The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

Pansharpening for Thin-Cloud Contaminated Remote Sensing Images: A Unified Framework and Benchmark Dataset

Songcheng Du[1]*, Yang Zou[1]*, Jiaxin Li[2], Mingxuan Liu[1], Ying Li[1]†, Changjing Shang[3], Qiang Shen[3]

[1]Northwest Polytechnical University
[2]Chongqing University of Posts and Telecommunications
[3]Aberystwyth University
dusongcheng@mail.nwpu.edu.cn, archerv2@mail.nwpu.edu.cn, lijx@cqupt.edu.cn, 2023302820@mail.nwpu.edu.cn,
lybyp@nwpu.edu.cn, cns@aber.ac.uk, qqs@aber.ac.uk

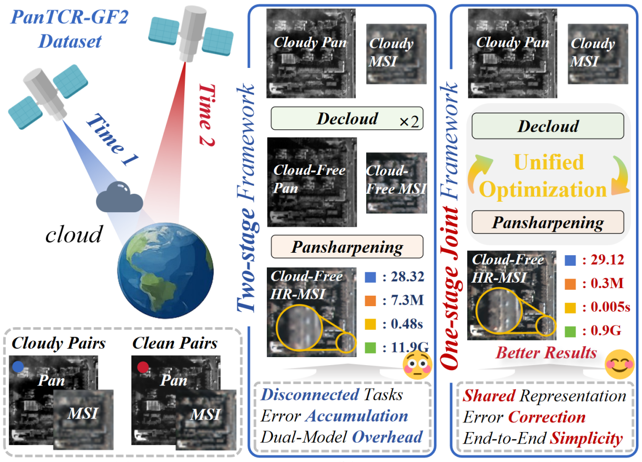
Abstract

Pansharpening under thin cloudy conditions is a practically significant yet rarely addressed task, challenged by simultaneous spatial resolution degradation and cloud-induced spectral distortions. Existing methods often address cloud removal and pansharpening sequentially, leading to cumulative errors and suboptimal performance due to the lack of joint degradation modeling. To address these challenges, we propose a Unified Pansharpening Model with Thin Cloud Removal (Pan-TCR), an end-to-end framework that integrates physical priors. Motivated by theoretical analysis in the frequency domain, we design a frequency-decoupled restoration (FDR) block that disentangles the restoration of multispectral image (MSI) features into amplitude and phase components, each guided by complementary degradation-robust prompts: the near-infrared (NIR) band amplitude for cloud-resilient restoration, and the panchromatic (PAN) phase for high-resolution structural enhancement. To ensure coherence between the two components, we further introduce an interactive inter-frequency consistency (IFC) module, enabling cross-modal refinement that enforces consistency and robustness across frequency cues. Furthermore, we introduce the first real-world thin-cloud contaminated pansharpening dataset (PanTCR-GF2), comprising paired clean and cloudy PAN-MSI images, to enable robust benchmarking under realistic conditions. Extensive experiments on real-world and synthetic datasets demonstrate the superiority and robustness of Pan-TCR, establishing a new benchmark for pansharpening under realistic atmospheric degradations.

Code — https://github.com/dusongcheng/PanTCR-GF2

Introduction

Pansharpening aims to generate high-resolution multispectral images (MSIs) by fusing the spectral information of low-resolution MSIs with the spatial details of panchromatic (PAN) images. This technique is crucial for a wide range of applications, including land use monitoring, precision agriculture, and environmental surveillance (Xi et al. 2025; Peng et al. 2025; Du et al. 2023; Li et al. 2022). However, real-world satellite images are frequently contaminated by cloud cover, which not only degrades the spatial textures but also distorts spectral reflectance in a non-uniform manner.



(a) (b) (c)

Figure 1: Collection of the first thin-cloud pansharpening dataset (a) and comparison of processing paradigms and performance between existing methods (b) and proposed approach (c). 🟦: PSNR, 🟧: Parameters, 🟨: Test Time, 🟩: FLOPs.

Despite the widespread occurrence of cloud, particularly thin-cloud, in optical satellite imagery, most existing pansharpening methods assume clear-sky conditions (Huang et al. 2025; Wang et al. 2025; Zhong et al. 2024; Du et al. 2024; Zou et al. 2026), making them unreliable under cloud-contaminated scenarios. A common workaround is to perform cloud removal and pansharpening sequentially, where cloud-free MSI and PAN are first reconstructed using two separate cloud removal networks, followed by conventional pansharpening (Li and Jing 2017; Meng et al. 2018; Wang et al. 2024b). As illustrated in Fig. 1(b), to implement such an approach, a state-of-the-art (SOTA) decloud algorithm (Zhao, Feng, and Li 2025) and a leading pansharpening method (Huang et al. 2025) may be employed. While intuitive, this two-stage design suffers from several drawbacks: 1) **Cascaded error accumulation.** Artifacts arising

---

*These authors contributed equally.
†Corresponding author.

from the cloud removal stage, including residual clouds and spectral inconsistencies, may be propagated and further amplified during pansharpening, degrading the final result. 2) **High computational overhead.** The cascade of two separate networks significantly raises inference time and memory consumption, making deployment impractical in real-time or resource-constrained environments.

Besides, following the cascade approach, thin-cloud removal and pansharpening are typically conducted by separate models trained under different settings and objectives, resulting in inconsistent intermediate representations and suboptimal final outputs. Although several recent works have proposed joint learning strategies for other multimodal fusion tasks (Ni and Zhang 2025; Feijoo et al. 2025; Su et al. 2025), few studies have explicitly addressed the unified modeling of thin-cloud and resolution degradations within a single pansharpening framework. This lack of research attention limits the robustness of existing methods in practical scenarios where thin-cloud contamination is the norm rather than the exception.

Therefore, designing an end-to-end framework that simultaneously addresses both resolution degradation and thin-cloud interference is essential. However, achieving this requires: (1) effectively leveraging complementary cues from heterogeneous inputs (e.g., PAN, near-infrared image (NIR)), and (2) disentangling the above two degradations in a principled manner. To this end, we propose **Pan-TCR**, the first unified pansharpening framework (see Fig. 1(c)) that explicitly models both degradations within a frequency-prompt guided formulation, offering a principled solution tailored to real-world satellite imagery.

Our key insight stems from the observation that the NIR band within MSI is inherently more robust to thin-cloud interference (Fig. 2 (a)), while the PAN image offers high-resolution structures yet is vulnerable to thin-cloud interference similar to visible bands. Motivated by their distinct frequency-domain characteristics (Fig. 2(b) and (c)), we treat the NIR amplitude and PAN phase as two degradation-robust frequency prompts, respectively guiding cloud-resilient amplitude restoration and structure-preserving phase enhancement.

To implement this design, we propose a **Frequency-Decoupled Restoration (FDR) Block** that performs component-wise restoration on amplitude and phase, enabling simultaneous suppression of cloud artifacts and recovery of structural details. To enhance consistency between these frequency components (highlighted in Fig. 2(b) and (c)), we embed an **Interactive inter-Frequency Consistency (IFC) Module** within the FDR block to enhance consistency between these components through interactive cross-modal refinement. Additionally, we introduce a **Modality-Adaptive Frequency Gating (MAFG) Module** to dynamically suppress modality-induced spectral outliers, enhancing cross-modal consistency between NIR and visible bands. To further refine spectral representations beyond the frequency domain, a **Spectral Enhancement (SE) Module** is incorporated to provide fine-grained spectral-wise correction complementary to frequency-based restoration.

