ONBOARD HYPERSPECTRAL IMAGE COMPRESSION WITH DEEP LINE-BASED PREDICTIVE ARCHITECTURES

Diego Valsesia, Tiziano Bianchi, Enrico Magli

[(1)]***Politecnico di Torino***
Department of Electronics and Telecommunications
C.so Duca degli Abruzzi 24
Torino, Italy
Email:diego.valsesia@polito.it

ABSTRACT

AI-driven onboard compression of hyperspectral images remains a challenging problem due to the need to balance the high computational requirements of neural networks and their representational capability. So far, autoencoder-based approaches have dominated the literature and offer interesting performance at very low bitrates. However, they do not scale well into the high-quality, high-bitrate regimes, requiring large amounts of compute and memory and are significantly outperformed by classic and simpler approaches such as CCSDS-123. In this paper, we present a novel approach towards AI-driven onboard compression, consisting of two fundamental ingredients: i) neural predictive coding; ii) line-based architectures. With the former, we depart from the autoencoder approach to develop a neural network that predicts the value of a pixel based on a causal spatial-spectral context of past pixels. As in classic predictive coding, this allows to easily control maximum distortion and is particularly suited to the high-quality regime. While in-loop quantization is possible, the use of a prequantizer followed by lossless predictive coding offers the higher throughput at minimal rate-distortion suboptimality. The other ingredient of our approach seeks to limit computational requirements and, in particular, memory usage by developing a neural network architecture that can process one line with all its spectral bands at a time. Several mechanisms for sequence processing can be used to achieve this goal such as recurrent neural networks. However, we show that a recent hybrid recurrent-attentive operation can overcome the limitations of recurrent neural networks and offer Transformer-like performance with no need to store the whole line sequence in memory. This approach greatly limits memory requirements and potentially allows for continuous operation with pushbroom sensors. Experiments show that the proposed approach outperforms CCSDS-123 in lossless and near-lossless compression.

INTRODUCTION

Hyperspectral satellite imagery is crucial for various Earth observation tasks, as it captures a wide range of wavelengths across the electromagnetic spectrum, but it generates very large amounts of data, making efficient onboard compression essential for space missions. Traditional methods often rely on either transform-based or predictive coding, with the CCSDS-123 standard being a notable example of the latter, using an adaptive filter for pixel prediction. Recently, AI-driven compression techniques [1,2,3,4,5,6] have demonstrated great success with standard images and videos, but applying these methods to hyperspectral data faces challenges. The complexity of neural network-based approaches, particularly in terms of memory and processing power, often renders them impractical for satellite applications.

Most neural network compression techniques use autoencoder designs, where the encoder reduces the image to a compact latent space, which can be encoded and later decoded by the decoder network. While effective, this approach is resource-intensive when handling large hyperspectral data cubes. To address these limitations, this paper introduces a new method centered on a specialized neural network architecture. This model adopts a line-by-line processing approach to limit computational and memory requirements, a feature crucial for onboard applications. Drawing inspiration from the RWKV [7] model, which recently appeared for language processing, it blends aspects of recurrent networks and attention mechanisms. This design allows for efficient training while maintaining a low-memory inference process, positioning it as a promising solution for hyperspectral image compression. Preliminary results show that it surpasses the performance of the CCSDS-123 standard.

BACKGROUND

Compressing hyperspectral images onboard spacecraft presents a difficult challenge in balancing between rate-distortion efficiency and computational cost. The CCSDS standards have gained popularity due to their low computational complexity, facilitating high throughput on dedicated FPGA hardware and delivering good rate-distortion results. Specifically, CCSDS 122.0-B-2 and CCSDS 122.1-B-1 use transform coding methods for both lossless and lossy compression, combining a 2D discrete wavelet transform with spectral transformation. More recently, CCSDS 123.0-B-2 [8] has adopted a predictive coding method for lossless and near-lossless compression, utilizing a spatial-spectral predictor with an adaptive filter.

