Technische Universität Berlin

Faculty of Electrical Engineering and Computer Science
Dept. of Computer Engineering and Microelectronics
Remote Sensing Image Analysis Group

Learning-Based Hyperspectral Image Compression Using A Spatio-Spectral Approach

---

Master of Science in Computer Science
August, 2023

Sprengel, Niklas

Matriculation Number: 380009

Supervisor: Prof. Dr. Begüm Demir

Advisor: Martin Fuchs

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt habe. Sämtliche benutzten Informationsquellen sowie das Gedankengut Dritter wurden im Text als solche kenntlich gemacht und im Literaturverzeichnis angeführt. Die Arbeit wurde bisher nicht veröffentlicht und keiner Prüfungsbehörde vorgelegt.

Hereby I declare that I wrote this thesis myself with the help of no more than the mentioned literature and auxiliary means.

Berlin, Date

. . . . . .

Sprengel Niklas

Abstract

This template is intended to give an introduction of how to write diploma and master thesis.

Contents

Acronyms

1D One-dimensional . . . . . 8 f.

2D Two-dimensional 8, 10

ANN Artifical neural network . . . . . 7 f., 11 f.

ASI Italian Space Agency 2

BTC Block Truncation Coding 4

CAF Cross-layer adaptive fusion 12

CCSDS Consultative Committee for Space Data Systems 8

CNN Convolutional neural network . . . . . 5, 7 f., 10–14

DCT Discrete cosine transform 7

DFT Discrete Fourier transform 7

DWT Discrete wavelet transform . . . . . 7 f.

EnMAP Environmental Mapping and Analysis Program . . . . . 1 f.

ESA European Space Agency 2

GAN Generative adversial network 7

GPU Graphical processor unit 11

GSD Ground sampling distance . . . . . 1 f.

HEIC High Efficiency File Format 4

JPEG JPEG . . . . . 4, 12 f.

KLT Karhunen-Loeve transform 7

LeakyReLU Leaky rectified linear unit 9

MLP Multi-layer perceptron 7, 12

NLP Natural Language Processing 14

PCA Principal component analysis 8

PNG Portable network graphics 12

PRISMA PRecursore IperSpettrale della Missione Applicativa 2

RGB Red, green and blue . . . . . 2, 4 f., 7, 11 ff.

S2 Sentinel-2 2

SVM Support vector machine 3, 7

SWIR Shortwave infrared . . . . . 1 f.

TD Tucker decomposition 8

VNIR Visible and near infrared . . . . . 1 f.

List of Figures

List of Tables

1 Introduction

Hyperspectral imaging is a quickly growing field. It is the technique of capturing images with a specialized camera in order to obtain a spectrum of many wavelengths of light for each pixel of a taken image. There are three categories of hyperspectral cameras that are used to capture such images. The first, push broom scanners, use a linear arrangement of spectroscopic sensors, that are sensors able to capture a spectrum of many wavelengths of light at once. The sensor arrangement is then moved over the subject of the image, for example by being attached to a satellite orbiting earth. The second category, whisk broom scanners, functions similarly to push broom scanners with the difference being that the stationary sensor arrangement is replaced by a moving mirror reflecting light into a single detector that collects the spectral information in a step-by-step manner. Lastly, snapshot hyperspectral imaging works by using a sensor array to capture a complete hyperspectral image with a single activation of the sensors.

Regardless of capture method, there are many applications for the use of hyperspectral images since much information can be extracted from the combination of spatial and spectral data in them. One such application is pixel-wise classification of materials, that is determining the material of specific pixels in a hyperspectral image. This is possible because of the high information present per pixel resulting from the large amount of spectral information. An instance of this approach can be seen in Zea et.al. [1]. They use classification of hyperspectral images in order to detect the presence of the toxic metal cadmium in soil, lessening the need for chemical methods that require the plant to be harvested to test for cadmium stress. Using hyperspectral imaging the detection of cadmium can be performed on living plants. Another example of pixel-wise classification materials is in Henriksen et.al. [2], where it is used to separate plastic waste by twelve kinds of plastics better than previously used methods such as near-infrared technology.

Another field where the use of hyperspectral imaging is rising in importance is in geology and environmental sciences. An example of this is the Environmental Mapping and Analysis Program (EnMAP) mission which launched a satellite into earths orbit in 2022 that takes images of the surface of the earth with a comparatively high ground sampling distance (GSD) of 30 square meters per pixel, allowing for regional geographic anal-

ysis [3]. In the spectral domain, the satellite yields high resolution. By combining a sensor in the visible and near infrared (VNIR) spectrum and a sensor in the shortwave infrared (SWIR) spectrum it can capture light with wavelengths between 420 and 2450 nm. Wavelengths between 420 and 1000 nm are captured by the VNIR sensor with a resolution of 6.5 nm per spectral band, while wavelengths between 900 and 2450 nm are captured by the SWIR sensor with a resolution of 10 nm per band. The EnMAP mission has already lead to many interesting studies including of glacier ice surface properties of the Ice Sheet in South-West Greenland, moisture content of soil below grassland and identification of specific crop traits such as the chlorophyll content or the leaf water content [4][5][6]. A dataset created from the images produced by the EnMAP mission is also the main dataset studied in this thesis.

