H-019

Semi-supervised Unmixing for Hyperspectral images

シャ　イケン　　　鎌田　清一郎
WEIXUAN XIE　　　Sei-ichiro Kamata

1. Abstract

Due to the limited spatial resolution of hyperspectral sensors, each pixel in a hyperspectral image (HSI) often contains a mixture of multiple materials, known as endmembers. Hyperspectral unmixing (HU) aims to extract these endmembers and estimate their corresponding fractional abundances from HSI data. In recent years, HU has seen significant progress with the rise of deep learning-based approaches. However, the localized receptive fields of convolutional neural networks (CNNs) and the high computational cost of Transformer-based models pose substantial challenges for HU tasks. These limitations highlight the need for image-level unmixing methods that can effectively capture long-range spatial–spectral dependencies while maintaining computational efficiency. Furthermore, given the high cost and difficulty of acquiring large amounts of manually labeled hyperspectral data, there is a growing interest in semi-supervised approaches that can leverage limited annotations to enhance unmixing performance. In this paper, we propose a spatial–spectral Mamba model for semi-supervised hyperspectral unmixing, which leverages the advantages of the Mamba state space model (Mamba-SSM) to efficiently capture long-range spatial–spectral dependencies with linear computational complexity. By utilizing a small number of labeled samples, our method achieves excellent image-level unmixing performance with fewer parameters and lower annotation costs.

2. Introduction

Hyperspectral imaging (HSI) provides rich spectral information across numerous contiguous bands, enabling detailed material identification at the pixel level. However, due to the inherently low spatial resolution of hyperspectral sensors, most pixels in an HSI scene are mixtures of multiple pure substances, referred to as endmembers. Hyperspectral unmixing (HU) aims to decompose each pixel into a set of endmembers and their corresponding fractional abundances. As a fundamental task in HSI analysis, HU has been widely applied in downstream applications such as change detection and material composition analysis.

With the rapid development of deep learning, HU methods have evolved significantly—from early fully connected networks (FCNs) and convolutional neural networks (CNNs) to more recent models based on attention mechanisms, such as Transformers. These methods typically follow a self-supervised learning paradigm, in which an encoder generates abundance of maps, and a decoder reconstructs the original input via a mixing model, such as the widely used linear mixing model (LMM). While CNN-based models are effective at capturing local spatial–spectral features, they lack the ability to model long-range dependencies. Transformers, on the other hand, can capture global spatial–spectral relationships, but suffer from quadratic computational complexity, limiting their scalability for large HSI data.

Recently, the emergence of Mamba, a selective state space model (SSM), offers a promising alternative. Mamba enables the modeling of long-range dependencies with linear computational complexity and efficient parallelization, making it well-suited for large-scale vision tasks. However, its potential in HU remains underexplored. Moreover, existing unsupervised or self-supervised approaches often suffer from instability and lack of physical interpretability.

To address these challenges, we propose a novel spatial–spectral dual-branch Mamba model within a semi-supervised learning framework. By leveraging the Mamba-SSM's capacity for efficient long-range dependency modeling and incorporating a limited amount of manually labeled data, our method achieves accurate and stable image-level unmixing with fewer parameters, lower computational costs, and improved interpretability.

3. Related Work

3.1 Deep Learning Methods for HU

Early HU models, such as fully connected networks (FCNs) [1] and convolutional neural networks (CNNs) [2][3], have been widely used to model the mapping between spectral signatures and abundance maps. For example, MLM-1DAE [4] utilizes spatial convolutions to extract local features. However, due to their inherently local receptive fields, CNN-based methods struggle to capture long-range spatial–spectral dependencies, which are critical for complex unmixing scenarios.

To address this limitation, Transformer-based methods have been introduced into HU tasks [5][6]. By leveraging self-attention mechanisms, these models can capture global spatial–spectral interactions more effectively. Methods such as DeepTrans-HSU [7] have demonstrated the potential of attention-based architectures in extracting rich contextual information. Nonetheless, Transformer models often suffer from quadratic computational complexity with respect to input size, which significantly limits their applicability to high-resolution hyperspectral data with large spatial dimensions.

