Erciyes University Journal of Institute Of Science and Technology

Research Article
10.65520/erciyesfen.1971600

Imprint:
Volume: 42(3)
Year: 2026
Page: 353-373

Huseyin Aydilek[a*]
Nihat İnanç[b]

[a] Asst. Prof. Dr., Kırıkkale University, huseyinaydilek@kku.edu.tr
[b] Prof. Dr., Halic University, nihatinanc@halic.edu.tr

* Corresponding Author

Received: 6/15/2026
Accepted: 8/15/2026

Citation:
Huseyin Aydilek, Nihat İnanç (2026). Hyperspectral Image Super-Resolution Using a Hybrid Method Combining Deep Learning and Sparse Representation. *Erciyes University Journal of Institute Of Science and Technology*, 42(3), 353-373. https://doi.org/10.65520/erciyesfen.*1971600*

Screened by iThenticate for Authors & Researchers

Hyperspectral Image Super-Resolution Using a Hybrid Method Combining Deep Learning and Sparse Representation

Abstract

Background/Aim: Hyperspectral images provide rich spectral information but often suffer from low spatial resolution due to hardware limitations and long imaging distances. This study aims to enhance the spatial resolution of low-resolution hyperspectral images while preserving their spectral information by proposing a hybrid super-resolution method that combines deep learning and sparse representation.

Methods: The proposed framework comprises two components. Spatial resolution is enhanced using deep learning models, while the spectral information of the original low-resolution hyperspectral cube is reintroduced into the reconstructed image through Bayesian sparse-representation-based dictionary learning. A deep convolutional network and a deep residual network were adapted to hyperspectral images. To reduce the convolutional workload, two-dimensional convolution was applied to one band selected from every three bands. The method was evaluated on the CAVE and Harvard datasets at scale factors of ×2, ×3, and ×4 using root mean square error (RMSE) and spectral angle mapper (SAM), and was compared with bicubic interpolation, GFPCA, and SCT_SDCNN.

Results: At the dataset-average level, the deep residual network-based hybrid model yielded the lowest RMSE and SAM values among the evaluated methods at all three scale factors. On CAVE, the RMSE/SAM values were 2.0146/2.8507, 2.1380/4.0742, and 2.4712/5.1211 at ×2, ×3, and ×4, respectively. The corresponding values on Harvard were 1.5753/2.1328, 1.7373/2.7135, and 2.1524/3.3576. Relative to SCT_SDCNN, the SAM reductions ranged from 19.12% to 24.80% on CAVE and from 21.93% to 26.02% on Harvard.

Conclusion: The findings suggest that combining deep-learning-based spatial enhancement with sparse-representation-based spectral fusion can support lower spatial and spectral reconstruction errors under the evaluated conditions. The need for training data with characteristics comparable to the target data remains a limitation.

Keywords: Hyperspectral imaging, Super-resolution, Deep learning, Sparse representation, Image fusion

Derin Öğrenme ve Seyrek Temsil Tabanlı Hibrit Bir Yöntem ile Hiperspektral Görüntü Süper Çözünürlüğü

Öz

Hiperspektral görüntüler, çok sayıda spektral bant içermeleri sayesinde nesnelere ait ayrıntılı bilgi sağlar; ancak görüntüleme sistemlerinin donanımsal kısıtları ve uzun görüntüleme mesafesi nedeniyle uzamsal çözünürlükleri çoğunlukla düşük kalmaktadır. Bu çalışmada, düşük uzamsal çözünürlüklü hiperspektral görüntülerin çözünürlüğünü artırmak amacıyla derin öğrenme ile seyrek temsili birleştiren hibrit bir süper çözünürlük yöntemi incelenmiştir.

Except where otherwise noted, content in this article is licensed under a Creative Commons 4.0 International license. Icons by Font Awesome.

Önerilen yapıda uzamsal çözünürlük derin öğrenme modelleriyle iyileştirilirken, özgün düşük çözünürlüklü hiperspektral kübteki spektral bilginin yeniden oluşturulan görüntüye aktarılması Bayesçi seyrek temsil tabanlı sözlük öğrenme ile desteklenmiştir. Hesaplama yükünü azaltmak amacıyla iki boyutlu evrişim, her üç banttan seçilen bir banda uygulanmıştır. Derin evrişimsel ağ ve derin kalıntı ağ olmak üzere iki model, CAVE ve Harvard veri kümelerinde ×2, ×3 ve ×4 ölçek faktörlerinde RMSE ve SAM ölçütleri kullanılarak değerlendirilmiştir. Veri kümesi ortalamalarına göre derin kalıntı ağ tabanlı hibrit model, değerlendirilen yöntemler arasında tüm ölçek faktörlerinde en düşük RMSE ve SAM değerlerini vermiştir. CAVE veri kümesinde RMSE/SAM değerleri ×2, ×3 ve ×4 için sırasıyla 2,0146/2,8507, 2,1380/4,0742 ve 2,4712/5,1211; Harvard veri kümesinde ise 1,5753/2,1328, 1,7373/2,7135 ve 2,1524/3,3576 olarak elde edilmiştir. SCT_SDCNN ile karşılaştırıldığında SAM değerlerindeki azalma CAVE için %19,12-%24,80, Harvard için %21,93-%26,02 aralığındadır. Bulgular, derin öğrenme tabanlı uzamsal iyileştirme ile seyrek temsil tabanlı spektral füzyonun birlikte kullanılmasının, değerlendirilen koşullarda uzamsal ve spektral hata değerlerinin azaltılmasına katkı sağlayabileceğini göstermektedir. Yöntemin hedef veriye benzer özelliklerde eğitim verisine ihtiyaç duyması temel sınırlılıklardan biridir.

Anahtar kelimeler: Hiperspektral görüntü, Süper çözünürlük, Derin öğrenme, Seyrek temsil, Görüntü füzyonu

1. INTRODUCTION

Conventional imaging devices mainly record light in the visible part of the electromagnetic spectrum, whereas hyperspectral sensors sample reflected energy across many adjacent wavelengths that may extend from the ultraviolet to the infrared [1]. The resulting set of spectral bands provides information about the material and structural properties of the observed scene. The reflectance curve measured across wavelengths is referred to as a spectral signature [2]. Since absorption, reflection, and transmission vary with the physical and chemical properties of materials, spectral signatures can differ among objects and materials. By combining imaging and spectroscopy, hyperspectral imaging represents both spatial and spectral information [3]. This capability has supported its use in detection, analysis, and classification tasks in areas including military applications, mining, medicine, pharmaceutical sciences, agriculture, and forestry [4-6].

Reliable interpretation of hyperspectral data generally requires an appropriate balance between spectral and spatial resolution. In practice, sensor limitations and long imaging distances can restrict spatial resolution. At low spatial resolution, the spectra of neighbouring pixels may become mixed, a phenomenon referred to as spectral mixing [7]. This may produce spectral observations influenced by multiple materials within the same region and may reduce the reliability of subsequent analyses. Hardware-based improvements can be costly and technically demanding; consequently, software-based resolution enhancement has become an active area of image-processing research.

Super resolution denotes the recovery of a high-resolution image by software means, using various image-processing algorithms and without upgrading the imaging hardware. Approaches in the literature can generally be grouped into two broad categories. The first category consists of image fusion-based techniques, whereas the second enhances resolution from a single image.

Image-fusion methods initially developed from spectral-unmixing and statistical-estimation approaches [8-10]. Subsequent studies considered Bayesian formulations [11,12], principal-component and wavelet techniques [13], spectral-decomposition methods [14,15], tensor-based schemes [16,17], sparse-representation methods [18-20], matrix factorization, and deep-learning-based fusion [21-24]. Component-based methods such as GFPCA have also been used for comparison [25]. Sparse representation is one of the model-based approaches investigated for hyperspectral fusion [18,26]. Conventional fusion formulations generally require a second observation of the same scene at higher spatial resolution and require a relationship to be established between the spatial and spectral inputs. Recent work in this direction has moved towards transformer-based fusion networks and general fusion frameworks guided by a high-resolution reference [27,28]. This requirement is structural rather than incidental: reviews of the field have repeatedly identified it, together with the sensitivity of the estimated spectral response to registration error, as the practical constraint that

CC: BY-NC-ND

limits the deployment of fusion-based super-resolution [29,30]. Where only a single hyperspectral acquisition exists, the family is therefore not directly applicable.

