Euler-Mamba: Learning Resolution-Invariant State Evolution in Complex Polar Frequency Space for Asymmetric Pansharpening

Mengting Ma[1]*, **Anqi Zhu**[2]*, **Yizhen Jiang**[2], **Chengfu Ding**[2], **Wei Zhang**[2]†

[1]School of Computer Science and Technology
[2]School of Software Technology
Zhejiang University, Hangzhou, 310027, China
{mtma, 22551006, 22451166, dingcf, cstzhangwei}@zju.edu.cn

Abstract

The goal of asymmetric pansharpening is to synthesize a high-resolution multispectral (HR-MS) image by integrating panchromatic (PAN) and low-resolution multispectral (LR-MS) images. While Mamba efficiently captures long-range dependencies, spatial scanning does not explicitly distinguish the structural and spectral roles of PAN and LR-MS features. To address this, we propose Euler-Mamba, a physics-inspired framework that combines modality-specific phase and magnitude modulation with frequency-ordered state propagation. The Explicit Phase Rotation (EPR) block learns PAN-guided phase offsets while preserving the pointwise LR-MS Fourier magnitude during phase adjustment. The Implicit Magnitude Mapping (IMM) block recalibrates the LR-MS magnitude through channel-dependent weights. The Eulerian Selective Scan (ESS) mechanism then orders the resulting complex features by frequency radius and angle and processes them with an input-dependent selective state space model. The outputs are restored to their original frequency positions and reconstructed by inverse Fourier transformation. This design combines interpretable frequency-domain modulation with long-range dependency modeling. Experiments demonstrate high reconstruction quality with a compact model, together with zero-shot generalization at the evaluated spatial scales and across sensors.

1 Introduction

Pansharpening is a critical task in remote sensing that aims to synthesize a high-resolution multispectral (HR-MS) image by fusing asymmetric modalities. This process involves a panchromatic (PAN) image and a low-resolution multispectral (LR-MS) image, where both sources exhibit an intrinsic frequency discrepancy due to physical sensor constraints [51]. This frequency mismatch fundamentally dictates the asymmetric spatial-spectral attributes observed between the two modalities. Specifically, the PAN modality primarily captures broad-band structural details at the cost of sacrificed spectral resolution due to energy integration across a wide wavelength range, whereas the LR-MS modality preserves refined narrow-band spectral energy while lacking precise spatial topology [21]. Consequently, the objective of pansharpening is to resolve this underlying frequency contrast by extracting the most reliable components from these mixed attributes, thereby reconstructing both spatial consistency and spectral fidelity within a unified complex-valued frequency representation.

*Equal contribution.
†Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Despite the success of deep learning [25, 7, 41], existing paradigms struggle to handle the intrinsic physical asymmetry between modalities, leading to two fundamental problems. **The first problem is the attribute entanglement within grid-based Cartesian representations.** Traditional component substitution (CS) [9, 38, 1] and multi-resolution analysis (MRA) [23, 39, 35] methods rely on rigid linear assumptions that often cause severe spectral distortion. While convolutional neural networks (CNNs) and Transformers have improved feature extraction, they primarily treat pansharpening as a spatial regression task [44, 37, 26]. Models like PNN [33], PanNet [49], and FusionNet [10] operate on discrete Cartesian grids where structural details and spectral energy are inextricably entangled. Because these architectures lack a principled mechanism to isolate spatial topology from spectral intensity, they inevitably suffer from spatial misalignment and spectral distortion when integrating high-frequency PAN cues with low-frequency LR-MS energy.

Beyond these physical limitations, **the second problem stems from the complexity-generalization bottleneck in grid-based modeling.** Traditional high-performance architectures, particularly Transformer-based models such as PanFormer [54], capture long-range dependencies at the cost of quadratic computational complexity, which significantly hinders their application to high-resolution remote sensing imagery. While State Space Models (SSMs) like Mamba [14] have emerged as a linear-complexity alternative, they remain largely confined to discrete grid-based pixel traversal. This symmetric and frequency-agnostic nature lacks the asymmetric modeling flexibility required to distinguish between spatial topology and spectral intensity within a continuous framework, which further limits the potential for resolution-invariant generalization across heterogeneous sensors. Conversely, although Neural Operators [27] offer a solution to such generalization by modeling features as continuous functions, their practical implementation in models like PINO [29] emulates the end-to-end electro-optical sensor process by learning a continuous radiance field through iterative neural integrals. Such computational inflexibility renders them ill-suited for large-scale deployments. Consequently, a critical void remains for a lightweight framework that can break the complexity barrier of Transformers and the generalization limits of grid-based Mamba, achieving practical synergy between Mamba's linear efficiency and asymmetric continuous representational power.

To resolve these problems, we propose **Euler-Mamba**, a physics-inspired framework for asymmetric phase-magnitude modulation and frequency-ordered state propagation. The complex polar frequency space denotes the Fourier representation parameterized by magnitude and phase. Our design assigns PAN-guided phase modulation to the Explicit Phase Rotation (EPR) block and LR-MS magnitude recalibration to the Implicit Magnitude Mapping (IMM) block (**Addressing Problem 1**). The Eulerian Selective Scan (ESS) mechanism orders the modulated coefficients by increasing radius and then angle before applying the Mamba selective scan [14]. The resulting hidden-state sequence captures dependencies across frequency coefficients. This frequency ordering is the task-specific design, while the selective scan retains its input-dependent state-space formulation. The experiments evaluate the resulting balance between reconstruction quality, model compactness, and cross-scale generalization (**Addressing Problem 2**). The major contributions of this work are summarized as follows:

- **PAN-Guided Phase Modulation:** The EPR block learns phase offsets from PAN features to guide structural reconstruction while preserving the pointwise LR-MS magnitude during phase adjustment.
- **Channel-Dependent Magnitude Recalibration:** The IMM block recalibrates LR-MS Fourier magnitude through channel-dependent weights. Together, EPR and IMM assign different operators to the structural and spectral roles of the two modalities.
- **Frequency-Ordered Selective State Propagation:** ESS applies an input-dependent selective SSM to complex features in radial-angular frequency order. The resulting framework achieves high reconstruction quality and cross-scale and cross-sensor generalization in our experiments, with 148.3 K parameters for the four-band configuration.

2 Related Works

2.1 Deep Learning for Asymmetric Pansharpening

Pansharpening is inherently an asymmetric fusion task requiring the alignment of high-resolution structures from PAN with spectral information from LR-MS. While CNN-based models like PNN [33] and Transformer architectures like PanFormer [54] achieve high accuracy, they typically adopt a

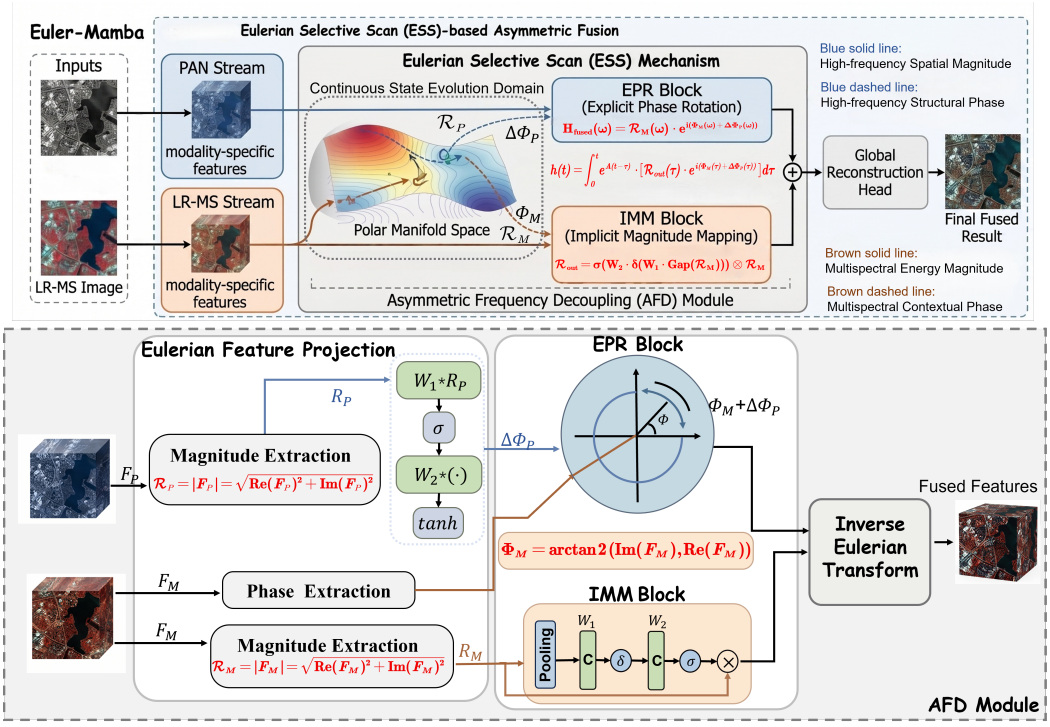

Figure 1: Overview of Euler-Mamba. PAN and LR-MS features are transformed by 2D FFT. EPR performs PAN-guided phase modulation, IMM recalibrates LR-MS magnitude, and ESS processes the resulting coefficients in radial-angular frequency order for reconstruction. In this schematic, "Polar Manifold Space" denotes the complex polar frequency representation, and "Inverse Eulerian Transform" denotes inverse frequency ordering, inverse shifting, and IFFT.

symmetric fusion paradigm by concatenating heterogeneous features into a unified space. This black-box approach ignores the fundamental asymmetry between geometric topology and spectral energy, often leading to phase drift. Furthermore, frequency-domain fusion methods explicitly integrate spatial and spectral information [37]. Our approach processes the resulting frequency coefficients in a defined state sequence for asymmetric fusion. S3FNet uses separate spatial, spectral, and fusion branches with similarity-guided reconstruction [46]. Related hyperspectral and multispectral fusion methods use low-rank attention and rank-aware routing to model spatial-spectral dependencies [30, 31]. Studies of incomplete multi-view clustering, shadow restoration, and multi-scale object detection address other remote-sensing tasks under heterogeneous or degraded observations [53, 5, 32].

