Unsupervised Pan-Sharpening via Mutually Guided Detail Restoration

Huangxing Lin[1], Yuhang Dong[2], Xinghao Ding[2], Tianpeng Liu[1]*, Yongxiang Liu[1]

[1]College of Electronic Science, National University of Defense Technology, China
[2]School of Informatics, Xiamen University, China
huangxinglin92@gmail.com, dongyh@stu.xmu.edu.cn, dxh@xmu.edu.cn, liutianpeng2004@nudt.edu.cn, lyx_bible@sina.com

Abstract

Pan-sharpening is a task that aims to super-resolve the low-resolution multispectral (LRMS) image with the guidance of a corresponding high-resolution panchromatic (PAN) image. The key challenge in pan-sharpening is to accurately modeling the relationship between the MS and PAN images. While supervised deep learning methods are commonly employed to address this task, the unavailability of ground-truth severely limits their effectiveness. In this paper, we propose a mutually guided detail restoration method for unsupervised pan-sharpening. Specifically, we treat pan-sharpening as a blind image deblurring task, in which the blur kernel can be estimated by a CNN. Constrained by the blur kernel, the pan-sharpened image retains spectral information consistent with the LRMS image. Once the pan-sharpened image is obtained, the PAN image is blurred using a pre-defined blur operator. The pan-sharpened image, in turn, is used to guide the detail restoration of the blurred PAN image. By leveraging the mutual guidance between MS and PAN images, the pan-sharpening network can implicitly learn the spatial relationship between the two modalities. Extensive experiments show that the proposed method significantly outperforms existing unsupervised pan-sharpening methods.

Introduction

Many Earth observation satellites can provide low-resolution multispectral (LRMS) images and high spatial resolution panchromatic (PAN) images simultaneously. The objective of pan-sharpening is to fuse the complementary information from the LRMS and PAN images to generate high spatial resolution multispectral (HRMS) images. HRMS images contain rich spectral and spatial information that is beneficial for various remote sensing tasks, such as land cover classification, change detection, and target tracking.

Pan-sharpening has consistently remained a vibrant subject of interest in the remote sensing community. Over the past few decades, numerous techniques have emerged to achieve improved pan-sharpening effects. Traditional pan-sharpening methods (Deng et al. 2022) primarily encompass three categories: component substitution (CS), multi resolution analysis (MRA) and variational optimization (VO) methods. These methods can be well-suited for various satellite images, but their effectiveness is hampered by the heavy reliance on hand-crafted features.

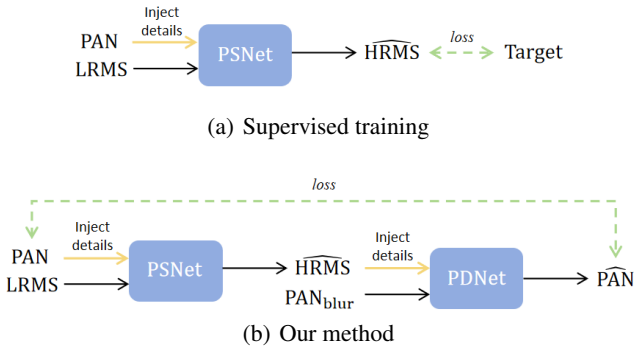


Figure 1: The conventional supervised training framework (a) and our proposed unsupervised training framework (b). PSNet and PDNet denote pan-sharpening network and PAN deblurring network, respectively. $\mathrm{PAN}_{\mathrm{blur}}$ is the blurry counterpart of the PAN image.

Recently, deep learning methods (Jin et al. 2022) have also been applied to pan-sharpening due to their powerful feature extraction capabilities. By utilizing a large amount of labeled data for training, supervised pan-sharpening methods significantly outperform traditional methods in restoring spectral and spatial details. However, the ground-truth for pan-sharpening does not exist. A commonly adopted alternative is to synthesize training data by reducing the size of PAN and LRMS images following the Wald's protocol (Wald, Ranchin, and Mangolini 1997), with the original LRMS images serving as the ground-truth. The networks trained using these reduced-resolution data overlook the scale variations in the test data, resulting in a degradation of their restoration quality for actual full-resolution samples (Ciotola et al. 2022).

To address the limitations of supervised methods, there has been a growing interest in unsupervised pan-sharpening methods that do not rely on ground-truth. Unsupervised pan-sharpening methods require some prior knowledge to guide the reconstruction toward more realistic high-resolution images. For instance, a simple way to ensure spectral fidelity is

*Corresponding author.

to downsample the fused images using a predefined degradation operator (*e.g.* bilinear + gaussian blur (Guo, Dian, and Li 2020)) and then minimize the distance with the LRMS image. The high-frequency part of the fused image is often assumed to be consistent with the PAN image. However, these prior assumptions may not hold true in practice. The predefined degradation operator may deviate from the real degradation model, and the statistical similarity of high-frequency components between fused and PAN images lacks empirical evidence. As a result, existing unsupervised pan-sharpening methods exhibit disappointing robustness across various satellite data.

For the above reasons, in this paper, we propose a novel Mutually Guided Detail Restoration method (PAN-MGDR) for unsupervised pan-sharpening. We regard pan-sharpening as a blind image deblurring problem. Particularly, the blur kernel is not assumed to be known as in other methods but is estimated using a CNN to satisfy non-negativity and equality constraints (Ren et al. 2020). The pan-sharpening network is constrained by the estimated blur kernel to preserve spectral fidelity. After the pan-sharpening process, the PAN image is blurred by a pre-defined degradation operator. Subsequently, we devise a detail restoration task for the blurred PAN image, where the fused image plays the role of the guiding image (as illustrated in Figure 1(b)). By minimizing the discrepancy between the detail-restored PAN image and the original PAN image, the fused image is compelled to contain similar spatial details to the original PAN image. In this manner, the pan-sharpening network learns to enhance the spatial details of LRMS images.

