Expert Systems With Applications 299 (2026) 129864

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Deep low-rank tensor embedded network for hyperspectral image super-resolution

Qiang Zhang [a], Xianpeng Zhang [a], Yi Xiao [b,c,d,*], Hongjie Xie [a]

[a] Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, 116026, Dalian, China
[b] School of Computer and Artificial Intelligence, Zhengzhou University, No. 100 Science Avenue, High-tech Zone, 450001, Zhengzhou, China
[c] Engineering Research Center of Intelligent Swarm Systems, Ministry of Education, No. 100 Science Avenue, High-tech Zone, 450001, Zhengzhou, China
[d] National Supercomputing Center, Changchun Road and Fengyang Street, High-tech Zone, 450001, Zhengzhou, China

ARTICLE INFO

Keywords:
Hyperspectral image super-resolution
Convolutional neural networks
Tensor decomposition

ABSTRACT

Recent efforts have witnessed significant progress in deep-learning-based hyperspectral image super-resolution (HSISR). However, most existing methods focus solely on spatial or spectral exploration, while lacks enough consideration of the intrinsic correlation between these aspects. This oversight limits the potential for collaborative optimization, leading to suboptimal feature representations of HSI. Moreover, they mainly engaged in super-resolve the pixel-wise spatial details, neglecting the vital spectral consistency. To mitigate these issues, this paper proposed LRTENet, a novel deep low-rank tensor embedding network for HSISR, which effectively bridges the optimization gap between spatial and spectral features with well-defined low-rank tensor decomposition. Specially, we introduce a low-rank embedding module (LREM) to extract low-rank dependencies across multiple directions facilitating a holistic mapping by adaptively integrating these tensors. This enables our model to generate discriminative spatial-spectral representations for accurate reconstruction. Furthermore, to better preserve the spectral consistency, we incorporate LREM after upsample operation to progressively refine and correct spectral distortion. Extensive experiments demonstrate that LRTENet achieves superior spatial reconstruction and spectral preservation performance, outperforming state-of-the-art methods on various benchmarks, including Chikusei, CAVE, and Pavia.

1. Introduction

Hyperspectral imaging can record multiple narrow and continuous spectral bands in the electromagnetic spectrum, ranging from visible light to near-infrared and even short-wave infrared. Benefiting from these unique advantages, hyperspectral images (HSI) provides rich spatial and spectral information and has been widely used in various applications (Ghamisi et al., 2017), including agriculture (Sahadevan, 2021) and environmental monitoring (Tan et al., 2020), mineral exploration (Hajaj et al., 2024), hyperspectral change detection (Zhou et al., 2025), etc. However, limited by the bandwidth of imaging sensors, HSI often strike back and forth between spatial and spectral resolution (Loncan et al., 2015). Generally, to obtain rich spectral information, it is inevitable to sacrifice spatial resolution, resulting in high-frequency information loss and posing challenges for downstream tasks (Gendy et al., 2023; Liang et al., 2018; Villa et al., 2013), such as spectral unmixing, classification, and object detection. Therefore, it is of practical significant to increase the spatial resolution of HSI.

HSISR offers a cost-effective alternative to hardware improvements, and can be broadly categorized into two typical approaches (Wang et al., 2023b) : single-image hyperspectral super-resolution and fusion-based hyperspectral super-resolution (Wang & Chen, 2024). Fusion-based methods enhance spatial resolution by incorporating high-resolution (HR) panchromatic (PAN) or multispectral images (MSI). Despite recovering richer spatial details, they often require laborious alignment process between HSI and MSI. What's worse, they suffer from severe performance drop when misalignment occurs, resulting in training instability and suboptimal fusion outcomes. In contrast, single-image HSISR could directly reconstruct HR hyperspectral images from low-resolution (LR) inputs, offering greater flexibility and practicality for real-world applications.

Furthermore, these methods can be divided into traditional (Bu et al., 2024) and deep learning-based approaches (Chen et al., 2023a; Yan et al., 2025). Traditional models usually rely on hand-craft priors, e.g., sparse (Akhtar et al., 2015; Dian et al., 2019; Dong et al., 2016; Xu et al., 2019) and low-rank (Dian et al., 2018; Wang et al., 2017; Xue et al.,

* Corresponding author.
E-mail addresses: qzhang95@dlmu.edu.cn (Q. Zhang), roc@dlmu.edu.cn (X. Zhang), yixiao@zzu.edu.cn (Y. Xiao), dlmuxhj@dlmu.edu.cn (H. Xie).

https://doi.org/10.1016/j.eswa.2025.129864
Received 25 March 2025; Received in revised form 2 September 2025; Accepted 22 September 2025
Available online 11 October 2025

2021) priors, to build a mapping between LR and HR HSI. These priors are often served as regularization terms to constrain the ill-posed reconstruction process iteratively. However, they are of limited representation (Ma et al., 2023) and often require substantial computational resources to tame the optimization instability. In contrast, deep learning-based methods, propelled by the success of convolutional neural networks, can directly learn the mapping (Lepcha et al., 2023) from LR to HR HSI using external training data. These methods excel in capturing the nonlinear relationships between spatial and spectral features, significantly outperforming traditional SR models. Nevertheless, most deep learning methods tend to focus solely on recovering pixel-level spatial details, often neglecting the critical spectral consistency (Hu et al., 2024) of HSI. Although some recent works achieve both spatial and spectral feature modeling (Chen et al., 2024a; Liu et al., 2024), they generally optimize these processes independently, failing to fully exploit the intrinsic correlation between spatial and spectral information. This oversight often leads to suboptimal reconstruction results. More specifically, spatial feature learning emphasizes enhancing pixel-wise resolution, while spectral feature learning focuses on restoring spectral bands. The inherent discrepancy poses significant challenge for joint representation and collaborative learning, leading to undesirable spectral inconsistency (Xie et al., 2024). The research motivation of this paper lies in the fact that existing deep learning methods in the task of hyperspectral image super-resolution have difficulties in effectively handling the separation of spatial and spectral information optimization and are unable to fully utilize the intrinsic correlation between them, resulting in poor reconstruction performance.

To address this issue, a straightforward solution is to construct a holistic representation of the spatial and spectral relationships. However, establishing such a mapping is challenging due to the high-rank (Chen et al., 2020; Xue et al., 2019; Zhang et al., 2019) nature of HSI. In this context, there are at least two key challenges: 1) **suboptimal exploration of high-rank data**, and 2) **inaccurate spatial-spectral representation**. More precisely, deep learning-based methods rely heavily on external data to extract high-rank HSIs, which inevitably complicates the learning process, especially in limited HSI data scenarios. Additionally, there is a lack of efficient scheme for modeling joint low-rank spatial and spectral dependencies. Based on these analyses, a natural question arises: *can we develop a model-driven spatial-spectral collaborative representation framework to enhance the reconstruction performance of data-driven networks?*

To answer this question, inspired by the tensor decomposition theory (Kolda & Bader, 2009), this paper proposed to decompose the high-dimensional HSI into multiple low-rank parts for efficient yet effective spatial-spectral representation. Recently, tensor regularization and canonical polyadic (CP) (Kolda & Bader, 2009) decomposition have demonstrated favorable advantages in representing high-rank data with multiple rank-one low-rank tensors. Based on this, we design a deep low-rank tensor embedding network (LRTENet) for HSISR, which effectively extracts the holistic spatial and spectral representation through the low-rank reconstruction method, thus improving the efficiency and accuracy of spatial-spectral information exploration for high-quality reconstruction.

Specifically, we extract multiple low-rank dependencies from the entire contextual HSI to facilitate the learning of the holistic mapping relationship. To achieve this, we develop a low-rank embedding module (LREM), which extracts discriminative rank-one tensors and constructs the mapping through a weighted fusion of these tensors. In LREM, a rank-one tensor generation module (ROM) was devised, which generates rank-one tensors by extracting features in multiple directions and enhancing contextual information from both spatial and channel dimensions. To further retain and utilize rich spectral features, ROM introduces the channel feature retention pooling (CFRP) strategy and employs the enhanced localization information (ELI) module for adaptively fusing channel features and acquiring high-quality spatial information. Furthermore, unlike conventional approaches that perform one-step upsampling directly, we integrate the LREM after multi-step upsampling to refine and correct spectral information, further alleviating the spectral distortion.

To sum up, the contributions of this paper are summarized as follows.

- This paper proposes a deep low-rank tensor embedding network for HSISR. The network decomposes the holistic relationship between spatial and spectral representations into multiple low-rank components through the CP decomposition theory, enabling accurate expression of complex relationships and thus bridging the optimization gap between spatial and spectral representations.
- To generate discriminative low-rank tensors, we design a rank-one tensor generation module (ROM) that constructs low-rank dependencies across multiple dimensions. This module adaptively fuses multi-channel features through attention mechanisms, enhancing the extraction of high-quality spatial details while preserving spectral fidelity.
- On Chikusei, CAVE, and Pavia datasets, LRTENet outperforms state-of-the-art HSISR methods in both quantitative and visual evaluations, validating its effectiveness.

2. Related work

2.1. Deep learning-based single hyperspectral image super-resolution

Inspired by the huge success of super-resolution convolutional neural network (SRCNN) (Dong et al., 2015), deep learning has also demonstrated remarkable potential for single hyperspectral image super-resolution (Xue et al., 2024). HSIs are characterized by their rich spectral information, which offers unique opportunities for leveraging both spatial and spectral features (Zhang et al., 2020, 2024b). Consequently, researchers are motivated to elaborate the network design to effectively exploit spatial-spectral features.

The grouped deep recursive residual network (GDRRN) (Li et al., 2018) employs residual connections and grouped recursive modules to mitigate redundancies in HSI data, thereby reducing computational overhead. Similarly, the 3D fully convolutional cascade cetwork (3DFCCN) (Mei et al., 2017) leverages 3D convolution to extract spatial and spectral context from adjacent channels. However, the inherent computational burden and parameter-intensive nature of 3D convolutions impose significant limitations, hindering optimal performance. To alleviate the challenges posed by high spectral dimensionality, SSPSR (Jiang et al., 2020) incorporates grouped convolutions with shared network parameters and adopts a progressive upsampling strategy. It further employs a channel attention mechanism to explore inter-spectral correlations. In contrast, MCNet (Li et al., 2020) and ERCSR (Li et al., 2021) combine the strengths of 2D and 3D convolutions, where 2D convolutions effectively capture spatial features and 3D convolutions are utilized for local spectral feature extraction, resulting in reduced computational complexity. Nonetheless, theirs feature extraction capability is constrained by the limited receptive field inherent to convolutions. Building upon SSPSR (Jiang et al., 2020), CLSCNet (Xu et al., 2024) integrates ConvLSTM-based (Shi et al., 2015) skip connections to suppress redundant features, while its convolutional modules enhance edge feature extraction, thereby achieving improved super-resolution accuracy. More recently, with the advent of vision transformers, which excel in capturing long-range dependencies, Chen et al. proposed MSDformer (Chen et al., 2023b). This hybrid framework employs CNNs to extract spatial features while leveraging a global spectral transformer to model dependencies across all spectral bands, overcoming the limitations of CNNs in capturing global context. SRDNet (Liu et al., 2024) and CST (Chen et al., 2024a) extend this concept by introducing independent transformers for spatial and spectral dimensions to explicitly capture long-range dependencies within each domain. EigenSR (Su et al., 2025) utilizes pre-trained RGB models to address the issue of data scarcity in HSI. This method is based on spatial-spectral decoupling and can

effectively utilize the pre-trained model while maintaining spectral fidelity. DSDCN (Muhammad et al., 2025) is designed as a lightweight depthwise separable dilated convolutional network. It combines depthwise separable convolutions, residual connections, and dilated convolution fusion to improve spatial resolution.

2.2. Low-rank tensor representation

Low-rank tensor representations have found widespread applications in computer vision tasks due to their ability to efficiently reduce dimensionality and extract meaningful features.

In work (Zhang et al., 2023b), the unsupervised denoising of HSIs based on tensor decomposition for mining spectral low-rank priors (Zhang et al., 2022) and using deep space priors is proposed. In work (Chen et al., 2024b), a denoising and recovery algorithm is proposed to mine image prior information by constructing low-rank tensor through deep learning and the synergistic effect of model-based framework. In work (Zhang et al., 2021a), low-rank tensor singular value decomposition and tensor product are used to excavate the structural properties of multi-temporal images, and depth priors are combined to remove thick clouds from time series images. The work (Xue et al., 2021) employs a novel subspace clustering method with structured sparse low-rank representation for fusion-based hyperspectral image super-resolution. It fully considers the spatial and spectral subspace low-rank relationships among the available HR-MSI, LR-HSI, and the latent HSI. The work (Yan et al., 2023) employs low-rank property embedding to minimize the impact of spectral variations and uses adaptive non-negative sparse coefficients derived from the corresponding HR-MSI to further reconstruct the desired HSI, thereby achieving spectral super-resolution.