†WEIXUAN XIE and Sei-ichiro Kamata are with the Image Media Laboratory, Graduate School of Information, Production and Systems, Waseda University, Kitakyushu 8080135, Japan.

3.2 Mamba-SSM in HU

Recently, Mamba—a selective state space model (SSM) [8]—has been proposed as a promising alternative to Transformer architectures. Mamba enables efficient modeling of long-range dependencies with linear computational complexity and supports highly parallelizable inference, making it suitable for dense visual data. While Mamba has shown strong performance in image classification and segmentation, its application in hyperspectral unmixing remains limited.

UNMamba [9] represents an early attempt to incorporate Mamba into HU. It adopts a cascaded architecture, where multiple Mamba blocks are stacked sequentially to extract hierarchical features. This structure demonstrates the feasibility of Mamba in modeling spatial–spectral correlations in HU tasks. However, the cascaded design used in UNMamba tends to entangle spatial and spectral dependencies within the same computational flow.

In contrast, we propose a novel dual-branch spatial–spectral Mamba model, in which spatial and spectral dependencies are explicitly decoupled and processed in parallel. This design allows the model to more effectively preserve and exploit the structural characteristics of hyperspectral data, while also improving interpretability.

3.3 Semi-supervised learning for HU

Acquiring abundant high-quality labeled hyperspectral data is both labor-intensive and costly. As a result, semi-supervised learning (SSL) has gained increasing attention in hyperspectral classification and segmentation. Approaches such as pseudo-labeling, consistency regularization, and entropy minimization have been explored to leverage both labeled and unlabeled data. However, most existing semi-supervised techniques focus on pixel-level classification tasks rather than unmixing, and few studies attempt to incorporate SSL into abundance estimation or endmember extraction.

Furthermore, self-supervised or unsupervised HU models—while useful in reducing annotation costs—often produce unstable results and suffer from limited physical interpretability. This motivates the development of HU models that can effectively integrate limited supervision while maintaining consistency with physical unmixing principles.

Our work addresses this gap by proposing a semi-supervised HU framework based on the dual-branch Mamba architecture. By leveraging both the efficiency of Mamba and the guidance from a small set of labeled samples, our model achieves accurate, stable, and interpretable unmixing results with significantly reduced annotation requirements.

4. Proposed Method

4.1 Dual-branch spatial–spectral Mamba model

To address the limitations of existing unmixing frameworks in capturing long-range spatial–spectral dependencies under limited supervision, we propose a dual-branch spatial–spectral Mamba network for semi-supervised hyperspectral unmixing. The overall architecture is illustrated in Figure 1.

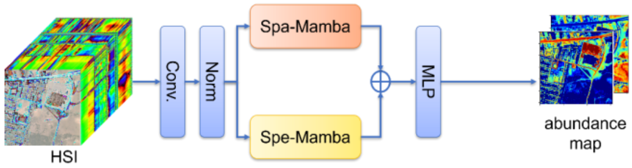

Figure 1. Dual-branch spatial-spectral Mamba model. Two specially designed branches based on Mamba-SSM are used to handle independent spatial and spectral dimensions.

Let $X \in \mathbb{R}^{B\times H\times W}$ denote the HSI with spatial dimensions H, W, and B spectral bands, $E \in \mathbb{R}^{R\times B}$ be the endmember matrix with R endmembers and $M \in \mathbb{R}^{R\times H\times W}$ the corresponding abundances.

The Linear Mixing Model [10] can be expressed as:

$$X = E^T M \tag{1}$$

We first apply a series of convolutional layers and normalization to extract low-level features that retain both spatial texture and spectral semantics denoted by $F_L \in \mathbb{R}^{C\times H\times W}$. These shared features F_L are then passed through two parallel Mamba-based branches as shown in Figure 2.