In the realm of deep learning, most research has focused on compressing 8-bit RGB or grayscale images. Only a few studies have addressed hyperspectral image compression, especially in the context of onboard applications where complexity is a major constraint. These studies often follow the design principles of Ballé et al [1]., which employ autoencoder networks. In these networks, an encoder creates a compressed latent representation of the image, which is then quantized and encoded using entropy. The decoder reconstructs the image from this latent space. These networks are trained using a rate-distortion approach to balance between reconstruction accuracy and compression efficiency. For hyperspectral images, however, the primary challenge lies in the high memory demands required to capture 3D spatial-spectral features. Some researchers have attempted to mitigate this by designing spectral autoencoders that reduce complexity while capturing spectral correlations [9], though they miss out on exploiting spatial redundancies. Others use spatio-spectral features but encounter scalability issues at high rates, where deep learning methods typically struggle to achieve very low distortion without significantly increasing complexity [10].

PROPOSED METHOD

A high-level overview of the proposed process is described in Figure 1. Essentially, a model operates as a non-linear predictor that compresses hyperspectral images by predicting pixel values from their spatial and spectral context. The goal is to predict pixel values using information from previous lines and bands, and then compress the prediction errors through entropy coding.

A key feature of the proposed method is its efficiency in memory use during inference. The model processes one line of pixels at a time, with all its spectral bands, to reduce memory consumption. This design is well-suited for satellite systems that have limited computational resources. The model uses an efficient memory mechanism to retain information about previous lines, which is critical for handling the spatial and spectral correlations in hyperspectral images.
In particular, we combine ideas from both recurrent neural networks (RNNs) and transformers. While RNNs excel in memory efficiency, they suffer from slow training due to their sequential nature. On the other hand, transformers have better representational power and can be trained in parallel but require significant memory, as they need to explicitly keep the past context. The proposed model utilizes the RWKV operation, a hybrid design from natural language processing, which allows for parallel training like transformers but maintains the memory-efficient, recurrent structure for inference.

In order to perform a causal non-linear prediction for compression of hyperspectral images, the proposed model includes several neural network components: a spatial encoder, a line predictor, a spectral predictor, and a pixel decoder, which are described in the following.

The spatial encoder captures spatial information from the previous line of pixels using 1D convolutional layers, producing a set of features for each pixel. This allows the model to capture the correlation between image columns and should be designed according to the expected inter-column redundancy patterns. A simple design for the spatial encoder is a sequence of 1D convolution, batch normalization and non-linear activation.

The line predictor then uses the features output for each pixel by the spatial encoder to predict the feature values of the pixels in the current line. This operation requires a memory mechanism in order to exploit information from lines older that last one, thus exploiting longer-range redundancies. This is where the RWKV operation is used. It consists o a sequence of “time-mixing” and “channel-mixing” layers to merge features of past lines in a recursive manner, with an operation that resembles the Transformer attention, while having linear complexity in the number of lines.