By leveraging tensor decomposition theories, high-dimensional tensors can be expressed as combinations of multiple low-rank sub-tensors, facilitating the representation of the most salient data components while suppressing redundancy. The core principle of tensor decomposition lies in breaking down the original tensor into smaller, more manageable components. Among the popular decomposition techniques, tucker decomposition represents a tensor as the product of multiple matrices and a core tensor, capturing its key structures. On the other hand, CP decomposition expresses a tensor as the sum of a set of rank one tensors, serving as a specific case of tucker decomposition. CP decomposition is particularly advantageous in scenarios requiring compact representations of high-dimensional data.

In work (Chen et al., 2020), a network based on CP decomposition to mine context features is proposed for semantic segmentation. In work (Zhang et al., 2021b), a tensor-low-rank prior learning network is proposed for snapshot hyperspectral imaging based on CP decomposition and generation of discriminative rank tensor. In work (Dian et al., 2024), multidimensionwise multihead self-attention is introduced in generating basis vectors to improve the ability of CP decomposition to convey information, and a spectral super-resolution network based on deep low-rank tensor representation is designed.

Building on these approaches, we are motivated to explore the potential of low-rank tensor representations in the domain of hyperspectral super-resolution. In this paper, we employ the CP decomposition framework to effectively approximate the complex mapping relationships by capturing a global set of rank-one tensors, thus enhancing the resolution of hyperspectral images.

3. Methodology

3.1. Overview

As illustrated in Fig. 1, the proposed LRTENet consists of three major components: shallow feature extraction module, deep feature extraction (DF) module, reconstruction module.Given a predefined scaling factor s and low-resolution HSI $I_{LR} \in \mathbb{R}^{C\times h\times w}$, LRTENet is designed to learn the mapping function $F_{\text{LRTENet}}(\cdot)$, to produce the high-resolution HSI $I_{SR} \in \mathbb{R}^{C\times sh\times sw}$. This process can be mathematically formulated as:

$$I_{SR} = F_{\text{LRTENet}}(I_{LR}, s) \tag{1}$$

where h,w, and C represent the height, width, and the number of spectral bands of the HSI, respectively. The overall model flow is detailed as follows.

The low-resolution HSI I_{LR} is first processed through a shallow feature extraction layer, represented by $F_{\text{Extraction}}(\cdot)$, which is a convolutional operation. The resulting feature x_0 is computed as:

$$x_0 = F_{\text{Extraction}}(I_{LR}) \tag{2}$$

The extracted shallow features x_0 are then sent through a cascade of LREM in the deep feature extraction function $F_{\text{DF}}(\cdot)$, resulting in deeper feature representations:

$$x_u = F_{\text{DF}}(x_0) \tag{3}$$

The high-resolution HSI I_{SR} is reconstructed by combining the deep features x_u and the upsampled shallow features $x_0\uparrow$ through the reconstruction module $F_{\text{Reconstruction}}(\cdot)$. This process allows for the integration of both deep and shallow information for better reconstruction.

$$I_{SR} = F_{\text{Reconstruction}}(x_u, x_0\uparrow) \tag{4}$$

Here, $x_0\uparrow$ refers to the upsampled shallow features obtained via the pixel shuffle layer. The reconstruction module consists of two convolutional layers, aiming to unifying the channel number of x_u and $x_0\uparrow$,

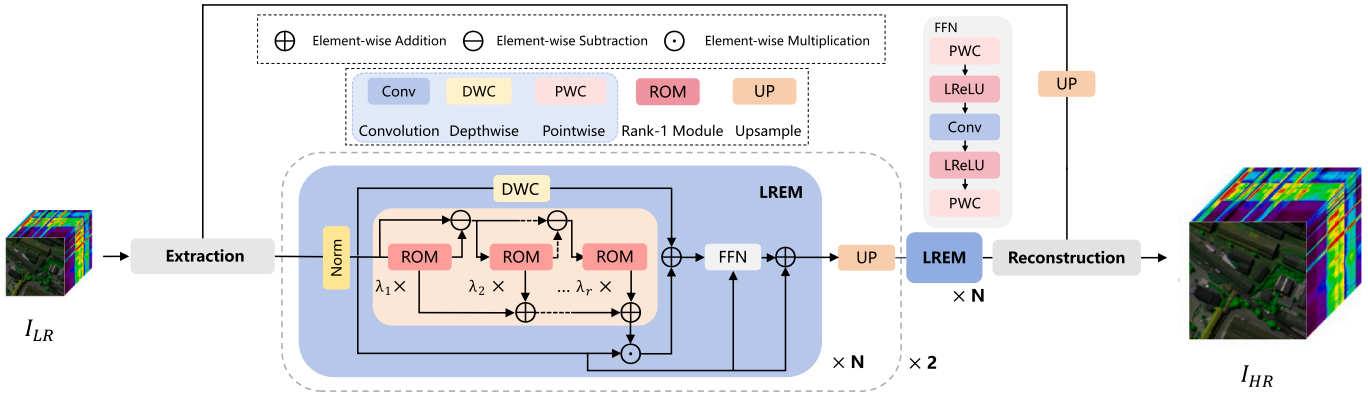


Fig. 1. Overview of the proposed LRTENet. (1) Shallow feature extraction module: This stage extracts the initial features from the input data. (2) Deep feature extraction module: This part consists of a cascade of LREM modules.LREM decomposes complex mappings into r low-rank mappings through the CP decomposition theory. The low-rank mappings are extracted by the ROM, fused by adding them with adaptive weights, and the fused features are enhanced by the FFN. Each LREM is designed to capture the overall mapping relationships of the features, facilitating the seamless integration of spatial and frequency information. (3) Reconstruction module: This final stage reconstructs the output.

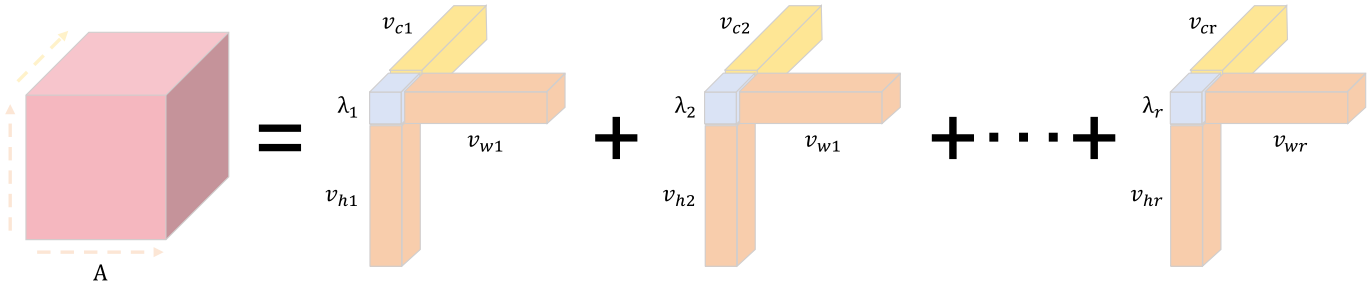


Fig. 2. Illustration of the third - order tensor CP decomposition, where the symbolic notations correspond to the respective equation symbols.

followed by residual learning to stabilize the optimization process. It is worth noting that during our experiments, we found that residual learning methods may vary for different datasets. More specific implementation details are provided in the experimental section.

3.2. Low-rank embedding module

Our objective is to maintain spectral consistency by jointly modeling spatial and spectral features. However, capturing such spatial-spectral mapping presents a significant challenge due to the high-rank nature of hyperspectral feature representations. This complexity arises from the high dimensionality of hyperspectral data and is further exacerbated by the expansion of feature channels during shallow feature extraction. Directly learning such complex mappings is computationally prohibitive and tend to reaching suboptimal performance. Inspired by the CP decomposition theory and recent advancements in tensor reconstruction (Chen et al., 2020; Zhang et al., 2021b), this study adopts a low-rank representation strategy to simplify the high-dimensional data modeling procedure. This approach not only mitigates the challenges of high-rank tensor learning but also facilitates a more efficient yet effective representation of the spatial-spectral correlations.

Before delving into the specific implementation of LREM, we first present the theoretical formulation of the CP decomposition for a 3rd-order tensor.

Assume that in the row, column, and spectral directions, there are $3r$ vectors $v_{ci} \in \mathbb{R}^c$,$v_{hi} \in \mathbb{R}^h$ and $v_{wi} \in \mathbb{R}^w$,where $i \in r$ and r is the predefined rank. These vectors are the CP decomposition components of the tensor $A \in \mathbb{R}^{c\times h\times w}$. The tensor A can then be reconstructed from these decomposition components as defined by:

$$A = \sum_{i=1}^{r} \lambda_i v_{ci} \otimes v_{hi} \otimes v_{wi} \tag{5}$$

$\otimes$ represents the Kronecker product operatio, and v_{ci}, v_{hi}, and v_{wi} are referred to as rank-one Kronecker basis vectors. The expression $v_{ci} \otimes v_{hi} \otimes v_{wi}$ represents a rank-one tensor. This formula indicates that a 3rd-order high-rank tensor can be expressed as a weighted sum of several low-rank tensors. λ_i is the weight factor for each rank-one tensor. This process is illustrated in Fig. 2, where the gray-blue color represents the weights, the vertical orange vector represents v_{hi}, the horizontal orange vector represents v_{wi}, and the yellow vector represents v_{ci}.

Based on the CP decomposition representation of third-order tensors mentioned above, high-rank tensors can be effectively approximated as a combination of multiple rank-one low-rank tensors. To leverage this property, we designed the LREM.

As illustrated in Fig. 1 and implementation details provided in Algorithm 1, the LREM comprises three components: feature normalization, low-rank feature reconstruction, and feature enhancement.

1) **Feature normalization**: This stage aims employing normalization (Zhang et al., 2025) layers to accelerate the optimization process (Wang et al., 2024). This operation can be mathematically expressed as:

$$x_{norm} = F_{\text{Norm}}(x_{input}) \tag{6}$$

Algorithm 1 Implementation of low-rank embedded modules.

Require: Input tensor x_{input}
Hyperparameter $r = 5$ (number of low-rank tensors)
Ensure: Output tensor x_{output}
Feature normalization:
$x_{norm} \leftarrow F_{\text{Norm}}(x_{input})$
Low-rank feature reconstruction:
Initialize $i = 1$, $x_r = 0$, $\text{mid}_1 = x_{norm}$, $\text{mid}_0 = 0$
for $i = 1 \rightarrow r$ **do**
  $\text{mid}_1 \leftarrow \text{mid}_1 - \text{mid}_0$
  $\text{mid}_0 \leftarrow F_{\text{ROM}_i}(\text{mid}_1)$ ▷ F_{ROM}: Rank-one Tensor Generation Module
  $x_r \leftarrow x_r + \lambda_i \cdot \text{mid}_0$ ▷ λ_i: learnable weight
end for
Feature enhancement:
$x_g \leftarrow x_r \odot x_{\text{norm}} + F_{\text{DWC}}(x_{\text{norm}})$ ▷ F_{DWC}: Depthwise Convolution
$x_{\text{output}} \leftarrow F_{\text{FFN}}(\text{Concat}(x_{\text{input}}, x_g)) + x_{\text{input}}$ ▷ F_{FFN}: Feed-Forward Network
return x_{output}

where F_{Norm} refers to layer normalization. x_{input} represents the input features of the LREM. The output features after normalization are denoted as x_{norm}.

2) **Low-rank feature extraction**: the high-rank mapping is decomposed into r low-rank subproblems. This approach learns the low-rank dependencies between joint spatial and spectral features and aggregates them with adaptive weights to reconstruct the high-rank mapping. Each of these r low-rank subproblems is addressed using rank-one tensor generation modules (ROM). The ROM extracts contextual information, incorporating both channel and spatial dimensions, to construct rank-one tensors a_i. Additionally, residual learning is employed within ROM to enhance the discriminability of individual rank-one tensors (Zhang et al., 2021b), ensuring that each tensor captures unique and significant features. The rank-one tensors are subsequently aggregated using learnable (Chen et al., 2020; Xiao et al., 2025b) parameters λ_i, allowing the model to reconstruct the target tensor with discriminative capability. This method, leveraging learnable parameters for aggregation, surpasses traditional convolution-based aggregation methods (Dian et al., 2024; Zhang et al., 2021b) by facilitating the exploration of richer rank-based features. The process is mathematically expressed as:

$$a_i = F_{\text{ROM}_i}\left(x_{norm} - \sum_{j=0}^{i-1} a_j\right), \tag{7}$$

$$x_r = \sum_{i=0}^{r-1} \lambda_i a_i \tag{8}$$

Eq. (7) describes the construction of discriminative rank-one tensors within the ROM, which extracts low-rank features by leveraging contextual spatial and spectral information. This process enhances the model's ability to capture unique and diverse feature

representations. Eq. (8) demonstrates the aggregation and mapping relationship for reconstructing the target tensor, where the learnable parameters dynamically adjust the contribution of each rank-one tensor to the final representation.