The contributions of this work are as follows:

- We propose a novel unsupervised pan-sharpening framework. Instead of imposing unfounded assumptions about spatial details, the proposed method leverages mutual guidance between two detail restoration tasks to implicitly learn the spatial relationship between MS and PAN images.
- We propose utilizing a CNN to adaptively estimate the blur kernel of LRMS images, enabling our method to adapt to images with unknown degradation kernels.
- Experiments on three satellite imagery datasets show that our method far outperforms other unsupervised pan-sharpening methods. PAN-MGDR can effectively restore the high spatial resolution image details while preserving the spectral information of the MS image.

Related Work

Traditional Pan-sharpening

Component substitution (CS) (Choi, Yu, and Kim 2010; Shah, Younan, and King 2008), multi resolution analysis (MRA) (Pradhan et al. 2006; Vivone, Marano, and Chanussot 2020), and variational optimization (VO) (Deng, Feng, and Tai 2019; Fang et al. 2013) are three representative categories of traditional methods. The CS method aims to enhance the spatial resolution by substituting the spatial details of the MS image with those of the PAN image in a transform domain. The main idea of the MRA method is to utilize multi-resolution decomposition techniques (*e.g.* Laplacian pyramid) to extract high-frequency spatial details from the PAN image and then inject them into the upsampled MS image. The VO method employs prior knowledge to constrain the optimization process and utilizes efficient algorithms to obtain the result of pan-sharpening. In general, traditional methods overly rely on hand-crafted features. Due to their limited representation capacity, the pan-sharpening results of traditional methods often exhibit undesirable spatial or spectral distortions.

Supervised Methods

With the rise of deep learning, supervised learning has become the mainstream method for pan-sharpening. Most supervised methods seek to enhance the pan-sharpening effect by exploring more advanced network architectures. Some high-performance networks include PanNet (Yang et al. 2017), MSDCNN (Yuan et al. 2018), DiCNN (He et al. 2019), SRPPPNN (Cai and Huang 2020), GPPNN (Xu et al. 2021), LAGConv (Jin et al. 2022), SFIIN (Zhou et al. 2022c), among others (Zhou et al. 2022b). Due to the lack of ground-truth for pan-sharpening, supervised methods are typically trained using reduced-resolution data. However, the scale variation leads to a significant degradation in the performance of supervised methods on full-resolution data.

Unsupervised Methods

Due to their independence from ground-truth, unsupervised pan-sharpening methods have become a trending topic in recent years. Unsupervised pan-sharpening methods (Liu et al. 2023) rely on assumptions about spectral and spatial details to guide the network in generating high-resolution multi-spectral images. For instance, both Z-PNN (Ciotola et al. 2022) and RMFF-UPGAN (Wu et al. 2023) assume that the degradation operator of the LRMS images is known to ensure spectral fidelity. UAP-Net (Xiong et al. 2023) utilizes a low-pass filter to maintain spectral consistency between the pan-sharpened image and the LRMS image. Given that the degradation kernel of LRMS images is typically unknown, the spectral fidelity will inevitably be compromised if the pre-set degradation operators in these methods deviate from the actual ones. On the other hand, CSFNet (Li et al. 2023) enhances the spatial details of the MS image by maximizing the local texture similarity between the pan-sharpened image and the PAN image. UCGAN (Zhou et al. 2022a) preserves spatial fidelity by minimizing the $L1$ distance between the high-frequency components of the output and the PAN image. However, the statistical correlation between HRMS images and PAN images is unclear. Indiscriminately minimizing the distance between the pan-sharpened image and the PAN image may lead to abnormal spatial details. By contrast, our PAN-MGDR possesses the capability to adaptively estimate the degradation kernel of LRMS images, enabling it to preserve spectral information effectively in the output. Additionally, PAN-MGDR leverages the mutual guidance between MS and PAN images to restore spatial details in the MS image. This efficient approach implicitly models the spatial correlation between MS and PAN data.

