

URBAN MATERIAL CLASSIFICATION USING SPECTRAL AND TEXTURAL FEATURES RETRIEVED FROM AUTOENCODERS

R. Ilehag*, J. Leitloff, M. Weinmann, A. Schenk

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Germany -
(rebecca.ilehag, jens.leitloff, martin.weinmann, andreas.schenk)@kit.edu

Commission I, WG I/1

KEY WORDS: Material classification, Spectral features, Textural features, Autoencoder, Compressed representation, Spectral library

ABSTRACT:

Classification of urban materials using remote sensing data, in particular hyperspectral data, is common practice. Spectral libraries can be utilized to train a classifier since they provide spectral features about selected urban materials. However, urban materials can have similar spectral characteristic features due to high inter-class correlation which can lead to misclassification. Spectral libraries rarely provide imagery of their samples, which disables the possibility of classifying urban materials with additional textural information. Thus, this paper conducts material classification comparing the benefits of using close-range acquired spectral and textural features. The spectral features consist of either the original spectra, a PCA-based encoding or the compressed spectral representation of the original spectra retrieved using a deep autoencoder. The textural features are generated using a deep denoising convolutional autoencoder. The spectral and textural features are gathered from the recently published spectral library KLUM. Three classifiers are used, the two well-established Random Forest and Support Vector Machine classifiers in addition to a Histogram-based Gradient Boosting Classification Tree. The achieved overall accuracy was within the range of 70 - 80% with a standard deviation between 2 - 10% across all classification approaches. This indicates that the amount of samples still is insufficient for some of the material classes for this classification task. Nonetheless, the classification results indicate that the spectral features are more important for assigning material labels than the textural features.

1. INTRODUCTION

Assessing the materials in the urban environment has in recent times increased in importance for several reasons and applications. For example, this information is useful for researchers and city planners who deal with city simulations or models where knowledge about the existing materials is important. This can include three-dimensional building models as given in CityGML (Kolbe et al., 2005) and the extended dataset called Buildings from OpenStreetMap (OpenStreetMap contributors, 2017). Furthermore, knowledge about the material of a cultural heritage object contributes to preservation and conservation of either buildings (Sánchez and Quirós, 2017; Yuan et al., 2020) or statues (Grilli and Remondino, 2019). Further knowledge about urban materials can be an indicator on how to tackle and handle the urban heat island effect which has an increasing occurrence in cities (Ward et al., 2016; Santamouris et al., 2011).

Observed in previous studies regarding urban material classification, the characteristic spectral features of different urban material classes can be similar which makes it challenging to distinguish them from each other (Ilehag et al., 2017b; Ouerghemmi et al., 2017; Deshpande et al., 2019). This is partly due to the high inter-class correlation which leads to misclassification. Furthermore, when different spectral libraries only provide the spectra and no imagery of their samples, it can be challenging to assure that the different spectral libraries contain materials that have been labeled in a consistent manner (Fairbairn Jr, 2013; Ilehag et al., 2019). Likewise, this can again lead to misclassification as the different materials might have the same labels. Few spectral libraries contain imagery of the samples (Kotthaus et al., 2014; Kokaly et al., 2017; Ilehag et al., 2019) which can provide additional information for material classification.

Combining spectral and textural features from satellite images for classification of e.g. tree species (Ferreira et al., 2019), agriculture (Mirzapour and Ghassemian, 2015; Ding et al., 2019) and urban areas (Yuan et al., 2013) has shown great potential. Thus, as urban materials are challenging to separate on a spectral level due to the similarity of the spectral characteristic features, added textural information could be beneficial. Assessment of materials for close-ranged approaches, such as scene analysis via unmanned aerial vehicles (UAV), could likewise benefit from combining spectral and textural features (Ilehag et al., 2017a). Imagery of material samples acquired from a close distance, where distinct textural features of the material can be detected, could contribute to improved material distinction. One way to extract textural features can be through the usage of an autoencoder (AE) (Kramer, 1991). An AE is an unsupervised machine learning method for compressing and decompressing data to retrieve important dimensionality-reduced features, which can include textural features (Das and Walia, 2019). The retrieved features can be used as input for classification tasks and have shown great potential (Geng et al., 2015; Li et al., 2016).

In this paper, we aim to classify urban materials based on spectral and textural features using the close-range acquired spectral library KLUM (Ilehag et al., 2019). In particular, we intend to determine the benefits of using either spectral or textural features, or the combination of both. The spectral features are either expressed as the original spectra, the encodings based on the first few principal components or the compressed spectral features acquired from a deep AE (DAE) (Vincent et al., 2010). The textural features are retrieved from a deep denoising convolutional AE (CAE) (Masci et al., 2011).

This paper is structured as follows. We briefly summarize the related literature about AEs and the approaches for urban material classification in Section 2. The used dataset, the spectral

*Corresponding author

library KLUM, is presented in Section 3. The proposed methodology is presented in Section 4, which consists of creating AEs for extraction of the compressed representation and classification. The classification results and further analyzes of the usefulness of AEs are presented in Section 5. Lastly, final remarks and suggestions for future work are provided in Section 6.

2. RELATED WORK

To provide an overview of related work, the architecture of an AE and its applications are presented in Subsection 2.1. Spectral and textural approaches to urban material classification are presented in Subsection 2.2.

2.1 Autoencoders

An AE is a data compression method that can be used to learn and extract important features in an unsupervised manner (Kramer, 1991). An AE consists of a compression function, an encoder, that reduces the dimensionality of the dataset and of a decompression function, a decoder, that restores the compressed dataset into its original form. The architecture of a typical AE can be seen in Figure 1. As visualized, the input layer is compressed using a number of hidden layers which the user has chosen for the particular task in mind. The outcome of the encoder is the compressed representation, which is a dimensionality-reduced representation of the original input. To restore the original input into the reconstructed input, that is, the output layer, the compressed representation is decompressed using the decoder.