Another ongoing hyperspectral imaging mission is the PRISMA (PRecursore IperSpettrale della Missione Applicativa) mission led by the Italian Space Agency (ASI) that was launched in 2019 [7]. Similar to the EnMAP mission, the PRISMA mission also consists of a satellite that observes earth using both a sensor for the VNIR spectrum and for the SWIR spectrum. It also has a GSD of 30 square meters per pixel and a slightly lower spectral resolution given as less than 12 nm per band [8][9]. In contrast to the EnMAP mission the satellite also carries a panchromatic camera, meaning that the camera only has a single spectral band. This camera however has a higher GSD of 5 square meters per pixel. The data from the PRISMA mission has been used to increase spatial resolution in hyperspectral images. This was done by combining the original hyperspectral images with multispectral images with a higher spatial resolution in a process called hyperspectral-multispectral fusion [10]. The term 'multispectral image' refers to images with more spectral bands than traditional red, green and blue (RGB) images, but less bands than hyperspectral images. Here the multispectral images contain 10 bands and are obtained from the Sentinel-2 (S2) project, a mission by the European Space Agency (ESA) that continually performs multispectral imaging with fine spatial resolution [11]. Another use of the PRISMA data has been in the disaster monitoring where it has been used to segment wildfires into areas of active fire, smoke, burned ground, bare soil and remaining vegetation [12]. Another related system developed using PRISMA data performs wildfire detection in real-time from satellites [13].

The increasing usage of hyperspectral imaging makes it essential to address the disadvantages of the technology. One major disadvantage of this type of imaging is the required amount of disk space for the resulting images. Because there are many more spectral bands compared to the three bands of red, green and blue in traditional photography the resulting file size rises accordingly. Furthermore, hyperspectral images often use high precision for the specific brightness values captured for each point in the pixel-band data cube. The EnMAP satellite images are composed of 32 bit values, while in

RGB imaging 8 bits per pixel per band are most common [3]. A hyperspectral image might for example have 300 bands. In combination with 32 bit values per pixel per band the resulting file size would be 400 times larger than an RGB image with the same pixel density. In addition to this it is important to notice that even RGB images need to be compressed for many use cases in order to be efficiently stored and transmitted, which is why there are many compression standards for RGB images in wide usage. Furthermore, compression of hyperspectral images is a complex problem. This arises from the combination of spatial image compression complexities and the added challenge of encoding information in the spectral dimension which RGB compression algorithms do not consider. The latter is one of the reasons why many RGB compression algorithms do not perform well on hyperspectral data as will be shown in this thesis. In addition to this there are more technical reasons for the difficulty of adapting algorithms for RGB compression that will also be explored in this thesis. For the above mentioned reasons it is clear that it is important to research compression algorithms for hyperspectral images.

Possible algorithms for image compression can be categorized into multiple broad groups. They can firstly be grouped by whether they compress the image losslessly or with loss. Lossless algorithms restore the compressed image exactly whereas lossy reconstructions cause distortion. The disadvantage of lossless compression is however that there are mathematical limits to the compression rate given by the entropy in the data that is to be compressed. This limit is given by Shannon's source coding theorem [14], which has been stated in MacKay 2003, pg. 81 [15] as follows:

> N independent identically distributed random variables each with entropy $H(X)$ can be compressed into more than $NH(X)$ bits with negligible risk of information loss, as $N \to \infty$; but conversely, if they are compressed into fewer than $NH(X)$ bits it is virtually certain that information will be lost.

This means that the average bitrate, the compression rate given by the quotient of input bits and output bits of the compression algorithm, that is possible to achieve using lossless compression algorithms is given by the entropy of the data that is to be compressed. In contrast, lossy compression methods can achieve much higher compression rates by allowing the introduction of distortion. Furthermore, they can also adapt their rate of compression based on either the desired amount of distortion or the target compression rate. They can also be combined with a lossless compression method to further optimise their compression rate. Another important factor that determines the trade-off between lossless and lossy compression methods is that many applications do not require the perfect reconstruction given by lossless methods. For example, García-Vílchez et.al. [16] showed that in hyperspectral image classification lossy compression of the image does not always reduce the performance of the classifier. For some lossy compression

algorithms the classifier even produced better results on the compressed data than the uncompressed data. This is explained by an introduction of smoothing from the compression which is advantageous for the support vector machine (SVM) classifier used in the experiment. For these reasons this thesis focuses on lossy compression methods.

The category of lossy compression algorithms can be further divided into traditional compression methods and the learning-based compression methods. Traditional compression methods employ algorithms such as the discrete cosine transform which is used in the JPEG (JPEG) image compression standard or the wavelet transform, often in combination with a lossless entropy coding method. Examples of traditional methods other than JPEG are Block Truncation Coding (BTC) which splits images into blocks and then rounds pixel values inside the blocks efficiently and the High Efficiency File Format (HEIC) standard, which uses wavelet transforms similar to JPEG [17][18]. Learning-based approaches use the powerful capability of artificial neural networks to universally approximate arbitrary functions using gradient descent [19]. These networks are then used to build models that learn to compress and decompress images. Lately these networks have been shown to outperform traditional compression methods for many applications, especially for RGB images using for example the hyperprior model given by Ballé et.al. which will be discussed in detail in section 2.1 [20][21][22]. Compression for hyperspectral images is in the early stages of research, however even in this domain there are promising results showing the capabilities of learned autoencoders for this task [23][24][25][26].

1.1 Objective

This thesis addresses the problem of hyperspectral image compression using machine learning models consisting of two parts, an encoder and a decoder.
The encoder learns to map the original image to a low-dimensional latent space, thereby performing compression. Analogously, the decoder is trained to reconstruct the input by mapping elements from this latent dimension to full-size images that are as close as possible to the original image.

Contributions to the research in on learned hyperspectral image compression are made in this thesis by introducing a new architecture that enables the use of spatial compression algorithms such as models that might be used for RGB image compression by combining them with a model that performs compression in the spectral domain while keeping spatial relationships intact. In this way, it is possible to compress the spatial information in hyperspectral images using models that are not ordinarily applicable to

these images.