Furthermore, to support the current development in this area, we construct the first real-world dataset for pansharpening under thin-cloud conditions, named **PanTCR-GF2** (Fig. 1(a)). It contains 15,603 high-quality, aligned pairs of cloudy/clean PAN and MSI images across five representative land cover categories, supporting comprehensive benchmarking under realistic atmospheric interference.

Our contributions made in this work are as follows:

- We propose **Pan-TCR**, the first unified end-to-end framework that jointly addresses thin-cloud removal and pansharpening within a frequency-prompt restoration formulation tailored for real-world remote sensing imagery.
- We introduce the **Frequency-Decoupled Restoration Block**, which leverages NIR amplitude and PAN phase as degradation-robust frequency prompts to perform targeted restoration on amplitude and phase components.
- We build **PanTCR-GF2**, the first real-world benchmark dataset for thin-cloud contaminated pansharpening, enabling systematic evaluation of fusion methods under atmospheric degradation.
- Our method achieves state-of-the-art performance across multiple benchmarks while being more efficient in terms of model size, FLOPs, and inference time.

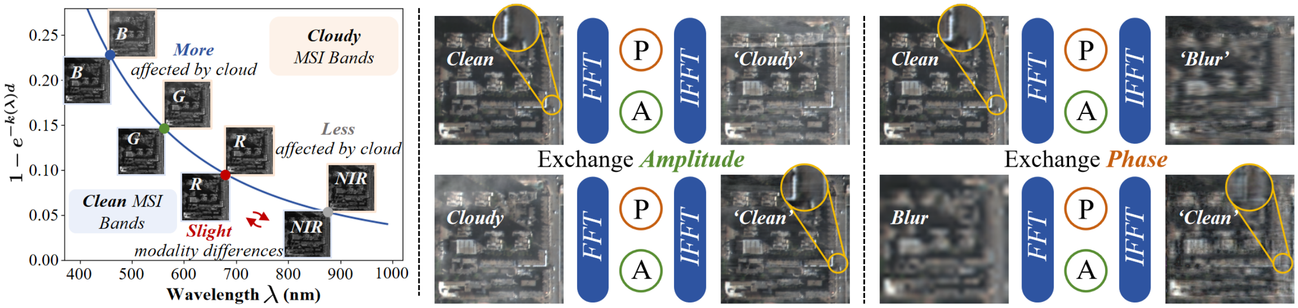
Related Work

Pan-sharpening

Early pansharpening techniques, such as component substitution (Carper et al. 1990), multi-resolution analysis (Aiazzi et al. 2006), and variational optimization (Ballester et al. 2006), are prone to spectral distortions due to their limited capacity for semantic modeling. With the advent of deep learning, CNN-based methods like PNN (Masi et al. 2016), PanNet (Yang et al. 2017), and BDPN (Zhang et al. 2019) marked a turning point, achieving significant fusion improvements. Later works introduced adaptive convolutions (e.g., LAGConv (Jin et al. 2022), CANNet (Duan et al. 2024)) and frequency-domain designs (e.g., SFINet (Zhou et al. 2024), BiMPan (Hou et al. 2023)) to further enhance spatial fidelity. FusionMamba (Peng et al. 2024) and Pan-Mamba (He et al. 2025) attempt to leverage the Mamba architecture for improving efficiency. Recently, diffusion models (Zhong et al. 2024; Kim et al. 2025) have been explored for their generative capacity. However, nearly all existing methods assume cloud-free conditions (Li et al. 2023), limiting their robustness in real-world degraded scenarios.

Thin-Cloud Removal

Unlike thick clouds that completely obscure the surface, thin clouds permit partial signal transmission, thereby allowing algorithmic restoration. Traditional methods rely on priors like spectral correlation (Zhang, Guindon, and Cihlar 2002), dark channel (He, Sun, and Tang 2010), or frequency filtering (Shen et al. 2014), but struggle with complex cloud patterns. Learning-based approaches offer robust solutions, utilizing residual learning (Li et al. 2019), cloud prior learning (e.g., ThiefCloud (Zhao, Feng, and Li 2025)), attention mechanisms (Xu et al. 2025), wavelet and Fourier



(a) (b) (c)

Figure 2: Motivations. (a) Spectral sensitivity to thin-cloud across different wavelengths λ, modeled via atmospheric scattering model $E(\lambda)$, where $E_0(\lambda)$, $L_\infty(\lambda)$ and $e^{-k(\lambda)d}$ denote clean surface radiance, airlight and transmittance, respectively, and $1 - e^{-k(\lambda)d}$ denotes cloud interference. **NIR band is less affected by cloud scattering** than shorter wavelengths (e.g., B, G, R). Although NIR is cloud-robust, its **modality difference** from visible bands presents challenges for spectral consistency. (b) **The amplitude component** primarily reflects global intensity distributions and **is highly sensitive to cloud-induced degradation**, whereas (c) **the phase component** encodes fine structural details and **is more affected by resolution degradation**. Moreover, directly exchanging amplitude or phase between images can result in **inter-frequency inconsistency artifacts** (highlighted by yellow circles).

decompositions (e.g., WaveCNN-CR (Zi et al. 2023)). Recent works also explore Mamba-based architectures (e.g., CR-Famba (Liu, Pan, and Shi 2025)), and transformer variants (Liu, Pan, and Shi 2024; Dai et al. 2024; Song et al. 2023; Li et al. 2025b) for improved feature extraction and global context modeling. Moreover, generative paradigms such as GANs (Tan et al. 2024), VAEs (Ding, Zi, and Xie 2022), and diffusion models (Li et al. 2025a; Sui et al. 2024; Yu, Idris, and Wang 2025), further enhance realism in reconstruction.

Proposed Method

Framework Overview

As illustrated in Fig. 3, we propose **Pan-TCR**, a unified end-to-end framework that simultaneously performs pansharpening and thin-cloud removal by leveraging complementary frequency and spectral restoration strategies. Given a cloud-induced low-resolution multispectral image (LR-MSI, $\mathcal{Y} \in \mathbb{R}^{\frac{H}{r} \times \frac{W}{r} \times C}$) and a cloud-induced high-resolution PAN image ($\mathcal{X} \in \mathbb{R}^{H \times W \times 1}$), the model generates a high-quality, cloud-free high-resolution MSI (HR-MSI, $\mathcal{Z} \in \mathbb{R}^{H \times W \times C}$).

To align spatial dimensions, the LR-MSI is first upsampled using bicubic interpolation and concatenated with the PAN image. The fused input is passed through a shallow residual block to extract initial features $\mathcal{F}_E^0$, which are then processed through a three-stage encoder–decoder pipeline. Each stage integrates two components: the frequency-decoupled restoration (FDR) block, which leverages PAN ($\mathcal{X}^i$, $i \in [0, 1, 2]$) and NIR ($\mathcal{I}^i$) as two degradation prompts for frequency-domain restoration across different resolutions, and the spectral enhancement (SE) module, which refines spectral consistency in the image domain.