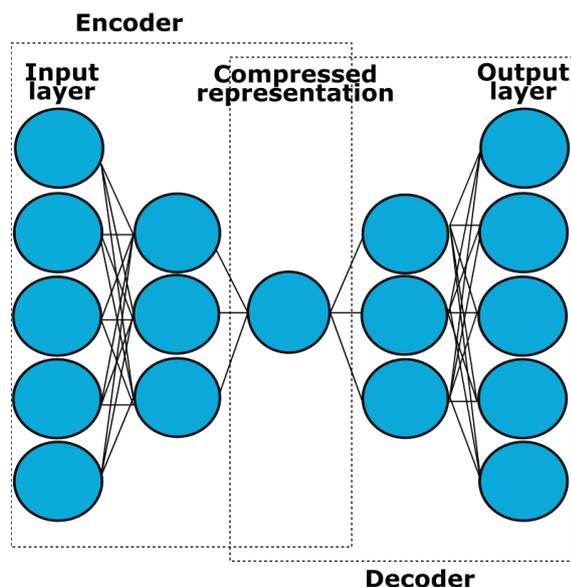


Figure 1. The general architecture of an AE.

An AE is built for specific sets of data, that is, it is only able to compress and decompress datasets comparable to what it was trained for. For example, if being trained on images of animals, it would perform poorly on images of cars. An AE requires both training and testing data to assess the quality of the encoder and the decoder, and can perform well on smaller datasets (Feng et al., 2019).

Depending on the dataset and the purpose in mind, different autoencoding architecture are available. We will here present those relevant for this paper. A DAE (Vincent et al., 2010) is an AE

that contains several hidden layers connected with each other. It means that the output from one hidden layer is the input of the next consecutive hidden layer. A CAE (Masci et al., 2011) utilizes convolutional layers to extract features, which is most commonly applied to imagery as it discovers localized repeated features over the input domain.

AEs can be used for different kinds of tasks, one being denoising of corrupted data, such as signals (Vincent et al., 2008; Vincent et al., 2010). A denoising AE is trained on both the corrupted and uncorrupted data. The uncorrupted data receives intentionally randomly added noise to corrupt the dataset. By adding noise to the uncorrupted dataset, it forces the AE to extract the most important features (Vincent et al., 2010) which makes it more robust. The compressed representation can be used as an input for classification tasks, such as land-cover classification (Li et al., 2016) or classification of a SAR image (Geng et al., 2015). As the compressed representation is a dimensionality-reduced representation, it can be a useful approach for high-dimensional data e.g. given when dealing with hyperspectral data (Zhou et al., 2019; Lan et al., 2019; Lin et al., 2013) or for extracting important features from images (Das and Walia, 2019).

2.2 Urban material datasets and classification approaches

The utilization of spectral libraries is a common procedure while dealing with material classification based on purely spectral features. They contain a large variation of material samples which can be used for training a classifier due to their unique spectral characteristic features. As libraries are compiled in different countries across the world, the provided samples can vary due to the regional differences. Spectral libraries are therefore generated to represent a particular region. Publicly available spectral libraries such as the Santa Barbara spectral library (Herold et al., 2004), ASTER (Baldrige et al., 2009), SLUM (Kotthaus et al., 2014), USGS spectral library version 7 (Kokaly et al., 2017) and KLUM (Ilehag et al., 2019) enable the possibility to extract spectral features for further studies. Since spectral libraries aim to provide only spectra of a variation of materials, imagery corresponding to the given samples is usually not provided.

In contrast to such spectral libraries, other material libraries have been presented where the focus is set on material classification based on textural properties. Respective material libraries include the CURET Database (Dana et al., 1999) and the KTH-TIPS Database (Hayman et al., 2004) that contain RGB images of a single representative per material class captured under different conditions (e.g. with respect to scale, illumination and view-point) in a controlled setting. Since these databases only contain a single material sample per class and thus do not take into account intra-class variations, an extension has been presented with the KTH-TIPS2 Database (Caputo et al., 2005). It contains further material classes and different samples of the same material class that have been acquired under different viewing and illumination conditions. A similar concept for data acquisition has been applied for acquiring the UBO2014 Database (Weinmann et al., 2014). However, one of the main limitations of such databases is the fact that they have been acquired in a lab environment in a controlled setting, thus not appropriately taking into account the complexity of real-world environment conditions, as e.g. illumination conditions strongly influence the appearance of materials. This has for instance been addressed with the Flickr Material Database (Sharan et al., 2009), the OpenSurfaces dataset (Bell et al., 2013) or the Materials in Context (MINC) Database (Bell et al., 2015). These contain images of materials acquired under uncontrolled viewing and illumination conditions beyond lab environments, thus accounting for a large intra-class variation

of material samples regarding their appearance in complex real-world scenarios. Further work is focused on the use of material representations digitized in lab environments to synthesize the respective material appearance variations as seen in real-world environments under different viewing and illumination conditions (Weinmann et al., 2014).

The combination of textural and spectral features for material classification is not as well-established as only using spectral features. However, due to hyperspectral data having a high dimension, classification can be challenging, especially if there is a lack of training data available. Thus, texture can provide additional information for classification of agricultural land using three-dimensional wavelet texture features (Qian et al., 2012), or with a Gray-Level Co-Occurrence Matrix to assess either building materials (Lerma et al., 2000) and meat quality (Yang et al., 2018). The texture can be described using different approaches and the algorithm is chosen to suit the needs.

3. DATASET

For this paper, we used the publicly available spectral library KLUM (Ilehag et al., 2019) that was acquired in-situ in 2018 in the city of Karlsruhe, Germany. The spectral samples were acquired with the high-resolution spectroradiometer ASD Field-Spec-4 Hi-Res¹ in the spectral range of 350–2500 nm. The spectroradiometer has three sensors which cover the ranges of 350–1000 nm, 1001–1800 nm and 1801–2500 nm respectively. A spectral sampling of 1.4 nm and a spectral resolution of 3 nm are used in the spectral range of 350–1000 nm, while the spectral sampling of 1.1 nm and a spectral resolution of 8 nm are used in the remaining spectral range. The spectroradiometer has in total 2151 channels and a wavelength accuracy of 0.5 nm.

