PanComplex: Leveraging Complex-Valued Neural Networks for Enhanced Pansharpening

Chunhui Luo[1], **Dong Li**[1*], **Xiaoliang Ma**[2], **Xin Lu**[1],
Zhiyuan Wang[1], **Jiangtong Tan**[1], **Xueyang Fu**[1]
[1]University of Science and Technology of China, Hefei, China
[2]Geovis, Hefei, China
{luochunhui, dongli6, luxion, ustcwangzhiyuan, jttan}@mail.ustc.edu.cn,
maxl@geovis.com.cn, xyfu@ustc.edu.cn

Abstract

Pansharpening combines panchromatic and low-resolution multispectral images to generate high-resolution multispectral images. Previous studies have explored the connection between pansharpening and the frequency domain, but mostly in the real-valued domain, leaving the complex domain relatively unexplored. To redefine the pansharpening task, we propose a complex-valued spatial-frequency dual-domain framework, PanComplex. To achieve this, we first establish complex representations and introduce basic complex operators tailored to pansharpening, enabling the transformation of multispectral real-valued signals into the complex domain for learning. We then model both spatial and frequency branches to capture global frequency features and local spatial features comprehensively. Finally, we employ a complex-based interaction module to fuse the spatial and frequency features, achieving complementary information across both domains. By using the representation power of the complex domain, PanComplex effectively extracts complementary features from PAN and MS images, thereby enhancing pansharpening performance. Experiments on multiple datasets demonstrate that our method achieves optimal performance with the fewest parameters and exhibits strong generalization ability to other tasks. The source code for this work is publicly available at https://github.com/lch-ustc/PanComplex.

1 Introduction

Multispectral (MS) images contain richer spectral information and have been widely applied in various fields, such as environmental monitoring, agriculture, and mapping services. However, the physical limitations of satellites hinder the direct acquisition of high-resolution multispectral (HRMS) images through sensors. As an alternative, only high-resolution panchromatic (PAN) images and corresponding low-resolution multispectral (LRMS) images can be obtained. Pansharpening involves merging LRMS and PAN images to generate high-resolution multispectral images. In recent years, with the rapid development of deep learning, neural network-based methods have made significant progress in pansharpening. Many methods achieve promising results by designing adaptive network architectures to capture spatial and spectral correlations between panchromatic and multispectral images [Zhou *et al.*, 2022d; Zhou *et al.*, 2022a; Yang *et al.*, 2017; Masi *et al.*, 2016]. However, most existing works are limited to operations in the real-valued domain. While some studies have explored the relationship between pansharpening and the frequency domain [Zhou *et al.*, 2022c; Tan *et al.*, 2024], exploration of the complex domain remains limited, though it naturally aligns with frequency analysis.

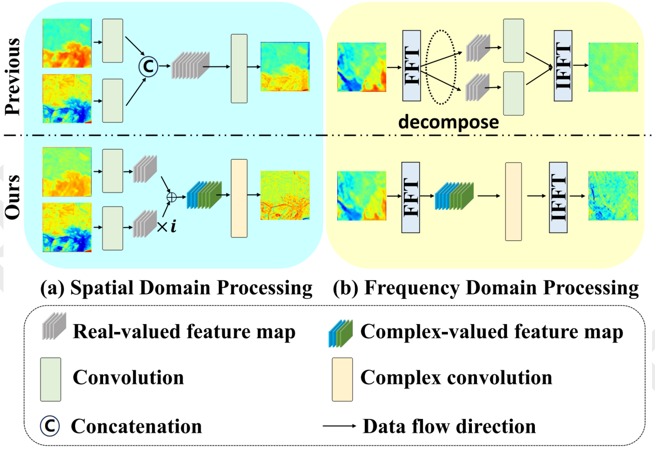

Figure 1: Comparison of our method with previous methods. (a) In the spatial domain, previous methods concatenate the feature maps of the two modalities along the channel dimension and then apply convolution. In contrast, our method combines the two feature maps into a complex-valued representation and processes them with complex convolution. (b) In the frequency domain, previous methods split the complex-valued feature map into real-valued feature maps for convolution. Our approach, however, directly applies complex convolution to the complex-valued feature map.

Complex-valued neural networks (CNNs) have gained widespread attention in recent years due to their performance in signal processing and computer vision tasks [Trabelsi *et al.*, 2017; Quan *et al.*, 2021a; Nguyen *et al.*, 2022]. Com-

[*] Corresponding author.

pared to traditional real-valued neural networks, complex-valued networks can effectively leverage the structural information inherent in the complex domain, providing richer representation capabilities that are applicable in both the frequency and spatial domains. In the frequency domain, Fourier transforms are widely used in pansharpening [Zhou *et al.*, 2022b; He *et al.*, 2023], converting images into the frequency domain and representing them in complex form. Previous methods, limited by real-valued networks, cannot directly process the Fourier spectra and instead decompose them into real and imaginary parts for separate processing. In contrast, complex operators, such as complex convolutions, can operate directly on the Fourier spectra, yielding more continuous and richer frequency representations, as shown in Figure 1. In the spatial domain, complex representations allow PAN and MS images to interact as a whole, facilitating more enriched fusion representations. As shown in Figure 1, complex-based spatial fusion preserves more texture details. Furthermore, Complex-valued neural networks offer compact representations and strong generalization capabilities [Quan *et al.*, 2021b], rendering them well-suited for the lightweight design of pansharpening methods and for enhancing their generalization ability. Based on this, we aim to reconstruct the pansharpening task using complex-valued neural networks to explore the potential of this powerful representation method in pansharpening.

