) and Wishart-based (right column) 2-dimensional embeddings derived with PCA, LE, t-SNE and UMAP.

Quantitative Evaluation. The performance of the tested dimension reduction methods regarding neighborhood preservation, as measured by Trustworthiness and Continuity, is presented in Figure 4.6. Results of Feature-based and Wishart-based dimension reduction are presented separately. In the Feature-based category, t-SNE and UMAP outperform PCA and LE in preserving local neighborhood structures, with t-SNE achieving the highest Trustworthiness scores for small local neighborhoods. For larger neighborhoods ($K > 50$), UMAP and t-SNE show similar Trustworthiness.

For Wishart-based dimension reduction, all methods consistently yield high Trustworthiness scores, approaching the maximum value of one for $K \leq 50$, indicating that the embeddings faithfully reflect the patterns present in the original data. For $K > 50$, the Trustworthiness of t-SNE embeddings decreases slightly, whereas the scores for LE and UMAP remain nearly steady, suggesting better preservation of global neighborhood structures.

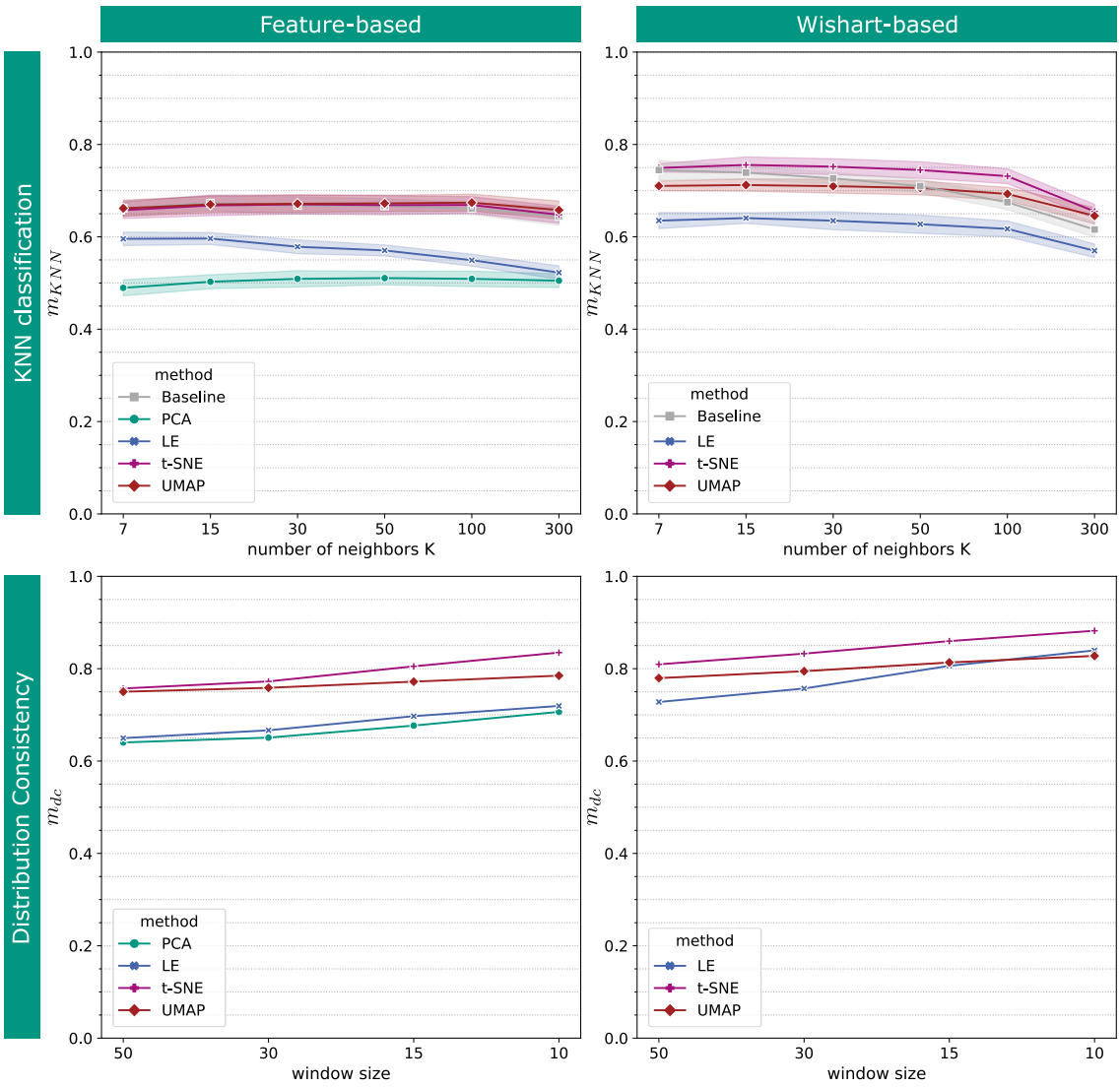


Figure 4.7.: Results for KNN classification m_{KNN} (top row) and Distance Consistency m_{dc} (bottom row) of Feature-based (left column) and Wishart-based (right column) 2-dimensional embeddings derived with PCA, LE, t-SNE and UMAP.

Regarding Continuity, the tested methods exhibit comparable effectiveness within each category—Feature-based and Wishart-based. When comparing between categories, Wishart-based approaches markedly outperform Feature-based methods in both Trustworthiness and Continuity. This suggests that the neighborhood structures based on the SRWD of coherency matrices are better preserved and visualized compared to those derived from the Euclidean distances of feature vectors.

KNN classification results are presented in Figure 4.7 (top). The plots illustrate the average (marked line) and standard deviation (transparent surrounding) of mIoU over five test runs from a 5-fold cross-validation. KNN classification results are shown for the classification of 2-dimensional embeddings and additionally for the classification conducted in the original high-dimensional space, denoted as the Baseline. For the Feature-based Baseline, the Euclidean distance of the original 84-dimensional feature representation is employed

in the classification process, while the SRWD from coherency matrices is used for the Wishart-based Baseline.

The juxtaposition of classification results of the original and the embedded data allows for assessing the extent to which class separability is retained after projection into 2-dimensional space. Results of Feature-based approaches suggest that class separability is effectively maintained through projections using t-SNE and UMAP, with both techniques achieving accuracy comparable to the Baseline. In contrast, the LE and PCA embeddings compromise discriminative information, resulting in a noticeable decrease in classification performance relative to the Baseline, with decreases ranging from 0.07 to 0.12 for LE and from 0.14 to 0.17 for PCA.

For Wishart-based approaches, t-SNE embeddings provide the best classification results, slightly surpassing the Baseline. UMAP results are slightly behind t-SNE, with a maximum difference in the mean mIoU of 0.03, but still exceed the results of LE. Comparing Feature-based to Wishart-based results, the baseline already indicates that leveraging the SRWD between coherency matrices is more suitable for KNN classification than using extracted features. Consequently, there are better classification results for the Wishart-based embeddings compared to the Feature-based embeddings.

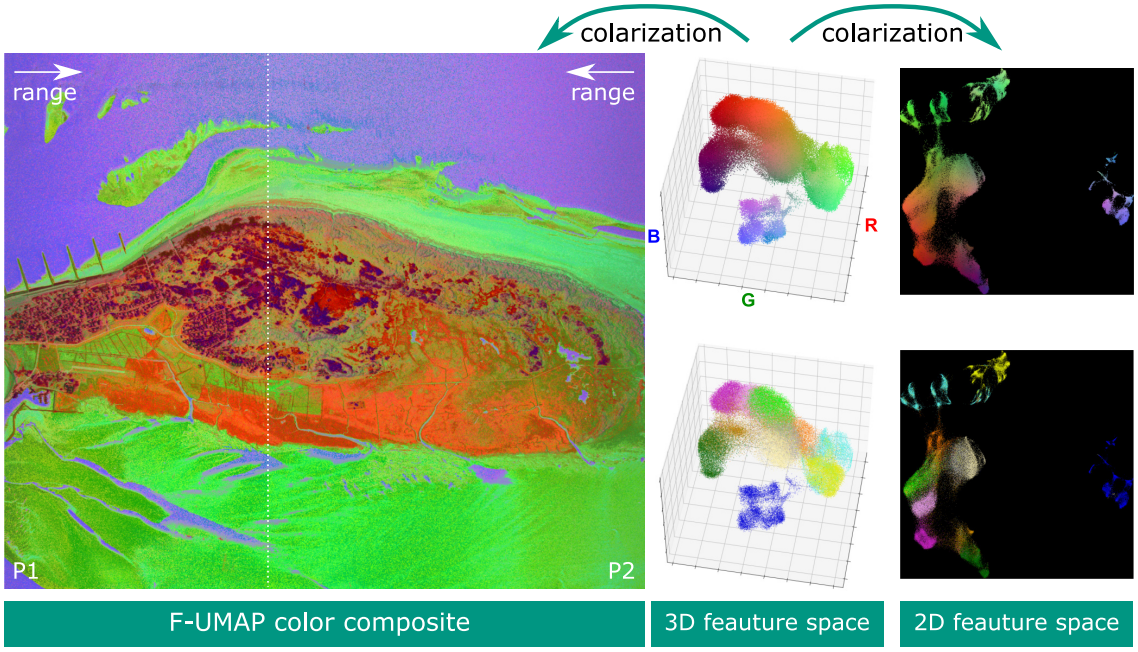
The final metric under discussion is Distribution Consistency, which evaluates the visual quality of the scatter plots (see Figure 4.4 and Figure 4.5). This metric serves as an indicator of the degree to which different classes overlap within the scatter plots. For Feature-based scatter plots, PCA and LE, closely aligned, show the poorest performance, indicative of substantial class overlaps. In comparison, UMAP and t-SNE achieve better results with comparable Distribution Consistency scores, with t-SNE particularly excelling when only small window sizes are considered. In the case of Wishart-based projections, the better class separability of the t-SNE embedding, already recognizable in the KNN classification, is also reflected in a higher Distribution Consistency compared to LE and UMAP. While the Distribution Consistency of Feature-based to Wishart-based LE projections improves significantly, the corresponding gain for UMAP and t-SNE is only small.

Overall, the results indicate that UMAP is well-suited for visualizing the data structure of PolInSAR data. In terms of preserving neighborhood structure and class separability, it surpasses PCA and LE methods and is comparable to t-SNE. The Wishart-based approach allows for the direct application of UMAP on PolInSAR coherency matrices, thereby avoiding the need for feature selection and extraction. The results show that the preservation of the neighborhood structure, the class separability, and the visual quality of the resulting scatter plots can be improved in this way.

4.5.2 3-dimensional Projections

The preceding analysis demonstrated that UMAP effectively preserves the essential information contained in multi-frequency PolInSAR data when projected into a low-dimensional feature space. Subsequently, F-UMAP and W-UMAP are employed to project data into a 3-dimensional feature space. The projections are the foundation for generating the color composites visualized in Figure 4.8. The color of each pixel reflects its corresponding position in the 3-dimensional feature space, which in turn correlates with specific land cover classes. The figure also presents the 3-dimensional feature space with points colorized according to their respective land cover classes to enhance the interpretability of these colors and their association with land cover classes. Considering that 3-dimensional scatter plots are more suited for interactive visualization rather than static, Figure 4.8 further includes corresponding 2-dimensional scatter plots: one with points colored based on their position in the 3-dimensional feature space and another with points colored by their respective land cover classes. This makes it easy to identify, for example, that in the F-UMAP color composite, the *Water* class appears in light violet tones, the *Tidal flat* class in green tones, and the *Peat bog* class in red tones.

The presented color composites provide a direct visual method for analyzing the distribution of various land cover classes and enhance the capability to discern subtle transitions between classes. Unlike traditional classification, where pixels are assigned a fixed class, this visualization approach facilitates the identification of fuzzy borders where one land cover class gradually transitions into another, e.g., from *Coastal shrub* colored in red tones to *High vegetation* colored in purple tones. The presented color composites capture a broad spectrum of information encoded in the PolInSAR coherency matrix. Consequently, this approach substantially enhances visual class separability compared to visualizations incorporating only three selected features. This enhancement is exemplified in Figure 4.9 by contrasting the F-UMAP visualizations with Pauli color composites of the scene. All vegetation types in the Pauli color composites appear bright green due to high backscatter intensities and a strong volume scattering component. In contrast, the UMAP color composite shows clear color differences between various vegetation types. Most notable is the difference in visual separability of the classes *Peat bog* and *High vegetation*. In the Pauli color composite of S-band data, visual separation is not possible, and in the Pauli color composite of L-band data, only barely recognizable color differences appear. In contrast, in the UMAP color composite, the red color representing *Peat bog* distinctly stands out from the purple tone characterizing *High vegetation* class, and the class borders can be easily detected. *High vegetation* can be also separated from *Coastal shrub* in the UMAP color composite. However, the color tone of *Coastal shrub* and *Peat bog* are similar, indicating that scattering properties are similar, which may also cause difficulties in automatic classification. A further improvement of the UMAP color composite compared to the Pauli color composites is the distinction between *Water* (violet) and *Sand* (green) (see upper right corner of Figure 4.9). The improved visual separability, compared to the



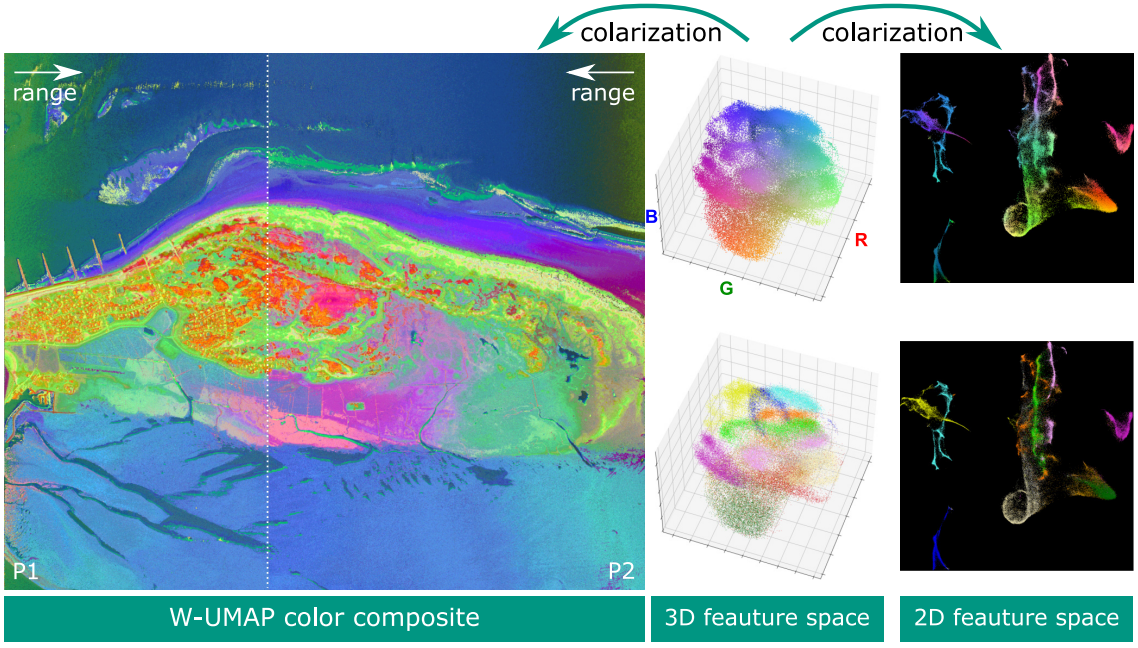
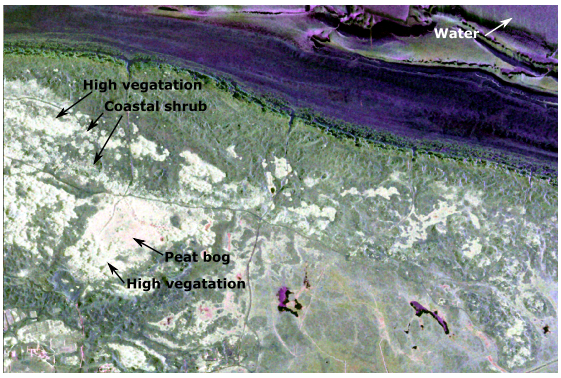
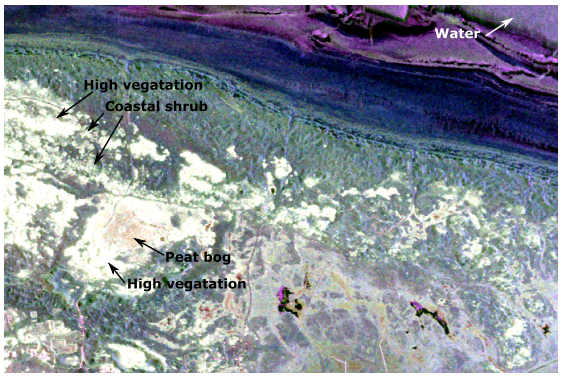


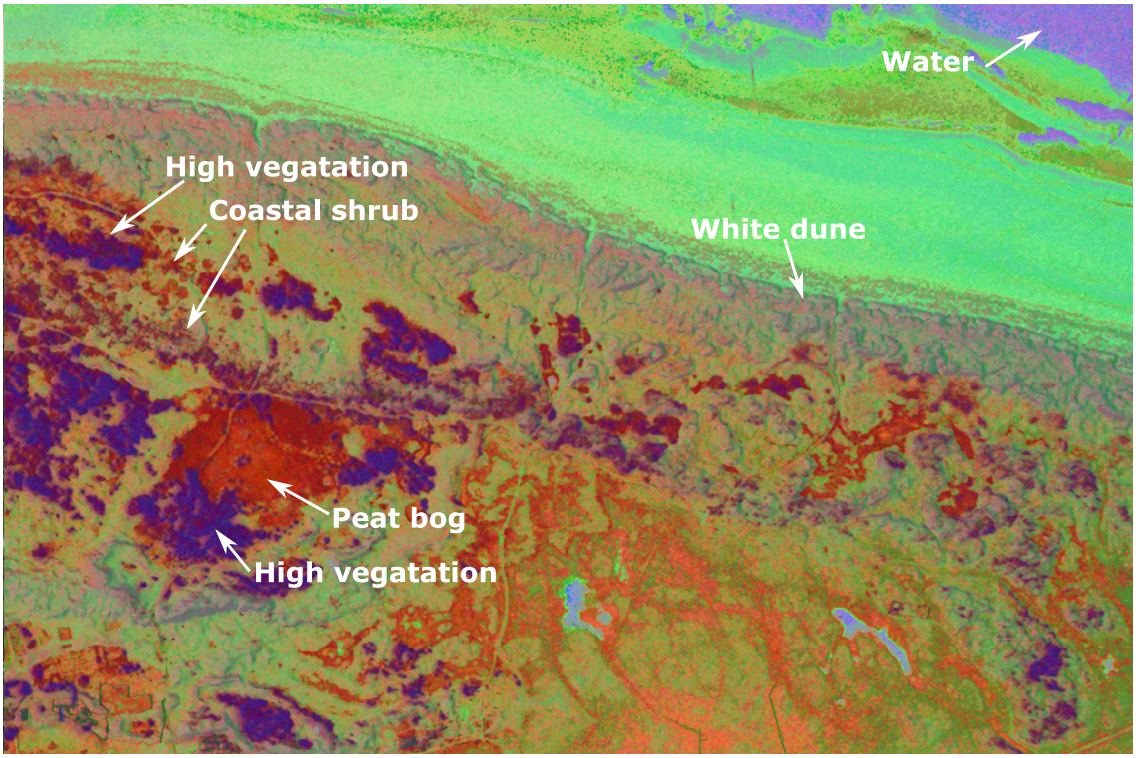
Figure 4.8.: RGB color composites resulting from 3-dimensional projections using F-UMAP (top) or W-UMAP (bottom). The color of each pixel reflects its position in the 3-dimensional feature space (middle).



(a) Pauli color composite (S-band).



(b) Pauli color composite (L-band).



(c) F-UMAP color composite.

Figure 4.9.: Comparison between Pauli color composite based on L-band data shown in (a) and F-UMAP color composite (b).

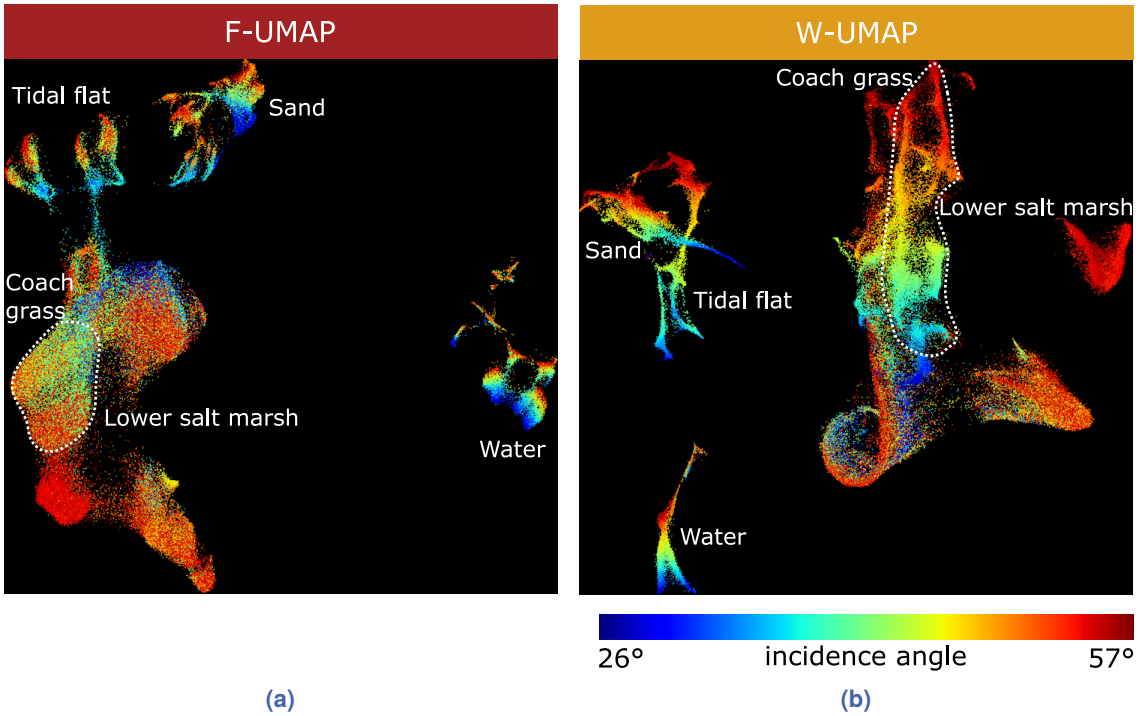


Figure 4.10.: Influence of the incidence angle on UMAP embeddings. The points of the 2-dimensional F-UMAP (a) and W-UMAP (b) embeddings are colored according to their incidence angle.

Pauli color composite, can be attributed to the inclusion of interferometric coherences. Due to the temporal variability of backscattering, there is low coherence over water areas and high coherence over *Tidal flat* and *Sand*.

When comparing the visual appearance of the F-UMAP with the W-UMAP color composite (Figure 4.8), differences can be observed regarding class-internal color gradients along the range direction. While these gradients are only slightly present in the F-UMAP color composite over the sea (north of the island) and over tidal flat (south of the island), strong range-dependent color changes appear in the W-UMAP image over the southern part of the island covered by *Lower salt marshes* and *Couch grass*, transitioning from light green in the near range to pink in the far range.

To further investigate this observation, Figure 4.10 illustrate the impact of the incidence angle on the 2-dimensional F-UMAP and W-UMAP embedding. Notably, the classes *Water*, *Tidal flat*, and *Sand* demonstrate a clear internal cluster organization by incidence angle in both the F-UMAP and W-UMAP scatter plots. As explored in depth in the study [78], this structuring is primarily attributed to the dependence of backscattering intensity on the incidence angle for scatterers dominated by surface scattering, which includes the aforementioned classes. Since surface backscattering intensity is reflected in polarimetric features (e.g., $\langle t_{11} \rangle$, $\langle t_{22} \rangle$, P_s, λ_i) used as input for F-UMAP as well as in elements of the coherency matrix (e.g., $\langle t_{11} \rangle$, $\langle t_{44} \rangle$) used as input for W-UMAP, both embeddings are affected. In contrast, for the classes *Lower salt marsh* and *Coach grass*, the incidence angle

dependencies exhibit differences between F-UMAP and W-UMAP. While the F-UMAP scatter plot shows no strict ordering by incidence angle within these classes, the W-UMAP scatter plot reveals a clear ordering of points corresponding to the incidence angle. Obviously, the incidence angle variation is reflected more strongly in the Wishart distance than in the Euclidean distance of extracted features. One possible explanation is that in the F-UMAP approach, also range invariant polarimetric features (e.g., H, p_i, ρ) are included. Furthermore, unwrapped phases are used in the F-UMAP approach, where the height of ambiguity, which changes with the incidence angle (see Equation 2.62), has less influence than on the wrapped phase contained in the coherency matrix. This is because one fringe interval $[-\pi, \pi]$ corresponds to changes in the unwrapped phase that are small compared to the entire range interval of the unwrapped phase.

Figure 4.11 shows that UMAP color composites not only reflect differences in land cover but also reflect topographic features of the scene. The figure compares Pauli, F-UMAP and W-UMAP color composites of a selected region. Additionally, a Digital Surface Model (DSM) shaded relief is included to visualize the actual scene topography based on data from an Airborne Laser Scanning (ALS) measurement conducted one day after the SAR acquisition.

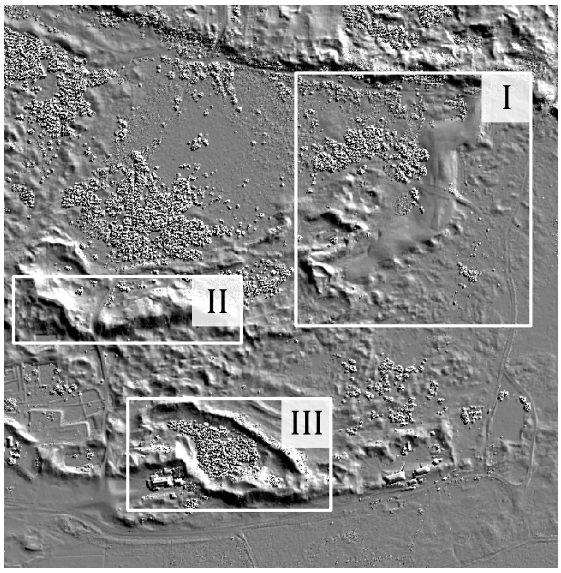
Three regions of interest (ROI), each exhibiting elevations due to dunes, are highlighted. The elevation in region I is distinctly visible only in the W-UMAP color composite due to its clear color separation from the surrounding flat area. The dunes in regions II and III do not stand out in the Pauli color composite, whereas they are recognizable in both the F-UMAP and W-UMAP color composites. In the F-UMAP color composite, these dunes are characterized by bright pink and light violet tones. In the W-UMAP color composite, the dune crests appear in bright orange and yellow tones, making them easily identifiable. The ability of UMAP color composites to depict topographic features obtained from interferometric measurements enhances visual differentiation of subtle landscape variations and thus contributes to a more complete understanding of the scene.

4.6 Applications

The UMAP-based visualization enables the analysis of the data structure using scatter plots as well as image-based scene analysis. In the following, example applications are given for both types of analysis.

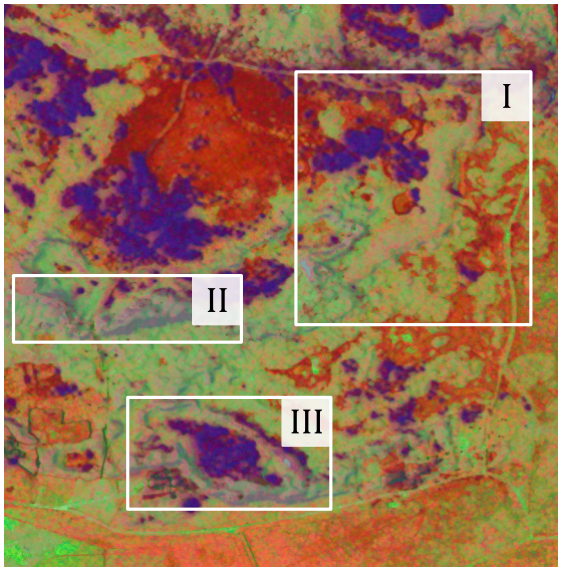
4.6.1 2-dimensional Projections

Figure 4.12 visualizes the data structure of single- and multi-frequency PolSAR (top row) and PolInSAR (bottom row) data using F-UMAP. As before, the Pol-InSAR-Island

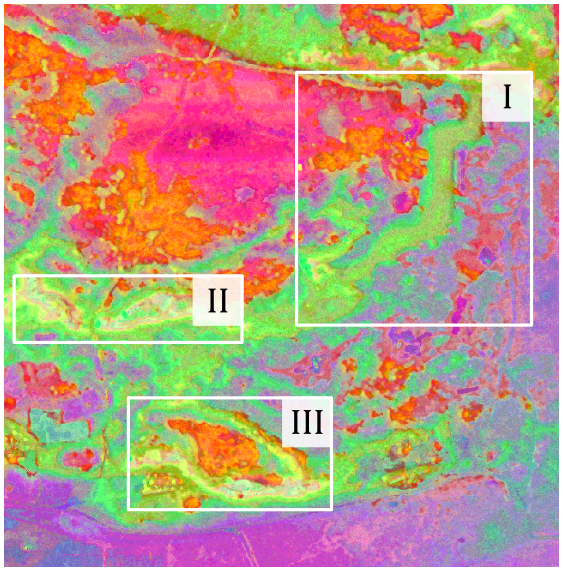


(a) DSM shaded relief.

(b) Pauli color composite (S-band).



(c) F-UMAP color composite.



(d) W-UMAP color composite.

Figure 4.11.: Visibility of topographic features: (a) shows the DSM, (b) the Pauli color composite, (c) the F-UMAP color composite and (d) the W-UMAP of a section of the scene.

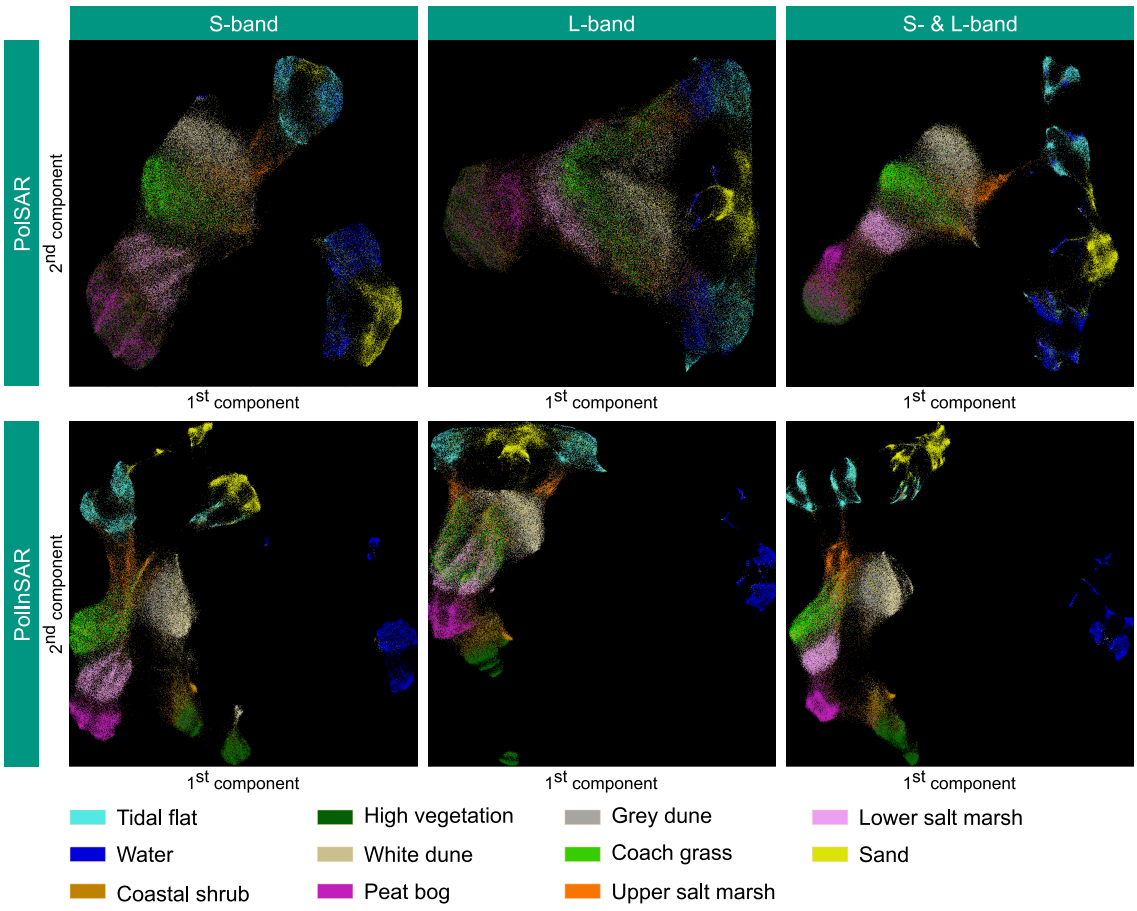


Figure 4.12.: Embedding results of F-UMAP for single- and multi-frequency PolSAR and PolInSAR data.

dataset forms the data basis. The features listed in Table 4.2 are used as input for the UMAP algorithm, whereby the interferometric features are omitted for the PolSAR data representation. The comparison of the depicted scatter plots shows how the addition of observables (i.e. combination of frequency bands and addition of interferometric features) affects the class separability. The visualization of the PolSAR S-band data indicates that no reliable separation of the classes *Coastal shrub*, *High vegetation* and *Peat bog* is possible. While a better separation of the class *Peat bog* is recognizable for the L-band data, there is a strong overlap of the classes *White dune*, *Grey dune*, *Lower salt marsh* and *Upper salt marsh* as well as the classes *Coastal shrub* and *High vegetation*. The combination of both frequency bands results in visually better separable clusters. The plot suggests that despite the improvement, the separation of the classes *Coastal shrub*, *High vegetation* and *Peat bog* using multi-frequency PolSAR data is still challenging. The visualization of the PolInSAR data shows that the addition of interferometric features achieves a further improvement. Furthermore, it can be seen that for PolInSAR data, the improvement potential by combining the frequency bands is lower than is the case for PolSAR data. This can be seen from the slight change in the class cluster based on single-frequency S-band or multi-frequency (S- and L-band) PolInSAR data.

4.6.2 3-dimensional Projections

Transfer Projection. To assess the transferability of learned UMAP embeddings to new data of a geographically distant region, UMAP projections trained on a subset of points from the Pol-InSAR-Island dataset are applied to a test scene on the neighboring island Langeoog, which exhibits similar land cover characteristics. The results are shown in Figure 4.13. The top section displays the previously established F-UMAP and W-UMAP color composites for Baltrum, while the bottom section shows the color composites for the test scene on Langeoog. The DSM and RGB ortho image displayed on the right-hand side serve as references for land cover and topography. Examination of these references reveals that the color scheme is consistently reproduced in the test scene, meaning identical land cover types are represented by the same colors in both the training and test areas. This consistency is evident in both the F-UMAP and W-UMAP approaches. For example, in the F-UMAP color composites, high vegetation is shown in purple and salt marshes in orange; in the W-UMAP color composites, high vegetation is displayed in red and elevated areas in bright colors within training and test areas.

In conclusion, the UMAP-based visualization approach facilitates coherent visual representations for data acquired with the same sensor across different locations. This is particularly beneficial for continuous monitoring, where new data must be consistently integrated and represented alongside existing data.

Visualizing Tidal Flats. The proposed visualization approach is particularly valuable in use cases for which PolInSAR data contains information that cannot be obtained from the visual analysis of optical data. One example of this type of application is the observation of tidal flats, which is presented in Figure 4.14 using a scene south of the East Frisian island of Spiekeroog. The underlying PolInSAR data were acquired in a measurement campaign in 2019. In contrast to the Pol-InSAR-Island dataset, polarimetric data were exclusively acquired in S-band during this survey. Interferometric processing (based on single- and repeat-pass measurement) was only carried out in VV polarization. Therefore, the feature set considered for the UMAP-based visualization (which can be found in [79]) slightly differs from the previously used feature set.

Tidal flats are made up of different types of mudflats, which differ in terms of their surface structure, grain size, and water content. As can be seen in Figure 4.14b, these properties cannot be distinguished by color using an optical RGB image. Even in the Pauli color composite shown in Figure 4.14c, there are only slight color differences over the entire area covered by tidal flats and water. In contrast, in the F-UMAP color composite in Figure 4.14d, these areas appear in different colors, indicating their different properties. Water-covered areas appear in dark blue, and the tidal flat area is divided into easily separable green-colored areas near the water channels and turquoise areas. The pink colored areas, which

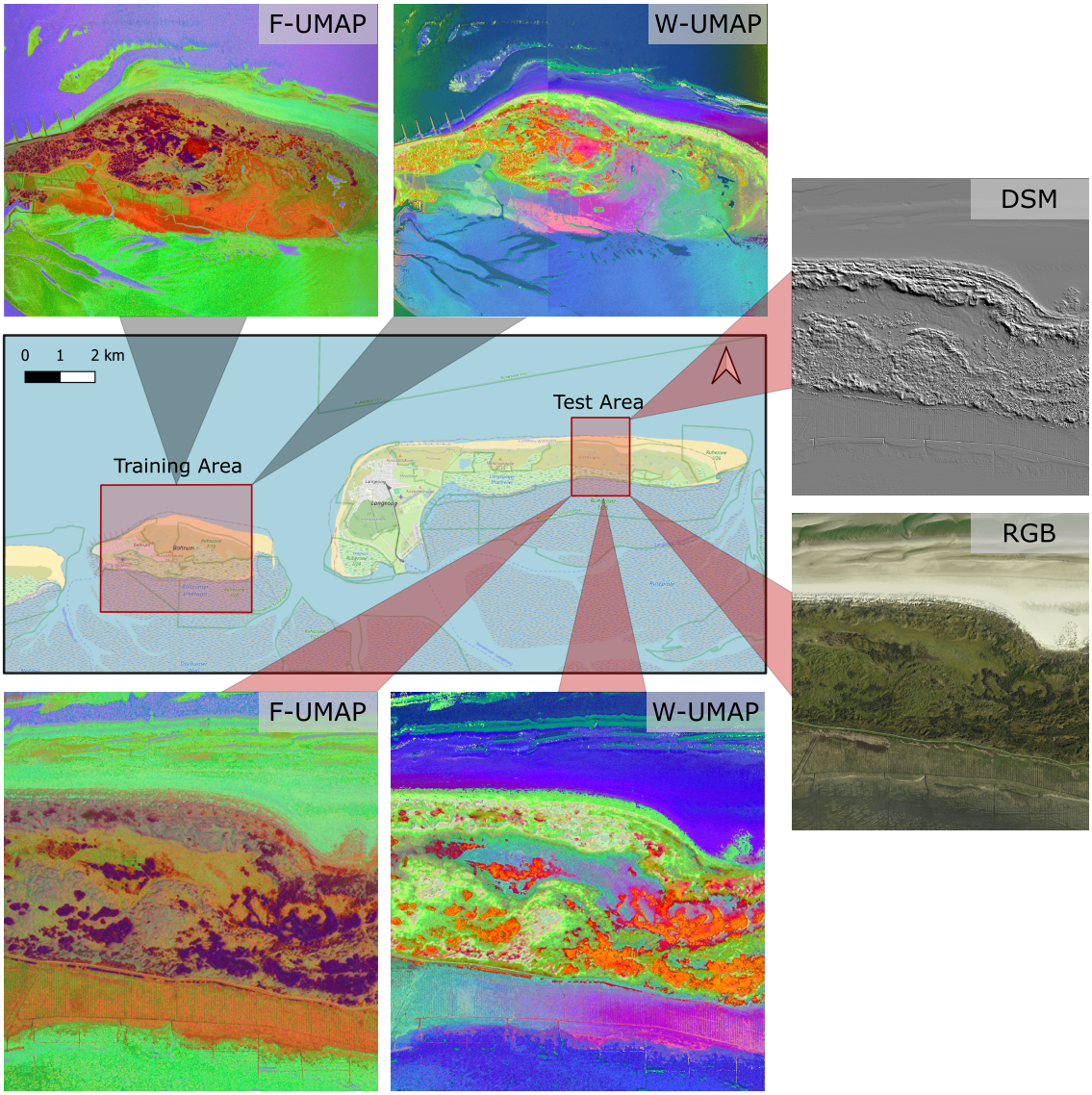
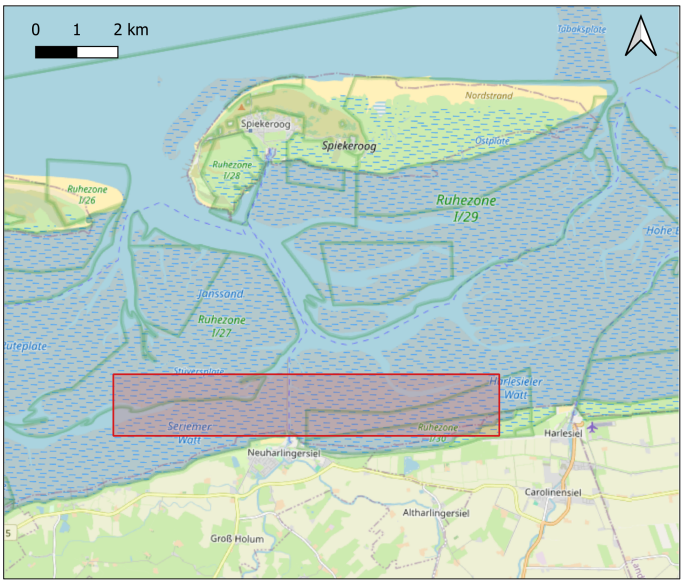
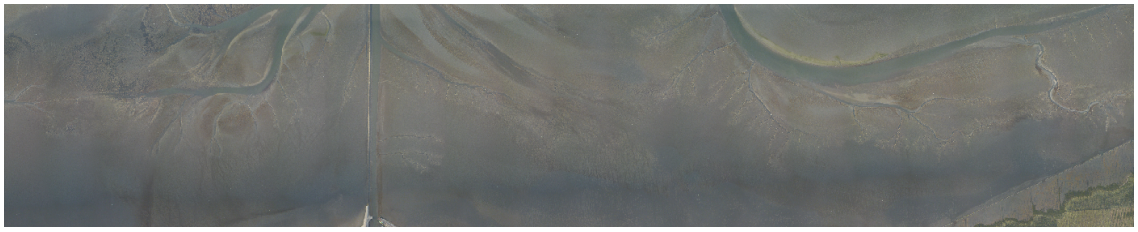


Figure 4.13.: The central map (OSM [187]) illustrates the geographic locations of interest. UMAP projections, trained on data from Baltrum (training area), are applied to a test area on Langeoog. The top section displays UMAP-based color composites of Baltrum, including training data points. The bottom section presents UMAP-based color composites for previously unseen data from Langeoog. On the right, the optical image (RGB) and the DSM based on ALS data are shown.

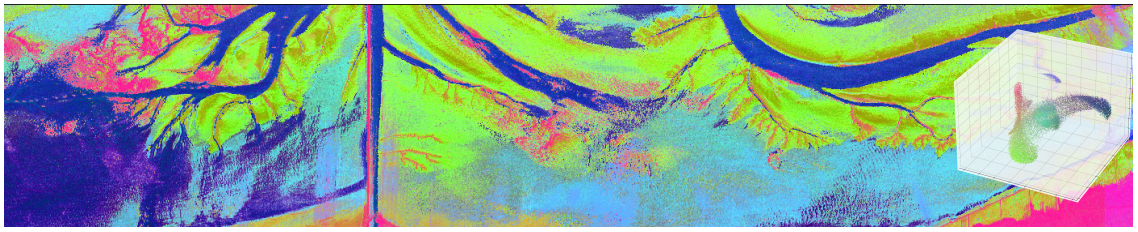


(a) Geographic location of sample scene (Basemap taken from OpenStreetMap [187]).



(b) Optical image (RGB).

(c) Pauli color composite.



(d) F-UMAP color composite.

Figure 4.14.: Visualization of PolInSAR data (S-band) acquired over a tidal flat area. The geographic location is shown in (a). An optical image of the scene, acquired on the same day as the SAR data, is presented in (b). Visualizations of the PolInSAR data are generated by (c) Pauli decomposition and (d) F-UMAP.

are also conspicuous in the Pauli color composite due to high backscatter intensities, represent mussel beds.

4.7 Discussion

This work explored the application of the dimension reduction method UMAP for visualizing multi-frequency PolInSAR data. Despite the large body of research surrounding the automated analysis of PolInSAR data, the domain of visualizing this data to enable human-driven visual analysis remains largely unexplored. This work addressed this gap by developing two distinctive UMAP-based visualization methodologies: F-UMAP and W-UMAP. The evaluation of these approaches was guided by two main questions:

1. How does UMAP compare to other dimension reduction methods in terms of creating a visualizable representation of multi-frequency PolInSAR that effectively preserves the data structure and class separability?
2. Which of the two methods, F-UMAP or W-UMAP, better preserves data structure and class separability?

UMAP Performance Evaluation. UMAP was found to outperform PCA and LE methods in preserving data structure and class separability. Its performance aligns closely with that of t-SNE, a leading approach for high-dimensional data visualization. The conclusion that UMAP is more suited for visualizing PolInSAR data compared to t-SNE is primarily due to UMAP's ability to project new data into a learned low-dimensional space and its more intuitive handling of hyperparameters. These outcomes align with findings from other domains, including single-cell analysis [28] and mass spectrometry analysis [27], that highlights UMAP's superior potential for visualizing multi-dimensional and large-scale datasets.

Comparing F-UMAP and W-UMAP. The F-UMAP approach aims to summarize a wide range of expert-designed features in a low-dimensional representation. On the other hand, W-UMAP applies the UMAP algorithm directly to the complex-valued PolInSAR coherency matrices, bypassing the need for prior feature selection and extraction. Each method offers unique advantages. F-UMAP allows for incorporating prior knowledge from physical models and the preselection of features, which may be beneficial when a specific application requires a known set of informative features. In contrast, W-UMAP is more universally applicable, free from biases of manual feature selection, and immune to potential artifacts that may occur during feature extraction.

The quantitative evaluation of neighborhood preservation indicates that the original neighborhood structure, based on the Wishart distance, is more accurately represented in a 2-dimensional scatter plot compared to the neighborhood structure of high-dimensional feature representations. However, it should be noted that for F-UMAP, the input dimension is 84, corresponding to the number of features, while for W-UMAP, the original dimension is only 42, aligning with the free parameters defining the coherency matrices of S- and L-band data.

Comparing class separability preservation, as measured by KNN classifier results, both F-UMAP and W-UMAP demonstrate good performance. There is minimal or no loss between the classification of original high-dimensional and embedded low-dimensional representations. This confirms that UMAP-based visualizations accurately represent class separability that can be achieved on the given PolInSAR data, aiding in identifying potential application areas. In a direct comparison of absolute values from KNN classification results, W-UMAP outperforms F-UMAP. This performance difference is attributed to the fact that SRWD between coherency matrices is better suited for distance-based classification than the Euclidean distance between extracted feature vectors. This assumption is supported by evidence from other research studies investigating various distance measures for PolSAR classification [172, 188].

Despite W-UMAP's superior quantitative results, a shortcoming compared to F-UMAP is evident in the color composite that reveals a stronger influence of the incidence angle. As a result, the same land cover classes appear in different colors, depending on whether they are captured in the near or far range.

Advantages. Overall, the quantitative and qualitative results show that UMAP is a very well-suited tool that can reliably represent and visualize the data structure of a multi-frequency PolInSAR dataset. One key advantage is the easy visual identification of clusters within the complex data, which represent different types and characteristics of the land cover. As shown in Section 4.6, data visualization through UMAP enables an intuitive assessment of the added value of additional frequency bands and interferometric information. This supports the development of more precise classification methods by targeting relevant features and aids in the strategic planning of sensor deployment and acquisition methods for future campaigns. The knowledge gained can thus contribute to optimization and increased efficiency in both automated data analysis and data acquisition.

