Adaptively Learning Low-high Frequency Information Integration for Pan-sharpening

Man Zhou
Hefei Institute of Physical Science, Chinese Academy of Sciences
University of Science and Technology of China
manman@mail.ustc.edu.cn

Jie Huang*
University of Science and Technology of China
hj0117@mail.ustc.edu.cn

Chongyi Li†
Nanyang Technological University
lichongyi25@gmail.com

Hu Yu
University of Science and Technology of China
yuhu520@mail.ustc.edu.cn

Keyu Yan
Hefei Institute of Physical Science, Chinese Academy of Sciences
University of Science and Technology of China
keyu@mail.ustc.edu.cn

Naishan Zheng
University of Science and Technology of China
nszheng@mail.ustc.edu.cn

Feng Zhao
University of Science and Technology of China
fzhao956@ustc.edu.cn

ABSTRACT

Pan-sharpening aims to generate high-spatial resolution multi-spectral (MS) image by fusing high-spatial resolution panchromatic (PAN) image and its corresponding low-spatial resolution MS image. Despite the remarkable progress, 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 propose a novel pan-sharpening framework by adaptively learning low-high frequency information integration in the spatial and frequency dual domains. It consists of three key designs: mask prediction sub-network, low-frequency learning sub-network and high-frequency learning sub-network. Specifically, the first is responsible for measuring the modality-aware frequency information difference of PAN and MS images and further predicting the low-high frequency boundary in the form of a two-dimensional mask. In view of the mask, the second adaptively picks out the corresponding low-frequency components of different modalities and then restores the expected low-frequency one by spatial and frequency dual domains information integration while the third combines the above refined low-frequency and the original high-frequency for the latent high-frequency reconstruction. In this way, the low-high frequency information is adaptively learned, thus leading to the pleasing results. Extensive experiments validate the effectiveness of the proposed network and demonstrate the favorable performance against other state-of-the-art methods. The source code will be released at https://github.com/manman1995/pansharpening.

CCS CONCEPTS

• **Computing methodologies** → **Hyperspectral imaging**.

KEYWORDS

low-high frequency, adaptively learning, pan-sharpening

ACM Reference Format:
Man Zhou, Jie Huang, Chongyi Li, Hu Yu, Keyu Yan, Naishan Zheng, and Feng Zhao. 2022. Adaptively Learning Low-high Frequency Information Integration for Pan-sharpening . In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3503161.3547924

1 INTRODUCTION

Pan-sharpening is the technique of spatially super-resolving low-resolution (LR) multi-spectral (MS) images in order to generate the expected high-resolution (HR) MS images based on the matching high-resolution PAN images. Pan-sharpening is, in essence, a PAN-guided super-resolution issue for MS images, which is solved 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, including military systems, environmental monitoring, and mapping services. However, owing to the limits of hardware equipment, such images are difficult to obtain. Both the image processing and remote sensing communities have placed a premium on the pan-sharpening technique for this reason.

*Both authors contributed equally to this research.
†Chongyi Li is the corresponding author.

MM '22, October 10–14, 2022, Lisboa, Portugal

ACM ISBN 978-1-4503-9203-7/22/10…$15.00
https://doi.org/10.1145/3503161.3547924

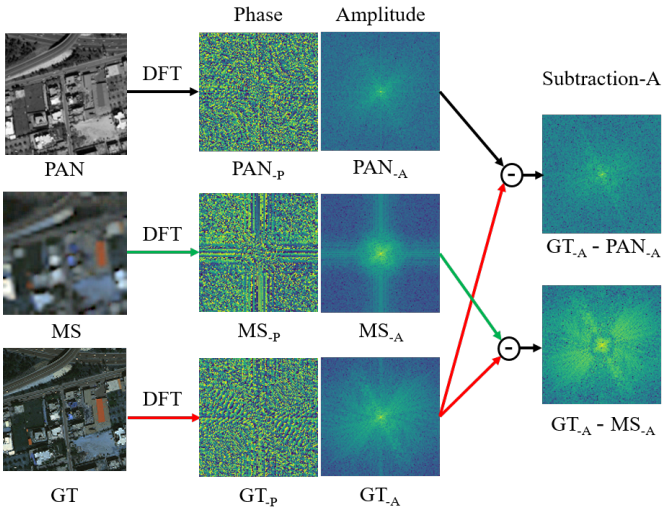

Figure 1: The frequency domain analysis of discrete Fourier transform (DFT) of PAN image, MS image and the corresponding ground truth (GT) where phase and amplitude are abbreviated as P and A respectively. The middle two columns represent the phase and amplitude components in Fourier space while the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

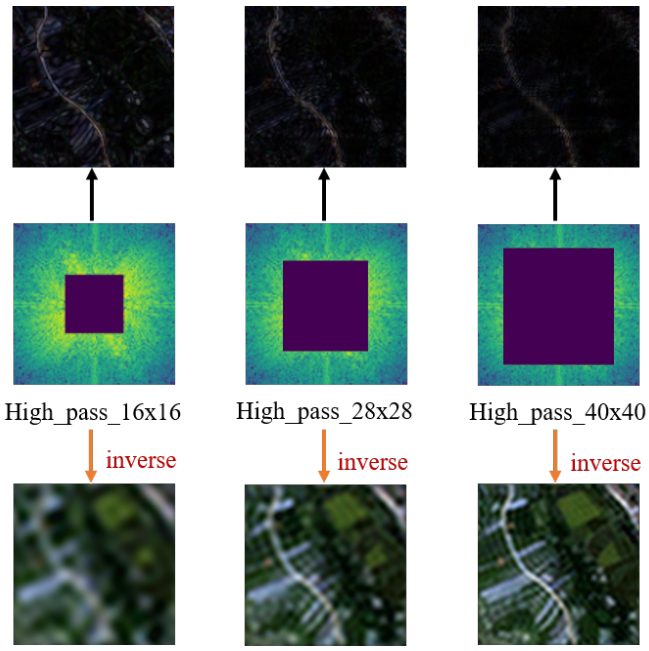

Figure 2: The low-high frequency analysis of discrete Fourier transform (DFT) of ground truth MS image. In the middle row, we mask out the different-size frequency around the center, acting as high-pass filtering. With the mask size increasing, the more high frequencies are removed, as validated in the first row. In the last row, the images represent the corresponding low-frequency parts of the first row by operating the inverse mask as the middle row.