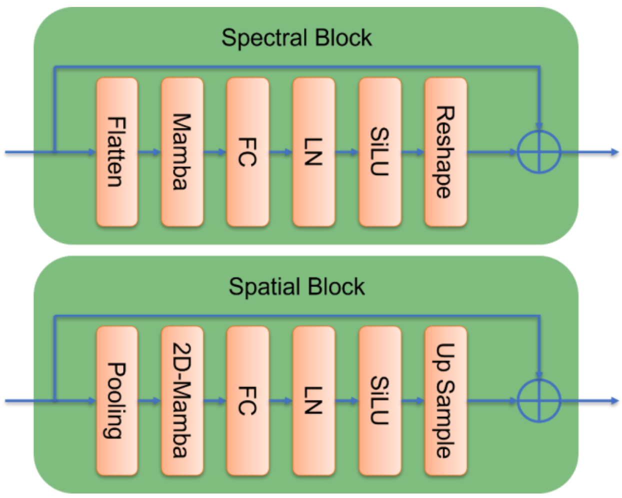

Figure 2. Two Mamba-based branches for spatial and spectral.

In the spatial block, after the average mean pooling layer, feature $F_L \in \mathbb{R}^{C\times H\times W}$ turns into $T_{spa} \in \mathbb{R}^{S\times S\times C}$ which will input to the 2D-Mamba block (Parallel optimized Mamba module).

In the spectral block, feature $F_L \in \mathbb{R}^{C\times H\times W}$ will be flattened to $F \in \mathbb{R}^{C\times(H\times W)}$ which will input to Mamba block for extracting the spectral dependencies.

The outputs of both branches are subsequently fused and forwarded to a lightweight MLP module, which estimates the abundance of maps of different endmembers.

4.2 Semi-supervised learning

Given the scarcity and high annotation cost of pixel-level abundance labels in hyperspectral images, we design a semi-supervised learning strategy that leverages a small set of annotated pixels to guide the training of our Spe-Mamba module and to estimate an initial set of endmembers. Specifically, we assume access to a small number of annotated samples, denoted as:

$$S = \{(x_i, p_i, a_i)\}_{i=1}^{N}, \tag{2}$$

where $x_i \in \mathbb{R}^B$ is the hyperspectral vector of i-th pixel (with B spectral bands), p_i is its spatial coordinate, and $a_i \in \mathbb{R}^R$ is the corresponding abundance vector with R components. We utilize the annotated data S to obtain an initial estimate of the matrix $E^{(0)} \in \mathbb{R}^{R\times B}$ based on LMM as a constrained least-squares problem. The estimation of endmembers will be updated on the training.

Meanwhile we set up multiple $Loss\ Function$ with different weights to achieve semi-supervised learning.

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \lambda_{\mathrm{sup}}\mathcal{L}_{sup} + \lambda_{sparse}\mathcal{L}_{sparse}$$

$$\mathcal{L}_{rec} = \frac{1}{N}\sum_{i=1}^{N}\left\|x_i - a_i^T E\right\|_2^2$$

$$\mathcal{L}_{sup} = \frac{1}{N}\sum_{i=1}^{N}\|\hat{a}_i - a_i\|_1 \tag{3}$$

Where $\mathcal{L}_{sparse}$ denotes the abundance sparsity term and λ_{sparse} is set to 0.0001 in the experiment.

5. Experiment

5.1 Dataset and Metrics

Two datasets were used for experiments: Urban [11] and APEX [12].

The Urban dataset is one of the most widely used hyperspectral data used in the hyperspectral unmixing study. In this dataset, there are 210 wavelengths ranging from 400 nm to 2500 nm, resulting in a spectral resolution of 10 nm. 307 × 307 spatial resolution and each pixel correspond $2 \times 2\, m^2$. After the pre-processing, 162 channels remain for HU analyses. 4 endmembers are given and Root Mean Squared Error (RMSE) is selected as evaluation metrics.

