Unsupervised Single-Generator CycleGAN-Based Pansharpening With Spatial-Spectral Degradation Modeling

Wenxiu Diao, Mengying Jin, Kai Zhang, *Member, IEEE*, and Liang Xiao, *Member, IEEE*

***Abstract*—Supervised pansharpening methods require the ground truth, which is generally unavailable. Therefore, the popularity of unsupervised pansharpening methods has increased. Generative adversarial networks (GANs) are often employed for unsupervised pansharpening, although achieving precise control over the generation process to capture rich spatial and spectral details is challenging. CycleGAN introduces cycle consistency loss and utilizes the cooperative training of two generators and two discriminators to learn the mapping between different domains. This approach partially addresses the issue of limited control over the generated results in traditional GANs. Therefore, CycleGAN also can be employed to accomplish unsupervised pansharpening tasks. However, it is complicated to directly apply the network structure of CycleGAN to pansharpening. To address this issue, we integrate a process model capable of simulating spatial and spectral degradations into a single-generator CycleGAN, which can learn the target distribution. Specifically, we propose an unsupervised CycleGAN for pansharpening based on spatial and spectral degradations and consists of one lightweight generator and two discriminators. Then, the low-resolution multispectral and panchromatic images are considered as the spatial and spectral degradations of the high-resolution multispectral images. Besides, unsupervised loss functions consisting of cycle consistency, adversarial, spectral angle mapper, and edge enhancement losses are designed to preserve spectral and spatial information. The experimental results on the QuickBird, GeoEye-1, and GF-2 datasets show that the qualitative and quantitative analysis of the proposed method is comparable with most supervised methods and superior to most unsupervised methods.**

***Index Terms*—Cycle consistency (CC), generative adversarial network (GAN), pansharpening, spatial and spectral degradations, unsupervised.**

I. Introduction

THE technology of remote sensing imaging has evolved tremendously. Many earth observation satellites have obtained remote sensing images, such as panchromatic (PAN) and multispectral (MS) images [1]. PAN images include comprehensive spatial information that is used for object detection and recognition [2]. MS images feature a variety of spectral bands that are applied to image classification [1], [3]. Nevertheless, due to the tradeoff between spatial and spectral resolutions in remote sensing satellites [4], acquiring high-resolution (HR) MS images is problematic. The task of combining the spatial and spectral information of multisource remote sensing images is known as pansharpening. It aims to generate a more complete description of ground scene information by fusing PAN images with low-resolution (LR) MS images [5].

There are four types of pansharpening methods, including component substitution (CS)-based methods [6], [7], [8], [9], [10], multiresolution analysis (MRA)-based methods [11], [12], [13], [14], [15], model-based methods [16], [17], [18], [19], [20], [21], and deep neural networks (DNNs)-based methods [22], [23], [24], [25], [26], [27], [28], [29], [30]. The DNN-based pansharpening methods have been a research focus in recent years.

With the remarkable progress of DNNs in various image processing tasks [31], [32], DNNs-based pansharpening methods have been widely used. Based on whether the ground truth is provided, these methods are classified as supervised and unsupervised types. In supervised pansharpening, the labeled training dataset is typically used to learn a mapping function between the LR MS and PAN images [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43]. This mapping function can be represented using DNNs. The DNNs learn from the labeled training data to improve resolution and enhance visual details. In natural environments, obtaining labeled datasets can be challenging or even impossible. While using simulated labeled datasets can help alleviate the challenges of limited labeled data, they may not completely replace the need for labeled examples in some cases. Although unsupervised pansharpening methods produce satisfactory results, there are numerous issues that require additional investigation. GAN-based pansharpening methods have achieved excellent fusion performance, however, it is challenging to achieve precise control over the generation process in GANs to capture accurate spatial and spectral information. Some unsupervised methods frequently use cycle consistency (CC) to overcome this problem [25], [26], [27], [28]. Existing unsupervised methods often do this by minimizing the reconstructed data and the reconstructed data created by

Manuscript received 31 July 2023; revised 25 September 2023; accepted 18 October 2023. Date of publication 26 October 2023; date of current version 23 November 2023. This work was supported in part by the National Natural Science Foundation of China under Grants 61871226 and 61571230, and in part by Jiangsu Geological Bureau Research Project under Grant 2023KY11. *(Corresponding author: Liang Xiao.)*

Wenxiu Diao, Mengying Jin, and Liang Xiao are with the School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: diaowx0920@163.com; jinmengying_maths@njust.edu.cn; xiaoliang@mail.njust.edu.cn).

Kai Zhang is with the School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China (e-mail: zhangkainuc@163.com).

The source code is available at https://github.com/DDXNJUST/SD-CycleGAN.

Digital Object Identifier 10.1109/JSTARS.2023.3327169

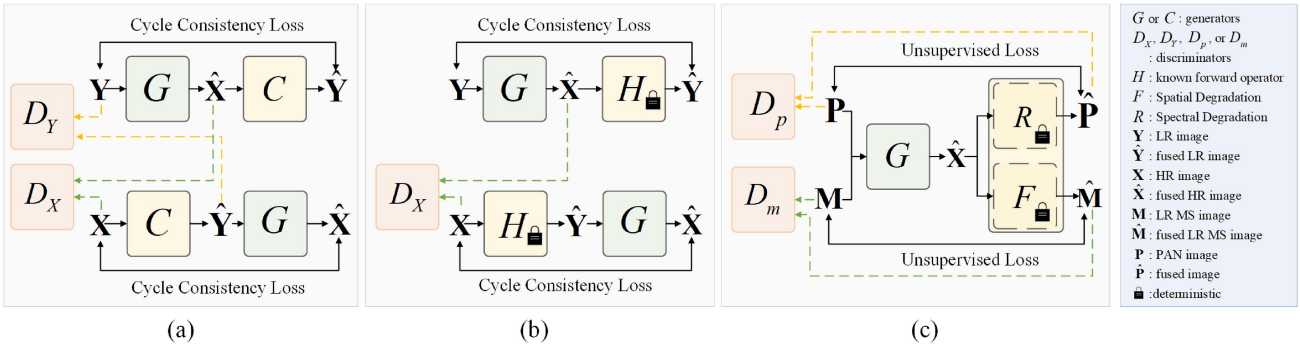

Fig. 1. Different network frameworks based on CycleGAN. (a) Classic CycleGAN network with two generators (G, C) and two discriminators D_Y and D_X, which are used to train tasks between two image domains. (b) CycleGAN with a known forward physics operator H, only a single pair of generator G and discriminator D_X is required. (c) Proposed simplified CycleGAN-based pansharpening framework with known degradation processes. As shown in (a) and (b), pansharpening based on CycleGAN requires labeled datasets, which means they need ground truth data for training. In comparison, our proposed model does not require ground truth data and operates in an unsupervised manner.

the model with the reconstructed data as input [26], or by evaluating the difference between the reconstructed data and the interpolated input data [25], [27], [28]. This can sometimes blur the features of the original image distribution, which may lead to a decrease in the visual quality of the generated images.

Considering the limitations raised above, we propose an unsupervised single-generator CycleGAN with spatial-spectral degradation modeling (SD-CycleGAN) for unsupervised pansharpening, as shown in Fig. 1(c). We use CycleGAN to fully exploit the cyclic consistency loss. A single generator is used to simplify the network. To achieve high-quality results, finding suitable methods for preserving spatial and spectral information is essential. The pansharpening of the inverse problem is divided into two degradation processes: spectral degradation, which means that the HR MS image is turned into a single-channel image that can be seen as a PAN image using the spectral response function, and spatial degradation, which means that the HR MS image is converted into a single-channel image that can be seen as a PAN image using the spatial response function. The proposed unsupervised method aims to fully leverage the degradation relationship between the original image pairs, as well as optimize the network and increase performance through the use of simulated spectral and spatial degradations. The base framework of this method is single-generator CycleGAN. Two degradation modules are utilized to model the two degradation processes on this basis. Furthermore, we offer a collection of unsupervised losses for constraining spatial and spectral information. For example, CC and adversarial generation losses maintain spatial and spectral consistency between input images and predicted HR MS images generated by the spectral and spatial degradation of fake-PAN and LR MS images. The spectral angle and edge enhancement (EE) losses restrict the disparities in spectral and spatial information between the input and output images, resulting in similar distributions.

The following are the contributions of the SD-CycleGAN.

1) The proposed method is the unsupervised method that is improved based on CycleGAN and fully exploits the cyclic consistency characteristics. The proposed method includes only one lightweight generator, enhancing the stability of training. At the same time, this method reasonably utilizes cyclic consistency.
2) The proposed network learns the generating process in a model-driven way, which entails mapping the pansharpening and corresponding degradation processes. The proposed method could generate images of excellent quality because of the stability of the degradation model.
3) Experiments on the QuickBird, GeoEye-1, and GF-2 datasets show that SD-CycleGAN is superior to the state-of-the-art unsupervised compared methods in terms of qualitative and quantitative evaluations.

The rest of this article is organized as follows. Section II provides a brief overview of existing pansharpening methods and the basic model of CycleGAN. In Section III, we introduce the proposed method along with the various modules and their functionalities. Section IV presents experimental results demonstrating the effectiveness of the proposed approach. Finally, Section V concludes this article and comments on this work.

II. Related Works

This section introduces existing traditional, DNN-based supervised and unsupervised sharpening methods, and describes the basic framework of CycleGAN.

A. CycleGAN

The degradation procedure of the observed LR and HR images can be formulated as

$$\mathbf{Y} = \mathbf{C}\mathbf{X} + \mathbf{N} \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{WH \times L}$ represents the HR image, and W, H, and L represent the width, height, and band number. And $\mathbf{Y} \in \mathbb{R}^{wh \times L}$ represents the LR image, where w and h represent the width and height. $\mathbf{C} \in \mathbb{R}^{wh \times WH}$ is the measurement matrix. $\mathbf{N}$ is the noise component. In practice, we cannot obtain $\mathbf{C}$ directly.[1]

[1]Here, for the convenience of description, we unfold the tensor of HR image $\mathbf{X} \in \mathbb{R}^{W \times H \times L}$ into a matrix with the size of $WH \times L$. Similar for $\mathbf{Y}$.

Fig. 1(a) depicts the standard CycleGAN structure with two generators $G : Y \rightarrow X$ and $C : X \rightarrow Y$, and two discriminators D_X and D_Y. The degradation of LR and HR MS images indicated in (1) can be accomplished by CycleGAN, as shown in Fig. 1(a). Generator G in Fig. 1(a) learns the process of converting LR images into HR images, whereas generator C in Fig. 1(a) learns the process of transferring LR images into HR images. Discriminator D_Y in Fig. 1(a) differentiates between the source LR MS images and fake LR MS images produced by generator C, whereas discriminator D_X differentiates between the source HR MS images and fake HR MS images generated by generator G. The competition between discriminators and generators provides an environment that improves network performance, resulting in higher quality image production. Furthermore, cyclic consistency imposes the one-to-one mapping condition between the two types of images, lowering the likelihood risk of the difficulty of controlling the generated results. The abovementioned structure can be viewed as a supervised super-resolution network. When using HR MS images as ground truth, it is necessary to provide supervision to this network.

