Domain-irrelevant Feature Learning for Generalizable Pan-sharpening

Yunlong Lin*
School of Informatics, Xiamen University, Xiamen, China
lyl047853860@gmail.com

Zhenqi Fu*
School of Informatics, Xiamen University, Xiamen, China
fuzhenqi@stu.xmu.edu.cn

Ge Meng
School of Informatics, Xiamen University, Xiamen, China
mengg@stu.xmu.edu.cn

Yingying Wang
Institute of Artificial Intelligence, Xiamen University, Xiamen, China
wangyingying7@stu.xmu.edu.cn

Yuhang Dong
School of Informatics, Xiamen University, Xiamen, China
dongyh@stu.xmu.edu.cn

Linyu Fan
School of Informatics, Xiamen University, Xiamen, China
flyannie@stu.xmu.edu.cn

Hedeng Yu
School of Informatics, Xiamen University, Xiamen, China
yuhedeng@stu.xmu.edu.cn

Xinghao Ding†
School of Informatics, Xiamen University, Xiamen, China
dxh@xmu.edu.cn

ABSTRACT

Pan-sharpening aims to spatially enhance the low-resolution multispectral image (LRMS) by transferring high-frequency details from a panchromatic image (PAN) while preserving the spectral characteristics of LRMS. Previous arts mainly focus on how to learn a high-resolution multispectral image (HRMS) on the i.i.d. assumption. However, the distribution of training and testing data often encounters significant shifts in different satellites. To this end, this paper proposes a generalizable pan-sharpening network via domain-irrelevant feature learning. On the one hand, a structural preservation module (STP) is designed to fuse high-frequency information of PAN and LRMS. Our STP is performed on the gradient domain because it consists of structure and texture details that can generalize well on different satellites. On the other hand, to avoid spectral distortion while promoting the generalization ability, a spectral preservation module (SPP) is developed. The key design of SPP is to learn a phase fusion network of PAN and LRMS. The amplitude of LRMS, which contains 'satellite style' information is directly injected in different fusion stages. Extensive experiments have demonstrated the effectiveness of our method against state-of-the-art methods in both single-satellite and cross-satellite scenarios. Code is available at: https://github.com/LYL1015/DIRFL.

CCS CONCEPTS

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

*Both authors contributed equally to this research.
†Xinghao Ding is the corresponding author.

MM '23, October 29–November 3, 2023, Ottawa, ON, Canada.

ACM ISBN 979-8-4007-0108-5/23/10…$15.00
https://doi.org//10.1145/3581783.3611894

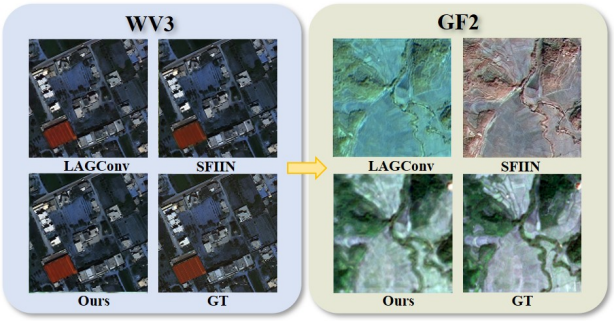

Figure 1: The generalization problem of pan-sharpening. Deep learning models are trained on the WV3 dataset and tested on both WV3 and GF2 datasets. The proposed method consistently achieves visually pleasing results. In contrast, existing methods (i.e., LAGConv [27] and SFIIN [61]) are sensitive to data distribution and show poor generalization performance.

KEYWORDS

Generalizable Pan-sharpening, Cross-satellite, Domain-irrelevant

ACM Reference Format:
Yunlong Lin, Zhenqi Fu, Ge Meng, Yingying Wang, Yuhang Dong, Linyu Fan, Hedeng Yu, and Xinghao Ding. 2023. Domain-irrelevant Feature Learning for Generalizable Pan-sharpening. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29–November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org//10.1145/3581783.3611894

1 INTRODUCTION

High-resolution multispectral images (HRMS) can be beneficial in a broad range of remote sensing tasks such as military systems, environmental monitoring, mapping services, and scene interpretation. Nevertheless, due to technological and physical limitations, satellites will often only capture a high-resolution panchromatic image (PAN) and a low-resolution multispectral image (LRMS). To obtain

HRMS, various pan-sharpening techniques have been developed to fuse LRMS and PAN [48, 57, 62, 63].

Traditional pan-sharpening methods adopt component substitution, multi-resolution analysis, and variational approaches to transform spatial details from PAN to LRMS. Due to the inaccessibility of sensor characteristics and improper modeling of prior knowledge, traditional methods commonly fail to restore precise spatial and spectral details of HRMS. Recently, deep convolutional neural networks have been introduced for pan-sharpening and have shown significant progress. The pioneering one refers to PNN [33], which adapts a three-layer convolution operation to directly learn a mapping from PAN and LRMS to HRMS. Due to the excellent ability in learning proper image features, PNN achieved a significant improvement compared with classical methods. Subsequently, the performance of pan-sharpening was further promoted by designing different network architectures.

Although deep convolutional neural networks can learn powerful representations from large quantities of annotated data, they cannot always generalize well when the input distribution changes. Unfortunately, most existing pan-sharpening networks train and evaluate the model on the same satellite dataset. These solutions are susceptible to the domain shift issue because the distribution of LRMS can be significantly different. For verification, we show an example of the domain shift issue in Figure 1, where all methods are trained on WV3 and evaluated on both WV3 and GF2. As can be seen, state-of-the-art (SOTA) approaches suffer from sharp performance degradation when their models are directly applied to new satellite data. Therefore, the generalization issue of deep learning-based pan-sharpening methods poses challenges to practical applications.

In this paper, we propose a novel pan-sharpening framework via learning domain-irrelevant representations to promote the generalization capability of the network. The proposed model consists of three distinct designs, i.e., a structural preservation module (STP), a spectral preservation module (SPP), and a complementary information fusion module (CIF). Specifically, STP aims at learning spatial structure information. STP is performed in the gradient domain to fuse the gradient features and enforce structural consistency between PAN and MS. Since gradient features are naturally consistent across different domains, STP is insensitive to the domain shift. To preserve spectral information, we propose SPP, which first decouples the phase and amplitude of PAN and MS in the frequency domain. Then, phase features are fused to further enhance the structural consistency. Note that SPP does not use the amplitude feature of PAN because it is uninformative for pan-sharpening. The amplitude of MS contains domain-specific features, which are directly injected into the network in different fusion stages. The final prediction of our method is obtained via a simple fusion module (i.e., CIF), which integrates the complementary information of structure and spectra. Note that CIF does not learn any amplitude features of MS. As a result, the proposed method can avoid learning domain-specific features and thus significantly improve the generalization ability.

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

- We propose a novel pan-sharpening framework that improves the generalization capability of CNN-based fusion networks. To the best of our knowledge, this is the first attempt to improve the generalization capability for Pan-sharpening.
- To learn generalized features, we use gradient and phase features to enforce structural consistency. Besides, the amplitude features of LRMS are directly injected into the network in different fusion stages to avoid learning domain-specific features.
- Extensive experiments on different satellite datasets demonstrate that the proposed method achieves SOTA performance in both single-satellite and cross-satellite scenarios.

