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

. . . .

Name Surname

Acknowledgements

Abstract

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

Contents

List of Acronyms

3GPP	3rd Generation Partnership Project
AJAX	Asynchronous JavaScript and XML
API	Application Programming Interface
AS	Application Server
CSCF	Call Session Control Function
CSS	Cascading Stylesheets
DHTML	Dynamic HTML
DOM	Document Object Model
FOKUS	Fraunhofer Institut fuer offene Kommunikationssysteme
GUI	Graphical User Interface
GPS	Global Positioning System
GSM	Global System for Mobile Communication
HTML	Hypertext Markup Language
HSS	Home Subscriber Server
HTTP	Hypertext Transfer Protocol
I-CSCF	Interrogating-Call Session Control Function
IETF	Internet Engineering Task Force
IM	Instant Messaging
IMS	IP Multimedia Subsystem
IP	Internet Protocol
J2ME	Java Micro Edition
JDK	Java Developer Kit
JRE	Java Runtime Environment
JSON	JavaScript Object Notation
JSR	Java Specification Request
JVM	Java Virtual Machine
NGN	Next Generation Network

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Hyperspectral imaging, the technique of capturing images with a camera that has a spectrometer for every pixel of the image, is a quickly growing field. There are many applications for the use of these images as much information can be extracted from the combination of spatial and spectral data in them.

One such application is in agriculture where it is used 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 [1].
Hyperspectral imaging is also being used to separate plastic waste where a machine learning model based on hyperspectral images can separate twelve kinds of plastics better than previously used methods such as near-infrared technology [2].

Another field where the use of hyperspectral imaging is rising in importance is in geology and environmental sciences. An example of this is the EnMAP mission which launched a satellite into earths orbit in 2022 that takes images of the surface of the earth with a comparatively high spatial resolution, allowing for regional geographic analysis [3].
This 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.
The increasing usage of hyperspectral imaging makes it essential to address the disadvantes of the technology. One major disadvantage of this type of imaging is the size of the resulting images. Because there are much more spectral bands compared to the three bands of red, green and blue in traditional photography the resulting file size rises

accordingly. A hyperspectral image might for example have 300 bands resulting in an image that is a hundred times larger than an RGB image with the same pixel density.
In addition to this it is important to notice that even RGB images are too large for many use cases 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 normal image compression complexities and the added challenge of encoding information in the spectral dimension which RGB compression algorithms rarely address. The latter is one of the reasons why many RGB compression algorithms do not perform well on hyperspectral data. In addition to this there are more technical reasons for the difficulty of adapting algorithms for RGB compression that will be explored in this thesis.
For these reasons it is clear that it is important to find good 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. Lossless algorithms restore the compressed image exactly whereas lossy algorithms introduce some distortion during the compression. 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. Lossy compression methods can instead 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 performance. 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 image compression standard or the wavelet transform, often in combination with a lossless entropy coding method.
Learning-based approaches use the powerful capability of artificial neural networks to universally approximate arbitrary functions using gradient descentCITATION. 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.
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.

1.2 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 smaller latent dimension. Analogously, the decoder is trained to map 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 ratios 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. TODO besser formulieren, bzw. überlegen wie viel hier erklärt werden sollte zu hyperprior etc.

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.

1.3 Outline

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

Chapter 2 is called 'Related Work'. In this chapter the current state of research is described. This relates to the research of hyperspectral image compression by itself, however, as it is a recent and relatively unexplored field of study, the adjacent research topic of 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, 'Theoretical Foundations' 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, 'Methodology', 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, 'Implementation', gives a brief overview of the frameworks and libraries used in the implementation of the methods used in the thesis. It also describes the overall code architecture and structure. Furthermore the tools used for structuring and viewing the experiments are explained.

Chapter 6, 'Experiments', 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 7, 'Conclusion and discussion', 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, as explained before, is a developing field of study. While there are some 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. [7] shows the discrepance 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.
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 [8][9][10]. Another model proposed by Guo et.al. [11] uses the hyperprior architecture, originally developed by Ballé et.al. [12], which also uses convolutional layers in an ANN but combines them with an arithmetic coder to improve compression rate.
Some other models that will be explored later are an SVM-based model by Aidini et.al. [13], a GAN model by Deng et.al. [14] and a model using a simple multi-layer perceptron by Kumar et.al. [15].

