Unmixing-Guided Convolutional Transformer for Spectral Reconstruction

Subjects: Computer Science, Artificial Intelligence

Contributor: Shiyao Duan , Jiaojiao Li , Rui Song , Yunsong Li , Qian Du

Specifically, transformer and ResBlock components are embedded in Paralleled-Residual Multi-Head Self-Attention (PMSA) to facilitate fine feature extraction guided by the excellent priors of local and non-local information from CNNs and transformers. Furthermore, the Spectral–Spatial Aggregation Module (S2AM) combines the advantages of geometric invariance and global receptive fields to enhance the reconstruction performance.

spectral reconstruction convolutional transformer hyperspectral unmixing

1. Introduction

Hyperspectral image (HSI) refers to a three-dimensional data cube generated through the collection and assembly of numerous contiguous electromagnetic spectrums, which are acquired via airborne or spaceborne hyperspectral sensors. Unlike regular RGB or grayscale images, HSI provides more information in the band dimension, which allows subsequent tasks to distinguish materials and molecular components that are difficult to distinguish from normal RGB through their stored explicit or implicit distinctions. As a result, HSI has distinct advantages in a variety of tasks, including object detection [1][2], water quality monitoring [3][4][5], intelligent agriculture [6][7][8], geological prospecting [9][10], etc.

However, hyperspectral imaging often requires long exposure times and various costs, making it unaffordable to collect sufficient data using sensors for many tasks with restricted budgets. Instead, acquiring a series of RGB or multispectral images is often a fast and cost-effective alternative. Therefore, using SR methods to inexpensively reconstruct the corresponding HSI from RGB or multispectral images (MSI) is a valuable solution. Currently, there are two main reconstruction approaches: the first involves fusing paired low-resolution hyperspectral (lrHS) and high-resolution multispectral (hrMS) images to produce a high-resolution hyperspectral (HrHs) image [11][12][13] with both high spatial and spectral resolutions, and the second approach generates the corresponding HSI by learning the inverse mapping from a single RGB image [14][15][16][17][18][19]. Commonly, image fusion-based methods [11][12][13] require paired images of the same scene, which can still be overly restrictive. Although reconstruction only from RGB images [14][15][16][20][21] is an ill-posed task due to the assumptions of inverse mapping, theoretical evidence demonstrates that feasible solutions exist under low-dimensional manifolds [22], and it provides sufficient cost-effectiveness.

Utilizing deep learning to model the inverse mapping in single-image reconstruction problems has been widely studied. Initially, numerous methods leveraged the excellent geometric feature extraction capabilities of CNNs [15]

[16][17][18][19] to achieve success in SR tasks. However, with the outstanding performance of transformers in various computer vision tasks, many transformer-based approaches [14][23][24] have recently emerged. These approaches take advantage of the transformer's global receptive field and sophisticated feature parsing abilities to achieve more refined HSI reconstruction. Nonetheless, current methods are predominantly limited to single-mechanism-driven frameworks, which often implies that the transformer architecture sacrifices the exceptional geometric invariance prior offered by CNNs. In fact, to ingeniously combine the advantages of both, numerous computer vision tasks have attempted to employ convolutional transformers to enhance the capability of feature extraction in their models, yielding highly impressive results [25][26][27][28]. Hence, employing a convolutional transformer to integrate the outstanding characteristics of both approaches is a clearly beneficial solution in SR.

Additionally, to achieve a higher signal-to-noise ratio in hyperspectral imaging, a trade-off between spectral resolution and spatial resolution is inevitable [29]. Most airborne hyperspectral sensors typically have a spatial resolution lower than 1 m/pixel [30][31], while satellite-based sensors, such as the Hyperion dataset of Ahmedabad, only have a 30 m/pixel resolution [32]. This significantly limits the effectiveness of HSI in capturing geographic spatial features. As a result, numerous approaches concentrate on employing mature CNNs or advanced transformer architectures to enhance feature extraction capabilities while overlooking the interpretability of the modeling itself and the pixel-mixing issues that arise during the imaging process.

2. Spectral Reconstruction (SR) with Deep Learning

Deep learning technology in SR task encompasses two distinct aspects. The first involves a fusion method based on paired images, while the second entails a direct reconstruction approach that leverages a single image such as those from CASSI or RGB systems. In the first category, a simultaneous capture of lrHS and hrMs images is employed, both possessing the same spectral and spatial resolution as HSIs separately. For example, Yao et al. [11] views hrMS as a degenerate representation of HSI in the spectral dimension and lrHS as a degenerate representation of HSI in the spatial dimension. It is suggested to use cross-attention in coupled unmixing nets based on the complementarities of the two features. Hu et al. [13], on the other hand, employed the Fusformer to obtain the implicit connection between global features and to solve the local neighborhood issue of the finite receptive field of the convolution kernel in the fusion problem using the transformer mechanism. The training process's data load is decreased by learning the spectral and spatial properties, respectively. However, the majority of the models' prior knowledge was created manually, which frequently results in a performance decrease when the domain is changed. Using the HSI denoising iterative spectral reconstruction approach based on deep learning, the MoG-DCN described by Dong et al. [33] has produced outstanding results in numerous datasets.