B. Traditional Pansharpening Methods

For CS-based methods, interpolated LR MS images are projected onto a new space [6], [7]. It is assumed in this space that the spatial and spectral components of an image are independent of one another. PAN images are thus utilized to replace the spatial components in MS images. Finally, an inverse transformation is applied to the reconstructed components to obtain the generated HR MS images. Popular methods include intensity-hue-saturation transformation [8], principal component analysis [9], and Gram–Schmidt transformation [10]. These methods extract spatial and spectral components from MS images that are difficult in practice and perform badly in terms of maintaining spectral information. The MRA-based methods [11], [12] assume that the missing spatial information in LR MS images can be extracted from PAN images. As a result, MRA extracts spatial information from PAN images and injects it into interpolated LR MS images. Image decomposition methods, such as wavelet transform [13], contourlet [14], and curvature [15] are beneficial when applied to MRA-based methods, but are sensitive to spatial correspondence. The model-based methods [16], [17] assume that LR MS and PAN images are spatial and spectral degraded versions of HR MS images. Model-based methods offer a systematic framework for tackling inverse problems by formulating them as optimization tasks. The search space is constrained by defining an appropriate energy function, which encodes the desired properties or constraints of the solution. These methods employ several priors, such as sparsity [18] and low-rank priors [19], to decrease the solution space of the model. The representative model-based methods are sparse representation [20] and variation [21]. These methods necessitate an appropriate model to contain spatial and geometrical information, which can result in significant computing complexity and parameter sensitivity.

C. Supervised Pansharpening Methods

For supervised methods, training datasets with ground truth are required. Masi et al. [33], for example, used PNN with a three-layer convolutional structure for pansharpening. Zhang et al. [34] designed an adaptive feature fusion module SSE-Net that combines features from different subnetworks, reducing feature redundancy. Zhang et al. [35] proposed an attention-based Tri-UNet, which uses two subnetworks to extract spectral and spatial features from both MS and PAN images. Zhou et al. [41] proposed SFIIN, which is a spatial–frequency information integration network that incorporates spatial and frequency domain information branches with bidirectional interactions. Zhou et al. [42] proposed a framework that encourages complementary information learning between PAN and MS images to reduce information redundancy. Generative adversarial network (GAN) [44] is a framework that achieved remarkable progress in pansharpening by pitting a generator and a discriminator against each other. Gastineau et al. [36] introduced a novel GAN with two spectral information discriminators to preserve the spatial resolution of source images. Liu et al. [37] proposed PSGAN for generating high-fidelity images. In addition, Transformer [45] is a deep learning model based on a self-attention mechanism, which can better capture the relationships between image patches [38], [39], [40], [43]. HyperTransformer [38] enhances performance through attention mechanisms that learn spatial correlations between PAN and LR hyperspectral images (HSI). Zhang et al. [40] constructed a multiscale subnetwork with a convolution-transformer encoder to extract local and global features at different scales from LR MS and PAN images. Zhou et al. [43] proposed a method that leverages both transformer and information-lossless invertible neural modules to enhance spatial and spectral resolution. Furthermore, model-driven methods leverage mathematical models to drive problem-solving, which can address the lack of interpretability in deep learning methods to some extent [46], [47], [48], [49]. Yan et al. [46] combined model-driven and data-driven methods by introducing deep priors as implicit regularization into the network. Xiang et al. [49] proposed a depth fusion network based on a detail injection model, treating pansharpening as a complex nonlinear detail learning and injection problem.

The supervised pansharpening methods are trained on labeled datasets. To obtain labeled datasets, the original PAN and MS images are artificially degraded to reduce their resolution, resulting in reduced-scale datasets. The original MS images are used as a reference ground truth [50]. During the degradation process, the images lose spectral and spatial details. DNN-based models trained on the reduced-scale datasets may not be able to fully recover the lost spectral and spatial details, leading to a significant domain discrepancy between the training and testing data. As a result, when applying the trained models to real-world data, the performance may not be as satisfactory as expected.

D. Unsupervised Pansharpening Methods

On the other hand, unsupervised methods can learn feature representations without the need for labeled information, allowing them to better adapt to different domains of data.

In unsupervised paradigms, various methods have been proposed. The unsupervised methods do not require paired training datasets. GAN is usually used in unsupervised methods [22], [23], [24], [25]. PanGAN [22] was a GAN-based pansharpening method, in which the generator set up an adversarial game with spectral and spatial discriminators. These spectral and spatial information were processed using two discriminators, respectively. PGMAN [23] utilized dual-stream generators to extract modality-specific features from the PAN and MS images and developed dual discriminators to preserve the spectral and spatial information during the fusion process. In addition, to enhance the stability of the generator and encourage it to learn the correspondence between domains, CC loss is commonly used in training the network [25], [26], [27], [28]. UCGAN [26] was a kind of unsupervised GAN that extracts spatial and spectral information from source images on full-scale images, along with a hybrid loss that combines CC and adversary principles. Li et al. [27] proposed a self-supervised GAN with CC, consisting of two generators and two discriminators. It employed CC loss to ensure the consistency between the input LR MS images and the fused images in terms of spectral information. Besides, unsupervised methods have difficulty in dealing with complex scenes with a wide range of materials, such as urban areas with mixed land cover, textural heterogeneity, or strong spectral variations. Therefore, they often rely on assumptions about image statistical characteristics [29], [30]. LDP-Net [29] was an unsupervised network with a learnable degradation process, which learns two degradation processes using two basic CNN modules. P2Sharpen [30] was a progressively growing pansharpening network with deep spectral transformation. It established a mapping from MS to PAN images using the U-Net framework during the pretraining phase.

In addition, researchers have also proposed numerous unsupervised methods to fuse HSI and multispectral image (MSI). HyCoNet [51] consists of three coupled unsupervised autoencoder networks that adaptively learn the parameters of the point spread function and spectral response function. Liu et al. [52] proposed an unsupervised implicit autoencoder network for the fusion of MSI and HSI images, treating each pixel as an individual sample.

Existing unsupervised pansharpening methods typically rely on evaluating the difference between the reconstructed data and interpolated data, which has drawbacks in preserving the characteristics of the original image distribution. As a result, the generated images may lack the fine details and overall visual quality present in the original images.

III. Proposed Method

A. Problem Setup

To overcome the limitations of unsupervised pansharpening methods, we employ CycleGAN to address this issue. The CC loss [53] compares the differences between the original images and the reconstructed images. This loss function is mainly designed to promote the learning of bidirectional mappings. CycleGAN [53] allows for unsupervised image-to-image translation between two different domains, making full use of the advantages of CC loss, as shown in Fig. 1(a). CycleGAN is typically used for processing between two image domains. If pansharpening is applied directly to the CycleGAN framework, there will be conversion issues between three image domains, including HR MS into PAN images, HR MS into LR MS images, and PAN and LR MS into HR MS images. The model is too complex to attain stability since it requires three generators and three discriminators. Therefore, a more complex model structure is needed to handle the transformation problem between multiple image domains in this case. Besides, as shown in Fig. 1(b) [54], if we have prior knowledge of the forward operator, we can simplify the CycleGAN architecture. In this case, only one generator G and one discriminator D_X are needed to accomplish the task. However, the learned features may not be sufficient with simple CNN modules, resulting in inferior pansharpening results. Therefore, we need to create modules capable of learning the degradation relationship between images.

For pansharpening, the LR MS and PAN images can be expressed as

$$\mathbf{M} = \mathbf{FX} + \mathbf{N_1}, \mathbf{P} = \mathbf{XR} + \mathbf{N_2} \tag{2}$$

where $\mathbf{M} \in \mathbb{R}^{wh \times L}$ and $\mathbf{P} \in \mathbb{R}^{WH \times 1}$ represent LR MS and PAN images. $\mathbf{X} \in \mathbb{R}^{WH \times L}$ represents the HR MS image. $\mathbf{F} \in \mathbb{R}^{wh \times WH}$ and $\mathbf{R} \in \mathbb{R}^{L \times 1}$ describe the spatial and spectral degradations, respectively. $\mathbf{N}_1$ and $\mathbf{N}_2$ are both noise components.

B. SD-CycleGAN

Unsupervised methods are utilized to explore patterns and structures within the data without relying on prior knowledge or guidance. The proposed unsupervised pansharpening method SD-CycleGAN is shown in Fig. 2. SD-CycleGAN learns three mappings $G : [\mathbf{P}, \mathbf{M}] \to \mathbf{X}$, $R : \mathbf{X} \to \mathbf{P}$, and $F : \mathbf{X} \to \mathbf{M}$. To simplify the network structure, the proposed pansharpening method based on CycleGAN only has one generator G. PAN and LR MS images are fed into generator G, which is used to generate image $\hat{X}$. Meanwhile, the generator G is lightweight. On this basis, to find efficient methods for preserving spatial and spectral information, we established spectral and spatial degradations. As shown in Fig. 2, spectral degradation R and spatial degradation F together constitute the inverse operator C in CycleGAN shown in Fig. 2. The degradation model is used to preserve spatial and spectral features. According to the spatial and spectral degradations, the generated image $\hat{\mathbf{X}}$ is fed into the spatial and spectral degradations to generate the fake LR MS image $\hat{\mathbf{M}}$ and fake PAN image $\hat{\mathbf{P}}$, respectively. In addition, we use the source images and the degraded PAN and LR MS images to achieve an unsupervised network as discriminant criteria and conditions to maintain cyclic consistency, respectively. Therefore, we require two discriminators to distinguish spatial and spectral information separately. Therefore, we present two discriminators D_p and D_m. D_p differentiates the spatial details between $\hat{\mathbf{P}}$ and $\mathbf{P}$, whereas D_m distinguishes the spectral information between $\hat{\mathbf{M}}$ and $\mathbf{M}$. Adversarial loss is important among them. We also use CC loss in CycleGAN to address the issue of difficult-to-control the generated results. Furthermore, both

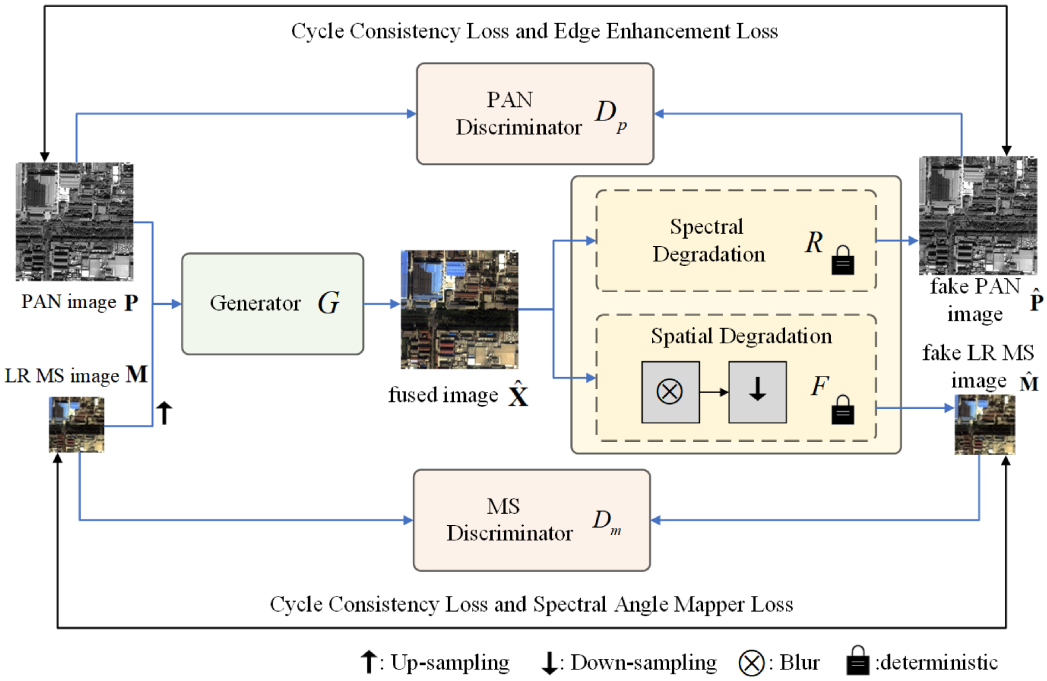

Fig. 2. Overview of the proposed SD-CycleGAN. In this CycleGAN-based framework, there is one lightweight generator G used to generate HR MS image $\hat{\mathbf{X}}$. In addition, spatial degradation F and spectral degradation R are employed to produce spatially and spectrally degraded images $\hat{\mathbf{M}}$ and $\hat{\mathbf{P}}$ of the generated image. The two discriminators D_p and D_m are responsible for assessing the realism of the generated images. The CC loss ensures that the spatial and spectral degradation information of the generated images remain consistent with the original LR MS and PAN images.