Using this architecture, multiple model combinations are designed and compared with each other as well as with the base versions of these models.
These models include models based on convolutional neural networks (CNNs) as well as transformer-based architectures. A model using a hyperprior architecture is also used for the spatial compression. This architecture yields much higher compression ratios than other models by using an arithmetic encoder in the bottleneck. With this model compression rates much higher than the current state of the art for hyperspectral image compression are achieved while distortion is only reduced by a comparatively small amount.

Additionally, the latent spaces of both the encoder in the spectral domain as well as the latent space of the spatial encoder will be analysed.

1.2 Outline

This section gives a brief introduction into the chapters in the thesis, which is split into 7 chapters.

Chapter 2 presents the related work. This relates to the research of learning-based hyperspectral image compression by itself. However, as learning-based compression for hyperspectral images, in contrast to traditional hyperspectral compression methods, is a recent and relatively unexplored field of study, the adjacent research topic of learned RGB image compression is also explored. There is also some exploration of the general studies done on convolutional neural networks as well as transformers.

Chapter 3 gives an overview of the theoretical ideas used in the proposed methods, such as CNNs, transformers, arithmetic coding as well as the hyperprior architecture for compression.

Chapter 4 details the individual submodels making up the models that are proposed to address the hyperspectral image compression problem as well as the models in their totality. It also gives an overview over the loss functions used for training the models, one of which is a loss function specifically designed for the models proposed in this thesis.

Chapter 5 explains the design of the experiments done, the dataset that is used for

these experiments as well as the results of these experiments. It also details some of the challenges encountered during the experimentation phase.

Chapter 6 gives a summary of the results from the thesis as well as possible improvements that could be made as well as ideas that could be explored in future research.

2 Related Work

Learned hyperspectral image compression is a developing field of study. While there are papers published on this topic, there are some problems making it difficult to assess and compare the results of these studies, as will be illuminated further in this chapter. An overview of hyperspectral image compression algorithms by Dua et.al. [27] shows the discrepancy between traditional transform-based and prediction-based compression techniques and techniques based on machine learning. The review contains 21 transform-based and 19 prediction based techniques but only five learning-based models. While the overview over the models based on machine learning is no longer complete since the review was done in 2020, it still shows that learning-based approaches for hyperspectral image compression have been a less researched field until recently. Lately however, research in this field has been increasing quickly SOURCE. Since learned RGB image compression is a closely related field and much more widely researched, the studies done regarding this topic are also analysed.

2.1 Hyperspectral Image Compression

Most learned hyperspectral image compression papers use a CNN-based model architecture to reduce the dimensions of the input image [23][24][25]. Another model proposed by Guo et.al. [26] uses a hyperprior architecture, originally developed by Ballé et.al. [20], which also uses convolutional layers in an artifical neural network (ANN) but combines them with an arithmetic coder to improve compression rate. Some other models that will be explored later are a model based on a SVM by Aidini et.al. [28], a generative adversial network (GAN) model by Deng et.al. [29] and a model using a simple multi-layer perceptron (MLP) by Kumar et.al. [30]. Before detailing these methods, an overview of some traditional hyperspectral image compression architectures is given.

2.1.1 Traditional Architectures

Dua et. al. [27] describes eight categories of hyperspectral image compression techniques, one of which are learning-based approaches. Of the traditional methods, that is methods not based on machine learning, the most researched categories are transform-based techniques and prediction-based techniques. Transform-based techniques work by transforming the spatial dimension, the spectral dimension or both into the frequency domain by using a transformation function that is applied to the input image [27]. Possible transformation functions are the Karhunen-Loeve transform (KLT), the discrete cosine transform (DCT), the discrete Fourier transform (DFT) or a discrete wavelet transform (DWT). The transformation can then be used to decorrelate the image in the spatial, spectral or both domains depending on the specific technique used. More specifically, the KLT directly decorrelates the data, while for the other transformation functions additional algorithms are commonly used to decorrelate the coefficients resulting from the transformation [31][32]. After decorrelation, the coefficients are quantized by removing coefficients that are close to zero [27]. The quantized coefficients are then compressed further using an entropy coding algorithm in the final step of compression. Decompression is possible because the used transformation functions are invertible, although loss is introduced during the quantization step. Some of these general steps may be modified for specific transform-based architectures.

Looking at a specific example of a transform algorithm, Karami et.al. [32] uses a two-dimensional DWT in combination with the Tucker decomposition (TD) algorithm, an extension of principal component analysis (PCA), in order to reduce correlation. They apply the 2DWT to each spectral band of the input image and obtain four tensors containing the coefficients for each band. The TD algorithm is then used to decorrelate the coefficients both in the spatial and spectral domain. The least significant coefficients are then removed using an iterative process before the remaining coefficients are compressed using the entropy coding technique of arithmetic coding.

The other common category of traditional hyperspectral image compression techniques are prediction-based approaches. These approaches compress images by predicting pixel values from the other values of the other pixels using some algorithm and then only storing the prediction residuals [27][33]. The current compression standard proposed by the Consultative Committee for Space Data Systems (CCSDS), CCSDS 123.0-B2, is a prediction-based approach [34]. The predictor for this standard predicts a pixel value by computing the sum of its neighbours [33]. The difference between these sums and the original pixel values are then stored in a local difference vector which can then be compressed using an entropy coder.