Visualizing multi-frequency PolInSAR data as a color composite image enhances human-driven visual scene analysis. Unlike color composites that rely on three selected features (e.g., Pauli color composite), UMAP color composites incorporate the entirety of data information, thereby providing a more comprehensive representation of the data. As demonstrated in Section 4.6, these composites enable the identification of potential applications such as differentiating tidal flat types. Furthermore, the ability to transfer a

learned projection to unseen PolInSAR image data, as demonstrated in Section 4.6, enables consistent evaluation across different spatial and temporal scales, which is a key feature for continuous monitoring tasks.

Limitations. A limitation of the study design is the use of a single dataset from a single location and sensor for evaluating the proposed approaches. As discussed in Section 3.4, this is due to the scarcity of freely available labeled PolInSAR datasets beyond the Pol-InSAR-Island dataset. This limitation restricts the generalizability of this study's findings. Nevertheless, the proposed approach for data visualization does not rely on any sensor- or location-specific assumptions, ensuring its applicability to additional datasets.

Another point that should be mentioned is that UMAP was used as a visualization tool to support the class aggregation in the generation of the Pol-InSAR-Island dataset. This potentially induces a bias that affects the evaluation of label-based quality metrics (KNN classification and Distance Consistency).

Considering the UMAP-based visualization approach itself, a disadvantage is the loss of physical interpretability of the feature dimensions. Unlike Pauli color composites, which correlate pixel colors to dominant scattering mechanisms, the color in UMAP color composites do not directly reveal scattering properties. Nevertheless, it is possible to analyze the relationship between physically interpretable features and the UMAP color assignment. The points in the scatter plot can be colored according to certain feature values to illustrate the influence of individual features on the data structure. This is even possible in the absence of the underlying original data, as the UMAP projection is invertible, i.e. individual feature values can be approximately reconstructed using the low-dimensional representation.

Future Work. Future work should explore several promising avenues to enhance the proposed UMAP-based visualization of PolInSAR data. To better exploit the information potential resulting from the combination of polarimetry and interferometry, the input of the F-UMAP method can be extended by further PolInSAR-specific features, such as coherence shape parameters proposed in [189] or polarimetric–interferometric model-based decomposition proposed in [83].

A further improvement of the proposed method can be realized by replacing the RGB color space with a perceptually uniform color space, such as the CIEL*a*b color space, for translating positions in the feature space into color. The advantage of utilizing the CIEL*a*b color space lies in its ability to represent distances in a manner that better aligns with the non-linear human visual system. Thus, distances in feature space are accurately reflected by perceived color differences. This may lead to better visual separability of land cover classes that correspond to neighboring clusters in the feature space.

5 CNN-based PolInSAR Image Classification

The test setup for comparing real-valued representations of PolSAR and PolInSAR data for CNN-based classification (including model architecture and model training) described in this chapter essentially follows the concept published in:

[190] Sylvia Hochstuhl, Niklas Pfeffer, Antje Thiele, Horst Hammer, and Stefan Hinz. “Your input matters—Comparing real-valued PolSAR data representations for CNN-based segmentation”. In: *Remote Sensing* 15.24 (2023), p. 5738

Adapted text passages are marked with a green line.

5.1 Introduction

Multi-frequency PolInSAR data provides a wealth of information, making it a uniquely powerful tool for Earth observation. The combination of polarimetric and interferometric measurements across multiple frequencies enables the derivation of detailed information about geophysical parameters, vertical structure, and topography of an observed scene. This enables the detailed differentiation of different vegetation types and other land covers in complex ecosystems. The challenge in the automated classification of this data is to extract the relevant information needed for a specific task. While learning-based land cover classification for PolSAR data is a well-explored research topic, the learning-based classification of multi-frequency PolInSAR data is much less considered in the literature.

The first approaches for land cover classification of PolSAR data are mainly based on statistical modeling. The Wishart classifier is a prominent representative of these approaches [175]. In later works, machine learning models such as Random Forest [13–16] and SVM [17–19] are used, which classify previously extracted, physically interpretable features, partly supplemented by manually designed spatial features. This approach shows superior classification performance compared to statistical methods. More recently, progress has been made by applying deep learning models, particularly CNNs, which can autonomously learn spatial features and thus recognize and classify complex patterns in PolSAR image data. The advantage of CNNs over traditional machine learning models

such as Random Forest or SVM is that they can automatically learn task-specific spatial image features during training, often better suited for classification than heuristic, human-designed features. It has already been demonstrated [20, 21, 23] that this advantage can lead to an improvement in PolSAR image classification

Despite these advances in PolSAR classification, the potential of CNNs for multi-frequency PolInSAR data classification still needs to be explored. Existing work is primarily based on statistical modeling [191–193] or the use of traditional machine learning models [194], which do not incorporate the potential of learning and employing more complex image features for classification.

CNNs are developed initially to analyze optical RGB data, where three real-valued components represent each pixel. In contrast, the pixels of PolInSAR image data are defined by the complex-valued coherency matrix. To process this type of data using CNNs, complex-valued CNNs have been developed (e.g., in [24, 195]), which perform CNN operations in the complex domain. Despite their potential, complex-valued CNNs have some disadvantages. These include the increased computational and resource requirements due to complex arithmetic operations and the limited research activity on complex-valued CNNs due to their low demand in common data analysis applications. In contrast, real-valued CNNs are researched more intensively, which leads to faster development and improvement of models and optimization techniques. To exploit the potential of this research advantage for PolInSAR image analysis, the question of how to best represent the data in the real domain must first be addressed.

In CNN-based PolSAR classification, two fundamental types of real-valued data representations are commonly used as input. The first type directly represents each pixel through the nine independent real-valued parameters of the coherency matrix, which are stacked into a nine-channel image. The rationale behind this approach is that the coherency matrix completely describes the underlying scattering process within the resolution cell, and the CNN model should learn to model the relationship between the coherency matrix and the class labels directly. The second type of representation relies on including physically interpretable polarimetric features, which have been shown to provide discriminative information for land cover classification. It can be assumed that CNNs cannot automatically learn this type of features, which require knowledge about physical processes. Thus, classification performance could be improved by feeding those features directly into CNNs.

Both approaches are implemented in varying forms in the CNN-based classification of PolSAR data. A comprehensive comparative analysis of which type of data representation is best suited to serve as CNN input can hardly be found in the literature. Most studies focus on analyzing different CNN architectures so that only one type of input data representation is usually used, which is often chosen without justification. However, the

chosen approach to represent PolSAR data by a real-valued representation is decisive for preserving information and ultimately affects the classification performance of a CNN.

This thesis, therefore, aims to identify the most suitable real-valued representation for CNN classification. In particular, it addresses whether CNNs can directly capture the relationship between PolSAR or PolInSAR data represented by the coherency matrix and the target label or whether prior extraction of expert-designed, physically interpretable features is beneficial.

Especially for the latter case, another challenge arises. To fully capture the information of PolInSAR data by real-valued features, a combination of several different polarimetric and interferometric features is required, the number of which doubles when data from different frequency bands are combined. Previous studies using traditional machine learning models (e.g., SVM) have shown that high-dimensional representations containing not only complementary but also redundant features increase the complexity of the classifier (and thus the computational cost), increase the risk of overfitting, or even lead to a loss of performance [196–200]. Whether CNNs, in contrast, can manage high-dimensional input data and selectively focus on relevant features remains to be analyzed. Additional mechanisms that learn and exploit the interrelation between feature channels might be necessary to equip CNNs with this capability.

A further approach that simplifies the handling of high-dimensional data is using dimension reduction methods that are generally categorized into automatic feature selection methods, which identify an optimal subset of features, and feature transformation methods, which project high-dimensional data into a low-dimensional space. Both strategies aim to maximize the preservation of relevant information and eliminate redundancies. Using low-dimensional data representations reduces resource consumption and can improve data understanding (e.g., by evaluating the importance of individual features) and the traceability of classification results (as it simplifies the analysis of input data). While dimension reduction has been shown to be effective for Random Forest and SVM classification for PolSAR and PolInSAR classification [197, 201, 202], its application to CNN-based classification remains under-researched.

To fill these research gaps, the following questions are addressed in this thesis: Are CNNs capable of effectively processing multi-channel images that contain both complementary and redundant information, or are there:

- specific CNN architectures tailored for cross-channel feature extraction or
- dimension reduction methods

that can enhance classification performance and efficiency?

To investigate these aspects, different approaches are compared in the thesis: Feature stacking, where multi-channel images are constructed from polarimetric and interferometric

features as CNN inputs; feature stacking in combination with CNN-internal cross-channel feature extraction that highlights relevant features; and the application of dimension reduction, either by automatic feature selection or feature transformation, before or within CNN classification.

After identifying the optimal representation of multi-frequency PolInSAR data for CNN-based classification and determining the best strategies for processing high-dimensional feature representations, the question of whether CNNs provide better classification performance than statistical methods and traditional machine learning models are addressed.

5.2 Fundamentals

The following describes the fundamentals of the classification approaches used in this thesis. This includes the description of the Wishart classifier [175], the basics of feature selection, commonly used as a preprocessing step in machine learning approaches, and the presentation of the traditional machine learning model Random Forest and the deep learning model CNN.

5.2.1 Wishart Classifier

The supervised Wishart classifier, proposed by Lee et al. in [175], is a statistical method widely used for the classification of PolSAR data. This classifier leverages the complex Wishart distribution (see Equation 4.8 in Section 4.3.3) to model the statistical properties of the coherency matrix. This method has been extended to PolInSAR data in [192] to incorporate the additional interferometric information.

In a supervised classification context, the aim is to assign each pixel, represented by its coherency matrix $\langle \mathbf{T} \rangle$, to one of the predefined classes based on the maximum likelihood criterion. The likelihood of a pixel x_i belonging to class c with a corresponding mean coherency matrix $\langle \mathbf{T} \rangle_c$, is:

$$p(\langle \mathbf{T} \rangle_i | c) = \frac{|\langle \mathbf{T} \rangle_i|^{n-q} \exp\left(-n\mathrm{tr}(\langle \mathbf{T} \rangle_c^{-1} \langle \mathbf{T} \rangle_i)\right)}{|\langle \mathbf{T} \rangle_c|^n \Gamma_q(n)}, \tag{5.1}$$

where q is the dimension of the coherency matrix (i.e, $q = 3$ for PolSAR and $q = 6$ for PolInSAR data), n is the number of looks, $\Gamma_q(n)$ is the multivariate Gamma function, $\mathrm{tr}(\cdot)$ denotes the matrix trace, and $|\cdot|$ denotes the matrix determinant. The mean coherency matrices $\langle \mathbf{T} \rangle_c$ for each target class c are estimated from the training data. These matrices are calculated as the average of the sample coherency matrices of all training pixels x_i that belong to that class:

$$\langle \mathbf{T} \rangle_c = \frac{1}{N_c} \sum_{x_i \in \text{class } c} \langle \mathbf{T} \rangle_i, \tag{5.2}$$

where N_c is the number of training pixels in class c, and $\langle \mathbf{T} \rangle_i$ is the covariance matrix of the pixel x_i that belongs to class c.

To classify a pixel, the log-likelihood is used due to numerical stability:

$$\ln p(\langle \mathbf{T} \rangle_i | c) = (n-q) \ln |\langle \mathbf{T} \rangle_i| - n \ln |\langle \mathbf{T} \rangle_c| - n \text{tr}(\langle \mathbf{T} \rangle_c^{-1} \langle \mathbf{T} \rangle_i) - \ln \Gamma_q(n). \tag{5.3}$$

Since the terms $(n-q) \ln |\langle \mathbf{T} \rangle_i|$ and $\ln \Gamma_q(n)$ are independent of the class c, they can be omitted when comparing the likelihoods for different classes. The relevant terms for the classification are:

$$\ln p(\langle \mathbf{T} \rangle_i | c) \propto - \ln |\langle \mathbf{T} \rangle_c| - \text{tr}(\langle \mathbf{T} \rangle_c^{-1} \langle \mathbf{T} \rangle_i). \tag{5.4}$$

The pixel x_i is assigned to the class c that maximizes this log-likelihood:

$$c^* = \arg\max_c \left(- \ln |\langle \mathbf{T} \rangle_c| - \text{tr} \left(\langle \mathbf{T} \rangle_c^{-1} \langle \mathbf{T} \rangle_i \right) \right). \tag{5.5}$$

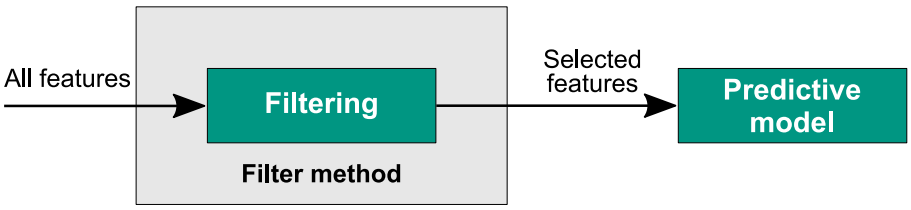
For multi-frequency PolInSAR data, the classification process can be extended based on the joint probability density function of the coherency matrices from multiple frequency bands. The joint likelihood of a pixel belonging to a specific class is the product of the likelihoods for each frequency band. Taking the logarithm of the joint likelihood results in the sum of the log-likelihoods. Therefore, the pixel x_i is assigned to the class c that maximizes this summed log-likelihood over F frequency bands:

$$c^* = \arg\max_c \sum_{f=1}^{F} \left[- \ln |\mathbf{T}_c^{(f)}| - \text{tr}(\mathbf{T}_c^{(f)-1} \mathbf{T}_i^{(f)}) \right]. \tag{5.6}$$

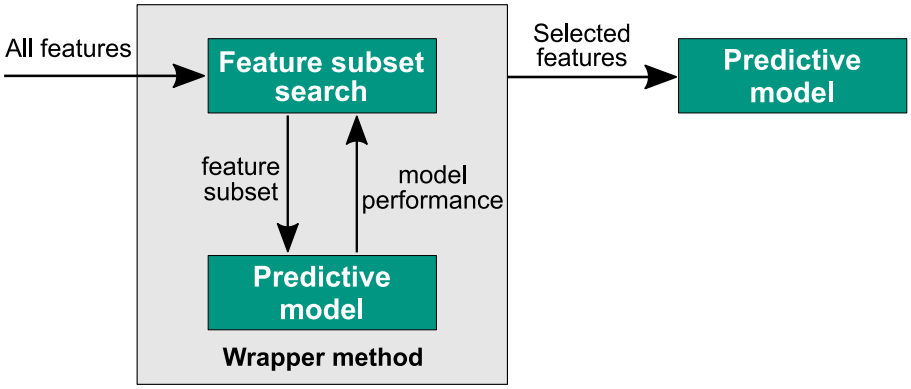
5.2.2 Feature Selection

A typical machine learning pipeline for the classification of PolSAR or PolInSAR data includes extraction of polarimetric (and interferometric) features, an optional step performing dimension reduction through automatic feature selection or feature transformation, followed by a classification model that needs to be trained using labeled data. The following introduces the basic concepts of the optional automatic feature selection step.

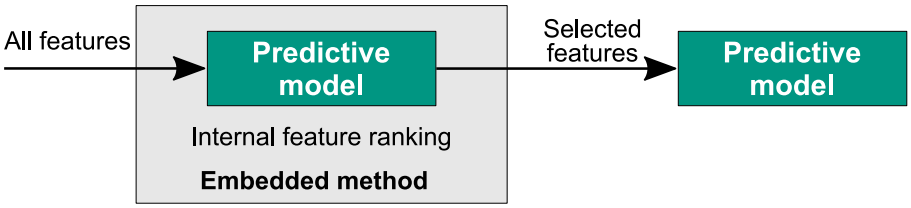
Feature selection, also called variable or attribute selection, is a crucial step in the learning-based analysis of high-dimensional data. It aims at selecting a discriminative subset of features for predictions or classification tasks. The main goal of feature selection is to eliminate irrelevant or redundant features and thus reduce the dimension of the data



(a) Filter method.



(b) Wrapper method.



(c) Embedded method.

Figure 5.1.: Schematic structure of (a) filter methods, (b) wrapper methods, and (c) embedded methods for feature selection.

to simplify and accelerate the training of learning-based models and increase model interpretability.

Depending on whether given target variables, such as land cover classes, are included in the feature selection process, a distinction is made between supervised and unsupervised methods. Since this thesis focuses on supervised classification methods, only supervised feature selection approaches are described in more detail. These can be divided into three main groups: filter, wrapper, and embedded methods [203]. The different functional principles of these three approaches are visualized in Figure 5.1.

Filter methods evaluate the relevance of the features independently of the subsequent prediction model used and are based on distance, consistency, similarity, or statistical measures. Features are evaluated and filtered either independently of each other (univariate filter methods) or as a feature subset (multivariate filter methods). The evaluation criteria of univariate methods, which have the advantage of being less computationally

intensive, include the χ^2-test [204], the Fisher score [205], Mutual Information (MI) [206], the Laplacian score [207] or Relief or ReliefF score [208]. Multivariate filter methods, such as the calculation of the generalized Fisher score [209], the minimum Redundancy Maximum Relevance (mRMR) method [210], or the fast correlation-based filter [211], were developed to include feature redundancies and feature interactions in the feature subset selection in addition to feature relevance. This allows potentially smaller and more suitable feature subsets to be found for prediction, but an increase in computational effort accompanies this. Both univariate and multivariate filter methods decouple feature selection and model training, generating model-agnostic feature subsets, which may prevent overfitting the overall system. However, it can be argued that ignoring the interaction between feature selection and classifier does not necessarily select the optimal feature subset for maximizing the prediction accuracy of the model [212].

In order to include the interaction between feature selection and classifier, wrapper methods can be used (see Figure 5.1b). These evaluate feature subsets based on the evaluation performance of a selected prediction model. Since testing all feasible feature combinations, which requires repeated training of the model, is impractical, feature subsets are selected using heuristic search methods. Frequently used search methods are sequential algorithms such as forward or backward selection or randomized search algorithms such as simulated annealing or genetic algorithms (a detailed overview can be found in [212]). Theoretically, any search strategy can be combined with any predictive model to select the best-performing feature subset. However, in practice, only the use of models that are fast to train is feasible due to the repeated model training.

Embedded methods integrate the feature selection step directly into the model training and thus aim to select model-specific features computationally efficiently. In this sense, embedded methods combine the advantages of filter and wrapper methods. The basic idea is to weigh the features during the training of the model according to their relevance for the predictive performance of the model. This is done by adding regularization terms to the cost function or using unique learning algorithms that inherently perform feature selection during training. Well-known representatives of the first approach are the Lasso regression [213], the bridge regression [214] and the Elastic Net algorithm [215]. An example of the second approach is the Random Forest algorithm [216].

5.2.3 Random Forest Classification

The Random Forest introduced by Breiman et al. [217] is a widely used algorithm in machine learning for classification and regression. It is an ensemble method combining several learning-based prediction models to generate a powerful, robust model.

The prediction models underlying the Random Forest are decision trees. A decision tree is a classifier that categorizes data using hierarchically structured decision rules. The

tree nodes represent individual logical rules that split the data; the tree leaves symbolize the decisions. The decision rules can either be defined manually or learned from labeled data.

For Random Forest classification, several decision trees are combined and trained to be as uncorrelated as possible. To ensure a low correlation between decision trees, they are trained independently of each other using randomly selected samples of the training data. These samples are generated by drawing with replacement, a process known as bootstrapping (or bagging). In addition, the random subspace method [218], also known as feature bagging, is used to train a decision tree. When learning individual decision rules, this involves considering only a randomly selected subset of all features.

The prediction of the Random Forest is obtained by majority voting on (for classification problems) or by averaging of (for regression problems) the predictions of the individual trees. This aggregation reduces the variance of the model, which improves its predictive power.

5.2.4 Convolutional Neural Networks

A CNN is a machine learning model that belongs to the group of deep learning models (see Figure 5.2) and has been developed to classify image data. Traditional machine learning models (sometimes referred to as shallow learning models), such as the Random Forest, usually use a two-stage approach for image classification, which consists of extracting hand-crafted features and subsequent classification. In contrast, the basic idea of deep learning is to automatically learn and classify suitable non-linear transformations to extract complex discriminative image features directly from given data. The underlying assumption is that this enables complex image patterns to be described in a way that is more suitable for machine evaluation than human-designed image descriptors.

The basic concept of CNNs was already developed in the 1980s and used in [219] to recognize handwritten digits. However, CNNs only gained prominent attention and importance in the mid-2000s, when higher computing capacities and efficient data processing became available thanks to advances in GPU development, making the application of CNNs practical.

CNNs are feedforward neural networks in which data is processed from the input layer through intermediate layers to an output representation without any feedback loop. In recent years, several different CNN model architectures have been developed, generally based on a sequence of convolutional, pooling, and activation layers followed by fully connected layers. The functionality of the operations performed in these layers is described below.

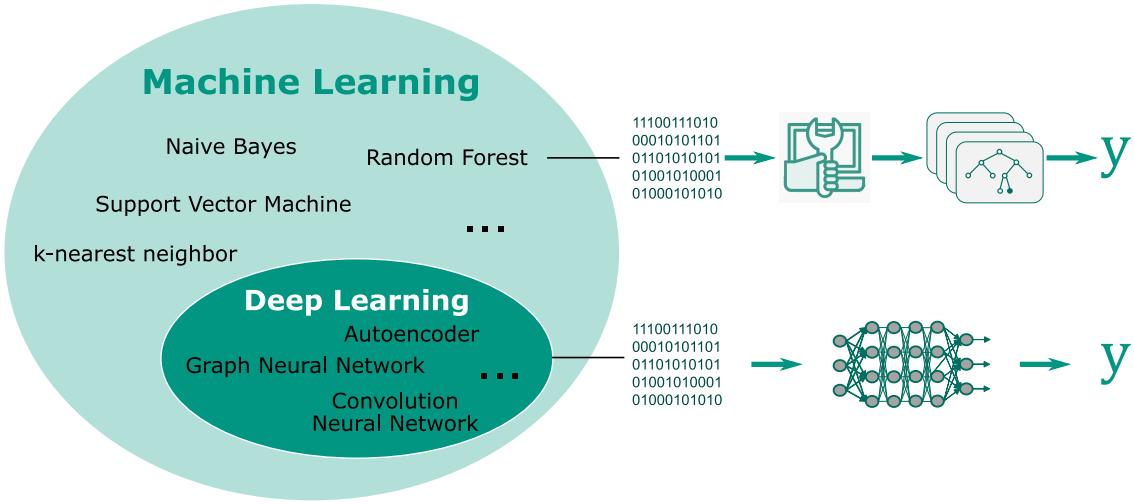
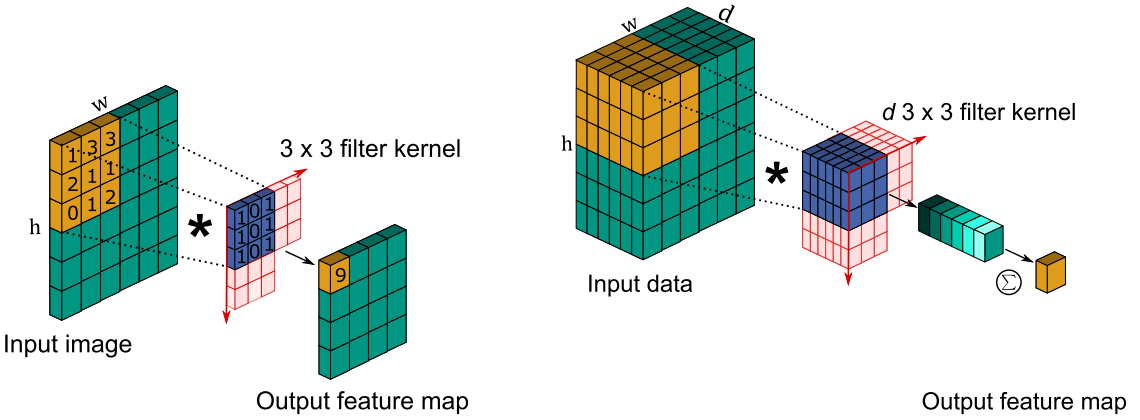


Figure 5.2.: Categorization of CNNs in the machine learning context.

Convolutional Layer. The convolutional layers are the main component of a CNN and are used for feature extraction by applying convolution operations to the input data. Although the concept of convolutional layers can be applied to 1-dimensional (e.g., time series), 2-dimensional (e.g., image data), and 3-dimensional (e.g., video data), it is explained below, focusing exclusively on 2-dimensional image data. As illustrated in Figure 5.3a, filter kernels, which typically have dimensions between 3×3 and 7×7, are shifted over the input image. At each point, a convolution is performed between the filter kernel $\mathbf{W}$ and the covered image region, which corresponds to the calculation of the weighted sum of the pixel values under consideration:

$$\hat{\mathbf{X}}_{i,j} = \sum_{k=1}^{m} \sum_{l=1}^{n} w_{k,l} \cdot \mathbf{X}_{i-k+a,j-l+b}, \tag{5.7}$$

where $\hat{\mathbf{X}}_{i,j}$ is the resulting filtered image pixel, $\mathbf{X}$ is the input image, a and b denotes the coordinates of the center point of the filter kernel and $w_{k,l}$ specifies the filter weight at the position (k,l) of the filter kernel, that has the dimension $m \times n$. The filter weights represent trainable model parameters that are adjusted during training. The filtered output image is referred to as a feature map. Typically, several filter kernels with different weights are applied within a convolutional layer, resulting in several feature maps. For a single-channel input (as illustrated in Figure 5.3a), the number of resulting feature maps is equal to the number of filter kernels. Figure 5.3b shows the concept of a convolution applied to a multi-channel input image. To filter an input image with d channels, d filter kernels of size $m \times n$ are applied to generate one feature map. The individual image channels are convolved independently with a specific filter kernel, and the resulting outputs are summed up to obtain a single feature map. Thus, to generate M feature maps based on a d-channel input image, $M \cdot d$ filter kernels of size $m \times n$ are required.



(a) Single-channel input. **(b)** Multi-channel input.

Figure 5.3.: Concept of a 2D convolution operation using a 3×3 filter kernel applied to (a) single-channel input and (b) multi-channel input.

Activation Layer. Before the resulting feature maps are further processed, an activation function $f(\cdot)$ is applied, which performs a non-linear transformation of the data that enables the learning of non-linear and, thus, complex features. A very commonly used activation function is Rectified Linear Unit (ReLU), defined by:

$$f(x) = \max(0, x). \tag{5.8}$$

Pooling Layer. Pooling layers are used to reduce the spatial dimension of the feature maps through aggregation. This operation aims to extract features that are robust against displacement and distortion of patterns in the input image. The aggregation is performed by calculating the average or maximum within small neighborhoods, denoted as Average Pooling and Max Pooling.

Batch Normalization. The batch normalization method was introduced by Ioffe et al. in [220] to improve training speed and stability. It consists of normalizing a mini-batch of feature maps to have a zero mean and a variance of one at each pixel location. In this context, a mini-batch represents a random subset of the entire training dataset. The normalized inputs are then scaled and shifted by trainable parameters. As a result, the so-called internal covariate shift is reduced by stabilizing the distribution of the feature maps. This means that the strong change in trainable parameters that is otherwise necessary due to constant input distribution changes is reduced, stabilizing and accelerating the training. In addition, batch normalization acts as a form of regularization that can reduce overfitting.

Fully Connected Layers. After the sequential application of several convolutional, activation, and pooling layers, which extract abstract image features, fully connected layers

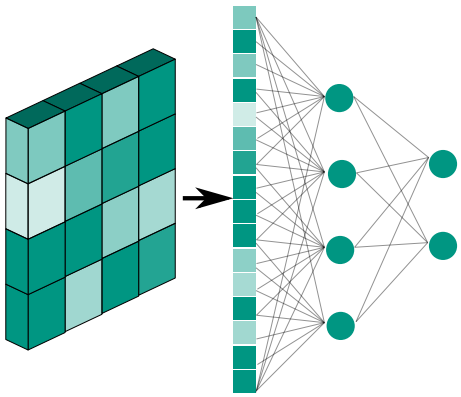


Figure 5.4.: Fully connected layers.

follow, which combine these features. As illustrated in Figure 5.4, a fully connected layer is applied to the feature maps previously flattened to a 1-dimensional representation. The length of the output vector of the last fully connected layers corresponds to the number of target classes of a classification task. To translate the resulting output vector z into a probability distribution, the Softmax activation function σ is used:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}. \tag{5.9}$$

Model Training. During model training, the parameters of the network, also known as weights, are adjusted with the aim of optimally matching the prediction of the model to the given classes of the training data. This task is formulated as the minimization of a loss function that quantifies the difference between model predictions and the actual class distributions. For classification tasks, for example, the cross entropy loss function can be used:

$$L = -\sum_i y_i \log(\hat{y}_i), \tag{5.10}$$

where y_i is the actual and $\hat{y}_i$ is the predicted class probability. A suitable optimization method is the Stochastic Gradient Descent (SGD) approach. The basic idea is to change the parameters (i.e. the weights of the network) in an iterative process along the direction of the steepest descent of the loss function in order to move towards its minimum. The weights are adjusted according to the equation:

$$w_{t+1} = w_t - \eta \nabla_w L(w_t), \tag{5.11}$$

where w_t are the current weights, w_{t+1} are the adjusted weights, η is the step size (learning rate) that indicates the amount of change applied to the weights and $\nabla_w L(w_t)$ is the

gradient of the loss function in relation to the weights. In the SGD method, instead of considering the entire dataset for the gradient calculation during one iteration step, only a small, randomly selected subset of the dataset (mini-batch) is considered. This reduces the calculation time and can also help to overcome local minima.

A method to efficiently calculate the gradients of the loss function with respect to each weight is backpropagation. First, an input is passed forward through the network to calculate the outputs and intermediate results of each layer. This step is called forward propagation. Once the loss has been calculated, the error gradients are calculated within the backward propagation. First, the gradient of the loss with respect to the output of the network is calculated. Subsequently, the gradients are propagated from the output layer back to the input layer. In this process, the partial derivative of the loss function is calculated with respect to each weight in the network. Once the gradients are calculated, the weights are updated using the gradient descent rule (Equation 5.11).

5.3 Related Work

For CNN-based PolInSAR classification, there is little literature available so far. However, several works deal with the related task of CNN-based PolSAR classification. Since the findings from the PolSAR classification are possibly transferable to the PolInSAR classification, the relevant studies from this area are discussed in this Section.

5.3.1 Input Representation

The challenge in applying CNNs to classifying PolSAR data is to deal with a multi-dimensional complex-valued signal. There are two general approaches to address this challenge. One is the development and use of complex-valued CNNs, as proposed in [24, 195], and the other is the transformation of the data into a real-valued representation that can be used as input to established CNNs. In this work, the latter variant is chosen. There are various approaches for creating a real-valued representation that captures the essential information of the PolSAR data. In this section, different real-valued representations used as CNN input in existing research are presented and discussed.

Feature Stacking

One approach for representing PolSAR data, referred to as feature stacking in this work, consists of extracting real-valued features and combining them into a multi-channel image. The following gives a structured overview of different representations used in the literature as CNN input. Table 5.1 categorizes these representations into those directly based on

the coherency matrix $\langle \mathbf{T} \rangle$ and those requiring polarimetric feature extraction by means of target decomposition.

The most commonly used representation is the concatenation of the real-valued diagonal elements and the real and imaginary parts of the off-diagonal elements of the polarimetric coherency matrix $\langle \mathbf{T} \rangle$. This representation is subsequently called $\mathrm{T}_{\mathrm{re,im}}$. It serves as input vector for autoencoders in [173, 244, 245], as multi-channel input image for 2D-CNNs in [147, 221, 222, 224, 226–230], and as multi-channel input image for 3D-CNNs in [223]. These studies focus primarily on improving the classification process or network architecture and do not address the choice of input or provide a rationale for their selection.

In contrast, Zhang et al. [231] question whether separating the complex-valued matrix elements into real and imaginary parts is optimal for CNN-based processing. They hypothesize that decoupling these components leads to information loss since they can only be interpreted in combination. Therefore, they propose to split the complex-valued elements of $\langle \mathbf{T} \rangle$ into their amplitude and phase ($\mathrm{T}_{\mathrm{amp,pha}}$) to better preserve the information content. Their results show that this real-valued representation improves the classification performance. Despite this finding, this representation is rarely used in the literature.

Widely used is the six-dimensional representation proposed by [20]. This consists of the total scattered power ($span_{\mathrm{dB}}$), normalized power ratios ($\langle t_{22} \rangle / span$ and $\langle t_{33} \rangle / span$) and polarimetric correlation coefficients ($\rho_{t_{12}}$, $\rho_{t_{13}}$, $\rho_{t_{23}}$). For CNN classification, it is used in [233–238]. According to Zhou et al. [20], the proposed representation is well-suited for neural networks due to the limited value range of power ratios and correlation coefficients. However, their work does not provide experimental evidence of the superiority of this data representation over other representations.

To support and enhance CNN-based classification, several studies [21, 239–243] suggest incorporating domain and model-based knowledge about target scattering mechanisms into the classification process. Chen et al. [21] were the first to propose this approach. In their work, they combine the total scattering power $span$, the well-established polarimetric features entropy H, anisotropy A, and α, as well as hidden polarimetric features describing target orientation diversity $\theta_{null\mathrm{Re}(\langle t_{12} \rangle)}$, $\theta_{null\mathrm{Im}(\langle t_{12} \rangle)}$, into a multi-channel image that serves as input for a CNN. Experiments on two different datasets show that this feature representation improves classification performance and training time, compared to using $\mathrm{T}_{\mathrm{re,im}}$ as input.

A similarly constructed feature set is used in [243] as input for a newly designed CNN, combining polarimetric features H, A, and α with the smallest eigenvalue λ_3 of $\langle \mathbf{T} \rangle$ and the $\mathrm{T}_{\mathrm{re,im}}$ representation, along with the total scattering power and three correlation coefficients. The selection of this feature set is justified by an initial feature selection process, though it is not extensively detailed.

Table 5.1.: Real-valued PolSAR data representations used as input for CNN-based image classification in the literature categorized into feature sets based directly on the polarimetric coherency matrix and feature sets that include target decomposition features (TDF).

Category	Name	Components	Related publications
$\langle \mathbf{T} \rangle$	$\mathrm{T}_{\mathrm{re,im}}$	$\langle t_{11} \rangle$, $\langle t_{22} \rangle$, $\langle t_{33} \rangle$, $\mathrm{Re}(\langle t_{12} \rangle)$, $\mathrm{Im}(\langle t_{12} \rangle)$, $\mathrm{Re}(\langle t_{13} \rangle)$, $\mathrm{Im}(\langle t_{13} \rangle)$, $\mathrm{Re}(\langle t_{23} \rangle)$, $\mathrm{Im}(\langle t_{23} \rangle)$	[147, 221–230]
	$\mathrm{T}_{\mathrm{amp,pha}}$	$\langle t_{11} \rangle$, $\langle t_{22} \rangle$, $\langle t_{33} \rangle$, $\lvert\langle t_{12} \rangle\rvert$, $\arg(\langle t_{12} \rangle)$, $\lvert\langle t_{13} \rangle\rvert$, $\arg(\langle t_{13} \rangle)$, $\lvert\langle t_{23} \rangle\rvert$, $\arg(\langle t_{23} \rangle)$	[231]
	Pauli	$\langle t_{11} \rangle$, $\langle t_{22} \rangle$, $\langle t_{33} \rangle$	[232]
	Zhou	$span_{\mathrm{dB}}$, $\langle t_{22} \rangle / span$, $\langle t_{33} \rangle / span$, $\lvert\rho_{t_{12}}\rvert$, $\lvert\rho_{t_{13}}\rvert$, $\lvert\rho_{t_{23}}\rvert$	[20, 233–238]
TDF	Yamaguchi	$P_{s,Y}$, $P_{d,Y}$, $P_{v,Y}$, $P_{h,Y}$	[239]
	Freeman Durden	$P_{s,FD}$, $P_{d,FD}$, $P_{v,FD}$	[239]
	H-A-α	H, A, α	[239]
	H-A-α & SERD	H, A, α, SERD	[239]
	Chen	H, A, α, $span$, $\theta_{null\mathrm{Re}(\langle t_{12} \rangle)}$, $\theta_{null\mathrm{Im}(\langle t_{12} \rangle)}$	[21]
	Li	$\mathrm{T}_{\mathrm{re,im}}$, H, A, α, λ_1, λ_2, λ_3	[240]
	Ahishali	$\mathrm{C}_{re,im}$, $\mathrm{T}_{\mathrm{re,im}}$, H, A, α, $\bar{\beta}$, $\bar{\gamma}$, $\bar{\delta}$, $span$, λ_1, λ_2, λ_3, $\lvert\rho_{c_{12}}\rvert$, $\lvert\rho_{c_{21}}\rvert$, $\lvert\rho_{c_{23}}\rvert$, $\lvert\rho_{t_{12}}\rvert$, $\lvert\rho_{t_{21}}\rvert$, $\lvert\rho_{t_{23}}\rvert$	[241]
	Wang	$\lvert t_{i,j}\rvert$, $\lvert c_{i,j}\rvert$, $r_{hv/vv}$, $r_{hv/hh}$, $r_{vv/hh}$, $r_{hv/hh+vv}$, H, A, α, λ_1, λ_2, λ_3, Yamaguchi, Van Zyl ($P_{s,VZ}$, $P_{d,VZ}$, $P_{v,VZ}$), Huynen (A_0, $B_0 + B$, $B_0 - B$, C, D, E, F, G, H)	[242]
	Qin	$\mathrm{T}_{\mathrm{re,im}}$, λ_3, A, α, $span_{db}$, $\lvert\rho_{t_{12}}\rvert$, $\lvert\rho_{t_{13}}\rvert$, $\lvert\rho_{t_{23}}\rvert$	[243]

One of the few studies that compare different feature sets as CNN input was conducted by Zhang et al. [239]. In this study, five feature sets (H-A-α, H-A-α combined with Single-bounce Eigenvalue Relative Difference (SERD), the conformity coefficient combined with polarimetric correlation coefficients, Freeman Durden decomposition parameters, and Yamaguchi decomposition parameters) are compared as inputs for a CNN. The target task is to detect oil slicks in PolSAR images. The Yamaguchi and the Freeman Durden feature set provide the best results in their analyses. However, since the study focuses on a specific two-class problem, the generalization of this statement to more complex land cover classification tasks remains questionable. Furthermore, it has not been investigated whether merging all feature sets into a comprehensive data representation might improve the results.

Larger feature sets that combine features derived directly from $\langle \mathbf{T} \rangle$ with features based on different target decompositions are presented by Wang et al. [242] and Ahishali et al. [241] as input for a 1D-CNN. Wang et al. [242] show that using an extensive feature set (with 45 features) achieves better results than the $\mathrm{T}_{\mathrm{re,im}}$ representation. However, a comparison with smaller feature sets is not performed. In contrast, Ahishali et al. [241] incrementally add polarimetric features to a representation based on $\langle \mathbf{T} \rangle$ and $\langle \mathbf{C} \rangle$ and show that classification accuracy increases as the number of features increases. This indicates the superiority of comprehensive feature sets over small feature sets. However, the applicability of this statement to 2D-CNNs remains open.

Summary. In the existing literature on CNN-based classification of PolSAR data, there is no consensus on the optimal strategy for representing the input data. The influence of the choice of representation on the classification often goes unnoticed. In many papers, $\mathrm{T}_{\mathrm{re,im}}$ is used as input without justifying this choice. Some works show the potential for improvement by integrating polarimetric features. However, a comprehensive comparison of different input data representations for classification using 2D-CNNs cannot be found. This research gap is addressed in this thesis. First, a systematic analysis regarding suitable representations for the CNN-based classification of PolSAR data is conducted, which subsequently guides the design for the development of CNN-based approaches for classifying multi-frequency PolInSAR data.

Automatic Feature Selection

The previous section indicated that exhaustive feature sets, representing the comprehensive information in complex-valued measurements, may enhance CNN classification. The disadvantage of using comprehensive feature sets, which comprise many features, is the increase in computing time and memory requirements. Dimension reduction methods can be applied to address this problem. The literature dealing with the use of feature transformation methods for dimension reduction in the context of PolSAR classification

has already been discussed in detail in the previous chapter (Section 4.2.2). Therefore, in the following, only related works are described in which automatic feature selection is used for dimension reduction.

While there is extensive literature on feature selection for PolSAR classification prior to SVMs and Random Forest classification, research focused on feature selection prior to CNN classification is relatively sparse. The following sections will discuss both cases.

Filter Methods for Traditional Machine Learning Models. One of the earliest studies applying automatic feature selection on an exhaustive PolSAR feature set was conducted by Maghsoudi et al. in [196]. The authors use a multivariate filter method with sequential forward selection and the Fisher class separability measure and introduce a new nonparametric measure for between-class and within-class scatter. The highest-ranked feature subset is subsequently used as input for an SVM to classify boreal forest types. Experiments show that SVM classification based on 14 selected features outperforms using the complete set containing 58 features.

Yin et al. [197] introduce a new feature similarity measure to realize unsupervised feature selection. This similarity is calculated through a two-step process: first, pairwise correlation between all features is computed; second, similar features are grouped and represented by the most frequent occurrence, while single dissimilar features are retained. This method reduces a feature set from 107 to 10 features without compromising SVM classification performance.

Huang et al. [198] propose a three-fold filter method to construct an optimal feature subset, considering both polarimetric and texture features. The feature selection is framed as an optimization problem with three objectives: maintaining feature consistency using Generalized Canonical Correlation Analysis (GCCA) [246], selecting features via $l_{2,1}$ norm sparsity regularization, and preserving the data structure through manifold regularization. Results show that this method improves the classification accuracy of KNN and Random Forest classifiers compared to the classification using the original feature set and compared to other filter methods.

Wrapper Methods for Traditional Machine Learning Models. Several studies demonstrate the efficacy of wrapper methods in feature selection for PolSAR data classification. Lardeux et al. [199] combine forward and backward feature selection with an SVM classifier, reducing a 54-feature set to 25 features and improving classification accuracy. To reduce the computation time of repeated SVM training, Maghsoudi et al. [201] propose a point sampling strategy for this wrapper method and demonstrate that this approach outperforms filter-based methods. Instead of using forward or backward feature selection, Salehi et al. [247] employ a genetic algorithm for feature subset generation combined with an SVM

classifier, reducing a 106-dimensional feature set to 16 components and achieve improved accuracy.

Embedded Methods for Traditional Machine Learning Models. Several approaches integrate feature selection directly into Random Forest classifiers. For example, Loosvelt et al. [14] employ Random Forest feature permutation importance to maintain classification accuracy while reducing prediction uncertainty in crop classification. Similarly, Deng et al. [200], and Hariharan et al. [248] utilize Random Forest feature importance to enhance land cover and crop growth stage classification, respectively. Decision tree-based feature importance has also been applied in PolInSAR classification and parameter estimation [249, 250], demonstrating improved interpretability and efficiency. Embedded feature selection within SVM classification of PolSAR data is proposed by Bai et al. [202]. Using the concept of the Feature Generating Machine [251] to learn a sparse solution with respect to input features allows to identify a reduced feature set that is suitable for classification without performance loss.

Feature Selection for CNN Models. The number of studies addressing feature selection prior to CNN-based PolSAR classification is limited. In the literature, only two filter-based methods have been identified. One method, proposed by Yang et al. [252], involves a CNN-based multivariate filter approach. Initially, a 1D-CNN is trained on an exhaustive feature set. Feature subset candidates are generated using the Kullback-Leibler distance and are then evaluated using classification accuracy obtained with the 1D-CNN. The best-performing subset is tested based on SVM and 2D-CNN classification. Compared to other dimension reduction techniques (mRMR, PCA) and deep learning methods (WAE, WCAE, WDBN, CV-CNN) that utilize the coherency matrix as input, the proposed approach shows superior results. However, a comparison to using the exhaustive feature set for 2D-CNN classification is not conducted.

Shang et al. [253] propose a multivariate filter method to select discriminative features for efficient CNN training. This method extends the PolSAR-specific ReliefF algorithm [254] by integrating the revised Wishart distance for neighborhood search. Results from CNN classification on various datasets show improvements in both classification accuracy and training time.

To date, no literature addresses CNN-based wrapper feature selection or embedded feature selection within CNN training for PolSAR classification. While wrapper methods are generally unsuitable due to the long training times of CNNs, embedded feature selection within CNN training presents a feasible approach. This concept has already been explored in hyperspectral image classification. In [255, 256], channel-independent 1×1 convolutions are applied, utilizing hard thresholding and l_1-norm weight regularization, respectively,

for feature selection. Both approaches demonstrate the advantage of jointly optimizing feature selection, spatial-spectral feature extraction, and classification.

Summary. The literature on feature selection for PolSAR classification using traditional machine learning models demonstrates that feature selection can reduce training time while maintaining or even improving classification accuracy. However, research on feature selection combined with CNN classification remains limited. The gap can be attributed to the fact that only the coherency matrix elements are often used as input for the CNN classification, whereby dimension reduction is not required. However, if an extensive feature representation is used as input, dimension reduction also gains importance in this area. Therefore, this thesis evaluates filter-based feature selection methods to create low-dimensional representations of multi-frequency PolInSAR data. Furthermore, inspired by advances in hyperspectral image classification, this research introduces and evaluates a feature selection method embedded into CNN training applied to multi-frequency PolInSAR data.

5.3.2 CNN Architectures

The methods employed in CNN-based PolSAR classification can be broadly categorized into three distinct approaches: pixel-based, patch-based, and image-based.

The pixel-based approach focuses on learning and extracting complex patterns along the feature dimension, treating each pixel independently from its neighbors. The patch-based approach incorporates spatial context by learning and extracting patterns considering each pixel in conjunction with its immediate neighbors. The image-based approach analyzes patterns across larger image regions. In addition, methods for cross-channel feature extraction have been developed, which extend patch- or image-based approaches to combine feature learning along the feature and image dimensions within the CNN.

Pixel-based Classification. One of the pioneering works in the domain of deep learning-based classification of PolSAR images is conducted by Xie et al. [244]. They propose the use of a stacked autoencoder for unsupervised feature extraction from the coherency matrix elements ($T_{re,im}$), followed by supervised classification using a shallow, fully connected network. This approach performs superior to the Wishart classifier and SVM-based methods. In [173], Xie et al. further enhance this technique by integrating PolSAR statistics into the loss function of the autoencoder, using the Wishart distance to quantify the reconstruction error, thereby improving classification accuracy. Extracting pixel-based features using convolutional layers is proposed by Wang et al. in [242]. They implement supervised feature learning and classification using 1D-CNNs, using a stack of polarimetric features as input. This approach achieves a classification accuracy that is comparable to that of

autoencoders. Similarly, Ahishali et al. [241] effectively adapt this 1D-CNN approach for multi-frequency PolSAR image classification.

Patch-based Classification. Patch-based methods differ from pixel-based techniques by incorporating spatial context. For each pixel, the surrounding image patch is considered for classification. Spatial features are extracted from this patch using a 2D-CNN to classify the central pixel. This method is applied across the entire PolSAR image by employing a sliding window approach.

One of the first patch-based classification approaches is proposed by Zhou et al. [20]. The authors propose a network architecture adapted from LeNet [219] that was originally developed for image classification. Feature extraction is performed on an 8×8 pixel image patch through convolutional layers, followed by ReLU activation and max pooling, with classification handled by a fully connected layer. Similar architectures are used in [21, 22] for PolSAR land cover classification.

Enhancements to the CNN architecture are explored to further improve classification. Instead of simple networks such as LeNet, more sophisticated architectures such as Residual Networks (ResNets) [257], which facilitate the training of deep networks, are examined in [258–260]. Another example, proposed in [232], is the employment of the Pyramid Scene Parsing Network [261], which enables the extraction and combination of multi-scale features. Moreover, attention mechanisms including Squeeze and Excitation (SE) [262] and Convolutional Block Attention Module (CBAM) [263], which focus on significant feature channels or image regions during CNN training, are explored for PolSAR classification in [243, 258, 264].