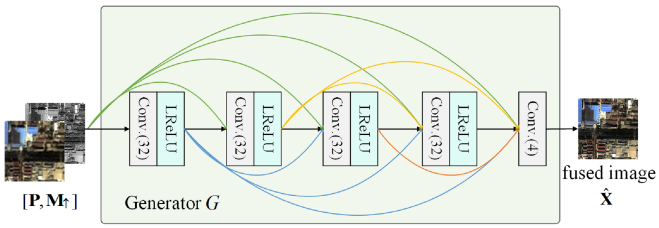

Fig. 3. Overview of the lightweight generator G, which is composed of a residual block. By stacking the PAN image $\mathbf{P}$ and interpolated LR MS image $\mathbf{M}_{\uparrow}$, the input is fed into G to generate HR MS image $\hat{\mathbf{X}}$.

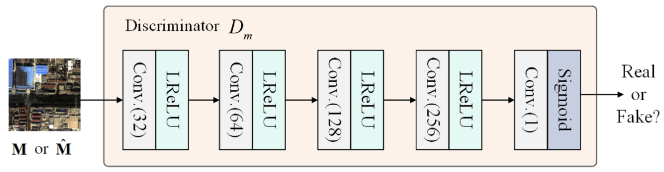

Fig. 4. Overview of the discriminator D_m, which is used to evaluate the differences between the source LR MS images $\mathbf{M}$ and the generated fake LR MS images $\hat{\mathbf{M}}$.

spectral angle loss and EE loss are used to preserve spectral and spatial information, respectively.

1) Generator: As shown in Fig. 3, the proposed method employs generator G, which includes a residual block.

As the only generator in CycleGAN, the input of G is the concatenation of $\mathbf{P}$ and the up-sampled LR MS image $\mathbf{M}_{\uparrow}$. The spatial size of $\mathbf{M}_{\uparrow}$ is the same as $\mathbf{P}$. The generator includes dense connections [55] to improve feature propagation across different convolutional layers and to stabilize the network. In Fig. 3, conv(n) represents the convolutional layer operator with n filters, and LReLU denotes the leaky ReLU activation function. The generator contains a residual module with five convolution layers and four Leaky ReLU activation functions. The first four convolutional layers are 32 channels, whereas the final convolutional layer is four channels. The filter size is 3×3.

2) Discriminators: The structure of discriminator D_m is shown in Fig. 4, which can capture the difference in distribution between $\mathbf{M}$ and $\hat{\mathbf{M}}$. The symbols in Fig. 4 represent the same as those in Fig. 3. The filter size is 3×3. The full convolution is used to effectively model the spatial and spectral information. The stride for the first three filters is 2, whereas the stride for the last two filters is 1. The architectures of D_p and D_m are similar, whereas $\mathbf{P}$ and $\hat{\mathbf{P}}$ are fed into D_p.

3) Spatial Degradation: One important aspect of spatial degradation is simulating the degradation of HR images into LR images. The consistency principle of Wald's Protocol [50] states that once the fused images are degraded to their original resolution, they should be similar to the original MS images. However, the spatial filter used for degradation remains an open question. To address this issue, filtering operators can be used, such as emulating a modulation transfer function (MTF). However, the MTF filters differ for each satellite sensor. Nevertheless, the filter gain at the Nyquist cutoff frequency can be obtained from on-orbit measurements. Utilizing this information, we assume that the frequency response of each filter approximately follows a Gaussian shape [56]. As a result, we can estimate the MTF

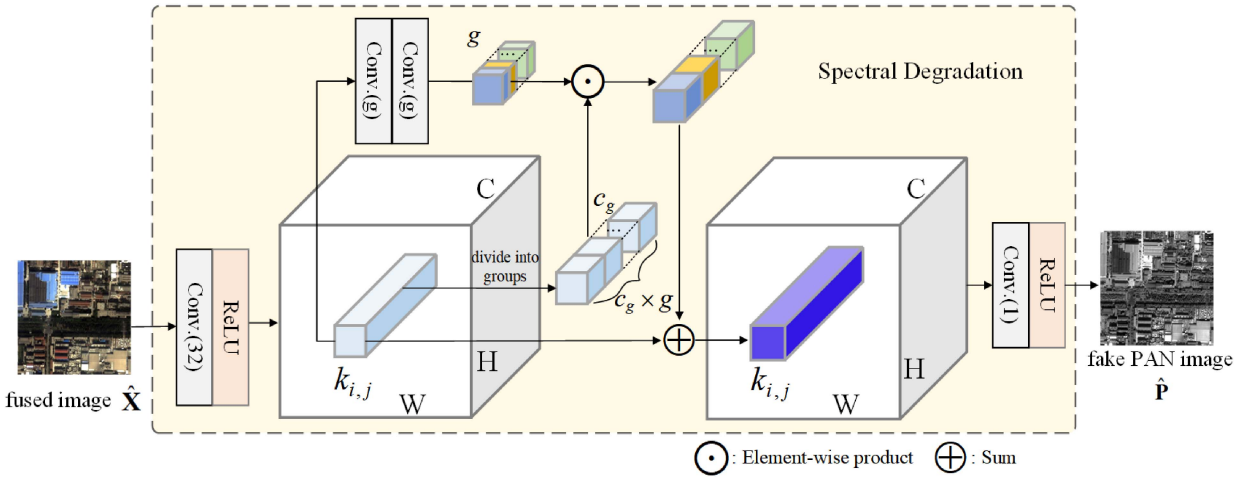

Fig. 5. Overview of spectral degradation R. To simulate the relationship between each spectral channel of MS and PAN images, we employ a self-attention mechanism to learn the pixelwise channel degradation mapping.

filters for each sensor of each satellite. Therefore, we substitute the MTF-shaped filters with approximated ideal Gaussian filters, which aims to ensure that the fused image maintains similarity with the original MS image even after being downscaled to its original resolution. In addition, [57] compared several operators such as truncated Shannon function, bicubic interpolation, pyramid-weighted averaging, and wavelet transform for different scenarios, showing that the relative differences between the results are only a few percentage points. Therefore, as long as the operator is sufficiently appropriate, the impact of the filtering operators can be kept minimal.

To more accurately describe the degradation model of HR and LR MS images, we introduce blurring operation and downsampling into the spatial degradation. According to [56], we approximate the blurring kernel by Gaussian filter with variance and mean of 2.28 and 5, respectively. The downsampling operation uses bilinear interpolation.

4) Spectral Degradation: We use the structure in Fig. 5 to focus on simulating spectral degradation from HR MS images into PAN images, which is equal to adaptively learning the spectral response function. The middle part of the proposed model differs from the standard convolutional layer. Convolutional operation is a feature map that shares convolutional kernel parameters at different spatial places in a single channel but utilizes a separate convolutional kernel for each channel. And our method is the opposite. We employed the same convolutional kernel for different channels, as inspired by [58], and constructed a shared convolutional kernel for channels. This can summarize the information of the front and rear channels within a spatial range and learn pixel-level spectral degradation.

According to [1], there is a relationship between the spectrums of PAN and MS images. To more accurately simulate the relationship among each spectrum corresponding to MS and PAN images, we introduce a self-attention mechanism on the channel dimension to approximate the spectral degradation R. First, we expand the HR MS image into a feature map of 32 channels through a layer of convolution and the ReLU activation function. As shown in Fig. 5, the single pixel point $k_{i,j}$ at the (i, j) in the feature map generates a tensor with a size of $(1, 1, g)$ by convolutions. The 1-D vector containing $K_{i,j}$ is divided into g groups, each having c_g channels. As the weight, the generated tensor multiplies the tensor with C channels of the pixel point $k_{i,j}$. Finally, the multiplied tensor is added as a residual to the original tensor to obtain the generated result at $k_{i,j}$ of the PAN image. By repeating this process for each point on the feature map, PAN images with one channel are generated. The proposed spectral degradation learns pixelwise channel degradation relationships.

To improve the optimization of SD-CycleGAN, the spectral degradation is pretrained by original LR MS images $\mathbf{M}$ and PAN images after spatial degradation $F(\mathbf{P})$ with the size of $64 \times 64 \times 4$ and 64×64. During the pretraining process, the regularization constraint used is the mean absolute error between $F(\mathbf{P})$ and $\hat{\mathbf{P}}$. This constraint helps to regularize the model and encourage consistency between the two variables $F(\mathbf{P})$ and $\hat{\mathbf{P}}$. Furthermore, pretraining provides a better initial state for subsequent optimization processes, allowing for a rapid improvement in overall performance.

C. Unsupervised Loss Functions

To train SD-CycleGAN for effective unsupervised learning, we propose the following unsupervised loss functions.

1) Cycle Consistency Loss: As shown in (2), pansharpening requires the simultaneous acquisition of spatial and spectral degradations. The CC of SD-CycleGAN requires that any $\mathbf{P}$ and $\mathbf{M}$ can be reconstructed after applying the generator operator $G(\cdot)$ and degradation operators $[R(\cdot), F(\cdot)]$ on $\mathbf{P}$ and $\mathbf{M}$ in turn. Any $\mathbf{X}$ can be reconstructed after applying $[R(\cdot), F(\cdot)]$ and $G(\cdot)$ on $\mathbf{X}$. That is, $R(G([\mathbf{P}, \mathbf{M}])) \approx \mathbf{P}$ and $F(G([\mathbf{P}, \mathbf{M}])) \approx \mathbf{M}$.

SD-CycleGAN uses CC to optimize the network. The spectral information of $\mathbf{M}$ and the spatial details of $\mathbf{P}$ are unsupervised information from the network training. The CC loss is expressed

as

$$L_{\text{cycle}} = \alpha \left\| \mathbf{P} - R\left(G\left([\mathbf{P}, \mathbf{M}]\right)\right)\right\|_1 + \beta \left\| \mathbf{M} - F\left(G\left([\mathbf{P}, \mathbf{M}]\right)\right)\right\|_1. \quad (3)$$

2) Adversarial Loss: In the proposed SD-CycleGAN, the adversarial learning between real and fake images is achieved by the WGAN-GP [59] loss, which is expressed as

$$L_{WGAN-GP} = D_p\left(\hat{\mathbf{P}}\right) - D_p\left(\mathbf{P}\right) + \left(\left\|\nabla D_p\left(\hat{\mathbf{P}}\right)\right\|_2 - 1\right)^2 + D_m\left(\hat{\mathbf{M}}\right) - D_m\left(\mathbf{M}\right) + \left(\left\|\nabla D_m\left(\hat{\mathbf{M}}\right)\right\|_2 - 1\right)^2 \quad (4)$$

where $\hat{\mathbf{P}} = R(G([\mathbf{P}, \mathbf{M}]))$, $\hat{\mathbf{M}} = F(G([\mathbf{P}, \mathbf{M}]))$, ∇ denotes the gradient operator. Furthermore, $D_m(\cdot)$ and $D_p(\cdot)$, respectively, represent the MS and PAN discriminator operators.

