SCALMU: Synthetically Trained Coupling of Adaptive Learned Multiplicative Updates for Hyperspectral–Multispectral Fusion

Xinxin Xu, Yann Gousseau, Christophe Kervazo, and Saïd Ladjal

***Abstract*—Hyperspectral–multispectral image (HSI–MSI) fusion aims to recover a high-resolution hyperspectral (HSI) image from a low-resolution HSI and a high-resolution multispectral (MSI) image. Classical methods such as coupled non-negative matrix factorization (CNMF) benefit from a strong physical interpretability but suffer from inferior results compared to their deep-learning counterparts. To address this limitation, we propose synthetically trained coupling of adaptive learned multiplicative updates (SCALMUs), a novel blind unrolled neural network architecture that integrates adaptive learnable matrices within the classical framework of CNMF multiplicative updates (MUs), improving its results. Due to its architectural proximity to CNMF, the resulting algorithm preserves physical interpretability and non-negativity constraints. To overcome the scarcity of supervised training data, we generate a synthetic HSI–MSI dataset using the dead leaves model and train SCALMU end-to-end under synthetic supervision. Experiments on several datasets show that SCALMU outperforms state-of-the-art (SOTA) methods and highlights the potential of blind fusion trained with synthetic data. The code is available at https://github.com/xinxinxu99/SCALMU.git**

***Index Terms*—Data fusion, hyperspectral (HSI) image, remote sensing, super-resolution, synthetic training data, unrolling.**

I. Introduction

HYPERSPECTRAL (HSI) and multispectral (MSI) images are 3-D data cubes with two spatial and one spectral dimensions. When both modalities are used to acquire the same scene, the resulting images differ in spectral richness and spatial detail. HSIs contain hundreds of contiguous narrow bands, enabling precise material identification and supporting diverse applications such as source separation [1], target detection [2], vegetation monitoring [3], and land cover classification [4]. However, due to physical constraints, there is a tradeoff between spatial and spectral resolutions, leading in HSIs to a low spatial resolution, which is further amplified in remote sensing due to large sensor-to-scene distances. In contrast, MSIs capture only a few broad spectral bands but at much higher spatial resolution, offering finer spatial structures and sharper details.

Combining these complementary modalities, namely, the high spectral precision of HSI and the great spatial details of MSI, defines the hyperspectral–multispectral image (HSI–MSI) fusion problem, also referred to as HSI image super-resolution. The goal is to reconstruct a high-resolution HSI image that preserves both the spectral fidelity of HSI and the spatial richness of MSI, a capability that has become increasingly important for modern remote sensing and environmental applications [5].

Existing HSI–MSI fusion methods can be broadly categorized into model-based and deep learning approaches [6]. Traditional model-based techniques, exemplified by the coupled non-negative matrix factorization (CNMF) [7], decompose the HSI into high spectral-resolution endmembers and leverage MSI-derived high spatial-resolution abundances through a linear unmixing model. This physically interpretable approach ensures spectral fidelity without requiring ground-truth (GT) HR-HSI data. However, CNMF requires prior knowledge of degradation operators and suffers from slow convergence through iterative multiplicative updates (MUs).

More recently, the advent of deep learning has enabled fully unsupervised and blind fusion methods that jointly estimate these degradation operators from data. Architectures like HyCoNet [8] and MIAE [9] integrate physical modeling within neural networks, learning degradation parameters alongside the fusion process to enhance the reconstruction quality and flexibility.

Despite these advances, deep learning approaches depend on scarce real training pairs with limited generalization. In this work, we propose to overcome this limitation by relying on synthetic training sets. For the super-resolution of RGB images, Achddou et al. [10] demonstrated the feasibility of training with purely synthetic images. Recently, we proposed a dead leaves model-based dataset for single HSI image super-resolution [11], establishing the potential of synthetic training for HSI restoration tasks.

In this article, in the framework of HSI–MSI fusion tasks, we introduce synthetically trained coupling of adaptive learned MUs (SCALMUs), which unrolls CNMF MUs into a learnable architecture trained exclusively on dead leaves synthetic data, combining physical interpretability with accurate data-driven reconstruction. The main contributions are as follows.

Received 27 March 2026; revised 3 July 2026; accepted 9 July 2026. Date of publication 13 July 2026; date of current version 22 July 2026. This work was supported by the Agence de l'Innovation de Défense (AID) via the Centre Interdisciplinaire d'Etudes pour la Défense et la Sécurité (CIEDS) under Project 2023—ALIA. *(Corresponding author: Xinxin Xu.)*

The authors are with LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France (e-mail: xinxin.xu@telecom-paris.fr; yann.gousseau@telecom-paris.fr; christophe.kervazo@telecom-paris.fr; said.ladjal@telecom-paris.fr).

Digital Object Identifier 10.1109/TGRS.2026.3712501

1558-0644 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

1) We propose complementary ingredients within a unified framework: adaptive learned MU (CALMU), an unrolled CNMF algorithm that integrates learnable components within the classical framework of MUs. This unrolled design preserves the interpretability of CNMF while requiring far fewer iterations and improving the super-resolution results.
2) We introduce a synthetic training set generation pipeline for HSI–MSI fusion based on the dead leaves model and HSI unmixing, allowing the creation of realistic spatial and spectral patterns with full GT control. By training on synthetic data, our method enables fully end-to-end learning without requiring GT high-resolution HSI images.
3) The main novelty of SCALMU lies in the combination of three CALMUs, synthetic training based on dead leaves generation, and blind estimation of the spatial and spectral degradation operators. Rather than introducing these components independently, our contribution is to show that their coupling yields an interpretable, trainable, and effective approach for HSI–MSI fusion.
4) We demonstrate that combining our physics-informed unrolled architecture (CALMU) with a purely synthetic training strategy yields superior super-resolution performance compared to state-of-the-art (SOTA) model-based and deep learning methods. Extensive experiments on multiple datasets validate that our approach generalizes exceptionally well to real-world scenarios, offering a highly efficient and practical solution for HSI–MSI fusion.

The rest of this article is organized as follows. In Section II, we give an overview of the related work. Section III details the proposed SCALMU framework, presenting its two main components: CALMU, the unrolled CNMF formulation, and the procedure for the Synthetic dataset generation leading from CALMU to SCALMU. Section IV reports experimental results, comparisons with SOTA methods, and ablation studies, demonstrating the effectiveness of our approach. Finally, conclusions and perspectives are drawn in Section V.

II. Related Work

HSI–MSI fusion has been extensively studied over the past decades, evolving from physically grounded model-based methods to data-driven deep learning approaches. This section reviews model-based fusion techniques (Section II-A), deep learning-based methods, addressing supervised and unsupervised paradigms (Section II-B), and, closely related to our work, recent advances in synthetic data generation to overcome training data scarcity (Section II-C).

A. Model-Based Fusion Method

Model-based methods for HSI–MSI fusion explicitly make use of physical observation models and handcrafted priors on the latent high-resolution HSI image. Early approaches benefited from spectral unmixing and Maximum A Posteriori estimation with stochastic mixing models. Pioneering works include Zhukov et al. [12], who proposed multisensor multiresolution fusion by linearly unmixing HSI into endmembers and using MSI high-resolution abundances; Eismann and Hardie [13], who applied stochastic mixing models assuming pixel-wise endmember variability for HSI resolution enhancement; and Hardie et al. [14], who developed MAP estimators fusing HSI with auxiliary high-resolution MSI via joint spatial–spectral regularization. These methods were extended to arbitrary spectral responses [15].

A major characteristic of these early model-based approaches was to represent HSI data in a low-dimensional subspace. Under this assumption, fusion can be formulated through matrix or tensor factorization, dictionary learning, or spectral unmixing into endmembers and abundances. For instance, dictionary-based methods decompose HSI into spectral atoms and sparse coefficients [16], while Akhtar et al. [17] learned overcomplete dictionaries by incorporating sparsity, non-negativity, and spatial priors through greedy pursuit. In these formulations, fusion amounts to jointly estimating spectral bases and spatial coefficients so that the reconstructed image, after spatial and spectral degradation, matches the observed HSI and MSI. Recent extensions also include noise-aware and robust matrix-factorization formulations, such as MixFus [18], which explicitly models mixed noise during HSI–MSI fusion, and GSFus [19], which introduces group sparsity and a plug-and-play denoising prior to handle localized discrepancies while improving robustness to noise.

Among these, CNMF [6], [7] stand out by decomposing each pixel into high spectral-resolution endmembers from HSI and high spatial-resolution abundances from MSI, enforcing non-negativity constraints and coupling both modalities through degradation operators to ensure spectral-spatial consistency without requiring GT HR-HSI. Despite their robustness and interpretability, these methods often require iterative optimization that can be computationally expensive and sensitive to initialization or degradation model inaccuracies.

B. Deep-Learning-Based Fusion Method

Deep-learning-based methods for HSI–MSI fusion have recently emerged as powerful alternatives to model-based frameworks, replacing handcrafted physical priors with data-driven representations learned end-to-end from paired HSI–MSI data [20]. Instead of explicitly modeling spectral mixing or imposing spatial regularization, deep networks directly learn nonlinear mappings from low-resolution HSI and high-resolution MSI to a high-resolution HSI, effectively capturing complex spatial–spectral correlations [8], [21]. This paradigm shift from analytical optimization to learned inference has led to notable improvements in reconstruction accuracy and robustness to unknown degradations [22], though it also introduces a strong dependence on the fidelity of training pairs and the realism of the simulated degradations.

Early supervised deep learning approaches primarily relied on residual convolutional neural networks (CNNs). Wang et al. [20] introduced a deep residual CNN that jointly processes HSI and MSI inputs, predicting high-frequency spatial details through skip connections while preserving global spectral structure. This residual learning paradigm inspired subsequent

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

architectures like Han et al.'s [23] SSF-CNN, which fuses spatial high-resolution and spectral low-resolution features within shared residual blocks. In the same vein, Dian et al. [24] regularized HSI–MSI fusion with a CNN denoiser prior, improving reconstruction quality while enhancing robustness to noise. Later developments emphasized multiscale feature extraction and high-resolution guidance: Xie et al.'s [21] MS/HS FusionNet combines unrolled model-based priors with residual fusion modules and progressive reconstruction, while Ran et al.'s [25] GuidedNet leverages a high-resolution guidance branch alongside feature reconstruction for enhanced spatial–spectral fidelity.

