ISPRS Journal of Photogrammetry and Remote Sensing 239 (2026) 876–892

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Prior-guided multi-domain mixture-of-experts for multimodal Earth observation data gaps

Chuang Liu [a], Jianhua Guo [b,c], Yingdong Pi [a], Xiao Wu [d], Zhiqi Zhang [e], Ru Chen [a], Xinyi Wang [a], Mi Wang [a,*]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, 430079, China
[b] International Research Center of Big Data for Sustainable Development Goals, Beijing, 100094, China
[c] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100094, China
[d] School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, 611731, China
[e] School of Computer Science, Hubei University of Technology, Wuhan, 430068, China

ARTICLE INFO

Dataset link: https://github.com/JUSTM0VE0N/RSMIF_Project

Keywords:
Earth observation data fusion
Hyperspectral image fusion
Pansharpening
Cloud removal
Mixture of experts
Sparse routing

ABSTRACT

Multimodal Earth observation (EO) often requires integrating sensors with complementary strengths in spatial detail, spectral richness, and weather robustness. Existing task-specific fusion networks typically use EO priors as passive cues, leaving them disconnected from the adaptive allocation of computation across heterogeneous scenes. To address these recurring EO data gaps, we propose DAMoE, a task-instantiated sparse multi-domain mixture-of-experts framework. DAMoE uses a reusable conditional-computation backbone centered on prior-guided sparse routing, where compact frequency-energy, spatial-structure, and spectral–statistical priors are injected into branch-specific routers to activate task-adaptive expert subsets. A frequency-domain front-end enhances wavelet subband representations, while parallel spatial and spectral expert branches refine local structures and bandwise consistency. Diversity and load-balancing regularizers further encourage complementary expert specialization and stable sparse routing. Experiments on five source-domain benchmarks and two transfer test sets show that DAMoE achieves 1%–12% improvement in key metrics (PSNR/SAM/ERGAS) across optical and hyperspectral fusion tasks, maintains competitive performance in SAR-assisted optical reconstruction (PSNR up to 29.78), and retains a favorable accuracy–efficiency trade-off with only 0.39M parameters and 0.013 s inference time (faster than state-of-the-art methods by 20%–50%). Cross-sensor and cross-dataset transfer experiments further evaluate its robustness under distribution shifts. These results indicate that prior-guided sparse routing offers an effective conditional-computation mechanism for adapting fusion pathways to recurring EO data gaps.

1. Introduction

1.1. Background and problem setting

Earth observation (EO) underpins a broad range of applications, including land-cover mapping, environmental monitoring, urban analysis, agricultural assessment, and disaster management (Cai et al., 2022). Yet multimodal EO fusion is ultimately driven by a persistent practical constraint: no single sensor can simultaneously deliver fine spatial detail, rich spectral information, and robust sensing capability under all observation conditions (Li et al., 2024). This limitation has made multimodal remote sensing image fusion a central topic in EO analysis, since it enables complementary information from heterogeneous sensors, such as panchromatic (PAN), multispectral (MSI), hyperspectral (HSI), and synthetic aperture radar (SAR), to be integrated within a common reconstruction process (Jozdani et al., 2022). The benefit of multimodal EO fusion is not limited to producing visually enhanced images (Li et al., 2025). More importantly, it raises a broader methodological question of how to address recurring EO data gaps across different sensing settings within a reusable conditional-computation backbone. In practice, these gaps are typically associated with missing spatial detail, insufficient spectral information, or reduced optical availability under adverse observation conditions. This requires EO fusion models to go beyond fixed reconstruction pipelines and adapt their computational pathways to heterogeneous spatial structures, spectral responses, and sensing conditions.

* Corresponding author.
E-mail address: wangmi@whu.edu.cn (M. Wang).

https://doi.org/10.1016/j.isprsjprs.2026.06.034
Received 7 April 2026; Received in revised form 22 June 2026; Accepted 24 June 2026
Available online 1 July 2026

1.2. Related work on multimodal EO fusion

Earlier studies addressed these challenges mainly through classical fusion paradigms, most notably component substitution (CS), multiresolution analysis (MRA), and variational optimization (VO) (Vivone et al., 2025). In optical–optical fusion, for example, CS (Choi et al., 2011) sharpens spatial content by replacing the spatial component of MSI with that of PAN, although this usually comes with spectral distortion. MRA (Restaino et al., 2016) injects PAN high-frequency information into MSI and thereby offers a more balanced treatment of spatial and spectral content, yet its performance remains tied to linear modeling assumptions that are often restrictive in complex scenes. VO-based methods (Palsson et al., 2014) instead impose spatial and spectral fidelity through explicit regularization, but the resulting optimization process generally incurs higher computational cost (Yokoya et al., 2012). Taken together, these methods remain important because they are interpretable and supported by mature evaluation protocols, even though their capacity becomes limited once multimodal relationships grow strongly nonlinear.

The rise of data-driven models has shifted this landscape substantially, and recent studies have increasingly turned to deep learning (DL) for multimodal EO fusion (Ciotola et al., 2022). As summarized in Fig. 1, existing DL approaches can be broadly organized into two dominant streams. One line emphasizes dense spatial–spectral modeling (*SS modeling*) directly in the image domain, as exemplified by multi-scale dual-stream networks (Liao et al., 2023), spatial–spectral Transformers (L. Chen et al., 2024), and attention-augmented residual frameworks (B. Wang et al., 2024). Multi-scale dual-stream networks enhance local spatial representation (Wang et al., 2025), Transformer-based models improve long-range spatial–spectral dependency modeling (T. Wen et al., 2025), and attention-augmented residual designs further strengthen feature reuse and spectral preservation (Li et al., 2023). However, these methods mainly operate in the image domain and may underexploit frequency-domain detail cues.

A second line introduces frequency-domain cues through spatial–frequency dual-domain guidance or multimodal dual-domain learning (*SF modeling*) (He et al., 2023), which is particularly helpful for preserving edges, textures, and fine structures (Zhou et al., 2022). These frequency-aware methods improve structural detail recovery by explicitly exploiting high-frequency (X. Wen et al., 2025) or wavelet-domain information (Huang et al., 2025). Nevertheless, frequency information is often used as an auxiliary enhancement cue rather than being adaptively coordinated with spatial and spectral modeling. Beyond this two-stream view, some methods in either category further incorporate observation models (Dong et al., 2023) or physically motivated priors (Y. Chen et al., 2024) to improve reconstruction consistency. Such prior-aware designs improve physical consistency by introducing degradation models, sensor-response assumptions, or domain constraints, but the priors are usually used as losses, constraints, or auxiliary inputs rather than as routing signals for adaptive computation (Cao et al., 2022; Wu et al., 2025). This progression suggests that EO fusion is evolving not only toward stronger nonlinear representation learning, but also toward tighter integration between learned models and domain knowledge (Zhu et al., 2023a).

1.3. Research gaps

Taken together, existing studies suggest that multimodal EO fusion is increasingly moving from interpretable classical models toward deep spatial–spectral architectures, frequency-aware representations, prior-informed learning, and sparse conditional computation (Liu et al., 2025b). Even so, several limitations remain unresolved. *A first issue is the lack of explicit coordination across representation domains.* Many current models are still dominated by image-domain spatial or spectral processing, while frequency cues are introduced only weakly or remain auxiliary (Zhu et al., 2023b). The result is a fragmented modeling strategy that makes it difficult to preserve fine spatial structures and spectral fidelity simultaneously in heterogeneous scenes.

A second issue concerns computational efficiency. Stronger reconstruction performance is often obtained by increasing network depth, channel width, or attention complexity (Jiang and Chen, 2025), which raises memory consumption and inference cost (B. Wang et al., 2024). Although mixture-of-experts (MoE) designs offer a plausible path toward sparse conditional computation (Guo et al., 2025), naive gating can repeatedly activate only a small subset of experts and thereby weaken the intended specialization effect.

A third issue lies in the role of EO priors. Existing prior-guided methods typically inject domain knowledge through losses (Cao et al., 2022), constraints (Liu et al., 2025b), or auxiliary inputs (W. Wang et al., 2024), rather than allowing such priors to directly shape routing or expert selection. Consequently, informative EO cues, including wavelet-energy statistics, texture descriptors, and spectral indices, are still not fully exploited as part of adaptive computation.

Against this background, we propose DAMoE, a task-instantiated sparse multi-domain MoE framework for recurring multimodal EO data gaps. The central idea is to make EO priors participate directly in conditional computation: instead of using frequency, spatial, and spectral cues only as auxiliary constraints, DAMoE injects compact domain priors into branch-specific routers to determine which experts should be activated for each input. This prior-guided routing is implemented within a frequency–spatial–spectral sparse MoE backbone, where a frequency-front branch enhances wavelet subband representations and parallel spatial/spectral branches refine structural details and bandwise consistency. Two auxiliary regularizers are further introduced to encourage expert diversity and balanced utilization, improving expert specialization under heterogeneous EO scenes.

1.4. Contributions

The main contributions are summarized as follows.

- We propose a prior-guided sparse routing mechanism for multimodal EO fusion. Compact frequency-energy, spatial-structure, and spectral–statistical priors are injected into branch-specific routers, enabling physically meaningful EO cues to guide adaptive expert activation rather than remaining passive auxiliary inputs.
- We design a frequency–spatial–spectral sparse MoE backbone to realize this routing mechanism. The frequency-front branch enhances wavelet-domain subband representations, while the parallel spatial and spectral branches refine local structures and bandwise consistency through lightweight expert modulation.
- We formulate recurring EO fusion tasks from an input–guidance–target data-gap perspective and instantiate DAMoE as a reusable conditional-computation backbone for MSI pansharpening, HSI pansharpening, HSI super-resolution, and SAR-assisted optical reconstruction.
- Extensive experiments across four EO fusion settings demonstrate that DAMoE achieves favorable reconstruction quality and accuracy–efficiency trade-offs. Cross-sensor and cross-dataset transfer analyses, together with downstream segmentation evaluation, further assess its robustness and structural preservation ability.

The remainder of this paper is organized as follows. Section 2 presents the proposed methodology. Section 3 introduces the study data and experimental design. Section 4 reports the results across representative EO fusion settings. Section 5 analyzes the model design and efficiency characteristics. Section 6 discusses the framework in relation to existing EO fusion literature, recurring EO data gaps, extensibility, and limitations. Section 7 concludes the paper.

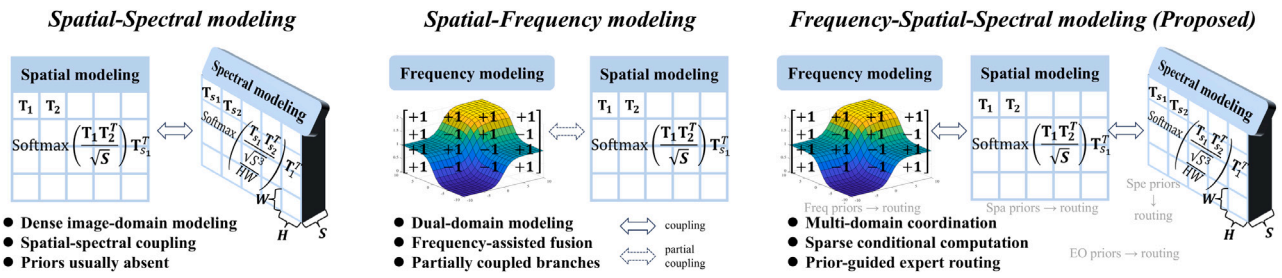

Fig. 1. Representative modeling paradigms for multimodal EO fusion: spatial–spectral (SS) modeling, spatial–frequency (SF) modeling, and the proposed frequency–spatial–spectral (FSS) modeling with prior-guided sparse routing.

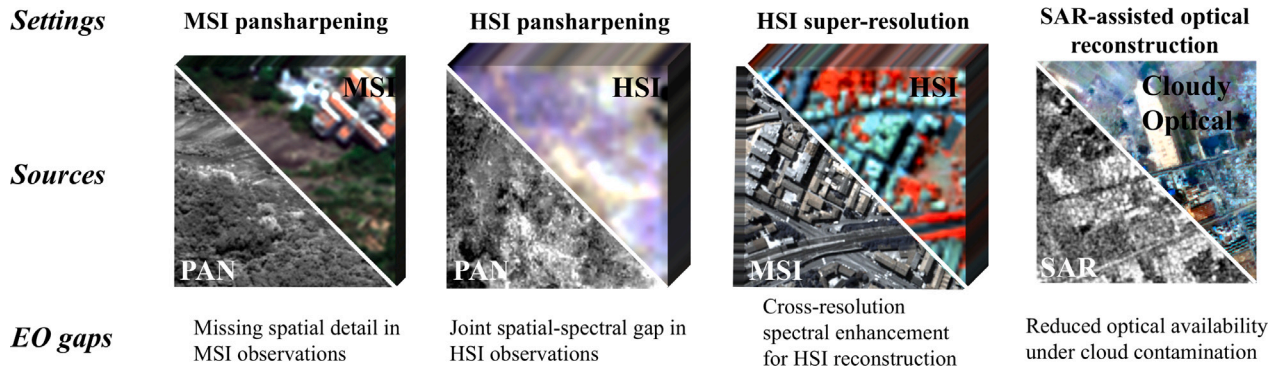

Fig. 2. Representative EO data-gap settings considered in this study under an input–guidance–target gap-recovery formulation.

2. Methodology

2.1. Task-instantiated EO fusion settings and overall framework

Rather than treating each EO fusion task as an isolated problem, we cast them under a sparse architectural template with task-specific instantiations. Let Y denote the degraded multiband observation, Z the higher-resolution or complementary guidance observation, and X the desired fused output. From an observation perspective, an EO data gap can be regarded as the task-dependent information deficit induced by incomplete, degraded, or indirect sensing of the target product. The primary and guidance observations are abstracted as

$$Y = \mathcal{A}(X) + n_Y, \qquad Z = \mathcal{B}(\xi) + n_Z, \tag{1}$$

where $\mathcal{A}$ denotes a task-specific degradation, masking, or spectral-response operator applied to the target product, $\mathcal{B}$ denotes the guidance observation operator, ξ represents the underlying land-surface state observed by different sensing modalities, and n_Y, n_Z are observation noise terms. For optical–optical fusion tasks, Z can be regarded as a spatial or spectral response of X; for cross-modal settings such as SAR-assisted optical reconstruction, Z is not a direct transformation of the optical target but provides complementary structural information about the same land surface.

