Spatial-Frequency Domain Information Integration for Pan-Sharpening

Man Zhou[1,2], Jie Huang[1], Keyu Yan[1,2], Hu Yu[1], Xueyang Fu[1], Aiping Liu[1], Xian Wei[3], and Feng Zhao[1(✉)]

[1] University of Science and Technology of China, Hefei, China
{manman,hj0117}@mail.ustc.edu.cn, fzhao956@ustc.edu.cn
[2] Hefei Institute of Physical Science, Chinese Academy of Sciences, Hefei, China
[3] MoE Engineering Research Center of Hardware/Software Co-design Technology and Application, East China Normal University, Shanghai, China

Abstract. Pan-sharpening aims to generate high-resolution multi-spectral (MS) images by fusing PAN images and low-resolution MS images. Despite its great advances, most existing pan-sharpening methods only work in the spatial domain and rarely explore the potential solutions in the frequency domain. In this paper, we first attempt to address pan-sharpening in both spatial and frequency domains and propose a Spatial-Frequency Information Integration Network, dubbed as SFIIN. To implement SFIIN, we devise a core building module tailored with pan-sharpening, consisting of three key components: spatial-domain information branch, frequency-domain information branch, and dual domain interaction. To be specific, the first employs the standard convolution to integrate the local information of two modalities of PAN and MS images in the spatial domain, while the second adopts deep Fourier transformation to achieve the image-wide receptive field for exploring global contextual information. Followed by, the third is responsible for facilitating the information flow and learning the complementary representation. We conduct extensive experiments to validate the effectiveness of the proposed network and demonstrate the favorable performance against other state-of-the-art methods.

Keywords: Pan-sharpening · Spatial-frequency domain

1 Introduction

Pan-sharpening is the process of super-resolving the low-resolution (LR) multi-spectral (MS) images in the spatial domain to generate the expected high-resolution (HR) MS images, conditioning on the paired high-resolution PAN images. In other words, pan-sharpening is essentially a PAN-guided MS image

M. Zhou and J. Huang—Co-first authors contributed equally.
F. Zhao—Corresponding author.

S. Avidan et al. (Eds.): ECCV 2022, LNCS 13678, pp. 274–291, 2022.
https://doi.org/10.1007/978-3-031-19797-0_16

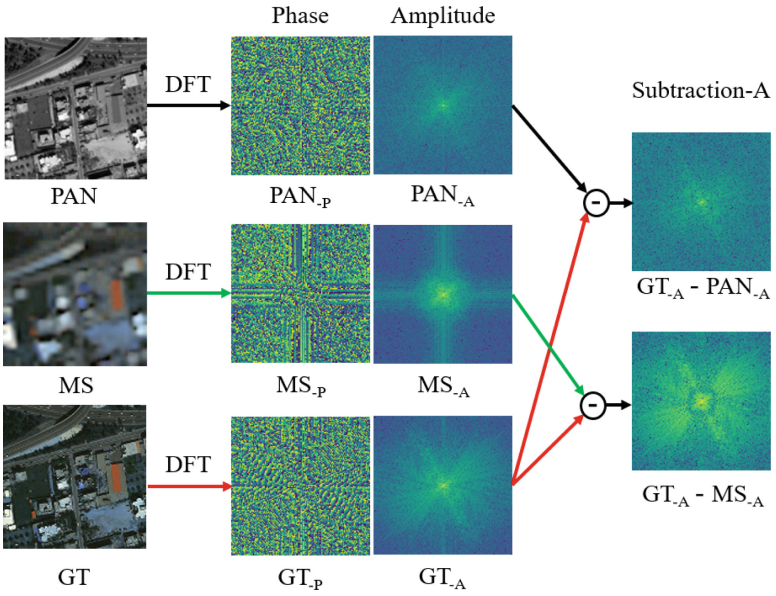

Fig. 1. The frequency domain analysis of discrete Fourier transform (DFT) of PAN image, MS image and the corresponding ground truth (GT) where phase and amplitude are abbreviated as P and A respectively. The middle two columns represent the phase and amplitude components in Fourier space while the last column shows the absolute value of the amplitude subtraction among the connected pairs.

super-resolution problem by learning the non-linear mapping between low- and high-resolution MS images. Both high-spectral and high-spatial images are desirable in the field of remote sensing for a variety of applications such as military systems, environmental monitoring, and mapping services. However, due to the limits of hardware devices, such images can hardly be obtained. To this end, pan-sharpening technique has drawn great attention from both image processing and remote sensing communities.

Inspired by the success of deep neural networks (DNN) over image processing, explosive DNN-based pan-sharpening methods [1,12,33,34] have been developed. The pioneering one refers to PNN [36], which only adapts three-layer convolution operation to account for the MS pan-sharpening learning motivated by the representative super-resolution model SRCNN [9]. Since then, more complicated and deeper architectures have been designed to improve the mapping capability of pan-sharpening. Despite the remarkable progress, existing pan-sharpening methods still suffer from the common limitation. All of them only focus on learning the pan-sharpening function in the spatial domain and rarely explore the potential solutions of pan-sharpening in the frequency domain, which deserves more attention in pan-sharpening. However, pan-sharpening is essentially a PAN-guided MS image super-resolution problem and super-resolution task is tightly coupled to the frequency domain due to the removal of high frequency information during

the down-sampling process, illustrated in [14]. Given this observation, we devote considerable effort to pan-sharpening in frequency domain.

Our Motivation. As shown in Fig. 1, we conduct the comprehensive frequency analysis of pan-sharpening by revisiting the properties of phase and amplitude components via discrete Fourier transformation and deepening into their difference of amplitude components. Targeting at pan-sharpening, there are two observations in frequency domain: 1) The phase of PAN is more similar with the phase of GT than that of MS, which is consistent with the spatial observation that PAN has more detailed textures than MS images. As well recognized, the phase component of the Fourier transformation characterizes the structure information. It is therefore natural to leverage the phase of PAN to support that of MS for approximating the phase of GT; 2) In the last column, it is noted that the amplitude difference of PAN and GT lies in low frequency while the amplitude difference of MS and GT lies in both low and high frequency. We can deduce that compared with GT, the missing frequency information of MS can be borrowed from that of PAN. In short, the frequency domain provides the more powerful tool to analyze and observe the degradation of pan-sharpening and it motivates us to explore the potential solution of pan-sharpening in both space and frequency domains. Besides, motivated by spectral convolution theorem [10], we note that learning in frequency information allows the image-wide receptive field that models the global contextual information. Therefore, leveraging the global frequency information complements the local information by pixel values in spatial domain with boosting the information representation and model capability.