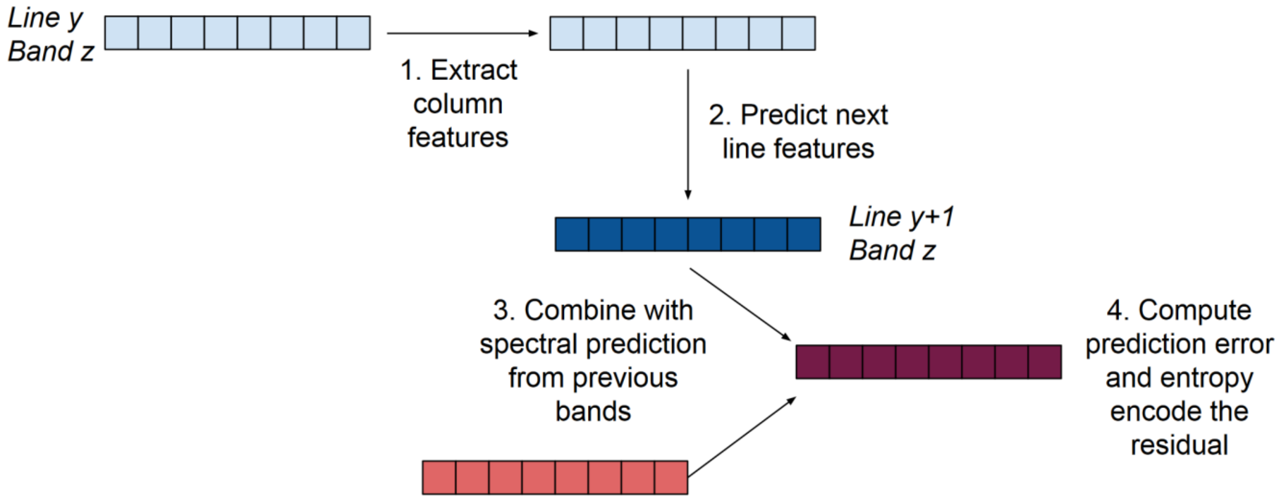

Figure 1. High-level overview of proposed scheme. A line at a time is processed. First, column features are extracted by neural network operating inside the line, then these features are further processed to predict the features of the next line. These are then combined with features derived from previous bands for the same line to obtain feature that can be decoded into the prediction for the line. The prediction residual is then entropy encoded via conventional methods.

The spectral predictor handles the spectral dimension by predicting pixels across different bands. In particular, the error between the predicted features of the current line as output by the line predictor and the spatial encoding of the true line pixels, forms a sequence of features along the spectral dimensions. A causal model is used to process this sequence, which could be a RNN, causal convolution, or again RWKV. The output is a feature vector per pixel, whose value ultimately depends on the pixels in the previous lines for the same band, and the pixels in previous bands on the same line. Notice that contrary to many predictor designs, such as CCSDS-123, the pixels in the same band and in the same line on the left of the current pixels are not used for prediction, as handling them would make the design of the neural architecture substantially more complex. Although this leaves some potential performance on the table.

Finally, the pixel decoder finally reconstructs the pixel values using the output of the spectral and spatial predictors. This can simply be one or more 1x1 convolutional layers interleaved with non-linearities. The decoded pixel value is the prediction performed by the model. The residual is then entropy coded with any standard technique.

Some special cases need to handled independently in order to ensure causality of the model. In particular, for the first line and the first spectral band of the image, a simpler differential pulse-code modulation (DPCM) technique is applied since no prior context is available for prediction. Subsequent lines and bands are encoded using the more advanced predictor.

During training, the L1 loss between the model prediction and the true pixel values is computed to update the model parameters. Notice that the model returns a real-valued prediction. At inference time, the predicted values are rounded to compute integer residuals, which are then entropy-coded for compression. The model supports both lossless and near-lossless compression, with the latter achieved through prequantization, making it a flexible option for various use cases. Although the primary focus is on lossless compression, further optimizations could be explored to improve near-lossless performance.

RESULTS

In this section, the paper presents experimental comparisons between the proposed model and the widely used CCSDS-123 standard for onboard hyperspectral image compression, as well as a recent AI-based compression method based on autoencoders (1D-CAE [9]).

The experiments are conducted using the HySpecNet-11k dataset [11], the largest available hyperspectral image dataset, consisting of 11,483 patches, each of size 128×128×202, captured by the EnMAP satellite. A predefined train-test split is used to ensure that the test patches come from different regions than those used for training.

The model used in these experiments consists of a spatial encoder with two convolutional layers, a line predictor built with RWKV blocks, a spectral predictor using causal convolutional layers, and a pixel decoder with one 1x1 convolutional layer. The model has around 110,000 trainable parameters. It was trained for approximately 2,000 epoch.

The results show that the proposed method achieves a noticeable reduction in the lossless compression rate compared to CCSDS-123. Specifically, it attains a rate of 5.593 bits per pixel (bpp), which is 0.208 bpp lower than the 5.801 bpp of CCSDS-123, demonstrating its superior prediction capability. This improvement is achieved while using the same sample-adaptive Golomb entropy encoder for fair comparison. We remark that the 1D-CAE AI method is not capable of lossless compression.

Rate distortion performance is shown in Figure 2, where it is clear that the proposed method outperforms CCSDS-123 and is significantly better than the existing autoencoder-based method. This can be attributed to the difficulty in scaling autoencoder methods to high quality levels.