Image-based Classification. The use of patch-based methods for image classification has the disadvantage of requiring high computing and memory requirements, because the classification is performed pixel-by-pixel. Further, this approach limits the ability of the network to capture global contextual information, often resulting in less coherent and precise segmentations. To solve these problems, Fully Convolutional Networks (FCNs) with an encoder-decoder structure were developed that allow dense predictions for the entire image in a single forward pass. This type of network consists of an encoder that progressively reduces the spatial dimensions while extracting spatial features, followed by a decoder that upsamples the feature maps to the original image size for pixel-wise classification. By processing the entire image at once, FCNs leverage local and global features to produce a more accurate and consistent segmentation while significantly reducing computational and memory requirements.

The concept of FCNs was first proposed by Long et al. [265] for image segmentation. The authors introduce an architecture that uses skip connections to combine high-level semantic information from deeper layers with low-level spatial details from shallower layers

(resulting in FCN-8/16/32). This method has been successfully applied in subsequent research for PolSAR classification in [160, 240, 266]. Specifically, Wang et al. [266] demonstrate that FCN-based classification using the polarimetric features H, A and α outperforms SVM-based classification. In [240], it is shown that FCN-based classification is also superior to the patch-based CNN classification in terms of classification performance and efficiency. Additionally, Wang et al. [160] combined polarimetric features with deep spatial features using an FCN. This approach achieves an improvement in classification accuracy over methods that rely solely on polarimetric features.

For fine-grained classification of complex wetlands, Mohammadimanesh et al. [267] propose an enhanced FCN architecture based on the SegNet [268]. The SegNet framework retains spatial detail by storing the max-pooling indices during downsampling and using these indices to guide the upsampling process, thereby preserving the spatial locations of features. As an extension to the original SegNet architecture, the authors of [267] incorporate Inception modules to facilitate multi-scale feature extraction. The network inputs include the amplitudes of three polarizations (HH, HV, VV) and the interferometric coherence in VV polarization resulting from a repeat-pass measurement. The proposed model exhibits superior classification performance compared to FCN-8/16/32 and SegNet without Inception modules. This study is among the few that integrate interferometric coherence into image-based CNN classification, although the impact of this additional information on classification outcomes remains unexplored in their work.

The U-Net presented by Ronneberger et al. [269] is a further development of the FCN architecture, which is suitable for training using small amounts of data. It enhances the basic FCN framework using a symmetric architecture with numerous skip connections between the encoder and decoder, allowing precise feature localization throughout the upsampling process. This architecture has also been applied successfully to PolSAR image analysis. In [230], the U-Net model performs better than patch-based approaches and architectures similar to FCN-8. Classification accuracy is further improved by incorporating attention mechanisms, which capture features at various scales. Liu et al. [270] extend U-Net to the complex domain for PolSAR classification, demonstrating superior performance compared to its real-valued counterpart (using $T_{re,im}$ as input). Further advancements in complex-valued U-Net are demonstrated by Yu et al. [271] through the integration of an embedded capsule network between the encoder and decoder, enhancing feature extraction capabilities.

Cross-channel Feature Extraction. A central difference between pixel-based and patch- or image-based deep learning approaches lies in how the features are extracted. Pixel-based methods recognize patterns within the polarimetric features of individual pixels, while patch- and image-based methods recognize image patterns. To combine both types of pattern recognition, Zhang et al. [231] propose using a 2D depthwise separable convolution. In contrast to conventional 2D convolutions that extract spatial features

across all input channels and sum the results channel by channel, depthwise separable convolutions apply convolution to each input channel individually, followed by a point-wise convolution to combine the channels. This approach achieves a slight improvement in classification accuracy. 2D depthwise separable convolutions are also used in [224]. However, a significant improvement in classification accuracy is not observed, but shorter training and testing times due to fewer trainable parameters.

To enhance patch- and image-based methods by incorporating cross-channel feature extraction, 3D convolutions have been proposed. These convolutions simultaneously extract features along both the channel and image dimensions. Zhang et al. [272] introduce this approach for patch-based PolSAR classification, showing a slight improvement over 2D counterparts. To improve efficiency, Dong et al. [223] propose lightweight 3D-CNNs for PolSAR classification using pseudo-3D or 3D depthwise separable convolutions. These lightweight models not only outperformed 2D-CNNs but also slightly exceeded the performance of standard 3D-CNNs while significantly reducing the number of trainable parameters.

Another approach involves channel attention mechanisms, which enhance the representational power of the network by learning weights for each channel, selectively emphasizing informative channels while suppressing less relevant ones. Dong et al. [264] applied this concept to recalibrate input image channels composed of various polarimetric features using the SE module [262]. This module compresses spatial information via global average pooling and models channel interrelationships using two fully connected layers to find optimal weighting parameters. Recalibration improved accuracy compared to CNNs without the attention module. Similarly, Yang et al. [258] employed the CBAM [263], which uses global average and max pooling followed by a fully connected layer for polarimetric feature channel weighting. Results demonstrate that incorporating this module in patch-based CNN classification improves performance and accelerates network convergence.

Summary. The literature review concerning different CNN models used for the classification of PolSAR data suggests that image-based models based on the concept of FCNs are best suited for land cover classification. Advantages over pixel-based and patch-based approaches include the ability to classify larger image regions in a single forward pass and to learn and integrate image patterns that capture a larger image context. Based on these findings, FCNs based on the U-Net structure are employed in this work for the classification of multi-frequency PolInSAR data.

Several studies indicate that extensions of the model structure that allow cross-channel feature extraction can further improve the classification results by recognizing patterns along the input feature dimension in addition to image patterns. These concepts are, therefore, integrated and tested in the U-Net models in this thesis.

One aspect rarely considered in the existing literature is the inclusion of interferometric features in CNN-based classification. One approach that includes this information is described in [267]. However, this approach only utilizes part of the polarimetric information. In contrast, this work aims to combine the full information potential of PolInSAR data within image-based CNN classification.

5.4 Methods

The classification of multi-frequency PolInSAR image data using CNNs follows the workflow illustrated in Figure 5.5. The process begins with multi-frequency PolInSAR image data, where each pixel is represented by the matrix $\mathbf{T}_6$. In the preprocessing stage, ensemble averages $\langle \mathbf{T}_6 \rangle$ are calculated using the Refined Lee filter (window size: 5×5, $\sigma_v = 1$), followed by radiometric calibration to γ^0. The complex-valued data is then transformed into a real-valued multi-channel image using one of the following methods:

- **Feature Stacking**: Extraction and stacking of predefined polarimetric and interferometric features.
- **Automatic Feature Selection**: Extraction of a comprehensive set of polarimetric and interferometric features, from which an optimal subset is automatically selected and stacked.
- **Feature Transformation**: Extraction of a comprehensive set of polarimetric and interferometric features, which is transformed into a low-dimensional representation.

The resulting multi-channel image is subsequently classified using an FCN with an encoder-decoder architecture. The output $\hat{\mathbf{y}}$ represents the predicted class probabilities for each pixel, and the predicted class corresponds to the class with the highest probability. The subsequent sections will detail the different methods for generating the input data compared in this study and the CNN models employed for the analysis.

5.4.1 Feature Stacking

To generate the input for the CNN model, predefined sets of real-valued features are extracted and stacked into multi-channel images. Although the primary focus of this research is the classification of multi-frequency PolInSAR data, a preliminary analysis is conducted comparing feature sets based solely on single-frequency PolSAR data. This initial analysis step is necessary because existing literature does not definitively indicate which data representation approach (using only the coherency matrix, only decomposition features, or a combination thereof) is most effective for CNN-based classification.

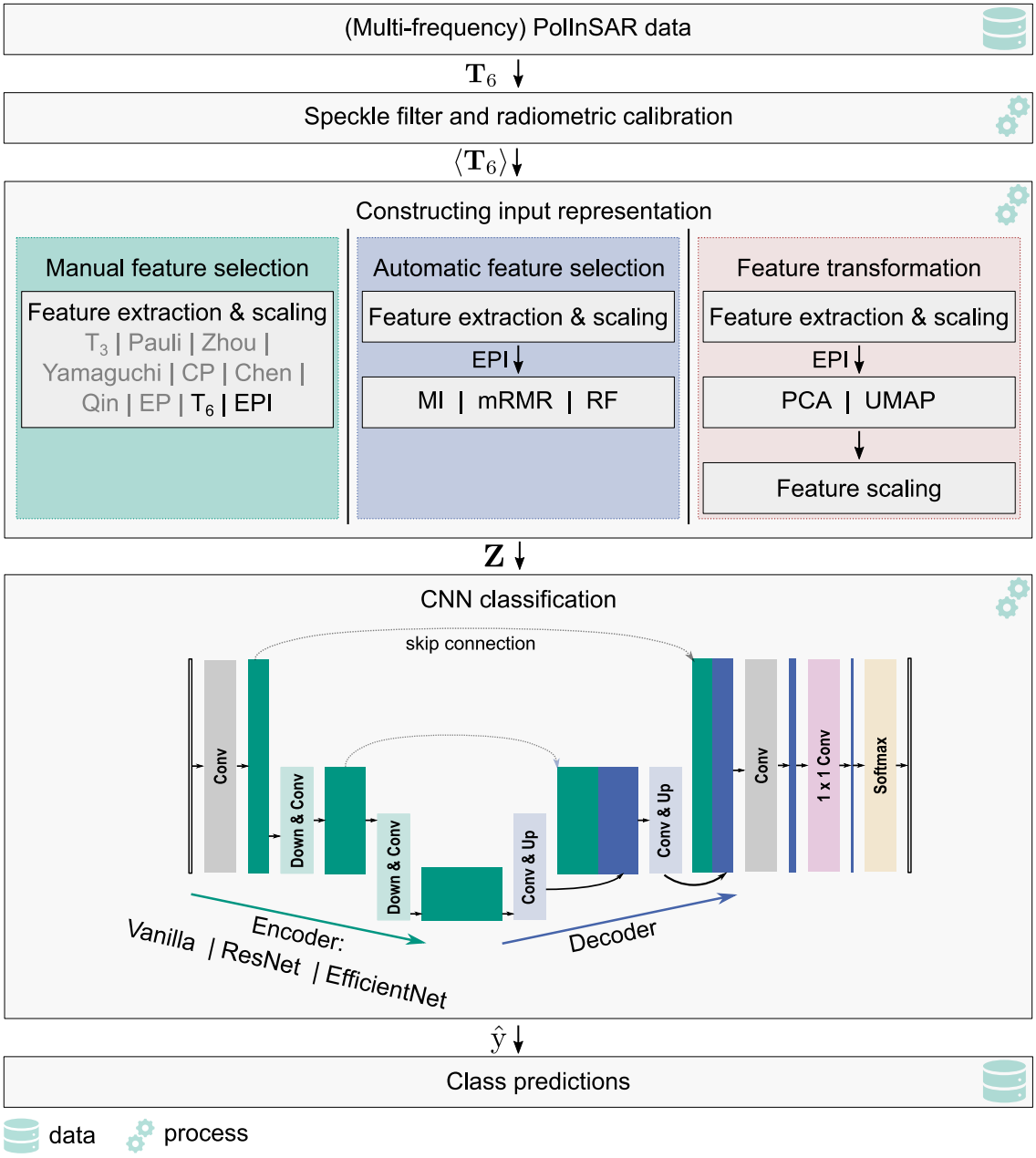


Figure 5.5.: Workflow for CNN-based classification of multi-frequency PolInSAR image data.

From the PolSAR representations employed in the literature (see Table 5.1), several feature sets are tested as input for a CNN classification. From the category of feature sets directly based on $\langle \mathbf{T}_{11} \rangle$, the following are evaluated: $\mathrm{T}_{\mathrm{re,im}}$, $\mathrm{T}_{\mathrm{amp,pha}}$, Pauli, and Zhou.

1. $\mathbf{T}_{\mathbf{re,im}}$: This representation is examined because it is the most widely used in the literature. It is a generic representation that requires no additional computational steps.

2. $\mathbf{T}_{\mathbf{amp,pha}}$: This is also a generic representation that involves only simple computational steps (calculating amplitude and phase). Although scarcely used in literature, a significant argument favoring this representation is that it prevents decoupling the real and imaginary parts of complex-valued off-diagonal elements.

3. **Pauli**: This representation, a subset of $\mathrm{T}_{\mathrm{re,im}}$ and $\mathrm{T}_{\mathrm{amp,pha}}$, is frequently employed for visualizing PolSAR images. Its use facilitates direct visual comparison between input images and classification outputs. Including this subset in the comparison also allows for evaluating whether using the off-diagonal elements of $\langle \mathbf{T}_{11} \rangle$ improves the classification.

4. **Zhou**: This feature set is included due to its widespread use in the literature as CNN input. The six elements of this representation are defined as:

$$span_{db} = 10 \log_{10}(\langle t_{11} \rangle + \langle t_{22} \rangle + \langle t_{33} \rangle) \tag{5.12}$$

$$r_2 = \langle t_{22} \rangle / (\langle t_{11} \rangle + \langle t_{22} \rangle + \langle t_{33} \rangle) \tag{5.13}$$

$$r_3 = \langle t_{33} \rangle / (\langle t_{11} \rangle + \langle t_{22} \rangle + \langle t_{33} \rangle) \tag{5.14}$$

$$|\rho_{t_{12}}| = |\langle t_{12} \rangle| / \sqrt{\langle t_{11} \rangle \cdot \langle t_{22} \rangle} \tag{5.15}$$

$$|\rho_{t_{13}}| = |\langle t_{13} \rangle| / \sqrt{\langle t_{11} \rangle \cdot \langle t_{33} \rangle} \tag{5.16}$$

$$|\rho_{t_{23}}| = |\langle t_{23} \rangle| / \sqrt{\langle t_{22} \rangle \cdot \langle t_{33} \rangle}. \tag{5.17}$$

From the category of feature sets that incorporate target decomposition features (TDF), the sets Yamaguchi, H-A-α, VanZyl, Chen, and Qin are chosen:

1. **Yamaguchi**: This feature set, whose discriminative power has been demonstrated across a wide range of applications, is included as a representative of model-based target decomposition methods.

2. **H-A-α**: Similarly, the H-A-α feature set, known for its discriminative power for land cover classification, is included as a representative of algebraic target decomposition methods.

3. **VanZyl**: The decomposition features proposed by VanZyl, are derived by a combined approach of model-based and algebraic decomposition.

4. **Chen**: This feature set was among the first to include a combination of polarimetric features in CNN classification. It has received significant attention in the literature, has been used in subsequent studies as a benchmark method, and is therefore included in this work. The included null-angle features are defined as:

$$\theta_{null\mathrm{Re}(\langle t_{12}\rangle)} = -\frac{1}{2}\arg(\mathrm{Re}(\langle t_{13}\rangle) + j\mathrm{Re}(\langle t_{12}\rangle)) \tag{5.18}$$

$$\theta_{null\mathrm{Im}(\langle t_{12}\rangle)} = -\frac{1}{2}\arg(\mathrm{Im}(\langle t_{13}\rangle) + j\mathrm{Im}(\langle t_{12}\rangle)) \tag{5.19}$$

5. **Qin**: This feature set combines features derived directly from the coherency matrix with features based on target decomposition. With 16 different features, it is one of the largest feature sets used as input for CNN-based PolSAR classification in the literature, aiming at providing a comprehensive representation of the data.

In addition to the PolSAR representations used in the literature, the exhaustive feature set introduced in Chapter 4 is incorporated into the analysis. Since the preliminary analysis focuses exclusively on single-frequency PolSAR data, this feature set —referred to as the exhaustive PolSAR (EP) feature set— contains the 36 polarimetric features listed in Table 4.2.

Informed by the results of the preliminary analysis on CNN-based PolSAR classification with varying inputs, which will be detailed in Section 5.6, only two feature sets are employed for the representation of multi-frequency PolInSAR data. One approach involves concatenating matrix elements (see Figure 5.6) of the PolInSAR coherency matrix $\langle \mathbf{T}_6 \rangle$, divided into amplitude and phase components. Given the high redundancy between $\langle \mathbf{T}_{11} \rangle$ and $\langle \mathbf{T}_{22} \rangle$ (each corresponding to the polarimetric coherency matrix of one PolSAR image of the interferometric image pair), only the elements from the upper triangle of $\langle \mathbf{T}_{11} \rangle$ are utilized. Additionally, elements from the block matrix $\langle \mathbf{\Omega}_{12} \rangle$, which contains interferometric information, are included. Although $\langle \mathbf{\Omega}_{12} \rangle$ is not positive semi-definite (e.g., $\langle t_{15} \rangle \neq \langle t_{24}^* \rangle$), the information content of the upper and lower triangular matrices is very similar. Therefore, only the upper triangle elements are considered here as well. This results in a 21-dimensional feature set, denoted as $\mathrm{T}_{6,\mathrm{amp,pha}}$.

The second representation of the PolInSAR data is the exhaustive PolInSAR (EPI) feature set. This extends the EP feature set by incorporating interferometric coherences and unwrapped interferometric phases (see Table 4.2), resulting in a total of 42 features.

For the representation of multi-frequency PolInSAR data, the corresponding feature images for each frequency band are calculated and then concatenated. This produces the 42-dimensional representation $\mathrm{T}_{6,\mathrm{amp,pha}_{SL}}$ and the 84-dimensional representation $\mathrm{EPI}_{S,L}$.

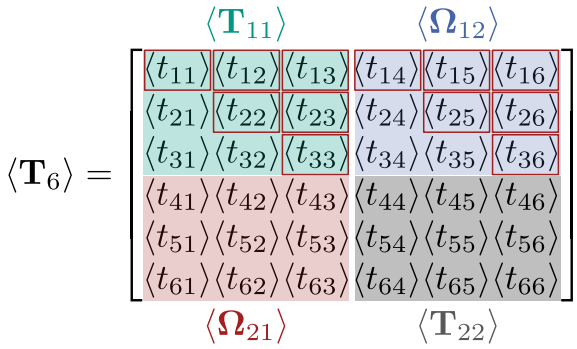


Figure 5.6.: Considered elements of the PolInSAR coherency matrix to construct the input representation $\mathrm{T}_{6,\mathrm{amp,pha}}$.

Some of the extracted features, such as the amplitudes of elements in $\langle \mathbf{T} \rangle$ and scattering power contributions, exhibit distributions that deviate significantly from a normal distribution and possess a high dynamic range. These characteristics can negatively influence the accuracy of CNN-based classification, considering that CNNs are originally designed to process normally distributed RGB images. To mitigate this issue, the affected features $[(|\langle t_{ij} \rangle|, \lambda_i, \underline{\lambda}, P_{s,Y}, P_{d,Y}, P_{v,Y}, P_{s,VZ}, P_{d,VZ}, P_{v,VZ})]$ are logarithmically scaled to approximate a normal distribution and reduce the dynamic range. Subsequently, all features are standardized to convert the various feature values into a common unit, making them comparable. In image processing, z-standardization is typically employed, which involves subtracting the mean and dividing by the standard deviation. However, the extracted PolInSAR features may contain outliers, such as unusually high backscatter values caused by artificial structures. These outliers can greatly influence the mean and standard deviation, making z-standardization inadequate for achieving balanced feature scales. To address this issue, a robust standardization method based on the median and quantile range is employed:

$$x_{ij,\mathrm{scaled}} = \frac{x_{ij} - \mathrm{median}(\mathbf{x}^j)}{p_{98\%}(\mathbf{x}^j) - p_{2\%}(\mathbf{x}^j)}, \tag{5.20}$$

where x_{ij} denotes the value of feature j of i^{th} pixel, $\mathbf{x}^j$ denotes the j^{th} column vector of $\mathbf{X}$, and $p_{98\%}(\mathbf{x}^j)$ and $p_{2\%}(\mathbf{x}^j)$ denote the 98^{th} and 2^{nd} percentiles of $\mathbf{x}^j$, respectively.

5.4.2 Automatic Feature Selection

The starting point for automatic feature selection is the 84-dimensional $\mathrm{EPI}_{S,L}$ feature set that contains features with high correlation and thus redundancy. Redundancy arises from the resemblance of features between frequency bands, the presence of features that are linear or non-linear combinations of others, and using contributions of scattering mechanisms calculated by different methods. To identify discriminative feature subsets with predefined sizes of three, 15, and 40, various filter-based feature selection methods, as well as a CNN-embedded feature selection method, are employed and tested. The selection

of these specific sizes is based on practical considerations. Due to the long training time required for CNN training and testing, it is impractical to test all possible feature set sizes. Therefore, only a very small feature set (three features), an intermediate size (15 features), and a comprehensive feature set that still reduces the number of features to less than half (40 features) are tested. This approach balances the need for thorough evaluation with the constraints of computational resources.

Since filter-based feature selection methods are independent of the classifier, established filter methods can be directly used as preprocessing steps for CNN-based classification. In this work, three approaches for feature weighting and filtering are employed: the univariate method MI, the multivariate method mRMR, and feature selection based on Random Forest Feature Importance (RF-FI).

Mutual Information. MI is a valuable tool for feature selection due to its ability to capture linear and non-linear relationships between a feature and the target class and its independence from any specific assumptions about the distribution of the features. These advantages make MI a more suitable choice than other methods, such as correlation or parametric tests, which often assume normally distributed data.

MI quantifies the dependency between a feature and the target class by measuring the reduction in uncertainty about the target class $\mathbf{y}$ given the feature $\mathbf{x}^j$. Mathematically, it is defined as:

$$MI(\mathbf{x}^j, \mathbf{y}) = \sum_{y \in \mathbf{y}} \sum_{x \in \mathbf{x}^j} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right), \tag{5.21}$$

where $p(x,y)$ is the joint probability distribution of the feature $\mathbf{x}^j$ and the target classes $\mathbf{y}$, and $p(x)$ and $p(y)$ are their marginal probability distributions. Selecting features with high MI values with the target class ensures that these features provide significant information for the given task.

Mimimum Redundancy Maximum Relevance. mRMR is a feature selection method that builds upon the strengths of MI while addressing some of its limitations. While MI effectively measures the dependency between individual features and the target class, it does not consider the redundancy among the selected features. mRMR improves this aspect by not only selecting features that are highly relevant to the target class (maximum relevance) but also ensuring that the selected features are minimally redundant with each other (minimum redundancy).

The relevance of a feature $\mathbf{x}^j$ to the target classes $\mathbf{y}$ is measured using MI, denoted as $MI(\mathbf{x}^j, \mathbf{y})$. The redundancy among features is also measured using MI, denoted as $MI(\mathbf{x}^j, \mathbf{x}^k)$ for features $\mathbf{x}^j$ and $\mathbf{x}^k$. The mRMR criterion optimizes the following objective:

$$\max \left(\frac{1}{|S|} \sum_{\mathbf{x}^j \in S} MI(\mathbf{x}^j, \mathbf{y}) - \frac{1}{|S|^2} \sum_{\mathbf{x}^j, \mathbf{x}^k \in S} MI(\mathbf{x}^j, \mathbf{x}^k) \right), \tag{5.22}$$

where S is the set of selected features. The first term ensures maximum relevance by averaging the MI between each feature and the target classes, while the second term ensures minimum redundancy by averaging the MI among the features themselves.

By considering both relevance and redundancy, mRMR provides a more balanced feature selection process, leading to a more informative and less redundant feature set. This may improve model performance compared to using MI alone.

Random Forest Feature Importance. RF-FI belongs to the category of embedded feature selection methods. However, this work employs it as a filter method since the subsequent classification is performed using a CNN rather than a Random Forest.

A strength of the Random Forest model is its ability to estimate the importance of each feature for classification. To quantify the feature importance, the impurity of the class labels within the nodes of the decision trees is measured by the gini index. The gini index for a node is defined as:

$$gini = 1 - \sum_{c \in C} p_c^2, \tag{5.23}$$

where p_c is the proportion of samples belonging to class c at a given node. The importance of a feature is computed by averaging the decrease in the gini index caused by splits on that feature across all trees in the forest.

Using RF-FI is chosen as feature selection method because it can automatically capture complex non-linear relationships and feature interactions without additional search strategies for feature subsets. Additionally, it does not rely on parametric assumptions.

Regularized Feature Weighting. Filter-based feature selection methods have potential drawbacks when applied to subsequent CNN classification. One concern is the decoupling of feature selection and classification processes. Previously introduced filter-based methods perform feature selection based on pixel-wise features without considering the spatial relationships between pixels. Since CNNs classification is primarily based on detecting image patterns, the features chosen by filter-based methods may not align well with the working concept of CNNs.

Given these considerations, a feature selection method is proposed embedded into CNN training. The method is adapted from the approach proposed in [256], which performs CNN-embedded band selection for hyperspectral image classification. The concept, referred to as Regularized Feature Weighting (RFW) in the following. An additional layer with trainable weights is inserted between the input and the first convolutional layer of the CNN. The number of trainable weights, each initialized to one, corresponds to the number of input features. This additional weight layer aims to mask out features irrelevant for classification by assigning them a weight of zero (or very close to zero).

An l_1-norm regularization term is added to the loss function, which is minimized during CNN training to promote sparsity in the weighting layer. The loss function is formulated as:

$$\text{Total Loss} = \text{Classification Loss} + \lambda_{rfw} |W|_1 \ , \tag{5.24}$$

where W denotes the weights of the additional layer and λ_{rfw} is a hyperparameter that adjusts the contribution of the regularization term. The l_1-norm regularization leads to sparsity by penalizing the absolute values of the weights, effectively pushing many of them towards zero.

After training the model, features are ranked based on the weights of the additional layer. The highest ranked features are then selected and used as input for a CNN (without additional weight layer), which is retrained to optimally adapt to the reduced feature set.

5.4.3 Feature Transformation

While feature selection methods choose a subset of the original features, feature transformation creates new features that individually carry more discriminative information. The starting point for feature transformation is the $\text{EPI}_{S,L}$ feature set or the coherency matrices $\langle \mathbf{T}_6 \rangle$ from S- and L-band, respectively. Based on findings from Chapter 4, the dimension reduction methods PCA and UMAP, including its variants F-UMAP (Section 4.3.2) and W-UMAP (Section 4.3.3), are employed to construct feature sets containing three, 15, or 40 features. PCA is included due to its low computational cost, while UMAP is selected for its superior performance in retaining discriminative information. The final input representation is constructed by scaling the transformed features using robust standardization following Equation 5.20.

5.4.4 CNN Architecture

To realize the classification of multi-frequency PolInSAR image data, this work utilizes an FCN with an encoder-decoder structure based on the widely used U-Net architecture. A

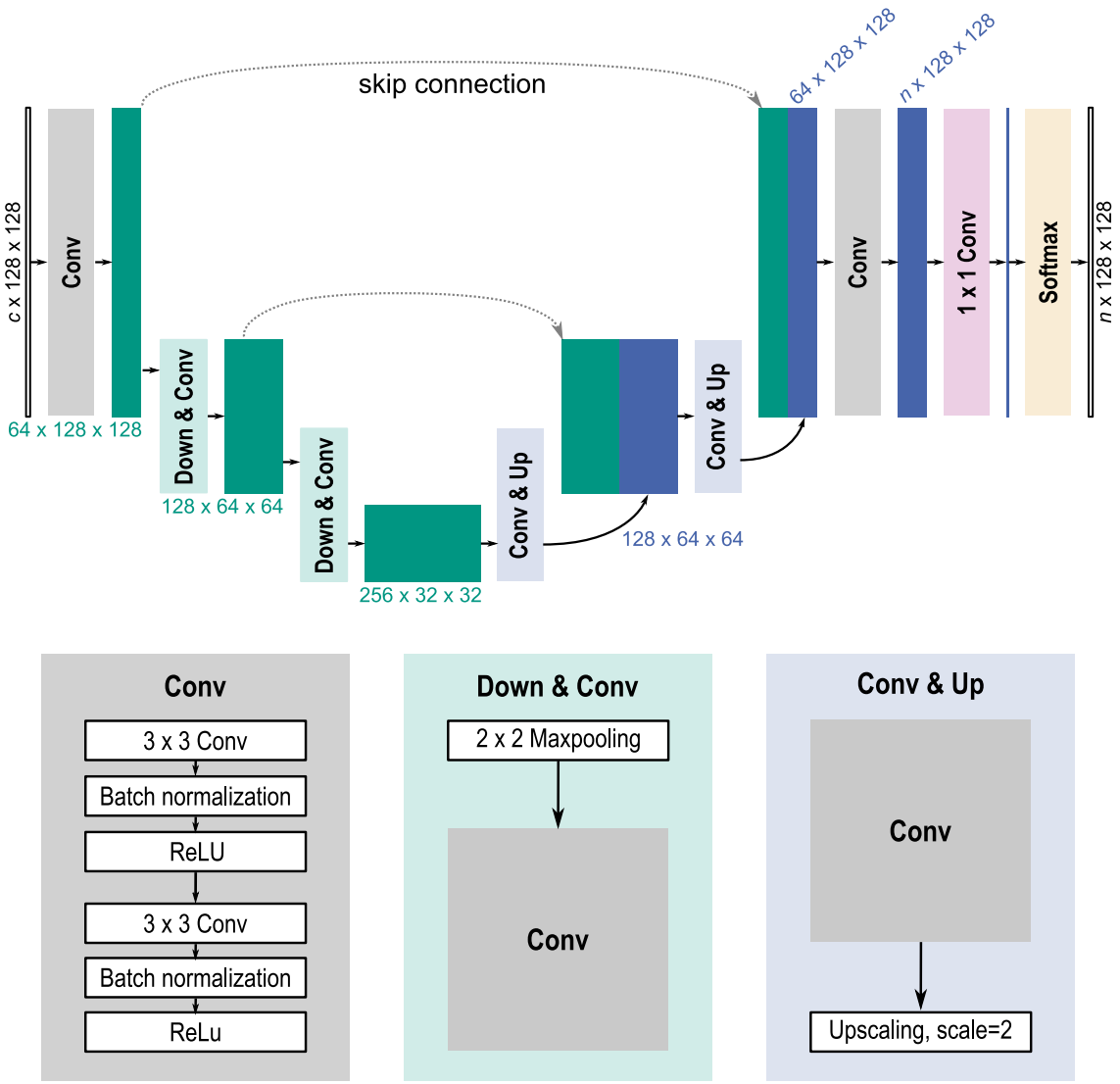


Figure 5.7.: Architecture of the three-stage Vanilla U-Net.

three-stage U-Net variant employed in this work is depicted in Figure 5.7 and is referred to as Vanilla U-Net. The network comprises a contracting path (encoder) and an expansive path (decoder). Within the encoder, hierarchical features are extracted at three resolution levels. Each level utilizes sequences of two 2D convolutional layers followed by batch normalization, ReLU activation layers, and a max-pooling layer, which spatially aggregates the resulting feature maps. While the spatial dimensions of the feature maps are progressively halved, the number of convolutional filter kernels (and thus resulting feature maps) is doubled at each step, resulting in coarse-resolution feature maps at the final level that encode the entire image information.

The symmetrically constructed decoder upsamples the feature maps using nearest neighbor interpolation followed by convolutional layers, ultimately generating feature maps with the same spatial resolution as the input image. Skip connections are employed to enhance spatial localization of the encoded image information, which fuse deep features of coarse

resolution with low-level features of finer spatial resolution through concatenation. At the end of the network, the feature dimension is reduced to the number of output classes using a 1×1 convolution. The softmax function is then applied to transform the feature vectors into a probability distribution over the classes, with each pixel assigned the class with the highest softmax score. The use of only three resolution levels, as opposed to the usual five, aims to reduce model parameters and mitigate the risk of overfitting, which is a concern due to the relatively small training dataset used in this work.

In addition to the Vanilla U-Net, two modified U-Net models are employed for classification in this work. These models, illustrated in Figure 5.8 and Figure 5.9, differ from the Vanilla U-Net in their encoder architectures, which are responsible for feature extraction. In this work, two common models, a ResNet proposed in [257] (Figure 5.8) and an EfficientNet proposed in [273] (Figure 5.9), are used as encoders. ResNets are specific deep CNNs that are proven to be very powerful for the classification of RGB images. The high performance of this model has been achieved by introducing the concept of residual learning, which allows for faster training and enables the training of very deep networks. Instead of direct mapping functions that transform an input into the desired output, residual functions are learned with reference to the layer inputs. This is realized using residual blocks (highlighted in orange in Figure 5.8). These contain so-called shortcut connections that perform an identity mapping of the input, which is added to the output of subsequent layers. This work uses the first three stages of ResNet-18 as the encoder for the U-Net model (Figure 5.8).

EfficientNet was proposed by Tan et al. in [273]. By scaling the network's depth, width, and resolution in a structured way, good performance can be achieved with low resource consumption. The network is mainly built using Mobile Inverted Bottleneck (MBConv) blocks introduced in [274] that are shown in Figure 5.9. This building block includes a so-called inverted residual block, which first employs point-wise convolution (1×1 convolution) to project an input feature map into a higher-dimensional space, subsequently performs depthwise convolution, and finally projects the resulting feature map back to a lower-dimensional space. The input feature map is added to the output feature map using a residual shortcut connection. The inverted residual block is extended by a Squeeze and Excitation (SE) block, which allows recalibration of channel-wise feature responses, enabling the network to emphasize relevant features. This work employs the first three stages of EfficientNet-b0 as the encoder for the U-Net model (Figure 5.9).

While other encoder architectures, such as VGG16 or Inception, could also be considered, identifying the optimal architecture is not the primary focus of this work. Therefore, an exhaustive comparison of different architectures is beyond the scope of this thesis. Using the three different models (Vanilla U-Net, ResNet U-Net, and EfficientNet U-Net) primarily aims to introduce variation in the experiments designed to determine the most suitable input for CNN models, to improve the generalization of the results.

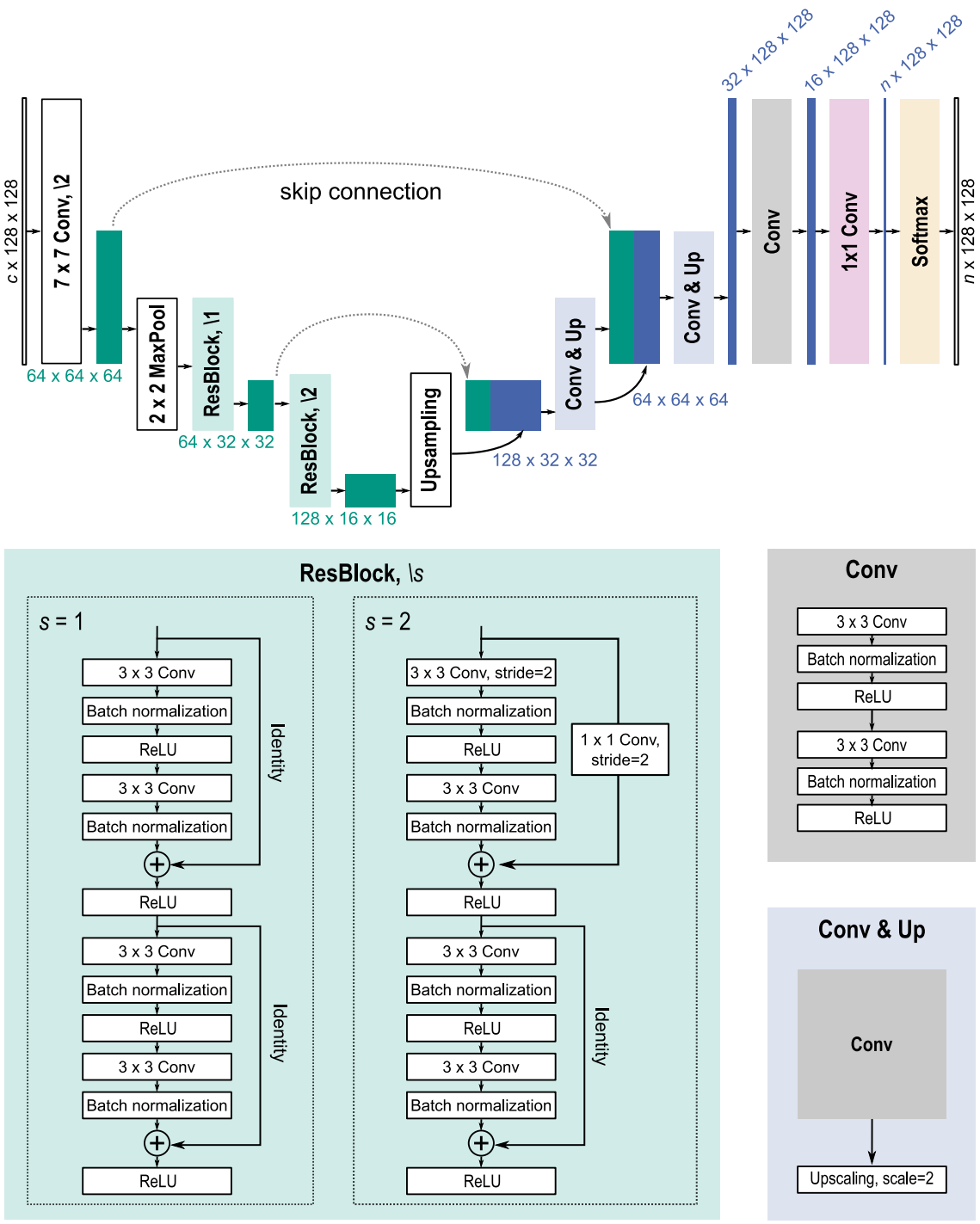


Figure 5.8.: Architecture of the three-stage ResNet U-Net.

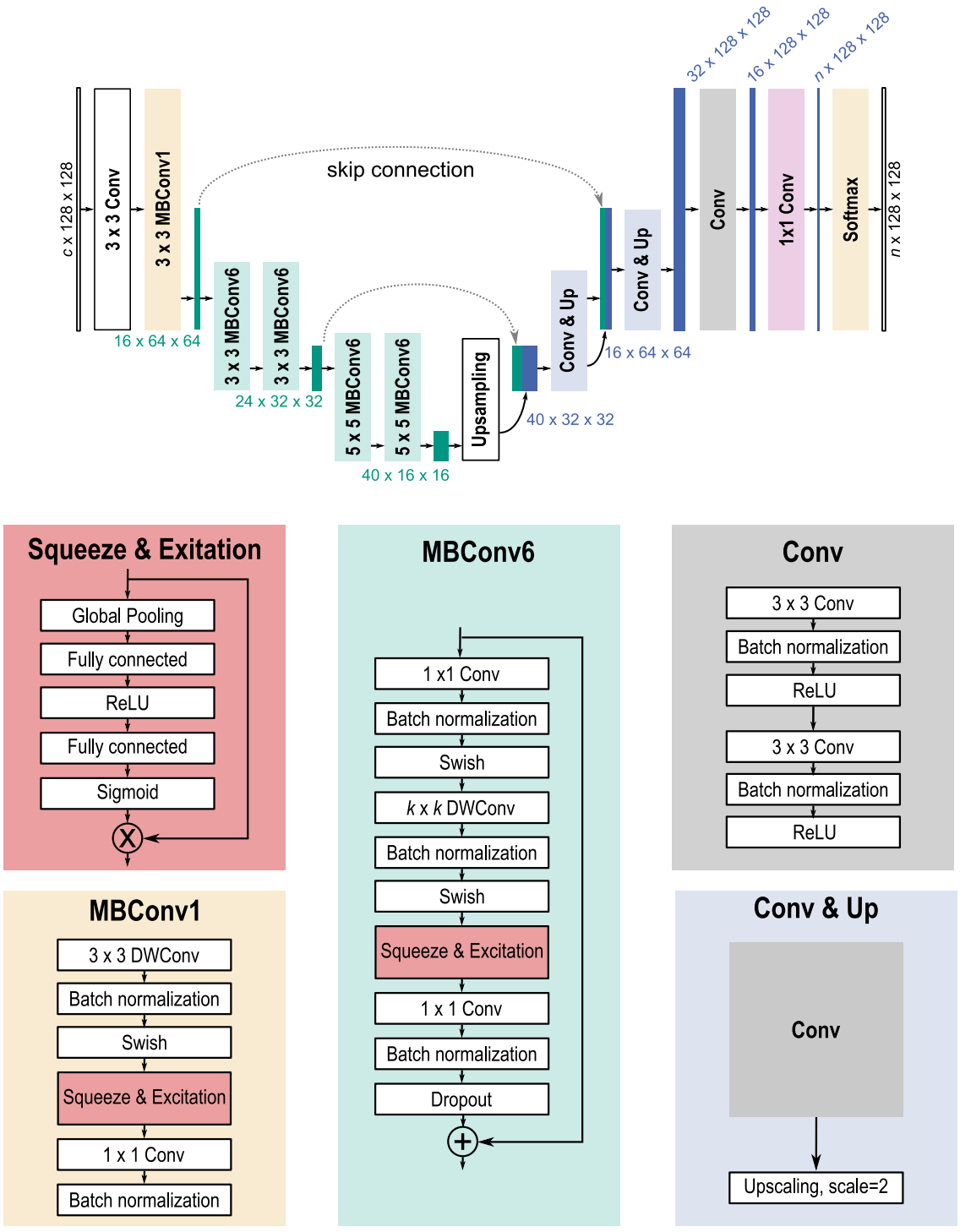
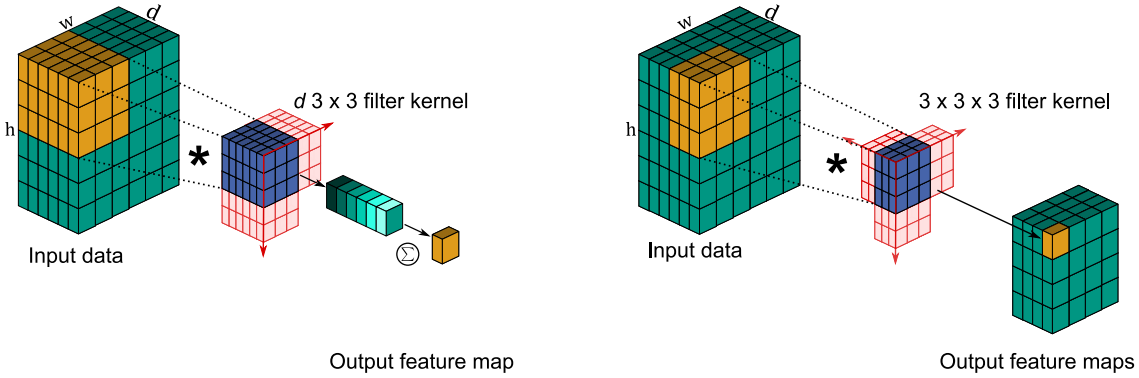


Figure 5.9.: Architecture of the three-stage EfficientNet U-Net.



(a) 2D convolution. **(b)** 3D convolution.

Figure 5.10.: Difference in working principal of 2D convolution (a) and 3D convolution (b) for multi-channel input data.

5.4.5 Cross-channel Feature Extraction

To enable CNNs to model not only complex spatial features but also non-linear relationships between the various polarimetric and interferometric features contained in the input layer, three different methods are tested. These methods include using 3D convolutions, employing cross-channel pooling via 1×1 convolution, and incorporating a Channel Attention Module.

3D Convolution. 2D convolutions and 3D convolutions both have the capability to process multi-channel data, but they differ fundamentally in how they handle inter-channel relationships. This is illustrated in Figure 5.10. In 2D convolutions (Figure 5.10a), the convolution operation is applied to each of the d channels of the input image separately using d filter kernels of dimensions $w \times h$. The resulting filtered image channels are then summed to produce a single-channel output map. While this method can capture spatial features within each channel, it is limited in its ability to model complex, non-linear interactions between channels.

In contrast, 3D convolutions (Figure 5.10b) extend the convolution operation to the channel dimension. This means that the filtering is performed not only across the spatial dimensions but also across the channel dimension. This allows 3D convolutions to extract complex, non-linear inter-channel features step by step through several consecutive convolution, activation, and pooling layers. In this work, 3D convolutional layers are integrated into the Vanilla U-Net architecture, replacing the 2D convolutions within the encoder with 3D convolutions. However, replacing all filter kernels would exponentially increase memory requirements for storing intermediate feature maps. This is because, instead of generating a single output feature map per filter kernel, a 3D convolution produces an entire output feature volume per filter. To manage the increased memory demands, the number of filter

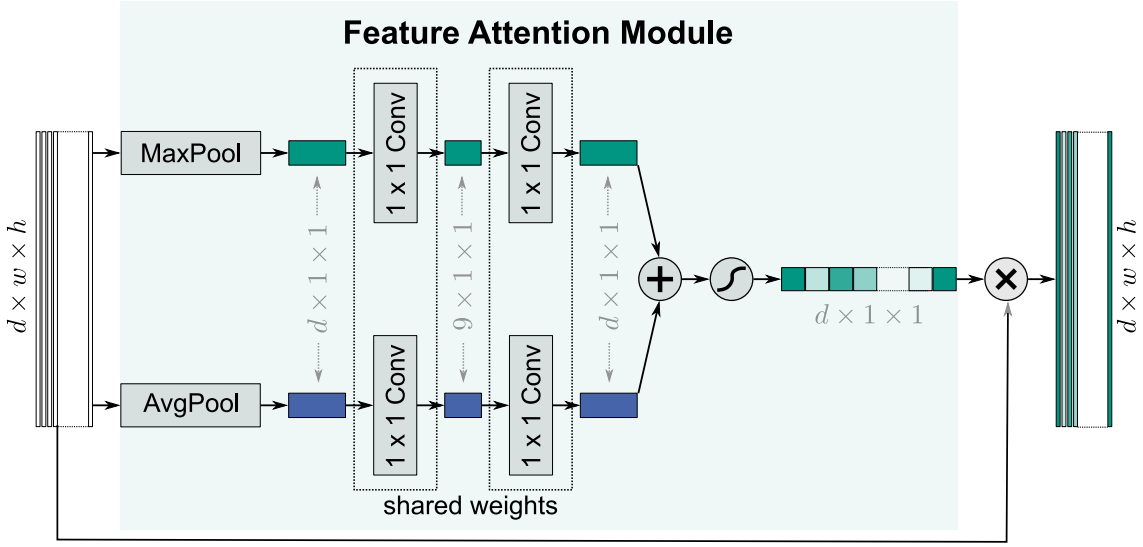


Figure 5.11.: Architecture of the Channel Attention module as proposed in [263].

kernels is reduced to eight in the first stage, 16 in the second stage, and 32 in the third stage.

Cross Channel Pooling. Cross-Channel Pooling, first proposed by Lin et al. in [275], is implemented using a 2D convolutional layer with a kernel size of 1×1 followed by ReLU activation. Placing a 1×1 convolutional layer directly after the input layer performs weighted summations of the input features, effectively combining them into new features that capture their interrelationships. The subsequent ReLU activation introduces non-linearity, enabling the network to learn complex interactions between polarimetric and interferometric features. This method can be interpreted as projecting input features into a latent feature space whose dimension is defined by the number of filter kernels used in the 1×1 convolutional layer. To ensure better comparability with other cross channel extraction approaches, the number of filters is chosen to be equal to the dimension of the input data. In this way, the channel dimension of the input data is retained.

Channel Attention Module. The motivation behind the Channel Attention Module proposed by Woo et al. [263], illustrated in Figure 5.11, is to enhance the feature representation by focusing on the most informative channels. By weighting the feature maps, the model can emphasize the channels that are more relevant to the task at hand. Originally, the Channel Attention Module is applied in the context of RGB image classification to recalibrate intermediate feature maps generated by convolutional filters. In contrast, in this work, the Channel Attention Module is applied directly to the multi-channel input image to recalibrate polarimetric and interferometric features before spatial filtering by subsequent convolutional layers of a CNN.

The Channel Attention Module generates a channel attention map which is then used to recalibrate the multi-channel input image. The steps are as follows:

1. **Global Average Pooling and Global Max Pooling**: The input multi-channel input image $\mathbf{I} \in \mathbb{R}^{d \times h \times w}$ is passed through two pooling operations: Global Average Pooling and Global Max Pooling. These operations generate two different context descriptors: $\mathbf{F}_{\text{avg}} \in \mathbb{R}^{d \times 1 \times 1}$ and $\mathbf{F}_{\text{max}} \in \mathbb{R}^{d \times 1 \times 1}$. Global Average Pooling aggregates each input channel to a scalar value by taking the average of all spatial elements, while Global Max Pooling aggregates each input channel to a scalar value by taking the maximum value among all spatial elements.