In this work, we establish complex representations for the pansharpening task and introduce basic complex operators. Based on this, we design a spatial-frequency dual-domain pansharpening framework, PanComplex. Given that the frequency domain provides powerful tools for analyzing pansharpening degradation and that the Fourier spectra in the frequency domain align naturally with the complex domain, PanComplex models complex signals in both the spatial and frequency branches to extract complementary features from PAN and MS. The frequency-domain branch directly operates on the Fourier spectra of PAN and MS images in the complex space, yielding more continuous and richer representations. The spatial-domain branch, based on the complex representation of MS and PAN, uses complex convolutions to further facilitate their fusion. In addition, we introduce a complex-based complementary learning mechanism to promote spatial-frequency interaction between PAN and MS. Through these methods, our approach leverages the representational power of the complex domain to enhance the interaction between PAN and MS, improving the complementary representation of both spatial and frequency information, thus significantly improving pansharpening performance. Our contributions are as follows:

- We propose the PanComplex framework based on complex-valued neural networks, the first work to explore complex-valued pansharpening.
- We design a complex-based spatial-frequency dual-domain structure that efficiently integrates spatial and spectral information from PAN and MS images.
- We validate the superiority of the proposed method on multiple datasets, demonstrating strong generalization ability and low computational cost.

Our experimental results show that PanComplex outperforms existing methods on several benchmark datasets, achieving optimal performance with the fewest parameters, and generalizes well to other fusion tasks, such as infrared-RGB fusion.

2 Related Work

2.1 Pansharpening

The pansharpening task aims to fuse low-resolution multispectral images with panchromatic images to obtain high-resolution multispectral images. Traditional pansharpening methods are generally classified into three categories: Component Substitution (CS), Multi-resolution Analysis (MRA), and Variational Optimization (VO). CS methods are simple and feasible but prone to spectral distortion. MRA methods are generally simple and effective, providing good spectral information fusion; however, they lack modeling of the beneficial prior knowledge embedded in the data's spatial dimensions, often leading to spatial distortion. VO methods heavily rely on manually designed prior structural knowledge and often fail to sufficiently model the complex structural priors inherent in the data.

With the success of deep learning methods in various computer vision tasks, deep learning has also been applied to pansharpening, leading to significant improvements. PNN [Masi *et al.*, 2016] first uses a three-layer convolutional neural network to learn the relationship between panchromatic images, low-resolution multispectral images, and high-resolution multispectral images. Subsequently, PANNet [Yang *et al.*, 2017] adopts the residual learning mechanism from ResNet. MSDCNN [Yuan *et al.*, 2018] adds a multiscale module on top of the residual connections. GPPNN [Xu *et al.*, 2021] improves interpretability using a deep unfolding approach. INNformer [Zhou *et al.*, 2022a] introduces the Transformer architecture to the field, effectively modeling long-range dependencies and feature fusion. FAME-Net [He *et al.*, 2024] combines MOE and frequency domain information, enabling the network to dynamically learn high-frequency information in remote sensing images. HFIN [Tan *et al.*, 2024] combines local Fourier and global Fourier information for image fusion.

2.2 Complex-Valued Learning

Complex-valued neural networks are a type of neural network that processes information using complex-valued parameters and variables [Hirose, 1994]. Biological researchers have discovered that, in addition to firing rates, the relative timing of neuronal spikes may also carry information about sensory inputs and dynamic networks [Geman, 2006]. The amplitude and phase of the complex-valued units in complex-valued neural networks, which are similar to biological neurons, have sparked interest among researchers. There have been efforts to extend real-valued network architectures to their complex-valued counterparts. For example, Trabelsi *et al.* [Trabelsi *et al.*, 2017] designed a complex-valued convolutional neural network, Sun *et al.* developed a complex-valued generative adversarial network [Sun *et al.*, 2019]. Based on these network structures, researchers have investigated the

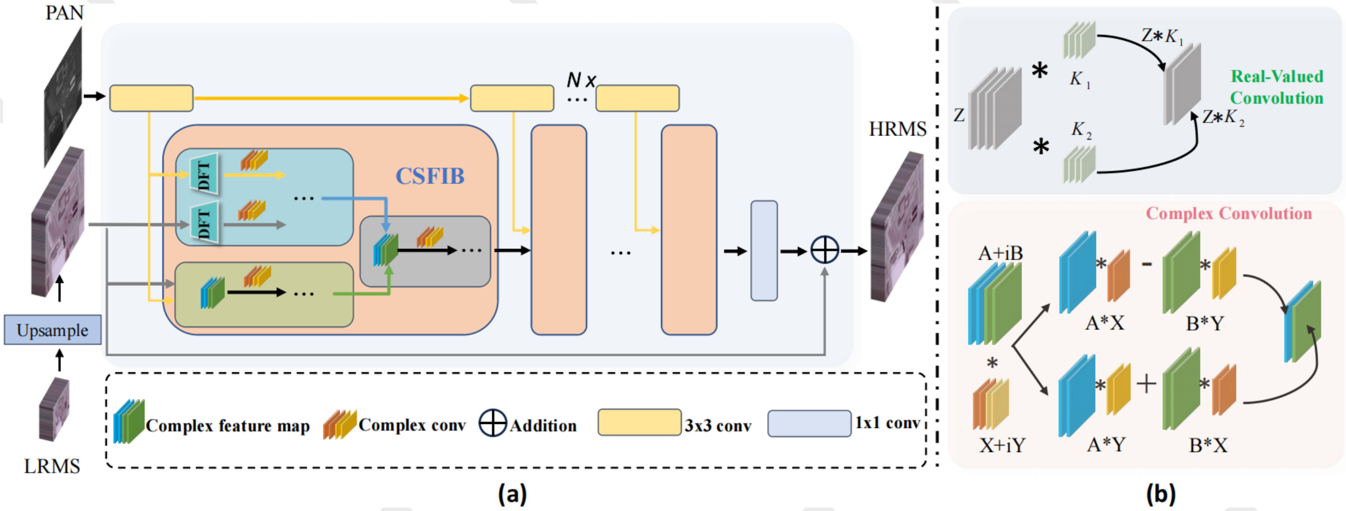

Figure 2: (a) The framework of our proposed pan-sharpening network PanComplex. It consists of several Complex Spatial and Frequency Interaction Blocks (CSFIB). The Complex Spatial and Frequency Interaction Block is shown in Figure 3. (b) The operation process of complex convolution and real-valued convolution.