Unsupervised/self-supervised methods emerged to address the need for paired training data, which comprises the HSI, the MSI, and the corresponding high-resolution GT HSI, incorporating physical modeling directly into neural architectures. Zheng et al.'s [8] uSDN employs a coupled encoder–decoder with sparse Dirichlet priors [26], while HyCoNet adopts a three-branch autoencoder capable of jointly estimating spatial and spectral degradations from observations. Similarly, MIAE [9] introduces model-inspired autoencoders, UDALN [27] explores blind degradation learning within an unsupervised deep fusion framework, BUSIFusion [28] proposes a blind unsupervised single-image fusion strategy for HSI and RGB observations, and EU2ADL [29] enhances degradation learning with attention mechanisms. CSSnet [30] leverages cross-scale nonlocal attention, and DCnet [31] proposes decoupled fusion separating common and sensor-specific components. More recently, advanced unsupervised methods have pushed the boundaries further. Zhang et al.'s [32] test-time adaptation enables handling unknown degradations through self-supervised fine-tuning at inference time. Symmetrical propagation [33] captures long-range spatial–spectral dependencies via bidirectional feature propagation. Transformer-based FusFormer [34] leverages global self-attention for enhanced cross-modality interactions, while FeafusFormer [35] further investigates blind fusion through complementary local and global feature modeling. HSR-Diff [36] introduces conditional diffusion models for iterative high-fidelity reconstruction. EDIP-Net [37] improves deep image priors with zero-shot input generation and dual U-Net architectures. Finally, OTIAS [38] proposes OcTree-based adaptive sampling to preserve fine details across scales.

C. Synthetic Data

Synthetic datasets are particularly relevant for HSI–MSI fusion as they provide paired training data with perfect GT control, overcoming the scarcity of real high-resolution HSI references. However, their use remains uncommon in the HSI imaging and remote sensing literature. In contrast, in the field of computer vision, the use of synthetic data has been commonplace, and many works have been devoted to this topic, mostly for classification, estimation, or detection tasks, relying, for instance, on simplified object generation [39] or more recently text-to-image models [40]. For image restoration tasks, various generative models have been proposed to simulate realistic image statistics and structures. Markov random fields [41] allow the modeling of local spatial dependencies, making them effective for capturing texture patterns. Wavelet-based models [42] are well-suited for representing images at multiple scales, enabling the synthesis of textures with both fine and coarse structures. Gaussian models [43] offer a mathematically tractable way to reproduce second-order statistics of natural scenes, which often suffice for basic texture synthesis. The dead leaves model offers a simple framework for simulating edges and homogeneous regions by overlapping opaque objects and effectively reproduces non-Gaussian statistics observed in natural images [44], [45], [46], [47]. Achddou et al. [10] demonstrated that CNNs trained exclusively on dead leaves images can achieve competitive performance for the denoising and super-resolution of natural images. Recently, we extended this approach to single HSI image restoration in [11], introducing a dead leaves synthetic dataset with abundance map generation that enables unsupervised single image super-resolution training.

III. Proposed Method

This section presents the SCALMU framework for blind HSI–MSI fusion. Section III-A formulates the problem, Section III-B presents CALMU (the unrolled CNMF network), Section III-C describes our synthetic data generation leading to SCALMU, and Section III-D details blind degradation estimation.

A. Problem Formulation

HSI–MSI fusion aims to estimate a super-resolved image, denoted as $Z \in \mathbb{R}^{B\times HW}$, from the same scene observed by an HSI, $X \in \mathbb{R}^{B\times hw}$, and an MSI, $Y \in \mathbb{R}^{b\times HW}$. Here, (h, w) and (H, W) represent the spatial dimensions of the HSI and MSI, respectively, with $h < H$ and $w < W$, while B and b denote the number of spectral bands in the HSI and MSI, respectively, satisfying $b < B$. By introducing the point spread function (PSF)[1] $\mathbf{P} \in \mathbb{R}^{hw\times HW}$ and the spectral response function (SRF) $\mathbf{R} \in \mathbb{R}^{b\times B}$, which model the spatial and spectral degradations of Z, respectively, the HSI X and the MSI Y can be expressed as

$$\begin{aligned} X &= Z\mathbf{P}^\top \\ Y &= \mathbf{R}Z. \end{aligned} \tag{1}$$

In parallel, under the linear HSI unmixing model [7], by introducing M as the number of materials, the observations X and Y can be expressed as $X \simeq S_h A_h$ and $Y \simeq S_m A_m$, where $S_h \in \mathbb{R}^{B\times M}$ and $S_m \in \mathbb{R}^{b\times M}$ denote the endmember matrices obtained from the HSI and MSI, respectively, while $A_h \in \mathbb{R}^{M\times hw}$ and $A_m \in \mathbb{R}^{M\times HW}$ denote the abundance maps obtained from the HSI and MSI, respectively. The fused image can then be obtained as $Z = S_h A_m$, where S_h and A_m encode the high spectral resolution (from the HSI) and high spatial resolution (from the MSI) information, respectively. Based on this model and (1), we obtain

$$X \simeq S_h A_h = S_h A_m \mathbf{P}^\top$$

[1] $\mathbf{P}$ is, in fact, the linear operation consisting of blurring an image with a PSF and then downsampling it. The linear operation notation is more convenient for our presentation and details about the determination of the actual PSF are given in Section III-D.

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

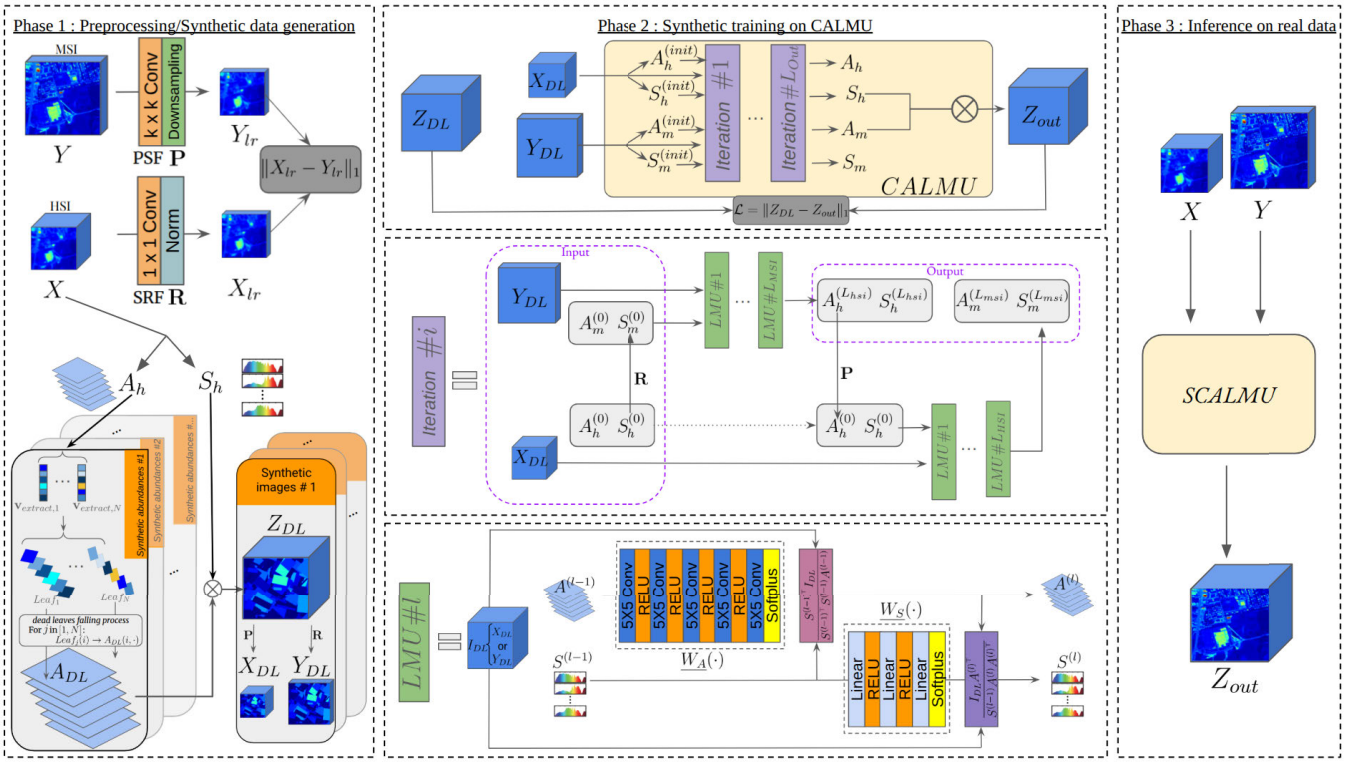


Fig. 1. Structure of the proposed SCALMU: Phase 1 consists of the blind determination of degradations **P** and **R** followed by the dead leaves dataset generation. Phase 2 (central part) is the training using the dead leaves dataset. Phase 3 corresponds to test time: merely replace X_{DL} and Y_{DL} by real hyper and multispectral images in phase 2, Z (high-resolution HSI image) being the output of the network.

$$\begin{aligned} Y &\simeq S_m A_m = \mathbf{R} S_h A_m \\ Z &= S_h A_m \end{aligned} \tag{2}$$

with the coupling relations $A_h = A_m \mathbf{P}^\top$ and $S_m = \mathbf{R} S_h$.

B. Proposed CALMU Structure

We propose CALMU, an unrolled version of the CNMF algorithm [7]. We recall that in this method, the fusion problem is formulated as a joint minimization, under non-negativity conditions, of the following cost functions:

$$\begin{aligned} &\underset{S_h, A_h \geq 0}{\operatorname{argmin}} \ \|X - S_h A_h\|_F^2 \\ &\underset{S_m, A_m \geq 0}{\operatorname{argmin}} \ \|Y - S_m A_m\|_F^2 \end{aligned} \tag{3}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. In the CNMF algorithm [7], the minimization is done iteratively using the well-known MUs [48] and leveraging the coupling constraints $A_h = A_m \mathbf{P}^\top$ and $S_m = \mathbf{R} S_h$. In the present work, the MU iterations of CNMF are unrolled, building on the recursive adaptive learned MUs (RALMUs) framework of Kervazo and Cohen [49] initially designed for HSI unmixing.

More precisely, we introduce four lightweight neural networks in CNMF updates to predict some adaptive learnable matrices $W_{A_m}(\cdot), W_{S_m}(\cdot), W_{A_h}(\cdot)$, and $W_{S_h}(\cdot)$, which dynamically adjust the MU update rules based on the input data characteristics, enabling to largely improve CNMF super-resolution results while dramatically reducing the number of required iterations. As illustrated in the *LMU* block of Fig. 1, the estimation of the abundance-related matrices is handled by two convolutional networks, one dedicated to the MSI branch $W_{A_m}(\cdot)$ and the other to the HSI branch $W_{A_h}(\cdot)$. Each of these networks is implemented as a five-layer Conv2D architecture with 5×5 convolutions, intermediate ReLU activations, and a final Softplus activation to enforce non-negativity. They process the current abundance estimates to generate spatially adaptive weights, allowing the model to refine abundance maps at each iteration. The endmember-related matrices are controlled by two MLPs that produce the corresponding spectral transformations, denoted as $W_{S_m}(\cdot)$ and $W_{S_h}(\cdot)$, each composed of three linear layers with ReLU activations between hidden layers and a final Softplus activation. These MLPs operate along the spectral dimension and learn to adjust the endmember updates according to the spectral distribution of the input. To maintain the non-negativity in the unrolled MUs, all predicted matrices are constrained through the final Softplus layer. Due to the structure of its alternating updates, the architecture ensures that its intermediate outputs maintain their physical meaning as abundance and endmember matrices, contrary to methods relying on abstract latent spaces.

