SIMILARITY-GUIDED HOMOGENEOUS-HETEROGENEOUS FUSION NETWORK FOR HYPERSPECTRAL IMAGE SUPER-RESOLUTION

Jin Xu[1], Shun'an Deng[2], Chunbo Jiang[2], Zheng Wang[3], Jinchang Ren[2,3,4], Jaime Zabalza[5], and Tariq Durrani[5]*

[1] International Joint Institute, Tianjin University, Fuzhou, 350205, China
[2] School of Artificial Intelligence, Guangdong Polytechnic Normal University, Guangzhou, 510665, China
[3] School of Artificial Intelligence, Tianjin University, Tianjin, 300072, China
[4] SoCET, Robert Gordon University, Aberdeen, AB10 7AQ, UK
[5] Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1XW, UK

ABSTRACT

Fusing a low-resolution hyperspectral image (LR-HSI) with a high-resolution multispectral image (HR-MSI) is an effective approach to hyperspectral image super-resolution (HSI-SR), as the two observations carry complementary spatial and spectral information. However, existing methods often fail to explicitly distinguish the information shared by the two modalities and source-specific, limiting their ability to exploit both observations fully. To this end, we propose a Similarity-Guided Homogeneous-Heterogeneous Fusion Network (SHHF). SHHF first encodes the LR-HSI and HR-MSI with separate feature extractors, and then maps the resulting features into homogeneous and heterogeneous representations using weight-sharing and modality-specific encoders, respectively. Two projection-based contrastive objectives are designed to train these representations: a cross-modal InfoNCE loss with a shared projection head aligns the homogeneous representations, while a subspace decorrelation loss keeps the heterogeneous representations source-specific. The four representations are subsequently concatenated and fed into a decoder to reconstruct the high-resolution hyperspectral image. Experimental results on the Washington DC Mall dataset demonstrate that SHHF outperforms advanced HSI-SR methods in both quantitative metrics and visual quality.

Index Terms— Hyperspectral image, Super-resolution, Contrastive learning, Deep learning.

1. INTRODUCTION

Hyperspectral images (HSIs) are three-dimensional data cubes comprising numerous contiguous spectral bands that capture both spatial and spectral information [1, 2]. Owing to their rich spectral information, HSIs have been widely used in various applications [3], including classification [4], change detection [5, 6] and anomaly detection [7].

However, acquiring high-spatial-resolution HSIs (HR-HSIs) remains challenging because of the physical constraints of hyperspectral imaging systems [8]. The resulting limitations in spatial resolution hinder the characterization of fine spatial details and restrict the applicability of HSIs. These challenges motivate the development of hyperspectral image super-resolution (HSI-SR) methods. Compared with HSIs, multispectral images (MSIs) generally offer higher spatial resolution but lower spectral resolution. Fusing a low-spatial-resolution HSI (LR-HSI) with a high-spatial-resolution MSI (HR-MSI) therefore provides a practical approach to reconstructing an HR-HSI by exploiting their complementary spatial and spectral information. Existing fusion-based HSI-SR methods can be broadly categorized into deep learning-based and non-deep-learning approaches.

Non-deep-learning methods commonly formulate HSI-MSI fusion as an inverse problem regularized by explicit structural priors, with matrix and tensor factorization representing two major approaches. Matrix factorization typically expresses the target HR-HSI as the product of a spectral basis and spatial coefficients. Berné et al. [9] employed nonnegative matrix factorization (NMF) for hyperspectral data fusion, while Yokoya et al. [10] developed coupled NMF to integrate the complementary information in HSIs and MSIs through spectral unmixing. Tensor factorization methods retain the multidimensional structure of hyperspectral data, allowing spatial and spectral dependencies to be modeled jointly. For example, Kanatsoulis et al. [11] formulated hyperspectral super-resolution using coupled canonical polyadic decomposition. Although these approaches offer interpretable reconstruction frameworks, their reliance on prescribed factorization models and handcrafted priors can limit their flexibility in representing complex spatial-spectral relationships.