The operator $\mathcal{A}$ is generally non-invertible, so Y alone cannot fully determine X. Let $\Pi(Y)$ denote a canonical lifting of Y to the target spatial and spectral grid. The corresponding EO data gap is defined as

$$\Delta = X - \Pi(Y), \tag{2}$$

where Δ represents the missing target information to be recovered, while Z provides complementary constraints for estimating this gap. Under this notation, fusion can be viewed as a gap-recovery process.

$$\hat{X} = \Pi(Y) + f_\theta(\Pi(Y), Z), \tag{3}$$

where f_θ estimates the missing component Δ from the lifted primary observation and the guidance observation. Based on this formulation, four representative EO fusion settings are considered in this study: MSI pansharpening, HSI pansharpening, HSI super-resolution, and SAR-assisted optical reconstruction. Although these tasks differ in sensor combinations and in the specific EO gaps to be addressed, they can all be interpreted through task-specific instantiations of the same input–guidance–target gap-recovery formulation in Eq. (3). In these settings, Δ corresponds respectively to missing spatial details, joint spatial–spectral degradation, insufficient spectral resolution, and cloud-obscured optical content. Fig. 2 summarizes their source configurations together with the corresponding EO data gaps.

At the architectural level, the proposed framework adopts a frequency-front and spatial/spectral-parallel design. Two lightweight encoders first project Y and Z into shallow feature spaces. Each encoder consists of two 1×1 convolution layers with a ReLU activation in between. The resulting features are concatenated to form a shared multimodal representation, which is then refined by a frequency-domain MoE module in the wavelet domain. The frequency-enhanced feature produced at this stage is subsequently delivered to two parallel expert branches. One branch is responsible for spatial refinement and focuses on local geometry and structural detail; the other performs spectral refinement and emphasizes bandwise consistency as well as material-dependent characteristics. Their outputs are concatenated in feature space and passed to a lightweight decoder, which follows the same design as the encoders and maps the fused representation back to the target output space.

Accordingly, the proposed model should be understood as a reusable conditional-computation backbone with task-specific instantiations, rather than as a single parameter-shared model applied identically to all EO tasks. What remains invariant across settings is the input–guidance–target formulation, the multi-domain sparse backbone, and the prior-guided routing principle. In contrast, the concrete input configuration, prior source, output head, and trained parameters are instantiated according to the target EO fusion scenario.

2.2. Prior-guided routing across frequency, spatial, and spectral domains

As illustrated in Fig. 3, the proposed framework organizes multimodal EO fusion through three dedicated MoE branches, corresponding

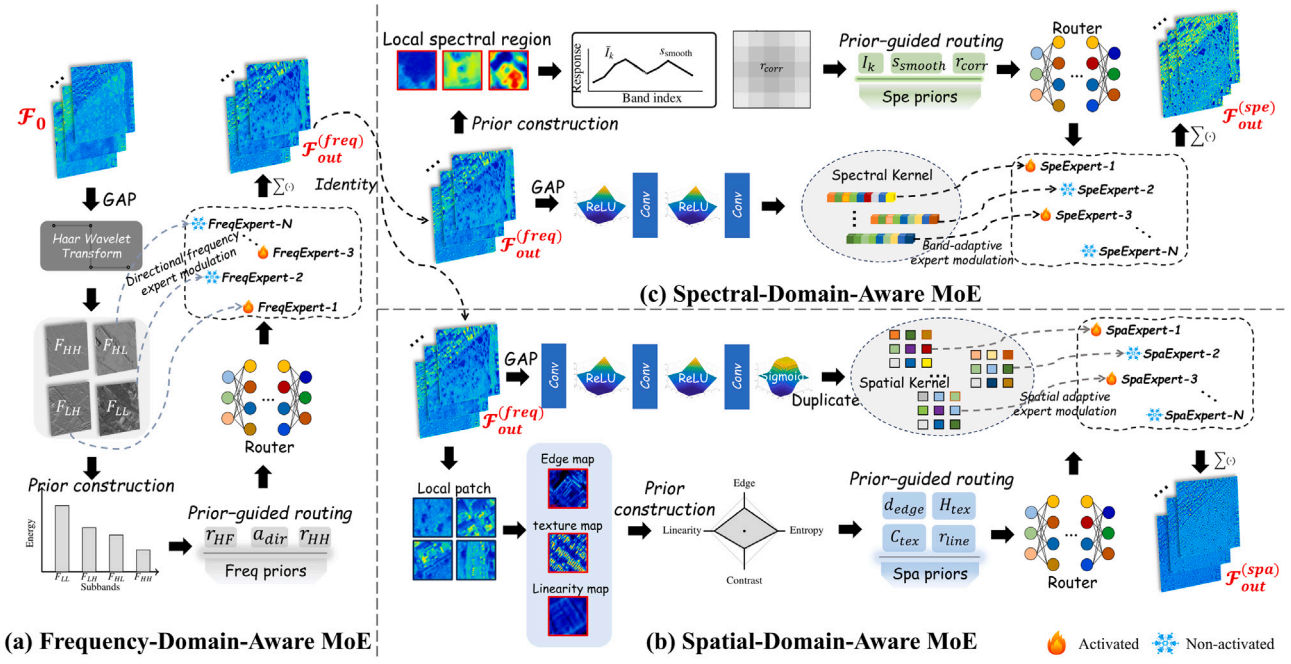

Fig. 3. Illustration of the three expert branches in the proposed framework: (a) the frequency-domain-aware branch, (b) the spatial-domain-aware branch, and (c) the spectral-domain-aware branch. These branches are designed to model complementary frequency, spatial, and spectral characteristics for prior-guided expert routing and multimodal EO fusion.

to the frequency, spatial, and spectral domains. Each branch maintains its own expert pool and lightweight gate, while routing is jointly informed by pooled deep features and compact EO priors. For each domain $b \in \{\text{freq}, \text{spa}, \text{spe}\}$, let $\mathbf{F}^{(b)}$ denote the corresponding feature representation and $\mathbf{d}^{(b)}$ the associated prior descriptor. The routing logits are computed as

$$\mathbf{z}^{(b)} = G^{(b)}\big([\text{GAP}(\mathbf{F}^{(b)}), \mathbf{d}^{(b)}]\big). \tag{4}$$

Each router $G^{(b)}$ is implemented as a two-layer MLP with ReLU activation. The expert pools are branch-specific and parameter-independent, with no weight sharing among frequency, spatial, and spectral experts. For notational convenience, GAP$(\cdot)$ is written in a unified form, although its pooling axis varies by branch: spatial and spectral routing use spatial pooling, whereas the frequency branch pools over subbands. The resulting logits are converted into sparse routing weights through softmax followed by a top-k selection.

$$\boldsymbol{\alpha}^{(b)} = \big[\alpha_1^{(b)}, \ldots, \alpha_{N_b}^{(b)}\big]^\top, \tag{5}$$

where only the top-k entries remain nonzero. The same top-k routing rule is used during training and inference, and the selected weights are renormalized after sparsification. Let $\{\mathcal{E}_e^{(b)}\}_{e=1}^{N_b}$ denote the expert pool of branch b. The routed output of that branch is then written as

$$\mathbf{F}_{\text{out}}^{(b)} = \sum_{e=1}^{N_b} \alpha_e^{(b)}\, \mathcal{E}_e^{(b)}(\mathbf{F}^{(b)}). \tag{6}$$

Here, $\mathcal{E}_e^{(b)}(\cdot)$ denotes the abstract expert mapping at the architectural level, and its branch-specific realization is detailed in the following subsections.

2.2.1. Frequency-domain-aware MoE branch

The frequency-domain-aware MoE branch appears at the front end of the framework. Starting from the shallow joint feature $\mathbf{F}_0$, it applies the 2D Haar wavelet transform to separate low- and high-frequency components along different directions. We define the frequency-domain representation as

$$\mathbf{F}_{\text{freq}} = \mathcal{H}_{\text{Haar}}(\mathbf{F}_0), \qquad \text{s.t.}\mathbf{F}_{\text{freq}} := (\mathbf{F}_{\text{LL}}, \mathbf{F}_{\text{LH}}, \mathbf{F}_{\text{HL}}, \mathbf{F}_{\text{HH}}) \tag{7}$$

where $\mathbf{F}_{\text{LL}}$ denotes the low-frequency component, and $\mathbf{F}_{\text{LH}}$, $\mathbf{F}_{\text{HL}}$, and $\mathbf{F}_{\text{HH}}$ represent horizontal, vertical, and diagonal high-frequency details, respectively.

Frequency prior construction. Routing in this branch is guided by a compact prior that characterizes how energy is distributed across wavelet subbands. We first compute the average absolute energies of the low- and high-frequency components.

$$\begin{cases} E_L = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \big|\mathbf{F}_{\text{LL}}(i,j)\big|, \\ E_H = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \big(|\mathbf{F}_{\text{LH}}(i,j)| + |\mathbf{F}_{\text{HL}}(i,j)| + |\mathbf{F}_{\text{HH}}(i,j)|\big), \end{cases} \tag{8}$$

where Ω denotes the spatial support over which the statistic is computed. Based on these quantities, the high-frequency ratio is defined as

$$r_{\text{HF}} = \frac{E_H}{E_L + \varepsilon}, \tag{9}$$

with $\varepsilon > 0$ introduced for numerical stability. To further quantify directional imbalance, we compute the horizontal and vertical high-frequency energies

$$E_H^x = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \big|\mathbf{F}_{\text{LH}}(i,j)\big|, \qquad E_H^y = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \big|\mathbf{F}_{\text{HL}}(i,j)\big|, \tag{10}$$

and define the directional anisotropy as

$$a_{\text{dir}} = \frac{|E_H^x - E_H^y|}{E_H + \varepsilon}. \tag{11}$$

To reflect diagonal-detail intensity, we additionally define

$$r_{\text{HH}} = \frac{\frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} |\mathbf{F}_{\text{HH}}(i,j)|}{E_H + \varepsilon}. \tag{12}$$

These quantities jointly form the frequency prior

$$\mathbf{d}^{(\text{freq})} = \big[r_{\text{HF}}, a_{\text{dir}}, r_{\text{HH}}\big]^\top. \tag{13}$$

Prior-guided routing. For expert selection, we first pool the frequency representation over the subband dimension

$$\mathbf{g}^{(\text{freq})} = \text{GAP}_{\text{freq}}\big(\{\mathbf{F}_{\text{LL}}, \mathbf{F}_{\text{LH}}, \mathbf{F}_{\text{HL}}, \mathbf{F}_{\text{HH}}\}\big), \tag{14}$$

and then concatenate it with the prior descriptor to obtain the routing logits

$$\mathbf{z}^{(\mathrm{freq})} = G^{(\mathrm{freq})}\left([\mathbf{g}^{(\mathrm{freq})}, \mathbf{d}^{(\mathrm{freq})}]\right) \in \mathbb{R}^{N_\mathrm{f}}. \tag{15}$$

Here, N_f denotes the number of frequency experts. The sparse routing weights are then obtained by applying softmax followed by a top-k sparsifier

$$\begin{aligned} &\boldsymbol{\alpha}^{(\mathrm{freq})} = \mathcal{T}_k\left(\mathrm{softmax}(\mathbf{z}^{(\mathrm{freq})})\right), \\ \text{s.t.}\quad &\boldsymbol{\alpha}^{(\mathrm{freq})} = \left[\alpha_1^{(\mathrm{freq})}, \dots, \alpha_{N_\mathrm{f}}^{(\mathrm{freq})}\right]^\top. \end{aligned} \tag{16}$$

Directional expert modulation and aggregation. Each frequency expert is intended to emphasize different directional frequency patterns. Since EO scenes can be dominated by horizontal boundaries, vertical edges, or diagonal textures, using the same response pattern for all inputs would reduce structural discriminability. To avoid this, the eth expert learns adaptive directional responses from the four subbands.

$$\mathbf{W}_e^{(\mathrm{freq})} = \mathcal{M}_e^{(\mathrm{freq})}(\mathbf{F}_{\mathrm{LL}}, \mathbf{F}_{\mathrm{LH}}, \mathbf{F}_{\mathrm{HL}}, \mathbf{F}_{\mathrm{HH}}), \tag{17}$$

where $\mathcal{M}_e^{(\mathrm{freq})}(\cdot)$ denotes the modulation function of expert e. These responses modulate the relative contributions of different subbands before feature extraction. In implementation, $\mathcal{M}_e^{(\mathrm{freq})}$ is realized as a subband-attention module, which generates adaptive weights from globally pooled Haar subbands. The output of the eth frequency expert is then written as

$$\mathbf{Z}_e^{(\mathrm{freq})} = \mathcal{E}_e^{(\mathrm{freq})}\left(\mathbf{F}_{\mathrm{LL}}, \mathbf{F}_{\mathrm{LH}}, \mathbf{F}_{\mathrm{HL}}, \mathbf{F}_{\mathrm{HH}}, \mathbf{W}_e^{(\mathrm{freq})}\right), \qquad e = 1, \dots, N_\mathrm{f}, \tag{18}$$

where $\mathcal{E}_e^{(\mathrm{freq})}(\cdot)$ applies subband-wise 3×3 depthwise convolution followed by 1×1 pointwise convolution. The processed subbands are then mapped back to the image-feature domain by inverse Haar reconstruction, allowing different frequency experts to specialize in low-frequency, horizontal, vertical, and diagonal detail patterns. The frequency-branch output is obtained by routing-weighted aggregation over all frequency expert outputs.

$$\mathbf{F}_{\mathrm{out}}^{(\mathrm{freq})} = \sum_{e=1}^{N_\mathrm{f}} \alpha_e^{(\mathrm{freq})}\, \mathbf{Z}_e^{(\mathrm{freq})}, \tag{19}$$

which is passed to both the spatial and spectral branches.

2.2.2. Spatial-domain-aware MoE branch

Built on top of $\mathbf{F}_{\mathrm{out}}^{(\mathrm{freq})}$, the spatial-domain-aware MoE branch targets local geometry and structural refinement. We set $\mathbf{F}^{(\mathrm{spa})} = \mathbf{F}_{\mathrm{out}}^{(\mathrm{freq})}$ and construct compact spatial priors to summarize region-level structural properties in EO imagery.

