Towards Deep Line-Based Architectures for Onboard Hyperspectral Image Super-Resolution

TOWARDS DEEP LINE-BASED ARCHITECTURES FOR ONBOARD HYPERSPECTRAL IMAGE SUPER-RESOLUTION

Davide Piccinini
Politecnico di Torino
Torino, Italy
davide.piccinini@polito.it

Diego Valsesia
Politecnico di Torino
Torino, Italy
diego.valsesia@polito.it

Enrico Magli
Politecnico di Torino
Torino, Italy
enrico.magli@polito.it

***Abstract*—Hyperspectral images can provide the fine spectral resolution needed for several tasks involving material identification, but they suffer from limited spatial resolution due to instrument constraints. The emerging paradigm of onboard computing seeks to address several detection problems of interest directly onboard of a satellite, in order to minimize latency due to the downlink and processing chain. In this paper, we design a novel lightweight architecture for onboard spatial super-resolution of hyperspectral images with the goal of improving onboard detection methods that rely on such images. The core of our contribution is a novel neural network architecture that works in a line-by-line fashion thanks to a recurrent-attentive mechanism in the along-track direction. This design greatly limits the memory requirements and computational complexity of the model, making it suitable for onboard usage. We show that this novel architecture is competitive with respect to the state of the art in term of super-resolution quality, while providing significant savings in complexity and memory.**

***Index Terms*—Super-resolution, onboard processing, hyperspectral images.**

I. Introduction

Hyperspectral imagery is increasingly more important in remote sensing thanks to the fine spectral resolution available. This enables precise material identification in tasks for environmental monitoring, urban planning, military, and more. However, the price to be paid is in terms of coarser spatial resolution compared to multispectral or panchromatic imagers, which may limit detection capabilities from hyperspectral images. For this reason, spatial super-resolution of hyperspectral images is a topic of great interest for the recent literature [1] , [2], [3], [4], particularly using deep neural networks.

At the same time, a recent trend in remote sensing is to study the potential of performing image analysis tasks onboard of satellites [5], [6], [7], [8]. This would allow to minimize detection latency in case of events of particular interest, such as natural disasters, without having to wait for download to a ground segment and for the completion of the processing pipeline. With this objective in mind, super-resolution is also a task that would be desirable to move onboard in order to preprocess hyperspectral images before using them for detection problems, which would benefit from the higher spatial resolution.

At the moment, the recent SOTA literature on hyperspectral super-resolution [1], [2], [3] is focused on maximizing image quality, at the expense of high computational complexity, and designing models suitable for onboard usage is still a largely unexplored area of research. In doing so, a point of attention is the memory requirements of such methods, since hyperspectral data cubes can be large due to the presence of possibly hundreds of spectral channels, and the super-resolution operation itself seeks to further increase their size.

In this paper, we investigate a novel design of a deep neural network for onboard single-image super-resolution of hyperspectral images. The core idea is to limit memory and complexity requirements by working one line, with all the spectral channels, at a time. A memory mechanism based on recently proposed deep state-space models, such as Mamba [9], allows to exploit a rich receptive field of past lines to super-resolve the current line. This allows significant memory savings since only features from the current line and an internal state are required for the super-resolution operation, compared to features from the full image needed by state-of-the-art architectures. The approach is also an excellent match to the acquisition mechanism of pushbroom sensors, possibly allowing direct pipelining of the super-resolution module.

We can summarize our contributions as follows:

- We propose a low-complexity neural network for hyperspectral single-image super-resolution, suitable for usage onboard of satellites.
- Our model mimics the line-by-line acquisition of pushbroom scanners, by super-resolving the image in a causal line-by-line fashion, exploiting deep Selective State Space Models (SSMs) to effectively model a memory of past lines.
- The proposed model is close in image quality to state-of-the-art models, at a fraction of memory and computational requirements.

II. Background

A. Super-resolution of hyperspectral images

The hyperspectral image super-resolution task is among the most critical challenges in remote sensing because it enables a finer inspection of satellite imagery, which often suffers from limited spatial resolution. Early methods addressed this task using a variety of techniques, such as low-rank and sparse modeling [10], spectral-spatial sub-pixel mapping [11], and low-rank tensor modeling combined with total variation regularization [12].

