Deep Learning for Reconstructing Hyperspectral VNIR-SWIR Data from Multispectral Sources

1st Michael Alibani
DII, University of Pisa
56122 Pisa, Italy
michael.alibani@unipi.it

2nd Martina Pastorino
DITEN, University of Genoa
16145 Geona, Italy
martina.pastorino@edu.unige.it.

3rd Gabriele Moser
DITEN, University of Genoa
16145 Geona, Italy
gabriele.moser@unige.it.

4th Nicola Acito
DII, University of Pisa
56122 Pisa, Italy
nicola.acito@unipi.it

***Abstract*—Hyperspectral satellite data are crucial for applications such as environmental monitoring and precision agriculture due to their rich spectral information. However, hyperspectral data typically suffer from limited spatial resolution and are less readily available compared to more common multispectral data. This study explores the use of attention-based spectral reconstruction techniques, specifically MST++ and Restormer, to simulate high spatial and spectral hyperspectral data in the VNIR-SWIR from multispectral imagery. High-resolution multispectral and hyperspectral image pairs generated from AVIRIS-NG aerial data were used for training, allowing for the reconstruction of hyperspectral data that closely matches the original measurements. These results suggest that spectral reconstruction techniques can significantly enhance the utility of existing multispectral datasets for hyperspectral dependent applications and predict high-resolution hyperspectral data from multispectral inputs. This approach is especially useful for generating simulated data for missions still in development. For instance, PRISMA-2G data can be modeled using Sentinel-2 data.**

***Index Terms*—Spectral Reconstruction, Restormer, MST++, AVIRIS-NG, PRISMA, Sentinel-2.**

I. Introduction

Hyperspectral (HS) satellite data are a valuable resource for understanding Earth's complex environmental systems due to its high spectral resolution, which enables precise identification and analysis of materials. Consequently, it plays a crucial role in applications such as environmental monitoring [1], precision agriculture [2], mineral exploration [3], and urban planning [4]. The importance of HS satellite data for Earth observation is demonstrated by current operational missions such as PRISMA [5], ENMAP [6] and the satellite programs planned for the near future such as PRISMA second generation (PRISMA-2G [7]) and CHIME [8].

In the preparatory activities for a new mission, having a data simulation tool is essential for consolidating mission requirements, validating the operational processor, and assessing the effects of various instrumental and environmental parameters. In addition to system design and analysis, simulating HS systems is crucial for developing effective data process techniques. By mimicking real-world conditions, simulations provide a controlled setting for designing, testing, and refining algorithms that extract valuable information from HS data. For the analysis of instrumental radiometric and spectral responses, simulated 1-D radiance spectra can serve as a basis. However, when evaluating parameters related to the instrument's spatial performance, such as the Point Spread Function (PSF), the keystone effect, or spectrometer coregistration, the spatial dimension must also be considered. The initial challenge in developing an end-to-end HS simulator involves generating a HS reflectance image with a spatial resolution higher than that of the system being simulated. To have realistic surface patterns for the simulated images, reflectance data from existing remote sensing missions can be used. The majority of existing satellite HS data have low spatial resolution, making them unsuitable for simulating reflectance images needed to assess the spatial performance of instruments. A possible solution to the problem is to use existing high spatial resolution multispectral (MS) satellite data along with Spectral Reconstruction (SR, [9]) techniques to infer the spectral information related to the target HS sensor.

Aiming to develop an end-to-end simulator for data of the PRISMA-2G HS mission, we intend to apply SR algorithms to obtain HS images from Sentinel-2 (S2) MS data. Specifically, our goal is to obtain HS images that span 230 spectral bands, starting with the ten S2 bands (B2-B8, B8a, B11, and B12) at higher spatial resolution. These S2 bands cover the spectral range observed by the PRISMA-2G instrument, from the Visible and Near InfraRed (VNIR) regions to the Short Wave InfraRed (SWIR) region. Deep Learning (DL) based SR techniques have shown remarkable success, particularly in the task of reconstructing HS images from data with lower spectral dimensionality, such as red, green, and blue (RGB) images that consist of three channels. Approaches such as AWAN [10], HSCNN [11] and DRCRNet [12], as well as transformer-based models [13], have demonstrated that high-quality HS information can be extracted from RGB images. Recent studies have shown that SR methods were also effective in reconstructing remotely sensed images from MS to HS however they are mainly limited to the VNIR spectral range [14][15]. In this work we test the effectiveness of a selection of DL based SR techniques to obtain high-resolution HS images from MS data considering the overall VNIR-SWIR spectral range that will characterize the PRISMA-2G data. Specifically, we consider two recent networks, Restormer [16] and MST++ [13], which have demonstrated strong performance in SR

