Decoupling spatial and spectral features for efficient hyperspectral image super-resolution

Machine Vision and Applications (2026) 37:99
https://doi.org/10.1007/s00138-026-01857-2

RESEARCH

Decoupling spatial and spectral features for efficient hyperspectral image super-resolution

Matteo Kolyszko[1] · **Marco Buzzelli**[1] · **Simone Bianco**[1] · **Raimondo Schettini**[1]

Received: 12 January 2026 / Revised: 13 May 2026 / Accepted: 24 May 2026

Abstract

Hyperspectral image super-resolution (HSI-SR) aims to reconstruct hyperspectral images at high spatial resolution, starting from low-resolution inputs, while preserving both spatial details and spectral fidelity. In this work, we propose the Efficient Spatial-Spectral Processing Network (ESSPN), a lightweight deep learning architecture designed to address the challenges of HSI-SR in a computationally efficient way. ESSPN is built around a novel Spatial-Spectral Block (SSB) that separately models spatial structures and spectral correlations through residual convolutional and attention mechanisms. The network head incorporates an efficient upsampling module based on pixel shuffle decomposition to produce high-resolution outputs without interpolation artifacts. Extensive experiments on two publicly available datasets, i.e., ARAD1K and StereoMSI demonstrate that ESSPN achieves competitive or superior performance compared to state-of-the-art methods when evaluated at scale factors of $\times 4$, $\times 6$ and $\times 8$. Notably, the model shows strong generalization across hyperspectral cameras with varying spectral responses, and across radiometric domains, covering both radiance and reflectance measurements, while requiring significantly fewer parameters and FLOPs compared to existing methods. These results position the proposed ESSPN as a practical and effective solution for high-quality hyperspectral image super-resolution in real-world applications.

Keywords Super-resolution · Lightweight · Hyperspectral imaging · Computer vision

1 Introduction

Hyperspectral imaging (HSI) has emerged as a powerful sensing modality capable of capturing detailed spectral information at each spatial location in a scene. Unlike conventional imaging systems that acquire only a few broadband channels (e.g., RGB), HSI captures tens to hundreds of narrow spectral bands, enabling precise material identification and fine-grained spectral analysis. This unique combination of spatial and spectral information makes hyperspectral imaging valuable across a wide range of applications, including remote sensing [1], precision agriculture [2], environmental monitoring [3], and biomedical imaging [4]. Despite its advantages, hyperspectral imaging is fundamentally constrained by the limited amount of incident energy, which forces a trade-off between spatial and spectral resolution in practical systems [5]. As the number of spectral bands increases, maintaining a high signal-to-noise ratio (SNR) often requires sacrificing spatial resolution. Consequently, obtaining high-resolution hyperspectral images remains a major challenge in the field. Enhancing spatial resolution without compromising spectral integrity remains a central challenge in hyperspectral imaging. Super-resolution (SR) techniques aim to recover high-resolution images from one or more low-resolution observations and provide a post-processing alternative to improve spatial detail without modifying the acquisition hardware. In the context of hyperspectral data, SR can be broadly categorized into two paradigms: fusion-based methods [6, 7] and single-image methods [8,

✉ Matteo Kolyszko
m.kolyszko@campus.unimib.it

Marco Buzzelli
marco.buzzelli@unimib.it

Simone Bianco
simone.bianco@unimib.it

Raimondo Schettini
raimondo.schettini@unimib.it

[1] Department of Informatics, Systems and Communication, University of Milano - Bicocca, Viale Sarca 336, 20125 Milan, Italy

Published online: 18 June 2026

9]. Fusion-based approaches leverage high-resolution auxiliary data, such as RGB or panchromatic images to guide the upsampling of the hyperspectral image; while effective, these methods typically assume that the hyperspectral and auxiliary images are spatially aligned, which is often difficult to guarantee in real-world conditions due to parallax, motion, or differences in sensor geometry. In contrast, single-image hyperspectral super-resolution (single-HSI-SR) methods attempt to enhance spatial resolution using only the information present in the low-resolution hyperspectral image itself. The task faced in the latter SR paradigm is significantly more challenging, as it must rely solely on the intrinsic correlations within the hyperspectral image, both spatially and spectrally. Early approaches employed sparse representation [10], dictionary learning [11], or low-rank modeling [12] to exploit spectral redundancy. However, such handcrafted priors often lack generalization and struggle to capture the full complexity of natural scenes. More recently, deep learning has shown strong potential in this field, with convolutional neural networks (CNNs) [13], and attention-based architectures [8] demonstrating impressive results in image reconstruction tasks.

Despite these advances, hyperspectral image super-resolution remains fundamentally challenging due to several intrinsic limitations of hyperspectral data. First, hyperspectral images are characterized by high spectral dimensionality, often consisting of tens to hundreds of contiguous bands, which leads to increased computational complexity and memory requirements [14].

Second, hyperspectral imaging systems inherently involve trade-offs between spatial resolution, spectral resolution, and signal-to-noise ratio, making accurate reconstruction particularly difficult [14].

Third, hyperspectral data are affected by various degradations, including noise, calibration errors, and illumination variability, which further complicate both spatial detail recovery and spectral fidelity [14].

These challenges highlight the need for efficient hyperspectral super-resolution methods that can effectively handle high-dimensional data while maintaining strong reconstruction performance.

To address these limitations, we propose the Efficient Spatial-Spectral Processing Network (ESSPN), a lightweight and modular architecture tailored for single-image hyperspectral super-resolution. Our design integrates spatial and spectral processing through distinct but complementary modules, enabling efficient joint feature learning through a novel Spatial-Spectral Block (SSB), and employs an efficient upsampling strategy based on pixel shuffle decomposition. By combining parameter sharing, spectral grouping, and lightweight residual structures, ESSPN achieves high reconstruction quality with fewer parameters and fewer FLOPs than existing methods.

We summarize the main contributions of this work as follows:

- We design a novel SSB that separates spatial enhancement and spectral attention into dedicated residual pathways, allowing effective feature extraction while preserving the consistency and smoothness of spectral signatures across bands.
- We introduce a lightweight and scalable architecture (ESSPN) that achieves state-of-the-art performance with significantly fewer parameters and lower inference cost compared to existing methods.
- We demonstrate that our model generalizes well across different datasets, hyperspectral cameras, and radiometric domains such as reflectance and radiance, validating its robustness and practical applicability.

The rest of the paper is organized as follows: Sect. 2 reviews related works in hyperspectral and super-resolution. Section 3 presents the architecture of ESSPN and its building blocks. Section 4 describes the datasets, evaluation metrics, and implementation details. Quantitative and qualitative results are discussed in Sect. 5, followed by conclusions in Sect. 6.

2 Related work

2.1 Gray/RGB image super-resolution

Single-image super-resolution for gray-scale and RGB imagery has advanced considerably with the development of deep convolutional architectures. Dong et al. [15] introduced SRCNN, a foundational three-layer CNN that achieved substantial improvements over prior non-deep learning techniques. Deeper residual networks like VDSR [16] and DRCN [17] further enhanced reconstruction quality by employing deeper layers with residual learning.

This trend continued with models such as LapSRN [18], DRRN [19], and EDSR [20], which introduced hierarchical and residual designs to improve performance and stability. More sophisticated architectures, such as RDN [21], DBPN [22], and NLRN [23], incorporated feedback mechanisms and non-local operations for improved context modeling.

Zhang et al. introduced the Residual Channel Attention Network (RCAN) [24], incorporating channel-wise attention modules inspired by the Squeeze-and-Excitation (SE) block [25] to emphasize informative spectral features. Dai et al. [26] further advanced this line of work by introducing

SAN, a Second-order Attention Network capable of capturing long-range spatial dependencies.

Despite these successes, applying RGB-based SR networks directly to hyperspectral images is nontrivial. These models are optimized for 1–3 input channels and rely on early feature expansion to dozens of channels. Applying such scaling to hyperspectral images, often containing over a hundred bands, would result in an exponential increase in parameters, requiring more data than is typically available in the HSI domain. Moreover, the lack of explicit modeling of inter-band spectral correlations in these networks leads to suboptimal performance when applied to hyperspectral data.

2.2 Spatial super-resolution for hyperspectral images

Spatial super-resolution for hyperspectral imagery is an active area of research due to its practical importance in applications such as remote sensing [1], precision agriculture [2], and medical imaging [4]. One of the earliest contributions in this domain was presented by Akgun et al. [27], who proposed reconstruction via a hyperspectral image formation model, alongside the Projection Onto Convex Sets (POCS) algorithm [28] to recover high-resolution representations. Later, Huang et al. [12] proposed a method that explicitly models unknown blur kernels i.e., spatially-invariant degradation filters by integrating low-rank priors and group-sparsity regularization.