Single-image super-resolution studies commonly use interpolation, statistical-analysis methods [31], and learning-based methods. Interpolation can provide limited reconstruction detail for hyperspectral data, while statistical methods often require assumptions or prior information about the data distribution. Learning-based approaches have therefore received increasing attention for modelling the relationship between low- and high-resolution images [32-37]. Methods first developed for natural RGB or grayscale images have also been adapted to hyperspectral data, including architectures based on two-dimensional and three-dimensional convolutions [38-45]. Two-dimensional convolutions mainly model spatial information and may not directly represent inter-band relationships, whereas three-dimensional convolutions model spatial and spectral dimensions jointly but increase computational and memory requirements. Processing all bands together may also introduce spectral mixing in some network designs [42,43,46]. Three developments have shaped the field since then. First, attention and transformer architectures have been introduced to model long-range dependencies across bands and across space, from interactive transformer-CNN designs [47] through multiscale deformable attention [48] and efficiency-oriented self-attention approximations [49] to cross-scope aggregation and explicit spectral-prior modelling [50,51], with more recent variants replacing quadratic attention by linear recurrent formulations [52]. Second, generative and state-space formulations have appeared, including conditional diffusion models [53] and selective state-space backbones [54], alongside strategies that reuse pre-trained RGB models through eigenimage decomposition [55]; hybrid state-space and convolutional architectures have since been proposed for the same purpose [56], and the rapid growth of the diffusion-based branch has itself become the subject of review [57]. Third, and in response to the practical fragility of the preceding two groups, attention has turned to blind degradation [58], meta-transfer across sensors [59], and test-time adaptation [60].

Read together, these studies indicate a consistent trade-off rather than a monotonic improvement. Reported accuracy gains have been obtained alongside increasing parameter counts, increasing dependence on the volume and representativeness of the training data, and, in the generative case, substantial inference cost. Spectral consistency in all of these designs is enforced implicitly, through the loss function and the architecture, so that the fidelity of the reconstructed spectral signature is a property learned from the training distribution rather than one guaranteed by construction. Surveys of the area have noted that this implicit treatment of the spectral dimension, and the limited transferability that follows from it, remain open problems [29,61].

A third line of work responds to that observation by retaining an explicit physical or statistical model inside a learning-based pipeline. Model-guided and unfolding networks embed the degradation model into trainable layers [62-64], a direction extended most recently by cross-domain-aware unfolding transformers [65]; interpretable and lightweight formulations pursue the same objective under an explicit stability constraint [66]; and hierarchical schemes fuse spatial and spectral information through sparse representation at successive scales [67]. Sparse representation continues to be used as a spectral prior in current remote-sensing super-resolution work [68], and Bayesian formulations remain in use where spectral variability must be modelled explicitly [69]. Two-stage designs that pair a learned spatial stage with an explicit spectral component continue to appear [70], as do fully unsupervised deep-prior formulations that dispense with paired training data altogether [71]. These studies establish that the model-assisted direction is active rather than superseded. Two characteristics, however, recur across them. Most still presuppose an auxiliary high-resolution observation, which returns the difficulty identified above; and where the model is unrolled into trainable layers, the model stage itself becomes a learned component, so that the training-data dependency the explicit model was intended to relax is partly reintroduced.

This study investigates a hybrid method that combines the complementary roles of deep learning and sparse representation within a unified framework. The approach uses deep learning to model the relationship between low- and high-spatial-resolution images, while sparse representation is employed to support the transfer of spectral information from the original hyperspectral image to

CC: BY-NC-ND

the reconstructed image.

A preliminary version of the convolutional network-based hybrid framework was presented in our earlier study [72]. Building upon that work, the present study extends the framework by incorporating a deep residual network architecture, evaluating the method on both the CAVE and Harvard datasets, and providing a more comprehensive experimental analysis using multiple scale factors and quantitative performance metrics.

The configuration that has received comparatively little systematic attention is the simplest one in this family: a strictly two-stage, non-unfolded pipeline in which a compact single-image network is responsible only for spatial enhancement, and a fully unsupervised Bayesian sparse-representation stage restores the spectral content from the observed low-resolution cube itself, with no auxiliary sensor and no training signal in the second stage. Such a configuration is of interest precisely because it separates the two roles: whatever spectral error the network introduces is addressed by a stage whose behaviour does not depend on the training set, and the cost of the network can be reduced independently of the spectral component. The present study examines this configuration and characterises the trade-off it entails.

The aim is deliberately narrower than a claim of state-of-the-art accuracy. Current transformer-, diffusion- and state-space-based methods [49,53,54] operate under different parameter and training budgets, and a like-for-like accuracy comparison with them would require a separate experimental protocol. What is examined here is whether an explicit, unsupervised spectral stage can compensate for spectral subsampling in a deliberately compact convolutional stage, and at what computational cost. Table 1 summarises the position of the proposed framework with respect to the principal methodological families in the literature. The contributions of this study are as follows.

(i) A two-stage framework is formulated in which spatial enhancement and spectral restoration are assigned to structurally distinct components, so that the spectral stage remains unsupervised and independent of the training distribution.

(ii) A band-selection strategy is introduced in which two-dimensional convolution is applied to one band in every three, reducing the convolutional workload by approximately 64 per cent, with the bands omitted from the convolutional stage remaining available to the fusion stage.

(iii) Two network architectures, a three-layer convolutional model and a densely connected residual model, are evaluated within the identical fusion pipeline, isolating the effect of the spatial stage on the final reconstruction.

(iv) The framework is assessed on two datasets with different acquisition characteristics, at three scale factors, under a single degradation model applied uniformly to all compared methods, and the resulting accuracy-cost trade-off is reported together with its boundaries.

CC: BY-NC-ND

Table 1. Positioning of the proposed framework with respect to the principal methodological families in hyperspectral super-resolution.

Methodological family	Representative recent work	Principal requirement	Relation to this study
Fusion of a hyperspectral image with an auxiliary high-resolution multispectral or RGB observation	Fusformer [27]; general high-resolution guidance frameworks [28]	A second, co-registered observation of the same scene at higher spatial resolution	Not applicable in the present setting, where only a single low-resolution cube is observed
End-to-end single-image networks with spatial-spectral priors	Spatial-spectral prior learning [40]; 2D/3D convolution studies [42,43]	Paired training data; spectral consistency learned implicitly from the loss	Shares the single-image setting; differs in that spectral consistency is imposed by an explicit model rather than learned
Transformer- and attention-based single-image architectures	Interactformer [47]; MSDformer [48]; ESSAformer [49]; cross-scope aggregation [50]; spectral prior [51]	Large parameter budgets and extensive training data; long-range attention across bands	Complementary rather than competing; the present study examines a compact configuration under the same single-image constraint
Generative and state-space formulations	Conditional diffusion [53]; selective state-space backbones [54]; pre-trained RGB reuse [55]	Iterative sampling or pre-trained backbones; substantial inference cost	Shares the position that iterative inference is acceptable when it buys spectral fidelity
Model-assisted and unfolding-based hybrids	Model-guided networks [62]; disentangled unfolding [63]; knowledge-driven unrolling [64]; interpretable lightweight formulation [66]; hierarchical sparse fusion [67]	The physical model is unrolled into trainable layers, so the model stage is itself learned	Closest family. The present framework keeps the two stages separate and leaves the spectral stage entirely unsupervised
Blind, transfer and test-time formulations	Blind degradation estimation [58]; meta-transfer learning [59]; test-time training [60]	Adaptation machinery for unknown degradation or unseen sensors	Addresses the training-data dependency identified as a limitation of the present framework

2. MATERIALS AND METHODS

2.1. General Structure of the Proposed Hybrid Method

The proposed approach is made up of two principal stages. In the first stage, a deep network model increases the spatial resolution of the low-spatial-resolution hyperspectral image. In the second stage, the spectral content of the original low-resolution image is transferred to the spatially enhanced

CC: BY-NC-ND

representation by means of Bayesian sparse-representation fusion. In this way, spatial enhancement is carried out through deep learning, while spectral preservation is supported through sparse representation.

In the image fed into the deep network model, spatial and spectral data are separated, and each feature is evaluated independently. To reduce the computational burden, the spectral data to be used are sampled by selecting one band from every three spectral bands with close wavelengths. Figure 1 presents the overall algorithmic flow of the proposed hybrid approach.

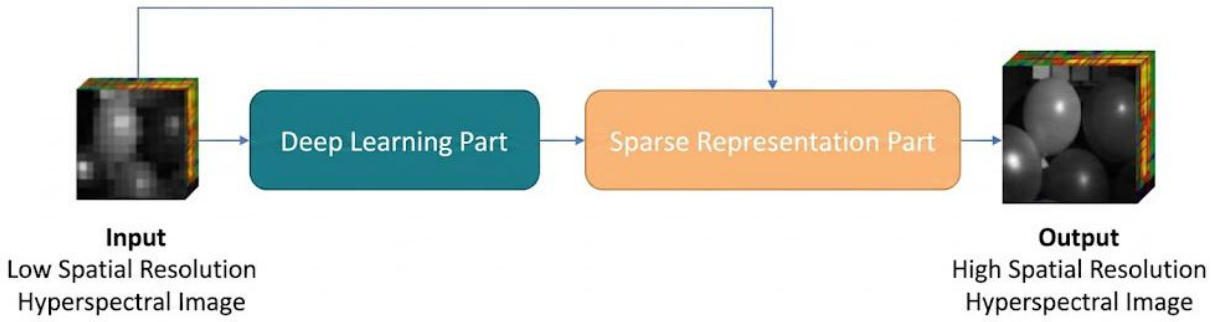

Figure 1. Algorithmic representation of the proposed hybrid method.

