Spectral Super-Resolution via Adversarial Unfolding and Data-Driven Spectrum Regularization: From Multispectral Satellite Data to NASA Hyperspectral Image

Supplementary Material

Si-Sheng Young[1], Chia-Hsiang Lin[1,2,*]
[1]National Cheng Kung University, [2]Industrial Technology Research Institute
* Corresponding author
q38121509@gs.ncku.edu.tw, chiahsiang.steven.lin@gmail.com

1. The Proposed PriorNet

The notation usages in the following description are consistent with the main article (see Section 3 of the main article). The overall pipeline of the proposed lightweight PriorNet is shown in Figure 1. PriorNet comprises two branches for a) performing the spatial resolution unification (SRU) to 5 m GSD, and b) providing the spectral prior matrix $\boldsymbol{P}$ to empower the spectrum regularization [see Eq. (1) of the main article]. Given a multiresolution Sentinel-2 data $\boldsymbol{S}$, the first branch initially performs a $2\times$ bicubic upsampling, followed by a short residual convolution process (composed of two convolutions and a multiscale module). This residual block performs primary spatial dimension alignment (SDA) for the subsequent encoder. A single convolution then projects the SDA features to a latent space, and four cascaded multiscale modules are applied for feature encoding. After the encoder, the bottleneck, composed of three multiscale modules and spe-spa attentions, contributes to the main signal processing. The bottlenecks not only process the encoded features sequentially but also provide informative multi-depth features for the following fusion-based decoder. We empirically observe that the multi-depth features significantly contribute to the performance of the output spatial prior image. Finally, the fusion-based decoder aggregates these multi-depth features with a long-range residual connection to output the final 5 m GSD spatial prior image $\boldsymbol{S}_u$.

In the second branch, we design the PriorNet to trade the spatial structure for precise cross-similarity of the target hyperspectral bands. In detail, an inevitable challenge in spectral super-resolution (SSR) is to preserve both spatial and spectral structure at the same time. Therefore, the second branch outputs a spatially compressed matrix $\widetilde{\boldsymbol{A}} \in \mathbb{R}^{186\times\frac{L}{64}}$, which preserves the similar corss-similarity patterns as $\boldsymbol{A}$, i.e., $\boldsymbol{A}\boldsymbol{A}^T \approx \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{A}}^T = \boldsymbol{P}$. Since we do not strictly demand the spatial details of $\widetilde{\boldsymbol{A}}$, an efficient design is sufficient to yield a promising approximation, $\boldsymbol{P} \approx \boldsymbol{A}\boldsymbol{A}^T$. Then, the spectral prior matrix $\boldsymbol{P}$ can be used to guide the overall UALNet, namely, data-driven spectrum regularization (see Section 3.1 of the main article). To this end, channel attention is first applied to the input Sentinel-2 data to adjust the energy of each spectral band. Then, the first convolution is adopted for the stage-one spectral upsampling (12-to-46), while the second one is for a trainable $2\times$ spatial downsampling. In addition, a $2\times$ maxpooling layer with channel attention is further incorporated. Consequently, we use two other convolutions for the stage-two spectral upsampling (46-to-186) and trainable $2\times$ spatial downsampling to output $\widetilde{\boldsymbol{A}}$.

For the loss function design, we first adopt the smooth ℓ_1-norm loss [1] between the reference $\boldsymbol{S}_u$ and the estimated result from PriorNet $\widehat{\boldsymbol{S}_u}$ as the data-fitting loss, i.e., $\mathcal{L}_1 :=$ SmoothL1$(\boldsymbol{S}_u, \widehat{\boldsymbol{S}_u})$. In addition, conventional distance-based loss functions (e.g., MSE and ℓ-1 loss) encourage low-frequency components [2], failing to obtain fine spatial details for $\widehat{\boldsymbol{S}_u}$. Accordingly, the frequency loss function is adopted to resolve this issue, i.e., $\mathcal{L}_2 :=$ SmoothL1$(|\mathcal{F}(\boldsymbol{S}_u)|, |\mathcal{F}(\widehat{\boldsymbol{S}_u}|)$, where $\mathcal{F}(\cdot)$ indicates the 2D Fourier transform, and $|\cdot|$ is the absolute value operator. Moreover, we preserve the spectral fidelity using the spectral angle mapper (SAM) loss, i.e., $\mathcal{L}_3 :=$ SAM$(\boldsymbol{S}_u, \widehat{\boldsymbol{S}_u})$. Consequently, we also use smooth ℓ-1 loss for the spectral prior matrix, $\mathcal{L}_4 :=$ SmoothL1$(\boldsymbol{A}\boldsymbol{A}^T, \boldsymbol{P})$. The overall loss function can be explicitly written as

$$\mathcal{L}_1 + \lambda_1\mathcal{L}_2 + \lambda_2\mathcal{L}_3 + \lambda_3\mathcal{L}_4,$$

where λ_1 :=2.5E-3, λ_2 :=2.5E-3, and λ_3 :=5E-4. Thus far, the introduction of PriorNet has been completed, and we refer readers to Section 4.2 of the main article for the training settings.