3) **Feature enhancement:** This component focuses on generating the holistic representations and refining them through a feedforward network (FFN). Specifically, the joint spatial and channel features are obtained by performing element-wise multiplication between the x_r and the initial input features. Considering that low-rank tensor representations may inherently discard some feature components, we draw inspiration from existing approaches (Han et al., 2024), and integrate depthwise convolution (DWC) modules. The DWC modules are utilized to preserve feature diversity and recover missing components effectively. The process can be expressed mathematically as:

$$x_g = x_r \odot x_{input} + F_{\mathrm{DWC}}(x_{input}) \tag{9}$$

FFN plays a pivotal role in enhancing feature representation (Chen et al., 2025; Han et al., 2024; Neupane et al., 2024; Zhang et al., 2023a; Zhou et al., 2024; Zhu & Liu, 2025). The architecture of the FFN is shown in Fig. 1. It consists of three parts: pointwise convolution (PWC), a 3×3 convolution, and leaky rectified linear unit (LReLU) activation. The features from the overall mapping are concatenated along the channel dimension with the input features of the LREM. This combined feature is subsequently passed through the FFN, allowing the network to adaptively refine and enhance the extracted features. Additionally, a skip connection is employed to directly link the input of the LREM to its output, facilitating efficient reconstruction and making it easier to retain valuable information during the SR process. This process can be mathematically expressed as:

$$x_{output} = F_{\mathrm{FFN}}(F_{\mathrm{Cat}}(x_{input}, x_g)) + x_{input} \tag{10}$$

where x_g denotes the joint spatial-spectral features extracted by the LREM. Cat refers to the concatenation operation,x_{output} is the output feature of the LREM.

Most existing networks employ upsampling layers that primarily target the spatial dimension, while largely overlooking the intricate interdependencies between spatial and spectral information in hyperspectral images. This limitation often compromises spectral consistency during reconstruction. To address this challenge, we incorporate multi-layer LREMs following the upsampling process. These modules are specifically designed to iteratively refine the spatial-spectral representations at the target resolution, thereby enhancing the spectral fidelity.

The deep feature extraction phase is composed of two sequential stages, which collaboratively process features to yield x_u.

The first stage is the upsampling phase, which employs a progressive upsampling strategy designed to decompose the one-step reconstruction into incremental steps, mitigating the difficulties of high-resolution feature reconstruction. Following upsampling, features at the target spatial resolution are further refined to enhance spectral consistency. This phase corrects and adjusts the spectral representations using stacked LREMs. The pixel shuffle layer is utilized within the upsampling layers for efficient upscale, while LREMs are applied iteratively to extract and refine deep features.

$$x_u = F_{\mathrm{SC}}(F_{\mathrm{UP}}(x_0)) \tag{11}$$

where F_{UP} represents the upsampling phase, which adopts a progressive upsampling strategy and thus includes two sub-stages of feature extraction. F_{SC} refers to the refinement and correction of spectral information, where the pixel shuffle method is applied in the upsampling layer. Deep features are extracted using LREM at each stage.

3.3. Rank-one tensor generation module

As shown in Fig. 3, ROM draws inspiration from prior works (Chen et al., 2020; Xiao et al., 2025b). We adopted a compromise method, which retained the channel information of the basis vectors on the basis of obtaining the basis vectors by global pooling in priorworks (Chen et al., 2020; Xiao et al., 2025b). Specifically, we transform the 3D feature into a low-dimensional feature representation and construct the decomposition framework of the ROM. Unlike existing approaches (Chen et al., 2020; Xiao et al., 2025b), which primarily utilize global pooling (Zhang et al., 2024a; Zhao et al., 2017) to extract coarse contextual information, our ROM considers the detailed extraction of row and column features, critical for hyperspectral image data that contains abundant spectral information. To address the limitations of global pooling in hyperspectral image processing, we adopt a channel-preserving pooling strategy. This method ensures that while global contextual features (Fang et al., 2024) are extracted through pooling operations for row and column dimensions, the spectral information, pivotal for hyperspectral image, can be carefully preserved. This dual focus on spatial and spectral information ensures robust feature extraction, specifically for HSISR tasks. The process of globally capturing features in different directions through channel-preserving pooling can be expressed as follows:

$$x_c, x_h, x_w = F_{\mathrm{CFRP}}(x_{input}) \tag{12}$$

Eq. (12) represents the process of globally capturing features in different directions through channel-preserving pooling. And this process can be demonstrated by the operation of CFRP shown in Fig. 3, yielding $x_c \in R^{C\times 1\times 1}$, $x_h \in R^{C\times H}$, $x_w \in R^{C\times W}$.

To further enhance the extraction of spatial features while preserving channel information, we introduce ELI module. This module adaptively fuses rich spectral features to extract the row and column basis vectors

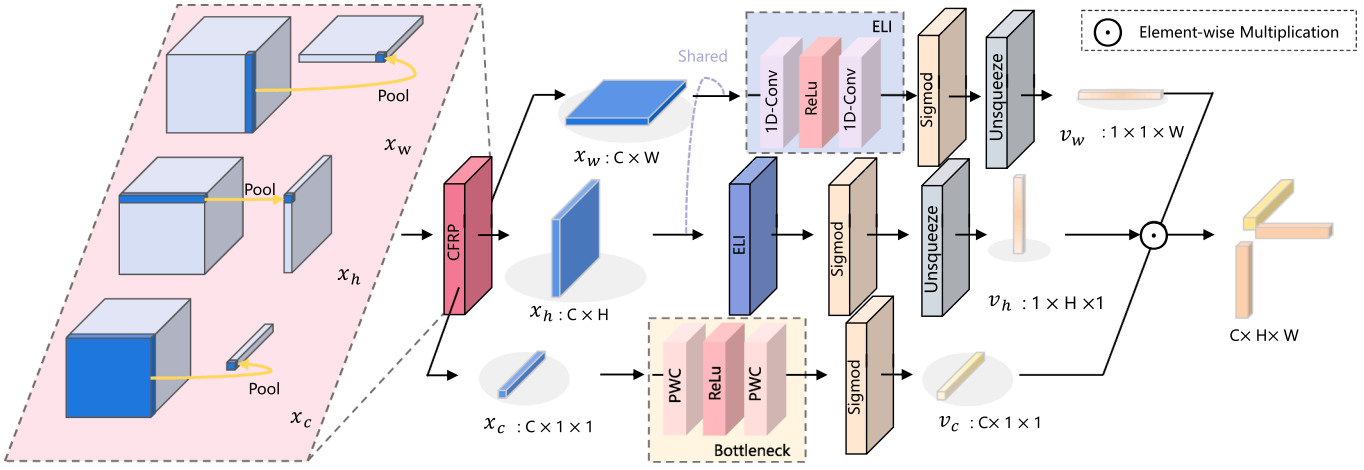


Fig. 3. The proposed rank-one tensor generation module (ROM).

in the spatial domain, as illustrated by the following equations:

$$\begin{cases} v_h = u(\sigma((F_{\mathrm{ELI}}(x_h)))), \\ v_w = u(\sigma((F_{\mathrm{ELI}}(x_w)))) \end{cases} \tag{13}$$

where F_{ELI} represents the ELI module, $\sigma(\cdot)$ denotes sigmoid activation, and $u(\cdot)$ means adding a tensor dimension to match subsequent calculations.

The ELI module employs a Conv1D-ReLU-Conv1D architecture, utilizing 1D convolutional kernels with a size of 7 rather than conventional 2D convolutions. This choice is motivated by the computational efficiency of 1D convolutions, which are significantly more lightweight (Shi et al., 2024). By stacking two large-kernel convolutional layers (Xiao et al., 2024) and adjusting the number of channels in the hidden layers, the ELI module progressively aggregates spatial information. This design effectively captures large-scale spatial features within HSI objects, while enhancing the interaction and localization capabilities of embedded spatial information. Furthermore, the ELI module facilitates the learning of pixel-level unique weights, enabling precise feature extraction.

Furthermore, since both row and column features are extracted in the spatial dimension, ELI serves as a shared (Xu & Wan, 2024) module for extracting these row and column features. Inspired by the SENet (Hu et al., 2018) architecture, which extracts channel weights to capture inter-channel relationships, we employ a Bottleneck structure following the global pooling operation to derive basis vectors along the channel dimension. This design reduces the number of channels in the hidden layers, enabling efficient extraction of channel-specific features (Nandi et al., 2023; Xiao et al., 2025a). The process can be mathematically expressed as:

$$v_c = \sigma((F_{\mathrm{Bottleneck}}(x_c))) \tag{14}$$

where $F_{\mathrm{Bottleneck}}$ denotes the bottleneck (Hu et al., 2018; Wang et al., 2023a) structure, which is specifically implemented as PWC-ReLU-PWC structure, and $\sigma(\cdot)$ denotes sigmoid activation.

To facilitate the construction of rank-one tensors within the network, a broadcasting mechanism is employed during the element-wise multiplication of tensors. This operation is analogous to the decomposition outlined in Eq. (5), enabling the generation of rank-one tensors. The process is defined as:

$$x_{output} = \lambda v_c \odot v_h \odot v_w \tag{15}$$

where $\odot$ denotes the Hadamard product.$v_c \in \mathbb{R}^{C\times1\times1}$, $v_h \in \mathbb{R}^{1\times H\times1}$, $v_w \in \mathbb{R}^{1\times1\times W}$ are the basis tensors obtained through the operations in Eqs. (14) and (13) in different directions. σ represents the sigmoid function used for feature normalization. x_{output} is the resulting rank-one tensor.

4. Experiments

4.1. Datasets and settings

4.1.1. Datasets

The Chikusei (Yokoya & Iwasaki, 2016) dataset was acquired using the Headwall Hyperspec-VNIR-C imaging sensor, capturing agricultural and urban areas in Chikusei, Ibaraki Prefecture, Japan. The dataset spans a spectral range of 363–1018 nm across 128 bands, with a spatial resolution of 2517 × 2335 pixels. The CAVE (Yasuma et al., 2010) dataset was collected using a cooled CCD camera and comprises diverse real-world materials and objects. It covers a spectral range of 400–700 nm across 31 spectral bands. Each hyperspectral image has a spatial resolution of 512 × 512 pixels, and the dataset includes 32 hyperspectral scenes.The Pavia[1] Center dataset was captured using a reflective optical system imaging spectrometer sensor. After removing water vapor absorption and noisy bands, the dataset contains 102 spectral bands from an original 115. The spatial resolution of the hyperspectral images is 1096 × 1096 pixels.

4.1.2. Implementation details

1) **Model details:** Unless explicitly stated, the convolution kernel size throughout the network is uniformly set to 3 × 3. For specific convolution kernel sizes, detailed descriptions are provided within the text or corresponding figs. The number of feature channels is set to 256. The deep feature extraction process is divided into two stages. The first stage employs a progressive upsampling strategy, further subdivided into two sub-stages, followed by a spectral-preserving refinement stage. Each stage incorporates three LREMs. The rank (r) for each ROM is set to 5, while the channel reduction ratio (d) in the Bottleneck and ELI modules is set to 16. For residual learning, distinct strategies are adopted based on dataset characteristics: 1) For the Chikusei and Pavia datasets, a shallow-feature pixel-shuffling upsampling approach is utilized, and 2) For the CAVE dataset, bicubic-interpolated upsampled images of the LR inputs are employed.
2) **Training details:** For hyperspectral image super-resolution, the loss function is typically defined using either the l_1 or the l_2. Since the l_2 often results in overly smoothed outputs, this paper adopts the l_1 as the loss function for the model, defined as:

$$Loss = \frac{1}{B}\sum_{i=1}^{B} \|X_i - X_i^{gt}\|_1 \tag{16}$$

Let B represent the batch size, and i represent the index of each image within the batch. $X_i$denotes the image generated by the model after super-resolution, while X_i^{gt} denotes the ground truth image.

