




PDDM: Prior-Guided Dual-Branch Diffusion Model for Pansharpening

Changjie Chen , Yong Yang , Senior Member, IEEE, Shuying Huang , Member, IEEE, Hangyuan Lu , Weiguo Wan , Shengna Wei , Wenying Wen , Member, IEEE, and Shuzhao Wang

Abstract—Pansharpening is to fuse a panchromatic (PAN) image with a multispectral (MS) image to obtain a high-spatial-resolution MS (HRMS) image. Although the denoising diffusion probabilistic model can generate high-quality image details, its inherent stochasticity can lead to spectral and spatial distortions in the pansharpening task, and the adding noise method for fixed-size images can weaken the generalization of the model at different scales. To address these issues, a novel pansharpening method based on prior-guided dual-branch diffusion model (PDDM) is proposed. First, a dual-branch diffusion model for different information flows from MS and PAN images is constructed to achieve the spatial and spectral fidelity, which is developed by a collaborative and adversarial learning strategy. Then, to guide detail recovery and reduce the uncertainty of the generated detail information, two pregeneration modules based on different prior information are designed for pixel-to-pixel reconstruction. Finally, a focus module is constructed to fuse the features from the dual-branch and improve the generalization of the proposed PDDM. Extensive experiments on multiple satellite datasets demonstrate that the proposed PDDM has superior performance compared to state-of-the-art methods.

Index Terms—Diffusion model, dual-branch, pansharpening, pregeneration module.

I. INTRODUCTION

MULTISPECTRAL (MS) images have been widely used in various fields, such as rescue, navigation, and geological exploration [1], [2], [3], [4], [5]. However, due to the current physical limitations of sensors, MS sensors capture images with low-spatial resolution, which are not conducive to practical applications [6], [7], [8], [9]. To address this issue, researchers fuse low-spatial-resolution MS (LRMS) image with

high-spatial-resolution panchromatic (PAN) image to obtain high-spatial-resolution MS (HRMS) image, namely, pansharpening. Although great achievements have been obtained in the field of pansharpening, there are still some challenges [10], [11], [12]. For instance, how to generate high-quality spatial details and preserve the similarity to the source images in terms of spatial and spectral information, are still hot research topics [13], [14], [15].

Currently, the existing pansharpening methods are mainly divided into four categories, i.e., component substitution methods [16], [17], [18], multiresolution analysis methods [12], [19], variational optimization methods [20], [21], [22], and deep learning (DL)-based methods [23], [24], [25]. The first three categories are traditional methods, which use symbolic computation to generate the HRMS images. Wang et al. [26] proposed a two-stage approach to generate prior information, which is then integrated into the pansharpening model to simulate the spatial-spectral degradation process. Wen et al. [27] developed a spatial fidelity term with a learnable nonlinear mapping to establish a nonlinear relationship between PAN and HRMS images. Wang et al. [28] introduced fog-line priors to correct the fog effect in source images, and then used a tensor completion technique to reconstruct HRMS images from the corrected source images. Traditional methods have the advantage of interpretability and do not rely on large amounts of data. However, the fusion quality of the traditional methods depends on the model design, and the settings of various parameters in traditional methods are uncertain, which may lead to inaccuracy in the constructed models [29].

Inspired by the performance of DL technique, Masi et al. [23] proposed a simple multilayer neural network, called PNN, extracting feature from the MS and PAN images and fusing them by convolution neural network (CNN), which is the first CNN-based solution applied in the field of pansharpening. Subsequently, several improved DL-based pansharpening methods were proposed. TFNet [24] utilized two encoding structures to extract features from MS and PAN images, respectively, ensuring that spectral and spatial features are not interfered with each other during the extraction process. FusionNet [25] absorbed the idea of detail injection, using a residual network to learn the injected information between the HRMS and LRMS images, thereby reducing the learning burden of the network. TDNet [30] used a multilevel, multibranch, and multidirectional architecture to fully explore spatial and spectral features. AWFLN [31] allowed the network to focus on the composite features of

Received 11 July 2024; revised 16 September 2024; accepted 4 October 2024. Date of publication 10 October 2024; date of current version 23 October 2024. This work was supported by the National Natural Science Foundation of China under Grant 62072218, Grant 62261025, and Grant 62362035. (Changjie Chen and Shuying Huang contributed equally to this work.) (Corresponding author: Yong Yang.)

Changjie Chen and Wenying Wen are with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: chencjpro@163.com; wenyngwen@sina.cn).

Yong Yang, Shengna Wei, and Shuzhao Wang are with the School of Computer Science and Technology, Tiangong University, Tianjin 300387, China (e-mail: greatyang@126.com; shengnaw@163.com; 794446715@qq.com).

Shuying Huang is with the School of Software, Tiangong University, Tianjin 300387, China (e-mail: shuyinghuang2010@126.com).

Hangyuan Lu is with the College of Information Engineering, Jinhua University of Vocational Technology, Jinhua 321007, China (e-mail: lhyhziee@163.com).

Weiguo Wan is with the School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: wanweiguo@jxufe.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3477593

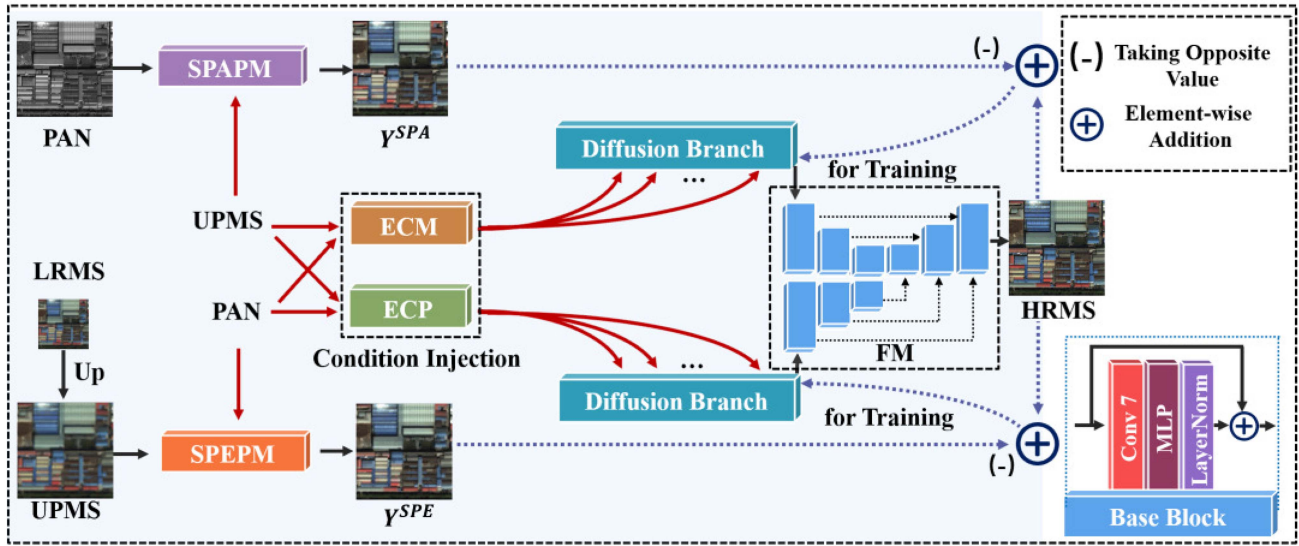


Fig. 1. Architecture of the proposed PDDM. SPAPM denotes the spatial pregeneration module, SPEPM denotes the spectral pregeneration module, ECM denotes encoding-MS module, ECP denotes the encoding-PAN module, and FM denotes the focus module.

spectral and spatial information by an adaptive spatial–spectral interleaved attention structure. Although the CNN methods have advantages in feature extraction, the significant modality difference between MS and PAN images makes their results difficult to preserve spatial information closed to the MS modality [32].

Currently, the denoising diffusion probabilistic model [33] can generate high-quality image details compared to the CNN-based model, and shows great potential in image denoising, generation, and other visual domains. In these studies, Rui et al. [34] developed a low-rank strategy to calculate the coefficient matrix, and then used this matrix to build an unsupervised low-rank diffusion method for pansharpening. Pang et al. [35] proposed an unsupervised HSI restoration framework called HIR-Diff, which uses a pretrained diffusion model to decompose the source images into the degraded images for image restoration. Cao et al. [36] introduced a diffusion model for pansharpening that effectively integrates high-frequency details and spectral information. Zhong et al. [37] separated the learning processes of spatial details and spectral features into distinct branches, and proposed a spatial–spectral integrated diffusion model for pansharpening. However, in pansharpening task, inherent stochasticity of denoising diffusion probabilistic model often leads to the loss of spatial and spectral fidelity from source images [33], and the loss is proportional to information quantity of learning objects. Furthermore, the adding noise modes in diffusion models generally aim at fixed-scale images during the training process, which weakens generalization of model and fusion performance at different scales. One approach to address this issue is to divide the test samples into patches and compute them separately. But this approach can lead to distortion due to the window effect. Another one is to calculate the average of the stacked patches using sliding windows [38], which significantly increases computational complexity compared to the original model.

To address the issues of denoising diffusion probabilistic model in pansharpening, a prior-guided dual-branch diffusion

model (PDDM) is proposed for pansharpening. First, a dual-branch diffusion model structure is constructed, with each branch focusing on the recovery of spectral information and spatial information, respectively. The two branches of the diffusion model are guided by a collaborative loss, enhancing the global perception of spectral and spatial features across these two branches. Meanwhile, adversarial constraints are supervised on the outputs of different branches to maintain the spatial and spectral fidelity. Then, to reduce the uncertainty of generated detail information, two pregeneration modules based on different prior information are designed for pixel-to-pixel reconstruction. Finally, a focus module is established, supervised by a joint multiscale variation detection loss, to fuse the generated features of two branches and improve the generalization of the PDDM at different scales. The contributions of this work are as follows.

- 1) A dual-branch diffusion model, named PDDM is constructed for pansharpening, which can collaboratively generate high-quality details through different information streams for adversarial fusion, ensuring the spatial and spectral fidelity of in each branch.
- 2) To reduce the uncertainty of the generated detail information, two pixel-to-pixel pregeneration modules are established based on spatial and spectral priors, which guide the generation of details in the diffusion process.
- 3) A focus module is constructed to fuse the generated detail information. In addition, a joint multiscale variation detection loss is defined to supervise the focus module to improve the generalization performance of the PDDM.

II. PROPOSED METHOD

In this section, the architecture of PDDM is presented, as shown in Fig. 1, which consists of a dual-branch diffusion model, pregeneration modules, and a focus module. In PDDM, each diffusion branch injects condition features through spatial–spectral information, respectively, which learns from two pregenerated

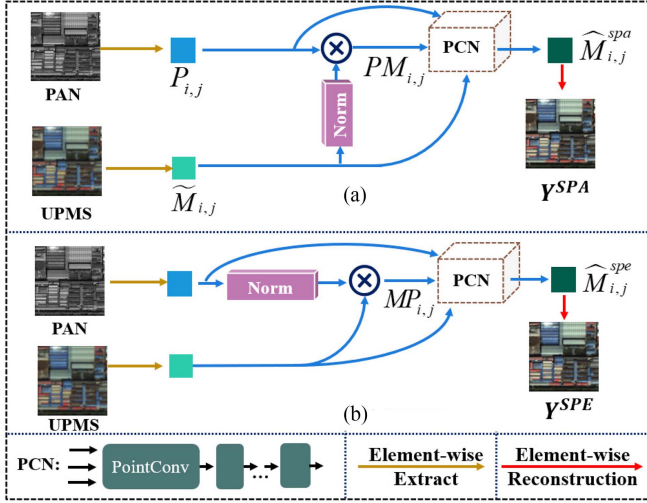


Fig. 2. Architectures of SPAPM and SPEPM. (a) SPAPM. (b) SPEPM.

pixel-to-pixel modules. Guided by prior information, two diffusion branches are constructed to generate fusion features. Finally, a focus module is built to generate the HRMS images. Each component in PDDM is introduced in detail as follows.