2 RELATED WORK

2.1 Classic pan-sharpening methods

Classic pan-sharpening methods can be roughly classified into three categories, including component substitution (CS) [5, 10, 35], multi-resolution analysis (MRA) [12, 36, 38], and variational optimization (VO) [11, 14]. PCA [29], IHS [5] and Brovey [19] are three typical CS methods. These methods improved the spatial resolution by projecting the MS image into a new space and replacing the spatial information with the PAN image. To solve the spectral distortion problems, MRA used multi-resolution decomposition methods such as laplacian pyramid [37] and decimated wavelet transform [32] to extract the spatial information of the PAN image, and then inject it into the up-sampled MS images. VO methods [2, 25] designed specific optimization functions based on various prior assumptions for pan-sharpening. Overall, classic pan-sharpening approaches rely on hand-crafted features. These methods often generate spatial and spectral distortions due to the limited representation ability of the applied priors.

2.2 Deep learning based methods

With the highly nonlinear mapping capability of deep convolutional neural networks (CNN), numerous researchers have explored the use of this technology for image restoration [7–9, 16, 17, 30, 31, 51–53], hyperspectral images [4, 15, 23, 24] and remote sensing images [47, 56, 58, 59]. Recently, the paradigm of pan-sharpening has gradually shifted to data-driven approaches based on deep learning. For example, Masi et al. [33] were the first to apply CNN to address the problem of pan-sharpening and achieved a significant improvement by comparison with the classical methods. Yang et al. [49] adopted the resblock [20] and trained the network in the high-frequency domain. The LRMS image is directly added to the network output. This method can generalize to new satellites to a certain degree. However, the spectral knowledge of LRMS is not exploited. Yuan et al. [54] added the multi-scale module into the fundamental CNN architecture, and Cai et al.[3] also refers to the design idea of the single-image super-resolution network SRCNN [13]. Wu et al. [43] deployed many parallel branches to continuously integrate features with varied sizes into the network's backbone. Besides, some model-driven methods with physical constraints have been presented. Xie et al. [44] and Xu et al. [46] first used prior knowledge to formulate optimization problems for pan-sharpening task. Then the authors replaced the steps in the algorithm with deep neural networks. Zhou et al. [60, 61] trained a pan-sharpening network in both spatial and frequency domains.

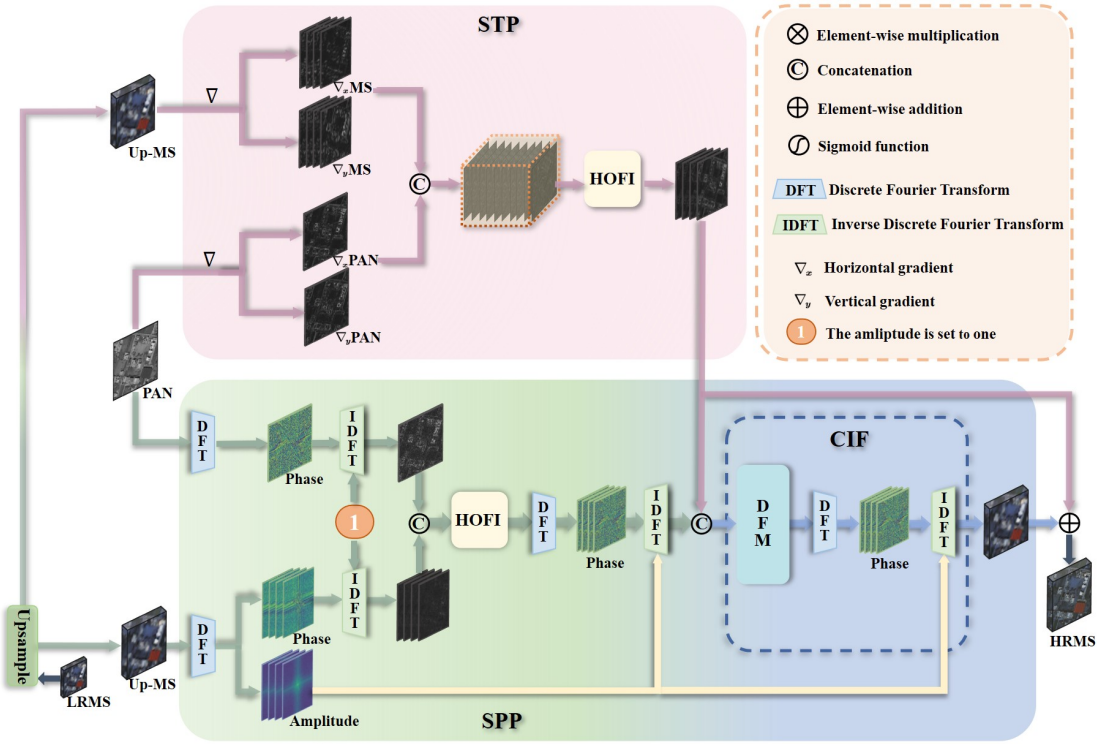

Figure 2: The overall framework of our proposed pan-sharpening framework. It consists of three distinct designs: 1) structural preservation mudule (STP), 2) spectral preservation module (SPP), and 3) complementary information fusion module (CIF). Specifically, STP is performed in the gradient domain to fuse the gradient features and enforce structural consistency between PAN and LRMS. To preserve spectral information, we propose SPP, which first decouples the phase and amplitude of PAN and LRMS in the Fourier domain. Then, phase features are fused to further enhance structural consistency. Meanwhile, the amplitude of LRMS contains domain-specific features, which is directly injected into the network at different stages. Finally, CIF integrates the complementary of structure and spectra to output the prediction.

3 METHODS

The goal of pan-sharpening is to combine the complementary information of PAN ($P \in R^{H\times W\times 1}$) and LRMS ($LM \in R^{H/r\times W/r\times C}$) image to generate HRMS ($HM \in R^{H\times W\times C}$) image, where H and W are the height and width of the image, C refers to spectral bands. The ratio of spatial resolution between PAN and the corresponding LRMS is equal to 4, i.e., $r = 4$. The overall framework of our proposed method is presented in Figure 2. It consists of three distinct designs: 1) structural preservation module, 2) spectral preservation module, and 3) complementary information fusion module.

3.1 Motivation

Deep learning-based approaches have achieved notable progress in pan-sharpening. However, those solutions commonly suffer from a lack of generalization ability. As shown in Figure 1, deep neural networks trained on WV3 perform poorly when evaluated on other satellites. We argue that the poor generalization performance of pan-sharpening networks may result from the training method, which leads to the model overfitting the 'styles information' in the LRMS image. Therefore, how to address the domain-specific features in LRMS is the key to generalizable pan-sharpening.