Spatial prior construction. The spatial prior is constructed from three complementary cues: edge strength, texture complexity, and linear structure. We first derive the edge-related cue. Specifically, the luminance-like projection is defined as the channel-averaged feature map.

$$L(i,j) = \frac{1}{C_f} \sum_{c=1}^{C_f} \mathbf{F}_c^{(\mathrm{spa})}(i,j), \tag{20}$$

where C_f denotes the channel dimension of the frequency-enhanced feature. We use the Sobel operator as the edge detector. Let S_x and S_y denote the horizontal and vertical Sobel filters. The gradient magnitude is computed as

$$G(i,j) = \sqrt{(S_x * L)^2(i,j) + (S_y * L)^2(i,j)}. \tag{21}$$

The edge-density descriptor is obtained by thresholding G with its patch-level mean.

$$d_{\mathrm{edge}} = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \mathbb{I}\left(G(i,j) > \frac{1}{|\Omega|} \sum_{(u,v)\in\Omega} G(u,v) \right). \tag{22}$$

Texture complexity is described through a gray-level co-occurrence matrix $\mathbf{P}$ obtained from a luminance projection of $\mathbf{F}^{(\mathrm{spa})}$. Entropy and contrast are then used as texture descriptors.

$$\begin{aligned} H_{\mathrm{tex}} &= -\sum_{u,v} P(u,v)\, \log P(u,v), \\ C_{\mathrm{tex}} &= \sum_{u,v} (u-v)^2 P(u,v). \end{aligned} \tag{23}$$

We compute P after patch-wise min–max normalization and 32-level quantization, using pixel distance 1 and four standard directions. To complement these statistics, we further estimate a linearity score $r_{\mathrm{line}} \in [0,1]$ from the eigenvalue ratio of the local structure tensor, which is intended to capture elongated structures such as roads and rivers. The resulting spatial prior vector is defined as

$$\mathbf{d}^{(\mathrm{spa})} = \left[d_{\mathrm{edge}}, H_{\mathrm{tex}}, C_{\mathrm{tex}}, r_{\mathrm{line}}\right]^\top. \tag{24}$$

Prior-guided routing. As shown in Fig. 3(b), the spatial feature is first pooled over spatial dimensions

$$\mathbf{g}^{(\mathrm{spa})} = \mathrm{GAP}_{\mathrm{spatial}}(\mathbf{F}^{(\mathrm{spa})}) \in \mathbb{R}^{C_\mathrm{f}}. \tag{25}$$

The routing logits are then produced by

$$\mathbf{z}^{(\mathrm{spa})} = G^{(\mathrm{spa})}\left([\mathbf{g}^{(\mathrm{spa})}, \mathbf{d}^{(\mathrm{spa})}]\right) \in \mathbb{R}^{N_\mathrm{s}}, \tag{26}$$

with N_s denoting the number of spatial experts. Sparse routing weights $\boldsymbol{\alpha}^{(\mathrm{spa})}$ are obtained through softmax and top-k selection.

Spatial adaptive expert modulation and aggregation. The role of each spatial expert is to respond adaptively to local heterogeneity in EO imagery. Since different regions can exhibit man-made edges, vegetation textures, or homogeneous backgrounds, fixed convolution weights are often insufficient to capture such variability. To address this issue, the eth expert generates adaptive spatial weights from the input feature.

$$\mathbf{W}_e^{(\mathrm{spa})} = \mathcal{M}_e^{(\mathrm{spa})}(\mathbf{F}^{(\mathrm{spa})}), \tag{27}$$

where $\mathcal{M}_e^{(\mathrm{spa})}(\cdot)$ denotes the corresponding modulation function. In implementation, $\mathcal{M}_e^{(\mathrm{spa})}$ generates a single-channel spatial attention map using two 1×1 convolutions followed by sigmoid activation. The attention map modulates the expert response spatially, enabling different experts to focus on heterogeneous structures such as sharp boundaries, textured vegetation, and homogeneous regions. The output of the eth spatial expert is then written as

$$\mathbf{Z}_e^{(\mathrm{spa})} = \mathcal{E}_e^{(\mathrm{spa})}\left(\mathbf{F}^{(\mathrm{spa})}, \mathbf{W}_e^{(\mathrm{spa})}\right), \qquad e = 1, \dots, N_\mathrm{s}, \tag{28}$$

where $\mathcal{E}_e^{(\mathrm{spa})}(\cdot)$ consists of a 3×3 depthwise convolution, a 1×1 pointwise convolution, and ReLU activation. The spatial expert outputs are aggregated according to the routing weights.

$$\mathbf{F}_{\mathrm{out}}^{(\mathrm{spa})} = \sum_{e=1}^{N_\mathrm{s}} \alpha_e^{(\mathrm{spa})}\, \mathbf{Z}_e^{(\mathrm{spa})}. \tag{29}$$

2.2.3. Spectral-domain-aware MoE branch

The spectral-domain-aware MoE branch is responsible for preserving per-pixel spectra and bandwise consistency. As illustrated in Fig. 3(c), the spectral-domain-aware MoE branch is built on the same frequency-enhanced feature $\mathbf{F}_{\mathrm{out}}^{(\mathrm{freq})}$ as the spatial branch, while its routing is guided by priors that characterize scene-level spectral behavior.

Spectral prior construction. We first set $\mathbf{F}^{(\mathrm{spe})} = \mathbf{F}_{\mathrm{out}}^{(\mathrm{freq})}$ and summarize the feature-space spectral response by spatially pooling $\mathbf{F}^{(\mathrm{spe})}$.

$$\mathbf{g}^{(\mathrm{spe})} = \mathrm{GAP}_{\mathrm{spatial}}(\mathbf{F}^{(\mathrm{spe})}) \in \mathbb{R}^{C_\mathrm{f}}. \tag{30}$$

To complement this pooled feature with physically meaningful information, we compute a compact spectral prior from the available multiband observation. For HSI-related settings, these priors are derived from the input HSI; for the remaining settings, analogous descriptors are extracted from the available multiband optical observation.

Let $\{\phi_k\}_{k=1}^{K}$ denote a set of representative spectral-index functions. The scene-level mean of each index is computed as

$$\bar{I}_k = \frac{1}{|\Omega|}\sum_{j\in\Omega}\phi_k\big(x_j(\lambda_1),\dots,x_j(\lambda_{N_\lambda})\big), \qquad k=1,\dots,K, \tag{31}$$

where N_λ denotes the number of bands. In our implementation, these spectral-index functions include NDVI, NDWI, the red–green normalized difference index, and the NIR/red ratio. For HSI data, the red, green, and NIR bands are selected by nearest wavelength matching; for MSI data, they follow the dataset-provided band definitions. Spectral smoothness is measured through the average absolute difference between adjacent bands.

$$s_{\text{smooth}} = \frac{1}{N_\lambda-1}\sum_{\ell=1}^{N_\lambda-1}\frac{1}{|\Omega|}\sum_{j\in\Omega}\left|x_j(\lambda_{\ell+1})-x_j(\lambda_\ell)\right|, \tag{32}$$

while inter-band dependency is summarized by

$$r_{\text{corr}} = \frac{1}{N_\lambda(N_\lambda-1)}\sum_{\substack{\ell_1,\ell_2=1\\ \ell_1\neq\ell_2}}^{N_\lambda}\mathbf{C}(\ell_1,\ell_2), \tag{33}$$

where $\mathbf{C}\in\mathbb{R}^{N_\lambda\times N_\lambda}$ is the band correlation matrix. These quantities form the spectral prior descriptor

$$\mathbf{d}^{(\text{spe})} = [\bar{I}_1,\dots,\bar{I}_K, s_{\text{smooth}}, r_{\text{corr}}]^\top. \tag{34}$$

Prior-guided routing. The spectral gate receives the concatenation of the pooled feature and the spectral prior

$$\mathbf{z}^{(\text{spe})} = G^{(\text{spe})}\big([\mathbf{g}^{(\text{spe})}, \mathbf{d}^{(\text{spe})}]\big)\in\mathbb{R}^{N_\text{p}}, \tag{35}$$

where N_p is the number of spectral experts. Softmax and top-k selection then yield the sparse routing weights $\boldsymbol{\alpha}^{(\text{spe})}$.

Band-adaptive expert modulation and aggregation. Each spectral expert is designed to emphasize different spectral bands according to local spectral characteristics. Since distinct targets in EO imagery are associated with different spectral curves, fixed channel responses can obscure informative bandwise variation. For this reason, the eth expert generates adaptive spectral weights through a channel-attention-like mechanism.

$$\mathbf{W}_e^{(\text{spe})} = \mathcal{M}_e^{(\text{spe})}(\mathbf{F}^{(\text{spe})}), \tag{36}$$

where $\mathcal{M}_e^{(\text{spe})}(\cdot)$ denotes the modulation function of expert e. These weights rescale spectral responses within the same spatial region, allowing the expert to adapt to local spectral patterns. In implementation, $\mathcal{M}_e^{(\text{spe})}$ follows a channel-attention design based on globally pooled features. Spatial context is further preserved during this modulation process to account for radiometric discrepancies across sensors. The output of the eth spectral expert is thus written as

$$\mathbf{Z}_e^{(\text{spe})} = \mathcal{E}_e^{(\text{spe})}\big(\mathbf{F}^{(\text{spe})}, \mathbf{W}_e^{(\text{spe})}\big), \qquad e=1,\dots,N_\text{p}, \tag{37}$$

where $\mathcal{E}_e^{(\text{spe})}(\cdot)$ first performs 1×1 spectral mixing, then applies a 3×3 depthwise convolution to preserve local spatial context, followed by a final 1×1 projection. This design allows different spectral experts to emphasize complementary band combinations while maintaining spatial consistency. The final spectral-domain output is obtained as

$$\mathbf{F}_{\text{out}}^{(\text{spe})} = \sum_{e=1}^{N_\text{p}}\alpha_e^{(\text{spe})}\mathbf{Z}_e^{(\text{spe})}. \tag{38}$$

2.3. Dual regularization for expert collaboration

Although the proposed multi-domain MoE architecture provides strong conditional capacity, recurring EO data gaps still induce substantial heterogeneity in spatial structures, spectral responses, and sensing conditions. Under such variability, different experts can drift toward similar behaviors or become unevenly activated during routing, reducing the effective benefit of sparse conditional computation. To mitigate this, we introduce two auxiliary regularizers: one encourages diversity across experts, and the other promotes balanced expert usage.

2.3.1. Expert diversity regularization

To encourage complementary expert representations within the same pool, we employ a kernel-based diversity regularizer. Let $\mathcal{P}_{e_1}$ and $\mathcal{P}_{e_2}$ denote the empirical distributions of features produced by experts e_1 and e_2, respectively, over a mini-batch. Their pairwise regularization term is defined as

$$\mathcal{R}_{\mathcal{H}}^2(\mathcal{P}_{e_1},\mathcal{P}_{e_2}) = \mathbb{E}_{\mathbf{r}\sim\mathcal{P}_{e_1},\hat{\mathbf{r}}\sim\mathcal{P}_{e_2}}\big[\kappa(\mathbf{r},\hat{\mathbf{r}})\big], \tag{39}$$

where

$$\kappa(\mathbf{r},\hat{\mathbf{r}}) = \exp\left(-\frac{\|\mathbf{r}-\hat{\mathbf{r}}\|^2}{2\sigma^2}\right) \tag{40}$$

is a Gaussian kernel defined in a reproducing kernel Hilbert space. The bandwidth σ is estimated adaptively from the average pairwise squared distance in the current mini-batch rather than fixed manually. Minimizing this term reduces similarity among expert outputs and encourages complementary specialization.

For a pool of N_b experts, the pairwise terms are aggregated as

$$\mathcal{R}_{\mathcal{H}}^2\big(\{\mathcal{E}_e\}_{e=1}^{N_b}\big) = \frac{2}{N_b(N_b-1)}\sum_{e<e'}\mathcal{R}_{\mathcal{H}}^2(\mathcal{P}_e,\mathcal{P}_{e'}), \tag{41}$$

and the same regularization is applied to the frequency, spatial, and spectral expert pools. Their contributions are summed to form the overall diversity regularization term.

2.3.2. Expert-usage balancing

To avoid routing collapse onto only a few experts, we additionally introduce an expert-usage balancing term. Let N_{mb} denote the number of training samples in a mini-batch, and let $\mathcal{M}_{j,e}^{(b)}\in\{0,1\}$ indicate whether expert $\mathcal{E}_e$ in branch b is selected by top-k routing for sample j. The hard utilization of expert e is then

$$p_e^{(b)} = \sum_{j=1}^{N_{\text{mb}}}\mathcal{M}_{j,e}^{(b)}. \tag{42}$$

Let $\mathbf{w}_e^{(b)}(\mathbf{F}^{(b),j};\boldsymbol{\theta})$ denote the corresponding soft routing weight for sample j. The cumulative soft utilization is defined as

$$\hat{p}_e^{(b)} = \sum_{j=1}^{N_{\text{mb}}}\mathbf{w}_e^{(b)}(\mathbf{F}_j^{(b)};\boldsymbol{\theta}). \tag{43}$$

Based on these quantities, the balancing loss is written as

$$\mathcal{L}_{\text{bal}} = \frac{1}{N_b}\sum_{e=1}^{N_b}(p_e^{(b)}\cdot\hat{p}_e^{(b)}), \tag{44}$$

which penalizes experts that are simultaneously selected too frequently and assigned overly large soft weights. Minimizing $\mathcal{L}_{\text{bal}}$ therefore encourages a more even routing distribution across experts and improves effective capacity usage. In implementation, both $p_e{}^{(b)}$ and $\hat{p}_e^{(b)}$ are normalized by the number of samples in each mini-batch, making the loss invariant to batch size. The load-balancing loss is computed within each branch and then averaged across the frequency, spatial, and spectral branches, rather than being computed from globally mixed expert-utilization statistics. Taken together, the diversity regularizer and the balancing term help maintain expert complementarity and routing stability under the heterogeneous conditions induced by recurring EO data gaps.

2.4. Objective function and training strategy

Given a training set of triplets $\{(Y_j,Z_j,X_j)\}_{j=1}^{N_{\text{tr}}}$, where X_j denotes the ground truth (GT), the network prediction is denoted by $\hat{X}_j$. The overall training objective combines the reconstruction term with the two auxiliary regularizers:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda\mathcal{L}_{\text{bal}} + \gamma\mathcal{R}_{\mathcal{H}}^2, \tag{45}$$

Table 1
Datasets, sensing combinations, scene characteristics, and EO settings used in this study.

Dataset	Scene type	EO setting	Protocol	Notes
GF2	Urban/river/built-up	MSI pansharpening	Reduced-scale + full-scale	4-band MSI + PAN
WV-III	Urban/cropland/mixed land cover	MSI pansharpening	Reduced-scale + full-scale	8-band MSI + PAN
Botswana	Rocky/mountainous	HSI pansharpening	Full-reference	128-band HSI + PAN
Pavia University	Urban/road/building	HSI super-resolution	Full-reference	92-band HSI + 4-band MSI
SEN2MS-CR	Cloud-affected scenes	SAR-assisted optical reconstruction	Paired evaluation	13-band MSI + SAR (VV + VH)
WV-II	Related optical sensor setting	Cross-sensor transfer	Test only	8-band MSI + PAN
Pavia Center	Urban/road/building	HSI pansharpening transfer	Zero-shot transfer only	102-band HSI + PAN

where $\mathcal{L}_{\text{recon}}$ is implemented as a pixel-wise ℓ_1 loss between $\hat{X}_j$ and X_j. Here, $\mathcal{L}_{\text{bal}}$ denotes the expert-usage balancing term, and $\mathcal{R}_{\mathcal{H}}^2$ denotes the diversity regularization aggregated across the three expert pools. The coefficients $\lambda, \gamma > 0$ control the strengths of the two regularizers, while the Gaussian kernel bandwidths are estimated adaptively from batch statistics. Unless otherwise specified, we set $N_f = N_s = N_p = 6$, $k = 5$, $\lambda = 0.1$, and $\gamma = 0.01$ according to the ablation results in Section 5. All components are optimized jointly in an end-to-end manner. In practice, we adopt Adam with an initial learning rate of 5×10^{-4}, a batch size of 64, and 1000 training epochs in total. The learning rate is halved every 200 epochs to promote stable convergence. Unless otherwise stated, the same training configuration is used across all evaluated tasks for fair comparison and simplified deployment.