applications of complex-valued neural networks across various tasks. Alan *et al.* [Oppenheim and Lim, 1981] demonstrated the importance of phase information in speech signals and images. Nguyen *et al.* [Nguyen *et al.*, 2022] applied complex-valued convolutional neural networks to iris recognition. Yadav *et al.* [Yadav and Jerripothula, 2023] designed a novel projection method from real-valued to complex-valued space for RGB images. Quan *et al.* [Quan *et al.*, 2021a] achieved state-of-the-art performance in image denoising using complex-valued convolutional neural networks. In recent years, remote sensing science and technology have seen significant advances [Tian *et al.*, 2024; Deng *et al.*, 2024]. Concurrently, complex-valued neural networks have also been applied to the research of Polarimetric Synthetic Aperture Radar images [Zhang *et al.*, 2017] and Interferometric Synthetic Aperture Radar images [Sunaga *et al.*, 2019]. In image classification, complex-valued convolutional neural networks have shown performance comparable to, or even exceeding, real-valued convolutional neural networks [Yadav and Jerripothula, 2023; Zhang *et al.*, 2017]. In the theoretical domain, studies have also confirmed the viability and potential benefits of complex-valued neural networks [Wu *et al.*, 2023; Zhang *et al.*, 2022]. Our research leverages complex-valued convolutional neural networks for the pansharpening task of remote sensing hyperspectral images.

3 Methodology

3.1 Motivation

In the pansharpening task, the benefit of spatial-frequency interaction has been demonstrated [Zhou *et al.*, 2022c]. Specifically, convolutions are used in the spatial domain to extract local information from both PAN and MS, and in the frequency domain, the Fourier spectrum is decomposed into real and imaginary parts (or amplitude and phase spectra), which are processed separately to extract global frequency information. After performing a 2D Fourier transform, the signal is complex-valued, and existing methods are limited by real-valued neural networks, which can only process the real and imaginary parts separately. While this approach has achieved some success, the internal correlations within the Fourier spectrum may be disrupted, potentially leading to a loss in the generalization ability of the features. In contrast, complex networks, such as those using complex convolution, can directly process signals in the frequency domain, resulting in a more continuous and richer representation space [Nguyen *et al.*, 2022]. As shown in Figure 1, by using complex convolution to process the Fourier spectrum of the image, we obtain better feature interaction effects compared to processing the real and imaginary parts (or amplitude and phase spectra) separately with two convolutions.

Recent studies on pansharping have focused on extracting local features in the spatial domain through convolution blocks. The challenge lies in extracting richer features from the two-modal images. In our work, we replace the convolution blocks in a relatively simple GPPNN model that interact with features from different modalities with complex convolutions, and study the feature maps obtained from both approaches. As shown in Figure 1, we find that the feature maps extracted using complex convolution capture more diverse features from the different modalities. At the same time, we conducted a quantitative analysis of the experiment and present the results in the supplementary materials. Furthermore, complex convolution offers significant advantages in weight compactness. As shown in Figure 2, The response of complex convolution includes two feature maps, which are treated as a single channel. Therefore, using complex convolution networks can halve the model parameters, which is crucial for lightweight panchromatic sharpening. Additionally, biological neurons, apart from their firing rate, encode important information in the relative timing of neural pulses [Geman, 2006]. The complex values in complex neural networks better simulate the output of biological neurons,

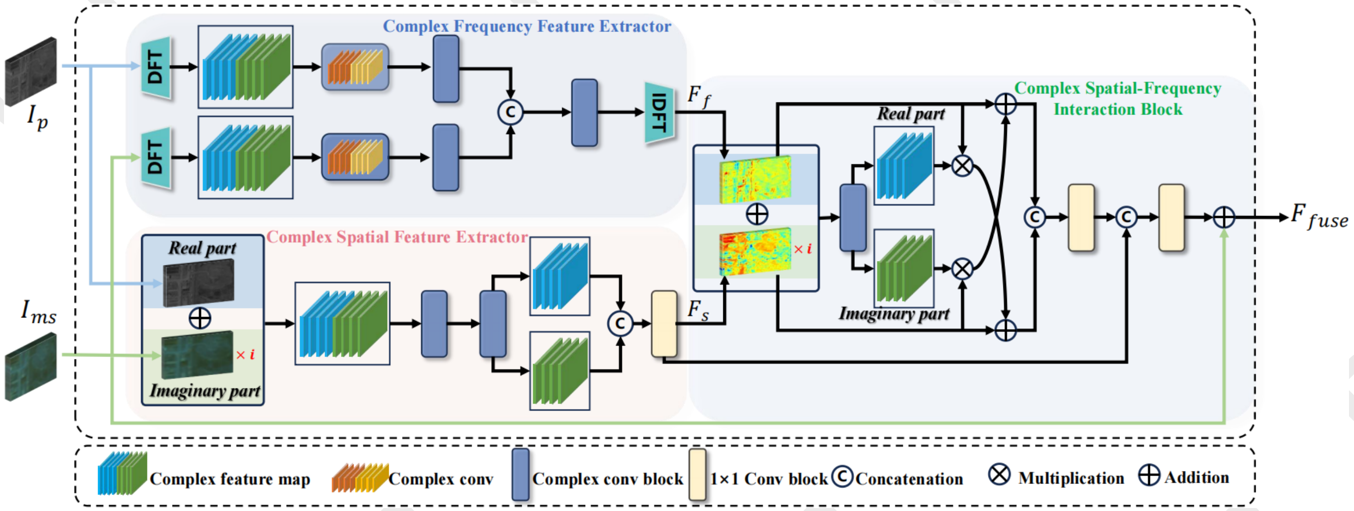

Figure 3: The detailed flowchart of the proposed core building module Complex Spatial and Frequency Interaction Block (CSFIB), consisting of three components: Complex Frequency Feature Extractor (CFFE), Complex Spatial Feature Extractor (CSFE) and Complex Spatial-Frequency Interaction (CSFI).

offering interpretability in the context of biological models [Reichert and Serre, 2013]. Complex neural networks also exhibit strong generalization abilities [Hirose and Yoshida, 2012], which is why our network performs exceptionally well in generalizing to two other image fusion tasks, achieving optimal results in each case.