Sparse coding and dictionary learning frameworks have been widely studied for super-resolution [10, 11], providing structured signal representations through learned sparse coefficients. However, these methods generally require solving iterative optimization problems at inference time, resulting in high computational cost. Furthermore, their performance is often limited by the reliance on hand-crafted priors or internal patch recurrence assumptions, which may fail to generalize to images with less self-similarity or complex textures. With the emergence of deep learning, several works have leveraged convolutional neural networks (CNNs) to model both spatial and spectral features. Early deep methods by Yuan et al. [10] and Xie et al. [11] applied deep CNNs followed by Nonnegative Matrix Factorization (NMF) to independently handle spatial enhancement and spectral consistency. However, the absence of end-to-end trainability mainly affects training efficiency, while inference remains unaffected.

Mei et al. [13] proposed a fully 3D CNN architecture to directly exploit spectral continuity, though this approach is computationally demanding due to the high dimensionality of hyperspectral data. Li et al. [8] addressed this issue by proposing a Grouped Deep Recursive Residual Network (GDRRN), which incorporates recursive blocks within a global residual learning framework. This design balances efficiency and performance by reusing parameters while maintaining representational power.

Recurrent architectures have also been explored to better exploit sequential dependencies across spectral bands. The Bidirectional 3D Quasi-Recurrent Neural Network (B3DQRNN) proposed by Fu et al. [29] models global spectral correlations while maintaining 3D spatial–spectral consistency, offering an efficient alternative to heavy 3D CNNs through quasi-recurrent gating mechanisms.

In subsequent research, feature pyramid architectures were introduced to capture multi-scale spatial information effectively [9]. Additionally, Sidorov et al. [30] adapted the Deep Image Prior paradigm (originally designed for natural image restoration) to the hyperspectral domain, demonstrating that CNNs can implicitly encode useful priors even without large-scale training data.

Recent transformer-based approaches have further advanced hyperspectral super-resolution by improving the modeling of long-range spatial and spectral dependencies. In particular, ESSAformer [31] introduces an efficient attention mechanism that reduces the computational burden of standard transformers while preserving strong spatial–spectral feature interactions, achieving competitive performance with significantly lower complexity.

More recently, alternative paradigms beyond convolutional and transformer-based models have been explored for image super-resolution. In particular, state space models have emerged as an efficient framework for sequence modeling. MambaHSISR [32] extends this paradigm to hyperspectral image super-resolution by leveraging selective state space mechanisms to capture long-range spatial–spectral dependencies with improved computational efficiency compared to attention-based approaches.

Building on this line of research, MambaX [33] introduces a dynamic state predictive control formulation, where the super-resolution process is modeled as a sequence of latent states governed by learnable control equations. This approach enables explicit modeling of the reconstruction dynamics and mitigates error accumulation across stages, while enhancing the representation capability through nonlinear state transitions and cross-domain interactions.

In parallel, neural operator-based methods have been introduced to model global image transformations in a continuous manner. KANO [34] proposes a Kolmogorov–Arnold Neural Operator for image super-resolution, which approximates complex mappings through functional decomposition, enabling strong representation capability with a different inductive bias compared to conventional deep networks.

Despite the strong progress in hyperspectral image super-resolution, computational efficiency remains a major limitation of current approaches. Recent high-performing methods, including deep CNNs [20], 3D convolutional networks [13], and transformer-based architectures [31], achieve excellent reconstruction quality but often at the cost of high computational complexity and long inference times. This issue is particularly critical in hyperspectral imaging, where the high dimensionality of spectral data significantly increases both memory usage and processing requirements. As a result, many state-of-the-art models are difficult to deploy in real-time or resource-constrained scenarios, such as embedded systems or onboard processing.

These limitations highlight the need for lightweight architectures that can maintain competitive reconstruction performance while significantly reducing computational cost. In this context, designing efficient mechanisms to process spatial and spectral information remains a key challenge.

3 Proposed method

3.1 Efficient spatial-spectral processing network (ESSPN)

We propose the Efficient Spatial-Spectral Processing Network as shown in Fig. 1 for spatial super-resolution. The architecture consists of a stack of residual SSBs, described in Sect. 3.2, which progressively refine spatial and spectral information, followed by an upsampling module that reconstructs the high-resolution (HR) output from the processed low-resolution (LR) input.

Given an input feature map $\mathrm{x} \in \mathbb{R}^{C\times H\times W}$, the network processes it through a sequence of N SSB modules, each composed of a spatial residual branch (Sect. 3.3) and a spectral attention branch (Sect. 3.4). The final output is generated by an efficient pixel shuffle-based Upsampler (Sect. 3.5).

$$\mathrm{y}_{\mathrm{HR}} = \mathrm{Upsampler}\left(\mathcal{F}^{(N)}_{\mathrm{SSB}} \circ \cdots \circ \mathcal{F}^{(1)}_{\mathrm{SSB}}(\mathrm{x})\right) \tag{1}$$

where y_{HR} denotes the reconstructed high-resolution image, x is the low-resolution input, and $\mathcal{F}^{(i)}_{\mathrm{SSB}}$ indicates the i-th Spatial-Spectral Block.

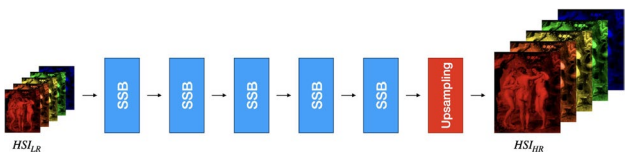

Fig. 1 The overall network architecture of the proposed ESSPN network

After feature refinement, the network applies a dedicated Upsampler block to increase the resolution by a desired scaling factor s. The upsampler uses pixel shuffle operations interleaved with convolutions to reconstruct the HR output efficiently.

The modular design of the proposed ESSPN model enables:

- Separation of spatial and spectral learning, with dedicated blocks for each;
- Scalability in depth by adjusting the number of SSB blocks;
- High-quality reconstruction, supported by efficient convolutional design and attention mechanisms further detailed in Sect. 3.3 and in Sect. 3.4;
- Improved inference efficiency compared to state-of-the-art methods using bicubic or transposed convolution upsampling, thanks to the use of pixel shuffle and a lightweight convolutional architecture. The latter increases modeling capacity during training while being re-parameterized into simpler operations at inference, preserving speed and efficiency.

The sequential application of spatial and spectral processing leverages the complementary nature of the two domains in hyperspectral images. Enhancing spatial structures like edges and textures first, and then refining spectral information through channel-wise attention, allows the model to focus on each aspect in a targeted and interpretable way.

This sequential spatial and spectral processing is made possible by the repeated use of the proposed SSB, which we describe in detail in the following section.

3.2 Spatial-spectral block

To better leverage the distinct characteristics of spatial and spectral information in hyperspectral data, we introduce the SSB, a dual-stage residual module inspired by [9]. As shown in Fig. 2, it first refines spatial features using residual convolutions, and then models local spectral dependencies through a lightweight attention mechanism, allowing for the extraction of complementary features sequentially.

The SSB consists of the following components:

- A Spatial Residual Block as described in Sect. 3.3;
- A Spectral Attention block as described in Sect. 3.4;

Let $\mathrm{x} \in \mathbb{R}^{C\times H\times W}$ be the input feature map. The output y of the SSB is defined as:

$$\mathrm{y} = \mathcal{F}_{\mathrm{SAB}}\left(\mathcal{F}_{\mathrm{SRB}}(\mathrm{x})\right), \tag{2}$$

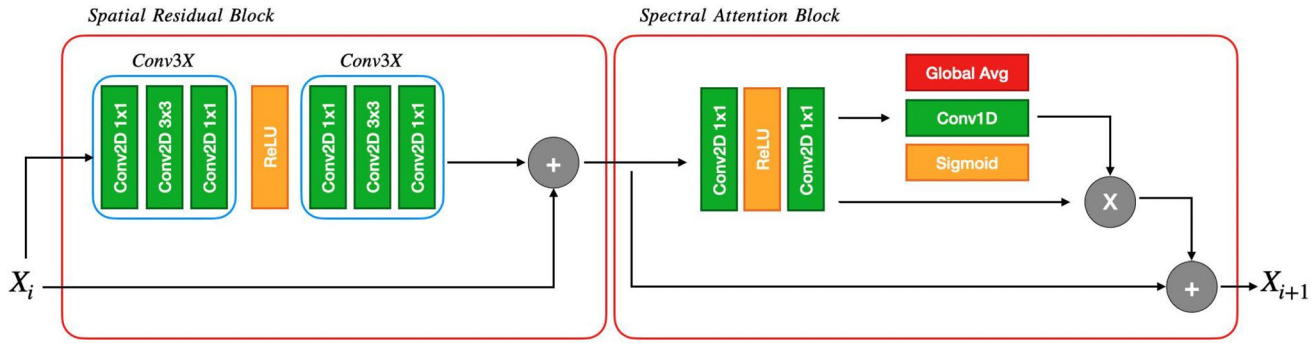

Fig. 2 The network architecture of the spatial-spectral block (SSB), which consists of a spatial residual module and a spectral attention residual module. "+" and "×" denote element-wise addition and element-wise multiplication, respectively

where $\mathcal{F}_{\text{SRB}}$ is the spatial residual block and $\mathcal{F}_{\text{SAB}}$ is the spectral attention block.