In recent years, with the availability of larger datasets, deep neural networks have revolutionized hyperspectral image super-resolution, delivering significantly improved performance. The initial wave of neural network-based models employed convolutional neural networks (CNNs) [13], [14], leveraging their capacity to extract and utilize the rich spectral information present in the numerous channels. These models demonstrated substantial promise, as confirmed also with more recent works [15], [16]. The rise of transformer architectures led to attention mechanisms being employed for hyperspectral super-resolution [17]. Hybrid models combining transformers and CNNs [18], [19], [1], [3] have recently emerged, achieving state-of-the-art results. These exploit the strengths of CNNs for local feature extraction along with an attention mechanisms for capturing long-range dependencies, marking a significant evolution in hyperspectral image super-resolution techniques.

B. State Space Models and Mamba

SSMs have gained popularity in the last few years thanks to their ability to handle sequential data; they are based on the following two state-space equations that define a sequence-to-sequence transformation from $\mathbf{x}(t) \in \mathbb{R}^M$ to $\mathbf{y}(t) \in \mathbb{R}^O$ through an implicit latent state $\mathbf{h}(t) \in \mathbb{R}^N$:

$$\mathbf{h}'(t) = \mathbf{A}\mathbf{h}(t) + \mathbf{B}\mathbf{x}(t), \tag{1}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{h}(t) + \mathbf{D}\mathbf{x}(t), \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ and $\mathbf{C} \in \mathbb{R}^{O\times N}$ control how the current state evolves over time and translates to the output, $\mathbf{B} \in \mathbb{R}^{N\times M}$ and $\mathbf{D} \in \mathbb{R}^{O\times M}$ describe how the input influences the state and the output, respectively. In [20] the authors showed that imposing specific structures on the SSM parameters, e.g., initializing $\mathbf{A}$ as diagonal, yields promising results.
Following [21], [22] and [23], Gu et al. [9] proved that letting the SSM parameters be functions of the input greatly improves performance because it allows the model to learn a selectivity mechanism to propagate or forget information along the sequence length dimension.

Furthermore, they incorporated the aforementioned Selective SSMs into a new neural block called Mamba Block, which was then used as the fundamental backbone of architectures for audio and language processing. A Mamba Block is composed of linear, convolutional and activation layers that wrap the Selective SSM. Mamba showed promising results to replace Transformers [24], critically thanks to linear complexity scaling with sequence length, compared to the quadratic bottleneck of Transformers. So far, Mamba has been successfully applied to detection problems in vision [25, 26], where a multi-directional scan of the flattened input was fed into the Selective SSM, and hyperspectral image classification [27, 28].

However, to the best of our knowledge, Mamba has not yet been used to process images in a line-by-line fashion. A similar idea was exploited for the task of hyperspectral image compression [29], but using RWKV [22], a different attentive-recursive mechanism.

III. Proposed Method

In this section, we present the methodology used in order to tackle the problem of onboard single-image super-resolution and the structure of our line-by-line architecture, named LineSR. A high-level overview of the concept is given in Fig. 1, while a detail of the LineSR neural network architecture is shown in Fig. 2.

A. Preliminaries and Proposed Strategy

The overall design goal is to super-resolve an image line-by-line in the along-track direction, to match a pushbroom acquisition system and minimize memory and computational requirements. We first need note that super-resolving a line, with all its spectral channels, of size $1 \times W \times C$ by a factor r produces an output of size $r \times rW \times C$, i.e., r lines in the along-track direction, expanded by a factor of r in the across-track direction.

More formally, we suppose to have the current low-resolution (LR) line $\mathbf{x}_y^{\text{LR}} \in \mathbb{R}^{W\times C}$ corresponding to along-track LR line location y and a memory of past LR lines encoded in a latent state and as features of a small number of previous LR lines. From this, we estimate $\{\mathbf{x}^{\text{SR}}_{r(y-1)}, \mathbf{x}^{\text{SR}}_{r(y-1)-1}, ..., \mathbf{x}^{\text{SR}}_{r(y-1)-(r-1)}\}$ with $y = 1, ...,$ i.e., the set of r super-resolved lines at the HR along-track spatial locations between the LR spatial locations $y-1$ (included) and y (excluded). It is important to remark that we are framing the problem with an *interpolation* setting, where line y is used to predict HR locations strictly preceding it. This is in contrast with an *extrapolation* setting which would predict spatial locations beyond y, and would be more challenging leading to worse image quality.