The entire CALMU network, described in Algorithm 1 and illustrated in Phase 2 of Fig. 1, is trained end-to-end using synthetic data by employing the L1 reconstruction loss $\mathcal{L} = \|Z - Z_{out}\|_1$, ensuring full control over the GT super-resolved images and the associated degradation processes. The procedure used to generate these synthetic data is detailed in Section III-C.

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 CALMU Algorithm

Require: HSI X, MSI Y, PSF $\mathbf{P}$, SRF $\mathbf{R}$, number of iterations L_{MS1}, L_{HS1}, L_{Out}

Step 1: Initial estimation with Multiplicative Update (MU)
$A_h^{(L_{HS1})}, S_h^{(L_{HS1})} \leftarrow MU(X)$
$A_m^{(L_{MS1})} \leftarrow rand_{[0,1]}(\cdot)$

Step 2: Multispectral update
$S_m^{(0)} \leftarrow \mathbf{R} S_h^{(L_{HS1})} \quad A_m^{(0)} \leftarrow A_m^{(L_{MS1})}$

for $l = 0$ **to** $L_{MS1} - 1$ **do**

$$A_m^{(l+1)} \leftarrow A_m^{(l)} \odot \underline{W_{A_m}}(A_m^{(l)}) \odot \frac{S_m^{(l)^\top} Y}{S_m^{(l)^\top} S_m^{(l)} A_m^{(l)}}$$

$$S_m^{(l+1)} \leftarrow S_m^{(l)} \odot \underline{W_{S_m}}(S_m^{(l)}) \odot \frac{Y A_m^{(l+1)^\top}}{S_m^{(l)} A_m^{(l+1)} A_m^{(l+1)^\top}}$$

end for

Step 3: Hyperspectral update
$A_h^{(0)} \leftarrow A_m^{(L_{MS1})} \mathbf{P}^T$,
$S_h^{(0)} \leftarrow S_h^{(L_{HS1})} \odot \frac{X A_h^{(0)^\top}}{S_h^{(L_{HS1})} A_h^{(0)} A_h^{(0)^\top}}$

for $l = 0$ **to** $L_{HS1} - 1$ **do**

$$A_h^{(l+1)} \leftarrow A_h^{(l)} \odot \underline{W_{A_h}}(A_h^{(l)}) \odot \frac{S_h^{(l)^\top} X}{S_h^{(l)^\top} S_h^{(l)} A_h^{(l)}}$$

$$S_h^{(l+1)} \leftarrow S_h^{(l)} \odot \underline{W_{S_h}}(S_h^{(l)}) \odot \frac{X A_h^{(l+1)^\top}}{S_h^{(l)} A_h^{(l+1)} A_h^{(l+1)^\top}}$$

end for

Step 4: Repeat Steps 2 and 3 for L_{Out} iterations

return: $Z_{out} = S_h^{(L_{HS1})} A_m^{(L_{MS1})}$

Notes: $\odot$ *denotes element-wise multiplication. The case where the four* $\underline{W^{(l)}}$ *matrices are set to 1 (nonlearnable) corresponds to the original CNMF updates.*

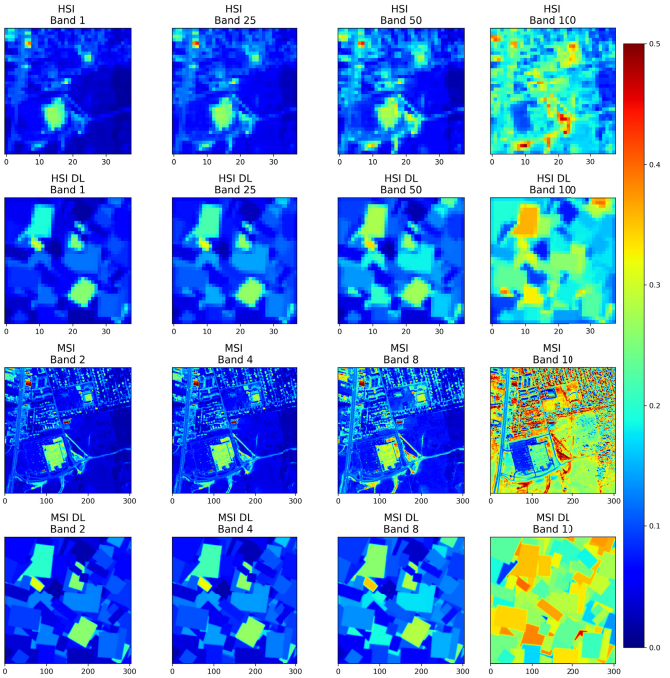


Fig. 2. Visual comparison between real and synthetic urban dataset observations. (Top two rows) HSI X and X_{DL} (synthetic) at spectral bands 1, 25, 50, and 100. (Bottom two rows) MSI Y and Y_{DL} (synthetic) at bands 2, 4, 8, and 10. The real observations X and Y were simulated from the original-resolution HSI, taken as the reference image Z at its native spatial resolution, using an $\times 8$ Gaussian PSF and Sentinel-2 SRF.

C. Synthetic Dataset Generation by the Dead Leaves Model

To train CALMU, we generate synthetic HSI–MSI pairs (X_{DL}, Y_{DL}) with synthetic GT Z_{DL} using abundances from the low-resolution HSI X and the dead leaves model. This stochastic model generates images through sequential superimposition of random shapes, referred to as leaves, at random positions [44], [47], reproducing key statistical properties of natural images [45], [50], [51]. Leaf shapes follow random geometric models with positions distributed via a stationary Poisson point process over the spatial domain. The process continues until a stationary state is reached, which can be efficiently achieved via perfect simulation techniques [52].

Taking inspiration from our previous work [11], we develop a methodology to generate pairs of synthetic HSI and MSI from the low-resolution HSI image X only. Here, again, our approach builds upon a first HSI unmixing preprocessing step (performed through the original MU), during which we estimate the endmembers and abundances $S_h, A_h = MU(X)$ from the HSI.

Starting from these low-resolution abundances, we construct the high-resolution synthetic abundances A_{DL} by superimposing rectangular leaf-shaped regions according to the dead leaves process. Following the same parameterization as in our previous work [11], each leaf is characterized by its width and height (a, b), orientation θ, center position (x, y), and an abundance vector randomly selected from A_h. The rectangle dimensions a and b are independently sampled from a uniform distribution over $[2r, \min(H, W)/3]$, where r denotes the super-resolution factor, while θ is uniformly drawn in $[0°, 45°]$ and the leaf centers positions (x, y) are uniformly sampled over the spatial domain. Leaves are sequentially superimposed until all pixels have been covered at least once. Each leaf inherits its abundance values from randomly selected vectors of A_h, ensuring consistency with the physical abundance proportions and preserving the non-negativity and sum-to-one constraints. By reusing the endmembers matrix S_h, we then reconstruct the high-resolution synthetic scene as $Z_{DL} = S_h A_{DL}$. The corresponding degraded HSI and MSI observations are subsequently obtained through the application of the spatial and spectral degradation operators: $X_{DL} = Z_{DL}\mathbf{P}^\top$ and $Y_{DL} = \mathbf{R}Z_{DL}$ which are automatically estimated (see Section III-D). Phase 1 of Fig. 1 shows the integration of the synthetic dataset creation into SCALMU's preprocessing step. An overview of this synthetic data generation is illustrated in Fig. 2, visually comparing the real observations X and Y with their synthetic counterparts X_{DL} and Y_{DL}. The detailed procedure is provided in Algorithm 2.

The use of an unmixing step has significant advantages compared to the direct generation of synthetic high-resolution HSIs. First, it reduces computational resources by making the dead leaves generation operate in a low-dimensional abundance space rather than using the full high-dimensional spectra. Second, it ensures sensor independence by decoupling spatial patterns from sensor-specific properties like varying band counts or spectral responses. Finally, it provides a physically interpretable decompositions that preserve non-negativity and sum-to-one constraints over the abundances, enhancing realism and consistency.

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

Algorithm 2 Synthetic Dataset Generation With the Dead Leaves Model

Require: Low-resolution HSI $X \in \mathbb{R}^{B\times hw}$, PSF $\mathbf{P}$, SRF $\mathbf{R}$, target resolution (H, W), number of endmembers M
Output: Synthetic high resolution image Z_{DL}, synthetic HSI X_{DL}, synthetic MSI Y_{DL}

Estimate endmembers and abundances $(S_h, A_h) \leftarrow MU(X)$
Initialize synthetic high-resolution abundances $A_{DL} \leftarrow 0 \in \mathbb{R}^{M\times H\times W}$
Initialize $Mask \leftarrow \emptyset$
while $\#Mask < HW$ **do**
  Randomly draw a, b (rectangle width and height)
  Randomly draw θ (orientation angle)
  Randomly draw (x, y) (center position)
  Sample $Leaf \leftarrow \text{Rect}(a, b, \theta) + (x, y)$
  Extract a material proportion vector $\mathbf{v}_{extract} \leftarrow \text{RandomPixel}(A_h) \in \mathbb{R}^M$
  for $(i, j) \in Leaf \backslash Mask$ **do**
    $A_{DL}[., i, j] \leftarrow \mathbf{v}_{extract}[.]$
  end for
  $Mask \leftarrow Mask \cup Leaf$
end while
Flatten $A_{DL} \in \mathbb{R}^{M\times H\times W}$ to $A_{DL} \in \mathbb{R}^{M\times HW}$
Compute synthetic $Z_{DL} = S_h A_{DL}$
Compute synthetic HSI $X_{DL} = Z_{DL}\mathbf{P}^\top$
Compute synthetic MSI $Y_{DL} = \mathbf{R}Z_{DL}$
return: X_{DL}, Y_{DL}, Z_{DL}

D. Blind Estimation Network

To enable the end-to-end blind use of SCALMU without the prior knowledge of the degradation operators $\mathbf{P}$ and $\mathbf{R}$, we use a blind estimation network that jointly estimates the PSF and SRF from the low-resolution HSI X and high-resolution MSI Y, following the unsupervised degradation estimation principles from recent works such as HyCoNet [8], MIAE [9], and EDIP-Net [37], as illustrated in Phase 1 of Fig. 1. The network parameterizes learnable SRF and PSF kernels via convolution layers without bias, optimizing an L1 loss $\mathcal{L} = \|X_{lr} - Y_{lr}\|_1$ between the spectrally degraded HSI X_{lr} and the spatially blurred-downsampled MSI Y_{lr}, while normalization enforces non-negativity and sum-to-one constraints. The resulting estimates $\mathbf{R}$ and $\mathbf{P}$ are then fed to CALMU and to the synthetic dataset generation process.

IV. Experiments

This section presents a comprehensive evaluation of SCALMU on standard HSI benchmarks and real satellite data. We first detail the used datasets and the experimental setup (Section IV-A). We then compare our approach to SOTA methods (Section IV-B), demonstrate real-data generalization (Section IV-C), and conduct thorough ablation studies on degradation estimation, adaptability, module contributions, and relevance of synthetic data (Section IV-D). We also present a preliminary result highlighting the potential of the proposed approach to generalize across synthetic datasets (Section IV-E), before analyzing its computational efficiency (Section IV-F).