For the second category, where only single images are input, the model will learn the inverse function of the camera response function of a sensor using a single RGB image as an example. It will separate the RGB image's hidden hyperspectral feature data from it and then combine it with the intact spatial data to reconstruct a fine HSI. Shi et al. [15], for instance, replaced leftover blocks with dense blocks to significantly deepen the network structure and achieved exceptional results in NTIRE 2018 [20]. The pixel-shuffling layer was employed by Zhao et al. [19] to achieve inter-layer interaction, and the self-attention mechanism was used to widen the perceptual field. Cai et al.

[14] presented a cascade-based visual transformer model, MST++, to address the numerous issues with convolution networks in SR challenges. Its designed S-MSA and other modules further improved the ability of model to extract spatial and spectral features and achieved outstanding results in a large number of experiments.

The aforementioned analysis reveals that most previous models predominantly focused on enhancing feature extraction capabilities while neglecting the interpretability of physical modeling. This oversight often resulted in diminished performance in practical applications. In response, an SR model with robust interpretability was developed, capitalizing on the autoencoder's prowess in feature extraction and the simplicity of LMM. By harnessing the ability of LMM to extract sub-pixel-level features, ample spatial information is concurrently gathered from RGB images. Subsequently, high-quality HSIs are restored during the reconstruction process.

3. Deep Learning-Based Hyperspectral Unmixing

Several deep learning models based on mathematical or physical modeling have been suggested recently and used in real-world tests with positive outcomes due to the growing demand for the interpretability of deep learning models. Among these, HU has made significant progress in tasks such as change detection (CD), SR, and other HSI processing tasks. Guo et al. [34] utilized HU to extract sub-pixel-level characteristics from HSIs to integrate the HU framework into a conventional CD task. In order to obtain the reconstructed HSI, Zou et al. [35] used the designed constraints and numerous residual blocks to obtain the endmember matrix and abundance matrix, respectively. Su et al. [36] used the paired lrHs and hrMs to learn the abundance matrix and endmember from the planned autoencoder network and then rearranged them into HSI using the fundamental LMM presumptions.

Moreover, deep learning-based techniques are frequently used to directly extract the abundance matrix or end endmembers from the HU mechanism. According to Hong et al. [37], EGU-Net can extract a pure-pixel directed abundance matrix extraction model and estimate the abundance of synchronous hyperspectral pictures by using the parameter-sharing mechanism and the two-stream autocoder framework. By utilizing the asymmetric autoencoder network and LSTM to capture spectral information, Zhao et al. [38] were able to address the issue of inadequate spectral and spatial information in the mixed model.

Based on the aforementioned research, utilizing the HU mechanism to drive the SR task evidently improves interpretability. In light of this, the method introduces a parallel feature fusion module that combines the rich geometric invariance present in the residual blocks with the global receptive field of the transformer. This approach ensures the generation of well-defined features and aligns the channel-wise information with the endmembers of the spectral library.

3. Convolutional Transformer Module

The transformer-based approach has achieved great success in the field of computer vision, but using it exclusively will frequently negate the benefits of the original CNN structure and add a significant amount of computing burden. Due to this, numerous studies have started fusing the two. Among these, Wu et al. [25] inserted

CNN into the conventional vision transformer block, replacing linear projection and other components, and improved the accuracy of various computer vision tasks. Guo et al. [26] linked the two in succession, created the CMT model with both benefits, and created the lightweight visual model. He et al. [27] created the parallel CNN and transformer feature fusion through the developed RAM module and the dual-stream feature extraction component.

The integration of CNN and transformer is inevitable because they are the two most important technologies in the field of image processing. Many performance comparisons between the two have produced their own upsides and downsides [39][40]. Important information will inevitably be lost when using a single module alone. It is crucial to understand how to incorporate the elements that can be derived from both. In order to perform feature fusion for the parallel structure of PMSA, the channel size of the CNN that lacks modeling [41] can be well constrained utilizing the channel information in the transformer.