3.2 Preliminary

Complex-valued convolution. For complex-valued images, each pixel value is represented as a complex number. A complex number z, defined as follows: $z = a + ib \in C, a \in R, b \in R, where\sqrt{-1} = i$,a is a real component and b is an imaginary component. The Complex number can also represented by a magnitude ,$r \in R$, and a phase, $\theta \in R$, as $z = re^{i\theta}$. In real-valued convolution, the convolution operation between a feature map Z and a kernel K is expressed as $Z * K$, where $*$ denotes the convolution operation. For complex-valued feature maps $Z = A + iB$ and complex-valued kernels $K = X + iY$, the convolution operation is denoted as

$$\begin{aligned} Z * K &= (A + iB) * (X + iY) \\ &= (A * X - B * Y) + i\,(A * Y + B * X)\,, \end{aligned} \tag{1}$$

which is expanded as shown in Figure 2. The complex convolution operation is represented as $\mathrm{C-Conv}\,(.)$

Fourier transformation of Images. The Discrete Fourier Transform (DFT) has long been used in image processing for analyzing the frequency content of images. The DFT can be represented in the following form: $x \in R^{H \times W \times C}$ is the given image, and $F\,(X)$ is the resulting complex-valued component, which is expressed as:

$$F\,(x)\,(u, v) = \frac{1}{\sqrt{HW}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} x\,(h, w)\, e^{-j2\pi\left(\frac{h}{H}u + \frac{w}{W}v\right)}. \tag{2}$$

The DFT is performed separately on each channel of the image. The magnitude and phase can be expressed through the following formulas:

$$A\,(x)\,(u, v) = \sqrt{[R^2\,(x)\,(u, v) + I^2\,(x)\,(u, v)]}, \tag{3}$$

$$P\,(x)\,(u, v) = \arctan\left[\frac{I\,(x)\,(u, v)}{R\,(x)\,(u, v)}\right], \tag{4}$$

where $R\,(x)$and $I\,(x)$ represent the real and imaginary parts of $F\,(x)$, respectively.The DFT operation is represented as $\mathrm{F}\,(.)$, and the IDFT operation is represented as $\mathrm{F}^{-1}\,(.)$.

3.3 Overview

We propose a novel pansharpening network PanComplex based on complex convolution, as shown in Figure 2. Given the panchromatic image (PAN) and the low-resolution multispectral image (LRMS), the LRMS is first upsampled using bicubic interpolation. Then, convolutional layers are applied to project both the panchromatic image and the upsampled multispectral image into feature maps of the same dimensionality. The PAN image is then passed through a series of cascaded convolutions to extract a sequence of informative features. Next, the obtained multimodal perceptual feature maps of both the multispectral and panchromatic images are jointly processed through N cascaded key modules, denoted as Complex Spatial and Frequency Interaction Block (CSFIB), for feature extraction and information integration. Finally, the complex feature maps obtained from the N modules are converted back to the image space through a 1×1 convolution, and then the residual of the upsampled LRMS is added to obtain the output image.In our training process, we employ the L_1 loss to optimize the network.

3.4 Complex Feature Extractor

As shown in Figure 3, the complex information extraction module consists of both a Complex Frequency Feature Extractor (CFFE) and a Complex Spatial Feature Extractor

Methods	Params (K)	WorldView II				WorldView III				GaoFen2			
		PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓	PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓	PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓
IHS	-	35.2962	0.9027	0.0461	2.0278	22.5579	0.5354	0.1266	8.3616	38.1754	0.9100	0.0243	1.5336
GS	-	35.6376	0.9176	0.0423	1.8774	22.5608	0.5470	0.1217	8.2433	37.2260	0.9034	0.0309	1.6736
PNN	68.9	40.7550	0.9624	0.0259	1.0646	29.9418	0.9121	0.0824	3.3206	43.1208	0.9704	0.0172	0.8528
GPPNN	119.8	41.1622	0.9684	0.0244	1.0315	30.1785	0.9175	0.0776	3.2593	44.2145	0.9815	0.0137	0.7361
MutNet	71.4	41.6773	0.9705	0.0224	0.9519	30.4907	0.9223	0.0749	3.1125	47.3042	0.9892	0.0102	0.5481
INNformer	70.6	41.6903	0.9704	0.0227	0.9514	30.5365	0.9225	0.0747	3.0997	47.3528	0.9893	0.0102	0.5479
HFIN	77.2	42.2319	0.9714	0.0215	0.8807	30.6147	0.9203	0.0742	3.0786	48.8783	**0.9898**	0.0093	**0.4591**
FAME-Net	140.8	42.0262	0.9723	0.0215	0.9172	30.9903	**0.9287**	0.0697	2.9531	47.6721	**0.9898**	0.0098	0.5242
Ours	**54.2**	**42.3058**	**0.9726**	**0.0210**	**0.8637**	**31.0463**	0.9267	**0.0686**	**2.9239**	**48.8937**	**0.9898**	**0.0092**	0.4607

Table 1: Experimental results of all the competing methods on the three benchmark datasets. The best and the second best values are highlighted in **bold** and underline, respectively.

(CSFE). The two branches extract effective information representations from the image's frequency domain and spatial domain through complex convolution, enabling dual-domain information fusion.