In response to the success of deep neural networks (DNN) in image processing, several DNN-based pan-sharpening algorithms have been created [1, 15, 40, 41]. The pioneering one is PNN [44], which uses just three-layer convolution to account for MS pan-sharpening learning driven by the representative super-resolution model SRCNN [12]. Since then, more intricate and deeper designs have been developed to improve the mapping capability of pan-sharpening. Despite the significant advancements, most existing pan-sharpening methods still suffer from a common constraint. Each of them focuses exclusively on learning the pan-sharpening function in the spatial domain, with little attention paid to the potential pan-sharpening solutions in the frequency domain, which needs greater attention in pan-sharpening. Pan-sharpening, on the other hand, is basically a PAN-guided MS image super-resolution issue, and the super-resolution job is inextricably linked to the frequency domain owing to the elimination of high-frequency information during the down-sampling process, as seen in [17]. Given this discovery, we dedicate substantial effort to the frequency domain pan-sharpening.

Our motivation. As illustrated in Figure 1, we undertake a detailed frequency analysis of pan-sharpening by revisiting the phase and amplitude component features using discrete Fourier transformation and delving into their amplitude component difference. Regarding pan-sharpening, there are two frequency domain observations: 1) PAN's phase is more comparable to GT's than MS's, which is compatible with the spatial finding that PAN images have more detailed textures. As is generally known, the Fourier transformation's phase component characterizes the structure information. Thus, it is logical to use the phase of PAN to support the phase of MS in order to approximate the phase of GT; 2) In the final column, it is noticed that the amplitude difference between PAN and GT is low frequency, whereas the amplitude difference between MS and GT is both low and high frequency. As with GT, we can determine that the missing frequency information for MS may be taken from PAN. Additionally, the degrees of missing low and high-frequency information is distinct. Therefore, they must be addressed differently. Additionally, we performed a low-high frequency analysis of the ground truth MS images using the discrete Fourier transform (DFT) in Figure 2. As can be seen, we use high-pass filtering to hide out the different-sized frequencies surrounding the center. Using the inverse mask as the middle row, as the mask size increases, the number of high-frequency elements such as edges decreases while the corresponding low-frequency elements carrying background information increase. As a result, we may determine that the low-to-high frequency border is imprecise and the method for determining the border must be delicately constructed. In short, the frequency domain provides a more powerful tool for analyzing and observing pan-sharpening deterioration, which pushes us to investigate the potential solution of pan-sharpening in both the space and frequency domains. Additionally, we emphasize, in light of the spectral convolution theorem [13], that learning in frequency information enables the creation of an image-wide receptive field that represents global contextual information. Thus, exploiting global frequency information complements the local information

contained in pixel values in the spatial domain, thereby boosting the information representation and model capability.

We provide a novel viewpoint on pan-sharpening in this study based on the above analysis. We first attempt to address this task in both spatial and frequency domains and then present a novel pan-sharpening framework by adaptively learning low-high frequency information integration in the spatial and frequency dual domains. It comprises three distinct designs: a mask prediction network, a low-frequency learning network, and a high-frequency learning network. Specifically, the first is responsible for assessing the modality-aware frequency information difference between PAN and MS images and then estimating the low-high frequency border using a two-dimensional mask. The second adaptively picks out the corresponding low-frequency components of various modalities in light of the mask. It then reconstructs the expected low-frequency component using spatial and frequency dual domain information integration. At the same time, the third combines the advanced low-frequency and the original high-frequency for latent high-frequency reconstruction. In this technique, the information with a low-high frequency is adaptively learned, resulting in appealing outcomes. We undertake comprehensive tests to evaluate the proposed network's efficacy and show its superior performance qualitatively and statistically against state-of-the-art methods while also demonstrating good generalization to real-world settings.

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

- A novel pan-sharpening framework by adaptively learning low-high frequency information integration in the spatial and frequency dual domains is proposed. To the best of our knowledge, this is the first attempt to explore the potential solution of pan-sharpening in the frequency domain.
- The adaptive low-high frequency boundary prediction mechanism is proposed in the form of a two-dimensional mask while the corresponding frequency loss function is devised to enable the network better learn the expected frequency.
- 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.

2 RELATED WORK

2.1 Traditional pan-sharpening methods

Component Substitution (CS), Multi-resolution Analysis (MRA), and Variational Optimization (VO) are the three classical pan-sharpening approaches [48, 49]. The most often used CS approaches are intensity hue saturation (IHS) fusion [10], principal component analysis (PCA) extraction [35, 47], Brovey transforms [19], and Gram-Schmidt (GS) orthogonalization method [37]. Additionally, researchers have proposed modifications based on the aforementioned methodologies, such as the nonlinear IHS (NIHS) method [18] for minimizing IHS's spectrum distortion and the GSA method [2] for enhancing the GS method's adaptive capabilities. While these convolutional neural networks are very fast to calculate, the produced images are prone to artifacts. When sharpening MS images, MRA procedures produce less spectral distortion than CS methods. The decimated wavelet transform (DWT) [43], the high-pass filter fusion (HPF) [46], indusion method [34], Laplacian pyramid (LP) [50] and atrous wavelet transform (ATWT) [45]. The first variational approach, P+XS pan-sharpening, assumes that the PAN image is made by linearly combining numerous HRMS bands, while the upsampled low resolution multi-spectral (LRMS) image is formed by upsampling the blurring HRMS image. Following that, the pan-sharpening task incorporates a variety of constraints, including the dynamic gradient sparsity property (SIRF) [11], local gradient constraint (LGC) [14], group low-rank constraint for texture similarity (ADMM) [49]. These various priors and constraints, which need human parameter modification, may only reflect the limited structural relations of the images, which may also result in degradation.

2.2 CNN-based pan-sharpening methods

Due to the fast growth of convolutional neural networks (CNNs) in computer vision, CNNs with strong learning capabilities have found widespread usage in hyperspectral images [9, 16, 21, 27–31, 52] and remote sensing images [5–8, 26, 32, 33, 41, 56, 57, 63–67]. Numerous CNN-based techniques for improving the fusion quality of pan-sharpening have been developed lately [42, 54, 61]. Masi et al. [44], for example, are the first to apply CNN to the issue of pan-sharpening. Although the structure is simple, the effect is much better than earlier methods. Then, utilizing resblock in [23], Yang et al. [58] developed a more complicated convolutional network. Meanwhile, Yuan et al. [59] extended the underlying CNN architecture with a multi-scale module. Cai et al. [4] and Wu *et al.* [53] subsequently created a similar concept, in which images of varied sizes are continuously sent into the backbone network. The difference between the two systems is that one utilizes PAN images and the other MS images. Numerous model-driven CNN models with clear physical meaning have emerged in recent years. The core principle is to construct optimization problems for computer vision tasks based on previous knowledge and then unfold the optimization algorithms into deep neural networks. For instance, Xu *et al.* [55] created the unfolding structure for pan-sharpening utilizing two separate priors of PAN and MS.

3 METHODS

In this section, we will first revisit the properties of Fourier Transform of images and then present an overview of the proposed pan-sharpening framework, illustrated in Figure 3. We further provide details of our method, containing three key components: (a) mask prediction network, (b) low-frequency learning network, (c) high-frequency learning network. Finally, we deepen into the newly-designed loss functions.