The SSB module decouples spatial and spectral processing, allowing each to be enhanced with dedicated mechanisms. This is particularly effective in spatial super-resolution, where:

- Spatial structures such as textures and edges benefit from deeper local context via residual convolutions;
- Spectral consistency across channels is preserved through the spectral attention mechanism, which captures local spectral dependencies via convolution on the spectral descriptor;
- Sequential refinement ensures spatial details are first extracted and then selectively modulated based on spectral importance.

The composition of spatial and spectral pathways in each SSB relies on custom-designed convolutional and attention mechanisms which we now examine individually.

3.3 Spatial residual block

In the spatial branch of each SSB block, we employ a compositional convolution module called Conv3XC as introduced by Wan et al. [35], which allows us to increase the representational capacity during training while maintaining inference efficiency.

Given an input $\mathrm{x} \in \mathbb{R}^{C\times H\times W}$, a residual block outputs:

$$\mathrm{y} = \mathrm{x} + \lambda \cdot \mathcal{F}(\mathrm{x}), \tag{3}$$

where $\mathcal{F}$ is composed of two convolutional layers and optional non-linearities or normalization and $\lambda \in (0, 1]$ is the residual scaling factor used to improve training stability.

The Conv3XC module replaces a standard 3×3 convolution with a sequence of three operations:

$$\mathcal{C}(\mathrm{x}) = \mathrm{Conv}^{\text{in}}_{1\times 1}\left(\mathrm{Conv}_{3\times 3}\left(\mathrm{Conv}^{\text{out}}_{1\times 1}(\mathrm{x})\right)\right), \tag{4}$$

where $\mathrm{Conv}^{\text{out}}_{1\times 1}$ increases the number of channels (channel expansion), $\mathrm{Conv}_{3\times 3}$ performs spatial filtering, and $\mathrm{Conv}^{\text{in}}_{1\times 1}$ reduces the number of channels back to the output size (channel reduction).

Compared to conventional spatial modeling strategies based on standard 3×3 convolutions, as commonly adopted in existing hyperspectral super-resolution methods [9, 20, 24], the proposed design uses Conv3XC as a more expressive alternative. Unlike standard convolutions, Conv3XC employs a re-parameterized structure that enhances the representational capacity during training while maintaining the same computational cost at inference time. From an architectural perspective, this modification enables a more effective extraction of spatial features without increasing complexity. From a representation standpoint, the richer parameterization improves the modeling of fine spatial details, which is crucial for high-quality reconstruction.

3.4 Spectral attention block

In the spectral branch of each SSB module, we adopt a lightweight attention mechanism designed to capture local dependencies across spectral bands. This is essential to preserve the continuity and smoothness of spectral signatures, which are crucial for accurate hyperspectral reconstruction.

Formally, the output is computed as:

$$\mathrm{Y} = \mathrm{X} + \lambda \cdot \mathcal{A}(\phi(\mathrm{X})), \tag{5}$$

where ϕ denotes a 1×1 convolution applied to the input $\mathrm{X} \in \mathbb{R}^{C\times H\times W}$, and $\mathcal{A}$ represents the spectral attention mechanism. The residual scaling factor $\lambda \in (0, 1]$, as in the Spatial Residual Block, improves training stability.

Given $\mathrm{X} \in \mathbb{R}^{C\times H\times W}$, we first compute a spectral descriptor via global average pooling:

$$z_c = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{c,i,j}, \quad \forall c = 1, \ldots, C, \tag{6}$$

resulting in $\mathrm{z} \in \mathbb{R}^C$.

We then apply a learnable 1D convolution over the spectral dimension:

$$\mathrm{s} = \sigma(\mathrm{z} * \mathrm{k}), \tag{7}$$

where $\mathrm{k} \in \mathbb{R}^K$ is a convolution kernel of size K, $*$ denotes a 1D convolution along the spectral dimension with appropriate padding to preserve the dimensionality (i.e., $\mathrm{s} \in \mathbb{R}^C$), and $\sigma(\cdot)$ is the sigmoid activation function.

More explicitly, each element s_c is computed as:

$$s_c = \sigma\left(\sum_{i=1}^{K} k_i \cdot z_{c+i-\lfloor K/2 \rfloor} \right), \tag{8}$$

where zero-padding is applied at the boundaries.

The feature map is then modulated channel-wise as:

$$Y_c = X_c \cdot s_c, \quad \forall c = 1, \ldots, C, \tag{9}$$

resulting in $\mathrm{Y} \in \mathbb{R}^{C \times H \times W}$.

Unlike the channel attention layer proposed by Jiang et al. [9], which relies on two fully connected layers and introduces a large number of parameters without modeling local spectral continuity, our design:

- Captures local spectral dependencies via a 1D convolution over the spectral descriptor;
- Requires fewer parameters (reducing the complexity from $2C^2/r$ to K, where K is the kernel size);
- Preserves the smoothness of spectral signatures by explicitly modeling local spectral continuity.

The primary objective of this block is to improve spectral fidelity rather than spatial sharpness. Therefore, its effectiveness is expected to be mainly reflected by spectral metrics such as Spectral Angle Mapper (SAM), while only marginal changes are expected in spatial or structural metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). This is particularly important in hyperspectral imaging, where accurate preservation of spectral signatures is crucial. In this context, reducing the SAM is essential, as it helps mitigate metamerism effects, where different spectral distributions may produce similar spatial or color responses. Therefore, improving SAM directly contributes to more reliable material discrimination and spectral analysis.

3.5 Efficient upsampling via pixel shuffle decomposition

To perform learnable upsampling without artifacts introduced by interpolation or transposed convolutions, we employ the PixelShuffle introduced by Shi et al. [36].

Given an input $\mathrm{X} \in \mathbb{R}^{C_{\text{in}} \times H \times W}$, a convolution is applied to expand the channels:

$$\mathrm{F} = \mathrm{W} * \mathrm{X} + \mathrm{b}, \quad \mathrm{W} \in \mathbb{R}^{r^2 \cdot C_{\text{out}} \times C_{\text{in}} \times k \times k} \tag{10}$$

where r is the desired upsampling factor and k the kernel size.

The expanded channels are spatially rearranged to increase resolution:

$$\mathrm{PixelShuffle}(\mathrm{F}) : \mathbb{R}^{r^2 \cdot C_{\text{out}} \times H \times W} \to \mathbb{R}^{C_{\text{out}} \times rH \times rW} \tag{11}$$

resulting in a high-resolution output $\mathrm{Y} \in \mathbb{R}^{C_{\text{out}} \times rH \times rW}$.

This upsampling strategy forms the final stage of the ESSPN, transforming enhanced low-resolution features into high-resolution outputs with minimal computational overhead and without introducing checkerboard artifacts.

4 Experimental setup

In this section, we describe the experimental setup used to evaluate the proposed method. We first introduce the datasets, followed by the evaluation metrics and implementation details.

4.1 Datasets

We evaluate our method on two widely used hyperspectral datasets that cover a broad range of acquisition devices, spectral resolutions and scene types: ARAD1K and StereoMSI dataset.

4.1.1 ARAD1K natural hyperspectral image dataset

The ARAD1K dataset, introduced in the NTIRE 2022 Spectral Recovery Challenge [37], is currently the largest publicly available collection of real-world natural hyperspectral images. It comprises 1,000 high-resolution hyperspectral scenes acquired using a Specim IQ push-broom hyperspectral camera.

Each image has an original spatial resolution of 512×512 pixels and covers 204 spectral bands in the 400–1000 nm range. For dataset construction, raw radiance measurements were radiometrically calibrated and uniformly resampled to

31 spectral bands within the visible spectrum (400–700 nm, 10 nm intervals), yielding images of size $482 \times 512 \times 31$.