Deep learning-based methods learn spatial-spectral representations through neural networks and can be broadly

This study was supported by the Dazhi Leading Scholarship of Guangdong Polytechnic Normal University (+National High-end Talents Program).*Corresponding author: Jinchang Ren (jinchang.ren@ieee.org).*

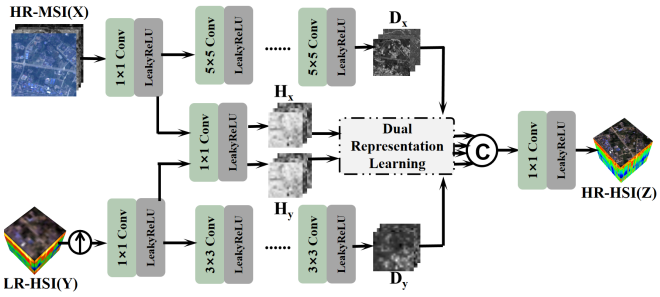


Fig. 1. The structure and details of the proposed SHHF.

categorized into supervised and unsupervised approaches. Palsson et al. [12] employed a three-dimensional convolutional neural network to extract joint spatial-spectral features. To capture global contextual information, Hu et al. [13] introduced Fusformer, which uses self-attention to model long-range dependencies. Xu et al. [14] introduced Graph-attention into coupled unmixing networks to facilitate information exchange between the two modalities. To account for the degradation process, Jia et al. [15] combined convolutional feature extraction with Transformer-based modeling in a multiscale spatial-spectral fusion network, integrating local information with global context. Despite their strong representation capabilities, supervised methods depend on paired training data with reference HR-HSIs, which are difficult to acquire in practical imaging scenarios. Liu et al. [16] combined degradation estimation with a model-inspired autoencoder for unsupervised HR-HSI reconstruction. Xu et al. [17] proposed a two-stream self-attention network based on different input data to achieve HSI-SR. Their later model-informed multistage framework further explored the integration of observation models and staged reconstruction [3], while their enhanced deep image prior approach investigated architectural priors for unsupervised hyperspectral super-resolution [18].

Although existing methods have achieved promising HSI-SR performance, modeling dependencies between the LR-HSI and HR-MSI does not necessarily ensure that their shared and source-specific information is explicitly distinguished and appropriately processed. This distinction motivates a fusion strategy that accounts for the different characteristics of homogeneous and heterogeneous features. Accordingly, we propose a Similarity-Guided Homogeneous–Heterogeneous Fusion Network (SHHF) for HSI-SR. SHHF first extracts latent representations from the LR-HSI and HR-MSI using separate feature extractors. It then maps the primary features into homogeneous representations (a shared encoder) and heterogeneous representations (modality-specific encoders), and trains them with two projection-based contrastive objectives: a cross-modal InfoNCE loss that aligns the shared information, and a subspace decorrelation loss that preserves the source-specific information.

2. METHODOLOGY

The proposed Similarity-Guided Homogeneous-Heterogeneous Fusion Network (SHHF) consists of three stages. First, two different encoders extract primary features from the HR-MSI and LR-HSI. Second, the primary features are mapped into two homogeneous and two heterogeneous representations, which are learned by two projection-based contrastive objectives: a cross-modal InfoNCE loss that aligns the homogeneous representations, and a subspace decorrelation loss that keeps the heterogeneous representations source-specific. Finally, the four representations are integrated to reconstruct the target HR-HSI. The overall architecture of SHHF is illustrated in Fig. 1.

2.1. Primary Feature Extraction

Let $\mathbf{X} \in \mathbb{R}^{H\times W\times c}$ and $\mathbf{Y} \in \mathbb{R}^{h\times w\times C}$ denote the observed HR-MSI and LR-HSI, respectively, where H and W are the spatial dimensions of the HR-MSI, h and w are those of the LR-HSI, and c and C denote their respective numbers of spectral bands. Generally, $H > h$, $W > w$, and $C > c$. The objective is to reconstruct an HR-HSI $\mathbf{Z} \in \mathbb{R}^{H\times W\times C}$.

Because the two observations have different spatial resolutions, the LR-HSI is first upsampled and then encoded together with the HR-MSI:

$$\begin{aligned}\mathbf{F}_X &= \mathcal{E}_X(\mathbf{X}),\\ \mathbf{F}_Y &= \mathcal{E}_Y(\mathbf{Y}\uparrow),\end{aligned} \tag{1}$$

where $\uparrow$ denotes the spatial upsampling operator, and the upsampled LR-HSI $\mathbf{Y}\uparrow \in \mathbb{R}^{H\times W\times C}$ has the same spatial resolution as $\mathbf{X}$. Moreover, $\mathcal{E}_X(\cdot)$ and $\mathcal{E}_Y(\cdot)$ are the MSI and HSI encoders, respectively. The resulting feature maps $\mathbf{F}_X, \mathbf{F}_Y \in \mathbb{R}^{H\times W\times d}$ have the same spatial dimensions and number of channels, where d denotes the feature dimension.