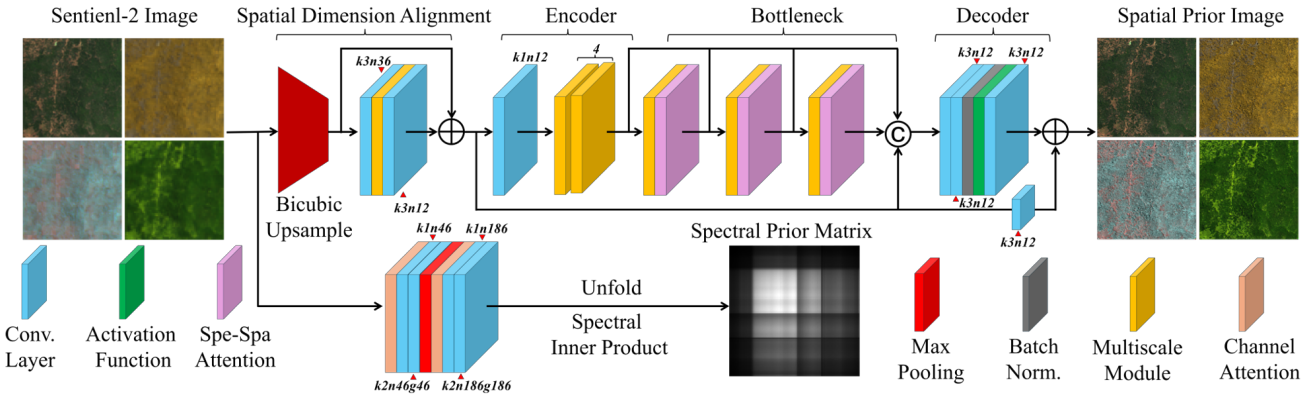

Figure 1. Overall pipeline of the proposed lightweight PriorNet, where the notation "ka-nb-gc" denotes a 2D convolution with a kernel size of a, output channel of b, and group number of c (a full convolution would not exhibit a group number additionally). The architectural details of Multiscale Module, Channel Attention, and Spe-Spa Attention are illustrated in Figure 2(a), Figure 2(b), and Figure 2(c), respectively. In our design, the fist branch of PriorNet perform a spatial resolution unification (SRU) to obtain the high and uniform resolution (5 m GSD) prior image $\boldsymbol{S}_u \in \mathbb{R}^{12\times L}$ from the input multiresolution (containing 60 m, 20 m, and 10 m GSD) Sentinel-2 image $\boldsymbol{S} \in \mathbb{R}^{12\times l}$, where $L = 4l$. The second branch provides a spectral prior matrix $\boldsymbol{P} \in \mathbb{R}^{186\times 186}$ that encodes the spectral cross-similarity of the target AVIRIS-level HSI $\boldsymbol{A} \in \mathbb{R}^{186\times L}$, i.e., $\boldsymbol{P} \approx \boldsymbol{A}\boldsymbol{A}^T$. The introduction of the architectural design is presented in Section 1, and Section 3 of the main article describes the role of PriorNet within the overall UALNet.

2. Analysis for Real Sentinel-2 Case Study

This section provides a detailed analysis to support the conclusion of Section 4.5 of the main article. In this case study, we employ real Sentinel-2 data as the input of our pretrained UALNet and then reconstruct the corresponding AVIRIS-level HSI. Nevertheless, as remarked in Section 4.1 of the main article, paired real Sentinel-2 data and AVIRIS-NG HSIs are hardly collected. Therefore, beyond simulated quantitative evaluations, we conduct an unmixing-based evaluation using real Sentinel-2 input to demonstrate the effectiveness and practicality of the proposed framework in real-world applications.