2.1.1 CNN-based architectures

The CNN-based architectures can be split into two categories, the first being the models using two-dimensional convolutional layers to learn the spatial dependencies of the hyperspectral images [10]. The second category are models using one-dimensional convolutional layers to learn the spectral dependencies of the input data [8][9]. 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. [11] does use both spatial and spectral dependencies, it does however use a hyperprior architecture and not a purely CNN-based model.
As said before, comparing the results from these papers directly is difficult. The reason for this is that the models use different data sets since there is currently no accepted standard data set for hyperspectral imaging. Furthermore, the models use different compression rates. Since a higher compression rate also leads to a higher compression error for the same model as described by rate-distortion theory [16], this makes it impossible to directly compare the results from papers that use both a different data set and different compression rates.
Testing these models with a common data set revealed that for this dataset, the one-dimensional CNN model by Kuester et.al. [8] had the best performance of all the compared hyperspectral compression models, even including the non-CNN-based architectures, making it the current state of the art. This model is trained per pixel of the input image, making training slow but also resulting in good reconstruction accuracy.

2.1.2 Other architectures

An architecture different from the CNN-based models is proposed by Aidini et.al. [13]. They use quantization to compress the original image, meaning that the resolution of the image is simply decreased. Then an algorithm tries to recover the original tensor values by trying 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 would be very difficult to adapt to the neural network based methodolgies used in the thesis.
Similarly, other models based on non-ANN algorithms such as the work by Ülkü et.al. [17] also using dictionary-learning and Zikiou et.al. [18] using a support vector machine were not applicable for the methods used in the thesis.

Another category of models use ANNs to determine parameters for lossless compression algorithms. Shen et.al. [19] 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. [11]. 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. [12], 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. However, a disadvantage of their model is that it is developed only for data sets 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.

Another model using neural networks is proposed by Kumar et.al. [15]. 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. [20] propose an interesting 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, the same does not hold for image compression of traditional RGB images. There are many traditional compression algorithms, some of which are installed on every modern operating system and browser. Examples of these are 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.
However, while these algorithms are very popular, learning-based image compression has outperformed methods such as JPEG in both compression ratio and distortion. 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] for the International Conference on Learning Representations (ICLR) 2018. This paper builds on a previous work 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 [12]. 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 esti-

mated probability distribution it uses for the data that it compresses. This area is where improvements were found. Ballé et.al. 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 the basis or a component of a large portion of modern learned RGB image compression models and also one hyperspectral image compression model that was discussed in Chapter 2.1.

One such example is a paper by Hu et.al. [22] 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. [23]. The first improvement is a generalisation in the structure of the probalitiy distribution from the original scale mixture of gaussians [24] 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. [25].

2.2.2 Transformer-based architectures

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 Spectral Autoencoder Methods

4.1.1 Pixel-Wise Convolutional Neural Network

4.1.2 Two-dimensional CNN with Low Kernel Size

4.1.3 Spectral Transformer Model

4.2 Spatial Autoencoder Methods

4.2.1 CNN-based Spatial Autoencoder

4.2.2 Hyperprior-based Spatial Autoencoder

4.2.3 Attention-based Model Using Hyperprior Architecture

4.3 The Combined Model

4.4 Loss functions

4.4.1 MSE Loss

4.4.2 Rate Distortion Loss

4.4.3 Dual MSE Loss

5 Implementation

This chapter describes the implementation of component X.

5.0.1 Frameworks and libraries used

5.0.2 Model implementation structure

5.0.3 Metrics and data set implementation

5.0.4 Weights and Biases for experiments

6 Experiments

This chapter describes the implementation of component X.

6.1 Description of the Data Set

6.2 Design of Experiments

6.2.1 Metrics

6.3 Results of the XX

6.4 Results of the YY

6.5 Results of the ZZ

7 Conclusion and Discussion

Bibliography

[1] Maria Zea et al. "Leveraging high-throughput hyperspectral imaging technology to detect cadmium stress in two leafy green crops and accelerate soil remediation efforts". In: *Environmental Pollution* 292 (Jan. 1, 2022), p. 118405. ISSN: 0269-7491. DOI: 10.1016/j.envpol.2021.118405. URL: https://www.sciencedirect.com/science/article/pii/S0269749121019874 (visited on 04/02/2023).