The model is implemented in PyTorch and optimized using the Adam optimizer. All experiments are conducted on the same machine with the following specifications: an i9 - 12900K CPU, 64 GB of RAM, a 3090 GPU with 24 GB of video memory, and CUDA version 12.6. During the training process, data augmentation was applied to enhance the model's generalization ability, and the specific implementation can be referred to in works such as SSPSR (Jiang et al., 2020).

4.1.3. Evaluation metrics

We evaluate the model's performance in both spatial and spectral domains using six widely adopted metrics: peak signal-to-noise ratio (PSNR), structural similarity (SSIM), spectral angle mapper (SAM), cross correlation (CC), erreur relative globale adimensionnelle de synthese (ERGAS), and root mean squared error (RMSE). The optimal values for these metrics are as follows: +∞, 1, 0, 1, 0, and 0.

4.2. Experimental results

We compare our approach against traditional bicubic interpolation and six representative deep learning-based methods: 3DFCNN (Mei et al., 2017), GDRRN (Li et al., 2018), MCNet (Li et al., 2020), EUNet (Liu et al., 2023), CST (Chen et al., 2024a), and SRDNet (Liu et al., 2024). All models were trained from scratch. The qualitative and visual results across various datasets are presented in Tables 1, 2, and 3 where our method consistently outperforms the others in terms of both spatial and spectral performance.

4.2.1. Experimental results on chikusei dataset

The original Chikusei dataset has dimensions of 2517 × 2335 × 128 pixels. To address edge artifacts, we crop the central region, yielding a sub-image of size 2304 × 2048×128 pixels. Following the partitioning strategy in SSPSR (Jiang et al., 2020), this sub-image is divided into a training set and a testing set. The testing set consists of four non-overlapping hyperspectral images, each with a size of 512 × 512 × 128 pixels. The remaining region of the sub-image is partitioned into image patches with overlapping regions (overlap size being half the patch size), which are used as high-resolution reference images during training. LR

[1] https://ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

Table 1
Quantitative comparison of different methods on the chikusei dataset.

Methods	Scale = 4					
	PSNR	SSIM	SAM	CC	ERGAS	RMSE
Bicubic	37.6377	0.8949	3.4040	0.9212	6.7564	0.0159
3DFCNN	38.5325	0.9154	3.1786	0.9349	6.0603	0.0141
GDRRN	39.9446	0.9385	2.5405	0.9524	5.1929	0.0118
MCNet	39.5699	0.9322	2.7359	0.9483	5.3762	0.0126
EUNet	39.8675	0.9383	2.4926	0.9515	5.2681	0.0119
CST	40.1551	0.9422	2.3637	0.9544	5.0711	0.0116
SRDNet	40.0837	0.9411	2.4274	0.9538	5.1310	0.0117
Ours	**40.6221**	**0.9474**	**2.2312**	**0.9589**	**4.7978**	**0.0110**
Methods	Scale = 8					
	PSNR	SSIM	SAM	CC	ERGAS	RMSE
Bicubic	34.5051	0.8069	5.0356	0.8313	9.6969	0.0223
3DFCNN	34.9175	0.8203	4.8227	0.8460	9.2113	0.0213
GDRRN	35.7307	0.8481	4.1867	0.8731	8.4221	0.0194
MCNet	35.4367	0.8368	4.4552	0.8643	8.6612	0.0201
EUNet	35.5846	0.8472	4.1237	0.8691	8.5942	0.0196
CST	35.7359	0.8494	4.1774	0.8738	8.4057	0.0193
SRDNet	35.6839	0.8490	4.1048	0.8726	8.4818	0.0194
Ours	**35.9645**	**0.8591**	**3.8733**	**0.8802**	**8.2064**	**0.0188**

Table 2
Quantitative comparison of different methods on the cave dataset.

Methods	Scale = 4					
	PSNR	SSIM	SAM	CC	ERGAS	RMSE
Bicubic	35.3132	0.9370	4.2665	0.9871	5.3941	0.0198
3DFCNN	37.0362	0.9487	4.1423	0.9908	4.3914	0.0165
GDRRN	37.8173	0.9528	3.9974	0.9922	4.0173	0.0150
MCNet	39.6099	**0.9645**	3.2556	0.9941	3.4307	0.0126
EUNet	38.6248	0.9601	3.5656	0.9932	3.6998	0.0138
CST	38.8304	0.9608	3.3091	0.9935	3.6326	0.0136
SRDNet	39.1078	0.9615	3.4591	0.9936	3.5477	0.0131
Ours	**40.1128**	**0.9645**	**3.1788**	**0.9943**	**3.2782**	**0.0122**
Methods	Scale = 8					
	PSNR	SSIM	SAM	CC	ERGAS	RMSE
Bicubic	30.7284	0.8632	5.9042	0.9672	6.2281	0.0240
3DFCNN	31.8507	0.8831	5.7864	0.9736	7.6642	0.0289
GDRRN	32.4870	0.8849	5.8649	0.9758	7.1145	0.0278
MCNet	34.3306	0.9148	4.6779	0.9814	6.0951	0.0233
EUNet	33.5809	0.9055	4.9633	0.9793	6.4811	0.0248
CST	33.7545	0.9072	4.6384	0.9805	6.3193	0.0244
SRDNet	33.9380	0.9084	4.9519	0.9806	6.2281	0.0240
Ours	**34.5240**	**0.9157**	**4.5259**	**0.9821**	**5.9550**	**0.0230**

HSI are generated by downsampling these patches and applying bicubic interpolation.For the experiments, we test scaling factors of 4× and 8×. In the case of the 4× scaling factor, the low-resolution images have an input resolution of 16×16 pixels, with an output resolution of 64×64 pixels. For the 8× scaling factor, the input resolution is 16×16 pixels, and the output resolution is 128×128 pixels.

Table 1 reports the average objective performance of all comparative algorithms on the test images, with boldface highlighting the best results and underscores indicating the second-best results. The performance metrics on the Chikusei dataset for both 4× and 8× scaling factors demonstrate that our method outperforms all others in both spatial and spectral domains, underscoring the effectiveness of jointly extracting spatial and spectral features to enhance spectral consistency. Moreover, a comparison of 2D and 3D network-based methods reveals that the 3D approach fails to fully capitalize on its potential to capture spectral features, particularly for datasets with a large number of bands. This limitation can be attributed to the model's capacity constraints.

To visually assess the performance of different methods, we conducted a visual evaluation, the results of which are presented in Fig. 4. The 3DFCNN method yields results akin to bicubic interpolation, retaining minimal details. In contrast, our method preserves finer details more effectively, particularly in the annotated region where a distinct curve intersects with a smaller curve. Our approach accurately reconstructs this intricate detail, which is missed by other methods, along with other subtle features. As a result, our method produces a more natural and detailed reconstruction.

4.2.2. Experimental results on CAVE dataset

To further validate the robustness and effectiveness of the proposed method, we conducted comparative experiments on natural scene hyperspectral images using the CAVE dataset. This dataset comprises 32 scene images, each with a resolution of 512×512 pixels and 31 spectral bands. We randomly selected 20 images for the training set. Similar to the Chikusei dataset, overlapping patches were extracted from the original images, which were treated as high-resolution references. These patches were then downsampled using bicubic interpolation to generate low-resolution images. In the experiment, we tested scaling factors of 4× and 8×, where the input low-resolution images had resolutions of 32×32 and 16×16 pixels, respectively, with output resolutions of 128×128 pixels for both scaling factors.

The results from the 4× and 8× experiments on the CAVE dataset, as presented in Table 2, demonstrate that our method outperforms others across spatial metrics. In terms of SSIM, the performance of MCNet is similar to ours. MCNet, which combines 2D and 3D convolutions, effectively captures local spatial-spectral features; however, it is limited in its ability to model global dependencies. In contrast, our method seamlessly integrates spatial and spectral features through a low-rank reconstruction strategy, enabling it to capture the global relationships between these features, thereby yielding superior spectral consistency. Furthermore, a comparison between 2D- and 3D-based methods reveals that 3D networks show an advantage on datasets with fewer spectral bands, whereas 2D networks tend to exhibit limitations in this regard. This is primarily due to the inability of traditional 2D methods to jointly capture spatial and spectral features at the same level of integration as 3D methods. Our approach, despite being based on 2D convolutions, uniquely integrates spatial and spectral information, setting it apart and highlighting its distinct advantages in achieving both spatial accuracy and spectral consistency.

Figs. 5 and 6 illustrate the visual outcomes of 4× and 8× SR on two test samples from the CAVE dataset, comparing results across various methods. In Fig. 5, the reconstructed letters within the highlighted region appear blurred and lack fine details when using other methods, whereas our approach accurately restores these details, underscoring its effectiveness in reconstructing intricate structures. Despite the inherent challenge of 8× SR, Fig. 6 demonstrates that our method successfully preserves a significant level of fine detail, further showcasing its robustness and superiority in high-magnification reconstruction tasks.

Moreover, a comparison of the visual results in Fig. 6 with the quantitative metrics reported in Table 2 reveals an important insight: while MCNet achieves competitive numerical scores, its visual outcomes remain suboptimal, lacking fidelity and detail. This inconsistency highlights the instability of MCNet in maintaining reconstruction quality across varying test scenarios, in contrast to the stability and reliability of our proposed method Fig. 7.

Figs. 6 and 8 present the error maps and spectral curves for different methods applied to the same test image, respectively. The error maps visualize the discrepancies between the reconstructed image and the ground truth, where darker blue regions indicate superior spatial reconstruction accuracy. As illustrated in Fig. 6, our method demonstrates significant advantages in both global reconstruction quality and fine detail accuracy. Notably, in the reconstruction of letters, the error maps generated by our method exhibit minimal contour discrepancies, while competing methods display pronounced contour errors, indicating greater deviations from the ground truth. These results affirm the superior spatial fidelity achieved by our approach.

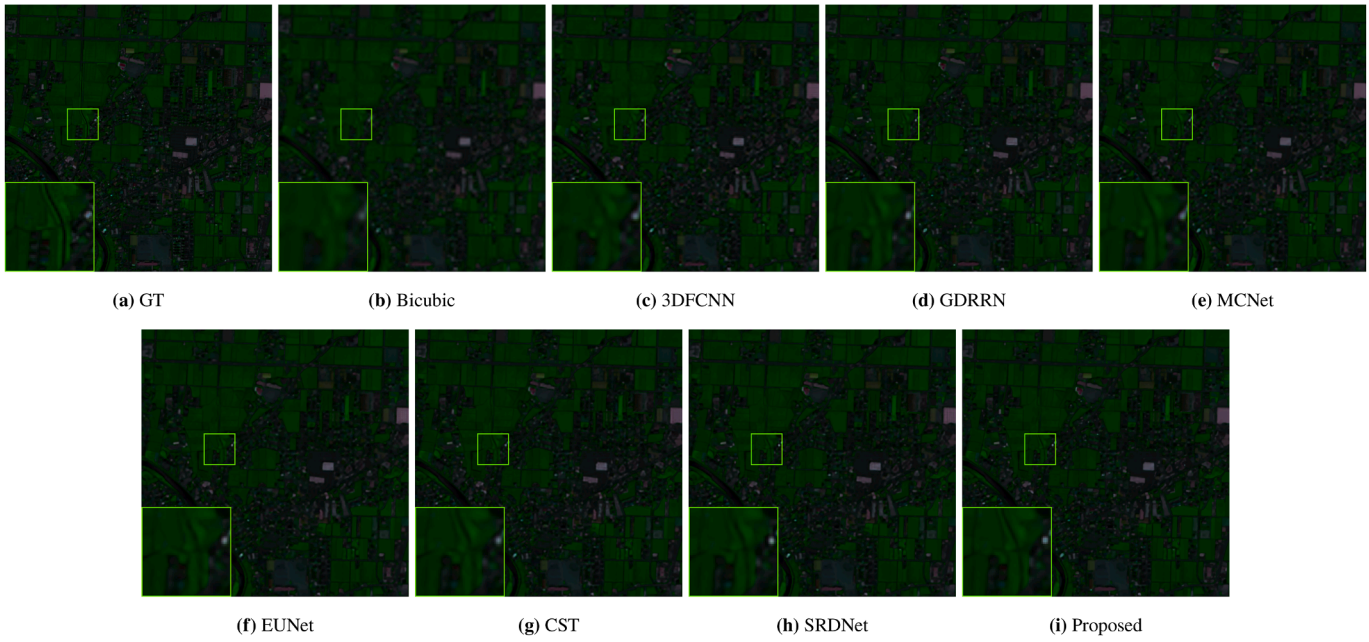


Fig. 4. Visualization of a test image from the Chikusei dataset when the upsampling factor is 4, where the spectral band combination of 31-98-61 is displayed as a false-color image.