A. Dataset and Setup

We evaluate our proposed method on four HSI datasets.

1) *Urban:* This dataset was captured by the hyperspectral digital image collection experiment (HYDICE) sensor over an urban area near Fort Hood, Texas, USA. It consists of 307 × 307 pixels with 210 spectral bands spanning 400–2500 nm at a spectral resolution of 10 nm. After discarding bands affected by water-vapor absorption and low signal quality, 162 bands are retained for evaluation.
2) *Pavia University (PaviaU):* Acquired by the ROSIS-3 airborne optical sensor in 2003, this dataset contains 610 × 340 pixels with a ground sampling distance of 1.3 m. It covers the 430–840-nm spectral range across 115 bands. After removing 12 bands affected by noise and water-vapor absorption, a 512×256 subregion corresponding to an urban area and containing the remaining 103 bands is used in our experiments.
3) *Chikusei:* This dataset was acquired by the VNIR-C HSI sensor over Chikusei, Japan. The original image consists of 2517 × 2335 pixels with 128 spectral bands covering the 363–1018-nm wavelength range. For our experiments, a 320 × 320 pixel subregion is randomly selected to represent a mixture of agricultural and urban areas.
4) *PRISMA-Paris:* The PRISMA dataset originates from the *Precursore Iperspettrale della Missione Operativa* (PRISMA) mission, launched by the Italian Space Agency in 2019 [53]. Focusing on urban scenes, we use a HSI image acquired over Paris, comprising 256 × 256 pixels and 230 spectral bands spanning the 400–2500-nm range.

In all experiments, we adopt a spatial scale factor of ×8 between the MSI and HSI. The original HSI at its native resolution serves as the reference super-resolved image. The low-resolution HSI X is simulated following Wald's protocol [54] by convolving the reference with a Gaussian PSF of standard deviation $\sigma = 3.5$ (using a 15 × 15 kernel), followed by bicubic downsampling. For the urban, Chikusei, and PRISMA-Paris datasets, the MSI Y is generated using the Sentinel-2A SRF, comprising 12 bands covering 443–2190 nm. For PaviaU, the MSI Y is generated via the Ikonos SRF, which includes four bands spanning 350–1035 nm.

Quantitative performance is assessed using standard HSI image quality metrics: root-mean-square error (RMSE), peak signal-to-noise ratio (PSNR), spectral angle mapper (SAM), Erreur Relative Globale Adimensionnelle de Synthése (ERGAS), and Universal Image Quality Index (UIQI) [6].

Synthetic training was performed using 1000 pairs of HSI–MSI image couples (X_{DL}, Y_{DL}) generated via the dead leaves model (Section III-C). SCALMU (Algorithm 1) was trained end-to-end using the Adam optimizer with learning rate $\eta = 1 \times 10^{-4}$, employing the L1 reconstruction loss $\mathcal{L} = \|Z_{DL} - Z_{out}\|_1$, where Z_{out}, denotes the SCALMU prediction. A grid search was then conducted on the PaviaU dataset over $L_{HSI}, L_{MSI}, L_{Out} \in [1, 10]$ to assess the influence of the unrolling depth on reconstruction performance. To limit the computational cost of this exploration,

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

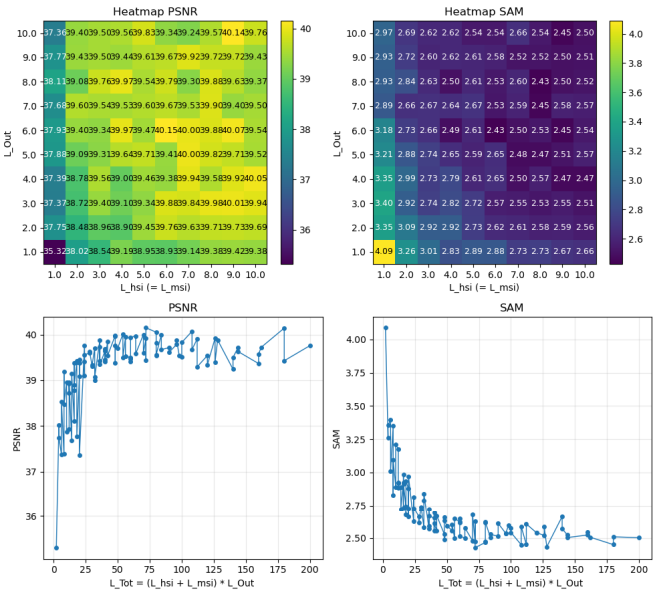


Fig. 3. Grid-search analysis of the unrolling depths on the PaviaU dataset. The top row reports heatmaps of PSNR and SAM as functions of $L_{HS1}(=L_{MS1})$ and L_{Out}, while the bottom row shows the corresponding evolution with respect to the total number of iterations.

the assumption $L_{HS1} = L_{MS1}$ was adopted. The detailed results are provided in Fig. 3, from which the configuration $L_{HS1} = L_{MS1} = L_{Out} = 6$ was retained as the best compromise between convergence speed and reconstruction quality.

B. Comparison With the SOTA Methods

We compare our method SCALMU, trained exclusively on synthetic data, against SOTA HSI–MSI fusion methods. Specifically, we evaluate against the classical CNMF baseline [7] and six SOTA unsupervised deep learning methods: HyCoNet [8], MIAE [9], EU2ADL [29], FeafusFormer [35], OTIAS [38], and EDIP-Net [37]. All competing methods were tested using the default configurations of the provided codes. While CNMF requires the exact Gaussian PSF ($\sigma = 3.5$, 15×15 kernel) and corresponding SRF provided during testing, SCALMU and all competing deep learning methods (HyCoNet, MIAE, EU2ADL, FeafusFormer, OTIAS, and EDIP-Net) operate in a blind configuration without access to degradation parameters, representing their realistic performance in practical scenarios.

Tables I–IV report the quantitative results on the urban, PaviaU, Chikusei, and PRISMA-Paris datasets at $\times 8$ scale factor. Across all datasets, SCALMU significantly surpasses the classical CNMF baseline, although CNMF uses the nonavailable in practice GT **P** and **S**. SCALMU also outperforms all its blind deep learning competitors. These results across diverse urban scenes robustly validate our synthetic-training strategy and unrolled CNMF architecture, achieving SOTA performance without requiring degradation model knowledge at test time. Visual comparisons in Figs. 4–7 confirm SCALMU's advantages through sharper details, better spectral fidelity, and reduced artifacts in error maps.

TABLE I

QUANTITATIVE EVALUATION ON **URBAN** DATASET ($\times 8$) COMPARING CNMF, HYCONET, MIAE, EU2ADL, FEAFUSFORMER, OTIAS, EDIP-NET, AND SCALMU (OURS). BEST/SECOND-BEST RESULTS IN BOLD/UNDERLINE

Method	RMSE ↓	PSNR ↑	SAM ↓	ERGAS ↓	UIQI ↑
CNMF	0.0126	37.99	3.30	1.10	0.9918
HyCoNet	0.0101	39.93	2.52	0.84	0.9941
MIAE	0.0089	41.03	2.19	1.21	0.9958
EU2ADL	0.0088	41.09	2.29	0.85	0.9959
FeafusFormer	0.0096	40.37	2.78	0.85	0.9946
OTIAS	0.0093	40.63	3.00	1.00	0.9944
EDIP-Net	0.0089	41.03	2.90	1.09	0.9950
SCALMU	**0.0084**	**41.49**	**2.17**	**0.82**	**0.9962**

TABLE II

QUANTITATIVE EVALUATION ON **PAVIAU** DATASET ($\times 8$) COMPARING CNMF, HYCONET, MIAE, EU2ADL, FEAFUSFORMER, OTIAS, EDIP-NET, AND SCALMU (OURS). BEST/SECOND-BEST RESULTS IN BOLD/UNDERLINE

Method	RMSE ↓	PSNR ↑	SAM ↓	ERGAS ↓	UIQI ↑
CNMF	0.0149	36.55	3.63	1.36	0.9854
HyCoNet	0.0138	37.17	3.40	0.95	0.9879
MIAE	0.0109	39.27	2.66	1.14	0.9912
EU2ADL	0.0115	38.76	3.05	1.06	0.9910
FeafusFormer	0.0122	38.24	3.14	1.12	0.9906
OTIAS	0.0112	38.98	2.93	0.96	0.9912
EDIP-Net	0.0108	39.32	2.81	0.94	0.9922
SCALMU	**0.0098**	**40.15**	**2.43**	**0.93**	**0.9930**

TABLE III

QUANTITATIVE EVALUATION ON **CHIKUSEI** DATASET ($\times 8$) COMPARING CNMF, HYCONET, MIAE, EU2ADL, FEAFUSFORMER, OTIAS, EDIP-NET, AND SCALMU (OURS). BEST/SECOND-BEST RESULTS IN BOLD/UNDERLINE

Method	RMSE ↓	PSNR ↑	SAM ↓	ERGAS ↓	UIQI ↑
CNMF	0.0063	44.05	3.70	0.89	0.9784
HyCoNet	0.0059	44.61	3.07	0.56	0.9798
MIAE	0.0041	47.79	2.75	0.63	0.9826
EU2ADL	0.0045	46.90	3.13	0.68	0.9814
FeafusFormer	0.0048	46.35	2.93	0.80	0.9823
OTIAS	0.0051	45.79	3.40	0.82	0.9844
EDIP-Net	0.0046	46.66	3.27	0.78	0.9837
SCALMU	**0.0040**	**48.03**	**2.65**	**0.56**	**0.9827**

TABLE IV

QUANTITATIVE EVALUATION ON **PRISMA-PARIS** DATASET ($\times 8$) COMPARING CNMF, HYCONET, MIAE, EU2ADL, FEAFUSFORMER, OTIAS, EDIP-NET, AND SCALMU (OURS). BEST/SECOND-BEST RESULTS IN BOLD/UNDERLINE

Method	RMSE ↓	PSNR ↑	SAM ↓	ERGAS ↓	UIQI ↑
CNMF	0.0123	38.22	5.51	0.97	0.8718
HyCoNet	0.0090	40.90	4.69	0.77	0.8785
MIAE	0.0072	42.90	4.87	0.58	0.8814
EU2ADL	0.0073	42.68	4.82	0.57	0.8820
FeafusFormer	0.0102	39.86	5.60	0.78	0.8699
OTIAS	0.0080	41.90	5.34	0.62	0.8803
EDIP-Net	0.0075	42.48	5.16	0.60	0.8818
SCALMU	**0.0069**	**43.24**	**4.42**	**0.52**	**0.8822**

C. Evaluation on Real Data

In this section, we evaluate SCALMU on a real-world dataset and compare it with two high-performing SOTA