We remark that this setting is constraining the model to work with only past lines without having access to future lines, in contrast to all state-of-the-art models that have access to the entire image. This is a deliberate design tradeoff. In fact, access to future lines may allow the model to exploit a richer context, as well as self-similar patterns, to enhance image quality. However, the line-based approach is desirable for low-complexity and low-latency systems as it limits memory requirements and follows the pushbroom acquisition of the image, which at the time to process line y has not yet acquired future lines.

B. LineSR architecture

The primary challenge in designing an architecture for the described line-based image processing system is the potential

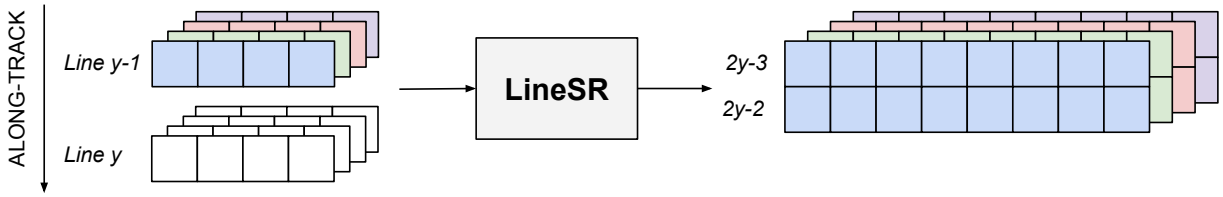

Fig. 1. LineSR super-resolves a hyperspectral image line-by-line, by using line $y-1$ and line y, as well as a memory of past lines to super-resolve line $y-1$ in the along- and across-track directions.

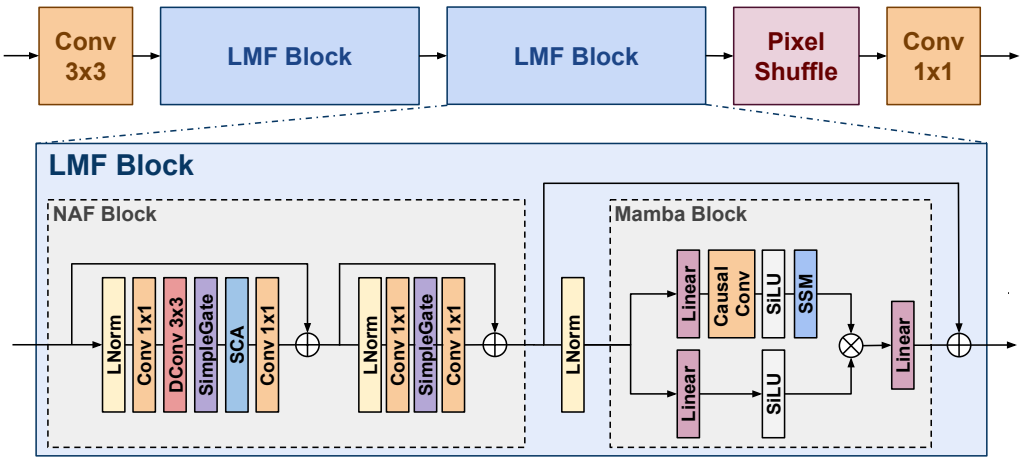

Fig. 2. LineSR processes a line with all its spectral channels with a sequence of LMF blocks composed of NAFBlocks to exact features from the across-track and spectral dimensions and Mamba Blocks to carry over memory of previous lines. 2D operations like convolution are over the across-track and spectral dimensions.

loss of detailed spatial context. At the same time, when considering an onboard environment with severely limited resources in terms of both computational power and memory capacity, the size of the model and its complexity become critical factors that must be carefully addressed. These considerations were a significant and integral part of the architectural design process, ensuring that the model is both effective and capable of operating within the given constraints. An overview of the architecture is presented in Fig. 2.