3) Spectral Angle Mapper (SAM) Loss: Spectral angle considers the spectrum of each pixel to be a high-dimensional vector and calculates the angle between the two vectors to determine the similarity of spectra. The smaller the angle is, the closer the two spectrums are. So, we use the SAM loss to constrain the spectral information, which is denoted as

$$L_{\text{SAM}} = \cos^{-1}\left(\frac{\langle F\left(G\left([\mathbf{P}, \mathbf{M}]\right)\right), \mathbf{M}\rangle}{\left\|F\left(G\left([\mathbf{P}, \mathbf{M}]\right)\right)\right\|_F \cdot \|\mathbf{M}\|_F}\right). \quad (5)$$

4) EE Loss: The edge and texture of the images have considerable influence on spatial information. Sharper image edges with higher spatial resolution, but blurrier image edges with lower spatial resolution. The Sobel operators are used to compute the gradient map in the x and y directions of images, and the extracted edge information is used as side information to compute the gradient difference between the faked PAN image and the real one to preserve the spatial characteristics of the generated images. The EE loss is denoted as

$$L_{\text{edge}} = \left\|\mathbf{P} \otimes \mathbf{G}_x - \hat{\mathbf{P}} \otimes \mathbf{G}_x\right\|_F^2 + \left\|\mathbf{P} \otimes \mathbf{G}_y - \hat{\mathbf{P}} \otimes \mathbf{G}_y\right\|_F^2 \quad (6)$$

where $\mathbf{G}_x$ and $\mathbf{G}_y$ are the Sobel operators of the x and y directions, respectively. $\otimes$ denotes convolution.

To summarize, the total loss functions are expressed as follows:

$$L = L_{WGAN-GP} + L_{\text{cycle}} + \delta L_{\text{SAM}} + \gamma L_{\text{edge}}. \quad (7)$$

SD-CycleGAN can be optimized by minimizing (7).

IV. Experiments

A. Experimental Settings

In this section, we compare and analyze the varied performance of the proposed method. BDSD [60], P+XS [61], PNN [33], PSGAN [37], M-GAN [36], PanGAN [22], LDP-Net [29], UCGAN [26], and SSCycleGAN [27] are the compared methods.

We conduct experiments on the QuickBird, GeoEye-1, and GF-2 datasets. The QuickBird dataset is collected in an urban area of Xi'an, which consists of buildings, roadways, and so on. The GeoEye-1 dataset includes some land, buildings, vegetation, and roads in the rural and urban districts of Hobart, Australia. Table I displays the spatial resolution of each satellite dataset along with the number of image pairs in the training, validation, and testing sets. The LR MS and PAN images have sizes of $64 \times 64 \times 4$ and 256×256.

Reduced-scale datasets can serve as a reference for quality evaluation. Wald's protocol [50] is utilized to generate the reduced-scale LR MS and PAN images by blurring and downsampling original MS and PAN images in datasets on which SD-CycleGAN is trained. Original MS images serve as reference images. Therefore, the generated images and the reference images have the same spatial size and can be compared. When evaluating fused images at the reduced-scale, we consider root-mean-squared error (RMSE), *Erreur Relative Globale Adimensionnelle de Synthèse* (ERGAS) [62], SAM [63], Q4 [64], and universal image quality index (UIQI) [65]. Full-scale datasets are another option to assess image quality without reference. While evaluating full-scale fused images, D_λ, D_S, and quality w/o reference (QNR) [66] are considered.

To train the network, we employ the Adam optimizer. The batch size is set to 2. The learning rate is set at 0.0001. The parameters α, β, δ, and γ in (3) and (7) are set to 50, 30, 100, and 100.

B. Reduced-Scale Dataset Experiments

Figs. 6–8 depict the fused images of SD-CycleGAN and the compared methods on the reduced-scale QuickBird, GeoEye-1, and GF-2 testing datasets. Building and road regions are selected from the fused images for further qualitative analysis in this section. The figures also show the absolute error mappings between the fused and reference images. The spectra in the BDSD fused images are distorted because BDSD cannot effectively estimate the gain parameters. The intensity in the P+XS fused images is overenhanced, and spectral distortion occurs in the road area. The spectral relationship among MS image bands is difficult to depict due to the simple structure of the PNN method. The PNN fused images have an obvious spectral distortion in the building area. The PSGAN fused images show some blurring effects because the method does not consider the loss of space enhancement. The M-GAN fused images produce spectral distortion, and the color of the magnified area becomes nearly gray, which is caused by the improper tradeoff between the two discriminators. The fused images of LDP-Net show spectral distortion, whereas the fused images of PanGAN and UCGAN have some fuzzy effects, owing to the difficulty of weighing the weight of spectral and spatial constraints in unsupervised methods. SSCycleGAN fused images have minor spectral distortion. The fused images of the proposed method show improved reconstruction performance and efficacy.

Tables II–IV show the quantitative analysis of full- and reduced-scale QuickBird testing datasets, with the best values of traditional, supervised, and unsupervised methods, marked in bold. These tables show that SD-CycleGAN performs well in both qualitative and quantitative effects. The average quantitative comparison results in these tables demonstrate that SD-CycleGAN outperforms most indexes of unsupervised methods, including RMSE, ERGAS, Q4, and UIQI.

TABLE I
DETAILS OF THE DATASETS USED IN THE FOLLOWING EXPERIMENTS

Sensor	Spatial resolution(m)		#Image pairs			
	LR MS	PAN	Training data	Validation data	Testing data	
					Reduced-scale	Full-scale
QuickBird	2.8	0.7	7720	10	155	395
GeoEye-1	2.0	0.5	6000	10	624	676
GF-2	4.0	1.0	6500	10	627	400

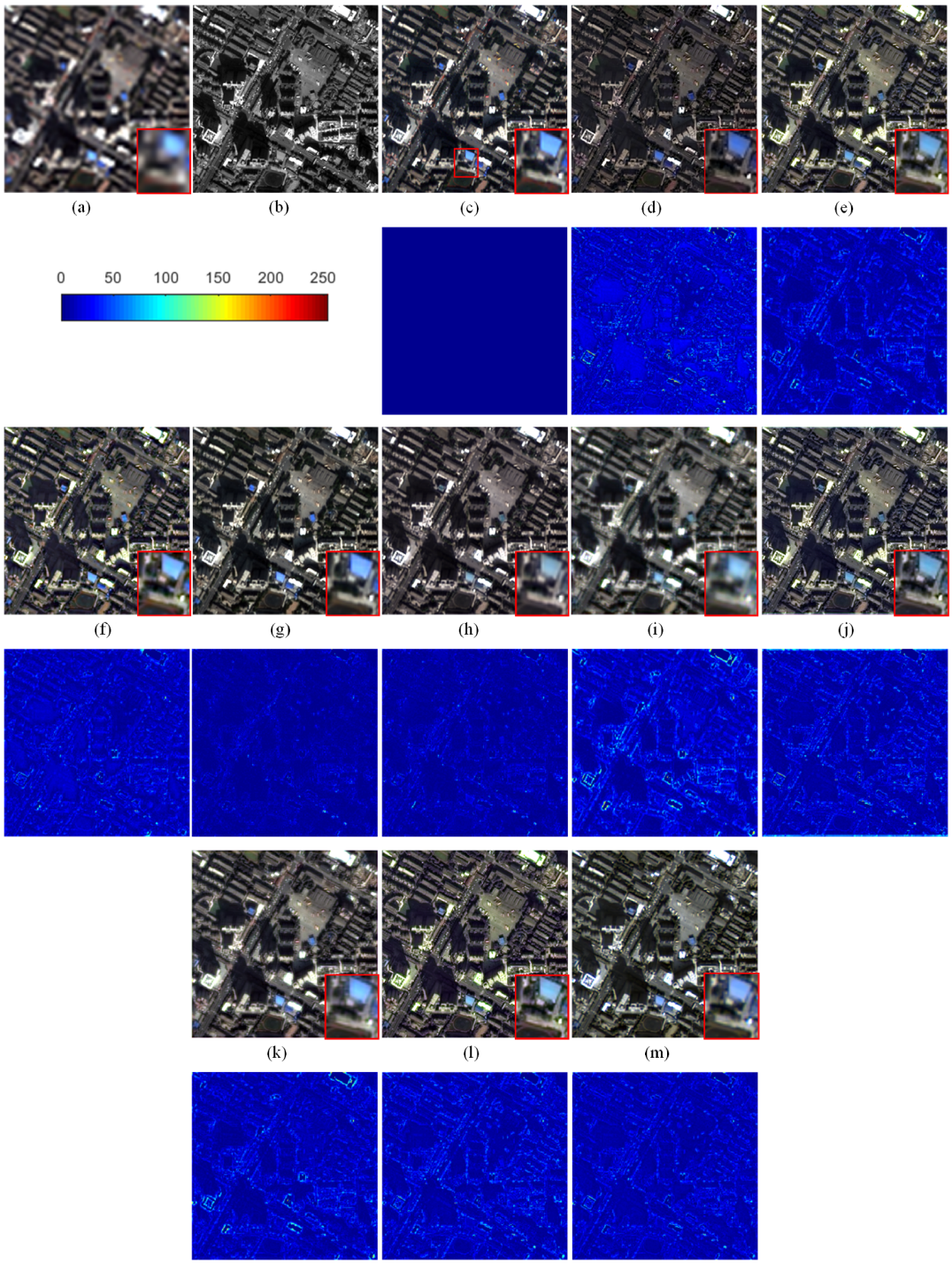

Fig. 6. Qualitative comparison of the reduced-scale QuickBird dataset fused images. (a) LR MS. (b) PAN. (c) Reference. (d) BDSD. (e) P+XS. (f) PNN. (g) PSGAN. (h) M-GAN. (i) PanGAN. (j) LDP-Net. (k) UCGAN. (l) SSCycleGAN. (m) SD-CycleGAN.

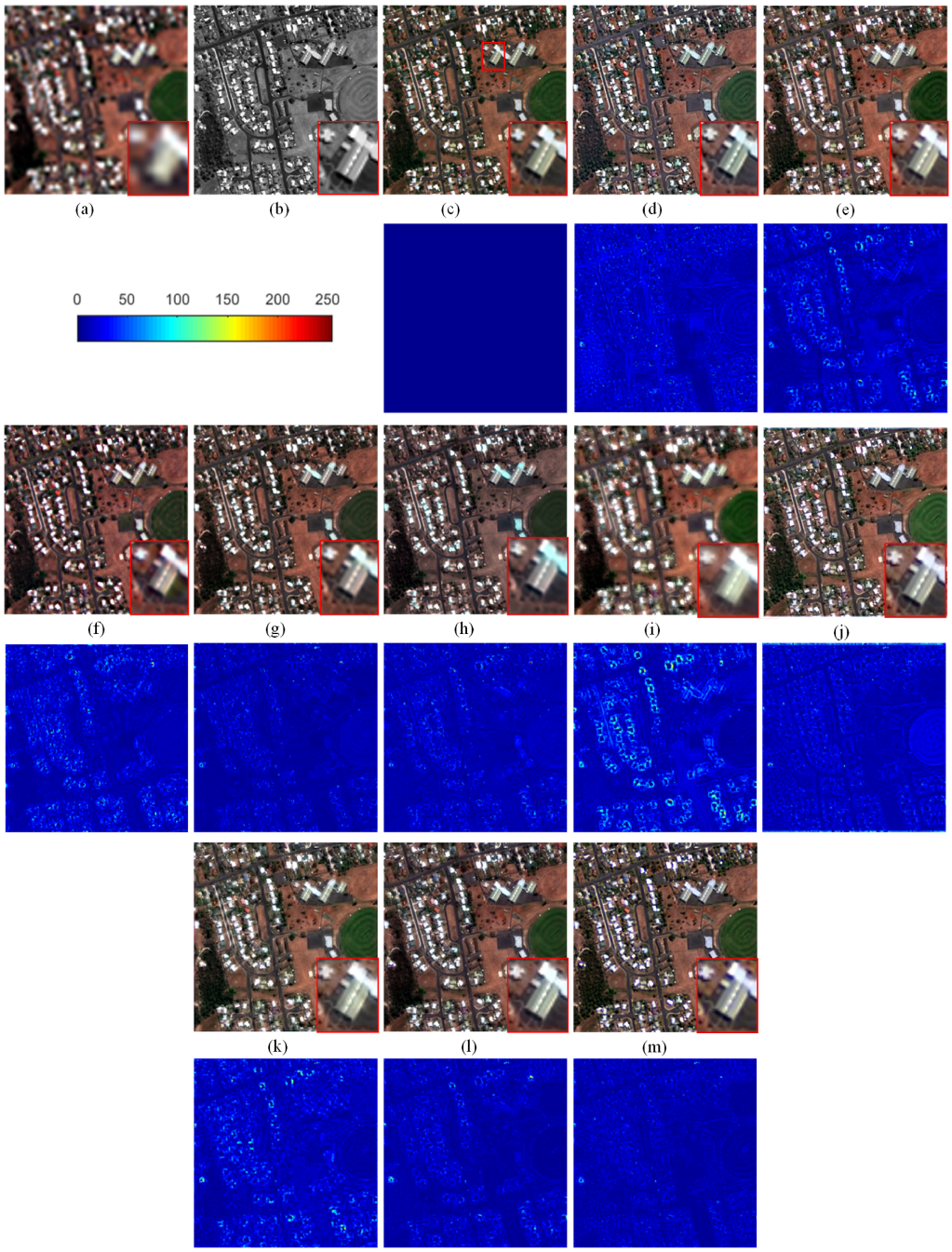

Fig. 7. Qualitative comparison of the reduced-scale GeoEye-1 dataset fused images. (a) LR MS. (b) PAN. (c) Reference. (d) BDSD. (e) P+XS. (f) PNN. (g) PSGAN. (h) M-GAN. (i) PanGAN. (j) LDP-Net. (k) UCGAN. (l) SSCycleGAN. (m) SD-CycleGAN.