Phase-magnitude decomposition is an established Fourier representation [34]. HFIN [37] separately fuses PAN and LR-MS amplitudes and phases through architecturally matched convolutional branches, followed by inverse Fourier transformation. Euler-Mamba assigns different operators to the two modalities: EPR predicts PAN-guided phase offsets, IMM recalibrates LR-MS magnitude, and ESS propagates states over the resulting frequency-ordered features. Its contribution lies in these modality-specific operators and state propagation mechanism.

2.2 State Space Models and Scanning Paradigms

Mamba-based models have emerged as a linear-complexity alternative for high-resolution remote sensing image fusion [8, 45]. Recent works like Pan-Mamba [16] utilize specific scanning patterns to capture long-range dependencies efficiently. For hyperspectral classification, PyS^{2}CF-Mamba combines spatial and spectral branches with competitive fusion [12]. Nevertheless, existing Mamba variants remain confined to symmetric pixel-traversal strategies that treat the image as a sequence of discrete tokens [36]. These discrete paradigms ignore the explicit separation of Fourier phase and magnitude used here for asymmetric modality alignment. While Mamba mitigates the efficiency

bottlenecks of self-attention, it lacks an explicit frequency-domain representation and remains sensitive to resolution changes across different sensor scales.

2.3 Continuous Representations and Function Space Modeling

To achieve resolution-invariant inference, continuous representations have been explored to learn mappings between infinite-dimensional function spaces [27, 15, 43]. Recent efforts, such as PINO [29], attempt to incorporate sensor physics through iterative integration processes, while NoDiff [19] leverages diffusion-based dynamics for complex distribution modeling. However, the practical scalability of these continuous frameworks is often hindered by prohibitive computational overhead. Specifically, the heavy reliance of integration-based methods on high-density kernel approximations and the dependence of generative approaches on multi-step sampling lead to excessive memory consumption at high resolutions. This leaves a void for a framework that preserves the continuous modeling strength of function-space methods while achieving an efficient, asymmetric reconstruction without the burden of iterative sampling or integration.

3 Methodology

3.1 Overall Architecture

As shown in Fig. 1, Euler-Mamba first extracts PAN and LR-MS features through two separate streams and transforms them into centered frequency grids using 2D FFT. The Asymmetric Frequency Decoupling (AFD) module then assigns different operations to the two modalities. EPR uses PAN features to predict phase offsets for the LR-MS Fourier coefficients, while IMM recalibrates their magnitudes using channel-dependent weights derived from LR-MS features. The adjusted phase and magnitude are combined into complex coefficients, incorporating PAN-derived structural guidance into the LR-MS frequency representation. ESS then arranges the Fourier coefficients from low to high frequencies for sequential processing. A selective SSM processes the resulting sequence to aggregate information across frequency locations. Finally, the output coefficients are restored to their original positions on the two-dimensional frequency grid and transformed back to spatial features by inverse FFT. A global reconstruction head integrates these features to produce the fused HR-MS image.

3.2 Eulerian Selective Scan Mechanism

Selective State Propagation. Following Mamba [14], a selective SSM uses input-dependent parameters $\mathbf{B}_k$, $\mathbf{C}_k$, and Δt_k to propagate hidden states. Here $k \in \{1, \ldots, L\}$ indexes the ordered frequency sequence, with $L = HW$ for an $H \times W$ feature grid. The input coefficient vector is x_k, the hidden state is h_k, $\mathbf{B}_k$ maps the input into the state, and $\mathbf{C}_k$ defines the state readout. The state recurrence is

$$h_k = \overline{\mathbf{A}}_k h_{k-1} + \overline{\mathbf{B}}_k x_k. \tag{1}$$

The hidden states $(h_1, \ldots, h_L)$ form the *state sequence*.

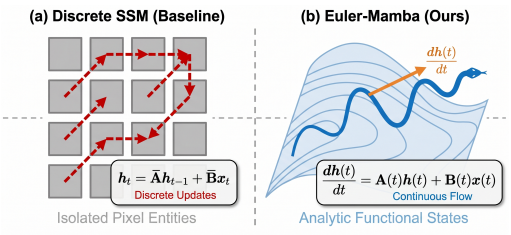

Figure 2: Comparison of state propagation orders. (a) Spatial-token ordering. (b) Frequency-ordered state propagation in Euler-Mamba.

At each step, the recurrent state combines information retained from earlier frequency coefficients with the current modulated input. Input dependence allows the selective SSM to vary its state update and readout with the coefficient being processed. Thus, the ordering specifies which frequency components precede the current input, while the selective parameters control how their information is propagated. These two aspects jointly define ESS: the radial-angular permutation organizes the sequence, and the selective SSM models dependencies within that sequence. The continuous-state formulation is

$$\frac{dh(t)}{dt} = \mathbf{A}(t)h(t) + \mathbf{B}(t)x(t), \tag{2}$$

where t is the continuous evolution variable. Appendix A.1 gives the first-order Euler approximation and the ESS algorithm.

Complex Polar Frequency Representation. We use the magnitude-phase parameterization of Fourier coefficients:

$$H(\omega) = \mathcal{R}(\omega)e^{i\Phi(\omega)} = \mathcal{R}(\omega)(\cos\Phi(\omega) + i\sin\Phi(\omega)). \tag{3}$$

Here ω denotes a frequency coordinate, $\mathcal{R}$ is the magnitude, and Φ is the phase. We call this representation the *complex polar frequency space*. It is a parameterization of complex Fourier coefficients, rather than a differential-geometric manifold.

Frequency Ordering and ESS. The PAN and LR-MS features are first transformed by FFT into centered frequency grids. EPR and IMM form the modulated representation

$$H_{\text{mod}} = \mathcal{R}_{\text{out}} \odot e^{i(\Phi_M + \Delta\Phi_P)}. \tag{4}$$

For each coefficient (u, v), define $r(u, v) = \sqrt{(u-u_0)^2 + (v-v_0)^2}$ and $\theta(u, v) = \text{atan2}(v - v_0, u - u_0) \bmod 2\pi$, where (u_0, v_0) is the frequency center. The permutation π sorts coefficients first by increasing radius and then by angle. With the batch dimension omitted, $H_{\text{mod}} \in \mathbb{C}^{d\times H\times W}$ has d feature channels, and its flattened ordered inputs are $x_k = H_{\text{mod}}[:, \pi(k)] \in \mathbb{C}^d$. The selective SSM produces y_k, which is restored to its original frequency position through $H_{\text{out}}[:, \pi(k)] = y_k$. Inverse shifting and IFFT reconstruct the ESS feature. Algorithm A.1 summarizes this pipeline. For fixed channel and state sizes, the scan is linear in L; the full pipeline also includes FFT/IFFT and the construction of the frequency ordering.

The two angular quantities used here have different meanings. The coordinate angle $\theta(u, v)$ describes the location of a coefficient relative to the center of the frequency grid and determines its position within the ordering. The Fourier phase $\Phi_M(u, v)$ describes the complex value at that location and is the quantity adjusted by EPR. Increasing radius organizes the scan from central, lower-frequency components toward peripheral, higher-frequency components; the angular key orders locations with equal radius. This provides a frequency-based context for state propagation while retaining the original coefficient locations for reconstruction.

The permutation is determined by the frequency-grid coordinates and is shared across feature channels on that grid. It rearranges the coefficients without resampling them or changing their values. After selective processing, writing y_k back to $\pi(k)$ reverses this rearrangement, so that each output again occupies the frequency location associated with its input. The subsequent inverse shift restores the FFT indexing convention before IFFT. These operations make the relationship between the two-dimensional frequency grid and the one-dimensional state sequence explicit.

3.3 Asymmetric Frequency Decoupling Module

The AFD module combines EPR and IMM to assign different operations to the two modalities. PAN provides high-resolution structural cues, while LR-MS supplies multispectral radiometric information. This sensing asymmetry motivates PAN-guided phase modulation and LR-MS magnitude recalibration.