A. Construction of Pregeneration Module

To reduce the uncertainty in the generation process of the diffusion model, two pregeneration modules based on different priors are designed, as shown in Fig. 2.

The obtained HRMS image \hat{M} is expected to be unique when the PAN images and UPMS images \tilde{M} are fixed, which can be illustrated by the following probability models:

$$P_{n1}(\hat{M}|\tilde{M}, f(\tilde{M}, P)) = P_{n2}(\hat{M}|P, f(P, \tilde{M})) \quad (1)$$

$$f(a, b) = \text{Norm}(a) * b \quad (2)$$

where $f(\cdot)$ is an operation to obtain the features of input images, $\text{Norm}(\cdot)$ is a normalization operation, and $*$ is a Hardman product. P_{n1} and P_{n2} are two learnable networks.

Based on the probability model, we construct two pregenerated pixel-to-pixel modules according to different interference ways for the two image features, namely, spectral pregeneration module (SPEPM) and spatial pregeneration module (SPAPM). These two modules are used for two diffusion model branches to generate pansharpening results. The specific process is as follows.

In the construction of SPAPM, given a pixel $P_{i,j}$ of PAN image, where i and j are the positional coordinates of the pixel, influenced by the pixel $\tilde{M}_{i,j}$ at the corresponding position in the UPMS image, the pixel $\hat{M}_{i,j}^{spa}$ is generated, as shown in the following formula:

$$\hat{M}_{i,j}^{spa} = f_{\text{SPAPM}}(P_{i,j}, \tilde{M}_{i,j}, PM_{i,j}) \quad (3)$$

where $f_{\text{SPAPM}}(\cdot)$ is the network for preconstructing pixel $\hat{M}_{i,j}^{spa}$, and $PM_{i,j}$ denotes an influence factor that can be obtained by

Algorithm 1: Training Process of the Pregeneration Modules.

Input: The UPMS images \tilde{M} , PAN images P , GTs, and epoch numbers.

Output: The parameters of $f_{\text{SPAPM}}(\cdot)$ and $f_{\text{SPEPM}}(\cdot)$.

for each epoch do

(1) Extract pixel $\tilde{M}_{i,j}$ and $P_{i,j}$ from \tilde{M} and P by elementwise operation.

(2) Input $\tilde{M}_{i,j}$ and $P_{i,j}$ into SPAPM and SPEPM to obtain $\hat{M}_{i,j}$, and then reconstruct $\hat{M}_{i,j}$ to output \hat{M} .

(3) Calculate $\mathcal{L}_{\text{base}} = |GT - \hat{M}|$ for each module.

(4) Update parameters of $f_{\text{SPAPM}}(\cdot)$ and $f_{\text{SPEPM}}(\cdot)$ modules, respectively.

end for

Return: The parameters of $f_{\text{SPAPM}}(\cdot)$ and $f_{\text{SPEPM}}(\cdot)$.

the following formula:

$$PM_{i,j} = \text{Norm}(\tilde{M}_{i,j}) \otimes P_{i,j}. \quad (4)$$

By elementwise multiplication operator \otimes , the interference factors $PM_{i,j}$ are obtained and used as inputs to the SPAPM to obtain reconstruction pixels $\hat{M}_{i,j}^{spa}$.

Similarly, assuming that pixels $\tilde{M}_{i,j}$ in the spectral channels are also influenced by $P_{i,j}$, an SPEPM is constructed to obtain the corresponding $\hat{M}_{i,j}^{spe}$ by the following formulas:

$$\hat{M}_{i,j}^{spe} = f_{\text{SPEPM}}(\tilde{M}_{i,j}, P_{i,j}, MP_{i,j}) \quad (5)$$

$$MP_{i,j} = \text{Norm}(P_{i,j}) \otimes \tilde{M}_{i,j}. \quad (6)$$

The training process of the pregeneration modules is shown in Algorithm 1. By applying two learnable pixel-to-pixel pregeneration modules to the entire image, approximate reconstructed images Y^{SPA} and Y^{SPE} can be quickly obtained, which are used to guide the forward process of the diffusion model.

B. Structure of Diffusion Branch

In the construction of pregeneration modules, although the reconstructed HRMS images Y^{SPA} and Y^{SPE} are close to the corresponding ground-truth (GT) image, the residual features between them are different due to the use of simple interactive interference. Therefore, to increase the consistency of the generated features in feature interaction, two diffusion branches are constructed for generating residual features $Y_{\text{res}}^{\text{SPA}}$ and $Y_{\text{res}}^{\text{SPE}}$ between Y^{SPA} and the GT image, as well as between Y^{SPE} and the GT image, respectively, by increasing the interaction between two image features at different scales, as shown in Fig. 3.

The mathematical derivations of deduction for each branch adhere to the fundamental version of the denoising diffusion probabilistic model. During the forward process, the input image X_0 is known. At each time step, an approximate standard normal distribution noise $\epsilon \in N(0, 1)$ is added to X_0 through the Markov chain process [39], and at time step t , the noisy image

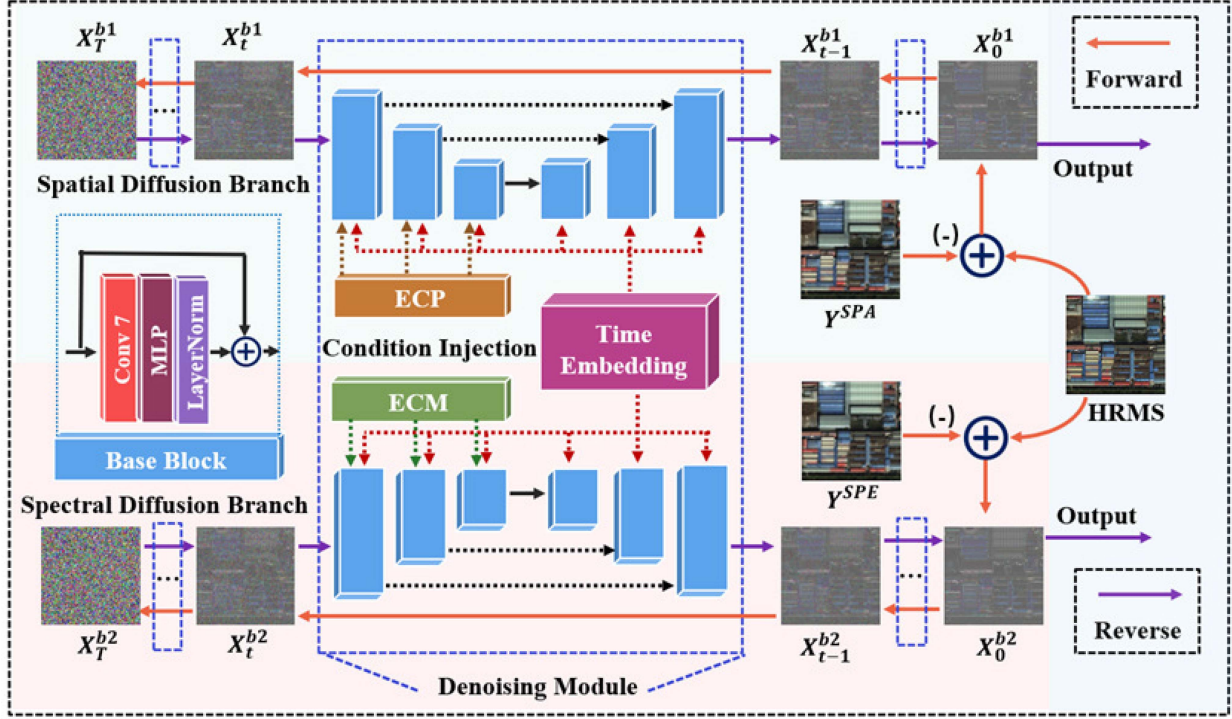


Fig. 3. Architecture of the proposed dual-branch diffusion model. ECP and ECM denote the encoding-PAN module and encoding-MS module, respectively.

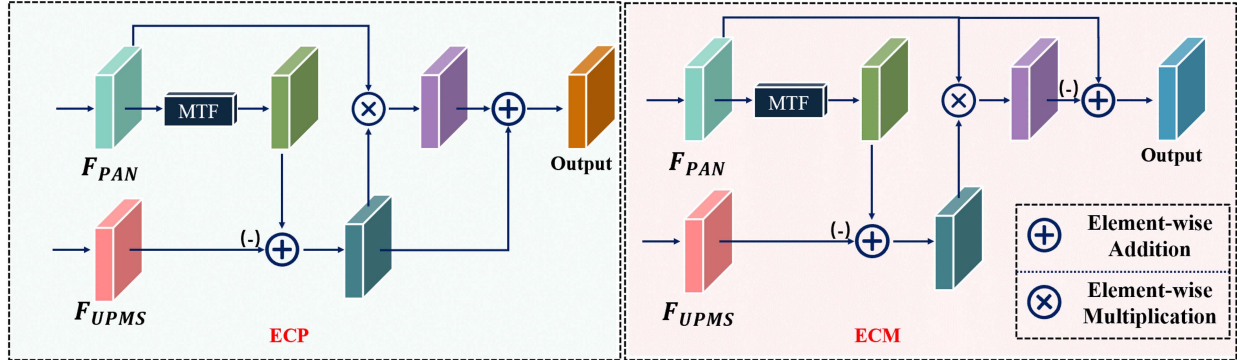


Fig. 4. Architectures of ECP and ECM. MTF denotes MTF, F_{PAN} denotes the features of PAN images, and F_{UPMS} denotes the features of UPMS images.

X_t can be obtained using the following formula:

$$X_t = \sqrt{\bar{\alpha}}X_0 + \sqrt{1 - \bar{\alpha}}\epsilon \quad (7)$$

where $\bar{\alpha}$ is a reparameterization operator for controlling the noise distribution so that the sampling process is easy to differentiate.

Due to the inherent randomness of denoising diffusion probabilistic model, the information of X_t is reduced to improve the stability in generated details. Therefore, in different branches, X_0 is replaced with Y_{res}^{SPA} and Y_{res}^{SPE} , respectively. In addition, as a deterministic conditional reconstruction diffusion model, the reencoded UPMS and PAN image features are injected as condition injections, guiding the recovery of the learning objects during the backward process, as shown in Algorithm 2.

In this work, the encoding-PAN (ECP) module and the encoding-MS (ECM) module are constructed to obtain the condition injections, as shown in Fig. 4. In these modules,

features of PAN images are first smoothed through a modulation transfer function (MTF) [40]. Then, by calculating the analogous spectral differences between the features of UPMS and smoothed PAN images, the spatial features of PAN images are reassigned to get the condition injections. The specific computing processes are as follows:

$$cd_1 = f_{ECP}(\tilde{M}, P) \quad (8)$$

$$cd_2 = f_{ECM}(\tilde{M}, P) \quad (9)$$

where cd_1 and cd_2 represent the condition injections of the dual-branch diffusion model; $f_{ECP}(\cdot)$ and $f_{ECM}(\cdot)$ denote the ECP and ECM modules, respectively.

The time embedding step is used to make the network aware of the current time step t . The noise images at different time

Algorithm 2: Training Process of the Dual-diffusion Branch.

Input: The \widetilde{M} , P , GTs, epoch number, T , parameters of $f_{\text{SPAPM}}(\cdot)$ and $f_{\text{SPEPM}}(\cdot)$.