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

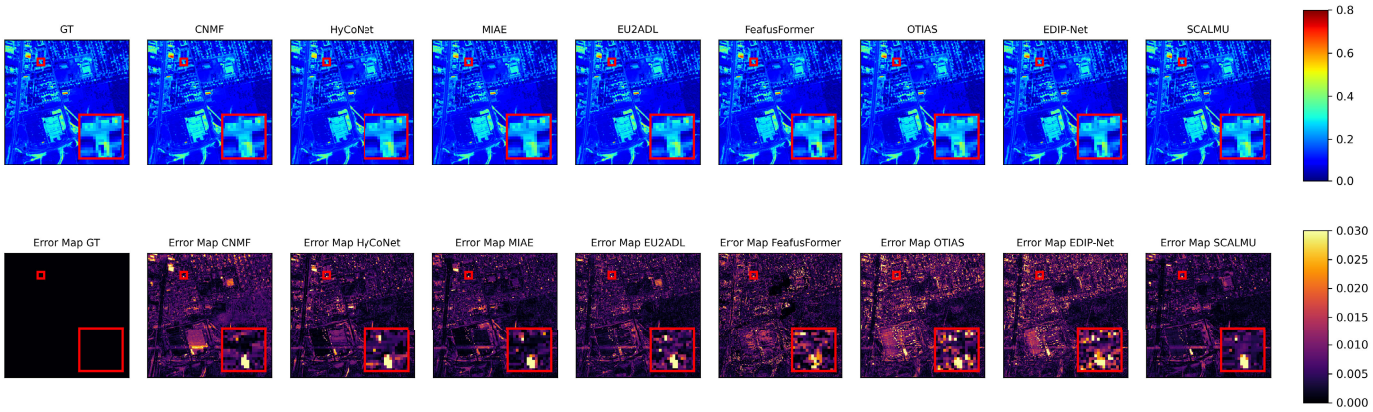


Fig. 4. Visual comparison on **urban** dataset (band 50). **Line 1**: reconstructed HSI images (GT, CNMF, HyCoNet, MIAE, EU2ADL, FeafusFormer, OTIAS, EDIP-Net, and SCALMU). **Line 2**: corresponding absolute error maps $|SR - GT|$. Red rectangles indicate regions with a zoomed-in view in the insets.

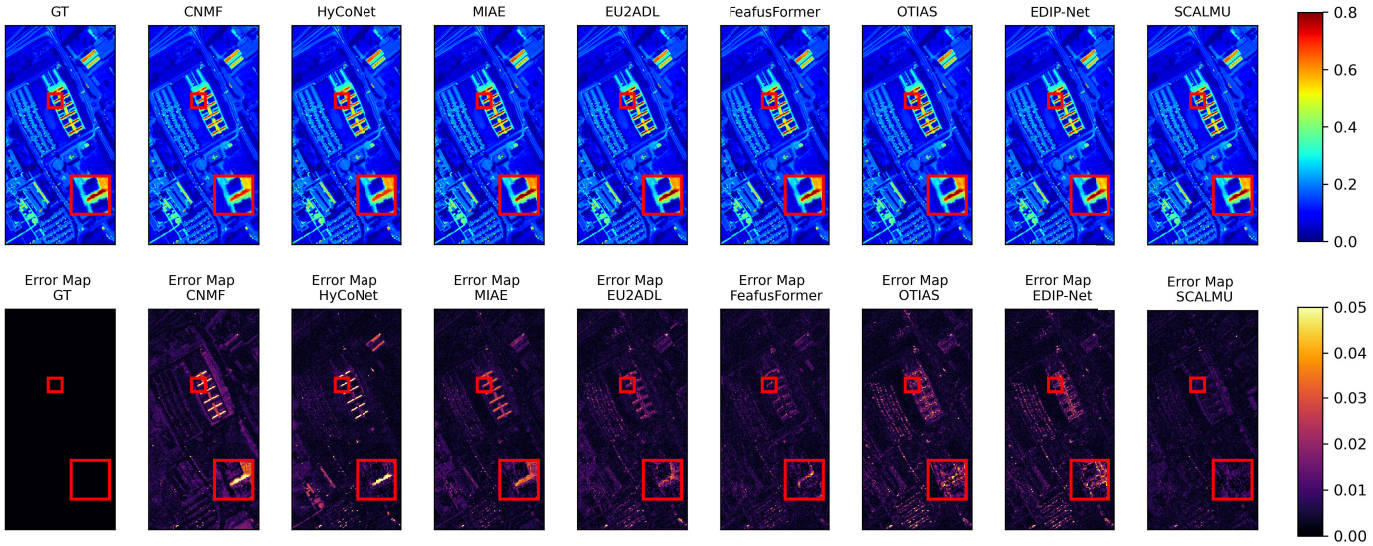


Fig. 5. Visual comparison on **PaviaU** dataset (band 20). **Line 1**: reconstructed HSI images (GT, CNMF, HyCoNet, MIAE, EU2ADL, FeafusFormer, OTIAS, EDIP-Net, and SCALMU). **Line 2**: corresponding absolute error maps $|SR - GT|$. Red rectangles indicate regions with a zoomed-in view in the insets.

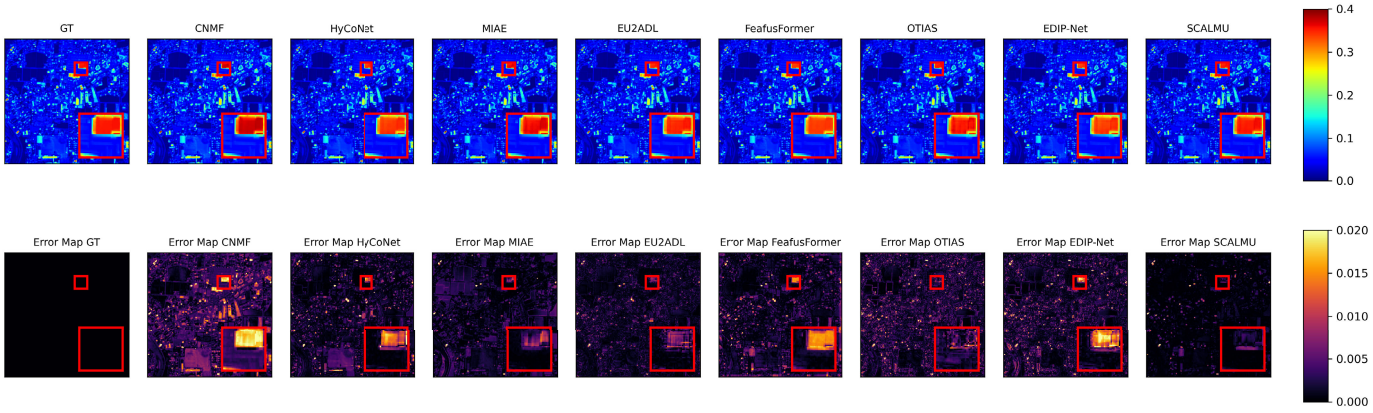


Fig. 6. Visual comparison on **chikusei** dataset (band 20). **Line 1**: reconstructed HSI images (GT, CNMF, HyCoNet, MIAE, EU2ADL, FeafusFormer, OTIAS, EDIP-Net, and SCALMU). **Line 2**: corresponding absolute error maps $|SR - GT|$. Red rectangles indicate regions with a zoomed-in view in the insets.

methods previously considered in the simulated experiments, namely, MIAE and EDIP-Net. This experiment aims to assess the practical effectiveness of the different fusion approaches in a realistic scenario, without relying on simulated degradation models. The dataset was acquired by the Ziyuan-1 02D satellite and includes HSI, MSI, and PAN images [55].

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

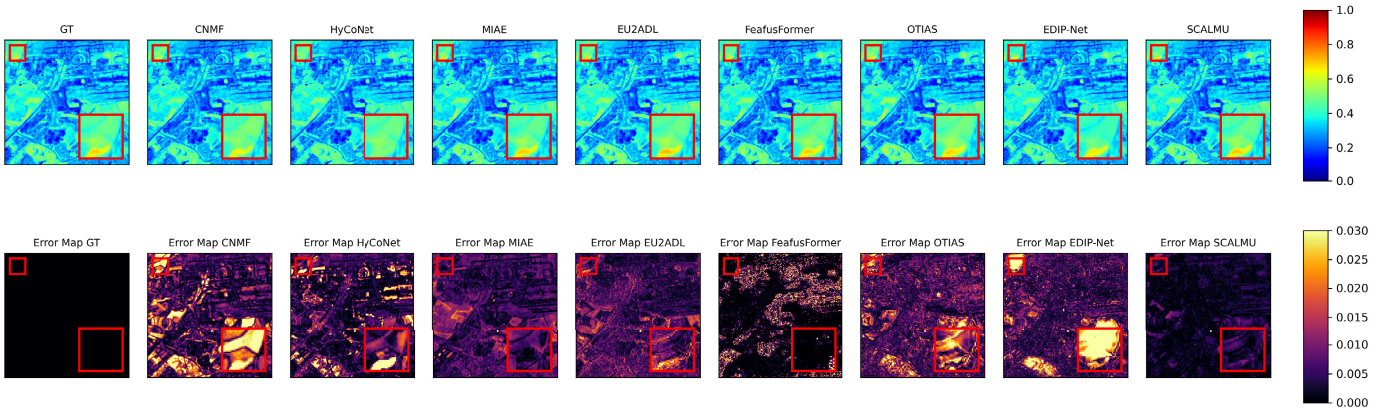


Fig. 7. Visual comparison on **PRISMA-Paris** dataset (band 50). **Line 1**: reconstructed HSI images (GT, CNMF, HyCoNet, MIAE, EU2ADL, FeafusFormer, OTIAS, EDIP-Net, and SCALMU). **Line 2**: corresponding absolute error maps $|SR - GT|$. Red rectangles indicate regions with a zoomed-in view in the insets.

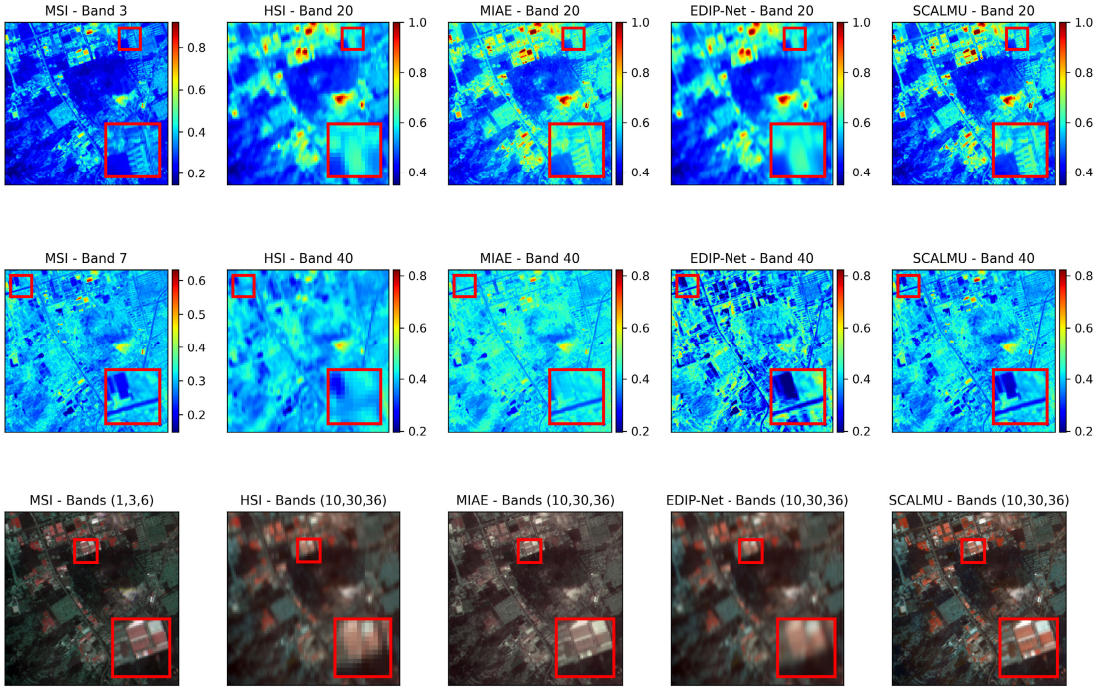


Fig. 8. Visual comparison on the real **Ziyuan-1 02D** dataset ($\times 3$). Column 1 (MSI): bands 3 and 7, and pseudo-color RGB (1, 3, 6). Column 2 (HSI): bands 21 and 40, and pseudo-color RGB (10, 30, 36). Columns 3–5 (SRI): the corresponding fused super-resolved HSI images obtained by MIAE, EDIP-Net, and the proposed SCALMU method.