2.2. Deep Learning Models

Hyperspectral images have a data structure that differs from natural RGB and grayscale images; therefore, super-resolution methods designed for natural images cannot always be transferred directly without adaptation. In this study, two architectures previously applied to natural-image super-resolution were adapted to hyperspectral images: a convolutional network-based model and a deep residual network-based model.

2.2.1. Deep Convolutional Network Model

A single-image super-resolution technique resting on a simple convolutional neural network architecture [73] was taken up and adapted to hyperspectral images. Once the spectral and spatial information has been separated from the low-resolution image, every band is rescaled by bicubic interpolation according to the chosen scaling factor. The model consists of three convolutional layers.

The first layer is responsible for extracting image patches and feature matrices.

$$F_1(Y) = \max(0,\ W_1 * Y + B_1) \tag{1}$$

In this expression, Y denotes the low-resolution image resized by bicubic interpolation, $F_1(Y)$ denotes the ReLU activated output of the first layer, W_1 denotes the convolution filter, and B_1 denotes the bias term. W_1 consists of n_1 filters.

The second layer generates high resolution features by performing nonlinear mapping.

$$F_2(Y) = \max(0,\ W_2 * F_1(Y) + B_2) \tag{2}$$

To form the final image, the third layer reconstructs the resulting high-resolution patches.

$$F(Y) = W_3 * F_2(Y) + B_3 \tag{3}$$

Figure 2 shows the architecture of the proposed model based on the deep convolutional network.

CC: BY-NC-ND

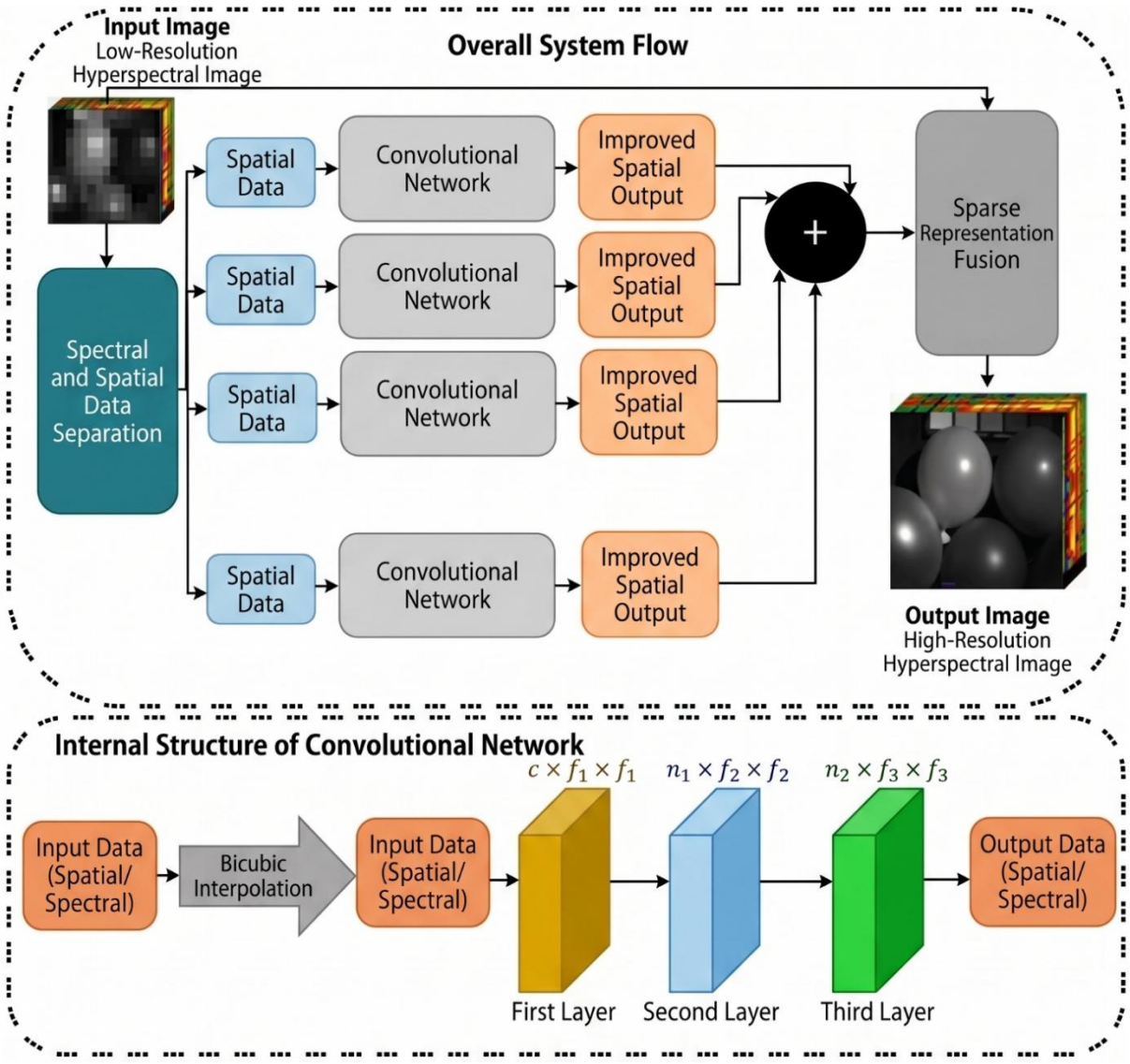

Figure 2. Representation of the proposed method using the deep convolutional network.

2.2.2. Deep Residual Network Model

The approach that combines dense skip connected networks [74] with deep residual networks [75] to incorporate the features lost during successive convolution operations into the reconstruction process [76] was adapted to hyperspectral images. Its underlying idea is to drive the main network through both local and global residual connections. Figure 3 shows the architecture of the proposed model based on the deep residual network.

CC: BY-NC-ND

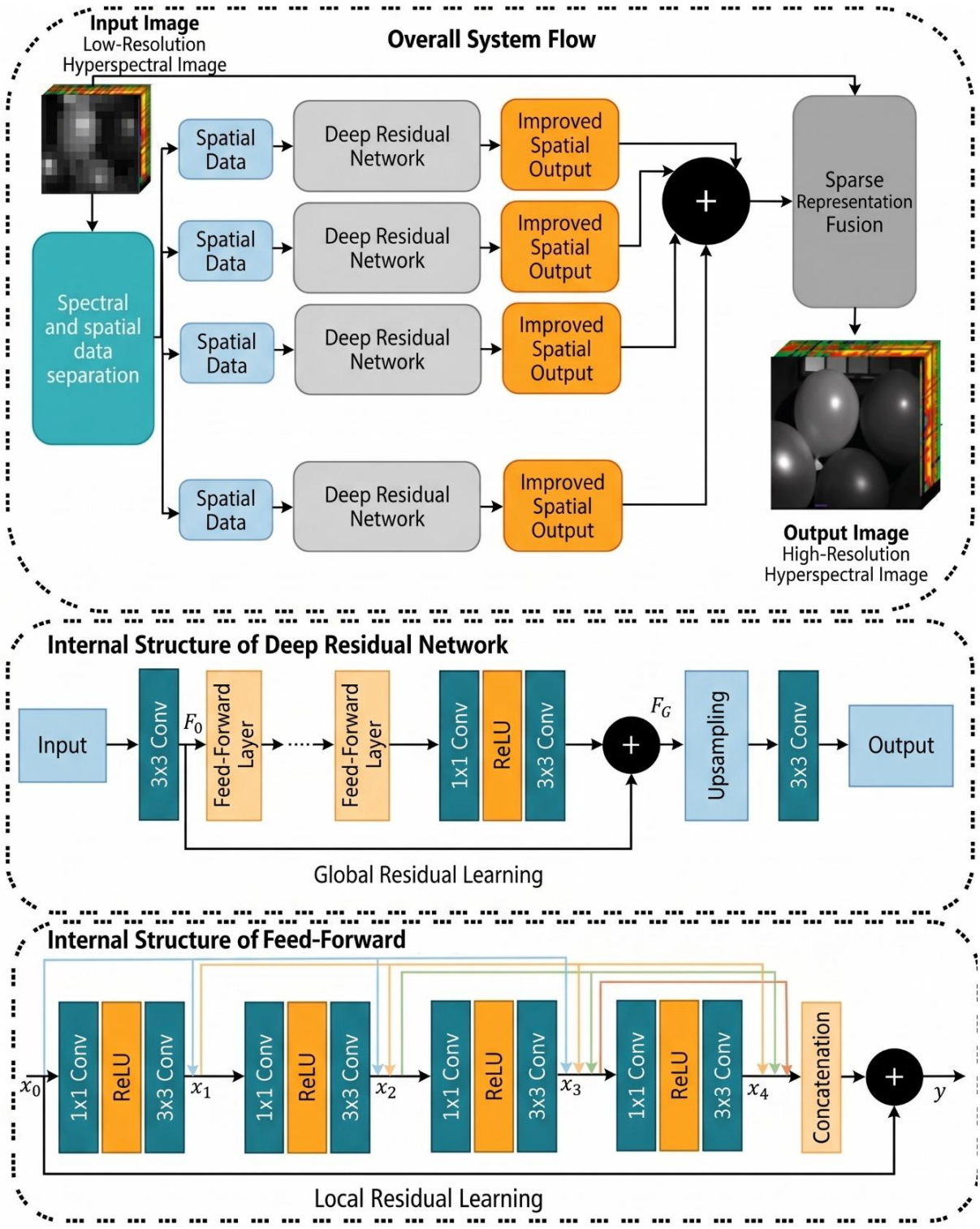

Figure 3. Depiction of the proposed model based on the deep residual network.