In Figure 3, to enforce structural consistency and promote generalization performance, PanNet is trained in the high-frequency domain, which is insensitive to the domain shift. Inspired by it, the gradient features of PAN and LRMS are utilized in our method to enhance structural consistency and learn domain-irrelevant features. For spectra and domain-specific features preservation, PanNet propagates the LRMS image to the network output. Nevertheless, directly mapping of the spectral information contained in the LRMS may lead to inaccurate predictions. This is because the spectral knowledge of LRMS is not well exploited and used. As the domain-specific features can be disentangled in the frequency domain[6, 41, 45, 50], which motivates us to design frequency-based feature integration networks for generalizable pan-sharpening. As illustrated in Figure 3, the domain-irrelevant knowledge (i.e., the phase) of LRMS is sufficiently leveraged in our method to enhance the structural consistency. Meanwhile, the amplitude of LRMS is directly injected in different fusion stages to ensure the generalization ability and avoid spectral distortion.

3.2 Structural Preservation

Since gradients contain rich image structure details and they are inherently consistent across domains, we perform the structural preservation on the gradient domain. Specifically, the gradients of

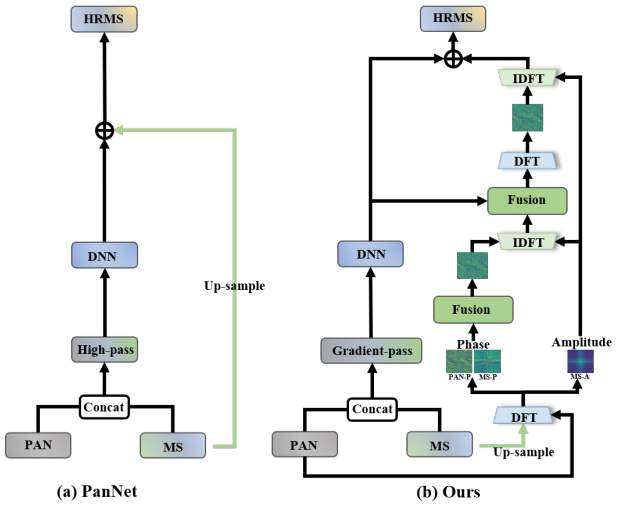

Figure 3: The flowchart of PanNet[49] and our method where DNN represents the deep neural networks. Note that the two fusion modules in our method are quite different.

PAN and LRMS are first calculated by:

$$\begin{aligned} P_x, P_y &= \nabla(P) \\ M_x, M_y &= \nabla\left(LM\uparrow\right) \end{aligned} \tag{1}$$

where P refers to the PAN image, and $LM\uparrow$ denotes the upsampled version of LRMS. $LM\uparrow$ and P have a same spatial resolution. ∇ indicates the gradient operation, P_x and P_y represent the horizontal and vertical gradient images of PAN, respectively. Similarly, M_x and M_y represent the horizontal and vertical gradient images of upsampled LRMS, respectively.

To facilitate the interaction between gradient features, a high-order information interaction module (HOFI) is employed (shown in Figure 4). Concretely, we first use a linear projection layer $\varphi_{\text{in}}(\cdot)$ to obtain a set of projected feature components K_0 and $\{Q_k\}_{k=0}^{n-1}$:

$$\left[K_0^{H\times W\times C_0}, Q_0^{H\times W\times C_0}, ..., Q_{n-1}^{H\times W\times C_{n-1}}\right] = \varphi_{\text{in}}(F_H) \tag{2}$$

We set the interaction orders (i.e., the n in g^nConv [34]) as 3 by default, F_H refers to the joint shallow gradient features of PAN and upsampled LRMS, $\varphi_{\text{in}}(F_H) \in \mathbb{R}^{H\times W\times(C_0+\sum_{0\le k\le n-1}C_k)}$. The channel dimension in each order as $C_k = \frac{C}{2^{n-k-1}}, 0 \le k \le n-1$. We then perform the gated convolution recursively by:

$$K_{k+1} = \begin{cases} f_k\left(Q_k\right) \odot K_k, & k = 0 \\ f_k\left(Q_k\right) \odot g_k\left(K_k\right), & 1 \leqslant k \leqslant n-1 \end{cases} \tag{3}$$

The coarse fused gradient component K_{k+1} is derived through the element-wise multiplication of $f_k(Q_k)$ and $g_k(K_k)$. Where f_k are a set of depth-wise convolutions (ie., 7×7 convolution) for deep feature extraction. $\{g_k\}$ are a set of 1×1 convolutions used to match the dimension in different orders.

Finally, we feed the last recursion step fused gradient component K_n to the projection layer $\varphi_{\text{out}}(\cdot)$ to obtain the result of high-order interactions. And we then model the relationship between feature channels by channel attention $CA(\cdot)$ to obtain a multispectral image $H_G \in \mathbb{R}^{H\times W\times C}$ that contains structural information:

$$H_G = CA\left(\varphi_{\text{out}}\left(K_n\right)\right) \tag{4}$$

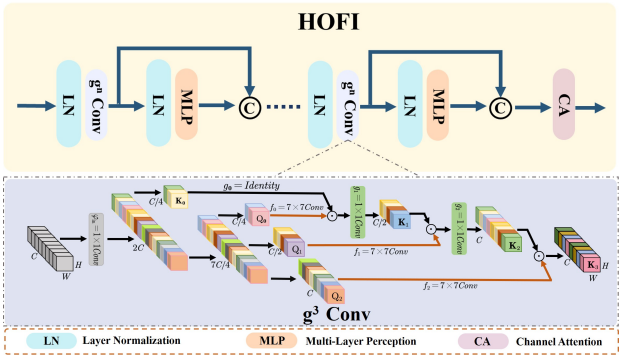

Figure 4: The architecture of our proposed high-order information interaction module (HOFI). We also provide the detailed implementation of g^nConv [34] with the interaction order $n = 3$ (below).

3.3 Spectral Preservation

In Figure 5, we perform inverse DFT to obtain phase-only reconstruction images in the spatial domain. As observed, The phase of PAN is more similar to the phase of GT than that of LRMS and exhibits richer structural information. Our results reveal that the phase of LRMS needs to be enhanced during the training phase. As CNN is not sensitive to changes in the phase domain [18], we explicitly fuse the phase features of PAN and LRMS to further promote structural consistency. Meanwhile, the amplitude features of LRMS are directly injected into the network in different fusion stages for spectral preservation. Formally, we first decouple the phase and amplitude of PAN (P) and upsampled LRMS ($LM\uparrow$) in the Fourier domain. The corresponding Fourier transform is expressed as:

$$\begin{aligned} \mathcal{A}\left(P\right), \mathcal{P}\left(P\right) &= \mathcal{F}\left(P\right) \\ \mathcal{A}\left(LM\uparrow\right), \mathcal{P}\left(LM\uparrow\right) &= \mathcal{F}\left(LM\uparrow\right) \end{aligned} \tag{5}$$

where $\mathcal{A}(\cdot)$ and $\mathcal{P}(\cdot)$ indicate the amplitude and phase respectively. Subsequently, we apply inverse DFT $\mathcal{F}^{-1}$ to the phase spectra of PAN (P) and upsample LRMS ($LM\uparrow$) to obtain the phase-only reconstruction image in the spatial domian:

$$\begin{aligned} P_u &= \mathcal{F}^{-1}\left(1, \mathcal{P}\left(P\right)\right) \\ L_u &= \mathcal{F}^{-1}\left(1, \mathcal{P}\left(LM\uparrow\right)\right) \end{aligned} \tag{6}$$

where P_u and L_u represent the phase-only reconstruction image of the PAN and the upsampled LRMS, respectively.