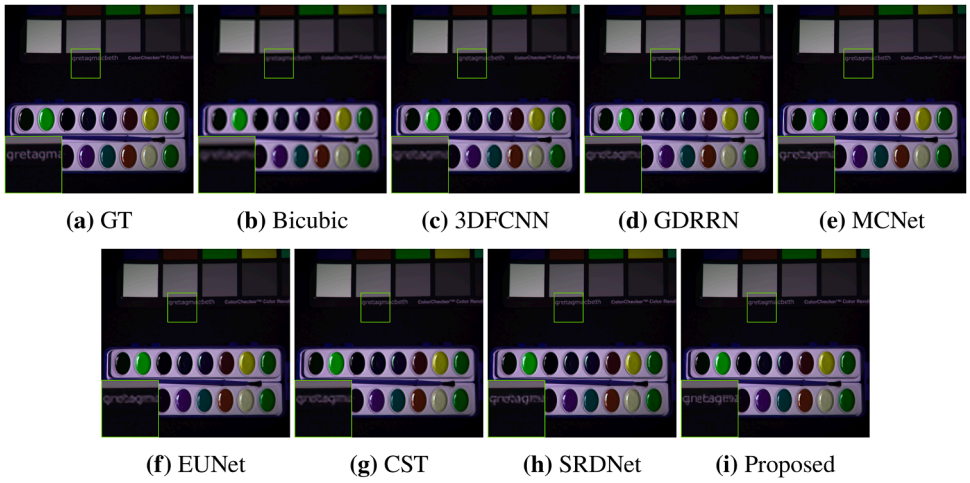


Fig. 5. Visualization of a test image from the CAVE dataset when the upsampling factor is 4, where the spectral band combination of 16-26-6 is displayed as a false-color image.

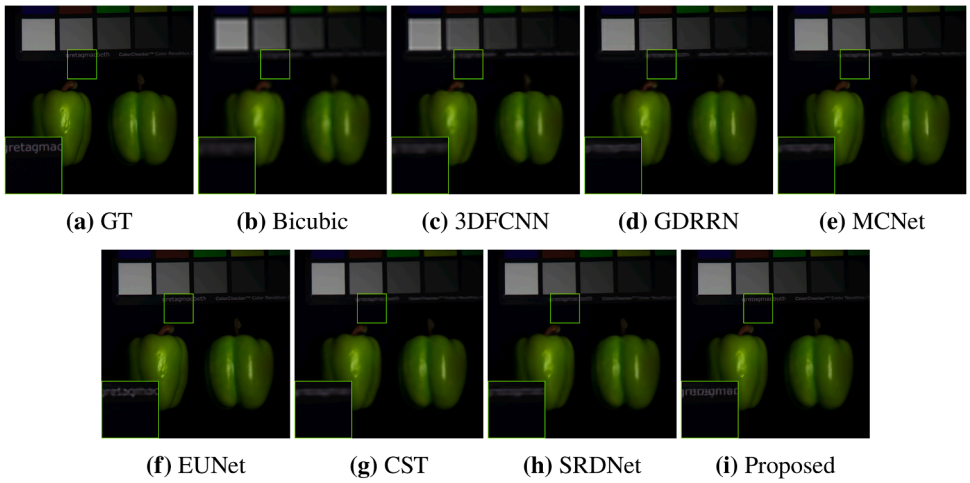


Fig. 6. Visualization of a test sample from the CAVE dataset with an upsampling factor of 8, where the spectral bands 16-26-6 are displayed as a false-color image.

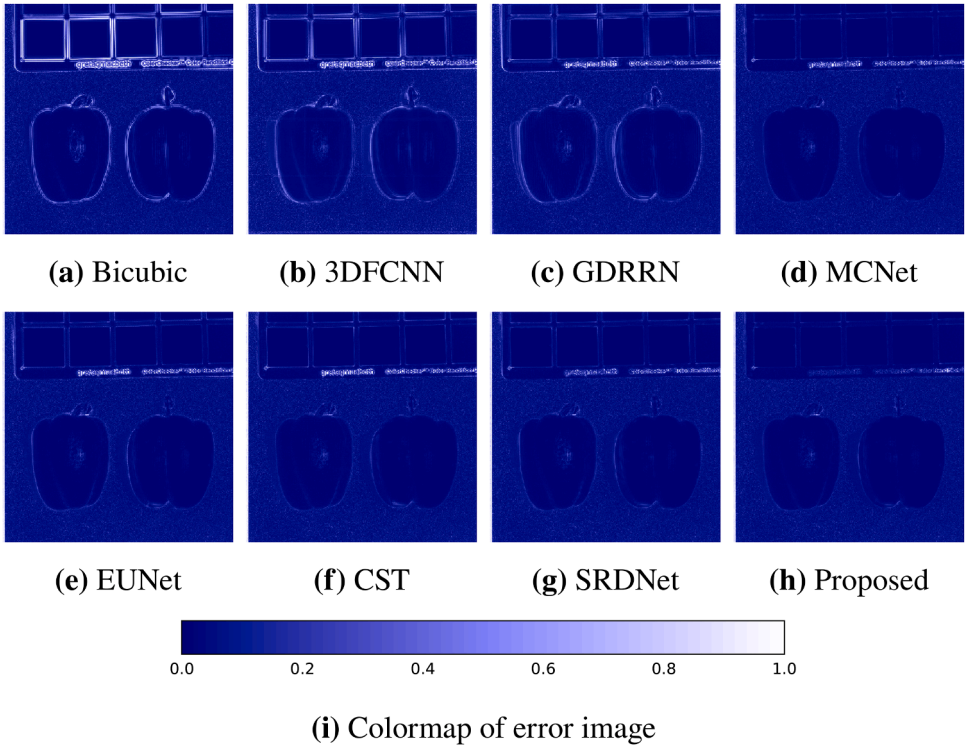
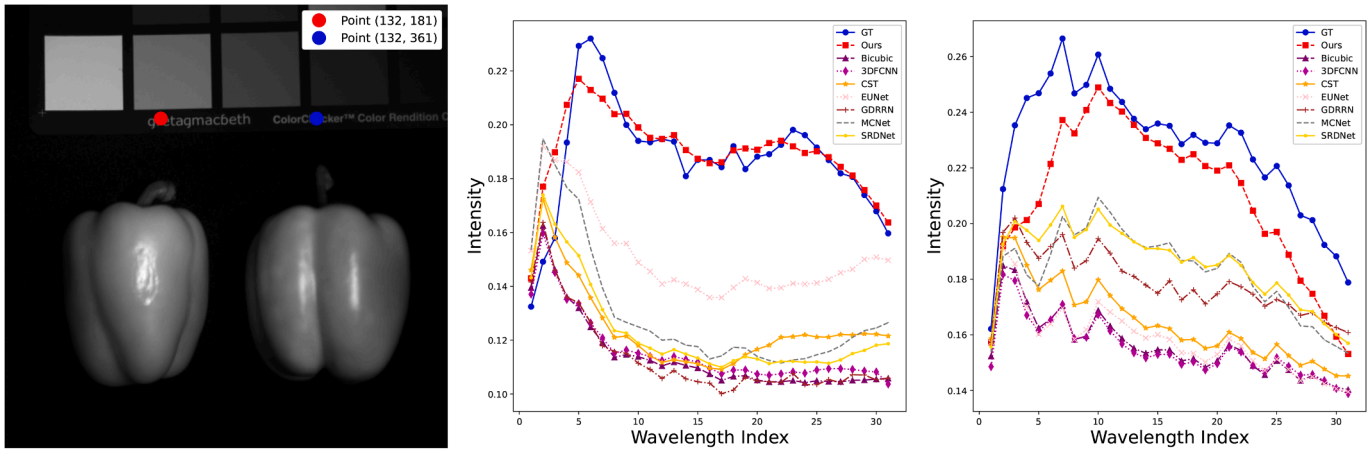


Fig. 7. Visualization of error maps for various methods on a test sample from the CAVE dataset with an upsampling factor of 4.

Fig. 8 highlights the methods' performance in preserving spectral information. The leftmost panel shows the ground truth, with the 21st spectral band rendered in grayscale. Two specific pixel locations, (132,181) and (132,371), are marked, corresponding to the spectral curves on the right. By comparing the error maps in Fig. 6 and the spectral curves in Fig. 8, it is evident that these pixel locations reside in areas with higher spatial errors, often associated with regions containing intricate textures. For these pixels, the spectral curves reconstructed by other methods deviate significantly from the ground truth, whereas our method accurately reproduces the original spectral profiles.

The exceptional performance of our method in both spatial reconstruction and spectral preservation can be attributed to its effective integration of spatial and spectral features. The interplay between these two domains is critical, as accurate spatial reconstruction directly influences spectral consistency, while the retention of spectral integrity enhances spatial representation. By seamlessly capturing and fusing spatial and spectral features, our method achieves superior outcomes, as evidenced by the spectral curves and error maps.

The Chikusei and CAVE test datasets differ significantly in spatial resolution, spectral resolution, and the number of bands, providing a robust benchmark for evaluating method performance. As summarized in Tables 1 and 2 , our method consistently achieves superior results across both datasets, demonstrating its robustness and adaptability. In contrast, other methods exhibit variable performance. For instance, CST performs competitively on the Chikusei dataset, ranking second to our method, but shows markedly lower performance on the CAVE dataset. This inconsistency likely stems from CST's independent extraction of spatial and spectral features, which, despite capturing long-range dependencies in both domains, fails to integrate them effectively.



(a) Original Image (Band 21) **(b)** Spectral Comparison at Pixel (132, 181) **(c)** Spectral Comparison at Pixel (132, 361)

Fig. 8. Comparison of spectral curves at two pixel points from a test sample in the CAVE dataset for various methods with an upsampling factor of 4.

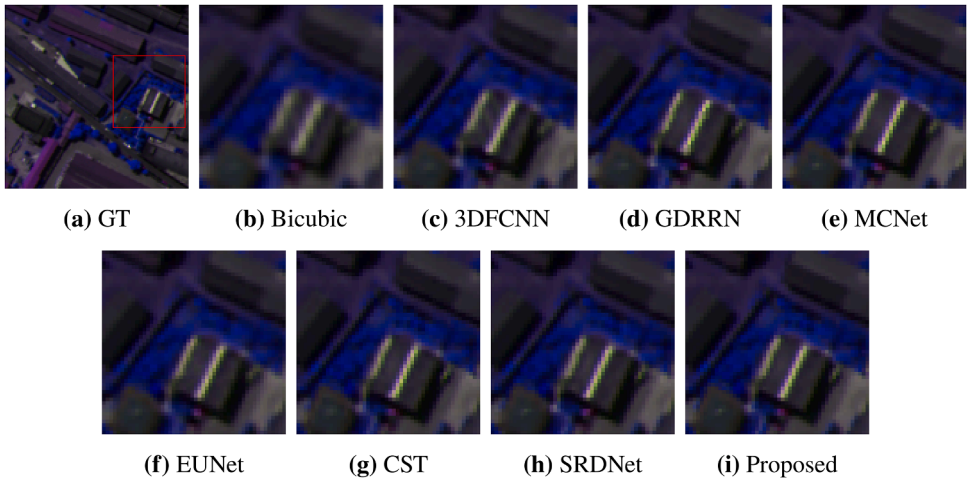


Fig. 9. Visualization of a test image from the Pavia dataset with an upsampling factor of 2, where the spectral band combination of 51-31-91 is displayed as a false-color imag.

Table 3
Quantitative comparison of different methods on the pavia dataset.

Methods	Scale = 2					
	PSNR	SSIM	SAM	CC	ERGAS	RMSE
Bicubic	33.2946	0.9155	3.9875	0.9549	3.9495	0.0225
3DFCNN	35.3005	0.9470	3.7107	0.9699	3.1703	0.0180
GDRRN	36.6871	0.9584	3.3757	0.9766	2.7617	0.0152
MCNet	36.7451	0.9589	3.3374	0.9770	2.7222	0.0153
EUNet	36.0794	0.9525	3.5749	0.9737	2.9302	0.0164
CST	37.8721	0.9667	3.0374	0.9812	2.4552	0.0132
SRDNet	37.4684	0.9637	3.1217	0.9796	2.5532	0.0140
Ours	**38.6412**	**0.9704**	**2.8700**	**0.9834**	**2.2766**	**0.0122**
Methods	Scale = 4					
	PSNR	SSIM	SAM	CC	ERGAS	RMSE
Bicubic	28.5279	0.7341	5.6917	0.8644	6.7753	0.0396
3DFCNN	29.3110	0.7808	5.4846	0.8849	6.1859	0.0362
GDRRN	29.9443	0.8112	5.2079	0.9014	5.7671	0.0336
MCNet	29.7459	0.8001	5.4308	0.8959	5.8807	0.0347
EUNet	29.6511	0.7984	5.2980	0.8942	5.9497	0.0348
CST	30.2462	0.8242	5.0222	0.9080	5.5683	0.0324
SRDNet	29.9707	0.8129	5.0356	0.9016	5.7390	0.0336
Ours	**30.4777**	**0.8372**	**4.7163**	**0.9125**	**5.4201**	**0.0316**

MCNet, while excelling on the CAVE dataset with its hybrid 2D-3D convolutional architecture, performs less effectively on the Chikusei dataset. Its focus on local spatial-spectral feature extraction suits datasets with fewer bands, such as CAVE, but the absence of global dependency modeling hinders its performance on datasets with higher spectral complexity, such as Chikusei. These findings underscore the stability and generalizability of our method, which effectively integrates spatial and spectral features to adapt to diverse data characteristics.