In densely skip-connected networks, layer i is fed the feature matrices of all the layers preceding it.

$$x_i = H_i([x_0, x_1, \dots, x_{i-1}]) \quad (4)$$

In this expression, x_i is the output of the i-th layer, $[x_0, x_1, \dots, x_{i-1}]$ denotes the concatenation of

CC: BY-NC-ND

the feature matrices produced by the preceding layers, and H_i is the composite function comprising operations such as convolution and ReLU. The output of the feeding block with the local residual connection is expressed as follows.

$$y = x_0 + [x_1, x_2, \dots, x_n] \quad (5)$$

Residual learning is performed through a skip connection that transfers the input features of one layer to another layer. This structure can help the network model the relationship between low- and high-resolution information, reduce the effect of vanishing gradients, and support reconstruction performance. Since the feed-forward blocks receive a large number of input channels, a bottleneck layer was employed to limit model complexity and the risk of overfitting. To this end, a 1×1 convolution layer was positioned ahead of the 3×3 convolution layer. Global residual learning is computed as follows.

$$F_G = F_0 + H_F(H_D(F_0)) \quad (6)$$

Here, F_0 denotes the result of the convolution applied for feature extraction, H_D denotes the output of the feeding blocks, and H_F denotes the reconstruction operation.

2.3. Sparse Representation Based Fusion

The spatially enhanced band outputs produced by the deep-learning models constituted the high-spatial-resolution input to the fusion stage, while the original low-spatial-resolution hyperspectral cube supplied the complete spectral information. These two inputs were combined using the four-stage Bayesian sparse-representation method [18], which is based on sparse-representation theory [77] and has previously been used for hyperspectral image fusion. Figure 4 illustrates the basic scheme of this method.

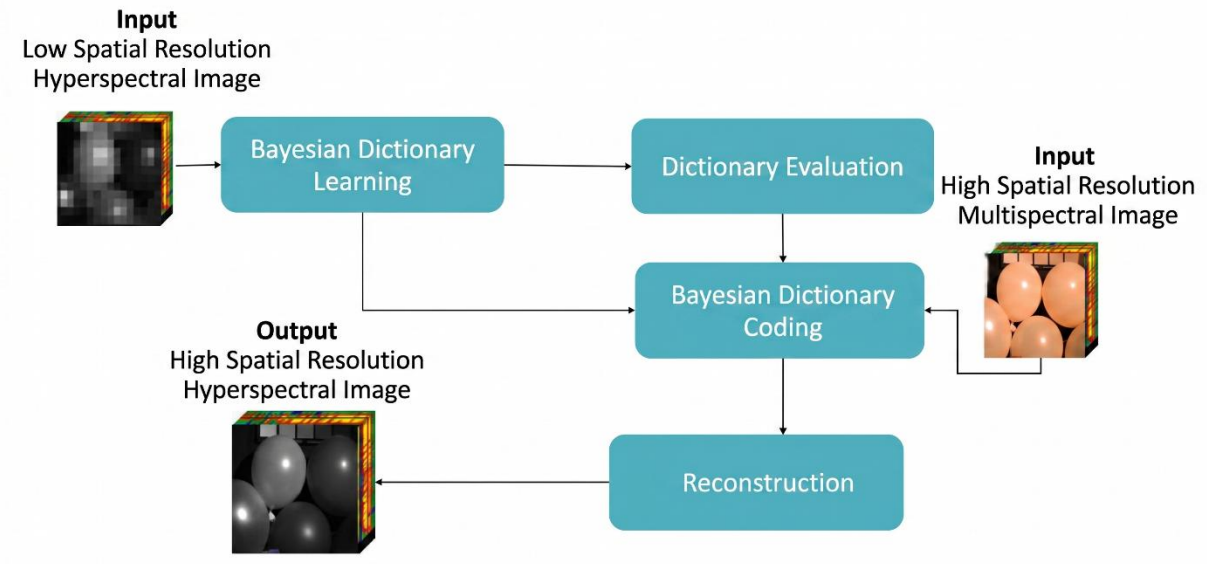

Figure 4. Algorithmic representation of super resolution using Bayesian sparse representation, adapted from [18].

In the first stage, the low-spatial-resolution image $Y^h \in \mathbb{R}^{m \times n \times L}$ is rearranged into a two-dimensional matrix of size $L \times (m \cdot n)$, and the number of spectral signatures it holds is estimated with the Bernoulli distribution. In the second stage, a dictionary is obtained from the distribution learned over the spectral data and is then adapted to the spectral quantization of the high-resolution image of the same scene. In the third stage, the high-spatial-resolution image is sparsely coded using the

CC: BY-NC-ND

dictionary and the distributions derived through Bayesian sparse coding. In the fourth stage, the final high-spatial-resolution hyperspectral image is produced from the dictionary and the sparse codes.

The columns of the low-resolution matrix are handled and described by a coefficient matrix A.

$$\overline{Y}^h = \Phi A \tag{7}$$

Likewise, for the high-resolution image, the expression is as follows.

$$\overline{Y} = \widehat{\Phi} B \tag{8}$$

Here, Φ and $\widehat{\Phi}$ are the dictionaries of the low- and high-resolution images, respectively, while A and B are their sparse-coding matrices. The dictionaries are related by $\widehat{\Phi} = \pi\Phi$, where π is the transformation matrix. Under Bayesian dictionary learning, the *i-th* pixel of Y^h is written using the sparse representation β_i^h over the dictionary and the generation error ϵ_i^h as follows.

$$y_i^h = \Phi\beta_i^h + \epsilon_i^h \tag{9}$$

2.4. Datasets

Two datasets commonly used in hyperspectral super-resolution studies were employed, namely CAVE and Harvard. The CAVE dataset was acquired with a cooled CCD camera equipped with a liquid-crystal tunable filter and contains 32 indoor scenes recorded under controlled illumination [78]. Each scene is stored as a 512×512×31 hyperspectral cube. The 31 spectral bands are sampled at 10 nm intervals over the 400-700 nm range, and each band is provided as a 16-bit reflectance image. The scenes contain a varied set of materials, including fabrics and textiles, paints, food and drinks, skin and hair, and real and artificial objects, thereby providing different spatial textures and spectral signatures within the collection.

The Harvard dataset contains 77 real-world scenes acquired with a commercial hyperspectral camera equipped with a liquid-crystal tunable filter [79]. Each cube has a size of 1392×1040×31, and the spectral axis is sampled at 10 nm intervals over the 420-720 nm range. The collection includes indoor and outdoor scenes: 50 were recorded under daylight illumination, while 27 were recorded under artificial or mixed illumination. The datasets differ in sensor, scene content, spatial size, and illumination conditions, while having the same number of bands and spectral sampling interval. Their joint use allows the method to be evaluated under a controlled, object-centred acquisition setting (CAVE) and a more varied, scene-centred setting (Harvard). All 31 available bands are retained in both datasets; no spectral interpolation or band extension is applied.

2.5. Implementation Details

Harvard images were cropped to 1024×1024×31 with the top-left corner used as the anchor, while CAVE images were kept at their native size. Low-spatial-resolution images were generated by applying a Gaussian blur followed by bicubic scaling at factors of 1, 0.75 and 0.5; rotations of 90°, 180° and 270° were additionally used to augment the training set. Each degraded image was then restored to its initial dimensions by bicubic interpolation.

The data were partitioned into 80% for training, 10% for validation, and 10% for testing. For CAVE, the model was run several times with varying train-validation-test combinations so that each evaluated image appeared in the test set at least once. For Harvard, a fixed set of 61 training, 8 validation, and 8 test images was used for both models. To expand the training data, the images were divided into sub-patches of size 32×32×L. Performance was evaluated at scale factors of ×2, ×3, and ×4.