Complex Frequency Feature Extractor. As shown in Figure 3, we first apply the discrete Fourier transform (DFT) to convert the panchromatic and multispectral images into the complex domain. Let the input features of the multispectral and panchromatic images be $\mathrm{I_p}$ and $\mathrm{I_{ms}}$, respectively, with the corresponding Fourier transforms denoted as

$$\mathrm{R}\left(\mathrm{I_p}\right), \mathrm{I}\left(\mathrm{I_p}\right) = \mathrm{F}\left(I_p\right), \tag{5}$$

$$\mathrm{R}\left(\mathrm{I_{ms}}\right), \mathrm{I}\left(\mathrm{I_{ms}}\right) = \mathrm{F}\left(I_{ms}\right), \tag{6}$$

where $\mathrm{R}\left(.\right)$ and $\mathrm{I}\left(.\right)$ represent the real and imaginary parts. Then, we pass these through a two-layer complex-valued 1x1 convolutional neural network with complex ReLU activations. The resulting complex feature maps from the two images are concatenated and reduced in dimensionality through a complex-valued 1x1 convolution. Finally, we apply the inverse discrete Fourier transform (IDFT) to convert the obtained complex feature map back to the spatial domain to produce the frequency-domain feature map $\mathrm{F_f}$. The entire process is formulated as follows:

$$\mathrm{F_f} = \mathrm{F}^{-1}\left(\mathrm{Cat}\left(\mathrm{C-Conv}\left(\mathrm{F}\left(\mathrm{I_p}\right)\right), \mathrm{C-Conv}\left(\mathrm{F}\left(\mathrm{I_{ms}}\right)\right)\right)\right). \tag{7}$$

Complex Spatial Feature Extractor. As shown in Figure 3, we combine the $\mathrm{I_p}$ and $\mathrm{I_{ms}}$ into a complex-valued representation, which results in better performance compared to working in the real domain. Based on this, we design a spatial-domain information interaction branch, Complex Spatial Feature Extractor (CSFE), which conducts deep feature interaction through complex convolution in the spatial domain between the PAN and LRMS features. Specifically, we treat the multispectral feature $\mathrm{I_{ms}}$ as the real part and the panchromatic feature $\mathrm{I_p}$ as the imaginary part, forming a complex feature map Z,which is formulated as: $\mathrm{Z} = \mathrm{I_{ms}} + \mathrm{iI_p}$. We then pass the complex feature map through several residual blocks with a 3×3 complex convolution layer for information interaction between the multispectral and panchromatic features, eventually obtaining the spatial-domain information $\mathrm{F_s}$.

3.5 Complex Spatial-Frequency Interaction

Spatial-frequency interaction is critical for pansharpening [Zhou *et al.*, 2022c]. Since the powerful capabilities of complex convolutions have been demonstrated in Section 3.1, we have designed a Complex Spatial-Frequency Interaction (CSFI) that leverages complex convolutions to perform complementary learning of spatial and frequency domain information, facilitating the deep fusion of panchromatic and multispectral image data. Specifically, we first combine the frequency-domain and spatial-domain features $\mathrm{F_f}$ and $\mathrm{F_s}$ into a complex-valued feature $\mathrm{F_c}$. Then, we use complex convolution to fully exploit the complementary information from both domains, outputting a set of complex-valued attention maps $\mathrm{A_{map}}$, which are weighted by the corresponding real and imaginary parts. To facilitate further interaction of complementary information, we add the weighted real part of the feature map to the imaginary part of the input, and the weighted imaginary part of the feature map to the real part of the input, resulting in a complex-valued feature map $\mathrm{F_{ca}}$ with stronger cross-domain interaction, which can be expressed as

$$\mathrm{R}\left(\mathrm{F_{ca}}\right) = \mathrm{R}\left(\mathrm{A_{map}}\right) \odot \mathrm{R}\left(\mathrm{F_c}\right) + \mathrm{I}(\mathrm{F_c}), \tag{8}$$

$$\mathrm{I}\left(\mathrm{F_{ca}}\right) = \mathrm{I}\left(\mathrm{A_{map}}\right) \odot \mathrm{I}\left(\mathrm{F_c}\right) + \mathrm{R}(\mathrm{F_c}). \tag{9}$$

Since the target image to be reconstructed is real-valued, we finally convert the complex-valued feature map back to the real domain and obtain the real-valued feature map $\mathrm{F_a}$. As the feature map from the spatial domain branch is closer to the HRMS that needs to be restored, we perform residual connection with $\mathrm{F_a}$ and $\mathrm{F_s}$ and pass the resulting feature map through a 1x1 convolution block to obtain the final output feature map $\mathrm{F_{fuse}}$.

$$\mathrm{F_{fuse}} = \mathrm{I_{ms}} + \mathrm{Conv}_{1\times1}\left(\mathrm{Cat}\left(\mathrm{F_s}, \mathrm{F_a}\right)\right). \tag{10}$$

4 Experiments

4.1 Dataset and Benchmarks

To evaluate the effectiveness of our network, we conduct experiments on three satellite datasets: WorldView-II (WV2), Gaofen2 (GF2) and WorldView-III (WV3). Each dataset contains a large number of paired low-resolution multispectral images and panchromatic images, which are divided into training, validation, and test sets. The dataset construction follows the methodology of previous studies, using the Wald protocol [Wald *et al.*, 1997] tool to generate training and

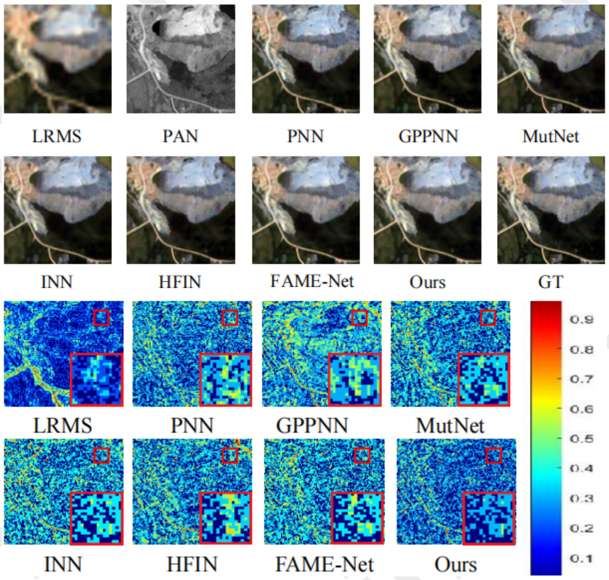

Figure 4: The visual comparisons between other methods and our method on WorldView-II satellite

testing data. To evaluate the effectiveness of our approach, we compare it against several state-of-the-art pansharpening methods, including PNN [Masi *et al.*, 2016], GPPNN [Xu *et al.*, 2021], MutNet [Zhou *et al.*, 2022d], INNformer [Zhou *et al.*, 2022a], HFIN [Tan *et al.*, 2024], and FAME-Net [He *et al.*, 2024], as well as traditional pansharpening techniques such as GS [Laben and Brower, 2000], and IHS [Haydn, 1982].