KLUM mainly contains building facade materials, consisting of 181 spectral samples. KLUM is split into 13 material classes and 33 material subclasses. The subclasses are split based on the color, the surface attributes (coating, structure and texture) and the state of the sample (e.g. new or old). The spectral library also provides images of the acquired samples.

For this paper, we choose a total of 61 samples from the six material classes *Asphalt*, *Ceramic*, *Conglomerate*, *Limestone*, *Plaster* and *Sandstone* (see Figure 2 for selected examples). These samples provide us with 606 spectra, each sample being represented by eight to ten continuously acquired spectra. The six chosen classes are selected to firstly focus on material classes with similar spectral features but with different textural features, such as *Asphalt* and *Conglomerate*. Secondly, we wanted to limit the amount of contaminated samples (e.g. painted surface) as these features can be observed in this spectral range (Ilehag et al., 2019). Lastly, we include material classes which have distinguished visible textural features, such as *Sandstone*, to investigate the impact of the textural information.

Due to this paper focusing on both spectral and textural information for material classification, we extend the imagery of KLUM by acquiring further high-resolution images of each sample. The images are acquired with a Nikon D810, a single-lens digital camera with 36.6 megapixels and a sensor size of 35.9 mm x 24 mm. The acquired images have a pixel-resolution of 6144 x 4080 pixels. For each sample, three to four images from different distances were acquired, ranging between 0.2–1.2 m. The extended image library consists of 185 images in total and is available online².

¹www.malvernpanalytical.com

²https://github.com/rebeccailehag/KLUM_extended_imagery

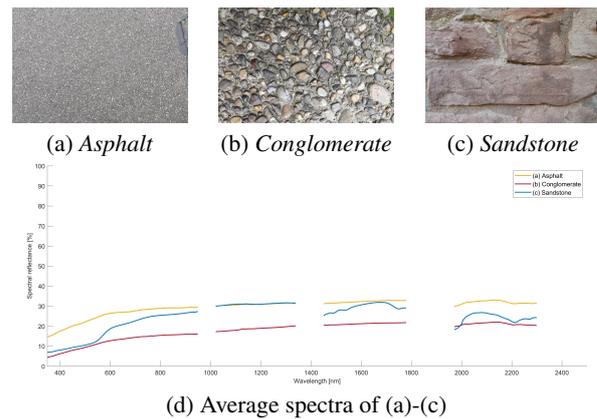


Figure 2. Imagery from the extended KLUM library; (a) *Asphalt*, (b) *Conglomerate* and (c) *Sandstone*, and the corresponding average spectra (d).

4. PROPOSED METHOD

To classify the given samples, a framework consisting of two major steps is utilized. We firstly address the feature extraction by separately deriving the compressed representation of both the spectral information and the textural information using a DAE and a CAE respectively (visualized as the workflow in Figure 3). Secondly, we perform the material classification using all extracted textural and spectral features as input to the supervised classifiers. For the assessment of the DAE, we perform a comparison with the alternative spectral representations by firstly using the original spectra and secondly employing the Principal Component Analysis (PCA) (Tipping and Bishop, 1999) on the original spectra.

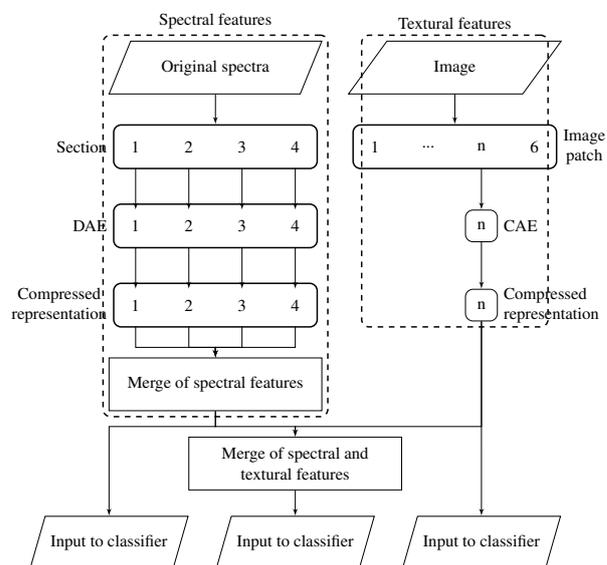


Figure 3. Data processing flow for one material sample, either for the spectral features, the textural features or the combination using the original spectra.

4.1 Compressed feature representation with autoencoders

To assess the performance of the encoder, a decoder, which decodes the compressed representation, and a loss function, that measures the information loss between the compressed and decompressed representation, are needed. The AE structures differ depending on the input, as it is either spectral data or imagery.

Some precautions and procedures remain the same for both types of AEs. Firstly, the input data and the compressed representation are normalized to the closed interval $[0, 1]$. Secondly, we use 20% of the dataset as testing data, 10% for validation and the remaining 70% for training. Lastly, due to the KLUM dataset containing a class imbalance, we assure that each class is represented in both the training and testing datasets.

The utilized models and the necessary pre-processing procedures are further explained in the following sections, firstly the DAE for the spectral compression and secondly the CAE for the textural compression.

Deep autoencoder for spectral information As the dataset was acquired in-situ, some of the spectral channels are removed due to atmospheric effects, resulting in spectral gaps. To counter this, we split the spectral range into four individual sections to reduce the impact of any undesired atmospheric feature. Thus, four spectral bands are defined: 350–949 nm, 1021–1339 nm, 1451–1779 nm and 1971–2299 nm. The bands are not equally sized and each individual DAE is therefore structured differently. As the KLUM dataset contains several spectra of the same sample, we take measures to assure that the training and testing data do not contain spectra from the same sample. That is, a sample's spectra are either present in the training or the testing dataset.