The original dataset split included 900 training images, 50 validation images, and 50 test images. However, because the official test split provides only RGB data without ground-truth hyperspectral labels, we redefined the dataset into 850 training images, 50 validation images, and 50 test images for our experiments.

4.1.2 StereoMSI dataset

The StereoMSI dataset was introduced in the PIRM 2018 Spectral Image Super-Resolution Challenge [38] to benchmark both single-image and RGB-guided spectral super-resolution methods. It contains 350 stereo image pairs, each consisting of a spatially aligned RGB and multispectral image captured across diverse environments (urban, industrial, office, rainforest, and desert scenes in Canberra, Australia).

Each pair comprises a high-resolution RGB image acquired with a XiQ MQ022CG-CM CMOS camera and a low-resolution multispectral image captured using a XiQ MQ022HG-IM-SM4x4 sensor equipped with a 4×4 filter array, providing 14 effective spectral bands in the 470–620 nm visible range.

Because of the sensor mosaic layout, RGB images have four times the spatial resolution of their multispectral counterparts. The processed image resolutions are 960×480 pixels for RGB and 480×240 pixels for spectral images.

4.2 Evaluation metrics

To quantitatively evaluate the performance of spectral super-resolution, we use three standard metrics: Spectral Angle Mapper (SAM) [39], Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [40]. Each metric provides a complementary perspective on reconstruction quality, covering spectral fidelity, signal-level error, and perceptual similarity.

4.2.1 Spectral angle mapper (SAM)

SAM is a commonly used metric in hyperspectral imaging to measure the angular difference between the predicted and ground-truth spectral vectors. For a pixel with spectral vectors r (ground truth) and $\hat{\mathrm{r}}$ (prediction), SAM is defined as:

$$\mathrm{SAM}(\mathrm{r}, \hat{\mathrm{r}}) = \cos^{-1}\left(\frac{\langle \mathrm{r}, \hat{\mathrm{r}} \rangle}{\|\mathrm{r}\|_2 \cdot \|\hat{\mathrm{r}}\|_2}\right) \tag{12}$$

The final SAM score is obtained by first computing the spectral angle (in radians) for each valid pixel in an image, then averaging over all valid pixels to obtain a per-image SAM value. The reported score is the mean SAM across all test images. Lower SAM values indicate better spectral similarity.

4.2.2 Peak signal-to-noise ratio (PSNR)

PSNR is used to measure the fidelity of the reconstructed image compared to the ground truth in terms of pixel-wise intensity values. Let $I \in \mathbb{R}^{H \times W \times C}$ be the ground-truth high-resolution image, and $\hat{I} \in \mathbb{R}^{H \times W \times C}$ the reconstructed image. The Mean Squared Error (MSE) is computed as:

$$\mathrm{MSE} = \frac{1}{HWC} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{c=1}^{C} \left(\hat{I}_{i,j,c} - I_{i,j,c}\right)^2, \tag{13}$$

where H, W, and C denote the height, width, and number of spectral channels, respectively.

Then, PSNR is computed as:

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{MAX^2}{\mathrm{MSE}}\right) \tag{14}$$

where MAX is the maximum possible pixel value (e.g., 1.0 for normalized data). Higher PSNR indicates lower reconstruction error.

4.2.3 Structural similarity index (SSIM)

Originally proposed for grayscale images, the Structural Similarity Index (SSIM) assesses perceptual image quality by comparing local patterns of luminance, contrast, and structure between two image patches. In our setting, where images have multiple spectral channels, SSIM is computed channel-wise and then averaged to produce a global score, following standard practice in multispectral and hyperspectral image quality assessment.

Given a local patch from the predicted image $\hat{I}$ and the corresponding patch from the ground truth image I, the SSIM is defined as:

$$\mathrm{SSIM}(I, \hat{I}) = \frac{(2\mu_I \mu_{\hat{I}} + C_1)(2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)}, \tag{15}$$

where μ_I, $\mu_{\hat{I}}$ are the local means, σ_I^2, $\sigma_{\hat{I}}^2$ the variances, and $\sigma_{I\hat{I}}$ the covariance between the patches. Constants C_1 and C_2 are used to stabilize the division and avoid numerical instability.

The SSIM score is computed using a sliding Gaussian window (typically of size 11×11) over each channel, and the final result is obtained by averaging across spatial

locations and channels. SSIM values range in $[0, 1]$, with higher scores indicating better structural fidelity.

4.3 Implementation details

The proposed ESSPN model was implemented in PyTorch and trained for 780 epochs on a single NVIDIA TITAN X GPU, requiring approximately 26 h.

The network is composed of $N = 5$ stacked Spatial-Spectral Blocks (SSBs). We optimized the model parameters using the Adam optimizer with an initial learning rate of $\eta = 5 \times 10^{-4}$ and a weight decay of 1×10^{-5}. The training was performed with a batch size of $B = 8$.

To adapt the learning rate during training, we employed a *ReduceLROnPlateau* scheduler that monitors the validation loss. The learning rate is reduced by a factor of 0.5 if the validation loss does not improve for 50 consecutive epochs.

For the generation of low-resolution (LR) inputs, high-resolution (HR) images are first convolved with a Gaussian kernel of size 5×5 and standard deviation $\sigma = 1.0$, followed by bicubic downsampling with scaling factors $s \in \{4, 6, 8\}$.

The training loss combines both spatial and spectral total variation (TV) terms to encourage smoothness in the reconstructed hyperspectral images:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{TV}}^{\text{spatial}} + \mathcal{L}_{\text{TV}}^{\text{spectral}}. \tag{16}$$

Spatial total variation loss The spatial TV loss promotes local smoothness along the spatial dimensions (height and width) of each image:

$$\mathcal{L}_{\text{TV}}^{\text{spatial}} = \frac{1}{B} \sum_{b=1}^{B} \|\nabla_h x_b\|_2^2 + \|\nabla_w x_b\|_2^2, \tag{17}$$

where B is the batch size, $x_b \in \mathbb{R}^{H \times W \times C}$ denotes the b-th sample in the batch, $\nabla_h x_b$ and $\nabla_w x_b$ represent the finite difference gradients of x_b along the height (H) and width (W) dimensions, respectively, and $\| \cdot \|_2$ denotes the Euclidean (ℓ_2) norm.

Spectral total variation loss To ensure spectral smoothness, the spectral TV loss penalizes variations between adjacent spectral bands:

$$\mathcal{L}_{\text{TV}}^{\text{spectral}} = \frac{1}{B} \sum_{b=1}^{B} \|\nabla_c x_b\|_2^2, \tag{18}$$

where $\nabla_c x_b$ denotes the finite difference gradient along the channel (spectral) dimension C.

Together, these losses act as regularizers to reduce high-frequency noise and preserve the inherent smoothness of natural hyperspectral data across both spatial and spectral domains.

5 Experimental results

In this section, we present a comprehensive evaluation of our proposed method on two hyperspectral datasets considered in this study. These datasets cover a wide range of acquisition conditions, spatial resolutions, and spectral characteristics, including both radiance and reflectance data. We compare our model against several state-of-the-art approaches, including traditional interpolation (Bicubic), general single-image super-resolution networks originally designed for RGB images, namely RCAN [24] and EDSR [20], and four dedicated hyperspectral super-resolution methods, namely SSPSR [9], ESSAFormer [31], Bi-3DQRNN [29] and MambaHSISR [32]. For each method, we tune the hyperparameters and configurations to ensure fair comparison and report results at scaling factors $\times 4$, $\times 6$ and $\times 8$.

5.1 Results on ARAD1K dataset

Table 1 reports the quantitative performance of all competing methods at $4\times$, $6\times$, and $8\times$ super-resolution on the ARAD1K dataset. MambaHSISR achieves the highest PSNR across all scales, while ESSAFormer attains the best SSIM at $4\times$ and $6\times$, and remains competitive at $8\times$.

Our method consistently demonstrates strong overall performance, achieving the lowest SAM values at all scales (0.026, 0.036, and 0.049 for $4\times$, $6\times$, and $8\times$, respectively), indicating superior spectral fidelity and better preservation of spectral consistency across wavelengths. In terms of structural quality, it ranks among the top-performing methods, maintaining competitive SSIM scores while remaining close to the best-performing approaches.

Importantly, our model is significantly more efficient than all competing methods. It requires only 9.56, 4.25, and 2.16 GFLOPs for $4\times$, $6\times$, and $8\times$ super-resolution, respectively, and achieves the fastest inference times (4.50 ms, 2.10 ms, and 1.04 ms), measured on a single NVIDIA TITAN X GPU. This highlights a favorable trade-off between performance and efficiency, making the proposed method particularly suitable for real-time and resource-constrained applications.