Output: Generated images \widehat{X}_0^{b1} and \widehat{X}_0^{b2} .

for each epoch **do**

- (1) Sample $t \sim \text{uniform}(\{1, \dots, T\})$.
- (2) Sample $\epsilon \sim \mathcal{N}(0, I)$.
- (3) Compute Y^{SPA} and Y^{SPE} by (3) and (5).
- (4) Compute residual images $X_0^{b1} = GT - Y^{\text{SPA}}$ and $X_0^{b2} = GT - Y^{\text{SPE}}$, respectively.
- (5) Input \widetilde{M} and P into $f_{\text{ECP}}(\cdot)$ and $f_{\text{ECM}}(\cdot)$ to obtain cd_1 and cd_2 .
- (6) Take gradient descent step on loss $\mathcal{L}_{b1} + \mathcal{L}_{\text{col}}$ and $\mathcal{L}_{b2} + \mathcal{L}_{\text{col}}$, respectively.
- (7) **if** epoch mod 100 = 0 **then**
 - a. Sample X_T^{b1} and $X_T^{b2} \sim \mathcal{N}(0, I)$.
 - b. **for** $t \leftarrow T$ **to** 1 **do**

Output \widehat{X}_{t-1}^{b1} and \widehat{X}_{t-1}^{b2} by using denoising module.
 - end for**
 - c. Output \widehat{X}_0^{b1} and \widehat{X}_0^{b2} .
 - d. Take gradient descent step on loss \mathcal{L}_{adv} .
 - e. Update parameters of $f_{\text{SPAPM}}(\cdot)$ and $f_{\text{SPEPM}}(\cdot)$ modules, respectively.

end for

Return: The latest values of \widehat{X}_0^{b1} and \widehat{X}_0^{b2} .

Algorithm 3: Training Process of the Focus Module.

Input: epoch number, \widehat{X}_0^{b1} , \widehat{X}_0^{b2} , and GTs.

Output: pansharpened image Y^{output} .

for each epoch **do**

- (1) Input \widehat{X}_0^{b1} and \widehat{X}_0^{b2} into the focus module and then output Y^{output} .
- (2) Take gradient descent step on loss \mathcal{L}_{msv} .
- (3) Update parameters of the focus module.

end for

Return: pansharpened image Y^{output} .

steps can be used to train a U-shaped denoising network. The losses of the branches $b1$ and $b2$ are as follows:

$$\mathcal{L}_{b1}(\theta) = E \left[\|\epsilon - \epsilon_\theta (\sqrt{\bar{\alpha}_t} Y_{\text{res}}^{\text{SPA}} + \sqrt{1 - \bar{\alpha}_t} \epsilon, \text{cd}_1, t) \|^2 \right] \quad (10)$$

$$\mathcal{L}_{b2}(\theta) = E \left[\|\epsilon - \epsilon_\theta (\sqrt{\bar{\alpha}_t} Y_{\text{res}}^{\text{SPE}} + \sqrt{1 - \bar{\alpha}_t} \epsilon, \text{cd}_2, t) \|^2 \right] \quad (11)$$

where θ represents parameters of denoising network. By constantly searching for the optimal network parameters, the diffusion model can generate learning objects from pregeneration modules.

C. Collaborative and Adversarial Losses

Each branch of PDDM focuses on spatial and spectral information, respectively. To interact and complement output details

in both spatial and spectral domains, a collaboration loss and an adversarial loss are proposed.

First, at each time step t , the diffusion branches from different information streams learn synchronously. Ideally, each branch should be able to reconstruct the HRMS image. However, different injection methods make each branch perceive spectral and spatial information with different intensity and generate different results. To enhance the interaction of the two branches, a collaborative loss is introduced during each synchronization process to perceive different information

$$\mathcal{L}_{\text{col}} = \|\mathcal{L}_{b1} - \mathcal{L}_{b2}\| \quad (12)$$

where \mathcal{L}_{col} is used in reverse process and works with \mathcal{L}_{b1} or \mathcal{L}_{b2} in different branches, promoting interaction and collaboration between the branches, leading to improved training effectiveness of the network.

Second, since the approximate reconstruction HRMS image can be directly obtained by the generation features of the diffusion model adding with $Y_{\text{res}}^{\text{SPA}}$ and $Y_{\text{res}}^{\text{SPE}}$, an adversarial loss is proposed to constrain the fused result. By monitoring the intermediate results and making them compete with each other, the ability of the pregeneration module is improved to interact with different information. The adversarial loss \mathcal{L}_{adv} is defined as follows:

$$\mathcal{L}_{\text{adv}} = \|X_0^{b1} + Y^{\text{SPA}}\| - \|X_0^{b2} + Y^{\text{SPE}}\| \quad (13)$$

where \mathcal{L}_{adv} is used to adversarially constrain the fused result, causing the updated residuals to produce an HRMS image.

D. Focus Module

To enhance the generalization of PDDM in the pansharpening task at different scales, we build a focus module to fuse the outputs X_0^{b1} and X_0^{b2} from the two branches. The focus module is constructed as a U-shaped connection structure. Besides, to improve the fusion performance of the model across different scales, a multiscale variation detection loss function is defined as follows:

$$\mathcal{L}_{\text{msv}} = |\text{Down}_n(GT) - \text{Down}_n(Y^{\text{output}})| \quad (14)$$

where $\text{Down}(\cdot)$ denotes a downsampling method, and n is an integer scale factor for downsampling. This loss is used to guide the fusion process of the focus module at different scales, improving the generalization of PDDM, as shown in Algorithm 3.

III. EXPERIMENTAL RESULTS AND ANALYSIS

To assess the efficacy of the proposed PDDM, both subjective and objective experiments were conducted on simulated and real satellite datasets, including IKONOS (four bands), Pléiades (four bands), and WorldView-3 (eight bands). The specific satellite parameters are shown in Table I. In the simulated dataset, LRMS images were generated based on the Wald's protocol [40], [41] using MTF and downsampling operations applied to HRMS images. The LRMS and PAN images have dimensions of 64×64 and 256×256 , respectively. For the real datasets, the LRMS and PAN images have dimensions of 200×200 and 800×800 , respectively. In this work, t is set to 500 empirically [33].

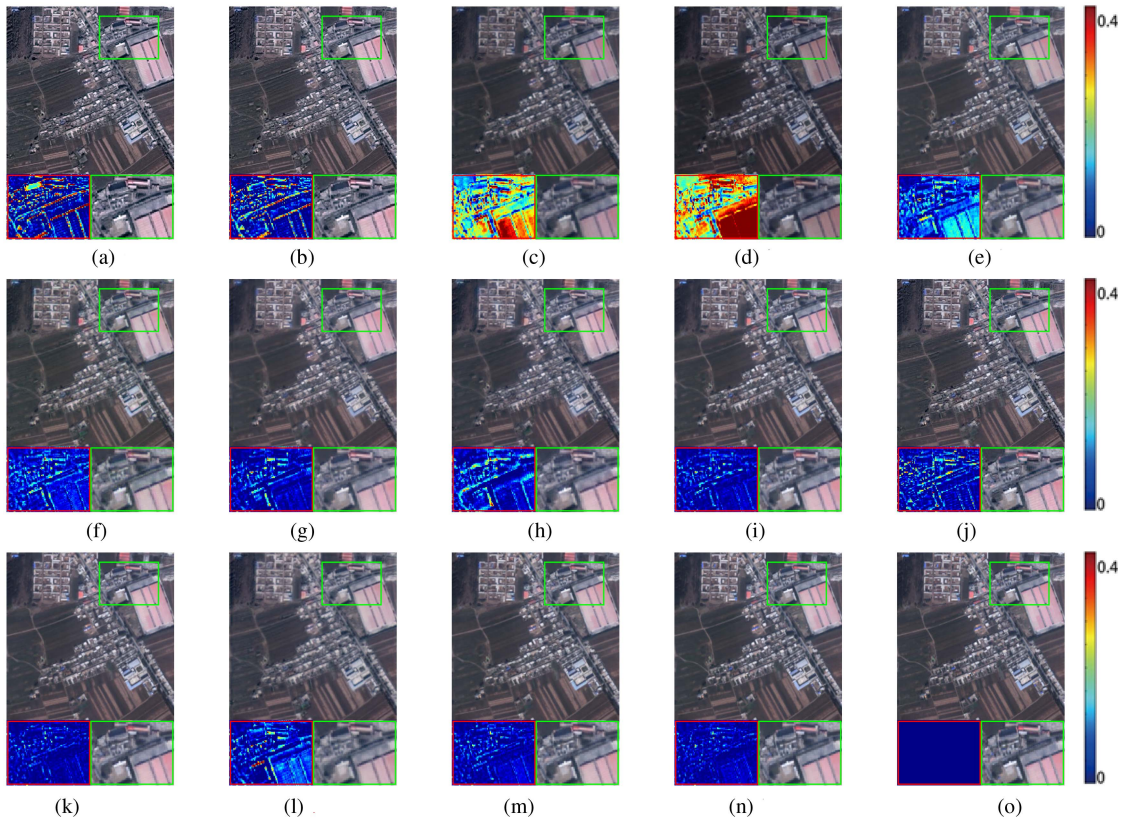


Fig. 5. Comparison of pansharpened images on simulated data from the IKONOS dataset. (a) GSA [16]. (b) GLP-REG [13]. (c) DRPNN [14]. (d) MSDCNN [19]. (e) PCDRN [42]. (f) TFNet [24]. (g) FusionNet [25]. (h) TDNet [30]. (i) AWFLN [31]. (j) LRTCFFPan [20]. (k) VOGTNet [26]. (l) PLR-Diff [34]. (m) Dif-PAN [36]. (n) PDDM (ours). (o) GT.

TABLE I
SPECIFICATIONS OF THE DATASETS

Sensors	Pléiades	IKONOS	WorldView-3
MS/PAN resolutions	0.5/2.0(m)	0.82/3.2(m)	0.31/1.24(m)
MS/PAN sizes (Simulated)	64×64×4 256×256×4	64×64×4 256×256×4	64×64×8 256×256×8
MS/PAN sizes (Real)	200×200×4 800×800×4	200×200×4 800×800×4	200×200×8 800×800×8
MS bands	Red(R), green(G), blue(B), and near infrared (NIR)	R, G, B, and NIR	R, G, B, NIR1, coastal blue, yellow, red edge, and NIR2

During the experiments, the performance of the proposed PDDM was compared with that of state-of-the-art methods. The comparative methods include GSA [16], GLP-REG [13], DRPNN [14], MSDCNN [19], PCDRN [42], TFNet [24], FusionNet [25], TDNet [30], AWFLN [31], LRTCFFPan [20], VOGTNet [26], PLR-Diff [34], and Dif-PAN [36].

Notably, all DL-based methods were retrained using the same datasets to ensure fairness and were tested on the NVIDIA GeForce RTX 3090 and INTEL 11700K hardware environment.

A. Experiments on Simulated Dataset

As shown in Fig. 5, the subjective fusion images of various comparison methods were examined using a pair of images from

the IKONOS dataset. Obviously, the fusion results of FusionNet, MSDCNN, DRPNN, PCDRN, TDNet, AWFLN, TFNet, and PLR-Diff exhibit more blurred edges compared to GTs. The pansharpened images of GSA and GLP-REG perform an over-injection phenomenon in the edge. The DRPNN and MSDCNN methods show significant spectral distortion, while our proposed method produces fusion results that are more closed to GTs. To provide a clearer demonstration of the differences between the fusion results and GTs, residual maps were calculated and displayed by an enlarged local area beneath each result. The residual maps clearly indicate that the comparison methods suffer from noticeable spectral distortion and loss of spatial details. In contrast, our method exhibits the least residual information, further confirming the effectiveness of the proposed approach.

Fig. 6 presents a group of pansharpened images of all comparison methods on the Pléiades dataset. From the figure, it can be observed that most comparison methods exhibit varying degrees of edge distortion in their fused results. It is evident that both VOGTNet and our method successfully reconstruct almost all curved edges. Compared to the result of VOGTNet, our result has less residual map and is closer to GT.