4.2.3. Experimental results on pavia dataset

Due to the absence of information in the central region of the Pavia dataset, we cropped this region following the methodology outlined in SSPSR (Jiang et al., 2020), resulting in a sub-image of size $1096 \times 715 \times 102$. This sub-image was subsequently partitioned into training and test sets. Specifically, the image was divided into a top and bottom region. The bottom region ($128 \times 715 \times 102$) was designated for testing, with center cropping (Liu et al., 2023) applied to both the left and right sides, generating four non-overlapping images, each of size $128 \times 128 \times 102$. For the remaining portion of the sub-image, we followed the procedure used for training data extraction from the Chikusei dataset. Overlapping patches were extracted from the original image, treated as high-resolution references, and downsampled using bicubic interpolation to generate corresponding low-resolution images. In this experiment, we tested scaling factors of 2× and 4×, with the input low-resolution images having resolutions of 32×32 and 16×16 pixels, respectively, and the output resolutions set to 64×64 pixels for both cases.

As shown in Table 3, the proposed method significantly outperforms all other approaches across all evaluation metrics. Compared to the second-best CST method, our method demonstrates a clear advantage at both 2× and 4× magnification. The dataset utilized in this study is smaller than the Chikusei and CAVE datasets, and due to the lower resolution of the test images, we present visual comparisons of the reconstructed images for different methods using a single test image (Fig. 9) and corresponding error maps to more intuitively highlight performance differences. As depicted in Fig. 10, our method is capable of reconstructing more fine details than the other methods. Furthermore, Fig. 10 reveals that our method produces smaller errors, particularly in the bright regions on the right of the error map, indicating superior spatial reconstruction.

To assess spectral preservation, we captured the spectral curves of two pixels from the same image (Fig. 11). The results show that the spectral curve for the proposed method (red) aligns more closely with the ground truth (blue), further validating the efficacy of our approach in jointly extracting spatial and spectral features. This demonstrates the method's superiority, particularly on small-scale datasets. Moreover, when comparing the performance of different methods on the Chikusei and Pavia datasets, it is evident that while 3D convolution-based methods can extract both spatial and spectral features, their performance is constrained by the receptive field and model capacity limitations (Liu et al., 2023). These constraints hinder their ability to achieve optimal results on datasets with hundreds of spectral bands.

5. Discussion

5.1. Ablation study

The LREM module serves as the key component of our framework, constructed through a ROM group generated by CP decomposition principles. The ROM group integrates four core components: CFRP, ELI, DWC, and Bottleneck. To systematically evaluate their impacts, we conducted ablation studies on the constituent modules of the ROM group. The quantitative results are summarized in Table 4. In addition, we also evaluated the effect of LREM insertion after up-sampling, and the quantitative results were summarized in Table 5.

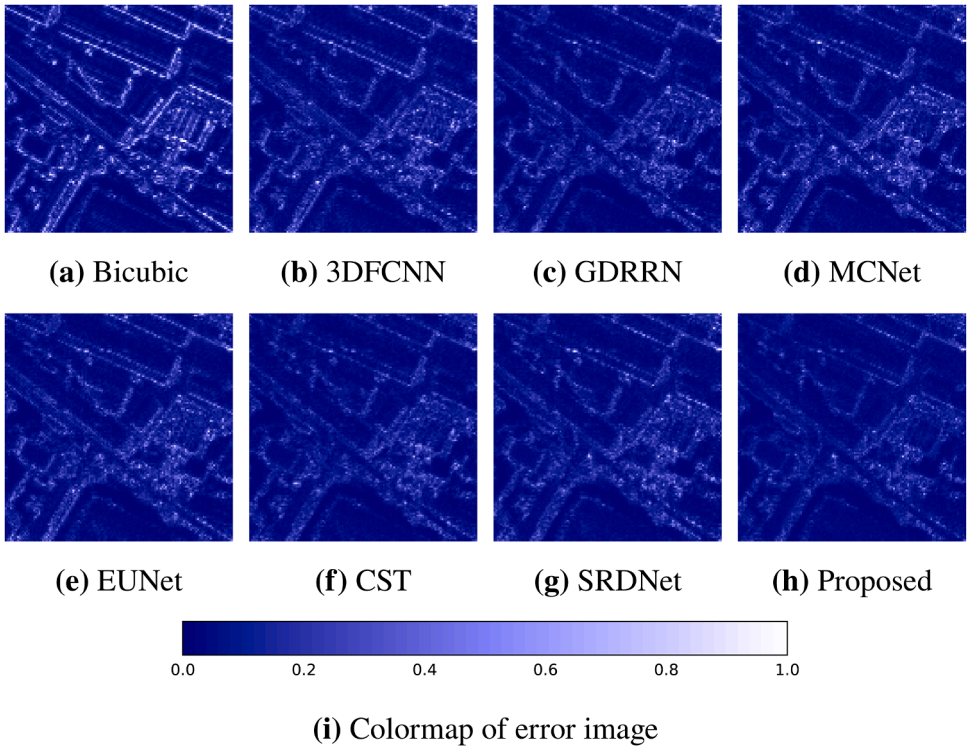


Fig. 10. Error map of a test image from the Pavia dataset with an upsampling factor of 2.

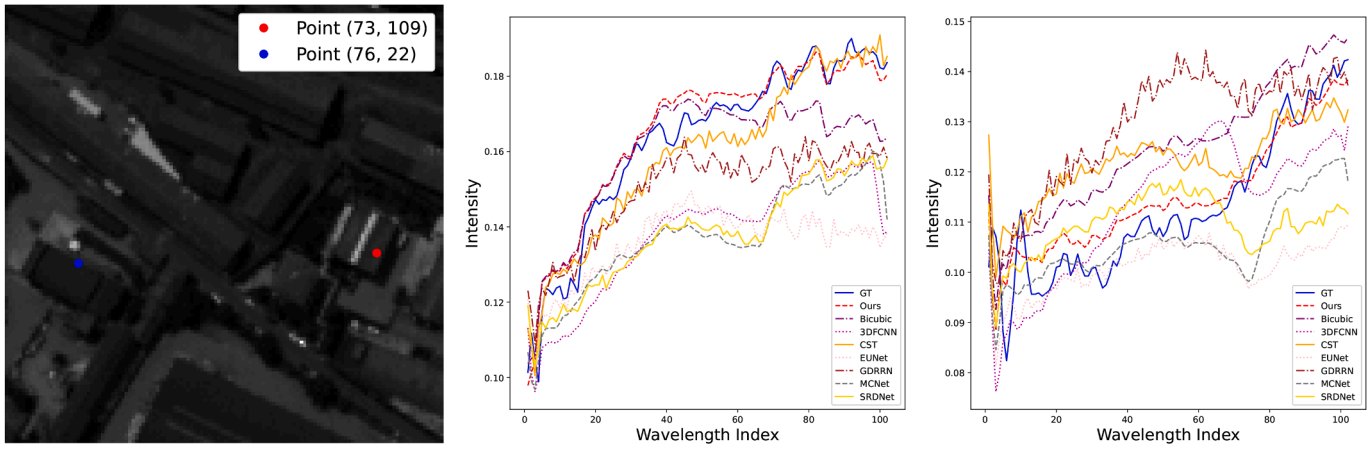
1) **Effectiveness of the CFRP strategy:** In the ROM, we retain channel information and further adaptively fuse it through the ELI module. In contrast, other computer vision tasks often employ global pooling to merge channel information for extracting row and column features. However, this method is not optimal for hyperspectral images, which contain rich channel-specific information. To demonstrate the effectiveness of our approach, we replaced the CFRP strategy with global pooling, which compresses and retains information only in the row and column directions. The experimental performance of the corresponding model is shown in rows 3 and 4 of Table 4. The results clearly show that the CFRP strategy significantly improves the spatial quality and spectral quality of the image. This improvement is attributed to the limitation of global pooling in capturing non-local dependencies, which leads to the loss of critical channel information and adversely affects the extraction of subsequent features.
2) **Effectiveness of the ELI module:** In the ROM, the ELI module adaptively fuses channel information through 1D large kernel convolutions, thereby enhancing information localization and establishing non-local dependencies. This fusion improves the accuracy in capturing overall dependencies. To assess the efficacy of this module, we replaced it with a 2D convolution using a 1×1 kernel. The experimental performance of the corresponding model is shown in rows 4 and 6 of Table 4. The results demonstrate that the ELI module significantly enhances the spatial information expression capability, underlining its contribution to improved performance.
3) **Effectiveness of the DWC module:** In the LREM module, the DWC module is employed to preserve feature diversity. To assess its effectiveness, we removed the DWC module. The experimental performance of the corresponding model is shown in rows 1 and 2 of Table 4. The results indicate that the DWC module preserves more detailed information and mitigates the loss of certain feature components associated with low-rank tensor representations, thereby enhancing the model's expressive capability.
4) **Effectiveness of the bottleneck module:** In the ROM group, the Bottleneck module is employed to statistically model channel-wise dependencies and exploit low-rank structural information (Yang et al., 2024). To assess the efficacy of this module, we replaced it with a 2D convolutional layer utilizing a 1×1 kernel. The experimental performance of the corresponding model configurations is illustrated in rows 5 and 6 of Table 4. The results demonstrate that the Bottleneck module improves both spatial and spectral metrics and enhances the model's representational capacity.
5) **Effectiveness of inserting the LREM after upsampling:** Progressive upsampling strategies have demonstrated efficacy in various studies (Jiang et al., 2020; Liu et al., 2024; Xu et al., 2024). To assess the impact of adding the LREM after upsampling, we conducted an experiment in which no operation was performed following upsampling. This model is referred to as "OPU" (only progressive upsampling), with $r = 5$ and 3 LREMs per stage. We evaluated the 4× super-resolution performance using the SAM metric on the CAVE dataset (see Table 5). The results show a significant decline in the

Table 4
Quantitative metrics of LTRN with/without CFRP, ELI, DWC, AND bottleneck.

CFRP	ELI	DWC	Bottleneck	PSNR	SAM
×	×	×	×	40.4879	2.3070
×	×	✓	×	40.5398	2.2523
×	×	✓	✓	40.5409	2.2501
✓	×	✓	✓	40.5777	2.2397
✓	✓	✓	×	40.5721	2.2279
✓	✓	✓	✓	**40.6053**	**2.2242**

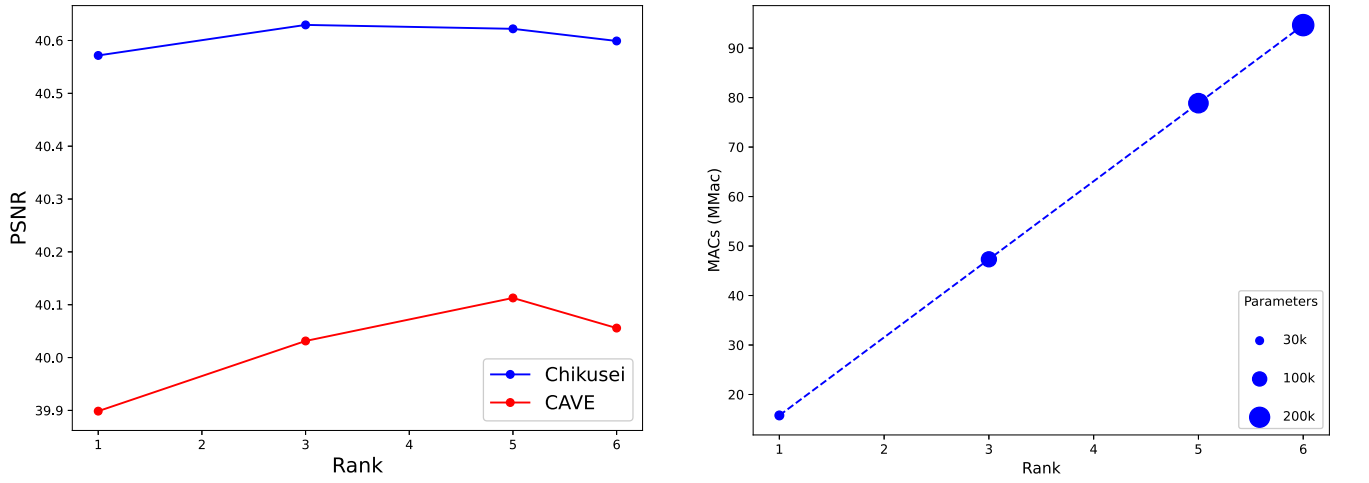
Table 5
Ablation quantitative metrics of inserting the LREM after upsampling.