After achieving the phase-only reconstruction images P_u and L_u, we perform the high-order information interaction by the HOIF module between them (similar to Eq.2-4). Aiming to obtain a fused phase-only reconstruction image H_{us} contains enhanced structural information:

$$H_{us} = \text{HOFI}(F_u) \in \mathbb{R}^{H\times W\times C} \tag{7}$$

where F_u denotes the joint shallow features of P_u and L_u.

After obtaining H_{us}, we conduct the first injection of amplitude from LRMS. To be specific, we perform an inverse DFT to the phase of H_{us} and the amplitude of upsampled LRMS ($LM\uparrow$):

$$\begin{aligned} \mathcal{A}\left(H_{us}\right), \mathcal{P}\left(H_{us}\right) &= \mathcal{F}\left(H_{us}\right) \\ H_S &= \mathcal{F}^{-1}\left(\mathcal{A}\left(LM\uparrow\right), \mathcal{P}\left(H_{us}\right)\right) \end{aligned} \tag{8}$$

To fuse cross-modal complementary features of H_G and H_S, we propose a feature fusion module, i.e., $DFM(\cdot)$, whcih can be

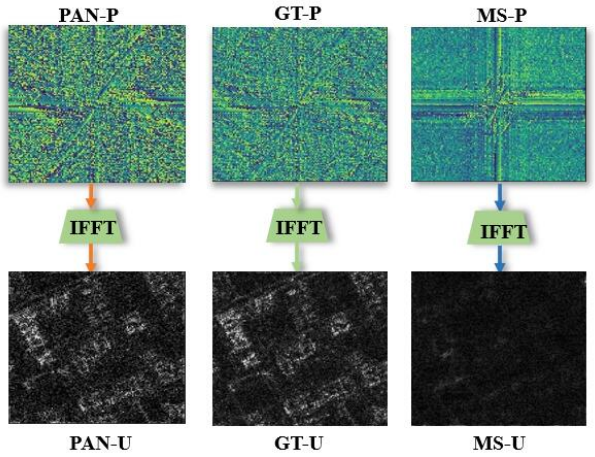

Figure 5: We applied an inverse DFT to the phase spectra of PAN image, GT image, and MS image to obtain the phase-only reconstruction image in the spatial domain. That means the amplitude of PAN, MS, and GT is set to 1.

expressed as:

$$H_f = DFM\left(H_G, H_S\right) \tag{9}$$

DFM is based on the residual channel attention mechanism and its detailed structure is presented in Figure 6. Notably, to avoid the network learning domain-specific features, we merely estimate the phase of the fused image H_f and discard the amplitude:

$$\mathcal{A}\left(H_f\right), \mathcal{P}\left(H_f\right) = \mathcal{F}\left(H_f\right) \tag{10}$$

Finally, we inject the amplitude of LRMS again via inverse DFT. Moreover, a residual learning mechanism is adopted by adding H_G to the final prediction HM:

$$HM = \mathcal{F}^{-1}\left(\mathcal{A}\left(LM\uparrow\right), \mathcal{P}\left(H_f\right)\right) + H_G \tag{11}$$

3.4 Loss Function

To generate high-quality results, we propose a joint structural-spectral loss to train the network. For structural consistency, we adopt the $L1$ loss:

$$\mathcal{L}_1 = \left\|HM - H_{gt}\right\|_1 \tag{12}$$

where HM and H_{gt} denote the network output and the corresponding ground truth, respectively. To further supervise the structural consistency, we employ the DFT to convert HM and H_{gt} into Fourier space, where the $L1$-norms of phase difference are calculated:

$$\mathcal{L}_p = \left\|\mathcal{P}\left(HM\right) - \mathcal{P}\left(H_{gt}\right)\right\|_1 \tag{13}$$

Simultaneously, we employ a spectral loss function to enhance the spectral information consistency among HM and H_{gt}. The spectral loss function is defined as follows:

$$\mathcal{L}_s = \frac{1}{N}\sum_{i=1}^{N} \arccos\left(\frac{HM \cdot H_{gt}}{\|HM\|_2 \left\|H_{gt}\right\|_2}\right) \tag{14}$$

where N denotes the number of pixels in each multispectral image band. Finally, the full objective function for our method is a weighted sum of all sub-loss terms:

$$\mathcal{L} = \mathcal{L}_1 + \lambda_1 \mathcal{L}_p + \lambda_2 \mathcal{L}_s \tag{15}$$

Table 1: The Composition of each satellite image dataset.

Satelite	WorldView-III	WorldView-II	GaoFen2
Training number	2150	-	-
Testing number	200	840	2912

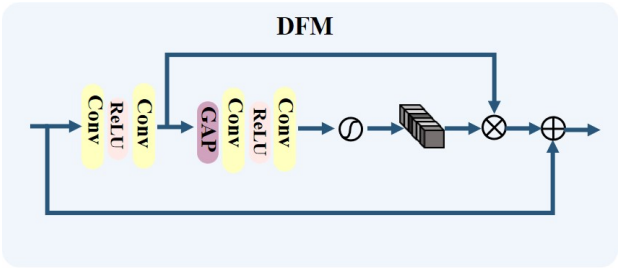

Figure 6: The architecture of feature fusion module (DFM).

where λ_1 and λ_2 denote the weights for balancing the three terms. λ_1, λ_2 are empirically set to 0.1 and 0.2 in all experiments.

4 EXPERIMENTS

4.1 Datasets

In order to show the effectiveness of the proposed model, we conduct experiments over the widely-used datasets including WorldView-II (WV2), GaoFen2 (GF2) and WorldView-III (WV3). As the paired training samples are not available, we construct the training datasets using the Wald protocol [40] to generate paired images. To be specific, given an origin MS image $HM \in R^{H\times W\times C}$ and its corresponding PAN image $P \in R^{rH\times rW\times C}$, both of them are downsampled with ratio r to obtain image pairs $LM \in R^{\frac{H}{r}\times\frac{W}{r}\times C}$ and $p \in R^{H\times W\times C}$. LM and p are treated as inputs, while HM serves as the ground truth. Furthermore, the PAN images are cropped into patches with the size of $128 \times 128 \times 1$, while the MS images are cropped into patches with the size of $32 \times 32 \times 4$.

For the WV3 satellite dataset, approximately 90% of the data are allocated for training and 10% for validation, whereas the remaining two satellite datasets are solely utilized for validation. The detailed composition of each dataset is reported in Table 1.

4.2 Implementation Details

All experiments are conducted on a single NVIDIA GeForce GTX 2080Ti GPU, and the PyTorch framework is used to construct our networks. During the training phase, we employ an ADAM optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ to update the network parameters for 1000 epochs with a batch size of 4. The learning rate is initialized with 8×10^{-4}. In parallel, a StepLR learning rate adjustment strategy is employed to reduce the learning rate by half after every 150 iterations.