Fig. 7 displays a group of pansharpened results from various comparison methods on the WorldView-3 dataset. It can be clearly seen from the figure that all methods perform well in reconstructing the gray areas. However, the fusion results of FusionNet, MSDCNN, DRPNN, PCDRN, TDNet, AWFLN, TFNet, and PLR-Diff exhibit varying degrees of spectral and spatial distortions in the white street areas. In contrast, our

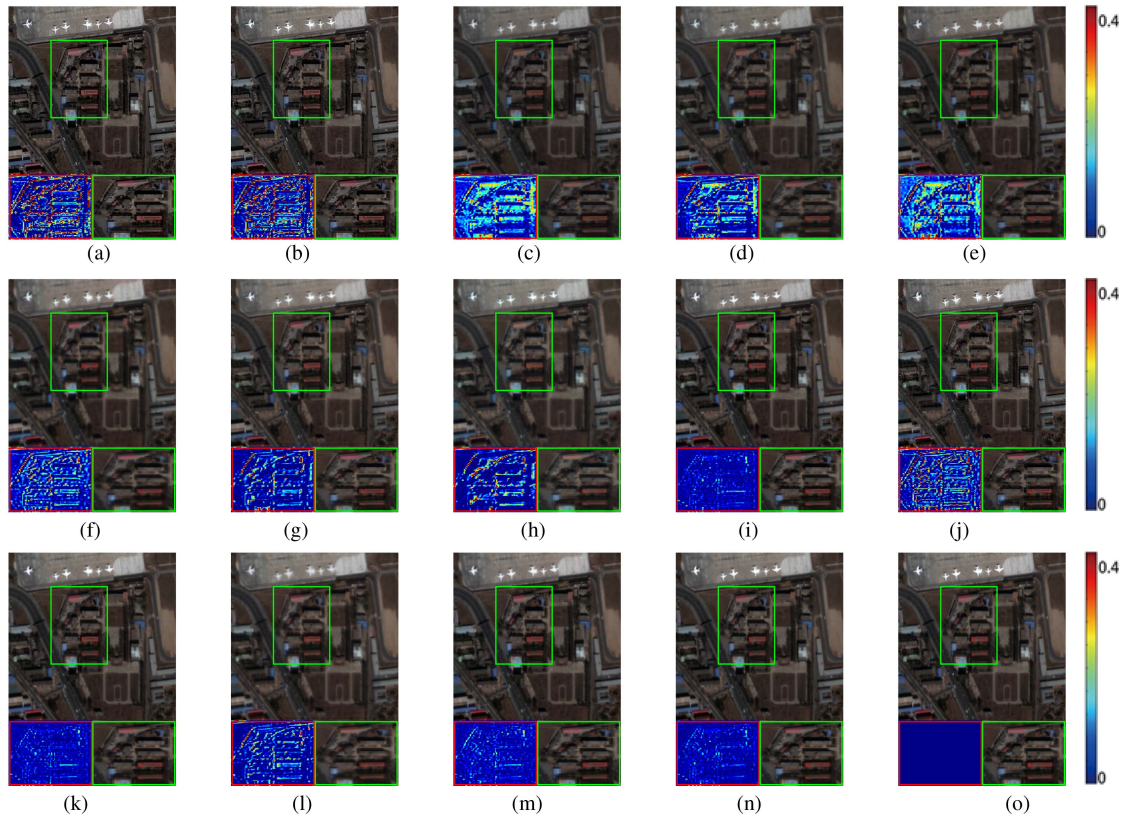


Fig. 6. Comparison of pansharpened images on simulated data from the Pléiades dataset. (a) GSA [16]. (b) GLP-REG [13]. (c) DRPNN [14]. (d) MSDCNN [19]. (e) PCDRN [42]. (f) TFNet [24]. (g) FusionNet [25]. (h) TDNet [30]. (i) AWFLN [31]. (j) LRTCFPan [20]. (k) VOGTNet [26]. (l) PLR-Diff [34]. (m) Dif-PAN [36]. (n) PDDM (ours). (o) GT.

method produces a reconstruction result that is closest to the GT. Furthermore, our result has the least residues compared to the results of other methods.

The evaluation metrics, including peak signal-to-noise ratio (PSNR), root mean square error, relative average spectral error, universal image quality index (UIQI), $Q2^n$, spectral angle mapper (SAM), erreur relative globale adimensionnelle de synthèse (ERGAS), and spatial cross correlation (SCC) were employed to objectively assess the performance of the different methods [43], [44]. Table II shows the quantitative evaluation metrics of each method across all three datasets, where the best results are highlighted in bold and the second-best results are underlined.

Table II shows that VOGTNet achieves superior objective results compared to GSA, GLP, and LRTCFPan among the traditional methods. In DL-based methods, AWFLN, Dif-PAN, and our PDDM perform better than others. In diffusion model-based methods, PDDM performs better than PLR-Diff and Dif-PAN. In summary, the results in Table II indicate that our method achieves the best performance on all the three datasets.

B. Experiments on Real Dataset

Fig. 8 presents the fusion results of a pair of images from the IKONOS dataset. To better distinguish the spectral and detail differences between the fusion results, two small regions were selected and magnified. As shown in the figure, compared to the

fusion results of other methods, GSA and GLP-REG methods fail to restore the spatial features of the PAN image adequately. The results of TFNet, TDNet, FusionNet, PLR-Diff, and Dif-PAN methods exhibit more serious color distortion than those of other methods. VOGTNet and our PDDM successfully recovered distinct texture features. However, compared to VOGTNet, our method presents clearer spatial details.

Fig. 9 shows a set of fusion results of various methods on the Pléiades dataset. The figure reveals that the results of most methods exhibit clear spatial texture details. Compared to other methods, AWFLN, LRTCFPan, VOGTNet, and our PDDM display abundant color information. However, the color information presented by PDDM is closest to that of the UPMS image.

Fig. 10 shows a pair of pansharpened images from the WorldView-3 datasets. As can be seen from the figures, the results of the GSA method have the problem of over-injection details and lack of spectral information. Compared with other results, the results of MSDCNN, DRPNN, and PCDRN methods display distinct color distortions. It can be clearly observed from the enlarged areas that our result has more abundant information than those of other comparison methods.

In the absence of GT, nonreference quantitative metrics, D_{λ}^k , D_s , and HQNR [45] are employed to evaluate the similarity of spectral and spatial details between the fusion images and source images. D_{λ}^k quantifies the spectral similarity between the fusion results and LRMS images, while D_s measures the spatial

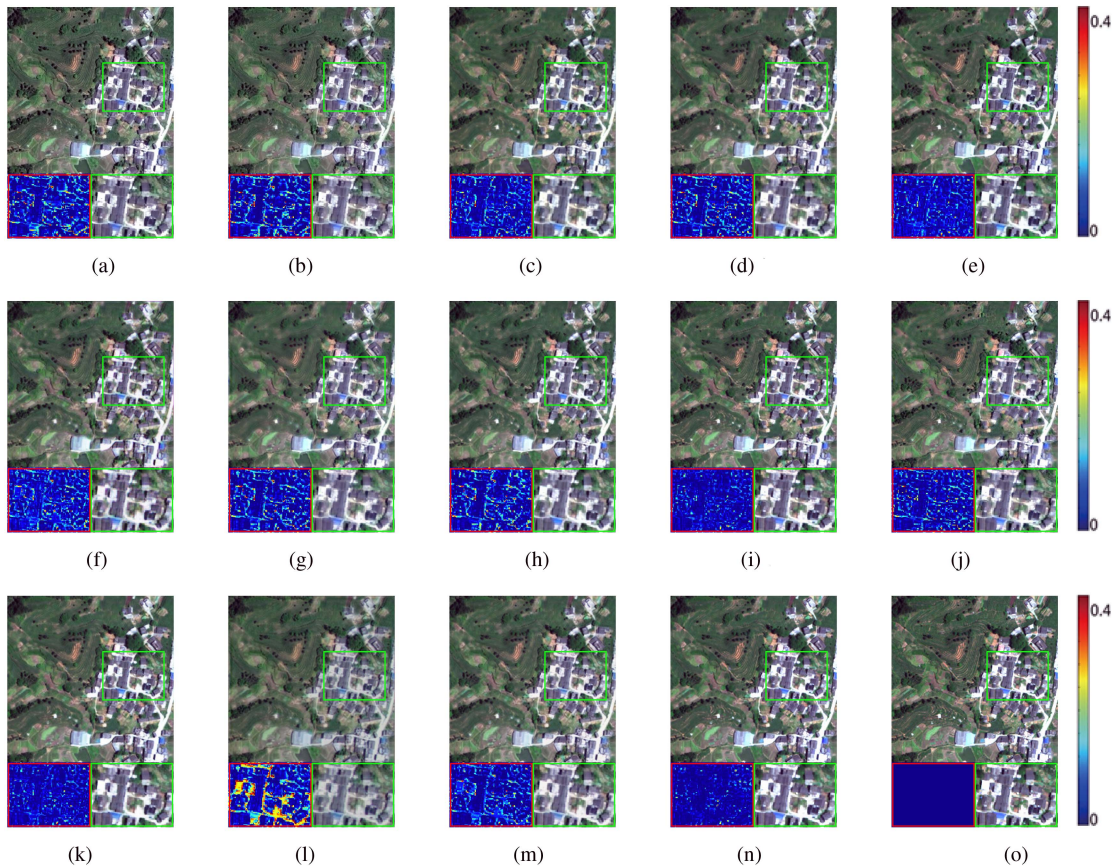


Fig. 7. Comparison of pansharpened images on simulated data from the WorldView-3 dataset. (a) GSA [16]. (b) GLP-REG [13]. (c) DRPNN [14]. (d) MSDCNN [19]. (e) PCDRN [42]. (f) TFNet [24]. (g) FusionNet [25]. (h) TDNet [30]. (i) AWFLN [31]. (j) LRTCFFPan [20]. (k) VOGTNet [26]. (l) PLR-Diff [34]. (m) Dif-PAN [36]. (n) PDDM (ours). (o) GT.

similarity between the fusion results and PAN images. HQNR calculates the overall similarity by incorporating both D_{λ}^k , D_s .

As presented in Table III, the traditional method LRTCFFPan outperforms several DL-based methods, such as DRPNN, MSDCNN, PCDRN, TFNet, PLR-Diff, and Dif-PAN, in the QNR index. Among the diffusion model-based methods, PLR-Diff and Dif-PAN perform worse than PDDM. One possible explanation is that the fixed-size noise addition approach in diffusion models limits the generalization of models at different scales. Among all comparison methods, our proposed PDDM achieves the highest QNR scores, indicating its superior ability to effectively preserve both spectral and spatial fidelity.

In summary, the proposed PDDM achieves advanced performance in both subjective visual effects and quantitative metrics compared with other methods.

C. Ablation Study

To verify the performance of each component of the proposed method, we conducted some ablation studies on the simulated IKONOS dataset to explore the effectiveness of the various components to the overall results.

1) *Effect of Structure of PDDM*: In the experiments of validating the PDDM structure, five models containing different modules are tested. Model 0 is only a single diffusion model

with neither any branch nor the proposed module. Models 1–3 are various combinations of the leave-one-out module approach, respectively.

Fig. 11 shows two examples of the pansharpened images on five different models. It is obvious that the results of Model 0 (without any proposed component) retain the most residues in the magnified rectangles. Our results are closer to GT images than those of other models, as shown in the magnified rectangles. Furthermore, Table IV shows the quantitative indexes obtained by the five models. It can be seen that the indicators obtained by our method with all modules are better than those of Models 0–3. Therefore, the performance of each module is verified.