3.1 Fourier Transform of Images

The Fourier transform, as previously stated, is commonly employed to examine the frequency content of pictures. The Fourier transform is computed and done independently for each color channel in pictures with multiple color channels. To keep things simple, we exclude the channel notation from formulae. Given an image$x \in R^{H\times W\times C}$, the Fourier transform $\mathcal{F}$ transfers it to Fourier space as the complex component $\mathcal{F}(x)$, which is expressed as

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

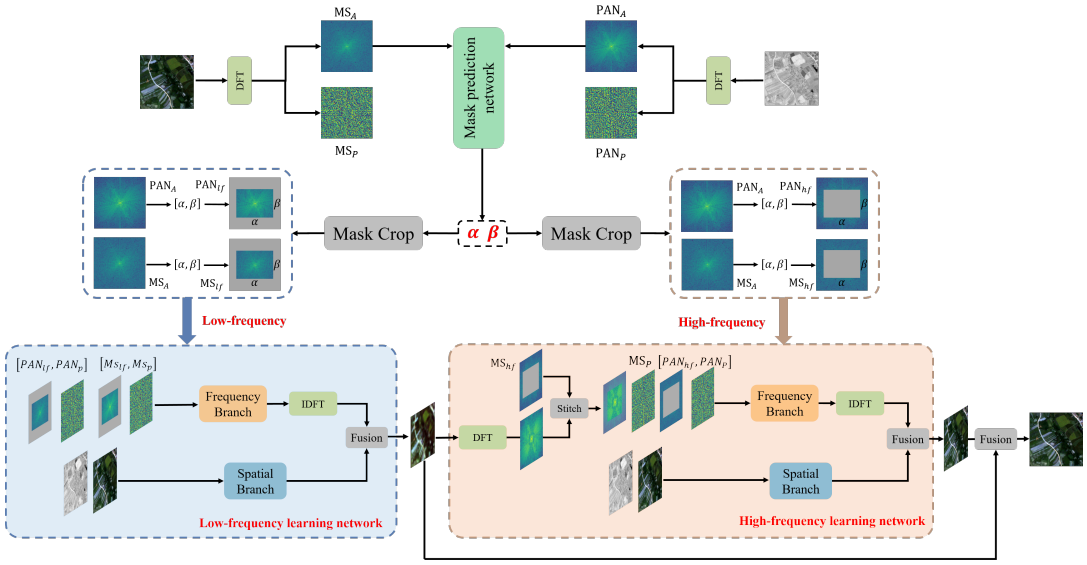

Figure 3: The pipeline of our proposed pan-sharpening framework. It consists of three key components: mask prediction network, low-frequency learning network and high-frequency learning network.

$\mathcal{F}^{-1}(x)$ defines the inverse Fourier transform accordingly. Both the Fourier transform and its inverse procedure can be efficiently implemented with the FFT algorithm in [13]. 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 parts of $\mathcal{F}(x)$. In our method, Fourier transform and the inverse procedure are computed independently on each channel of feature maps.

Targeting at pan-sharpening, there are several observations in the frequency domain as shown in Figure 1 and 2: 1) The phase of PAN is more similar to 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 transform 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; 3) Furthermore, we can figure out that the degrees of the missing low and high frequency are different, and they thus need to be treated discriminately. That is, the low-high frequency boundary is fuzzy and difficult to determine constantly for all the images. Therefore, how to decide the boundary needs to be delicately designed, and it motivates us to learn the low-high frequency boundary for each image adaptively.

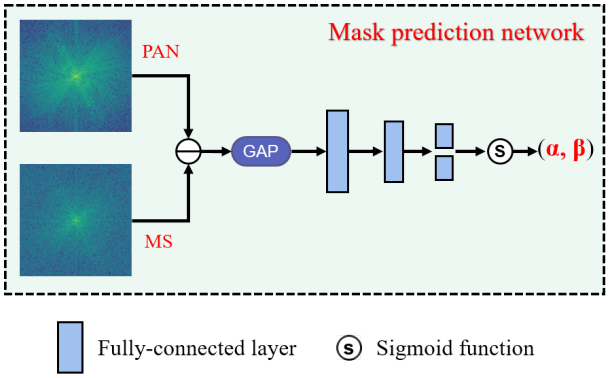

Figure 4: The detail of mask prediction network.

3.2 Framework

We further provide details of the fundamental components of our method, containing three key components: (a) mask prediction network, (b) low frequency learning network, (c) high frequency learning network.

Mask prediction network. Mask prediction network aims to generate the low-high frequency boundary in the form of a two-dimensional mask, detailed in Figure 3 and 4. 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 P and the r-times L by Bibubic upsampling into shallow feature representations. Next, the obtained modality-aware feature maps of MS and PAN are transformed into frequency domain as F_p and F_{ms} by Fourier Transform

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

where $\mathcal{A}(.)$ and $\mathcal{P}(.)$ indicate the amplitude and phase respectively. Further, we calculate the frequency difference F_d between the amplitude of F_p and F_{ms} and exploit F_d to predict the mask. Specifically, the F_d is firstly mapped into one-dimensional vector by global average pooling and then pass through the fully-connected layers to generate two scalars α and β in the range of 0 to 1. Finally, the mask $M \in R^{H \times W}$ is obtained by setting the corresponding bounding box with the size of $[\frac{H}{2} - \frac{\alpha}{2}H : \frac{H}{2} + \frac{\alpha}{2}H, \frac{W}{2} - \frac{\beta}{2}W : \frac{W}{2} + \frac{\beta}{2}W]$ as 1 and the remaining as 0

$$F_d = \mathcal{A}(F_p) - \mathcal{A}(F_{ms}), \tag{5}$$

$$\alpha, \beta = sigmoid(FC(GAP(F_d))), \tag{6}$$

$$M[\frac{H}{2} - \frac{\alpha}{2}H : \frac{H}{2} + \frac{\alpha}{2}H, \frac{W}{2} - \frac{\beta}{2}W : \frac{W}{2} + \frac{\beta}{2}W] = 1, \tag{7}$$

where *sigmoid* and *GAP* denote the sigmoid activation function and global average pooling operation respectively. In view of the predicted mask, we can adaptively screen out the low-high frequency for further low-high frequency learning.