C. *Full-Scale DataSet Experiments*

Furthermore, we conducted comparative experiments on the full-scale QuickBird, GeoEye-1, and GF-2 datasets. The qualitative and quantitative comparisons of the proposed and compared methods are depicted in Figs. 9–11 and Tables V–VII, respectively.

In Figs. 9–11, a detailed analysis of the fused images reveals certain characteristics for each method employed. For BDSD fused images, it is observed that the spectral details are excessively enhanced, leading to an overemphasis on the spectral information. Some spectral distortion is evident in the P+XS and LDP-Net fused images. This distortion indicates the presence of inconsistencies or inaccuracies in capturing the true spectral properties of the scene. The PNN fused images exhibit obvious spectral distortion, suggesting a significant deviation from the original spectral content. Spatial artifacts are noticeable in the PSGAN fused images, which may indicate imperfect integration of spatial information from the input images. Significant spectral distortion is observed in the M-GAN fused images, indicating

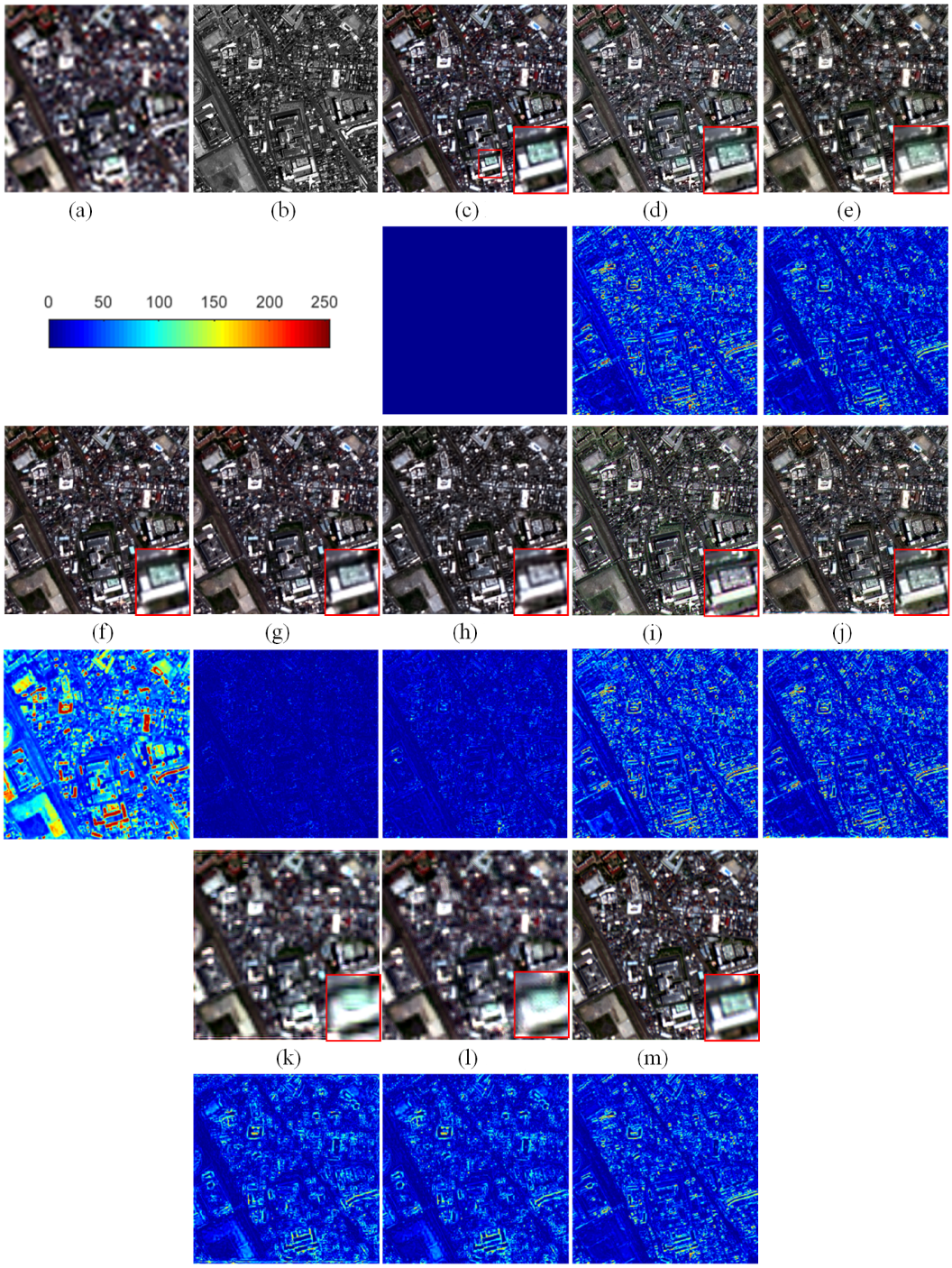

Fig. 8. Qualitative comparison of the reduced-scale GF-2 dataset fused images. (a) LR MS. (b) PAN. (c) Reference. (d) BDSD. (e) P+XS. (f) PNN. (g) PSGAN. (h) M-GAN. (i) PanGAN. (j) LDP-Net. (k) UCGAN. (l) SSCycleGAN. (m) SD-CycleGAN.

substantial alterations or deviations in the spectral characteristics of the scene. In the enlarged areas of the PanGAN fused images, the building edges appear fuzzy, suggesting a loss of sharpness or clarity in the representations of these edges. Comparatively, the fused images obtained from UCGAN exhibit some spatial differences when compared with the reference images. Subtle spectral distortion is present in the fused images produced by SSCycleGAN, implying minor deviations or inconsistencies in the spectral representations of the scenes. When compared with other methods, the proposed SD-CycleGAN fused images have better visual performance. This suggests that the SD-CycleGAN effectively preserves the spectral and spatial information, resulting in fused images with improved overall quality and fidelity.

The best values of traditional, supervised, and unsupervised methods are indicated in bold, respectively. Tables V–VII show that the fused images of SD-CycleGAN perform better in spatial

TABLE II
AVERAGE QUANTITATIVE ANALYSIS OF REDUCED-SCALE QUICKBIRD DATASETS FUSED IMAGES

	Index	RMSE	ERGAS	SAM	Q4	UIQI
Traditional Methods	**BDSD [60]**	**41.3444**	**1.8226**	**3.1447**	**0.8664**	**0.9215**
	P+XS [61]	121.6095	5.3649	4.3113	0.8449	0.8598
Supervised Methods	**PNN [33]**	36.4301	1.6553	4.2041	0.8741	0.9346
	PSGAN [37]	**15.7709**	**0.6959**	**1.7358**	**0.9479**	**0.9857**
	M-GAN [36]	35.5564	1.5736	2.7618	0.8736	0.9205
Unsupervised Methods	**PanGAN [22]**	34.9118	1.5237	3.1995	0.8574	0.9224
	LDP-Net [29]	35.8977	1.5984	3.0972	0.8864	0.9219
	UCGAN [26]	35.7197	1.5755	2.8379	0.8558	0.9186
	SSCycleGAN [27]	31.2334	1.3718	**2.6660**	0.8866	0.9391
	Proposed SD-CycleGAN	**28.9953**	**1.2741**	3.1213	**0.9114**	**0.9538**
Ideal value		0	0	0	1	1

TABLE III
AVERAGE QUANTITATIVE ANALYSIS OF REDUCED-SCALE GEOEYE-1 DATASETS FUSED IMAGES

	Index	RMSE	ERGAS	SAM	Q4	UIQI
Traditional Methods	**BDSD [60]**	**29.6742**	**1.8975**	**5.9940**	**0.7815**	**0.9365**
	P+XS [61]	44.4760	2.9445	5.8446	0.7748	0.9222
Supervised Methods	**PNN [33]**	33.0988	2.1108	7.0581	0.7556	0.9245
	PSGAN [37]	**22.9820**	**1.4597**	**4.4784**	**0.8054**	**0.9556**
	M-GAN [36]	41.6747	2.7468	5.2310	0.7065	0.8378
Unsupervised Methods	**PanGAN [22]**	31.9981	2.0682	**4.9724**	0.7156	0.9026
	LDP-Net [29]	29.4963	1.8719	5.8026	0.7758	0.9197
	UCGAN [26]	28.2518	1.8095	5.4077	0.7625	0.9262
	SSCycleGAN [27]	29.1689	1.8669	5.3442	0.7663	0.9186
	Proposed SD-CycleGAN	**26.2875**	**1.6543**	5.4913	**0.8071**	**0.9405**
Ideal value		0	0	0	1	1

TABLE IV
AVERAGE QUANTITATIVE ANALYSIS OF REDUCED-SCALE GF-2 DATASETS FUSED IMAGES

	Index	RMSE	ERGAS	SAM	Q4	UIQI
Traditional Methods	**BDSD [60]**	**117.6551**	**3.6265**	**3.5713**	**0.5123**	**0.6406**
	P+XS [61]	129.0614	4.0632	3.5352	0.6470	0.7659
Supervised Methods	**PNN [33]**	28.0379	0.8693	1.9420	0.9037	0.9740
	PSGAN [37]	**22.3323**	**0.6883**	**1.5872**	**0.9119**	**0.9827**
	M-GAN [36]	43.1395	1.3322	2.7196	0.8693	0.9371
Unsupervised Methods	**PanGAN [22]**	128.8774	3.9461	5.0036	0.5252	0.6481
	LDP-Net [29]	97.5132	3.0806	3.5774	0.5662	0.6915
	UCGAN [26]	73.1148	2.2275	2.9771	0.6434	0.7948
	SSCycleGAN [27]	67.4087	2.0736	2.7390	0.6273	0.7945
	Proposed SD-CycleGAN	**53.2287**	**1.6717**	**2.5103**	**0.7664**	**0.8848**
Ideal value		0	0	0	1	1

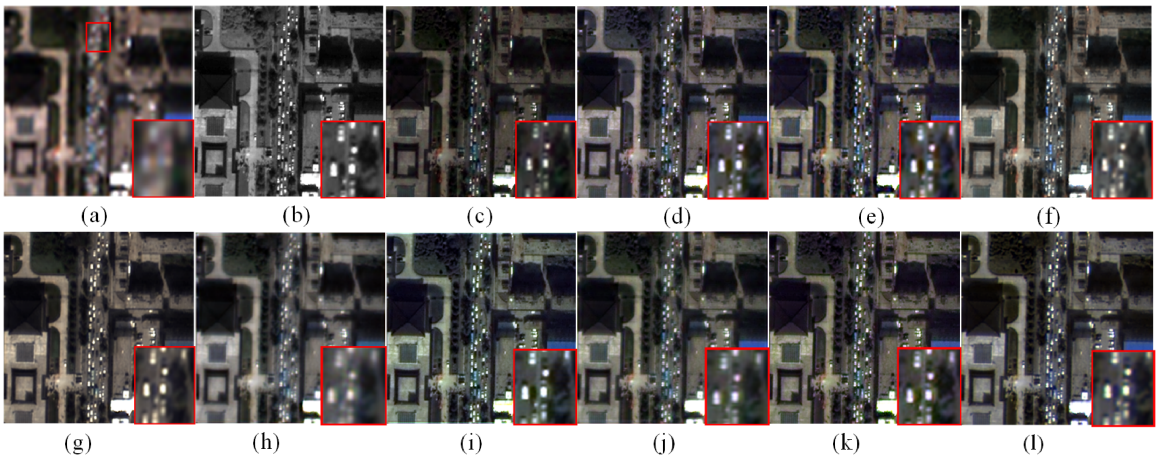

Fig. 9. Qualitative comparison of the full-scale QuickBird dataset fused images. (a) LR MS. (b) PAN. (c) BDSD. (d) P+XS. (e) PNN. (f) PSGAN. (g) M-GAN. (h) PanGAN. (i) LDP-Net. (j) UCGAN. (k) SSCycleGAN. (l) SD-CycleGAN.