2) *Effect of Loss Functions*: Loss functions play a key role in network training. To prove the performance of each loss in PDDM, the ablation experiments using different losses are performed. Model 4 only has the basic loss $\mathcal{L}_{\text{base}}$, which is supervised by HRMS images. Model 5 adds a collaborative loss \mathcal{L}_{col} to the previous Model 4 and Model 6 adds an adversarial loss \mathcal{L}_{adv} to Model 5. Fig. 12 displays two examples of the pansharpened images on different loss functions. It can be seen that our method with all loss functions can better maintain the details and spectral features of the source images to the fused results, and our results have the least residual information. In addition, Table V shows the objective results on different loss functions. From the table, we can find that the proposed model

TABLE II
AVERAGE QUANTITATIVE RESULTS ON THE SIMULATED DATA FROM IKONOS, PLÉIADES, AND WORLDVIEW-3 DATASETS

Sensors	Methods	PSNR(\uparrow)	UIQI(\uparrow)	$Q2^n$ (\uparrow)	SAM(\downarrow)	ERGAS(\downarrow)	SCC(\uparrow)
IKONOS	GSA [16]	27.0956	0.9352	0.8363	4.6886	3.9925	0.8956
	GLP-REG [13]	27.2915	0.9378	0.8409	4.5549	3.7941	0.8974
	DRPNN [14]	26.9798	0.9273	0.8154	4.0383	3.7677	0.9464
	MSDCNN [19]	27.1694	0.9310	0.8239	4.0121	3.6587	0.9463
	PCDRN [42]	28.8464	0.9516	0.8650	3.5955	3.0352	0.9616
	TFNet [24]	30.9652	0.9702	0.9227	3.2528	2.4465	0.9485
	FusionNet [25]	28.1699	0.9561	0.8763	4.4868	4.0201	0.9394
	TDNet [30]	29.6391	0.8783	0.8737	3.7885	3.0256	0.9235
	AWFLN [31]	34.0407	0.9812	0.9522	2.3388	1.6913	0.9600
	LRTCFPan [20]	29.4280	0.9598	0.9006	3.5787	2.9326	0.9283
	VOGTNet [26]	<u>34.2703</u>	<u>0.9856</u>	<u>0.9542</u>	<u>2.2067</u>	1.6513	<u>0.9740</u>
	PLR-Diff [34]	31.7013	0.9345	0.8302	2.8545	2.2666	0.9642
Dif-PAN [36]	34.1767	0.9850	0.9537	2.2686	<u>1.6027</u>	0.9702	
PDDM (ours)	34.6573	0.9870	0.9581	2.1705	1.5412	0.9759	
Pléiades	GSA [16]	27.9760	0.9530	0.8757	3.5046	3.5725	0.9118
	GLP-REG [13]	27.9586	0.9522	0.9067	3.4788	3.5613	0.9115
	DRPNN [14]	30.1216	0.9680	0.8636	2.6809	2.9884	0.9601
	MSDCNN [19]	29.2428	0.9629	0.8638	2.7117	3.2713	0.9432
	PCDRN [42]	31.5914	0.9767	0.9004	2.3488	2.6322	0.9739
	TFNet [24]	32.8707	0.9805	0.9452	2.9289	2.0643	0.9568
	FusionNet [25]	31.3848	0.9794	0.9420	2.5486	2.5394	0.9209
	TDNet [30]	31.3226	0.9164	0.9117	3.5946	2.6642	0.9340
	AWFLN [31]	35.0748	0.9883	0.9652	2.2029	1.5431	0.9545
	LRTCFPan [20]	30.1723	0.9697	0.9195	2.8051	2.7336	0.9110
	VOGTNet [26]	<u>35.8613</u>	<u>0.9898</u>	<u>0.9684</u>	<u>2.1842</u>	<u>1.3394</u>	<u>0.9803</u>
	PLR-Diff [34]	33.5731	0.9809	0.9586	2.4908	1.8425	0.9430
Dif-PAN [36]	35.7482	0.9887	0.9661	2.1937	1.4311	0.9715	
PDDM (ours)	36.2443	0.9922	0.9708	2.1010	1.2594	0.9853	
WorldView-3	GSA [16]	27.2705	0.9358	0.8688	6.6912	4.3425	0.9136
	GLP-REG [13]	27.1812	0.9327	0.8630	6.4110	4.3466	0.9146
	DRPNN [14]	25.6762	0.9224	0.8152	7.9339	6.1439	0.9518
	MSDCNN [19]	25.2981	0.9171	0.8033	7.7049	6.2930	0.9370
	PCDRN [42]	27.8324	0.9460	0.8681	5.3713	4.0870	0.9628
	TFNet [24]	28.2917	0.9047	0.8921	5.3725	3.7714	0.9335
	FusionNet [25]	27.2539	0.8999	0.8818	5.1426	4.5063	0.9258
	TDNet [30]	30.1461	0.8936	0.8504	5.3467	3.7361	0.9330
	AWFLN [31]	31.7700	0.9712	0.9352	4.8546	2.7522	0.9651
	LRTCFPan [20]	28.6875	0.9549	0.8987	5.3327	3.6266	0.9137
	VOGTNet [26]	<u>31.7854</u>	<u>0.9727</u>	<u>0.9365</u>	<u>4.1002</u>	<u>2.8223</u>	<u>0.9664</u>
	PLR-Diff [34]	30.7318	0.9319	0.8582	5.2807	3.5260	0.9425
Dif-PAN [36]	31.3554	0.9629	0.9323	4.6597	2.9882	0.9642	
PDDM (ours)	32.0058	0.9867	0.9395	4.0620	2.5707	0.9787	

The best quantitative results are marked with bold fonts, and the second-best quantitative results are underlined.

TABLE III
AVERAGE QUANTITATIVE RESULTS ON THE REAL DATA FROM IKONOS, PLÉIADES, AND WORLDVIEW-3 DATASETS

Method	IKONOS			Pléiades			WorldView-3		
	D_λ^k (\downarrow)	D_S (\downarrow)	HQNR(\uparrow)	D_λ^k (\downarrow)	D_S (\downarrow)	HQNR(\uparrow)	D_λ^k (\downarrow)	D_S (\downarrow)	HQNR(\uparrow)
GSA [16]	0.0479	0.1239	0.8369	0.0621	0.1619	0.7816	0.0405	0.1195	0.8471
GLP-REG [13]	0.0226	0.0941	0.8859	0.0185	0.0990	0.8807	0.0273	0.0903	0.8859
DRPNN [14]	0.0505	0.0793	0.8761	0.0377	0.0363	0.9289	0.1118	0.0869	0.8123
MSDCNN [19]	0.0421	0.0915	0.8705	0.0402	0.0462	0.9327	0.1161	0.0944	0.8073
PCDRN [42]	0.0301	0.0849	0.8863	0.0247	0.0382	0.9379	0.0453	0.0650	0.8909
TFNet [24]	0.0121	0.0609	0.9272	0.0080	0.0310	0.9589	0.0352	0.0464	0.9194
FusionNet [25]	0.0187	0.0637	0.9190	0.0275	0.0285	0.9431	0.0308	0.0468	0.9241
TDNet [30]	0.0174	0.0557	0.9280	0.0116	0.1006	0.8901	0.0211	0.0870	0.9026
AWFLN [31]	0.0110	<u>0.0505</u>	0.9397	0.0113	<u>0.0233</u>	0.9604	0.0202	0.0470	0.9359
LRTCFPan [20]	<u>0.0093</u>	0.0529	0.9389	<u>0.0064</u>	0.0426	0.9513	<u>0.0199</u>	<u>0.0332</u>	<u>0.9474</u>
VOGTNet [26]	0.0110	0.0465	<u>0.9432</u>	0.0101	0.0237	<u>0.9630</u>	<u>0.0251</u>	<u>0.0326</u>	0.9455
PLR-Diff [34]	0.0236	0.0762	0.9022	0.0246	0.0514	0.9248	0.0359	0.0483	0.9192
Dif-PAN [36]	0.0271	0.0659	0.9076	0.0289	0.0420	0.9276	0.0381	0.0420	0.9234
PDDM (ours)	0.0076	0.0448	0.9481	0.0056	0.0218	0.9652	0.0153	0.0296	0.9531

The best quantitative results are marked with bold fonts, and the second-best quantitative results are underlined.

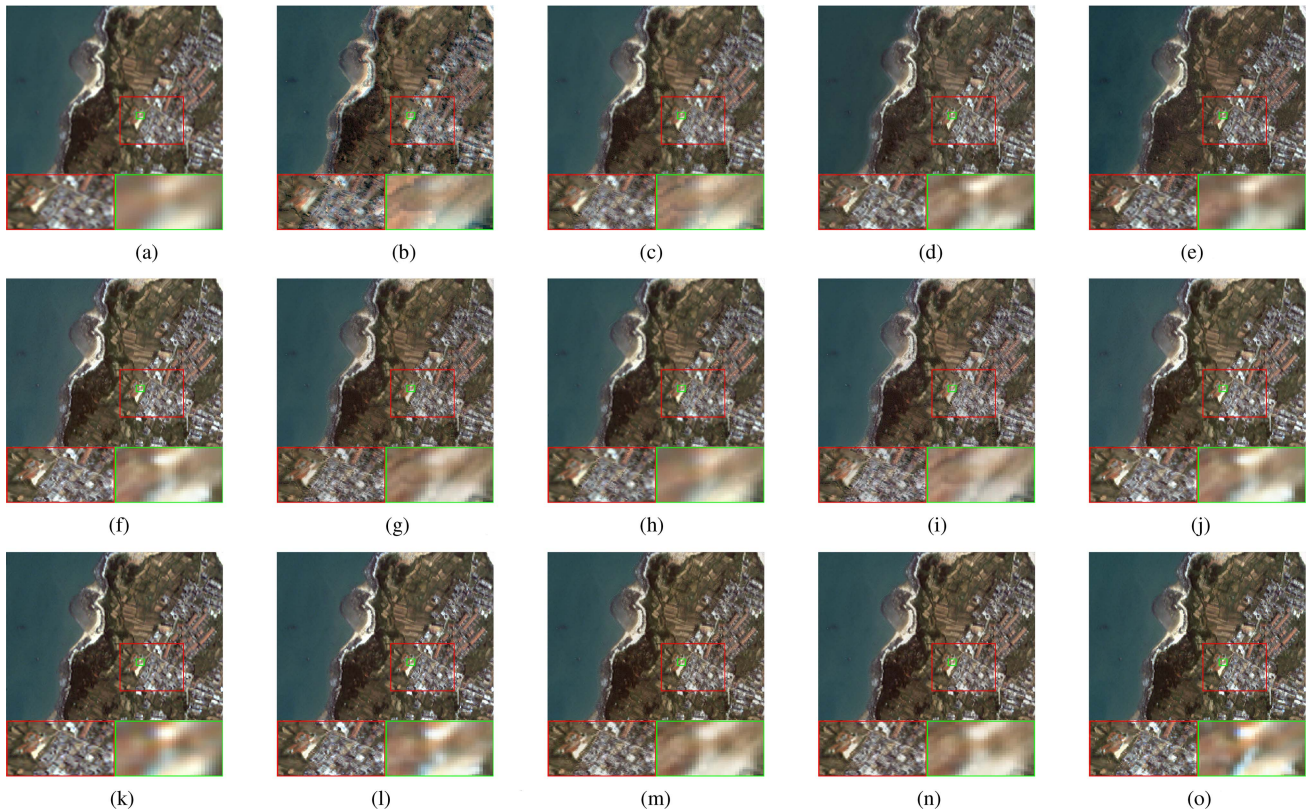


Fig. 8. Comparison of pansharpened images on real data from the IKONOS dataset. (a) UPMS. (b) GSA [16]. (c) GLP-REG [13]. (d) DRPNN [14]. (e) MSDCNN [19]. (f) PCDRN [42]. (g) TFNet [24]. (h) FusionNet [25]. (i) TDNet [30]. (j) AWFLN [31]. (k) LRTCf [20]. (l) VOGTNet [26]. (m) PLR-Diff [34]. (n) Dif-PAN [36]. (o) PDDM (ours).