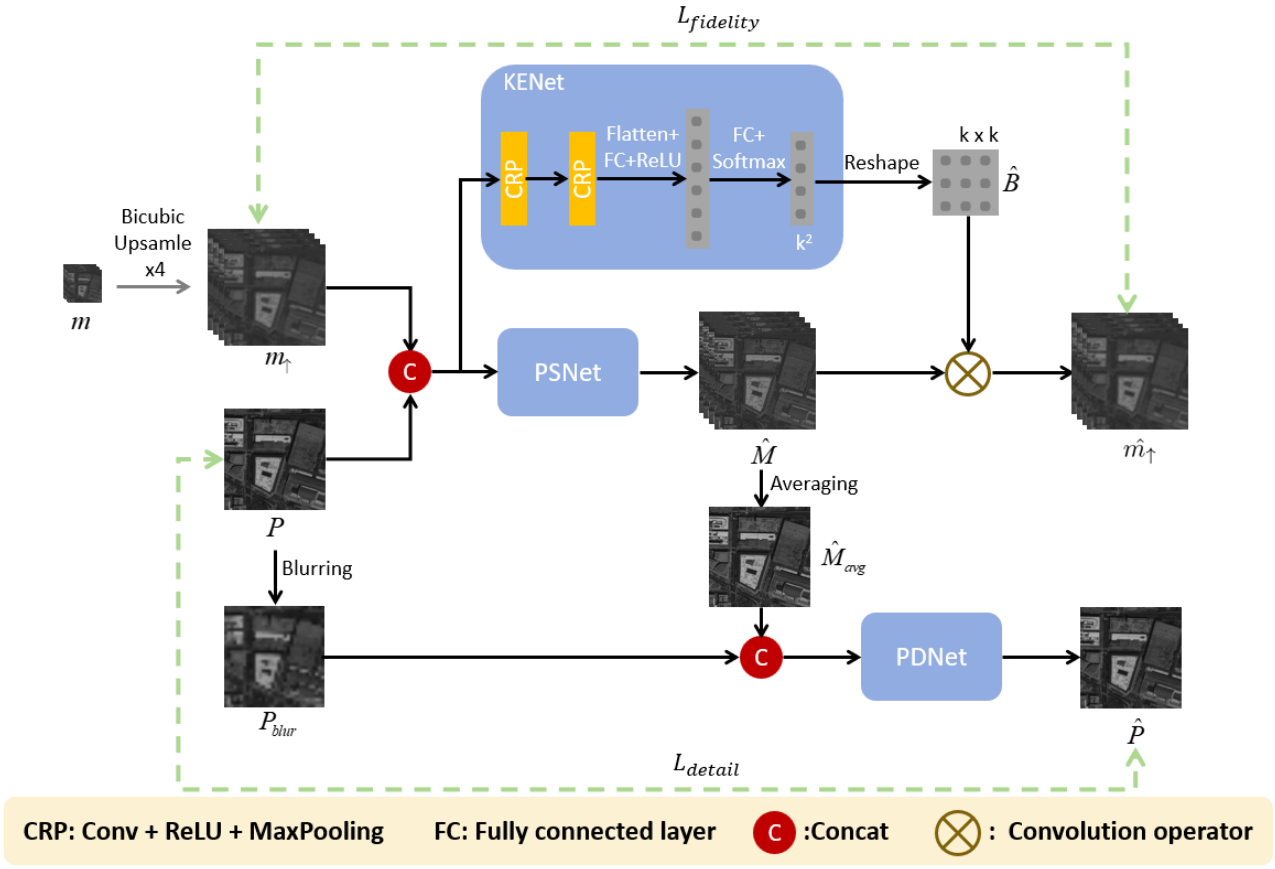


Figure 2: Illustration of PAN-MGDR training scheme. The pan-sharpening network (PSNet) aims to restore high-resolution spatial details of the MS image with the guidence of the PAN image. Constrained by the blur kernel $\hat{B}$ derived from the kernel estimation network (KENet), the pan-sharpening network preserves the spectral fidelity of its output $\hat{M}$. In turn, a PAN deblurring network (PDNet) utilizes $\hat{M}$ as guidance to restore the spatial details of P_{blur}. To minimize the disparity between P and $\hat{P}$, the pan-sharpening network must inject high-resolution spatial details consistent with the PAN image P into $\hat{M}$.

Methodology

Degradation Model

The observed LRMS image can be modeled as

$$m = (B * M)_{\downarrow}, \tag{1}$$

where $m \in R^{H/r \times W/r \times C}$ and $M \in R^{H \times W \times C}$ are LRMS and HRMS images, respectively. $B \in R^{k \times k}$ refers to the blur kernel. $*$ denotes the convolution operator. $\downarrow$ is the downsampling operator. H and W are the height and width of the image, C represents spectral bands and the ratio r is equal to 4. Here, we do not take into account the interference of noise. If the LRMS image is upsampled by interpolation (*e.g.* bicubic), then solving the HRMS image can be viewed as a blind image deblurring task in which the blur kernel B is unknown. This task can be solved using the Maximum A Posteriori (MAP) estimation model,

$$\arg\min_{M,B} \|B * M - m_{\uparrow}\|_2^2 + \lambda_1 \phi(M) + \lambda_2 \varphi(B) \tag{2}$$

where $\|B * M - m_{\uparrow}\|_2^2$ represents the fidelity term, $m_{\uparrow}$ is the upsampled version of m, λ_1 and λ_2 are the trade-off parameters, $\phi(M)$ and $\varphi(B)$ are regularizations incorporating prior knowledge about the HRMS image and the blur kernel, respectively.

The purpose of pan-sharpening is to utilize a high-resolution PAN image ($P \in R^{H \times W \times 1}$) to guide the detail enhancement of the LRMS image. From the perspective of MAP, the PAN image can be regarded as prior knowledge about the spatial details of the HRMS image. The quality of the fused image depends on the chosen regularization terms.

Blur Kernel Estimation

Most unsupervised pan-sharpening methods (Ciotola et al. 2022) assume that the blur kernel of the LRMS image is known (such as Gaussian blur) to constrain the spectral fidelity. However, due to the intricate nature of image acquisition processes, the blur kernel of MS images is often unknown, and images captured by different sensors exhibit unique characteristics. If the employed blur kernel does not match the actual point spread function of the imaging system, the performance of unsupervised pan-sharpening meth-

ods can deteriorate significantly. Therefore, unlike other methods that employ a fixed blur kernel, our PAN-MGDR utilizes a kernel estimation network (KENet) to adaptively estimate the blur kernel from the input MS image. KENet is a lightweight neural network consisting of 2 CRP modules and 2 fully connected layers, as shown in Figure 2. Each CRP module comprises a convolutional layer, a ReLU activation function, and a MaxPooling layer. The number of feature maps in each convolutional layer is 32. The feature map produced by the second CRP module is flattened into a $1D$ vector, which is then fed into a fully connected layer. The output of the second fully connected layer is a vector containing k^2 nodes. This vector is reshaped into a $2D$ $k \times k$ blur kernel $\hat{B}$. By convolving the fused image with $\hat{B}$, we can obtain a reconstructed blurred MS image $\hat{m_\uparrow}$,