Figures 3 and 4 provide a qualitative comparison of the reconstructed hyperspectral images and their corresponding error maps at representative wavelengths (480 nm, 580 nm, and 680 nm). In the spatial reconstructions of Fig. 3, Bicubic interpolation naturally lacks the ability to generate

Table 1 Quantitative comparison on the ARAD1K dataset at $4\times$, $6\times$ and $8\times$ super-resolution scales.

Scale	Method	PSNR↑	SSIM↑	SAM↓	GFLOPs↓	Runtime(ms)↓
4×	Bicubic	33.21	0.886	0.040	N/A	N/A
	RCAN	36.61	0.922	0.036	245.2	476.7
	EDSR	37.18	0.927	0.033	776.7	312.5
	SSPSR	36.56	0.929	0.032	902.6	216.1
	ESSAFormer	37.25	**0.941**	0.028	97.30	46.35
	Bi-3DQRNN	36.90	0.931	0.030	122.7	57.83
	MambaHSISR	**37.90**	0.936	0.027	11.19	16.80
	Ours	37.22	0.935	**0.026**	**9.56**	**4.50**
6×	Bicubic	30.99	0.833	0.050	N/A	N/A
	RCAN	33.51	0.882	0.046	116.3	232.8
	EDSR	33.62	0.881	0.043	434.4	189.5
	SSPSR	33.59	0.884	0.038	736.5	132.6
	ESSAFormer	34.02	**0.890**	0.037	48.64	24.20
	Bi-3DQRNN	33.70	0.886	0.039	61.35	28.10
	MambaHSISR	**34.08**	0.887	0.037	5.60	8.40
	Ours	34.04	0.888	**0.036**	**4.25**	**2.10**
8×	Bicubic	27.10	0.742	0.068	N/A	N/A
	RCAN	29.45	0.801	0.061	116.3	232.8
	EDSR	29.62	0.806	0.058	434.4	189.5
	SSPSR	29.71	0.811	0.054	736.5	132.6
	ESSAFormer	30.08	**0.823**	0.051	48.64	24.20
	Bi-3DQRNN	29.80	0.815	0.053	61.35	28.10
	MambaHSISR	**30.09**	0.820	0.050	3.80	4.53
	Ours	30.05	0.819	**0.049**	**2.16**	**1.04**

Metrics include PSNR, SSIM, SAM, computational complexity (GFLOPs), and runtime. Best results are shown in bold, and second-best results are underlined

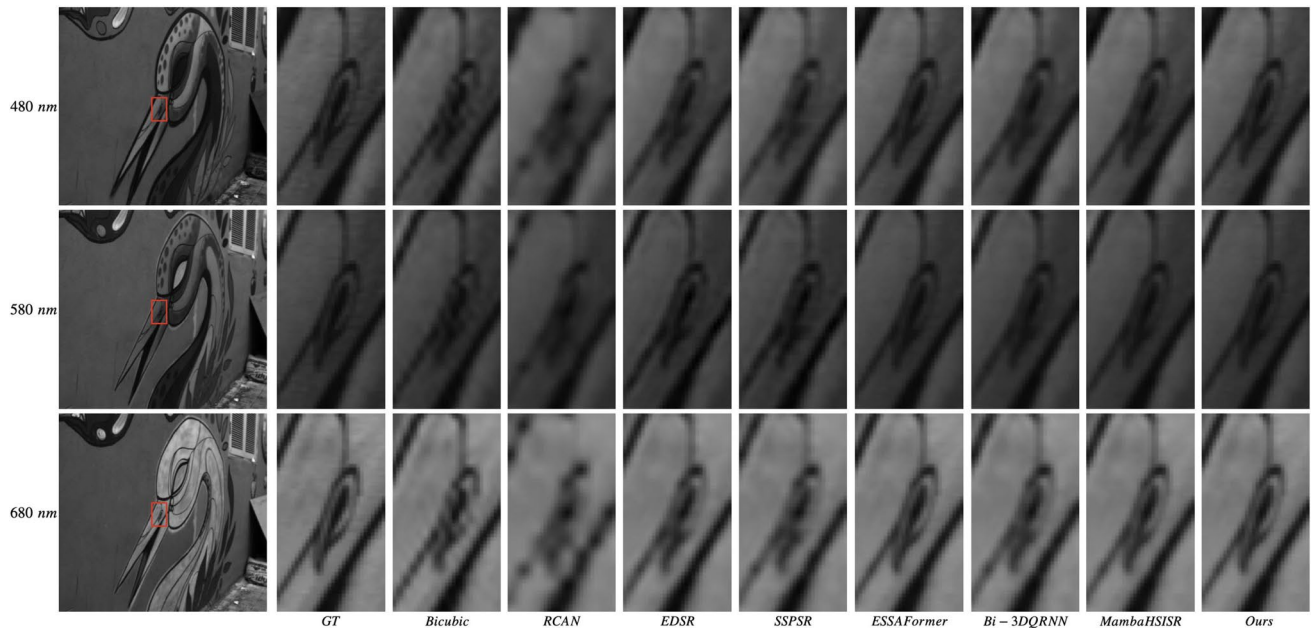

Fig. 3 Visual comparison of reconstructed hyperspectral images from ARAD1K at wavelengths 480 nm, 580 nm, and 680 nm, under $\times 4$ super-resolution. Rows correspond to spectral bands; columns to different methods. All images are displayed using the same grayscale range to allow consistent visual comparison across methods

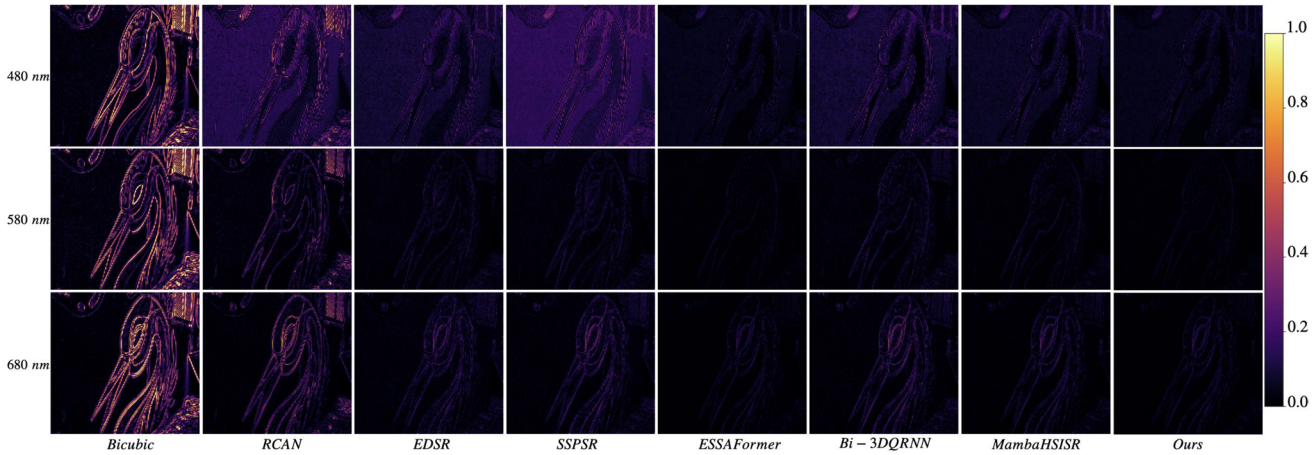

Fig. 4 Error maps from ARAD1K at wavelengths 480 nm, 580 nm, and 680 nm, under $\times 4$ super-resolution. Rows correspond to spectral bands; columns to different methods. Brighter regions represent higher reconstruction errors, highlighting differences in spatial accuracy among the models

high-frequency details, as it can only interpolate from the available low-resolution content. While RCAN and EDSR introduce noticeable distortions and fail to accurately recover thin contours. SSPSR improves the sharpness but still shows irregularities in low-contrast regions. ESSAFormer achieves the visually cleanest reconstructions among all baselines, recovering sharp edges and maintaining good local contrast across wavelengths. Bi-3DQRNN shows smoother structures but tends to oversmooth fine details. Our method produces reconstructions that remain competitive with the best-performing baselines, reconstructing fine structural details effectively.

The error maps in Fig. 4 further highlight these differences. Traditional CNN-based approaches leave strong residual edges and textured artifacts, while SSPSR and Bi-3DQRNN reduce error intensity but still retain visible discrepancies around curved edges. ESSAFormer yields the cleanest error distribution, with lower residuals particularly in high-frequency regions. Our method achieves low and spatially consistent errors across wavelengths. These visual results align with the quantitative SSIM and SAM scores, confirming that ESSAFormer provides the most accurate spatial reconstruction, while our model offers competitive quality with significantly lower computational cost.