Eulerian Feature Projection. For the LR-MS stream, the Fourier features F_M are expressed by magnitude $\mathcal{R}_M = |F_M|$ and phase Φ_M. The phase is obtained as:

$$\Phi_M = \text{arctan2}(\text{Im}(F_M), \text{Re}(F_M)), \tag{5}$$

where $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ denote the Cartesian components of the LR-MS feature in the frequency domain. This projection ensures that Φ_M preserves the original spectral topology as the structural anchor. Simultaneously, the PAN stream is disentangled into the high-frequency spatial magnitude $\mathcal{R}_P$ and a *predictive spatial phase shift* $\Delta\Phi_P$. Unlike the intrinsic Φ_M, the phase delta $\Delta\Phi_P$ is explicitly learned to capture the structural residuals required to align LR-MS features with the PAN geometry. We define this extraction process as:

$$\Delta\Phi_P = \mathcal{G}_{ext}(\mathcal{R}_P) = \tanh(\mathbf{W}_2 * \sigma(\mathbf{W}_1 * \mathcal{R}_P)), \tag{6}$$

where $\mathcal{G}_{ext}$ represents a lightweight convolutional extractor, and the $\tanh$ activation constrains the predicted phase shift to $[-1, 1]$ radians. This asymmetric design ensures that $\Delta\Phi_P$ serves as a purely structural compensator derived from the PAN resolution foundation $\mathcal{R}_P$.

Explicit Phase Rotation (EPR) Block. The EPR block uses the learned PAN-guided phase offset $\Delta\Phi_P$ to modulate LR-MS Fourier phase. It implements the rotation operator:

$$H_{fused}(\omega) = \mathcal{R}_M(\omega) \cdot e^{i(\Phi_M(\omega) + \Delta\Phi_P(\omega))}. \tag{7}$$

Since $|e^{i\Delta\Phi_P(\omega)}| = 1$, the EPR output satisfies $|H_{fused}(\omega)| = \mathcal{R}_M(\omega)$ pointwise. This preserves LR-MS magnitude during phase adjustment; IMM separately recalibrates the magnitude.

Equivalently, EPR multiplies the original complex coefficient by $e^{i\Delta\Phi_P}$, a factor with unit modulus. The learned PAN guidance therefore changes the coefficient angle while retaining its distance from the origin in the complex plane. This gives the structural-guidance branch a specific, interpretable action on the LR-MS representation. The phase offset is predicted from PAN features, while the original LR-MS phase remains the reference to which that offset is added.

Implicit Magnitude Mapping (IMM) Block. The IMM block is designed to calibrate the spectral intensity through a channel-wise refinement of the magnitude $\mathcal{R}_M$. We employ a spectrally aware attention mechanism to regulate the spectral magnitude response across wavelength bands, ensuring global color consistency:

$$\mathcal{R}_{out} = \sigma(\mathbf{W}_2 \cdot \delta(\mathbf{W}_1 \cdot \text{Gap}(\mathcal{R}_M))) \otimes \mathcal{R}_M, \tag{8}$$

where $\text{Gap}(\cdot)$ denotes global average pooling, σ is the sigmoid function, δ is the rectified linear unit (ReLU) activation, and $\otimes$ denotes channel-wise broadcast multiplication. The resulting channel weights recalibrate the LR-MS magnitude. The refined magnitude and modulated phase form H_{mod} before frequency ordering and selective state propagation.

Global average pooling summarizes the magnitude response over the frequency grid for each feature channel. The mapping through $\mathbf{W}_1$, δ, and $\mathbf{W}_2$, followed by the sigmoid, converts this summary into channel-dependent weights. Broadcasting each weight over its channel recalibrates the corresponding magnitude map while preserving its within-channel relative frequency profile. Together, the EPR phase offset and IMM magnitude weight specify complementary adjustments to each complex coefficient. This separation makes the modality roles explicit before ESS integrates information across frequency locations.

Continuous-State Notation. For this continuous-state representation, we assume a constant transition matrix and a zero initial state, with the input projection absorbed into the modulated input. The modulated magnitude and phase can also be expressed in the continuous-state notation:

$$h(t) = \int_0^t e^{\mathbf{A}(t-\tau)} \cdot \left[\mathcal{R}_{out}(\tau) \cdot e^{i(\Phi_M(\tau)+\Delta\Phi_P(\tau))}\right] d\tau. \tag{9}$$

Here t is the continuous evolution variable and τ is the integration variable. The discrete processing order used to form the ESS state sequence is specified in Algorithm A.1.

3.4 Loss Function

To optimize Euler-Mamba for high-fidelity reconstruction, we employ a hybrid loss function $\mathcal{L}_{total}$ that balances spectral consistency and geometric integrity:

$$\mathcal{L}_{total} = \|I_{fused} - I_{GT}\|_1 + \lambda\|\nabla I_{fused} - \nabla I_{GT}\|_1, \tag{10}$$

where λ is a balancing hyper-parameter, set to 0.1 in our experiments.

The first term, *Spectral Consistency Loss*, utilizes the ℓ_1-norm to minimize radiometric discrepancy. By penalizing pixel-wise intensity differences, it ensures that the IMM block accurately reconstructs spectral signatures without radiometric bias. The second term, *Geometric Topology Loss*, focuses on high-frequency alignment. By constraining spatial gradients ∇, it encourages the EPR block to prioritize structural boundaries and fine textures consistent with the PAN modality. This joint optimization enables a superior balance between physical interpretability and reconstruction fidelity.

4 Experiments

Datasets and Evaluation Metrics. Following the Wald protocol [40], we evaluate Euler-Mamba on the benchmark datasets curated by LGTEUN [26], which comprise imagery from the GaoFen-2 (GF2, 4-band), WorldView-2 (WV2, 4-band), and WorldView-3 (WV3, 8-band) satellites. For training and testing, samples are downsampled by a factor of 4 to generate 32×32 LR-MS and 128×128 PAN/GT patch pairs. Performance is quantified using five reference-based metrics: PSNR [18], SSIM [42], Q_{2^n} [13], SAM [50] (presented in radians), and ERGAS [3]. To assess the generalization, we

Table 1: Quantitative results on GF2 and WV2 datasets. Best and second best values are highlighted in **bold** and blue underlining, respectively.

	Methods	Params (K)	OPs (G)	PSNR ↑	SSIM ↑	Q_4 ↑	SAM ↓	ERGAS ↓	D_λ ↓	D_s ↓	QNR ↑
GaoFen-2 dataset	SFIM [28]	—	—	34.772	0.857	0.458	0.066	4.207	0.069	0.162	0.725
	BICUBIC [20]	—	—	35.308	0.851	0.363	0.060	4.038	0.067	0.214	0.734
	Wavelet [24]	—	—	33.921	0.820	0.403	0.070	4.645	0.131	0.181	0.704
	IHS [11]	—	—	35.232	0.884	0.522	0.066	3.991	0.078	**0.090**	0.721
	SRPPNN [6]	1711.4	21.1	45.562	0.982	0.893	0.026	1.233	0.066	0.205	0.741
	GPPNN [47]	119.8	11.2	43.598	0.976	0.866	0.033	1.522	0.072	0.173	0.800
	MutInf [55]	185.5	10.0	44.830	0.980	0.884	0.028	1.339	0.076	0.176	0.761
	HSIT [4]	75857.0	51.4	44.239	0.972	0.864	0.031	1.474	0.073	0.164	0.776
	PanFormer [54]	1530.3	12.0	45.130	0.981	0.886	0.026	1.293	0.065	0.200	0.748
	MSDDN [17]	2850.2	8.5	45.789	0.983	0.895	0.027	1.192	0.069	0.158	0.784
	MDCUN [48]	983.0	15.0	45.479	0.982	0.891	0.026	1.235	0.067	0.233	0.715
	PAPS [22]	1141.6	20.2	44.564	0.980	0.880	0.027	1.362	0.073	0.106	0.745
	LGTEUN [26]	402.2	4.3	45.954	0.983	0.898	0.025	1.171	0.075	0.147	0.790
	WINet [52]	2173.6	7.8	45.954	0.983	0.899	0.026	1.168	0.068	0.217	0.728
	PINO [29]	2522.8	13.6	45.980	0.983	0.895	0.027	1.160	0.070	0.135	0.795
	FusionMamba [45]	730.0	31.0	39.351	0.944	0.764	0.058	2.382	0.094	0.240	0.687
	Pan-Mamba [16]	578.4	4.3	45.250	0.981	0.887	0.026	1.269	0.065	0.208	0.740
	LE-Mamba [8]	741.2	4.2	45.643	0.984	**0.901**	0.026	1.203	0.068	0.215	0.730
	Ours	148.3	11.9	**46.188**	**0.985**	**0.901**	**0.024**	**1.137**	**0.063**	0.149	**0.803**
WorldView-2 dataset	SFIM [28]	—	—	32.633	0.873	0.516	0.060	3.192	0.074	0.090	0.844
	BICUBIC [20]	—	—	32.296	0.824	0.328	0.055	3.541	0.063	0.141	0.806
	Wavelet [24]	—	—	32.199	0.850	0.458	0.064	3.380	0.097	0.102	0.813
	IHS [11]	—	—	32.825	0.878	0.531	0.064	3.059	0.087	0.119	0.805
	SRPPNN [6]	1711.4	21.1	42.420	0.976	0.836	0.024	0.957	0.064	0.086	0.857
	GPPNN [47]	119.8	11.2	40.509	0.970	0.801	0.028	1.206	0.067	0.079	0.861
	MutInf [55]	185.5	10.0	41.953	0.976	0.826	0.023	1.015	0.065	0.079	0.864
	HSIT [4]	75857.0	51.4	41.603	0.974	0.823	0.024	1.043	0.064	0.086	0.856
	PanFormer [54]	1530.3	12.0	41.579	0.974	0.825	0.024	1.050	0.063	0.083	0.861
	MSDDN [17]	2850.2	8.5	42.197	0.977	0.834	0.022	0.977	0.064	0.084	0.866
	MDCUN [48]	983.0	15.0	42.322	0.977	0.836	0.022	0.962	0.066	0.083	0.857
	PAPS [22]	1141.6	20.2	42.252	0.977	0.833	0.022	0.980	0.062	0.078	0.865
	LGTEUN [26]	402.2	4.3	42.543	0.977	0.838	**0.021**	0.941	0.065	0.077	0.864
	WINet [52]	2173.6	7.8	42.408	0.976	0.837	0.022	0.957	0.064	0.076	0.857
	PINO [29]	2522.8	13.6	42.560	0.977	0.837	**0.021**	0.946	0.068	**0.072**	0.862
	FusionMamba [45]	730.0	31.0	33.916	0.901	0.618	0.056	2.248	0.119	0.092	0.801
	Pan-Mamba [16]	578.4	4.3	42.393	0.978	0.836	0.022	0.956	0.065	0.076	0.865
	LE-Mamba [8]	741.2	4.2	41.914	0.977	0.832	0.024	0.995	0.065	0.081	0.860
	Ours	148.3	11.9	**42.629**	**0.979**	**0.841**	**0.021**	**0.932**	**0.061**	0.075	**0.869**

further conduct full-resolution testing using non-reference metrics: D_λ, D_s, and QNR [2]. Due to page limitations, implementation details and further experimental results (e.g., cross-sensor and cross-scale experiments) are provided in the Supplementary Material.