3. Data and experimental design

3.1. Study data and EO settings

We evaluated the proposed framework on five source-domain public benchmarks and two additional transfer test sets spanning four representative EO fusion settings: MSI pansharpening, HSI pansharpening, HSI super-resolution, and SAR-assisted optical reconstruction. Specifically, GF2 and WV-III (Deng et al., 2022) were used for MSI pansharpening, Botswana (Zhuo et al., 2022) for HSI pansharpening, Pavia University (Xu et al., 2019) for HSI super-resolution, and SEN2MS-CR (Ebel et al., 2021) for SAR-assisted optical reconstruction. In addition, WV-II and Pavia Center were used only for cross-sensor and cross-dataset transfer evaluation. Together, these datasets cover urban, agricultural, rocky, and cloud-affected scenes, providing a suitable test bed for assessing whether a shared sparse MoE framework can generalize across heterogeneous EO conditions.

For the MSI and HSI fusion datasets, we followed the original or previously published train/validation/test partitioning protocols, with a 9:1 ratio between the numbers of training and validation samples. For SAR-assisted optical reconstruction, we used the publicly released paired Sentinel-1 and Sentinel-2 data. The full dataset contains 122,218 paired SAR and MSI cloudy/cloud-free samples, each with a fixed patch size of 256 × 256 pixels. Following Ebel et al. (2021), six non-overlapping regions of interest were selected for training, validation, and testing, containing 3166, 718, and 780 patches, respectively. Since the cloudy optical image and the reference image are not strictly simultaneous, the SAR-assisted optical reconstruction results should be interpreted with this temporal mismatch in mind. All datasets used in this study are publicly available through their original publications and official websites. Table 1 summarizes the datasets, sensing combinations, scene characteristics, EO settings, and evaluation protocols used in this work.

3.2. Evaluation protocol and metrics

Task-specific evaluation protocols were adopted for the four EO fusion settings considered in this study. For MSI pansharpening, both reduced-scale and full-scale evaluations were conducted. Reduced-scale experiments were performed on simulated test sets and evaluated using ERGAS, SAM, and Q2n (Arienzo et al., 2022). Full-scale experiments were conducted on real test sets and evaluated using D_λ, D_λ^F, D_s, QNR (Garzelli and Nencini, 2009), and HQNR (Arienzo et al., 2022). Since QNR is derived from D_λ and D_s, and HQNR from D_λ^F and D_s, D_λ and D_λ^F are omitted from the quantitative result tables. For HSI pansharpening and HSI super-resolution, full-reference evaluation was adopted, using PSNR, SAM, and ERGAS to assess pixel-wise fidelity, spectral consistency, and global reconstruction error. For SAR-assisted optical reconstruction, PSNR, SAM, and SSIM were used. Since the cloudy optical image and the reference image in SEN2MS-CR are not strictly simultaneous, these full-reference metrics should be interpreted as reference-based similarity measures rather than exact physical reconstruction error. To assess whether the performance margins are statistically meaningful, we further performed paired Wilcoxon signed-rank tests using per-sample metric values. The tests were conducted for GF2, WV-III, WV-II, and SEN2MS-CR, where the numbers of independent test samples are sufficient for paired comparison. For HSI benchmarks and HSI cross-dataset transfer, results are reported descriptively because of the limited number of independent test samples. To further assess whether fusion quality transfers to downstream analysis, we also performed unsupervised segmentation using GMM-SS (Fauvel et al., 2013) and SLIC-Agglo (Achanta et al., 2012), and reported overall accuracy together with confusion-pattern analysis (Liu et al., 2025a).

3.3. Baselines and implementation details

Representative methods validated for the corresponding EO fusion settings were selected as baselines and grouped into three categories: traditional methods, SS modeling methods, and SF modeling methods. For MSI pansharpening, the compared methods include FS (Vivone et al., 2018), BDSD-PC (Vivone, 2019), ADKNet (Peng et al., 2022), SSAFF (Yang et al., 2023), SSUN-Net (Fang and Gan, 2025), FAME (He et al., 2024), RAMSF (Liu et al., 2025b), and PanComplex (Luo et al., 2025). For HSI pansharpening, the compared methods include CNMF (Yokoya et al., 2012), MTF-GLP (Vivone et al., 2015), MDANet (Guan and Lam, 2022), PSRT (Deng et al., 2023), MCIFNet (Zhu et al., 2025a), Hyper-DSNet (Zhuo et al., 2022), DCINN (W. Wang et al., 2024), and SFIGNet (Zhu et al., 2025b). For HSI super-resolution, the compared methods include SFIM (Liu, 2000), GFPCA (Liao et al., 2015), DHIFNet (Huang et al., 2022), 3DT-Net (Ma et al., 2023), MCIFNet, QIS-GAN (Zhu et al., 2023b), DCINN, and SFIGNet. For SAR-assisted optical reconstruction, the compared methods include U-Net3D (Rustowicz et al., 2019), DSen2-CR (Meraner et al., 2020), GLF-CR (Xu et al., 2022), CF2N (Liu et al., 2025), and RAMSF. In all experiments, we followed the official implementations or published settings of the compared methods as closely as possible. All trainable deep baselines were retrained and evaluated under the same experimental pipeline, using the same public data splits, preprocessing protocols, input generation, normalization, and evaluation scripts as DAMoE. Traditional methods were evaluated with their recommended parameters under the same preprocessing and metric-computation settings. For reduced-scale and full-scale evaluations, all methods used identical testing samples and metric implementations. Experiments were conducted on a single compute node with six Intel Xeon Gold 6133 CPUs and two GeForce RTX 4090 GPUs. Runtime was measured under the same protocol for all compared deep models

Table 2
Quantitative evaluation on the reduced-scale testing sets of GF2 and WV-III. Indicators marked with ↑ indicate higher and better values, while ↓ indicate lower and better values.

Data	Metrics	Traditional		SS modeling			SF modeling			FSS modeling
		FS	BDSD-PC	ADKNet	SSAFF	SSUN-Net	FAME	RAMSF	PanComplex	DAMoE (Ours)
GF2	ERGAS (↓)	1.6201 ±0.3526	1.6954 ±0.3896	0.8215 ±0.1149	0.8134 ±0.1416	0.8908 ±0.1368	0.8151 ±0.1393	0.7441 ±0.1259	0.8364 ±0.1187	**0.6564***** ±0.1226
	SAM (↓)	1.6807 ±0.3394	1.7243 ±0.3118	0.8830 ±0.1509	0.8780 ±0.1566	1.1265 ±0.1889	0.8958 ±0.2331	0.8082 ±0.1475	0.8912 ±0.1434	**0.7105**** ±0.1335
	Q2n (↑)	0.8904 ±0.0250	0.8847 ±0.0300	0.9721 ±0.0097	0.9737 ±0.0087	0.9695 ±0.0107	0.9734 ±0.0092	0.9768 ±0.0101	0.9657 ±0.0101	**0.9822***** ±0.0069
WV-III	ERGAS (↓)	4.6450 ±1.4062	4.6499 ±1.4270	2.2905 ±0.5504	2.3878 ±0.5330	2.3865 ±0.5613	2.3211 ±0.4828	2.2265 ±0.4605	2.2344 ±0.4772	**2.1726**** ±0.4875
	SAM (↓)	5.3228 ±1.6112	5.4643 ±1.6708	3.1382 ±0.5618	3.2079 ±0.5778	3.2323 ±0.5206	3.2271 ±0.6020	2.9863 ±0.5296	3.0436 ±0.5472	**2.9361**** ±0.5266
	Q2n (↑)	0.8177 ±0.0989	0.8117 ±0.1036	0.9051 ±0.0834	0.9015 ±0.0873	0.9015 ±0.0866	0.9014 ±0.0853	0.9162 ±0.0896	0.9068 ±0.0831	**0.9165*** ±0.0783

Bold and underline denote the best and second-best results, respectively. The markers *, **, and *** indicate statistically significant improvement of DAMoE over the strongest competing method under a paired Wilcoxon signed-rank test, with $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

Table 3
Quantitative evaluation on full-scale testing sets of GF2 and WV-III.

Data	Metrics	Traditional		SS modeling			SF modeling			FSS modeling
		FS	BDSD-PC	ADKNet	SSAFF	SSUN-Net	FAME	RAMSF	PanComplex	DAMoE (Ours)
GF2	D_s (↓)	0.0524 ±0.0173	0.0559 ±0.0196	0.0253 ±0.0100	0.0384 ±0.0132	**0.0241** ±0.0135	0.0495 ±0.0178	0.0333 ±0.0117	0.0311 ±0.0125	0.0253 ±0.0084
	QNR (↑)	0.9254 ±0.0273	0.9319 ±0.0260	0.9663 ±0.0152	0.9524 ±0.0180	0.9665 ±0.0132	0.9530 ±0.0385	0.9585 ±0.0165	0.9615 ±0.0178	**0.9700*** ±0.0139
	HQNR (↑)	0.9121 ±0.0210	0.8678 ±0.0368	0.9534 ±0.0135	0.9362 ±0.0155	0.9505 ±0.0155	0.9509 ±0.0125	0.9471 ±0.0129	0.9435 ±0.0155	**0.9571*** ±0.0120
WV-III	D_s (↓)	0.0851 ±0.0307	0.0912 ±0.0362	0.0484 ±0.0190	0.0674 ±0.0398	0.0337 ±0.0138	0.0680 ±0.0221	0.0433 ±0.0165	0.0436 ±0.0208	**0.0279*** ±0.0067
	QNR (↑)	0.8973 ±0.0432	0.8973 ±0.0432	0.9312 ±0.0236	0.9016 ±0.0695	0.9468 ±0.0232	0.9155 ±0.0307	0.9219 ±0.0286	0.9226 ±0.0299	**0.9490***** ±0.0243
	HQNR (↑)	0.8971 ±0.0352	0.8528 ±0.0509	0.9338 ±0.0181	0.9113 ±0.0567	0.9397 ±0.0201	0.9130 ±0.0261	0.9386 ±0.0191	0.9363 ±0.0240	**0.9454*** ±0.0201

Bold and underline denote the best and second-best results, respectively. The markers *, **, and *** indicate statistically significant improvement of DAMoE over the strongest competing method under a paired Wilcoxon signed-rank test, with $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

and includes prior extraction, since prior descriptors are embedded in the forward pass. For reproducibility, the released repository provides training and evaluation scripts, preprocessing routines, dataset split scripts, configuration files, fixed random seeds, and pretrained weights.

4. Results across representative EO fusion settings

4.1. Multispectral pansharpening results

Table 2 and Figs. 4–5 summarize the fusion results on GF2 and WV-III. On GF2, DAMoE shows the clearest advantage in the error-related metrics, improving SAM and ERGAS by about 12% over the strongest competing method while further increasing Q2n. On WV-III, the gains are smaller but remain consistent, with about 2% improvements in both SAM and ERGAS together with a further gain in Q2n. The qualitative comparisons in Figs. 4 and 5 support the same trend: DAMoE reconstructs cleaner building boundaries, river edges, and vegetation regions, with darker absolute error maps (AEMs) than the compared baselines.

This advantage is maintained under full-scale evaluation. As shown in Table 3, DAMoE remains among the strongest methods on GF2 and achieves the best overall performance on WV-III. The gain is most evident in the distortion-related metric on WV-III, where D_s is reduced by more than 17% relative to the strongest competing method, while QNR and HQNR are also further improved. On GF2, the improvements are smaller but remain consistent in the two comprehensive metrics. The paired Wilcoxon signed-rank tests further confirm that most reduced-scale and full-scale MSI improvements are statistically significant, except for GF2 full-scale D_s, where DAMoE obtains a second-best value and is therefore not marked as significant superiority. Overall, these results indicate that the proposed sparse multi-domain design improves both spectral consistency and overall reconstruction quality under simulated and real MSI pansharpening.

4.2. Hyperspectral pansharpening results

Table 4 and Fig. 6 report the results on the Botswana HSI pansharpening benchmark. DAMoE achieves the best overall performance, improving PSNR by about 1%, reducing SAM by about 4%, and lowering ERGAS by nearly 9% relative to the strongest competing method. The qualitative results in Fig. 6 show the same trend. In rocky and mountainous regions, DAMoE reconstructs spatial and spectral details more faithfully than the compared baselines. This suggests that coordinated frequency-, spatial-, and spectral-domain modeling is particularly effective for HSI pansharpening.

4.3. Hyperspectral super-resolution results

Table 5 and Fig. 7 report the Pavia University HSI super-resolution results. DAMoE achieves the best performance across all three full-reference metrics, improving PSNR by about 0.8%, reducing SAM by about 2%, and lowering ERGAS by about 1.4% relative to the strongest competing method. These consistent gains indicate stable effectiveness across HSI fusion settings. Fig. 7 shows the same trend. In urban regions with buildings and roads, DAMoE preserves sharper boundaries while keeping the reconstructed spectra closer to the reference. This suggests that the proposed framework improves both spatial detail and spectral fidelity in HSI super-resolution.

4.4. Spectral profile analysis on the HSI benchmarks

Fig. 8 further examines spectral fidelity at representative pixels on the Botswana and Pavia University benchmarks. On Botswana, the spectral curves reconstructed by DAMoE remain consistently closest to the reference across the two selected pixels, especially around the major peaks, valleys, and abrupt band transitions. On Pavia University, the same trend is observed in urban regions with more complex material mixtures. DAMoE better follows the reference spectral trajectory over the full band range, while several competing methods exhibit visible deviations in amplitude or shape. These results show that the advantage of DAMoE is not limited to aggregate metrics such as PSNR, SAM, and ERGAS, but also extends to pixel-level spectral consistency, which is critical for faithful HSI reconstruction.

4.5. SAR-assisted optical reconstruction results

Table 6 and Fig. 9 report the results on the SEN2MS-CR benchmark. DAMoE achieves the best overall performance, with the highest PSNR and SSIM and competitive SAM. The qualitative results are consistent

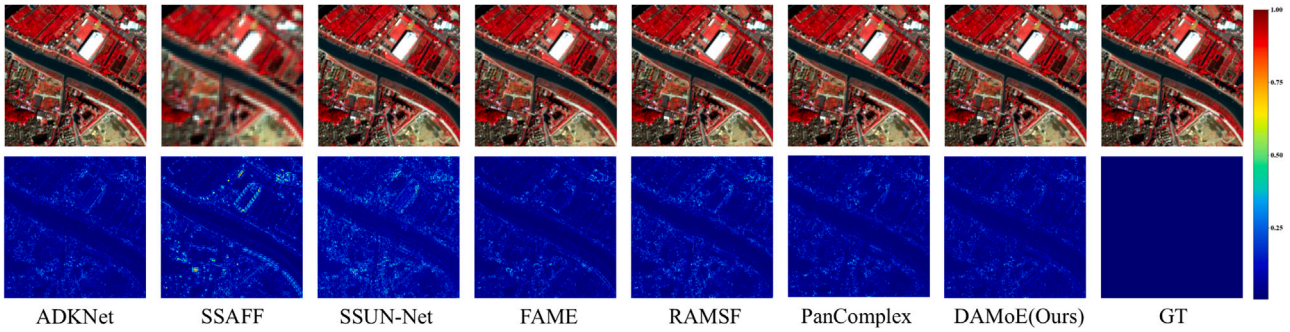

Fig. 4. Representative qualitative comparison on the GF2 reduced-scale MSI pansharpening test set. Top row: fused false-color images. Bottom row: absolute error maps (AEMs) with respect to the ground truth (GT).