2. **Shared Network**: Both context descriptors are passed through a shared network consisting of two consecutive 1×1 convolutional layers.

3. **Combination and Sigmoid Activation**: The outputs from the convolutional layers are combined using element-wise summation. This combined output is then passed through a sigmoid activation function to produce the channel attention map of size $d \times 1 \times 1$.

4. **Feature Recalibration**: The original multi-channel input image $\mathbf{I}$ is multiplied (using element-wise multiplication) by the channel attention map to produce the recalibrated multi-channel image.

5.4.6 Model Training

The training strategy for the CNN models involves selecting the loss function and optimization algorithm to ensure effective learning. This section details the loss function used to measure the discrepancy between predicted and reference labels and the optimization algorithm employed to minimize this loss and update the parameters of the network.

Loss Function. Instead of using the standard categorical cross-entropy loss used in U-Net-based classification, this work employs the Focal Tversky Loss (FTL) proposed in [276]. This decision is driven by the significant class imbalance observed in the training dataset, as shown in Table 3.2. Initial tests using categorical cross-entropy loss revealed that the model disregards the underrepresented classes *Peat bog* and *High vegetation* after a few training epochs. Even a class-specific weighted categorical cross-entropy loss function did not lead to stable training results. The adoption of FTL is based on findings from a survey on loss functions for semantic classification conducted in [277]. FTL incorporates the Tversky similarity index, which quantifies the overlap between predicted and true classes

and facilitates the balancing of False Positives and False Negatives. The Tversky similarity index is defined as:

$$TI_c = \frac{\sum_{i=1}^{N} \hat{y}_{ic} y_{ic}}{\sum_{i=1}^{N} \hat{y}_{ic} y_{ic} + \alpha \sum_{i=1}^{N} \hat{y}_{i\bar{c}} y_{ic} + \beta \sum_{i=1}^{N} \hat{y}_{ic} y_{i\bar{c}} + \epsilon}, \tag{5.25}$$

where $\hat{y}$ and y denote the predicted and true probabilities of a pixel belonging to class c or $\bar{c}$, The constant ϵ prevents division by zero to ensure numeric stability. N is the number of considered pixels. The hyperparameters α and β, which sum up to 1, determine the weighting between False Positives and False Negatives in the Tversky similarity index. [For $\alpha = \beta = 0.5$, the Tversky similarity simplifies to the Dice loss. Setting $\beta > \alpha$ allows the training focus to shift towards minimizing False Negatives, which helps prevent the neglect of underrepresented classes.] The FTL extends the Tversky similarity index by introducing a focal parameter γ, which further enables the model to prioritize samples that are hard to classify. The loss function is defined as:

$$FTL_c = \sum_{c} (1 - TI_c)^{\frac{1}{\gamma}}. \tag{5.26}$$

Based on results from [276], the hyperparameters of the loss function in all experiments are set to $\alpha = 0.3$, $\beta = 0.7$, and $\gamma = 0.75$.

Optimization. Model training is conducted for a maximum of 300 epochs, with early termination if no improvement in validation accuracy is observed after 30 epochs. The learning rate, a crucial parameter influencing training convergence, is adjusted using a dynamic approach based on the cosine annealed warm restart learning scheduler proposed in [278]. This approach reduces the learning rate following a cosine function and includes multiple warm restarts. A warm restart is a technique where the learning rate is periodically reset to a higher value, simulating a restart of the training process but without resetting the weights of the model. This can help the model escape local minima and continue improving. During each warm restart, the initial learning rate is decreased by a factor of 0.5 to avoid abrupt changes in training weights during advanced stages. Additionally, the number of training epochs until the next restart increases by a factor of 1.2. This work uses the implementation provided by [279]. The optimization of the training loss is performed using SGD with a momentum of 0.9 and weight decay of 0.0005, following the proposed training strategy in [278].

While the described training strategy is suitable for most tested CNN models, it fails when adding the RFW layer (used for feature selection). Empirical observations (see Appendix A) indicate that for CNNs with an additional RFW layer, there is a significant reduction in test performance each time a warm restart is performed. An alternative learning rate adjustment strategy is implemented to address this issue and ensure better model convergence for CNNs with an additional RFW layer. Initially, the learning rate is set

to 0.001. If the test metric stagnates or decreases for six consecutive epochs, the learning rate is reduced to 20% of its current value. This reduction process continues iteratively until a minimum learning rate of 10^{-6} is reached.

5.5 Experimental Setup

Data Preparation. All experiments are conducted on the Pol-InSAR-Island dataset, which has been pre-divided into training and test sets using a chessboard grid structure (see Section 3.3.4). Given the necessity of making several design decisions during model selection and the choice of appropriate training strategies (e.g., selecting the optimal loss function, learning rate scheduler, and training termination criterion), the training data is further split into training and validation subsets. Specifically, each training tile was subdivided into four (geographically separate) sub-tiles, with three sub-tiles allocated for training and one for validation.

The validation data is used to make design decisions without compromising the integrity of the test data. This ensures that the test data remains a completely unseen benchmark, used exclusively for evaluating the performance of the final trained models under fixed settings.

For classification setups that include automatic feature selection or feature transformation prior to CNN classification, the training process includes learning projection functions to map high-dimensional representations to low-dimensional representations. These projection functions are learned exclusively from the training data and are subsequently applied to the entire dataset.

For the training, validation, and testing of CNN models, multi-channel image patches with a spatial dimension of 128×128 pixels are cropped from the respective areas. The chosen patch size is motivated by the spatial concentration of the class *Peat bog*, which occupies an area of approximately $300\,\text{m} \times 200\,\text{m}$. Larger patch sizes would result in an insufficient number of patches encompassing this class.

Validation and test patches are cropped without overlap, resulting in 85 validation patches and 221 test patches. Conversely, training patches are cropped with a spatial overlap of 64 pixels in both horizontal and vertical directions to enlarge the training dataset, resulting in 657 patches. Enlarging the training dataset is crucial to ensure robust model training and to mitigate overfitting. To further increase the number of training samples to 1,000 patches, data augmentation techniques are applied to randomly selected training patches. Common augmentation methods used in computer vision include rotation, flipping, changing brightness, and adding Gaussian noise. However, caution is necessary when augmenting SAR images, as it may lead to unrealistic and physically impossible data, potentially deteriorating model performance. This issue is particularly relevant for

scenes with elevated objects causing SAR-specific effects such as shadowing, layover, and foreshortening. Given that such objects are rare in the considered scene and geocoded images from various acquisition directions ensure that shadows do not have a preferential direction, the following data augmentation methods can be applied:

- Horizontal flipping,
- Vertical flipping, and
- Rotation by 90°, 180° or 270°.

Evaluation. To assess training, validation, and test accuracy, class-wise Intersection-over-Union (IoU) and mean Intersection-over-Union (mIoU) (already introduced in Section 4.4.3) are employed. Given the non-deterministic nature of CNN training processes, where different training runs can lead to different model weights, classification performance is evaluated based on 10 training and subsequent test runs to ensure well-founded evaluation. During testing, the model weights corresponding to the training stage, for which the model achieves the highest mIoU score on the validation data, are used.

To compare the classification performance of CNNs to statistical methods and traditional machine learning models, two common approaches are implemented and tested on the Pol-InSAR-Island dataset. Specifically, a supervised Wishart classifier and a Random Forest classifier are trained and evaluated. The Wishart classifier uses the complex-valued coherency matrices ($\langle \mathbf{T}_{11} \rangle$ for PolSAR and $\langle \mathbf{T}_6 \rangle$ for PolInSAR data) as input, while Random Forest models are trained on the EP and EPI feature sets for both single- and multi-frequency data. To select a suitable number of decision trees in the Random Forest, models are trained with 50, 100, and 300 trees and evaluated on the validation data. The number of decision trees is then fixed based on the highest mIoU scores obtained on the validation data. These optimized models are subsequently tested and evaluated on the held-out test data.

The conducted experiments are structured as follows:

1. **Baseline classification**: Training and evaluation of supervised Wishart and Random Forest classifiers.
2. **Preliminary analysis**: Training and evaluation of CNN-based classification of single-frequency PolSAR data.
3. **Main analysis**: Training and evaluation of CNN-based classification of multi-frequency PolInSAR data, analyzing:
    - **Feature stacking**: using $\mathrm{T}_{6,\mathrm{amp,pha}_{SL}}$ and EPI_{SL} as input,

- **Feature stacking and cross-channel feature extraction**: using EPI_{SL} as input for CNN models extended by 3D convolutions, Channel Attention Module or Cross Channel Pooling,
- **Automatic feature selection**: using three-, 15- and 40-channel input images based on MI, mRMR, RF-FI and RFW,
- **Feature transformation**: using three-, 15- and 40-channel input images based on PCA, F-UMAP, and W-UMAP.

5.6 Results

In the following, the classification results of the various tested approaches are presented. First, the results of the Baseline classifiers are shown for single- and multi-frequency PolSAR and PolInSAR image data. Second, the results of CNN-based classification of single-frequency PolSAR data are presented, comparing different input data representations. Third, the results of CNN-based classification of multi-frequency PolInSAR data are outlined, comparing feature stacking, feature selection and feature transformation approaches.

5.6.1 Baseline Results

The baseline results achieved by applying the supervised Wishart classifier and the Random Forest classifier to single-frequency PolSAR data are presented in Table 5.2. The Wishart classifier shows poor performance, with mIoU values of 0.28 for S-band and 0.24 for L-band data. Notably, the *Settlement* class achieves an IoU value of less than 0.01. This poor performance is expected due to the heterogeneity of this class, which encompasses diverse land covers such as houses, meadows, and trees, resulting in a non-representative cluster centroid in the classification process. But even less heterogeneous classes are classified unreliably, with IoU values often significantly below 0.5.

In contrast, the Random Forest classifier exhibits superior performance, achieving mIoU of 0.43 for both bands. Significant improvements compared to the Wishart classifier are observed across almost all classes, with the largest increase in IoU (+0.45) for the *Water* class in L-band data. Despite these enhancements, most classes still exhibit IoU values below 0.5.

Table 5.2.: Classification results of supervised Wishart and Random Forest classification for **single-frequency PolSAR** data quantified by class-wise IoU and mIoU.

			TF	W	CS	HV	WD	PB	GD	CG	US	LS	SA	SE	mIoU
PolSAR	Wishart	S	0.2972	0.5820	0.1951	0.2218	0.1604	0.2141	0.1212	0.3522	0.1717	0.3940	0.5156	0.0886	0.2762
		L	0.3440	0.3638	0.1825	0.2800	0.1650	0.0967	0.0832	0.2520	0.1356	0.3831	0.5041	0.0942	0.2404
	Random Forest	S	0.6243	0.8097	0.2623	0.2114	0.1329	0.2131	0.5326	0.4524	0.3251	0.5271	0.8117	0.2834	0.4322
		L	0.7430	0.8142	0.2683	0.2813	0.1663	0.1704	0.4750	0.3378	0.3415	0.4818	0.8085	0.3022	0.4325

Table 5.3.: Classification results of supervised Wishart and Random Forest classification for **single- and multi-frequency PolInSAR** data quantified by class-wise IoU and mIoU.

			TF	W	CS	HV	WD	PB	GD	CG	US	LS	SA	SE	mIoU
PolInSAR	Wishart	S	0.3987	0.6896	0.1928	0.2712	0.1570	0.3341	0.1232	0.2671	0.1754	0.1923	0.5991	0.0948	0.2913
		L	0.7167	0.7225	0.2289	0.3363	0.2793	0.4827	0.1266	0.3204	0.1264	0.3238	0.6048	0.1305	0.3666
		SL	0.5978	0.7519	0.2372	0.3267	0.2586	0.5836	0.1442	0.3546	0.2273	0.4770	0.6781	0.1149	0.3960
	Random Forest	S	0.8888	0.9882	0.4046	0.3628	0.4340	0.7947	0.6348	0.6118	0.5220	0.6778	0.9549	0.5654	0.6533
		L	0.9215	0.9837	0.4271	0.3737	0.5100	0.8215	0.5884	0.5181	0.4765	0.5551	0.9257	0.5710	0.6394
		SL	0.9504	0.9916	0.5169	0.4487	0.5318	0.8559	0.6807	0.6930	0.6132	0.7741	0.9693	0.6825	0.7257

The impact of incorporating interferometric and multi-frequency measurements on classification performance is detailed in Table 5.3. For the Wishart classifier, adding interferometry results in only a slight improvement for S-band data, with an increase of the mIoU from 0.28 to 0.29. For L-band data, the mIoU increases more significantly from 0.24 to 0.37, with the strongest improvements observed for the classes *Tidal flat* (+0.37), *Water* (+0.36), and *Peat bog* (+0.39). The enhanced discrimination between *Water* and *Tidal flat* can be attributed to the discriminative information provided by interferometric coherence, which is low for *Water* and high for *Tidal flat* due to different temporal stability of backscattering processes.

Conversely, the Random Forest classifier shows substantial performance gains with the addition of interferometry for both bands. For S-band data, the mIoU increases from 0.43 to 0.65, and for L-band data, from 0.43 to 0.64, reflecting improvements across all classes. Based on the UMAP-based visual analysis of the data structure presented in Chapter 4 (see Figure 4.12), these improvements in class separability were expected.

Incorporating both frequency bands into the classification process (denoted as SL in the tables) raises the mIoU of the Wishart classifier by 0.03, yet it remains relatively low at 0.40. For the Random Forest classifier, this combination results in an improvement of 0.07, resulting in a satisfactory mIoU of 0.73. Reliable recognition (IoU > 0.95) for classes *Tidal flat*, *Water*, and *Sand* is observed. However, certain classes, including *Coastal shrub*, *High vegetation*, and *White dune*, continue to pose classification challenges. These results also align with the UMAP-based visual representation of the data structure, which is shown again at this point in Figure 5.12 for convenience. Points belonging to *Tidal flat Water*, and *Sand* each form well-separated clusters, while strong overlaps are observable between *White dune* and *Grey dune* and between *Coastal shrub* and *High vegetation*.

To provide a comprehensive depiction of classification results based on multi-frequency PolInSAR data, confusion matrices for the Wishart and Random Forest classifiers are given in Figure 5.13 and Figure 5.14, respectively. The rows of these matrices represent the reference classes, while the columns indicate the predicted classes. The cell coloration reflects the row-normalized values.

For the Wishart classifier, three frequently occurring misclassifications, highlighted with red-dashed boxes in Figure 5.13, are evident. These include confusion between *Coastal shrub* and *High vegetation*, confusion among *Grey dune*, *Coach grass*, *Upper salt marsh*, and *Lower salt marsh*, frequent False Positives for *Sand*, and False Negatives for *Settlement*.

In the Random Forest confusion matrix (Figure 5.14), similar confusion occurs between *Coastal shrub* and *High vegetation*. However, confusion among *Grey dune*, *Coach grass*, *Upper salt marsh*, and *Lower salt marsh* is significantly reduced. The False Positive rate for *Sand* is also considerably lower. While the False Negative rate for *Settlement* decreases, the False Positive rate increases due to confusion with *High vegetation*, *Grey dune*, and *Coastal shrub*.

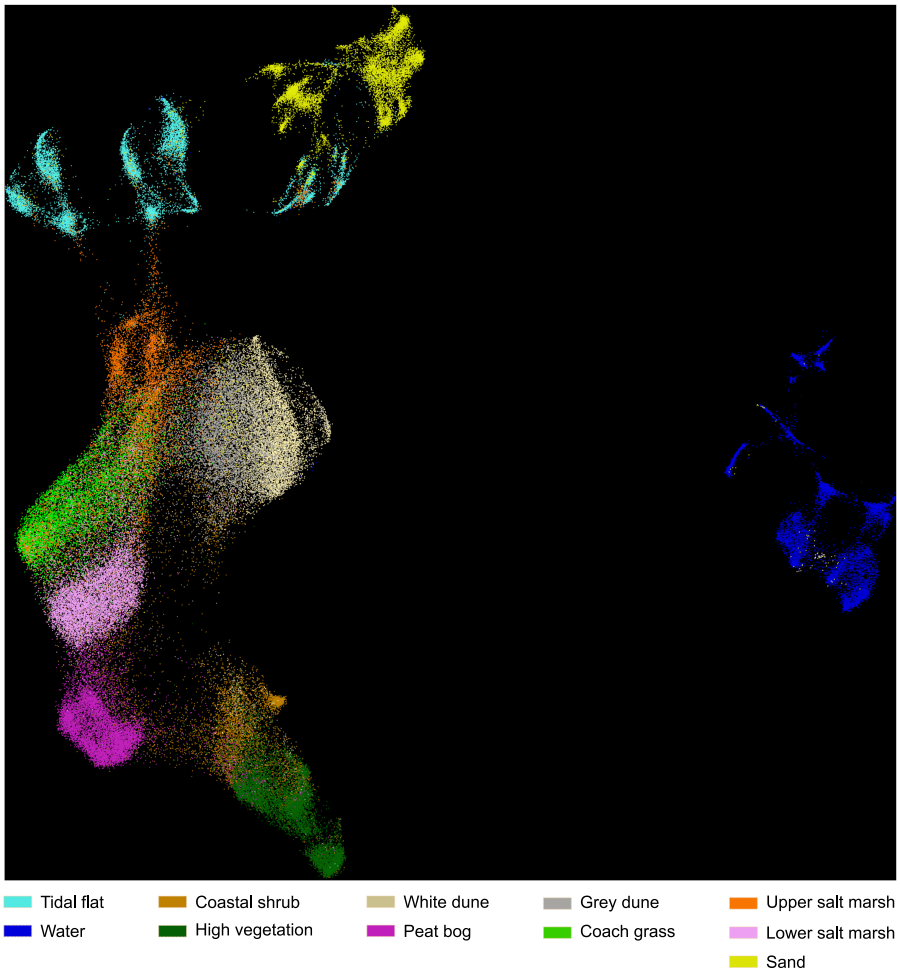


Figure 5.12.: F-UMAP scatter plot for multi-frequency PolInSAR data.

To further illustrate the classification results, Figure 5.15 displays the resulting land cover classification maps and the reference class map. The training areas are marked with red hatching on the reference class map. It is apparent that the chessboard pattern used for partitioning the data into training and test sets is not ideal due to the significant correlation among neighboring pixels. Nevertheless, as explained in Section 3.3.4, this partitioning strategy was necessary to ensure sufficient representation of data diversity, specifically regarding class coverage and incidence angle, in both subsets. Alongside the classification maps, UMAP color composites are provided in Figure 5.15b and Figure 5.15d to visualize the data underlying the classifications (Wishart distance between coherency matrices $\langle \mathbf{T}_6 \rangle$ and the $\mathrm{EPI}_{S,L}$ feature set).

Observations indicate that the confusion among *Coach grass*, *Upper salt marsh*, and *Lower salt marsh* in the Wishart classification can be attributed to the incidence angle dependence of the data. This conclusion is drawn by comparing the spatial class distribution in the classification map to the color transitions in the UMAP color composite. Further misclassifications are evident in the near-range, particularly over *Tidal flats* and *Water*.

Figure 5.13.: Classification results of the Wishart classifier for multi-frequency PolInSAR data given as confusion matrix.

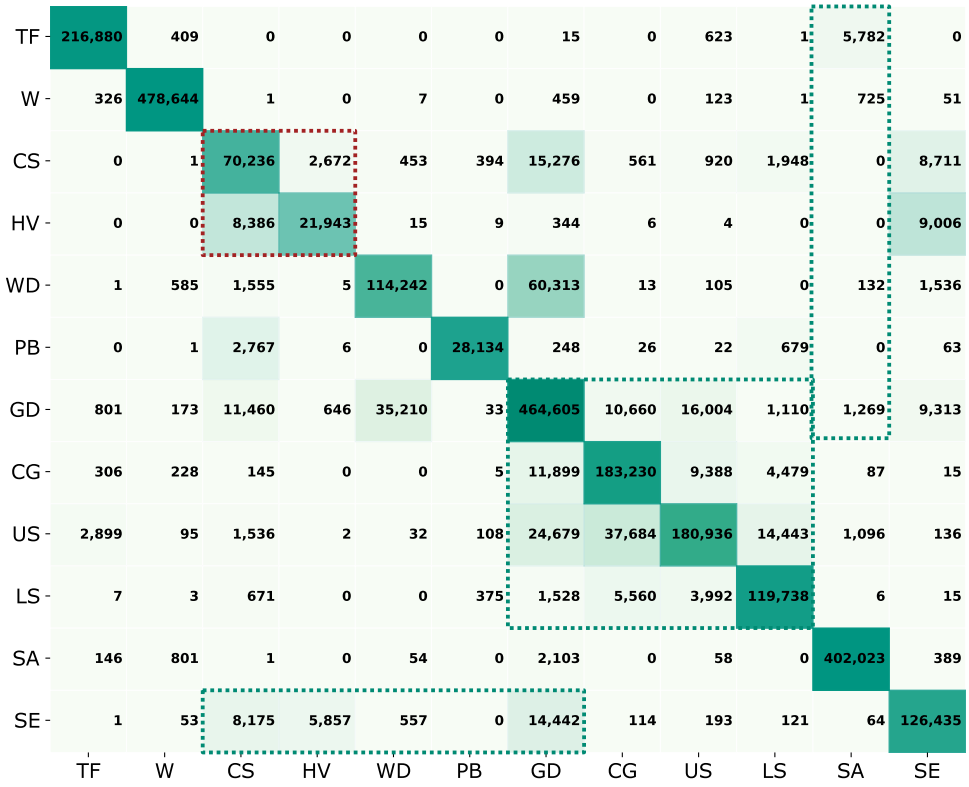
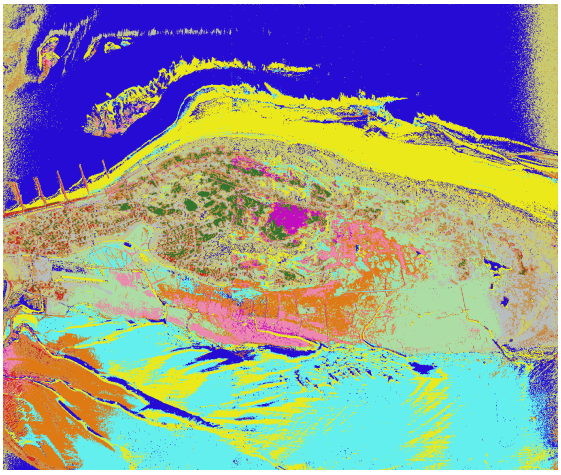
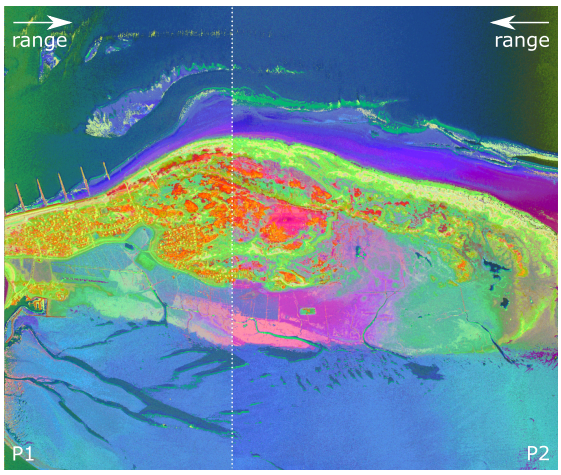


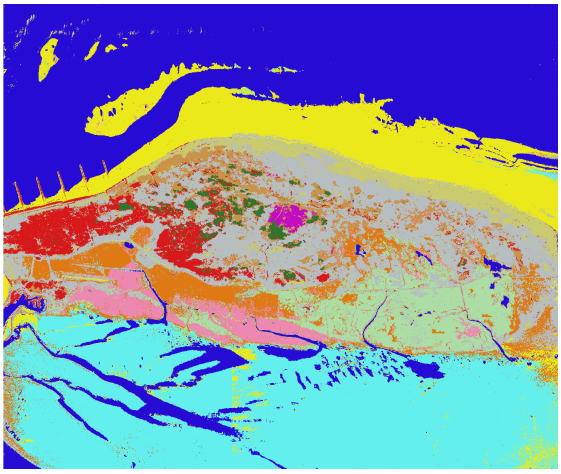
Figure 5.14.: Classification results of the Random Forest classifier for multi-frequency PolInSAR data given as confusion matrix.



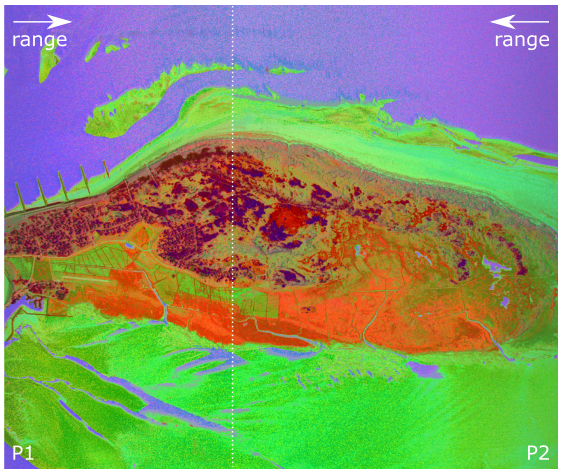
(a) Wishart classifier result.



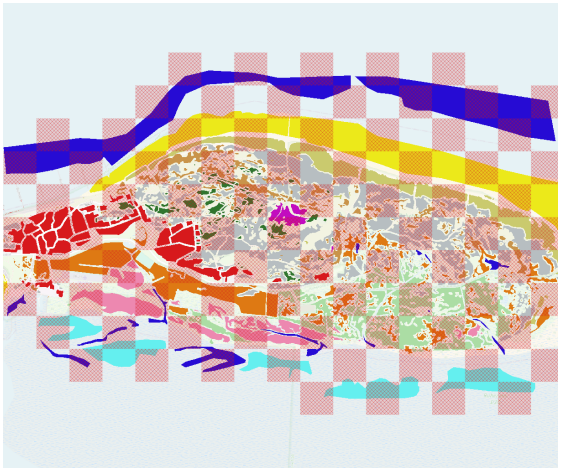
(b) W-UMAP color composite.



(c) Random Forest result.



(d) F-UMAP color composite.



(e) Reference class map.

Figure 5.15.: Land cover classification maps derived by Wishart classifier (a) and Random Forest classifier (c) based on multi-frequency PolInSAR data alongside UMAP color composites in (b) and (d) and the reference map in (e).

Additionally, False Negatives for *Sand* and *Tidal flat* are observed in the interior of the island that is actually covered by *Grey dunes*.

The Random Forest classification result aligns more closely with the reference class map, with a significantly lower proportion of misclassification in the near range. Nonetheless, misclassifications remain, which are examined again in more detail as part of the comparison with CNN results in Section 5.6.3.

5.6.2 Single-frequency PolSAR Classification

This section examines the suitability of different real-valued representations of single-frequency PolSAR data as input for CNNs. The nine representations of PolSAR data described in Section 5.4.1 are evaluated using six setups that vary by frequency bands (S and L) and CNN architectures (Vanilla U-Net, ResNet U-Net, and EfficientNet U-Net). This variation prevents a data- and model-specific bias in the evaluation process.

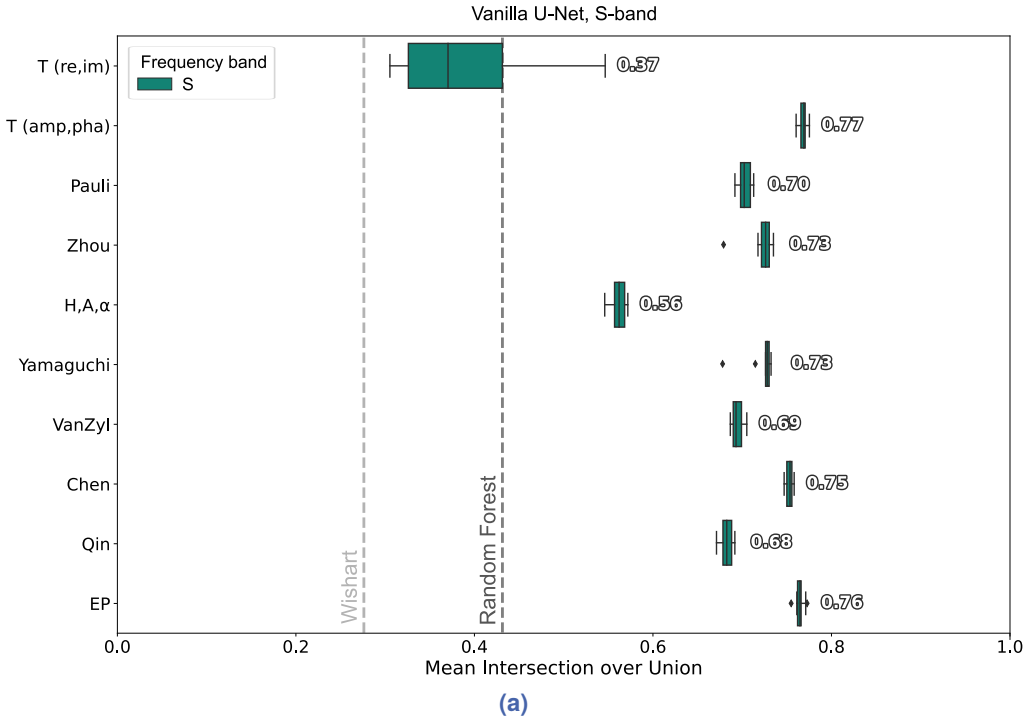
The results for each test setup are illustrated in Figures 5.16 to 5.18. The figures show box plots that present the test result statistics based on the mIoU for each input representation across 10 test runs. Aggregated results across all test setups are shown in Figure 5.19.

In the box plots:

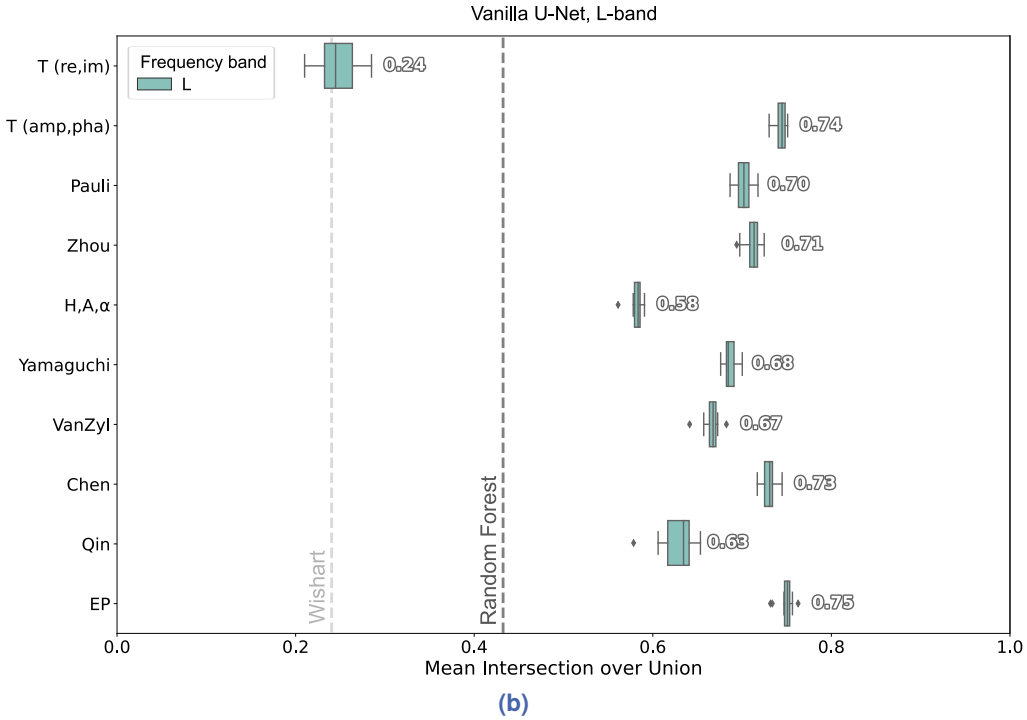
- The central line indicates the median mIoU value.
- The box spans from the lower quartile (25^{th} percentile) to the upper quartile (75^{th} percentile), representing the interquartile range.
- The whiskers extend to 1.5 times the interquartile range from the quartiles.
- Outliers beyond the whiskers are depicted as individual points.

Without going into the specific results for every test setup, the following paragraphs describe the observed general trends.

Impact of Test Configurations. The results reveal that the classification performance varies depending on the frequency band and the used CNN architecture. Notably, the S-band data consistently outperforms the L-band data, indicating its higher suitability for the PolSAR classification task posed in this study. Among the CNN architectures, EfficientNet U-Net demonstrates inferior classification performance compared to ResNet and Vanilla U-Net, which can be attributed to its lower number of trainable parameters. While ResNet and Vanilla U-Net show comparable results, the latter shows more consistent performance across different training runs, indicating its robustness to variations in initialization.

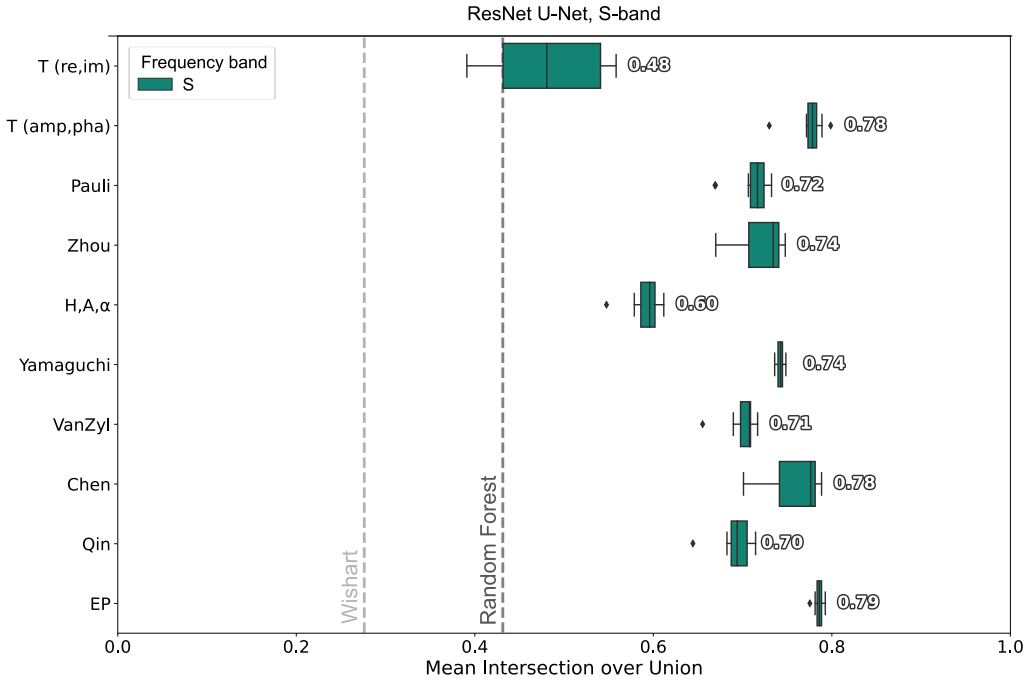


(a)

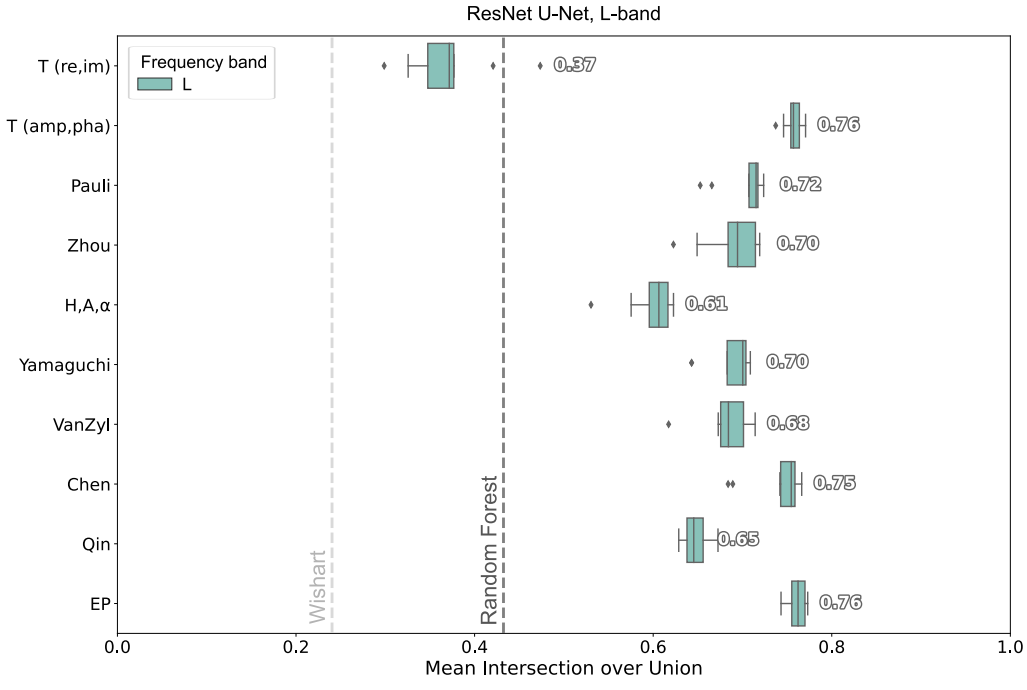


(b)

Figure 5.16.: Classification results achieved with Vanilla U-Net using different PolSAR representations as input data. The boxes show the test statistics of mIoU based on 10 training and test runs using (a) S-band or (b) L-band data.



(a)



(b)

Figure 5.17.: Classification results achieved with ResNet U-Net using different PolSAR representations as input data. The boxes show the test statistics of mIoU based on 10 training and test runs using (a) S-band or (b) L-band data.

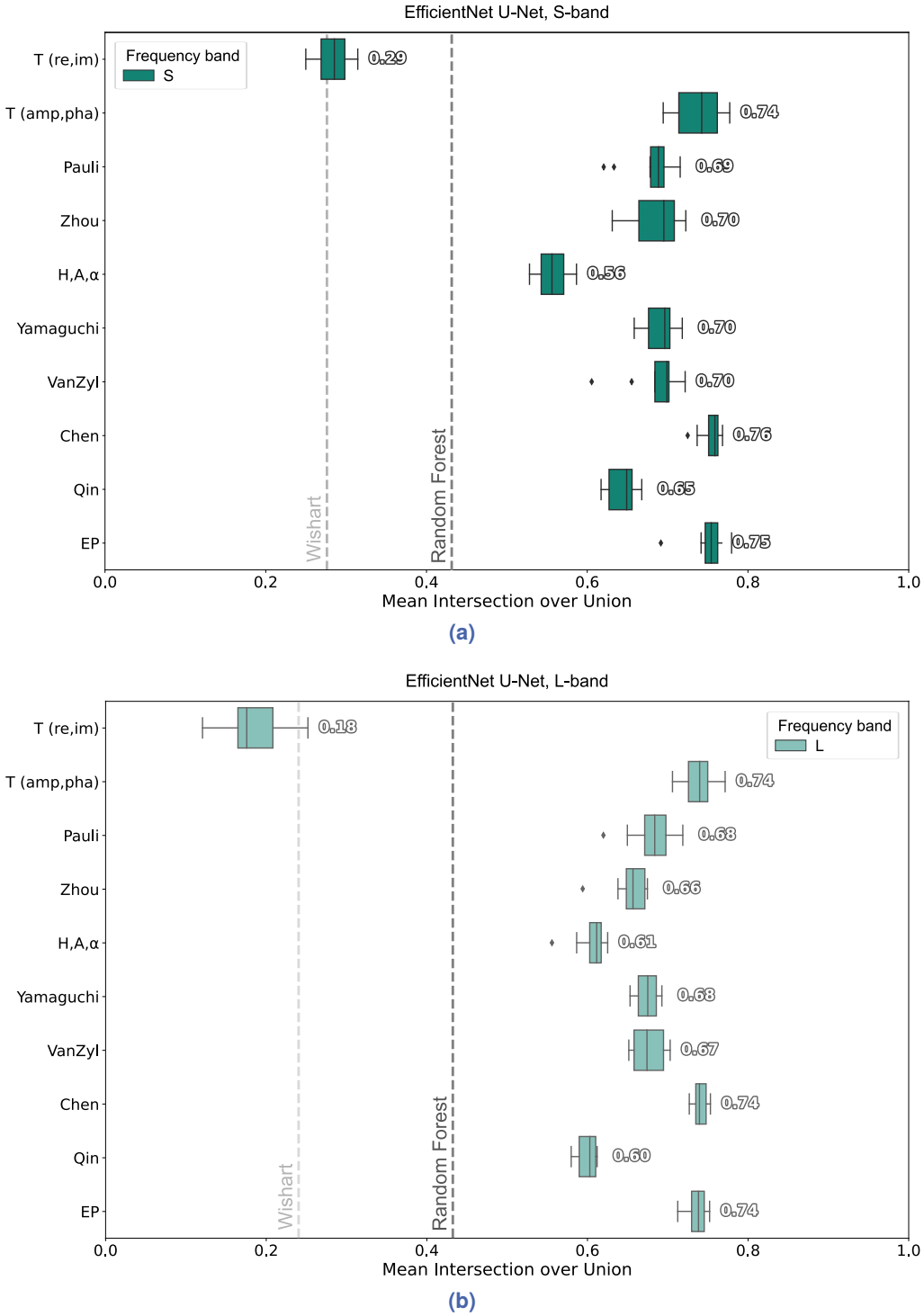


Figure 5.18.: Classification results achieved with EfficientNet U-Net using different PolSAR representations as input data. The boxes show the test statistics of mIoU based on 10 training and test runs using (a) S-band or (b) L-band data.

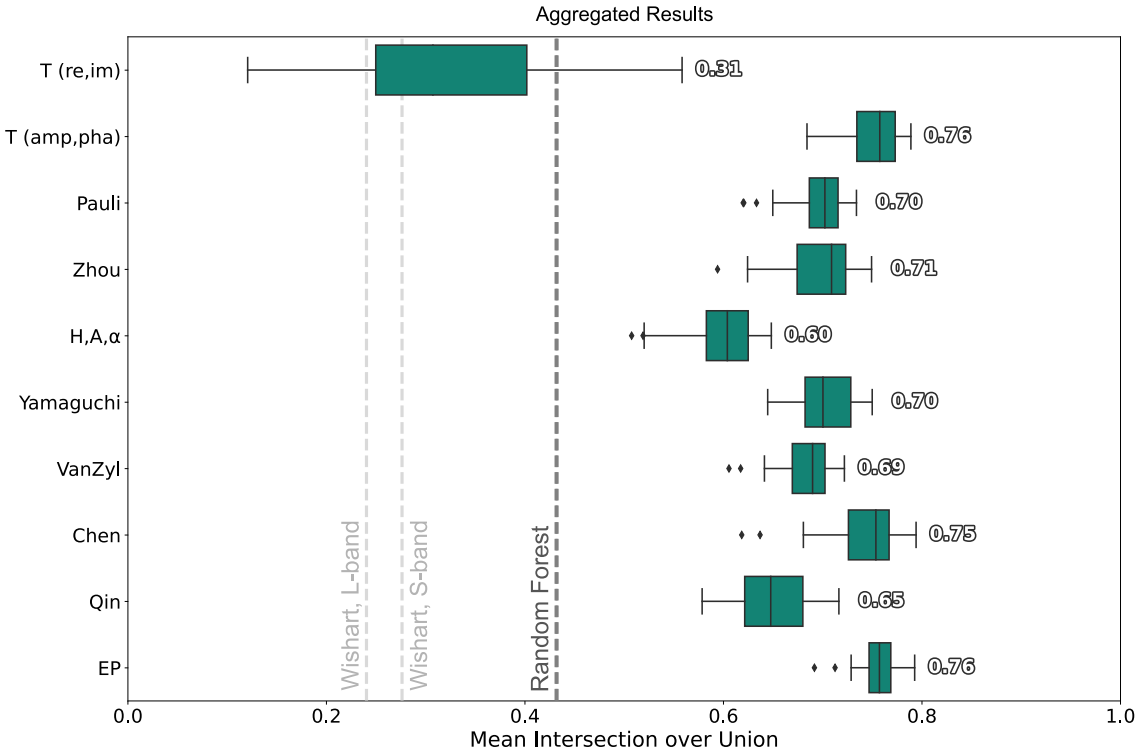


Figure 5.19.: Test statistics of mIoU aggregated across frequency bands and network architectures for different input representations.

Impact of Input Representation. While the frequency band and CNN architecture clearly affect classification results, the results reveal that selecting the appropriate input representation has an even more profound impact on performance.

The $T_{re,im}$ representation shows very poor performance across all test setups, with mIoU median values consistently below the Random Forest baseline (except for ResNet U-Net on S-band data). The poor performance may be attributed to the separate processing of real and imaginary parts, resulting in information loss. In contrast, the H-A-α representation yields better results, outperforming the Random Forest classification, despite relying only on three selected features. The improved performance over Random Forest can be attributed to the ability of CNNs to learn spatial features that are more powerful for class discrimination than additional polarimetric features. However, the H-A-α representation still ranks among the least effective feature sets for CNN classification across all test setups. A possible reason that leads to its poorer performance compared to other input representations may be the lack of absolute intensity information.

The Qin representation performs slightly better than H-A-α but remains among the lower-performing feature sets, despite utilizing an extensive data description with 16 features. Presumably, the inclusion of $T_{re,im}$ has a negative effect on the classification performance. This assumption is confirmed by subsequent experiments (see Appendix B), in which replacing $T_{re,im}$ with $T_{amp,pha}$ leads to significant improvements.

The Zhou, Pauli, Yamaguchi, and VanZyl feature sets show comparable results, ranking in the middle range. The similar performance of Pauli, Yamaguchi, and VanZyl is expected, as they all represent PolSAR data by quantifying surface, double-bounce, and volume scattering intensity.

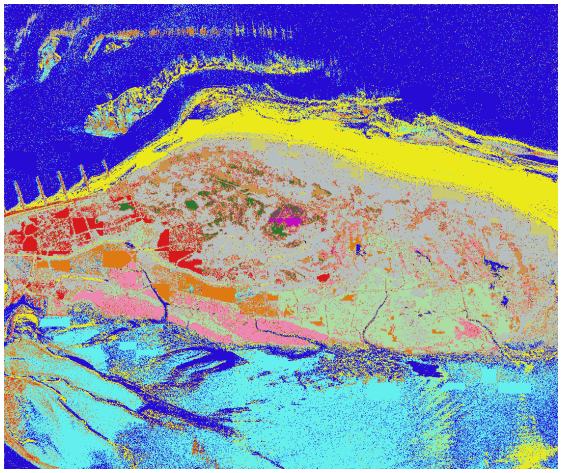
The Chen, EP, and $T_{amp,pha}$ feature sets consistently rank among the top three representations across all test setups. The superior performance of the Chen representation compared to the H-A-α highlights the importance of intensities (represented by the span feature) and target orientation (represented by null-angle features) to distinguish land cover classes. The strong performance of the EP feature set, containing 42 features, demonstrates the ability of the CNNs to identify relevant information from numerous channels and manage redundancies effectively. However, the comparable performance of the generic representation $T_{amp,pha}$ indicates that extracting numerous polarimetric features is not necessary for achieving good classification performance. Although CNNs are not capable of extracting physically interpretable polarimetric features, they can reliably classify unseen data by learning and recognizing patterns in amplitude and phase images. Thus, CNNs sufficiently well model the relationship between the image data represented by coherency matrices and the target classes.