The quantitative metrics and qualitative visualizations jointly demonstrate that our approach achieves the best balance between reconstruction accuracy, spectral consistency, and computational efficiency on ARAD1K.

5.2 Results on StereoMSI dataset

Table 2 summarizes the quantitative results on the StereoMSI dataset at $4\times$, $6\times$, and $8\times$ super-resolution scales. As expected, overall performance is lower than on ARAD1K due to the reduced spectral dimensionality and the more challenging acquisition conditions of StereoMSI, while a consistent ranking among methods is largely preserved across scales.

MambaHSISR achieves the highest PSNR at $4\times$ and $8\times$, while our method attains the best PSNR at $6\times$ and remains highly competitive at the other scales. In terms of structural similarity, ESSAFormer consistently provides the highest SSIM across all scaling factors, with both MambaHSISR and our approach closely following.

Regarding spectral fidelity, our method demonstrates clear advantages by achieving the lowest SAM values at all scales (0.025, 0.035, and 0.049 for $4\times$, $6\times$, and $8\times$, respectively), indicating superior preservation of spectral consistency. MambaHSISR also shows strong performance in this metric, consistently ranking among the top methods.

From an efficiency perspective, the differences between models become even more pronounced. Our method requires only 6.80, 3.10, and 1.02 GFLOPs for $4\times$, $6\times$, and $8\times$, respectively, and achieves the fastest inference times (3.20 ms, 1.35 ms, and 0.99 ms), measured on a single NVIDIA TITAN X GPU. While MambaHSISR offers a favorable balance between performance and efficiency compared to heavier models, it remains noticeably more demanding than our approach. In contrast, RCAN, EDSR, and SSPSR exhibit substantially higher computational costs, confirming the efficiency advantage of lightweight designs.

Figures 5 and 6 provide qualitative evidence consistent with the numerical results. ESSAFormer yields the sharpest spatial details and the cleanest error maps. Our model produces visually competitive reconstructions with uniformly low spectral error, though slightly less sharp along high-frequency boundaries. Traditional CNN-based baselines show blurred structures and pronounced residual artifacts.

An additional observation concerns the preservation of relative band intensities. As visible in Fig. 5, the ground truth

Table 2 Quantitative comparison on the StereoMSI dataset at $4\times$, $6\times$ and $8\times$ super-resolution scales

Scale	Method	PSNR↑	SSIM↑	SAM↓	GFLOPs↓	Runtime(ms)↓
4×	Bicubic	32.95	0.882	0.042	N/A	N/A
	RCAN	35.40	0.909	0.038	198.5	392.0
	EDSR	36.00	0.915	0.036	645.7	260.2
	SSPSR	35.85	0.920	0.034	775.9	180.0
	ESSAFormer	36.30	**0.934**	0.029	80.50	37.60
	Bi-3DQRNN	35.60	0.923	0.032	95.30	41.20
	MambaHSISR	**36.45**	0.930	0.028	11.20	15.80
	Ours	36.28	0.928	**0.025**	**6.80**	**3.20**
6×	Bicubic	29.80	0.812	0.054	N/A	N/A
	RCAN	32.40	0.865	0.048	88.70	198.4
	EDSR	32.62	0.864	0.045	355.0	149.1
	SSPSR	32.58	0.870	0.042	628.0	101.5
	ESSAFormer	33.10	**0.878**	0.039	39.65	18.90
	Bi-3DQRNN	32.75	0.873	0.041	48.20	21.80
	MambaHSISR	33.12	0.874	0.038	5.60	8.10
	Ours	**33.13**	0.876	**0.035**	**3.10**	**1.35**
8×	Bicubic	25.80	0.720	0.072	N/A	N/A
	RCAN	28.20	0.790	0.064	88.70	198.4
	EDSR	28.45	0.795	0.061	355.0	149.1
	SSPSR	28.60	0.802	0.058	628.0	101.5
	ESSAFormer	28.99	**0.815**	0.054	39.65	18.90
	Bi-3DQRNN	28.70	0.808	0.056	48.20	21.80
	MambaHSISR	**29.06**	0.813	0.050	3.80	4.53
	Ours	29.05	0.812	**0.049**	**1.02**	**0.99**

Metrics include PSNR, SSIM, SAM, computational complexity (GFLOPs), and runtime. Best results are shown in bold, and second-best results are underlined

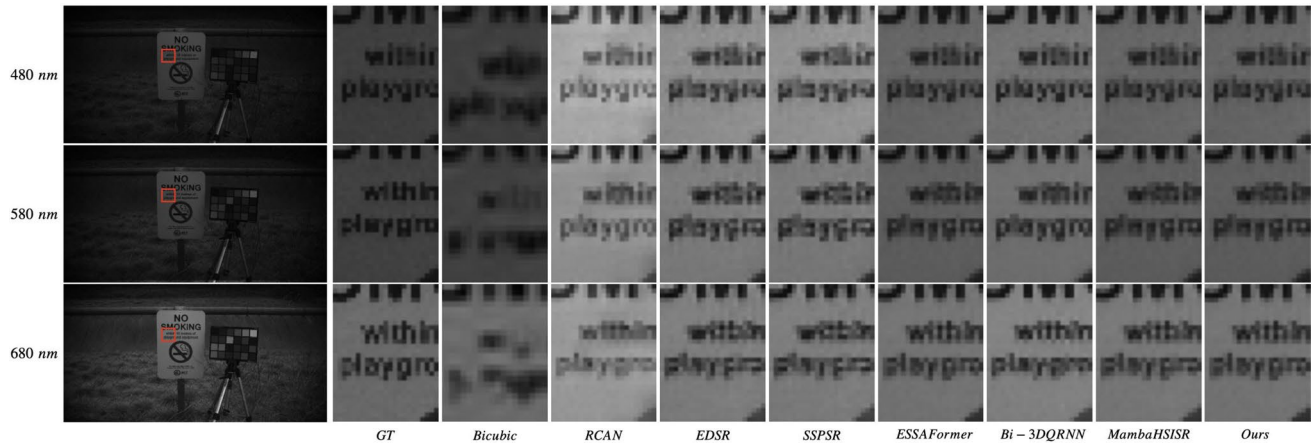

Fig. 5 Visual comparison of reconstructed hyperspectral images from StereoMSI at wavelengths 477 nm, 537 nm, and 617 nm, under × 4 super-resolution. Rows correspond to spectral bands; columns to different methods. All images are displayed using the same grayscale range to allow consistent visual comparison across methods

exhibits a clear brightness pattern across the three bands, which Bicubic interpolation preserves and our method reconstructs consistently. Several CNN-based approaches, while generating sharper details, tend to alter these relative intensity relationships, whereas our method maintains them more consistently.

StereoMSI results highlight the strengths of our method in spectral accuracy and computational efficiency.

Overall, while some competing methods achieve slightly higher spatial reconstruction accuracy, they do so at a significantly higher computational cost. In contrast, the proposed method provides a more favorable trade-off between reconstruction quality and efficiency. It maintains competitive PSNR and SSIM values, achieves the best spectral fidelity, and requires substantially fewer GFLOPs and lower inference time.

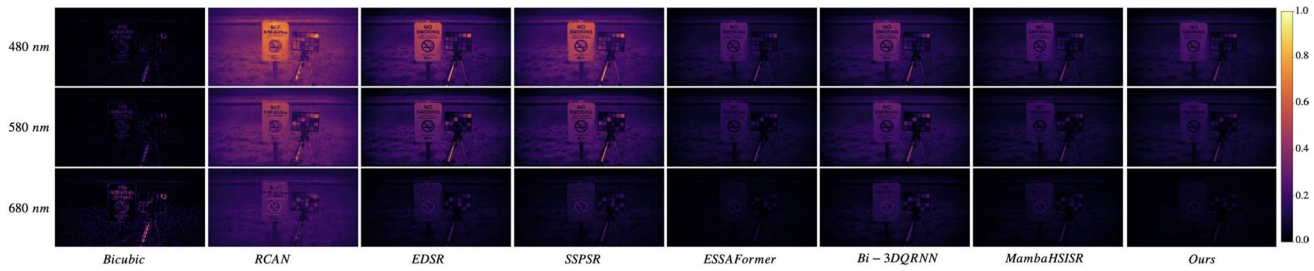

Fig. 6 Error maps from StereoMSI at wavelengths 477 nm, 537 nm, and 617 nm, under $\times 4$ super-resolution. Rows correspond to spectral bands; columns to different methods. Brighter regions represent higher reconstruction errors, highlighting differences in spatial accuracy among the models

Table 3 Ablation study on the impact of the spatial convolution design at $4\times$ super-resolution

Conv3XC	PSNR↑	SSIM↑	SAM↓	Params(M)↓	GFLOPs(G)↓	Runtime(ms)↓
×	36.92	0.930	0.023	**0.98**	9.56	4.50
✓	**38.62**	**0.941**	**0.020**	2.84	9.56	4.50

Conv3XC is compared with a standard 3×3 Conv2D block. ✓ indicates the use of Conv3XC, while × denotes the standard Conv2D

This makes the proposed approach a more practical choice for real-world hyperspectral imaging applications, where both reconstruction quality and computational efficiency are critical.