This work is carried out within ASI-funded project agreement "SIM4PRISMA2G" n. 2023-2-HB.0, CUP n. F53C23000620001.

using RGB images. We apply these networks to our case of reconstructing PRISMA-2G data from S2 images to produce data with high spatial and spectral resolution that is useful for simulation purposes. For both training and testing the networks, S2 and PRISMA-2G image pairs are simulated starting from airborne AVIRIS-NG imagery [17] using the nominal Spectral Response Functions (SRFs) of S2 [18] and PRISMA-2G [5] instruments, respectively.

The work is organized as follows: Section II introduces spectral reconstruction with a focus on MST++ and Restormer. Section III describes the generation of the dataset and the application of spectral reconstruction methods to it. Finally, Section IV presents both qualitative and quantitative evaluations of the spectral reconstruction.

II. Spectral Reconstruction

The process of recovering HS images from MS images, is a complex and highly ill-posed problem. This challenge comes from the need to estimate a high-dimensional spectrum from a lower-dimensional observation, which can lead to significant difficulties in achieving accurate results. Over the years, researchers have developed various methods to address this issue, achieving remarkable results in reconstructing HS images from images that has a lower spectral dimensionality. SR approaches can be categorized into two classes: prior-based methods and data-driven methods.

This work focuses on data-driven approaches that have gained significant traction in recent years within the computer vision community. Neural networks can automatically learn complex patterns from data and often outperform traditional methods in both accuracy and generalizability. For example, HSCNN+ [11] uses dense blocks and a novel fusion scheme to enhance spectral recovery. AWAN [10] improves spectral detail with an Adaptive Weighted Channel Attention (AWCA) module. Zhao et al.'s HRNet [19] employs a PixelShuffle layer to boost inter-level interactions and accuracy in HS images reconstruction. Additionally, MiRNet [20] leverages a multi-scale residual learning framework to effectively capture and refine features at different scales, further enhancing the quality of HS images reconstruction. Here, we apply two state-of-the-art methods from the NITRE Spectral Recovery challenge [13]: Multi-Scale Transformer++ (MST++) and Restoration Transformer (Restormer) that represent cutting-edge approaches to HS images reconstruction, each offering unique capabilities tailored to different aspects of image restoration.

MST++ is specifically designed for HS data and employs a Transformer-based architecture optimized for spectral information. The model utilizes multi-head self-attention to capture both long-range dependencies and self-similarities within the spectral domain. A key innovation of MST++ is its treatment of each spectral channel as an individual token within the self-attention mechanism. This approach allows MST++ to perform sophisticated inter-channel comparisons, which is crucial for accurately reconstructing spectral features. The multi-scale design of MST++ enables it to analyze features at different resolutions, enhancing its ability to resolve fine spectral details and adapt to varying levels of data granularity. Additionally, MST++ benefits from a high level of spectral sensitivity, which is critical for tasks that require detailed spectral discrimination.

Restormer, on the other hand, is designed with a broader scope, focusing on high-quality image restoration across various domains, including HSI. Restormer employs self-attention mechanisms to effectively integrate both local and global information, addressing spatial and spectral details simultaneously. Unlike MST++, which is heavily focused on the spectral dimension, Restormer uses a dual approach to balance the preservation of spatial structures and spectral accuracy. Its architecture includes advanced mechanisms for spatial feature extraction and noise reduction, making it highly versatile for diverse restoration tasks. Restormer's design also incorporates sophisticated normalization techniques and attention layers that improve its performance in fine-grained image restoration and enhance its robustness to different types of image degradation.

III. Reconstructing Hyperspectral Data from Multispectral

As discussed in Section I, the objective of this work is to evaluate the effectiveness of state-of-the-art MST++ and Restormer networks in retrieving HS remote sensed data from corresponding MS data. Specifically, this case study focuses on extrapolating a high-resolution PRISMA-2G data, which consists of 230 spectral bands, from S2 data, which has only 10 spectral bands.

It is important to note that real S2 images have four spectral channels at 10 m Ground Sampling Distance (GSD) and six channels at 20 m GSD, while the PRISMA-2G instrument in the stripmap operational setting will provide data over 230 spectral channels in the range $400 - 2505$ nm with a 30 m GSD.