In our work, we focus exclusively on the HSI and MSI modalities. The HSI has a native spatial resolution of 30 m (76 bands retained after removing noisy bands), while the MSI offers 10-m resolution across eight bands. For experiments, we select a 300×300 urban scene region from the MSI, with its corresponding 100×100 HSI patch ($\times 3$ scale factor).

Since no high-resolution HSI GT is available for real data, the evaluation relies primarily on no-reference quality assessment metrics. In particular, we use the quality with no reference (QNR) index and its associated spectral and spatial distortion components, D_λ and D_S, respectively, following [56]. For completeness, we also report consistency measures in the HSI and MSI domains through PSNR_{HSI} and PSNR_{MSI}, which can be computed by reprojecting the fused image into the observed domains using the degradation operators **P** and **R** estimated by each method.

TABLE V
QUANTITATIVE EVALUATION ON **ZIYUAN-1 02D** DATASET ($\times 3$) COMPARING MIAE, EDIP-NET, AND SCALMU (OURS). BEST RESULTS IN BOLD

Method	$QNR \uparrow$	$D_\lambda \downarrow$	$D_S \downarrow$	$PSNR_{HSI} \uparrow$	$PSNR_{MSI} \uparrow$
MIAE	0.9483	0.0333	0.0190	34.87	35.33
EDIP-Net	0.9236	**0.0310**	0.0468	34.96	32.57
SCALMU	**0.9578**	0.0323	**0.0102**	**35.00**	**37.81**

As shown in Table V and Fig. 8, SCALMU delivers the best overall performance on the real Ziyuan-1 02D dataset, both quantitatively and visually, achieving the best trade-

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
QUANTITATIVE EVALUATION OF SCALMU (BLIND) AND (NONBLIND) ON **URBAN** DATASET (×8). BEST RESULTS IN BOLD

Method	RMSE ↓	PSNR ↑	SAM ↓	ERGAS ↓	UIQI ↑
SCALMU(non-Blind)	**0.0083**	**41.52**	**2.16**	**0.81**	**0.9963**
SCALMU (Blind)	0.0084	41.49	2.17	0.82	0.9962

off between spatial reconstruction and spectral preservation. Although EDIP-Net achieves the lowest D_λ, its significantly higher D_S reveals weaker spatial reconstruction, which is also apparent in the visual results. In contrast, MIAE better preserves spatial structures, but at the expense of spectral fidelity. Visually, the super-resolved images produced by SCALMU effectively combine the sharp spatial details from the MSI with the rich spectral information from the HSI, yielding convincing fusion results on this real dataset. Fine urban structures are clearly reconstructed, while spectral signatures remain consistent across the displayed bands and pseudo-color images. Overall, these quantitative and visual comparisons confirm that MIAE tends to emphasize spatial structures at the expense of spectral consistency, whereas EDIP-Net favors spectral fidelity but reconstructs spatial details less effectively. SCALMU provides the best balance between these two aspects, demonstrating its practical relevance for real-world HSI image super-resolution fusion tasks.

D. Ablation Studies

We conduct ablation studies on four key aspects in this section: degradation estimation (Section IV-D1), comparing GT versus estimated PSF/SRF; adaptability of the learned updates (Section IV-D2), comparing adaptive SCALMU with nonadaptive NALMU; contribution of each adaptive module (Section IV-D3), analyzing abundance and spectral adaptive modules for HSI and MSI; and relevance of synthetic dead leaves training data (Section IV-D4), comparing synthetic versus real PRISMA training.

1) Degradation Estimation: As described in Section III-D, we designed a degradation estimation network to jointly estimate the PSF and SRF by minimizing the spectral and spatial low-resolution inconsistencies between HSI and MSI. When used upstream of SCALMU, we show that this network produces sufficiently accurate PSF and SRF estimates for our fusion task.

Figs. 9 and 10 illustrate both the visual comparison and the quantitative estimation errors between the GT and estimated PSF and SRF on the urban and PaviaU datasets (×8), respectively. On the urban dataset, we further conducted a quantitative ablation study comparing SCALMU using GT degradation parameters (SCALMU nonblind) versus estimated parameters (SCALMU blind), as reported in Table VI. The results show nearly identical performances, with only marginal degradation, confirming the effectiveness of our estimator for practical blind fusion scenarios.

2) Adaptability Analysis: Unlike nonadaptive unrolled methods using fixed learned parameters, SCALMU uses

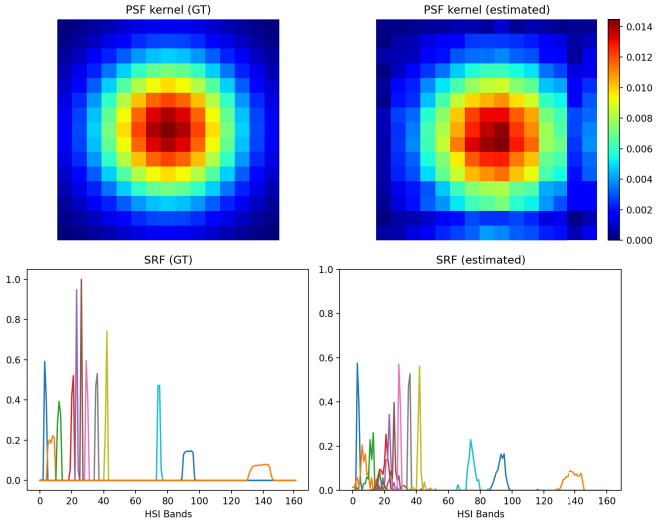


Fig. 9. Visual comparison between real and estimated PSF and SRF on **urban** dataset (×8). The corresponding quantitative estimation errors are PSF RMSE = 0.0011, SRF RMSE = 0.0207, and SRF SAD = 0.3492 rad.

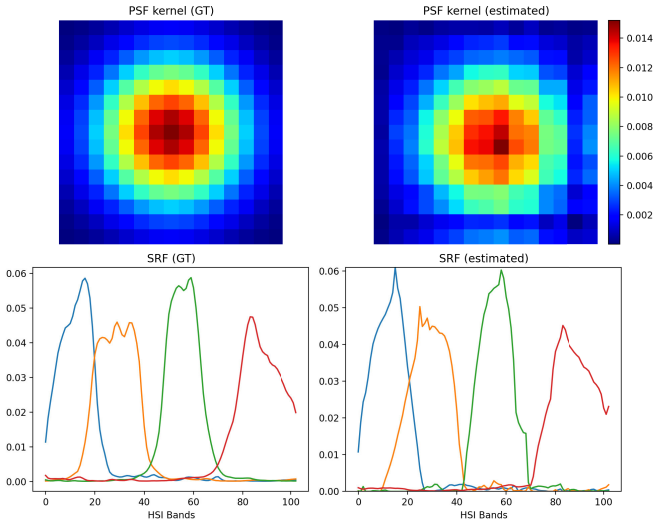


Fig. 10. Visual comparison between real and estimated PSF and SRF on **PaviaU** dataset (×8). The corresponding quantitative estimation errors are PSF RMSE = 0.0010, SRF RMSE = 0.0019, and SRF SAD = 0.0965 rad.

lightweight neural networks $\underline{W}(\cdot)$ that dynamically generate the update matrices from input data characteristics, enabling adaptation to unseen spatial–spectral patterns and degradation profiles. Specifically, SCALMU builds upon RALMU [49], which uses input-dependent neural networks for NMF update acceleration. To isolate the benefits of this adaptive design, we perform an ablation by replacing SCALMU's four adaptive neural modules with the fixed matrices of nonadaptive learned MUs (NALMUs) [57]. We consider NALMU here as a representative baseline for classical unrolling, since its unfolded updates rely on fixed learned parameters rather than input-adaptive operators. In NALMU, the four update factors $\underline{W}_{A_m}, \underline{W}_{A_h}, \underline{W}_{S_m}$, and $\underline{W}_{S_h}$ are directly learned as iteration-dependent parameters, with dimensions matching those of A_m, A_h, S_m, and S_h, respectively. More precisely, for

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

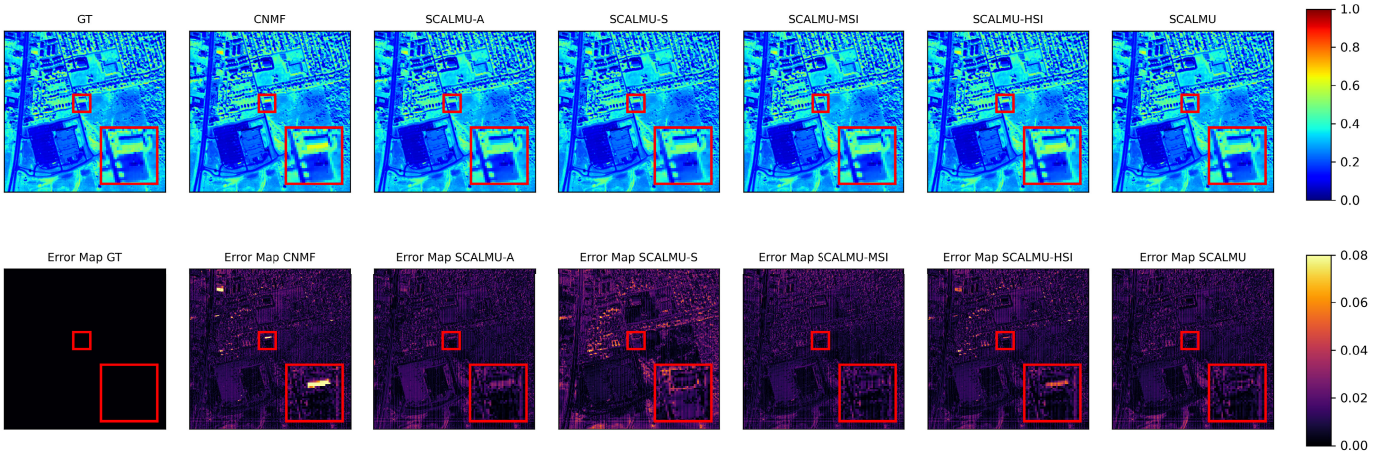


Fig. 11. Visual comparison on **urban** dataset (band 80). **Line 1**: reconstructed HSI images (GT, CNMF, SCALMU-A, SCALMU-S, SCALMU-MSI, SCALMU-HSI, and SCALMU). **Line 2**: corresponding absolute error maps $|SR - GT|$. Red rectangles indicate regions with a zoomed-in view in the insets.