Low frequency learning network. Low-frequency learning network aims to restore the expected low-frequency information. To be specific, based on the obtained mask, we filter out the low-frequency parts of PAN and MS modalities where the operation is performed over amplitude, and the phase is kept unchanged

$$\mathcal{A}(L_p), \mathcal{A}(L_{ms}) = M \odot \mathcal{A}(F_p), M \odot \mathcal{A}(F_{ms}), \tag{8}$$

where $\odot$ indicates the element-wise multiplication. As shown in Figure 2, the network contains two branches: the spatial branch $OS(.)$ and the frequency branch $OF(.)$. Referring to the filtered $\mathcal{A}(L_p)$ and $\mathcal{A}(L_{ms})$, we integrate them in the spatial and frequency domains. In the frequency branch, the operation is

$$\begin{aligned} \mathcal{A}(F_L) &= \mathcal{OF}(Cat[\mathcal{A}(L_p), \mathcal{A}(L_{ms})]), \\ \mathcal{P}(F_L) &= \mathcal{OF}(Cat[\mathcal{P}(F_p), \mathcal{P}(F_{ms})]), \\ F_L &= F^{-1}(\mathcal{A}(F_L), \mathcal{P}(F_L)). \end{aligned} \tag{9}$$

For the spatial branch, we directly perform the spatial information integration by spatial convolution operation over the MS image P and PAN image L

$$S_L = OS(P, L). \tag{10}$$

Followed by above spatial and frequency information integration, we further employ the fusion mechanism $FM(.)$ for the expected MS low-frequency reconstruction

$$H_L = FM(F_L, S_L). \tag{11}$$

High frequency learning sub-network. Referring to the restored low-frequency part and the predicted mask, we focus on the high-frequency part. As detailed in Figure 2, it combines the above-refined low-frequency and the original high-frequency for the latent high-frequency reconstruction. Specifically, we first transform the output of the low-frequency learning network into a frequency domain by Fourier Transform

$$\mathcal{A}(H_L), \mathcal{P}(H_L) = \mathcal{F}(H_L). \tag{12}$$

Then, we add the low-frequency with the original high-frequency of MS modality while screening out the high-frequency part of PAN modality

$$\mathcal{A}(H_{ms}) = \mathcal{A}(H_L) + (1 - M) \odot \mathcal{A}(F_{ms}), \tag{13}$$

$$\mathcal{A}(H_p) = (1 - M) \odot \mathcal{A}(F_p). \tag{14}$$

Since then, we follow the same operation process as low-frequency sub-network by spatial and frequency dual domain information integration. In the frequency branch, the operation is

$$\begin{aligned} \mathcal{A}(F_H) &= \mathcal{OF}(Cat[\mathcal{A}(H_p), \mathcal{A}(H_{ms})]), \\ \mathcal{P}(F_H) &= \mathcal{OF}(Cat[\mathcal{P}(F_p), \mathcal{P}(F_{ms})]), \\ F_H &= F^{-1}(\mathcal{A}(F_H), \mathcal{P}(F_H)). \end{aligned} \tag{15}$$

For the spatial branch, we perform the spatial information integration by spatial convolution operation over the MS image P and PAN images L

$$S_H = OS(P, L). \tag{16}$$

Followed by above spatial and frequency information integration, we further employ the fusion mechanism $FM(.)$ for the expected MS low-frequency reconstruction

$$H_H = FM(F_H, S_H). \tag{17}$$

In view of above low and high frequency learning networks, we further employ the fusion mechanism $FM(.)$ to integrate the two-stage outputs and add it with the up-sampling input MS image as the final output

$$H_O = FM(H_H, H_L) + L. \tag{18}$$

where $FM(.)$ is implemented by the residual channel attention *RCAB* [62].

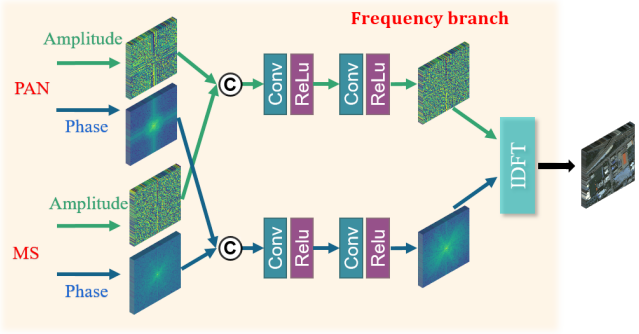

Figure 5: The detail of frequency branch.

3.3 Network implementation

As shown in Figure 5, we deepen into the designs of two key elements of the low and high frequency learning networks: (a) frequency branch, (b) spatial branch.

Frequency branch. In the frequency branch, based on above equations 9 and 15, the $OF(.)$ is implemented as the same operation. To be specific, we use 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

$$\mathcal{A}(F_L) = \mathcal{OA}(Cat[\mathcal{A}(L_p), \mathcal{A}(L_{ms})]), \tag{19}$$

$$\mathcal{P}(F_L) = \mathcal{OP}(Cat[\mathcal{P}(L_p), \mathcal{P}(L_{ms})]), \tag{20}$$

where Cat indicates the concatenation operation by channel dimension. In the frequency learning network, it is expressed as

$$\mathcal{A}(F_H) = \mathcal{OA}(Cat[\mathcal{A}(H_p), \mathcal{A}(H_{ms})]), \tag{21}$$

$$\mathcal{P}(F_H) = \mathcal{OP}(Cat[\mathcal{P}(H_p), \mathcal{P}(H_{ms})]), \tag{22}$$

According to spectral convolution theorem in Fourier theory, processing information of Fourier space captures the global frequency representation in the frequency domain. In short, the frequency branch generates the global information representation.

Spatial branch. For the spatial operation $OS(.)$, the spatial branch first adopts a residual block [24] with 3×3 convolution layers to integrate information of PAN and MS features P_L and MS_L in low-frequency network and P_H and MS_H in the high-frequency network, then generate the space representation F_{spa} in the spatial domain. It is well recognized that ordinary convolution focuses on learning local representations in the spatial domain. In short, the spatial branch provides the local information representation S_L and S_H in low and high-frequency networks.

3.4 Loss function

Let H_L, H_H, H_O and GT, M denote the low-frequency sub-network, high-frequency sub-network, the whole network outputs, the corresponding ground truth, and the mask of mask prediction subnetwork, respectively. To generate pleasing pan-sharpening results, we propose a joint spatial-frequency domain loss to supervise network training. In the spatial domain, we adopt the $L1$ loss

$$\mathcal{L}_{spa} = \|H_O - GT\|_1 + \eta \|H_H - GT\|_1 . \tag{23}$$

In the frequency domain, we first employ the DFT to convert H_L, H_O, 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, H_O, and GT are summed to produce the total frequency loss

$$\mathcal{L}_{freL} = \|M \odot (\mathcal{A}(H_L) - \mathcal{A}(GT))\|_1 + \|\mathcal{P}(H_L) - \mathcal{P}(GT)\|_1 , \tag{24}$$

$$\mathcal{L}_{freH} = \|\mathcal{A}(H_H) - \mathcal{A}(GT)\|_1 + \|\mathcal{P}(H_H) - \mathcal{P}(GT)\|_1 . \tag{25}$$

Finally, the overall loss function is formulated as follows

$$\mathcal{L} = \mathcal{L}_{spa} + \lambda(\mathcal{L}_{freL} + \mathcal{L}_{freH}), \tag{26}$$

where λ and η are the weight factors and set to 0.1 empirically.