This work starts from the assumption that it is possible to obtain 10-bands S2 data at 20 m GSD by downsampling the higher-resolution bands, or to obtain 10-bands S2 data at 10 m GSD using one of the well-established spatial super-resolution methods [21]. S2 images are good candidates for producing synthetic PRISMA-2G reflectance data with a spatial resolution higher than the nominal one. Therefore to demonstrate the feasibility of producing corresponding HR PRISMA-2G data from S2, we used AVIRIS-NG imagery for simulating the HR image pairs that are necessary for training, validation and testing.

In our experiments, we begin by implementing the MST++ and Restormer networks as described in [22]. These networks were originally designed to reconstruct a 31-band HS image from its RGB version. For our application, we adapted these networks to reconstruct a 230-band HS image from a 10-band MS image by adapting the input and output interfaces, as well as the dimensions of the multi-head attention layers to accommodate the increased spectral dimensionality. The image pairs for training and testing the networks are generated as follow.

A. Dataset Generation from AVIRIS-NG

Training DL models requires a large and sufficiently diverse dataset for effective performance. In this study, we utilize the AVIRIS-NG product to generate a suitable dataset composed of high-resolution MS and HS image pairs. AVIRIS-NG is an advanced aircraft-borne HS sensor that utilizes HR spectroscopy to acquire precise measurements of solar-reflected radiation. It operates across a wavelength range from 380 nm to 2510 nm, sampling the spectrum at approximately 5 nm. The sensor utilizes a pushbroom configuration to collect data as spectral images, with the spatial sampling determined by flight altitude. As a result, AVIRIS-NG is capable of achieving ground resolutions ranging from approximately 0.3 m at low altitudes to around 7.0 m at higher altitudes. The characteristics of this dataset make it well-suited for generating high-resolution image pairs for our application. The comprehensive AVIRIS-NG dataset is freely accessible via NASA's AVIRIS Data Portal [23].

We selected images obtained from AVIRIS-NG sensors across diverse regions worldwide, covering various years between 2018 and 2022. These images encompass a variety of landscapes, including coastal areas, urban centers, mountainous regions, and rural settings. During the selection process, we excluded regions that had cloud cover, corrupted data, or were completely covered by water. Using the AVIRIS-NG dataset as our baseline, we simulated high-resolution data for both S2 and PRISMA-2G by applying the SRF of each sensor. This process involved convolving the AVIRIS-NG data with the SRF of S2 and PRISMA to replicate how these sensors would capture the same scene at high resolution. From the dataset, we extracted approximately 3000 image tiles, each measuring 128×128 pixels, which were divided into a training set (80%) and a test set (20%). Each image tile consists of atmospherically corrected spectral reflectances with a 12-bit radiometric resolution, appropriately scaled within the range [0,1]. The training data are augmented with random rotation and flipping.

The training process was conducted on a single machine equipped with a NVIDIA GeForce RTX 4090 GPU with 24GB RAM. The networks were trained using Adam optimizer using the following optimization parameters: $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The learning rate is initialized as $l_r = 10^{-4}$ and a Cosine Annealing scheme is adopted to dynamically adjust the learning rate during training. The models were trained for 200 epochs each. The duration of training for each network model was approximately 48 hours.

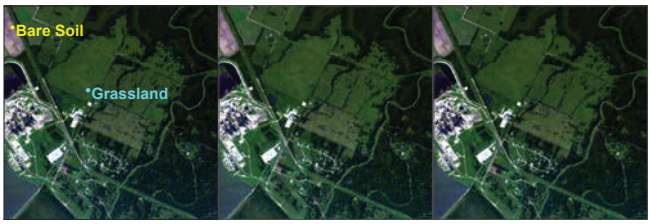
IV. RESULTS

This section presents the results of the two SR methods for generating high-resolution HS images. Both qualitative and quantitative analyzes are conducted: the qualitative evaluation assesses the visual fidelity of the reconstructed images, while the quantitative analysis provides objective performance metrics.