2.1.2 Learning-based Architectures

In contrast to the traditional methods, learning-based architectures use ANNs trained using gradient descent to learn a mapping from the space of input images to a latent space [19]. In image compression, this latent space is lower-dimensional than the input space, thereby performing compression. The most common approach for learned hyperspectral compression are CNN-based architectures. These architectures can be split into two categories, the first being the models using two-dimensional (2D) convolutional layers to learn the spatial dependencies of the hyperspectral images [25]. The second category are models using one-dimensional (1D) convolutional layers to learn the spectral dependencies of the input data [23][24]. Prior to the release of this thesis there are no purely CNN-based papers using both the spatial and the spectral dependencies of hyperspectral images for compression. The model proposed by Guo et.al. [26] does use both spatial and spectral dependencies, it does however use a hyperprior architecture and not a purely CNN-based model. This model as well as the hyperprior architecture will be detailed later in this chapter. As mentioned above, comparing the results from these papers directly is difficult. The reason for this is that the models use different datasets since there is currently no accepted standard dataset for hyperspectral imaging. Furthermore, the models are designed for different compression rates. Since a higher compression rate also leads to a higher distortion for the same model in most cases as described by rate-distortion theory [35], this makes it hard to directly compare the results from papers that use both a different dataset and different compression rates, therefore requiring a reimplementation of the models in order to be able to fairly compare them.

A model exploiting only the spectral dependencies in the input images is provided by Kuester et.al. [23][24]. They use a model consisting of a 1D convolutional layer to learn spectral dependencies followed by a max pooling layer applied to the spectral dimension to reduce the dimensionality of the image. This block of 1D convolutional layer and a max pooling layer is repeated, using leaky rectified linear unit (LeakyReLU) as the activation function. Following that, two more 1D convolutional layers are added, now without max pooling layers inbetween. Each convolutional layer uses less filters than the layer before it. Finally, the last layer uses only one filter, thereby providing the bottleneck of the compression algorithm with a fixed compression rate of 4. Decompression is performed using a reversed model, symmetrical to the encoder. The decoder only differs by replacing the blocks consisting of a convolutional and a max pooling layer by transposed 1D convolutional layers with a stride of two to increase the spectral dimensionality instead of reducing it [24]. In an earlier variation of the model, upsampling layers are used instead for this purpose [23]. In contrast to most other models, this model is trained and applied not to a whole image at once, but rather each pixel individ-

ually. This has the advantage of reducing the complexity of the task the model has to perform. Instead of compressing a complete hyperspectral image, the model only learns to compress the spectral signatures of arbitrary pixels from these images. The number of training samples also increases dramatically, as each hyperspectral image consists of many pixels, although each training sample contains much less information than it does for a model that trains on complete images at once. A disadvantage of the approach is that modelling spatial dependencies is not possible using a per-pixel training approach. Furthermore, training is much slower when compared to other models. The reasons for this and more detailed comparisons with other model architectures are explored in Chapter 5.

In contrast to the aforementioned model, La Grassa et.al. [25] propose a model that focuses on the spatial dependencies in the image data. This model is trained and applied on complete hyperspectral images in one step. They use a combination of 2D convolutional layers and max pooling layers applied to the spatial dimension of a input image to reduce the spatial dimensionality while simultaneously increasing the number of filters in the convolutional layers to keep the amount of stored information stable. After multiple of these blocks, the output of the last block is flattened into a single dimension, followed by a linear layer that maps that output into a one-dimensional latent space of the chosen size. The decoder is again a symmetrical, reversed version of the encoder, replacing only the blocks of 2D convolutional and max pooling layers by transposed convolutional layers with a stride of 2. This model has the advantaqe of having the ability to set a precise bit rate by changing the output dimensions of the final linear layer. However, applying the model to hyperspectral images with many bands is difficult. This is because in the proposed model, the first convolutional layer uses 64 filters. Since this layer is applied to the complete input image, an input image with dimensions (C,H,W) where C is the number of spectral bands and H and W are the height and width of the image respectively is transformed into the dimensions (64,H,W). If C is significantly higher than 64, this leads to large information loss in the first layer. This is common in hyperspectral datasets. This can be seen in the HySpecNet11k dataset CITATION, the main dataset used in this thesis, which uses 202 channels. Possible solutions to this problem are explored in Chapter 4.

There are also some approaches that do not use a CNN-based architecture. While some of these architectures use models that contain a CNN in addition to other parts, some architectures use different model types completely. Among these are both different types of neural networks as well as learning-based algorithms not using neural networks. One such architecture is proposed by Aidini et.al. [28]. They use quantization to compress the original image, meaning that the precision of the spatial and spectral values of the image are decreased. Then an algorithm tries to recover the original tensor values by try-

ing to reconstruct low-rank tensors as a constrained optimization problem. Afterwards a spatial super-resolution algorithm using trained dictionary learning is used to increase the resolution of this image, after which a classifier is trained on these super-resolved images. While this architecture is interesting, it is not directly used in this thesis as its methodology is very different to the neural network based methodologies used in the thesis.

Another category of models use ANNs to determine parameters for lossless compression algorithms. Shen et.al. [36] use a deep belief network to determine the optimal parameters for golomb-rice coding, a lossless coding algorithm that normally assumes a geometric underlying distribution. Using a neural network to determine the parameter removes that necessity. This core strategy is also used by Guo et.al. [26]. They use a hyperprior architecture to compress hyperspectral images. Hyperprior models use an ANN-based model to transform the image data into a latent space that is commonly lower-dimensional than the input image. Then a second ANN is trained on the latent space to determine parameters for an arithmetic coder, a lossless coding algorithm. In both Guo et.al. and the original paper introducing the hyperprior architecture for RGB images, Ballé et.al. [20], both ANNs are CNNs and the latent space is indeed a lower-dimensional space.