Benchmark. To evaluate the efficiency of our method, we compare it against eighteen representative pansharpening models on benchmark datasets including four classical methods SFIM [28], BICUBIC [20], Wavelet [24], and IHS [11] as well as fourteen deep learning frameworks. These advanced models span from CNN-based architectures such as SRPPNN [6], GPPNN [47], MutInf [55], MSDDN [17], MDCUN [48], and PAPS [22] to Transformer-based paradigms such as HSIT [4], PanFormer [54], LGTEUN [26], and WINet [52]. Furthermore, we evaluate our model against the most recent developments including the operator-based PINO [29] and Mamba-based variants such as FusionMamba [45], Pan-Mamba [16], and LE-Mamba [8]. For a fair comparison, all learning-based competitors are retrained and evaluated on our benchmark datasets using the configurations specified in their original papers.

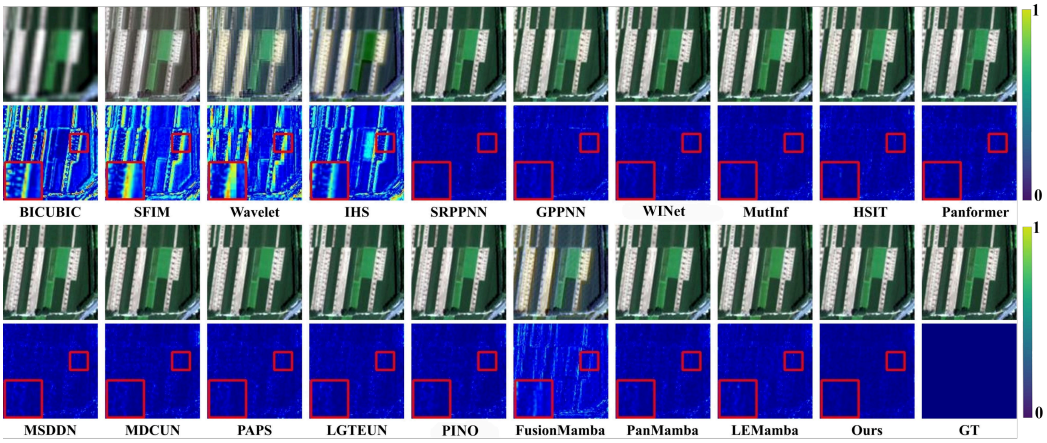

Figure 3: Visual comparison and absolute errors of our method versus other representative pansharpening methods on the reduced-resolution GF2 dataset.

4.1 Comparison With SOTA Methods

Reduced Resolution. Euler-Mamba achieves the highest PSNR and SSIM on both GF2 and WV2 datasets (Tab. 1). Superior metrics like Q_4 and SAM confirm that our asymmetric phase-magnitude decoupling effectively reconstructs structural topologies while suppressing spectral distortion. Visually (Fig. 3), Euler-Mamba yields the cleanest error maps with minimal artifacts compared to WINet or PanFormer. Zoomed-in regions highlight its ability to preserve fine-grained building boundaries and eliminate edge aliasing, demonstrating a robust balance between spatial sharpening and spectral consistency.

Full Resolution. Quantitative evaluations on full-resolution GF2 and WV2 scenes show that Euler-Mamba achieves the highest QNR scores (0.803 and 0.869) in Tab. 1. Euler-Mamba achieves the lowest spectral distortion D_λ on both datasets, while its highest QNR scores indicate a favorable overall balance between spectral and spatial fidelity. Visually (Fig. 4), our model produces the most natural fusion results. While IHS and SFIM introduce spectral distortion or blurriness, and PanFormer or FusionMamba show residual artifacts on roof textures, Euler-Mamba precisely recovers high-frequency boundaries and fine-grained structures while suppressing spectral bleeding, confirming the robustness of ESS for practical large-scale applications.

Performance-Efficiency Comparison. As detailed in Tab. 1, Euler-Mamba achieves an optimal balance between fusion quality and model compactness. With only 148.3 K parameters, it is the most lightweight architecture among high-performing competitors. While its OPs (11.9 G) are not the absolute lowest, this trade-off reflects the ESS mechanism's priority on modeling long-range dependencies over local convolutions. Despite this moderate cost, Euler-Mamba delivers state-of-the-art performance (PSNR of 46.188 dB on GF2 and 42.629 dB on WV2). This combination of minimal parameters and superior accuracy validates that asymmetric modulation and frequency-ordered state propagation provide an efficient yet powerful framework for pansharpening.

4.2 Ablation Study and Discussion

We conduct extensive ablation experiments on the GF2 dataset to evaluate the individual contributions of the modulation blocks (i.e., EPR block and IMM block), the complex-valued state representation, and the ESS mechanism. To evaluate the proposed framework, we define several variants: **V1** serves as the real-valued SSM baseline; **V2** extends **V1** with complex-valued state representation; **V3** and **V4** are the full model excluding the EPR block and IMM block, respectively; **V5** replaces the ESS with standard scanning; and **V6** is our complete Euler-Mamba.

Effectiveness of Asymmetric Modulation Components. Ablation studies on the EPR (**V3**) and IMM (**V4**) blocks in Tab. 2 demonstrate their essential roles in the full model. Removing the EPR

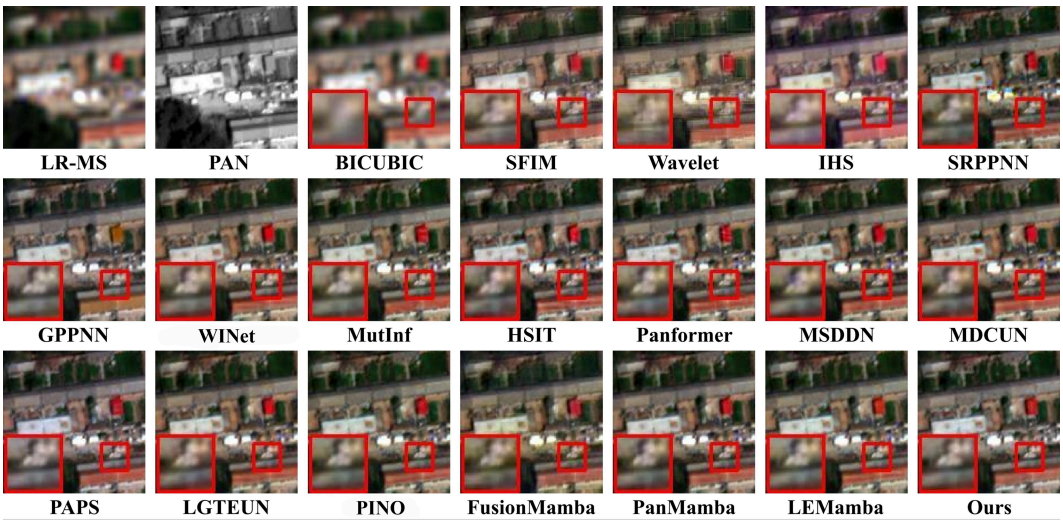

Figure 4: Visual comparison of our method versus other representative pansharpening methods on the full-resolution GF2 dataset.

block (**V3**) leads to a noticeable performance drop, where D_s increases from 0.149 to 0.176 and SSIM decreases from 0.985 to 0.980, confirming that phase information from EPR is critical for preserving structural sharpness and fine-grained spatial details. Conversely, discarding the IMM block (**V4**) primarily degrades spectral fidelity, with SAM rising from 0.024 to 0.027 and PSNR dropping by 0.505 dB, highlighting the necessity of the IMM block for magnitude calibration to maintain radiometric intensity and color consistency. Their integration enables the full model to suppress both spatial artifacts and spectral distortion, validating the effectiveness of our asymmetric frequency decoupling strategy.