The Apex dataset is recently introduced for unmixing validation and contains 110 × 110 spatial resolution and 285 spectral bands. 4 endmembers are given and RMSE is selected as evaluation metrics.

DeepTrans-HSU and UNMamba were employed in comparison with PyTorch framework and RTX 3090 GPU. Both dataset will choose 10 sample points as the manually annotation information.

5.2 Performance comparison

Method	**Endmember/RMSE**				
	Asphalt	Grass	Tree	Roof	Ave.
DeepTrans-HSU	0.153	0.119	0.129	0.104	0.126
UNMamba	0.094	0.114	0.724	0.088	0.092
Ours	0.088	0.111	0.063	0.067	0.082

Table 1. Quantitative comparison on Urban Dataset.

Table 1 presents the results of endmember and abundance estimation for four materials. For each endmember, our method got a lower RMSE which meant a better result.

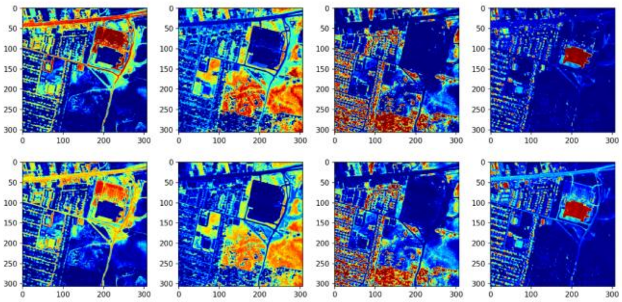

Figure 3. Abundance map comparison between Ground Truth and result on Urban dataset.

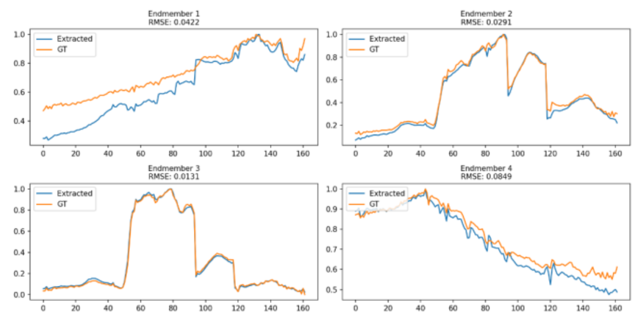

Figure 4. Visualizing spectral comparison between Ground Truth and result on Urban dataset.

Method	**Endmember/RMSE**				
	Asphalt	Grass	Tree	Roof	Ave.
DeepTrans-HSU	0.192	0.107	0.127	0.086	0.128
UNMamba	0.166	0.098	0.107	0.119	0.123
Ours	0.136	0.133	0.104	0.120	0.123

Table 2. Quantitative comparison on APEX Dataset.

Table 2 presents the results that our method performance is not good at each endmember. One possible reason is that the sample points are unevenly distributed.

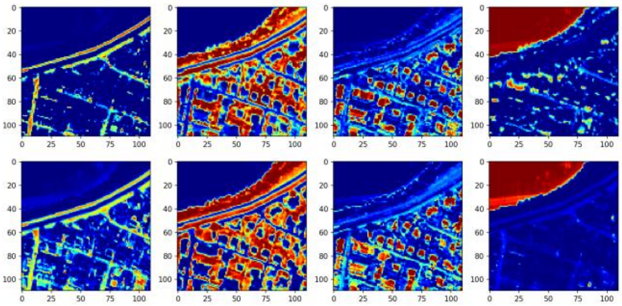

Figure 5. Abundance map comparison between Ground Truth and result on APEX dataset.