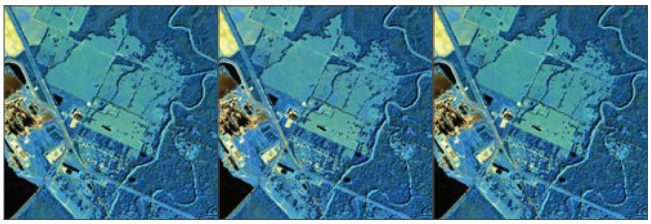
To quantitatively evaluate the performance of our proposed methods, we employed three metrics, two of which were recommended by the 2022 Spectral Recovery Challenge [24]. The first metric is the Mean Relative Absolute Error (MRAE), which calculates the pixel-wise relative difference between the reconstructed and ground-truth images across all spectral channels, providing a measure of the relative error at each pixel. The second metric is the Root Mean Square Error (RMSE), which quantifies the average magnitude of the errors

TABLE I
MEDIAN RRMSE, RMSE, AND MRAE VALUES FOR EVALUATING THE PERFORMANCE OF MST++ AND RESTORMER NETWORKS ON THE TEST SET. THE ARROWS INDICATE THE DIRECTION OF THE BETTER VALUE. THE BOLD INDICATES THE BEST RESULTS.

Method	MRAE (↓)	RMSE (↓)	RRMSE (↓)
MST++ [13]	0.0485	0.0076	0.0359
Restormer [16]	**0.0473**	**0.0067**	**0.0324**

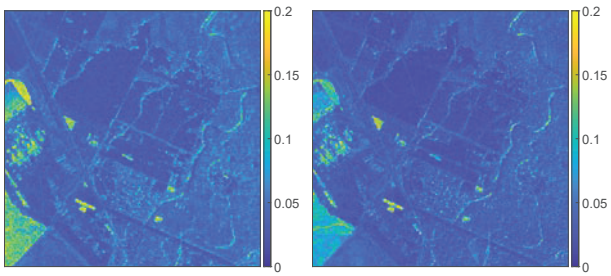


(a) RGB - GT (b) RGB - MST++ (c) RGB - Res.



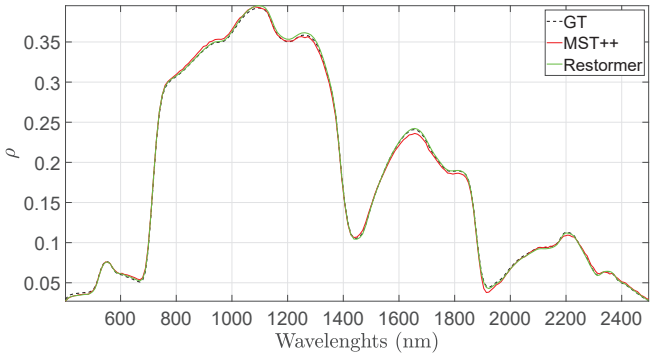
(d) FC - GT (e) FC - MST++ (f) FC - Res.

Fig. 1. Visual comparison of image reconstruction methods in both RGB and FC representations. Wavelengths 470.75 nm, 546.27 nm, and 631.90 nm are utilized for RGB, whereas wavelengths 1008.07 nm, 1646.92 nm, and 2143.18 nm are used for FC: (a) Ground truth simulated HR PRISMA image (RGB); (b) Image reconstructed by MST++ (RGB); (c) Image reconstructed by Restormer (RGB); (d) Ground truth simulated HR PRISMA image (FC); (e) Image reconstructed by MST++ (FC); (f) Image reconstructed by Restormer (FC).

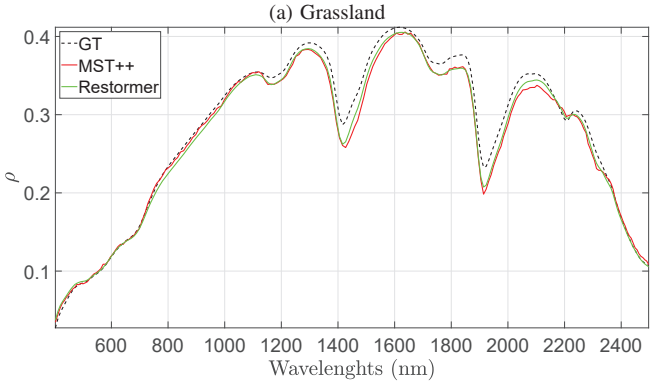


(a) Error Map - MST++ (b) Error Map - Restormer

Fig. 2. RRMSE maps of the reconstructed images with respect to the Ground Truth: (a) Relative error map of MST++; (b) Relative error map of Restormer.