CC: BY-NC-ND

Root mean square error and spectral angle mapper were used as loss terms in the networks. Optimization was performed with Adam using $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a learning rate of 10^{-4} for all layers. The settings of the deep convolutional neural network are listed in Table 2. The deep residual network contained 16 feed-forward blocks, each including four units based on 3×3 convolution, ReLU, and 1×1 convolution operations, with a growth rate of k = 16. The neural-network output and the low-resolution input image were fused using Bayesian dictionary learning with 1000 dictionary-learning iterations and 25 iterations for each sparse-code estimation. All compared methods were evaluated on the same test images generated through the degradation process described above.

Band selection strategy. Applying two-dimensional convolution to one band out of every three was adopted as a design choice for three reasons. First, the spectral axes of both datasets are sampled at 10 nm intervals, and neighbouring bands generally contain substantially correlated spatial structures, although their intensity values vary according to the spectral signature. Second, the deep network in the proposed framework is used mainly for spatial enhancement, whereas the Bayesian sparse-representation stage operates on the complete 31-band cube and reintroduces spectral information from all bands. Thus, bands not processed in the convolutional stage remain available to the fusion stage. Third, processing 11 rather than 31 bands reduces the convolutional workload by approximately two-thirds and facilitated the evaluation of two network architectures on two datasets and three scale factors.

The possible effect of this band-selection strategy on spectral fidelity should nevertheless be considered. If a narrow spectral feature occurs in a band that is not processed by the convolutional stage, the fine spatial detail of that band is recovered mainly through the fusion stage, which may attenuate some high-frequency spatial content. The spectral angle mapper (SAM) results reported in Section 3 suggest that this effect was limited for the datasets evaluated here, since the proposed models yielded the lowest dataset-average spectral errors among the compared methods at each scale factor. These findings suggest that sparse-representation-based fusion may partly compensate for spectral subsampling in the convolutional stage for densely sampled visible-range hyperspectral data. For sensors with coarser spectral sampling, wider spectral coverage, or pronounced narrow absorption features, a smaller stride or an adaptive band-selection strategy may be more appropriate.

Computational complexity. The cost of the framework can also be expressed analytically, independently of the hardware. For an image of H×W pixels, the three-layer convolutional network defined in Table 2 requires 9×9×1×64 + 1×1×64×32 + 5×5×32×1 = 8,032 multiply–accumulate operations per pixel and per processed band, so that its overall cost is O(H·W·B), where B is the number of processed bands. Processing 11 bands instead of all 31 therefore lowers the cost of the convolutional stage from about 2.49×10^5 to about 8.84×10^4 multiply–accumulate operations per pixel, i.e. a reduction of roughly 64% with respect to processing every band. The fusion stage is independent of the network and of the scale factor: Bayesian dictionary learning performs 1000 iterations for dictionary estimation and 25 iterations for each sparse-code estimation.

Table 2. Parameters used in the deep convolutional neural network.

Layer	**# of Channel**	**Filter Size**	**# of Filter**
First Layer	1	9×9	64
Second Layer	64	1×1	32
Third Layer	32	5×5	1

2.6. Evaluation Metrics

The quality of the super-resolved images was assessed with two complementary metrics: the root mean square error (RMSE), which quantifies the spatial reconstruction error, and the SAM, which quantifies the spectral similarity between the reconstructed pixels and the reference ones [80]. Let X denote the reference hyperspectral cube and $\hat{X}$ the reconstructed cube, both of size M×N×L, where M and N are the spatial dimensions and L is the number of spectral bands. RMSE is defined as

CC: BY-NC-ND

$$RMSE = \sqrt{\left((1 / (M \cdot N \cdot L)) \cdot \Sigma_{i=1}^{M} \Sigma_{j=1}^{N} \Sigma_{l=1}^{L} [X(i, j, l) - \hat{X}(i, j, l)]^2 \right)} \quad (10)$$

where X(i, j, l) and $\hat{X}(i, j, l)$ are, respectively, the reference and the reconstructed intensity of the pixel located at (i, j) in the l-th band. RMSE is expressed in the same intensity units as the images, and a value of zero corresponds to a perfect reconstruction.

SAM evaluates the reconstruction in the spectral domain. For every spatial position (i, j), the reference spectrum $x_{ij} \in \mathbb{R}^L$ and the reconstructed spectrum $\hat{x}_{ij} \in \mathbb{R}^L$ are treated as vectors, and the angle between them is computed. The metric is then obtained by averaging this angle over all pixels:

$$SAM = (1 / (M \cdot N)) \cdot \Sigma_{i=1}^{M} \Sigma_{j=1}^{N} \arccos\left(\langle x_{ij} , \hat{x}_{ij} \rangle / (\|x_{ij}\|_2 \cdot \|\hat{x}_{ij}\|_2) \right) \quad (11)$$

where $\langle \cdot , \cdot \rangle$ denotes the inner product and $\|\cdot\|_2$ the Euclidean norm. The resulting angle is reported in degrees. Because it depends only on the direction of the spectral vector and not on its magnitude, SAM is insensitive to multiplicative illumination changes and therefore measures the preservation of the spectral signature rather than that of the absolute intensity. It is thus the metric of primary interest for the fusion stage of the proposed framework.

3. RESULTS

The proposed methods were compared with bicubic interpolation, GFPCA [25], and SCT_SDCNN [36], which were established comparison methods within the literature and technical scope of the study when the experiments were conducted. The reported values were not copied directly from the corresponding publications because degradation models and evaluation settings can differ between studies. To improve comparability, all competing methods were run under the same degradation model, input data, and evaluation protocol used in this study, following the procedures described in their original publications. Tables 3 and 4 report the mean RMSE and SAM values obtained on the CAVE and Harvard datasets, respectively, at three scale factors.

Table 3. Comparison of the proposed methods against the other methods in terms of mean RMSE and SAM values on the CAVE dataset.

Method	x2 RMSE	x2 SAM	x3 RMSE	x3 SAM	x4 RMSE	x4 SAM
Bicubic	3.0751	7.7547	3.1290	10.2485	3.3475	11.4764
GFPCA	2.9098	6.6691	2.9307	9.1212	3.1352	10.4437
SCT_SDCNN	2.0625	3.6156	2.3320	5.4177	2.7023	6.3318
Convolutional Neural Network Based Proposed Method	2.3435	3.2836	2.6275	4.6418	3.0912	5.8737
Deep Residual Network Based Proposed Method	**2.0146**	**2.8507**	**2.1380**	**4.0742**	**2.4712**	**5.1211**

Table 4. Comparison of the proposed methods with the other methods in terms of mean RMSE and SAM values on the Harvard dataset.

Method	**x2 RMSE**	**x2 SAM**	**x3 RMSE**	**x3 SAM**	**x4 RMSE**	**x4 SAM**
Bicubic	2.6287	6.2398	2.9034	7.0563	3.3343	8.7641
GFPCA	2.5172	5.0676	2.4595	6.2308	2.8997	7.5734
SCT_SDCNN	1.6542	2.7320	1.9421	3.6680	2.4770	4.4290
Convolutional Neural Network Based Proposed Method	1.8471	2.4663	2.1623	3.1216	2.7447	3.8504
Deep Residual Network Based Proposed Method	1.5753	2.1328	1.7373	2.7135	2.1524	3.3576

CC: BY-NC-ND

At the dataset-average level, the deep residual network-based hybrid model yielded the lowest RMSE and SAM values among the evaluated methods on both datasets and at all three scale factors. On CAVE, its RMSE/SAM values were 2.0146/2.8507, 2.1380/4.0742, and 2.4712/5.1211 at ×2, ×3, and ×4, respectively. On Harvard, the corresponding values were 1.5753/2.1328, 1.7373/2.7135, and 2.1524/3.3576. Compared with SCT_SDCNN, the residual model reduced SAM by 21.16%, 24.80%, and 19.12% on CAVE and by 21.93%, 26.02%, and 24.19% on Harvard at ×2, ×3, and ×4, respectively. The more consistent reduction in SAM than in RMSE suggests that the sparse-representation stage contributed mainly to spectral-signature preservation under the evaluated conditions.

The error values generally increased as the scale factor grew. Nevertheless, the deep residual network-based model retained the lowest dataset-average RMSE and SAM values at each evaluated scale factor. When the two proposed models were compared at the dataset-average level, the deep residual network-based model produced lower RMSE and SAM values than the convolutional neural network-based model for both CAVE and Harvard.

4. DISCUSSION

This study examines the reconstruction of high-spatial-resolution hyperspectral images from low-spatial-resolution inputs using super-resolution methods. The proposed framework brings together single-image and image-fusion approaches that are often considered separately in the literature. Spatial resolution is enhanced using deep-learning models, while spectral information from the original hyperspectral cube is transferred through sparse representation.