Necessity of Complex-valued State Representation. To justify operating in the complex polar frequency space, we compare our framework against a real-valued baseline with an equivalent parameter budget. As shown in Tab. 2, transitioning from real-valued SSM (**V1**) to complex-valued representation (**V2**) yields a substantial performance leap, with PSNR increasing from 42.562 dB to 44.894 dB. This improvement demonstrates that the complex-valued state provides a more robust hypothesis space for representing heterogeneous features of LR-MS and PAN images, avoiding the representational entanglement where spectral and structural information interfere with each other in the real-valued counterpart.

Evaluation of Eulerian Selective Scan Mechanism. To verify the effectiveness of the proposed ESS, we compare the full model (**V6**) with a variant (**V5**) that replaces ESS with standard multi-directional scanning while maintaining the asymmetric modulation components. As shown in Tab. 2, although **V5** achieves competitive results, our ESS-equipped model (**V6**) consistently outperforms it across all spatial and spectral metrics. Specifically, QNR improves from 0.801 to 0.803, and ERGAS decreases from 1.156 to 1.137. These results indicate that standard scanning patterns cannot fully exploit latent correlations between complex-valued states and modulated frequency features, whereas ESS provides a frequency-ordered sequence of the modulated coefficients for state propagation.

5 Conclusion and Future Work

This paper introduces **Euler-Mamba**, a physics-inspired complex-valued state space framework for remote sensing pansharpening. By leveraging frequency-ordered state sequences, we achieve an effective decoupling of structural phase and spectral magnitude within a complex polar frequency space. The integration of the EPR and IMM blocks with ESS combines interpretable frequency-domain modulation with efficient long-range dependency modeling. Experimental results across multiple satellite datasets demonstrate that Euler-Mamba achieves state-of-the-art fusion quality with

Table 2: Ablation study of the proposed components on the GF2 dataset. The best values are highlighted in **bold**.

Method	PSNR ↑	SSIM ↑	Q_4 ↑	SAM ↓	ERGAS ↓	D_λ ↓	D_s ↓	QNR ↑
V1	42.562	0.965	0.863	0.038	1.482	0.082	0.205	0.732
V2	44.894	0.978	0.887	0.031	1.305	0.074	0.188	0.764
V3	45.210	0.980	0.891	0.029	1.254	0.071	0.176	0.779
V4	45.683	0.982	0.895	0.027	1.198	0.068	0.165	0.791
V5	45.921	0.984	0.898	0.025	1.156	0.066	0.152	0.801
V6 (Ours)	**46.188**	**0.985**	**0.901**	**0.024**	**1.137**	**0.063**	**0.149**	**0.803**

superior parameter efficiency, offering an optimal balance between performance and computational overhead.

In the future, we plan to extend the frequency-ordered state propagation from the spatial-spectral domain to the temporal dimension. By incorporating temporal consistency into the complex-valued state evolution, we aim to leverage Mamba's sequential modeling strengths for multi-temporal pansharpening and change detection. This integration will enhance the model's capacity to capture dynamic land-cover transitions while preserving high-fidelity physical signatures.

Acknowledgements and Disclosures

This work was supported by the Natural Science Foundation of Ningbo City (Grant No. 2025J004) and the Deep Earth Probe and Mineral Resources Exploration - National Science and Technology Major Project (2025ZD1007008). The authors declare no competing interests.

References

[1] B. Aiazzi, S. Baronti, and M. Selva. Improving component substitution pansharpening through multivariate regression of ms + pan data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3230–3239, 2007.

[2] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva. Multispectral and panchromatic data fusion assessment without reference. *Photogrammetric Engineering & Remote Sensing*, 74(2):193–200, 2008.

[3] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce. Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data-fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3012–3021, 2007.

[4] W. G. C. Bandara and V. M. Patel. Hypertransformer: A textural and spectral feature fusion transformer for pansharpening. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1767–1777, 2022.

[5] Z.-Y. Bo, W. Lu, H. Chen, S.-B. Chen, and B. Luo. SARU: A shadow-aware and removal unified framework for remote sensing images with new benchmarks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 239:39–53, 2026.

[6] J. Cai and B. Huang. Super-resolution-guided progressive pansharpening based on a deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):5206–5220, 2020.

[7] Q. Cao, L.-J. Deng, W. Wang, J. Hou, and G. Vivone. Zero-shot semi-supervised learning for pansharpening. *Information Fusion*, 101:102001, 2024.

[8] Z. Cao, X. Wu, L.-J. Deng, and Y. Zhong. A novel state space model with local enhancement and state sharing for image fusion. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 1235–1244, 2024.

[9] J. Choi, K. Yu, and Y. Kim. A new adaptive component-substitution-based satellite image fusion by using partial replacement. *IEEE transactions on geoscience and remote sensing*, 49(1):295–309, 2010.

[10] L.-J. Deng, G. Vivone, C. Jin, and J. Chanussot. Detail injection-based deep convolutional neural networks for pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 59(8):6995–7010, 2020.

[11] O. S. Faragallah. Enhancing multispectral imagery spatial resolution using optimized adaptive pca and high-pass modulation. *International journal of remote sensing*, 39(20):6572–6586, 2018.

[12] Y. Gao, J. Li, Y. Wu, L. Zhang, D. Lei, L. Kong, and L. Yue. PyS^{2}CF-Mamba: A pyramid spatial–spectral competitive fusion mamba network for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 23:5507805, 2026.

[13] A. Garzelli and F. Nencini. Hypercomplex quality assessment of multi/hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 6(4):662–665, 2009.

[14] A. Gu and T. Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

[15] J. Guibas, M. Mardani, Z. Li, A. Tao, A. Anandkumar, and B. Catanzaro. Adaptive fourier neural operators: Efficient token mixers for transformers. *arXiv preprint arXiv:2111.13587*, 2021.

[16] X. He, K. Cao, J. Zhang, K. Yan, Y. Wang, R. Li, C. Xie, D. Hong, and M. Zhou. Pan-mamba: Effective pan-sharpening with state space model. *Information Fusion*, 115:102779, 2025.

[17] X. He, K. Yan, J. Zhang, R. Li, C. Xie, M. Zhou, and D. Hong. Multiscale dual-domain guidance network for pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–13, 2023.

[18] A. Hore and D. Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010.

[19] J. Hou, R. Ran, S. Chen, Z. Chen, X. Cong, J. Li, and L.-J. Deng. Nodiff: Neural operator diffusion for multispectral image fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 4753–4761, 2026.

[20] Z. Huang and L. Cao. Bicubic interpolation and extrapolation iteration method for high resolution digital holographic reconstruction. *Optics and Lasers in Engineering*, 130:106090, 2020.

[21] F. D. Javan, F. Samadzadegan, S. Mehravar, A. Toosi, R. Khatami, and A. Stein. A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery. *ISPRS journal of photogrammetry and remote sensing*, 171:101–117, 2021.

[22] Y. Jia, Q. Hu, R. Dian, J. Ma, and X. Guo. Paps: Progressive attention-based pan-sharpening. *IEEE/CAA Journal of Automatica Sinica*, 11(2):391–404, 2024.

[23] M. M. Khan, J. Chanussot, L. Condat, and A. Montanvert. Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique. *IEEE Geoscience and Remote Sensing Letters*, 5(1):98–102, 2008.

[24] R. L. King and J. Wang. A wavelet based algorithm for pan sharpening landsat 7 imagery. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217)*, volume 2, pages 849–851. IEEE, 2001.

[25] J. Li, K. Zheng, L. Gao, L. Ni, M. Huang, and J. Chanussot. Model-informed multistage unsupervised network for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–17, 2024.

[26] M. Li, Y. Liu, T. Xiao, Y. Huang, and G. Yang. Local-global transformer enhanced unfolding network for pan-sharpening. *arXiv preprint arXiv:2304.14612*, 2023.

[27] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

[28] J. Liu. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of remote sensing*, 21(18):3461–3472, 2000.

[29] X. Liu, J. Hou, C. Wu, X. Cong, Z. Chen, S. Deng, J. Li, L.-J. Deng, and B. Liu. Physics-informed neural operator for pansharpening. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[30] Y. Liu, R. Dian, and S. Li. Low-rank transformer for high-resolution hyperspectral computational imaging. *International Journal of Computer Vision*, 133(2):809–824, 2025.

[31] Y. Liu, R. Dian, Z. Zhao, and S. Li. Rank-aware routing decomposition for hyperspectral and multispectral image fusion. *Information Fusion*, 136:104498, 2026.

[32] W. Lu, H.-D. Li, C. Wang, S.-B. Chen, C. H. Q. Ding, J. Tang, and B. Luo. UnravelNet: A backbone for enhanced multi-scale and low-quality feature extraction in remote sensing object detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 231:431–442, 2026.

[33] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7):594, 2016.

[34] A. V. Oppenheim and J. S. Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.

[35] R. A. Schowengerdt. *Remote sensing: models and methods for image processing*. elsevier, 2006.

[36] H. Sun, L. Lv, P. Zhang, T. Tang, F. Tian, W. Sun, and H. Lu. Spatial-frequency enhanced mamba for multi-modal image fusion. *IEEE Transactions on Image Processing*, 34:7684–7696, 2025.

[37] J. Tan, J. Huang, N. Zheng, M. Zhou, K. Yan, D. Hong, and F. Zhao. Revisiting spatial-frequency information integration from a hierarchical perspective for panchromatic and multi-spectral image fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25922–25931, 2024.

[38] G. Vivone. Robust band-dependent spatial-detail approaches for panchromatic sharpening. *IEEE transactions on Geoscience and Remote Sensing*, 57(9):6421–6433, 2019.