Guo et.al. innovates on the original approach in the fact that they assume a student's T distribution instead of a gaussian distribution for the arithmetic coder and in the design of the first CNN. Their version includes both a spatial and a spectral part in the main CNN, making it the only model to learn both spatial and spectral information for hyperspectral image compression. They achieve this by using 2-D convolutional layers for the spatial domain and 3-D convolutional layers with a kernel size of one to include the spectral dependencies. However, a disadvantage of their model is that it is developed only for datasets with a low amount of channels compared to other hyperspectral datasets with the highest having 102 spectral channels. The approach is also not easily adaptable to datasets with a much higher number of channels. This is because the first convolutional layer of the model uses 192 filters and a stride of two, therefore transforming an input image with the dimensions (C, H, W), where C is the number of spectral channels and H and W are the height and width respectively, into the dimensions (192, H/2, W/2). For one of the two datasets studied in Guo et.al., the KAIST dataset, which contains 31 spectral channels, this is a large increase in information. For the other dataset, ROSIS-Pavia, with 102 to 103 spectal channels depending on the scene, this is already a loss of information, requiring the first layer to already encode some of the spatial information of the image appropriately in order to not introduce large amounts of distortion. For datasets with more channels this problem increases. Furthermore, because of graphical processor unit (GPU) memory limitations, increasing the amount

of filters appropriately is not always possible.

Another model using neural networks is proposed by Kumar et.al. [30]. Instead of CNNs they use a simple multi-layer perceptron as the decoder for the reason that they use this model for real-time onboard image compression which requires a much more simple model for faster execution speed. Another uncommon trait of their architecture is that they do not use a symmetric model, meaning that the encoder and decoder are mirrors of one another. Instead, they only use an ANN for the decoder and use a low-complexity encoder based on matrix multiplication. Hong et.al. [37] propose an novel architecture as well. They use a transformer that works on a spectral embedding by linearly projecting groups of neighbouring spectral bands to an embedding vector. This improves the capabilities of the network since neighbouring bands in hyperspectral images capture detailed changes in the absorption of the underlying material and therefore contain important information, especially for classification tasks. In addition to this they implement cross-layer adaptive fusion (CAF) to improve exchange of information in the transformer section of the model. This means that they use multiple transformer layers and, in addition to the direct connection between adjacent transformer layers, add connections that skip one layer and connect with the layer after using a special CAF module. The transformers can be applied either per-pixel or for small patches. However, their work is not used in the context of image compression but rather image classification and therefore only contains an encoder combined with an MLP head to classify hyperspectral image pixels or patches based on categories such as ”Corn”, ”Grass Pasture” or ”Wheat”. This means that an application of this architecture to the task of image compression would require substantial additions to the model.

2.2 RGB Image Compression

While the study of hyperspectral image compression strongly increased in recent years, RGB image compression has been a widely researched field for many years. There are many traditional compression algorithms, some of which are installed on every modern operating system and browser. Examples of these are portable network graphics (PNG), a lossless image compression format as well as JPEG, a collection of compression algorithms, the most common of which performs lossy compression of images. An improved version of JPEG, JPEG 2000, also exists, offering increased compression rate for similar distortion values [38]. However, while these algorithms are very popular, learning-based image compression has outperformed methods such as JPEG in both compression ratio and distortion [20][21]. Furthermore, many of the ideas in learned hyperspectral image compression originate from RGB image compression studies and

many of the ideas in RGB image compression have not yet been adapted to the hyperspectral realm. For these reasons RGB image compression papers are directly relevant even for a thesis that only concerns itself with the hyperspectral domain.

2.2.1 The Hyperprior Architecture

One of the most important recent works in RGB image compression was released by Ballé et.al. [21]. The model proposed in this builds on a previous work [20] using a CNN to reduce the dimensionality of the input image, followed by a quantization of the resulting latent and the usage of an arithmetic autoencoder to losslessly compress the quantized latent. The output of the arithmetic autoencoder is then decoded by a CNN that is symmetrical to the encoder CNN, similar to the models for hyperspectral image compression that were already discussed. This model already outperformed JPEG and the improved JPEG 2000 on the tested data.

The performance of the arithmetic autoencoder depends on the accuracy of the estimated probability distribution that is used to compress the given data. This area is where improvements were found. Ballé et.al. [21] used a smaller, separate CNN that learns to extract parameters for a good probability distribution estimate from the latent resulting from the main CNN. This network also uses an autoencoder structure, meaning that the estimate can also be transmitted in compressed form as side channel information using only a small amount of space. Training a network using gradient descent for this purpose would not ordinarily be possible since the quantized latents have discrete values resulting in zero gradients everywhere. For this reason the quantization is substituted during training by addition of a small amount of uniform noise to dediscretize the latents.
This hyperprior architecture is part of a large portion of modern learned RGB image compression models and also the hyperspectral image compression model by Guo et.al. [26] that was discussed in Chapter 2.1.

One such example is a paper by Hu et.al. [39] where the model is generalised to include not only two but an adaptable number of CNNs, where each CNN learns the probability distribution of the CNN before it. This leads to slight improvements in the bitrate of the model while not changing the distortion. The distortion remains unchanged since the only loss occurs within the first CNN which compresses the input image . The other CNNs are only used to improve the performance of the lossless arithmetic coders.

The architecture was also further improved in two ways by Minnen et.al. [22]. The first improvement is a generalisation in the structure of the probalitiy distribution from the

original scale mixture of gaussians [40] to a gaussian mixture model. This means that the second CNN generating the probability distribution parameters has to predict both a scale and a mean instead of only a scale as before. This allows for a better modeling of the true underlying distribution and the paper shows that the increase in necessary side channel information is lesser than the improvement created by the improvement in probability distribution prediction. The second new idea is the addition of an autoregressive model over the latents of the first, main CNN. This also improves compression performance as the more structure from the latents can be exploited, it does however also increase the computational costs of the network as autoregressive models cannot be trained in parallel.