Comparison to Random Forest. To evaluate the differences in classification results between pixel-based Random Forest and image-based CNN classification in more detail, the class-wise IoU values are presented in Table 5.4, and the land cover maps of both approaches are shown in Figure 5.20 for the entire region and in Figure 5.21 for a smaller sample section to show differences in detail. Additionally, evaluation maps in Figure 5.22 indicate the spatial distribution of correct predictions (green) and false predictions (red). For the CNN-based classification, the test results correspond to models that achieve the best mIoU on the validation data during the training process. Specifically, for S-band data, the ResNet U-Net model with the $T_{amp,pha}$ representation as input is used, and for L-band data, the ResNet U-Net model with the EP feature set as input is employed. Using the CNN approach, significant improvements appear across all classes for both frequency bands (see Table 5.4). As illustrated in Figure 5.21 and Figure 5.22, the classes *Peat bog* and *White dune*, which are scarcely detected by the Random Forest method (i.e., high False Negative rate), are much more reliably identified by the CNN. This shows, in particular, that the CNN is able to correctly identify land covers that cannot be clearly separated using polarimetric and interferometric data alone (see also overlap in the UMAP visualization in Figure 5.12). Furthermore, in the Random Forest prediction map, contiguous clusters of False Positives of the classes *Settlement* (red) and *Lower salt marsh* (rose) occur in the dune area, which, in fact, consists of the classes *Grey dune* and *Coastal shrub*. These misclassifications are completely eliminated using the CNN model. Moreover, the improvement in classification results is reflected in a more consistent land cover map, where neighboring pixels are generally assigned to the same class. This results in homogeneously classified areas and

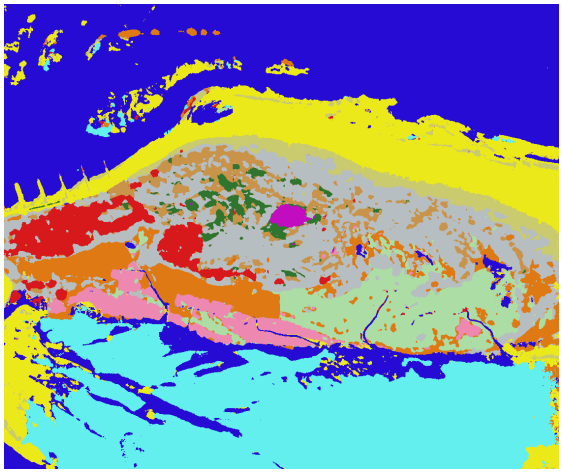
prevents isolated misclassifications. This improvement in consistency was expected since Random Forest classification is separately applied to single pixels, while in CNN classification, spatial neighborhoods are considered through learning and extracting spatial features in the image domain.

Table 5.4.: Classification results of Random Forest and CNN classification for **single-frequency PolSAR** data quantified by class-wise IoU and mIoU.

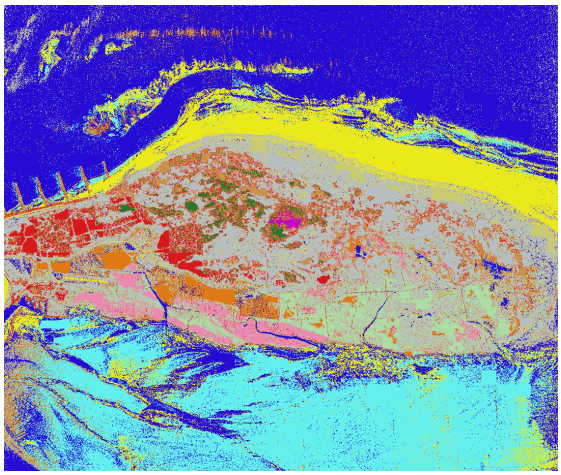
			TF	W	CS	HV	WD	PB	GD	CG	US	LS	SA	SE	mIoU
PolSAR	S	Random Forest	0.6243	0.8097	0.2623	0.2114	0.1329	0.2131	0.5326	0.4524	0.3251	0.5271	0.8117	0.2834	0.4322
		CNN	0.9813	0.9712	0.6045	0.4567	0.6916	0.9296	0.7903	0.7465	0.6919	0.8631	0.9759	0.8815	0.7987
	L	Random Forest	0.7430	0.8142	0.2683	0.2813	0.1663	0.1704	0.4750	0.3378	0.3415	0.4818	0.8085	0.3022	0.4325
		CNN	0.9425	0.9605	0.6027	0.5008	0.5672	0.9045	0.7289	0.7055	0.6484	0.8396	0.9724	0.8957	0.7724



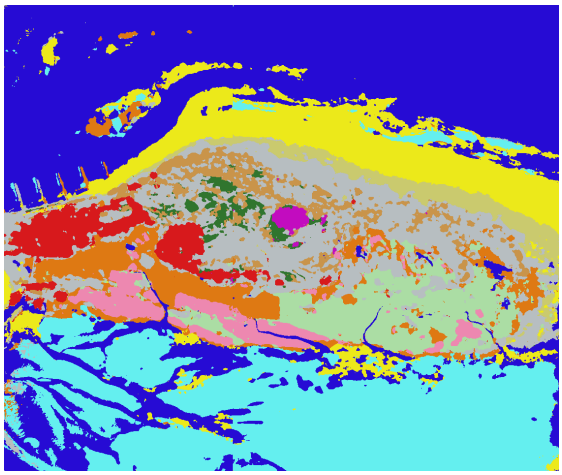
(a) Random Forest result (S-band).



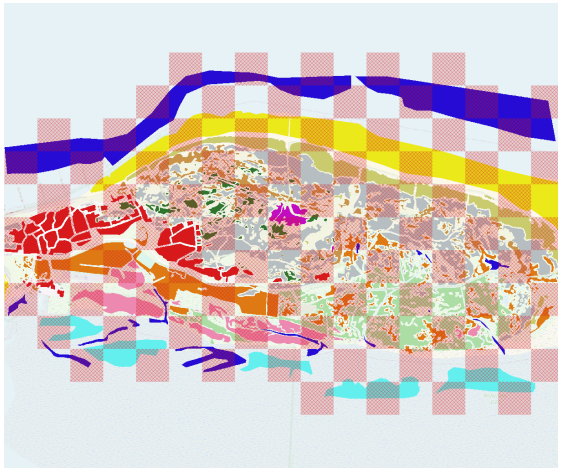
(b) CNN result (S-band).



(c) Random Forest result (L-band).

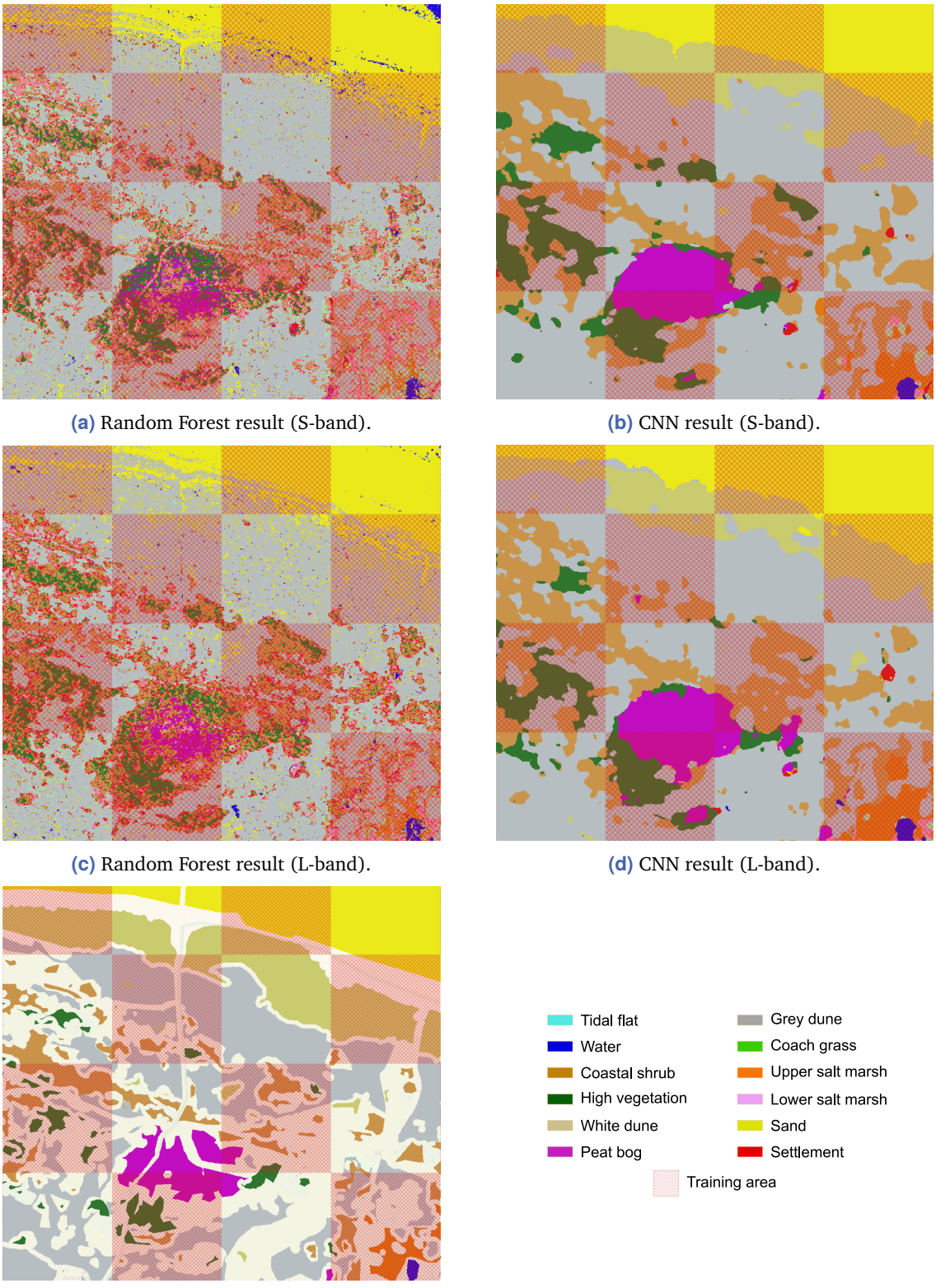


(d) CNN result (L-band).



(e) Reference class map.

Figure 5.20.: Land cover classification maps derived by Random Forest or CNN classification based on single-frequency PolSAR data for S-band data in (a) and (b) and for L-band data in (c) and (d). The reference map with marked training areas is given in (e).



(a) Random Forest result (S-band).

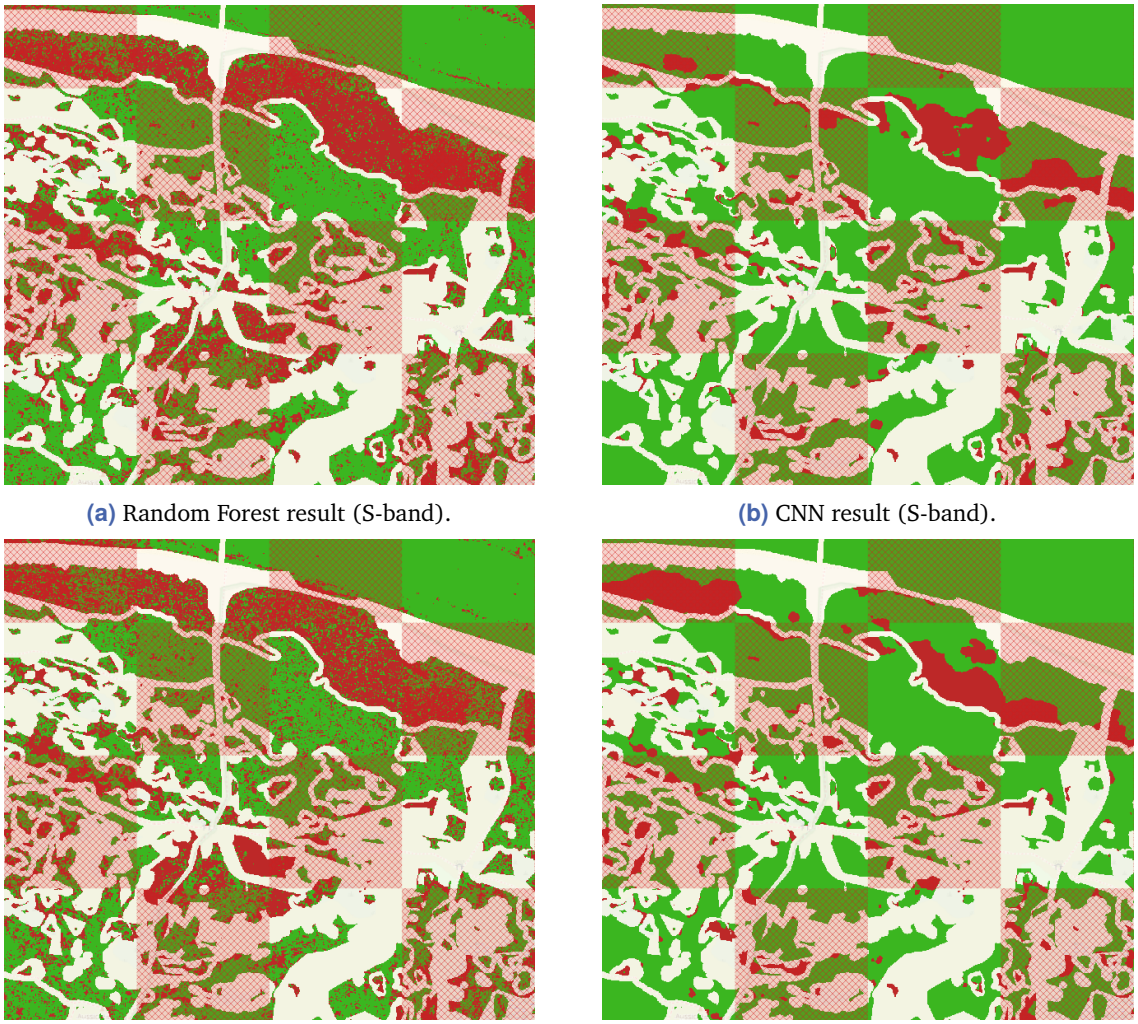
(b) CNN result (S-band).

(c) Random Forest result (L-band).

(d) CNN result (L-band).

(e) Reference class map.

Figure 5.21.: Land cover classification maps of sample section derived by Random Forest or CNN classification based on single-frequency PolSAR data for S-band data in (a) and (b) and for L-band data in (c) and (d). The reference map with marked training areas is given in (e).



(a) Random Forest result (S-band).

(b) CNN result (S-band).

(c) Random Forest result (L-band).

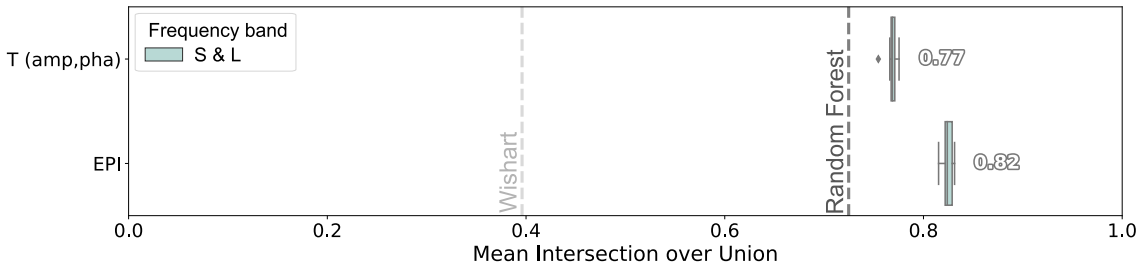
(d) CNN result (L-band).

Figure 5.22.: Evaluation maps of sample section based on Random Forest or CNN classification maps based on single-frequency PolSAR data for S-band data in (a) and (b) and for L-band data in (c) and (d). Correct predictions are colored in green, and false predictions in red.

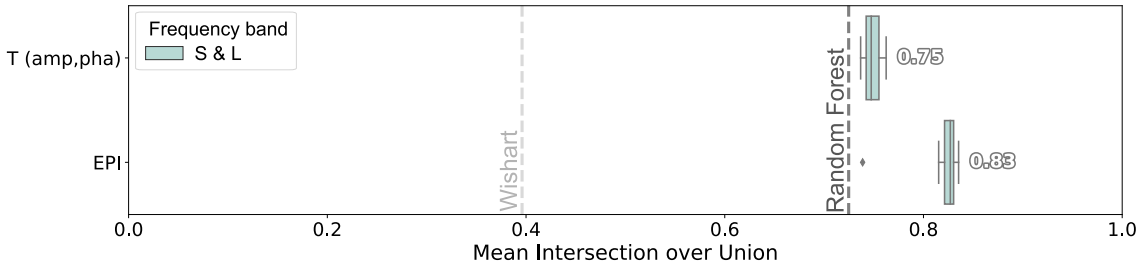
5.6.3 Multi-frequency PolInSAR Classification - Feature Stacking

In the following, the results of multi-frequency PolInSAR image classification based on feature stacking are presented. As input for the CNNs, the representations $\mathrm{T}_{6,\mathrm{amp,pha}_{SL}}$ and the full $\mathrm{EPI}_{S,L}$ feature set are used and compared.

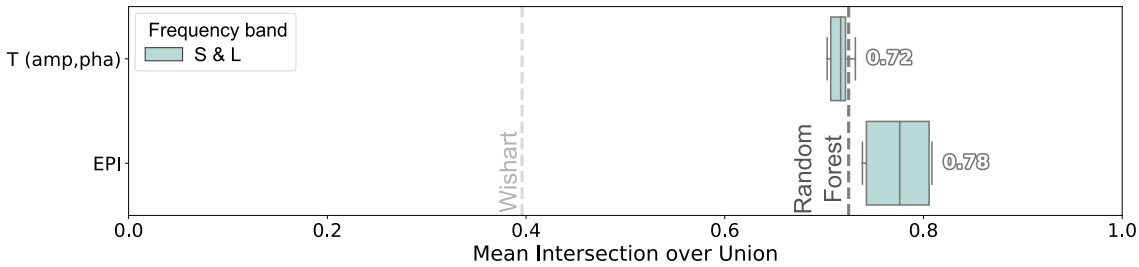
Impact of Input Representation. The results attained using varying CNN architectures are depicted in Figure 5.23. Across all test setups, using $\mathrm{EPI}_{S,L}$ yields significantly better results than $\mathrm{T}_{6,\mathrm{amp,pha}_{SL}}$, with a performance gain between 0.05 and 0.08 in mIoU (median). This contrasts with the results of CNN-based single-frequency PolSAR classification, where both representations EP and $\mathrm{T}_{\mathrm{amp,pha}}$ show comparable performance. This leads to the



(a) Results using Vanilla U-Net.



(b) Results using ResNet U-Net.



(c) Results using EfficientNet U-Net.

Figure 5.23.: Classification results achieved with Vanilla U-Net (a), ResNet U-Net (b) and EfficientNet U-Net (c) using $T_{6,\text{amp,pha},S,L}$ or $EPI_{S,L}$ to represent multi-frequency PolInSAR data. The boxes show the test statistics of mIoU based on 10 training and test runs.

conclusion that simply adding the amplitudes and phases of the interferometric block matrix $\langle \mathbf{\Omega_{12}} \rangle$ to the input representation does not allow the exploitation of the additional information from interferometry. The additional information remains unused and even negatively influences the results, leading to performances that fall short of single-frequency PolSAR classification. In contrast, the $EPI_{S,L}$ feature set achieves good mIoU with median values up to 0.83. This shows that for a successful classification of PolInSAR data, the extraction of interferometric features is necessary to effectively leverage interferometric information.

Added Value of Interferometry and Multi-frequency. To further investigate the improvements gained from incorporating interferometric information and considering multi-frequency data, Figure 5.24 presents the results of CNN classifications based on single-frequency PolSAR data (using the EP feature set) and single- and multi-frequency PolInSAR data (using the EPI and $EPI_{S,L}$ feature sets). For each setup, the CNN model with the highest median mIoU reached on validation data during training is used. Observations indicate

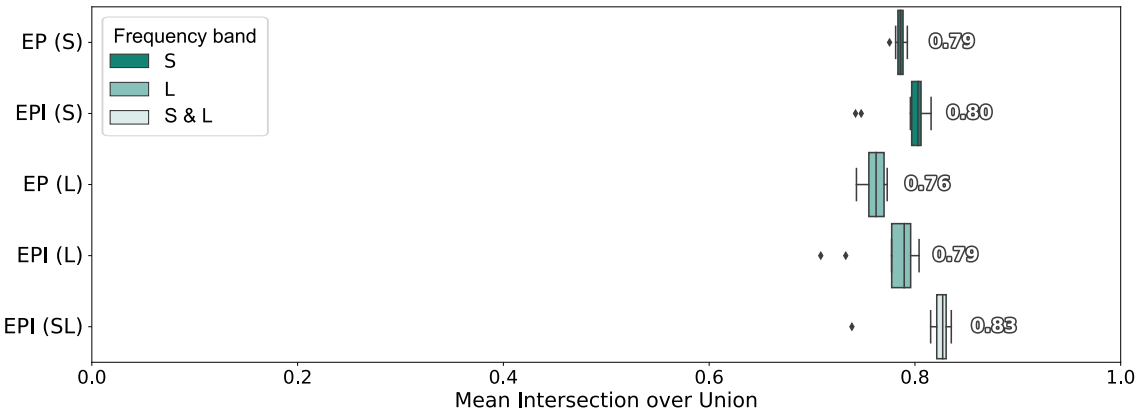
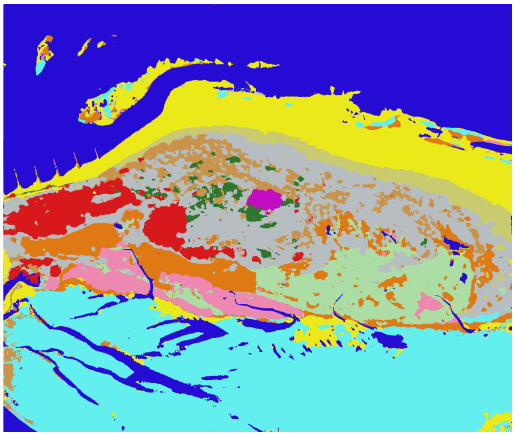


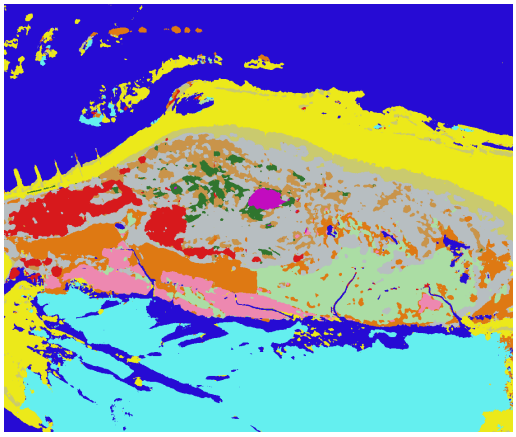
Figure 5.24.: Comparison between test statistics of CNN results based on single-frequency PolSAR (EP (S), EP (L)), single-frequency PolInSAR (EPI (S), EPI (L)) and multi-frequency PolInSAR data (EPI (SL)).

a slight improvement of +0.01 median mIoU for S-band data and +0.03 median mIoU for L-band data when adding interferometric features. Further improvement is observed when concatenating features of both frequencies, resulting in an additional increase of +0.03 in median mIoU.

The land cover map of the best-performing model for multi-frequency PolInSAR data, chosen based on performance on validation data, is shown in Figure 5.25, which is contrasted with the prediction based on single-frequency (S-band) PolSAR data. The qualitative comparison reveals improvements that are not reflected in the quantitative results due to the sparsely labeled reference map. Improvements can be seen, for example, in the more precise separation of water and land areas. This enhancement is attributed to the integration of interferometric coherence, exemplified by $|\gamma_{opt,1}|$ in Figure 5.25c. The delineation of the water-land boundary is clearly visible due to low coherence values over *Water* and high coherence values over *Tidal flat* or *Sand*. As illustrated in the zoomed regions shown in Figures 5.25e to 5.25j, the predicted water-land boundary based on multi-frequency PolInSAR data accurately follows the boundary between low and high coherence values, thus resulting in better segmentation of water areas compared to using only single-frequency PolSAR data. These improvements are expected, as results from baseline classifiers have already demonstrated that interferometry and multi-frequency data provide information complementary to polarimetric features. However, the improvement for CNN classification is less pronounced than that observed for Random Forest classification. This is illustrated in Figure 5.26 showing differences in predictions between using single-frequency PolSAR and multi-frequency PolInSAR data for Random Forest and CNN classification. A plausible explanation of this observation is that classes benefiting from added interferometric features in Random Forest classification (e.g., *Water*, *Tidal flat* and *Peat bog*) are already reliably detected without interferometric features by the CNN. This ability may be due to the capability of CNN to learn and extract image patterns, which mitigates the need for additional interferometric information, thereby diminishing the potential for further enhancement.

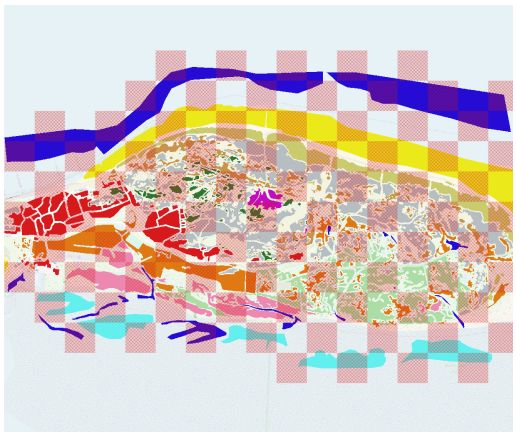


(a) Multi-frequency PolInSAR.

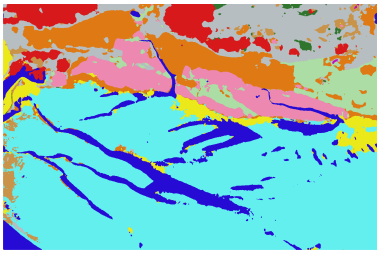


(b) Single-frequency PolSAR.

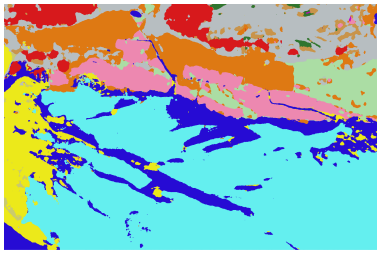
(c) Interferometric coherence.



(d) Reference class map.



(e) Multi-frequency PolInSAR.

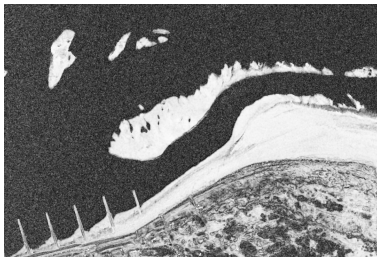


(f) Single-frequency PolSAR.

(g) Interferometric coherence.

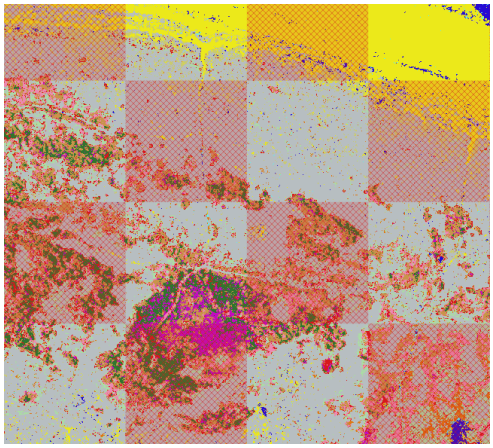
(h) Multi-frequency PolInSAR.

(i) Single-frequency PolSAR.

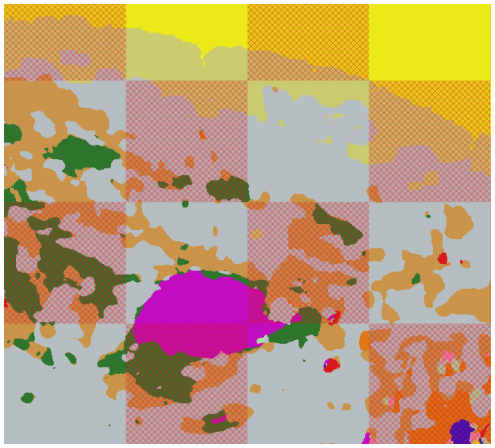


(j) Interferometric coherence.

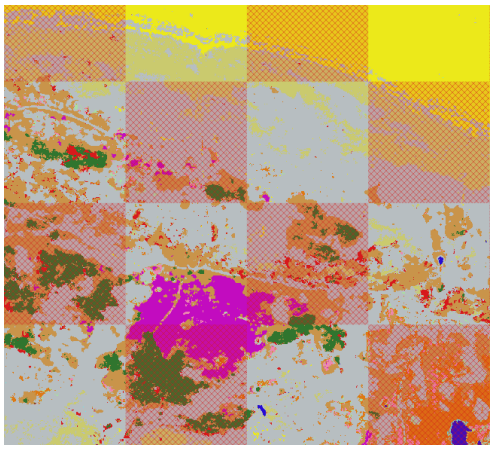
Figure 5.25.: Predicted land cover classification maps derived by CNN-based classification using multi-frequency PolInSAR data shown in (a) and single-frequency S-band PolSAR data shown in (b), alongside interferometric coherence of S-band data ($|\gamma_{opt,1}|$) in (c) and the reference map with marked training areas in (d). Subfigures (e) - (j) present close-up details.



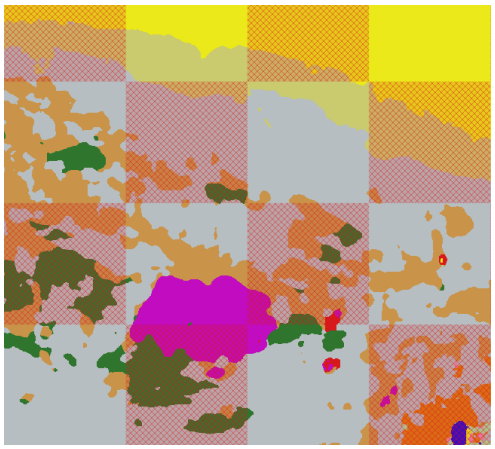
(a) Random Forest (single-frequency PolSAR).



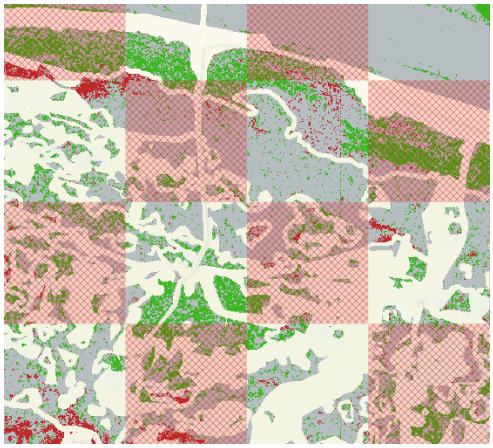
(b) CNN (single-frequency PolSAR).



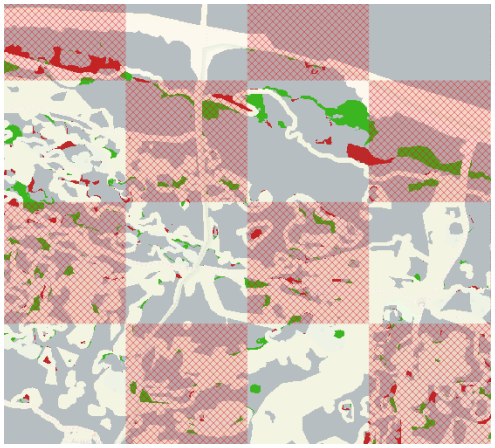
(c) Random Forest (multi-frequency PolInSAR).



(d) CNN (multi-frequency PolInSAR).



(e) Random Forest change map.



(f) CNN change map.

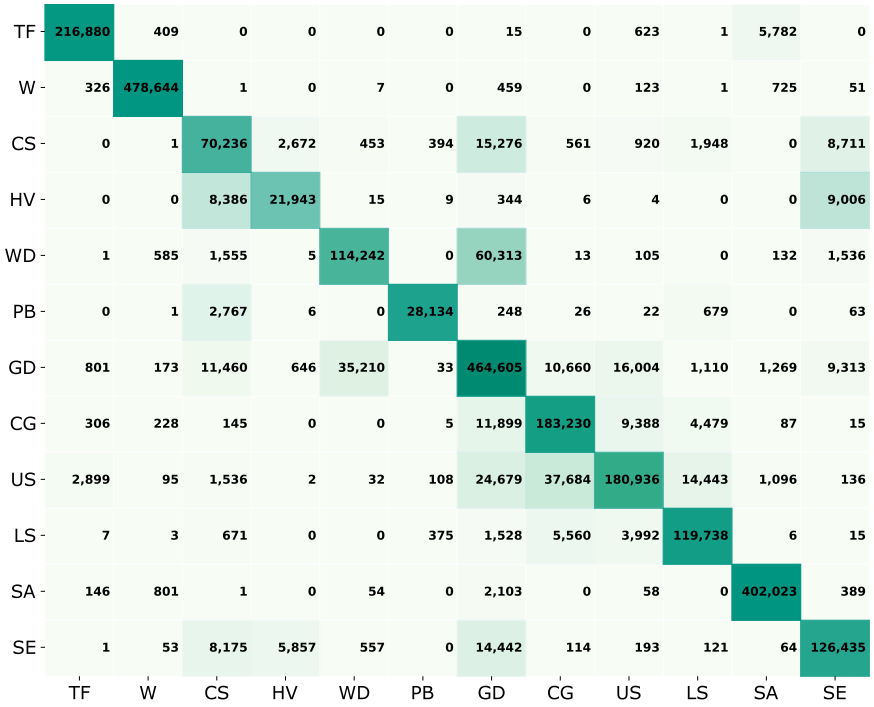
Figure 5.26.: Improvements of classification results achieved by adding interferometric and multi-frequency information. Predictions are presented in (a) - (d), while changes are indicated in (e) and (f). Grey represents no change, green indicates a shift towards correct predictions, and red denotes a shift towards incorrect predictions.

Comparison to Random Forest. When using $EPI_{S,L}$ as input, CNN classification, measured by mIoU, consistently outperforms Random Forest classification, irrespective of the CNN architecture (see Figure 5.23). Detailed class-wise IoU for the best CNN model (ResNet U-Net) compared to Random Forest results are presented in Table 5.5. Confusion matrices in Figure 5.27 further illustrate the distribution of misclassifications over the classes. The CNN classifier shows improvements in IoU for all classes, leading to an increase in mIoU from 0.73 to 0.83. Specifically, the CNN confusion matrix reveals minimal misses and almost no misclassification of the class *Peat bog*, in contrast to the Random Forest results. Additionally, there is significantly less confusion between *Coastal shrub* and *High vegetation* with *Settlement*. This improvement is primarily attributed to the inclusion of image context, enhancing the recognition of the heterogeneous class *Settlement*. The confusion between *White dune* and *Grey dune* is reduced, although distinguishing these two classes remains challenging. Another persistent issue is correctly differentiating between *Coastal shrub* and *High vegetation*.

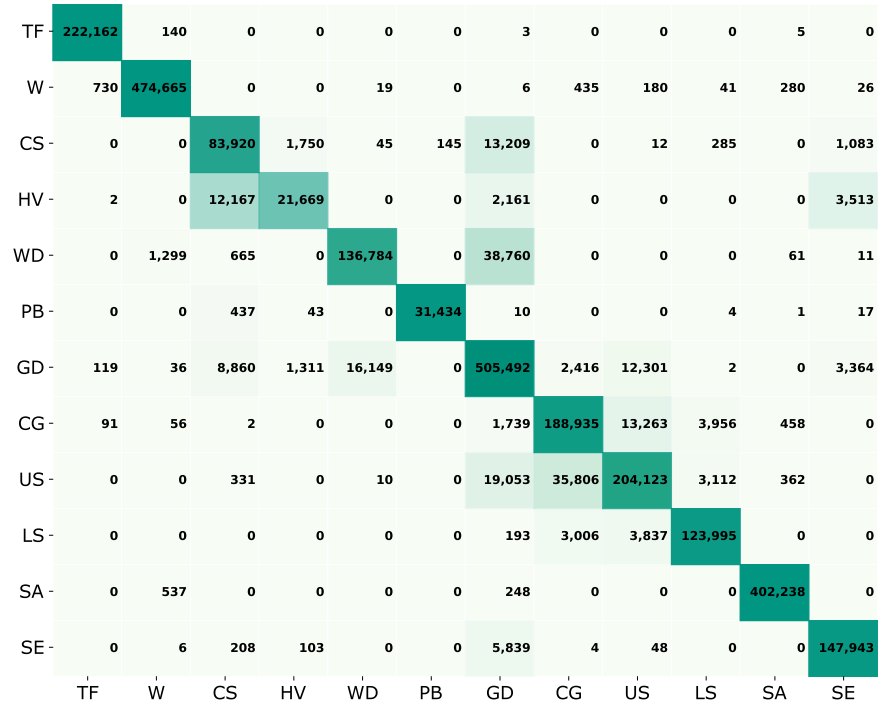
A side-by-side comparison of predicted land cover maps is shown in Figure 5.28, and the highlighted ROI (I, II) are shown in Figures 5.29 and 5.30. For a better visual evaluation, the UMAP color composites are shown in addition to the reference map to illustrate which information serves as input for the classifiers. Figure 5.29 (ROI I) shows an example of misclassifications that occur with the Random Forest and are improved by the CNN classification. In the upper left corner of the image region, the Random Forest falsely predicts *Lower salt marsh* (rose), and in the lower right corner, the classes *White dune* (dark yellow), *Settlement* (red) and *Coastal shrub* (light green) are falsely predicted. It is important to note that the misclassifications in the Random Forest classification are not limited to misclassifications of isolated pixels, so simple post-processing could probably not suppress the error sufficiently.

Table 5.5.: Classification results of Random Forest and ResNet U-Net classification for multi-frequency PolInSAR data quantified by class-wise IoU and mIoU.

	TF	W	CS	HV	WD	PB	GD	CG	US	LS	SA	SE	mIoU
Random Forest	0.9504	0.9916	0.5169	0.4487	0.5318	0.8559	0.6807	0.6930	0.6132	0.7741	0.9693	0.6825	0.7257
ResNet U-Net	0.9951	0.9921	0.6816	0.5072	0.7058	0.9795	0.8008	0.7552	0.6980	0.8957	0.9952	0.9123	0.8265

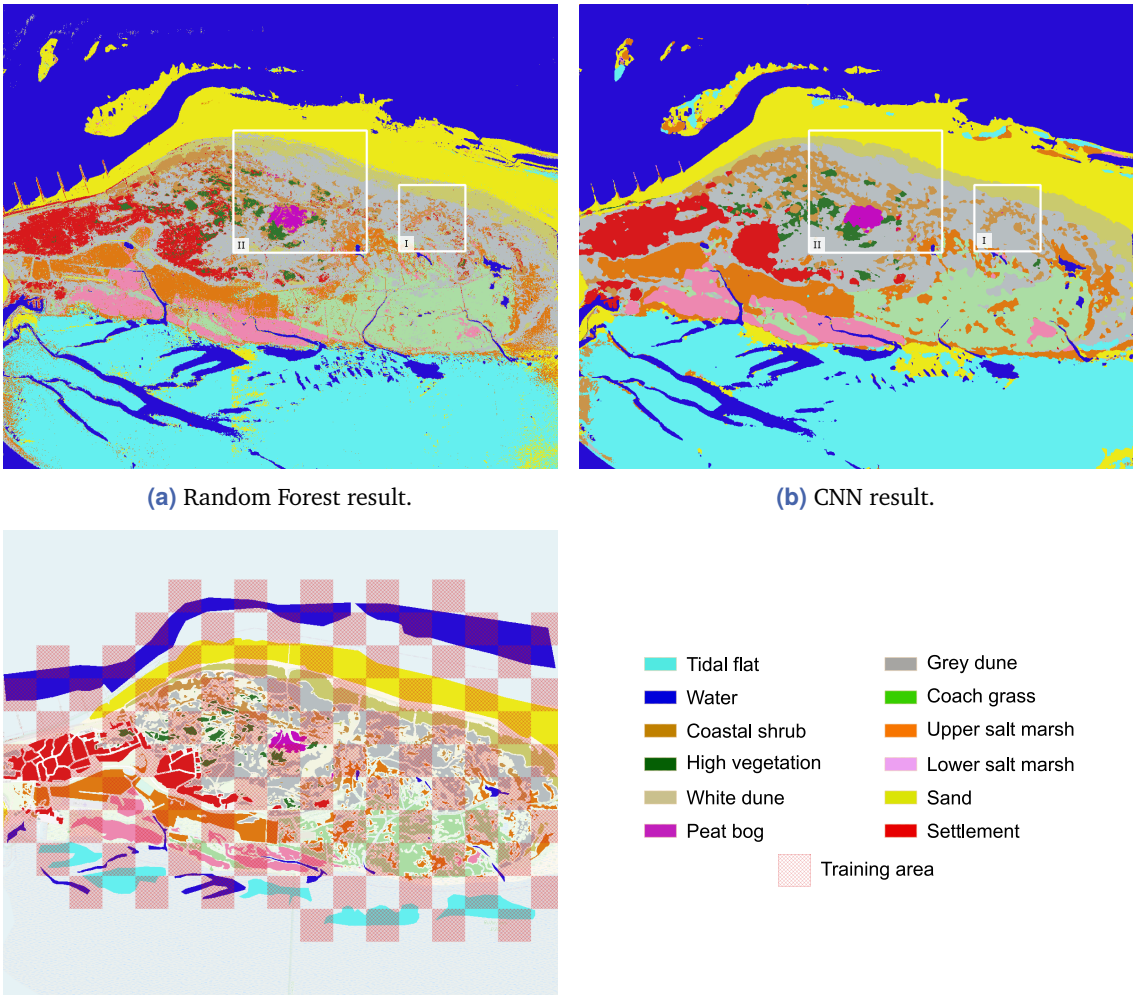


(a) Results using Random Forest.



(b) Results using ResNet U-Net.

Figure 5.27.: Confusion matrix resulting from Random Forest (a) and ResNet U-Net (b) based on multi-frequency PolInSAR data.



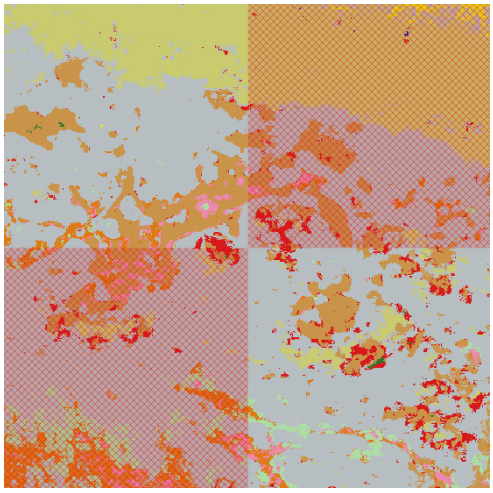
(a) Random Forest result.

(b) CNN result.

(c) Reference map.

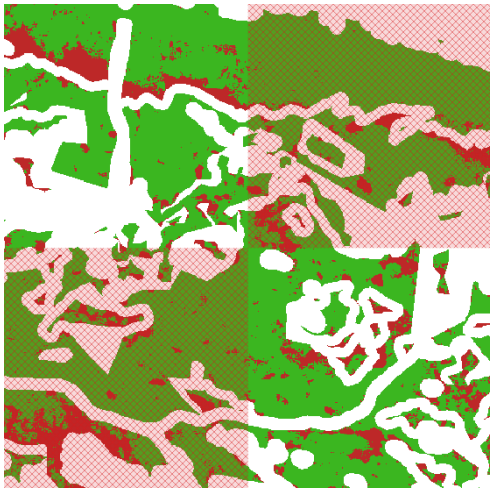
Figure 5.28.: Land cover classification maps derived by Random Forest (a) or CNN classification (b) based on multi-frequency PolInSAR data. The reference map with marked training areas is given in (c).

The predicted land cover classes of ROI II, presented in Figure 5.30 show a high degree of similarity. However, in the CNN classification map, adjacent pixels are typically assigned the same class, resulting in more coherent regions and minimizing misclassifications, while in the Random Forest classification map, isolated patches of misclassified pixels are noticeable. One advantage of Random Forest classification, on the other hand, is the better preservation of details, evident through more accurate mapping of class boundaries. This can be seen, for example, in the prediction in the *Peat bog* area (purple). The UMAP color composite indicates the presence of vegetation, denoted by the violet color, at the northern tip and in the central region of the *Peat bog* area. While the boundaries between *Peat bog* and vegetation (*Coastal shrub* or *High vegetation*) in the Random Forest classification correspond precisely with the color boundaries in the UMAP color composite, this is less accurate for the CNN classification.



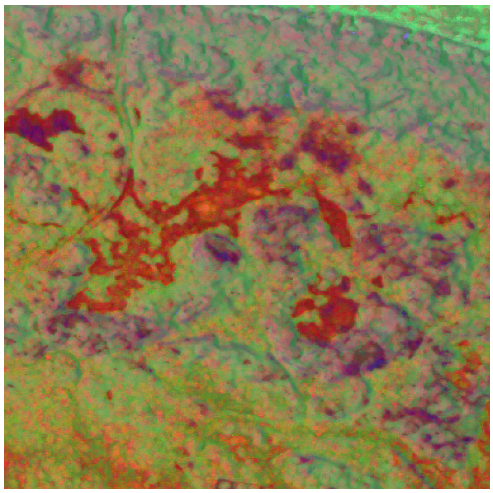
(a) Random Forest result.

(b) CNN result.

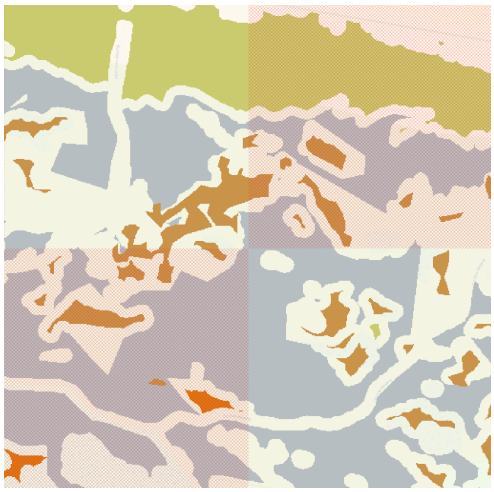


(c) Random Forest evaluation map.

(d) CNN evaluation map.

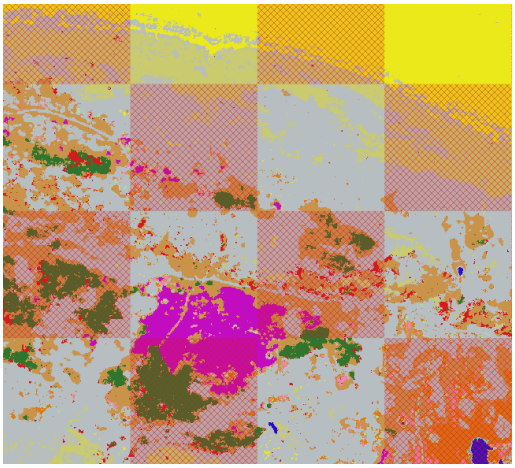


(e) UMAP color composite.

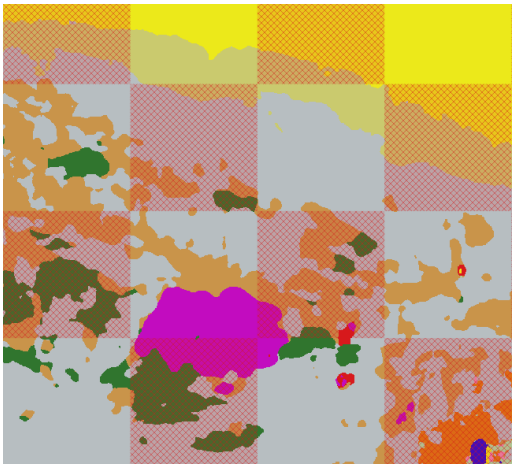


(f) Reference class map.

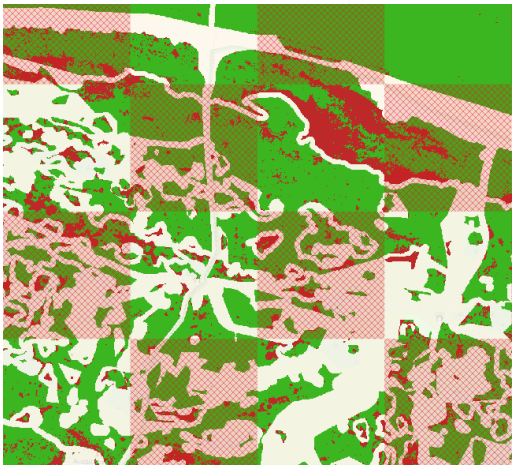
Figure 5.29.: Land cover classification maps of ROI I derived by (a) Random Forest or (b) CNN classification based on multi-frequency PolInSAR data. For reference, the UMAP color composite based on $EPI_{S,L}$ is shown in (c) and the reference map with marked training areas is given in (d).