2.2. Homogeneous and Heterogeneous Representations

The primary feature maps $\mathbf{F}_X$ and $\mathbf{F}_Y$ contain both consistent information shared by the two modalities and distinctive information caused by their different spatial and spectral characteristics. To separate these two types of information, the primary features are mapped into homogeneous and heterogeneous representations. The homogeneous representations are produced with a single weight-sharing encoder:

$$\begin{aligned}\mathbf{H}_X &= \mathcal{E}_h(\mathbf{F}_X),\\ \mathbf{H}_Y &= \mathcal{E}_h(\mathbf{F}_Y),\end{aligned} \tag{2}$$

where $\mathcal{E}_h(\cdot)$ is a weight-sharing homogeneous encoder, and $\mathbf{H}_X, \mathbf{H}_Y \in \mathbb{R}^{H\times W\times d_r}$ are the homogeneous representations of the HR-MSI and LR-HSI, respectively, with d_r the representation dimension.

In contrast, two modality-specific encoders produce the heterogeneous representations:

$$\begin{aligned}\mathbf{D}_X &= \mathcal{E}_X^d(\mathbf{F}_X),\\ \mathbf{D}_Y &= \mathcal{E}_Y^d(\mathbf{F}_Y),\end{aligned} \tag{3}$$

where $\mathcal{E}_X^d(\cdot)$ and $\mathcal{E}_Y^d(\cdot)$ are independent heterogeneous encoders without parameter sharing, and $\mathbf{D}_X, \mathbf{D}_Y \in \mathbb{R}^{H\times W\times d_r}$ retain the source-specific spatial and spectral characteristics of the HR-MSI and LR-HSI, respectively.

2.3. Dual Representation Learning

The two types of representations are trained by two distinct contrastive objectives, each computed through a dedicated projection head. The projection heads are used only during training to compute the losses; the fusion stage operates on the original representations $\mathbf{H}$ and $\mathbf{D}$.

Homogeneous contrastive alignment. To align the shared information, $\mathbf{H}_X$ and $\mathbf{H}_Y$ are divided into N non-overlapping patches at the same spatial locations, where $N = HW/b^2$ and b denotes the patch size. Each patch embedding is mapped by a shared projection head $\Phi_{\text{hom}}(\cdot)$, into a lower-dimensional space:

$$\begin{aligned}\mathbf{u}_i^X &= \text{Norm}\left(\Phi_{\text{hom}}\left(\text{vec}\left(\mathcal{P}_i(\mathbf{H}_X)\right)\right)\right),\\ \mathbf{u}_i^Y &= \text{Norm}\left(\Phi_{\text{hom}}\left(\text{vec}\left(\mathcal{P}_i(\mathbf{H}_Y)\right)\right)\right),\end{aligned} \tag{4}$$

where $\mathcal{P}_i(\cdot)$ extracts the ith spatial patch, $\text{vec}(\cdot)$ denotes vectorization, $\text{Norm}(\cdot)$ performs ℓ_2 normalization, and $\mathbf{u}_i^X, \mathbf{u}_i^Y \in \mathbb{R}^{d_p}$. The homogeneous loss is a symmetric cross-modal InfoNCE that pulls the corresponding patch pair together and pushes the non-corresponding pairs apart:

$$\mathcal{L}_{\text{hom}} = -\frac{1}{2N}\sum_{i=1}^{N}[\log\frac{e^{\langle \mathbf{u}_i^X, \mathbf{u}_i^Y\rangle/\tau}}{\sum_{j=1}^{N} e^{\langle \mathbf{u}_i^X, \mathbf{u}_j^Y\rangle/\tau}} + \log\frac{e^{\langle \mathbf{u}_i^Y, \mathbf{u}_i^X\rangle/\tau}}{\sum_{j=1}^{N} e^{\langle \mathbf{u}_i^Y, \mathbf{u}_j^X\rangle/\tau}}], \tag{5}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product and $\tau > 0$ is a fixed temperature.