The proposed approach has three intended features. First, deep learning is used to model the relationship between low- and high-spatial-resolution images. Second, the convolutional workload is reduced by processing one band selected from every three bands. Third, sparse representation is used to support the transfer of spectral information to the reconstructed image. The results suggest that combining single-image super-resolution with model-based fusion can provide lower reconstruction errors than the comparison methods under the experimental protocol used in this study. The work is therefore presented as an integrated hybrid implementation and evaluation within this defined scope.

Deep-learning-based hyperspectral super-resolution methods considered in the literature include three-dimensional or mixed 2D/3D convolutions for joint spatial-spectral modelling [42,43,46], as well as attention- or adversarial-based mechanisms that learn spectral consistency from data [40,41,44,45]. Once trained, such end-to-end models can provide efficient inference; however, their performance depends on the selected loss functions, network design, and training data, while three-dimensional convolutions generally require additional parameters and memory.

The framework examined here follows a model-assisted hybrid strategy. Spectral consistency is supported explicitly through a fusion stage in which information from the original low-resolution cube is reintroduced by Bayesian sparse representation. This design has three practical characteristics. First, it does not rely on an independently acquired high-resolution multispectral or panchromatic image or on a co-registered second sensor, because the spatially enhanced input is generated by the deep network from the observed cube. Second, the network uses two-dimensional convolutions on a subset of bands, which limits the convolutional workload relative to processing all bands with a three-dimensional architecture. Third, the spectral component is handled through an explicit model-based stage. The lower SAM values observed in the experiments are consistent with, but do not by themselves prove, a contribution from this spectral-fusion component.

The design also has limitations. The framework is not end-to-end: the network and fusion stages are optimized separately, and the reconstruction error of the first stage is not propagated to the dictionary-learning stage. Bayesian dictionary learning is iterative, which increases inference cost relative to a trained end-to-end network, although the convolutional stage processes only a subset of the bands. The reported conclusions should therefore be interpreted within the evaluated datasets, methods, and degradation settings.

CC: BY-NC-ND

The comparison set used in this study consists of bicubic interpolation, GFPCA and SCT_SDCNN, and reflects the literature context in which the experiments were designed and implemented. Since that time the single-image branch has advanced considerably, and transformer-, diffusion- and state-space-based methods now define the accuracy frontier on the CAVE and Harvard benchmarks [48,49,53,54]. Those methods, however, differ from the present framework in parameter budget, training-data requirement and inference procedure, and a comparison against them under the protocol used here would not be informative about the question this study addresses. The results reported in Section 3 should therefore be read as characterising the behaviour of an explicit, unsupervised spectral stage relative to methods of comparable scale, and not as a claim of state-of-the-art accuracy. Positioning the framework against current high-capacity architectures under a matched protocol is a distinct question, and one that the design of this study does not answer.

The lower dataset-average errors of the deep residual network-based model may be related to the use of local and global residual connections, which reintroduce features from earlier layers into the reconstruction process. This mechanism may reduce information loss during successive convolution operations and may help explain the more stable average results observed at higher scale factors.

A limitation of the method is its dependence on training data before application. Although CAVE and Harvard contain different scenes and acquisition conditions, both consist of densely sampled visible-range hyperspectral data with 31 bands. Application to data with substantially different sensors, spectral ranges, or scene characteristics may therefore require training samples with comparable properties.

Future studies may examine alternative deep-learning architectures and hyperparameter settings within the same hybrid framework. Adjusting the number of layers, band-selection strategy, and dictionary-learning settings may offer different trade-offs between reconstruction error and processing cost. The general structure may also be adapted to other combinations of single-image super-resolution and image-fusion methods.

5. CONCLUSION

This study evaluated a hybrid hyperspectral super-resolution method that combines deep learning with sparse representation. In the proposed framework, deep-learning models are used for spatial enhancement, while Bayesian sparse-representation-based dictionary learning supports the transfer of spectral information from the original low-resolution cube to the reconstructed image. Two-dimensional convolution is applied to selected bands in order to reduce the convolutional workload, while the complete spectral cube remains available to the fusion stage.

The method was evaluated on the CAVE and Harvard datasets using RMSE and SAM at scale factors of ×2, ×3, and ×4. Within the set of methods compared, the deep residual network-based hybrid model yielded the lowest dataset-average RMSE and SAM values at all scale factors. Its RMSE/SAM values were 2.0146/2.8507, 2.1380/4.0742, and 2.4712/5.1211 on CAVE and 1.5753/2.1328, 1.7373/2.7135, and 2.1524/3.3576 on Harvard. The SAM reductions relative to SCT_SDCNN ranged from 19.12% to 24.80% on CAVE and from 21.93% to 26.02% on Harvard. These results suggest that the hybrid structure can support spectral preservation while maintaining competitive spatial reconstruction under the evaluated protocol.

The findings are limited to the datasets, degradation process, scale factors, and comparison methods used in this study. Since the method contains a learning-based stage, application to data with different characteristics may require appropriate training samples. The iterative dictionary-learning stage also introduces additional computational cost. Alternative architectures, band-selection strategies, and model parameters may be investigated in future work to examine different accuracy-cost trade-offs.

CC: BY-NC-ND

Peer-review: External, Independent.

Acknowledgements:
The authors would like to thank the researchers who made the CAVE and Harvard hyperspectral image datasets publicly available. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. This study was derived from H.A.'s doctoral dissertation, conducted under the supervision of N.İ.

Declarations:

1. Statement of Originality:
This manuscript represents original work and has not been published in its current form, nor is it under consideration for publication elsewhere. A preliminary version of the convolutional-network-based hybrid framework was presented in an earlier study cited as Reference [72]. The present study substantially extends that preliminary work by incorporating a deep residual network architecture, evaluating the proposed framework on both the CAVE and Harvard datasets, considering multiple scale factors, and providing a more comprehensive experimental and comparative analysis.

2. Author Contributions:
Concept: HA,Nİ; ***Conceptualization:*** HA,Nİ; ***Literature Search:*** HA; ***Data Collection:*** HA; ***Data Processing:*** HA; ***Analysis:*** HA,Nİ; ***Writing – original draft:*** HA; ***Writing – review & editing:*** Nİ.

3. Ethics approval:
This study used publicly available hyperspectral image datasets and did not involve human participant recruitment, clinical interventions, animal experiments, or the collection of identifiable personal data. Therefore, ethics committee approval was not required.

4. Funding/Support:
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

5. Competing Interests:
The authors declare no competing interests.

6. GenAI Usage Statement:
Generative artificial intelligence tools were used during manuscript preparation for language refinement, stylistic editing, and improving the clarity and organization of the text. These tools were not used to generate experimental data or results. All scientific content, analyses, interpretations, references, and conclusions were reviewed and verified by the authors, who take full responsibility for the final content of the manuscript.

7. Sustainable Development Goals:

REFERENCES

[1] Goel, M., Whitmire, E., Mariakakis, A., Saponas, T. S., Joshi, N., Morris, D., Guenter, B., Gavriliu, M., Borriello, G., & Patel, S. N. (2015). HyperCam: Hyperspectral imaging for ubiquitous computing applications. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and

CC: BY-NC-ND

Ubiquitous Computing (UbiComp '15) (pp. 145–156). Osaka, Japan: ACM. https://doi.org/10.1145/2750858.2804282

[2] Jensen, J. R. (2005). Introductory digital image processing: A remote sensing perspective (3rd ed.). Upper Saddle River, NJ: Prentice Hall. ISBN 978-0-13-145361-6.

[3] Shaw, G. A., & Burke, H. K. (2003). Spectral imaging for remote sensing. Lincoln Laboratory Journal, 14(1), 3–28.

[4] Adão, T., Hruška, J., Pádua, L., Bessa, J., Peres, E., Morais, R., & Sousa, J. J. (2017). Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. Remote Sensing, 9(11), 1110. https://doi.org/10.3390/rs9111110

[5] Briottet, X., Boucher, Y., Dimmeler, A., Malaplate, A., Cini, A., Diani, M., Bekman, H., Schwering, P., Skauli, T., Kasen, I., Renhorn, I., Klasén, L., Gilmore, M., & Oxford, D. (2006). Military applications of hyperspectral imagery. In Targets and Backgrounds XII: Characterization and Representation (Proc. SPIE, Vol. 6239, Art. 62390B, pp. 82–89). Bellingham, WA: SPIE. https://doi.org/10.1117/12.672030

[6] Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: A review. Journal of Biomedical Optics, 19(1), 010901. https://doi.org/10.1117/1.JBO.19.1.010901

[7] Keshava, N., & Mustard, J. F. (2002). Spectral unmixing. IEEE Signal Processing Magazine, 19(1), 44–57. https://doi.org/10.1109/79.974727

[8] Eismann, M. T., & Hardie, R. C. (2004). Application of the stochastic mixing model to hyperspectral resolution enhancement. IEEE Transactions on Geoscience and Remote Sensing, 42(9), 1924–1933. https://doi.org/10.1109/TGRS.2004.830644

[9] Hardie, R. C., Eismann, M. T., & Wilson, G. L. (2004). MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor. IEEE Transactions on Image Processing, 13(9), 1174–1184. https://doi.org/10.1109/TIP.2004.829779