4 EXPERIMENTS AND RESULTS

4.1 Baseline methods

We compare our method's performance to that of many representative pansharpening techniques to illustrate its efficacy: 1) five cutting-edge deep learning techniques, including PNN [44], PANNET [58], MSDCNN [60], SRPPNN [4], GPPNN [55] and BAM [68]; 2) five promising traditional methods, namely SFIM [39], Brovey [20], GS [36], IHS [22], and GFPCA [38].

4.2 Implementation details

On a single NVIDIA GeForce GTX 2080Ti GPU running on a personal computer, we built all our networks in Python using PyTorch. Adam optimizer is used to train them with 1000 epochs. The initial investment of with $8{\times}10^{-4}$ starts the learning rate. The learning rate is halved every 200 epochs. The Wald protocol tool [51] was used to construct the training set because of a lack of sufficient ground-truth pan-sharpened images. So, the MS image $H \in R^{M \times N \times C}$ and the PAN image $P \in R^{rM \times rN \times b}$ are both downsampled using the ratio r and are designated as $L \in R^{M/r \times N/r \times C}$ and $p \in R^{M \times N \times b}$. As opposed to L and p, H is the ground truth.

Three satellite image datasets from Worldview-II, Worldview-III, and GaoFen2 are examined. The PAN images are cropped into 128×128 pixel patches for each database, but the matching MS patches are 32×32 pixel patches. Several widely-used image quality assessment (IQA) measures, such as the relative dimensionless global error in synthesis (ERGAS) [3], the peak signal-to-noise ratio (PSNR), the structural similarity (SSIM), SAM [25], have been employed for performance evaluation.

We construct an additional full-resolution real-world dataset of 200 samples over the newly selected GaoFen2 satellite to conduct the model generalization comparison. To be more exact, the additional dataset is constructed in full-resolution mode, which creates PAN and MS images without downsampling, with PAN images having a resolution of 32×32 and MS images having a resolution of 128×128. Due to the scarcity of ground-truth MS images, we assess the model's performance using three commonly used no-reference IQA metrics: the spectral distortion index D_λ, the spatial distortion index D_S, the quality without reference (QNR).

4.3 Comparison with state-of-the-art methods

Evaluation on reduced-resolution scene. Table 1 summarizes the assessment metrics for three datasets, with the best results highlighted in red. Our method clearly beats earlier comparison algorithms on three satellite datasets across all assessment metrics. Our method increases PSNR by 0.24 dB, 0.16 dB, and 0.11 dB, respectively, when compared to the second-best results on the WorldView-II, GaoFen2, and WorldView-III datasets. Similar improvements in other indicators may be observed in addition to PSNR. We get significantly superior results to state-of-the-art deep learning-based systems, hence demonstrating the proposed method's effectiveness.

Additionally, we compare the visual findings in Figure 6 and Figure 7 to representative samples from the WorldView-II and GaoFen2 datasets to illustrate the technique's efficacy. The last row depicts the MSE residual between the pan-sharpened results and the ground truth. Our model displays minor spatial and spectral aberrations in contrast to other comparison methods. This conclusion is evident from a review of MSE maps. It's worth mentioning that our proposed strategy is more accurate than other comparison methods in terms of MSE residues. As a consequence, our technology beats existing competing pan-sharpening techniques.

Evaluation on full-resolution scene To assess the performance of our network in the full resolution case and the model generalization ability, we apply a pre-trained model built on GaoFen2 data to some unseen full-resolution GaoFen2 satellite datasets. The experimental results of all the methods are summarized in Table 2.

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
BAM	41.3527	0.9671	0.0239	0.9932	45.7419	0.9836	0.0134	0.6267	30.3845	0.9188	0.0773	3.1679
Ours	**41.6981**	**0.9723**	**0.0226**	**0.9514**	**47.3564**	**0.9896**	**0.0103**	**0.5475**	**30.5439**	**0.9228**	**0.0745**	**3.1097**

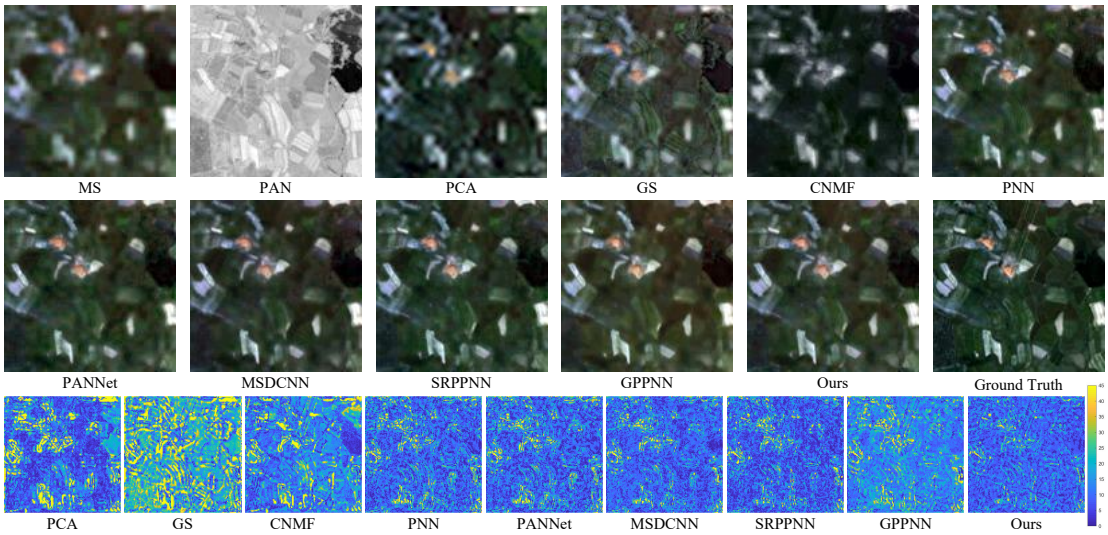

Figure 6: The visual comparisons between other pan-sharpening methods and our method on WorldView-II satellite.

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	BAM	**Ours**
D_λ↓	0.0822	0.0696	0.1378	0.0770	0.0914	0.0746	0.0737	0.0734	0.0767	0.0782	0.0755	**0.0685**
D_s↓	**0.1087**	0.2456	0.2605	0.2985	0.1635	0.1164	0.1224	0.1151	0.1162	0.1253	0.1159	0.1121
QNR↑	0.8214	0.7025	0.6390	0.6485	0.7615	0.8191	0.8143	0.8251	0.8173	0.8073	0.8211	**0.8463**

From Table 2, we can observe that our proposed method performs almost the best in terms of all the indexes, which indicates that our method has better generalization ability compared with other traditional and deep learning-based methods.