The FDR block is designed to separately address cloud-induced amplitude degradation and resolution-related phase degradation. Different from previous frequency-based methods (e.g., DarkIR (Feijoo et al. 2025), FourLLIE (Wang, Wu, and Jin 2023), and FECNet (Huang et al. 2022)) that process amplitude and phase sequentially, risking error accumulation, FDR block employs a dual-branch structure for parallel processing. The NIR band amplitude serves as a cloud-robust prompt to guide cloud-contaminated restoration, while the PAN phase acts as a resolution-robust prompt for spatial enhancement. Both guidance prompts ($\mathcal{X}$, $\mathcal{I}$) are spatially aligned with the input feature $\mathcal{F}$. Using Fast Fourier Transform (FFT), we decouple amplitude and phase components for targeted restoration in the frequency domain.

Phase Prompt-Guided Branch. To model resolution degradation while minimizing thin-cloud interference, we first extract high-frequency structures from the PAN image using a contrast-aware learnable high-pass filter ($\mathcal{H}(\cdot)$):

$$\mathcal{X}_{\text{filtered}} = \mathcal{X} \otimes \sigma\left(|\mathcal{X} - \text{AvgPool}(\mathcal{X})|\right), \tag{1}$$

where σ, $\otimes$, AvgPool$(\cdot)$, and $|\cdot|$ denote the sigmoid function, dot product, average pooling, and absolute operation, respectively. This enhances local structural contrast while suppressing low-frequency noise such as clouds.

Next, we apply FFT to extract phase components $\mathcal{P}_\mathcal{X}$ and $\mathcal{P}_\mathcal{F}$ from $\mathcal{X}_{\text{filtered}}$ and input features $\mathcal{F}$. The concatenated phase inputs are fed into a lightweight Degradation Aware Module (DAM), comprising three convolutional layers with ReLU activations, to obtain phase residual $\mathcal{P}_{deg}$, and the enhanced phase representation can be obtained:

$$\hat{\mathcal{P}} = \mathcal{P}_\mathcal{F} + \mathcal{P}_{deg}. \tag{2}$$

By using $\mathcal{P}_\mathcal{X}$ as a prompt, unlike complex architectures in prior works (e.g., DarkIR (Feijoo et al. 2025), FourierISP (He et al. 2024), FourLLIE (Wang, Wu, and Jin 2023)), our lightweight design efficiently captures localized phase discrepancies, ensuring accurate spatial restoration with reduced computational complexity.

Amplitude Prompt-Guided Branch. This branch mitigates cloud-induced degradation using the NIR band amplitude as a cloud-robust prompt. We apply FFT to obtain

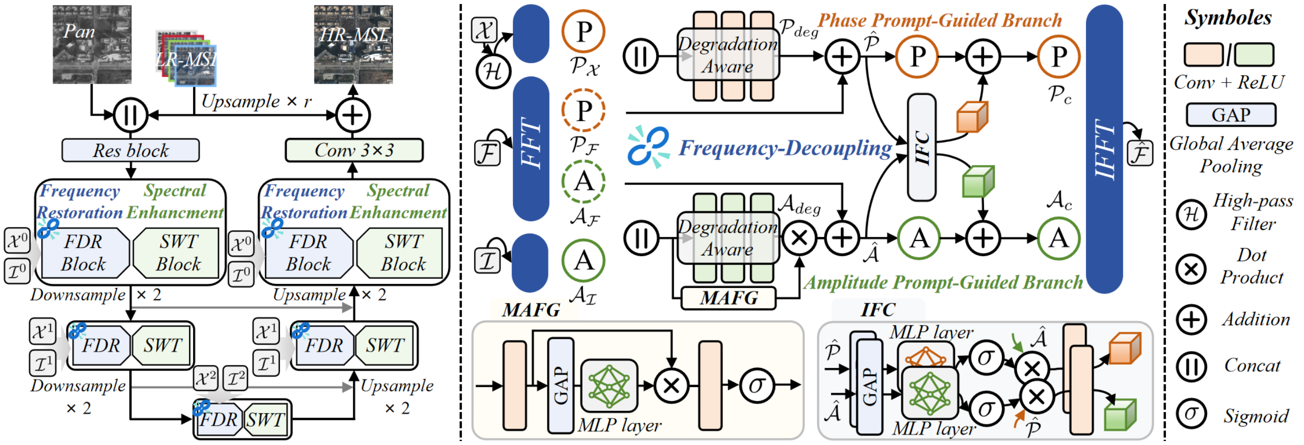


Figure 3: Overview of proposed framework. The left panel illustrates the overall architecture, which takes a stacked input of LR-MSI and PAN images and progressively reconstructs cloud-free high-resolution MSI (HR-MSI) outputs. The central panel illustrates the core frequency-decoupled restoration (FDR) block, which incorporates key sub-modules, including modality-aware frequency gating (MAFG) and interactive inter-frequency consistency (IFC) to guide amplitude and phase restoration via degradation-robust frequency prompts. The right panel provides the symbols used throughout the pipeline.

amplitude components from the cloud-robust NIR band $\mathcal{I}$ and the degraded input feature map $\mathcal{F}$, denoted as $\mathcal{A}_{\mathcal{I}}$ and $\mathcal{A}_{\mathcal{F}}$. Concatenated inputs are passed through a DAM residual predictor to estimate the amplitude residual $\mathcal{A}_{deg}$.

Due to inherent modality discrepancies between NIR and visible bands, direct residual modeling may introduce spectral artifacts. To mitigate this issue, we propose the modality-adaptive frequency gating (MAFG) module, which performs channel-wise spectral gating to suppress modality-inconsistent features and enhance cross-modal compatibility. Specifically, the concatenated amplitude features $[\mathcal{A}_{\mathcal{I}}, \mathcal{A}_{\mathcal{F}}]$ are compressed via a 1×1 convolution, followed by global average pooling to capture spectral statistics. Further, an MLP with sigmoid activation generates a gating vector that reflects the spectral relevance and reliability of each channel in the amplitude domain. This gating vector is subsequently used to modulate $\mathcal{A}_{\mathrm{deg}}$, adaptively suppressing noisy or inconsistent components while preserving cloud-resilient and spectrally aligned information. The enhanced amplitude representation is given by:

$$\hat{\mathcal{A}} = \mathcal{A}_{\mathcal{F}} + f_{gate}([\mathcal{A}_{\mathcal{I}}, \mathcal{A}_{\mathcal{F}}]) \otimes \mathcal{A}_{deg}, \tag{3}$$

where $[\cdot, \cdot]$ and $\otimes$ represent the concatenation and dot product operation, $f_{gate}(\cdot)$ denotes the MAFG function.

Interactive inter-Frequency Consistency (IFC) Module. The FDR block employs a parallel dual-branch strategy to recover amplitude and phase features, guided by distinct degradation-robust priors. However, this independent processing may introduce subtle inconsistencies between the two frequency components, potentially limiting the fidelity of the final reconstruction.

To promote frequency-domain coherence, we embed the IFC module as an interactive refinement unit. It employs a bidirectional cross-modulation mechanism that allows the amplitude and phase branches to guide each other via global feature cues.