Model	PSNR	SAM
Proposed	40.1128	3.1788
OPU	39.9324	3.2402



(a) Original Image (Band 21) **(b)** Spectral Comparison at Pixel (73,109) **(c)** Spectral Comparison at Pixel (76,22)

Fig. 11. Spectral curves of two pixels from a test image in the Pavia dataset with an upsampling factor of 2.



(a) The effect of Rank size on reconstruction results. **(b)** The effect of Rank size on the complexity and parameters of CP decomposition.

Fig. 12. Effects of rank size: (a) Effect on reconstruction results; (b) Effect on complexity and parameters of CP decomposition.

SAM metric upon removal of the LREM after upsampling. The upsampling layer primarily performs feature expansion in the spatial domain by increasing the spatial resolution of the feature map, facilitating the recovery of spatial details. However, it does not substantially enhance the representation of spectral information. By neglecting the refinement and correction of spectral information after upsampling, the crucial relationship between spatial and spectral features is weakened. This underscores the necessity of joint spatial and spectral feature extraction, validating the effectiveness of inserting the LREM post-upsampling.

5.2. Hyperparameter analysis

The main hyperparameters of the proposed LRTENet include the number of LREMs (N) and the rank (r) of the ROM in each LREM. Given the large number of possible combinations of these parameters, we adopted a systematic approach by fixing one hyperparameter and examining the effects of varying the other to ensure scientific rigor. Initially, we investigated the impact of the low-rank tensor rank (r) on the performance of the model, with N fixed at 3, implying that each stage comprises 3 LREMs. We conducted this analysis on both the CAVE and Chikusei datasets, exploring a range of values for the rank parameter and assessing its influence on reconstruction quality through a series of controlled experiments. Furthermore, we have added two aspects, namely the multiplication-accumulation operation (MAC) and the number of parameters, to measure the influence of rank size in CP decomposition. The results, presented in Fig. 12, indicate that optimal performance is achieved when the rank is selected within the range of 3 to 5, and the complexity and the number of parameters are also appropriate at this time. It can be seen that the size of the r has a significant impact on the parameters and performance. In future work, we may consider adapting the size of the r, and the size of the r in each LREM needs to be further explored.

Subsequently, we investigated the impact of the number of LREMs (N) on model performance, using both the CAVE and Chikusei datasets, while keeping the rank (r) fixed at 5. Specifically, we tested the floating-point operations (FLOPs), Parameters, PSNR, and SSIM of different models on 128×128×C-size hyperspectral images, as shown in Table 6. The results indicate that increasing N enhances the model's representational capability and improves performance, but it also significantly increases model complexity. For the Chikusei and CAVE datasets, we observed an improvement in reconstruction metrics when N was set to 3 compared to $N = 2$, with the most notable enhancement occurring on the CAVE dataset. However, for the Chikusei dataset, the improvement was

Table 6
The effect of the number of LREMS on reconstruction results.

N	Chikusei		CAVE		Complexity	
	PSNR	SSIM	PSNR	SSIM	FLOPs	Parameters
1	40.4887	0.9457	39.5621	0.9625	3.1T	17.7M
2	40.6054	0.9473	39.9911	0.9642	5.8T	29.5M
3	**40.6221**	**0.9474**	**40.1128**	**0.9645**	8.6T	41.9M

Table 7
Comparison of computational complexity.

Methods	GPU memory (GB)	FLOPs (G)	PSNR (dB)	SAM
MCNet	1.5	459.02	29.75	5.4308
SRDNet	0.9	160.23	29.97	5.0356
GDRNN	0.4	11.84	29.94	5.2079
3DFCNN	1.1	8.23	29.31	5.4846
CST	0.5	39.74	30.25	5.0222
EUNet	0.6	22.59	29.65	5.2980
LRTENet (Ours)	0.9	541.33	**30.48**	**4.7163**

marginal, suggesting that further experiments with higher values of N were unnecessary. Based on these findings, we recommend selecting $N = 3$ as the optimal configuration.

5.3. Complexity analysis

To further assess the computational complexity, we report both FLOPs and GPU memory consumption, and compare our method with state-of-the-art approaches. The results are summarized in Table 7. As shown, the memory usage of LRTENet is 0.9 GB, which is lower than that of MCNet (1.5 GB) and 3DFCNN (1.1 GB), while slightly higher than GDRNN (0.4 GB) and CST (0.5 GB). This demonstrates that the proposed method delivers superior performance without imposing excessive memory overhead. Although our approach involves higher FLOPs than most competing methods, it should be emphasized that fine-grained spectral modeling in hyperspectral super-resolution inherently entails a certain computational cost. Through an efficient feature extraction design, our method allocates computation strategically to the critical stage of spatial–spectral joint optimization, thereby striking an effective balance between computational investment and performance gain. For example, compared with CST, which requires only 39.74 GFLOPs, our method introduces a reasonable increase in computation while achieving a performance improvement to a SAM of 0.3059.

6. Conclusion

This paper proposed a novel deep low-rank tensor embedded network (LRTENet) for hyperspectral image super-resolution. By leveraging low-rank tensor decomposition, the proposed LRTENet efficiently approximates the complex mapping relationships inherent in hyperspectral data, facilitating the seamless fusion of spatial and spectral features. Another key innovation lies in the inclusion of the low-rank tensor embedding module, which addresses the spectral distortion commonly associated with upsampling processes. Comprehensive evaluations on multiple benchmark datasets demonstrate that the proposed LRTENet consistently outperforms state-of-the-art approaches, achieving superior performance in both spatial fidelity and spectral accuracy.

CRediT authorship contribution statement

Qiang Zhang: Conceptualization, Methodology, Writing – original draft; **Xianpeng Zhang:** Data curation, Validation; **Yi Xiao:** Supervision, Resources, Investigation, Writing – review & editing; **Hongjie Xie:** Visualization, Validation.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests. Qiang Zhang reports financial support was provided by Open Fund of State Key Laboratory of Remote Sensing Science. Yi Xiao reports financial support was provided by National Natural Science Foundation of China. Qiang Zhang reports financial support was provided by Fundamental Research Funds for the Central Universities. Qiang Zhang reports financial support was provided by Dalian Science and Technology Talent Innovation Supporting Project. Qiang Zhang reports financial support was provided by China Postdoctoral Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study is supported in part by the National Natural Science Foundation of China under Grant 62401095 and 423B2104; And in part by the Natural Science Foundation of Inner Mongolia Autonomous Region under Grant 2023LHMS04007; And in part by the Dalian Science and Technology Talent Innovation Supporting Project under Grant 2024RQ028; And in part by the China Postdoctoral Science Foundation under Grant 2023M740460; And in part by the Natural Science Foundation of Liaoning Province under Grant 2025-BS-0236; And in part by the Fundamental Research Funds for the Central Universities under Grant 3132025267.

References

Akhtar, N., Shafait, F., & Mian, A. (2015). Bayesian sparse representation for hyperspectral image super resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3631–3640).

Bu, Y., Zhao, Y., Xue, J., Yao, J., & Chan, J. C.-W. (2024). Transferable multiple subspace learning for hyperspectral image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, *21*, 1–5. https://doi.org/10.1109/LGRS.2023.3339505

Chen, C., Wang, Y., Zhang, N., Zhang, Y., & Zhao, Z. (2023a). A review of hyperspectral image super-resolution based on deep learning. *Remote Sensing*, *15*(11), 2853.

Chen, S., Zhang, L., & Zhang, L. (2023b). Msdformer: Multiscale deformable transformer for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–14. https://doi.org/10.1109/TGRS.2023.3315970

Chen, S., Zhang, L., & Zhang, L. (2024a). Cross-scope spatial-spectral information aggregation for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, *33*, 5878–5891. https://doi.org/10.1109/TIP.2024.3468905

Chen, W., Zhu, X., Sun, R., He, J., Li, R., Shen, X., & Yu, B. (2020). Tensor low-rank reconstruction for semantic segmentation. In *Proceedings of the European conference on computer vision* (pp. 52–69).

Chen, Y., Wei, M., & Chen, Y. (2025). A method based on hybrid cross-multiscale spectral-spatial transformer network for hyperspectral and multispectral image fusion. *Expert systems with applications*, *263*, 125742. https://www.sciencedirect.com/science/article/pii/S0957417424026095. https://doi.org/10.1016/j.eswa.2024.125742

Chen, Y., Zhang, H., Wang, Y., Yang, Y., & Wu, J. (2024b). Flex-DLD: Deep low-rank decomposition model with flexible priors for hyperspectral image denoising and restoration. *IEEE Transactions on Image Processing*, *33*, 1211–1226. https://doi.org/10.1109/TIP.2024.3360902

Dian, R., Li, S., Fang, L., & Bioucas-Dias, J. (2018). Hyperspectral image super-resolution via local low-rank and sparse representations. In *Proceedings of the IEEE international geoscience and remote sensing symposium* (pp. 4003–4006). https://doi.org/10.1109/IGARSS.2018.8519213

Dian, R., Li, S., Fang, L., & Wei, Q. (2019). Multispectral and hyperspectral image fusion with spatial-spectral sparse representation. *Information Fusion*, *49*, 262–270. https://doi.org/10.1016/j.inffus.2018.11.012

Dian, R., Liu, Y., & Li, S. (2024). Spectral super-resolution via deep low-rank tensor representation. *IEEE Transactions on Neural Networks and Learning Systems*, (pp. 1–11). https://doi.org/10.1109/TNNLS.2024.3359852

Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307.

Dong, W., Fu, F., Shi, G., Cao, X., Wu, J., Li, G., & Li, X. (2016). Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Transactions on Image Processing*, *25*(5), 2337–2352. https://doi.org/10.1109/TIP.2016.2542360

Fang, M., Tan, Z., Tang, Y., Chen, W., Huang, H., Dananjayan, S., He, Y., & Luo, S. (2024). Pest-conformer: A hybrid CNN-transformer architecture for large-scale multi-class crop pest recognition. *Expert Systems with Applications*, *255*, 124833. https://doi.org/10.1016/j.eswa.2024.124833

Gendy, G., He, G., & Sabor, N. (2023). Lightweight image super-resolution based on deep learning: State-of-the-art and future directions. *Information Fusion*, *94*, 284–310. https://doi.org/10.1016/j.inffus.2023.01.024

Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., & Plaza, A. (2017). Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 37–78. https://doi.org/10.1109/MGRS.2017.2762087

Hajaj, S., El Harti, A., Pour, A. B., Jellouli, A., Adiri, Z., & Hashim, M. (2024). A review on hyperspectral imagery application for lithological mapping and mineral prospecting: Machine learning techniques and future prospects. *Remote Sensing Applications: Society and Environment*, *35*, 101218. https://doi.org/10.1016/j.rsase.2024.101218

Han, D., Ye, T., Han, Y., Xia, Z., Pan, S., Wan, P., Song, S., & Huang, G. (2024). Agent attention: On the integration of softmax and linear attention. In *Proceedings of the European conference on computer vision* (pp. 124–140).

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7132–7141). https://doi.org/10.1109/CVPR.2018.00745

Hu, Q., Wang, X., Jiang, J., Zhang, X.-P., & Ma, J. (2024). Exploring the spectral prior for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, *33*, 5260–5272. https://doi.org/10.1109/TIP.2024.3460470

Jiang, J., Sun, H., Liu, X., & Ma, J. (2020). Learning spatial-spectral prior for super-resolution of hyperspectral imagery. *IEEE Transactions on Computational Imaging*, *6*, 1082–1096. https://doi.org/10.1109/TCI.2020.2996075

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*(3), 455–500.