4.4 Ablation experiments

To investigate the contribution of the devised modules in our proposed network, we have conducted comprehensive ablation studies on the WorldView-II satellite dataset of the Pan-sharpening task. Specifically, the adaptively high-low frequency boundary learning

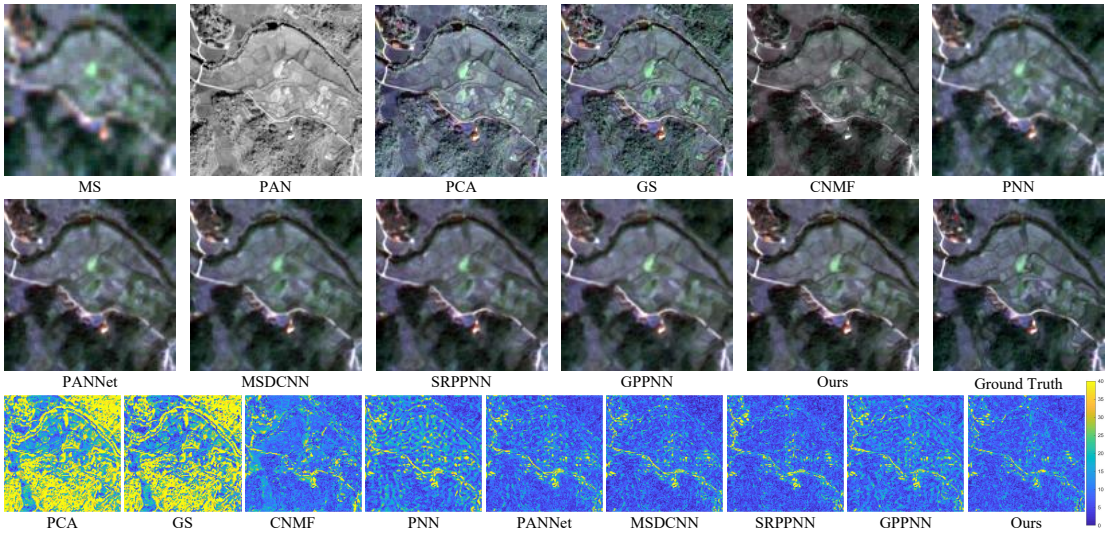

Figure 7: The visual comparisons between other pan-sharpening methods and our method on GaoFen2 satellite.

Table 3: Ablation studies comparison about the adaptively frequency boundary learning. The best performance is highlighted in red bold.

$\alpha = \beta$	PSNR↑	SSIM↑	SAM↓	ERGAS↓
0.25	41.5485	0.9683	0.0231	0.9542
0.5	41.5779	0.9690	0.0229	0.9536
Ours	**41.6981**	**0.9723**	**0.0226**	**0.9514**

Table 4: Ablation studies comparison about the frequency loss of $\mathcal{L}_{freL}$ and $\mathcal{L}_{freH}$.

Config	$\mathcal{L}_{freL}$	$\mathcal{L}_{freH}$	PSNR↑	SSIM↑	SAM↓	ERGAS↓
(I)	✗	✓	41.6341	0.9716	0.0229	0.9519
(II)	✓	✗	41.6127	0.9716	0.0229	0.9517
(III)	✗	✗	41.5995	0.9710	0.0229	0.9538
Ours	✓	✓	**41.6981**	**0.9723**	**0.0226**	**0.9514**

and the newly-designed frequency loss are the two core designs. All the experimental results are measured by the widely-used IQA metrics, i.e., ERGAS [3], PSNR, SSIM, and SAM.

The adaptively high-low frequency boundary learning. To explore its impact, we experiment the proposed network with the fixed boundary setting. The corresponding quantitative comparison with the fixed setting as $\alpha = \beta = 0.25$ and $\alpha = \beta = 0.5$ is reported in Table 3. Observing the results from Table 3, it can be clearly figured out that the model performance has obtained considerable degradation in terms of almost all the IQAs when removing the adaptively learning mechanism. It is because the low-high frequency boundary of each image is different and the fixed setting constraints the model flexibility, thus degrading the performance.

The frequency loss. The newly-designed frequency loss aims to directly emphasize global frequency information optimization. In the experiment of Table 4, we delete it to examine its effectiveness. The results in Table 4 demonstrate that removing it will severally degrade all metrics dramatically, indicating its significant role in our network. In contrast to existing methods that usually adopt pixel losses with local guides in the spatial domain, we propose the frequency domain supervision loss via Fourier transformation calculated on the global frequency components. Motivated by the spectral convolution theorem, direct emphasis on the frequency content is capable of better reconstructing the global information, thus improving the pan-sharpening performance.

5 CONCLUSION

In this paper, we propose a novel pan-sharpening framework from the frequency domain perspective. The core idea is to adaptively learn low-high frequency information integration in the spatial and frequency dual domains. In the framework, the low-high frequency boundary is first predicted in the form of a two-dimensional mask by referring to the modality-aware frequency information difference between PAN and MS images. Followed by the mask, the low and high frequency is adaptively screened out and then employed to restore the expected low and high-frequency components by spatial and frequency dual domains information integration in a coarse-to-fine manner. Extensive experiments over different satellite datasets demonstrate the effectiveness of our proposed method.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of MindSpore, CANN, and Ascend AI Processor used for this research. We also acknowledge the support of 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] Paolo Addesso, Gemine Vivone, Rocco Restaino, and Jocelyn Chanussot. 2020. A Data-Driven Model-Based Regression Applied to Panchromatic Sharpening. *IEEE Transactions on Image Processing* 29 (2020), 7779–7794.

[2] Bruno Aiazzi, Stefano Baronti, and Massimo Selva. 2007. Improving component substitution pansharpening through multivariate regression of MS + Pan data. *IEEE Transactions on Geoscience and Remote Sensing* 45, 10 (2007), 3230–3239.

[3] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce. 2007. Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data Fusion Contest. *IEEE Transactions on Geoscience and Remote Sensing* 45, 10 (2007), 3012–3021.

[4] Jiajun Cai and Bo Huang. 2021. Super-Resolution-Guided Progressive Pansharpening Based on a Deep Convolutional Neural Network. *IEEE Transactions on Geoscience and Remote Sensing* 59, 6 (2021), 5206–5220.

[5] Xiangyong Cao, Yang Chen, Qian Zhao, Deyu Meng, Yao Wang, Dong Wang, and Zongben Xu. 2015. Low-Rank Matrix Factorization under General Mixture Noise Distributions. In *2015 IEEE International Conference on Computer Vision (ICCV)*. 1493–1501.