First of all, we decouple the along-track dimension from the across-track and spectral dimensions by extracting the line x_y^{LR} including all its spectral channels. The latter two are then treated as two "spatial" dimensions in order to extract joint spatial-spectral features via convolution-like layers. The sequence of spatial-spectral features corresponding to multiple lines is instead represented by Selective SSMs, i.e., Mamba [9]. This leads to a structure having an initial convolutional layer to extract shallow spatial-spectral features of a line, followed by a backbone composed of two Line-Features Mixer Blocks (LMFBs) and, finally, an upsampler module to output the lines at the chosen resolution. For LR along-track location y, the input $\mathbf{x}_y^{\text{LR}}$ line is fed to the network, where it then follows this pipeline:

$$\mathbf{z}_y = \text{Conv}(\mathbf{x}_y^{\text{LR}}), \tag{3}$$

$$\mathbf{w}_y = \text{LMFB}(\text{LMFB}(\mathbf{z}_y)), \tag{4}$$

$$\mathbf{x}^{\text{SR}}_{r(y-1)}, ..., \mathbf{x}^{\text{SR}}_{r(y-1)-(r-1)} = \text{Upsampler}(\mathbf{w}_y). \tag{5}$$

We remark that the initial 3×3 convolution works on the across-track and spectral dimensions, leading to $\mathbf{h}_y \in \mathbb{R}^{W \times C \times F}$, being F the desired number of features. The core modules of LineSR are the two LMFBs, which are responsible for both extracting rich joint features in the across-track and spectral dimensions, and modeling their evolution in the along-track direction, for each pixel. Therefore, each LMFB has been designed as the cascade of two sub-blocks, addressing the two tasks, respectively. In particular a NAFBlock extracts rich features from the across-track and spectral dimensions, then followed by a Mamba Block to model inter-line evolution:

$$\text{LMFB}(\mathbf{z}) = \text{MambaBlock}(\text{NAFB}(\mathbf{z})). \tag{6}$$

The NAFBlock uses a sequence of layers inspired by the building block of the NAFNet architecture [30], which provides near state-of-the-art performance for image restoration problems. In particular, this block uses a SimpleGate and a Simplified Channel Attention operation to implement an attention-like mechanism of input-dependent feature extraction, while keeping a lightweight design. We refer the reader to [30] for more details on these operations.

The decision to incorporate the Mamba Block is primarily driven by the requirement to process one line at a time. The output of the NAFBlock consists of one feature vector for each of the WC pixels in the line, which can be regarded

TABLE I
QUANTITATIVE COMPARISON OF DIFFERENT METHODS ON THE HYSPECNET-11K DATASET. 4× SUPER-RESOLUTION.

Method	MPSNR (dB)↑	SAM↓	MSSIM↑	RMSE↓	Params
Bicubic	39.34	4.1767	0.9005	0.0173	
GDRRN [13]	40.02	4.0718	0.9156	0.0156	612.86 K
MSDformer [1]	41.44	3.5992	0.9397	0.0126	16.49 M
LineSR	40.21	3.7685	0.9168	0.0155	95.11 K

TABLE II
MEMORY AND COMPLEXITY COMPARISON FOR INPUT SIZE (202,256,256).

Model	Memory	FLOPs/pixel
MSDformer [1]	8203 MiB	617.9 K
GDRRN [13]	4987 MiB	284.0 K
LineSR	1772 MiB	165.5 K

as a token whose evolution is modeled over the along-track direction by means of Mamba. Critically, Mamba uses the Selective SSM formulation introduced in Sec. II-B to process the features based on the latent state which models a memory of past lines. The latent state is a vector of length N for each feature of the line pixels. Moreover, we remark that in addition to the SSM, the Mamba Block leverages a causal convolution operation, also in the along-track direction, to further process the sequence before feeding it to the SSM. This means that the Mamba Block requires to store: i) its latent state for each feature of each pixel, i.e. a tensor of size $W \times C \times F \times N$ and ii) the receptive field of the causal convolution operation, consisting of feature maps of a few past lines, i.e., a tensor of size $K-1 \times W \times C \times F$, being K the kernel size.