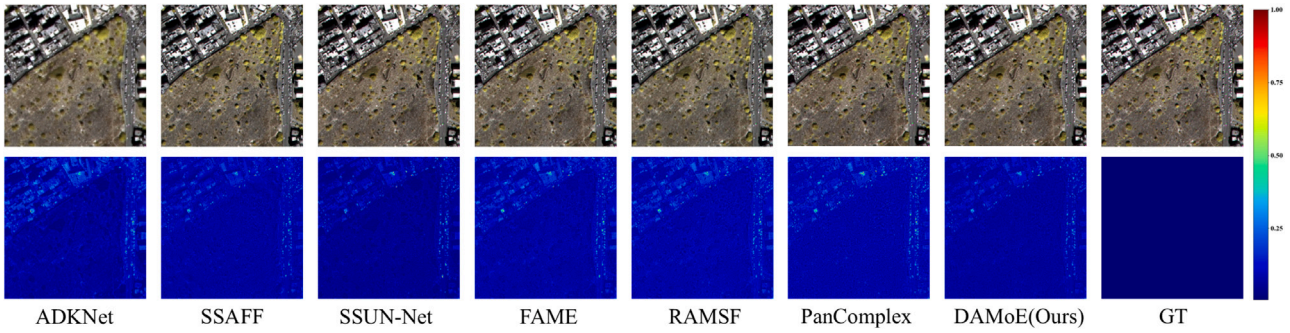

Fig. 5. Representative qualitative comparison on the WV-III reduced-scale MSI pansharpening test set.

Table 4
Quantitative comparison on the Botswana HSI pansharpening benchmark.

Metrics	Traditional		SS modeling			SF modeling			FSS modeling
	CNMF	MTF-GLP	MDANet	PSRT	MCIFNet	DSNet	DCINN	SFIGNet	DAMoE (Ours)
PSNR (↑)	31.9540 ±1.0506	33.1565 ±1.2652	37.8565 ±2.6100	37.5537 ±2.6206	38.1266 ±3.0501	37.0110 ±2.4849	37.9243 ±2.3557	37.5018 ±2.2962	**38.5618** ±2.5487
SAM (↓)	2.0324 ±0.2435	1.9279 ±0.2223	1.5151 ±0.2022	1.5498 ±0.1986	1.4886 ±0.2046	1.6464 ±0.2302	1.5912 ±0.1705	1.5957 ±0.1835	**1.4289** ±0.1745
ERGAS (↓)	2.8813 ±0.4748	2.5634 ±0.3519	1.9192 ±0.5234	1.8140 ±0.5508	1.7634 ±0.6233	1.8512 ±0.5023	1.7959 ±0.4837	1.8565 ±0.4935	**1.6104** ±0.5263

Bold and underline denote the best and second-best results, respectively.

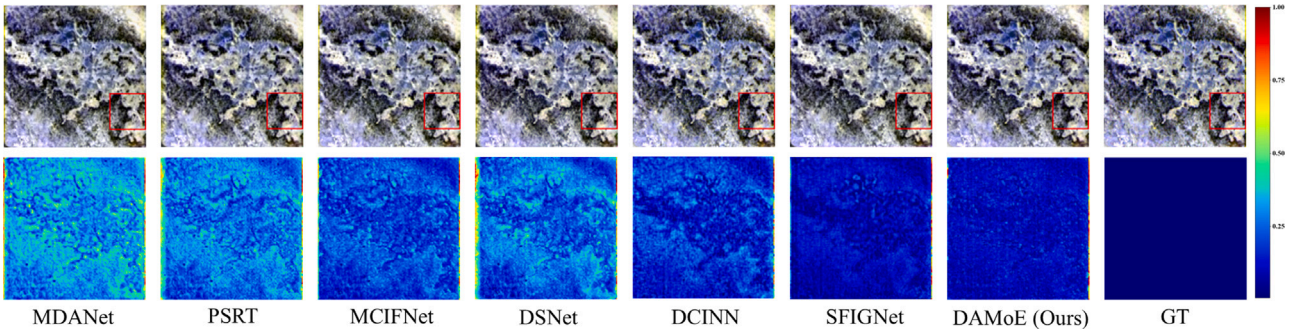

Fig. 6. Representative qualitative comparison on the Botswana HSI pansharpening test set.

Table 5
Quantitative comparison on the Pavia University HSI super-resolution benchmark.

Metrics	Traditional		SS modeling			SF modeling			FSS modeling
	SFIM	GFPCA	DHIFNet	3DT-Net	MCIFNet	QIS-GAN	DCINN	SFIGNet	DAMoE (Ours)
PSNR (↑)	26.2683 ±0.5149	25.7571 ±0.5754	47.4610 ±1.3416	47.2101 ±1.5826	47.4694 ±2.0256	47.3713 ±1.3344	44.6116 ±0.9054	47.1652 ±1.1862	**47.8569** ±1.6415
SAM (↓)	7.0703 ±0.6646	7.6052 ±0.7573	1.5594 ±0.1152	1.5549 ±0.1091	1.4988 ±0.1190	1.5631 ±0.1236	1.9501 ±0.1726	1.6220 ±0.1288	**1.4696** ±0.1241
ERGAS (↓)	8.4280 ±0.7749	8.8297 ±0.7883	0.8632 ±0.1606	0.8958 ±0.1744	0.8749 ±0.2163	0.8719 ±0.1608	1.1001 ±0.1681	0.8922 ±0.1703	**0.8508** ±0.1817

Bold and underline denote the best and second-best results, respectively.

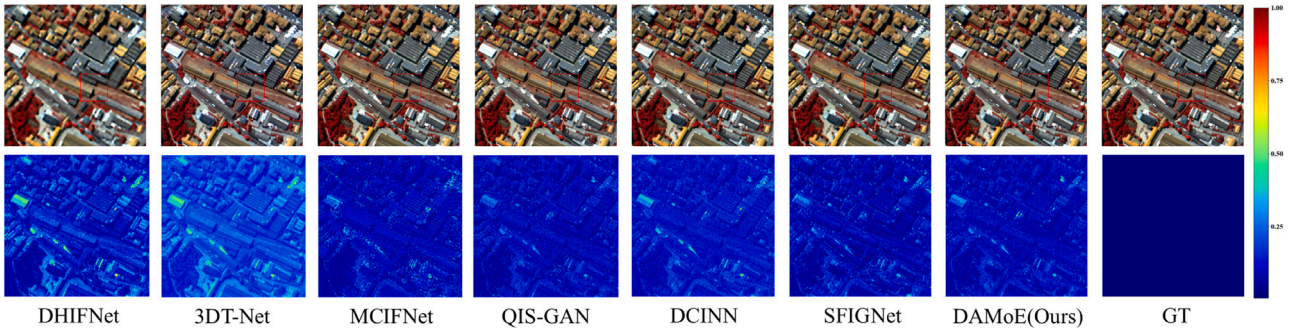

Fig. 7. Representative qualitative comparison on the Pavia University HSI super-resolution test set.

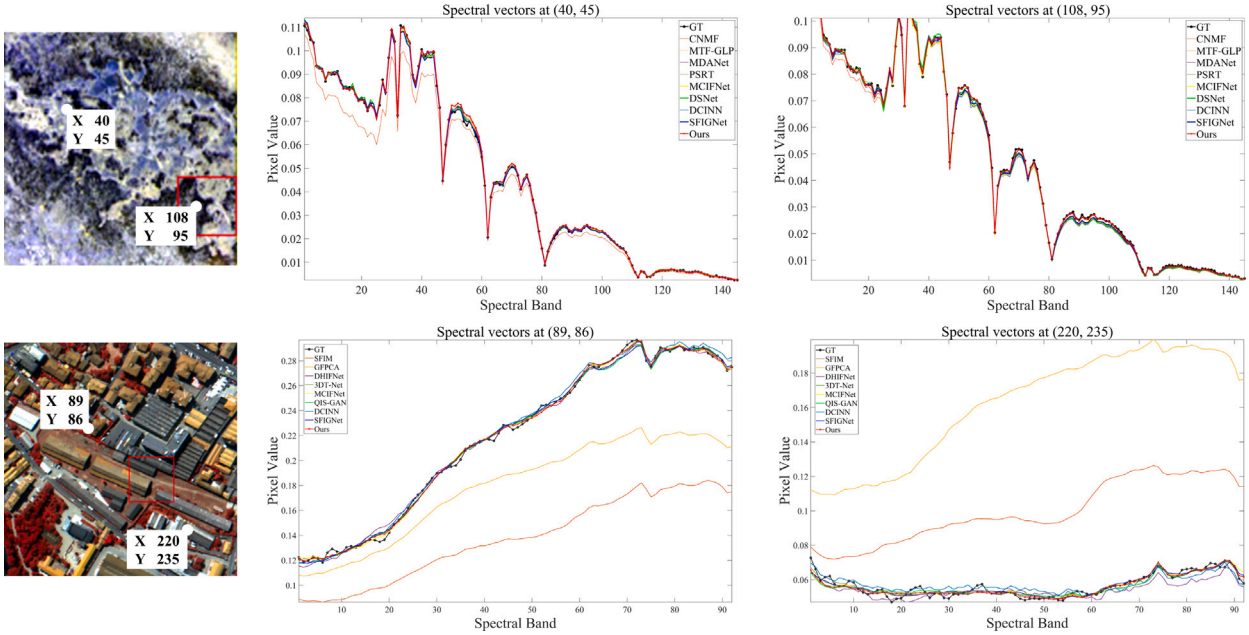

Fig. 8. Spectral profiles at representative pixels from the two HSI benchmarks.

Table 6
Quantitative comparison on the SEN2MS-CR SAR-assisted optical reconstruction benchmark.

Data	Metrics	SS modeling			SF modeling		FSS modeling
		U-Net3D	DSen2-CR	GLF-CR	CF2N	RAMSF	DAMoE (Ours)
SEN2MS-CR	PSNR (↑)	24.6062 ±1.5872	28.1201 ±2.3621	27.1886 ±2.2097	29.0263 ±1.7701	<u>29.0856</u> ±1.8954	**29.7849***** ±1.7847
	SAM (↓)	12.2276 ±2.7467	**6.6350** ±1.9170	7.5923 ±2.4394	6.8566 ±2.0214	6.8842 ±2.2262	<u>6.7726</u> ±1.9847
	SSIM (↑)	0.7683 ±0.0349	0.8655 ±0.0477	<u>0.8759</u> ±0.0397	0.8561 ±0.0433	0.8652 ±0.0459	**0.8871***** ±0.0439

Bold and <u>underline</u> denote the best and second-best results, respectively. The markers *, **, and *** indicate statistically significant improvement of DAMoE over the strongest competing method under a paired Wilcoxon signed-rank test, with $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

with the quantitative comparison. DAMoE reconstructs cleaner structures and yields lower reconstruction error in cloud-affected urban regions, whereas the compared baselines show more visible blurring, spectral distortion, or local artifacts. Since the cloudy optical input and the reference image are not strictly simultaneous in SEN2MS-CR, these results should be interpreted with this temporal mismatch in mind. Even under this condition, DAMoE remains effective when optical availability is reduced.

4.6. Cross-sensor and cross-dataset transfer

To further evaluate the generalization ability of DAMoE beyond source-domain testing, we conduct transfer experiments on both MSI and HSI data. The MSI experiment examines related-sensor transfer from WV-III to WV-II, while the HSI experiment evaluates a more challenging zero-shot cross-dataset transfer from Botswana to Pavia Center.

For MSI pansharpening, the cross-sensor generalization of DAMoE is evaluated by directly testing the models trained on WV-III on WV-II. Although WV-III and WV-II are related optical sensors, they still differ in spectral response functions, spatial resolution, radiometric characteristics, and imaging conditions, making this setting a realistic related-sensor generalization scenario. As reported in Table 7, DAMoE achieves the best overall performance in terms of SAM, ERGAS, and Q2n, with improvements of 5.67%, 13.41%, and 3.13% over the second-best method, respectively. Fig. 10 shows the same trend. Most learning-based methods exhibit clear degradation after transfer, whereas DAMoE maintains the smallest overall degradation across the three metrics. Although the two traditional methods appear more stable, their source-domain performance is already limited. Taken

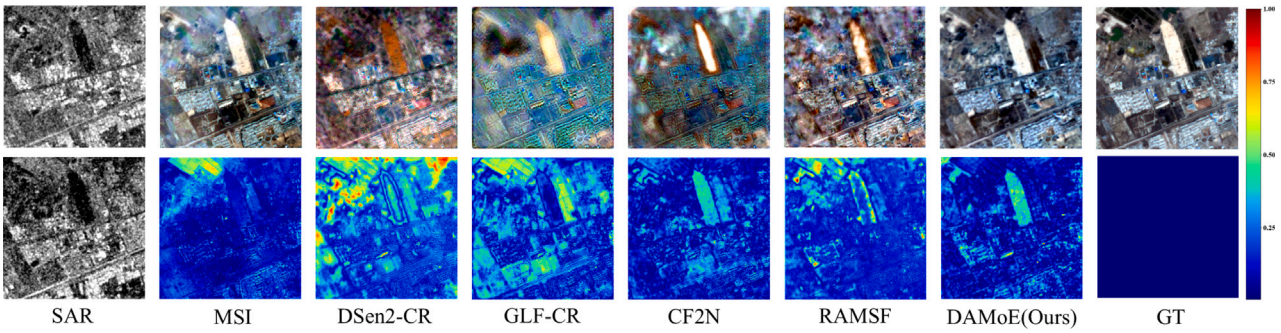

Fig. 9. Qualitative comparison on the SEN2MS-CR SAR-assisted optical reconstruction test set.

Table 7
Cross-sensor quantitative comparison on the WV-II MSI pansharpening test set using models trained on WV-III.

Metrics	Traditional		SS modeling			SF modeling			FSS modeling
	FS	BDSD-PC	ADKNet	SSAFF	SSUN-Net	FAME	RAMSF	PanComplex	DAMoE (Ours)
ERGAS (↓)	4.5552 ±0.5262	4.6486 ±1.5920	4.7658 ±0.4129	4.5874 ±0.3448	4.5062 ±0.3552	6.0343 ±0.3771	4.4668 ±0.3826	5.4427 ±0.3670	**3.8679***** ±0.3183
SAM (↓)	6.2085 ±0.8160	6.0894 ±0.8967	5.6921 ±0.6092	6.0407 ±0.5950	5.9784 ±0.3713	7.2734 ±0.3997	5.5011 ±0.5570	7.0594 ±0.6865	**5.1893**** ±0.4669
Q2n (↑)	0.8095 ±0.0896	0.8217 ±0.0960	0.8173 ±0.0759	0.8234 ±0.0785	0.8395 ±0.0826	0.7862 ±0.0800	0.8345 ±0.0805	0.8018 ±0.0693	**0.8658**** ±0.0804

Bold and underline denote the best and second-best results, respectively. The markers *, **, and *** indicate statistically significant improvement of DAMoE over the strongest competing method under a paired Wilcoxon signed-rank test, with $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

Table 8
Cross-dataset zero-shot quantitative comparison on the Pavia Center HSI pansharpening test set using models trained on Botswana.