Concretely, given the enhanced phase $\hat{\mathcal{P}} \in \mathbb{R}^{H \times W \times C}$ and amplitude $\hat{\mathcal{A}} \in \mathbb{R}^{H \times W \times C}$, we first extract global representations using global average pooling, followed by MLP-based projections and sigmoid activation to obtain modulation weights:

$$\mathbf{w}_{\mathcal{P}} = \sigma(\mathrm{MLP}_{\mathcal{P}}(\mathrm{GAP}(\hat{\mathcal{P}}))), \mathbf{w}_{\mathcal{A}} = \sigma(\mathrm{MLP}_{\mathcal{A}}(\mathrm{GAP}(\hat{\mathcal{A}}))), \tag{4}$$

where $\mathbf{w}_{\mathcal{P}} \in \mathbb{R}^{1 \times 1 \times C}$, $\mathbf{w}_{\mathcal{A}} \in \mathbb{R}^{1 \times 1 \times C}$ serve as cross-frequency modulation weights that encode amplitude-to-phase and phase-to-amplitude guidance, respectively, and are then used to perform channel-wise modulation on the opposite component to capture inter-frequency residuals:

$$\mathcal{P}_{res} = \mathbf{w}_{\mathcal{A}} \otimes \hat{\mathcal{P}}, \mathcal{A}_{res} = \mathbf{w}_{\mathcal{P}} \otimes \hat{\mathcal{A}}, \tag{5}$$

where $\otimes$ denotes channel-wise multiplication via broadcasting. Finally, these residuals are added back to their respective original components to yield inter-frequency consistent representations:

$$\mathcal{P}_c = \mathcal{P}_{res} + \hat{\mathcal{P}}, \mathcal{A}_c = \mathcal{A}_{res} + \hat{\mathcal{A}}. \tag{6}$$

This cross-frequency interaction harmonizes the amplitude and phase domains, mitigating internal inconsistencies and enhancing robustness. The refined components ($\mathcal{P}_c$,$\mathcal{A}_c$) are subsequently combined via Inverse Fast Fourier transform (IFFT) to reconstruct a spatially consistent and cloud-free image representation $\hat{\mathcal{F}}$.

Spectral Enhancement (SE) Module

While frequency-domain processing effectively addresses global degradation patterns such as cloud-induced amplitude suppression and resolution-related phase distortions, it often falls short in modeling fine-grained spectral dependencies and semantic inter-band relationships. Although the proposed MAFG module introduces channel-wise gating to mitigate modality discrepancies in the amplitude spectrum, its low-rank and frequency-based nature inherently limits its ability to perform context-aware spectral refinement.

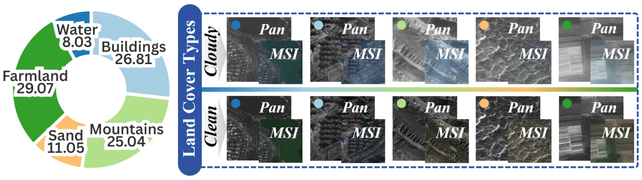


Figure 4: Overview of PanTCR-GF2 dataset: Distribution of land cover types with representative paired examples.

To maintain efficiency while addressing this limitation, we introduce a lightweight SE module in the image domain to refine the spectral fidelity of the restored feature $\hat{\mathcal{F}}$. Using a simple spectral-wise transformer (SWT) (Cai et al. 2022), it models long-range dependencies and adaptively recalibrates spectral responses, providing complementary spectral refinement beyond frequency-guided restoration.

PanTCR-GF2 Dataset

To support research on pansharpening under realistic atmospheric conditions, we construct the first real-world dataset tailored for this task, named **PanTCR-GF2**.

Data Collection. We construct the PanTCR-GF2 dataset using dual-temporal observations captured from the GaoFen-2 satellite, which provides 1-meter PAN and 4-meter LR-MSI images covering four bands: Blue, Green, Red, and NIR. Specifically, we collect 20 scene pairs from identical geographic locations, captured within a one-month interval, one with thin-cloud cover and the other cloud-free, ensuring robust thin-cloud pansharpening evaluation.

Data Preparation. To construct aligned image pairs, we first extract spatially overlapping regions from the dual-temporal scenes, followed by sub-pixel registration across four images (clean MSI, clean PAN, cloudy MSI, and cloudy PAN). Histogram matching is then applied to mitigate radiometric discrepancies caused by illumination variations or seasonal changes. After registration, the images are partitioned into fixed-size patches, and samples exhibiting significant land surface changes, cloudless or thick clouds are excluded to ensure data quality.

The final dataset contains 15,603 precisely aligned cloudy and clean PAN-MSI pairs, including 14,065 for training, 1,438 for validation, 80 reduced-resolution test patches (128×128), and 20 full-resolution test scenes (512×512), following the PanCollection processes. Fig.4 presents statistical distributions of land cover categories, along with representative examples.

Experiment

Datasets and Benchmark

To evaluate the proposed method under realistic and diverse conditions, we conduct experiments on both real and simulated datasets. Specifically, we utilize the real-world PanTCR-GF2 dataset and construct a synthetic dataset by introducing thin-cloud degradation with four distinct morphologies and varying thickness levels onto the widely used WorldView-3 (WV3) pansharpening dataset, collected from the PanCollection.

We compare our method with cutting-edge deep learning methods, including SRPPNN (Cai and Huang 2020), BiMPan (Hou et al. 2023), SFINet (Zhou et al. 2024), PreMix (Cui et al. 2025), DISPNet (Wang et al. 2024a), FusionMamba (Peng et al. 2024), PanMamba (He et al. 2025), SSDiff (Zhong et al. 2024), ARConv (Wang et al. 2025), and WFANet (Huang et al. 2025), and two traditional-based methods (TV (Palsson, Sveinsson, and Ulfarsson 2013) and LRTCFPan (Wu et al. 2023)).

We evaluate comparison methods under two settings: (1) **Two-stage setting**, where a state-of-the-art (SOTA) thin-cloud removal model (ThiefCloud (Zhao, Feng, and Li 2025)) preprocesses both panchromatic (PAN) and low-resolution multi-spectral image (LR-MSI), respectively, followed by the compared pansharpening methods. Specifically, the cloudy LR-MSI and cloudy PAN serve as the input to the decloud stage, with their corresponding clean versions (clean LR-MSI and PAN) as ground truth (GT), respectively. The output decloud PAN and LR-MSI are then used as inputs for pansharpening, supervised by the clean high-resolution MSI (HR-MSI). (2) **End-to-end setting**, where deep learning-based methods are directly applied to the cloudy images as unified pansharpening solutions, supervised by the clean HR-MSI. Note that the end-to-end setting does not apply to traditional methods.

We evaluate pansharpening methods by assessing only their final HR-MSI outputs in reduced- and full-resolution settings. For the reduced-resolution setting, we use quantitative reference-based metrics, including PSNR, SSIM, SAM (Alparone et al. 2007), and ERGAS (Wald 2002), which are commonly adopted for both pansharpening and thin-cloud removal tasks. For the full-resolution setting, we employ widely used no-reference pansharpening metrics: D_λ, D_s, and HQNR (Alparone et al. 2008). To further evaluate the perceptual cloud removal quality of the final HR-MSI, we also report BRISQUE (Mittal, Moorthy, and Bovik 2012), and NIQE (Mittal, Soundararajan, and Bovik 2012), three no-reference metrics commonly used for assessing image degradation and restoration quality.