IV. EXPERIMENTAL RESULTS

A. Experimental setting

All experiments have been conducted on the HySpecNet-11k dataset, a large-scale hyperspectral benchmark dataset consisting of 11,483 non-overlapping image patches. Each patch has 128×128 pixels, with 202 usable spectral bands, and corresponds to a ground sampling distance of 30m. The training set comprises 70% of the patches, the validation set includes 20% of the patches, and the test set contains the remaining 10%. Moreover, the dataset provides two types of splits: the easy split, where patches from the same tile can appear in different sets (patchwise splitting), and the hard split, where all patches from a single tile are restricted to the same set (tilewise splitting). All experiments in this work were conducted using the hard split configuration. We also used classic augmentations in order to increase the number of training samples by a factor of 7. We downsampled the images by a factor of 4 before feeding them into the models, resulting in inputs with dimensions $32 \times 32 \times 202$. For LineSR, we use $F = 64$ features to achieve a balance between performance and computational complexity. Additionally, while the training phase processes the input as a whole by leveraging the parallelism capability of Mamba, the evaluation is performed by processing each image line by line. The expansion factor in both Mamba blocks is set to 1 and the dimension of the latent state is set to $N = 16$, following standard practice. The final upsampler consists of a Pixel Shuffle operation followed by a convolution with a kernel size of 1.

TABLE III
ARCHITECTURE ABLATION.

Model	MPSNR	Params	FLOPs/pixel
CausalConv	40.06	95.36 K	177.9 K
Mamba	40.21	95.11 K	165.5 K

B. Main results

We compare our line-based approach against the state-of-the-art method for single-image hyperspectral super-resolution MSDformer [1]. However, it is important to note that this method processes the images as a whole rather than line-by-line, leading to significantly higher complexity. We also compare against GDRRN [13], a traditional efficient benchmark method. Additionally, we report baseline results obtained using bicubic interpolation when processing the entire image. All models have been retrained on HySpecNet-11k using the authors' code.

Table I reports the main image quality results for LineSR in comparison with state-of-the-art architecture. We can notice that LineSR is close to its performance while having significantly fewer parameters. This result is very promising when also contextualized in terms of memory requirements and floating point operations needed for each pixel, which are reported in Table II. When testing on more real-world like image sizes, we can see that LineSR uses about less than a fifth of the memory of MSDformer [1] and a third of the memory of GDRRN [13] during inference thanks to the line-by-line approach; the more the image size increases, the bigger becomes the difference in memory requirement. Moreover, LineSR requires about a fifth of the FLOPs/pixel of MSDformer [1] and almost half of the ones of GDRRN [13].

C. Ablation experiments

We include an ablation study of the architecture, where we replace Mamba with a simple causal convolution in order to assess the effectiveness of the SSM. The result in Table III shows that indeed Mamba is more effective than simple causal convolution.

V. CONCLUSIONS

We presented a novel design for a neural network for hyperspectral image super-resolution. Thanks to the proposed mechanism where an image is processed line-by-line, the novel architecture is extremely lightweight in terms of memory and computational complexity, which makes it suitable for deployment onboard of satellites. At the same time, the super-resolution quality is competitive with high-complexity state-of-the-art models. Future work will focus on refining the design to further improve efficiency and image quality and testing it on low-power hardware.

REFERENCES

[1] S. Chen, L. Zhang, and L. Zhang, "Msdformer: Multi-scale deformable transformer for hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[2] Y. Xu, J. Hou, X. Zhu, C. Wang, H. Shi, J. Wang, Y. Li, and P. Ren, "Hyperspectral image super-resolution with convlstm skip-connections," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[3] H. Wang, Q. Zhang, T. Peng, Z. Xu, X. Cheng, Z. Xing, and T. Li, "Ssaformer: Spatial–spectral aggregation transformer for hyperspectral image super-resolution," *Remote Sensing*, vol. 16, no. 10, p. 1766, 2024.

[4] Q. Hu, X. Wang, J. Jiang, X.-P. Zhang, and J. Ma, "Exploring the spectral prior for hyperspectral image super-resolution," *IEEE Transactions on Image Processing*, 2024.

[5] G. Giuffrida, L. Fanucci, G. Meoni, M. Batič, L. Buckley, A. Dunne, C. Van Dijk, M. Esposito, J. Hefele, N. Vercruyssen *et al.*, "The ϕ-sat-1 mission: The first on-board deep neural network demonstrator for satellite earth observation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[6] V. Ržička, A. Vaughan, D. De Martini, J. Fulton, V. Salvatelli, C. Bridges, G. Mateo-Garcia, and V. Zantedeschi, "Ravæn: unsupervised change detection of extreme events using ml on-board satellites," *Scientific reports*, vol. 12, no. 1, p. 16939, 2022.