$$\begin{aligned} \hat{m_\uparrow} &= \hat{B} * \hat{M} \\ &= reshape\left(G_{\theta_{KE}}\left(m_\uparrow, P\right)\right) * F_{\theta_{PS}}\left(m_\uparrow, P\right), \end{aligned} \tag{3}$$

where $\hat{M} = F_{\theta_{PS}}\left(m_\uparrow, P\right)$ is the fused image, $F_{\theta_{PS}}$ represents the pan-sharpening network (PSNet) with parameters θ_{PS}, $G_{\theta_{KE}}$ denotes the KENet with parameters θ_{KE}. Both the pan-sharpening network and KENet take as input a concatenation of $m_\uparrow$ and P. The output of KENet is reshaped into a $2D$ blur kernel $\hat{B}$. Based on the MAP model Eq. (2), we formulated a fidelity loss function in the following manner for updating the pan-sharpening network and KENet,

$$L_{fidelity} = \sum_i \left\| \hat{m_\uparrow}^{(i)} - m_\uparrow^{(i)} \right\|_1, \tag{4}$$

where i represents the i-th training sample. Here, we adopt the $L1$ loss. In addition to $L_{fidelity}$, we require some constraints to ensure the validity of the estimated blur kernel by KENet. Thus, we introduce non-negativity and equality constraints, both of which are widely applied in the task of image deblurring (Ren et al. 2020; Pan et al. 2017). The non-negativity constraint demands that all entries in the blur kernel $\hat{B}$ have values greater than 0. Meanwhile, the equality constraint ensures that the sum of all entries in blur kernel $\hat{B}$ equals 1, *i.e.*

$$\hat{B}_j \geq 0, \quad \sum_j \hat{B}_j = 1, \quad \forall j, \tag{5}$$

where j represents the j-th element in $\hat{B}$. To ensure the consistent satisfaction of these two constraints, we apply a Softmax activation function (Ren et al. 2020) after the final output layer of KENet. In this way, KENet is able to adaptively estimate the blur kernel from the input LRMS image.

Mutually Guided Detail Restoration

Training with the fidelity loss $L_{fidelity}$, the pan-sharpening network learns to preserve the spectral information consistent with the LRMS image in its output. However, the pan-sharpening network falls short in restoring high-resolution spatial details due to the absence of constraints on spatial details. To address this issue, we incorporate an additional PAN image deblurring task to implicitly impose constraints related to spatial details. Specifically, the PAN image is downsampled (bicubic) by a factor of 4 and then upsampled back

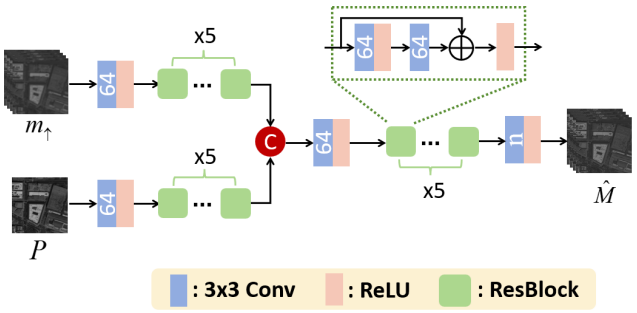


Figure 3: Network Architecture. The numerical value in each convolutional layer represents the number of output channels.

to the original size to generate a blurred PAN image P_{blur}. During the scaling process, numerous high-frequency spatial details are lost. Our aim is to utilize the fused image $\hat{M}$ to guide the restoration of spatial details in the blurred PAN image P_{blur}. However, to mitigate the potential overfitting of the neural network to specific bands within $\hat{M}$, we opt to employ the inter-band average $\hat{M}_{avg}$ of $\hat{M}$ rather than directly inputting $\hat{M}$ into the network. $\hat{M}_{avg}$ can be expressed as

$$\hat{M}_{avg} = \frac{1}{n} \sum_{i=1}^{n} \hat{M}_i, \tag{6}$$

where n is the number of bands of $\hat{M}$, i denotes the i-th band, respectively.

Similar to the pan-sharpening process, we concatenate $\hat{M}_{avg}$ with P_{blur} as input to the PAN deblurring network (PDNet). PDNet integrates complementary information from both $\hat{M}_{avg}$ and P_{blur} to produce a PAN image with finer details. The loss function for PAN image detail restoration is given by

$$\begin{aligned} L_{detail} &= \sum_i \left\| \hat{P}^{(i)} - P^{(i)} \right\|_1 \\ &= \sum_i \left\| H_{\theta_{PD}}\left(\hat{M}_{avg}^{(i)}, P_{blur}^{(i)}\right) - P^{(i)} \right\|_1, \end{aligned} \tag{7}$$

where $\hat{P} = H_{\theta_{PD}}\left(\hat{M}_{avg}, P_{blur}\right)$ is the PAN image restored by PDNet, $H_{\theta_{PD}}$ denotes the PDNet with parameters θ_{PD}, respectively. The original PAN image P serves as the ground truth. Note that the gradients produced by L_{detail} are utilized not only for updating PDNet but also propagated back to the pan-sharpening network.

Detail restoration is a difficult task. PDNet leverages the spatial details from $\hat{M}$ as a reference to restore the details of P_{blur}. The closer the spatial details of $\hat{M}$ resemble the original PAN image P, the more closely the image restored by PDNet approaches the desired outcome. As a result, minimizing the objective function L_{detail} will compel the pan-sharpening network to include high-resolution spatial details in its output $\hat{M}$ that closely resemble those of the original PAN image P.