[6] Xiangyong Cao, Lin Xu, Deyu Meng, Qian Zhao, and Zongben Xu. 2017. Integration of 3-dimensional discrete wavelet transform and Markov random field for hyperspectral image classification. *Neurocomputing* 226 (2017), 90–100.

[7] Xiangyong Cao, Zongben Xu, and Deyu Meng. 2019. Spectral-Spatial Hyperspectral Image Classification via Robust Low-Rank Feature Extraction and Markov Random Field. *Remote. Sens.* 11, 13 (2019), 1565.

[8] Xiangyong Cao, Jing Yao, Zongben Xu, and Deyu Meng. 2020. Hyperspectral Image Classification With Convolutional Neural Network and Active Learning. *IEEE Transactions on Geoscience and Remote Sensing* 58, 7 (2020), 4604–4616.

[9] Xiangyong Cao, Feng Zhou, Lin Xu, Deyu Meng, Zongben Xu, and John Paisley. 2018. Hyperspectral Image Classification With Markov Random Fields and a Convolutional Neural Network. *IEEE Transactions on Image Processing* 27, 5 (2018), 2354–2367.

[10] Wjoseph Carper, Thomasm Lillesand, and Ralphw Kiefer. 1990. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogrammetric Engineering and remote sensing* 56, 4 (1990), 459–467.

[11] Chen Chen, Yeqing Li, Wei Liu, and Junzhou Huang. 2015. SIRF: Simultaneous Satellite Image Registration and Fusion in a Unified Framework. *IEEE Transactions on Image Processing* 24, 11 (2015), 4213–4224.

[12] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2016. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 2 (2016), 295–307.

[13] M. Frigo and S. G. Johnson. 1998. FFTW: An adaptive software architecture for the FFT. *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on* 3 (1998).

[14] Xueyang Fu, Zihuang Lin, Yue Huang, and Xinghao Ding. 2019. A variational pan-sharpening with local gradient constraints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10265–10274.

[15] Xueyang Fu, Wu Wang, Yue Huang, Xinghao Ding, and John Paisley. 2021. Deep Multiscale Detail Networks for Multiband Spectral Image Sharpening. *IEEE Transactions on Neural Networks and Learning Systems* 32, 5 (2021), 2090–2104.

[16] Ying Fu, Zhiyuan Liang, and Shaodi You. 2021. Bidirectional 3D Quasi-Recurrent Neural Network for Hyperspectral Image Super-Resolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), 2674–2688.

[17] Dario Fuoli, Luc Van Gool, and Radu Timofte. 2021. Fourier Space Losses for Efficient Perceptual Image Super-Resolution. arXiv:2106.00783 [eess.IV]

[18] Morteza Ghahremani and Hassan Ghassemian. 2016. Nonlinear IHS: A promising method for pan-sharpening. *IEEE Geoscience and Remote Sensing Letters* 13, 11 (2016), 1606–1610.

[19] Alan R Gillespie, Anne B Kahle, and Richard E Walker. 1987. Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques. *Remote Sensing of Environment* 22, 3 (1987), 343–365.

[20] A. R. Gillespie, A. B. Kahle, and R. E. Walker. 1987. Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques - ScienceDirect. *Remote Sensing of Environment* 22, 3 (1987), 343–365.

[21] Juan Mario Haut, Mercedes E. Paoletti, Javier Plaza, Jun Li, and Antonio Plaza. 2018. Active Learning With Convolutional Neural Networks for Hyperspectral Image Classification Using a New Bayesian Approach. *IEEE Transactions on Geoscience and Remote Sensing* 56, 11 (2018), 6440–6461.

[22] R. Haydn, G. W. Dalke, J. Henkel, and J. E. Bare. 1982. Application of the IHS color transform to the processing of multisensor data and image enhancement. *National Academy of Sciences of the United States of America* 79, 13 (1982), 571–577.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778.

[25] A. F. Goetz J. R. H. Yuhas and J. M. Boardman. 1992. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. *Proc. Summaries Annu. JPL Airborne Geosci. Workshop* (1992), 147–149.

[26] Junjun Jiang, Jiayi Ma, Chen Chen, Zhongyuan Wang, Zhihua Cai, and Lizhe Wang. 2018. SuperPCA: A Superpixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 56, 8 (2018), 4581–4593.

[27] Junjun Jiang, Jiayi Ma, and Xianming Liu. 2020. Multilayer Spectral-Spatial Graphs for Label Noisy Robust Hyperspectral Image Classification. *IEEE Transactions on Neural Networks and Learning Systems* (2020), 1–14.

[28] Junjun Jiang, Jiayi Ma, Zheng Wang, Chen Chen, and Xianming Liu. 2019. Hyperspectral Image Classification in the Presence of Noisy Labels. *IEEE Transactions on Geoscience and Remote Sensing* 57, 2 (2019), 851–865.

[29] Junjun Jiang, He Sun, Xianming Liu, and Jiayi Ma. 2020. Learning Spatial-Spectral Prior for Super-Resolution of Hyperspectral Imagery. *IEEE Transactions on Computational Imaging* 6 (2020), 1082–1096.

[30] Kui Jiang, Zhongyuan Wang, Peng Yi, and Junjun Jiang. 2018. A Progressively Enhanced Network for Video Satellite Imagery Superresolution. *IEEE Signal Processing Letters* 25, 11 (2018), 1630–1634.

[31] Kui Jiang, Zhongyuan Wang, Peng Yi, Junjun Jiang, Guangcheng Wang, Zhen Han, and Tao Lu. 2019. GAN-Based Multi-level Mapping Network for Satellite Imagery Super-Resolution. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*. 526–531.

[32] Kui Jiang, Zhongyuan Wang, Peng Yi, Junjun Jiang, Emily Xiao, and Yuan Yao. 2018. Deep Distillation Recursive Network for Remote Sensing Imagery Super-Resolution. *Remote Sensing* 10 (10 2018), 1700.

[33] Kui Jiang, Zhongyuan Wang, Peng Yi, Guangcheng Wang, Tao Lu, and Junjun Jiang. 2019. Edge-Enhanced GAN for Remote Sensing Image Superresolution. *IEEE Transactions on Geoscience and Remote Sensing* 57, 8 (2019), 5799–5812.

[34] Muhammad Murtaza Khan, Jocelyn Chanussot, Laurent Condat, and Annick Montanvert. 2008. Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique. *IEEE Geoscience and Remote Sensing Letters* 5, 1 (2008), 98–102.

[35] P Kwarteng and A Chavez. 1989. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogrammetric Engineering and remote sensing* 55, 339-348 (1989), 1.

[36] C.A. Laben and B.V. Brower. 2000. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening. *US Patent 6011875A* (2000).

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

[38] W. Liao, H. Xin, F. V. Coillie, G. Thoonen, and W. Philips. 2017. 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*.