[2] Martin L. Henriksen et al. "Plastic classification via in-line hyperspectral camera analysis and unsupervised machine learning". In: *Vibrational Spectroscopy* 118 (Jan. 1, 2022), p. 103329. ISSN: 0924-2031. DOI: 10.1016/j.vibspec.2021.103329. URL: https://www.sciencedirect.com/science/article/pii/S0924203121001247 (visited on 04/02/2023).

[3] Luis Guanter et al. "The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation". In: *Remote Sensing* 7.7 (July 2015). Number: 7 Publisher: Multidisciplinary Digital Publishing Institute, pp. 8830–8857. ISSN: 2072-4292. DOI: 10.3390/rs70708830. URL: https://www.mdpi.com/2072-4292/7/7/8830 (visited on 04/02/2023).

[4] Niklas Bohn et al. "Glacier Ice Surface Properties in South-West Greenland Ice Sheet: First Estimates From PRISMA Imaging Spectroscopy Data". In: *Journal of Geophysical Research: Biogeosciences* 127.3 (2022). _eprint: https://onlinelibrary.wiley.com/doi/pdf/ e2021JG006718. ISSN: 2169-8961. DOI: 10.1029/2021JG006718. URL: https://onlinelibrary.wiley.com/doi/abs/10.1029/2021JG006718 (visited on 04/02/2023).

[5] Ana B. Pascual-Venteo et al. "Prototyping Crop Traits Retrieval Models for CHIME: Dimensionality Reduction Strategies Applied to PRISMA Data". In: *Remote Sensing* 14.10 (Jan. 2022). Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, p. 2448. ISSN: 2072-4292. DOI: 10.3390/rs14102448. URL: https://www.mdpi.com/2072-4292/14/10/2448 (visited on 04/02/2023).

[6] Veronika Döpper et al. "Estimating soil moisture content under grassland with hyperspectral data using radiative transfer modelling and machine learning". In: *International Journal of Applied Earth Observation and Geoinformation* 110

(June 1, 2022), p. 102817. ISSN: 1569-8432. DOI: 10.1016/j.jag.2022.102817. URL: https://www.sciencedirect.com/science/article/pii/S156984322200019X (visited on 04/02/2023).

[7] Yaman Dua, Vinod Kumar, and Ravi Shankar Singh. “Comprehensive review of hyperspectral image compression algorithms”. In: *Optical Engineering* 59.9 (Sept. 2020). Publisher: SPIE, p. 090902. ISSN: 0091-3286, 1560-2303. DOI: 10.1117/1.OE.59.9.090902. URL: https://www.spiedigitallibrary.org/journals/optical-engineering/volume-59/issue-9/090902/Comprehensive-review-of-hyperspectral-image-compression-algorithms/10.1117/1.OE.59.9.090902.full (visited on 04/10/2023).

[8] J. 1 Kuester et al. “1d-Convolutional Autoencoder Based Hyperspectral Data Compression”. In: ISSN: 16821750 Num Pages: 15-21. Copernicus GmbH, 2021, pp. 15–21. DOI: 10.5194/isprs-archives-XLIII-B1-2021-15-2021. URL: https://www.proquest.com/docview/2585318626 (visited on 10/01/2022).

[9] Jannick Kuester et al. *Transferability of Convolutional Autoencoder Model for Lossy Compression to Unknown Hyperspectral PRISMA Data*. TU Berlin. 2022. URL: https://tubcloud.tu-berlin.de/s/dnAzTNsC5762HHp (visited on 10/01/2022).

[10] Riccardo La Grassa et al. “Hyperspectral Data Compression Using Fully Convolutional Autoencoder”. In: *Remote Sensing* 14.10 (Jan. 2022). Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, p. 2472. ISSN: 2072-4292. DOI: 10.3390/rs14102472. URL: https://www.mdpi.com/2072-4292/14/10/2472 (visited on 10/01/2022).

[11] Yuanyuan Guo et al. “Learned Hyperspectral Compression Using a Student’s T Hyperprior”. In: *Remote Sensing* 13.21 (Jan. 2021). Number: 21 Publisher: Multidisciplinary Digital Publishing Institute, p. 4390. ISSN: 2072-4292. DOI: 10.3390/rs13214390. URL: https://www.mdpi.com/2072-4292/13/21/4390 (visited on 10/01/2022).