Based on above analysis, we introduce a novel perspective for pan-sharpening in this paper. Specifically, we first attempt to address pan-sharpening in both spatial-frequency domain and propose a Spatial-Frequency Information Integration Network, dubbed as SFIIN. To implement SFIIN, the fundamental building module called SFIB is devised, which consists of three key components: spatial-domain information branch, frequency-domain information branch and dual domain information interaction. The spatial branch employs the ordinary convolution to exploit the local information of two modalities of PAN and MS images in spatial domain while the frequency branch is responsible for extracting and transforming the global frequency information via deep Fourier transformation over amplitude and phase components in frequency domain. Motivated by spectral convolution theorem, we argue that the frequency information branch allows the image-wide receptive field that models the global contextual information, thus boosting the model capability. Followed by, the dual domain information interaction is performed to facilitate the information flow and learn the complementary representation in spatial and frequency domain. We conduct extensive experiments to analyze the effectiveness of the proposed network and demonstrate the favorable performance against state-of-the-art methods qualitatively and quantitatively while generalizing well to real-world scenes.

In summary, the contributions of this work are as follows:

– To the best of our knowledge, this is the first attempt to explore the potential solution of pan-sharpening in both spatial and frequency domain. In this paper, a Spatial-Frequency Information Integration Network is proposed and substantially improves the Pan-sharping performance.
– We devise a core building module tailored with pan-sharpening, consisting of three key components: spatial domain information branch, frequency domain information one and dual domain information interaction. It enables the local spatial information and global frequency information to learn the complementary representation, thus boosting the model capability.
– Extensive experiments over different satellite datasets demonstrate that our proposed method performs the best qualitative and quantitative while generalizing well to real-world full-resolution scenes.

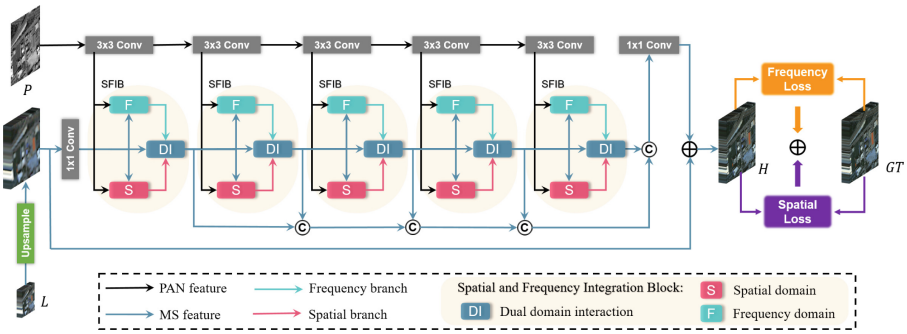

Fig. 2. The framework of our proposed pan-sharpening network. The network is equipped with the core building module SFIB which is tailored with pan-sharpening and consists of three key components: spatial domain information S, frequency domain information F and dual domain interaction DI. Therefore, it is capable of effectively exploring the space and frequency domain information of MS and PAN images. In addition, we design the new loss function in both spatial and frequency domain to better optimize the proposed network.

2 Related Work

2.1 Traditional Pan-Sharpening Methods

Component Substitution (CS), Multi-resolution Analysis (MRA), and Variational Optimization (VO) are the three categories under which traditional pan-sharpening techniques are categorized [38,39]. The principal component analysis (PCA) methods [28,37], Brovey transforms [16], and the Gram-Schmidt (GS) orthogonalization approach [30] are the most used CS techniques. Researchers have also suggested various enhancements to the approaches mentioned above, such as the nonlinear IHS (NIHS) method [15], which reduces the spectrum distortion of IHS, and the GSA method, which has adaptive capabilities for

the GS method [2]. These CS algorithms are quite quick to calculate, however the artifacts that are produced in the photos are quite common. When sharpening MS images, MRA approaches produce less spectral distortion than CS methods. Decimated wavelet transform (DWT), high-pass filter fusion (HPF), indusion method [27], Laplacian pyramid (LP) [41], and atrous wavelet transform (ATWT) are examples of common MRA techniques. The first variational method, P+XS pan-sharpening approach [4], makes the assumption that PAN images are created by linearly combining different bands of HRMS images, while upsampled low resolution multi-spectral (LRMS) images are created from fuzzy HRMS images. The pan-sharpening task is then subjected to a variety of constraints, including the dynamic gradient sparsity property (SIRF) [8], the local gradient constraint (LGC) [11], the group low-rank constraint for texture similarity (ADMM) [39], and others. These numerous priors and restrictions, which call for manual parameter setup, can only imperfectly reflect the images' restricted structural relationships, which can also lead to degradation.

2.2 CNN-Based Pan-Sharpening Methods

Convolutional neural networks (CNN), which have strong nonlinear fitting and feature extraction capabilities, have rapidly developed in computer vision and have been frequently used in hyperspectral images [7,13,17,22–26,43] and remote sensing images [48,49,53–57]. Recently, a number of CNN-based techniques have been proposed to support the fusion quality of pan-sharpening [35,46,52]. For instance, Masi *et al.* [36] are the first to apply CNN to address the problem of pan-sharpening. Even though the structure is straightforward, the results are far superior to those of conventional techniques. Then, Yang *et al.* [50] then used resblock in [19] to create a deeper convolutional network. In the meantime, Yuan *et al.* [51] added the multi-scale module into the fundamental CNN design. Later, Cai *et al.* [5] and Wu *et al.* [44] had a similar idea: continuously introducing images of various scales into the backbone network. The two methods differ in that the former uses PAN images and the latter MS images. A few model-driven CNN models with obvious physical meaning have recently surfaced. The fundamental concept is to create optimization issues for computer vision tasks using previous knowledge, and then to develop the optimization algorithms into deep neural networks. For instance, to build the unfolding structure for pan-sharpening, Xu *et al.* [47] constructed two distinct priors of PAN and MS. The model-driven approaches are comprehensible and have obvious physical significance. CNN was updated with an alternative optimization approach by Cao *et al.* [6]. Variational optimization and deep residual CNN were integrated by Tian *et al.* [40] and Wu *et al.* [45].