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

[40] Xiaochen Lu, Junping Zhang, Dezheng Yang, Longting Xu, and FengDe Jia. 2021. Cascaded Convolutional Neural Network-Based Hyperspectral Image Resolution Enhancement via an Auxiliary Panchromatic Image. *IEEE Transactions on Image Processing* 30 (2021), 6815–6828.

[41] Jiayi Ma, Han Xu, Junjun Jiang, Xiaoguang Mei, and Xiao-Ping Zhang. 2020. DDcGAN: A Dual-Discriminator Conditional Generative Adversarial Network for Multi-Resolution Image Fusion. *IEEE Transactions on Image Processing* 29 (2020), 4980–4995.

[42] Jiayi Ma, Wei Yu, Chen Chen, Pengwei Liang, Xiaojie Guo, and Junjun Jiang. 2020. Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Information Fusion* 62 (2020), 110–120.

[43] SG Mallat. 1989. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 7 (1989), 674–693.

[44] Giuseppe Masi, Davide Cozzolino, Luisa Verdoliva, and Giuseppe Scarpa. 2016. Pansharpening by Convolutional Neural Networks. *Remote Sensing* 8, 7 (2016).

[45] Jorge Nunez, Xavier Otazu, Octavi Fors, Albert Prades, Vicenc Pala, and Roman Arbiol. 1999. Multiresolution-based image fusion with additive wavelet decomposition. *IEEE Transactions on Geoscience and Remote sensing* 37, 3 (1999), 1204–1211.

[46] Robert A Schowengerdt. 1980. Reconstruction of multispatial, multispectral image data using spatial frequency content. *Photogrammetric Engineering and Remote Sensing* 46, 10 (1980), 1325–1334.

[47] Vijay P. Shah, Nicolas H. Younan, and Roger L. King. 2008. An Efficient Pan-Sharpening Method via a Combined Adaptive PCA Approach and Contourlets. *IEEE Transactions on Geoscience and Remote Sensing* 46, 5 (2008), 1323–1335.

[48] Xin Tian, Yuerong Chen, Changcai Yang, Xun Gao, and Jiayi Ma. 2020. A Variational Pansharpening Method Based on Gradient Sparse Representation. *IEEE Signal Processing Letters* 27 (2020), 1180–1184.

[49] Xin Tian, Yuerong Chen, Changcai Yang, and Jiayi Ma. 2021. Variational Pansharpening by Exploiting Cartoon-Texture Similarities. *IEEE Transactions on Geoscience and Remote Sensing* (2021), 1–16.

[50] Gemine Vivone, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio A Licciardi, Rocco Restaino, and Lucien Wald. 2014. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing* 53, 5 (2014), 2565–2586.

[51] Lucien Wald, Thierry Ranchin, and Marc Mangolini. 1997. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing* 63 (11 1997), 691–699.

[52] Xinya Wang, Jiayi Ma, and Junjun Jiang. 2021. Hyperspectral Image Super-Resolution via Recurrent Feedback Embedding and Spatial-Spectral Consistency Regularization. *IEEE Transactions on Geoscience and Remote Sensing* (2021), 1–13.

[53] Xiao Wu, Ting-Zhu Huang, Liang-Jian Deng, and Tian-Jing Zhang. 2021. Dynamic Cross Feature Fusion for Remote Sensing Pansharpening. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 14687–14696.

[54] Han Xu, Jiayi Ma, Zhenfeng Shao, Hao Zhang, Junjun Jiang, and Xiaojie Guo. 2021. SDPNet: A Deep Network for Pan-Sharpening With Enhanced Information Representation. *IEEE Transactions on Geoscience and Remote Sensing* 59, 5 (2021), 4120–4134.

[55] Shuang Xu, Jiangshe Zhang, Zixiang Zhao, Kai Sun, Junmin Liu, and Chunxia Zhang. 2021. Deep Gradient Projection Networks for Pan-sharpening. In *IEEE Conference on Computer Vision and Pattern Recognition*. 1366–1375.

[56] Keyu Yan, Man Zhou, Liu Liu, Chengjun Xie, and Danfeng Hong. 2022. When Pansharpening Meets Graph Convolution Network and Knowledge Distillation. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–15. https://doi.org/10.1109/TGRS.2022.3168192

[57] Gang Yang, Man Zhou, Keyu Yan, Aiping Liu, Xueyang Fu, and Fan Wang. 2022. Memory-Augmented Deep Conditional Unfolding Network for Pan-Sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1788–1797.

[58] Junfeng Yang, Xueyang Fu, Yuwen Hu, Yue Huang, Xinghao Ding, and John Paisley. 2017. PanNet: A deep network architecture for pan-sharpening. In *IEEE International Conference on Computer Vision*. 5449–5457.

[59] Qiangqiang Yuan, Yancong Wei, Xiangchao Meng, Huanfeng Shen, and Liangpei Zhang. 2018. 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 (2018), 978–989.

[60] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang. 2018. 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 (2018), 978–989.

[61] Hao Zhang and Jiayi Ma. 2021. GTP-PNet: A residual learning network based on gradient transformation prior for pansharpening. *ISPRS Journal of Photogrammetry and Remote Sensing* 172 (2021), 223–239.

[62] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. 2018. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision*. 286–301.

[63] Man Zhou, Xueyang Fu, Jie Huang, Feng Zhao, Aiping Liu, and Rujing Wang. 2022. Effective Pan-Sharpening With Transformer and Invertible Neural Network. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–15. https://doi.org/10.1109/TGRS.2021.3137967

[64] Man Zhou, Jie Huang, Yanchi Fang, Xueyang Fu, and Aiping Liu. 2022. Pan-Sharpening with Customized Transformer and Invertible Neural Network. AAAI Press.

[65] Man Zhou, Zeyu Xiao, Xueyang Fu, Aiping Liu, Gang Yang, and Zhiwei Xiong. 2021. Unfolding Taylor's Approximations for Image Restoration. In *NeurIPS*.

[66] Man Zhou, Keyu Yan, Jie Huang, Zihe Yang, Xueyang Fu, and Feng Zhao. 2022. Mutual Information-Driven Pan-Sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1798–1808.

[67] Man Zhou, Keyu Yan, Jinshan Pan, Wenqi Ren, Qiaokang Xie, and Xiangyong Cao. 2022. Memory-augmented Deep Unfolding Network for Guided Image Super-resolution. *ArXiv* abs/2203.04960 (2022).

[68] Tian-Jing Zhang Xiaoxu Jin Zi-Rong Jin, Liang-Jian Deng. 2021. BAM: Bilateral Activation Mechanism for Image Fusion. *Proceedings of the 29th ACM International Conference on Multimedia (ACM MM)* (2021), DOI: 10.1145/3474085.3475571.