Method	GaoFen2				QuickBird				Worldview-3			
	SAM↓	ERGAS↓	Q4↑	PSNR↑	SAM↓	ERGAS↓	Q4↑	PSNR↑	SAM↓	ERGAS↓	Q8↑	PSNR↑
Brovey	1.816	2.288	0.825	31.33	8.502	9.502	0.726	29.78	5.715	5.573	0.750	29.16
IHS	1.824	2.353	0.837	31.83	8.805	9.591	0.720	29.66	6.395	5.734	0.740	28.91
GS	2.060	2.409	0.828	31.52	8.645	9.543	0.721	29.73	6.049	5.519	0.764	29.31
GFPCA	2.415	2.455	0.739	31.70	9.844	11.16	0.581	28.39	6.522	6.478	0.635	27.89
PanGAN	2.654	2.712	0.823	30.38	10.29	8.461	0.806	30.67	9.905	6.706	0.736	28.22
PGMAN	2.177	2.609	0.828	31.70	8.970	11.29	0.663	28.26	6.369	7.481	0.655	26.58
LDP-Net	1.913	1.848	0.875	30.99	9.056	8.212	0.794	30.72	7.946	5.893	0.722	28.56
UCGAN	2.707	3.363	0.745	28.75	9.472	12.06	0.654	27.65	7.432	8.618	0.628	25.42
Ours	**1.428**	**1.484**	**0.924**	**35.71**	**8.233**	**7.728**	**0.821**	**31.49**	**5.066**	**4.518**	**0.833**	**31.03**

Table 1: Quantitative comparison on reduced-resolution data. The best results are marked in bold. ↑ indicates that higher values correspond to better performance, while ↓ signifies the opposite.

Loss Function

Our PAN-MGDR also incorporates a gradient consistency loss, *i.e.*

$$L_{grad} = \sum_i \left\| \nabla \left(\hat{M}_{avg}^{(i)} \right) - \nabla \left(P^{(i)} \right) \right\|_2^2, \tag{8}$$

where ∇ denotes the image gradient. Combining Eqs. (4), (7) and (8), the complete loss function of our PAN-MGDR is

$$L = L_{fidelity} + \lambda L_{detail} + \beta L_{grad}, \tag{9}$$

where λ and β are trade-off parameters. Since the gradients of HRMS images may not align with those of PAN images, the gradient consistency loss L_{grad} might compromise the quality of the fused image. Nonetheless, L_{grad} plays a role in preventing collapsed solutions and stabilizing PAN-MGDR training. To mitigate the adverse effects of L_{grad}, we set β to 1 and λ to 3. This strategic adjustment ensures that the quality of the fused image primarily depends on L_{detail} rather than $L_{fidelity}$ and L_{grad}.

In our method, the pan-sharpening network utilizes the PAN image as the guiding image, whereas the PDNet reciprocally uses the MS image as the guide. Through the mutual guidance between the MS and PAN images, the pan-sharpening network implicitly learns the spatial relationship between these two modalities. Consequently, it demonstrates the ability to efficiently restore high-quality, high-resolution multispectral images.

Network Architecture

The architecture of the pan-sharpening network in our method is shown in Figure 3. This network comprises two branches for feature extraction, along with a fusion branch. Each feature extraction branch consists of a convolutional layer and 5 ResBlocks. The extracted multispectral image features and PAN image features are concatenated and input into the fusion branch. The fusion branch leverages the features of these two modalities to produce a high-resolution multispectral image. While simple, we empirically find this network architecture to be highly efficient. The PDNet in our method also follows the architecture shown in Figure 3, with the only distinction being the variation in the number of input and output channels.

Experiments

Datasets

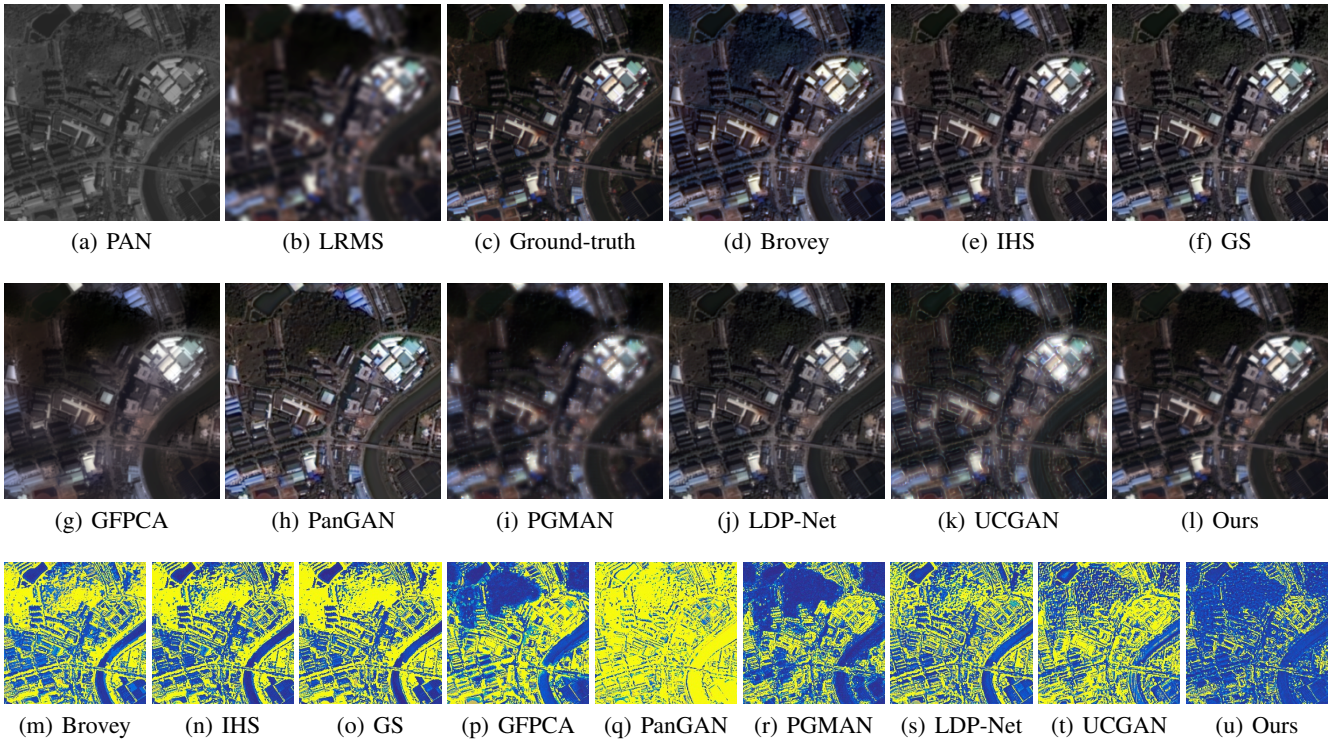
We utilize three satellite datasets, namely GaoFen2 (GF2), QuickBird (QB), and WorldView-3 (WV3), to assess the effectiveness of the proposed method. The experiments are conducted on both reduced-resolution and full-resolution data. The reduced-resolution data is synthesized following the Wald's protocol, wherein LRMS and PAN images are downsampled by a factor of 4, with the original LRMS image then serving as the ground truth.