[39] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot. Contrast and error-based fusion schemes for multispectral image pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 11(5):930–934, 2013.

[40] L. Wald, T. Ranchin, and M. Mangolini. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric engineering and remote sensing*, 63(6):691–699, 1997.

[41] W. Wang, L.-J. Deng, R. Ran, and G. Vivone. A general paradigm with detail-preserving conditional invertible network for image fusion. *International Journal of Computer Vision*, 132(4):1029–1054, 2024.

[42] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[43] M. Wei and X. Zhang. Super-resolution neural operator. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18247–18256, 2023.

[44] X. Wu, T.-Z. Huang, L.-J. Deng, and T.-J. Zhang. Dynamic cross feature fusion for remote sensing pansharpening. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14687–14696, 2021.

[45] X. Xie, Y. Cui, T. Tan, X. Zheng, and Z. Yu. Fusionmamba: Dynamic feature enhancement for multimodal image fusion with mamba. *Visual Intelligence*, 2(1):37, 2024.

[46] J. Xiong, Y. Zhang, X. Wang, and L. Zhang. Spatial-spectral similarity-guided fusion network for pansharpening. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 2089–2097, 2025.

[47] S. Xu, J. Zhang, Z. Zhao, K. Sun, J. Liu, and C. Zhang. Deep gradient projection networks for pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1366–1375, 2021.

[48] G. Yang, M. Zhou, K. Yan, A. Liu, X. Fu, and F. Wang. Memory-augmented deep conditional unfolding network for pan-sharpening. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1788–1797, 2022.

[49] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley. Pannet: A deep network architecture for pan-sharpening. In *Proceedings of the IEEE international conference on computer vision*, pages 5449–5457, 2017.

[50] R. H. Yuhas, A. F. Goetz, and J. W. Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In *JPL, summaries of the third annual JPL airborne geoscience workshop. volume 1: AVIRIS workshop*, 1992.

[51] B. Zhang, Y. Wu, B. Zhao, J. Chanussot, D. Hong, J. Yao, and L. Gao. Progress and challenges in intelligent remote sensing satellite systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:1814–1822, 2022.

[52] J. Zhang, X. He, K. Yan, K. Cao, R. Li, C. Xie, M. Zhou, and D. Hong. Pan-sharpening with wavelet-enhanced high-frequency information. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024.

[53] Y. Zhang, X. Lin, L. Zhang, and Z. Cai. Orthogonal spatial-aware multi-view anchor graph clustering for incomplete remote sensing data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20532–20541, 2026.

[54] H. Zhou, Q. Liu, and Y. Wang. Panformer: A transformer based model for pan-sharpening. In *2022 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2022.

[55] M. Zhou, K. Yan, J. Huang, Z. Yang, X. Fu, and F. Zhao. Mutual information-driven pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1798–1808, 2022.

A Appendix

A.1 ESS Formulation and Algorithm

This section specifies the continuous and discrete state notation, the ESS processing order, and the magnitude-preservation property of EPR.

Continuous Physics-Inspired Evolution. The pansharpening process is modeled as a continuous state evolution within a complex polar frequency space. Let $h(t) \in \mathbb{C}^N$ represent the latent physical state. The continuous-state notation uses the following linear differential equation:

$$\frac{dh(t)}{dt} = \mathbf{A}(t)h(t) + \mathbf{B}(t)x(t), \tag{11}$$

where $\mathbf{A}(t) \in \mathbb{C}^{N\times N}$ is the dynamic matrix defining internal dynamics, and $\mathbf{B}(t)$ maps the source input $x(t)$ (LR-MS or PAN features) into the latent frequency representation. To achieve frequency decoupling, we parameterize $\mathbf{A}$ using its polar representation, where eigenvalues $\lambda = \sigma + j\omega$ represent energy dissipation (σ) and structural rotation (ω).

Derivation of Explicit Phase Rotation (EPR). The EPR block is derived as a unitary transformation to align spatial structures without distorting energy. In the complex plane, this is expressed as:

$$\mathcal{R}(\theta) = \exp(j\theta) = \cos(\theta) + j\sin(\theta). \tag{12}$$

Since $|\exp(j\theta)| = 1$, the EPR phase rotation preserves pointwise input magnitude. This identity describes the EPR operation before IMM recalibration and selective state propagation.

Discretization and ESS Implementation. Let $k = 1, \ldots, HW$ index the frequency-ordered inputs. The discrete state recurrence is:

$$h_k = \overline{\mathbf{A}}_k h_{k-1} + \overline{\mathbf{B}}_k x_k. \tag{13}$$

Under the first-order Euler approximation stated in the main formulation, the parameters are:

$$\overline{\mathbf{A}}_k = \mathbf{I} + \Delta t_k \mathbf{A} \approx \exp(\Delta t_k \mathbf{A}), \quad \overline{\mathbf{B}}_k = \Delta t_k \mathbf{B}_k. \tag{14}$$

The parameters $\mathbf{B}_k$, $\mathbf{C}_k$, and Δt_k depend on the current input. $\mathbf{C}_k$ defines the hidden-state readout; the full input/output projections are included in the selective SSM operation below.

Transition Eigenvalues and Phase Rotation. For a fixed transition matrix, eigenvalues inside the unit circle characterize asymptotic stability of the homogeneous recurrence. If $\lambda_i(\mathbf{A}) = \sigma_i + j\omega_i$ with $\sigma_i < 0$ and $\Delta t > 0$, the exponential parameterization gives

$$|\exp(\Delta t(\sigma_i + j\omega_i))| = e^{\Delta t \sigma_i} < 1. \tag{15}$$

For a fixed Euler transition, the corresponding condition is $|1 + \Delta t \lambda_i(\mathbf{A})| < 1$. These statements concern the specified transition matrix. The EPR identity $|e^{j\theta} z| = |z|$ establishes pointwise magnitude preservation during phase rotation.

Algorithm A.1 Eulerian Selective Scan (ESS)

Input: PAN feature X_P, LR-MS feature X_M.
Output: ESS feature X_{ESS}.

$F_P \leftarrow \text{fftshift}(\text{FFT2}(X_P))$;
$F_M \leftarrow \text{fftshift}(\text{FFT2}(X_M))$.
$\mathcal{R}_P \leftarrow |F_P|$; $\mathcal{R}_M \leftarrow |F_M|$; $\Phi_M \leftarrow \arg(F_M)$.
$\Delta\Phi_P \leftarrow \mathcal{G}_{ext}(\mathcal{R}_P)$ using the EPR phase-offset extractor.
Compute $\mathcal{R}_{out}$ using the IMM magnitude-recalibration equation.
$H_{mod} \leftarrow \mathcal{R}_{out} \odot \exp(i(\Phi_M + \Delta\Phi_P))$.
$r(u,v) \leftarrow \sqrt{(u-u_0)^2 + (v-v_0)^2}$.
$\theta(u,v) \leftarrow \text{atan2}(v-v_0, u-u_0) \bmod 2\pi$.
$\pi \leftarrow \text{argsort}((u,v),\ \text{key} = (r(u,v), \theta(u,v)))$.
$x_k \leftarrow H_{mod}[:, \pi(k)],\ k = 1, \ldots, HW$.
$(y_1, \ldots, y_{HW}) \leftarrow \text{SelectiveSSM}(x_1, \ldots, x_{HW})$.
$H_{out}[:, \pi(k)] \leftarrow y_k,\ k = 1, \ldots, HW$.
$X_{ESS} \leftarrow \text{IFFT2}(\text{ifftshift}(H_{out}))$.

A.2 Implementation Details

The Euler-Mamba framework is optimized using the AdamW optimizer with a weight decay of 1×10^{-8} and a batch size of 16. The training process spans a total of 1000 epochs, governed by a cosine annealing learning-rate scheduler that starts from an initial value of 5×10^{-4}. To ensure the physical consistency of the frequency-ordered state propagation, the integration step Δt is treated as a trainable parameter, initialized and updated dynamically to adapt to the input-dependent state updates. Despite its sophisticated mathematical foundation, the framework maintains a highly efficient and lightweight profile with 148.3 K parameters for four-band inputs and 148.5 K parameters for eight-band inputs.

All experiments, including the training and evaluation of the Eulerian Selective Scan (ESS) mechanism and the Asymmetric Frequency Decoupling (AFD) module, were executed on a high-performance workstation equipped with an Intel Core i9-12900K CPU and an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM. The software environment was built upon Ubuntu 20.04 LTS, utilizing Python 3.8, PyTorch 1.10.1, and CUDA 11.3 for accelerated computation. Benefiting from the streamlined design of the proposed architecture, the training process is notably efficient, typically reaching convergence within 4 to 6 hours on standard remote sensing datasets such as GF2 or WV2.