References

1. Li, Y.; Shi, Y.; Wang, K.; Xi, B.; Li, J.; Gamba, P. Target detection with unconstrained linear mixture model and hierarchical denoising autoencoder in hyperspectral imagery. IEEE Trans. Image Process. 2022, 31, 1418–1432.
2. Chhapariya, K.; Buddhiraju, K.M.; Kumar, A. CNN-Based Salient Object Detection on Hyperspectral Images Using Extended Morphology. IEEE Geosci. Remote Sens. Lett. 2022, 19, 6015705.
3. Liu, H.; Yu, T.; Hu, B.; Hou, X.; Zhang, Z.; Liu, X.; Liu, J.; Wang, X.; Zhong, J.; Tan, Z.; et al. Uav-borne hyperspectral imaging remote sensing system based on acousto-optic tunable filter for water quality monitoring. Remote Sens. 2021, 13, 4069.
4. Niroumand-Jadidi, M.; Bovolo, F.; Bruzzone, L. Water quality retrieval from PRISMA hyperspectral images: First experience in a turbid lake and comparison with sentinel-2. Remote Sens. 2020, 12, 3984.
5. Niu, C.; Tan, K.; Jia, X.; Wang, X. Deep learning based regression for optically inactive inland water quality parameter estimation using airborne hyperspectral imagery. Environ. Pollut. 2021, 286, 117534.
6. Li, K.Y.; Sampaio de Lima, R.; Burnside, N.G.; Vahtmäe, E.; Kutser, T.; Sepp, K.; Cabral Pinheiro, V.H.; Yang, M.D.; Vain, A.; Sepp, K. Toward automated machine learning-based hyperspectral image analysis in crop yield and biomass estimation. Remote Sens. 2022, 14, 1114.
7. Arias, F.; Zambrano, M.; Broce, K.; Medina, C.; Pacheco, H.; Nunez, Y. Hyperspectral imaging for rice cultivation: Applications, methods and challenges. AIMS Agric. Food 2021, 6, 273–307.

8. Khan, A.; Vibhute, A.D.; Mali, S.; Patil, C. A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. Ecol. Inform. 2022, 69, 101678.
9. Chakraborty, R.; Kereszturi, G.; Pullanagari, R.; Durance, P.; Ashraf, S.; Anderson, C. Mineral prospecting from biogeochemical and geological information using hyperspectral remote sensing-Feasibility and challenges. J. Geochem. Explor. 2022, 232, 106900.
10. Pan, Z.; Liu, J.; Ma, L.; Chen, F.; Zhu, G.; Qin, F.; Zhang, H.; Huang, J.; Li, Y.; Wang, J. Research on hyperspectral identification of altered minerals in Yemaquan West Gold Field, Xinjiang. Sustainability 2019, 11, 428.
11. Yao, J.; Hong, D.; Chanussot, J.; Meng, D.; Zhu, X.; Xu, Z. Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution. In Proceedings of the Computer Vision—ECCV 2020: 16th European Conference (Part XXIX 16), Glasgow, UK, 23–28 August 2020; pp. 208–224.
12. Hu, J.F.; Huang, T.Z.; Deng, L.J.; Jiang, T.X.; Vivone, G.; Chanussot, J. Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks. IEEE Trans. Neural Netw. Learn. Syst. 2021, 33, 7251–7265.
13. Hu, J.F.; Huang, T.Z.; Deng, L.J.; Dou, H.X.; Hong, D.; Vivone, G. Fusformer: A transformer-based fusion network for hyperspectral image super-resolution. IEEE Geosci. Remote Sens. Lett. 2022, 19, 6012305.
14. Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; Pfister, H.; Timofte, R.; Van Gool, L. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 745–755.
15. Shi, Z.; Chen, C.; Xiong, Z.; Liu, D.; Wu, F. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 939–947.
16. Li, J.; Wu, C.; Song, R.; Li, Y.; Liu, F. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 462–463.
17. Hu, X.; Cai, Y.; Lin, J.; Wang, H.; Yuan, X.; Zhang, Y.; Timofte, R.; Van Gool, L. Hdnet: High-resolution dual-domain learning for spectral compressive imaging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17542–17551.