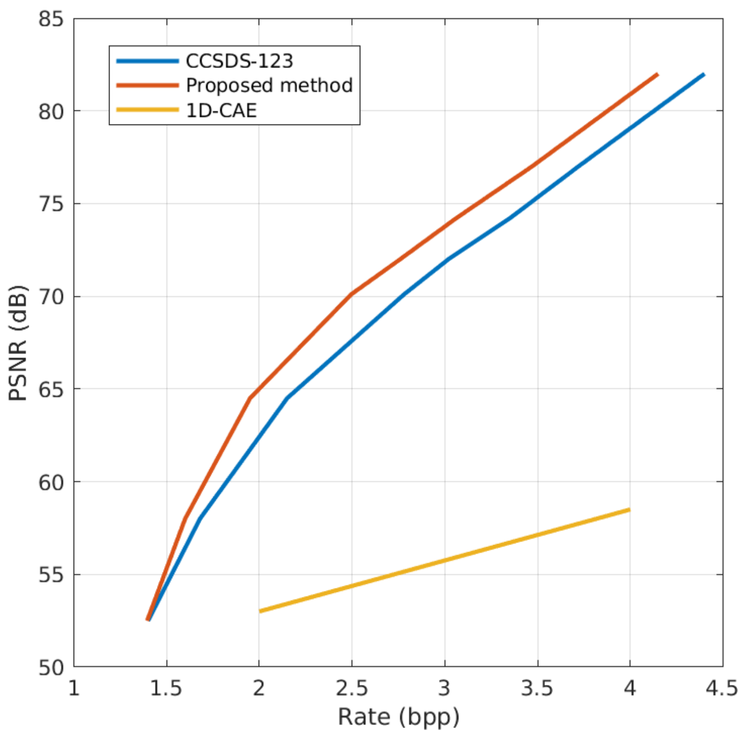

Figure 2: Rate-distortion performance.

CONCLUSIONS

We have shown a preliminary design of an AI-based compression method for hyperspectral images that is capable of outperforming CCSDS-123 while limiting computational complexity thanks to the line-based design.

ACKNOWLEDGEMENT

This study was carried out within the FAIR – Future Artificial Intelligence Research and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1555 11/10/2022, PE00000013). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

REFERENCES

[1] J. Balle, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," arXiv preprint arXiv:1611.01704, 2016
[2] J. Balle, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," in ICLR, 2018
[3] L. Zhou, C. Cai, Y. Gao, S. Su, and J. Wu, "Variational autoencoder for low bit-rate image compression," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018
[4] W. Jiang, J. Yang, Y. Zhai, P. Ning, F. Gao, and R. Wang, "MLIC: Multi-reference entropy model for learned image compression," in Proceedings of the 31st ACM International Conference on Multimedia, 2023
[5] J. Balle, P. A. Chou, D. Minnen, S. Singh, N. Johnston, E. Agustsson, S. J. Hwang, and G. Toderici, "Nonlinear transform coding," IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 2, 2020
[6] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. V. Gool, "Practical full resolution learned lossless image compression," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019
[7] B. Peng, E. Alcaide, Q. Anthony, A. Albalak, S. Arcadinho, H. Cao, X. Cheng, M. Chung, M. Grella, K. K. GV et al., "RWKV: Reinventing RNNs for the Transformer Era," arXiv preprint arXiv:2305.13048, 2023
[8] Consultative Committee for Space Data Systems (CCSDS), "Low-Complexity Lossless and Near-Lossless Multispectral and Hyperspectral Image Compression," Blue Book, no. 2, 2019
[9] J. Kuester, W. Gross, and W. Middelmann, "1d-convolutional autoencoder based hyperspectral data compression," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 43, 2021
[10] S. Mijares i Verdu, J. Balle, V. Laparra, J. Bartrina-Rapesta, M. Hernandez-Cabronero, and J. Serra-Sagrista, "A scalable reduced-complexity compression of hyperspectral remote sensing images using deep learning," Remote Sensing, vol. 15, no. 18, 2023
[11] M. H. P. Fuchs and B. Demir, "Hyspecnet-11k: A large-scale hyperspectral dataset for benchmarking learning-based hyperspectral image compression methods," arXiv preprint arXiv:2306.00385, 2023