Metrics	Traditional		SS modeling			SF modeling			FSS modeling
	CNMF	MTF-GLP	MDANet	PSRT	MCIFNet	DSNet	DCINN	SFIGNet	DAMoE (Ours)
PSNR (↑)	**29.4520** ±0.3558	26.2928 ±0.0333	23.3978 ±0.1224	6.9585 ±0.3926	8.2061 ±0.1258	21.5439 ±0.0078	26.3736 ±0.0249	15.9474 ±0.0328	27.6125 ±0.1448
SAM (↓)	**6.6936** ±0.1034	9.9042 ±0.2641	19.6452 ±0.0886	59.5461 ±0.5865	93.7230 ±0.0688	21.3701 ±0.5020	10.9245 ±0.0994	90.9419 ±0.1531	9.5018 ±0.0607
ERGAS (↓)	**6.3537** ±0.1907	8.8554 ±0.1007	3353.1491 ±3305.9315	255.9441 ±100.9980	260.6839 ±84.6981	179.4286 ±126.3841	13.4440 ±0.7859	2163.4294 ±1129.3995	8.7262 ±0.5285

Bold and underline denote the best and second-best results, respectively.

together, the absolute results and degradation analysis indicate that DAMoE provides a better balance between reconstruction accuracy and cross-sensor robustness.

For HSI pansharpening, Table 8 reports a more challenging zero-shot transfer experiment, where the models trained on the Botswana benchmark are directly evaluated on the Pavia Center test set without fine-tuning. This setting involves a severe domain shift: the spatial resolution changes from approximately 30 m to 1.2 m, and the scene content changes from coarse natural landscapes to fine-grained urban structures. As a result, most learning-based methods suffer substantial performance degradation, indicating that models trained on one HSI distribution may not directly generalize to another dataset with markedly different spatial scale, spectral dimensionality, and land-cover granularity. Traditional matrix-factorization and MTF-based methods are less affected by learned domain bias, with CNMF achieving the best overall scores in this stress-test setting. Nevertheless, among all trainable deep models, DAMoE achieves the best PSNR, SAM, and ERGAS, and ranks second overall across all three metrics. This suggests that the proposed prior-guided sparse routing does not eliminate extreme cross-dataset degradation, but provides better robustness than other learning-based baselines under a large HSI domain shift.

4.7. Downstream segmentation validation

We further examine whether the fusion gains of DAMoE extend beyond image-level reconstruction by using unsupervised segmentation as a downstream validation task. Fig. 11 shows that the segmentation maps derived from DAMoE are more spatially coherent and better aligned with major scene structures than those obtained from the compared fusion results, indicating better preservation of scene structure for subsequent interpretation.

Table 9 summarizes the quantitative results in terms of overall accuracy. DAMoE achieves the best overall accuracy under both GMM-SS and SLIC-Agglo, with a clearer advantage under SLIC-Agglo. This suggests that region-level segmentation is more sensitive than pixel-level clustering to differences among fusion methods, and that the fused outputs of DAMoE retain more useful semantic information for downstream analysis. Since the segmentation is unsupervised, the obtained cluster or region labels are matched to the reference classes in a post hoc manner before evaluation.

Fig. 12 further compares the confusion matrices of DAMoE and MCIFNet. Under GMM-SS, the two methods show broadly similar confusion patterns. Under SLIC-Agglo, DAMoE exhibits cleaner diagonal responses and fewer off-diagonal confusions, especially for Classes 3 and 6, while maintaining strong recognition for Classes 4, 5, 7, and 8. Although Class 2 remains challenging for both methods, the overall confusion structure indicates that DAMoE provides supportive evidence that the fused images preserve structures useful for unsupervised downstream analysis.

5. Model analysis

5.1. Effect of the multi-domain backbone

Fig. 13(a)–(b) shows the contribution of the proposed multi-domain backbone. When the model is reduced to a single-branch dense variant without MoE (*Single-Branch w/o MoE*), both ERGAS and HQNR degrade noticeably, indicating that increasing depth alone is insufficient for multimodal EO fusion. Splitting the network into spatial and spectral branches without MoE (*Spa-SpeBranch w/o MoE*) already improves performance, suggesting that decoupled spatial–spectral processing is beneficial. The effect of the frequency branch is even more evident.

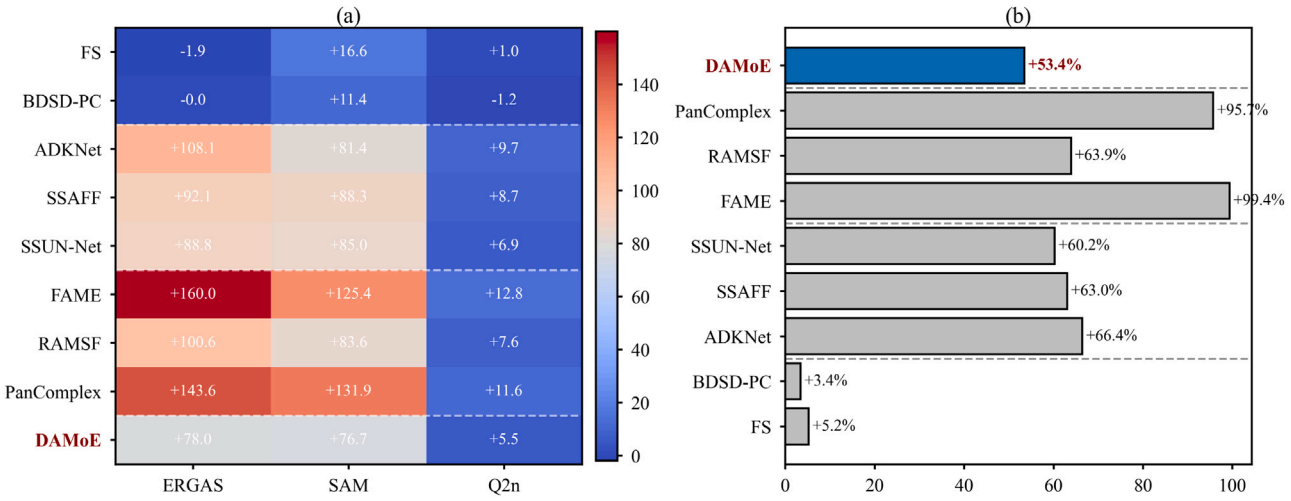

Fig. 10. Cross-sensor generalization from WV-III to WV-II. (a) Metric-wise performance degradation for ERGAS, SAM, and Q2n relative to source-domain results. (b) Mean degradation across the three metrics. Positive values indicate degradation and negative values indicate improvement. DAMoE is highlighted for comparison.

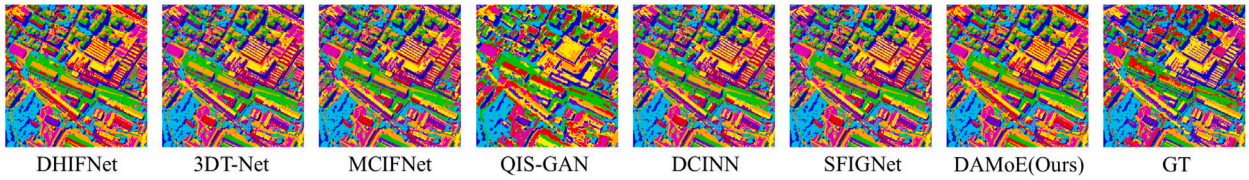

Fig. 11. Downstream unsupervised segmentation maps generated from fused outputs using the GMM-SS and SLIC-Agglo algorithms.

Table 9
Overall segmentation accuracy under GMM-SS and SLIC-Agglo for downstream validation of fused outputs.

	Traditional			SS modeling			SF modeling			FSS modeling
Class	SFIM	CNMF	GFPCA	DHIFNet	3DT-Net	MCIFNet	QIS-GAN	DCINN	SFIGNet	DAMoE
Overall	0.4467/0.5584	0.6044/0.6145	0.4257/0.4924	0.6804/0.6779	0.6884/0.6732	0.6906/0.7136	0.6885/0.7103	0.6887/0.7278	0.6934/0.7250	**0.6936/0.7405**

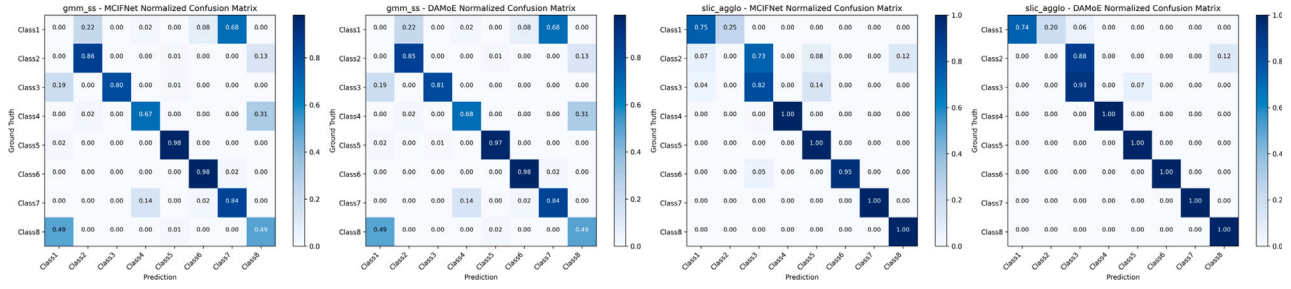

Fig. 12. Normalized confusion matrices for downstream segmentation validation of MCIFNet and DAMoE under GMM-SS and SLIC-Agglo. Rows denote reference classes and columns denote predicted classes.

Removing it (*w/o Freq*) degrades performance relative to the full model, whereas keeping only the frequency branch (*only Freq*) fails to preserve both spatial sharpness and spectral fidelity. By contrast, combining the frequency branch with either the spatial or spectral branch (*Freq+Spa* and *Freq+Spe*) consistently improves both metrics, and the full three-branch model achieves the best results. This indicates that the three branches are complementary rather than interchangeable. Fig. 13(b) further examines branch ordering. Driving all three branches in parallel from the shallow feature (*All Parallel*) is suboptimal, and placing the frequency branch after the spatial–spectral branches (*Freq after Spa/Spe*) remains inferior to our design. The best performance is obtained when the frequency MoE operates first and the spatial/spectral MoEs refine a shared frequency-enhanced representation, supporting the proposed *frequency-front, spatial–spectral-back* design.

5.2. Effect of prior-guided routing

Fig. 13(c)–(f) evaluates the effect of prior-guided routing. Removing all priors (*w/o Prior*) markedly degrades both ERGAS and HQNR, while replacing them with random vectors (*Random Prior*) yields only limited recovery and remains clearly inferior to the full model. This suggests that the gain does not come from a larger gating network alone, but from physically meaningful EO priors that guide expert selection. The branch-wise ablations show the same pattern. In Fig. 13(d), removing all frequency priors (*No Freq Priors*) or keeping only one of r_{HF}, a_{dir}, and r_{HH} consistently weakens performance, whereas their combination gives the best result. Fig. 13(e) shows a similar trend for the spatial branch: removing the full spatial prior or any individual component degrades performance. The same holds in Fig. 13(f) for the spectral

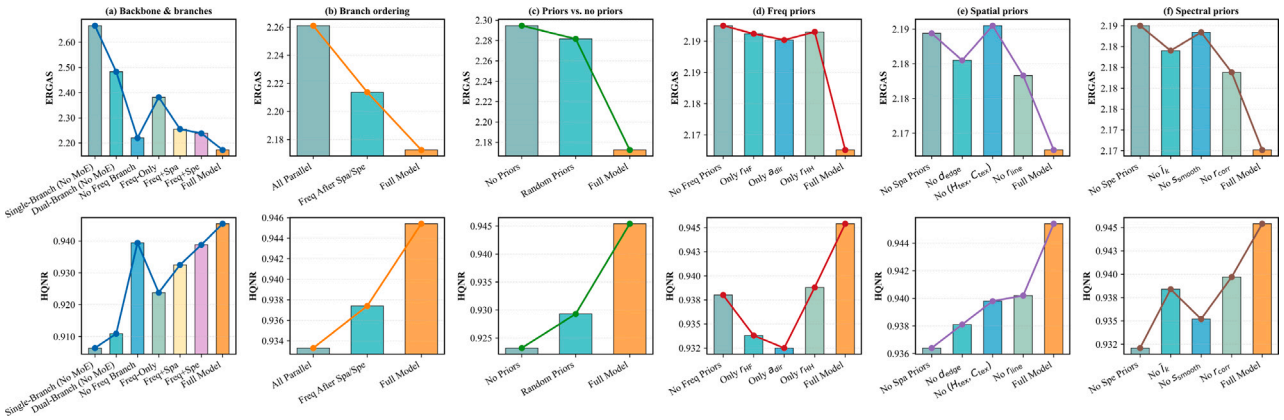

Fig. 13. Ablation study of DAMoE on the multispectral pansharpening setting. (a) Backbone and branch composition. (b) Branch ordering. (c) Effect of removing or randomizing all priors. (d) Frequency priors. (e) Spatial priors. (f) Spectral priors. Bars and lines report ERGAS and HQNR, respectively.

branch, where excluding the full spectral prior or individual terms, including the mean indices $\bar{I}_k$, smoothness s_{smooth}, and correlation r_{corr}, also reduces performance. Taken together, these results indicate that frequency, spatial, and spectral priors provide complementary routing cues and consistently improve domain-aware expert selection.

5.3. Effect of expert capacity and routing sparsity

Fig. 14(a) examines the effect of expert capacity and routing sparsity. As the number of experts N increases from 1 to about 5 and the top-k routing size grows from 2 to 4, ERGAS decreases and HQNR increases steadily. Beyond this range, the gains gradually saturate and may slightly deteriorate. This suggests that the proposed architecture benefits from a moderate number of experts and active routes, whereas overly large N or k introduces diminishing returns and may reintroduce redundancy similar to dense models. Together with the results in Fig. 13(a), where MoE-based variants consistently outperform dense counterparts under comparable backbones, these observations support the intended accuracy–efficiency trade-off of conditional computation and indicate that the model is not overly sensitive to the exact choice of N and k within a reasonable range.