Four DAEs are used in this paper and they are constructed by stacking either three or five hidden layers, depending on the size of the input spectra (see Figure 4 for the DAE structure for the first spectral range). We utilize densely connected layers and the rectified linear unit (ReLU) activation, defined as $y = \max(0, x)$, for each hidden layer except for the last output layer where we use a linear activation, defined as $y = x$. Additionally, we add a sparsity constraint of 10^{-5} to prevent that the hidden layers are learning an approximation of PCA, since this is not desired in this paper. To assemble the DAE, we use the stochastic optimizer Adam (Kingma and Ba, 2014) and the mean squared error to assess the loss. For each of the four DAEs, we retrieve a compressed representation as a vector of its corresponding spectral section. Over the complete range, we receive a compression representation with 32 elements (each DAE retrieves a vector with the size eight).

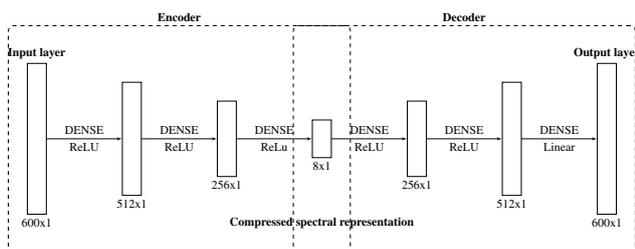


Figure 4. The structure of the DAE for the spectral range 350–949 nm.

Deep denoising convolutional autoencoder for textural information To extract the textural features from the extended imagery, we utilize a CAE. We perform a few image pre-processing procedures to enhance and improve the ability to encode and decode the images. Firstly, as the acquired imagery has a high pixel-resolution of 6144 x 4080 pixels, we scale down the pixel-resolution with a factor of 2 as such a high level of detail is neither desired nor necessary. Secondly, we extract six randomly located patches of 256 x 256 pixels of each image to extend and increase the variation of the dataset. Thirdly, the patches are transformed into a grey-scale using $Y' = 0.2125 \cdot R + 0.7154 \cdot G + 0.0721 \cdot B$

as we do not require information from three color channels to retrieve textural features. Lastly, we add normally distributed noise to increase robustness.

The structure used for the CAE consists firstly of the encoder and secondly the decoder. The structure of the encoder can be seen in Figure 5 and consists of the combination of a 2D convolutional layer with a linear activation, a batch normalization and an additional ReLU activation layer. The last layer in the encoder consists of reshaping the final encoder output into the compressed representation, a vector with 32 elements. The decoder consists of the inverse stacks to decode the compressed representation into the same shape as the encoded input but with a sigmoid activation, defined as $y = \frac{1}{1+e^{-x}}$ to retrieve a value between 0 and 1. We train the CAE with the stochastic optimizer Nadam (Dozat, 2016) and use the mean squared error to assess the loss. Figure 6 displays the distribution of extracted compressed textural features for two samples. As the compressed textural representations differ between the two examples, it indicates that we have retrieved distinct textural features which can be used for material class assignment.

4.2 Classification

To assess the material classes, we perform supervised classification using three different classifiers which are available in the Python toolbox *scikit-learn* (Pedregosa et al., 2011): Random Forest (RF) (Breiman, 2001), Histogram-based Gradient Boosting Classification Tree (HGB) (Pedregosa et al., 2011) and Support Vector Machine (SVM) (Wu et al., 2004). The RF classifier consists of an ensemble of randomly trained decision trees, where each decision tree uses a random subset of the training data for training, which results in a set of decision trees that are different from each other. The parameters of the RF are set to 100 decision-trees and a maximum tree depth of 100. The HGB classifier, which is available as an experimental approach in the toolbox *scikit-learn*, is implemented based on the classifier LightGBM (Ke et al., 2017), a gradient boosting framework that uses a tree-based approach. HGB bases the tree splits on the potential gain and splits the samples into integer-valued bins (histogram bins), which reduces the amount of splitting points. Here, we set the maximum number of iterations to 100, a maximum depth of 10, a loss function based on categorical cross-entropy and a learning rate of 0.5. The SVM classifier bases its classification on the scheme one-vs-one, that is, it separates two classes of interest at a time in a high-dimensional space and considers the distance between the nearest feature of the two classes (Wu et al., 2004). We use a radial basis function kernel for the separation between the classes as a simple linear separation would not be enough for high-dimensional data. Furthermore, the regularization parameter is set to 100, the kernel coefficient for the radial basis function to $1/(num(samples) \cdot var(dataset))$ and the regularization parameter for each class as $num(samples)/(num(classes) \cdot bincount(classes))$ to account for class imbalances.

The material classification is carried out for three kinds of input data: either only spectral features, only textural features that describe the image patches or a combination of both. We use the CAE features to represent the image patches while the spectral features are represented as either the compressed representation that we received from the DAE, the original spectra or the PCA-based encodings of the original spectra. All in all, we define seven different scenarios of input features. We can therefore assess both the contribution of combining spectral and textural features, and the usefulness of the DAE.