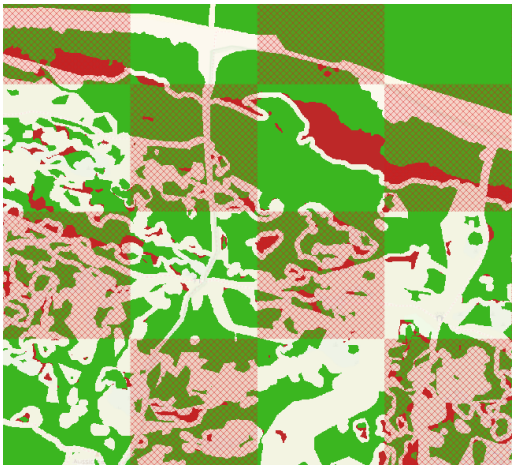
(a) Random Forest result.



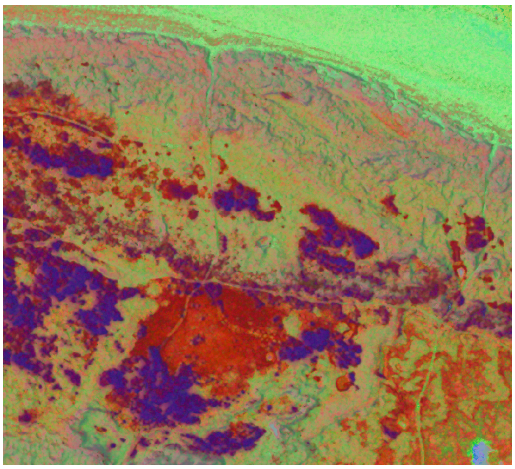
(b) CNN result.



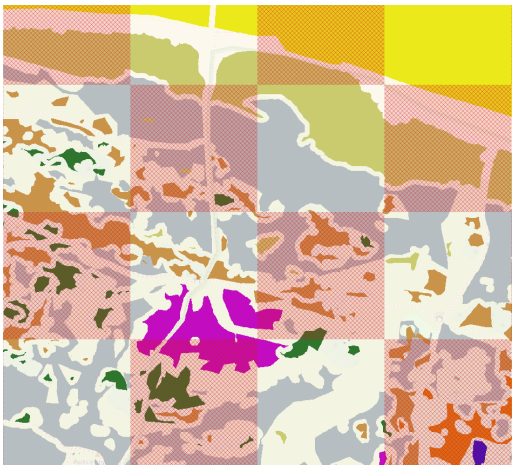
(c) Random Forest evaluation map.



(d) CNN result.



(e) UMAP color composite.



(f) Reference class map.

Figure 5.30.: Land cover classification maps of ROI II derived by (a) Random Forest or (b) CNN classification based on multi-frequency PolInSAR data. For reference, the UMAP color composite based on $EPI_{S,L}$ is shown in (c) and the reference map with marked training areas is given in (d).

5.6.4 Multi-frequency PolInSAR Classification with Cross-channel Feature Extraction

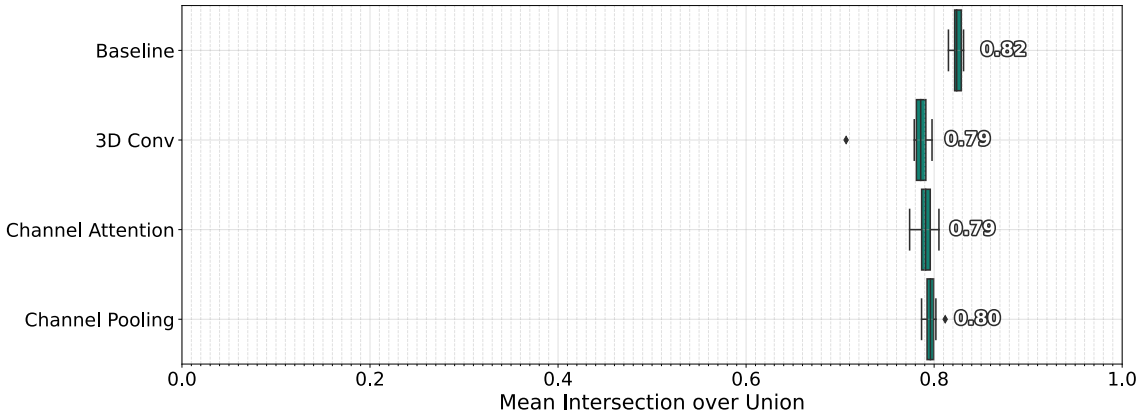
In the following, the impact of cross-channel feature extraction methods on the performance of multi-frequency PolInSAR data classification is analyzed. The evaluation focuses on the use of 3D convolutions, Cross-Channel Pooling, and the Channel Attention Module to enhance the ability of CNNs to model inter-channel relationships within the $\text{EPI}_{S,L}$ feature set.

The results illustrated in Figure 5.31 reflect the performance of the different approaches. Baseline results, representing CNN classification outcomes without any inter-channel relationship enhancements, are included for comparative purposes. While Cross-Channel Pooling and the Channel Attention Module are integrated into all three CNN models (Vanilla U-Net, ResNet U-Net and EfficientNet U-Net), 3D convolutions are exclusively evaluated with the Vanilla U-Net architecture.

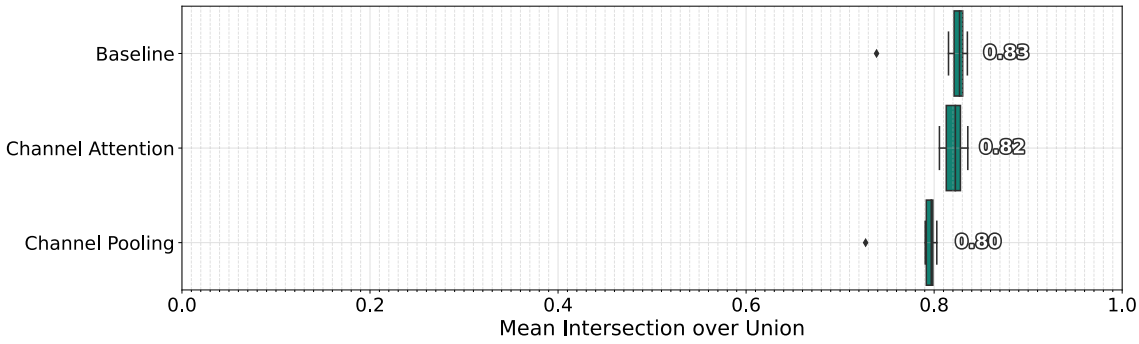
Comparative analysis reveals that none of the tested methods improve classification performance. The use of the Channel Attention Module within the ResNet U-Net, as well as the Channel Attention Module and Cross-Channel Pooling within the EfficientNet U-Net, yield results comparable to their respective baselines. The remaining configurations even demonstrate a performance decline of up to three percentage points. The poor performance when using 3D convolution can be explained by the lower number of filter kernels (compared to the baseline 2D-CNN model), as this reduces the diversity of spatial features. One explanation for the performance loss when using Cross-Channel Pooling is that potentially noisy feature images are combined at pixel level, resulting in noise in the feature maps. 2D convolutions that perform spatial filtering before channel-wise summation may better suppress this noise. The problem with using the Channel Attention Module could be the information aggregation applied within the module across the entire spatial dimension. For the used patch size of 128×128, the aggregated information is no longer representative, and the resulting recalibration vector is, therefore, not suitable.

5.6.5 Multi-frequency PolInSAR Classification with Dimension Reduction

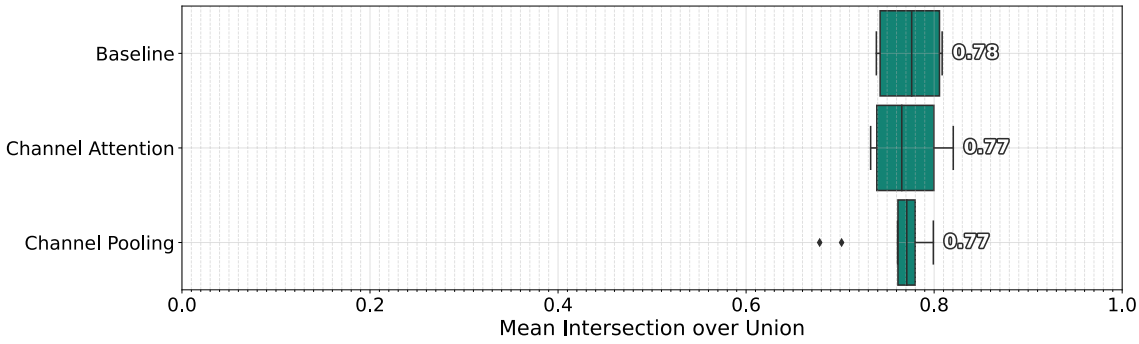
In the following, the impact of various dimension reduction techniques, based on automatic feature selection or feature transformation, on the classification results is analyzed. For this purpose, automatically generated feature sets comprising three, 15, and 40 features are evaluated. The top 40 ranked features selected using the filter-based methods MI, mRMR, and RF-FI are presented in Table 5.6. Features that are among the top 15 in all methods are highlighted in blue, while those ranking in the top 15 for two methods are shown in gray. Notably, the interferometric phases $\Psi_{opt,1}$ and $\Psi_{opt,2}$ from both frequency bands appear in all subsets, suggesting a high information content complementary to polarimetric features.



(a) Results using Vanilla U-Net.



(b) Results using ResNet U-Net.



(c) Results using EfficientNet U-Net.

Figure 5.31.: Classification results using 3D convolutions, Channel Attention Module or Cross-Channel Pooling within (a) Vanilla U-Net, (b) ResNet U-Net and (c) EfficientNet U-Net.

Table 5.6.: Top 40 ranked features by MI, mRMR and Random Forest (RF) feature selection

MI	mRMR	RF
λ_3 (S)	$\lvert\gamma_{opt,1}\rvert$ (S)	$\Psi_{opt,2}$ (L)
$P_{v,VZ}$ (S)	$\lvert\rho_{12}\rvert$ (L)	$\lvert\gamma_{opt,1}\rvert$ (L)
$\lvert\langle t_{33}\rangle\rvert$ (S)	$P_{s,Y}$ (S)	$\lvert\gamma_{opt,1}\rvert$ (S)
λ_2 (S)	$\Psi_{opt,2}$ (S)	$\Psi_{opt,1}$ (L)
$P_{v,Y}$ (S)	$\arg(\langle t_{12}\rangle)$ (L)	$\Psi_{opt,2}$ (S)
λ_2 (L)	$\lvert\rho_{13}\rvert$ (L)	$\Psi_{opt,1}$ (S)
λ_3 (L)	$\lvert\rho_{12}\rvert$ (S)	$\lvert\gamma_{opt,2}\rvert$ (S)
$\lvert\langle t_{22}\rangle\rvert$ (S)	A (L)	$\lvert\gamma_{opt,2}\rvert$ (L)
$P_{v,VZ}$ (L)	$\Psi_{opt,1}$ (L)	$\lvert\gamma_{opt,3}\rvert$ (L)
$\lvert\langle t_{33}\rangle\rvert$ (L)	$\lvert\langle t_{12}\rangle\rvert$ (L)	$\lvert\langle t_{22}\rangle\rvert$ (L)
$\Psi_{opt,1}$ (L)	$\arg(\langle t_{23}\rangle)$ (L)	$\Psi_{opt,3}$ (L)
$P_{d,VZ}$ (L)	$\Psi_{opt,1}$ (S)	λ_3 (S)
$\Psi_{opt,2}$ (S)	r_{hvvv} (L)	$\lvert\langle t_{22}\rangle\rvert$ (S)
$P_{d,VZ}$ (S)	$\lvert\rho_{13}\rvert$ (S)	$\Psi_{opt,3}$ (S)
$\Psi_{opt,1}$ (S)	$\arg(\langle t_{23}\rangle)$ (S)	$P_{v,VZ}$ (S)
$\Psi_{opt,2}$ (L)	$\arg(\langle t_{13}\rangle)$ (S)	$\lvert\langle t_{33}\rangle\rvert$ (S)
$\lvert\gamma_{opt,2}\rvert$ (S)	$\lvert\gamma_{opt,2}\rvert$ (L)	$\lvert\gamma_{opt,3}\rvert$ (S)
$\lvert\gamma_{opt,1}\rvert$ (S)	$\arg(\langle t_{13}\rangle)$ (L)	$P_{d,VZ}$ (S)
$\Psi_{opt,3}$ (L)	$\lvert\rho_{23}\rvert$ (L)	λ_2 (S)
$\lvert\gamma_{opt,1}\rvert$ (L)	$\lvert\rho_{23}\rvert$ (S)	λ_3 (L)
$\lvert\langle t_{22}\rangle\rvert$ (L)	p_2 (L)	λ_2 (L)
$\Psi_{opt,3}$ (S)	$\Psi_{opt,3}$ (L)	$\lvert\langle t_{33}\rangle\rvert$ (L)
$\lvert\gamma_{opt,3}\rvert$ (S)	$\arg(\langle t_{12}\rangle)$ (S)	$P_{v,Y}$ (S)
$\lvert\gamma_{opt,2}\rvert$ (L)	$P_{s,Y}$ (L)	$P_{v,VZ}$ (L)
$P_{v,Y}$ (L)	$\lvert\gamma_{opt,2}\rvert$ (S)	$r_{hh/vv}$ (S)
$\lvert\gamma_{opt,3}\rvert$ (L)	A (S)	H (S)
$r_{hh/vv}$ (S)	$\Psi_{opt,2}$ (L)	H (L)
$\bar{H}_d$ (S)	$r_{hh/hv}$ (S)	$P_{d,VZ}$ (L)
d_p (S)	$\Psi_{opt,3}$ (S)	$\lvert\rho_{12}\rvert$ (L)
d_{idx} (S)	$\lvert\langle t_{12}\rangle\rvert$ (S)	$\arg(\langle t_{12}\rangle)$ (L)
$\bar{H}_p$ (S)	r_{hvvv} (S)	$\bar{H}_d$ (S)
H (S)	$\lvert\gamma_{opt,1}\rvert$ (L)	$\arg(\langle t_{12}\rangle)$ (S)
p_1 (S)	$r_{hh/hv}$ (L)	$\lvert\langle t_{12}\rangle\rvert$ (L)
$\lvert\langle t_{23}\rangle\rvert$ (S)	p_2 (S)	α (L)
H (L)	$\lvert\gamma_{opt,3}\rvert$ (S)	λ_1 (S)
$\bar{H}_d$ (L)	$P_{s,VZ}$ (L)	$\lvert\langle t_{23}\rangle\rvert$ (L)
d_{idx} (L)	p_3 (L)	d_{idx} (S)
$\bar{H}_p$ (L)	$P_{d,Y}$ (S)	$r_{hh/hv}$ (L)
d_p (L)	$\lvert\gamma_{opt,3}\rvert$ (L)	$P_{v,Y}$ (L)
$r_{hh/vv}$ (L)	$r_{hh/vv}$ (L)	$\bar{H}_p$ (L)

Among the important polarimetric features, those ranked in the top 15 by MI and RF-FI include λ_3 (S), reflecting diffuse scattering, $|\langle t_{22} \rangle|$ (S), indicating depolarization or volume scattering contribution, and $P_{v,VZ}$ (S), representing volume scattering contribution.

Feature subsets selected by MI or RF-FI include redundant features. This is expected for MI, a univariate method evaluating each feature in isolation based on the information it provides about the target variable. The selected features predominantly relate to volume scattering, including volume scattering contributions based on target decomposition such as $P_{v,VZ}$, $P_{v,Y}$, $|\langle t_{33} \rangle|$, and the eigenvalues λ_2 and λ_3, which are relatively higher in volume scattering compared to λ_1. The selection of those features can be explained by the dataset composition, primarily consisting of natural land cover with varying vegetation densities, reflected in features related to volume scattering. Although RF-FI takes the interactions between the features into account in the selection, a high redundancy is also observed among the top-rated features, which are mostly interferometric ones. As expected, the most diverse feature subsets are selected using the mRMR method, which explicitly minimizes redundancy.

The features selected by the embedded RFW method are listed in Table 5.7 for each CNN architecture. Notably, the feature selection demonstrates a high degree of consistency across all architectures. Features identified as important for classification, similar to those selected by filter-based methods, include interferometric phases and the smallest eigenvalue λ_3. However, unlike filter-based selection, the inclusion of polarimetric phases is observed. This can be attributed to the fact that phase information is randomly distributed over certain land cover classes and homogeneous over others. This variability is significant for a classifier that considers the pixel neighborhood, as is done in CNNs, but not in pixel-based feature selection methods.

The results of CNN classification using the selected and transformed feature subsets as input are presented in Figure 5.32 for the Vanilla U-Net, in Figure 5.33 for the ResNet U-Net, and in Figure 5.34 for the EfficientNet U-Net. Results for feature subsets generated by automatic feature selection are marked in blue, and feature subsets generated by feature transformation are marked in red. Additionally, the results of CNN classification using the entire 84-dimensional $\text{EPI}_{S,L}$ are shown in green.

Vanilla U-Net Results. When using only three features, the features selected by MI and RF-FI show poor performance, with mIoU values of 0.57 and 0.6, respectively. These results are even inferior to the use of the three Pauli decomposition components of single-frequency (S-band) data (mIoU of 0.7, see Figure 5.16a). This poor performance can be attributed to high redundancy among the selected features, which capture only a fraction of all relevant information of the data. In contrast, the mRMR method achieves a median mIoU score of 0.74, and the RFW method achieves a median mIoU of 0.77, suggesting that these feature subsets are more effective in capturing relevant information. Nevertheless,

Table 5.7.: Top 40 ranked features by RFW within Vanilla, ResNet or EfficientNet U-Net

Vanilla	ResNet	EfficientNet
λ_3 (S)	$\Psi_{opt,1}$ (L)	λ_3 (S)
α (S)	$\Psi_{opt,1}$ (S)	$\Psi_{opt,3}$ (L)
r_{hvvv} (L)	α (S)	α (S)
$\lvert\langle t_{13}\rangle\rvert$ (L)	$\lvert\langle t_{12}\rangle\rvert$ (S)	r_{hvvv} (L)
$\Psi_{opt,3}$ (L)	$P_{v,Y}$ (L)	$\lvert\langle t_{13}\rangle\rvert$ (L)
α (L)	α (L)	α (L)
$\lvert\langle t_{12}\rangle\rvert$ (S)	$\lvert\langle t_{23}\rangle\rvert$ (S)	$P_{d,Y}$ (S)
$\arg(\langle t_{13}\rangle)$ (L)	r_{hvvv} (L)	$\Psi_{opt,1}$ (S)
$\arg(\langle t_{13}\rangle)$ (S)	$\arg(\langle t_{12}\rangle)$ (L)	$\arg(\langle t_{12}\rangle)$ (S)
$\Psi_{opt,1}$ (S)	$\arg(\langle t_{12}\rangle)$ (S)	$\arg(\langle t_{23}\rangle)$ (S)
$\arg(\langle t_{23}\rangle)$ (S)	$\lvert\langle t_{13}\rangle\rvert$ (L)	$\arg(\langle t_{12}\rangle)$ (L)
$\Psi_{opt,1}$ (L)	$\arg(\langle t_{23}\rangle)$ (S)	$\arg(\langle t_{13}\rangle)$ (L)
$\arg(\langle t_{12}\rangle)$ (S)	$\arg(\langle t_{13}\rangle)$ (L)	$\lvert\langle t_{12}\rangle\rvert$ (S)
$\arg(\langle t_{12}\rangle)$ (L)	$\lvert\langle t_{13}\rangle\rvert$ (S)	$\Psi_{opt,1}$ (L)
$\Psi_{opt,3}$ (S)	$P_{v,Y}$ (S)	$\lvert\langle t_{12}\rangle\rvert$ (L)
$\arg(\langle t_{23}\rangle)$ (L)	λ_3 (S)	$\arg(\langle t_{13}\rangle)$ (S)
$P_{d,Y}$ (L)	$P_{d,VZ}$ (L)	$\arg(\langle t_{23}\rangle)$ (L)
$\lvert\langle t_{12}\rangle\rvert$ (L)	$P_{s,Y}$ (L)	$\Psi_{opt,3}$ (S)
$P_{d,Y}$ (S)	$P_{v,VZ}$ (L)	$P_{d,Y}$ (L)
$\Psi_{opt,2}$ (S)	$P_{d,VZ}$ (S)	$\lvert\langle t_{23}\rangle\rvert$ (L)
$\lvert\langle t_{23}\rangle\rvert$ (L)	λ_1 (L)	λ_3 (L)
$\lvert\langle t_{22}\rangle\rvert$ (L)	$\lvert\langle t_{33}\rangle\rvert$ (S)	$P_{d,VZ}$ (S)
$\bar{\lambda}$ (L)	$P_{d,Y}$ (S)	λ_1 (S)
λ_1 (L)	$P_{s,Y}$ (S)	$\lvert\langle t_{22}\rangle\rvert$ (S)
$r_{hh/hv}$ (L)	$\lvert\langle t_{33}\rangle\rvert$ (L)	$P_{d,VZ}$ (L)
r_{hvvv} (S)	$\lvert\langle t_{22}\rangle\rvert$ (S)	$P_{v,VZ}$ (S)
$P_{s,Y}$ (L)	$\bar{\lambda}$ (L)	$\lvert\langle t_{13}\rangle\rvert$ (S)
p_2 (L)	λ_1 (L)	$P_{v,VZ}$ (L)
$\lvert\rho_{12}\rvert$ (L)	r_{hvvv} (S)	$\bar{\lambda}$ (L)
$\lvert\gamma_{opt,1}\rvert$ (S)	A (L)	$P_{s,VZ}$ (S)
p_1 (L)	$\lvert\rho_{23}\rvert$ (S)	$\lvert\langle t_{23}\rangle\rvert$ (S)
$\lvert\gamma_{opt,1}\rvert$ (L)	d_p (S)	λ_1 (S)
$\lvert\gamma_{opt,3}\rvert$ (L)	$\lvert\gamma_{opt,1}\rvert$ (S)	$\lvert\langle t_{33}\rangle\rvert$ (S)
$\lvert\rho_{13}\rvert$ (L)	conformity (S)	$\bar{\lambda}$ (S)
p_3 (S)	p_1 (S)	$r_{hh/vv}$ (L)
p_3 (L)	p_1 (L)	$P_{s,Y}$ (L)
conformity (S)	p_3 (S)	$\lvert\langle t_{11}\rangle\rvert$ (L)
$\lvert\rho_{23}\rvert$ (L)	$\lvert\rho_{13}\rvert$ (S)	$\lvert\langle t_{11}\rangle\rvert$ (S)
A (S)	$\lvert\rho_{13}\rvert$ (L)	λ_1 (L)
d_p (S)	$r_{hh/vv}$ (L)	p_2 (L)

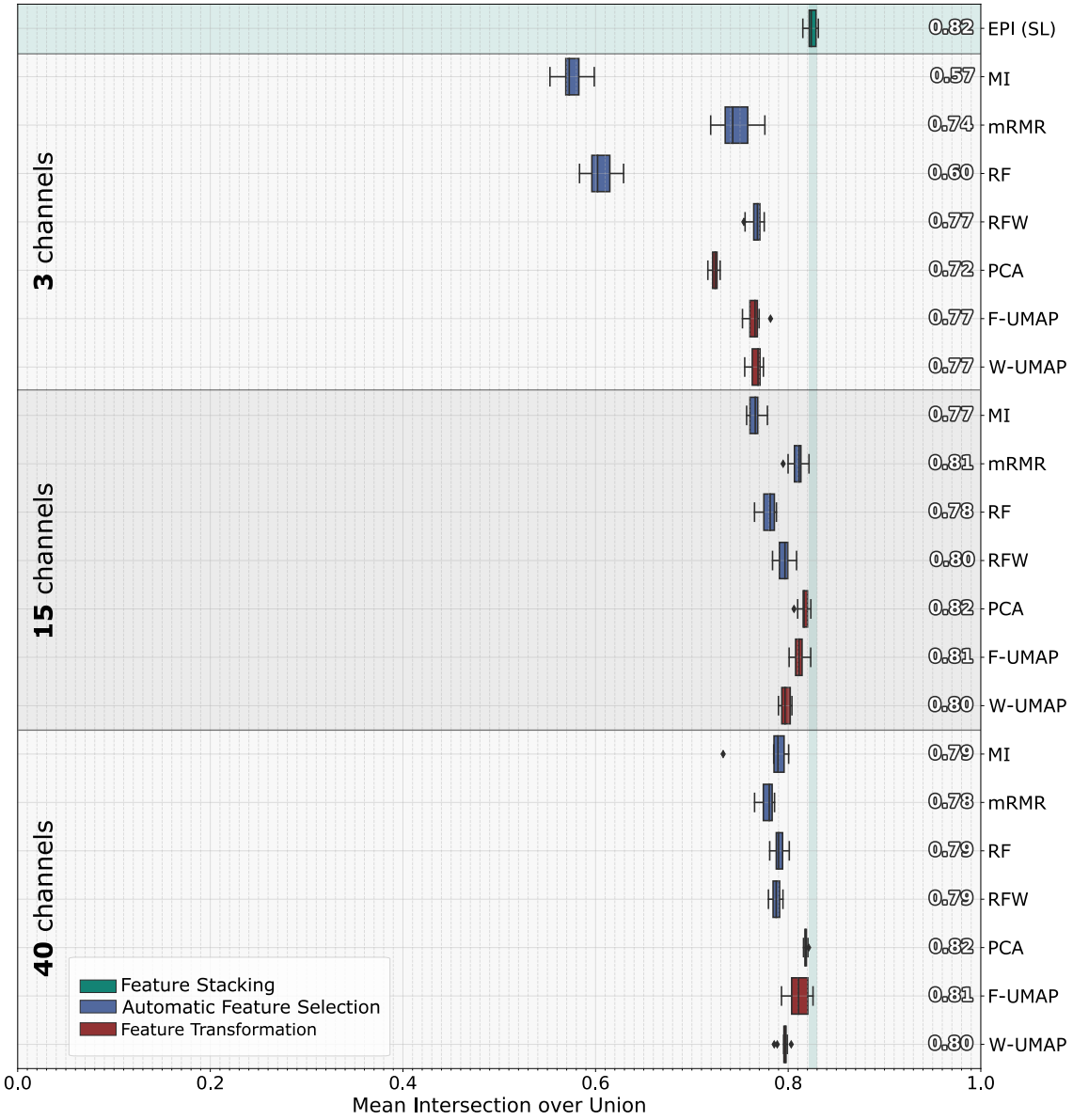


Figure 5.32.: Comparison of Vanilla U-Net classification results using feature stacking, automatic feature selection or feature transformation.

there remains a discrepancy of 0.08 and 0.05 in mIoU, respectively, when compared to the full feature set. When employing feature transformation to generate a three-channel image, both UMAP methods outperform PCA (+0.05 mIoU) and are comparable to the best feature selection method (RFW).

Increasing the feature dimension of the CNN input to 15 channels leads to substantial improvements across all methods. The most effective feature selection method is mRMR, which achieves a median mIoU of 0.81, closely approaching the performance of the full feature set (median mIoU of 0.82). For feature transformation, PCA slightly outperforms both UMAP approaches, with a median mIoU of 0.82, thereby achieving classification results on par with the full feature set.

Further increasing the feature dimension of the CNN input to 40 channels leads, in the best case, to a minimal improvement, but for mRMR, even a performance decrease is observed. This decrease may be attributed to adding features that are non-redundant to previous features but not beneficial for CNN classification.

ResNet U-Net Results. When using three channels, the results show a trend similar to the Vanilla U-Net: MI and RF-FI show poor performance, while RFW emerges as the best feature selection method, performing on par with mRMR. UMAP consistently outperforms PCA as well as all feature selection methods. Increasing the number of features to 15 leads to improvements across all methods, with mRMR and PCA achieving the best results that are comparable to the full feature set. When the feature dimension is increased to 40, a different pattern emerges compared to the Vanilla U-Net test setup: the mRMR feature subset shows further improvement, and the RF-FI feature set exhibits significant performance gains, achieving accuracies comparable to the full feature set.

Efficient U-Net Results. Similar to the previous results, MI and RF-FI perform poorly on three channels. mRMR performs well but is outperformed by the feature transformation methods. Again, the results of the UMAP methods outperform the results of PCA. Different from Vanilla and ResNet U-Net results, the three-channel RFW feature set performs worse than mRMR. A further distinction is that the utilization of three-channel inputs generated by UMAP is sufficient to attain classification outcomes comparable to those obtained with the complete feature set. This indicates that EfficientNet, unlike Vanilla U-Net and ResNet U-Net, may handle high-dimensional input data less effectively, making dimension reduction beneficial. When using 15 and 40 features, all dimension reduction methods find feature subsets that are equally or even better suited than the entire feature set. Feature transformation using PCA yields the best results.

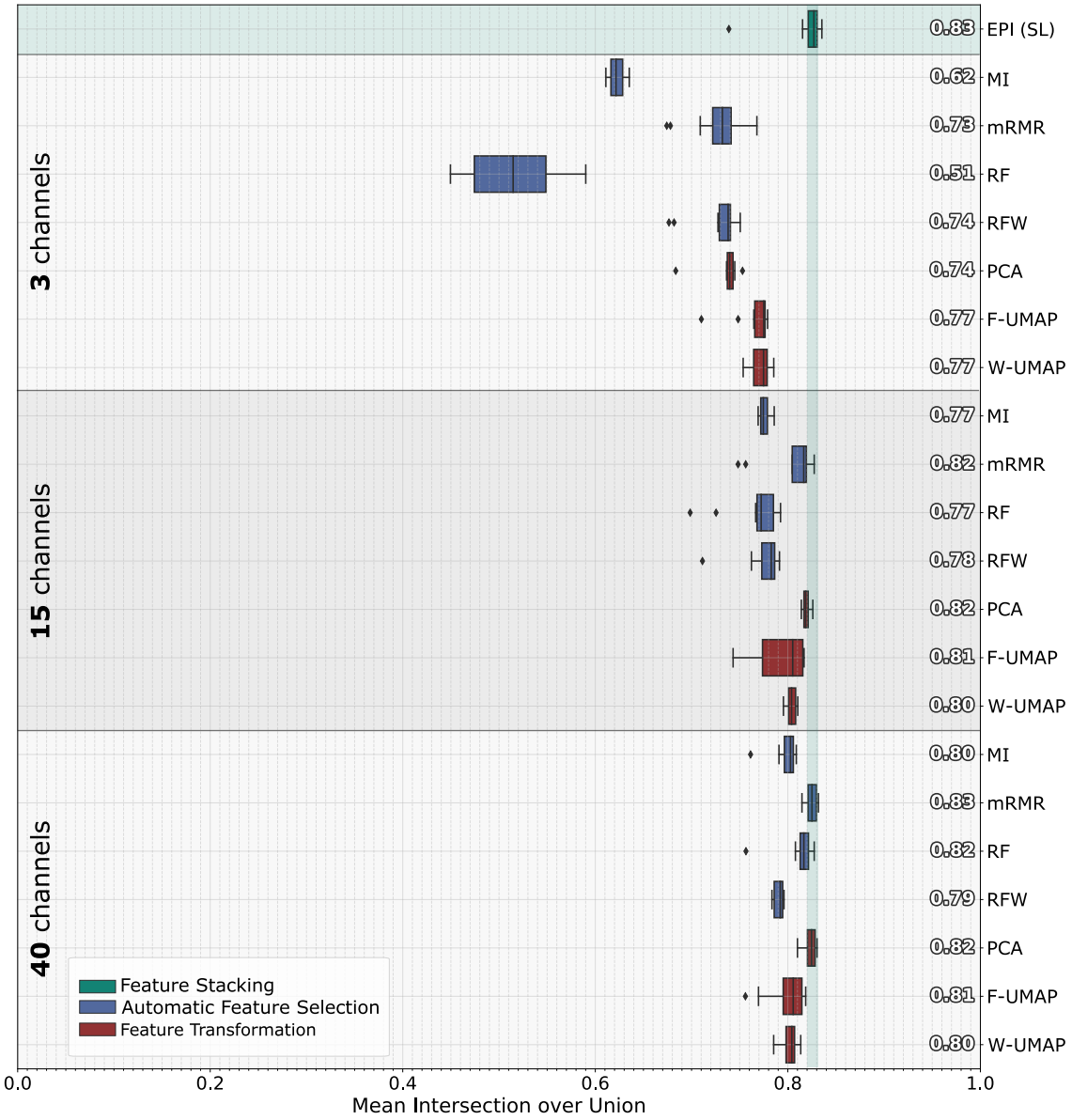


Figure 5.33.: Comparison of ResNet U-Net classification results using feature stacking, automatic feature selection or feature transformation.

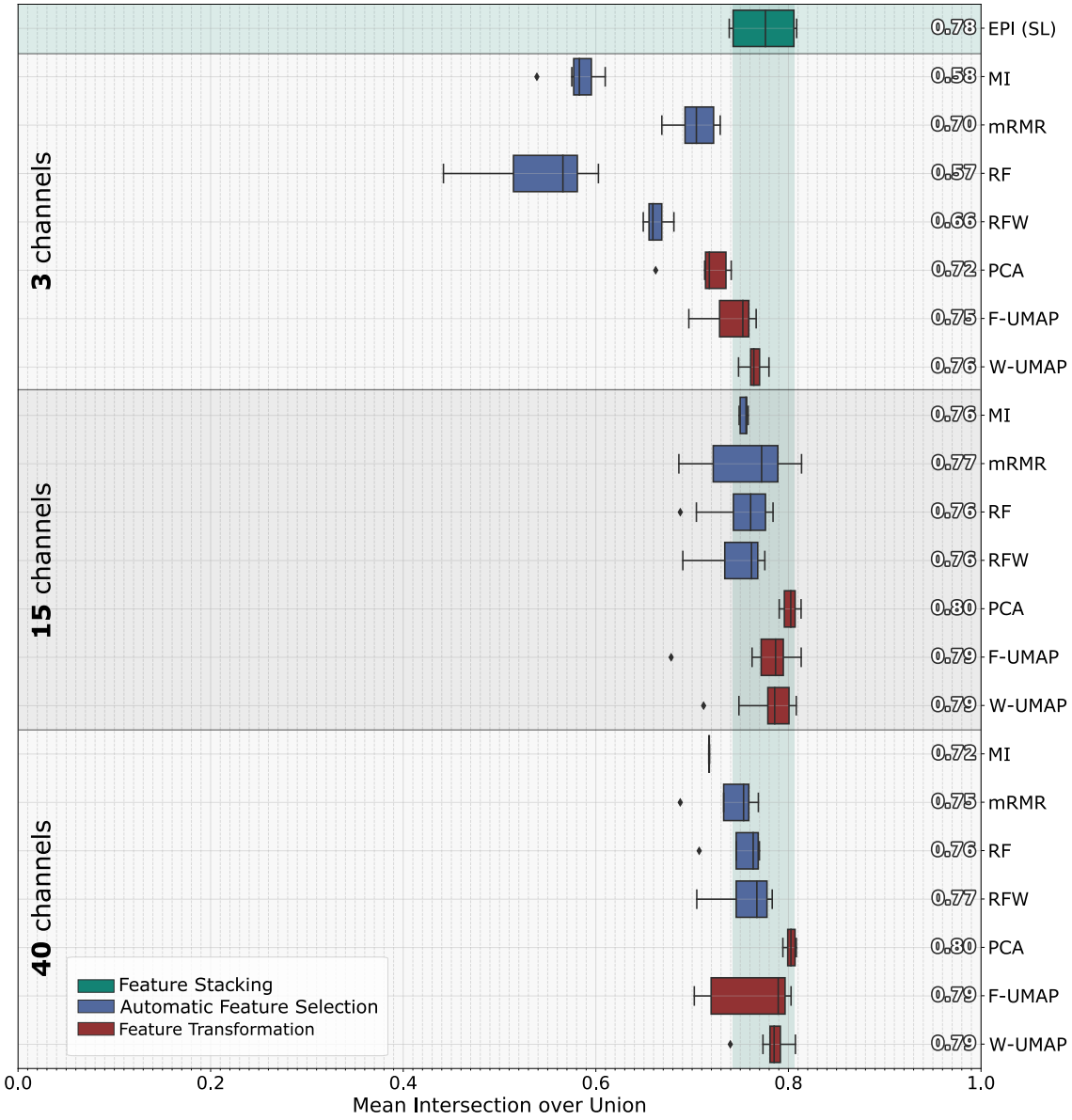


Figure 5.34.: Comparison of EfficientNet U-Net classification results using feature stacking, automatic feature selection or feature transformation.

Summary of Results. If the number of input channels of the CNN is limited to three, UMAP proves to be the best method for generating the reduced input data representation. However, the high level of compression of the original 84-dimensional feature set leads to a significant loss of information, whereby there remains a gap in CNN classification performance compared to using the full feature set. In contrast, using a 15-channel image, the feature selection method mRMR and the feature transformation method PCA turn out to be the means of choice, which generate feature subsets that are sufficient to achieve classification accuracy comparable to the use of the full feature set. Increasing the feature dimension to 40 cannot further improve these results. Overall, this shows that the dimension of the 84-dimensional feature set can be reduced to 15 without or with only minimal loss of classification accuracy of the CNNs. However, an improvement that could have been expected due to a reduction in the redundancy of the input data does not occur.

Runtime Analysis. An advantage of feature selection is that it reduces the number of features that need to be extracted, stored, and processed. Another significant advantage, which applies not only to feature selection but also to feature transformation, is the reduction in training time for the subsequent CNN model.

The reduction in training time when using input images with fewer channels is primarily due to the reduced number of required filter kernels and, thus, the reduced number of convolution operations in the first convolution layer. Reducing the filter kernels at this point of the network has a strong influence on the computing time since the spatial resolution of the input to which the convolutions are applied is maximal.

Run time tests are conducted using a single NVIDIA Tesla V100 (32 GB VRAM, 1230 MHz, 5120 CUDA-kernels) and four Intel(R) Xeon(R) CPUs (each with 12 kernels and 24 threads). Table 5.8 shows the average training times per epoch, measured over 30 epochs, and the number of input channels for each network architecture. While the difference in training time between using three and 15 channels is relatively small, the difference becomes substantial when comparing 15 channels to the full feature set of 84 channels, with a factor of about two. When using the suggested 300 training epochs, this leads to a reduction in the total training time from around four hours to two hours.

Considering previous findings, this implies that employing appropriate dimension reduction methods, such as mRMR or PCA, can halve the training time with negligible loss in classification performance.

Table 5.8.: Number of trainable parameters and average training times per epoch depending on the number of channels of the input image and the network architecture used.

Channel		Vanilla	ResNet	EfficientNet
3	#params	1,862,849	2,853,409	333,173
	train time [s]	25.67 ($\pm$0.45)	22.88 ($\pm$0.28)	23.89 ($\pm$0.37)
15	#params	1,869,761	2,855,137	334,9017
	train time [s]	28.00 ($\pm$0.71)	24.63 ($\pm$0.59)	26.12 ($\pm$0.54)
40	#params	1,884,161	2,858,737	338,501
	train time [s]	38.55 ($\pm$0.82)	34.23 ($\pm$0.63)	36.01 ($\pm$0.96)
84	#params	1,909,505	2,865,073	344,837
	train time [s]	55.66 ($\pm$1.25)	51.85 ($\pm$0.79)	52.99 ($\pm$0.98)

5.7 Discussion

The presented analyses aim to assess the potential of image-based CNN for the land cover classification of multi-frequency PolInSAR data. While CNN classification for PolSAR images has been explored in the literature, showing improvement over statistical methods and traditional machine learning models, the inclusion of interferometric and multi-frequency information in CNN classification remains largely unexplored.

A central question for PolSAR and PolInSAR data classification, which has not been thoroughly investigated in either domain, is how to represent multi-dimensional complex-valued data to fully exploit its informational potential within CNN classification. To address this issue, the following questions were analyzed in this work:

1. **Data representation**: Can CNNs directly capture the relationship between PolSAR or PolInSAR data represented by the coherency matrix and the target label, or is prior extraction of expert-designed, physically interpretable features advantageous?
2. **High-dimensional data handling**: Are CNNs capable of effectively processing multi-channel images that contain redundant information, or are there:
   - specific CNN architectures tailored for cross-channel feature extraction, or
   - dimension reduction methods

   that can enhance classification performance and efficiency?
3. **Advantage of deep learning models**: Do CNNs provide superior classification performance compared to statistical methods and traditional machine learning models?

5.7.1 Data Representation

The following discusses the results concerning the selection of a suitable data representation for single-frequency PolSAR and multi-frequency PolInSAR data.

Representing Single-frequency PolSAR Data. To address the first question, various input representations were tested to represent single-frequency PolSAR data as input for different U-Net models. The results highlight the importance of choosing a suitable representation, evidenced by significant differences in mIoU values between the most and least suitable data representations.

The representation that proves to be the least suitable across all experiments is $T_{re,im}$. This can be attributed to the lack of significance of the real and imaginary parts of a complex value when considered separately and the loss of relevant information due to separate processing (i.e., convolution with different kernel weights). The result is nevertheless surprising, given that $T_{re,im}$ is one of the most widely used representations of PolSAR data in the context of CNN classification (see Table 5.1).

The widely used six-dimensional representation proposed by Zhou in [20] also fails to achieve optimal results. This is possibly due to the lack of polarimetric phase information. This aligns with findings by Chen et al. [21], and Zhang et al. [280], which show that alternative representations (e.g., using polarimetric features) outperform the Zhou representation as input for patch-based CNNs.

Optimal results are achieved by employing the EP feature set. This suggests that CNNs can extract relevant information from large feature sets, which not only comprise complementary but also redundant features. This capability leads to superior classification performance compared to using representations with fewer polarimetric features such as H-A-α, Yamaguchi, and VanZyl. Nonetheless, similar classification results can also be achieved using the simple and more generic representation $T_{amp,pha}$. Thus, the demonstrated ability of CNNs to extract all relevant information directly from the elements of the coherency matrix makes the inclusion of additional polarimetric features obsolete. Due to its proximity to the original data and the elimination of time-consuming, task-specific feature selection and feature calculation, $T_{amp,pha}$ is considered the most suitable representation for CNN inputs. Taking into account that this representation is rarely used in CNN-based classification, this fundamental insight offers the potential to significantly improve existing approaches by replacing $T_{re,im}$ with $T_{amp,pha}$. Furthermore, this finding provides a robust foundation for developing new approaches.

Multi-frequency PolInSAR Classification. For CNN-based multi-frequency PolInSAR classification, the direct utilization of the PolInSAR coherency matrix, which encompasses polarimetric and interferometric information represented by $T_{6,amp,pha}$, does not prove to

be the optimal approach. Using this representation, the classification cannot successfully exploit the additional interferometric information. The suitability of $T_{amp,pha}$ for PolSAR, but not for PolInSAR data, can be explained as follows. In PolSAR data, important observables that play a crucial role in the separation of land cover are the contribution of different scattering mechanisms such as surface, double-bounce, and volume scattering. These are reflected directly, albeit not optimally, in the amplitudes of the diagonal coherency matrix elements, which are captured in $T_{amp,pha}$. In contrast, for SAR interferometry, an important observable that enhances class separability is interferometric coherence, as demonstrated, for example, in [6]. This parameter is not directly represented by an element of the PolInSAR coherency matrix but is calculated using correlation coefficients between the elements. Since this operation cannot be modeled by a sequence of convolutions, the CNN is unable to learn and utilize interferometric coherence effectively. A similar challenge arises with the interferometric phases. Although these are directly contained in $T_{6,amp,pha}$ (phases of main diagonal elements of $\langle \mathbf{\Omega_{12}} \rangle$), they are uniquely defined only within the interval from $-\pi$ to $+\pi$. This results in artificial discontinuities in the image at the transition from $-\pi$ to $+\pi$, complicating feature learning. To extract relevant information, such as topographic features, phase unwrapping is required. CNNs that are optimized in terms of classification performance are not capable of learning phase unwrapping inherently. Thus, in contrast to single-frequency PolSAR classification, the integration of pre-extracted features significantly enhances classification performance compared to employing only $T_{6,amp,pha}$. By granting the CNN direct access to optimal interferometric coherences and unwrapped interferometric phases contained in $EPI_{S,L}$, the additional information provided by interferometric measurements can be fully utilized. This also leads to an improvement in CNN classification performance over using solely polarimetric information. The strong classification results highlight the ability of CNNs to effectively extract relevant information from the high-dimensional input data representation, enabling accurate classification.

Conclusion. Based on the discussed results, the question posed in the introduction of whether CNNs can directly capture the relationship between Pol(In)SAR data represented by the coherency matrix and the target classes or whether the prior extraction of physically interpretable, expert-designed features is beneficial, can be answered. The results show that a distinction must be made between the classification of PolSAR and PolInSAR data. For PolSAR data, CNNs can derive the relevant information directly from the coherency matrix if it is represented by $T_{amp,pha}$. Thus, in this case, optimal classification results can be achieved without the need for prior feature extraction. In contrast, for CNN-based classification of PolInSAR data, prior extraction of polarimetric and especially interferometric features is advantageous. Although the PolInSAR coherency matrix contains the entire measurement information, CNNs cannot fully utilize this information. In contrast, including interferometric coherences and unwrapped phases in the input data is an effective approach to utilize the additional information from interferometry which further improves the classification compared to the classification of PolSAR data. These findings offer

potential improvements to existing CNN-based PolSAR classification methods and form the basis for the further development of CNN-based classification of multi-frequency PolInSAR data.

5.7.2 Handling High-dimensional Data

While the results are promising and suggest that redundancies along the channel dimension do not hinder successful CNN classification, further investigations were conducted to determine if PolInSAR features could be more effectively utilized to enhance the classification performance. Additionally, it was examined whether all features of the high-dimensional input representation are necessary or if a reduced set of features could be sufficient, thereby decreasing computational and memory requirements.

Cross-channel Feature Extraction. As the results of pixel-based Random Forest classification using $\mathrm{EPI}_{S,L}$ feature set show, a significant part of the decisive information for class separation directly resides in the feature dimension. This work, therefore, investigated whether 3D convolutions that extract high-level features across spatial and channel dimensions, a Channel Attention Module that emphasizes relevant features, or Cross-Channel Pooling that performs non-linear feature transformation can improve the classification performance of CNNs. However, all approaches fail to increase the classification accuracy and, in some cases, even show inferior results compared to standard 2D-CNNs. Thus, it can be concluded that 2D convolutions that perform spatial filtering and channel summation within one step are sufficient for feature extraction and are preferable due to their simplicity.

These results are in contrast to the findings of the studies [223], in which the classification performance of CNNs shows improvements through 3D convolutions, and to [264], in which an increase in performance is achieved through attention mechanisms. One possible explanation for the different results with 3D convolution is that in the cited studies, the 3D-CNNs have a significantly higher number of trainable parameters than the 2D-CNNs used for comparison in their evaluations. This is not the case in the present work, as the number of spatial filters must be reduced in order to keep the memory requirements manageable. In addition, in [223] and [272] $\mathrm{T}_{\mathrm{re,im}}$ is used as input, where the relationship between the channels plays a greater role, as the prevention of decoupling of the real and imaginary parts is necessary.

The deviating results regarding the use of channel attention mechanisms can be attributed to the different sizes of the input images. Dong et al. [264] use a patch-based approach with small image patches of 15×15 pixels, while in this work, an image-based approach with larger image patches of 128×128 pixels is tested. Channel attention operations include the aggregation of a feature image over the spatial dimension into a single value.

When the spatial dimension is high, this aggregation may not adequately represent the feature information. In contrast, for lower spatial dimensions (as in the patch-based approach), the aggregation becomes more representative and facilitates the calculation and application of an appropriate local recalibration factor. Consequently, the channel attention mechanism used in [264] is not beneficial for image-based CNN classification.