Implementation Details

We use the Pytorch framework and the Adam optimizer to train the networks. All experiments are performed on a single NVIDIA GeForce GTX 3090 GPU. In PAN-MGDR, the size of the blur kernel output by KENet is set to 11×11. The large kernel size enables KENet to adapt to images degraded by blur kernels of different sizes. During training, PAN images are cropped into 64×64-sized image blocks, while MS image blocks are sized at 16×16. During the initial half of the training process, the learning rate is fixed at 0.0002, and in the latter half of the training, the learning rate linearly decays to 0.

Comparison Methods and Evaluation Metrics

We compare PAN-MGDR with several state-of-the-art pan-sharpening methods, including four traditional methods: Brovey (Gillespie, Kahle, and Walker 1987), IHS (Carper et al. 1990), GS (Laben and Brower 2000), GFPCA (Liao et al. 2015); as well as four unsupervised methods: PanGAN (Ma et al. 2020), PGMAN (Zhou, Liu, and Wang 2021), LDP-Net (Ni et al. 2022), UCGAN (Zhou et al. 2022a). To ensure a fair evaluation, all deep learning methods are retrained using their officially released code.



(a) PAN (b) LRMS (c) Ground-truth (d) Brovey (e) IHS (f) GS

(g) GFPCA (h) PanGAN (i) PGMAN (j) LDP-Net (k) UCGAN (l) Ours

(m) Brovey (n) IHS (o) GS (p) GFPCA (q) PanGAN (r) PGMAN (s) LDP-Net (t) UCGAN (u) Ours

Figure 4: PAN-sharpened images produced by different methods on the reduced-resolution GF2 dataset. The MSE residuals between the fused images and the ground-truth are displayed in the last row.

Method	**GaoFen2**			**QuickBird**			**Worldview-3**		
	$D_\lambda \downarrow$	$D_S\downarrow$	HQNR↑	$D_\lambda \downarrow$	$D_S\downarrow$	HQNR↑	$D_\lambda \downarrow$	$D_S\downarrow$	HQNR↑
Brovey	0.222	0.197	0.627	0.264	0.176	0.607	0.111	0.073	0.825
IHS	0.212	0.208	0.625	0.271	0.209	0.578	0.133	0.0797	0.799
GS	0.218	0.208	0.622	0.242	0.197	0.610	0.093	0.065	0.849
GFPCA	0.131	0.025	0.848	0.211	0.071	0.641	0.146	**0.039**	0.820
PanGAN	0.257	0.156	0.628	0.239	0.218	0.597	0.178	0.108	0.737
PGMAN	0.116	0.026	0.861	0.109	0.066	0.832	**0.025**	0.084	0.802
LDP-NET	0.168	0.165	0.696	0.398	0.350	0.391	0.125	0.114	0.777
UCGAN	0.357	0.039	0.618	0.182	**0.064**	0.766	0.192	0.183	0.620
Ours	**0.028**	**0.019**	**0.953**	**0.053**	0.090	**0.862**	0.037	0.041	**0.905**

Table 2: Quantitative comparison on full-resolution data. The best results are marked in bold. ↑ indicates that higher values correspond to better performance, while ↓ signifies the opposite.

Four image quality assessment metrics, peak signal-to-noise ratio (PSNR), spectral angle mapper (SAM), relative global synthesis errors (ERGAS) and the multi-band extension of the universal image quality index (Q2n), are adopted to compare the performance of each method on reduced-resolution data. Among these, lower values of SAM and ERGAS (with an ideal value of 0) indicate better results, while higher values of Q2n (with an ideal value of 1) and PSNR are more desirable. For full-resolution data, as reference MS images are unavailable, we selected three no-reference image quality assessment metrics: spectral distortion index D_λ, spatial distortion index D_S, and hybird QNR (HQNR) (Aiazzi et al. 2014). Among these metrics, smaller values for D_λ and D_S are preferable (with an ideal value of 0), while a higher HQNR value is desirable (with an ideal value of 1). For more details about the metrics, please refer to (Arienzo et al. 2022).