Unmixing [i.e., blind source separation for remote sensing (RS) imagery] is a critical image processing technology for RS material identification [3, 4]. As typical RS images often suffer from the mixed-pixel phenomenon (MPP), the unmixing algorithms attempt to recover N pure spectral signatures (i.e., endmembers) and their corresponding spatial proportions (i.e., abundance maps). Mathematically, given a M-band RS images with L pixels $\boldsymbol{X} \in \mathbb{R}^{M\times L}$, the unmixing algotithms factorize it into the spectral signature matrix $\boldsymbol{E} \in \mathbb{R}^{M\times N}$ and spatial abundance matrix $\boldsymbol{S} \in \mathbb{R}^{N\times L}$ such that $\boldsymbol{X} \approx \boldsymbol{E}\boldsymbol{S}$, where N is typically assumed known. Since the AVIRIS-level HSI is reconstructed from its Sentinel-2 counterpart over the same spatial location, they naturally refer to the consistent abundance maps. Accordingly, we compare the abundance maps (i.e., spatial proportions of pure materials) between a real Sentinel-2 MSI and its AVIRIS-level reconstruction for cross-validation.

The ROI (see Figure 7 of the main article) for this case study is located in Washington (WA), USA, and was captured in August 2019. This ROI covers a spatial size of 256×256, and primarily consists of river, farmland, and mountainous areas. We illustrate the true-color and false-color compositions of the real Sentinel-2 and the spatial prior image to evaluate their spectral relations, as demonstrated in Figure 7 of the main article. The high-resolution and multiresolution compositions exhibit consistent color distributions, validating the effectiveness of the SRU provided by PriorNet. In addition, we demonstrate that the high-resolution spatial prior images preserve a meaningful data structure, rather than being "pretty pictures". Specifically, before performing the unmixing algorithm on the Sentinel-2 data and its AVIRIS-level reconstruction, the number of sources is estimated using the unsupervised model order selection method, namely minimum description length (MDL) [5]. Developed by information theory, the MDL algorithm determines the optimal model order N that results in the shortest code length for both real Sentinel-2 data and spatial prior image. As shown in Figure 7 of the main article, the code length with respect to the number of sources follows a consistent trend on both real Sentinel-2 data and spatial prior image, substantiating that the spatial prior image maintains a realistic multispectral structure.

With the consistently shortest code length at $N := 4$, we employ a theoretically guaranteed unmixing algorithm [i.e., hyperplane-based Craig's simplex identification (HyperCSI) [3]] to recover four endmembers/abundance maps from both real Sentinel-2 data and its AVIRIS-level recon-

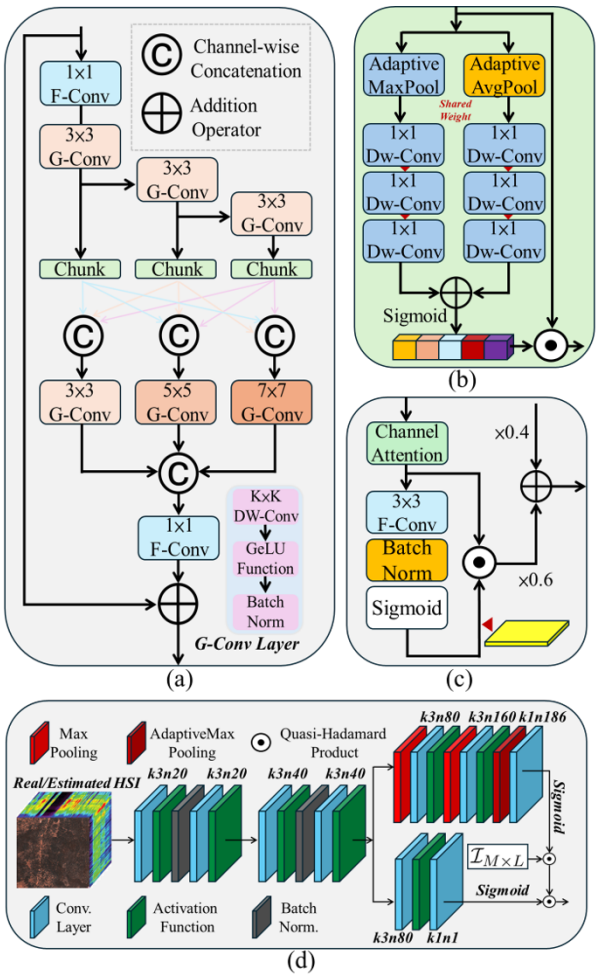

Figure 2. Schematic diagrams of the network architectures, including (a) Multiscale Module, (b) Channel Attention, (c) Spe-Spa Attention, and (d) Discriminator.