Another application of the hyperprior architecture is found in Cheng et.al. [41]. They improve on the hyperprior model including a joint autoregressive model given by Minnen et.al. [22] by introducing self-attention modules into the encoder and the decoder CNN. Self-attention modules are an idea taken from the field of Natural Language Processing (NLP). There it is used as an integral part of the novel transformer architecture, which yields state of the art results in machine translation and other NLP fields [42]. In computer vision, it has been used in order to generate attention masks that enable the model to use more bits for the more important parts of an image such as foreground or high-contrast elements and less bits to the less important parts of an image [43][44][45]. This approach of using self-attention modules is also the one used by Cheng et.al. They model the architecture of their self-attention blocks by using a simplified version of the architecture proposed in Liu et.al. [43]. The simplification is achieved by removing the non-local blocks, which reduces the training time by four times in their testing while achieving a similar loss [41].

3 Theoretical Foundations

This chapter describes the implementation of component X.

3.1 Convolutional Neural Networks

3.2 Transformer Models

3.3 Arithmetic Coding

3.4 Hyperprior Architecture

4 Methodology

4.1 The Combined Model

4.2 Spectral Autoencoder Methods

4.2.1 Pixel-Wise Convolutional Neural Network

4.2.2 Two-dimensional Spectral CNN

4.2.3 Spectral Transformer Model

4.3 Spatial Autoencoder Methods

4.3.1 CNN-based Spatial Autoencoder

4.3.2 Hyperprior-based Spatial Autoencoder

4.3.3 Attention-based Model Using Hyperprior Architecture

5 Experiments

This chapter describes the implementation of component X.

5.1 Description of the Data Set

5.2 Design of Experiments

5.3 Loss Functions and Metrics

5.3.1 MSE Loss

5.3.2 Rate Distortion Loss

5.3.3 Dual MSE Loss

5.3.4 Metrics

5.4 Results of the XX

5.5 Results of the YY

5.6 Results of the ZZ

6 Discussion and Conclusion

Bibliography

[1] M. Zea, A. Souza, Y. Yang, L. Lee, K. Nemali, and L. Hoagland, "Leveraging high-throughput hyperspectral imaging technology to detect cadmium stress in two leafy green crops and accelerate soil remediation efforts", *Environmental Pollution*, vol. 292, p. 118 405, Jan. 1, 2022, ISSN: 0269-7491. DOI: 10.1016/j.envpol.2021.118405.

[2] M. L. Henriksen, C. B. Karlsen, P. Klarskov, and M. Hinge, "Plastic classification via in-line hyperspectral camera analysis and unsupervised machine learning", *Vibrational Spectroscopy*, vol. 118, p. 103 329, Jan. 1, 2022, ISSN: 0924-2031. DOI: 10.1016/j.vibspec.2021.103329.

[3] L. Guanter, H. Kaufmann, K. Segl, S. Foerster, C. Rogass, S. Chabrillat, T. Kuester, A. Hollstein, G. Rossner, C. Chlebek, C. Straif, S. Fischer, S. Schrader, T. Storch, U. Heiden, A. Mueller, M. Bachmann, H. Mühle, R. Müller, M. Habermeyer, A. Ohndorf, J. Hill, H. Buddenbaum, P. Hostert, S. Van der Linden, P. J. Leitão, A. Rabe, R. Doerffer, H. Krasemann, H. Xi, W. Mauser, T. Hank, M. Locherer, M. Rast, K. Staenz, and B. Sang, "The EnMAP spaceborne imaging spectroscopy mission for earth observation", *Remote Sensing*, vol. 7, no. 7, pp. 8830–8857, Jul. 2015, Number: 7 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2072-4292. DOI: 10.3390/rs70708830.

[4] N. Bohn, B. Di Mauro, R. Colombo, D. R. Thompson, J. Susiluoto, N. Carmon, M. J. Turmon, and L. Guanter, "Glacier ice surface properties in south-west greenland ice sheet: First estimates from PRISMA imaging spectroscopy data", *Journal of Geophysical Research: Biogeosciences*, vol. 127, no. 3, e2021JG006718, 2022, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021JG006718, ISSN: 2169-8961. DOI: 10.1029/2021JG006718.

[5] A. B. Pascual-Venteo, E. Portalés, K. Berger, G. Tagliabue, J. L. Garcia, A. Pérez-Suay, J. P. Rivera-Caicedo, and J. Verrelst, "Prototyping crop traits retrieval models for CHIME: Dimensionality reduction strategies applied to PRISMA data", *Remote Sensing*, vol. 14, no. 10, p. 2448, Jan. 2022, Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2072-4292. DOI: 10.3390/rs14102448.

[6] V. Döpper, A. D. Rocha, K. Berger, T. Gränzig, J. Verrelst, B. Kleinschmit, and M. Förster, "Estimating soil moisture content under grassland with hyperspectral data using radiative transfer modelling and machine learning", *International Journal of Applied Earth Observation and Geoinformation*, vol. 110, p. 102 817, Jun. 1, 2022, ISSN: 1569-8432. DOI: 10.1016/j.jag.2022.102817.

[7] R. Loizzo, M. Daraio, R. Guarini, F. Longo, R. Lorusso, L. Dini, and E. Lopinto, "Prisma Mission Status and Perspective", in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, ISSN: 2153-7003, Jul. 2019, pp. 4503–4506. DOI: 10.1109/IGARSS.2019.8899272.