3 Method

In this section, we will first revisit the properties of Fourier Transformation of images and then present an overview of the proposed pan-sharpening network, illustrated in Fig. 2 and Fig. 3. We further provide details of the fundamental building block of our method, containing three key elements: (a) frequency

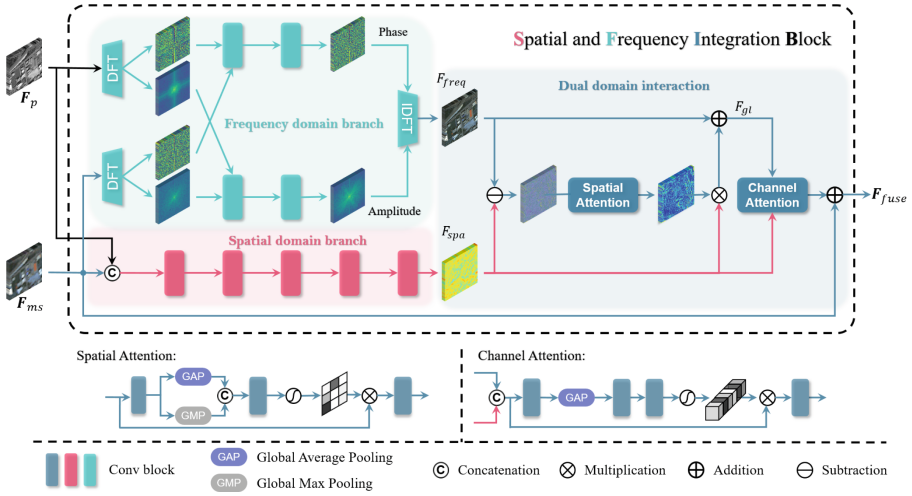

Fig. 3. The detailed flowchart of the proposed core building module SFIB, consisting of three components: frequency domain branch, spatial domain branch and dual domain interaction.

domain information branch for extracting the global frequency information representations via DFT, (b) spatial domain information branch to explore the local information via ordinary convolution, (c) dual domain information interaction to facilitate the information flow and learn the complementary representation. Finally, we deepen into the newly-designed loss functions.

3.1 Fourier Transformation of Images

As recognized, the Fourier transform is widely used to analyze the frequency content of images. For the images of multiple color channels, the Fourier transform is calculated and performed for each channel separately. For simplicity, we eliminate the notation of channels in formulas. Given a image $x \in R^{H \times W \times C}$, the Fourier transform $\mathcal{F}$ converts it to Fourier space as the complex component $\mathcal{F}(x)$, which is expressed as:

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

$\mathcal{F}^{-1}(x)$ defines the inverse Fourier transform accordingly. Both the Fourier transform and its inverse procedure can be efficiently implemented with the FFT algorithm in [10]. The amplitude component $\mathcal{A}(x)(u,v)$ and phase component $\mathcal{P}(x)(u,v)$ are expressed as:

$$\begin{aligned} \mathcal{A}(x)(u,v)) &= \sqrt{R^2(x)(u,v)) + I^2(x)(u,v))}, \\ \mathcal{P}(x)(u,v)) &= \arctan[\frac{I(x)(u,v))}{R(x)(u,v))}], \end{aligned} \tag{2}$$

where $R(x)$ and $I(x)$ represent the real and imaginary part of $\mathcal{F}(x)$ respectively. In our method, the Fourier transformation and inverse procedure is computed independently on each channel of feature maps.

Targeting at pan-sharpening, we employ Fourier transformation to conduct the detailed frequency analysis of MS, PAN and GT images by revisiting the properties of phase and amplitude components, as shown in Fig. 1. There are two observations in frequency domain: 1) The phase of PAN has more similar appearance with that of GT than that of MS. This claim also keeps consistent with the spatial observation that PAN has the more detailed textures than MS images. With the well-known properties of the Fourier transformation, the phase component characterizes the structure information. 2) The amplitude difference of PAN and GT lies in low frequency while the amplitude difference of MS and GT lies in both low and high frequency. We can deduce that compared with GT, the missing frequency information of MS can be borrowed from that of PAN to restore that of GT. It motivates us to explore the potential solution of pan-sharpening in both spatial and frequency domains.

3.2 Framework

Structure Flow. Based on above analysis, we introduce a novel spatial-frequency information integration-based perspective for pan-sharpening, detailed in Fig. 2. Given PAN image $P \in R^{H\times W\times 1}$ and MS image $L \in R^{H/r\times W/r\times C}$, the network first applies the convolution layer to project the r-times L by Bibubic upsampling into shallow feature representations while P is progressively fed into multiple cascaded convolution to extract the series of informative features. Next, the obtained modality-aware feature maps of MS and PAN are jointly pass through N number of the core building module SFIB with space-frequency information extraction and integration, yielding the effective feature representation. Next, we apply a convolution layer to transform the collected features from all N SFIBs back to image space and then combine with the input L as the output image.

Supervision Flow. Orthogonal to structure design, we also introduce a newly-designed loss functions to enable the network for better optimization, thus reconstructing the more pleasing results in both spatial and frequency domains. As shown in Fig. 2, it consists of two parts: spatial domain loss and frequency domain loss. In contrast to existing methods that usually adopt pixel losses with local guidance in the spatial domain, we additionally propose the frequency domain supervision loss via Fourier transformation that is calculated on the global frequency components. Motivated by spectral convolution theorem, direct emphasis on the frequency content is capable of better reconstructing the global information, thus improving the pan-sharpening performance.

3.3 The Core Building Block

As shown in Fig. 3, the fundamental building block of our method contains three key elements: (a) frequency domain information branch for extracting the global

frequency information representation via deep Fourier transform, (b) spatial domain information branch to explore the local information via ordinary convolution, (c) dual domain information interaction to facilitate the information flow and learn the complementary representation.