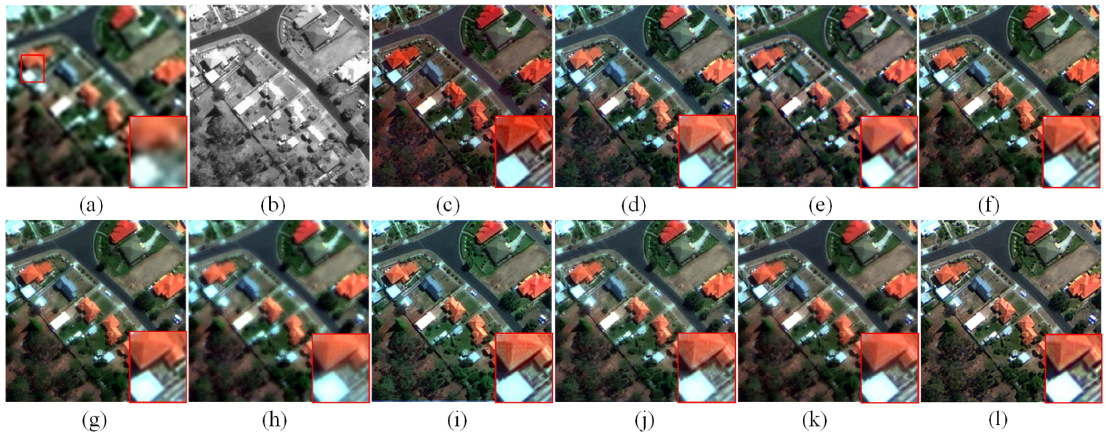

Fig. 10. Qualitative comparison of the full-scale GeoEye-1 dataset fused images. (a) LR MS. (b) PAN. (c) BDSD. (d) P+XS. (e) PNN. (f) PSGAN. (g) M-GAN. (h) PanGAN. (i) LDP-Net. (j) UCGAN. (k) SSCycleGAN. (l) SD-CycleGAN.

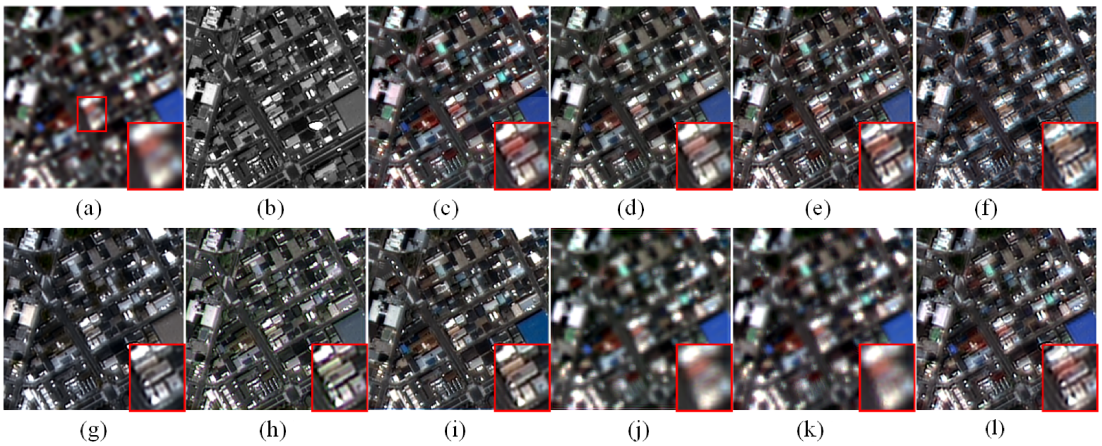

Fig. 11. Qualitative comparison of the full-scale GF-2 dataset fused images.(a) LR MS. (b) PAN. (c) BDSD. (d) P+XS. (e) PNN. (f) PSGAN. (g) M-GAN. (h) PanGAN. (i) LDP-Net. (j) UCGAN. (k) SSCycleGAN. (l) SD-CycleGAN.

TABLE V
AVERAGE QUANTITATIVE ANALYSIS OF FULL-SCALE QUICKBIRD DATASETS FUSED IMAGES

	Index	D_λ	D_S	QNR
Traditional Methods	**BDSD [60]**	0.0593	**0.0325**	**0.9102**
	P+XS [61]	0.1489	0.1227	0.7496
Supervised Methods	**PNN [33]**	0.0475	**0.0440**	**0.9109**
	PSGAN [37]	0.0412	0.0775	0.8846
	M-GAN [36]	**0.0409**	0.1325	0.8321
Unsupervised Methods	**PanGAN [22]**	0.0593	0.0692	0.8773
	LDP-Net [29]	0.0593	**0.0455**	0.8986
	UCGAN [26]	**0.0370**	0.1009	0.9037
	SSCycleGAN [27]	0.0508	0.0723	0.8806
	Proposed SD-CycleGAN	0.0435	0.0531	**0.9059**
Ideal value		0	0	1

TABLE VI
AVERAGE QUANTITATIVE ANALYSIS OF FULL-SCALE GEOEYE-1 DATASETS FUSED IMAGES

	Index	D_λ	D_S	QNR
Traditional Methods	**BDSD [60]**	**0.0594**	**0.0530**	**0.8925**
	P+XS [61]	0.1543	0.0979	0.7753
Supervised Methods	**PNN [33]**	**0.0567**	0.0621	**0.8859**
	PSGAN [37]	0.0880	**0.0545**	0.8656
	M-GAN [36]	0.0796	0.0777	0.8595
Unsupervised Methods	**PanGAN [22]**	**0.0526**	0.1033	0.8513
	LDP-Net [29]	0.9135	0.0870	0.8392
	UCGAN [26]	0.1105	0.0959	0.8165
	SSCycleGAN [27]	0.0904	0.0831	**0.8754**
	Proposed SD-CycleGAN	0.0796	**0.0656**	0.8646
Ideal value		0	0	1

and spectral information. At the same time, the performance of the proposed SD-CycleGAN outperforms both most traditional and some supervised methods.

D. Investigation on Network Architecture of the Spectral and Spatial Degradations

Spatial and spectral degradations are essential factors influencing fusion performance in the proposed method. To evaluate the effectiveness of the proposed network, we compare the spatial and spectral degradations employed in our method with the learnable degenerate modules of LDP-Net [29], as illustrated in Fig. 13. We investigate the effect of spatial and spectral degradations on the performance of image fusion by replacing the proposed spatial and spectral degradations with the learnable degenerate modules of LDP-Net and analyzing the resulting fusion performance. Table VIII shows the quantitative analysis.

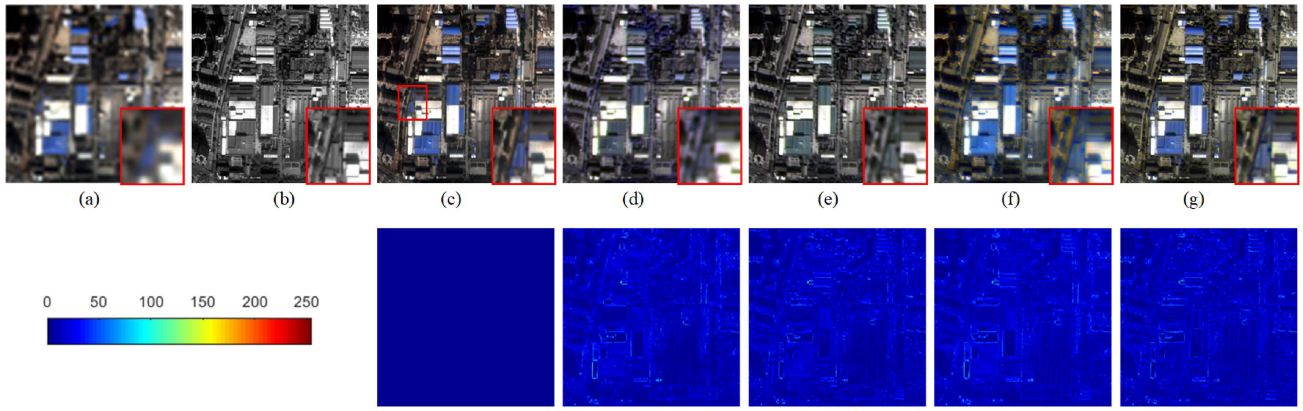

Fig. 12. Qualitative analysis of network structures for different spatial and spectral degradation modules. (a) LR MS. (b) PAN. (c) Reference. (d) Proposed method with Spectral and Spatial Modules of LDP-Net. (e) Proposed method with spectral module of LDP-Net. (f) Proposed method with spatial module of LDP-Net. (g) Proposed method.

TABLE VII
AVERAGE QUANTITATIVE ANALYSIS OF FULL-SCALE GF-2 DATASETS FUSED IMAGES

	Index	D_λ	D_S	QNR
Traditional Methods	**BDSD [60]**	**0.0338**	**0.0584**	**0.9101**
	P+XS [61]	0.1179	0.0884	0.8060
Supervised Methods	**PNN [33]**	**0.0266**	0.1721	**0.8058**
	PSGAN [37]	0.0368	**0.1870**	0.7828
	M-GAN [36]	0.0443	0.1589	0.8035
Unsupervised Methods	**PanGAN [22]**	0.0946	0.1612	0.7618
	LDP-Net [29]	0.0948	0.1438	0.7771
	UCGAN [26]	**0.0480**	0.3187	0.6487
	SSCycleGAN [27]	0.0643	0.3148	0.6413
	Proposed SD-CycleGAN	0.0575	**0.1283**	**0.8215**
Ideal value		0	0	1

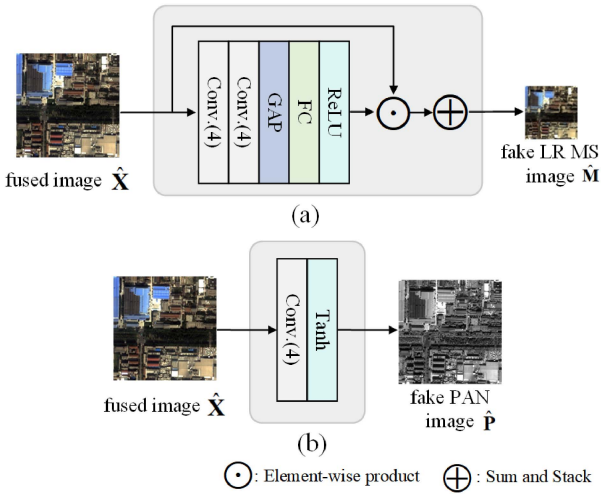

Fig. 13. Different network structures of the spectral and spatial degradations of LDP-Net [29]. (a) Spatial degradation of LDP-Net [29]. (b) Spectral degradation of LDP-Net [29].