In summary, the simplicity of 2D-CNNs, which learn and apply channel-specific convolution kernels and sums convolved feature images, proves to be the best approach for feature extraction from a multi-channel image. This result highlights the effectiveness of simpler models in dealing with high-dimensional input data and is an important finding for the development of future approaches.

One limitation of the presented analysis on cross-channel feature extraction concerns the fact that it was based exclusively on the approach of early fusion of polarimetric and interferometric features (i.e. the stacking of features within the input layer). Thus, it remains to be tested whether alternative fusion methods might be able to better exploit the information contained in the features to further improve the classification. Possible approaches that can be investigated in future work are mid-level fusion, also known as deep fusion and late fusion. In these approaches, instead of direct fusion in the input layer, the input features are initially processed separately and combined in deeper layers or at the end of the network. Which strategy is best suited for the fusion of multi-frequency polarimetric and interferometric features for classification remains to be analyzed.

Dimension Reduction. This thesis investigated whether dimension reduction using automatic feature selection methods (MI, mRMR, RF-FI, RFW) or feature transformation techniques (PCA, F-UMAP, W-UMAP) prior to CNN classification can increase efficiency and classification performance. It has been shown that these preprocessing methods cannot improve classification performance but can significantly reduce memory requirements and training time while maintaining classification performance.

Among the evaluated feature selection methods, mRMR proves to be the most effective. This superiority is attributed to the ability of mRMR, in contrast to MI and RF-FI, to avoid redundancy within selected feature subsets. The embedded RFW method, which is intended to generate CNN-specific feature subsets, leads to worse results than the use of mRMR. The poorer performance can be attributed to the simultaneous optimization of the weights for weighting the input features and learning the convolutional filter kernels during CNN training. After reducing the learning rate, selected features no longer change, resulting in the CNN only adapting the spatial filters to this potentially suboptimal feature selection from this point onwards. Improvement approaches could be to use different learning rates for the feature weighting layer and the convolutional filters or to integrate an additional classifier to evaluate the feature selection performance during training. To summarize, mRMR is recommended due to its effectiveness in identifying a medium-sized

feature set, which, when used as CNN classification input, provides comparable results to using the full 84-dimensional $EPI_{S,L}$ feature set.

Among feature transformation techniques, PCA proves to be the best method for preprocessing prior to the CNN classification and outperforms the best feature selection method, mRMR, in terms of classification performance. While UMAP generates the most suitable 3-dimensional feature sets - underlining its exceptional utility for data visualization - it fails to embed data into higher dimensions in a suitable way. PCA is, therefore, not only simpler (in terms of computational complexity) but also better suited to embedding data in multi-dimensional spaces.

One advantage of using a low-dimensional input representation is that human analysts can more easily understand the information that is fed into a CNN since inspecting a few features through visual and quantitative analysis is easier than inspecting the entire feature set. This is an important aspect of model development, as the quality and type of input data directly influence model performance and prediction accuracy. Artifacts or errors in the input data can not only lead to misinterpretations but also may cause the model to learn irrelevant or incorrect patterns and thus become unusable. A profound understanding of the input data helps to recognize and correct distortions and anomalies at an early stage. A further advantage of dimension reduction before CNN classification is the significant reduction in training time. Since the optimal selection of the CNN architecture and the optimization of the hyperparameters often require frequent retraining, the efficiency of the training plays a crucial role in development and application. However, it is important to note that the reduction in training time observed here is specific to the comparatively small models used in this work. When using deeper models or models with a high number of convolutional filter kernels, the impact of the input dimension on the training time is likely to be smaller.

Although dimension reduction improves efficiency, it does not improve classification accuracy. This is in contrast to other work that observes an improvement in classification accuracy through dimension reduction in conjunction with SVM classifiers [196, 199, 201, 247]. A possible explanation for this discrepancy is that the SVM class assignment is derived directly from the polarimetric and interferometric features without first extracting high-level features. Therefore, redundant or irrelevant features may affect the performance. A problem that can be solved by dimension reduction, which improves accuracy. In contrast, CNNs learn to extract suitable features from raw data before making a class assignment. During training, a CNN can learn to ignore irrelevant input features, making it less prone to redundancy. Thus, dimension reduction cannot further increase the classification performance.

A limitation of the research design used in this work to investigate dimension reduction methods is that only three feature set sizes were tested (three, 15, and 40 dimensions). This limited range may not include the feature set size that achieves an optimal balance

between training time and classification accuracy. The exhaustive search for the optimal feature set size was not considered in this work, as this would have required training a CNN with all sizes for each method and, thus, an immense, unfeasible computational effort.

Conclusion. This work investigated whether CNNs can effectively process multi-channel images containing redundant information or whether certain CNN architectures or dimension reduction methods improve performance and efficiency.

The presented analysis shows that standard 2D-CNNs are able to extract relevant information from multi-channel images and use it for classification. More sophisticated architectures, including 3D convolutions, a Channel Attention Module, or Cross-channel Pooling, do not improve performance and sometimes even decrease it. This result suggests that 2D convolutions are both sufficient and preferable for feature extraction from multi-channel images.

Regarding dimension reduction, although it cannot improve the classification accuracy of CNNs, it can improve interpretability and speed up the training by using low-dimensional input data representation. In particular, feature selection by mRMR and feature transformation by PCA effectively reduces the dimension of the 84-dimensional feature representation to 15 dimensions with minimal loss of classification accuracy, thus optimizing computational and memory requirements.

To summarize, if the primary goal is to achieve the best classification performance without regard to time and storage efficiency, representing PolInSAR data by stacking a wide range of polarimetric and interferometric features and using them as input to a 2D-CNN is a viable solution. However, the best trade-off between classification performance and efficiency regarding time and storage space is achieved by dimension reduction with mRMR or PCA.

5.7.3 Comparison of Random Forest and CNN Classification

CNNs have been established as powerful tools for classifying and segmenting optical image data due to their ability to recognize complex spatial patterns. However, their application for land cover classification based on PolInSAR image data is still little explored. To evaluate the potential of CNNs in this area, this work compared its performance with that of supervised Wishart and Random Forest classifications. Since the Random Forest classification surpasses the Wishart classification significantly, a detailed comparison was only performed between CNN and Random Forest classification.

Using single-frequency PolSAR data, Random Forest only achieved unsatisfactory results, while CNNs significantly improved classification outcomes. This improvement is reflected

in higher IoU values across all classes and is visually evident in the predicted land cover maps. These results align with previous work on
acpolsar land cover classification [20, 21, 23] that demonstrate the superior performance of CNNs compared to traditional machine learning classifiers. Notably, the performance gap between Random Forest and CNN classification demonstrated in this work is even more pronounced compared to other studies. This discrepancy can be attributed to the complexity of the classification task posed by the Pol-InSAR-Island dataset, which is more challenging than commonly used datasets such as Flevoland and PolSF.

Incorporating interferometric and multi-frequency features into both classifiers enhances performance, with CNN maintaining its superiority. However, the Random Forest classifier benefits more from this additional data, narrowing the performance gap. One possible explanation is that in the CNN-based classification of single-frequency PolSAR data, the use of spatial features is already sufficient for the detection of certain classes (e.g. *White dune* and *Peat bog*) and the additional information provided by interferometric and multi-frequency features are not needed. In contrast, for Random Forest classification, these additional features are necessary for accurate class differentiation, leading to substantial improvement. However, several misclassifications remain, resulting in a higher error rate compared to CNN. The quantitative and qualitative evaluation revealed that CNNs are particularly advantageous for classification tasks involving the recognition of complex classes that are composed of different land covers (e.g. *Settlement*) or have similar backscatter properties (e.g, *White dune* and *Grey dune*) so that they can only be correctly assigned by a suitable integration of spatial context. The ability of CNNs to autonomously learn appropriate spatial features is the key success factor in this regard.

Furthermore, CNNs can learn that adjacent pixels often belong to the same class, producing smoother maps without isolated pixel predictions that are common in pixel-based Random Forest classification. While this characteristic generally reduces misclassifications, it also has the drawback of producing overly smoothed results, causing a loss of fine details and an inability to precisely map class borders. The underlying causes are twofold: first, the spatial dimension is reduced within the encoder during the extraction of high-level features, leading to a loss of positional information. Even the employed skip connections are insufficient for accurate relocalization of this information. Second, the sparse reference map used for training does not adequately represent boundary areas between classes, leading to insufficient learning in these regions.

Visual analysis shows that Random Forest classification is advantageous over CNN in accurately mapping class borders and preserving fine structures. This superiority is due to the pixel-based classification approach of Random Forests, which do not inherently filter the data and thus maintain the granularity needed for detailed boundary mapping. This notable advantage is not fully reflected in the quantitative results, as class boundaries are often masked in the reference data used for evaluation, preventing an accurate assessment at these crucial locations.

Recognizing the strengths of both methodologies, an enhancement to CNN classification could involve integrating Random Forest predictions to combine the lower error rate and context awareness of CNNs with the detail preservation capabilities of a pixel-based classification. Alternatively, a purely CNN-based improvement could consist of incorporating boundary refinement mechanisms, as suggested in recent studies [281, 282].

Conclusion. In summary, the comparison of CNNs with statistical methods and traditional machine learning models shows that CNNs achieves better classification performance in terms of quantitative and qualitative criteria. The superiority of CNNs is primarily due to their ability to combine polarimetric and interferometric features with automatically learned spatial features and thus exploit the image context for successful classification. In particular, this enables the identification of complex and heterogeneous land cover types.

Despite the advantages of CNNs, traditional machine learning methods retain their relevance and usefulness, especially in applications where class discrimination unambiguously results from the combination of polarimetric and interferometric features, and accurate boundary detection is more important than the correct classification of each pixel. An example of such a task is the detection of the water-land boundary.

However, many real-world remote sensing applications require not only the differentiation of land cover but also the differentiation of land use or the detection and analysis of complex natural structures. Examples of land use classification include the identification of coastal protection structures, residential areas, or complex infrastructure such as airports or bridges. Examples of the classification of natural structures include the recognition of river deltas, mountain ranges, complex wetlands, various forest areas, and glaciers. The inclusion of image context is essential for these tasks, and CNNs provides the appropriate method for such a comprehensive and context-aware analysis. There is considerable untapped potential, especially in combination with the rich information provided by PolInSAR data. This work represents the first fundamental step towards exploiting this potential and sets the stage for further research and development.

6 Conclusion and Outlook

6.1 Conclusion

This thesis focused on the analysis of multi-frequency PolInSAR data. The central questions addressed were how the multi-layered information content of this data can be represented to firstly realize the visualization of the data to support human-driven data analysis and secondly to exploit its full potential in the context of automatic CNN-based classification.

Human-driven Visual Data Analysis. To address the limitation of traditional visualization methods for PolInSAR data, which typically rely on feature selection leading to information loss, this thesis proposed and evaluated a novel approach based on the dimension reduction method UMAP. This approach allows the preservation of a broader spectrum of information from the data, enriching the descriptive detail of the resulting visual representations. The analysis results show that UMAP can generate low-dimensional representations that preserve the data structure and class separability, thus the information content of multi-frequency PolInSAR data to a high degree. In this respect, UMAP is superior to the comparison methods PCA and LE and on par with the t-SNE method. A significant advantage of UMAP over t-SNE is its ability to transfer a projection learned on initial data to newly added unseen data. This allows the desired consistent visualization and analysis across multi-temporal or multi-location datasets. To tailor UMAP to PolInSAR data, two approaches, F-UMAP and W-UMAP, were developed in this thesis, each with distinct advantages. F-UMAP incorporates expert knowledge by relying on a pre-selected set of polarimetric and interferometric features, making it suitable for targeted applications. W-UMAP offers a more generalized approach by exploiting data-specific statistical distributions, which eliminates biases from manual feature selection.

All in all, the proposed UMAP-based approach provides a valuable tool for PolInSAR data visualization. It condenses complex multi-layered information into easily interpretable scatter plots and color images. The scatter plots help detect clusters and patterns within the complex data, while the color images enable visual scene interpretation, enhancing human-driven data analysis.

Automatic Land Cover Classification. Once a deeper understanding of the information contained in the data has been attained through human-interpretable data visualization, the natural next step involves developing algorithms that leverage this information for

automatic semantic interpretation. This automatic analysis is crucial for realizing data analysis on a global scale. In this work, the promising technology of CNNs was transferred to the classification of multi-frequency PolInSAR data. The presented analyses aimed to answer the research question of how the multi-layered information of the data can be represented in the real domain to fully exploit it within the classification process. The results demonstrate that using a comprehensive set of polarimetric and interferometric features to represent the data is an effective approach. Using this representation as input for CNNs outperforms the direct employment of the coherency matrix as input. Furthermore, this way of integrating interferometric and multi-frequency information also improves the classification performance compared to classification based solely on PolSAR data.

The stacking of numerous features into a multi-channel input image, processed by a CNN (with standard 2D convolutions), achieves the best results in the conducted experiments. However, a disadvantage of this approach is an increased resource requirement due to the storage and processing of the high-dimensional feature representation. Various dimension reduction methods that transform this representation of multi-frequency PolInSAR data into a low-dimensional representation have been analyzed to address this issue. Good results are achieved with the feature selection method mRMR and the feature transformation method PCA. These methods effectively reduce the dimension of the original 84-dimensional feature representation to 15 dimensions with minimal loss of classification accuracy while significantly reducing training time and memory requirements.

To investigate the performance of CNNs against traditional machine learning models, a comparison with the widely applied Random Forest model was performed. The results show that if a distinction of the target classes already uniquely results from the combination of polarimetric and interferometric features, traditional machine learning models are sufficiently powerful to achieve a high classification performance. Whether a simple separation is possible can be easily assessed using UMAP-based visualizations of the feature space, in which separable classes form well-separated clusters. However, if classes need to be distinguished that overlap in the feature space and can only be unambiguously identified using the broader image context or complex image features, CNNs demonstrate superior suitability. This applies in particular to classes that correspond not only to land cover but rather to land use or represent complex natural structures.

All in all, the comprehensive analysis on applying CNNs to multi-frequency PolInSAR data represents a fundamental first step in unlocking the potential of combining this highly informative image data with the powerful feature extraction capabilities of CNNs. In anticipation of the upcoming NISAR and Rose-L satellite missions, this advancement opens up opportunities for a wide range of applications, including detailed vegetation, forest and habitat mapping, precision agriculture, and crop monitoring or urban planning.

6.2 Outlook

This final section discusses potential improvements to the studies presented in this thesis and suggests directions for future research.

UMAP-based Scene Visualization. The proposed UMAP-based scene visualization is a versatile tool that not only offers the potential for analyzing PolInSAR data but can be transferred to any kind of multi-dimensional remote sensing data, including hyperspectral data, multi-sensor data, or spatiotemporal data. To improve the approach further, future work should investigate using different color spaces for mapping low-dimensional representations to color. In particular, perceptually uniform color spaces should be considered, which can represent distances in the feature space so that they correspond to the color differences perceived by the human eye. Another interesting future research question related to this work is how to directly represent the relationship between color and land cover for a more straightforward interpretation. One approach that could be tested is integrating RGB images using manifold alignment, which could help colorize scenes in natural color tones.

CNN-based Classification. This thesis provided insights into the most suitable data representation for CNN-based classification of PolSAR or PolInSAR data. The results indicate that the commonly used representation in the literature in the context of PolSAR data classification (coherency matrix elements in real and imaginary parts) is not optimal. It would be interesting to test whether representing the coherency matrix elements in magnitude and phase, which proved to be optimal in the studies of this work, can also enhance other existing approaches for PolSAR image classification.

The results of the CNN-based classification of PolInSAR data demonstrate that extracting polarimetric and interferometric features before inputting them into a CNN model is advantageous. The approach of stacking real-valued features (optionally followed by dimension reduction) allows the direct use of established CNN models, which have been successfully utilized to analyze optical data. Thus, in addition to segmentation networks such as U-Net, this approach enables the application and testing of CNN models for other tasks such as object classification, object detection or parameter estimation in future work.

Transfer to Spaceborne Data. Forthcoming work should build on the results of this thesis and investigate the transferability of the proposed methods to different sensor systems, especially with regard to the upcoming spaceborne missions. Due to limited data availability, all analyses in this work were carried out exclusively on the Pol-InSAR-Island dataset, which comprises data acquired by a single (airborne) sensor system over a single location.

Therefore, the transferability of the findings collected in this work to other PolInSAR systems that provide different spatial resolutions, acquisition geometry, and interferometric baselines remains to be tested. Nevertheless, since the proposed data visualization and classification approaches make no implicit assumptions about the data, their application to data acquired by other systems is generally straightforward. The UMAP-based visualization of different data sets will allow analysts to understand the information contained in the data and identify potential target applications. In terms of classification, however, CNN models trained on F-SAR data are not expected to provide good classification results for other sensor data. However, the classification process, including suitable data representation, as well as CNN model architecture and training design, can be adapted to train a model on new data to learn sensor and task-specific features. Therefore, the most significant challenge will be to collect reliable reference data to enable such training. To overcome the limitation of missing reference data, self-supervised learning, which aims to learn semantically meaningful and task-agnostic features from unlabeled data, could be explored, representing an exciting future research topic.

Declaration of Supportive Resources

During the implementation of the presented algorithms and throughout the writing process of this document, additional resources were utilized to aid in the completion and enhancement of the work. The following outlines the resources used and describes how each was employed to support both technical and writing tasks.

List of Software Packages and Toolboxes:

- **PolSARpro v6.0 (Biomass Edition) Toolbox - STEP** [283, 284]: This tool was used in processing the F-SAR PolInSAR data into a usable format for further analysis. It was used for filtering and extracting polarimetric and interferometric features.
- **PyTorch and PyTorch-Lightning** [285, 286]: These frameworks were utilized to generate, train, and test CNN models.
- **Scikit-learn** [183, 287]: This library was used to implement dimension reduction techniques and machine learning classifiers.
- **Umap-learn** [184]: This Python package was used to implement the UMAP algorithm.
- **Pandas [288] and Seaborn [289]**: These Python packages were employed for the visualization of results. These libraries helped in generating informative graphical representations of numerical results.

List of Tools Supporting the Writing Process:

- **DeepL** [290]: This translation tool was used for converting sections of the text written in German to English.
- **Grammarly** [291]: This tool provided grammar and spelling checks, as well as stylistic suggestions to enhance the readability and clarity of the text.
- **FhGenie, based on GPT-4 and GPT-4.o** [292]: The Large Language Models were used for grammar and spelling suggestions, as well as stylistic improvements. It should be noted that only self-authored texts were used as input, and only suggestions from these tools were evaluated and partly incorporated to augment the readability and comprehension of the text.

Bibliography

[1] Jong-Sen Lee and Eric Pottier. *Polarimetric radar imaging: From basics to applications*. CRC Press, 2017.

[2] Shane R. Cloude and Konstantinos P. Papathanassiou. "Polarimetric SAR interferometry". In: *IEEE Transactions on Geoscience and Remote Sensing* 36.5 (1998), pp. 1551–1565.

[3] Shashi Kumar, Unmesh G. Khati, Shreya Chandola, Shefali Agrawal, and Satya P. S. Kushwaha. "Polarimetric SAR interferometry based modeling for tree height and aboveground biomass retrieval in a tropical deciduous forest". In: *Advances in Space Research* 60.3 (2017), pp. 571–586.

[4] Akshay Patil, Shradha Mohanty, and Gulab Singh. "Snow depth and snow water equivalent retrieval using X-band PolInSAR data". In: *Remote Sensing Letters* 11.9 (2020), pp. 817–826.

[5] Qi Zhang, Linlin Ge, Scott Hensley, Graciela Isabel Metternicht, Chang Liu, and Ruiheng Zhang. "PolGAN: A deep-learning-based unsupervised forest height estimation based on the synergy of PolInSAR and LiDAR data". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 186 (2022), pp. 123–139.

[6] Michal Shimoni, Dirk Borghys, Roel Heremans, Christiaan Perneel, and Marc Acheroy. "Fusion of PolSAR and PolInSAR data for land cover classification". In: *International Journal of Applied Earth Observation and Geoinformation* 11.3 (2009), pp. 169–180.

[7] Fariba Mohammadimanesh, Bahram Salehi, Masoud Mahdianpari, Brian Brisco, and Mahdi Motagh. "Multi-temporal, multi-frequency, and multi-polarization coherence and SAR backscatter analysis of wetlands". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 142 (2018), pp. 78–93.

[8] Sylvia Schmitz, Martin Weinmann, and Antje Thiele. "Incorporating interferometric coherence into LULC classification of airborne PolSAR images using Fully Convolutional Networks". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XLIII-B1-2020. 2020, pp. 115–122.

[9] Kun-Shan Chen, WP Huang, DH Tsay, and Faouzi Amar. "Classification of multifrequency polarimetric SAR imagery using a dynamic learning neural network". In: *IEEE Transactions on Geoscience and Remote Sensing* 34.3 (1996), pp. 814–820.

[10] Laurent Ferro-Famil, Eric Pottier, and Jong-Sen Lee. "Classification and interpretation of polarimetric interferometric SAR data". In: *2002 IEEE International Geoscience and Remote Sensing Symposium*. 2002, pp. 635–637.

[11] Varsha Turkar, Rinki Deo, Y. S. Rao, Shiv Mohan, and Anup Das. "Classification accuracy of multi-frequency and multi-polarization SAR images for various land covers". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5.3 (2012), pp. 936–941.

[12] Kent Kellogg, Pamela Hoffman, Shaun Standley, Scott Shaffer, Paul Rosen, Wendy Edelstein, Charles Dunn, Charles Baker, Phillip Barela, Yuhsyen Shen, Ana Maria Guerrero, Peter Xaypraseuth, V Raju Sagi, C V Sreekantha, Nandini Harinath, Raj Kumar, Rakesh Bhan, and C. V. H. S. Sarma. "NASA-ISRO Synthetic Aperture Radar (NISAR) mission". In: *2020 IEEE Aerospace Conference*. 2020, pp. 1–21.

[13] Peijun Du, Alim Samat, Björn Waske, Sicong Liu, and Zhenhong Li. "Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 105 (2015), pp. 38–53.

[14] Lien Loosvelt, Jan Peters, Henning Skriver, Bernard De Baets, and Niko E. C. Verhoest. "Impact of reducing polarimetric SAR input on the uncertainty of crop classifications based on the Random Forests Algorithm". In: *IEEE Transactions on Geoscience and Remote Sensing* 50.10 (2012), pp. 4185–4200.

[15] Wensheng Wang, Xiaofeng Yang, Xiaofeng Li, Kunshan Chen, Guihong Liu, Ziwei Li, and Martin Gade. "A fully polarimetric SAR imagery classification scheme for mud and sand flats in intertidal zones". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.3 (2017), pp. 1734–1742.

[16] Xiaotong Zhang, Jia Xu, Yuanyuan Chen, Kang Xu, and Dongmei Wang. "Coastal wetland classification with GF-3 polarimetric SAR imagery by using oject-oriented Random Forest Algorithm". In: *Sensors* 21.10 (2021), p. 3395.

[17] Chengquan Huang, L.S. Davis, and J.R.G. Townshend. "An assessment of Support Vector Machines for land cover classification". In: *International Journal of Remote Sensing* 23.4 (2002), pp. 725–749.

[18] Seisuke Fukuda and Haruto Hirosawa. "Support Vector Machine classification of land cover: Application to polarimetric SAR data". In: *2001 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 1. 2001, pp. 187–189.

[19] Cédric Lardeux, P-L Frison, J-P Rudant, J-C Souyris, Céline Tison, and Benoıt Stoll. "Use of the SVM classification with polarimetric SAR data for land use cartography". In: *2006 IEEE International Symposium on Geoscience and Remote Sensing*. 2006, pp. 493–496.

[20] Yu Zhou, Haipeng Wang, Feng Xu, and Ya-Qiu Jin. "Polarimetric SAR image classification using deep convolutional neural networks". In: *IEEE Geoscience and Remote Sensing Letters* 13.12 (2016), pp. 1935–1939.

[21] Si-Wei Chen and Chen-Song Tao. "PolSAR image classification using polarimetric-feature-driven deep Convolutional Neural Network". In: *IEEE Geoscience and Remote Sensing Letters* 15.4 (2018), pp. 627–631.

[22] Fei Gao, Teng Huang, Jun Wang, Jinping Sun, Amir Hussain, and Erfu Yang. "Dual-Branch Deep Convolution Neural Network for Polarimetric SAR Image Classification". In: *Applied Sciences* 7.5 (2017), p. 447.

[23] Xu Liu, Licheng Jiao, Xu Tang, Qigong Sun, and Dan Zhang. "Polarimetric Convolutional Network for PolSAR Image Classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 57.5 (2019), pp. 3040–3054.

[24] Adugna G. Mullissa, Claudio Persello, and Alfred Stein. "PolSARNet: A deep Fully Convolutional Network for polarimetric SAR image classification". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.12 (2019), pp. 5300–5309.

[25] Hemani Parikh, Samir Patel, and Vibha Patel. “Classification of SAR and PolSAR images using deep learning: a review”. In: *International Journal of Image and Data Fusion* 11.1 (2020), pp. 1–32.

[26] Leland McInnes, John Healy, and James Melville. “UMAP: Uniform Manifold Approximation and Projection for dimension reduction”. In: *arXiv:1802.03426 [cs, stat]* (2018).

[27] Tina Smets, Nico Verbeeck, Marc Claesen, Arndt Asperger, Gerard Griffioen, Thomas Tousseyn, Wim Waelput, Etienne Waelkens, and Bart De Moor. “Evaluation of distance metrics and spatial autocorrelation in Uniform Manifold Approximation and Projection applied to mass spectrometry imaging data”. In: *Analytical Chemistry* 91.9 (2019), pp. 5706–5714.

[28] Etienne Becht, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W. H. Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W. Newell. “Dimensionality reduction for visualizing single-cell data using UMAP”. In: *Nature Biotechnology* 37.1 (2019), pp. 38–44.

[29] Evgeny Myasnikov. “Using UMAP for dimensionality reduction of hyperspectral data”. In: *2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*. 2020, pp. 1–5.

[30] Alexander Pletl, Michael Fernandes, Nicolas Thomas, Angelo Pio Rossi, and Benedikt Elser. “Spectral Clustering of CRISM datasets in Jezero Crater using UMAP and k-Means”. In: *Remote Sensing* 15.4 (2023), p. 939.

[31] Helmut Klausing and Wolfgang Holpp. *Radar mit realer und synthetischer Apertur: Konzeption und Realisierung*. Walter de Gruyter, 2010.

[32] Richard Bamler. “Principles of Synthetic Aperture Radar”. In: *Surveys in Geophysics* 21.2 (2000), pp. 147–157.

[33] Arsenios Tsokas, Maciej Rysz, Panos M. Pardalos, and Kathleen Dipple. “SAR data applications in earth observation: An overview”. In: *Expert Systems with Applications* 205 (2022), p. 117342.

[34] Christian Wolff. *Radar Basics*. URL: https://www.radartutorial.eu/20.airborne/ab06.de.html (visited on 09/12/2024).

[35] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P. Papathanassiou. “A tutorial on Synthetic Aperture Radar”. In: *IEEE Geoscience and Remote Sensing Magazine* 1.1 (2013), pp. 6–43.

[36] Joseph W. Goodman. “Some fundamental properties of speckle”. In: *Journal of the Optical Society of America* 66.11 (1976), pp. 1145–1150.

[37] Jong-Sen Lee, L Jurkevich, Piet Dewaele, Patrick Wambacq, and André Oosterlinck. “Speckle filtering of Synthetic Aperture Radar images: A review”. In: *Remote Sensing Reviews* 8.4 (1994), pp. 313–340.

[38] Fabrizio Argenti, Alessandro Lapini, Tiziano Bianchi, and Luciano Alparone. “A tutorial on speckle reduction in Synthetic Aperture Radar images”. In: *IEEE Geoscience and Remote Sensing Magazine* 1.3 (2013), pp. 6–35.

[39] Loïc Denis, Emanuele Dalsasso, and Florence Tupin. “A review of deep-learning techniques for SAR image restoration”. In: *2021 IEEE International Geoscience and Remote Sensing Symposium*. 2021, pp. 411–414.

[40] Irena Hajnsek and Yves-Louis Desnos. *Polarimetric Synthetic Aperture Radar: Principles and application*. Springer Nature, 2021.

[41] Eric Pottier. *PolSAR theory, 6th ESA advanced course on radar polarimetry 2021*. 2021. URL: https://eo4society.esa.int/resources/6th-advanced-polarimetry-2021/ (visited on 09/17/2024).

[42] Robert Clark Jones. "A new calculus for the treatment of optical systems. IV." In: *Journal of the Optical Society of America* 32.8 (1942), pp. 486–493.

[43] Chris Oliver and Shaun Quegan. *Understanding Synthetic Aperture Radar images*. SciTech Publishing, 2004.

[44] Samuel Foucher and Carlos López-Martínez. "Analysis, evaluation, and comparison of polarimetric SAR speckle filtering techniques". In: *IEEE Transactions on Image Processing* 23.4 (2014), pp. 1751–1764.

[45] Jong-Sen Lee, Mitchell R Grunes, and Gianfranco De Grandi. "Polarimetric SAR speckle filtering and its implication for classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 37.5 (1999), pp. 2363–2373.

[46] Jong-Sen Lee. "Refined filtering of image noise using local statistics". In: *Computer Graphics and Image Processing* 15.4 (1981), pp. 380–389.

[47] Jong-Sen Lee. "Digital image enhancement and noise filtering by use of local statistics". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-2.2 (1980), pp. 165–168.

[48] Ernst Krogager. "New decomposition of the radar target scattering matrix". In: *Electronics Letters* 18.26 (1990), pp. 1525–1527.

[49] William L. Cameron and Ling K. Leung. "Feature motivated polarization scattering matrix decomposition". In: *IEEE International Conference on Radar*. 1990, pp. 549–557.

[50] Konstantinos Karachristos, Georgia Koukiou, and Vassilis Anastassopoulos. "A review on PolSAR decompositions for feature extraction". In: *Journal of Imaging* 10.4 (2024), p. 75.

[51] Shane R. Cloude and Eric Pottier. "A review of target decomposition theorems in radar polarimetry". In: *IEEE Transactions on Geoscience and Remote Sensing* 34.2 (1996), pp. 498–518.

[52] Shane R. Cloude and Eric Pottier. "An entropy based classification scheme for land applications of polarimetric SAR". In: *IEEE Transactions on Geoscience and Remote Sensing* 35.1 (1997), pp. 68–78.

[53] Anthony Freeman and Stephen L. Durden. "A three-component scattering model for polarimetric SAR data". In: *IEEE Transactions on Geoscience and Remote Sensing* 36.3 (1998), pp. 963–973.

[54] Yoshio Yamaguchi, Yuki Yajima, and Hiroyoshi Yamada. "A four-component decomposition of PolSAR images based on the coherency matrix". In: *IEEE Geoscience and Remote Sensing Letters* 3.3 (2006), pp. 292–296.

[55] Jakob J. van Zyl. "Application of Cloude's target decomposition theorem to polarimetric imaging radar data". In: *Radar Polarimetry*. Vol. 1748. 1993, pp. 184–191.

[56] My-Linh Truong-Loi, Anthony Freeman, Pascale C. Dubois-Fernandez, and Eric Pottier. "Estimation of soil moisture and faraday rotation from bare surfaces using compact polarimetry". In: *IEEE Transactions on Geoscience and Remote Sensing* 47.11 (2009), pp. 3608–3615.

[57] Jaan Praks, Elise Colin Koeniguer, and Martti T. Hallikainen. "Alternatives to target entropy and alpha angle in SAR polarimetry". In: *IEEE Transactions on Geoscience and Remote Sensing* 47.7 (2009), pp. 2262–2274.

[58] José Jorge Gil and Eusebio Bernabeu. "A depolarization criterion in Mueller Matrices". In: *Optica Acta: International Journal of Optics* 32.3 (1985), pp. 259–261.

[59] Richard Bamler and Philipp Hartl. "Synthetic Aperture Radar interferometry". In: *Inverse Problems* 14.4 (1998), R1.

[60] Kazuyoshi Itoh. "Analysis of the phase unwrapping algorithm". In: *Applied Optics* 21.14 (1982), pp. 2470–2470.

[61] Hanwen Yu, Yang Lan, Zhihui Yuan, Junyi Xu, and Hyongki Lee. "Phase unwrapping in InSAR : A review". In: *IEEE Geoscience and Remote Sensing Magazine* 7.1 (2019), pp. 40–58.

[62] Richard M. Goldstein, Howard A. Zebker, and Charles L. Werner. "Satellite radar interferometry: Two-dimensional phase unwrapping". In: *Radio Science* 23.4 (1988), pp. 713–720.

[63] Ming Zhao, Lei Huang, Qican Zhang, Xianyu Su, Anand Asundi, and Qian Kemao. "Quality-guided phase unwrapping technique: comparison of quality maps and guiding strategies". In: *Applied Optics* 50.33 (2011), pp. 6214–6224.

[64] David Small, Paolo Pasquali, and Stefan Fuglistaler. "A comparison of phase to height conversion methods for SAR interferometry". In: *1996 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 1. 1996, pp. 342–344.

[65] Thomas Flynn, Mark Tabb, and Richard Carande. "Coherence region shape extraction for vegetation parameter estimation in polarimetric SAR interferometry". In: *2002 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 5. 2002, pp. 2596–2598.

[66] Elise Colin, C Titin-Schnaider, and W Tabbara. "Coherence optimization methods for scattering centers separation in polarimetric interferometry". In: *Journal of Electromagnetic Waves and Applications* 19.9 (2005), pp. 1237–1250.

[67] Shane R. Cloude and Konstantinos P. Papathanassiou. "Three-stage inversion process for polarimetric SAR interferometry". In: *IEEE Proceedings - Radar, Sonar and Navigation* 150.3 (2003), pp. 125–134.

[68] Sylvia Hochstuhl, Niklas Pfeffer, Antje Thiele, Stefan Hinz, Joel Amao-Oliva, Rolf Scheiber, Andreas Reigber, and Holger Dirks. "Pol-InSAR-Island - A benchmark dataset for multi-frequency Pol-InSAR data land cover classification". In: *ISPRS Open Journal of Photogrammetry and Remote Sensing* 10 (2023), p. 100047.

[69] Xiao Xiang Zhu, Sina Montazeri, Mohsin Ali, Yuansheng Hua, Yuanyuan Wang, Lichao Mou, Yilei Shi, Feng Xu, and Richard Bamler. "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives". In: *IEEE Geoscience and Remote Sensing Magazine* 9.4 (2021), pp. 143–172.

[70] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. “ImageNet: A large-scale hierarchical image database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–255.

[71] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. “The Cityscapes dataset for semantic urban scene understanding”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 3213–3223.

[72] Martin a. M. Vissers and Joost J. van der Sanden. *Groundtruth collection for the JPL-SAR and ERS-1 campaign in Flevoland and the Veluwe (NL) 1991*. Tech. rep. 31. Landbouwuniversiteit Wageningen, 1992.

[73] Xu Liu, Licheng Jiao, Fang Liu, Dan Zhang, and Xu Tang. “PolSF: PolSAR image datasets on san Francisco”. In: *International Conference on Intelligence Science*. Springer. 2022, pp. 214–219.

[74] Zhirui Wang, Xuan Zeng, Zhiyuan Yan, Jian Kang, and Xian Sun. “AIR-PolSAR-Seg: A large-scale data set for terrain segmentation in complex-scene PolSAR images”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022), pp. 3830–3841.

[75] Ralf Horn, Anton Nottensteiner, Andreas Reigber, Jens Fischer, and Rolf Scheiber. “F-SAR — DLR’s new multifrequency polarimetric airborne SAR”. In: *2009 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 2. 2009, pp. 902–905.

[76] Muriel Pinheiro, Rolf Scheiber, Joel Amao-Oliva, Marc Jaeger, Ralf Horn, Jens Fischer, Martin Keller, and Andreas Reigber. “Airborne SAR interferometry for large scale mapping of tidal flats: the GeoWAM project”. In: *13th European Conference on Synthetic Aperture Radar*. 2021, pp. 1–6.

[77] Muriel Pinheiro, Joel Amao-Oliva, Rolf Scheiber, Marc Jaeger, Ralf Horn, Martin Keller, Jens Fischer, and Andreas Reigber. “Dual-frequency airborne SAR for large scale mapping of tidal flats”. In: *Remote Sensing* 12.11 (2020), p. 1827.

[78] Sylvia Schmitz, Antje Thiele, Horst Hammer, and Stefan Hinz. “Impact of highly varying incidence angle on the feature space of PolSAR images”. In: *14th European Conference on Synthetic Aperture Radar*. 2022, pp. 1–6.

[79] Sylvia Schmitz, Uwe Weidner, Horst Hammer, and Antje Thiele. “Evaluating Uniform Manifold Approximation and Projection for dimension reduction and visualization of PolInSAR features”. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. V-1-2021. 2021, pp. 39–46.

[80] Sylvia Schmitz, Horst Hammer, and Antje Thiele. “Multi-frequency PolInSAR data are advantageous for land cover classification – A visual and quantitative analysis”. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. V-1-2022. 2022, pp. 49–56.

[81] Sylvia Hochstuhl, Horst Hammer, Antje Thiele, and Stefan Hinz. “Wishart-UMAP- Translating pixel similarity of PolSAR images to real-valued feature descriptors”. In: *2023 IEEE International Geoscience and Remote Sensing Symposium*. 2023, pp. 5966–5969.

[82] J. David Ballester-Berman and Juan M. Lopez-Sanchez. “Applying the Freeman–Durden decomposition concept to polarimetric SAR interferometry”. In: *IEEE Transactions on Geoscience and Remote Sensing* 48.1 (2010), pp. 466–479.

[83] Si-Wei Chen, Xue-Song Wang, Yong-Zhen Li, and Motoyuki Sato. “Adaptive model-based polarimetric decomposition using PolInSAR coherence”. In: *IEEE Transactions on Geoscience and Remote Sensing* 52.3 (2014), pp. 1705–1718.

[84] Shahid Shuja Shafai and Shashi Kumar. “PolInSAR coherence and entropy-based hybrid decomposition model”. In: *Earth and Space Science* 7.10 (2020), e2020EA001279.

[85] Lamei Zhang, Liangjie Sun, Bin Zou, and Wooil Moon. “Fully polarimetric SAR image classification via sparse representation and polarimetric features”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (2015), pp. 3923–3932.

[86] K. P. Papathanassiou, S. R. Cloude, M. Pardini, M. J. Quinones, D. Hoekman, L. Ferro-Famil, D. Goodenough, H. Chen, S. Tebaldini, M. Neumann, L. M. H. Ulander, and M. J. Soja. “Forest Applications”. In: *Polarimetric Synthetic Aperture Radar: Principles and Application*. Ed. by Irena Hajnsek and Yves-Louis Desnos. Springer International Publishing, 2021, pp. 59–117.

[87] Donato Amitrano, Gerardo Di Martino, Antonio Iodice, Daniele Riccio, and Giuseppe Ruello. “A new framework for SAR multitemporal data RGB representation: Rationale and products”. In: *IEEE Transactions on Geoscience and Remote Sensing* 53.1 (2015), pp. 117–133.

[88] Silvana G. Dellepiane and Elena Angiati. “A new method for cross-normalization and multitemporal visualization of SAR images for the detection of flooded areas”. In: *IEEE Transactions on Geoscience and Remote Sensing* 50.7 (2012), pp. 2765–2779.

[89] Abeer Mazher. “Visualization framework for high-dimensional spatio-temporal hydrological gridded datasets using machine-learning techniques”. In: *Water* 12.2 (2020), p. 590.

[90] Daniel Sousa and Christopher Small. “Joint characterization of Sentinel-2 reflectance: Insights from manifold learning”. In: *Remote Sensing* 14.22 (2022), p. 5688.

[91] Tulsi Patel, Mark W. Jones, and Thomas Redfern. “Manifold Explorer: Satellite image labelling and clustering tool with using deep Convolutional Autoencoders”. In: *Algorithms* 16.10 (2023), p. 469.

[92] P. Ready and Paul Wintz. “Information extraction, SNR improvement, and data compression in multispectral imagery”. In: *IEEE Transactions on Communications* 21.10 (1973), pp. 1123–1131.

[93] Xiuping Jia and John A Richards. “Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification”. In: *IEEE Transactions on Geoscience and Remote Sensing* 37.1 (1999), pp. 538–542.

[94] J. Scott Tyo, Athanasios Konsolakis, David I. Diersen, and Richard Christopher Olsen. “Principal-components-based display strategy for spectral imagery”. In: *IEEE Transactions on Geoscience and Remote Sensing* 41.3 (2003), pp. 708–718.

[95] Vassilis Tsagaris and Vassilis Anastassopoulos. “Multispectral image fusion for improved RGB representation based on perceptual attributes”. In: *International Journal of Remote Sensing* 26.15 (2005), pp. 3241–3254.

[96] Qian Du, Nareenart Raksuntorn, Shangshu Cai, and Robert J. Moorhead. “Color display for hyperspectral imagery”. In: *IEEE Transactions on Geoscience and Remote Sensing* 46.6 (2008), pp. 1858–1866.

[97] Charles M. Bachmann, Thomas L. Ainsworth, and Robert A. Fusina. “Exploiting manifold geometry in hyperspectral imagery”. In: *IEEE Transactions on Geoscience and Remote Sensing* 43.3 (2005), pp. 441–454.

[98] Jun He, Lei Zhang, Qing Wang, and Zigang Li. “Using diffusion geometric coordinates for hyperspectral imagery representation”. In: *IEEE Geoscience and Remote Sensing Letters* 6.4 (2009), pp. 767–771.

[99] Johannes Jordan and Elli Angelopoulou. “Hyperspectral image visualization with a 3-D self-organizing map”. In: *2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*. 2013, pp. 1–4.

[100] Thomas Villmann and Erzsébet Merényi. “Extensions and modifications of the Kohonen-SOM and applications in remote sensing image analysis”. In: *Self-Organizing Neural Networks: Recent Advances and Applications*. 2002, pp. 121–144.

[101] Yi Long, Heng-Chao Li, Turgay Celik, Nathan Longbotham, and William J. Emery. “Pairwise-distance-analysis-driven dimensionality reduction model with double mappings for hyperspectral image visualization”. In: *Remote Sensing* 7.6 (2015), pp. 7785–7808.

[102] Safa A. Najim, Ik Soo Lim, Peter Wittek, and Mark W. Jones. “FSPE: Visualization of hyperspectral imagery using faithful stochastic proximity embedding”. In: *IEEE Geoscience and Remote Sensing Letters* 12.1 (2015), pp. 18–22.

[103] Danping Liao, Yuntao Qian, Jun Zhou, and Yuan Yan Tang. “A manifold alignment approach for hyperspectral image visualization with natural color”. In: *IEEE Transactions on Geoscience and Remote Sensing* 54.6 (2016), pp. 3151–3162.

[104] Dmitry Kobak and Philipp Berens. “The art of using t-SNE for single-cell transcriptomics”. In: *Nature Communications* 10.1 (2019), p. 5416.

[105] George C. Linderman and Stefan Steinerberger. “Clustering with t-SNE, provably”. In: *SIAM Journal on Mathematics of Data Science* 1.2 (Jan. 2019), pp. 313–332.

[106] Frederik Otzen Bagger, Savvas Kinalis, and Nicolas Rapin. “BloodSpot: A database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles”. In: *Nucleic Acids Research* 47.D1 (2019), pp. 881–885.

[107] Maria Kuksin, Daphné Morel, Marine Aglave, François-Xavier Danlos, Aurélien Marabelle, Andrei Zinovyev, Daniel Gautheret, and Loïc Verlingue. “Applications of single-cell and bulk RNA sequencing in onco-immunology”. In: *European Journal of Cancer* 149 (2021), pp. 193–210.

[108] Malte D. Luecken and Fabian J. Theis. “Current best practices in single-cell RNA-seq analysis: a tutorial”. In: *Molecular Systems Biology* 15.6 (2019), e8746.

[109] Beate Vieth, Swati Parekh, Christoph Ziegenhain, Wolfgang Enard, and Ines Hellmann. “A systematic evaluation of single cell RNA-seq analysis pipelines”. In: *Nature Communications* 10.1 (2019), p. 4667.

[110] Yang Yang, Hongjian Sun, Yu Zhang, Tiefu Zhang, Jialei Gong, Yunbo Wei, Yong-Gang Duan, Minglei Shu, Yuchen Yang, Di Wu, and Di Yu. “Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data”. In: *Cell Reports* 36.4 (2021), p. 109442.

[111] Alex Diaz-Papkovich, Luke Anderson-Trocmé, and Simon Gravel. “A review of UMAP in population genetics”. In: *Journal of Human Genetics* 66.1 (Jan. 2021), pp. 85–91.

[112] Liang Zhou, Mengjie Fan, Charles Hansen, Chris R. Johnson, and Daniel Weiskopf. “A review of three-dimensional medical image visualization”. In: *Health Data Science* (2022), p. 9840519.

[113] Mohammed Ali, Mark W. Jones, Xianghua Xie, and Mark Williams. “TimeCluster: Dimension reduction applied to temporal data for visual analytics”. In: *The Visual Computer* 35.6 (2019), pp. 1013–1026.

[114] Fernando Vieira Paulovich and Rosane. Minghim. “Text Map Explorer: A tool to create and explore document maps”. In: *Tenth International Conference on Information Visualisation*. 2006, pp. 245–251.

[115] Rosane Minghim, Fernando Vieira Paulovich, and Alneu de Andrade Lopes. “Content-based text mapping using multi-dimensional projections for exploration of document collections”. In: *Visualization and Data Analysis 2006*. Vol. 6060. 2006, pp. 259–270.

[116] Karl Pearson. “LIII. On lines and planes of closest fit to systems of points in space”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1901), pp. 559–572.

[117] D. N. Lawley and A. E. Maxwell. “Factor Analysis as a statistical method”. In: *Journal of the Royal Statistical Society* 12.3 (1962), pp. 209–229.

[118] Christian Jutten and Jeanny Herault. “Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture”. In: *Signal Processing* 24.1 (1991), pp. 1–10.

[119] Daniel Lee and H. Sebastian Seung. “Algorithms for Non-negative Matrix Factorization”. In: *Advances in Neural Information Processing Systems*. Vol. 13. MIT Press, 2000.

[120] Xiaofei He and Partha Niyogi. “Locality Preserving Projections”. In: *Advances in Neural Information Processing Systems*. Vol. 16. MIT Press, 2003.

[121] Kilian Q Weinberger, John Blitzer, and Lawrence Saul. “Distance metric learning for large margin nearest neighbor classification”. In: *Advances in Neural Information Processing Systems*. Vol. 18. 2005.

[122] Tianhao Zhang, Jie Yang, Deli Zhao, and Xinliang Ge. “Linear local tangent space alignment and application to face recognition”. In: *Neurocomputing* 70.7 (2007), pp. 1547–1553.

[123] John W. Sammon. “A nonlinear mapping for data structure analysis”. In: *IEEE Transactions on Computers* C-18.5 (1969), pp. 401–409.

[124] Joseph B. Kruskal and Myron Wish. *Multidimensional Scaling*. SAGE, 1978.

[125] Teuvo Kohonen. “The self-organizing map”. In: *Proceedings of the IEEE* 78.9 (1990), pp. 1464–1480.

[126] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. “Kernel principal component analysis”. In: *International Conference on Artificial Neural Networks*. Springer. 1997, pp. 583–588.

[127] Pierre Demartines and Jeanny Hérault. “Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets”. In: *IEEE Transactions on Neural Networks* 8.1 (1997), pp. 148–154.

[128] Gaston Baudat and Fatiha Anouar. “Generalized discriminant analysis using a kernel approach”. In: *Neural Computation* 12.10 (2000), pp. 2385–2404.

[129] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. “A global geometric framework for nonlinear dimensionality reduction”. In: *Science* 290.5500 (2000), pp. 2319–2323.

[130] Neil Lawrence. “Gaussian process latent variable models for visualisation of high dimensional data”. In: *Advances in Neural Information Processing Systems*. Vol. 16. 2003.