Some precautions and procedures are the same for all classification tasks. Firstly, we assure once again that the training and

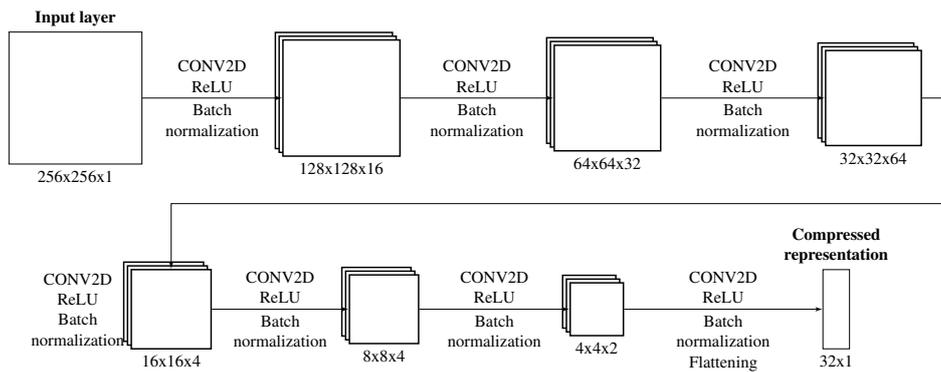


Figure 5. The encoder structure of the CAE for an image patch.

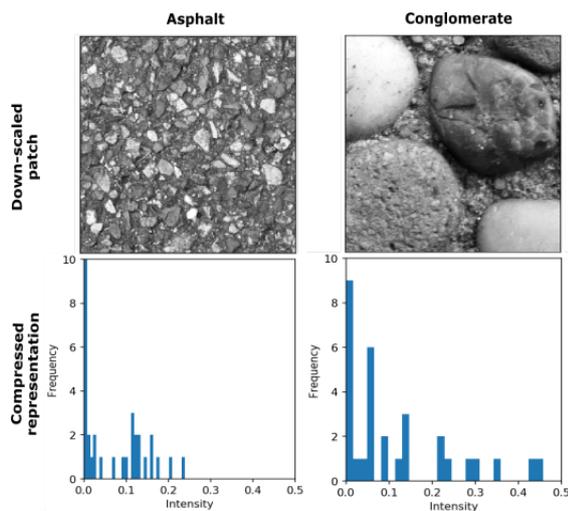


Figure 6. Examples for the two material classes *Conglomerate* and *Asphalt*. Upper row: the down-scaled image patches. Lower row: The corresponding compressed textural representation visualizing the intensity distribution.

testing data do not contain spectra and image patches from the same sample. Secondly, we take care that each class is present in both the training and the testing datasets due to the class imbalance. We utilize a stratified k -fold cross-validation approach to preserve the class distribution for training and testing the classifiers. A k -fold cross-validation approach allows us to perform the classification several times using different splits, which reduces the chance of overfitting during training. We use $k = 4$, which is based on the number of samples in the smallest class to make sure that each split contains enough samples.

Lastly, to evaluate the performance of the classifiers, we consider different measures. We determine the overall accuracy OA, which indicates the overall performance of the classifier, and the κ -value, which indicates how good classes can be separated from each other, in addition to the average recall-score \bar{R} , the average precision-score \bar{P} and the average F_1 -score across all classes. As we use a k -fold cross-validation, we provide the mean evaluation measures in addition to the standard deviation σ .

Classification based on spectral features Hyperspectral data contains redundant information due to their broad spectral range. We apply therefore the PCA algorithm (Tipping and Bishop, 1999) from the Python toolbox *scikit-learn* (Pedregosa et al., 2011) on the original spectra to reduce the spectral dimensionality and retrieve the few first principal components that cover 99.99% of the

variability of the given data.

For the spectral classification, we either use the original spectra, the PCA-based encoding or the compressed representation of the original spectra from the four DAEs. As the compressed representations of the original spectra were split into four, we merge those in the corresponding spectral order. The number of spectral features is 1577 for the original spectra, 10 for the PCA-based encoding and 32 for the merged DAEs.

Classification based on textural features 32 textural features are retrieved from the compressed representations of the image patches and are used as input for the textural classification. The imagery database contains images acquired from different acquisition distances (0.2 – 1.2 m) and the extracted features represent information from different detail levels. To retrieve uniform textural features, only the images acquired from 0.2 m are further used for the classification task.

Classification based on spectral and textural features To combine the spectral and textural features for each sample, we assign each image patch a corresponding set of spectral features randomly chosen from the available spectra for the same sample. We create three different spectral and textural combinations. They consist of the compressed textural features from an image patch combined with either the original spectra, the first few principal components or the compressed spectral representation, from the same sample.

5. RESULTS AND DISCUSSION

In comparison to the PCA method, a DAE requires both training and testing data to retrieve a compressed spectral representation. Depending on the size of the dataset and the structure of the DAE, the compressed spectral representation requires significantly more computing time and resource than the standard PCA computation. In Figure 7, we can observe the ability of the DAE to encode and decode the original spectra in the spectral range of 350 – 949 nm. A difference can be noted while comparing the original spectra with their decompressed correspondence since AEs are lossy. The general spectral characteristics remain the same but smaller deviations can be observed. Furthermore, the compressed spectral representations of the two samples are significantly distinct, which indicates that they can be utilized for material distinction.

The material classification results using the three classifiers reveal, in general, an OA within the range of 70 – 80% as seen in Table 1. However, as spectral characteristic features are similar for urban materials and an inter-class correlation exists, the OA

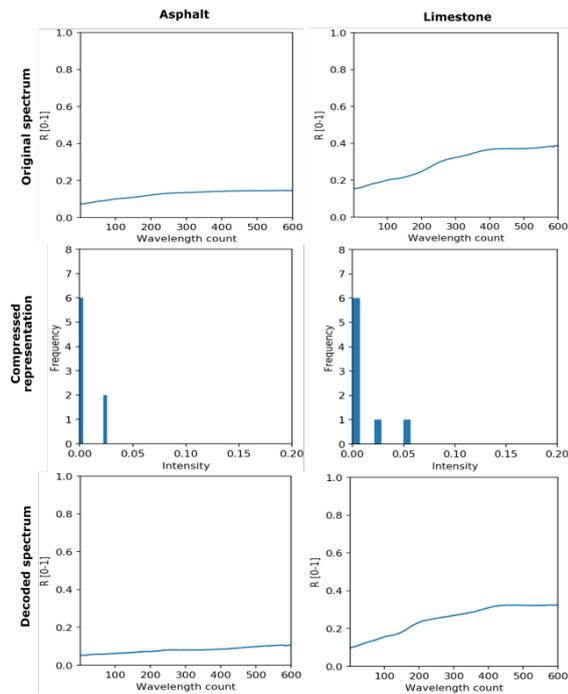


Figure 7. Two examples of the original spectra and their corresponding decompression in addition to the compressed representation visualized as a vector in the spectral range of 350–949 nm for the two material classes *Asphalt* and *Limestone*.

around this range is expected. Furthermore, the compressed spectral representation received from the DAEs does not contribute to substantially better classification results than the standard PCA approach. In general, we receive a rather large σ across all classification approaches, which could indicate that we have a small dataset for this challenging classification task.