A.3 More Experimental Results

Quantitative results on WV3 dataset. The quantitative evaluation on the WV3 dataset presented in Tab. 3 demonstrates the superior performance of the proposed Euler-Mamba framework across nearly all metrics. Traditional methods such as SFIM and IHS exhibit significant performance gaps compared to deep learning approaches particularly in terms of PSNR and SSIM. Among the neural network baselines, PINO and Pan-Mamba show strong competitive results with PINO achieving the best SSIM of 0.960 and D_s of 0.029 while Pan-Mamba maintains excellent spectral fidelity with a SAM of 0.060. Other Mamba-based variants like LE-Mamba and FusionMamba demonstrate the potential of state-space models but remain limited by higher spectral distortion or lower structural consistency compared to our approach. The proposed method achieves the top rank in PSNR at 32.589 and Q_8 at 0.953 while simultaneously obtaining the optimal SAM of 0.059 and QNR of 0.950. Notably, our framework achieves these state-of-the-art results with a highly efficient design utilizing only 148.5 K parameters which is significantly lower than the 2527.8 K required by PINO and the 580.8 K used by Pan-Mamba. This comparison highlights that Euler-Mamba provides a more balanced and efficient solution for high-fidelity remote sensing image fusion compared to existing state-of-the-art methods.

More visual results. The qualitative assessment across the WV2 and WV3 datasets at both reduced and full resolutions demonstrates the clear superiority of the Euler-Mamba framework in preserving both spatial and spectral integrity. In the reduced-resolution evaluations shown in Fig. 5 and Fig. 7, the absolute error maps reveal that our method achieves the lowest residual values compared to the GT as evidenced by the predominantly dark blue visualization which indicates a near-perfect

Table 3: Quantitative results on 8-band WV3 dataset. Best and second best values are highlighted in **bold** and blue underlining, respectively.

Methods	Params (K)	OPs (G)	PSNR ↑	SSIM ↑	Q_8 ↑	SAM ↓	ERGAS ↓	D_λ ↓	D_s ↓	QNR ↑
SFIM [28]	—	—	21.415	0.542	0.453	0.115	8.855	**0.009**	0.106	0.885
BICUBIC [20]	—	—	20.505	0.357	0.236	0.116	9.829	0.018	0.063	0.928
Wavelet [24]	—	—	21.446	0.566	0.527	0.150	9.155	0.055	0.133	0.819
IHS [11]	—	—	22.345	0.613	0.571	0.171	7.944	0.018	0.122	0.862
SRPPNN [6]	1722.6	21.5	31.538	0.947	0.941	0.067	2.850	0.040	0.045	0.918
GPPNN [47]	238.6	22.2	31.698	0.951	0.945	0.065	2.811	0.020	0.048	0.933
MutInf [55]	240.1	14.5	31.829	0.952	0.947	0.064	2.753	0.016	0.042	0.942
HSIT [4]	75913.0	51.7	31.135	0.951	0.947	0.071	2.965	0.033	0.045	0.924
PanFormer [54]	1532.8	24.3	31.134	0.944	0.938	0.067	2.962	0.021	0.044	0.936
MSDDN [17]	2855.1	8.6	32.055	0.954	0.948	0.063	2.683	0.016	0.034	0.940
MDCUN [48]	1350.3	18.5	31.993	0.953	0.948	0.063	2.703	0.041	0.035	0.926
PAPS [22]	1148.5	20.3	32.323	0.956	0.951	0.061	2.599	0.024	0.039	0.938
LGTEUN [26]	540.0	4.8	32.079	0.954	0.949	0.060	2.663	0.017	0.033	0.945
WINet [52]	2175.9	7.8	32.260	0.955	0.950	0.062	2.624	0.019	0.040	0.942
FusionMamba [45]	735.6	45.8	26.475	0.838	0.828	0.173	6.056	0.084	0.047	0.873
PINO [29]	2527.8	13.7	32.485	**0.960**	0.950	0.061	2.537	0.020	**0.029**	0.945
Pan-Mamba [16]	580.8	4.3	32.468	0.957	0.951	0.060	2.553	0.017	0.035	0.941
LE-Mamba [8]	743.5	4.2	32.187	0.954	0.949	0.060	2.648	0.017	0.051	0.932
Ours	148.5	12.8	**32.589**	0.958	**0.953**	**0.059**	**2.524**	0.014	0.033	**0.950**

reconstruction of urban and vegetated landscapes. While competitive models such as PINO and Pan-Mamba successfully recover sharp boundaries, they often exhibit higher error concentrations around fine-grained textures where the asymmetric modeling of Euler-Mamba proves more robust. The practical utility of the proposed approach is further validated through the full-resolution experiments illustrated in Fig. 6 and Fig. 8 where no reference images are available. Our framework effectively injects high-frequency geometric details from the PAN image into the LR-MS spectral bands without introducing the ringing artifacts or spectral distortion commonly observed in traditional methods like Wavelet or IHS. The zoomed-in regions highlight that Euler-Mamba maintains natural color fidelity and sharp structure textures, confirming that the frequency-ordered state propagation in the complex polar frequency space supports the proposed frequency-domain formulation for real-world satellite image fusion.

Visual Analysis of Ablation Variants. As illustrated in Fig. 9, the visual comparisons provide intuitive evidence for the contributions of the EPR and IMM blocks. In the absence of the EPR block (**V3**), the reconstructed images suffer from compromised structural integrity, characterized by blurred contours and noticeable double-edge artifacts in the building regions of Patch A. This highlights the EPR's role in using PAN-guided phase offsets for high-frequency detail injection. Conversely, the variant without the IMM block (**V4**) exhibits significant spectral distortion, manifested as localized color shifts and chromatic aberrations in the vegetation and soil textures of Patch B. By contrast, the full Euler-Mamba (**V6**) effectively mitigates these artifacts, achieving the highest fidelity in both sharp edges and natural color representation, thereby validating our asymmetric frequency decoupling strategy.

A.4 Further Generalization Analysis

A.4.1 Comparative Analysis of Cross-Scale Generalization

The quantitative comparison in Tab. 4 evaluates the scale-invariant properties of the proposed Euler-Mamba framework against the competitive LE-Mamba model on the WV3 dataset. While LE-Mamba leverages the selective scan mechanism to capture spatial dependencies, it exhibits visible performance fluctuations when the spatial resolution deviates from the $4\times$ training configuration. In contrast, the proposed framework maintains superior structural and spectral stability across both $2\times$ and $8\times$ testing scenarios, which can be attributed to the frequency-ordered state propagation defined in the complex

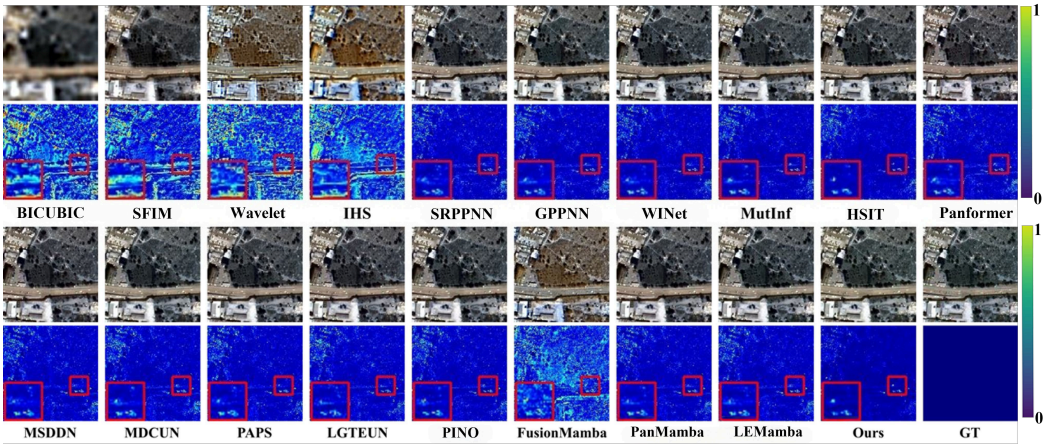

Figure 5: Visual comparison and absolute errors of our method versus other representative pansharpening methods on the reduced-resolution WV3 dataset.

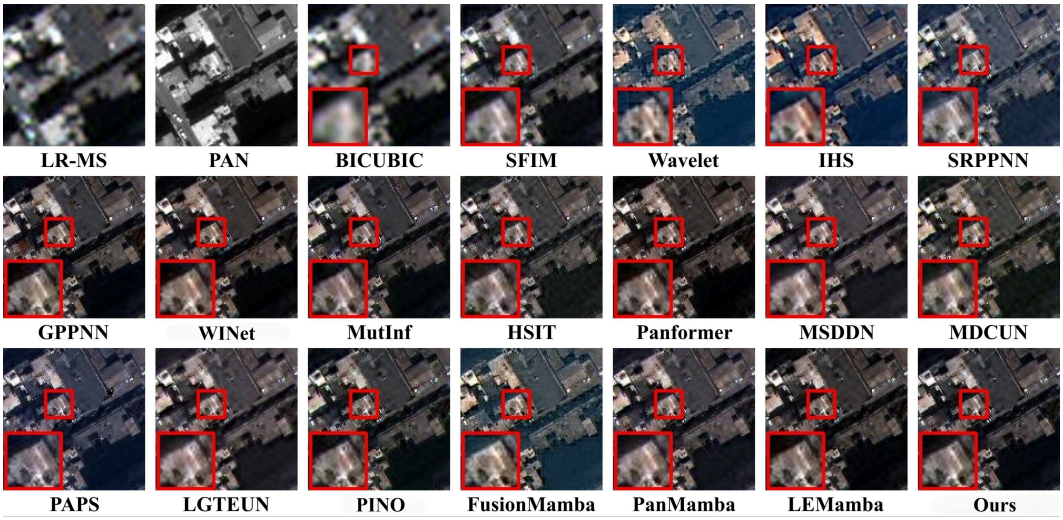

Figure 6: Visual comparison of our method versus other representative pansharpening methods on the full-resolution WV3 dataset.

polar frequency space. The results support cross-scale generalization under the evaluated sampling configurations. These findings confirm that our method provides a more consistent, physics-inspired solution for multi-resolution pansharpening tasks.