[8] R. Guarini, R. Loizzo, F. Longo, S. Mari, T. Scopa, and G. Varacalli, "Overview of the prisma space and ground segment and its hyperspectral products", in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, ISSN: 2153-7003, Jul. 2017, pp. 431–434. DOI: 10.1109/IGARSS.2017.8126986.

[9] R. Guarini, R. Loizzo, C. Facchinetti, F. Longo, B. Ponticelli, M. Faraci, M. Dami, M. Cosi, L. Amoruso, V. De Pasquale, N. Taggio, F. Santoro, P. Colandrea, E. Miotti, and W. Di Nicolantonio, "Prisma Hyperspectral Mission Products", in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, ISSN: 2153-7003, Jul. 2018, pp. 179–182. DOI: 10.1109/IGARSS.2018.8517785.

[10] N. Acito, M. Diani, and G. Corsini, "PRISMA Spatial Resolution Enhancement by Fusion With Sentinel-2 Data", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 62–79, 2022, Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, ISSN: 2151-1535. DOI: 10.1109/JSTARS.2021.3132135.

[11] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services", *Remote Sensing of Environment*, The Sentinel Missions - New Opportunities for Science, vol. 120, pp. 25–36, May 15, 2012, ISSN: 0034-4257. DOI: 10.1016/j.rse.2011.11.026.

[12] D. Spiller, S. Amici, and L. Ansalone, "Transfer Learning Analysis For Wildfire Segmentation Using Prisma Hyperspectral Imagery And Convolutional Neural Networks", in *2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, ISSN: 2158-6276, Sep. 2022, pp. 1–5. DOI: 10.1109/WHISPERS56178.2022.9955054.

[13] D. Spiller, K. Thangavel, S. T. Sasidharan, S. Amici, L. Ansalone, and R. Sabatini, "Wildfire segmentation analysis from edge computing for on-board real-time alerts using hyperspectral imagery", in *2022 IEEE International Conference on*

Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE), Oct. 2022, pp. 725–730. DOI: 10.1109/MetroXRAINE54828.2022.9967553.

[14] C. E. Shannon, "A mathematical theory of communication", *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948, ISSN: 1538-7305. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[15] D. J. C. MacKay, *Information theory, inference, and learning algorithms*. Cambridge, UK ; New York: Cambridge University Press, 2003, 628 pp., ISBN: 978-0-521-64298-9.

[16] F. Garcia-Vilchez, J. Munoz-Mari, M. Zortea, I. Blanes, V. Gonzalez-Ruiz, G. Camps-Valls, A. Plaza, and J. Serra-Sagrista, "On the impact of lossy compression on hyperspectral image classification and unmixing", *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 2, pp. 253–257, Mar. 2011, ISSN: 1545-598X, 1558-0571. DOI: 10.1109/LGRS.2010.2062484.

[17] E. Delp and O. Mitchell, "Image Compression Using Block Truncation Coding", *IEEE Transactions on Communications*, vol. 27, no. 9, pp. 1335–1342, Sep. 1979, Conference Name: IEEE Transactions on Communications, ISSN: 1558-0857. DOI: 10.1109/TCOM.1979.1094560.

[18] M. M. Hannuksela, J. Lainema, and V. K. Malamal Vadakital, "The High Efficiency Image File Format Standard [Standards in a Nutshell]", *IEEE Signal Processing Magazine*, vol. 32, no. 4, pp. 150–156, Jul. 2015, Conference Name: IEEE Signal Processing Magazine, ISSN: 1558-0792. DOI: 10.1109/MSP.2015.2419292.

[19] S. Ruder, *An overview of gradient descent optimization algorithms*, Jun. 15, 2017. DOI: 10.48550/arXiv.1609.04747. arXiv: 1609.04747[cs].

[20] J. Ballé, V. Laparra, and E. P. Simoncelli, *End-to-end Optimized Image Compression*, Number: arXiv:1611.01704, Mar. 3, 2017. DOI: 10.48550/arXiv.1611.01704. arXiv: 1611.01704[cs,math].

[21] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, *Variational image compression with a scale hyperprior*, Number: arXiv:1802.01436, May 1, 2018. DOI: 10.48550/arXiv.1802.01436. arXiv: 1802.01436[cs,eess,math].

[22] D. Minnen, J. Ballé, and G. D. Toderici, "Joint Autoregressive and Hierarchical Priors for Learned Image Compression", in *Advances in Neural Information Processing Systems*, vol. 31, Curran Associates, Inc., 2018.

[23] J. 1. Kuester, W. 1. Gross, W. 1. 1. F. I. Middelmann, G. Image Exploitation, E. Fraunhofer IOSB, and G. Image Exploitation, "1d-Convolutional Autoencoder Based Hyperspectral Data Compression", ISSN: 16821750 Num Pages: 15-21, Copernicus GmbH, 2021, pp. 15–21. DOI: 10.5194/isprs-archives-XLIII-B1-2021-15-2021.

[24] J. Kuester, W. Gross, S. Schreiner, M. Heizmann, and W. Middelmann, "Transferability of Convolutional Autoencoder Model For Lossy Compression to Unknown Hyperspectral Prisma Data", in *2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, ISSN: 2158-6276, Sep. 2022, pp. 1–5. DOI: 10.1109/WHISPERS56178.2022.9955109.

[25] R. La Grassa, C. Re, G. Cremonese, and I. Gallo, "Hyperspectral data compression using fully convolutional autoencoder", *Remote Sensing*, vol. 14, no. 10, p. 2472, Jan. 2022, Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2072-4292. DOI: 10.3390/rs14102472.

[26] Y. Guo, Y. Chong, Y. Ding, S. Pan, and X. Gu, "Learned hyperspectral compression using a student's t hyperprior", *Remote Sensing*, vol. 13, no. 21, p. 4390, Jan. 2021, Number: 21 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2072-4292. DOI: 10.3390/rs13214390.