Heterogeneous subspace decorrelation. To ensure that the heterogeneous branch encodes only the source-specific information not captured by the homogeneous branch, a shared projection head $\Phi_{\text{het}}(\cdot)$ (a two-layer MLP) maps both $\mathbf{H}$ and $\mathbf{D}$ position-wise into a common d_p-dimensional space. Let $\bar{\mathbf{H}}_X, \bar{\mathbf{H}}_Y \in \mathbb{R}^{d_p\times HW}$ and $\bar{\mathbf{D}}_X, \bar{\mathbf{D}}_Y \in \mathbb{R}^{d_p\times HW}$ be the flattened projections of $\mathbf{H}_X, \mathbf{H}_Y, \mathbf{D}_X, \mathbf{D}_Y$, respectively. The heterogeneous loss pushes the shared and private subspaces to be orthogonal within each modality:

$$\mathcal{L}_{\text{het}} = \left\|\bar{\mathbf{H}}_X\bar{\mathbf{D}}_X^\top\right\|_F^2 + \left\|\bar{\mathbf{H}}_Y\bar{\mathbf{D}}_Y^\top\right\|_F^2, \tag{6}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Minimizing $\mathcal{L}_{\text{het}}$ drives $\mathbf{D}$ to discard whatever is predictable from $\mathbf{H}$, thereby preserving only the distinctive information of each modality. To prevent the trivial solution in which the heterogeneous branch collapses to zero, the columns of $\bar{\mathbf{H}}$ and $\bar{\mathbf{D}}$ are ℓ_2-normalized before computing the Gram matrices.

2.4. Representation Fusion

These four tensors are concatenated along the channel dimension:

$$\mathbf{F}_{\text{cat}} = \text{Cat}\left[\mathbf{H}_X, \mathbf{H}_Y, \mathbf{D}_X, \mathbf{D}_Y\right], \tag{7}$$

where $\text{Cat}[\cdot]$ denotes channel-wise concatenation. The concatenated feature tensor $\mathbf{F}_{\text{cat}} \in \mathbb{R}^{H\times W\times 4d_r}$ contains the homogeneous and heterogeneous information learned from both input modalities.

Finally, the target HR-HSI is reconstructed as

$$\mathbf{Z} = \mathcal{D}_{\text{rec}}\left(\mathbf{F}_{\text{cat}}\right), \tag{8}$$

where $\mathcal{D}_{\text{rec}}(\cdot)$ denotes the reconstruction decoder and $\mathbf{Z} \in \mathbb{R}^{H\times W\times C}$ is the reconstructed HR-HSI.

2.5. Loss Function

To preserve consistency between the reconstructed image and the two observations, $\mathbf{Z}$ is projected into the LR-HSI and HR-MSI domains:

$$\begin{aligned}\widetilde{\mathbf{Y}} &= \mathcal{D}_s\left(\mathbf{Z}\right),\\ \widetilde{\mathbf{X}} &= \mathcal{D}_\lambda\left(\mathbf{Z}\right),\end{aligned} \tag{9}$$

where $\mathcal{D}_s(\cdot)$ denotes spatial degradation through point spread function filtering and downsampling, while $\mathcal{D}_\lambda(\cdot)$ denotes spectral degradation according to the spectral response function. The degraded images $\widetilde{\mathbf{Y}} \in \mathbb{R}^{h\times w\times C}$ and $\widetilde{\mathbf{X}} \in \mathbb{R}^{H\times W\times c}$ have the same dimensions as $\mathbf{Y}$ and $\mathbf{X}$, respectively.

The observation consistency loss is defined as

$$\mathcal{L}_{\text{obs}} = \left\|\mathbf{Y} - \widetilde{\mathbf{Y}}\right\|_F^2 + \left\|\mathbf{X} - \widetilde{\mathbf{X}}\right\|_F^2, \tag{10}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and $\mathcal{L}_{\text{obs}} \in \mathbb{R}$ measures the spatial and spectral consistency of the reconstructed HR-HSI.

The complete objective function is formulated as

$$\mathcal{L} = \mathcal{L}_{\text{obs}} + \alpha_1\mathcal{L}_{\text{hom}} + \alpha_2\mathcal{L}_{\text{het}}, \tag{11}$$

where $\mathcal{L}$ is the total training loss, and α_1 and α_2 are two hyperparameters controlling the contributions of the homogeneous alignment loss and the heterogeneous decorrelation loss, respectively.

3. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we validate the performance of the proposed method with Washington DC Mall dataset. We perform a simulation experiment by intercepting a $200\times200\times191$ subgraph

Table 1. Performance of different HSI-SR methods. The best one is shown in bold.

Methods	$PSNR\uparrow$	$SAM\downarrow$	$ERGAS\downarrow$
CNMF	34.1646	2.4640	2.0397
GSA	34.4956	2.6856	2.5449
NonRegSRNet	36.1375	4.6195	2.1254
UDALN	36.9354	1.3492	1.9997
EU2ADL	37.9119	1.2145	1.9953
SHHF	**38.3391**	**0.9985**	**1.5746**

from the original dataset to obtain the input data. The HR-MSI data contains 8 bands and the frame contains 200×200 pixels, and the LR-HSI data contains 191 bands and the frame contains 25×25 pixels. Our network is implemented in the PyTorch framework using the Adam optimizer. The number of training times is set to 10000 and the learning rate starts at 0.003 and gradually decreases to 0.

We selected five representative HSI super-resolution methods as baselines, including CNMF [10], GSA [19], NonRegSRNet [20], UDALN [21], and EU2ADL [22]. Among them, CNMF and GSA are conventional non-deep-learning methods, whereas NonRegSRNet, UDALN, and EU2ADL are deep-learning-based approaches.

The reconstruction performance of these methods was evaluated quantitatively and visually. For quantitative evaluation, we adopted three widely used metrics: peak signal-to-noise ratio (PSNR), spectral angle mapper (SAM), and erreur relative globale adimensionnelle de synthèse (ERGAS). PSNR evaluates spatial reconstruction fidelity, SAM measures spectral distortion, and ERGAS reflects the overall reconstruction error across spectral bands. As reported in Table 1, SHHF achieves the best performance for all three metrics, while UDALN and EU2ADL provide the most competitive results among the compared methods.

Fig. 2 presents the visual comparison, including RGB composites and three types of error maps: SAM maps, MRAE maps, and residual maps of the 40th spectral band. Although all methods generate visually plausible RGB images, their reconstruction differences become more apparent in the error maps, where darker regions indicate larger errors. The deep-learning-based methods generally produce lower reconstruction errors than the conventional methods. Among all the compared approaches, SHHF yields the smallest spatial and spectral errors, which is consistent with the quantitative results in Table 1.

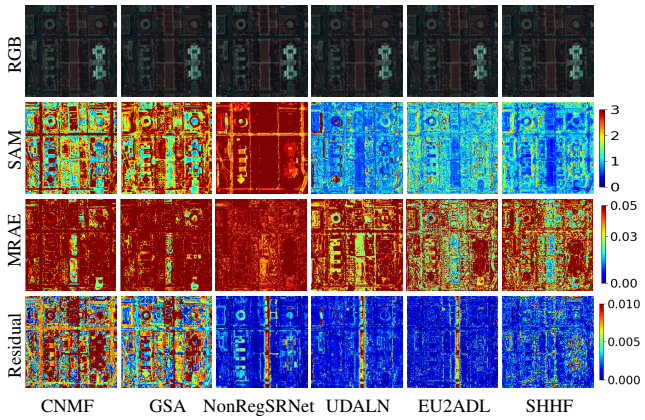


Fig. 2. Visual comparisons of different methods on the Washington DC Mall dataset. The first row shows the RGB images of Hr-HSI reconstructed by different methods (R:55, G:25 and B:15). The second line is the heatmap of the SAM. The third line is the heatmap of MRAE. The fourth line is the residual heatmap at band 40.

4. CONCLUSION

In this paper, we propose a Similarity-Guided Homogeneous-Heterogeneous Fusion Network (SHHF) for fusing an LR-HSI and an HR-MSI to reconstruct an HR-HSI. SHHF employs two projection-based contrastive objectives to learn the two types of representations: a cross-modal InfoNCE loss with a shared projection head aligns the homogeneous information, and a subspace decorrelation loss keeps the heterogeneous representations source-specific. Future work will investigate hard-negative mining for the InfoNCE objective, learnable projection heads, and more efficient representation learning to further improve the generalization capability of the proposed method.

5. REFERENCES

[1] Behnood Rasti, Danfeng Hong, Renlong Hang, Pedram Ghamisi, Xudong Kang, Jocelyn Chanussot, and Jon Atli Benediktsson, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, 2020.