4.3 Evaluation Metrics

We evaluate the algorithm performance using the following four widely used image quality assessment metrics: peak signal-to-noise ratio (PSNR) [22], structural similarity index (SSIM) [42], relative dimensionless global error in synthesis (ERGAS) [39], spectral angle mapper (SAM) [55]. The first three metrics measure the spatial distortion and the fourth one measures the spectral distortion. An image is better if its PSNR and SSIM are higher, and SAM and

Table 2: Quantitative comparison. Highlighted in red and underlined respectively indicate the first and second bset results.

Method	Worldview III				Worldview II				GaoFen2				#Param	#GFLOPs
	PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓	PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓	PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓		
PNN [33]	29.9418	0.9121	0.0824	3.3206	26.0051	0.8212	0.1461	6.5339	24.1737	0.7709	0.1556	4.4998	1.129M	0.068G
PanNet [49]	29.6840	0.9072	0.0851	3.4263	30.3104	0.8369	0.0707	3.3140	30.9604	0.7701	0.0502	3.0355	1.127M	0.069G
MSDCNN [54]	30.3038	0.9184	0.0782	3.1884	28.0123	0.8076	0.1025	5.0103	27.9701	0.7621	0.0736	4.2616	2.390M	3.916G
DiCNN [21]	29.7445	0.9090	0.0830	3.3902	26.1971	0.8176	0.1237	6.3263	28.4187	0.7869	0.0635	3.4095	0.412M	0.689G
SRPPNN [3]	30.4346	0.9202	0.0770	3.1553	28.2124	0.8183	0.1005	4.9073	28.5698	0.7622	0.0817	3.3878	17.114M	21.106G
GPPNN [46]	30.1785	0.9175	0.0776	3.2593	26.1912	0.8146	0.1069	6.0099	25.0855	0.6960	0.1100	5.2558	1.198M	1.397G
BAM [26]	30.3845	0.9188	0.0773	3.1670	27.4929	0.8127	0.1086	5.3810	27.9112	0.7237	0.0656	3.9072	0.971M	1.552G
LAGConv [27]	30.2933	0.9147	0.0782	3.2050	30.1302	0.8365	0.0761	3.8513	24.0323	0.7262	0.1404	7.9625	0.540M	0.511G
SFIIN [61]	30.5439	0.9228	0.0745	3.1097	29.5121	0.8360	0.0693	4.0178	28.2132	0.7029	0.0646	3.6545	0.921M	1.304G
Ours	30.4397	0.9207	0.0768	3.1478	31.1664	0.8464	0.0643	3.0163	32.8621	0.7874	0.0500	3.0117	1.091M	1.543G

Table 3: Quantitative comparison. Highlighted in red and underlined respectively indicate the first and second bset results.

Method	WorldView-III			WorldView-II			GaoFen2		
	D_λ ↓	D_S ↓	QNR ↑	D_λ ↓	D_S ↓	QNR ↑	D_λ ↓	D_S ↓	QNR ↑
PNN [33]	0.0460	0.0933	0.8654	0.1186	0.1228	0.7741	0.2778	0.1487	0.6153
PanNet [49]	0.0474	0.0942	0.8634	0.1092	0.1227	0.7826	0.1517	0.1253	0.7412
MSDCNN [54]	0.0432	0.0878	0.8732	0.1329	0.1228	0.7621	0.2916	0.1270	0.6191
DiCNN [21]	0.0469	0.0910	0.8666	0.1098	0.1156	0.7880	0.2564	0.1669	0.6204
SRPPNN [3]	0.0414	0.0909	0.8719	0.1371	0.1273	0.7543	0.2390	0.1092	0.6779
GPPNN [46]	0.0438	0.0936	0.8671	0.1193	0.1196	0.7764	0.2343	0.1391	0.6594
BAM [26]	0.0469	0.0910	0.8666	0.1314	0.1207	0.7649	0.3320	0.1292	0.5823
LAGConv [27]	0.0444	0.0886	0.8711	0.1248	0.1162	0.7743	0.2574	0.1864	0.6045
SFIIN [61]	0.0413	0.0876	0.8722	0.1209	0.1196	0.7751	0.3076	0.1201	0.6097
Ours	0.0420	0.0877	0.8716	0.0995	0.1195	0.7937	0.1798	0.0774	0.7562

ERGAS are lower. Furthermore, due to the lack of ground-truth MS images, we also quantify the model performance with three no-reference image quality assessment measures, i.e., the spectral distortion index D_λ [28], the spatial distortion index D_S and the quality with (QNR) [1].

4.4 Comparison with state-of-the-art methods

To prove the superior generalization capability of our approach, we compare its performance with several representative deep learning-based pan-sharpening methods: including PNN [33], PanNet [49], MSDCNN [54], DiCNN [21], SRPPNN [3], GPPNN [46], BAM [26], LAGConv [27], and SFIIN [61].

We've motivated the proposed method as being more robust to differences across satellites because it focuses on gradient and phase features for structural consistency, and the amplitude features of upsampled LRMS are directly injected into the network for spectral preservation. This strategy can help the network learn domain-irrelevant knowledge, and so networks trained on one satellite can generalize better to new satellites. To empirically show this, we train all comparison methods on WorldView-III dataset, but test them on multiple types of satallite datasets (e.g., WorldView-III, WorldView-II and GaoFen2).

Quantitative Comparison. Table 2 and Table 3 summarize the average assessment metrics for three datasets, with the best results highlighted in red. The results show that our model outperforms all the other models in terms of generalization performance on the WorldView-II and GaoFen2 datasets. Compared to the second-best results, our method improves PSNR by 0.86dB and 1.90dB across the two datasets, respectively. Furthermore, our method also leads all algorithms by a large margin in other metrics in addition to PSNR. As the domain shift increases, our model exhibits a slower performance degradation and thus demonstrates better generalization. In contrast, other models suffer from significant performance drops when dealing with more severe domain discrepancies. At the same time, our approach achieves comparable performance with state-of-the-art methods on the WorldView-III dataset.