Comparison on Reduced-Resolution Data

The first experiment is conducted on reduced-resolution data. Quantitative results are reported in Table 1. Traditional pan-sharpening methods rely on hand-crafted features, which limit their performance. Additionally, other unsupervised methods face challenges in adequately modeling the spatial relationship between MS and PAN, consequently yielding poor fusion outcomes. Our method significantly outperforms other unsupervised and traditional pan-sharpening methods across all metrics on the three datasets. These results demonstrate the superiority and effectiveness of our PAN-MGDR.

Figure 4 illustrates the visual results produced by each method on the GaoFen2 dataset. The results obtained from Brovey, IHS, and GS methods exhibit pronounced spectral distortions. GFPCA and PGMAN failed to inject high-resolution spatial details into the MS image. UCGAN resulted in anomalous artifacts within the fused image. In contrast, the result of our PAN-MGDR exhibits clear spatial structures while effectively preserving spectral information. The visual quality of our result surpasses that of other unsupervised and traditional methods by a significant margin.

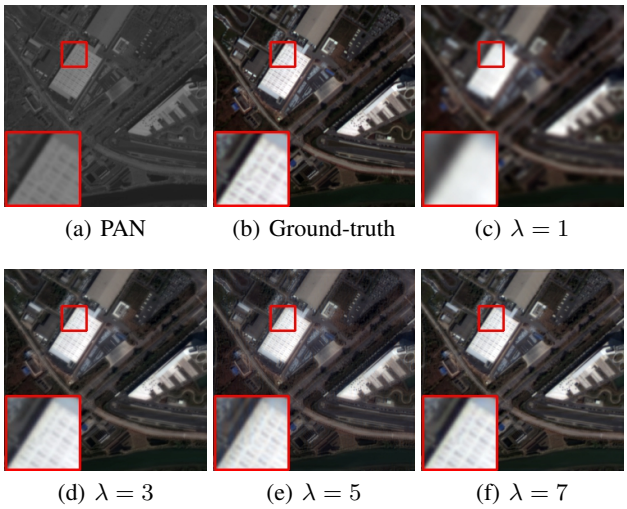
At the bottom of Figure 4, we showcase the MSE residual maps for each method. These residual maps depict the deviation between the fused image and the ground truth. Yellow pixels represent larger deviations, while blue pixels represent smaller ones. Notably, our PAN-MGDR displays more blue pixels in the residual map than other methods. This suggests that PAN-MGDR induces considerably fewer spectral and spatial distortions in the fused image than other methods.

Comparison on Full-Resolution Data

We further conduct comparative experiments on full-resolution data. We retrain all unsupervised methods using the full-resolution data. Upon completing the training, the same training data was directly employed as test data. Quantitative results are reported in Table 2. Lower values of D_λ and D_S indicate smaller spectral and spatial distortions, respectively. Higher HQNR corresponds to better image quality. Our method achieves the overall best performance among all compared methods. This demonstrates that our PAN-NGDR is a more effective solution compared to other methods for practical pan-sharpening tasks.

Ablation Study

Hyperparameter λ In the objective function Eq. (9), we utilize hyperparameters λ and β to balance the effects of spectral fidelity loss $L_{fidelity}$, detail restoration loss L_{detail}, and gradient consistency loss L_{grad}. We choose values from $\{1, 3, 5, 7\}$ as options for λ to assess its influence on pan-sharpening, keeping β constant at 1. We conduct experiments on the GaoFen2 dataset, and the test results are reported in Table 3 and Figure 5. As λ increases, the spatial details in the pan-sharpened results become clearer. This is because a larger λ strengthens the effect of the detail restoration loss, causing the pan-sharpening network to focus more on enhancing high-resolution spatial details. We have observed that setting λ to 3 enables PAN-MGDR to strike a favorable balance among the three loss functions, leading to superior fusion outcomes. Hence, we establish the default value of λ as 3.



(a) PAN (b) Ground-truth (c) $\lambda = 1$

(d) $\lambda = 3$ (e) $\lambda = 5$ (f) $\lambda = 7$

Figure 5: Pan-sharpened results with different λ.

	GaoFen2			
	SAM↓	ERGAS↓	Q4↑	PSNR↑
$\lambda = 1$	1.831	2.323	0.819	32.08
$\lambda = 3$	**1.428**	**1.484**	**0.924**	**35.71**
$\lambda = 5$	1.477	1.529	0.921	35.52
$\lambda = 7$	1.636	1.572	0.918	35.38

Table 3: Quantitative results of PAN-MGDR with different λ. The best results are marked in bold.

Conclusion

We propose to exploit an additional PAN deblurring task to impose a constraint on pan-sharpening. We demonstrate that through the mutually guided detail restoration between MS and PAN images, the pan-sharpening network can implicitly learn the spatial relationship between the two modalities. As a result, the network can learn to restore high-resolution multispectral images in an unsupervised manner. We validate the effectiveness and wide applicability of the proposed method on three satellite datasets. We anticipate that this method can be beneficial not only for pan-sharpening but also for other image fusion tasks.

Acknowledgements

This work was supported in part by the National Outstanding Youth Science Fund Project under Grant 62022091, and in part by the National Natural Science Foundation of China under Grant 61921001.

References

Aiazzi, B.; Alparone, L.; Baronti, S.; Carlà, R.; Garzelli, A.; and Santurri, L. 2014. Full-scale assessment of pansharpening methods and data products. In *Image and Signal Processing for Remote Sensing XX*. SPIE.