Frequency Domain Information Branch. In the frequency branch, we first adopt Fourier transform to convert the modality-aware features of MS and PAN images and generate the amplitude and phase components. Suppose that the features of MS and PAN images denote as F_p and F_{ms}, the corresponding Fourier transform is expressed as

$$\mathcal{A}(F_p), \mathcal{P}(F_p) = \mathcal{F}(F_p), \tag{3}$$
$$\mathcal{A}(F_{ms}), \mathcal{P}(F_{ms}) = \mathcal{F}(F_{ms}), \tag{4}$$

where $\mathcal{A}(.)$ and $\mathcal{P}(.)$ indicate the amplitude and phase respectively. Then we uses two groups of independent operation $OA(\cdot)$ and $OP(\cdot)$, consisting of 1×1 convolution and $Relu$ activation function to integrate the corresponding amplitude and phase components for providing the enhanced global frequency representations

$$\mathcal{A}(F) = \mathcal{OA}(Cat[\mathcal{A}(F_p), \mathcal{A}(F_{ms})]), \tag{5}$$
$$\mathcal{P}(F) = \mathcal{OP}(Cat[\mathcal{P}(F_p), \mathcal{P}(F_{ms})]), \tag{6}$$

where Cat indicates the concatenation operation by channel dimension. Next, we apply the inverse DFT to transform the fused amplitude and phase components of $\mathcal{A}(F)$ and $\mathcal{P}(F)$ back to spatial domain

$$F_{fre} = F^{-1}(\mathcal{A}(F), \mathcal{P}(F)). \tag{7}$$

According to spectral convolution theorem in Fourier theory, processing information of Fourier space is capable of capturing the global frequency representation in frequency domain. In short, the frequency branch generates the global information representation F_{fre}.

Spatial Domain Information Branch. In contrast, the spatial branch first adopts a residual block [20] with 3×3 convolution layers to integrate information of MS and PAN features and generate the space representation F_{spa} in spatial domain. It is well recognized that the ordinary convolution focuses on learning local representations in spatial domain. In short, the spatial branch provides the local information representation F_{spa}. Based on the above spatial and frequency domain branches, we note that the generated information representation from both branches is complementary. Therefore, interacting and integrating them is beneficial to compensate each other and provide more informative representation.

Dual Domain Information Interaction. The schematic of Dual domain information interaction mainly consists of information compensation and information integration part. **(a) information compensation:** Owing the complementary property of the frequency and spatial representation F_{fre} and F_{spa}, this motivates us to extract the distinguished components of local spatial information

F_{spa} to compensate the global frequency information F_{fre}. Therefore, we first calculate the absolute difference among them and then employ the spatial attention mechanism SA to exploit the inter-spatial dependencies. It outputs the spatial attention map and multiplies it over F_{spa} to select the more informative content, impose it over global frequency representation F_{fre} to the enriched representation F_{gl}

$$F_{gl} = F_{fre} + SA(F_{fre} - F_{spa}) \times F_{spa}. \tag{8}$$

(b) information integration: When obtaining the enhanced global frequency feature F_{gl}, we combine it with the local feature F_{spa} and then perform the channel attention to exploit the inter-channel relationship, thus facilitating the complementary learning and providing the more informative feature representation F_{fuse}. Finally, the residual learning mechanism is adopted by adding the input MS feature F_{ms} to the fused one

$$F_{fuse} = CA([F_{gl}, F_{spa}]) + F_{ms}. \tag{9}$$

Equipped with the core building block, our proposed network is capable of modeling and integrating the global and local information representation by exploring the potential of spatial and frequency dual domains.

3.4 Loss Function

Let H_L and GT denote the network output and the corresponding ground truth respectively. To generate the pleasing pan-sharpening results, we propose a joint spatial-frequency domain loss for supervising the network training. In spatial domain, we adopt the $L1$ loss

$$\mathcal{L}_{spa} = \|H_L - GT\|_1 . \tag{10}$$

In frequency domain, we first employ the DFT to convert H_L and GT into Fourier space where the amplitude and phase components are calculated. Then, the $L1$-norms of amplitude difference and phase difference between H_L and GT are summed to produce the total frequency loss

$$\mathcal{L}_{fre} = \|\mathcal{A}(H_L) - \mathcal{A}(GT)\|_1 + \|\mathcal{P}(H_L) - \mathcal{P}(GT)\|_1 . \tag{11}$$

Finally, the overall loss function is formulated as follows

$$\mathcal{L} = \mathcal{L}_{spa} + \lambda \mathcal{L}_{fre}, \tag{12}$$

where λ is weight factor and set to 0.1 empirically.

4 Experiments

4.1 Baseline Methods

To demonstrate the efficacy of our proposed method, we compare its performance to that of several representative pansharpening methods: 1) five cutting-edge deep-learning methods, such as PNN [36], PANNET [50], MSDCNN [51], SRPPNN [5], and GPPNN [47]; 2) five promising traditional methods, including SFIM [32], Brovey [16], GS [29], IHS [18], and GFPCA [31].

4.2 Implementation Details

On a personal computer with a single NVIDIA GeForce GTX 2080Ti GPU, the PyTorch framework is used to construct each of our networks. During the training phase, the Adam optimizer will optimize them using four-epoch batches over a total of 2000 iterations. 8×10^{-4} yields the initial value for the learning rate. After 200 epochs, the pace of learning will begin to decrease by a factor of two. Due to the lack of pan-sharpened ground-truth images, we generate the training set by employing the Wald protocol tool [42], as was done in previous studies. Specifically, given the MS image $H \in R^{M \times N \times C}$ and the PAN image $P \in R^{rM \times rN \times b}$, both of them are downsampled with ratio r, and the resulting images are denoted by $L \in R^{M/r \times N/r \times C}$ and $p \in R^{M \times N \times b}$ respectively. In the training set, L and p are regarded as the inputs, whereas H is the ground truth.

We have chosen to assess the worldview II, worldview III, and GaoFen2 satellite image datasets for this study. The PAN images for each database are cropped into patches measuring 128×128 pixels, while the corresponding MS patches are 32×32 pixels. Image quality assessment (IQA) metrics such as the relative dimensionless global error in synthesis (ERGAS) [3], the peak signal-to-noise ratio (PSNR), the structural similarity (SSIM) and SAM [21], are used for performance evaluation. These measures are frequently employed in the pan-sharpening field.