Visual Comparison. Additionally, we also provide the visual comparison with other advanced algorithms on WorldView-III, WorldView-II and GaoFen2 datasets, respectively, as shown in Figure 7, Figure 8 and Figure 9. The last row of the figure shows the mean squared error (MSE) between the ground truth and the pan-sharpened images. In Figures 8 and 9, we can clearly observe that our method achieves the best visually appealing results, and the fused high-quality pan-sharpened image is closer to ground truth even though it is trained only on the WorldView-III dataset. Specifically, upon amplifying local regions, we observe that our proposed method preserves spectral information (less color difference in smooth regions) and achieves better spatial resolution (clearer structures around edge regions) compared to other methods. It's worth mentioning that our proposed strategy is more accurate than other comparison methods in terms of MSE residues, thus further demonstrating the ability of our proposed method to generalize effectively across multiple satellites. These results confirm and support our motivation to learn domain-irrelevant representations to

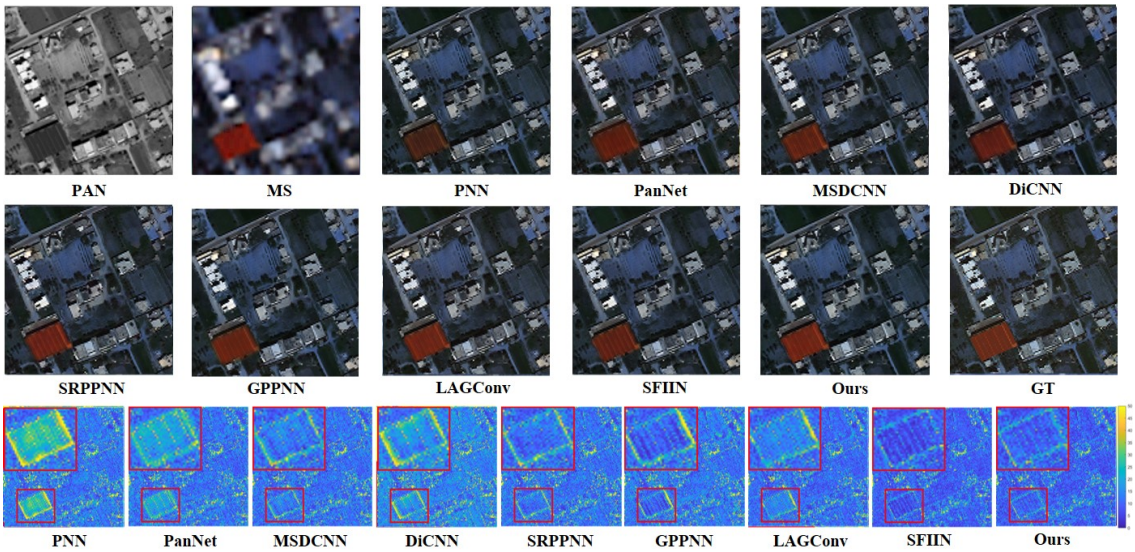

Figure 7: Visual comparison of all methods on WorldView-III.

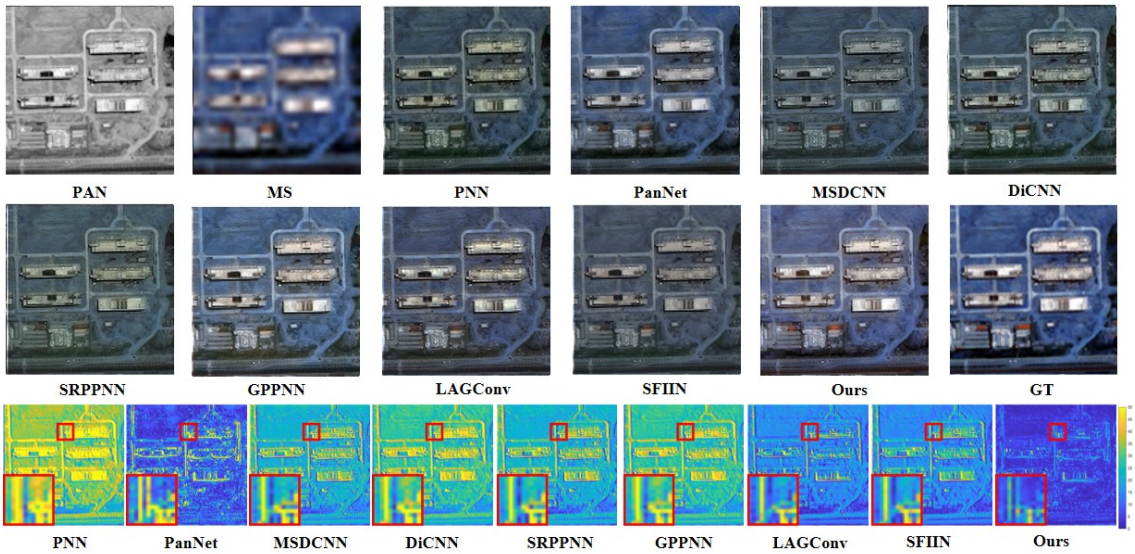

Figure 8: Visual comparison of all methods on WorldView-II.

promote the generalization capability of the network. In contrast, existing models tend to overfit the training satellite data distribution and fail when facing unseen satellites. More quantitative and qualitative results can be found in the supplementary material.

4.5 Comparison Over Full-Resolution Scenes

To compare the generalization of models in full-resolution scenes, we apply a pre-trained model built on WorldView-III data to an additional real-world full-resolution GaoFen2 dataset for evaluation. Table 4 provides an overview of the experimental findings for all approaches. From Table 4, we can observe that our devised technique performs almost at the top of all the indices, which suggests that it has superior generalization capacity than other deep learning-based methods. Please refer to the supplementary material to see visual results.

4.6 Ablation experiments

In this subsection, several ablation experiments are performed to verify the effectiveness of the proposed key insights, including: a) the phase and amplitude of PAN and LRMS are disentangled in the frequency domain, which can help network learn domain-irrelevant features and improve the model generalization capability; b) the gradient features are insensitive to the domain shift, which can also promote the model generalization ability. We train our model on the WorldView-III dataset, and evaluate the performance on the GaoFen2 dataset. The commonly used IQA measures, such as PSNR, SSIM, SAM, ERGAS index, D_λ, D_S, and QNR, are utilized to evaluate all of the experimental data.

(I) Effect of phase-amplitude decoupling in the frequency domain: In the spectral preservation phase, compared to previous

Table 4: Evaluation on the real-world full-resolution scenes from GaoFen2. Highlighted in red indicates the bset results.

Metrics	PNN [33]	PanNet [49]	MSDCNN [54]	DiCNN [21]	SRPPNN [3]	GPPNN [46]	BAM [26]	LAGConv [27]	SFIIN [61]	Ours
$D_\lambda\downarrow$	0.2443	0.1959	0.1825	0.1643	0.1601	0.2454	0.262	0.2455	0.1401	0.0696
$D_S\downarrow$	0.1419	0.2743	0.3226	0.3595	0.3246	0.3252	0.2983	0.2969	0.3258	0.2456
QNR↑	0.5090	0.6907	0.5563	0.5379	0.5700	0.5117	0.5200	0.5317	0.5825	0.7025

Table 5: Ablation studies comparison on GaoFen2. Highlighted in red indicates the bset results.