The combination of spectral and textural features contributes to unexpected results, as the OA does not significantly increase in comparison to only using spectral features. Contrary, it remains around the same interval. This hints that the spectral features are more important for material classification. This is further supported by observing the most important according to the RF classifier for these combinations. We receive an indication that about 75% of the most important features are spectral features.

The difference between the seven scenarios is minor as the evaluation scores are close to each other. The noticeable difference is however, in general, a larger variation across all scenarios. This implies either that the spectral and textural features are not comparable or that the assignment of spectral and textural feature pairs is not suitable since they contradict themselves. Nonetheless, this does indicate that adding the textural features to a spectral dataset does not improve the distinction of classes with similar spectral features, such as *Asphalt* and *Conglomerate*.

As we perform a stratified k -fold cross-validation approach, we retrieve several confusion matrices for each classifier. The confusion matrices for one cross-validation cycle for SVM can be seen in Figure 8. Here, the cycle with the highest average overall accuracy OA was chosen for comparison of the confusion matrices between the scenarios relying on original spectra, textural features and the combination of the two. The class distributions across these scenarios are similar, except for the scenario with textural features (which is likewise explained by the evaluation scores in Table 1). There is some confusion between the classes

Asphalt and *Conglomerate* in addition to *Ceramic*, *Sandstone* and *Plaster*. The textural and spectral features of *Limestone* appear more distinct as the class has fewer wrongly predicted samples.

In general, the occasionally occurring large σ , the lower evaluation scores and an overall accuracy OA within the range of 70–80% across all classification approaches may be attributed to the limited size of dataset. Furthermore, the dataset contains a certain class imbalance which is further prominent during training of either an AE or a classifier. If the class itself contains a distinct spectral or textural variation, this cannot be trained with the splits containing few samples. A rigorous balancing of all classes would either further decrease the number of samples or the number of considered classes. The conclusions drawn from this paper are specific to this setup. Future work should address the acquisition of a larger dataset including more variation of both spectral and textural features. Nonetheless, this paper can imply that the combination of spectral and textural features does not significantly improve the possibility of distinguishing materials with similar spectral features.

6. CONCLUSION AND OUTLOOK

In this paper, we have classified urban materials by utilizing a close-ranged acquired spectral library by using either spectral features, textural features or a combination of both. We have used two types of AEs to retrieve a compressed spectral representation of the original spectra and a compressed textural representation of the imagery. To assess the compressed spectral representation and its ability as input for classification, we also used the original spectra and a PCA-based encoding. The summary in Table 1 does not show a clear trend whether PCA-based or DAE-based features provide better classification results. While for the RF case the DAE features perform better than those from PCA, there are no consistent results for the HGB and SVM classifiers. Additionally, more steps are required to retrieve the compressed spectral features, which are more time and resource consuming than to retrieve the principal components. Three classification approaches were used: RF, HGB and SVM. The classification approaches perform similar and a clear ranking is not evident. In general, the achieved OA implies that material classification is heavily based on the spectral features as the retrieved results were worse when using only textural features. By analyzing the most important RF features for the combination of spectral and textural features, we receive a hint that implies that the spectral features are more important. The acquired images that were used for the textural study can be downloaded from the dataset 'KLUM extended imagery' (https://github.com/rebeccailehag/KLUM_extended_imagery) for further analysis studies.

The combination of spectral and textural features needs to be further explored. On the one hand, a larger dataset needs to be acquired which includes more spectral samples and corresponding imagery. Such a dataset should likewise cover a larger variation of both spectral and textural features to account for variance and thus more representatively cover a wide range of urban materials. On the other hand, a comparison between CAEs and classical texture analysis approaches could be of interest to retrieve different kinds of textural features and to see if the combined approach provides improved classification results.

REFERENCES

Baldrige, A. M., Hook, S. J., Grove, C. I. and Rivera, G., 2009. The ASTER spectral library version 2.0. *Remote Sens. Environ.* 113, pp. 711–715.

		Spectral features			Textural features	Spectral and textural features		
		Original spectra	PCA	DAE	CAE	Original spectra + CAE	PCA + CAE	DAE + CAE
RF	OA [%]	70±7	77±4	82±3	58±5	71±5	78±7	81±6
	κ [%]	59±10	66±11	74±4	38±12	60±8	68±10	73±9
	\bar{R} [%]	54±14	59±13	65±9	39±4	56±12	63±14	65±14
	\bar{P} [%]	62±19	74±12	72±15	56±8	60±17	74±15	68±16
	\bar{F}_1 [%]	52±12	57±8	62±8	37±4	52±10	63±16	63±15
HGB	OA [%]	72±10	75±8	77±2	60±5	73±5	81±9	77±4
	κ [%]	60±13	65±11	68±3	44±6	63±7	73±12	67±5
	\bar{R} [%]	50±17	58±12	65±13	47±2	54±8	67±18	60±12
	\bar{P} [%]	55±34	70±15	68±15	53±8	52±9	78±16	63±15
	\bar{F}_1 [%]	48±15	58±8	63±13	46±4	51±8	69±17	57±11
SVM	OA [%]	78±5	77±6	76±6	64±6	78±3	78±3	73±8
	κ [%]	71±7	69±8	67±8	50±8	70±3	70±3	61±11
	\bar{R} [%]	69±6	67±8	67±12	51±6	67±8	67±11	57±14
	\bar{P} [%]	65±8	63±7	68±7	51±7	66±11	65±13	59±12
	\bar{F}_1 [%]	64±7	62±8	63±13	49±6	63±9	63±11	56±12

Table 1. The achieved classification results using either RF, HGB or SVM. The evaluation measures display the mean value in addition to $\pm\sigma$.