(a) Grassland



(b) Bare Soil

Fig. 3. Comparison of reflectance values obtained after SR: spectral signatures are extracted from the pixels highlighted by cyan and yellow dots in Fig. 1a. (a) Spectral signatures of a grassland pixel, compared between MST++, Restormer, and the ground truth; (b) Spectral signatures of a bare soil pixel, compared between MST++, Restormer, and the GT.

between the reconstructed and ground-truth by computing the square root of the mean squared differences. Finally, the third metric is Root Relative Mean Square Error (RRMSE), which expresses the reconstruction error normalized by the root mean square value of the reference image.

Table I displays the median values of the proposed metrics to evaluate the performance of the MST++ and Restormer networks on the test dataset. All metrics indicate that Restormer outperforms MST++ (even if with a small margin) in reconstructing the 230-bands HS data from 10-bands MS input, even though MST++ resulted the top-performing network on the ARAD1K dataset during the 2022 Spectral Reconstruction Challenge.

Fig. 1 presents the results of SR when applied to an image comprising both natural and man-made elements, having a spatial dimension of 512×512. It is important to note that the network was fed with an MS image derived from an AVIRIS-NG image that was not included in the training set, thereby serving as a tangible test of generalization. In order to improve the visual comparison, we propose both an RGB representation (Figs. 1a - 1c) and a False Color (FC) representation (Figs. 1d - 1f) of the reconstructed images in relation to the corresponding Ground Truth (GT) images. Wavelengths 470,75 nm, 546,27 nm, and 631,90 nm are utilized for RGB, whereas wavelengths 1008,07 nm, 1646,92 nm, and 2143,18 nm are used for FC. A visual inspection of the generated images reveals no perceivable macroscopic artifacts. However, relying solely on RGB and FC visual comparisons is insufficient for a comprehensive evaluation. Figure 2 presents RRMSE maps that illustrate the distribution of reconstruction errors and pinpoint regions where the networks underperform in accurately reconstructing the image. These maps enable a detailed, pixel-by-pixel analysis of the RRMSE values across the entire image. RRMSE values thus represent a relative error with respect to the reference pixel, where 0% represents no error and 100% represents maximal error. Fig. 2 illustrates maps limited on a scale [0, 0.2], in order to facilitate improved visualisation, ranging from blue to yellow, respectively. Although the error distribution across the maps is quite similar, map in Fig. 2b shows a stronger shift towards blue, indicating a generally better reconstruction. This is further supported by the average error values, which are 0.072 for Fig. 2a and 0.059 for Fig. 2b. Notably, regions with lower Signal-to-Noise Ratio (SNR), such as water bodies, exhibit higher reconstruction errors. For instance, see the bottom left-hand corner of Fig. 2a.

To further analyze the study's results, the authors extracted reflectance values from two pixels in the GT image and their corresponding pixels in the reconstructed image. This enabled a direct comparison of the overall spectral signatures. The selected samples are indicated in Fig. 1a by a cyan marker (representing a grassland sample) and a yellow marker (representing a bare soil sample). Fig. 3 shows the spectral comparison, with Fig. 3a displaying the reflectance values for the grassland sample, and Fig. 3b for the bare soil sample.

The results indicate that both networks successfully reconstruct the spectral profile of the HR HS ground truth data across the entire VNIR-SWIR spectral range. However, a closer inspection of the Restormer profile (green line), compared to the MST++ profile (red line) for both selected samples, reveals a closer match to the ground truth (black line), thereby confirming previous findings. Notably, in the bare soil case, both networks exhibit higher errors beyond $\lambda = 1100$ nm. This can be attributed to the fact that the HR S2 data contains 8 spectral bands within the VNIR region, but only 2 bands within the SWIR, making it more challenging for the networks to accurately interpolate data in this spectral interval.

V. CONCLUSION

In conclusion, this work explored the use of MST++ and Restormer DL networks for the spectral reconstruction in the VNIR-SWIR spectral interval of high-resolution HS data from MS input. The data pairs used in this work to train the networks were generated from AVIRIS-NG using proper SRFs to simulate high-resolution S2 and PRISMA-2G datasets. Both networks proved to be effective in extrapolating detailed spectral information, demonstrating their potential to generate high quality HS data (from MS inputs) that could be used for simulation purposes of future missions. Future work will focus on applying these methods to real-world MS datasets and enhancing their capability to predict HS data that will be available in the future.