Implement Details

Our algorithm was implemented in the PyTorch framework, with all experiments conducted on an NVIDIA 4090D GPU. We adopt the l_1 loss function between the predicted high-resolution MSI and the corresponding GT for optimization and utilize the Adam optimizer (Kingma and Ba 2014) with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. A cosine annealing schedule is employed to decay the learning rate from $3e-3$ to $1e-6$ over 500 epochs (Wang et al. 2024c).

Comparison with State-of-the-Art Methods

Table 1 and Table 2 report quantitative results on the PanTCR-GF2 and WV3 datasets under both reduced-resolution and full-resolution settings, evaluating both a **two-stage pipeline** and an **end-to-end setting**.

From the results, we observe that deep learning-based methods with strong generalization capacity, such as

Method	Param (M)↓	FLOPs (G)↓	Time (s)↓	PanTCR-GF2				WV3			
				PSNR(↑)	SSIM(↑)	SAM(↓)	ERGAS(↓)	PSNR(↑)	SSIM(↑)	SAM(↓)	ERGAS(↓)
TV	-	-	24.07	25.50 (-)	0.725 (-)	3.05 (-)	4.85 (-)	30.28 (-)	0.815 (-)	6.14 (-)	6.23 (-)
LRTCFPan	-	-	38.24	23.96 (-)	0.633 (-)	3.19 (-)	5.74 (-)	31.71 (-)	0.876 (-)	5.46 (-)	5.49 (-)
SRPPNN	0.898	21.080	0.006	27.64 (26.95)	0.823(0.782)	2.86 (3.06)	3.83 (4.10)	37.41 (37.35)	0.965 (0.963)	3.45 (3.50)	2.67 (2.48)
BiMPan	0.603	6.480	0.011	27.75 (27.39)	0.827 (0.816)	2.79 (3.01)	3.77 (3.92)	37.22 (34.27)	0.963 (0.937)	3.56 (4.79)	2.59 (3.58)
FusionMa.	0.726	1.360	0.430	27.69 (28.12)	0.822 (0.836)	2.77 (2.98)	3.80 (3.67)	37.00 (37.37)	0.959 (0.963)	3.61 (3.53)	2.62 (2.51)
DISPNet	1.567	27.668	0.009	27.59 (26.77)	0.819 (0.776)	2.78 (3.08)	3.84 (4.19)	37.59 (37.57)	0.964 (0.963)	4.38 (3.35)	3.40 (2.45)
SSDiff	1.288	35.180	4.376	27.58 (27.81)	0.818 (0.828)	2.79 (3.06)	3.86 (3.70)	37.51 (37.62)	0.964 (0.967)	3.42 (3.49)	2.54 (2.41)
SFINet	1.770	11.452	0.015	27.74 (27.36)	0.822 (0.810)	2.87 (3.01)	3.79 (3.89)	35.45 (35.77)	0.947 (0.948)	4.52 (4.52)	3.19 (3.04)
PanMamba	0.478	3.008	0.025	27.65 (27.95)	0.820 (0.831)	2.79 (2.95)	3.82 (3.67)	36.97 (37.82)	0.958 (0.964)	3.35 (3.40)	2.76 (2.44)
PreMix	2.194	21.406	0.014	27.67 (27.72)	0.822 (0.830)	2.78 (2.98)	3.82 (3.62)	37.75 (38.13)	0.966 (0.967)	3.25 (3.36)	2.38 (2.31)
ARConv	4.762	15.35	0.111	27.67 (27.87)	0.819 (0.838)	2.82 (3.06)	3.82 (3.75)	37.30 (37.74)	0.962 (0.966)	3.34 (3.60)	2.52 (2.47)
WFANet	0.507	3.496	0.043	27.93 (28.32)	0.829 (0.852)	2.81 (2.88)	3.73 (3.54)	37.88 (38.37)	0.966 (0.970)	3.18 (3.34)	2.34 (2.22)
Ours	**0.310**	**0.901**	**0.005**	**29.08**	**0.874**	**2.75**	**3.23**	**38.77**	**0.972**	**3.04**	**2.11**

Table 1: Quantitative results on PanTCR-GF2 and WV3 with reduced-resolution test data. Metrics are reported for two-stage processing (outside brackets) and end-to-end processing (inside brackets). **Note:** All two-stage methods need to go through two decloud networks first, so the **"Param"** and **"FLOPs"** columns should add **6.823M** and **8.435G**, respectively. The best and second best results are **highlighted** and underlined.

WFANet and ARConv, tend to perform well in the end-to-end setting on structure-focused metrics (e.g., PSNR, SSIM). However, due to the absence of explicit cloud removal mechanisms, these methods exhibit inferior performance on quality-aware or spectral fidelity metrics (e.g., SAM, NIQE) compared with the two-stage setting, reflecting their limited ability to address cloud-induced degradation. Conversely, lightweight models like SRPPNN and BiMPan suffer a more significant performance drop in the end-to-end mode, likely due to insufficient representational capacity for modeling complex degradations. In contrast, our proposed unified framework consistently achieves superior results across all major metrics, both in reduced- and full-resolution settings.

Furthermore, qualitative comparisons in Fig. 5 further support our findings. Residual maps (MSE between outputs and GT) show that our results have the darkest residuals, indicating superior fidelity over all baselines.

Ablation Experiments

To comprehensively evaluate the effectiveness of each component in our framework, we conduct ablation studies from four key perspectives: (1) degradation modeling, (2) frequency-domain modeling, (3) spectral enhancement, and (4) learning strategy. As summarized in Table 3, experiments are performed on the PanTCR-GF2 dataset using both reduced- and full-resolution test data, with metrics including PSNR, SSIM, and SAM (reduced-resolution), and HQNR and BRISQUE (full-resolution).

1) **Degradation Modeling.** We first remove the two Degradation Aware Modules (DAMs) equipped in the frequency decoupled restoration (FDR) block, and directly fuse amplitude ($\mathcal{A}_{\mathcal{I}}$,$\mathcal{A}_{\mathcal{F}}$) and phase ($\mathcal{P}_{\mathcal{X}}$,$\mathcal{P}_{\mathcal{F}}$) components using concatenation (denoted as 'w/o Deg. Aware'). This results in a consistent degradation across all metrics, indicating that explicit modeling of degradation priors is essential for robust cloud removal and spatial detail preservation. To further investigate the role of degradation-robust prompts, we conduct three additional variants: removing the PAN or NIR individually ('w/o Pan'), only the NIR amplitude guidance ('w/o NIR'), and both ('w/o Pan&NIR'). All variants lead to performance drops, especially in SAM and ERGAS, validating the complementary nature of NIR's cloud robustness and PAN's spatial richness in guiding the restoration.

2) **Frequency-Domain Modeling.** We first assess the effectiveness of the high-pass filter $\mathcal{H}(\cdot)$, which is designed to extract high-frequency details from PAN and suppress thin-cloud interference ('w/o $\mathcal{H}(\cdot)$'), proving critical for preserving structural fidelity under cloud-contaminated conditions. We then evaluate the effectiveness of frequency-specific branches by individually disabling the phase prompt-guided branch ('w/o Pha. Branch') and the amplitude prompt-guided branch ('w/o Amp. branch'), while retaining the cross-modulation from the interactive inter-frequency consistency (IFC) module. Removing the phase prompt-guided Branch leads to a noticeable loss in structural fidelity (↓PSNR, ↓SSIM), whereas removing the amplitude prompt-guided branch affects spectral consistency and cloud suppression (↑SAM, ↑BRISQUE). This confirms the necessity of decoupled processing for targeted restoration.