Ablation	GaoFen2						
	PSNR ↑	SSIM ↑	SAM ↓	ERGAS ↓	$D_\lambda\downarrow$	$D_S\downarrow$	QNR ↑
(I)	26.6153	0.6887	0.0677	4.2745	0.3379	0.1189	0.5838
(II)	28.6974	0.7325	0.0809	3.7741	0.2457	0.1143	0.6684
Ours	32.8621	0.7874	0.0500	3.0117	0.1798	0.0774	0.7562

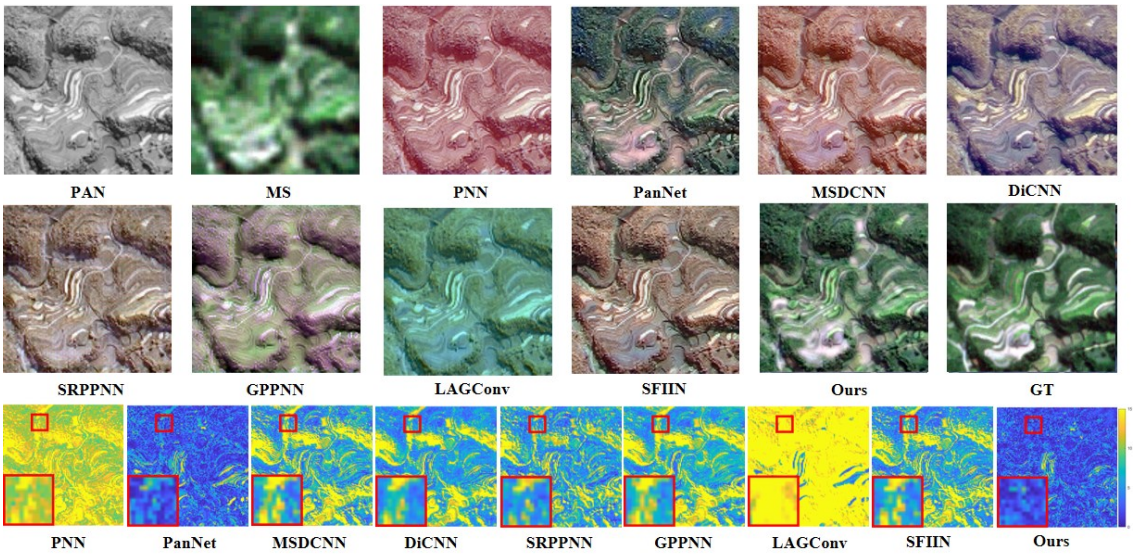

Figure 9: Visual comparison of all methods on Gaofen2.

works, we aim to enhance model generalization ability by decoupling the phase and amplitude of PAN and LRMS in the frequency domain. To demonstrate its effectiveness, we compare it to a variant model that directly fuses MS and PAN images in the spatial domain to preserve spectral information. As shown in Table 5, due to training-testing inconsistency, the quantitative performance of variant model degrades significantly, which nearly loses its generalization ability. This can be attributed to the direct fusion of LRMS and PAN images in the spatial domain, which inevitably guides the network parameters to learn domain-specific features contained in the amplitude of LRMS, thus degrading the generalization performance.

(II) The gradient features are insensitive to the domain shift: In the structural preservation phase, the gradient features of PAN and MS are utilized to enforce the structural consistency and learn the domain-irrelevant features. To assess its impact, we conduct an experiment that directly fuse the PAN and upsampled LRMS in the spatial domain instead of the gradient domain for structural consistency. Observing the results from Table 5, it can be clearly figured out that the variant model generalization performance has obtained considerable degradation in terms of all the IQAs when removing the gradient domain. It is because the gradient features are naturally consistent across different domains and are insensitive to the domain shift.

5 CONCLUSION

In this paper, we propose a generalizable pan-sharpening framework. Two core designs are devised to equip the network, i.e., a structural preservation module (STP) and a spectral preservation module (SPP). Concretely, STP is designed to fuse gradient information and enforce structural consistency between PAN and LRMS. SPP is presented to learn a phase fusion network of PAN and LRMS. The amplitude of LRMS that contains domain-specific features is directly injected into the network, which benefits the generalization capability of the network. Extensive experiments demonstrate that our proposed framework achieves better generalization performance than existing state-of-the-art pan-sharpening methods over multiple satellite datasets.

ACKNOWLEDGMENTS

This work was supported partly by the National Natural Science Foundation of China under Grants (No.82172033, No.U19B2031, No.61971369, No.52105126).

REFERENCES

[1] Luciano Alparone, Bruno Aiazzi, Stefano Baronti, Andrea Garzelli, Filippo Nencini, and Massimo Selva. 2008. Multispectral and panchromatic data fusion assessment without reference. *Photogrammetric Engineering & Remote Sensing* 74, 2 (2008), 193–200.

[2] Coloma Ballester, Vicent Caselles, Laura Igual, Joan Verdera, and Bernard Rougé. 2006. A variational model for P+ XS image fusion. *International Journal of Computer Vision* 69, 1 (2006), 43.

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

[4] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley. 2018. Hyperspectral image classification with Markov random fields and a convolutional neural network. *IEEE Trans. Image Process.* 27 (2018). https://doi.org/10.1109/TIP.2018.2799324

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

[6] Guangyao Chen, Peixi Peng, Li Ma, Jia Li, Lin Du, and Yonghong Tian. 2021. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 458–467.

[7] Sixiang Chen, Tian Ye, Yun Liu, Erkang Chen, Jun Shi, and Jingchun Zhou. 2022. Snowformer: Scale-aware transformer via context interaction for single image desnowing. *arXiv preprint arXiv:2208.09703* (2022).

[8] Sixiang Chen, Tian Ye, Yun Liu, Taodong Liao, Yi Ye, and Erkang Chen. 2022. MSP-Former: Multi-Scale Projection Transformer for Single Image Desnowing. *arXiv preprint arXiv:2207.05621* (2022).

[9] Sixiang Chen, Tian Ye, Jun Shi, Yun Liu, JingXia Jiang, Erkang Chen, and Peng Chen. 2023. DEHRFormer: Real-Time Transformer for Depth Estimation and Haze Removal from Varicolored Haze Scenes. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1–5.

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

[11] Liang-Jian Deng, Gemine Vivone, Weihong Guo, Mauro Dalla Mura, and Jocelyn Chanussot. 2018. A variational pansharpening approach based on reproducible kernel Hilbert space and heaviside function. *IEEE Transactions on Image Processing* 27, 9 (2018), 4330–4344.

[12] Minh N Do and Martin Vetterli. 2005. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on image processing* 14, 12 (2005), 2091–2106.

[13] 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. https://doi.org/10.1109/TPAMI.2015.2439281

[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] Y. Fu, Z. Liang, and S. You. 2021. Bidirectional 3D quasi-recurrent neural network for hyperspectral image super-resolution. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 14 (2021). https://doi.org/10.1109/JSTARS.2021.3057936

[16] Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. 2022. Uncertainty inspired underwater image enhancement. In *European Conference on Computer Vision*. Springer, 465–482.

[17] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. 2023. Learning a Simple Low-Light Image Enhancer From Paired Low-Light Instances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 22252–22261.

[18] Dario Fuoli, Luc Van Gool, and Radu Timofte. 2021. Fourier space losses for efficient perceptual image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2360–2369.

[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] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[21] Lin He, Yizhou Rao, Jun Li, Jocelyn Chanussot, Antonio Plaza, Jiawei Zhu, and Bo Li. 2019. Pansharpening via detail injection based convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12, 4 (2019), 1188–1204.

[22] Quan Huynh-Thu and Mohammed Ghanbari. 2008. Scope of validity of PSNR in image/video quality assessment. *Electronics letters* 44, 13 (2008), 800–801.

[23] 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* 33, 2 (2020), 839–852.