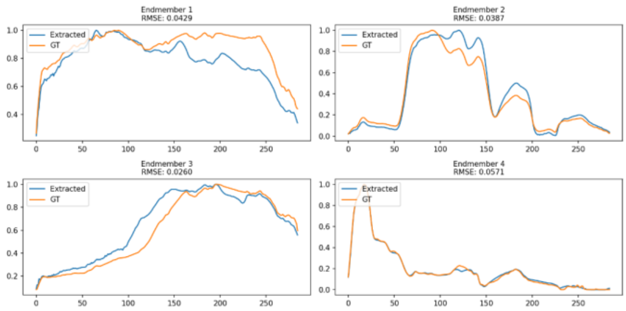

Figure 6. Visualizing spectral comparison between Ground Truth and result on APEX dataset.

5.3 Ablation Study

Table 3 presents the results of an ablation study evaluating the contribution of semi-supervised learning with 10 sample points. After adding the semi-supervised learning part, RMSE improve from 0.090 to 0.082, which can prove that SSL method is useful.

Method	**RMSE**
	Ave.
Baseline	0.090
+semi-supervised	0.082

Table 3. an ablation study on Urban dataset.

6. Conclusion

In this paper, we propose a novel semi-supervised hyperspectral unmixing framework based on a dual-branch spatial-spectral Mamba model. Extensive experiments on benchmark hyperspectral datasets demonstrate that our model achieves competitive or superior performance compared to existing CNN- and Transformer-based approaches. The proposed method provides a promising direction for scalable and physically meaningful hyperspectral unmixing in practical scenarios where labeled data is scarce.

Acknowledgement

The authors declare no conflict of interest. WEIXUAN XIE: Conceptualization, Methodology, Research, Experiment, Validation. Sei-ichiro Kamata: Conceptualization, Research, Validation. I would like to express my great appreciation to my supervisor Sei-ichiro Kamata for his consecutive suggestions.

References

[1] Palsson B, Sigurdsson J, Sveinsson J R, et al. Hyperspectral unmixing using a neural network autoencoder[J]. IEEE Access, 2018, 6: 25646-25656.
[2] Su Y, Li J, Plaza A, et al. DAEN: Deep autoencoder networks for hyperspectral unmixing[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(7): 4309-4321.
[3] Khajehrayeni F, Ghassemian H. Hyperspectral unmixing using deep convolutional autoencoders in a supervised scenario[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 567-576.
[4] T. Fang, F. Zhu, and J. Chen, “Hyperspectral unmixing based on multilinear mixing model using convolutional autoencoders,” IEEE Trans. Geosci. Remote Sens., vol. 62, 2024, Art. no. 5507316.
[5] Alexey D. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv: 2010.11929, 2020.
[6] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
[7] Ghosh P, Roy S K, Koirala B, et al. Hyperspectral unmixing using transformer network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-16.
[8] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces[J]. arXiv preprint arXiv:2312.00752, 2023.
[9] Chen D, Zhang J, Li J. UNMamba: Cascaded Spatial-Spectral Mamba for Blind Hyperspectral Unmixing[J]. IEEE Geoscience and Remote Sensing Letters, 2025.
[10] X. Tao, M. E. Paoletti, Z. Wu, J. M. Haut, P. Ren, and A. Plaza, “An abundance-guided attention network for hyperspectral unmixing,” IEEE Trans. Geosci. Remote Sens., vol. 62, 2024, Art. no. 5505414.
[11] Linda S. Kalman and Edward M. Bassett III "Classification and material identification in an urban environment using HYDICE hyperspectral data", Proc. SPIE 3118, Imaging Spectrometry III, (31 October 1997); https://doi.org/10.1117/12.283843
[12] Schaepman, M., Jehle, M., Hueni, A., D'Odorico, P., Damm, A., Weyermann, J., Schneider, F. D., Laurent, V., Popp, C., Seidel, F. C., Lenhard, K., Gege, P., Küchler, C., Brazile, J., Kohler, P., Vos, L. D., Meuleman, K., Meynart, R., Schläpfer, D. and Itten, K. I. (2015). "Advanced radiometry measurements and Earth science applications with the Airborne Prism Experiment (APEX)." Remote Sensing of Environment 158: 207-219.