Fig. 12 displays fused images under various architectures for qualitative analysis.

The performance of the proposed method exhibits significant superiority compared with other structures, as demonstrated in Table VIII. The results indicate that when the spectral degradation module of LDP-Net [29] is combined with our proposed structures, the fused image experiences a degradation in spectral information. In spectral degradation, the model can dynamically learn the spectral relationships between different bands. By introducing a pixelwise self-attention mechanism in the channel dimension, this module can better preserve and utilize the correlations among multiple channels. Similarly, the fusion of the learnable spatial degradation modules from LDP-Net [29] with our proposed structures leads to a blurring of spatial details in the fused image. This can be attributed to the rationale behind our proposed modules, which are designed to better simulate both spectral and spatial degradation processes. In spatial degradation, by using an approximate ideal Gaussian filter, it is possible to maintain the similarity between the degraded fused images and the original MS images as much as possible after the fused images degrade to their original resolution. This can accurately simulate the process of degrading the HR images to LR images, thereby providing a more accurate degradation model for image fusion and further enhancing the fusion effect. In comparison with the learnable degradation module in LDP-Net [29], our proposed method consistently outperforms alternative structures in terms of fusion quality and preserving spectral and spatial features.

E. Ablation Study

Unsupervised loss functions are utilized in the unsupervised network to preserve image spatial and spectral information. This section focuses on examining the impact of each loss function mentioned in Section III-C on the performance of the proposed method. To carry out this analysis, an ablation experiment is conducted on the reduced-scale testing dataset, as illustrated in Fig. 14. "w/o" is an abbreviation for "without", indicating that it does not contain an item. The CC loss helps ensure that the mapping from one domain to another and back again is consistent, thereby preserving the spectral and spatial characteristics of the fused image. Without CC loss, the fused image exhibits severe distortion both in spectral and spatial information. This indicates

TABLE VIII
QUANTITATIVE ANALYSIS OF NETWORK STRUCTURES FOR DIFFERENT SPATIAL AND SPECTRAL DEGRADATION MODULES

Spectral Module of LDP-Net	Spatial Module of LDP-Net	Spectral Module of proposed SD-CycleGAN	Spatial Module of proposed SD-CycleGAN	RMSE	ERGAS	SAM	Q4	UIQI
✓	✓			36.6187	1.6102	4.3539	0.8908	0.9182
	✓	✓		37.3834	1.6732	3.7742	0.8995	0.9324
✓			✓	37.5379	1.6467	4.3744	0.8940	0.8921
		✓	✓	**28.9953**	**1.2141**	**3.1213**	**0.9114**	**0.9538**

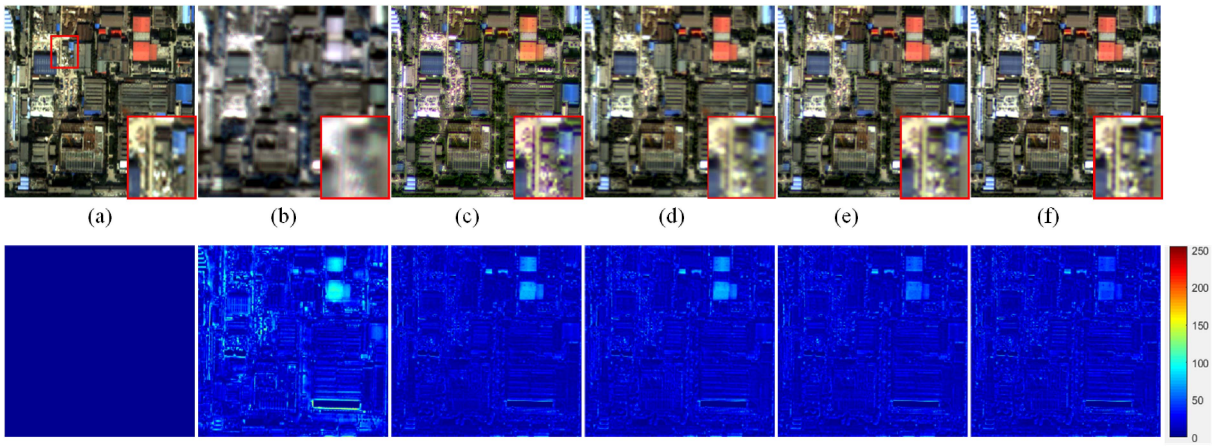

Fig. 14. Visual comparison of the fused images with different loss functions. (a) Reference. (b) w/o CC loss. (c) w/o SAM loss. (d) w/o EE loss. (e) w/o Adversarial loss. (f) SD-CycleGAN.

TABLE IX
ABLATIONS STUDY ON THE REDUCED-SCALE QUICKBIRD DATASETS

Index	RMSE	ERGAS	SAM	Q4	UIQI
w/o CC loss	439.9318	19.1122	8.3093	0.3725	0.4719
w/o SAM loss	31.1110	1.3818	3.5631	0.9107	0.9344
w/o EE loss	31.7742	1.3992	3.5962	0.9099	0.9415
w/o Adversarial loss	29.1527	1.2856	3.3436	**0.9148**	0.9466
Proposed SD-CycleGAN	**28.9953**	**1.2741**	**3.1213**	0.9114	**0.9538**

The bold entities indicate the best quantitative analysis values for each evaluation metric among the ablation studies, respectively.

that the CC loss plays a crucial role in preserving the spectral and spatial characteristics of the fused image. When the SAM loss is excluded, spectral distortion becomes apparent in the fused image, indicating that the SAM loss helps in eliminating spectral artifacts generated during the fusion process. The EE loss term aims to enhance the spatial details and edges in the fused image, preventing blurring or loss of fine structural information during the fusion process. In the absence of the EE loss, the fused image tends to appear blurred, highlighting the contribution of the EE loss in enhancing the spatial details of the fused image. Furthermore, the application of the adversarial loss proves effective in reducing spectral distortion and artifacts in the fused image. The experimental results demonstrate the effectiveness of the proposed unsupervised loss functions through both qualitative and quantitative analyses, showcasing their ability to preserve the spatial and spectral information of the fused images.

The effectiveness of each unsupervised loss function is further supported by the quantitative analysis presented in Table IX. Each unsupervised loss function contributes significantly to network performance.

F. Analysis of Parameter Setting in Unsupervised Loss Functions

In this section, we investigate the impact of adjusting the parameters in the unsupervised loss function L within the proposed SD-CycleGAN. Specifically, we analyze the effects of varying the values of parameters α, β, δ, and γ on the fused images. These parameters control different aspects of the unsupervised loss functions in terms of spatial and spectral terms. The spatial term of the CC loss is controlled by parameter α, whereas the spectral term is controlled by β. Increasing α enhances the spatial details, but may introduce blurriness in the fused images. On the other hand, reducing β can lead to spectral distortion in the fusion results. Parameters δ and γ are responsible for preserving additional spectral and spatial information, respectively. Gradually increasing δ enhances the spectral information, but may result in excessive enhancement. Similarly, increasing γ improves the spatial details, but it may cause instability in the spectral information. To examine the effects of these parameter adjustments, experiments are conducted by using the QuickBird reduced-scale testing dataset. The optimal values for the different parameters are highlighted in bold. As shown in Fig. 15, when the value of α is small, the spatial details of the fused images are relatively blurry. When the value of β is small, the fusion result exhibits spectral distortion. As δ increases gradually, the spectral information of the fused images is excessively enhanced. When γ increases gradually, the spatial details become richer, but the spectral information becomes unstable. Similarly, as shown in Table X, when these parameters are in a relatively balanced state, most of the quality evaluation

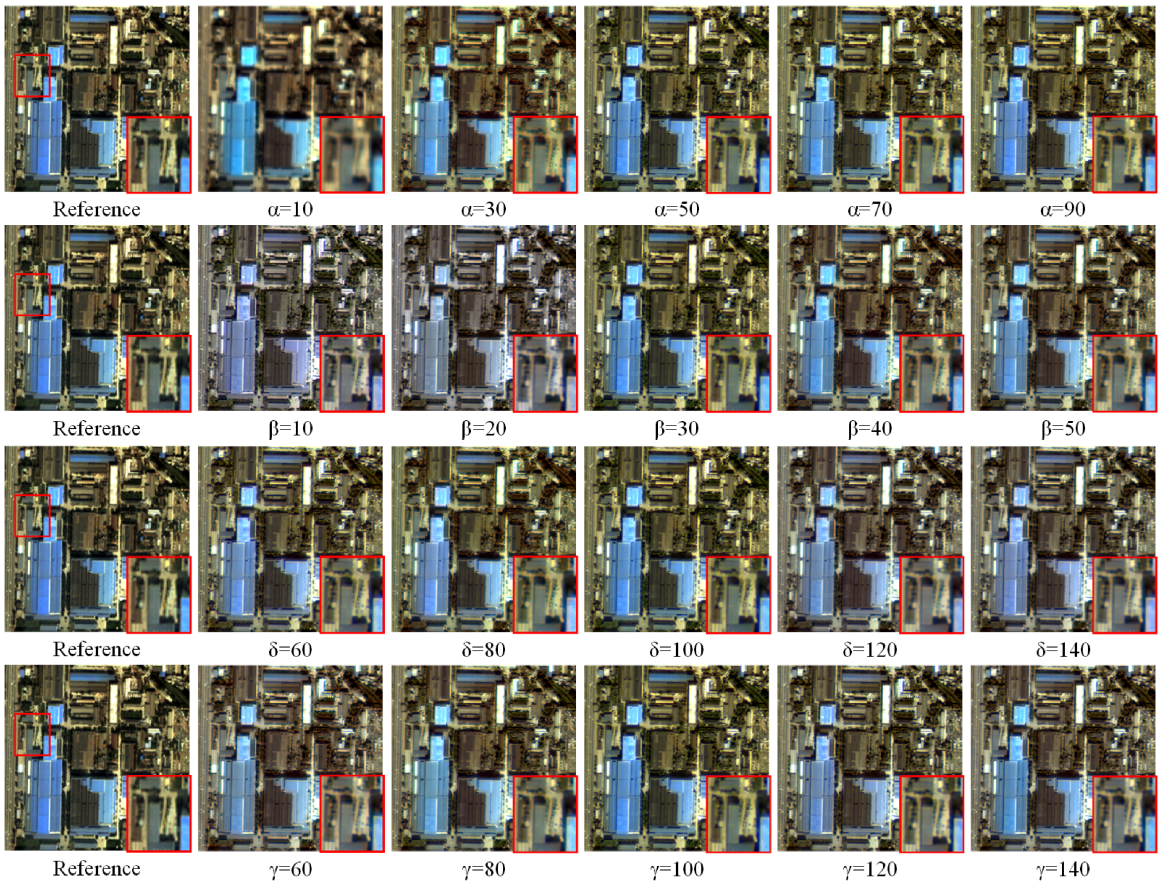

Fig. 15. Proposed SD-CycleGAN visual analysis in relation to different parameters.