5.3 Ablation studies

In this section, we conduct a series of ablation experiments to analyze the contribution of the key components in the proposed architecture. All experiments are performed at $4\times$ super-resolution for simplicity, as this setting provides a representative evaluation of the model behavior while keeping the computational cost manageable. Preliminary experiments at other scales exhibit consistent trends.

We focus on three main aspects: (i) the impact of the Conv3XC module as a global replacement for standard convolutions, (ii) the role of the Spectral Attention Block (SAB) in modeling spectral dependencies, and (iii) the effect of the number of Spatial-Spectral Blocks (SSBs) on the trade-off between reconstruction performance and computational efficiency.

These ablations are designed to provide a comprehensive understanding of the architectural choices behind the proposed model and to justify the final design configuration.

5.3.1 Impact of global Conv3XC replacement

Table 3 evaluates the impact of replacing all standard 3×3 convolution layers with the proposed Conv3XC module at $4\times$ super-resolution. This experiment compares two architectural variants of the network, rather than a local modification.

The results show that the Conv3XC-based model consistently improves reconstruction quality across all metrics. In particular, PSNR increases from 36.92 to 38.62 dB, and SSIM improves from 0.930 to 0.941, indicating enhanced structural fidelity. At the same time, the spectral distortion is reduced, with SAM decreasing from 0.022 to 0.020, demonstrating improved preservation of spectral information.

These improvements are achieved thanks to the higher representational capacity of Conv3XC, which enables more effective modeling of spatial–spectral correlations. While this architectural modification increases the number of parameters, the computational complexity and inference time remain unchanged due to the re-parameterization strategy.

Overall, the results confirm that Conv3XC is a key component of the proposed architecture, providing a better trade-off between reconstruction performance and efficiency.

Table 4 Ablation study on the impact of the Spectral Attention Block (SAB) at $4\times$ super-resolution. We compare SAB with a standard channel attention mechanism (SE)

Attention	PSNR↑	SSIM↑	SAM↓
None	38.59	0.940	0.031
SE	38.60	0.940	0.026
SAB (ours)	**38.62**	**0.941**	**0.020**

5.3.2 Impact of the spectral attention block on spectral fidelity

Table 4 evaluates the impact of the Spectral Attention Block (SAB) at $4\times$ super-resolution, and compares it with a standard channel attention mechanism (SE) used in [9]. The results show that removing attention entirely leads to only marginal changes in spatial reconstruction metrics, with PSNR and SSIM remaining nearly unchanged. Introducing SE produces a slight improvement, but the overall impact on spatial fidelity remains limited.

Table 5 Ablation study on the number of Spatial-Spectral Blocks (SSBs) at $4\times$ super-resolution. We analyze the trade-off between reconstruction performance and computational efficiency

#SSBs	PSNR↑	SSIM↑	SAM↓	Params(M)↓	GFLOPs(G)↓	Runtime(ms)↓
2	35.95	0.916	0.033	**0.94**	**4.10**	**2.10**
4	36.96	0.929	0.028	1.88	7.80	3.60
5	38.62	0.941	0.020	2.84	9.56	4.50
8	38.67	0.944	0.019	4.55	13.10	6.90
10	**38.70**	**0.946**	**0.018**	5.77	15.90	8.60

In contrast, the impact on spectral accuracy is substantial. The SAM value decreases from 0.031 without attention to 0.026 with SE, and further to 0.020 with SAB, highlighting a progressive improvement in spectral consistency. This demonstrates that while generic channel attention mechanisms can partially model inter-band relationships, the proposed SAB is significantly more effective in capturing spectral dependencies. The inclusion of SAB leads to a substantial reduction in SAM, decreasing from 0.031 to 0.020, which corresponds to an improvement of approximately 35%.

These results confirm that SAB plays a crucial role in preserving spectral information, with its benefits primarily reflected in SAM rather than in spatial reconstruction metrics such as PSNR and SSIM.

This behavior is expected, since SAB is explicitly designed to refine inter-band relationships and preserve spectral consistency rather than to enhance spatial details. Consequently, its contribution is primarily reflected in SAM, which directly measures spectral angular distortion, rather than in PSNR or SSIM.

Overall, these results confirm that the Spectral Attention Block is essential for achieving high spectral fidelity, which is a key requirement in hyperspectral image super-resolution.

5.3.3 Impact of the number of spatial-spectral blocks

Table 5 analyzes the impact of the number of Spatial-Spectral Blocks (SSBs) on reconstruction performance and computational efficiency at $4\times$ super-resolution. As expected, increasing the number of blocks consistently improves PSNR, SSIM, and SAM, confirming that deeper architectures provide stronger representational capacity for modeling spatial–spectral correlations.

In particular, significant performance gains are observed when increasing the number of SSBs from 2 to 5. PSNR improves from 35.95 to 38.62 dB, SSIM increases from 0.916 to 0.941, and SAM decreases from 0.033 to 0.020, highlighting the importance of sufficient network depth for accurate reconstruction.

However, further increasing the number of blocks beyond 5 results in diminishing returns. While configurations with 8 and 10 SSBs achieve slightly better quantitative results, the improvements are marginal compared to the substantial increase in the number of parameters, computational cost, and inference time. For example, moving from 5 to 10 SSBs leads to a significant increase in model complexity with only negligible gains in reconstruction performance.

Therefore, we adopt 5 SSBs in the final architecture, as it provides the best trade-off between reconstruction accuracy and computational efficiency.

6 Conclusion

Overall, ESSPN represents a step forward toward efficient, accurate, and deployable hyperspectral super-resolution.

Despite these promising results, several challenges remain open. First, preserving spectral fidelity under complex degradations remains a critical issue, especially in real-world scenarios where noise, sensor-specific responses, and non-ideal imaging conditions can distort spectral signatures. While ESSPN achieves low SAM values, further improvements are needed to guarantee physically consistent reconstruction across diverse acquisition settings.

Second, cross-sensor generalization remains a key challenge. Although our method demonstrates good generalization across datasets and radiometric domains, differences in spectral response functions, calibration procedures, and noise characteristics across sensors can still limit performance. Future research should investigate domain adaptation strategies and sensor-aware modeling to improve robustness in cross-device scenarios.

Third, the integration of large-scale foundation models represents a promising yet largely unexplored direction in hyperspectral imaging. Recent advances in vision transformers and multimodal pretraining suggest that leveraging large-scale data and learned priors could significantly enhance both spatial reconstruction and spectral consistency. However, adapting such models to hyperspectral data remains challenging due to the high dimensionality and limited availability of labeled datasets.

Finally, practical deployment constraints must be further addressed. Although ESSPN is computationally efficient, real-time and onboard deployment in edge devices (e.g., UAVs or satellites) imposes strict limitations in terms of memory, power consumption, and latency. Future work will focus on model compression, quantization, and

hardware-aware optimization to enable efficient deployment in real-world applications.

In addition, future research will explore extensions of the proposed framework to cross-modal fusion settings, incorporating auxiliary modalities such as RGB or panchromatic images to further enhance reconstruction quality.

Author contributions Conceptualization, M.K. and S.B.; methodology, M.K.; software, M.K.; validation, M.K., M.B. and S.B; formal analysis, M.K.; investigation, M.K.; resources, R.S.; data curation, M.K.; writing—original draft preparation, M.K.; writing—review and editing, M.K., M.B. and S.B.; visualization, M.K. and M.B.; supervision, R.S.; project administration, R.S. All authors have read and agreed to the published version of the manuscript

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement.