To compare the generalization of models, we create an additional real-world full-resolution dataset of 200 samples over the newly-selected GaoFen2 satellite for evaluation. Specifically, the additional dataset is generated when the full-resolution setting is used to generate the PAN and MS images as aforementioned manner without performing the down-sampling with PAN images of 32×32 and MS images of 128×128 resolutions. Due to the lack of ground-truth MS images, we measure the model's performance using the three commonly-used IQA metrics: the spectral distortion index D_λ, the spatial distortion index D_S, and the quality without reference (QNR).

4.3 Comparison with State-of-the-Art Methods

Evaluation on Reduced-Resolution Scene. Table 1 displays the assessment metrics for three datasets, with the red-highlighted values denoting the best results. On three satellite datasets, it is evidently discovered that our technique outperforms other comparison algorithms in all assessment measures. Specifically, on the WorldView-II, GaoFen2, and WorldView-III datasets, our technique improves PSNR by 0.27 dB, 0.28 dB, and 0.16 dB over the second-best findings, respectively. Similar gains may be observed in the other parameters in addition to PSNR. We greatly outperform the most recent deep learning-based algorithms, demonstrating the viability of the suggested approach.

We also compare the visual results to demonstrate the efficacy of our methodology using typical samples from the WorldView-II and GaoFen2 datasets in Fig. 4 and Fig. 5, respectively. The MSE residual between the pan-sharpened findings and the actual data are shown by the images in the last row. Our model

Table 1. Quantitative comparison. The best values are highlighted by the red bold. The up or down arrow indicates higher or lower metric corresponding to better images.

Method	Worldview II				GaoFen2				Worldview III			
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	SAM↓	EGAS↓	PSNR↑	SSIM↑	SAM↓	EGAS↓
SFIM	34.1297	0.8975	0.0439	2.3449	36.9060	0.8882	0.0318	1.7398	21.8212	0.5457	0.1208	8.9730
Brovey	35.8646	0.9216	0.0403	1.8238	37.7974	0.9026	0.0218	1.372	22.5060	0.5466	0.1159	8.2331
GS	35.6376	0.9176	0.0423	1.8774	37.2260	0.9034	0.0309	1.6736	22.5608	0.5470	0.1217	8.2433
IHS	35.2962	0.9027	0.0461	2.0278	38.1754	0.9100	0.0243	1.5336	22.5579	0.5354	0.1266	8.3616
GFPCA	34.5581	0.9038	0.0488	2.1411	37.9443	0.9204	0.0314	1.5604	22.3344	0.4826	0.1294	8.3964
PNN	40.7550	0.9624	0.0259	1.0646	43.1208	0.9704	0.0172	0.8528	29.9418	0.9121	0.0824	3.3206
PANNET	40.8176	0.9626	0.0257	1.0557	43.0659	0.9685	0.0178	0.8577	29.6840	0.9072	0.0851	3.4263
MSDCNN	41.3355	0.9664	0.0242	0.9940	45.6874	0.9827	0.0135	0.6389	30.3038	0.9184	0.0782	3.1884
SRPPNN	41.4538	0.9679	0.0233	0.9899	47.1998	0.9877	0.0106	0.5586	30.4346	0.9202	0.0770	3.1553
GPPNN	41.1622	0.9684	0.0244	1.0315	44.2145	0.9815	0.0137	0.7361	30.1785	0.9175	0.0776	3.2593
Ours	**41.7244**	**0.9725**	**0.0220**	**0.9506**	**47.4712**	**0.9901**	**0.0102**	**0.5462**	**30.5971**	**0.9236**	**0.0741**	**3.0798**

Table 2. Evaluation on the real-world full-resolution scenes from GaoFen2 dataset. The best results are highlighted in **bold**.

Metrics	SFIM	GS	Brovey	IHS	GFPCA	PNN	PANNET	MSDCNN	SRPPNN	GPPNN	**Ours**
$D_\lambda \downarrow$	0.0822	0.0696	0.1378	0.0770	0.0914	0.0746	0.0737	0.0734	0.0767	0.0782	**0.0681**
$D_s \downarrow$	**0.1087**	0.2456	0.2605	0.2985	0.1635	0.1164	0.1224	0.1151	0.1162	0.1253	0.1119
QNR↑	0.8214	0.7025	0.6390	0.6485	0.7615	0.8191	0.8143	0.8251	0.8173	0.8073	**0.8466**

exhibits very slight spectral and spatial aberrations as compared to other competing techniques. It is obvious to draw from the examination of MSE maps. Regarding the MSE residues, it has been observed that our suggested technique is more accurate than previous comparison methods. We can thus state with confidence that our approach outperforms other competing pan-sharpening methods.

Evaluation on Full-Resolution Scene We apply a pre-trained model created using GaoFen2 data to some unused full-resolution GaoFen2 satellite datasets in order to evaluate the performance of our network in the full resolution situation and the generalizability of the model. Table 2 provides an overview of the experimental findings for all approaches. Table 2 shows that our devised technique performs almost at the top of all the indices, which suggests that it has superior generalization capacity than other conventional and deep learning-based methods.

4.4 Parameter Numbers vs Model Performance

In order to conduct a more in-depth analysis of the methods, we will analyze the complexity of the suggested technique by looking at the number of floating-point operations (FLOPs) and the number of parameters (by 10 million) in Table 3. Compared to other deep learning-based methods, our network achieves the highest performance with the fewer parameters and storage space.

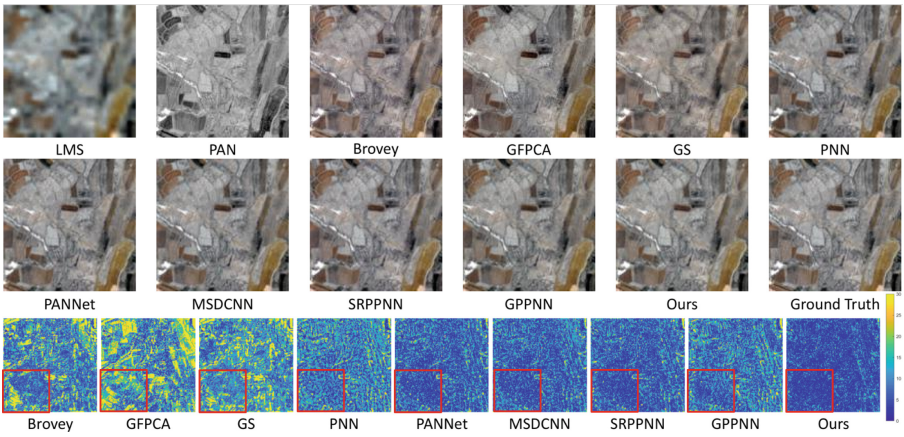

Fig. 4. The visual comparisons between other pan-sharpening methods and our method on WorldView-II satellite.