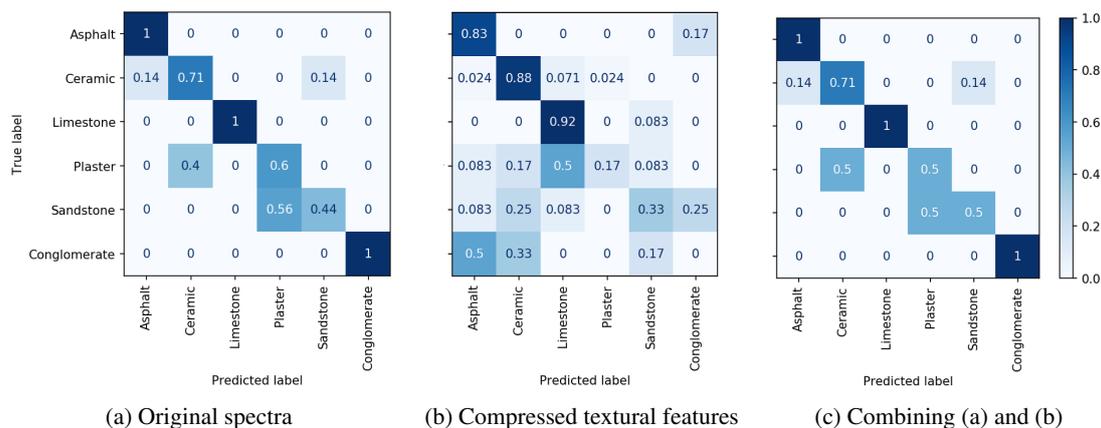


Figure 8. Confusion matrix for the classification results of the RF classifier using: (a) the original spectra, (b) the compressed textural features of the image patches and (c) the combination of the original spectra and the textural features.

Bell, S., Upchurch, P., Snavey, N. and Bala, K., 2013. OpenSurfaces: A richly annotated catalog of surface appearance. *ACM Trans. Graph.* 32(4), pp. 111:1–111:11.

Bell, S., Upchurch, P., Snavey, N. and Bala, K., 2015. Material recognition in the wild with the Materials in Context Database. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3479–3487.

Breiman, L., 2001. Random forests. *Mach. Learn.* 45(1), pp. 5–32.

Caputo, B., Hayman, E. and Mallikarjuna, P., 2005. Class-specific material categorisation. In: *IEEE International Conference on Computer Vision*, pp. I:1597–I:1604.

Dana, K. J., van Ginneken, B., Nayar, S. K. and Koenderink, J. J., 1999. Reflectance and texture of real-world surfaces. *ACM Trans. Graph.* 18(1), pp. 1–34.

Das, R. and Walia, E., 2019. Partition selection with sparse autoencoders for content based image classification. *Neural Comput. Appl.* 31(3), pp. 675–690.

Deshpande, S. S., Inamdar, A. B. and Vin, H. M., 2019. Spectral library and discrimination analysis of Indian urban materials. *J. Indian Soc. Remote Sens.* 47(5), pp. 867–877.

Ding, H., Wang, X., Wang, Y. and Luo, H., 2019. Ensemble classification of hyperspectral images by integrating spectral and texture features. *J. Indian Soc. Remote Sens.* 47(1), pp. 113–123.

Dozat, T., 2016. Incorporating Nesterov momentum into Adam. In: *ICLR Workshop 2016*, Vol. 1.

Fairbairn Jr, K. G., 2013. Visible-near infrared (VNIR) and shortwave infrared (SWIR) spectral variability of urban materials. Technical report, Naval Postgraduate School Monterey CA.

Feng, S., Zhou, H. and Dong, H., 2019. Using deep neural network with small dataset to predict material defects. *Materials & Design* 162, pp. 300–310.

Ferreira, M. P., Wagner, F. H., Aragão, L. E., Shimabukuro, Y. E. and de Souza Filho, C. R., 2019. Tree species classification in tropical forests using visible to shortwave infrared WorldView-3 images and texture analysis. *ISPRS J. Photogramm. Remote Sens.* 149, pp. 119–131.

Geng, J., Fan, J., Wang, H., Ma, X., Li, B. and Chen, F., 2015. High-resolution SAR image classification via deep convolutional autoencoders. *IEEE Geosci. Remote Sens. Lett.* 12(11), pp. 2351–2355.

Grilli, E. and Remondino, F., 2019. Classification of 3D digital heritage. *Remote Sens.* 11(7), pp. 847:1–847:23.

Hayman, E., Caputo, B., Fritz, M. and Eklundh, J.-O., 2004. On the significance of real-world conditions for material classification. In: *European Conference on Computer Vision*, pp. 253–266.

Herold, M., Roberts, D. A., Gardner, M. E. and Dennison, P. E., 2004. Spectrometry for urban area remote sensing—Development and analysis of a spectral library from 350 to 2400 nm. *Remote Sens. Environ.* 91, pp. 304–319.