[24] J. Jiang, H. Sun, X. Liu, and J. Ma. 2020. Learning spatial-spectral prior for super-resolution of hyperspectral imagery. *IEEE Trans. Comput. Imaging* 6 (2020). https://doi.org/10.1109/TCI.2020.2996075

[25] Yiyong Jiang, Xinghao Ding, Delu Zeng, Yue Huang, and John Paisley. 2015. Pan-sharpening with a hyper-Laplacian penalty. In *Proceedings of the IEEE International Conference on Computer Vision*. 540–548.

[26] Zi-Rong Jin, Liang-Jian Deng, Tian-Jing Zhang, and Xiao-Xu Jin. 2021. BAM: Bilateral activation mechanism for image fusion. In *Proceedings of the 29th ACM International Conference on Multimedia*. 4315–4323.

[27] Zi-Rong Jin, Tian-Jing Zhang, Tai-Xiang Jiang, Gemine Vivone, and Liang-Jian Deng. 2022. LAGConv: Local-context adaptive convolution kernels with global harmonic bias for pansharpening. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 1113–1121.

[28] Muhammad Murtaza Khan, Luciano Alparone, and Jocelyn Chanussot. 2009. Pansharpening quality assessment using the modulation transfer functions of instruments. *IEEE transactions on geoscience and remote sensing* 47, 11 (2009), 3880–3891.

[29] P Kwarteng and A Chavez. 1989. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens* 55, 1 (1989), 339–348.

[30] Yun Liu, Zhongsheng Yan, Sixiang Chen, Tian Ye, Wenqi Ren, and Erkang Chen. 2023. NightHazeFormer: Single Nighttime Haze Removal Using Prior Query Transformer. *arXiv preprint arXiv:2305.09533* (2023).

[31] Yun Liu, Zhongsheng Yan, Aimin Wu, Tian Ye, and Yuche Li. 2022. Nighttime image dehazing based on variational decomposition model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 640–649.

[32] Stephane G Mallat. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence* 11, 7 (1989), 674–693.

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

[34] Yongming Rao, Wenliang Zhao, Yansong Tang, Jie Zhou, Ser Nam Lim, and Jiwen Lu. 2022. Hornet: Efficient high-order spatial interactions with recursive gated convolutions. *Advances in Neural Information Processing Systems* 35 (2022), 10353–10366.

[35] 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.

[36] Jean-Luc Starck, Emmanuel J Candès, and David L Donoho. 2002. The curvelet transform for image denoising. *IEEE Transactions on image processing* 11, 6 (2002), 670–684.

[37] 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.

[38] Gemine Vivone, Rocco Restaino, Giorgio Licciardi, Mauro Dalla Mura, and Jocelyn Chanussot. 2014. Multiresolution analysis and component substitution techniques for hyperspectral pansharpening. In *2014 IEEE Geoscience and Remote Sensing Symposium*. IEEE, 2649–2652.

[39] Lucien Wald. 2002. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES.

[40] 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, 6 (1997), 691–699.

[41] Jingye Wang, Ruoyi Du, Dongliang Chang, Kongming Liang, and Zhanyu Ma. 2022. Domain Generalization via Frequency-domain-based Feature Disentanglement and Interaction. In *Proceedings of the 30th ACM International Conference on Multimedia*. 4821–4829.

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

[43] 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*. 14687–14696.

[44] Q. Xie, M. Zhou, Q. Zhao, Z. Xu, and D. Meng. 2022. MHF-Net: An Interpretable Deep Network for Multispectral and Hyperspectral Image Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 3 (2022), 1457–1473. https://doi.org/10.1109/TPAMI.2020.3015691

[45] Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. 2021. A fourier-based framework for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14383–14392.

[46] S. Xu, J. Zhang, Z. Zhao, K. Sun, J. Liu, and C. Zhang. [n. d.]. Deep Gradient Projection Networks for Pan-sharpening. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1366–1375. https://doi.org/10.1109/CVPR46437.2021.00142

[47] K. Yan, M. Zhou, L. Liu, C. Xie, and D. Hong. 2022. When pansharpening meets graph convolution network and knowledge distillation. *IEEE Trans. Geosci. Remote Sens.* 60 (2022). https://doi.org/10.1109/TGRS.2022.3168192

[48] Keyu Yan, Man Zhou, Li Zhang, and Chengjun Xie. 2022. Memory-Augmented Model-Driven Network for Pansharpening. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*. Springer, 306–322.

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

[50] Yanchao Yang and Stefano Soatto. 2020. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4085–4095.

[51] Tian Ye, Sixiang Chen, Yun Liu, Yi Ye, Jinbin Bai, and Erkang Chen. 2022. Towards real-time high-definition image snow removal: Efficient pyramid network with asymmetrical encoder-decoder architecture. In *Proceedings of the Asian Conference on Computer Vision*. 366–381.

[52] Tian Ye, Sixiang Chen, Yun Liu, Yi Ye, Erkang Chen, and Yuche Li. 2022. Underwater light field retention: Neural rendering for underwater imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 488–497.

[53] Tian Ye, Yunchen Zhang, Mingchao Jiang, Liang Chen, Yun Liu, Sixiang Chen, and Erkang Chen. 2022. Perceiving and modeling density for image dehazing. In *European Conference on Computer Vision*. Springer, 130–145.

[54] 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.

[55] Roberta H Yuhas, Alexander FH Goetz, and Joe W Boardman. 1992. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop*.

[56] H. Zhang and J. Ma. 2021. GTP-PNet: a residual learning network based on gradient transformation prior for pansharpening. *ISPRS J. Photogramm. Remote. Sens.* 172 (2021). https://doi.org/10.1016/j.isprsjprs.2020.12.014

[57] Kaiwen Zheng, Jie Huang, Man Zhou, Danfeng Hong, and Feng Zhao. 2023. Deep Adaptive Pansharpening via Uncertainty-aware Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing* (2023).

[58] M. Zhou, X. Fu, J. Huang, F. Zhao, A. Liu, and R. Wang. 2022. Effective pan-sharpening with transformer and invertible neural network. *IEEE Trans. Geosci. Remote Sens.* 60 (2022). https://doi.org/10.1109/TGRS.2021.3137967

[59] Man Zhou, Jie Huang, Yanchi Fang, Xueyang Fu, and Aiping Liu. 2022. Pan-sharpening with customized transformer and invertible neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 3553–3561.

[60] 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*. 3375–3384.

[61] Man Zhou, Jie Huang, Keyu Yan, Hu Yu, Xueyang Fu, Aiping Liu, Xian Wei, and Feng Zhao. 2022. Spatial-frequency domain information integration for pan-sharpening. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVIII*. Springer, 274–291.

[62] Man Zhou, Jie Huang, Feng Zhao, and Danfeng Hong. 2022. Modality-aware Feature Integration for Pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing* (2022).

[63] Man Zhou, Keyu Yan, Xueyang Fu, Aiping Liu, and Chengjun Xie. 2023. PAN-Guided Band-Aware Multi-Spectral Feature Enhancement for Pan-Sharpening. *IEEE Transactions on Computational Imaging* 9 (2023), 238–249.