5.4. Effect of dual regularization

Fig. 14(b) shows the effect of the two regularization weights λ and γ, which control the expert-usage balancing term and the expert-diversity regularization, respectively. When either weight is too small, the model becomes under-constrained, and both ERGAS and HQNR degrade due to expert collapse or homogenization. When either is too large, the regularization term begins to dominate the reconstruction objective and again harms performance. The best results appear in a moderate regime around $\lambda \approx 10^{-1}$ and $\gamma \approx 10^{-1}$–10^{-2}, where ERGAS reaches its minimum and HQNR its maximum. The relatively flat performance surface around this region further suggests that dual regularization improves not only final accuracy, but also training stability over a broader range of hyperparameter settings.

5.5. Efficiency analysis

Table 10 and Fig. 15 show that DAMoE achieves a favorable efficiency–performance trade-off. As reported in Table 10, it is among the lightest compared deep models and delivers the fastest inference. Fig. 15 further confirms this result from a five-dimensional view. In both the reduced-scale and full-scale comparisons, DAMoE remains highly competitive in reconstruction quality while maintaining low model size, low computational cost, and short runtime. Although RAMSF is slightly more compact, its full-scale quality is lower, whereas the other baselines are inferior in one or more dimensions. Overall, these results support the advantage of sparse conditional computation in achieving strong fusion quality without a large dense backbone.

6. Discussion

6.1. Relation to existing EO fusion literature

DAMoE is distinct from existing EO fusion methods in how it organizes computation. Relative to *SS modeling* methods, it reduces reliance on uniformly dense processing through conditional computation. Relative to *SF modeling* methods, it uses a frequency-enhanced representation as the shared basis for subsequent spatial and spectral refinement rather than as an auxiliary cue. Relative to other mixed methods, its key difference is that lightweight EO priors directly influence routing decisions. The present results therefore suggest that priors are most effective when they guide computation itself.

6.2. Implications for recurring EO data gaps

The evaluated tasks can be interpreted under one common view of recurring EO data gaps, including missing spatial detail, joint spatial–spectral gaps, cross-resolution spectral enhancement, and reduced optical availability. Under this view, the most consistent result is that DAMoE is especially effective when structural detail and spectral fidelity must be recovered simultaneously. This explains why the gains are strongest on the HSI tasks, remain clear on MSI pansharpening, and are still encouraging, though more cautiously interpreted, on SAR-assisted optical reconstruction.

6.3. Extensibility to larger-area applications and emerging EO data sources

The broader value of DAMoE lies in its potential as a reusable sparse-computation strategy for multimodal EO integration. Since different image regions may require different computational pathways, prior-guided sparse routing is naturally suitable for large-scene and heterogeneous-area processing. The same principle can also be extended to new sensing combinations by re-instantiating the prior descriptors, encoders, and output heads. Therefore, DAMoE should be viewed as a reusable conditional-computation paradigm rather than a fixed task-specific model.

A direct extension is to incorporate modality- and application-specific priors. The current implementation mainly uses generic frequency-energy, spatial-structure, and spectral–statistical cues to maintain consistency across the four evaluated tasks. For SAR-assisted optical reconstruction, SAR-specific descriptors such as VV/VH backscatter

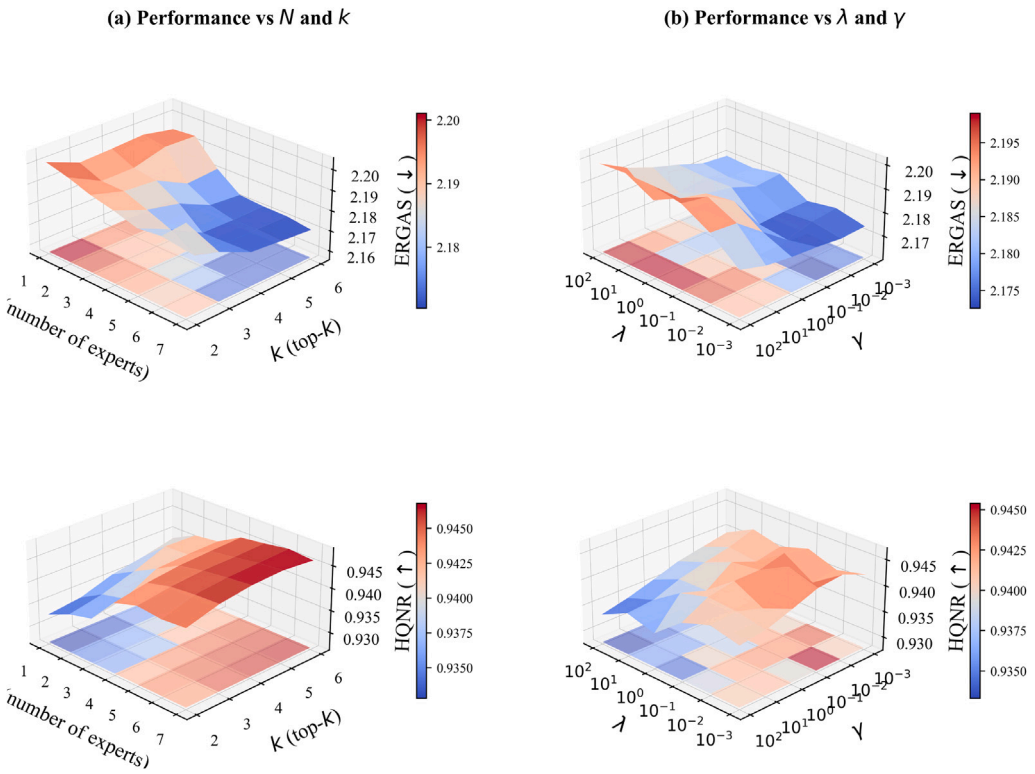

Fig. 14. Hyperparameter sensitivity of DAMoE. (a) ERGAS and HQNR under different numbers of experts (N) and active experts (k). (b) ERGAS and HQNR under different values of the load-balancing weight (λ) and diversity weight (γ).

Table 10
Computational cost in terms of FLOPs, the number of parameters (Params), and inference time (Time).

Metrics	Traditional		SS modeling			SF modeling			FSS modeling
	FS	BDSD-PC	ADKNet	SSAFF	SSUN-Net	FAME	RAMSF	PanComplex	DAMoE (Ours)
FLOPs (G)	–	–	0.72	8.84	2.84	4.72	2.60	1.58	1.54
Params (M)	–	–	0.60	1.08	0.68	0.58	0.24	0.46	0.39
Time (s)	–	–	0.036	0.018	0.400	0.033	0.017	0.020	0.013

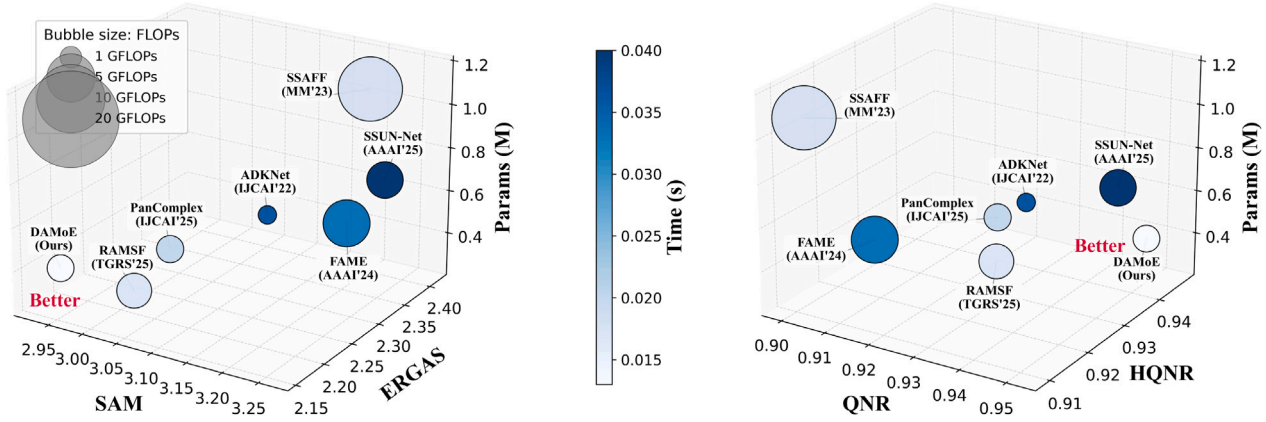

Fig. 15. Comparison of RST/FST performance, model parameters, FLOPs, and inference time. In the left panel, the three axes represent SAM, ERGAS, and Params; in the right panel, they represent QNR, HQNR, and Params. Bubble size denotes FLOPs, and color denotes inference time.

statistics, polarization ratios, local coefficient of variation, speckle-related texture, and roughness measures could provide additional routing cues when optical observations are degraded. For HSI-oriented applications, spectral priors can also be tailored to the target interpretation problem, such as vegetation and red-edge indices for vegetation monitoring, water-related normalized differences for water mapping, built-up indicators for urban analysis, and absorption-depth or continuum-removed descriptors for mineral mapping. These extensions would enrich the prior pool with more explicit domain knowledge while keeping the general routing mechanism unchanged.

The sparse-routing paradigm may also be extended to other EO and geospatial modalities, such as thermal infrared, nighttime light, LiDAR

or point-cloud data, street-view imagery, and crowdsourced social sensing data. In these cases, lightweight priors could be designed from modality-specific cues, including thermal contrast, brightness stability, height or roughness statistics, semantic distributions, geolocation uncertainty, and sampling density. For data with mismatched resolutions or irregular structures, modality-specific encoders could project them into a common latent grid or token space before applying prior-guided routing. These possibilities are beyond the scope of this study and require future empirical validation.

6.4. Limitations and reproducibility

Several limitations remain. The current prior descriptors are handcrafted to keep the routing process compact, stable, and physically interpretable. This design allows expert activation to be explicitly related to frequency-energy, spatial-structure, and spectral–statistical cues, which is an important motivation of DAMoE. Nevertheless, the optimality of fixed priors is not guaranteed under stronger cross-task or cross-sensor distribution shifts, especially when spectral response functions, spatial resolutions, noise characteristics, or observation geometries differ substantially. Learnable prior encoders may provide greater adaptivity, but they may also reduce the physical interpretability of the routing process if the learned descriptors are not properly constrained. Therefore, future work will not simply replace handcrafted priors with fully learned ones, but will explore interpretable adaptive priors, for example by combining physically defined descriptors with weakly learnable calibration or sensor-specific weighting modules. We will further examine this issue using broader real hyperspectral observations, including data from the forthcoming Dongfang Huiyan intelligent hyperspectral satellites developed by our team, so as to evaluate the interpretability, robustness, and transferability of DAMoE under new sensor conditions. The current routing mechanism also relies on compact scene- or patch-level descriptors and may not capture finer local heterogeneity. In addition, transfer evidence is still limited, and the SAR-assisted setting remains intrinsically harder to interpret because of temporal mismatch between cloudy inputs and reference images. Even so, the study remains reproducible, supported by public datasets, a consistent experimental pipeline, and released code.

7. Conclusions

This study presents DAMoE, a prior-guided sparse multi-domain framework for multimodal Earth observation data gaps. Across MSI pansharpening, HSI pansharpening, HSI super-resolution, and SAR-assisted optical reconstruction, DAMoE achieves consistently strong fusion quality while maintaining a favorable accuracy–efficiency trade-off. Cross-sensor/cross-dataset transfer and downstream segmentation results further provide supportive evidence for structural preservation. More broadly, the value of DAMoE lies in its potential as a reusable sparse-computation paradigm for multimodal EO integration under recurring data gaps. Nevertheless, DAMoE is currently most suitable for co-registered multimodal observations with complementary spatial, spectral, or all-weather information, and its reliability may decrease under severe misregistration, large temporal mismatch, or strong sensor-response differences. Future work will focus on learnable and uncertainty-aware prior modeling, broader cross-sensor and cross-dataset generalization evaluation, and extension to additional EO sensing combinations beyond the optical–HSI–SAR settings considered here.

CRediT authorship contribution statement

Chuang Liu: Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Jianhua Guo:** Writing – review & editing, Visualization, Formal analysis, Funding acquisition. **Yingdong Pi:** Validation, Software, Writing – review & editing, Funding acquisition. **Xiao Wu:** Writing – review & editing, Validation, Investigation. **Zhiqi Zhang:** Validation, Software. **Ru Chen:** Validation, Software, Formal analysis. **Xinyi Wang:** Validation, Formal analysis. **Mi Wang:** Resources, Project administration, Funding acquisition, Data curation, Conceptualization.

Funding

This research was supported in part by the National Science Fund for Distinguished Young Scholars (Grant No. 62425102), the National Natural Science Foundation of China Major Program (Grant No. 42192583), the CAS Hundred Talents Program (Grant No. E5Z105020F), and LIESMARS Special Research Funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Code (including training and evaluation scripts, configuration files, and fixed random seeds) is available at: https://github.com/JUSTM0VE0N/RSMIF_Project. We will archive the exact code snapshot upon acceptance. All datasets used in this study are publicly available from their official providers; preprocessing and dataset split scripts are provided in the repository to reproduce the reported experiments. Pretrained weights and inference demos are provided in the same repository.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34 (11), 2274–2282. http://dx.doi.org/10.1109/TPAMI.2012.120.

Arienzo, A., Vivone, G., Garzelli, A., Alparone, L., Chanussot, J., 2022. Full-resolution quality assessment of pansharpening: Theoretical and hands-on approaches. IEEE Geosci. Remote. Sens. Mag. 10 (3), 168–201. http://dx.doi.org/10.1109/MGRS.2022.3170092.

Cai, J., Huang, B., Fung, T., 2022. Progressive spatiotemporal image fusion with deep neural networks. Int. J. Appl. Earth Obs. Geoinf. 108, 102745. http://dx.doi.org/10.1016/j.jag.2022.102745.

Cao, X., Fu, X., Hong, D., Xu, Z., Meng, D., 2022. PanCSC-net: A model-driven deep unfolding method for pansharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–13. http://dx.doi.org/10.1109/TGRS.2021.3115501.

Chen, Y., Li, Y., Wang, T., Chen, Y., Fang, F., 2024. DPDU-net: Double prior deep unrolling network for pansharpening. Remote. Sens. 16 (12), http://dx.doi.org/10.3390/rs16122141.

Chen, L., Vivone, G., Qin, J., Chanussot, J., Yang, X., 2024. Spectral–spatial transformer for hyperspectral image sharpening. IEEE Trans. Neural Netw. Learn. Syst. 35 (11), 16733–16747. http://dx.doi.org/10.1109/TNNLS.2023.3297319.

Choi, J., Yu, K., Kim, Y., 2011. A new adaptive component-substitution-based satellite image fusion by using partial replacement. IEEE Trans. Geosci. Remote Sens. 49 (1), 295–309. http://dx.doi.org/10.1109/TGRS.2010.2051674.

Ciotola, M., Vitale, S., Mazza, A., Poggi, G., Scarpa, G., 2022. Pansharpening by convolutional neural networks in the full resolution framework. IEEE Trans. Geosci. Remote Sens. 60, 1–17. http://dx.doi.org/10.1109/TGRS.2022.3163887.