struction, as demonstrated in Figure 3. In the results, we can tell that the most significant distinction lies in their spectral resolutions of recovered pure signatures (i.e., endmembers). Due to the limited spectral resolution of the Sentinel-2 data (only 12 bands), it is challenging to identify materials using their discrete spectral signatures, even with the help of the unmixing algorithm. In contrast, the AVIRIS-level HSI reconstructed by the proposed UALNet contains 186 high-quality and densely sampled bands across a broad spectrum range. Such abundant spectral representation enables HyperCSI to recover the continuous hyperspectral endmembers, thereby facilitating more faithful RS applications To further demonstrate the reliability and effectiveness of the proposed UALNet, we evaluate the spatial consistency of the recovered abundance maps between the Sentinel-2 data and its AVIRIS-level counterpart. Because they correspond to the same geographic area,

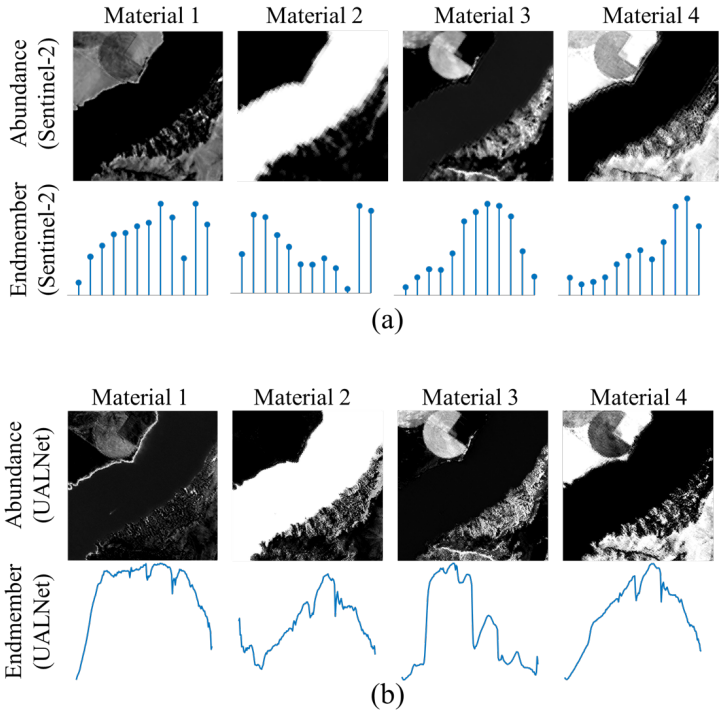

Figure 3. Qualitative comparisons of the abundance maps and endmembers recovered from (a) Real Sentinel-2 MSI and (b) the corresponding AVIRIS-level reconstruction from the proposed UALNet.

the reconstructed HSI in the high-spatial-resolution domain is expected to preserve material distributions aligned with the Sentinel-2 data in the coarse-spatial-resolution domain, rather than producing some spurious patterns artificially. To verify this quantitatively, the abundance maps recovered from the AVIRIS-level reconstruction (512×512) are resized back to 256×256 to match the spatial dimension with the real Sentinel-2 data, ensuring a quantitative evaluation. Subsequently, we compute the cross-correlation coefficient for each pair of abundance maps, which is more robust and less sensitive (compared to distance-based metrics) to the scale distortions caused by resizing. The averaged correlation across all abundance pairs (excluding the outlier minimum correlation) achieves 93.5388%, validating the strong spatial consistency and the spatial reliability of the proposed UALNet. In a nutshell, the proposed framework, namely unfolding adversarial learning network (UALNet), for the Sentinel-2 to AVIRIS-level reconstruction is experimentally validated to be applicable for real-world RS applications.

3. Practical Applicability: Real-Data Analysis

Real-world data pairs are further collected to verify the simulation-to-real (S2R) generalization capability of the proposed UALNet. Although the proposed UALNet is