Further, we explore two structural variants: (1) replacing the entire FDR block with a spatial attention module ('FDR→Spa-Atten.'), and (2) removing the FDR block entirely ('w/o FDR'). Both result in sharp performance degradation across all metrics, highlighting the superiority of frequency-based modeling over purely spatial baselines.

We also ablate the modality-adaptive frequency gating (MAFG) unit ('w/o MAFG'), which leads to reduced fusion quality due to inadequate suppression of modality-induced spectral bias. Additionally, we validate the necessity of the IFC module: replacing it with vanilla channel attention ('IFC→Cha-Atten.') or removing it completely ('w/o IFC') results in degraded performance and misaligned frequency fusion. We also test a serial configuration of the two parallel branches ('Parallel→Series'), confirming that our parallel design enables better decoupling and efficiency.

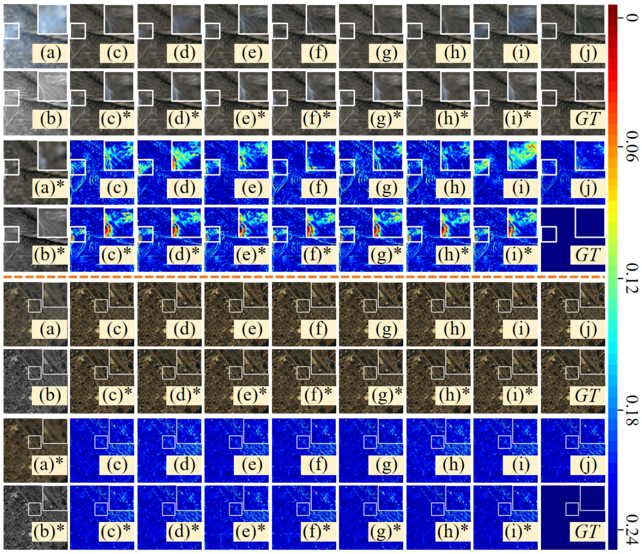
3) **Spectral Enhancement.** To assess the role of the spectral enhancement (SE) module, we remove it ('w/o SE')

Method	PanTCR-GF2					WV3				
	D_λ(↓)	D_s(↓)	HQNR(↑)	BRISQUE(↓)	NIQE(↓)	D_λ(↓)	D_s(↓)	HQNR(↑)	BRISQUE(↓)	NIQE(↓)
SRPPNN	0.033 (0.048)	0.178 (0.147)	0.795 (0.812)	44.05 (44.59)	5.13 (5.43)	0.038 (0.029)	0.041 (0.055)	0.923 (0.918)	34.06 (35.11)	4.22 (4.25)
BiMPan	0.034 (0.042)	0.158 (0.128)	0.813 (0.835)	46.04 (45.46)	5.58 (5.65)	0.032 (0.055)	0.044 (0.060)	0.926 (0.889)	31.91 (**29.50**)	4.26 (4.60)
FusionMa.	0.039 (0.041)	0.177 (0.129)	0.791 (0.836)	46.32 (46.04)	5.32 (5.01)	0.033 (**0.027**)	0.046 (0.055)	0.923 (0.919)	47.83 (34.85)	5.64 (4.19)
DISPNet	0.035 (0.041)	0.179 (0.139)	0.792 (0.826)	48.79 (44.84)	5.67 (4.96)	0.037 (0.032)	0.043 (0.068)	0.922 (0.903)	35.33 (34.42)	4.29 (4.42)
SSDiff	0.037 (0.041)	0.157 (0.120)	0.812 (0.844)	44.98 (44.49)	5.50 (5.64)	0.032 (0.039)	0.045 (0.045)	0.925 (0.918)	34.33 (36.08)	4.53 (4.47)
SFINet	**0.031** (0.045)	0.179 (0.125)	0.796 (0.836)	47.00 (47.20)	5.74 (5.21)	0.033 (0.048)	0.042 (0.044)	0.926 (0.910)	33.98 (34.51)	4.32 (4.34)
PanMamba	0.033 (0.039)	0.165 (0.144)	0.807 (0.823)	44.35 (44.73)	4.96 (4.93)	0.036 (0.029)	0.045 (0.057)	0.921 (0.916)	32.34 (32.69)	**4.07** (4.19)
PreMix	0.035 (0.046)	0.175 (0.142)	0.796 (0.819)	44.88 (45.54)	5.51 (5.48)	0.029 (0.030)	0.041 (0.047)	0.931 (0.924)	34.51 (37.76)	4.22 (4.51)
ARConv	0.036 (0.043)	0.188 (0.127)	0.782 (0.837)	46.50 (46.06)	4.92 (5.10)	0.031 (0.028)	0.041 (0.046)	0.929 (0.927)	32.95 (34.84)	4.28 (4.34)
WFANet	0.036 (0.040)	0.167 (0.118)	0.803 (0.847)	45.26 (45.30)	5.16 (5.36)	0.029 (0.030)	0.044 (0.059)	0.928 (0.913)	32.14 (32.61)	4.32 (4.41)
Ours	0.033	**0.117**	**0.854**	**43.91**	**4.84**	0.028	**0.040**	**0.933**	31.04	4.13

Table 2: Quantitative results on the PanTCR-GF2 and WV3 with full-resolution test data. Metrics are reported for two-stage processing (outside brackets) and end-to-end processing (inside brackets).

	Configurations	PSNR	SSIM	SAM	HQNR	BRISQUE
1	w/o Deg. Aware	27.96	0.843	2.95	0.826	45.32
	w/o Pan	28.32	0.856	2.86	0.838	44.43
	w/o NIR	28.46	0.862	2.93	0.842	44.96
	w/o Pan&NIR	28.03	0.848	2.96	0.823	45.65
2	w/o $\mathcal{H}(\cdot)$	28.33	0.853	2.79	0.844	44.07
	w/o Pha. Branch	27.94	0.838	2.81	0.828	44.77
	w/o Amp. Branch	28.38	0.852	2.92	0.839	45.48
	FDR→ Spa-Atten.	27.94	0.840	2.96	0.825	46.06
	w/o FDR	27.86	0.823	3.07	0.826	46.33
	w/o MAFG	28.45	0.859	2.88	0.848	45.19
	w/o IFC	28.03	0.848	2.90	0.829	45.43
	IFC→Cha-Atten.	28.56	0.866	2.84	0.845	43.96
	Parallel→ Series	28.09	0.846	3.04	0.838	45.06
3	w/o SE	27.96	0.841	3.06	0.829	44.89
	SE →Cha-Atten.	28.22	0.853	2.88	0.839	44.43
4	Unified→Two-stage	28.36	0.853	2.78	0.842	44.03
	Full model	**29.08**	**0.874**	**2.75**	**0.854**	**43.91**

Table 3: Ablation study results with the best results in **bold**.