Table 3. Comparisons of FLOPs (G) and parameters number (M). 'Param' denotes parameters number.

	PNN	PANNET	MSDCNN	SRPPNN	GPPNN	Ours
Param	0.0689	0.0688	0.2390	1.7114	0.1198	0.0871
FLOPs	1.1289	1.1275	3.9158	21.1059	1.3967	1.2558

4.5 Ablation Experiments

We have performed thorough ablation investigations using the WorldView-II satellite dataset of the Pan-sharpening task to examine the contribution of the created modules in our suggested network. The two fundamental designs are, more precisely, the frequency information branch in the core building module and the frequency loss in the optimization function. Additionally, research is done about the quantity of fundamental building modules used in the network. The commonly used IQA measures, such as ERGAS [3], PSNR, SSIM, SCC, Q index, SAM [21], D_λ, D_S, and QNR, are utilized to evaluate all of the experimental data.

The Number of the Core Building Modules. We experiment the suggested network with various numbers of the core building module to examine the influence of the number of the core building modules. Table 4 gives the equivalent quantitative figures K comparison from 1 to 5. It is evident from the findings in Table 4 that as the number of IQAs rises, the model performance has significantly improved at the expense of computation for nearly all of them. Performance and computational complexity were balanced in this work by using the default option of $K = 5$.

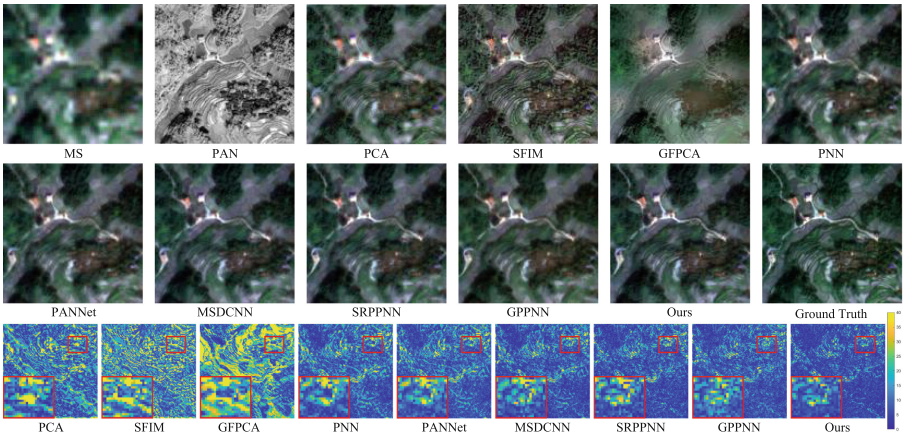

Fig. 5. The visual comparisons between other pan-sharpening methods and our method on GaoFen2 satellite.

Table 4. Average performance comparison on the WorldView-II datasets as the number of SFIB increases. The best performance is shown in red **bold**.

Number (K)	PSNR↑	SSIM ↑	SAM ↓	ERGAS ↓	SCC ↑	Q ↑	D_λ ↓	D_S ↓	QNR ↑
1	41.1343	0.9644	0.0257	1.0218	0.9651	0.7548	0.0639	0.1188	0.8249
2	41.2566	0.9650	0.0249	1.0126	0.9661	0.7554	0.0635	0.1176	0.8264
3	41.4781	0.9677	0.0242	0.9841	0.9696	0.7681	0.0627	0.1170	0.8276
4	41.6287	0.9690	0.0226	0.9527	0.9711	0.7699	0.0618	0.1168	0.8286
5	**41.7244**	**0.9725**	**0.0220**	**0.9506**	**0.9720**	**0.7751**	**0.0613**	**0.1167**	**0.8290**

The Frequency Information Branch. To assess the influence of the frequency information, we merely substitute the spatial information branch in the core building module with the frequency information branch in the first experiment of Table 5. The results in Table 5 show that eliminating it will impair the performance of our network. The global frequency information modeling will be broken if it is deleted, which would worsen the pan-sharpening results.

The Frequency Loss. The newly developed frequency loss intends to clearly highlight the optimization of global frequency information. We erase it in the second trial of Table 5 to test its efficacy. The findings in Table 5 show that deleting it would significantly worsen all metrics, demonstrating its importance to our network.

4.6 Visualization of Feature Maps in Dual Domains

To verify the effect of the designed dual domain information integration mechanism, we deepen into the feature maps of F_{ms}, F_{fre}, F_{spa}, $F_{fre} - F_{spa}$, F_{gl}, F_{fuse}. As illustrated in Sect. 3.3, the frequency feature F_{fre} and the spatial feature F_{spa}

Table 5. Ablation studies comparison on the WorldView-II datasets. The best performance is shown in **bold**.

Config	FSB	FSF	PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓	SCC ↑	Q ↑	D_λ ↓	D_S ↓	QNR ↑
(I)	✗	✓	41.2664	0.9651	0.0253	1.0117	0.9658	0.7553	0.0633	0.1181	0.8260
(II)	✓	✗	41.6766	0.9698	0.0227	0.9524	0.9747	0.7746	0.0621	0.1174	0.8267
(III)	✓	✓	**41.7244**	**0.9725**	**0.0220**	**0.9506**	**0.9720**	**0.7751**	**0.0613**	**0.1167**	**0.8290**