A.4.2 Comparative Analysis of Cross-Sensor Generalization

The quantitative results in Tab. 5 further illustrate the robust transferability of the Euler-Mamba framework compared to the LE-Mamba baseline under a zero-shot cross-sensor configuration. When models pre-trained on the GF2 dataset are directly deployed to evaluate WV3 imagery, our approach maintains significantly higher performance margins in both spectral fidelity and structural consistency. While LE-Mamba captures spatial dependencies through the selective scan mechanism, it remains more susceptible to the discrepancies in spectral response functions and imaging geometries between different satellite platforms. In contrast, the asymmetric frequency decoupling module within Euler-Mamba successfully disentangles universal geometric phase shifts from sensor-specific spectral magnitudes, allowing the model to adapt to new sensor distributions without any parameter updates. The superior SAM and QNR values achieved by our method confirm that modeling the pansharpening

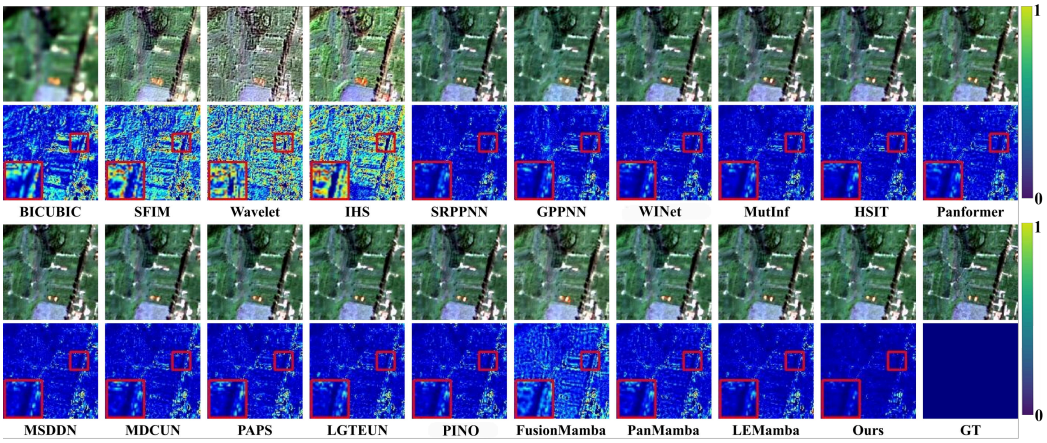

Figure 7: Visual comparison and absolute errors of our method versus other representative pansharpening methods on the reduced-resolution WV2 dataset.

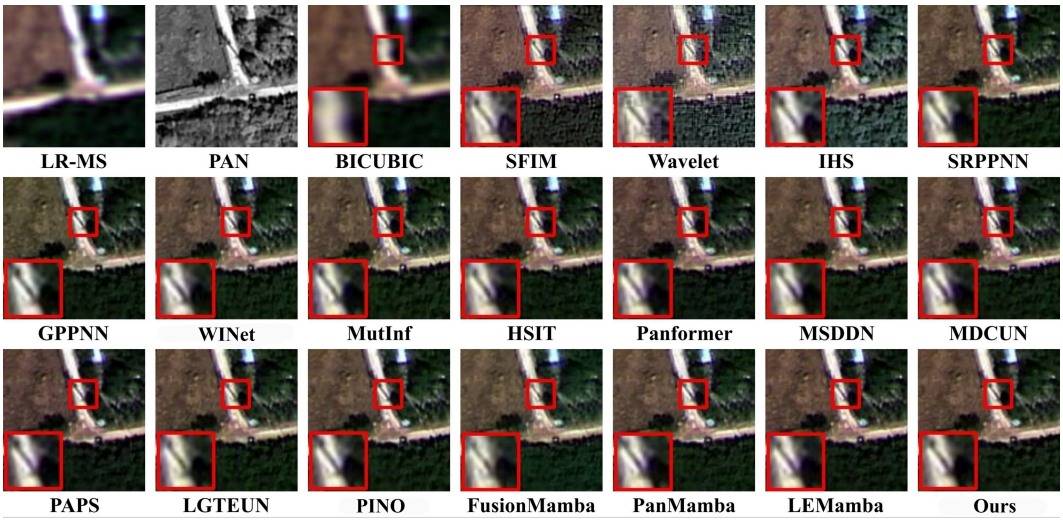

Figure 8: Visual comparison of our method versus other representative pansharpening methods on the full-resolution WV2 dataset.

Table 4: Comparative analysis of cross-scale generalization on WV3 dataset. Both models are trained at $4\times$ scale and directly evaluated at $2\times$ and $8\times$ scales to verify resolution-invariant properties.

Scale	Method	PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓	QNR ↑
$2\times$	LE-Mamba	33.154	0.962	0.054	2.189	0.948
	Ours	**35.124**	**0.978**	**0.042**	**1.845**	**0.968**
$8\times$	LE-Mamba	26.542	0.865	0.088	4.312	0.902
	Ours	**28.456**	**0.912**	**0.078**	**3.892**	**0.924**

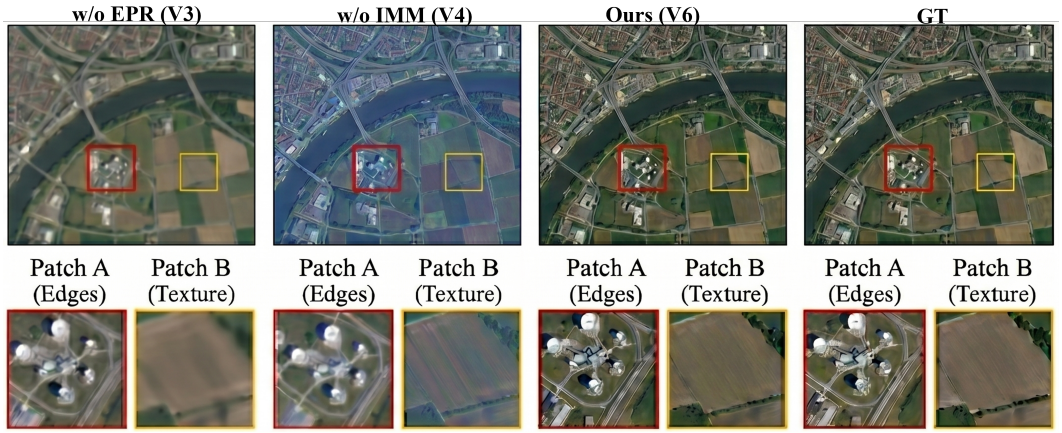

Figure 9: Visual comparison of ablation study variants on the GF2 dataset. Patch A (Edges) highlights that the absence of the EPR module (**V3**) leads to blurred contours and double-edge artifacts. Patch B (Texture) demonstrates that excluding the IMM module (**V4**) results in significant spectral distortion and localized color shifts. The full Euler-Mamba (**V6**) achieves the best balance between structural sharpness and spectral fidelity.

process as frequency-ordered state propagation effectively captures the underlying physical principles of remote sensing rather than merely overfitting to source-domain statistics. These findings emphasize that Euler-Mamba provides a more generalized solution with frequency-domain interpretability for practical multi-sensor deployment.

Table 5: Quantitative comparison of cross-sensor generalization performance on GF-2 and WV-3 datasets.

Method	Training Set	Test Set	PSNR ↑	SSIM ↑	SAM ↓	QNR ↑
LE-Mamba	GF-2	WV-3	30.142	0.924	0.082	0.915
Ours	GF-2	WV-3	**31.874**	**0.948**	**0.065**	**0.938**
Reference*	WV-3	WV-3	32.589	0.958	0.059	0.950

*Note: Reference refers to our model trained and tested on the same sensor distribution.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction sections offer a comprehensive discussion of the manuscript's context, intuition, and ambitions, as well as its contributions.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.

- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [No]

   Justification: The paper evaluates the method on the reported datasets and sampling configurations, but does not explicitly discuss the limitations of the proposed approach.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: For each theoretical result, the paper provides the full set of assumptions and a complete (and correct) proof.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The pipeline of the methods and the details of experiments are presented with corresponding reproducible credentials.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: The code is not publicly available yet. We plan to release the implementation and reproduction instructions through `https://github.com/Florina2333`.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).

- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The pipeline of the methods and the details of experiments are presented with corresponding reproducible credentials.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: The paper reports point estimates of the evaluation metrics, without error bars, confidence intervals, or statistical significance tests.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: The details of experiments are presented with corresponding reproducible credentials.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The research conducted in the paper conforms with the NeurIPS Code of Ethics

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: The paper focuses on the methodology and experimental evaluation of satellite image fusion and does not include a dedicated discussion of its potential positive and negative societal impacts.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The proposed method is a task-specific pansharpening model for satellite image fusion. This work does not introduce assets with the high-risk release characteristics targeted by this question.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [No]

    Justification: The sources of the datasets and baseline methods are cited, but the specific licenses and terms of use of the existing assets are not explicitly documented in the paper.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: No new code, model checkpoints, or datasets have been released at this stage. We plan to release the implementation with accompanying documentation through `https://github.com/Florina2333`.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.