TABLE IV
ABLATION STUDY RESULTS OF THE STRUCTURE OF PDDM ON THE SIMULATED DATA FROM IKONOS DATASET

	Dual-Branch	Pre-generation Module	Focus Module	PSNR(\uparrow)	UIQI(\uparrow)	SAM(\downarrow)	ERGAS(\downarrow)
Model 0	×	×	×	32.7458	0.9792	2.5377	1.9578
Model 1	×	✓	✓	33.5810	0.9834	2.4052	1.8133
Model 2	✓	×	✓	33.6755	0.9839	2.3975	1.7430
Model 3	✓	✓	×	34.3823	0.9857	2.2280	1.6040
Proposed	✓	✓	✓	34.6573	0.9870	2.1705	1.5412

The best quantitative results are marked with bold fonts.

TABLE V
ABLATION EXPERIMENT OF EACH LOSS FUNCTION ON THE SIMULATED DATA FROM IKONOS DATASET

Methods	\mathcal{L}_{base}	\mathcal{L}_{col}	\mathcal{L}_{adv}	\mathcal{L}_{msv}	PSNR(\uparrow)	UIQI(\uparrow)	SAM(\downarrow)	ERGAS(\downarrow)
Model 4	✓	×	×	×	33.4460	0.9823	2.3975	1.8027
Model 5	✓	✓	×	×	34.1438	0.9848	2.2658	1.6528
Model 6	✓	✓	✓	×	34.4411	0.9862	2.1862	1.5429
Proposed	✓	✓	✓	✓	34.6573	0.9870	2.1705	1.5412

The best quantitative results are marked with bold fonts.

with all losses obtains the best performance than other models, which indicates the effectiveness of the loss functions in PDDM.

3) *Effect of the Proposed Injection Conditions*: The injection conditions play an important role in the diffusion branches, which are obtained by ECP and ECM in our work. To demonstrate the effectiveness of the proposed injection conditions, we built two other models that use different structures to obtain

injection conditions for comparison. Model 7 directly concatenates UPMS and PAN images as injection conditions for the diffusion model. Model 8 uses convolution residual blocks to generate injection conditions.

Fig. 13 shows two examples of the pansharpened images on different injection conditions. It clearly shows that the results of Model 7 have more spectral and spatial distortions than those

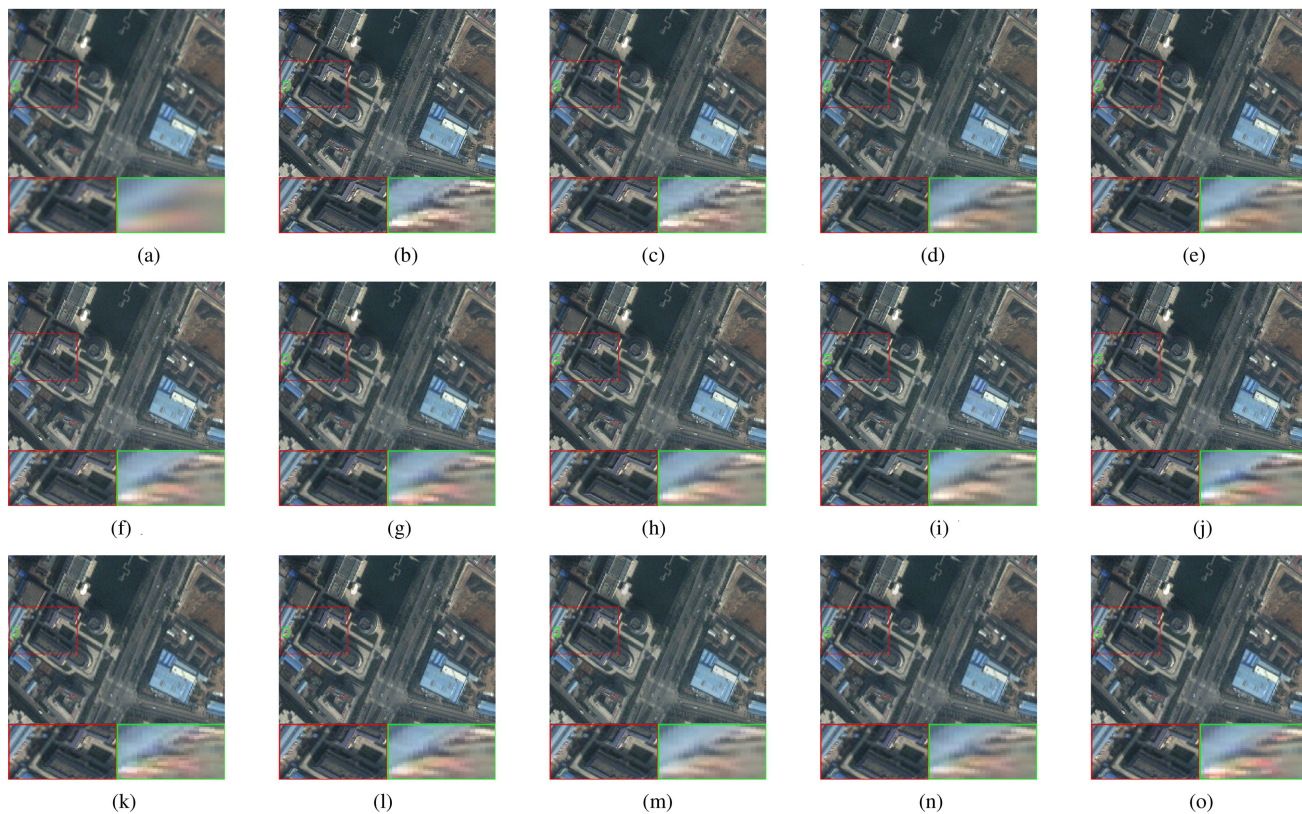


Fig. 9. Comparison of pansharpened images on real data from the Pléiades dataset. (a) UPMS. (b) GSA [16]. (c) GLP-REG [13]. (d) DRPNN [14]. (e) MSDCNN [19]. (f) PCDRN [42]. (g) TFNet [24]. (h) FusionNet [25]. (i) TDNet [30]. (j) AWFLN [31]. (k) LRTCF [20]. (l) VOGTNet [26]. (m) PLR-Diff [34]. (n) Dif-PAN [36]. (o) PDDM (ours).

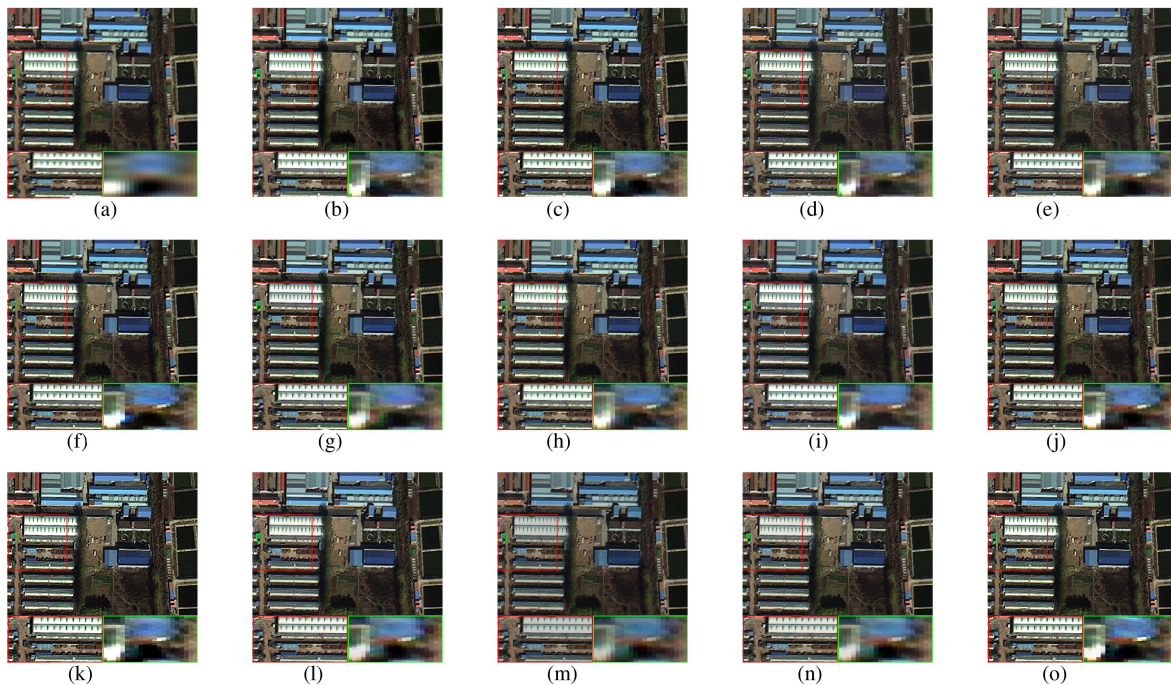


Fig. 10. Comparison of pansharpened images on real data from the WorldView-3 dataset. (a) UPMS. (b) GSA [16]. (c) GLP-REG [13]. (d) DRPNN [14]. (e) MSDCNN [19]. (f) PCDRN [42]. (g) TFNet [24]. (h) FusionNet [25]. (i) TDNet [30]. (j) AWFLN [31]. (k) LRTCF [20]. (l) VOGTNet [26]. (m) PLR-Diff [34]. (n) Dif-PAN [36]. (o) PDDM (ours).

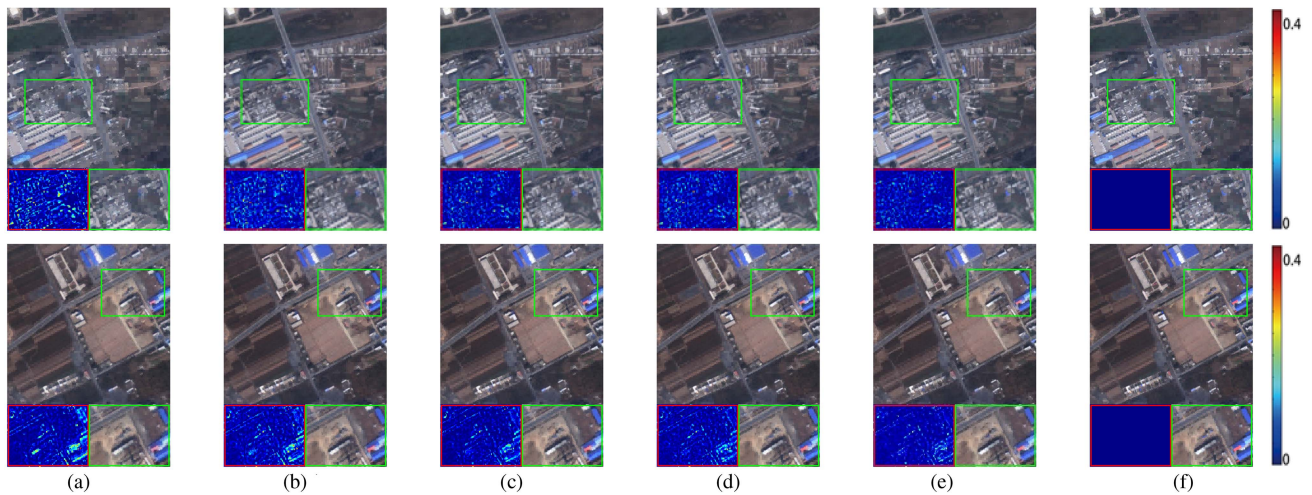


Fig. 11. Comparison of pansharpened images by the models with different structures on simulated data from IKONOS dataset. (a) Model 0. (b) Model 1. (c) Model 2. (d) Model 3. (e) Proposed. (f) GT.