[131] Santosh S. Vempala. *The Random Projection Method*. American Mathematical Soc., 2005.

[132] Ronald R. Coifman and Stéphane Lafon. “Diffusion Maps”. In: *Applied and Computational Harmonic Analysis* 21.1 (2006), pp. 5–30.

[133] Diederik P. Kingma. “Auto-encoding variational bayes”. In: *arXiv preprint arXiv:1312.6114* (2013).

[134] Sam T. Roweis and Lawrence K. Saul. “Nonlinear dimensionality reduction by Locally Linear Embedding”. In: *Science* 290.5500 (2000), pp. 2323–2326.

[135] Matthew Brand. “Charting a manifold”. In: *Advances in Neural Information Processing Systems*. Vol. 15. 2002.

[136] Mikhail Belkin and Partha Niyogi. “Laplacian Eigenmaps for dimensionality reduction and data representation”. In: *Neural Computation* 15.6 (2003), pp. 1373–1396.

[137] Zhenyue Zhang and Hongyuan Zha. “Principal manifolds and nonlinear dimensionality reduction via Tangent Space Alignment”. In: *SIAM Journal on Scientific Computing* 26.1 (2004), pp. 313–338.

[138] Amir Globerson and Sam Roweis. “Metric learning by collapsing classes”. In: *Advances in Neural Information Processing Systems*. Vol. 18. MIT Press, 2005.

[139] Laurens Van der Maaten and Geoffrey Hinton. “Visualizing data using t-SNE”. In: *Journal of Machine Learning Research* 9.11 (2008).

[140] Fernando V. Paulovich, Luis G. Nonato, Rosane Minghim, and Haim Levkowitz. “Least Square Projection: A fast high-precision multidimensional projection technique and its application to document mapping”. In: *IEEE Transactions on Visualization and Computer Graphics* 14.3 (2008), pp. 564–575.

[141] Jian Tang, Jingzhou Liu, Ming Zhang, and Qiaozhu Mei. “Visualizing large-scale and high-dimensional data”. In: *25th International Conference on World Wide Web*. 2016, pp. 287–297.

[142] Laurens Van Der Maaten, Eric O Postma, H Jaap Van Den Herik, et al. “Dimensionality reduction: A comparative review”. In: *Journal of Machine Learning Research* 10.66-71 (2009), p. 13.

[143] Luis Gustavo Nonato and Michaël Aupetit. “Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment”. In: *IEEE Transactions on Visualization and Computer Graphics* 25.8 (2019), pp. 2650–2673.

[144] Rizgar Zebari, Adnan Abdulazeez, Diyar Zeebaree, Dilovan Zebari, and Jwan Saeed. “A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction”. In: *Journal of Applied Science and Technology Trends* 1.2 (2020), pp. 56–70.

[145] Mateus Espadoto, Rafael M. Martins, Andreas Kerren, Nina S. T. Hirata, and Alexandru C. Telea. “Toward a quantitative survey of dimension reduction techniques”. In: *IEEE Transactions on Visualization and Computer Graphics* 27.3 (2021), pp. 2153–2173.

[146] Shan Carter, Zan Armstrong, Ludwig Schubert, Ian Johnson, and Chris Olah. “Activation Atlas”. In: *Distill* 4.3 (2019), e15.

[147] Lamei Zhang, Siyu Zhang, Bin Zou, and Hongwei Dong. “Unsupervised deep representation learning and few-shot classification of PolSAR images”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2020), pp. 1–16.

[148] Ruizhi Xiang, Wencan Wang, Lei Yang, Shiyuan Wang, Chaohan Xu, and Xiaowen Chen. “A comparison for dimensionality reduction methods of single-cell RNA-seq data”. In: *Frontiers in Genetics* 12 (2021), p. 646936.

[149] Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shaposhnik. “Understanding how dimension reduction tools work: An empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data visualization”. In: *The Journal of Machine Learning Research* 22.1 (2021), pp. 9129–9201.

[150] Thomas Di Martino, Regis Guinvarc’h, Laetitia Thirion-Lefevre, and Elise Colin Koeniguer. “Beets or cotton? Blind extraction of fine agricultural classes using a Convolutional Autoencoder applied to temporal SAR signatures”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–18.

[151] Laurane Charrier, Thomas di Martino, E Colin Koeniguer, Flora Weissgerber, and Aurélien Plyer. “Extracting relevance from SAR temporal profiles on a glacier and an alpine watershed by a deep autoencoder”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 43 (2022), pp. 1309–1316.

[152] Thomas Di Martino, Elise Colin Koeniguer, Laetitia Thirion-Lefevre, and Regis Guinvarc’h. “Modelling of agricultural SAR time series using Convolutional Autoencoder for the extraction of harvesting practices of rice fields”. In: *14th European Conference on Synthetic Aperture Radar*. 2022, pp. 1–6.

[153] Standard C.I.E. *Colorimetry-part 4: CIE 1976 L* a* b* colour space*. 2007.

[154] Ming Cui, Anshuman Razdan, Jiuxiang Hu, and Peter Wonka. “Interactive hyperspectral image visualization using convex optimization”. In: *IEEE Transactions on Geoscience and Remote Sensing* 47.6 (2009), pp. 1673–1684.

[155] Danping Liao, Yuntao Qian, and Yuan Yan Tang. “Constrained manifold learning for hyperspectral imagery visualization”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.4 (2018), pp. 1213–1226.

[156] Thomas L. Ainsworth and Jong-Sen Lee. “Polarimetric SAR image classification-exploiting optimal variables derived from multiple-image datasets”. In: *2004 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 1. 2004.

[157] Shang Tan Tu, Jia Yu Chen, Wen Yang, and Hong Sun. “Laplacian Eigenmaps-based polarimetric dimensionality reduction for SAR image classification”. In: *IEEE Transactions on Geoscience and Remote Sensing* 50.1 (2012), pp. 170–179.

[158] Lei Shi, Lefei Zhang, Jie Yang, Liangpei Zhang, and Pingxiang Li. “Supervised Graph Embedding for polarimetric SAR image classification”. In: *IEEE Geoscience and Remote Sensing Letters* 10.2 (2013), pp. 216–220.

[159] Bo Ren, Biao Hou, Zaidao Wen, Wen Xie, and Licheng Jiao. “PolSAR image classification via multimodal sparse representation-based feature fusion”. In: *International Journal of Remote Sensing* 39.22 (2018), pp. 7861–7880.

[160] Yan Wang, Chu He, Xinlong Liu, and Mingsheng Liao. “A hierarchical Fully Convolutional Network integrated with sparse and low-rank subspace representations for PolSAR imagery classification”. In: *Remote Sensing* 10.2 (2018), p. 342.

[161] Sinong Quan, Deliang Xiang, Wei Wang, Boli Xiong, and Gangyao Kuang. “Scattering feature-driven superpixel segmentation for polarimetric SAR images”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), pp. 2173–2183.

[162] Chu He, Mingxia Tu, Dehui Xiong, and Mingsheng Liao. “Nonlinear manifold learning integrated with fully convolutional networks for PolSAR image classification”. In: *Remote Sensing* 12.4 (2020), p. 655.

[163] Jiao Guo, Henghui Li, Jifeng Ning, Wenting Han, Weitao Zhang, and Zheng-Shu Zhou. “Feature dimension reduction using stacked sparse auto-encoders for crop classification with multi-temporal, quad-Pol SAR data”. In: *Remote Sensing* 12.2 (2020), p. 321.

[164] Jun Xu, Lei Xiang, Qingshan Liu, Hannah Gilmore, Jianzhong Wu, Jinghai Tang, and Anant Madabhushi. “Stacked Sparse Autoencoder (SSAE) for nuclei detection on breast cancer histopathology images”. In: *IEEE Transactions on Medical Imaging* 35.1 (2016), pp. 119–130.

[165] Mingliang Tao, Feng Zhou, and Zijing Zhang. “Tensor based dimension reduction for polarimetric SAR data”. In: *2014 IEEE Geoscience and Remote Sensing Symposium*. 2014, pp. 2802–2805.

[166] Hongying Liu, Zhi Wang, Fanhua Shang, Shuyuan Yang, Shuiping Gou, and Licheng Jiao. “Semi-supervised tensorial Locally Linear Embedding for feature extraction using PolSAR data”. In: *IEEE Journal of Selected Topics in Signal Processing* 12.6 (2018), pp. 1476–1490.

[167] Xiayuan Huang, Xiangli Nie, Hong Qiao, and Bo Zhang. “Supervised Polsar image classification by combining multiple features”. In: *2019 IEEE International Conference on Image Processing*. 2019, pp. 634–638.

[168] Bo Ren, Biao Hou, Jocelyn Chanussot, and Licheng Jiao. “PolSAR feature extraction via Tensor Embedding Framework for land cover classification”. In: *IEEE Transactions on Geoscience and Remote Sensing* 58.4 (2020), pp. 2337–2351.

[169] Chun Liu, Junjun Yin, Jian Yang, and Wei Gao. “Classification of multi-frequency polarimetric SAR images based on multi-linear subspace learning of tensor objects”. In: *Remote Sensing* 7.7 (2015), pp. 9253–9268.

[170] Paul R Kersten, Jong-Sen Lee, and Thomas L Ainsworth. “Unsupervised classification of polarimetric Synthetic Aperture Radar images using fuzzy clustering and EM clustering”. In: *IEEE Transactions on Geoscience and Remote Sensing* 43.3 (2005), pp. 519–527.

[171] Stian Normann Anfinsen, Robert Jenssen, and Torbjørn Eltoft. “Spectral clustering of polarimetric SAR data with Wishart-derived distance measures”. In: *Proceedings of the PoLInSAR Workshop 2007*. Vol. 7. 2007, pp. 1–9.

[172] Mohammed Dabboor, John Yackel, Mosharraf Hossain, and Alexander Braun. “Comparing matrix distance measures for unsupervised PolSAR data classification of sea ice based on agglomerative clustering”. In: *International Journal of Remote Sensing* 34.4 (2013), pp. 1492–1505.

[173] Wen Xie, Licheng Jiao, Biao Hou, Wenping Ma, Jin Zhao, Shuyin Zhang, and Fang Liu. "PolSAR image classification via Wishart-AE model or Wishart-CAE model". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.8 (2017), pp. 3604–3615.

[174] Ronny Hänsch and Olaf Hellwich. "Skipping the real world: Classification of PolSAR images without explicit feature extraction". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 140 (2018), pp. 122–132.

[175] Jong-Sen Lee, Mitchell R Grunes, and Ronald Kwok. "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution". In: *International Journal of Remote Sensing* 15.11 (1994), pp. 2299–2311.

[176] Bo Ren, Biao Hou, Jin Zhao, and Licheng Jiao. "Sparse subspace clustering-based feature extraction for PolSAR imagery classification". In: *Remote Sensing* 10.3 (2018), p. 391.

[177] Xiaowen Dong, Pascal Frossard, Pierre Vandergheynst, and Nikolai Nefedov. "Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds". In: *IEEE Transactions on Signal Processing* 62.4 (2014), pp. 905–918.

[178] Yu Song and Yiquan Wu. "Laplacian regularized low rank subspace clustering". In: *arXiv preprint arXiv:1610.07488* (2016).

[179] Xiayuan Huang, Xiangli Nie, and Hong Qiao. "PolSAR image feature extraction based on co-regularization". In: *2020 IEEE International Geoscience and Remote Sensing Symposium*. 2020, pp. 2727–2730.

[180] Miguel Arevallilo Herráez, David R. Burton, Michael J. Lalor, and Munther A. Gdeisat. "Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path". In: *Applied Optics* 41.35 (2002), pp. 7437–7444.

[181] N. R. Goodman. "Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction)". In: *The Annals of Mathematical Statistics* 34.1 (1963), pp. 152–177.

[182] Muni S. Srivastava. "On the complex Wishart Dstribution". In: *The Annals of Mathematical Statistics* 36.1 (1965), pp. 313–315.

[183] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. "Scikit-learn: Machine learning in Python". In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830.

[184] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger. "UMAP: Uniform Manifold Approximation and Projection". In: *The Journal of Open Source Software* 3.29 (2018), p. 861.

[185] Jarkko Venna and Samuel Kaski. "Neighborhood preservation in nonlinear projection methods: An experimental study". In: *Artificial Neural Networks — ICANN 2001*. 2001, pp. 485–491.

[186] Mike Sips, Boris Neubert, John P. Lewis, and Pat Hanrahan. "Selecting good views of high-dimensional data using class consistency". In: *Computer Graphics Forum* 28.3 (2009), pp. 831–838.

[187] OpenStreetMap contributors. *Planet dump retrieved from https://planet.osm.org*. https://www.openstreetmap.org. 2017.

[188] Ronny Hänsch and Olaf Hellwich. "A comparative evaluation of polarimetric distance measures within the Random Forest Framework for the classification of PolSAR images". In: *2018 IEEE International Geoscience and Remote Sensing Symposium*. 2018, pp. 8440–8443.

[189] Maxim Neumann, Andreas Reigber, and Laurent Ferro-Famil. "Data classification based on PolInSAR coherence shapes". In: *2005 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 7. Seoul, Korea, 2005, pp. 4852–4855.

[190] Sylvia Hochstuhl, Niklas Pfeffer, Antje Thiele, Horst Hammer, and Stefan Hinz. "Your input matters—Comparing real-valued PolSAR data representations for CNN-based segmentation". In: *Remote Sensing* 15.24 (2023), p. 5738.

[191] Jong See. Lee, M.R. Grunes, Thomas L. Ainsworth, Irena Hajnsek, Tobias Mette, and Konstantinos P. Papathanassiou. "Forest classification based on L-band polarimetric and interferometric SAR data". In: *Proceedings of PolInSAR Workshop 2005*. 2005, p. 6.

[192] Laurent Ferro-Famil, Florian Kugler, Eric Potier, and Jong-Sen Lee. "Forest mapping and classification at L-band using Pol-InSAR optimal coherence set statistics". In: *European Conference on Synthetic Aperture Radar)*. Dresden, Germany, 2006, pp. 1–4.

[193] Marc Jager, Maxim Neumann, StÉphane Guillaso, and Andreas Reigber. "A self-initializing PolInSAR classifier using interferometric phase differences". In: *IEEE Transactions on Geoscience and Remote Sensing* 45.11 (2007), pp. 3503–3518.

[194] Huiran Jin, Giorgos Mountrakis, and Stephen V. Stehman. "Assessing integration of intensity, polarimetric scattering, interferometric coherence and spatial texture metrics in PALSAR-derived land cover classification". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 98 (2014), pp. 70–84.

[195] Zhimian Zhang, Haipeng Wang, Feng Xu, and Ya-Qiu Jin. "Complex-valued Convolutional Neural Network and its application in polarimetric SAR image classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.12 (2017), pp. 7177–7188.

[196] Yasser Maghsoudi, Michael Collins, and Donald G. Leckie. "Polarimetric classification of boreal forest using nonparametric feature selection and multiple classifiers". In: *International Journal of Applied Earth Observation and Geoinformation* 19 (2012), pp. 139–150.

[197] Qiang Yin, Wen Hong, Fan Zhang, and Eric Pottier. "Optimal combination of polarimetric features for vegetation classification in PolSAR image". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.10 (2019), pp. 3919–3931.

[198] Xiayuan Huang and Xiangli Nie. "Multi-view feature selection for PolSAR image classification via l2,1 sparsity regularization and manifold regularization". In: *IEEE Transactions on Image Processing* 30 (2021), pp. 8607–8618.

[199] Cedric Lardeux, Pierre-Louis Frison, Céline Tison, Jean-Claude Souyris, Benoit Stoll, Benedicte Fruneau, and Jean-Paul Rudant. "Support Vector Machine for multifrequency SAR polarimetric data classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 47.12 (2009), pp. 4143–4152.

[200] Lei Deng, Ya-nan Yan, and Cuizhen Wang. "Improved PolSAR image classification by the use of multi-feature combination". In: *Remote Sensing* 7.4 (2015), pp. 4157–4177.

[201] Yasser Maghsoudi, Michael J. Collins, and Donald G. Leckie. “Radarsat-2 polarimetric SAR data for boreal forest classification using SVM and a wrapper feature selector”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6.3 (2013), pp. 1531–1538.

[202] Yu Bai, Dongqing Peng, Xiangli Yang, Lijun Chen, and Wen Yang. “Supervised feature selection for polarimetric SAR classification”. In: *2014 12th International Conference on Signal Processing*. 2014, pp. 1006–1010.

[203] Isabelle Guyon and André Elisseeff. “An introduction to variable and feature selection”. In: *The Journal of Machine Learning Research* 3 (2003), pp. 1157–1182.

[204] Huan Liu and Rudy Setiono. “Chi2: feature selection and discretization of numeric attributes”. In: *1995 7th IEEE International Conference on Tools with Artificial Intelligence*. 1995, pp. 388–391.

[205] Peter E. Hart, David G. Stork, and Richard O. Duda. *Pattern classification*. Wiley Hoboken, 2000.

[206] Daphne Koller and Mehran Sahami. “Toward optimal feature selection”. In: *13th International Conference on International Conference on Machine Learning*. 1996, pp. 284–292.

[207] Xiaofei He, Deng Cai, and Partha Niyogi. “Laplacian score for feature selection”. In: *Advances in Neural Information Processing Systems*. Vol. 18. 2005.

[208] Kenji Kira and Larry A. Rendell. “A practical approach to feature selection”. In: *Machine Learning Proceedings 1992*. Morgan Kaufmann, 1992, pp. 249–256.

[209] Quanquan Gu, Zhenhui Li, and Jiawei Han. “Generalized Fisher score for feature selection”. In: *arXiv preprint arXiv:1202.3725* (2012).

[210] Hanchuan Peng, Fuhui Long, and C. Ding. “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (2005), pp. 1226–1238.

[211] Lei Yu and Huan Liu. “Feature selection for high-dimensional data: A fast correlation-based filter solution”. In: *20th International Conference on Machine Learning*. 2003, pp. 856–863.

[212] Ron Kohavi and George H. John. “Wrappers for feature subset selection”. In: *Artificial Intelligence* 97.1 (1997), pp. 273–324.

[213] Robert Tibshirani. “Regression shrinkage and selection via the Lasso”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996), pp. 267–288.

[214] Wenjiang Fu and Keith Knight. “Asymptotics for Lasso-type estimators”. In: *The Annals of Statistics* 28.5 (2000), pp. 1356–1378.

[215] Hui Zou and Trevor Hastie. “Regularization and variable selection via the Elastic Net”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67.2 (2005), pp. 301–320.

[216] Robin Genuer, Jean-Michel Poggi, and Christine Tuleau-Malot. “Variable selection using random forests”. In: *Pattern Recognition Letters* 31.14 (2010), pp. 2225–2236.

[217] Leo Breiman. “Random Forests”. In: *Machine Learning* 45.1 (2001), pp. 5–32.

[218] Tin Kam Ho. “The random subspace method for constructing decision forests”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.8 (1998), pp. 832–844.

[219] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[220] Sergey Ioffe and Christian Szegedy. “Batch Normalization: Accelerating deep network training by reducing internal covariate shift”. In: *arXiv:1502.03167 [cs]* (2015).

[221] Lei Wang, Xin Xu, Hao Dong, Rong Gui, and Fangling Pu. “Multi-pixel simultaneous classification of PolSAR image using convolutional neural networks”. In: *Sensors* 18.3 (2018), p. 769.

[222] Dongmei Song, Zongjin Zhen, Bin Wang, Xiaofeng Li, Le Gao, Ning Wang, Tao Xie, and Ting Zhang. “A novel marine oil spillage identification scheme based on Convolution Neural Network feature extraction from fully polarimetric SAR imagery”. In: *IEEE Access* 8 (2020), pp. 59801–59820.

[223] Hongwei Dong, Lamei Zhang, and Bin Zou. “PolSAR image classification with lightweight 3D convolutional networks”. In: *Remote Sensing* 12.3 (2020), p. 396.

[224] Ronghua Shang, Jianghai He, Jiaming Wang, Kaiming Xu, Licheng Jiao, and Rustam Stolkin. “Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification”. In: *Knowledge-Based Systems* 194 (2020), p. 105542.

[225] Hongmiao Wang, Cheng Xing, Junjun Yin, and Jian Yang. “Land cover classification for polarimetric SAR images based on Vision Transformer”. In: *Remote Sensing* 14.18 (2022), p. 4656.

[226] Wen Xie, Ruonan Wang, Xin Yang, and Wenqiang Hua. “Depthwise separable residual network based on UNet for PolSAR images classification”. In: *2022 IEEE International Geoscience and Remote Sensing Symposium*. 2022, pp. 1039–1042.

[227] Rong Yang, Zhentao Hu, Yiming Liu, and Zhen Xu. “A novel polarimetric SAR classification method integrating pixel-based and patch-based classification”. In: *IEEE Geoscience and Remote Sensing Letters* 17.3 (2020), pp. 431–435.

[228] Yixin Zuo, Jiayi Guo, Yueting Zhang, Yuxin Hu, Bin Lei, Xiaolan Qiu, and Chibiao Ding. “Winner takes all: A superpixel aided voting algorithm for training unsupervised PolSAR CNN classifiers”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–19.

[229] Feng Zhao, Min Tian, Wen Xie, and Hanqiang Liu. “A new parallel dual-channel Fully Convolutional Network via semi-supervised FCN for PolSAR image Classification”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), pp. 4493–4505.

[230] Zheng Fang, Gong Zhang, Qijun Dai, Biao Xue, and Peng Wang. “Hybrid attention-based encoder–decoder Fully Convolutional Network for PolSAR image classification”. In: *Remote Sensing* 15.2 (2023), p. 526.

[231] Lamei Zhang, Hongwei Dong, and Bin Zou. “Efficiently utilizing complex-valued PolSAR image data via a multi-task deep learning framework”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 157 (2019), pp. 59–72.

[232] Rui Zhang, Jie Chen, Liang Feng, Shuang Li, Wei Yang, and Ding Guo. “A refined Pyramid Scene Parsing network for polarimetric SAR image semantic segmentation in agricultural areas”. In: *IEEE Geoscience and Remote Sensing Letters* 19 (2022), pp. 1–5.

[233] Jie Geng, Ru Wang, and Wen Jiang. “Polarimetric SAR image classification based on feature enhanced superpixel hypergraph neural network”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–12.

[234] Haixia Bi, Jian Sun, and Zongben Xu. “A graph-based semisupervised deep learning model for PolSAR image classification”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.4 (2019), pp. 2116–2132.

[235] Haixia Bi, Feng Xu, Zhiqiang Wei, Yong Xue, and Zongben Xu. “An active deep learning approach for minimally supervised PolSAR image classification”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.11 (2019), pp. 9378–9395.

[236] Jiaqiu Ai, Feifan Wang, Yuxiang Mao, Qiwu Luo, Baidong Yao, He Yan, Mengdao Xing, and Yanlan Wu. “A fine PolSAR terrain classification algorithm using the texture feature fusion-based improved Convolutional Autoencoder”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–14.

[237] Wenqiang Hua, Cong Zhang, Wen Xie, and Xiaomin Jin. “Polarimetric SAR image classification based on ensemble dual-branch CNN and superpixel algorithm”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022), pp. 2759–2772.

[238] Lingjuan Yu, Zhaoxin Zeng, Ao Liu, Xiaochun Xie, Haipeng Wang, Feng Xu, and Wen Hong. “A lightweight complex-valued DeepLabv3+ for semantic segmentation of PolSAR image”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022), pp. 930–943.

[239] Jin Zhang, Hao Feng, Qingli Luo, Yu Li, Jujie Wei, and Jian Li. “Oil spill detection in quad-polarimetric SAR images using an advanced Convolutional Neural Network based on superpixel model”. In: *Remote Sensing* 12.6 (2020), p. 944.

[240] Yangyang Li, Yanqiao Chen, Guangyuan Liu, and Licheng Jiao. “A novel deep Fully Convolutional Network for PolSAR image classification”. In: *Remote Sensing* 10.12 (2018), p. 1984.

[241] Mete Ahishali, Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. “Multifrequency PolSAR image classification using dual-band 1D convolutional neural networks”. In: *2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium*. 2020, pp. 73–76.

[242] Xianyuan Wang, Zongjie Cao, Zongyong Cui, Nengyuan Liu, and Yiming Pi. “PolSAR image classification based on deep polarimetric feature and contextual information”. In: *Journal of Applied Remote Sensing* 13.3 (2019), p. 1.

[243] Rui Qin, Xiongjun Fu, and Ping Lang. “PolSAR image classification based on low-frequency and contour subbands-driven polarimetric SENet”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), pp. 4760–4773.

[244] Huiming Xie, Shuang Wang, Kun Liu, Shaopeng Lin, and Biao Hou. “Multilayer feature learning for polarimetric synthetic radar data classification”. In: *2014 IEEE Geoscience and Remote Sensing Symposium*. 2014, pp. 2818–2821.

[245] Biao Hou, Hongda Kou, and Licheng Jiao. “Classification of polarimetric SAR images using multilayer autoencoders and superpixels”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.7 (2016), pp. 3072–3081.

[246] Paul Horst. *Generalized Canonical Correlations and their application to experimental data*. 14. Journal of Clinical Psychology, 1961.

[247] Maryam Salehi, Mahmod Reza Sahebi, and Yasser Maghsoudi. “Improving the accuracy of urban land cover classification using Radarsat-2 PolSAR data”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.4 (2014), pp. 1394–1401.

[248] Siddharth Hariharan, Dipankar Mandal, Siddhesh Tirodkar, Vineet Kumar, Avik Bhattacharya, and Juan Manuel Lopez-Sanchez. “A novel phenology based feature subset selection technique using Random Forest for multitemporal PolSAR crop classification”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.11 (2018), pp. 4244–4258.

[249] Zhixin Qi, Anthony Gar-On Yeh, Xia Li, and Zheng Lin. “A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data”. In: *Remote Sensing of Environment* 118 (2012), pp. 21–39.

[250] Sadhana Yadav, Hitendra Padalia, Sanjiv K. Sinha, Ritika Srinet, and Prakash Chauhan. “Above-ground biomass estimation of Indian tropical forests using X band Pol-InSAR and Random Forest”. In: *Remote Sensing Applications: Society and Environment* 21 (2021), p. 100462.

[251] Mingkui Tan, Li Wang, and Ivor W. Tsang. “Learning sparse SVM for feature selection on very high dimensional datasets”. In: *27th International Conference on International Conference on Machine Learning*. 2010, pp. 1047–1054.

[252] Chen Yang, Biao Hou, Bo Ren, Yue Hu, and Licheng Jiao. “CNN-based polarimetric decomposition feature selection for PolSAR Image classification”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.11 (2019), pp. 8796–8812.

[253] Ronghua Shang, Keyao Zhu, Jie Feng, Chao Wang, Licheng Jiao, and Songhua Xu. “Reg-superpixel guided convolutional neural network of PolSAR image classification based on feature selection and receptive field reconstruction”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2023), pp. 1–16.

[254] Igor Kononenko. “Estimating attributes: Analysis and extensions of RELIEF”. In: *European Conference on Machine Learning*. 1994, pp. 171–182.

[255] Jie Feng, Jiantong Chen, Qigong Sun, Ronghua Shang, Xianghai Cao, Xiangrong Zhang, and Licheng Jiao. “Convolutional neural network based on bandwise-independent convolution and hard thresholding for hyperspectral band selection”. In: *IEEE Transactions on Cybernetics* 51.9 (2021), pp. 4414–4428.

[256] Deshuai Yuan, Jinbao Jiang, Ziqiang Gong, Cheng Nie, and Yilin Sun. “Moldy peanuts identification based on hyperspectral images and point-centered convolutional neural network combined with embedded feature selection”. In: *Computers and Electronics in Agriculture* 197 (2022), p. 106963.

[257] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 770–778.

[258] Zhongao Yang, Qingyuan Zhang, Weidong Chen, and Chang Chen. “PolSAR image classification based on Resblock combined with attention model”. In: *2021 IEEE 6th International Conference on Signal and Image Processing*. 2021, pp. 340–344.

[259] Lei Ding, Kai Zheng, Dong Lin, Yuxing Chen, Bing Liu, Jiansheng Li, and Lorenzo Bruzzone. “MP-ResNet: Multipath residual network for the semantic segmentation of high-resolution PolSAR images”. In: *IEEE Geoscience and Remote Sensing Letters* 19 (2022), pp. 1–5.

[260] Amir Hossein Ghazvinizadeh, Maryam Imani, and Hassan Ghassemian. “Residual network based on entropy-anisotropy-alpha target decomposition for polarimetric SAR image classification”. In: *Earth Science Informatics* 16.1 (2023), pp. 357–366.

[261] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. “Pyramid Scene Parsing Network”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 2881–2890.

[262] Jie Hu, Li Shen, and Gang Sun. “Squeeze-and-Excitation networks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 7132–7141.

[263] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. “CBAM: Convolutional block attention module”. In: *Proceedings of the European Conference on Computer Vision*. 2018, pp. 3–19.

[264] Hongwei Dong, Lamei Zhang, Da Lu, and Bin Zou. “Attention-based polarimetric feature selection convolutional network for PolSAR image classification”. In: *IEEE Geoscience and Remote Sensing Letters* 19 (2022), pp. 1–5.

[265] Jonathan Long, Evan Shelhamer, and Trevor Darrell. “Fully convolutional networks for semantic segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 3431–3440.

[266] Yuanyuan Wang, Chao Wang, and Hong Zhang. “Integrating H-A-alpha with Fully Convolutional Networks for fully PolSAR classification”. In: *2017 International Workshop on Remote Sensing with Intelligent Processing*. 2017, pp. 1–4.

[267] Fariba Mohammadimanesh, Bahram Salehi, Masoud Mahdianpari, Eric Gill, and Matthieu Molinier. “A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem”. en. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 151 (2019), pp. 223–236.

[268] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. “SegNet: A deep convolutional encoder-decoder architecture for image segmentation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.12 (2017), pp. 2481–2495.

[269] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional networks for biomedical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015, pp. 234–241.

[270] Ao Liu, Lingjuan Yu, Zhaoxin Zeng, Xiaochun Xie, Yuting Guo, and Qiqi Shao. “Complex-valued U-Net for PolSAR image semantic segmentation”. In: *Journal of Physics: Conference Series* 2010.1 (2021), p. 012102.

[271] Lingjuan Yu, Qiqi Shao, Yuting Guo, Xiaochun Xie, Miaomiao Liang, and Wen Hong. “Complex-valued U-Net with capsule embedded for semantic segmentation of PolSAR image”. In: *Remote Sensing* 15.5 (2023), p. 1371.

[272] Lamei Zhang, Zexi Chen, Bin Zou, and Ye Gao. “Polarimetric SAR terrain classification using 3D Convolutional Neural Network”. In: *2018 IEEE International Geoscience and Remote Sensing Symposium*. 2018, pp. 4551–4554.

[273] Mingxing Tan. “Efficientnet: Rethinking model scaling for convolutional neural networks”. In: *arXiv preprint arXiv:1905.11946* (2019).

[274] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. “Mobilenetv2: Inverted residuals and linear bottlenecks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 4510–4520.

[275] Min Lin, Qiang Chen, and Shuicheng Yan. *Network In Network*. 2014. arXiv: 1312.4400 [cs.NE].

[276] Nabila Abraham and Naimul Mefraz Khan. “A novel focal Tversky loss function with improved Attention U-Net for lesion segmentation”. In: *2019 IEEE 16th International Symposium on Biomedical Imaging*. 2019, pp. 683–687.

[277] Shruti Jadon. “A survey of loss functions for semantic segmentation”. In: *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*. 2020, pp. 1–7.

[278] Ilya Loshchilov and Frank Hutter. *SGDR: Stochastic Gradient Descent with warm restarts*. 2017. arXiv: 1608.03983 [cs.LG].

[279] Jeremy Howard. *Ross Wightman’s timm image model library*. 2022. URL: https://github.com/fastai/timmdocs (visited on 08/23/2024).

[280] Shuaiying Zhang, Lizhen Cui, Yue Zhang, Tian Xia, Zhen Dong, and Wentao An. “Research on input schemes for polarimetric SAR classification using deep learning”. In: *Remote Sensing* 16.11 (2024), p. 1826.

[281] Bo Wang, Wei Wei, Shuang Qiu, Shengpei Wang, Dan Li, and Huiguang He. “Boundary aware U-Net for retinal layers segmentation in optical coherence tomography images”. In: *IEEE Journal of Biomedical and Health Informatics* 25.8 (2021), pp. 3029–3040.

[282] Hoin Jung, Han-Soo Choi, and Myungjoo Kang. “Boundary enhancement semantic segmentation for building extraction from remote sensed image”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–12.

[283] Eric Pottier, Laurent Ferro-Famil, Magdalena Fitrzyk, and Yves-Louis Desnos. “PolSARpro-BIO: The new Scientific Toolbox for ESA & third party fully Polarimetric SAR Missions”. In: *EUSAR 2018; 12th European Conference on Synthetic Aperture Radar*. 2018, pp. 1–4.

[284] Eric Pottier. *ESA PolSARpro v6.0 (Biomass Edition)*. Version 6.0.3. URL: https://ietr-lab.univ-rennes1.fr/polsarpro-bio/.

[285] Adam Paszke et al. “PyTorch: An imperative style, high-performance deep learning library”. In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019.

[286] William Falcon and The PyTorch Lightning team. *PyTorch Lightning*. Version 1.9. 2019. URL: https://github.com/Lightning-AI/lightning.

[287] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61.

[288] The pandas development team. *pandas-dev/pandas: Pandas*. Version 1.5.1. 2020. URL: https://doi.org/10.5281/zenodo.3509134.

[289] Michael L. Waskom. “seaborn: statistical data visualization”. In: *Journal of Open Source Software* 6.60 (2021), p. 3021.

[290] DeepL SE. *DeepL*. URL: https://www.deepl.com/de/translator.

[291] Grammarly. *Grammarly*. Version v.1.2.101.1482. URL: https://app.grammarly.com/.

[292] Ingo Weber, Hendrik Linka, Daniel Mertens, Tamara Muryshkin, Heinrich Opgenoorth, and Stefan Langer. "FhGenie: A custom, confidentiality-preserving Chat AI for corporate and scientific use". In: *2024 IEEE 21st International Conference on Software Architecture Companion*. 2024, pp. 26–31.

List of Figures

2.1 SLAR Geometry . . . 9
2.2 SAR Geometry . . . 10
2.3 SAR Focussing . . . 11
2.4 Polarization Ellipse . . . 13
2.5 Polarization States . . . 14
2.6 Canonical Bodies . . . 16
2.7 Refined Lee Filter Windows . . . 21
2.8 Across-track Interferometry . . . 30

3.1 Pol-InSAR-Island - Geographic Location . . . 38
3.2 Pol-InSAR-Island Dataset . . . 39
3.3 Biotope Type Map . . . 40
3.4 Data Labeling Workflow . . . 41
3.5 UMAP Projection - Six Classes . . . 43
3.6 UMAP Projection - Salt Marshes, Dunes Landscape . . . 44
3.7 UMAP Projection - 12 Classes . . . 45
3.8 Target Class Map . . . 46
3.9 Pol-InSAR-Island Class Map . . . 47

4.1 UMAP Workflow . . . 61
4.2 Projection Errors . . . 72
4.3 Distance Consisteny . . . 74
4.4 Pol-InSAR-Island Feature Embeddings . . . 76
4.5 Pol-InSAR-Island Wishart Embeddings . . . 77
4.6 Results for Trustworthiness and Continuity . . . 78
4.7 Results for KNN Classification and Distance Consistency . . . 79
4.8 UMAP RGB Color Composites . . . 82
4.9 Comparing UMAP to Pauli Visualization . . . 83
4.10 Influence of the Incidence Angle on UMAP Embeddings . . . 84
4.11 Topographic Features in UMAP Color Composites . . . 86
4.12 Analyzing Data Structure of PolSAR and PolInSAR Data . . . 87
4.13 UMAP Projection Transferred to Langeoog . . . 89
4.14 UMAP Color Composite of Tidal Flats . . . 90

5.1 Feature Selection Schemes . . . 100
5.2 CNNs in the Context of Machine Learning . . . 103

5.3 Convolutional Layer . . . 104
5.4 Fully Connected Layers . . . 105
5.5 PolInSAR Image Classification Workflow . . . 117
5.6 PolInSAR Coherency Matrix . . . 120
5.7 Vanilla U-Net Architecture . . . 124
5.8 ResNet U-Net Architecture . . . 126
5.9 EfficientNet U-Net Architecture . . . 127
5.10 3D Convolution . . . 128
5.11 Feature Attention Module . . . 129
5.12 F-UMAP Scatter Plot for Multi-frequency PolInSAR Data . . . 137
5.13 Confusion Matrix for Wishart Classifier . . . 138
5.14 Confusion Matrix for Random Forest . . . 138
5.15 Land Cover Classification Maps (Wishart and Random Forest) . . . 139
5.16 Vanilla U-Net Classification Results - PolSAR . . . 141
5.17 ResNet U-Net Classification Results - PolSAR . . . 142
5.18 EfficientNet U-Net Classification Results - PolSAR . . . 143
5.19 Aggregated U-Net Classification Results - PolSAR . . . 144
5.20 Land Cover Classification Maps (Random Forest and CNN) - PolSAR . . . 148
5.21 Zoom in Land Cover Classification Maps (Random Forest and CNN) - PolSAR 149
5.22 Land Cover Evaluation Maps (Random Forest and CNN) - PolSAR . . . 150
5.23 Multi-frequency PolInSAR CNN Classification Result - Feature Stacking . . . 151
5.24 PolSAR vs. Multi-frequency PolInSAR CNN Classification Results . . . 152
5.25 PolSAR vs. Multi-frequency PolInSAR CNN Classification Land Cover Maps . 153
5.26 Classification Improvements Using Interferometry and Multi-frequency . . . 154
5.27 Confusion Matrix for Random Forest and CNN Classification - PolInSAR Data 156
5.28 Land Cover Classification Maps (Random Forest and CNN) . . . 157
5.29 Land Cover Classification Maps (Random Forest and CNN) - ROI I . . . 158
5.30 Land Cover Classification Maps (Random Forest and CNN) - - ROI II . . . 159
5.31 Cross-channel Feature Extraction Results . . . 161
5.32 Vanilla U-Net Classification Result for Multi-frequency PolInSAR Data- Dimension Reduction . . . 165
5.33 ResNet U-Net Classification Result for Multi-frequency PolInSAR Data- Dimension Reduction . . . 167
5.34 EfficientNet U-Net Classification Result for Multi-frequency PolInSAR Data- Dimension Reduction . . . 168

A.1 Analysis of Learning Rate Scheduler . . . 215

B.1 CNN Classification Results Based on the Qin Representation . . . 217

List of Tables

2.1 Frequency Bands of Radar Remote Sensing 8

3.1 Polarimetric and Interferometric Features for Visualization 42
3.2 Class Distribution of Reference Data. 48

4.1 Dimension Reduction Methods 55
4.2 List of Polarimetric and Interferometric Features 65

5.1 Representations of PolSAR Data Used as CNN Input 108
5.2 Wishart and Random Forest Classification Results - PolSAR 135
5.3 Wishart and Random Forest Classification Results - PolInSAR 135
5.4 Random Forest vs. CNN Classification Results - PolSAR 147
5.5 Class-wise IoU for Random Forest and CNN Classification - PolInSAR Data . 156
5.6 Selected Features - MI, mRMR, RF 162
5.7 Selected Features - RFW 164
5.8 Runtime Analysis 170

List of Abbreviations

ALS Airborne Laser Scanning

CAE Convolutional Autoencoder

CBAM Convolutional Block Attention Module

CCA Curvilinear Component Analysis

CNN Convolutional Neural Network

DLR German Aerospace Center

DSM Digital Surface Model

EP exhaustive PolSAR

EPI exhaustive PolInSAR

F-UMAP Feature-UMAP

FA Factor analysis

FCN Fully Convolutional Network

FTL Focal Tversky Loss

GCCA Generalized Canonical Correlation Analysis

GDA Generalized Discriminant Analysis

GPLVM Gaussian Process Latent Variable Model

ICA Independent Component Analysis

IDMAP Interactive Document Maps

InSAR Interferometric SAR

IoU Intersection-over-Union

KNN *K*-Nearest-Neighbor

KPCA Kernel PCA

LDA Linear Discriminant Analysis

LDE Local Discriminant Embedding

LE Lapalcian Eigenmaps

LLE Locally Linear Embedding

LLTSA Linear Local Tangent Space Alignment

LMNN Large-Margin Nearest Neighbor

LPP Locality Preserving Projection

LSP Least Square Projection

LTSA Local Tangent Space Alignment

MBConv Mobile Inverted Bottleneck

MCML Maximally Collapsing Metric Learning

MDS Multi Dimensional Scaling

MI Mutual Information

mIoU mean Intersection-over-Union

mRMR minimum Redundancy Maximum Relevance

NISAR NASA-ISRO SAR

NLPV Lower Saxon Wadden Sea National Park Authority

NLWKN Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency

NMF Nonnegative Matrix Factorization

PBC Projection by Clustering

PCA Principal Component Analysis

PolInSAR Polarimetric Interferometric SAR

PolSAR Polarimetric Synthetic Aperture Radar

PolSAR Polarimetric SAR

ReLU Rectified Linear Unit

ResNet Residual Network

RF Random Forest

RF-FI Random Forest Feature Importance

RFW Regularized Feature Weighting

ROI region of interest

Rose-L Radar Observatory System for Europe in L-band

RP Random Projection

SAR Synthetic Aperture Radar

SE Squeeze and Excitation

SERD Single-bounce Eigenvalue Relative Difference

SGD Stochastic Gradient Descent

SLAR Side-Looking Airborne Radar

SOM Self-Organizing-Maps

SRWD Symmetric Revised Wishart Distance

SVM Support Vector Machine

t-SNE t-distributed Stochastic Neighbor Embedding

TDF target decomposition features

TMAP Trilateral Monitoring and Assessment Program

UMAP Uniform Manifold Approximation and Projection

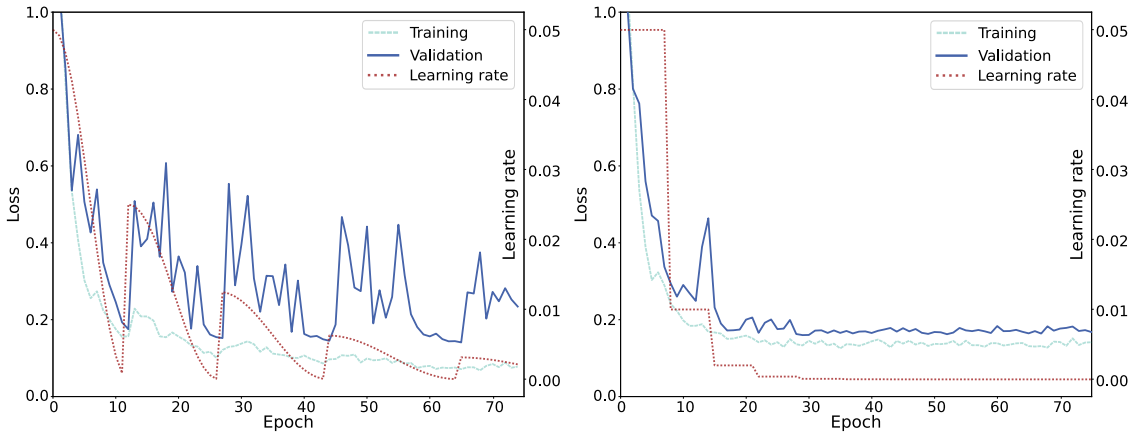
VAE Variational Autoencoder

W-UMAP Wishart-UMAP

A

Analysis of Learning Rate Schedule

When using CNNs, which include an additional layer between the input layer and the first convolution layer as part of the RFW feature selection method, the proposed optimization method (SGD optimization with adaptive learning rate according to a cosine learning rate scheduler with warm restarts) cannot achieve stable training. As shown in Figure A.1a, the warm restarts each cause a strong increase in the validation loss. A more stable training is achieved if the learning rate is gradually reduced when the validation loss stagnates over several iterations. This is shown in Figure A.1b.



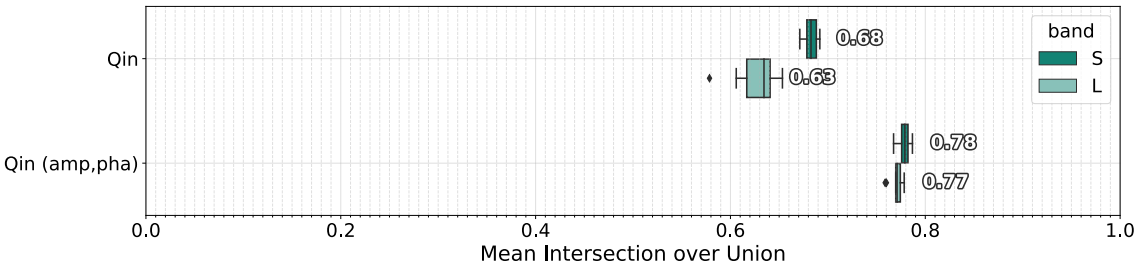
(a) Cosine learning rate scheduler.

(b) Reduce learning rate on plateau.

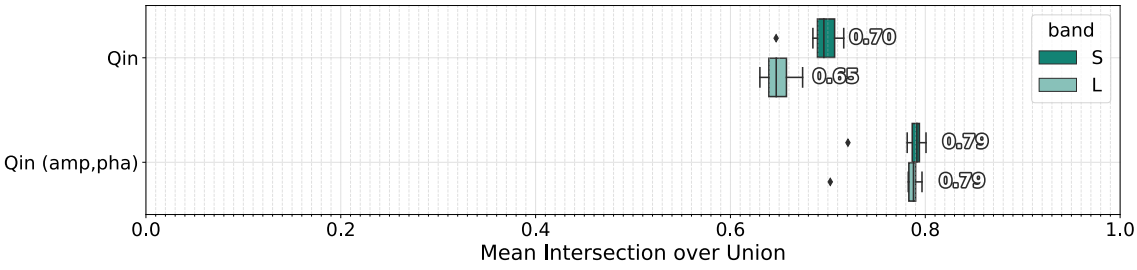
Figure A.1.: CNN training process using different learning rate adaption schemes. The plots show the training and validation loss as well as the learning rate over training epochs.

B Analysis of CNN Input Data Representations

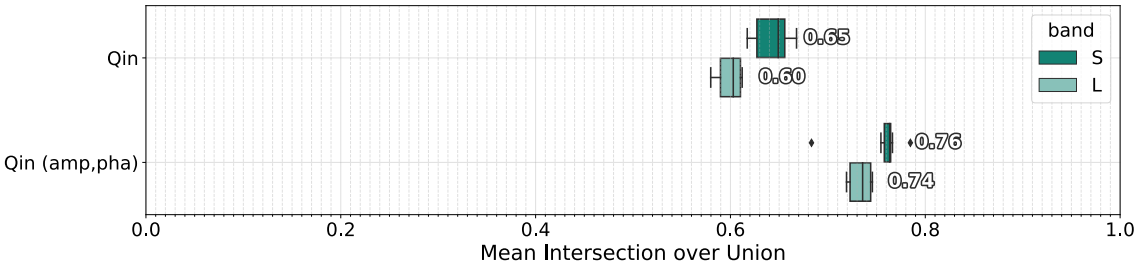
To test the hypothesis that the use of real and imaginary parts of the off-diagonal elements of the coherency matrix negatively influences the CNN classification results even in combination with other polarimetric features, the following experiment was performed. The original input representation Qin proposed in [243] was tested against an adapted representation in which the real and imaginary parts of the off-diagonal elements of the coherency matrix were replaced by their amplitudes and phases. The results in Figure B.1 show that this adaptation improves the classification results for all test configurations (defined by CNN architecture and frequency band) and thus verifies the hypothesis.



(a) Vanilla U-Net.



(b) ResNet U-Net



(c) EfficientNet U-Net

Figure B.1.: CNN Classification results using the original representation "Qin" and the modified representation "Qin (amp, pha)". The boxes show the test statistics of mIoU based on 10 training and test runs using S-band or L-band data.

Colophon

This thesis was typeset with LaTeX 2$_\varepsilon$. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at http://cleanthesis.der-ric.de/.