4.2 Implementation Details

In our experiments, all deep learning models are implemented using PyTorch and trained on an NVIDIA GeForce GTX 3090 GPU. For each dataset, the multispectral (MS) images are cropped into patches of size 32×32, while the corresponding panchromatic (PAN) images are resized to 128×128. During the training phase, the networks are optimized using the Adam optimizer with an initial learning rate of 1×10^{-4}. After 200 epochs, the learning rate is reduced by half. We adopt standard evaluation metrics, including PSNR, SSIM, SAM, and ERGAS. Furthermore, three widely-used no-reference image quality assessment metrics, namely D_λ, D_S and QNR [Vivone *et al.*, 2020], are employed to evaluate real-world full-resolution scenes.

4.3 Comparison With State-of-the-Art Methods

Evaluation on reduced-resolution scenes. The comparison results across the three datasets are presented in Table 1, where the best results for each metric are highlighted in red, and the second-best results are indicated in blue. The results demonstrate that our method outperforms state-of-the-art approaches on the majority of metrics, with the only two exceptions being metrics where our method ranks second. Overall, our approach achieves the best performance across all three datasets. In particular, with respect to the Peak Signal-to-Noise Ratio (PSNR), our method surpasses all other methods, indicating the high consistency between the images processed by our approach and the real high-resolution images. We also conducted qualitative experiments, and Figure 4 shows samples from the WV2 dataset. The first two rows compare the experimental results, while the last row displays the mean squared error (MSE) between the results of each network and the ground truth. It is evident that our results are closest to the ground truth, with minimal differences.

	PNN	GPPNN	INNformer	MutNet	HFIN	FAME-Net	Ours
$D_\lambda \downarrow$	0.0746	0.0782	0.0697	0.0694	0.0710	0.0674	**0.06622**
$D_s \downarrow$	0.1164	0.1253	0.1128	0.1118	0.1098	0.1121	**0.1040**
QNR ↑	0.8191	0.8073	0.8253	0.8247	0.8261	0.8291	**0.8375**

Table 2: Non-reference metrics on full-resolution dataset.

Evaluation on full-resolution scenes. We also conduct a full-resolution analysis in real-world scenarios to further validate the generalization capability of our method. Experiments are carried out on an additional 200 sets from the GF2 dataset. Since high-resolution multispectral images are unavailable for real-world scenarios, we utilize three commonly employed no-reference metrics, namely D_λ, D_S, and QNR, for evaluation. The experimental results, presented in Table 2, show that our method achieves superior performance across all three metrics.

Evaluation on other fusion tasks. To further assess the generalization capability of our proposed method, we apply it to two additional tasks: depth image super-resolution using the NYU v2 dataset and infrared-RGB fusion on the RoadScene dataset. As for depth image super-resolution, we compare our proposed method with nine state-of-the-art deep image super-resolution methods: GF [He *et al.*, 2012], DMSG [Hui *et al.*, 2016], DJFR [Li *et al.*, 2019], DSRNet [Guo *et al.*, 2018], PacNet [Su *et al.*, 2019] and HFIN [Tan *et al.*, 2024]. As for infrared-RGB fusion, we compared our proposed method with ten state-of-the-art visible and infrared image fusion methods: DDcGAN [Ma *et al.*, 2020], DIDFuse

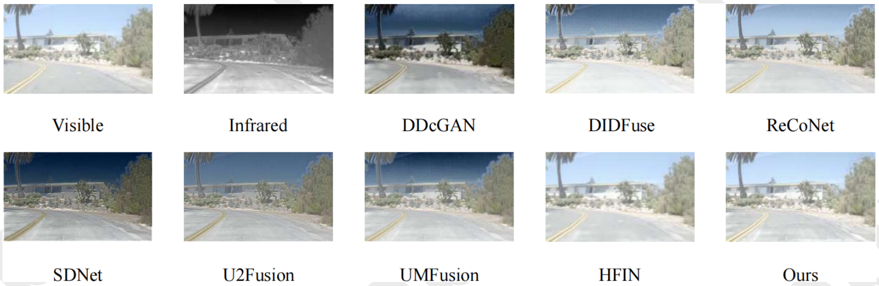

Figure 5: The visual comparisons between other infrared-RGB fusion methods and our method on RoadScene dataset

[Zhao *et al.*, 2020], ReCoNet [Huang *et al.*, 2022], SDNet [Zhang and Ma, 2021], U2Fusion [Xu *et al.*, 2020], UMFusion [Wang *et al.*, 2022] and HFIN [Tan *et al.*, 2024]. For the infrared-RGB fusion task, we evaluate the quality of the generated images using three metrics: MI, VIF, and FMI. For the depth image super-resolution task, we use RMSE as the metric to assess the quality of the generated images. The experimental results, shown in Table 3, highlight the best results in red and the second-best in blue. These results demonstrate that our method outperforms competing approaches in the depth image super-resolution task. This further supports the robustness and versatility of our approach, suggesting that our network can be effectively extended to other fusion tasks as well. We also conduct a qualitative analysis on the infrared-visible image fusion task. Figure 5 presents samples from the RoadScene dataset, showing the results of other state-of-the-art methods alongside those of our approach. It is evident that the images generated by our method exhibit no distortion relative to the visible light image, while also capturing the texture details of the infrared image.

Methods	RoadScene		
	MI↑	VIF↑	FMI↑
DDcGAN	2.6177	0.5945	0.859
DIDFuse	3.1840	0.8274	0.853
ReCoNet	3.1594	0.7955	0.858
SDNet	3.4225	0.8207	0.863
U2Fusion	2.8109	0.7401	0.861
UMFusion	3.2018	0.7912	0.866
HFIN	4.8114	0.8670	0.878
Ours	**4.9121**	**0.8754**	**0.881**

(a) Results of the infrared-visible image fusion task.