18. Koundinya, S.; Sharma, H.; Sharma, M.; Upadhyay, A.; Manekar, R.; Mukhopadhyay, R.; Karmakar, A.; Chaudhury, S. 2D-3D CNN based architectures for spectral reconstruction from RGB images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 844–851.
19. Zhao, Y.; Po, L.M.; Yan, Q.; Liu, W.; Lin, T. Hierarchical regression network for spectral reconstruction from RGB images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 422–423.
20. Arad, B.; Ben-Shahar, O.; Timofte, R.N.; Van Gool, L.; Zhang, L.; Yang, M.N. Challenge on spectral reconstruction from RGB images. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 18–22.
21. Arad, B.; Timofte, R.; Ben-Shahar, O.; Lin, Y.T.; Finlayson, G.D. Ntire 2020 challenge on spectral reconstruction from an rgb image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 446–447.
22. Arad, B.; Ben-Shahar, O. Sparse recovery of hyperspectral signal from natural RGB images. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference (Part VII 14), Amsterdam, The Netherlands, 11–14 October 2016; pp. 19–34.
23. He, J.; Yuan, Q.; Li, J.; Xiao, Y.; Liu, X.; Zou, Y. DsTer: A dense spectral transformer for remote sensing spectral super-resolution. Int. J. Appl. Earth Obs. Geoinf. 2022, 109, 102773.
24. Yuan, D.; Wu, L.; Jiang, H.; Zhang, B.; Li, J. LSTNet: A Reference-Based Learning Spectral Transformer Network for Spectral Super-Resolution. Sensors 2022, 22, 1978.
25. Wu, H.; Xiao, B.; Codella, N.; Liu, M.; Dai, X.; Yuan, L.; Zhang, L. Cvt: Introducing convolutions to vision transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 22–31.
26. Guo, J.; Han, K.; Wu, H.; Tang, Y.; Chen, X.; Wang, Y.; Xu, C. Cmt: Convolutional neural networks meet vision transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12175–12185.
27. He, X.; Zhou, Y.; Zhao, J.; Zhang, D.; Yao, R.; Xue, Y. Swin transformer embedding UNet for remote sensing image semantic segmentation. IEEE Trans. Geosci. Remote Sens. 2022, 60, 4408715.
28. Liu, Z.; Luo, S.; Li, W.; Lu, J.; Wu, Y.; Sun, S.; Li, C.; Yang, L. Convtransformer: A convolutional transformer network for video frame synthesis. arXiv 2020, arXiv:2011.10185.
29. He, J.; Yuan, Q.; Li, J.; Zhang, L. PoNet: A universal physical optimization-based spectral super-resolution network for arbitrary multispectral images. Inf. Fusion 2022, 80, 205–225.

30. Xu, Y.; Du, B.; Zhang, L.; Cerra, D.; Pato, M.; Carmona, E.; Prasad, S.; Yokoya, N.; Hänsch, R.; Le Saux, B. Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2019, 12, 1709–1724.
31. Liu, L.; Li, W.; Shi, Z.; Zou, Z. Physics-informed hyperspectral remote sensing image synthesis with deep conditional generative adversarial networks. IEEE Trans. Geosci. Remote Sens. 2022, 60, 5528215.
32. Mishra, K.; Garg, R.D. Single-Frame Super-Resolution of Real-World Spaceborne Hyperspectral Data. In Proceedings of the 2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Rome, Italy, 13–16 September 2022; pp. 1–5.
33. Dong, W.; Zhou, C.; Wu, F.; Wu, J.; Shi, G.; Li, X. Model-guided deep hyperspectral image super-resolution. IEEE Trans. Image Process. 2021, 30, 5754–5768.
34. Guo, Q.; Zhang, J.; Zhong, C.; Zhang, Y. Change detection for hyperspectral images via convolutional sparse analysis and temporal spectral unmixing. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2021, 14, 4417–4426.
35. Zou, C.; Huang, X. Hyperspectral image super-resolution combining with deep learning and spectral unmixing. Signal Process. Image Commun. 2020, 84, 115833.
36. Su, L.; Sui, Y.; Yuan, Y. An Unmixing-Based Multi-Attention GAN for Unsupervised Hyperspectral and Multispectral Image Fusion. Remote Sens. 2023, 15, 936.
37. Hong, D.; Gao, L.; Yao, J.; Yokoya, N.; Chanussot, J.; Heiden, U.; Zhang, B. Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing. IEEE Trans. Neural Netw. Learn. Syst. 2021, 33, 6518–6531.
38. Zhao, M.; Yan, L.; Chen, J. LSTM-DNN based autoencoder network for nonlinear hyperspectral image unmixing. IEEE J. Sel. Top. Signal Process. 2021, 15, 295–309.
39. Zhou, H.Y.; Lu, C.; Yang, S.; Yu, Y. ConvNets vs. Transformers: Whose visual representations are more transferable? In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2230–2238.
40. Heo, B.; Yun, S.; Han, D.; Chun, S.; Choe, J.; Oh, S.J. Rethinking spatial dimensions of vision transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 11936–11945.
41. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 18–23 June 2018; pp. 7132–7141.