TABLE VII
ABLATION STUDY ON SCALMU ADAPTABILITY (URBAN DATASET, $\times 8$). BEST RESULTS IN BOLD

Method	RMSE ↓	PSNR ↑	SAM ↓	ERGAS ↓	UIQI ↑
SCALMU (NALMU)	0.0093	40.61	2.44	0.89	0.9954
SCALMU	**0.0084**	**41.49**	**2.17**	**0.82**	**0.9962**

each unfolding iteration and for each branch, the model learns one fixed abundance-related matrix and one fixed endmember-related matrix, which are stored as trainable parameters and reused at every forward pass. These matrices are initialized with ones, optimized jointly with the other model parameters during training, and constrained to remain positive through the clamping operations used in the MUs. At test time, they are expanded along the batch dimension and applied identically to all samples. Therefore, unlike SCALMU, NALMU relies on fixed learned update operators that depend only on the training set and not on the specific spatial–spectral content of the input pair. Equivalently, one can view NALMU as learning a set of global multiplicative preconditioners for the unfolded MUs of a given dataset, without adapting to the spatial structures or degradation characteristics of individual HSI–MSI pairs.

Table VII compares these variants on the urban dataset ($\times 8$). SCALMU consistently outperforms NALMU across all metrics, confirming that our adaptive design delivers superior train set/test set generalization capacity and reconstruction quality compared to fixed-parameter unrolled approaches.

3) Model Analysis: To evaluate the contribution of each adaptive module $\underline{W(\cdot)}$ in SCALMU, we conduct a comprehensive ablation study systematically removing components while keeping the unrolled NMF backbone fixed (Table VIII, urban dataset, $\times 8$).

CNMF uses only the classical fixed unrolling without any learned modules, serving as our baseline. SCALMU-A activates abundance-adaptive modules $\underline{W_{A_m}(\cdot)}$ (MSI) and $\underline{W_{A_h}(\cdot)}$ (HSI). Spectral module (SCALMU-S) enables spectral-adaptive modules $\underline{W_{S_m}(\cdot)}$ (MSI) and $\underline{W_{S_h}(\cdot)}$ (HSI). SCALMU-MSI combines both MSI modules $[\underline{W_{A_m}(\cdot)}, \underline{W_{S_m}(\cdot)}]$. SCALMU-HSI uses both HSI modules $[\underline{W_{A_h}(\cdot)}, \underline{W_{S_h}(\cdot)}]$. Full SCALMU integrates all four modules.

The results in Table VIII and Fig. 11 reveal distinct contributions. Removing MSI abundance-adaptive modules $\underline{W_{A_m}(\cdot)}$ causes the most substantial performance degradation, as expected: these modules effectively help extract high-resolution spatial details from MSI essential for $\times 8$ super-resolution. SCALMU-S $[\underline{W_{S_m}(\cdot)}, \underline{W_{S_h}(\cdot)}]$ provides consistent improvements but has a lesser overall impact, primarily enhancing spectral fidelity rather than spatial reconstruction. Notably, as observed in the error maps (Fig. 11), relying solely on SCALMU-S fails to accurately recover structural information, resulting in pronounced errors particularly around the edges and boundaries of structures. MSI-focused modules yield larger gains than HSI ones, reflecting MSI's critical role in providing high-resolution spatial guidance. Only the full SCALMU configuration achieves optimal synergistic performance across all metrics, validating our modular adaptive design.

4) Relevance of Synthetic Data: To validate the effectiveness of our dead leaves synthetic training set, which is generated relying exclusively on the marginal distributions estimated from the PRISMA-Paris scene, we created a real training dataset from PRISMA HSI images over multiple cities (Athens, Copenhagen, Dubai, and Rome), keeping the Paris scene as the test set. We extracted 256×256 crops across all 230 spectral bands from the other cities, generating 1000 reference images. Applying the same PSF $\mathbf{P}$ and SRF $\mathbf{R}$ as PRISMA-Paris ensures identical degradation conditions and dataset size ($N = 1000$) as our synthetic dead leaves training set, enabling fair comparison.

We trained SCALMU separately on both datasets and evaluated on PRISMA-Paris ($\times 8$), as shown in Table IX. The real PRISMA dataset yields slightly superior performance as expected, but the gap with the dead leaves synthetic training remains marginal across all metrics. This demonstrates the remarkable effectiveness of our proposed synthetic data generation process, achieving near-real-data performance while eliminating dependency on scarce GT pairs. It should be noted

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

TABLE VIII

ABLATION ANALYSIS OF THE PROPOSED SCALMU WITH DIFFERENT MODULE COMBINATIONS ON **URBAN** DATASET ($\times 8$). BEST RESULTS IN BOLD

Method	$W_{A_m}(\cdot)$	$W_{S_m}(\cdot)$	$W_{A_h}(\cdot)$	$W_{S_h}(\cdot)$	RMSE ↓	PSNR ↑	SAM ↓	ERGAS ↓	UIQI ↑
CNMF	✗	✗	✗	✗	0.0145	36.79	3.84	1.32	0.9877
SCALMU-A	✓	✗	✓	✗	0.0102	39.85	2.79	0.94	0.9939
SCALMU-S	✗	✓	✗	✓	0.0141	37.02	3.39	1.30	0.9890
SCALMU-MSI	✓	✓	✗	✗	0.0094	40.58	2.73	0.92	0.9948
SCALMU-HSI	✗	✗	✓	✓	0.0111	39.11	2.79	0.94	0.9927
SCALMU	✓	✓	✓	✓	**0.0084**	**41.49**	**2.17**	**0.82**	**0.9962**

TABLE IX

QUANTITATIVE EVALUATION OF SCALMU TRAINED ON THE SYNTHETIC **DEAD LEAVES** DATASET AND ON THE REAL **PRISMA** DATASET, TESTED ON THE **PRISMA-PARIS** SCENE ($\times 8$). THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Training dataset	RMSE ↓	PSNR ↑	SAM ↓	ERGAS ↓	UIQI ↑
dead leaves	0.0069	43.24	4.42	0.52	0.8822
PRISMA	**0.0067**	**43.52**	**4.27**	**0.51**	**0.8823**

TABLE X

QUANTITATIVE EVALUATION OF SCALMU TRAINED ON THE SYNTHETIC **DEAD LEAVES** DATASET AND ON THE REAL **PRISMA** DATASET, TESTED ON THE INTERPOLATED **URBAN** SCENE ($\times 8$). THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Train. dataset	reestim. PR	RMSE ↓	PSNR ↑	SAM ↓	ERGAS ↓	UIQI ↑
dead leaves	✗	**0.0118**	**38.58**	**2.47**	**0.98**	**0.9939**
PRISMA	✗	0.0139	37.16	3.55	1.15	0.9891
dead leaves	✓	**0.0112**	**39.03**	**2.38**	**0.96**	**0.9945**
PRISMA	✓	0.0132	37.62	3.25	1.09	0.9910

TABLE XI

COMPUTATIONAL COST FOR DIFFERENT LEARNING-BASED FUSION METHODS ON THE URBAN DATASET (SCALE $\times 8$)

Method	Dataset Gen.	#Params	Running time*
HyCoNet	–	400K	45min
MIAE	–	200K	20min
EU2ADL	–	1M	53min
FeafusFormer	–	38M	1h10 | 4s
OTIAS	–	3M	1h50 | 4s
EDIP-Net	–	12M	1h40
SCALMU	11min	2M	5h40 | 3s

**If 2 running times are given as $t_1 \mid t_2$, t_1 corresponds to training time and t_2 to inference time. Otherwise, the time corresponds to the processing time needed to fuse one image pair.*

that this ablation study specifically aims to demonstrate the practical value of synthetic datasets. Indeed, assuming access to a substantial number of real HSI–MSI pairs along with their corresponding high-resolution GT is highly unrealistic in real-world remote sensing scenarios. Consequently, the performance achieved with the real PRISMA training dataset should be regarded as a theoretical upper bound rather than a practically attainable baseline.

E. Preliminary Study of Cross-Sensor Generalization Capacity

To further investigate the cross-sensor generalization capacity of synthetic images compared to real ones, we conducted a preliminary experiment by directly reusing the two SCALMU models trained in Section IV-D4. These models, trained, respectively, on the synthetic dead leaves dataset *generated from PRISMA-Paris abundance marginals* and on the real PRISMA dataset (acquired over multiple cities by PRISMA mission), are tested on the URBAN HSI scene acquired by the HYDICE sensor. HYDICE captures 210 spectral bands within the 400–2500-nm range without spectral overlap, which is comparable to PRISMA's 230 nonoverlapping bands over the same spectral interval. For consistency, the URBAN dataset was linearly interpolated from 210 to 230 bands, degraded using the same PSF **P** and SRF **R**, and used as input for both trained models. The fused outputs were then reinterpolated back to 210 bands, and noisy bands were removed to evaluate the result against the standard 162 bands URBAN reference.

As summarized in Table X, the SCALMU model trained on synthetic dead leaves data outperforms its PRISMA-trained counterpart across all metrics. This holds true regardless of whether the sensor's PSF and SRF are reestimated or directly reused from the PRISMA estimation. These preliminary results suggest that while real-data training may lead to a very good fit to specific sensor characteristics, synthetic supervision exhibits superior cross-sensor generalization. In particular, it appears to improve robustness to differences in spectral responses and modality gaps between heterogeneous HSI and MSI systems. The role of noise robustness may also be a limiting factor in such transfer settings, but it is not the primary focus of this study and will be investigated in future work. Although further extensive validation is required, these early findings highlight the strong potential of synthetic datasets as a robust, sensor-robust training paradigm.

F. Computational Efficiency

Table XI presents the computational costs (dataset generation, parameter count, and running times on V100 GPU) for learning-based fusion methods on the urban dataset ($\times 8$). The "Dataset Gen." column indicates the time required to generate the 1000 synthetic training images used by SCALMU, which are produced using the dead leaves model. Despite longer initial training for its end-to-end synthetic learning, it achieves fast inference per image pair, matching OTIAS while using fewer parameters and delivering better reconstruction quality, whereas single-time autoencoder methods require lengthy retraining per pair, which prevents generalization; SCALMU's one-time training, thus, might enable rapid deployment across datasets with minimal inference overhead and optimal parameter efficiency.