References

[1] G. Corsini, N. Acito, M. Alibani, *et al.*, "Hyperhealth - environmental impact assessment on human health: Advanced methods for hyperspectral prisma data exploitation," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023, pp. 1752–1755. DOI: 10.1109/IGARSS52108.2023.10282724.

[2] P. K. Sethy, C. Pandey, Y. K. Sahu, and S. K. Behera, "Hyperspectral imagery applications for precision agriculture-a systemic survey," *Multimedia Tools and Applications*, vol. 81, no. 2, pp. 3005–3038, 2022.

[3] E. Bedini, "The use of hyperspectral remote sensing for mineral exploration: A review," *Journal of Hyperspectral Remote Sensing*, vol. 7, no. 4, pp. 189–211, 2017.

[4] C. Weber, R. Aguejdad, X. Briottet, *et al.*, "Hyperspectral imagery for environmental urban planning," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2018, pp. 1628–1631.

[5] R. Loizzo, R. Guarini, F. Longo, *et al.*, "Prisma: The italian hyperspectral mission," in *IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium*, IEEE, 2018, pp. 175–178.

[6] E. O. C. E. of DLR, *Enamp*, https://www.enmap.org/, Accessed: Sep. 19, 2024, 2024.

[7] L. Ansalone, *Prisma2gen*, https://hyperspectral2022.esa.int/iframe-agenda/files/presentation-187.pdf/, Accessed: Sep. 19, 2024, 2024.

[8] J. Nieke, L. Despoisse, A. Gabriele, *et al.*, "The copernicus hyperspectral imaging mission for the environment (chime): An overview of its mission, system and planning status," *Sensors, Systems, and Next-Generation Satellites XXVII*, vol. 12729, pp. 21–40, 2023.

[9] J. Zhang, R. Su, Q. Fu, W. Ren, F. Heide, and Y. Nie, "A survey on computational spectral reconstruction methods from rgb to hyperspectral imaging," *Scientific reports*, vol. 12, no. 1, p. 11 905, 2022.

[10] J. Li, C. Wu, R. Song, Y. Li, and F. Liu, "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 (CVPR) Workshops*, Jun. 2020.

[11] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, and F. Wu, "Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 518–525. DOI: 10.1109/ICCVW.2017.68.

[12] J. Li, S. Du, C. Wu, Y. Leng, R. Song, and Y. Li, "Drcr net: Dense residual channel re-calibration network with non-local purification for spectral super resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jun. 2022, pp. 1259–1268.

[13] Y. Cai, J. Lin, Z. Lin, *et al.*, "Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jun. 2022, pp. 745–755.

[14] D. Du, Y. Gu, T. Liu, and X. Li, "Spectral reconstruction from satellite multispectral imagery using convolution and transformer joint network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023. DOI: 10.1109/TGRS.2023.3285893.

[15] T. Li and Y. Gu, "Progressive spatial–spectral joint network for hyperspectral image reconstruction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022. DOI: 10.1109/TGRS.2021.3079969.

[16] S. W. Zamir, A. Arora, S. Khan, *et al.*, "Learning enriched features for real image restoration and enhancement," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, Springer, 2020, pp. 492–511.

[17] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The airborne visible/infrared imaging spectrometer (aviris)," *Remote sensing of environment*, vol. 44, no. 2-3, pp. 127–143, 1993.

[18] F. Gascon, C. Bouzinac, O. Thépaut, *et al.*, "Copernicus sentinel-2a calibration and products validation status," *Remote Sensing*, vol. 9, no. 6, p. 584, 2017.

[19] J. Wang, K. Sun, T. Cheng, *et al.*, "Deep high-resolution representation learning for visual recognition," *TPAMI*, 2019.

[20] S. W. Zamir, A. Arora, S. Khan, *et al.*, "Learning enriched features for real image restoration and enhancement," in *ECCV*, 2020.

[21] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 146, pp. 305–319, 2018.

[22] Y. Cai, J. Lin, Z. Lin, *et al.*, *Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction*, https://github.com/caiyuanhao1998/MST-plus-plus, Accessed: Sep. 19, 2024, 2024.

[23] J. P. Laboratory, *Aviris-ng data portal*, https://avirisng.jpl.nasa.gov/dataportal/, Accessed: Sep. 19, 2024, 2024.

[24] B. Arad, R. Timofte, R. Yahel, *et al.*, "Ntire 2022 spectral recovery challenge and data set," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 863–881.