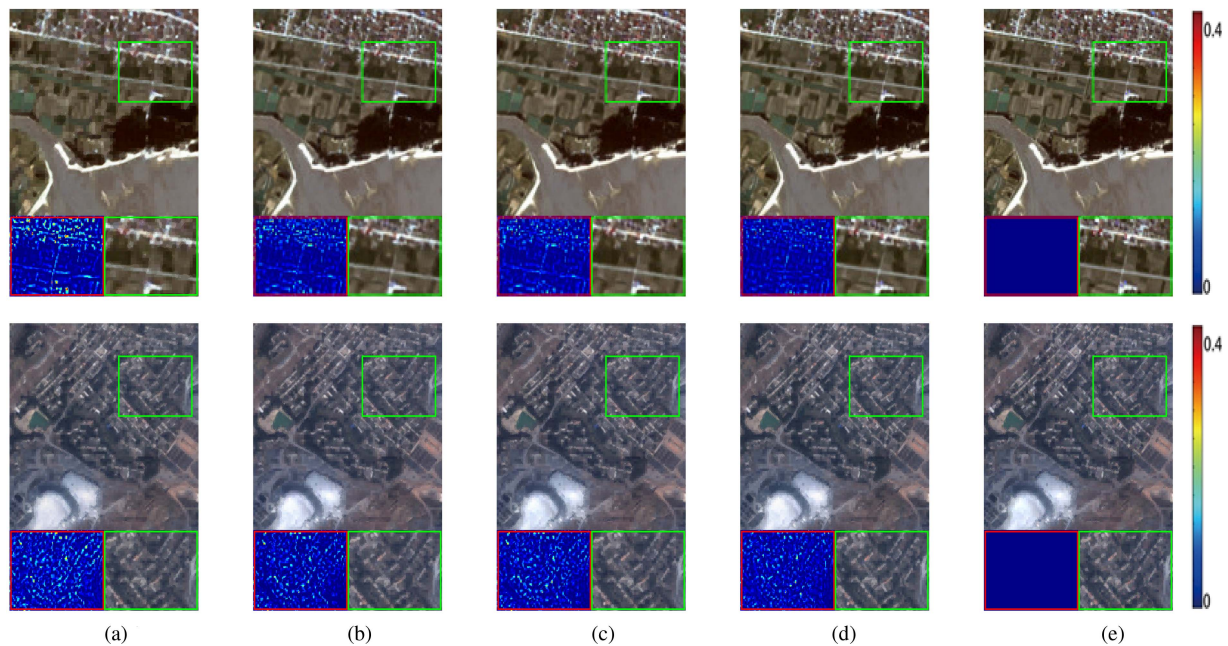


Fig. 12. Comparison of pansharpened images by the models with different loss functions on simulated data from IKONOS dataset. (a) Model 4. (b) Model 5. (c) Model 6. (d) Proposed. (e) GT.

of Model 8. Our results have the best visual effects and least residues compared to those of Models 7 and 8. Besides, the quantitative results are displayed in Table VI. The proposed injection conditions obtained by ECP and ECM performs better than Models 7 and 8, confirming its value to the method.

4) *Application Experiment*: To assess the applicability of all comparison methods, image classification experiments were conducted on the fusion results. In [21], the ENVI tool was employed for classification purpose. Initially, The GT images are first fed to the classification model to obtain classification reference images for other methods. As shown in Fig. 14, the subjective classification results of various comparison methods

were examined, and the residual images of two different selection regions were displayed at the bottom of each image. It is obvious that the result of our approach is the closest to the GT. In addition, the quantitative classification results are presented in Table VII. It is also evident from Table VII that the proposed PDDM outperforms the other comparison methods in terms of classification accuracy.

D. Discussion

One of the primary advantages of our method is its advanced performance in pansharpening, as assessed both subjectively

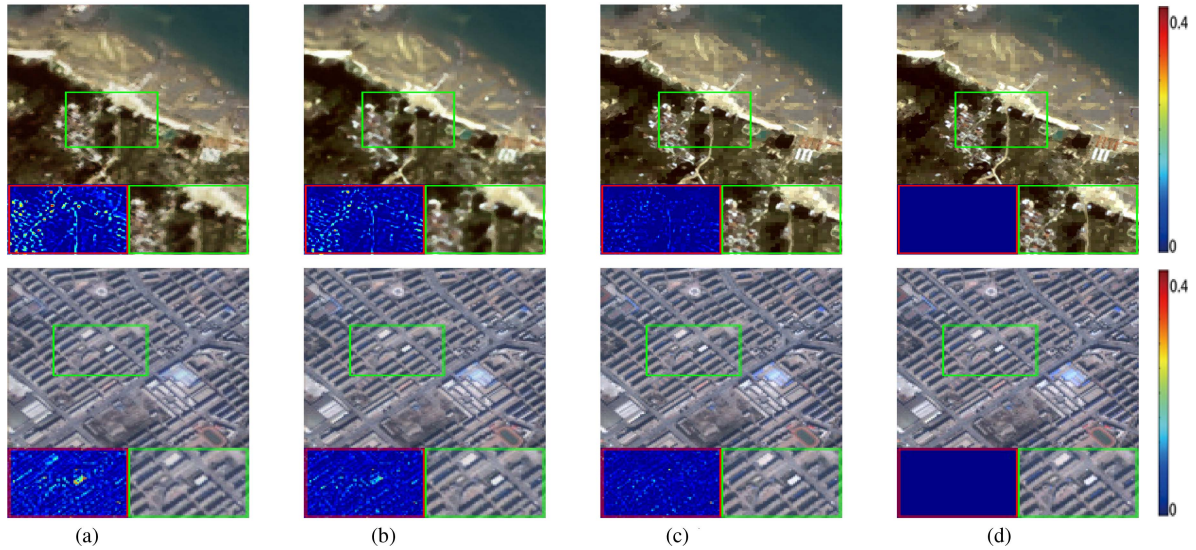


Fig. 13. Comparison of pansharpened images by the models with different injection conditions on simulated data from IKONOS dataset. (a) Model 7. (b) Model 8. (c) Proposed. (d) GT.

TABLE VI
ABLATION STUDY RESULTS OF DIFFERENT INJECTION CONDITIONS ON THE SIMULATED DATA FROM IKONOS DATASET

Methods	Injection Conditions	PSNR(\uparrow)	UIQI(\uparrow)	SAM(\downarrow)	ERGAS(\downarrow)
Model 7	Concatenation	31.8224	0.9719	2.8914	2.4749
Model 8	Convolutional Residual Block	33.0950	0.9783	2.3322	1.9744
Proposed	ECM & ECP	34.6573	0.9870	2.1705	1.5412

The best quantitative results are marked with bold fonts.

TABLE VII
QUANTITATIVE CLASSIFIED RESULTS OF FIG. 14

Methods	K(\uparrow)	OA(\uparrow)	CE(\downarrow)	OE(\downarrow)	PA(\uparrow)	UA(\uparrow)
GSA [16]	0.4831	0.5815	0.4504	0.4518	0.5482	0.5496
GLP-REG [13]	0.5033	0.5976	0.4348	0.4361	0.5639	0.5652
DRPNN [14]	0.6698	0.7321	0.3064	0.3041	0.6959	0.6936
MSDCNN [19]	0.6846	0.7442	0.2945	0.2909	0.7091	0.7056
PCDRN [42]	0.7597	0.8052	0.2272	0.2254	0.7747	0.7728
TFNet [24]	0.7150	0.7688	0.2667	0.2612	0.7388	0.7333
FusionNet [25]	0.6580	0.7212	0.3167	0.3027	0.6973	0.6833
TDNet [30]	0.6369	0.7047	0.3336	0.3252	0.6749	0.6664
AWFLN [31]	0.7554	0.8019	0.2296	0.2299	0.7701	0.7704
LRTCFPan [20]	0.6260	0.6973	0.3334	0.3361	0.6639	0.6666
VOGTNet [26]	0.7475	0.7953	0.2369	0.2350	0.7651	0.7631
PLR-Diff [34]	0.6597	0.7233	0.3161	0.3070	0.6930	0.6839
Dif-PAN [36]	0.7744	0.8172	0.2119	0.2115	0.7885	0.7881
PDDM (ours)	0.7987	0.8369	0.1887	0.1881	0.8120	0.8113

The best quantitative results are marked with bold fonts.

and objectively. Through a prior image-guided approach, our diffusion branches mitigate spatial and spectral distortions arising from uncertainties when applying diffusion models to pansharpening. In addition, adversarial and collaborative learning strategies are designed to fully interact spectral and spatial information across different branches. Furthermore, a focus module is developed to enhance the generalization of PDDM at different scales. Moreover, the applicability of our method has been validated.

Although our method performs exceptionally in pansharpening, there are still certain limitations. The diffusion model requires separate training of the denoising module, which interrupts the end-to-end training process. In addition, due to multiple noise sampling steps, more inference time of PDDM is required compared to traditional DL-based modules. Future research will focus on optimizing sampling techniques to enhance the inference speed of the diffusion model.

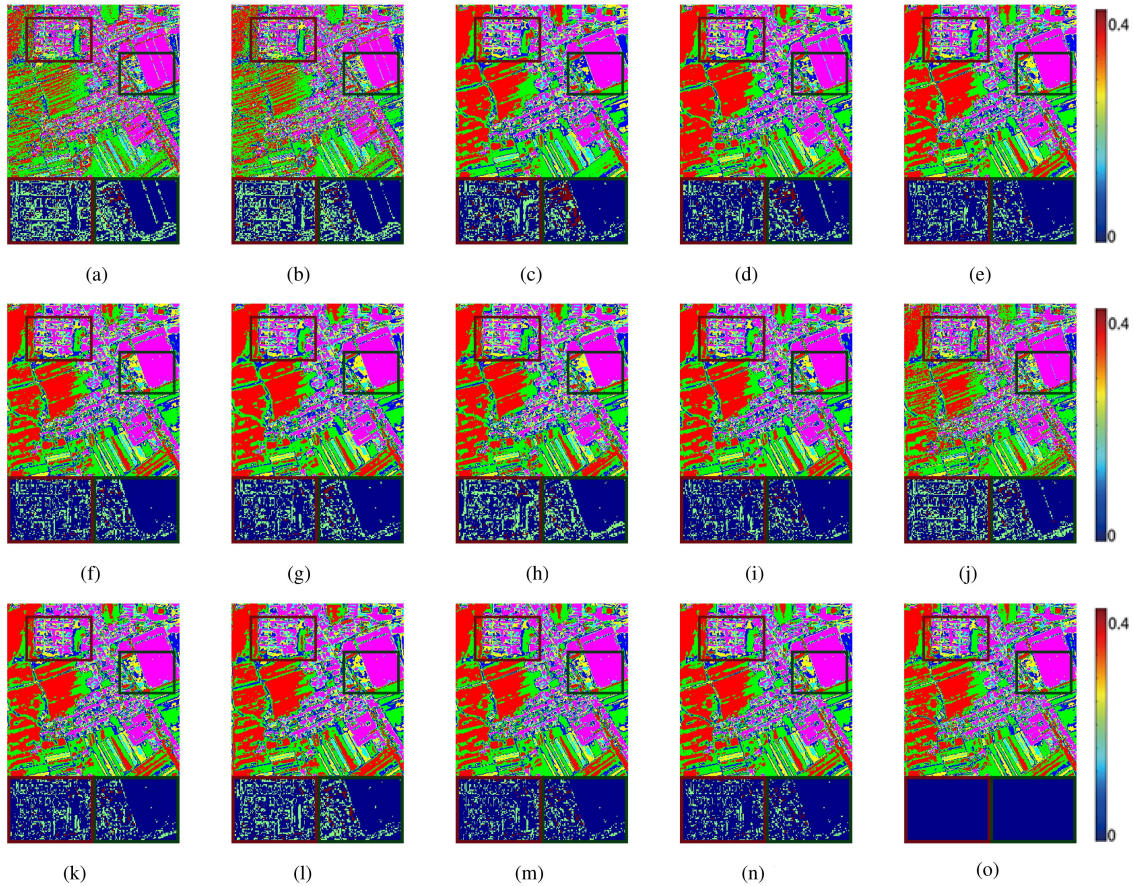


Fig. 14. Comparison of classified results of Fig. 5. (a) GSA [16]. (b) GLP-REG [13]. (c) DRPNP [14]. (d) MSDCNN [19]. (e) PCDRN [42]. (f) TFNet [24]. (g) FusionNet [25]. (h) TDNet [30]. (i) AWFLN [31]. (j) LRTCFFan [20]. (k) VOGTNet [26]. (l) PLR-Diff [34]. (m) Dif-PAN [36]. (n) PDDM (ours). (o) GT.