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

V. Conclusion

This article introduces SCALMU, a novel blind HSI–MSI fusion framework that unrolls CNMF into adaptive learned layers. Its core CALMU structure augments classical MUs with four lightweight neural modules that dynamically predict input-dependent learned matrices for MSI/HSI abundances and spectra, preserving CNMF's physical interpretability while enabling data-driven refinement. A lightweight degradation estimation subnetwork jointly predicts PSF and SRF from input pairs during preprocessing. Trained exclusively on dead leaves synthetic data simulating realistic spatial–spectral statistics, SCALMU eliminates paired real-data requirements. It achieves SOTA performance on several HSI datasets at $\times 8$ scale across all HSI quality metrics, surpassing classical and deep learning methods while generalizing robustly to real-world data. Future work will focus on extending SCALMU's generalization across heterogeneous sensors through the development of a unified fusion framework that remains robust and consistent under diverse sensor characteristics and imaging conditions.

References

[1] M. Fahes, C. Kervazo, J. Bobin, and F. Tupin, "Unrolling PALM for sparse semi-blind source separation," in *Proc. Int. Conf. Learn. Represent.*, 2021.

[2] X. Wu, D. Hong, and J. Chanussot, "UIU-Net: U-Net in U-Net for infrared small object detection," *IEEE Trans. Image Process.*, vol. 32, pp. 364–376, 2023.

[3] S. J. Mills et al., "Evaluation of aerial remote sensing techniques for vegetation management in power-line corridors," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3379–3390, Sep. 2010.

[4] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, "Extended vision transformer (ExViT) for land use and land cover classification: A multimodal deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5514415.

[5] B. Lu, P. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, p. 2659, Aug. 2020.

[6] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.

[7] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[8] K. Zheng et al., "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2487–2502, Mar. 2021.

[9] J. Liu, Z. Wu, L. Xiao, and X.-J. Wu, "Model inspired autoencoder for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5522412.

[10] R. Achddou, Y. Gousseau, and S. Ladjal, "Fully synthetic training for image restoration tasks," *Comput. Vis. Image Understand.*, vol. 233, Aug. 2023, Art. no. 103723.

[11] X. Xu, Y. Gousseau, C. Kervazo, and S. Ladjal, "Synthetic abundance maps for unsupervised super-resolution of hyperspectral remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 19, pp. 13507–13520, 2026.

[12] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999.

[13] M. T. Eismann and R. C. Hardie, "Application of the stochastic mixing model to hyperspectral resolution enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 9, pp. 1924–1933, Sep. 2004.

[14] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, Sep. 2004.

[15] M. T. Eismann and R. C. Hardie, "Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 455–465, Mar. 2005.

[16] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 2329–2336.

[17] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 63–78.

[18] X. Fu, H. Liang, and S. Jia, "Mixed noise-oriented hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5526916.

[19] X. Fu, S. Jia, M. Xu, J. Zhou, and Q. Li, "Fusion of hyperspectral and multispectral images accounting for localized inter-image changes," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517218.

[20] C. Wang, Y. Liu, X. Bai, W. Tang, P. Lei, and J. Zhou, "Deep residual convolutional neural network for hyperspectral image super-resolution," in *Proc. Int. Conf. Image Graph.*, 2017, pp. 370–380.

[21] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by MS/HS fusion net," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1585–1594.

[22] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3070–3079.

[23] X.-H. Han, B. Shi, and Y. Zheng, "SSF-CNN: Spatial and spectral fusion with CNN for hyperspectral image super-resolution," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2506–2510.

[24] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by CNN denoiser," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1124–1135, Mar. 2021.

[25] R. Ran, L.-J. Deng, T.-X. Jiang, J.-F. Hu, J. Chanussot, and G. Vivone, "GuidedNet: A general CNN fusion framework via high-resolution guidance for hyperspectral image super-resolution," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4148–4161, Jul. 2023.

[26] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse Dirichlet-Net for hyperspectral image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2511–2520.

[27] J. Li, K. Zheng, J. Yao, L. Gao, and D. Hong, "Deep unsupervised blind hyperspectral and multispectral data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[28] J. Li, Y. Li, C. Wang, X. Ye, and W. Heidrich, "BUSIFusion: Blind unsupervised single image fusion of hyperspectral and RGB images," *IEEE Trans. Comput. Imag.*, vol. 9, pp. 94–105, 2023.

[29] L. Gao, J. Li, K. Zheng, and X. Jia, "Enhanced autoencoders with attention-embedded degradation learning for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5509417.

[30] S. Li, Y. Tian, C. Wang, H. Wu, and S. Zheng, "Hyperspectral image super-resolution network based on cross-scale nonlocal attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509615.

[31] D. Hong, J. Yao, C. Li, D. Meng, N. Yokoya, and J. Chanussot, "Decoupled-and-coupled networks: Self-supervised hyperspectral image super-resolution with subpixel fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5527812.

[32] L. Zhang, J. Nie, W. Wei, and Y. Zhang, "Unsupervised test-time adaptation learning for effective hyperspectral image super-resolution with unknown degeneration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 7, pp. 5008–5025, Jul. 2024.

[33] Q. Li, M. Gong, Y. Yuan, and Q. Wang, "Symmetrical feature propagation network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5536912.

[34] J.-F. Hu, T.-Z. Huang, L.-J. Deng, H.-X. Dou, D. Hong, and G. Vivone, "Fusformer: A transformer-based fusion network for hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[35] X. Cao, Y. Lian, K. Wang, C. Ma, and X. Xu, "Unsupervised hybrid network of transformer and CNN for blind hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5507615.

[36] C. Wu, D. Wang, Y. Bai, H. Mao, Y. Li, and Q. Shen, "HSR-diff: Hyperspectral image super-resolution via conditional diffusion models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 7060–7070.

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.

[37] J. Li, K. Zheng, L. Gao, Z. Han, Z. Li, and J. Chanussot, "Enhanced deep image prior for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5504218.
[38] S. Deng, J. Ma, L.-J. Deng, and P. Wei, "OTIAS: OcTree implicit adaptive sampling for multispectral and hyperspectral image fusion," in *Proc. AAAI Conf. Artif. Intell.*, 2025, vol. 39, no. 3, pp. 2708–2716.
[39] A. Dosovitskiy et al., "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2758–2766.
[40] Y. Tian, L. Fan, P. Isola, H. Chang, and D. Krishnan, "StableRep: Synthetic images from text-to-image models make strong visual representation learners," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, pp. 48382–48402.
[41] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vols. PAMI–5, no. 1, pp. 25–39, Jan. 1983.
[42] D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis," in *Proc. 22nd Int. ACM Conf. Comput. Graph. Interact. Techn.*, 1995, pp. 229–238.
[43] B. Galerne, Y. Gousseau, and J.-M. Morel, "Micro-texture synthesis by phase randomization," *Image Process. Line*, vol. 1, pp. 213–237, Sep. 2011.
[44] G. Matheron, "Modele séquentiel de partition aléatoire," Centre de Morphologie Mathématique, École des Mines de Paris, Fontainebleau, France, Tech. Rep., 1968.
[45] L. Alvarez, Y. Gousseau, and J.-M. Morel, "The size of objects in natural and artificial images," in *Advances in Imaging and Electron Physics*, vol. 111. Amsterdam, The Netherlands: Elsevier, 1999, pp. 167–242.
[46] Y. Gousseau and F. Roueff, "The dead leaves model: General results and limits at small scales," 2003, *arXiv:math/0312035*.
[47] C. Bordenave, Y. Gousseau, and F. Roueff, "The dead leaves model: A general tessellation modeling occlusion," *Adv. Appl. Probab.*, vol. 38, no. 1, pp. 31–46, Mar. 2006.
[48] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
[49] C. Kervazo and J. E. Cohen, "Unrolled multiplicative updates for nonnegative matrix factorization applied to hyperspectral unmixing," Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France, Tech. Rep., 2026.
[50] A. B. Lee, D. Mumford, and J. Huang, "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model," *Int. J. Comput. Vis.*, vol. 41, nos. 1–2, pp. 35–59, Jan. 2001.
[51] Y. Gousseau and F. Roueff, "Modeling occlusion and scaling in natural images," *Multiscale Model. Simul.*, vol. 6, no. 1, pp. 105–134, Jan. 2007.
[52] W. S. Kendall and E. Thönnes, "Perfect simulation in stochastic geometry," *Pattern Recognit.*, vol. 32, no. 9, pp. 1569–1586, Sep. 1999.
[53] S. Cogliati et al., "The PRISMA imaging spectroscopy mission: Overview and first performance analysis," *Remote Sens. Environ.*, vol. 262, Sep. 2021, Art. no. 112499.
[54] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.
[55] S. Li, "Real HSI-MSI-pan image dataset for the hyperspectral/multispectral/panchromatic image fusion and super-resolution fields," 2024, *arXiv:2407.02387*.
[56] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogrammetric Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.
[57] C. Kervazo, A. Chetoui, and J. E. Cohen, "Deep unrolling of the multiplicative updates algorithm for blind source separation, with application to hyperspectral unmixing," in *Proc. 32nd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2024, pp. 656–660.



Xinxin Xu received the Engineering degree from the Institut d'Optique Graduate School, Palaiseau, France, in 2023. He is currently pursuing the Ph.D. degree with LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France, under the supervision of Yann Gousseau, Christophe Kervazo, and Saïd Ladjal.

His research focuses on hyperspectral image super-resolution, with applications in remote sensing.



Yann Gousseau received the B.Eng. degree from the École Centrale de Paris, Châtenay-Malabry, France, the Part III of the Mathematical Tripos degree from the University of Cambridge, Cambridge, U.K., in 1995, and the Ph.D. degree in applied mathematics from the University of Paris-Dauphine, Paris, France, in 2000.

He is currently a Professor with Télécom Paris, Palaiseau, France. His research interests include the mathematical modeling of natural images and textures, generative models, computer vision, image, and video processing.



Christophe Kervazo received the Engineering degree from Supélec, Gif-sur-Yvette, France, in 2015, the Master of Science degree from Georgia Institute of Technology, Atlanta, GA, USA, in 2016, and the Ph.D. degree from CEA Saclay, Gif-sur-Yvette, France, in 2019.

He is currently an Associate Professor at Télécom Paris, Palaiseau, France, where he works on imaging applications. His research focuses on interpretable deep learning, both from the neural network architecture point of view and the reliability of their outputs. His applications include remote sensing and medical imaging.



Saïd Ladjal received the Diplôme de Magister and master's degrees in computer science degree from École Normale Supérieure in 2000, the Engineering degree from Télécom Paris, Palaiseau, France, in 2002, and the Ph.D. degree in applied mathematics from École Normale Supérieure de Cachan, Cachan, France, in 2005.

He is currently a Professor with Télécom Paris. His research interests are in mathematical modeling for images and computational photography with applications to remote sensing, general image restoration, and medical imaging.

Authorized licensed use limited to: Telecom Paris. Downloaded on September 23,2026 at 12:58:13 UTC from IEEE Xplore. Restrictions apply.