Method	NYU v2		
	×4	×8	×16
Bicubic	4.71	8.29	13.17
GF	5.84	7.86	12.41
DMSG	3.02	5.38	9.17
DJFR	2.38	4.94	9.18
DSRNet	3.00	5.16	8.41
PacNet	1.89	3.33	6.78
HFIN	1.53	3.19	6.44
Ours	**1.51**	**3.09**	**6.25**

(b) Results of the depth image super-resolution task.

Table 3: Quantitative comparisons with other fusion methods

4.4 Ablation Experiments

We conducted an ablation study using the WorldView-II dataset to further demonstrate the effectiveness of our method. CFFE and CSFE are the core components of the proposed method, and thus, ablation studies were performed for each of them. Additionally, we also conducted an ablation study on the CSFI of the spatial-frequency fusion.

To validate the effectiveness of the complex-based spatial-frequency branches, we replaced the complex representations and operations with their real-valued counterparts, as shown in Table 4. The results in Table 4 indicate that the removal of either of the complex-based components leads to a performance degradation, thereby confirming the powerful representational ability of complex numbers in the spatial-frequency dual domain for pansharpening.

Configuration	CFFE	CSFE	PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓
I	✗	✓	41.9394	0.9701	0.0219	0.9145
II	✓	✗	42.1111	0.9710	0.0217	0.8945
Ours	✓	✓	**42.3058**	**0.9726**	**0.0210**	**0.8637**

Table 4: Ablation experiment results for CFFE and CSFE on the WorldViewII are presented, with the best values highlighted in **bold**.

We further conducted an ablation study on the complex operations used in the spatial-frequency interaction module to evaluate the impact of complex operations on spatial-frequency interaction. Specifically, we replaced the CSFI with two alternatives: a convolution block with residual connections, and a convolution block with residual connections following the spatial attention mechanism (SA), and compared the results with our method. The results, as shown in Table 5, indicate that switching to complex operations enhances the performance of spatial-frequency collaboration.

Configuration	PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓
Conv	41.9394	0.9701	0.0219	0.9145
SA + Conv	42.2142	0.9712	0.0215	0.8830
CSFI	**42.3058**	**0.9726**	**0.0210**	**0.8637**

Table 5: Ablation experiment results for CSFI on the WorldViewII are presented, with the best values highlighted in **bold**.

5 Conclusion

In conclusion, we present PanComplex, a novel complex-valued neural network framework for pansharpening, marking the first attempt to leverage the rich potential of the complex domain for this task. By introducing a dual-domain spatial-frequency structure, we effectively capture and fuse both spatial and spectral features from panchromatic and multispectral images. This framework not only exploits the natural alignment between the complex domain and the frequency domain but also introduces complex convolutions and complementary learning mechanisms to enhance the fusion process. Our experimental results demonstrate that PanComplex achieves superior performance with fewer parameters compared to existing methods, while also exhibiting strong generalization capabilities for other fusion tasks.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC) under Grants 62422609 and 62476260.

References

[Deng *et al.*, 2024] Jie Deng, Wei Wang, Huiqiang Zhang, Tao Zhang, and Jun Zhang. Polsar ship detection based on superpixel-level contrast enhancement. *IEEE Geoscience and Remote Sensing Letters*, 2024.

[Geman, 2006] Stuart Geman. Invariance and selectivity in the ventral visual pathway. *Journal of Physiology-Paris*, 100(4):212–224, 2006.

[Guo *et al.*, 2018] Chunle Guo, Chongyi Li, Jichang Guo, Runmin Cong, Huazhu Fu, and Ping Han. Hierarchical features driven residual learning for depth map super-resolution. *IEEE Transactions on Image Processing*, 28(5):2545–2557, 2018.

[Haydn, 1982] R Haydn. Application of the ihs color transform to the processing of multisensor data and image enhancement. In *Proc. of the International Symposium on Remote Sensing of Arid and Semi-Arid Lands, Cairo, Egypt, 1982*, 1982.

[He *et al.*, 2012] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6):1397–1409, 2012.

[He *et al.*, 2023] Xuanhua He, Keyu Yan, Rui Li, Chengjun Xie, Jie Zhang, and Man Zhou. Pyramid dual domain injection network for pan-sharpening. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12908–12917, 2023.

[He *et al.*, 2024] Xuanhua He, Keyu Yan, Rui Li, Chengjun Xie, Jie Zhang, and Man Zhou. Frequency-adaptive pan-sharpening with mixture of experts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 2121–2129, 2024.

[Hirose and Yoshida, 2012] Akira Hirose and Shotaro Yoshida. Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. *IEEE Transactions on Neural Networks and learning systems*, 23(4):541–551, 2012.

[Hirose, 1994] Akira Hirose. Applications of complex-valued neural networks to coherent optical computing using phase-sensitive detection scheme. *Information Sciences-Applications*, 2(2):103–117, 1994.

[Huang *et al.*, 2022] Zhanbo Huang, Jinyuan Liu, Xin Fan, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Reconet: Recurrent correction network for fast and efficient multi-modality image fusion. In *European conference on computer Vision*, pages 539–555. Springer, 2022.

[Hui *et al.*, 2016] Tak-Wai Hui, Chen Change Loy, and Xiaoou Tang. Depth map super-resolution by deep multi-scale guidance. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pages 353–369. Springer, 2016.

[Laben and Brower, 2000] Craig A Laben and Bernard V Brower. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening, January 4 2000. US Patent 6,011,875.

[Li *et al.*, 2019] Yijun Li, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Joint image filtering with deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1909–1923, 2019.

[Ma *et al.*, 2020] Jiayi Ma, Han Xu, Junjun Jiang, Xiaoguang Mei, and Xiao-Ping Zhang. Ddcgan: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. *IEEE Transactions on Image Processing*, 29:4980–4995, 2020.

[Masi *et al.*, 2016] Giuseppe Masi, Davide Cozzolino, Luisa Verdoliva, and Giuseppe Scarpa. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7):594, 2016.

[Nguyen *et al.*, 2022] Kien Nguyen, Clinton Fookes, Sridha Sridharan, and Arun Ross. Complex-valued iris recognition network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):182–196, 2022.