[2] Yijun Yan, Jinchang Ren, Barry Harrison, Oliver Lewis, Emanuele Trucco, Guofang Wang, Yi Ma, and Yutang Ma, "Hyperspectral imaging and machine learning for non-destructive phenolic compounds measurement in peat toward smart whisky manufacturing," *IEEE Trans. Ind. Inform.*, 2025.

[3] Jiaxin Li, Ke Zheng, Lianru Gao, Li Ni, Min Huang, and Jocelyn Chanussot, "Model-informed multi-stage

unsupervised network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, 2024.

[4] Fangyuan Lei, Xuejiao Liao, Xun Liu, Zhaoling Liu, RongJun Chen, Alex Hay-Man Ng, and Jinchang Ren, "Spectral-spatial fusion using separable large kernel attention convolution and simplified transformer for hyperspectral image classification," *Pattern Recognit.*, p. 114795, 2026.

[5] Chengle Zhou, Zhi He, Jian Dong, Yunfei Li, Jinchang Ren, and Antonio Plaza, "Low-rank and sparse representation meet deep unfolding: A new interpretable network for hyperspectral change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, pp. 1–16, 2025.

[6] Erlei Zhang, Xinyu Li, He Zong, Yuan Gao, and Jinchang Ren, "Disentangling representation and cross-domain adaptation for unsupervised change detection in multimodal remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 237, pp. 22–33, 2026.

[7] Yu Huo, Chenhao Wang, Min Zhang, Jinchang Ren, and Hai Wang, "Ddbmnet: Dual-domain background modeling network with anomaly suppression for hyperspectral anomaly detection," *IEEE Trans. Circuits Syst. Video Technol.*, 2026.

[8] Yuanchao Su, Sheng Li, Yicong Zhou, Lianru Gao, Mengying Jiang, Xu Sun, Haiwei Li, and Enke Hou, "Dilated transformation-guided unsupervised multimodal learning for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, pp. 1–15, 2025.

[9] Olivier Berné, A Helens, Paolo Pilleri, and Christine Joblin, "Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy," in *Proc. Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens. (WHISPERS)*. IEEE, 2010, pp. 1–4.

[10] Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, 2011.

[11] Charilaos I Kanatsoulis, Xiao Fu, Nicholas D Sidiropoulos, and Wing-Kin Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, 2018.

[12] Frosti Palsson, Johannes R Sveinsson, and Magnus O Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, 2017.

[13] Jin-Fan Hu, Ting-Zhu Huang, Liang-Jian Deng, Hong-Xia Dou, Danfeng Hong, and Gemine Vivone, "Fusformer: A transformer-based fusion network for hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[14] Jin Xu, Yuanchao Su, Xu Sun, Jiaxin Li, Haiwei Li, Jianjian Gao, Xiaohua Feng, and Ronghua Liu, "Fusgat: Graph attention-based fusion network for unsupervised hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 22, pp. 1–5, 2025.

[15] Sen Jia, Zhichao Min, and Xiyou Fu, "Multiscale spatial–spectral transformer network for hyperspectral and multispectral image fusion," *Inf. Fusion.*, vol. 96, pp. 117–129, 2023.

[16] Jianjun Liu, Zebin Wu, Liang Xiao, and Xiao-Jun Wu, "Model inspired autoencoder for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.

[17] Jin Xu, Yuanchao Su, Sheng Li, He Sun, Ke Zheng, and Mengying Jiang, "Tssa-net: two-stream self-attention network for hyperspectral image super-resolution," *J. Appl. Remote Sens.*, vol. 18, no. 4, pp. 042605–042605, 2024.

[18] Jiaxin Li, Ke Zheng, Lianru Gao, Zhu Han, Zhi Li, and Jocelyn Chanussot, "Enhanced deep image prior for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, 2025.

[19] Gemine Vivone, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio A Licciardi, Rocco Restaino, and Lucien Wald, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, 2014.

[20] Ke Zheng, Lianru Gao, Danfeng Hong, Bing Zhang, and Jocelyn Chanussot, "NonRegSRNet: A nonrigid registration hyperspectral super-resolution network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2021.

[21] Jiaxin Li, Ke Zheng, Jing Yao, Lianru Gao, and Danfeng Hong, "Deep unsupervised blind hyperspectral and multispectral data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[22] Lianru Gao, Jiaxin Li, Ke Zheng, and Xiuping Jia, "Enhanced autoencoders with attention-embedded degradation learning for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–17, 2023.