[12] Johannes Ballé, Valero Laparra, and Eero 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]. URL: http://arxiv.org/abs/1611.01704 (visited on 10/01/2022).

[13] A. Aidini et al. “Hyperspectral Image Compression and Super-Resolution Using Tensor Decomposition Learning”. In: *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. 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. URL: https://ieeexplore.ieee.org/document/9048735/ (visited on 04/10/2023).

[14] Chubo Deng, Yi Cen, and Lifu Zhang. "Learning-Based Hyperspectral Imagery Compression through Generative Neural Networks". In: *Remote Sensing* 12.21 (Jan. 2020). Number: 21 Publisher: Multidisciplinary Digital Publishing Institute, p. 3657. ISSN: 2072-4292. DOI: 10.3390/rs12213657. URL: https://www.mdpi.com/2072-4292/12/21/3657 (visited on 04/10/2023).

[15] Saurabh Kumar et al. "Onboard Hyperspectral Image Compression Using Compressed Sensing and Deep Learning". In: *Computer Vision – ECCV 2018 Workshops*. Ed. by Laura Leal-Taixé and Stefan Roth. 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. URL: https://link.springer.com/10.1007/978-3-030-11012-3_3 (visited on 04/10/2023).

[16] Toby 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. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/0471219282.eot142 (visited on 04/10/2023).

[17] İrem Ülkü and Ersin Kizgut. "Large-scale hyperspectral image compression via sparse representations based on online learning". In: *International Journal of Applied Mathematics and Computer Science* 28.1 (Mar. 1, 2018), pp. 197–207. ISSN: 2083-8492. DOI: 10.2478/amcs-2018-0015. URL: https://www.sciendo.com/article/10.2478/amcs-2018-0015 (visited on 04/10/2023).

[18] Nadia Zikiou, Mourad Lahdir, and David Helbert. "Support vector regression-based 3D-wavelet texture learning for hyperspectral image compression". In: *The Visual Computer* 36.7 (July 2020), pp. 1473–1490. ISSN: 0178-2789, 1432-2315. DOI: 10.1007/s00371-019-01753-z. URL: http://link.springer.com/10.1007/s00371-019-01753-z (visited on 04/10/2023).

[19] Hongda Shen et al. "Golomb-Rice coding parameter learning using deep belief network for hyperspectral image compression". In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth, TX: IEEE, July 2017, pp. 2239–2242. ISBN: 978-1-5090-4951-6. DOI: 10.1109/IGARSS.2017.8127434. URL: http://ieeexplore.ieee.org/document/8127434/ (visited on 04/10/2023).

[20] Danfeng Hong et al. "SpectralFormer: Rethinking Hyperspectral Image Classification with Transformers". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–15. ISSN: 0196-2892, 1558-0644. DOI: 10.1109/TGRS.2021.3130716. arXiv: 2107.02988[cs]. URL: http://arxiv.org/abs/2107.02988 (visited on 10/01/2022).

[21] Johannes Ballé et al. *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]. URL: http://arxiv.org/abs/1802.01436 (visited on 10/01/2022).

[22] Yueyu Hu, Wenhan Yang, and Jiaying Liu. "Coarse-to-Fine Hyper-Prior Modeling for Learned Image Compression". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.7 (Apr. 3, 2020). Number: 07, pp. 11013–11020. ISSN: 2374-3468. DOI: 10.1609/aaai.v34i07.6736. URL: https://ojs.aaai.org/index.php/AAAI/article/view/6736 (visited on 04/10/2023).

[23] David Minnen, Johannes Ballé, and George D Toderici. "Joint Autoregressive and Hierarchical Priors for Learned Image Compression". In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., 2018. URL: https://proceedings.neurips.cc/paper/2018/hash/53edebc543333dfbf7c5933af792 Abstract.html (visited on 10/01/2022).

[24] Martin J Wainwright and Eero P Simoncelli. "Scale Mixtures of Gaussians and the Statistics of Natural Images". In: (1999).

[25] Zhengxue Cheng et al. *Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules*. Mar. 30, 2020. DOI: 10.48550/arXiv.2001.01568. arXiv: 2001.01568[eess]. URL: http://arxiv.org/abs/2001.01568 (visited on 04/16/2023).

Appendix