Deng, S.Q., Deng, L.J., Wu, X., Ran, R., Hong, D., Vivone, G., 2023. PSRT: Pyramid shuffle-and-reshuffle transformer for multispectral and hyperspectral image fusion. IEEE Trans. Geosci. Remote Sens. 61, 1–15. http://dx.doi.org/10.1109/TGRS.2023.3244750.

Deng, L.j., Vivone, G., Paoletti, M.E., Scarpa, G., He, J., Zhang, Y., Chanussot, J., Plaza, A., 2022. Machine learning in pansharpening: A benchmark, from shallow to deep networks. IEEE Geosci. Remote. Sens. Mag. 10 (3), 279–315. http://dx.doi.org/10.1109/MGRS.2022.3187652.

Dong, R., Zhang, L., Li, W., Yuan, S., Gan, L., Zheng, J., Fu, H., Mou, L., Zhu, X.X., 2023. An adaptive image fusion method for sentinel-2 images and high-resolution images with long-time intervals. Int. J. Appl. Earth Obs. Geoinf. 121, 103381. http://dx.doi.org/10.1016/j.jag.2023.103381.

Ebel, P., Meraner, A., Schmitt, M., Zhu, X.X., 2021. Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery. IEEE Trans. Geosci. Remote Sens. 59 (7), 5866–5878. http://dx.doi.org/10.1109/TGRS.2020.3024744.

Fang, S., Gan, H., 2025. SSUN-net: Spatial-spectral prior-aware unfolding network for pan-sharpening. AAAI, In: Proc. AAAI Conf. Artif. Intell., vol. 39, pp. 32296–32305. http://dx.doi.org/10.1609/aaai.v39i3.32296.

Fauvel, M., Tarabalka, Y., Benediktsson, J.A., Chanussot, J., Tilton, J.C., 2013. Advances in spectral-spatial classification of hyperspectral images. Proc. IEEE 101 (3), 652–675. http://dx.doi.org/10.1109/JPROC.2012.2197589.

Garzelli, A., Nencini, F., 2009. Hypercomplex quality assessment of multi/hyperspectral images. IEEE Geosci. Remote. Sens. Lett. 6 (4), 662–665. http://dx.doi.org/10.1109/LGRS.2009.2022650.

Guan, P., Lam, E.Y., 2022. Multistage dual-attention guided fusion network for hyperspectral pansharpening. IEEE Trans. Geosci. Remote Sens. 60, 1–14. http://dx.doi.org/10.1109/TGRS.2021.3114552.

Guo, Z., Lei, J., Zhou, S., Wang, B., Kasabov, N.K., 2025. A multispectral pansharpening method based on CNN-DI network with mixture of experts. Appl. Soft Comput. 182, 113499. http://dx.doi.org/10.1016/j.asoc.2025.113499.

He, X., Yan, K., Li, R., Xie, C., Zhang, J., Zhou, M., 2023. Pyramid dual domain injection network for pan-sharpening. In: Proc. IEEE/CVF Int. Conf. Comput. Vis.. ICCV, pp. 12862–12871. http://dx.doi.org/10.1109/ICCV51070.2023.01186.

He, X., Yan, K., Li, R., Xie, C., Zhang, J., Zhou, M., 2024. Frequency-adaptive pan-sharpening with mixture of experts. AAAI, In: Proc. AAAI Conf. Artif. Intell., vol. 38, pp. 2121–2129. http://dx.doi.org/10.1609/aaai.v38i3.27984.

Huang, T., Dong, W., Wu, J., Li, L., Li, X., Shi, G., 2022. Deep hyperspectral image fusion network with iterative spatio-spectral regularization. IEEE Trans. Comput. Imaging 8, 201–214. http://dx.doi.org/10.1109/TCI.2022.3152700.

Huang, J., Huang, R., Xu, J., Peng, S., Duan, Y., Deng, L.J., 2025. Wavelet-assisted multi-frequency attention network for pansharpening. Proc. AAAI Conf. Artif. Intell. (AAAI) 39 (4), 3662–3670. http://dx.doi.org/10.1609/aaai.v39i4.32381.

Jiang, H., Chen, Z., 2025. Hyperspectral pansharpening with transformer-based spectral diffusion priors. In: Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops. WACVW, pp. 544–553. http://dx.doi.org/10.1109/WACVW65960.2025.00066.

Jozdani, S., Chen, D., Pouliot, D., Alan Johnson, B., 2022. A review and meta-analysis of generative adversarial networks and their applications in remote sensing. Int. J. Appl. Earth Obs. Geoinf. 108, 102734. http://dx.doi.org/10.1016/j.jag.2022.102734.

Li, J., Zheng, K., Gao, L., Han, Z., Li, Z., Chanussot, J., 2025. Enhanced deep image prior for unsupervised hyperspectral image super-resolution. IEEE Trans. Geosci. Remote Sens. 63, 1–18. http://dx.doi.org/10.1109/TGRS.2025.3531646.

Li, J., Zheng, K., Gao, L., Ni, L., Huang, M., Chanussot, J., 2024. Model-informed multistage unsupervised network for hyperspectral image super-resolution. IEEE Trans. Geosci. Remote Sens. 62, 1–17. http://dx.doi.org/10.1109/TGRS.2024.3391014.

Li, J., Zheng, K., Liu, W., Li, Z., Yu, H., Ni, L., 2023. Model-guided coarse-to-fine fusion network for unsupervised hyperspectral image super-resolution. IEEE Geosci. Remote. Sens. Lett. 20, 1–5. http://dx.doi.org/10.1109/LGRS.2023.3309854.

Liao, W., Huang, X., Van Coillie, F., Gautama, S., Pižurica, 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 J. Sel. Top. Appl. Earth Obs. Remote. Sens. 8 (6), 2984–2996. http://dx.doi.org/10.1109/JSTARS.2015.2420582.

Liao, D., Shi, C., Wang, L., 2023. A spectral–spatial fusion transformer network for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 61, 1–16. http://dx.doi.org/10.1109/TGRS.2023.3286950.

Liu, J.G., 2000. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. Int. J. Remote Sens. 21 (18), 3461–3472. http://dx.doi.org/10.1080/014311600750037499.

Liu, C., Sun, Y., Zhang, X., Xu, Y., Lei, L., Kuang, G., 2025a. OSHFNet: A heterogeneous dual-branch dynamic fusion network of optical and SAR images for land use classification. Int. J. Appl. Earth Obs. Geoinf. 141, 104609. http://dx.doi.org/10.1016/j.jag.2025.104609.

Liu, C., Zhang, Z., Wang, M., 2025b. RAMSF: A novel generic framework for optical remote sensing multimodal spatial-spectral fusion. IEEE Trans. Geosci. Remote Sens. 63, 1–22. http://dx.doi.org/10.1109/TGRS.2025.3552937.

Liu, C., Zhang, Z., Wang, M., Xiang, S., Xie, G., 2025. A novel cross fusion model with fine-grained detail reconstruction for remote sensing image pan-sharpening. Geo-Spat. Inf. Sci. 28 (4), 1520–1548. http://dx.doi.org/10.1080/10095020.2024.2416899.

Luo, C., Li, D., Ma, X., Lu, X., Wang, Z., Tan, J., Fu, X., 2025. PanComplex: Leveraging complex-valued neural networks for enhanced pansharpening. In: Proc. Int. Joint Conf. Artif. Intell.. IJCAI, International Joint Conferences on Artificial Intelligence Organization, pp. 1702–1710. http://dx.doi.org/10.24963/ijcai.2025/190.

Ma, Q., Jiang, J., Liu, X., Ma, J., 2023. Learning a 3D-CNN and transformer prior for hyperspectral image super-resolution. Inf. Fus. 100, 101907. http://dx.doi.org/10.1016/j.inffus.2023.101907.

Meraner, A., Ebel, P., Zhu, X.X., Schmitt, M., 2020. Cloud removal in sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. ISPRS J. Photogramm. Remote Sens. 166, 333–346. http://dx.doi.org/10.1016/j.isprsjprs.2020.05.013.

Palsson, F., Sveinsson, J.R., Ulfarsson, M.O., 2014. A new pansharpening algorithm based on total variation. IEEE Geosci. Remote. Sens. Lett. 11 (1), 318–322. http://dx.doi.org/10.1109/LGRS.2013.2257669.

Peng, S., Deng, L.J., Hu, J.F., Zhuo, Y., 2022. Source-adaptive discriminative kernels based network for remote sensing pansharpening. In: Proc. Int. Joint Conf. Artif. Intell.. IJCAI, International Joint Conferences on Artificial Intelligence Organization, pp. 1283–1289. http://dx.doi.org/10.24963/ijcai.2022/179.

Restaino, R., Vivone, G., Dalla Mura, M., Chanussot, J., 2016. Fusion of multispectral and panchromatic images based on morphological operators. IEEE Trans. Image Process. 25 (6), 2882–2895. http://dx.doi.org/10.1109/TIP.2016.2556944.

Rustowicz, R.M., Cheong, R., Wang, L., Ermon, S., Burke, M., Lobell, D., 2019. Semantic segmentation of crop type in africa: A novel dataset and analysis of deep learning methods. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops. CVPRW.

Vivone, G., 2019. Robust band-dependent spatial-detail approaches for panchromatic sharpening. IEEE Trans. Geosci. Remote Sens. 57 (9), 6421–6433. http://dx.doi.org/10.1109/TGRS.2019.2906073.

Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G.A., Restaino, R., Wald, L., 2015. A critical comparison among pansharpening algorithms. IEEE Trans. Geosci. Remote Sens. 53 (5), 2565–2586. http://dx.doi.org/10.1109/TGRS.2014.2361734.

Vivone, G., Deng, L.J., Deng, S., Hong, D., Jiang, M., Li, C., Li, W., Shen, H., Wu, X., Xiao, J.L., Yao, J., Zhang, M., Chanussot, J., García, S., Plaza, A., 2025. Deep learning in remote sensing image fusion: Methods, protocols, data, and future perspectives. IEEE Geosci. Remote. Sens. Mag. 13 (1), 269–310. http://dx.doi.org/10.1109/MGRS.2024.3495516.

Vivone, G., Restaino, R., Chanussot, J., 2018. Full scale regression-based injection coefficients for panchromatic sharpening. IEEE Trans. Image Process. 27 (7), 3418–3431. http://dx.doi.org/10.1109/TIP.2018.2819501.

Wang, B., Chen, J., Wang, H., Tang, Y., Chen, J., Jiang, Y., 2024. A spectral and spatial transformer for hyperspectral remote sensing image super-resolution. Int. J. Digit. Earth 17 (1), 2313102. http://dx.doi.org/10.1080/17538947.2024.2313102.

Wang, W., Deng, L.J., Ran, R., Vivone, G., 2024. A general paradigm with detail-preserving conditional invertible network for image fusion. Int. J. Comput. Vis. 132, 1029–1054. http://dx.doi.org/10.1007/s11263-023-01924-5.

Wang, X., Yin, S., Xu, X., Mei, Y., Huang, Y., Tan, K., 2025. MHFu-former: A multispectral and hyperspectral image fusion transformer. Int. J. Appl. Earth Obs. Geoinf. 143, 104843. http://dx.doi.org/10.1016/j.jag.2025.104843.

Wen, X., Ma, H., Li, L., 2025. A three-branch pansharpening network based on spatial and frequency domain interaction. Remote. Sens. 17 (1), http://dx.doi.org/10.3390/rs17010013.

Wen, T., Wang, H., Wang, L., 2025. Dual-branch spatial spectral transformer with similarity propagation for hyperspectral image classification. Remote. Sens. 17 (14), http://dx.doi.org/10.3390/rs17142386.

Wu, Y., Dai, J., Ma, Z., Zhang, T., 2025. A diffusion model for hyperspectral and multispectral fusion guided by prior knowledge. Int. J. Appl. Earth Obs. Geoinf. 144, 104923. http://dx.doi.org/10.1016/j.jag.2025.104923.

Xu, Y., Du, B., Zhang, L., Cerra, D., Pato, M., Carmona, E., Prasad, S., Yokoya, N., Hänsch, R., Le Saux, B., 2019. Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 12 (6), 1709–1724. http://dx.doi.org/10.1109/JSTARS.2019.2911113.

Xu, F., Shi, Y., Ebel, P., Yu, L., Xia, G.S., Yang, W., Zhu, X.X., 2022. GLF-CR: SAR-enhanced cloud removal with global-local fusion. ISPRS J. Photogramm. Remote Sens. 192, 268–278. http://dx.doi.org/10.1016/j.isprsjprs.2022.08.002.

Yang, Y., Li, M., Huang, S., Lu, H., Tu, W., Wan, W., 2023. Multi-scale spatial-spectral attention guided fusion network for pansharpening. In: Proc. 31st ACM Int. Conf. Multimed.. ACM MM '23, Association for Computing Machinery, New York, NY, USA, pp. 3346–3354. http://dx.doi.org/10.1145/3581783.3613814.

Yokoya, N., Yairi, T., Iwasaki, A., 2012. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE Trans. Geosci. Remote Sens. 50 (2), 528–537. http://dx.doi.org/10.1109/TGRS.2011.2161320.

Zhou, M., et al., 2022. Spatial-frequency domain information integration for pan-sharpening. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (Eds.), Proc. 17th Eur. Conf. Comput. Vis. (ECCV 2022). In: LNCS, vol. 13678, Springer, Cham, pp. 274–291. http://dx.doi.org/10.1007/978-3-031-19797-0_16.

Zhu, C., Dai, R., Gong, L., Gao, L., Ta, N., Wu, Q., 2023a. An adaptive multi-perceptual implicit sampling for hyperspectral and multispectral remote sensing image fusion. Int. J. Appl. Earth Obs. Geoinf. 125, 103560. http://dx.doi.org/10.1016/j.jag.2023.103560.

Zhu, C., Deng, S., Song, X., Li, Y., Wang, Q., 2025a. Mamba collaborative implicit neural representation for hyperspectral and multispectral remote sensing image fusion. IEEE Trans. Geosci. Remote Sens. 63, 1–15. http://dx.doi.org/10.1109/TGRS.2025.3537638.

Zhu, C., Deng, S., Zhou, Y., Deng, L.J., Wu, Q., 2023b. QIS-GAN: A lightweight adversarial network with quadtree implicit sampling for multispectral and hyperspectral image fusion. IEEE Trans. Geosci. Remote Sens. 61, 1–15. http://dx.doi.org/10.1109/TGRS.2023.3332176.

Zhu, C., Song, X., Li, Y., Deng, S., Zhang, T., 2025b. A spatial-frequency dual-domain implicit guidance method for hyperspectral and multispectral remote sensing image fusion based on Kolmogorov–Arnold network. Inf. Fus. 123, 103261. http://dx.doi.org/10.1016/j.inffus.2025.103261.

Zhuo, Y.W., Zhang, T.J., Hu, J.F., Dou, H.X., Huang, T.Z., Deng, L.J., 2022. A deep-shallow fusion network with multidetail extractor and spectral attention for hyperspectral pansharpening. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 15, 7539–7555. http://dx.doi.org/10.1109/JSTARS.2022.3202866.