Ilehag, R., Schenk, A. and Hinz, S., 2017a. Concept for classifying facade elements based on material, geometry and thermal

- radiation using multimodal UAV remote sensing. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XLII-2/W6, pp. 145–151.
- Ilehag, R., Schenk, A., Huang, Y. and Hinz, S., 2019. KLUM: An urban VNIR and SWIR spectral library consisting of building materials. *Remote Sens.* 11(18), pp. 2149:1–2149:28.
- Ilehag, R., Weinmann, M., Schenk, A., Keller, S., Jutzi, B. and Hinz, S., 2017b. Revisiting existing classification approaches for building materials based on hyperspectral data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XLII-3/W3, pp. 65–71.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y., 2017. LightGBM: A highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems*, pp. 3146–3154.
- Kingma, D. P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kokaly, R. F., Clark, R. N., Swayze, G. A., Livo, K. E., Hoefen, T. M., Pearson, N. C., Wise, R. A., Benzel, W. M., Lowers, H. A., Driscoll, R. L. et al., 2017. USGS spectral library version 7. Technical report, US Geological Survey.
- Kolbe, T. H., Gröger, G. and Plümer, L., 2005. CityGML: Interoperable access to 3D city models. In: *Geo-Information for Disaster Management*, Springer, pp. 883–899.
- Kotthaus, S., Smith, T. E. L., Wooster, M. J. and Grimmond, C. S. B., 2014. Derivation of an urban materials spectral library through emittance and reflectance spectroscopy. *ISPRS J. Photogramm. Remote Sens.* 94, pp. 194–212.
- Kramer, M. A., 1991. Nonlinear principal component analysis using autoassociative neural networks. *Am. Inst. Chem. Eng. J.* 37(2), pp. 233–243.
- Lan, R., Li, Z., Liu, Z., Gu, T. and Luo, X., 2019. Hyperspectral image classification using k-sparse denoising autoencoder and spectral-restricted spatial characteristics. *Appl. Soft Comput.* 74, pp. 693–708.
- Lerma, J. L., Ruiz, L. and Buchón, F., 2000. Application of spectral and textural classifications to recognize materials and damages on historic building facades. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XXXIII-B5, pp. 480–484.
- Li, W., Fu, H., Yu, L., Gong, P., Feng, D., Li, C. and Clinton, N., 2016. Stacked autoencoder-based deep learning for remote-sensing image classification: A case study of African land-cover mapping. *Int. J. Remote Sens.* 37(23), pp. 5632–5646.
- Lin, Z., Chen, Y., Zhao, X. and Wang, G., 2013. Spectral-spatial classification of hyperspectral image using autoencoders. In: *2013 9th International Conference on Information, Communications & Signal Processing*, IEEE, pp. 1–5.
- Masci, J., Meier, U., Cireşan, D. and Schmidhuber, J., 2011. Stacked convolutional auto-encoders for hierarchical feature extraction. In: *International Conference on Artificial Neural Networks*, Springer, pp. 52–59.
- Mirzapour, F. and Ghassemian, H., 2015. Improving hyperspectral image classification by combining spectral, texture, and shape features. *Int. J. Remote Sens.* 36(4), pp. 1070–1096.
- OpenStreetMap contributors, 2017. Planet dump retrieved from <https://planet.osm.org>. <https://www.openstreetmap.org>.
- Ouerghemmi, W., Gadal, S., Mozgeris, G., Jonikavičius, D. and Weber, C., 2017. Urban objects classification by spectral library: Feasibility and applications. In: *2017 Joint Urban Remote Sensing Event (JURSE)*, IEEE, pp. 1–4.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, pp. 2825–2830.
- Qian, Y., Ye, M. and Zhou, J., 2012. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. *IEEE Trans. Geosci. Remote Sens.* 51(4), pp. 2276–2291.
- Sánchez, J. and Quirós, E., 2017. Semiautomatic detection and classification of materials in historic buildings with low-cost photogrammetric equipment. *J. Cult. Herit.* 25, pp. 21–30.
- Santamouris, M., Synnefa, A. and Karlessi, T., 2011. Using advanced cool materials in the urban built environment to mitigate heat islands and improve thermal comfort conditions. *Sol. Energy* 85(12), pp. 3085–3102.
- Sharan, L., Rosenholtz, R. and Adelson, E., 2009. Material perception: What can you see in a brief glance? *J. Vis.* 9(8), pp. 784.
- Tipping, M. E. and Bishop, C. M., 1999. Probabilistic principal component analysis. *J. Royal Stat. Soc.: Series B (Statistical Methodology)* 61(3), pp. 611–622.
- Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P.-A., 2008. Extracting and composing robust features with denoising autoencoders. In: *The 25th International Conference on Machine Learning*, pp. 1096–1103.
- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. and Manzagol, P.-A., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* 11(Dec), pp. 3371–3408.
- Ward, K., Lauf, S., Kleinschmit, B. and Endlicher, W., 2016. Heat waves and urban heat islands in Europe: A review of relevant drivers. *Sci. Total Environ.* 569, pp. 527–539.
- Weinmann, M., Gall, J. and Klein, R., 2014. Material classification based on training data synthesized using a BTF database. In: *European Conference on Computer Vision*, pp. 156–171.
- Wu, T., Lin, C. and Weng, R. C., 2004. Probability estimates for multi-class classification by pairwise coupling. *J. Mach. Learn. Res.* 5(Aug), pp. 975–1005.
- Yang, Y., Wang, W., Zhuang, H., Yoon, S.-C. and Jiang, H., 2018. Fusion of spectra and texture data of hyperspectral imaging for the prediction of the water-holding capacity of fresh chicken breast filets. *Appl. Sci.* 8(4), pp. 640:1–640:18.
- Yuan, J., Wang, D. and Li, R., 2013. Remote sensing image segmentation by combining spectral and texture features. *IEEE Trans. Geosci. Remote Sens.* 52(1), pp. 16–24.
- Yuan, L., Guo, J. and Wang, Q., 2020. Automatic classification of common building materials from 3D terrestrial laser scan data. *Autom. Constr.* 110, pp. 103017.
- Zhou, P., Han, J., Cheng, G. and Zhang, B., 2019. Learning compact and discriminative stacked autoencoder for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 57(7), pp. 4823–4833.