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

[Quan *et al.*, 2021a] Yuhui Quan, Yixin Chen, Yizhen Shao, Huan Teng, Yong Xu, and Hui Ji. Image denoising using complex-valued deep cnn. *Pattern Recognition*, 111:107639, 2021.

[Quan *et al.*, 2021b] Yuhui Quan, Peikang Lin, Yong Xu, Yuesong Nan, and Hui Ji. Nonblind image deblurring via deep learning in complex field. *IEEE Transactions on Neural Networks and Learning Systems*, 33(10):5387–5400, 2021.

[Reichert and Serre, 2013] David P Reichert and Thomas Serre. Neuronal synchrony in complex-valued deep networks. *arXiv preprint arXiv:1312.6115*, 2013.

[Su *et al.*, 2019] Hang Su, Varun Jampani, Deqing Sun, Orazio Gallo, Erik Learned-Miller, and Jan Kautz. Pixel-adaptive convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11166–11175, 2019.

[Sun *et al.*, 2019] Qigong Sun, Xiufang Li, Lingling Li, Xu Liu, Fang Liu, and Licheng Jiao. Semi-supervised complex-valued gan for polarimetric sar image classification. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 3245–3248. IEEE, 2019.

[Sunaga *et al.*, 2019] Yuki Sunaga, Ryo Natsuaki, and Akira Hirose. Land form classification and similar land-shape discovery by using complex-valued convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10):7907–7917, 2019.

[Tan *et al.*, 2024] Jiangtong Tan, Jie Huang, Naishan Zheng, Man Zhou, Keyu Yan, Danfeng Hong, and Feng Zhao. Revisiting spatial-frequency information integration from a hierarchical perspective for panchromatic and multispectral image fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25922–25931, 2024.

[Tian *et al.*, 2024] Zhuangzhuang Tian, Wei Wang, Kai Zhou, Xiaoxiang Song, Yilong Shen, and Shengqi Liu. Weighted pseudo-labels and bounding boxes for semisupervised sar target detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:5193–5203, 2024.

[Trabelsi *et al.*, 2017] Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J Pal. Deep complex networks. *arXiv preprint arXiv:1705.09792*, 2017.

[Vivone *et al.*, 2020] Gemine Vivone, Mauro Dalla Mura, Andrea Garzelli, Rocco Restaino, Giuseppe Scarpa, Magnus O Ulfarsson, Luciano Alparone, and Jocelyn Chanussot. A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. *IEEE Geoscience and Remote Sensing Magazine*, 9(1):53–81, 2020.

[Wald *et al.*, 1997] Lucien Wald, Thierry Ranchin, and Marc 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.

[Wang *et al.*, 2022] Di Wang, Jinyuan Liu, Xin Fan, and Risheng Liu. Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration. *arXiv preprint arXiv:2205.11876*, 2022.

[Wu *et al.*, 2023] Jin-Hui Wu, Shao-Qun Zhang, Yuan Jiang, and Zhi-Hua Zhou. Complex-valued neurons can learn more but slower than real-valued neurons via gradient descent. *Advances in Neural Information Processing Systems*, 36:23714–23747, 2023.

[Xu *et al.*, 2020] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):502–518, 2020.

[Xu *et al.*, 2021] Shuang Xu, Jiangshe Zhang, Zixiang Zhao, Kai Sun, Junmin Liu, and Chunxia 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.

[Yadav and Jerripothula, 2023] Saurabh Yadav and Koteswar Rao Jerripothula. Fccns: Fully complex-valued convolutional networks using complex-valued color model and loss function. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10689–10698, 2023.

[Yang *et al.*, 2017] Junfeng Yang, Xueyang Fu, Yuwen Hu, Yue Huang, Xinghao Ding, and John Paisley. Pannet: A deep network architecture for pan-sharpening. In *Proceedings of the IEEE international conference on computer vision*, pages 5449–5457, 2017.

[Yuan *et al.*, 2018] Qiangqiang Yuan, Yancong Wei, Xiangchao Meng, Huanfeng Shen, and Liangpei Zhang. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3):978–989, 2018.

[Zhang and Ma, 2021] Hao Zhang and Jiayi Ma. Sdnet: A versatile squeeze-and-decomposition network for real-time image fusion. *International Journal of Computer Vision*, 129(10):2761–2785, 2021.

[Zhang *et al.*, 2017] Zhimian Zhang, Haipeng Wang, Feng Xu, and Ya-Qiu Jin. Complex-valued convolutional neural network and its application in polarimetric sar image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12):7177–7188, 2017.

[Zhang *et al.*, 2022] Shao-Qun Zhang, Wei Gao, and Zhi-Hua Zhou. Towards understanding theoretical advantages of complex-reaction networks. *Neural Networks*, 151:80–93, 2022.

[Zhao *et al.*, 2020] Zixiang Zhao, Shuang Xu, Chunxia Zhang, Junmin Liu, Pengfei Li, and Jiangshe Zhang. Didfuse: Deep image decomposition for infrared and visible image fusion. *arXiv preprint arXiv:2003.09210*, 2020.

[Zhou *et al.*, 2022a] Man Zhou, Jie Huang, Yanchi Fang, Xueyang Fu, and Aiping Liu. Pan-sharpening with customized transformer and invertible neural network. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 3553–3561, 2022.

[Zhou *et al.*, 2022b] Man Zhou, Jie Huang, Chongyi Li, Hu Yu, Keyu Yan, Naishan Zheng, and Feng Zhao. Adaptively learning low-high frequency information integration for pan-sharpening. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 3375–3384, 2022.

[Zhou *et al.*, 2022c] Man Zhou, Jie Huang, Keyu Yan, Hu Yu, Xueyang Fu, Aiping Liu, Xian Wei, and Feng Zhao. Spatial-frequency domain information integration for pan-sharpening. In *European conference on computer vision*, pages 274–291. Springer, 2022.

[Zhou *et al.*, 2022d] Man Zhou, Keyu Yan, Jie Huang, Zihe Yang, Xueyang Fu, and Feng Zhao. Mutual information-driven pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1798–1808, 2022.