(a) *LR-MSI* (c) *DISPNet* (e) *FusionMamba* (g) *PreMix* (i) *WFANet*
(b) *PAN* (d) *SSDiff* (f) *PanMamba* (h) *ARConv* (j) *Ours*

Figure 5: Qualitative results on PanTCR-GF2 and WV3 datasets with reduced-resolution, including restored images and residual error maps. Inputs ("LR-MSI", "PAN") or methods marked with "*" indicate two-stage pipelines, where cloud removal is performed prior to pansharpening.

and replace it with channel attention ('SE→Cha-Atten.'), while retaining the FDR block. Both modifications lead to notable performance degradation, particularly on SAM and ERGAS, suggesting that although frequency-based decoupling effectively restores structure and handles degradation, it may fail to recover fine-grained spectral fidelity. The SE module, built on spectral-wise transformer (SWT), complements frequency restoration by introducing localized spectral corrections, thus improving spectral coherence.

4) **Unified Framework.** Finally, we convert our end-to-end unified framework with a cascaded two-stage variant ('Unified→Two-stage'). Despite increased model complexity, the performance declines, especially on cloud-relevant and perceptual metrics, emphasizing the importance of unified joint modeling in enhancing overall performance.

Conclusion

In this paper, we present a unified and lightweight framework that simultaneously addresses thin-cloud removal and pansharpening, effectively tackling the dual degradations caused by cloud contamination and spatial resolution limitations. By leveraging frequency-domain priors from PAN and NIR modalities, our method explicitly decouples amplitude and phase components and performs targeted restoration, enabling accurate cloud suppression and structural detail enhancement within a single end-to-end network. In contrast to conventional two-stage pipelines, our approach offers a computationally efficient solution while achieving superior reconstruction quality. To further support this task, we construct **PanTCR-GF2**, the first real-world dataset tailored for thin-cloud pansharpening, which fills a critical gap in remote sensing benchmarks where cloud-related degradations have long been underexplored. Extensive experiments demonstrate the superiority of our unified approach over existing algorithms in both accuracy and efficiency.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62271400.

References

Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; and Selva, M. 2006. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogrammetric Engineering & Remote Sensing*, 72(5): 591–596.

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

Alparone, L.; Wald, L.; Chanussot, J.; Thomas, C.; Gamba, P.; and Bruce, L. M. 2007. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. *IEEE TGRS*, 45(10): 3012–3021.

Ballester, C.; Caselles, V.; Igual, L.; Verdera, J.; and Rougé, B. 2006. A variational model for P+ XS image fusion. *IJCV*, 69: 43–58.

Cai, J.; and Huang, B. 2020. Super-resolution-guided progressive pansharpening based on a deep convolutional neural network. *IEEE TGRS*, 59(6): 5206–5220.

Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; Pfister, H.; Timofte, R.; and Van Gool, L. 2022. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *CVPR*, 745–755.

Carper, W.; Lillesand, T.; Kiefer, R.; et al. 1990. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogrammetric Engineering and remote sensing*, 56(4): 459–467.

Cui, Y.; Liu, P.; Ma, Y.; Chen, L.; Xu, M.; and Guo, X. 2025. Pansharpening via predictive filtering with element-wise feature mixing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 219: 22–37.

Dai, J.; Shi, N.; Zhang, T.; and Xu, W. 2024. TCME: Thin Cloud removal network for optical remote sensing images based on Multi-dimensional features Enhancement. *IEEE TGRS*.

Ding, H.; Zi, Y.; and Xie, F. 2022. Uncertainty-based thin cloud removal network via conditional variational autoencoders. In *ACCV*, 469–485.

Du, S.; Leng, Y.; Liang, X.; Li, J.; Liu, W.; and Du, Q. 2023. Degradation aware unfolding network for spectral super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 21: 1–5.

Du, S.; Zou, Y.; Wang, Z.; Li, X.; Li, Y.; Shang, C.; and Shen, Q. 2024. Unsupervised Hyperspectral and Multispectral Image Fusion via Self-Supervised Modality Decoupling. *arXiv preprint arXiv:2412.04802*.

Duan, Y.; Wu, X.; Deng, H.; and Deng, L.-J. 2024. Content-adaptive non-local convolution for remote sensing pansharpening. In *CVPR*, 27738–27747.

Feijoo, D.; Benito, J. C.; Garcia, A.; and Conde, M. V. 2025. Darkir: Robust low-light image restoration. In *CVPR*, 10879–10889.

He, K.; Sun, J.; and Tang, X. 2010. Single image haze removal using dark channel prior. *IEEE TPAMI*, 33(12): 2341–2353.

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

He, X.; Hu, T.; Wang, G.; Wang, Z.; Wang, R.; Zhang, Q.; Yan, K.; Chen, Z.; Li, R.; Xie, C.; et al. 2024. Enhancing RAW-to-sRGB with decoupled style structure in Fourier domain. In *AAAI*, volume 38, 2130–2138.

Hou, J.; Cao, Q.; Ran, R.; Liu, C.; Li, J.; and Deng, L.-j. 2023. Bidomain modeling paradigm for pansharpening. In *ACM MM*, 347–357.

Huang, J.; Huang, R.; Xu, J.; Peng, S.; Duan, Y.; and Deng, L.-J. 2025. Wavelet-Assisted Multi-Frequency Attention Network for Pansharpening. In *AAAI*, volume 39, 3662–3670.

Huang, J.; Liu, Y.; Zhao, F.; Yan, K.; Zhang, J.; Huang, Y.; Zhou, M.; and Xiong, Z. 2022. Deep fourier-based exposure correction network with spatial-frequency interaction. In *ECCV*, 163–180. Springer.

Jin, Z.-R.; Zhang, T.-J.; Jiang, T.-X.; Vivone, G.; and Deng, L.-J. 2022. LAGConv: Local-context adaptive convolution kernels with global harmonic bias for pansharpening. In *AAAI*, volume 36, 1113–1121.

Kim, S.; Do, J.; Lee, J.; and Kim, M. 2025. U-Know-DiffPAN: An Uncertainty-aware Knowledge Distillation Diffusion Framework with Details Enhancement for PAN-Sharpening. In *CVPR*, 23069–23079.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Li, H.; and Jing, L. 2017. Improvement of a pansharpening method taking into account haze. *IEEE JSTARS*, 10(11): 5039–5055.

Li, J.; Du, S.; Song, R.; Li, Y.; and Du, Q. 2023. Progressive spatial information-guided deep aggregation convolutional network for hyperspectral spectral super-resolution. *IEEE TNNLS*.

Li, J.; Du, S.; Song, R.; Wu, C.; Li, Y.; and Du, Q. 2022. HASIC-Net: Hybrid attentional convolutional neural network with structure information consistency for spectral super-resolution of RGB images. *IEEE TGRS*, 60: 1–15.

Li, J.; Wang, Y.; Sheng, Q.; Wu, Z.; Wang, B.; Ling, X.; Liu, X.; Du, Y.; Gao, F.; Camps-Valls, G.; et al. 2025a. CloudRuler: Rule-based transformer for cloud removal in Landsat images. *Remote Sensing of Environment*, 328: 114913.