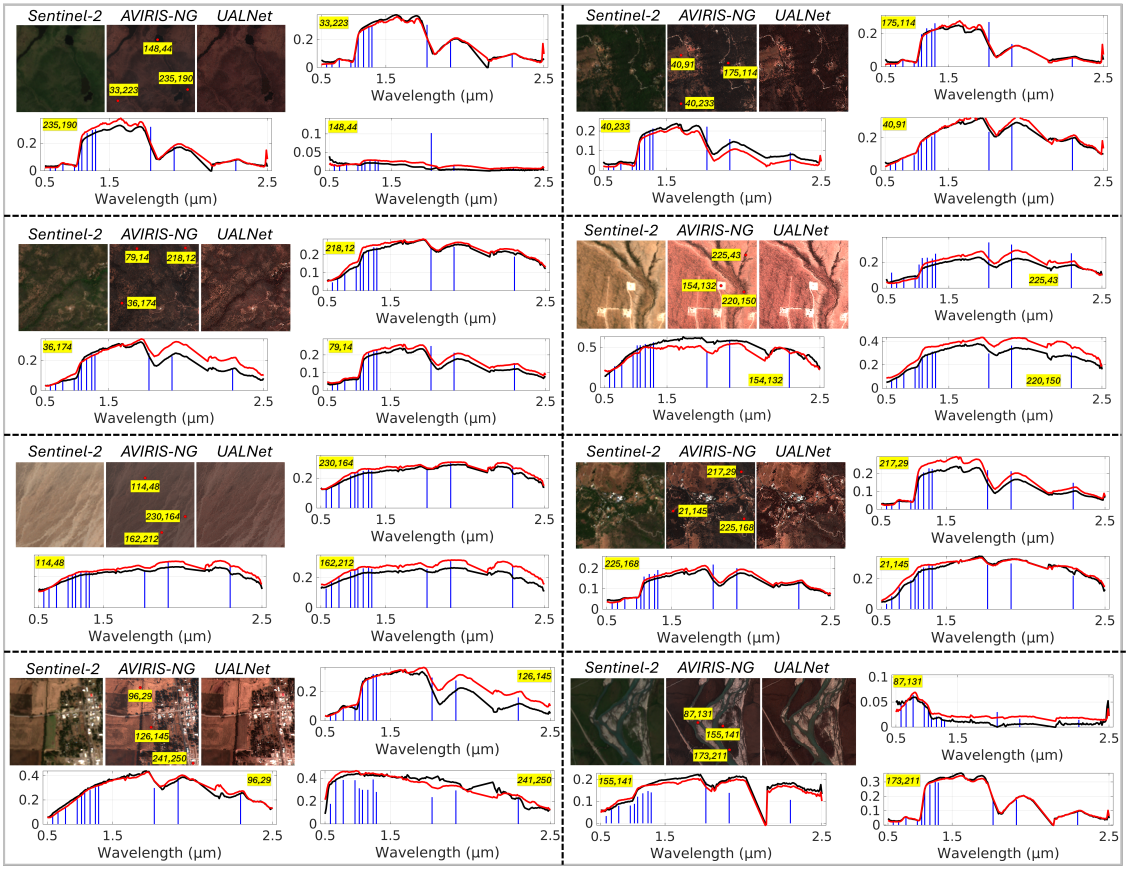

Figure 4. Qualitative evaluations (bands 25 12 8 as RGB) on various real-world datasets, including Inuvik (Northwest Territories, Canada), Valley Springs (Calaveras, USA), Death Valley National Park (Inyo, USA), Pauls Valley (Garvin, USA), North Slope Borough (Alaska, USA), Carlsbad Caverns National Park (Eddy, USA), and Shingle Springs (El Dorado, USA). The blue impulses denote Sentinel-2 signatures, while the black and red continuous curves are the real AVIRIS-NG and the UALNet-reconstructed signatures, respectively. The UALNet trained on simulated data can yield promising reconstructions for real-world inputs.

trained solely on simulated datasets, we apply it on real-world Sentinel-2 inputs (with 128×128 pixels) **without additional fine-tuning**. Then, we compare both false-color compositions and spectral signatures between the UALNet-reconstructed hyperspectral images and the corresponding real-world AVIRIS-NG dataset collected from diverse regions. As shown in Figure 4, the reconstructed signatures (i.e., red curves) exhibit strong consistency across all significantly different real-world datasets. The rigorous simulation training strategy enables robust S2R generalization. These observations will be included in the main article during the revision stage.

References

[1] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, pp. 142–158, 2015. 1

[2] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (Salt Lake City, U.S.), pp. 586–595, 2018. 1

[3] C.-H. Lin, C.-Y. Chi, Y.-H. Wang, and T.-H. Chan, "A fast hyperplane-based minimum-volume enclosing simplex algorithm for blind hyperspectral unmixing," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1946–1961, Dec. 2015. 2

[4] H. Zeng, J. Cao, K. Zhang, Y. Chen, H. Luong, and W. Philips, "Unmixing diffusion for self-supervised hyperspectral image denoising," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27820–27830, 2024. 2

[5] C.-H. Lin, C.-Y. Chi, L. Chen, D. J. Miller, and Y. Wang, "Detection of sources in non-negative blind source separation by minimum description length criterion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 9, pp. 4022–4037, 2018. 2