IV. CONCLUSION

In this article, a novel pansharpening method named PDDM is proposed, which employs a dual-branch diffusion model to extract different information to improve the spatial and spectral fidelity. The collaborative and adversarial loss functions ensure the fusion of complementary information from the source images. Furthermore, to enhance detail recovery and reduce the uncertainty of generated detail information, two pregeneration modules are constructed to leverage various prior information for pixel-to-pixel reconstruction. In addition, to improve the generalization capability of PDDM at different scales, a focus module supervised by a joint multiscale variation detection loss is designed. Extensive experiments on three satellite datasets demonstrate that our PDDM outperforms state-of-the-art pansharpening methods.

REFERENCES

- [1] Y. Zhang, Z. Guo, Y. Li, and B. Wu, "Structure tensor-driven block-based adaptive variational pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, Jan. 2024, Art. no. 5001005.
- [2] K. Zhang et al., "Panchromatic and multispectral image fusion for remote sensing and Earth observation: Concepts, taxonomy, literature review, evaluation methodologies and challenges ahead," *Inf. Fusion*, vol. 93, pp. 227–242, May. 2023.
- [3] L.-J. Deng et al., "Machine learning in pansharpening: A benchmark, from shallow to deep networks," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 3, pp. 279–315, Sep. 2022.
- [4] H. Tao, J. Li, Z. Hua, and F. Zhang, "DUDB: Deep unfolding-based dual-branch feature fusion network for pan-sharpening remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Dec. 2024, Art. no. 5400417.
- [5] J. Zhao, S. Tian, C. Geiß, L. Wang, Y. Zhong, and H. Taubenböck, "Spectral-spatial classification integrating band selection for hyperspectral imagery with severe noise bands," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1597–1609, Apr. 2020.
- [6] Z. Li, J. Li, L. Ren, and Z. Chen, "Transformer-based dual-branch multiscale fusion network for pan-sharpening remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 614–632, Nov. 2024.
- [7] H. Dai, Y. Yang, S. Huang, W. Wan, H. Lu, and X. Wang, "Pansharpening based on fuzzy logic and edge activity," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, Jan. 2024, Art. no. 5001205.
- [8] C. Chen et al., "MFITN: A multilevel feature interaction transformer network for pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, Mar. 2024, Art. no. 5003505.
- [9] Y. Yang, J. Wu, S. Huang, Y. Fang, P. Lin, and Y. Que, "Multimodal medical image fusion based on fuzzy discrimination with structural patch decomposition," *IEEE J. Biomed. Health. Inf.*, vol. 23, no. 4, pp. 1647–1660, Jul. 2019.
- [10] Q. Cao, L.-J. Deng, W. Wang, H. Junming, and G. Vivone, "Zero-shot semi-supervised learning for pansharpening," *Inf. Fusion*, vol. 101, Jan. 2024, Art. no. 102001.
- [11] G. Hassan, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, Nov. 2016.
- [12] B. Aiuzzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.

- [13] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3418–3431, Jul. 2018.
- [14] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.
- [15] G. Vivone et al., "A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 53–81, Mar. 2021.
- [16] B. Aiuzzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.
- [17] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001.
- [18] Q. Xu, B. Li, Y. Zhang, and L. Ding, "High-fidelity component substitution pansharpening by the fitting of substitution data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7380–7392, Nov. 2014.
- [19] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth convolutional neural network for remote sensing imagery pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 978–989, Mar. 2018.
- [20] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J. Huang, J. Chanussot, and G. Vivone, "LRTCFFan: Low-rank tensor completion based framework for pansharpening," *IEEE Trans. Image Process.*, vol. 32, pp. 1640–1655, Feb. 2023.
- [21] H. Lu, Y. Yang, S. Huang, W. Tu, and W. Wan, "A unified pansharpening model based on band-adaptive gradient and detail correction," *IEEE Trans. Image Process.*, vol. 31, pp. 918–933, Dec. 2022.
- [22] S. Parisotto, L. Calatroni, A. Bugeau, N. Papadakis, and C.-B. Schönlieb, "Variational osmosis for non-linear image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 5507–5516, Apr. 2020.
- [23] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, Jul. 2016, Art. no. 594.
- [24] X. Liu, Q. Liu, and Y. Wang, "Remote sensing image fusion based on two-stream fusion network," *Inf. Fusion*, vol. 55, pp. 1–15, Mar. 2020.
- [25] L.-J. Deng, G. Vivone, C. Jin, and J. Chanussot, "Detail injection-based deep convolutional neural networks for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6995–7010, Aug. 2021.
- [26] P. Wang, Z. He, B. Huang, M. D. Mura, H. Leung, and J. Chanussot, "VOGTNet: Variational optimization-guided two-stage network for multispectral and panchromatic image fusion," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, 2024, doi: [10.1109/TNNLS.2024.3409563](https://doi.org/10.1109/TNNLS.2024.3409563).
- [27] R. Wen, L.-J. Deng, Z.-C. Wu, X. Wu, and G. Vivone, "A novel spatial fidelity with learnable nonlinear mapping for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5401915.
- [28] P. Wang et al., "Low-rank tensor completion pansharpening based on haze correction," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, May. 2024, Art. no. 5405720.
- [29] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J.-F. Hu, and G. Vivone, "VO+ Net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5401016.
- [30] T.-J. Zhang, L.-J. Deng, T.-Z. Huang, J. Chanussot, and G. Vivone, "A triple-double convolutional neural network for panchromatic sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 9088–9101, Nov. 2023.
- [31] H. Lu et al., "AWFLN: An adaptive weighted feature learning network for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 5400815.
- [32] C. Zhu, S. Deng, Y. Zhou, L.-J. Deng, and Q. Wu, "QIS-GAN: A lightweight adversarial network with quadtree implicit sampling for multispectral and hyperspectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Nov. 2023, Art. no. 5531115.
- [33] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10850–10869, Sep. 2023.
- [34] X. Rui, X. Cao, L. Pang, Z. Zhu, Z. Yue, and D. Meng, "Unsupervised hyperspectral pansharpening via low-rank diffusion model," *Inf. Fusion*, vol. 107, 2024, Art. no. 102325.
- [35] L. Pang et al., "HIR-Diff: Unsupervised hyperspectral image restoration via improved diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2024, pp. 3005–3014.
- [36] Z. Cao, S. Cao, L.-J. Deng, X. Wu, J. Hou, and G. Vivone, "Diffusion model with disentangled modulations for sharpening multispectral and hyperspectral images," *Inf. Fusion*, vol. 104, 2024, Art. no. 102158.
- [37] Y. Zhong, X. Wu, L.-J. Deng, and Z. Cao, "SSDiff: Spatial-spectral integrated diffusion model for remote sensing pansharpening," 2024, [arXiv:2404.11537](https://arxiv.org/abs/2404.11537).
- [38] L. Guo et al., "A joint framework for denoising and estimating diffusion kurtosis tensors using multiple prior information," *IEEE Trans. Med. Imag.*, vol. 41, no. 2, pp. 308–319, Feb. 2022.
- [39] G. França and J. Bento, "Markov chain lifting and distributed ADMM," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 294–298, Mar. 2017.
- [40] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?," in *Proc. Fusion Earth Data: Merging Point Meas., Raster Maps Remotely Sens. Images*, 2000, pp. 99–103.
- [41] G. Vivone et al., "Pansharpening based on semiblind deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Sep. 2015.
- [42] Y. Yang, W. Tu, S. Huang, and H. Lu, "PCDRN: Progressive cascade deep residual network for pansharpening," *Remote Sens.*, vol. 12, no. 4, p. 676, Feb. 2020.
- [43] G. Vivone et al., "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May. 2015.
- [44] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, Nov. 1998.
- [45] X. Guan, F. Li, X. Zhang, M. Ma, and S. Mei, "Assessing full-resolution pansharpening quality: A comparative study of methods and measurements," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 6860–6875, Jul. 2023.



Changjie Chen received the B.S. degree in electronic information science and technology from Nanchang Hangkong University, Nanchang, China, in 2016. He is currently working toward the Ph.D. degree in management science and engineering with the Jiangxi University of Finance and Economics, Nanchang.

His research interests include image fusion and deep learning.



Yong Yang (Senior Member, IEEE) received the Ph.D. degree in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 2005.

From 2009 to 2010, he was a Postdoctoral Research Fellow with Chonbuk National University, Jeonju, South Korea. He is currently a Distinguished Professor with the School of Computer Science and Technology, Tiangong University, Tianjin, China. His research interests include image fusion, image super-resolution reconstruction, medical image processing and analysis, and deep learning.

Dr. Yang is an Associate Editor for IEEE ACCESS and an Editor for the *KSII Transactions on Internet and Information Systems*.



Shuying Huang (Member, IEEE) received the Ph.D. degree in computer application technology from the Ocean University of China, Qingdao, China, in 2013.

She is currently a Professor with the School of Software, Tiangong University, Tianjin, China. Her research interests include image and signal processing, and pattern recognition.



Hangyuan Lu received the Ph.D. degree in management science and engineering from the Jiangxi University of Finance and Economics, Nanchang, China, in 2021.

He is currently a Professor with the Jinhua University of Vocational Technology, Jinhua, China. His research interests include remote sensing image fusion and deep learning.



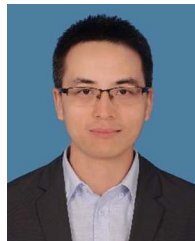
Wenyong Wen (Member, IEEE) received the Ph.D. degree in computational mathematics from Chongqing University, Chongqing, China, in 2013.

She is currently a Professor with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. Her research interests include image processing, multimedia security, and artificial intelligence security.



Weiguang Wan received the B.S. degree in mathematics and applied mathematics from Jiangxi Normal University, Nanchang, China, in 2014, and the Ph.D. degree in computer science and engineering from Jeonbuk National University, Jeonju, South Korea, in 2020.

He is currently a Lecturer with the School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang. His research interests include computer vision, deep learning, face sketch synthesis and recognition, and remote sensing image fusion.



Shuzhao Wang received the master's degree in mechanical manufacturing and automation from the Jiangxi University of Science and Technology, Ganzhou, China, in 2014. He is currently working toward the Ph.D. degree in electronic information with Tiangong University, Tianjin, China.

His research interests include image fusion, deep learning, and smart and intelligent mining.



Shengna Wei received the B.S. degree in software engineering from the East China University of Technology, Nanchang, China, in 2015. She is currently working toward the Ph.D. degree in electronic information with Tiangong University, Tianjin, China.

Her research interests include image fusion and deep learning.