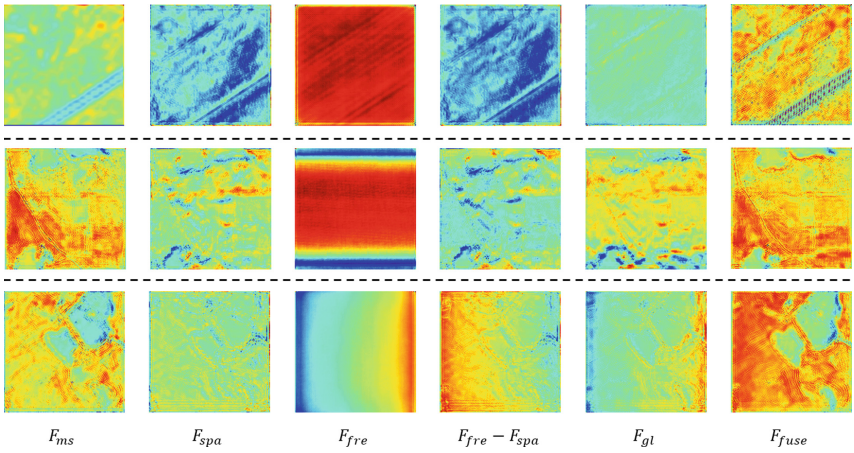

Fig. 6. The Visualization of feature maps in dual domains.

are complementary. In Fig. 6, it is clearly seen that the frequency feature F_{fre} characterizes the global information while the spatial feature F_{spa} focuses on the local content. With integrating them, the response of F_{fuse} is more informative. It demonstrates the powerful capability of the core module.

5 Conclusion

In this paper, we propose a spatial-frequency information integration network for pan-sharpening. To implement the network, we devise a core building module tailored with pan-sharpening to learn the complementary information representation of spatial and frequency domains, thus boosting the model capability. To the best of our knowledge, this is the first attempt to explore the potential solution of pan-sharpening in both spatial-frequency domain. Extensive experiments demonstrate that the proposed network performs favorably against state-of-the-art methods while generalizing well to real-world full-resolution scenes.

Acknowledgements. This work was supported by the Anhui Provincial Natural Science Foundation under Grant 2108085UD12. We acknowledge the support of GPU cluster built by MCC Lab of Information Science and Technology Institution, USTC.

References

1. Addesso, P., Vivone, G., Restaino, R., Chanussot, J.: A data-driven model-based regression applied to panchromatic sharpening. IEEE Trans. Image Process. **29**, 7779–7794 (2020)
2. Aiazzi, B., Baronti, S., Selva, M.: Improving component substitution pansharpening through multivariate regression of ms + pan data. IEEE Trans. Geosci. Remote Sens. **45**(10), 3230–3239 (2007)
3. Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., Bruce, L.M.: Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data fusion contest. IEEE Trans. Geosci. Remote Sens. **45**(10), 3012–3021 (2007)
4. Ballester, C., Caselles, V., Igual, L., Verdera, J., Rougé, B.: A variational model for p+ xs image fusion. Int. J. Comput. Vision **69**(1), 43–58 (2006)
5. Cai, J., Huang, B.: Super-resolution-guided progressive pansharpening based on a deep convolutional neural network. IEEE Trans. Geosci. Remote Sens. **59**(6), 5206–5220 (2021)
6. Cao, X., Fu, X., Hong, D., Xu, Z., Meng, D.: PanCSC-net: a model-driven deep unfolding method for pansharpening. IEEE Trans. Geosci. Remote Sens. 1–13 (2021)
7. Cao, X., Zhou, F., Xu, L., Meng, D., Xu, Z., Paisley, J.: Hyperspectral image classification with Markov random fields and a convolutional neural network. IEEE Trans. Image Process. **27**(5), 2354–2367 (2018)
8. Chen, C., Li, Y., Liu, W., Huang, J.: SIRF: simultaneous satellite image registration and fusion in a unified framework. IEEE Trans. Image Process. **24**(11), 4213–4224 (2015)
9. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016)
10. Frigo, M., Johnson, S.G.: FFTW: an adaptive software architecture for the FFT. In: 1988 International Conference on Acoustics, Speech, and Signal Processing. ICASSP-88, vol. 3 (1998)
11. Fu, X., Lin, Z., Huang, Y., Ding, X.: A variational pan-sharpening with local gradient constraints. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10265–10274 (2019)
12. Fu, X., Wang, W., Huang, Y., Ding, X., Paisley, J.: Deep multiscale detail networks for multiband spectral image sharpening. IEEE Trans. Neural Netw. Learn. Syst. **32**(5), 2090–2104 (2021)
13. Fu, Y., Liang, Z., You, S.: Bidirectional 3D quasi-recurrent neural network for hyperspectral image super-resolution. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. **14**, 2674–2688 (2021)
14. Fuoli, D., Gool, L.V., Timofte, R.: Fourier space losses for efficient perceptual image super-resolution (2021)
15. Ghahremani, M., Ghassemian, H.: Nonlinear IHS: a promising method for pansharpening. IEEE Geosci. Remote Sens. Lett. **13**(11), 1606–1610 (2016)
16. Gillespie, A.R., Kahle, A.B., Walker, R.E.: Color enhancement of highly correlated images. ii. channel ratio and "chromaticity" transformation techniques - sciencedirect. Remote Sens. Environ. **22**(3), 343–365 (1987)
17. Haut, J.M., Paoletti, M.E., Plaza, J., Li, J., Plaza, A.: Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach. IEEE Trans. Geosci. Remote Sens. **56**(11), 6440–6461 (2018)