Lepcha, D. C., Goyal, B., Dogra, A., & Goyal, V. (2023). Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Information Fusion*, *91*, 230–260. https://doi.org/10.1016/j.inffus.2022.10.007

Li, Q., Wang, Q., & Li, X. (2020). Mixed 2d/3d convolutional network for hyperspectral image super-resolution. *Remote Sensing*, *12*(10), 1660. https://doi.org/10.3390/rs12101660

Li, Q., Wang, Q., & Li, X. (2021). Exploring the relationship between 2d/3d convolution for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(10), 8693–8703. https://doi.org/10.1109/TGRS.2020.3047363

Li, Y., Zhang, L., Dingl, C., Wei, W., & Zhang, Y. (2018). Single hyperspectral image super-resolution with grouped deep recursive residual network. In *Proceedings of the IEEE fourth international conference on multimedia big data* (pp. 1–4). https://doi.org/10.1109/BigMM.2018.8499097

Liang, J., Zhou, J., Tong, L., Bai, X., & Wang, B. (2018). Material based salient object detection from hyperspectral images. *Pattern Recognition*, *76*, 476–490. https://doi.org/10.1016/j.patcog.2017.11.024

Liu, D., Li, J., Yuan, Q., Zheng, L., He, J., Zhao, S., & Xiao, Y. (2023). An efficient unfolding network with disentangled spatial-spectral representation for hyperspectral image super-resolution. *Information Fusion*, *94*, 92–111. https://doi.org/10.1016/j.inffus.2023.01.018

Liu, T., Liu, Y., Zhang, C., Yuan, L., Sui, X., & Chen, Q. (2024). Hyperspectral image super-resolution via dual-domain network based on hybrid convolution. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–18. https://doi.org/10.1109/TGRS.2024.3370107

Loncan, L., de Almeida, L. B., Bioucas-Dias, J. M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G. A., Simões, M., Tourneret, J.-Y., Veganzones, M. A., Vivone, G., Wei, Q., & Yokoya, N. (2015). Hyperspectral pansharpening: A review. *IEEE Geoscience and Remote Sensing Magazine*, *3*(3), 27–46. https://doi.org/10.1109/MGRS.2015.2440094

Ma, Q., Jiang, J., Liu, X., & Ma, J. (2023). Learning a 3d-CNN and transformer prior for hyperspectral image super-resolution. *Information Fusion*, *100*, 101907. https://www.sciencedirect.com/science/article/pii/S1566253523002233. https://doi.org/10.1016/j.inffus.2023.101907

Mei, S., Yuan, X., Ji, J., Zhang, Y., Wan, S., & Du, Q. (2017). Hyperspectral image spatial super-resolution via 3d full convolutional neural network. *Remote Sensing*, *9*(11), 1139.

Muhammad, U., Laaksonen, J., & Mihaylova, L. (2025). Towards lightweight hyperspectral image super-resolution with depthwise separable dilated convolutional network. arXiv:2505.00374.

Nandi, U., Roy, S. K., Hong, D., Wu, X., & Chanussot, J. (2023). Tattmsrecnet:triplet-attention and multiscale reconstruction network for band selection in hyperspectral images. *Expert Systems with Applications*, *212*, 118797. https://doi.org/10.1016/j.eswa.2022.118797

Neupane, B., Aryal, J., & Rajabifard, A. (2024). Cnns for remote extraction of urban features: A survey-driven benchmarking. *Expert Systems with Applications*, *255*, 124751. https://doi.org/10.1016/j.eswa.2024.124751

Sahadevan, A. S. (2021). Extraction of spatial-spectral homogeneous patches and fractional abundances for field-scale agriculture monitoring using airborne hyperspectral images. *Computers and Electronics in Agriculture*, *188*, 106325. https://doi.org/10.1016/j.compag.2021.106325

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., & Woo, W.-c. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems* (pp. 802–810).

Shi, Y., Zhang, Q., Kang, S., Yin, C., Liu, X., & He, X. (2024). Fgrc-net: A high-information interactive convolutional neural network for identifying ink spectral information. *Expert Systems with Applications*, *235*, 121167. https://doi.org/10.1016/j.eswa.2023.121167

Su, X., Shen, X., Wan, M., Nie, J., Chen, L., Liu, H., & Zhou, X. (2025). EigenSR: Eigenimage-bridged pre-trained RGB learners for single hyperspectral image super-resolution. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 7033–7041). (vol. *39*).

Tan, K., Wang, H., Chen, L., Du, Q., Du, P., & Pan, C. (2020). Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest. *Journal of Hazardous Materials*, *382*, 120987. https://doi.org/10.1016/j.jhazmat.2019.120987

Villa, A., Chanussot, J., Benediktsson, J. A., Jutten, C., & Dambreville, R. (2013). Unsupervised methods for the classification of hyperspectral images with low spatial resolution. *Pattern Recognition*, *46*(6), 1556–1568. https://doi.org/10.1016/j.patcog.2012.10.030

Wang, Q., & Chen, Z. (2024). Parallel wavelet networks incorporating modality adaptation for hyperspectral image super-resolution. *Expert Systems with Applications*, *235*, 121299. https://doi.org/10.1016/j.eswa.2023.121299

Wang, Q., Jin, X., Jiang, Q., Wu, L., Zhang, Y., & Zhou, W. (2023a). Dbct-net:a dual branch hybrid cnn-transformer network for remote sensing image fusion. *Expert Systems with Applications*, *233*, 120829. https://doi.org/10.1016/j.eswa.2023.120829

Wang, X., Hu, Q., Cheng, Y., & Ma, J. (2023b). Hyperspectral image super-resolution meets deep learning: A survey and perspective. *IEEE/CAA Journal of Automatica Sinica*, *10*(8), 1668–1691.

Wang, Y., Chen, X., Han, Z., & He, S. (2017). Hyperspectral image super-resolution via nonlocal low-rank tensor approximation and total variation regularization. *Remote Sensing*, *9*(12), 1286. https://doi.org/10.3390/rs9121286

Wang, Y., Li, Y., Wang, G., & Liu, X. (2024). Multi-scale attention network for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5950–5960).

Xiao, Y., Yuan, Q., Jiang, K., Chen, Y., Wang, S., & Lin, C.-W. (2025a). Multi-axis feature diversity enhancement for remote sensing video super-resolution. *IEEE Transactions on Image Processing*, *34*, 1766–1778. https://doi.org/10.1109/TIP.2025.3547298

Xiao, Y., Yuan, Q., Jiang, K., Chen, Y., Zhang, Q., & Lin, C.-W. (2025b). Frequency-assisted mamba for remote sensing image super-resolution. *IEEE Transactions on Multimedia*, *27*, 1783–1796. https://doi.org/10.1109/TMM.2024.3521798

Xiao, Y., Yuan, Q., Jiang, K., He, J., Lin, C.-W., & Zhang, L. (2024). Ttst: A top-k token selective transformer for remote sensing image super-resolution. *IEEE Transactions on Image Processing*, *33*, 738–752. https://doi.org/10.1109/TIP.2023.3349004

Xie, W., Liu, T., & Gu, Y. (2024). Intrinsic hyperspectral image recovery for UAV strips stitching. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–13. https://doi.org/10.1109/TGRS.2024.3433024

Xu, W., & Wan, Y. (2024). Ela: Efficient local attention for deep convolutional neural networks. arXiv:2403.01123.

Xu, Y., Hou, J., Zhu, X., Wang, C., Shi, H., Wang, J., Li, Y., & Ren, P. (2024). Hyperspectral image super-resolution with convLSTM skip-connections. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–16. https://doi.org/10.1109/TGRS.2024.3401843

Xu, Y., Wu, Z., Chanussot, J., & Wei, Z. (2019). Nonlocal patch tensor sparse representation for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, *28*(6), 3034–3047. https://doi.org/10.1109/TIP.2019.2893530

Xue, J., Zhao, Y., Liao, W., & Chan, J. C.-W. (2019). Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(7), 5174–5189.

Xue, J., Zhao, Y.-Q., Bu, Y., Liao, W., Chan, J. C.-W., & Philips, W. (2021). Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, *30*, 3084–3097. https://doi.org/10.1109/TIP.2021.3058590

Xue, J., Zhao, Y.-Q., Wu, T., & Chan, J. C.-W. (2024). Tensor convolution-like low-rank dictionary for high-dimensional image representation. *IEEE Transactions on Circuits and Systems for Video Technology*, *34*(12), 13257–13270. https://doi.org/10.1109/TCSVT.2024.3442295

Yan, H.-F., Zhao, Y.-Q., Chan, J. C.-W., & Kong, S. G. (2023). Spectral super-resolution based on dictionary optimization learning via spectral library. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–16. https://doi.org/10.1109/TGRS.2022.3229439

Yan, H.-F., Zhao, Y.-Q., Chan, J. C.-W., Kong, S. G., EI-Bendary, N., & Reda, M. (2025). Hyperspectral and multispectral image fusion: When model-driven meet data-driven strategies. *Information Fusion*, *116*, 102803. https://doi.org/10.1016/j.inffus.2024.102803

Yang, Y., Wang, Y., Wang, H., Zhang, L., Zhao, E., Song, M., & Yu, C. (2024). Spectral-enhanced sparse transformer network for hyperspectral super-resolution reconstruction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 17278–17291. https://doi.org/10.1109/JSTARS.2024.3457814

Yasuma, F., Mitsunaga, T., Iso, D., & Nayar, S. K. (2010). Generalized assorted pixel camera: Post-capture control of resolution, dynamic range and spectrum. *IEEE Transactions on Image Processing*, *99*.

Yokoya, N., & Iwasaki, A. (2016). Airborne hyperspectral data over chikusei. Space Appl. Lab., Univ. Tokyo, Japan, Tech. Rep. SAL-2016-05-27, *5*(5), 5.

Zhang, H., Zhang, H., Wang, C., & Xie, J. (2019). Co-occurrent features in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 548–557). https://doi.org/10.1109/CVPR.2019.00064

Zhang, M., Zhang, C., Zhang, Q., Guo, J., Gao, X., & Zhang, J. (2023a). Essaformer: Efficient transformer for hyperspectral image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 23073–23084).

Zhang, Q., Dong, Y., Yuan, Q., Song, M., & Yu, H. (2023b). Combined deep priors with low-rank tensor factorization for hyperspectral image restoration. *IEEE Geoscience and Remote Sensing Letters*, *20*, 1–5.

Zhang, Q., Dong, Y., Zheng, Y., Yu, H., Song, M., Zhang, L., & Yuan, Q. (2024a). Three-dimension spatial–spectral attention transformer for hyperspectral image denoising. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–13. https://doi.org/10.1109/TGRS.2024.3458174

Zhang, Q., Yuan, Q., Li, J., Li, Z., Shen, H., & Zhang, L. (2020). Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *162*, 148–160.

Zhang, Q., Yuan, Q., Li, Z., Sun, F., & Zhang, L. (2021a). Combined deep prior with low-rank tensor SVD for thick cloud removal in multitemporal images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *177*, 161–173. https://doi.org/10.1016/j.isprsjprs.2021.04.021

Zhang, Q., Yuan, Q., Song, M., Yu, H., & Zhang, L. (2022). Cooperated spectral low-rankness prior and deep spatial prior for HSI unsupervised denoising. *IEEE Transactions on Image Processing*, *31*, 6356–6368. https://doi.org/10.1109/TIP.2022.3211471

Zhang, Q., Zheng, Y., Yuan, Q., Song, M., Yu, H., & Xiao, Y. (2024b). Hyperspectral image denoising: From model-driven, data-driven, to model-data-driven. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(10), 13143–13163. https://doi.org/10.1109/TNNLS.2023.3278866

Zhang, Q., Zhu, J., Dong, Y., Zhao, E., Song, M., & Yuan, Q. (2025). 10-Minute forest early wildfire detection: Fusing multi-type and multi-source information via recursive transformer. *Neurocomputing*, *616*, 128963. https://doi.org/10.1016/j.neucom.2024.128963

Zhang, S., Wang, L., Zhang, L., & Huang, H. (2021b). Learning tensor low-rank prior for hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12001–12010). https://doi.org/10.1109/CVPR46437.2021.01183

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6230–6239). https://doi.org/10.1109/CVPR.2017.660

Zhou, C., He, Z., Dong, J., Li, Y., Ren, J., & Plaza, A. (2025). Low-rank and sparse representation meet deep unfolding: A new interpretable network for hyperspectral change detection. *IEEE Transactions on Geoscience and Remote Sensing*, *63*, 1–16. https://doi.org/10.1109/TGRS.2025.3564996

Zhou, C., He, Z., Lou, A., & Plaza, A. (2024). Rgb-to-hsv: A frequency-spectrum unfolding network for spectral super-resolution of rgb videos. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–18. https://doi.org/10.1109/TGRS.2024.3361929

Zhu, P., & Liu, J. (2025). Joint u-nets with hierarchical graph structure and sparse transformer for hyperspectral image classification. *Expert Systems with Applications*, *275*, 127046. https://doi.org/10.1016/j.eswa.2025.127046