Li, M.; Xu, Q.; Li, K.; and Li, W. 2025b. DecloudFormer: Quest the key to consistent thin cloud removal of wide-swath multi-spectral images. *Pattern Recognition*, 166: 111664.

Li, W.; Li, Y.; Chen, D.; and Chan, J. C.-W. 2019. Thin cloud removal with residual symmetrical concatenation network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 153: 137–150.

Liu, J.; Pan, B.; and Shi, Z. 2024. Cascaded memory network for optical remote sensing imagery cloud removal. *IEEE TGRS*.

Liu, J.; Pan, B.; and Shi, Z. 2025. CR-Famba: A Frequency-Domain Assisted Mamba for Thin Cloud Removal in Optical Remote Sensing Imagery. *IEEE TMM*.

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

Meng, X.; Shen, H.; Yuan, Q.; Li, H.; Zhang, L.; and Sun, W. 2018. Pansharpening for cloud-contaminated very high-resolution remote sensing images. *IEEE TGRS*, 57(5): 2840–2854.

Mittal, A.; Moorthy, A. K.; and Bovik, A. C. 2012. No-reference image quality assessment in the spatial domain. *IEEE TIP*, 21(12): 4695–4708.

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3): 209–212.

Ni, N.; and Zhang, L. 2025. Hazy Low-Quality Satellite Video Restoration Via Learning Optimal Joint Degradation Patterns and Continuous-Scale Super-Resolution Reconstruction. In *CVPR*, 12690–12699.

Palsson, F.; Sveinsson, J. R.; and Ulfarsson, M. O. 2013. A new pansharpening algorithm based on total variation. *IEEE Geoscience and Remote Sensing Letters*, 11(1): 318–322.

Peng, D.; Liu, X.; Zhang, Y.; Guan, H.; Li, Y.; and Bruzzone, L. 2025. Deep learning change detection techniques for optical remote sensing imagery: Status, perspectives and challenges. *International Journal of Applied Earth Observation and Geoinformation*, 136: 104282.

Peng, S.; Zhu, X.; Deng, H.; Deng, L.-J.; and Lei, Z. 2024. Fusionmamba: Efficient remote sensing image fusion with state space model. *IEEE TGRS*.

Shen, H.; Li, H.; Qian, Y.; Zhang, L.; and Yuan, Q. 2014. An effective thin cloud removal procedure for visible remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 96: 224–235.

Song, Y.; He, Z.; Qian, H.; and Du, X. 2023. Vision transformers for single image dehazing. *IEEE TIP*, 32: 1927–1941.

Su, D.; Zhang, Y.; Li, H.; Li, J.; and Liu, Y. 2025. UniFuse: A Unified All-in-One Framework for Multi-Modal Medical Image Fusion Under Diverse Degradations and Misalignments. *arXiv preprint arXiv:2506.22736*.

Sui, J.; Ma, Y.; Yang, W.; Zhang, X.; Pun, M.-O.; and Liu, J. 2024. Diffusion enhancement for cloud removal in ultra-resolution remote sensing imagery. *arXiv preprint arXiv:2401.15105*.

Tan, Z. C.; Du, X. F.; Man, W.; Xie, X. Z.; Wang, G. S.; and Nie, Q. 2024. Unsupervised remote sensing image thin cloud removal method based on contrastive learning. *IET Image Processing*, 18(7): 1844–1861.

Wald, L. 2002. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES.

Wang, C.; Wu, H.; and Jin, Z. 2023. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *ACM MM*, 7459–7469.

Wang, H.; Gong, M.; Mei, X.; Zhang, H.; and Ma, J. 2024a. Deep unfolded network with intrinsic supervision for pansharpening. In *AAAI*, volume 38, 5419–5426.

Wang, P.; Su, Y.; Huang, B.; Zhu, D.; Liu, W.; Nedzved, A.; Krasnoproshin, V. V.; and Leung, H. 2024b. Low-rank tensor completion pansharpening based on haze correction. *IEEE TGRS*, 62: 1–20.

Wang, X.; Zheng, Z.; Shao, J.; Duan, Y.; and Deng, L.-J. 2025. Adaptive Rectangular Convolution for Remote Sensing Pansharpening. In *CVPR*, 17872–17881.

Wang, Z.; Zhang, C.; Chen, Z.; Hu, W.; Lu, K.; Ge, L.; and Wang, Z. 2024c. ACR-Net: Learning High-Accuracy Optical Flow via Adaptive-Aware Correlation Recurrent Network. *IEEE TCSVT*, 34(10): 9064–9077.

Wu, Z.-C.; Huang, T.-Z.; Deng, L.-J.; Huang, J.; Chanussot, J.; and Vivone, G. 2023. LRTCFPan: Low-rank tensor completion based framework for pansharpening. *IEEE TIP*, 32: 1640–1655.

Xi, B.; Zhang, W.; Li, J.; Song, R.; and Li, Y. 2025. HyperTaFOR: Task-Adaptive Few-Shot Open-Set Recognition With Spatial-Spectral Selective Transformer for Hyperspectral Imagery. *IEEE TIP*, 34: 4148–4160.

Xu, X.; He, W.; Xia, Y.; Zhang, H.; Wu, Y.; Jiang, Z.; and Hu, T. 2025. TANet: Thin Cloud Aware Network for Cloud Detection in Optical Remote Sensing Image. *IEEE TGRS*.

Yang, J.; Fu, X.; Hu, Y.; Huang, Y.; Ding, X.; and Paisley, J. 2017. PanNet: A deep network architecture for pansharpening. In *ICCV*, 5449–5457.

Yu, Z.; Idris, M. Y. I.; and Wang, P. 2025. DC4CR: When Cloud Removal Meets Diffusion Control in Remote Sensing. *arXiv preprint arXiv:2504.14785*.

Zhang, Y.; Guindon, B.; and Cihlar, J. 2002. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. *Remote Sensing of Environment*, 82(2-3): 173–187.

Zhang, Y.; Liu, C.; Sun, M.; and Ou, Y. 2019. Pansharpening using an efficient bidirectional pyramid network. *IEEE TGRS*, 57(8): 5549–5563.

Zhao, A.; Feng, R.; and Li, X. 2025. ThiefCloud: A Thickness Fused Thin Cloud Removal Network for Optical Remote Sensing Image With Self-Supervised Learnable Cloud Prior. *IEEE TCSVT*.

Zhong, Y.; Wu, X.; Cao, Z.; Dou, H.-X.; and Deng, L.-J. 2024. Ssdiff: Spatial-spectral integrated diffusion model for remote sensing pansharpening. *NeurIPS*, 37: 77962–77986.

Zhou, M.; Huang, J.; Yan, K.; Hong, D.; Jia, X.; Chanussot, J.; and Li, C. 2024. A general spatial-frequency learning framework for multimodal image fusion. *IEEE TPAMI*.

Zi, Y.; Ding, H.; Xie, F.; Jiang, Z.; and Song, X. 2023. Wavelet integrated convolutional neural network for thin cloud removal in remote sensing images. *Remote Sensing*, 15(3): 781.

Zou, Y.; Chen, Z.; Zhang, Z.; Li, X.; Ma, L.; Liu, J.; Wang, P.; and Zhang, Y. 2026. Contourlet refinement gate framework for thermal spectrum distribution regularized infrared image super-resolution. *IJCV*.