18. Haydn, R., Dalke, G.W., Henkel, J., Bare, J.E.: Application of the IHS color transform to the processing of multisensor data and image enhancement. Natl. Acad. Sci. USA **79**(13), 571–577 (1982)
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
21. J. R. H. Yuhas, A.F.G., Boardman, J.M.: Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In: Proceedings of Summaries Annual JPL Airborne Geoscience Workshop, pp. 147–149 (1992)
22. Jiang, J., Ma, J., Liu, X.: Multilayer spectral-spatial graphs for label noisy robust hyperspectral image classification. IEEE Trans. Neural Netw. Learn. Syst. 1–14 (2020)
23. Jiang, J., Ma, J., Wang, Z., Chen, C., Liu, X.: Hyperspectral image classification in the presence of noisy labels. IEEE Trans. Geosci. Remote Sens. **57**(2), 851–865 (2019)
24. Jiang, J., Sun, H., Liu, X., Ma, J.: Learning spatial-spectral prior for super-resolution of hyperspectral imagery. IEEE Trans. Comput. Imaging **6**, 1082–1096 (2020)
25. Jiang, K., Wang, Z., Yi, P., Jiang, J.: A progressively enhanced network for video satellite imagery superresolution. IEEE Sig. Process. Lett. **25**(11), 1630–1634 (2018)
26. Jiang, K., et al.: GAN-based multi-level mapping network for satellite imagery super-resolution. In: 2019 IEEE International Conference on Multimedia and Expo (ICME), pp. 526–531 (2019)
27. Khan, M.M., Chanussot, J., Condat, L., Montanvert, A.: Indusion: fusion of multispectral and panchromatic images using the induction scaling technique. IEEE Geosci. Remote Sens. Lett. **5**(1), 98–102 (2008)
28. Kwarteng, P., Chavez, A.: Extracting spectral contrast in Landsat thematic mapper image data using selective principal component analysis. Photogramm. Eng. Remote. Sens. **55**(339–348), 1 (1989)
29. Laben, C., Brower, B.: Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. US Patent 6011875A (2000)
30. Laben, C.A., Brower, B.V.: Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. US Patent 6,011,875 (2000)
31. Liao, W., Xin, H., Coillie, F.V., Thoonen, G., Philips, W.: Two-stage fusion of thermal hyperspectral and visible RGB image by PCA and guided filter. In: Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (2017)
32. Liu., J.G.: Smoothing filter-based intensity modulation: a spectral preserve image fusion technique for improving spatial details. Int. J. Remote Sens. **21**(18), 3461–3472 (2000)
33. Lu, X., Zhang, J., Yang, D., Xu, L., Jia, F.: Cascaded convolutional neural network-based hyperspectral image resolution enhancement via an auxiliary panchromatic image. IEEE Trans. Image Process. **30**, 6815–6828 (2021)
34. Ma, J., Xu, H., Jiang, J., Mei, X., Zhang, X.P.: DDCGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Trans. Image Process. **29**, 4980–4995 (2020)

35. Ma, J., Yu, W., Chen, C., Liang, P., Guo, X., Jiang, J.: Pan-GAN: an unsupervised pan-sharpening method for remote sensing image fusion. Inf. Fusion **62**, 110–120 (2020)
36. Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G.: Pansharpening by convolutional neural networks. Remote Sens. **8**(7) (2016)
37. Shah, V.P., Younan, N.H., King, R.L.: An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. IEEE Trans. Geosci. Remote Sens. **46**(5), 1323–1335 (2008)
38. Tian, X., Chen, Y., Yang, C., Gao, X., Ma, J.: A variational pansharpening method based on gradient sparse representation. IEEE Sig. Process. Lett. **27**, 1180–1184 (2020)
39. Tian, X., Chen, Y., Yang, C., Ma, J.: Variational pansharpening by exploiting cartoon-texture similarities. IEEE Trans. Geosci. Remote Sens. 1–16 (2021)
40. Tian, X., Li, K., Wang, Z., Ma, J.: VP-Net: an interpretable deep network for variational pansharpening. IEEE Trans. Geosci. Remote Sens. 1–16 (2021)
41. Vivone, G., et al.: A critical comparison among pansharpening algorithms. IEEE Trans. Geosci. Remote Sens. **53**(5), 2565–2586 (2014)
42. Wald, L., Ranchin, T., Mangolini, M.: Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images. Photogram. Eng. Remote Sens. **63**, 691–699 (1997)
43. Wang, X., Ma, J., Jiang, J.: Hyperspectral image super-resolution via recurrent feedback embedding and spatial-spectral consistency regularization. IEEE Trans. Geosci. Remote Sens. 1–13 (2021)
44. Wu, X., Huang, T.Z., Deng, L.J., Zhang, T.J.: Dynamic cross feature fusion for remote sensing pansharpening. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 14687–14696, October 2021
45. Wu, Z.C., Huang, T.Z., Deng, L.J., Hu, J.F., Vivone, G.: Vo+net: an adaptive approach using variational optimization and deep learning for panchromatic sharpening. IEEE Trans. Geosci. Remote Sens. 1–16 (2021)
46. Xu, H., Ma, J., Shao, Z., Zhang, H., Jiang, J., Guo, X.: SDPNet: a deep network for pan-sharpening with enhanced information representation. IEEE Trans. Geosci. Remote Sens. **59**(5), 4120–4134 (2021)
47. Xu, S., Zhang, J., Zhao, Z., Sun, K., Liu, J., Zhang, C.: Deep gradient projection networks for pan-sharpening. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1366–1375, June 2021
48. Yan, K., Zhou, M., Liu, L., Xie, C., Hong, D.: When pansharpening meets graph convolution network and knowledge distillation. IEEE Trans. Geosci. Remote Sens. **60**, 1–15 (2022). https://doi.org/10.1109/TGRS.2022.3168192
49. Yang, G., Zhou, M., Yan, K., Liu, A., Fu, X., Wang, F.: Memory-augmented deep conditional unfolding network for pan-sharpening. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1788–1797, June 2022
50. Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., Paisley, J.: PanNet: a deep network architecture for pan-sharpening. In: IEEE International Conference on Computer Vision, pp. 5449–5457 (2017)
51. Yuan, Q., Wei, Y., Meng, X., Shen, H., Zhang, L.: A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. **11**(3), 978–989 (2018)
52. Zhang, H., Ma, J.: GTP-PNet: a residual learning network based on gradient transformation prior for pansharpening. ISPRS J. Photogramm. Remote. Sens. **172**, 223–239 (2021)

53. Zhou, M., Fu, X., Huang, J., Zhao, F., Liu, A., Wang, R.: Effective pan-sharpening with transformer and invertible neural network. IEEE Trans. Geosci. Remote Sens. **60**, 1–15 (2022). https://doi.org/10.1109/TGRS.2021.3137967
54. Zhou, M., Huang, J., Fang, Y., Fu, X., Liu, A.: Pan-Sharpening with Customized Transformer and Invertible Neural Network. AAAI Press, Palo Alto (2022)
55. Zhou, M., Xiao, Z., Fu, X., Liu, A., Yang, G., Xiong, Z.: Unfolding Taylor's approximations for image restoration. In: NeurIPS (2021)
56. Zhou, M., Yan, K., Huang, J., Yang, Z., Fu, X., Zhao, F.: Mutual information-driven pan-sharpening. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1798–1808 (June 2022)
57. Zhou, M., Yan, K., Pan, J., Ren, W., Xie, Q., Cao, X.: Memory-augmented deep unfolding network for guided image super-resolution. arXiv abs/2203.04960 (2022)