Data availability The data presented in this study are openly available in public repositories. Specifically, the ARAD1K dataset is available at https://github.com/boazarad/ARAD_1K, and the StereoMSI dataset is available at https://data.csiro.au/collection/csiro:40743.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Blackburn, G.A.: Hyperspectral remote sensing of plant pigments. J. Exp. Bot. **58**(4), 855–867 (2007)
2. Misbah, K., Laamrani, A., Khechba, K., Dhiba, D., Chehbouni, A.: Multi-sensors remote sensing applications for assessing, monitoring, and mapping npk content in soil and crops in african agricultural land. Remote Sens. **14**(1), 81 (2022)
3. Stuart, M.B., McGonigle, A.J.S., Willmott, J.R.: Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems. Sensors **19**(14), 3071 (2019)
4. Bjorgan, A., Randeberg, L.L.: Towards real-time medical diagnostics using hyperspectral imaging technology. In: European Conference on Biomedical Optics, p. 953712 (2015). Optica Publishing Group
5. Laparrcr, V., Santos-Rodríguez, R.: Spatial/spectral information trade-off in hyperspectral images. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1124–1127. IEEE (2015)
6. Fu, Y., Zhang, T., Zheng, Y., Zhang, D., Huang, H.: Hyperspectral image super-resolution with optimized rgb guidance. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11661–11670 (2019)
7. Zhou, Y., Rangarajan, A., Gader, P.D.: An integrated approach to registration and fusion of hyperspectral and multispectral images. IEEE Trans. Geosci. Remote Sens. **58**(5), 3020–3033 (2019)
8. Li, J., Liu, J., Tong, X., Gao, J., Shen, H.: Hyperspectral image super-resolution with grouped deep recursive residual network. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6 (2018)
9. Jiang, J., Sun, H., Liu, X., Ma, J.: Learning spatial-spectral prior for super-resolution of hyperspectral imagery. IEEE Trans. Comput. Imaging **6**, 1082–1096 (2020)
10. Yuan, Q., Zhang, L., Shen, H.: Hyperspectral image superresolution by transfer learning. IEEE J. Select. Topics Appl. Earth Observ. Remote Sens. **11**(5), 1234–1245 (2018)
11. Xie, Q., Zhang, L., Yuan, Q., Shen, H.: Hyperspectral image super-resolution with deep feature consistent network. Remote Sens. **11**(15), 1816 (2019)
12. Huang, B., Song, H., Cui, H., Peng, J., Xu, X.: Spatial and spectral image fusion using sparse matrix factorization. IEEE Trans. Geosci. Remote Sens. **53**(12), 6678–6690 (2015)
13. Mei, S., Huang, Y., Wang, Z., Ma, L., Fan, X., Zhang, L.: Hyperspectral image spatial super-resolution via 3d full convolutional neural network. In: Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 2344–2348. (2017)
14. Hong, D., Li, C., Yokoya, N., Zhang, B., Jia, X., Plaza, A., Gamba, P., Benediktsson, J.A., Chanussot, J.: Hyperspectral imaging. Nat. Rev. Methods Primers **6**(1), 19 (2026)
15. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 184–199. (2014)
16. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1646–1654 (2016)
17. Kim, J., Lee, J.K., Lee, K.M.: Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1637–1645 (2016)
18. Lai, W., Huang, J., Ahuja, N., Yang, M.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5835–5843. (2017)
19. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3147–3155. (2017)
20. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 136–144. (2017)
21. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2472–2481. (2018)
22. Haris, M., Shakhnarovich, G., Ukita, N.: Deep back-projection networks for super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1664–1673. (2018)
23. Zhang, K., Zuo, W., Zhang, L.: Deep plug-and-play super-resolution for arbitrary blur kernels. In: Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1671–1681 (2019)
24. Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 286–301 (2018)
25. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132–7141. (2018)
26. Dai, T., Cai, J., Zhang, Y., Xia, S., Zhang, L.: Second-order attention network for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11065–11074 (2019)
27. Akgun, T., Altunbasak, Y., Mersereau, R.M.: Super-resolution reconstruction of hyperspectral images. IEEE Trans. Image Process. **14**(11), 1860–1875 (2005)
28. Stark, H., Yang, Y., Yang, Y.: Vector Space Projections: a Numerical Approach to Signal and Image Processing, Neural Nets, and Optics. John Wiley & Sons Inc, New York, NY, USA (1998)
29. Fu, Y., Liang, Z., You, S.: Bidirectional 3d quasi-recurrent neural network for hyperspectral image super-resolution. IEEE J. Select. Topics Appl. Earth Observ. Remote Sens. **14**, 2674–2688 (2021). https://doi.org/10.1109/JSTARS.2021.3057936
30. Sidorov, K., Dai, Y., Porikli, F.: Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6945–6954. (2019)
31. Zhang, M., Zhang, C., Zhang, Q., Guo, J., Gao, X., Zhang, J.: ESSAformer: Efficient Transformer for Hyperspectral Image Super-resolution (2023). https://arxiv.org/abs/2307.14010
32. Xu, Y., Wang, H., Zhou, F., Luo, C., Sun, X., Rahardja, S., Ren, P.: Mambahsisr: Mamba hyperspectral image super-resolution. IEEE Trans. Geosci. Remote Sens. **63**, 1–16 (2025). https://doi.org/10.1109/TGRS.2025.3560632
33. Li, C., Hong, D., Zhang, B., Pan, Z., Yokoya, N., Chanussot, J.: Mambax: Image super-resolution with state predictive control, (2025). arXiv:2511.18028 arXiv preprint
34. Li, C., Hong, D., Zhang, B., Pan, Z., Chanussot, J.: Kano: Kolmogorov-arnold neural operator for image super-resolution, (2025). arXiv preprint arXiv:2512.22822
35. Wan, C., Yu, H., Li, Z., Chen, Y., Zou, Y., Liu, Y., Yin, X., Zuo, K.: Swift parameter-free attention network for efficient super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 6246–6256 (2024)
36. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1874–1883. (2016)
37. Arad, B., Timofte, R., Yahel, R., Morag, N., Bernat, A., Cai, Y., Lin, J., Lin, Z., Wang, H., Zhang, Y., Pfister, H., Gool, L.V., Liu, S., Li, Y., Feng, C., Lei, L., Li, J., Du, S., Wu, C., Leng, Y., Song, R., Zhang, M., Song, C., Zhao, S., Lang, Z., Wei, W., Zhang, L., Dian, R., Shan, T., Guo, A., Feng, C., Liu, J., Agarla, M., Bianco, S., Buzzelli, M., Celona, L., Schettini, R., He, J., Xiao, Y., Xiao, J., Yuan, Q., Li, J., Zhang, L., Kwon, T., Ryu, D., Bae, H., Yang, H.-H., Chang, H.-E., Huang, Z.-K., Chen, W.-T., Kuo, S.-Y., Chen, J., Li, H., Liu, S., Sabarinathan, S., Uma, K., Bama, B.S., Roomi, S.M.M.: Ntire 2022 spectral recovery challenge and data set. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 862–880 (2022)
38. Shoeiby, M., Robles-Kelly, A., Wei, R., Timofte, R.: Pirm 2018 challenge on spectral image super-resolution: Dataset and study, (2019). arXiv preprint arXiv:1904.00540
39. Yuhas, R.H., Goetz, A.F., Boardman, J.W.: Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In: JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop, Volume 1: AVIRIS Workshop, pp. 147–149 (1992)
40. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

Matteo Kolyszko received his BSc degree in Computer Science from the University of Milano-Bicocca, Italy, in 2021, and his MSc degree in Data Science from the same university in 2023. He is currently pursuing a PhD in Artificial Intelligence at the University of Milano-Bicocca. His research interests include hyperspectral image processing, spatial and spectral super-resolution of hyperspectral images, and computational photography.

Marco Buzzelli received his PhD in Computer Science from the University of Milano-Bicocca, Italy, in 2019, where he is currently an Assistant Professor. His research interests include signal, image, and video processing and understanding, machine learning, and color imaging. He has collaborated with several European institutions and is a member of ELLIS.

Simone Bianco received his PhD in Computer Science from the University of Milano-Bicocca, Italy, in 2010, after obtaining his BSc and MSc degrees in Mathematics from the same university. He is currently an Associate Professor at the University of Milano-Bicocca. His research interests include computer vision, artificial intelligence, machine learning, optimization algorithms, and multimodal and multimedia applications. He is a member of ELLIS.

Raimondo Schettini is a Full Professor at the University of Milano-Bicocca, Italy, where he leads the Imaging and Vision Lab. His research interests include color imaging, image processing and analysis, pattern recognition, image and video understanding, and retrieval. He has authored more than 400 refereed papers and holds 12 patents. He has coordinated several national and industrial research projects and supervised more than 10 PhD students. He is Editor-in-Chief of the MDPI Journal of Imaging, Fellow of IAPR and AAIA, and a member of ELLIS.