Arienzo, A.; Vivone, G.; Garzelli, A.; Alparone, L.; and Chanussot, J. 2022. Full-resolution quality assessment of pansharpening: Theoretical and hands-on approaches. *IEEE Geoscience and Remote Sensing Magazine*, 10(3): 168–201.

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

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

Choi, J.; Yu, K.; and Kim, Y. 2010. A new adaptive component-substitution-based satellite image fusion by using partial replacement. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1): 295–309.

Ciotola, M.; Vitale, S.; Mazza, A.; Poggi, G.; and Scarpa, G. 2022. Pansharpening by convolutional neural networks in the full resolution framework. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–17.

Deng, L.-J.; Feng, M.; and Tai, X.-C. 2019. The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior. *Information Fusion*, 52: 76–89.

Deng, L.-J.; Vivone, G.; Paoletti, M. E.; Scarpa, G.; He, J.; Zhang, Y.; Chanussot, J.; and Plaza, A. 2022. Machine Learning in Pansharpening: A benchmark, from shallow to deep networks. *IEEE Geoscience and Remote Sensing Magazine*, 10(3): 279–315.

Fang, F.; Li, F.; Shen, C.; and Zhang, G. 2013. A variational approach for pan-sharpening. *IEEE Transactions on Image Processing*, 22(7): 2822–2834.

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

Guo, A.; Dian, R.; and Li, S. 2020. Unsupervised blur kernel learning for pansharpening. In *IGARSS*. IEEE.

He, L.; Rao, Y.; Li, J.; Chanussot, J.; Plaza, A.; Zhu, J.; and Li, B. 2019. Pansharpening via detail injection based convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(4): 1188–1204.

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

Laben, C. A.; and Brower, B. V. 2000. Process for enhancing the spatial resolution of multispectral imagery using pansharpening. US Patent 6,011,875.

Li, S.; Tian, Y.; Wang, C.; Wu, H.; and Zheng, S. 2023. Cross spectral and spatial scale non-local attention based unsupervised pansharpening network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Liao, W.; Huang, X.; Van Coillie, F.; Thoonen, G.; Pižurica, A.; Scheunders, P.; and Philips, W. 2015. Two-stage fusion of thermal hyperspectral and visible RGB image by PCA and guided filter. In *WHISPERS Workshop*, 1–4. Ieee.

Liu, Q.; Meng, X.; Shao, F.; and Li, S. 2023. Supervised-unsupervised combined deep convolutional neural networks for high-fidelity pansharpening. *Information Fusion*, 89: 292–304.

Ma, J.; Yu, W.; Chen, C.; Liang, P.; Guo, X.; and Jiang, J. 2020. Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Information Fusion*, 62: 110–120.

Ni, J.; Shao, Z.; Zhang, Z.; Hou, M.; Zhou, J.; Fang, L.; and Zhang, Y. 2022. LDP-Net: An unsupervised pansharpening network based on learnable degradation processes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 5468–5479.

Pan, J.; Sun, D.; Pfister, H.; and Yang, M.-H. 2017. Deblurring images via dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(10): 2315–2328.

Pradhan, P. S.; King, R. L.; Younan, N. H.; and Holcomb, D. W. 2006. Estimation of the number of decomposition levels for a wavelet-based multiresolution multisensor image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 44(12): 3674–3686.

Ren, D.; Zhang, K.; Wang, Q.; Hu, Q.; and Zuo, W. 2020. Neural blind deconvolution using deep priors. In *CVPR*.

Shah, V. P.; Younan, N. H.; and King, R. L. 2008. An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5): 1323–1335.

Vivone, G.; Marano, S.; and Chanussot, J. 2020. Pansharpening: Context-based generalized Laplacian pyramids by robust regression. *IEEE Transactions on Geoscience and Remote Sensing*, 58(9): 6152–6167.

Wald, L.; Ranchin, T.; and Mangolini, M. 1997. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing*, 63(6): 691–699.

Wu, Y.; Li, Y.; Feng, S.; and Huang, M. 2023. Pansharpening using unsupervised generative adversarial networks with recursive mixed-scale feature fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Xiong, Z.; Liu, N.; Wang, N.; Sun, Z.; and Li, W. 2023. Unsupervised pansharpening method using residual network with spatial texture attention. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–14.

Xu, S.; Zhang, J.; Zhao, Z.; Sun, K.; Liu, J.; and Zhang, C. 2021. Deep gradient projection networks for pansharpening. In *CVPR*.

Yang, J.; Fu, X.; Hu, Y.; Huang, Y.; Ding, X.; and Paisley, J. 2017. PanNet: A deep network architecture for pan-sharpening. In *ICCV*.

Yuan, Q.; Wei, Y.; Meng, X.; Shen, H.; and Zhang, L. 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): 978–989.

Zhou, H.; Liu, Q.; and Wang, Y. 2021. PGMAN: An unsupervised generative multiadversarial network for pansharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 6316–6327.

Zhou, H.; Liu, Q.; Weng, D.; and Wang, Y. 2022a. Unsupervised cycle-consistent generative adversarial networks for Pan sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–14.

Zhou, M.; Huang, J.; Fang, Y.; Fu, X.; and Liu, A. 2022b. Pan-sharpening with customized transformer and invertible neural network. In *AAAI*.

Zhou, M.; Huang, J.; Yan, K.; Yu, H.; Fu, X.; Liu, A.; Wei, X.; and Zhao, F. 2022c. Spatial-frequency domain information integration for pan-sharpening. In *ECCV*.