[10] Zhukov, B., Oertel, D., Lanzl, F., & Reinhackel, G. (1999). Unmixing-based multisensor multiresolution image fusion. IEEE Transactions on Geoscience and Remote Sensing, 37(3), 1212–1226. https://doi.org/10.1109/36.763276

[11] Wei, Q., Dobigeon, N., & Tourneret, J.-Y. (2014). Bayesian fusion of multispectral and hyperspectral images with unknown sensor spectral response. In 2014 IEEE International Conference on Image Processing (ICIP) (pp. 698–702). Paris, France. https://doi.org/10.1109/ICIP.2014.7025140

[12] Zhang, Y., De Backer, S., & Scheunders, P. (2009). Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images. IEEE Transactions on Geoscience and Remote Sensing, 47(11), 3834–3843. https://doi.org/10.1109/TGRS.2009.2017737

[13] Palsson, F., Sveinsson, J. R., Ulfarsson, M. O., & Benediktsson, J. A. (2015). Model-based fusion of multi- and hyperspectral images using PCA and wavelets. IEEE Transactions on Geoscience and Remote Sensing, 53(5), 2652–2663. https://doi.org/10.1109/TGRS.2014.2363477

[14] Bendoumi, M. A., He, M., & Mei, S. (2014). Hyperspectral image resolution enhancement using high-resolution multispectral image based on spectral unmixing. IEEE Transactions on Geoscience and Remote Sensing, 52(10), 6574–6583. https://doi.org/10.1109/TGRS.2014.2298056

[15] Yokoya, N., Yairi, T., & Iwasaki, A. (2012). Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE Transactions on Geoscience and Remote Sensing, 50(2), 528–537. https://doi.org/10.1109/TGRS.2011.2161320

CC: BY-NC-ND

[16] Li, S., Dian, R., Fang, L., & Bioucas-Dias, J. M. (2018). Fusing hyperspectral and multispectral images via coupled sparse tensor factorization. IEEE Transactions on Image Processing, 27(8), 4118–4130. https://doi.org/10.1109/TIP.2018.2836307

[17] Zhang, K., Wang, M., Yang, S., & Jiao, L. (2018). Spatial–spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(4), 1030–1040. https://doi.org/10.1109/JSTARS.2017.2785411

[18] Akhtar, N., Shafait, F., & Mian, A. (2015). Bayesian sparse representation for hyperspectral image super resolution. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3631–3640). Boston, MA, USA. https://doi.org/10.1109/CVPR.2015.7298986

[19] Huang, B., Song, H., Cui, H., Peng, J., & Xu, Z. (2014). Spatial and spectral image fusion using sparse matrix factorization. IEEE Transactions on Geoscience and Remote Sensing, 52(3), 1693–1704. https://doi.org/10.1109/TGRS.2013.2253612

[20] Wei, Q., Bioucas-Dias, J. M., Dobigeon, N., & Tourneret, J.-Y. (2014). Fusion of multispectral and hyperspectral images based on sparse representation. In 2014 22nd European Signal Processing Conference (EUSIPCO) (pp. 1577–1581). Lisbon, Portugal.

[21] Han, X.-H., & Chen, Y.-W. (2019). Deep residual network of spectral and spatial fusion for hyperspectral image super-resolution. In 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM) (pp. 266–270). Singapore. https://doi.org/10.1109/BigMM.2019.00-13

[22] Han, X.-H., Shi, B., & Zheng, Y. (2018). SSF-CNN: Spatial and spectral fusion with CNN for hyperspectral image super-resolution. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 2506–2510). Athens, Greece. https://doi.org/10.1109/ICIP.2018.8451142

[23] Palsson, F., Sveinsson, J. R., & Ulfarsson, M. O. (2017). Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network. IEEE Geoscience and Remote Sensing Letters, 14(5), 639–643. https://doi.org/10.1109/LGRS.2017.2668299

[24] Zhou, F., Hang, R., Liu, Q., & Yuan, X. (2019). Pyramid fully convolutional network for hyperspectral and multispectral image fusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(5), 1549–1558. https://doi.org/10.1109/JSTARS.2019.2910990

[25] Liao, W., Huang, X., Van Coillie, F., Gautama, S., Pizurica, A., Philips, W., Liu, H., Zhu, T., Shimoni, M., Moser, G., & Tuia, D. (2015). Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 IEEE GRSS Data Fusion Contest. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(6), 2984–2996. https://doi.org/10.1109/JSTARS.2015.2420582

[26] Simsek, M., & Polat, E. (2021). Performance evaluation of pan-sharpening and dictionary learning methods for sparse representation of hyperspectral super-resolution. Signal, Image and Video Processing, 15, 1099–1106. https://doi.org/10.1007/s11760-020-01836-8

[27] Hu, J.-F., Huang, T.-Z., Deng, L.-J., Dou, H.-X., Hong, D., & Vivone, G. (2022). Fusformer: A transformer-based fusion network for hyperspectral image super-resolution. IEEE Geoscience and Remote Sensing Letters, 19, Article 6012305, 1–5. https://doi.org/10.1109/LGRS.2022.3194257

[28] Ran, R., Deng, L.-J., Jiang, T.-X., Hu, J.-F., Chanussot, J., & Vivone, G. (2023). GuidedNet: A general CNN fusion framework via high-resolution guidance for hyperspectral image super-resolution. IEEE Transactions on Cybernetics, 53(7), 4148–4161.

CC: BY-NC-ND

https://doi.org/10.1109/TCYB.2023.3238200

[29] Wang, X., Hu, Q., Cheng, Y., & Ma, J. (2023). Hyperspectral image super-resolution meets deep learning: A survey and perspective. IEEE/CAA Journal of Automatica Sinica, 10(8), 1668–1691. https://doi.org/10.1109/JAS.2023.123681

[30] Dian, R., Li, S., Sun, B., & Guo, A. (2021). Recent advances and new guidelines on hyperspectral and multispectral image fusion. Information Fusion, 69, 40–51. https://doi.org/10.1016/j.inffus.2020.11.001

[31] Mianji, F. A., Gu, Y., Zhang, Y., & Zhang, J. (2011). Enhanced self-training superresolution mapping technique for hyperspectral imagery. IEEE Geoscience and Remote Sensing Letters, 8(4), 671–675. https://doi.org/10.1109/LGRS.2010.2102334

[32] Hu, J., Li, Y., & Xie, W. (2017). Hyperspectral image super-resolution by spectral difference learning and spatial error correction. IEEE Geoscience and Remote Sensing Letters, 14(10), 1825–1829. https://doi.org/10.1109/LGRS.2017.2737637

[33] Irmak, H., Akar, G. B., & Yuksel, S. E. (2018). A MAP-based approach for hyperspectral imagery super-resolution. IEEE Transactions on Image Processing, 27(6), 2942–2951. https://doi.org/10.1109/TIP.2018.2814210

[34] Jia, J., Ji, L., Zhao, Y., & Geng, X. (2018). Hyperspectral image super-resolution with spectral–spatial network. International Journal of Remote Sensing, 39(22), 7806–7829. https://doi.org/10.1080/01431161.2018.1471546

[35] Lei, S., Shi, Z., & Zou, Z. (2017). Super-resolution for remote sensing images via local–global combined network. IEEE Geoscience and Remote Sensing Letters, 14(8), 1243–1247. https://doi.org/10.1109/LGRS.2017.2704122

[36] Li, Y., Hu, J., Zhao, X., Xie, W., & Li, J. (2017). Hyperspectral image super-resolution using deep convolutional neural network. Neurocomputing, 266, 29–41. https://doi.org/10.1016/j.neucom.2017.05.024

[37] Yuan, Y., Zheng, X., & Lu, X. (2017). Hyperspectral image superresolution by transfer learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(5), 1963–1974. https://doi.org/10.1109/JSTARS.2017.2655112

[38] Arun, P. V., Buddhiraju, K. M., Porwal, A., & Chanussot, J. (2020). CNN-based super-resolution of hyperspectral images. IEEE Transactions on Geoscience and Remote Sensing, 58(9), 6106–6121. https://doi.org/10.1109/TGRS.2020.2973370

[39] Fu, Y., Liang, Z., & You, S. (2021). Bidirectional 3D quasi-recurrent neural network for hyperspectral image super-resolution. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 2674–2688. https://doi.org/10.1109/JSTARS.2021.3057936

[40] 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

[41] Li, J., Cui, R., Li, B., Song, R., Li, Y., Dai, Y., & Du, Q. (2020). Hyperspectral image super-resolution by band attention through adversarial learning. IEEE Transactions on Geoscience and Remote Sensing, 58(6), 4304–4318. https://doi.org/10.1109/TGRS.2019.2962713

[42] 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

[43] Li, Q., Wang, Q., & Li, X. (2021). Exploring the relationship between 2D/3D convolution for