[27] Y. Dua, V. Kumar, and R. S. Singh, "Comprehensive review of hyperspectral image compression algorithms", *Optical Engineering*, vol. 59, no. 9, p. 090 902, Sep. 2020, Publisher: SPIE, ISSN: 0091-3286, 1560-2303. DOI: 10.1117/1.OE.59.9.090902.

[28] A. Aidini, M. Giannopoulos, A. Pentari, K. Fotiadou, and P. Tsakalides, "Hyperspectral image compression and super-resolution using tensor decomposition learning", in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA: IEEE, Nov. 2019, pp. 1369–1373, ISBN: 978-1-72814-300-2. DOI: 10.1109/IEEECONF44664.2019.9048735.

[29] C. Deng, Y. Cen, and L. Zhang, "Learning-based hyperspectral imagery compression through generative neural networks", *Remote Sensing*, vol. 12, no. 21, p. 3657, Jan. 2020, Number: 21 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2072-4292. DOI: 10.3390/rs12213657.

[30] S. Kumar, S. Chaudhuri, B. Banerjee, and F. Ali, "Onboard hyperspectral image compression using compressed sensing and deep learning", in *Computer Vision – ECCV 2018 Workshops*, L. Leal-Taixé and S. Roth, Eds., vol. 11130, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2019, pp. 30–42, ISBN: 978-3-030-11011-6 978-3-030-11012-3. DOI: 10.1007/978-3-030-11012-3_3.

[31] J. A. Saghri, "Adaptive two-stage karhunen-loeve-transform scheme for spectral decorrelation in hyperspectral bandwidth compression", *Optical Engineering*, vol. 49, no. 5, p. 057 001, May 1, 2010, ISSN: 0091-3286. DOI: 10.1117/1.3425656.

[32] A. Karami, M. Yazdi, and G. Mercier, "Compression of Hyperspectral Images Using Discerete Wavelet Transform and Tucker Decomposition", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 444–450, Apr. 2012, Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, ISSN: 2151-1535. DOI: 10.1109/JSTARS.2012.2189200.

[33] M. Conoscenti, R. Coppola, and E. Magli, "Constant SNR, Rate Control, and Entropy Coding for Predictive Lossy Hyperspectral Image Compression", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7431–7441, Dec. 2016, Conference Name: IEEE Transactions on Geoscience and Remote Sensing, ISSN: 1558-0644. DOI: 10.1109/TGRS.2016.2603998.

[34] M. Hernández-Cabronero, A. B. Kiely, M. Klimesh, I. Blanes, J. Ligo, E. Magli, and J. Serra-Sagristà, "The CCSDS 123.0-B-2 "Low-Complexity Lossless and Near-Lossless Multispectral and Hyperspectral Image Compression" Standard: A comprehensive review", *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 4, pp. 102–119, Dec. 2021, Conference Name: IEEE Geoscience and Remote Sensing Magazine, ISSN: 2168-6831. DOI: 10.1109/MGRS.2020.3048443.

[35] T. Berger, "Rate-distortion theory", in *Wiley Encyclopedia of Telecommunications*, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471219282.eot142, John Wiley & Sons, Ltd, 2003, ISBN: 978-0-471-21928-6. DOI: 10.1002/0471219282.eot142.

[36] H. Shen, W. D. Pan, Y. Dong, and Z. Jiang, "Golomb-rice coding parameter learning using deep belief network for hyperspectral image compression", in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, TX: IEEE, Jul. 2017, pp. 2239–2242, ISBN: 978-1-5090-4951-6. DOI: 10.1109/IGARSS.2017.8127434.

[37] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "SpectralFormer: Rethinking Hyperspectral Image Classification with Transformers", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022, ISSN: 0196-2892, 1558-0644. DOI: 10.1109/TGRS.2021.3130716. arXiv: 2107.02988[cs].

[38] J. E. Fowler and J. T. Rucker, "Three-dimensional wavelet-based compression of hyperspectral imagery", in *Hyperspectral Data Exploitation*, Section: 14 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470124628.ch14, John Wi-

ley & Sons, Ltd, 2007, pp. 379–407, ISBN: 978-0-470-12462-8. DOI: 10.1002/9780470124628.ch14.

[39] Y. Hu, W. Yang, and J. Liu, "Coarse-to-fine hyper-prior modeling for learned image compression", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 11 013–11 020, Apr. 3, 2020, Number: 07, ISSN: 2374-3468. DOI: 10.1609/aaai.v34i07.6736.

[40] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images", 1999.

[41] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, *Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules*, Mar. 30, 2020. DOI: 10.48550/arXiv.2001.01568. arXiv: 2001.01568[eess].

[42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention Is All You Need*, Dec. 5, 2017. DOI: 10.48550/arXiv.1706.03762. arXiv: 1706.03762[cs].

[43] H. Liu, T. Chen, P. Guo, Q. Shen, X. Cao, Y. Wang, and Z. Ma, *Non-local Attention Optimized Deep Image Compression*, version: 1, Apr. 22, 2019. DOI: 10.48550/arXiv.1904.09757. arXiv: 1904.09757[cs,eess].

[44] M. Li, W. Zuo, S. Gu, D. Zhao, and D. Zhang, *Learning Convolutional Networks for Content-weighted Image Compression*, Sep. 19, 2017. DOI: 10.48550/arXiv.1703.10553. arXiv: 1703.10553[cs].

[45] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool, *Conditional Probability Models for Deep Image Compression*, Jun. 4, 2019. DOI: 10.48550/arXiv.1801.04260. arXiv: 1801.04260[cs].

Appendix A