[7] M. Ziaja, P. Bosowski, M. Myller, G. Gajoch, M. Gumiela, J. Protich, K. Borda, D. Jayaraman, R. Dividino, and J. Nalepa, "Benchmarking deep learning for on-board space applications," *Remote Sensing*, vol. 13, no. 19, p. 3981, 2021.

[8] G. Inzerillo, D. Valsesia, and E. Magli, "Efficient onboard multi-task ai architecture based on self-supervised learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[9] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[10] H. Huang, A. G. Christodoulou, and W. Sun, "Super-resolution hyperspectral imaging with unknown blurring by low-rank and group-sparse modeling," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 2155–2159.

[11] X. Xu, X. Tong, J. Li, H. Xie, Y. Zhong, L. Zhang, and D. Song, "Hyperspectral image super resolution reconstruction with a joint spectral-spatial sub-pixel mapping model," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2016, pp. 6129–6132.

[12] S. He, H. Zhou, Y. Wang, W. Cao, and Z. Han, "Super-resolution reconstruction of hyperspectral images via low rank tensor modeling and total variation regularization," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2016, pp. 6962–6965.

[13] Y. Li, L. Zhang, C. Dingl, W. Wei, and Y. Zhang, "Single hyperspectral image super-resolution with grouped deep recursive residual network," in *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*. IEEE, 2018, pp. 1–4.

[14] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3d full convolutional neural network," *Remote Sensing*, vol. 9, no. 11, p. 1139, 2017.

[15] D. Liu, J. Li, Q. Yuan, L. Zheng, J. He, S. Zhao, and Y. Xiao, "An efficient unfolding network with disentangled spatial-spectral representation for hyperspectral image super-resolution," *Information Fusion*, vol. 94, pp. 92–111, 2023.

[16] P. V. Arun, K. M. Buddhiraju, A. Porwal, and J. Chanussot, "Cnn-based super-resolution of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 9, pp. 6106–6121, 2020.

[17] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1082–1096, 2020.

[18] Y. Liu, J. Hu, X. Kang, J. Luo, and S. Fan, "Interactformer: Interactive transformer and cnn for hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[19] M. Zhang, C. Zhang, Q. Zhang, J. Guo, X. Gao, and J. Zhang, "Essaformer: Efficient transformer for hyperspectral image super-resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 23 073–23 084.

[20] A. Gu, K. Goel, and C. Ré, "Efficiently modeling long sequences with structured state spaces," *arXiv preprint arXiv:2111.00396*, 2021.

[21] D. Y. Fu, T. Dao, K. K. Saab, A. W. Thomas, A. Rudra, and C. Re, "Hungry hungry hippos: Towards language modeling with state space models," in *The Eleventh International Conference on Learning Representations*.

[22] B. Peng, E. Alcaide, Q. G. Anthony, A. Albalak, S. Arcadinho, S. Biderman, H. Cao, X. Cheng, M. N. Chung, L. Derczynski *et al.*, "Rwkv: Reinventing rnns for the transformer era," in *The 2023 Conference on Empirical Methods in Natural Language Processing*.

[23] M. Poli, S. Massaroli, E. Nguyen, D. Y. Fu, T. Dao, S. Baccus, Y. Bengio, S. Ermon, and C. Ré, "Hyena hierarchy: Towards larger convolutional language models," in *International Conference on Machine Learning*. PMLR, 2023, pp. 28 043–28 078.

[24] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[25] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision mamba: Efficient visual representation learning with bidirectional state space model," in *Forty-first International Conference on Machine Learning*.

[26] A. Hatamizadeh and J. Kautz, "Mambavision: A hybrid mamba-transformer vision backbone," *arXiv preprint arXiv:2407.08083*, 2024.

[27] J. Yao, D. Hong, C. Li, and J. Chanussot, "Spectralmamba: Efficient mamba for hyperspectral image classification," *arXiv preprint arXiv:2404.08489*, 2024.

[28] L. Huang, Y. Chen, and X. He, "Spectral-spatial mamba for hyperspectral image classification," *arXiv preprint arXiv:2404.18401*, 2024.

[29] D. Valsesia, T. Bianchi, and E. Magli, "Onboard deep lossless and near-lossless predictive coding of hyperspectral images with line-based attention," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[30] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *European conference on computer vision*. Springer, 2022, pp. 17–33.