CC: BY-NC-ND

hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 59(10), 8693–8703. https://doi.org/10.1109/TGRS.2020.3047363

[44] Wang, Q., Li, Q., & Li, X. (2021). Hyperspectral image superresolution using spectrum and feature context. IEEE Transactions on Industrial Electronics, 68(11), 11276–11285. https://doi.org/10.1109/TIE.2020.3038096

[45] Wang, X., Ma, J., & Jiang, J. (2022). Hyperspectral image super-resolution via recurrent feedback embedding and spatial–spectral consistency regularization. IEEE Transactions on Geoscience and Remote Sensing, 60, Article 5503113, 1–13. https://doi.org/10.1109/TGRS.2021.3064450

[46] Zheng, K., Gao, L., Ran, Q., Cui, X., Zhang, B., Liao, W., & Jia, S. (2019). Separable-spectral convolution and inception network for hyperspectral image super-resolution. International Journal of Machine Learning and Cybernetics, 10, 2593–2607. https://doi.org/10.1007/s13042-018-00911-4

[47] Liu, Y., Hu, J., Kang, X., Luo, J., & Fan, S. (2022). Interactformer: Interactive transformer and CNN for hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 60, Article 5531715, 1–15. https://doi.org/10.1109/TGRS.2022.3183468

[48] Chen, S., Zhang, L., & Zhang, L. (2023). MSDformer: Multiscale deformable transformer for hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 61, Article 5525614, 1–14. https://doi.org/10.1109/TGRS.2023.3315970

[49] Zhang, M., Zhang, C., Zhang, Q., Guo, J., Gao, X., & Zhang, J. (2023). ESSAformer: Efficient transformer for hyperspectral image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 23016–23027). Paris, France. https://doi.org/10.1109/ICCV51070.2023.02109

[50] Chen, S., Zhang, L., & Zhang, L. (2024). 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

[51] 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

[52] Yang, X., Li, S., Cao, W., Tang, D., Ban, Y., & Zhou, Y. (2026). RWKVSR: Receptance weighted key-value network for hyperspectral image super-resolution. IEEE Transactions on Circuits and Systems for Video Technology, 36(4), 4165–4180. https://doi.org/10.1109/TCSVT.2025.3626779

[53] Wu, C., Wang, D., Bai, Y., Mao, H., Li, Y., & Shen, Q. (2023). HSR-Diff: Hyperspectral image super-resolution via conditional diffusion models. In 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 7060–7070). Paris, France. https://doi.org/10.1109/ICCV51070.2023.00652

[54] Xu, Y., Wang, H., Zhou, F., Luo, C., Sun, X., Rahardja, S., & Ren, P. (2025). MambaHSISR: Mamba hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 63, Article 5511216, 1–16. https://doi.org/10.1109/TGRS.2025.3560632

[55] 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. Proceedings of the AAAI Conference on Artificial Intelligence, 39(7), 7033–7041. https://doi.org/10.1609/aaai.v39i7.32755

[56] Gendy, G., & AlMarzouqi, H. (2026). Remote sensing image super-resolution via hybrid Mamba

CC: BY-NC-ND

and convolutional architectures. Neurocomputing, 683, 133561. https://doi.org/10.1016/j.neucom.2026.133561

[57] Cao, H., Tan, Z., Zhu, B., Xiong, H., & Lv, Q. (2026). A review on super-resolution reconstruction of single-frame remote sensing images via diffusion models. Remote Sensing, 18(11), 1702. https://doi.org/10.3390/rs18111702

[58] Yang, P., Ma, Y., Mei, X., Chen, Q., Wu, M., & Ma, J. (2025). Deep blind super-resolution for hyperspectral images. Pattern Recognition, 157, 110916. https://doi.org/10.1016/j.patcog.2024.110916

[59] Cheng, Y., Wang, X., Ma, Y., Mei, X., Wu, M., & Ma, J. (2025). General hyperspectral image super-resolution via meta-transfer learning. IEEE Transactions on Neural Networks and Learning Systems, 36(4), 6134–6147. https://doi.org/10.1109/TNNLS.2024.3387970

[60] Li, K., Van Gool, L., & Dai, D. (2025). Test-time training for hyperspectral image super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(9), 7231–7242. https://doi.org/10.1109/TPAMI.2024.3461807

[61] Chen, C., Wang, Y., Zhang, N., Zhang, Y., & Zhao, Z. (2023). A review of hyperspectral image super-resolution based on deep learning. Remote Sensing, 15(11), 2853. https://doi.org/10.3390/rs15112853

[62] Dong, W., Zhou, C., Wu, F., Wu, J., Shi, G., & Li, X. (2021). Model-guided deep hyperspectral image super-resolution. IEEE Transactions on Image Processing, 30, 5754–5768. https://doi.org/10.1109/TIP.2021.3078058

[63] 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

[64] Wang, K., Liao, X., Li, J., Meng, D., & Wang, Y. (2023). Hyperspectral image super-resolution via knowledge-driven deep unrolling and transformer embedded convolutional recurrent neural network. IEEE Transactions on Image Processing, 32, 4581–4594. https://doi.org/10.1109/TIP.2023.3293768

[65] Cao, X., Wang, X., Dun, X., Lian, Y., Cheng, X., & Hao, X. (2026). Cross-domain-aware deep unfolding transformer for hyperspectral image super-resolution. Pattern Recognition, 172(Part A), 112374. https://doi.org/10.1016/j.patcog.2025.112374

[66] Guo, W.-J., Xie, W., Jiang, K., Li, Y., Lei, J., & Fang, L. (2023). Toward stable, interpretable, and lightweight hyperspectral super-resolution. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 22272–22281). Vancouver, BC, Canada. https://doi.org/10.1109/CVPR52729.2023.02133

[67] Yang, J., Wu, C., You, T., Wang, D., Li, Y., Shang, C., & Shen, Q. (2023). Hierarchical spatio-spectral fusion for hyperspectral image super resolution via sparse representation and pre-trained deep model. Knowledge-Based Systems, 260, 110170. https://doi.org/10.1016/j.knosys.2022.110170

[68] Deka, B., Mullah, H. U., Barman, T., & Datta, S. (2023). Joint sparse representation-based single image super-resolution for remote sensing applications. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 2352–2365. https://doi.org/10.1109/JSTARS.2023.3244069

[69] Ye, F., Wu, Z., Xu, Y., Liu, H., & Wei, Z. (2022). Bayesian hyperspectral image super-resolution in the presence of spectral variability. IEEE Transactions on Geoscience and Remote Sensing, 60,

CC: BY-NC-ND

Article 5545613, 1–13. https://doi.org/10.1109/TGRS.2022.3228313

[70] Muhammad, U., & Laaksonen, J. (2025). Hybrid deep learning for hyperspectral single-image super-resolution. IEEE Geoscience and Remote Sensing Letters, 22, Article 5509905, 1–5. https://doi.org/10.1109/LGRS.2025.3616772

[71] Li, J., Zheng, K., Gao, L., Han, Z., Li, Z., & Chanussot, J. (2025). Enhanced deep image prior for unsupervised hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 63, Article 5504218, 1–18. https://doi.org/10.1109/TGRS.2025.3531646

[72] Aydilek, H., & İnanç, N. (2021). Learning Based Super Resolution Application for Hyperspectral Images. International Scientific and Vocational Studies Journal, 5(2), 210-217. https://doi.org/10.47897/bilmes.1049338

[73] Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281

[74] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2261–2269). Honolulu, HI, United States. https://doi.org/10.1109/CVPR.2017.243

[75] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). Las Vegas, NV, United States. https://doi.org/10.1109/CVPR.2016.90

[76] Wen, R., Fu, K., Sun, H., Sun, X., & Wang, L. (2018). Image superresolution using densely connected residual networks. IEEE Signal Processing Letters, 25(10), 1565–1569. https://doi.org/10.1109/LSP.2018.2861989

[77] Elad, M. (2010). Sparse and redundant representations: From theory to applications in signal and image processing. Springer. https://doi.org/10.1007/978-1-4419-7011-4

[78] Yasuma, F., Mitsunaga, T., Iso, D., & Nayar, S. K. (2010). Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. IEEE Transactions on Image Processing, 19(9), 2241–2253. https://doi.org/10.1109/TIP.2010.2046811

[79] Chakrabarti, A., & Zickler, T. (2011). Statistics of real-world hyperspectral images. In CVPR 2011 (pp. 193–200). Colorado Springs, CO, United States. https://doi.org/10.1109/CVPR.2011.5995660

[80] Girouard, G., Bannari, A., El Harti, A., & Desrochers, A. (2004). Validated spectral angle mapper algorithm for geological mapping: Comparative study between QuickBird and Landsat-TM. In Proceedings of the 20th ISPRS Congress: Geo-Imagery Bridging Continents (pp. 599–604). Istanbul, Turkey.

CC: BY-NC-ND