TABLE X
PERFORMANCE OF THE PROPOSED SD-CYCLEGAN AGAINST SEVERAL PARAMETERS

α	β	δ	γ	RMSE	ERGAS	SAM	Q4	UIQI
10	30	100	100	33.8761	1.5359	4.1147	0.8959	0.9406
30	30	100	100	30.1638	1.3376	3.3574	**0.9127**	0.9427
50	30	100	100	**28.9953**	**1.2741**	**3.1213**	0.9114	**0.9538**
70	30	100	100	30.3361	1.3299	3.4385	0.9109	0.9427
90	30	100	100	30.3971	1.3508	3.4793	0.9107	0.9428
50	10	100	100	31.4724	1.3908	3.7614	0.9081	0.9373
50	20	100	100	29.9614	1.3154	3.3343	0.9102	0.9466
50	**30**	100	100	**28.9953**	**1.2741**	**3.1213**	0.9114	**0.9538**
50	40	100	100	30.7927	1.3544	3.4797	0.9128	0.9367
50	50	100	100	30.6850	1.3754	3.4568	**0.9156**	0.9432
50	30	60	100	33.4497	1.4426	3.9586	0.9073	0.9319
50	30	80	100	30.3852	1.3317	3.3091	0.9081	0.9464
50	30	**100**	100	**28.9953**	**1.2741**	**3.1213**	0.9114	**0.9538**
50	30	120	100	30.1053	1.3240	3.3772	**0.9144**	0.9355
50	30	140	100	29.9127	1.3158	3.3710	0.9142	0.9425
50	30	100	60	29.3617	1.3122	3.4008	0.9126	0.9460
50	30	100	80	30.0358	1.3126	3.3653	0.9127	0.9417
50	30	100	**100**	**28.9953**	**1.2741**	**3.1213**	0.9114	**0.9538**
50	30	100	120	30.2588	1.3319	3.4508	**0.9137**	0.9367
50	30	100	140	31.0962	1.3702	3.3971	0.9077	0.9300
Ideal Value				0	0	0	1	1

metrics reach their optimal values. This analysis demonstrates the importance of fine-tuning these parameter values to achieve optimal performance in terms of spectral and spatial fidelity in the fused images. As shown in Table X and Fig. 15, the best fusion results are obtained when α, β, δ, and γ are set to 50, 30, 100, and 100, respectively.

G. Investigation on Network Architecture of the Generator

We represent the structure of the generator G in Fig. 3 as a block with dense connections. We believe that using dense connections and the number of blocks are two important factors influencing the fusion performance of G in the proposed SD-CycleGAN. We discuss the impact of different structures of G on the fused images. For example, using a single block without dense connections or multiple blocks with dense connections.

Table XII presents the variations in quantitative analysis metrics for different structures of G. From Table XII, it is observed that the proposed SD-CycleGAN outperforms other architectures in terms of the quantitative analysis metrics compared with the different structures of G mentioned above. With the introduction of dense connections, there is a significant improvement in the quantitative analysis values in Table XII. Although the computational complexity increases significantly with the number of blocks, the quantitative analysis values do not improve. The qualitative analysis of different network structures of G is shown in Fig. 16. Compared with the fused image from SD-CycleGAN, fused images without dense connections exhibit noticeably poorer spectral and spatial information. In addition, it can be observed that with an increased number of blocks, most of the spectral information is lost in the fused images. Table XII and Fig. 16 demonstrate the superior performance of the proposed

TABLE XI
TIME COMPARISON AND MODEL SIZE ANALYSIS

Index	PNN	PSGAN	M-GAN	PanGAN	LDP-Net	UCGAN	SSCycleGAN	Proposed SD-CycleGAN
Test time (ms)	0.8	7.2	11.7	3.6	6.1	8.6	8.7	2.2
#Para. (M)	0.08	3.02	15.51	0.88	0.11	2.83	0.37	0.85

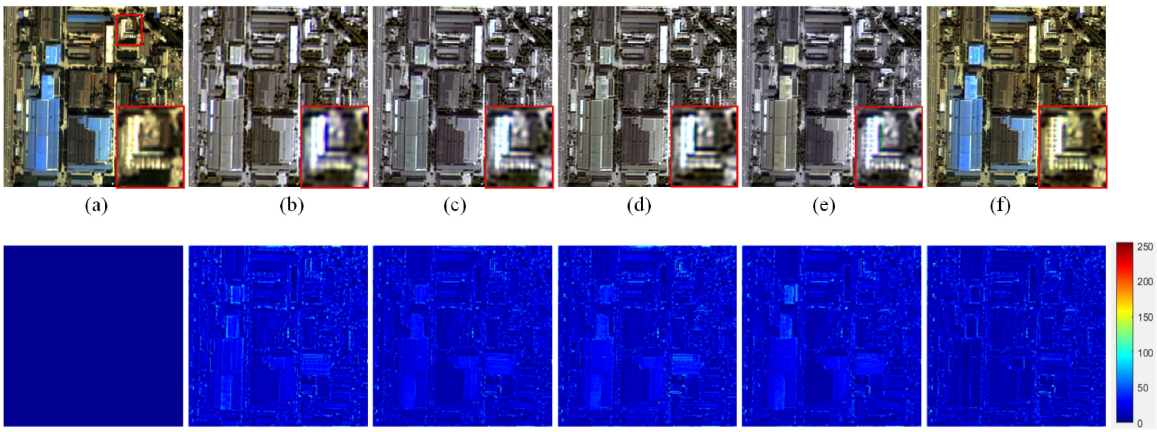

Fig. 16. Qualitative comparison of the fused images with different generator structures. (a) Reference. (b) w/o Dense connection. (c) 2 Blocks. (d) 3 Blocks. (e) 4 Blocks. (f) SD-CycleGAN.

TABLE XII
QUANTITATIVE COMPARISON OF THE FUSED IMAGES WITH DIFFERENT NETWORK STRUCTURES

Index	RMSE	ERGAS	SAM	Q4	UIQI
w/o Dense Connection	31.7333	1.3986	3.5500	0.9073	0.9356
2 Blocks	29.6008	1.3218	3.5149	**0.9130**	0.9492
3 Blocks	32.3228	1.4363	3.5456	0.8993	0.9424
4 Blocks	31.9297	1.4457	3.9499	0.9063	0.9455
Proposed SD-CycleGAN	**28.9953**	**1.2741**	**3.1213**	0.9114	**0.9538**

The bold entities highlight the best quantitative analysis values for each evaluation metric among the various generator network structures.

generator structure compared with others. Therefore, the architecture used in SD-CycleGAN is a good choice for modeling spatial and spectral information.

H. Running Time and Model Size

In Table XI, we analyze the complexity of SD-CycleGAN and DNN-based comparison methods, which are trained and tested on NVIDIA GeForce RTX 3090 and Intel (R) Core (TM) i7-9700KF CPU @3.60 GHz. Based on the values in Table XI, it is feasible to conclude that SD-CycleGAN has fewer parameters and less testing time than the compared methods. As a result, the computational complexity and model size of SD-CycleGAN perform well.

V. CONCLUSION

To generate improved fusion images, we propose an unsupervised single-generator CycleGAN containing the spatial and spectral degradation processes, called SD-CycleGAN. This method is based on the well-established CycleGAN framework. To simplify the unsupervised CycleGAN framework, our proposed method utilizes only one generator. In addition, the method incorporates modules that can simulate spatial and spectral degradation processes to facilitate the unsupervised learning process. Moreover, to preserve spatial details and spectral information, we introduce a set of unsupervised losses to enhance the spatial details and reduce spectral distortion in the fused images. Compared with state-of-the-art methods, experimental results on the QuickBird, GeoEye-1, and GF-2 datasets demonstrate that the fusion images produced by SD-CycleGAN contain more spatial and spectral information.

REFERENCES

[1] 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, 2023.
[2] X. Wei and M. Yuan, "Adversarial pan-sharpening attacks for object detection in remote sensing," *Pattern Recognit.*, vol. 139, 2023, Art. no. 109466.
[3] L. Sun, C. He, Y. Zheng, Z. Wu, and B. Jeon, "Tensor cascaded-rank minimization in subspace: A unified regime for hyperspectral image low-level vision," *IEEE Trans. Image Process.*, vol. 32, pp. 100–115, 2023.
[4] L. Wang, Z. Xiong, G. Shi, W. Zeng, and F. Wu, "Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 10, pp. 2104–2111, Oct. 2017.
[5] X. Meng et al., "A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 18–52, Mar. 2021.
[6] M. Zhou, K. Yan, J. Huang, Z. Yang, X. Fu, and F. Zhao, "Mutual information-driven pan-sharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1788–1798.
[7] W. J. Carper, T. M. Lillesand, and R. W. Kiefer, "The use of intensity hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogrammetric Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, 1990.
[8] P. S. Chavez, S. C. Slides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogrammetric Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.

[9] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent US06011875A, Jan. 4, 2000.

[10] 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, 1988.

[11] P. Wang, H. Yao, C. Li, G. Zhang, and H. Leung, "Multiresolution analysis based on dual-scale regression for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5406319.

[12] P. Wang, H. Yao, B. Huang, H. Leung, and P. Liu, "Multiresolution analysis pansharpening based on variation factor for multispectral and panchromatic images from different times," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5401217.

[13] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3 pp. 1204–1211, May 1999.

[14] K. P. Upla, M. V. Joshi, and P. P. Gajjar, "An edge preserving multiresolution fusion: Use of contourlet transform and MRF prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3210–3220, Jun. 2015.

[15] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inf. Fusion*, vol. 8, pp. 143–156, 2007.

[16] 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, 2022.

[17] H. Yin, "Panchromatic side sparsity model-based deep unfolding network for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5406715.

[18] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9 pp. 4779–4789, Sep. 2013.

[19] S. Yang, K. Zhang, and M. Wang, "Learning low-rank decomposition for pan-sharpening with spatial-spectral offsets," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3647–3657, Aug. 2018.

[20] M. Ghahremani and H. Ghassemian, "A compressed-sensing-based pansharpening method for spectral distortion reduction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2194–2206, Apr. 2016.

[21] F. Palsson, M. Ulfarsson, and J. Sveinsson, "Model-based reduced rank pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 656–660, Apr. 2020.

[22] J. Ma et al., "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inf. Fusion*, vol. 62, pp. 110–120, 2020.

[23] H. Zhou, Q. Liu, and Y. Wang, "PGMAN: An unsupervised generative multi adversarial network for pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6316–6327, 2021.

[24] Q. Xu, Y. Li, J. Nie, Q. Liu, and M. Guo, "UPanGAN: Unsupervised pansharpening based on the spectral and spatial loss constrained generative adversarial network," *Inf. Fusion*, vol. 91, pp. 31–46, Mar. 2023.

[25] Y. Wang, Y. Xie, Y. Wu, K. Liang, and J. Qiao, "An unsupervised multi-scale generative adversarial network for remote sensing image pan-sharpening," in *Proc. Int. Conf. Multimedia Model.*, 2022, pp. 356–368.

[26] H. Zhou, Q. Liu, D. Weng, and Y. Wang, "Unsupervised cycle-consistent generative adversarial networks for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5408814.

[27] J. Li, W. Sun, M. Jiang, and Q. Yuan, "Self-supervised pansharpening based on a cycle-consistent generative adversarial network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5511805.

[28] S. Luo, S. Zhou, Y. Feng, and J. Xie, "Pansharpening via unsupervised convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4295–4310, 2020.

[29] J. Ni et al., "LDP-Net: An unsupervised pansharpening network based on learnable degradation processes," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5468–5479, 2022.

[30] H. Zhang, H. Wang, X. Tian, and J. Ma, "P2Sharpen: A progressive pansharpening network with deep spectral transformation," *Inf. Fusion*, vol. 91, pp. 103–122, 2023.

[31] G. Zhao, Q. Ye, L. Sun, Z. Wu, C. Pan, and B. Jeon, "Joint classification of hyperspectral and LiDAR data using a hierarchical CNN and transformer," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5500716.

[32] Y. Fang, Q. Ye, L. Sun, Y. Zheng, and Z. Wu, "Multiattention joint convolution feature representation with lightweight transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5513814.

[33] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 594.

[34] K. Zhang, A. Wang, F. Zhang, W. Diao, J. Sun, and L. Bruzzone, "Spatial and spectral extraction network with adaptive feature fusion for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5410814.

[35] W. Zhang, J. Li, and Z. Hua, "Attention-based Tri-UNet for remote sensing image pan-sharpening," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3719–3732, 2021.

[36] A. Gastineau, J. Aujol, Y. Berthoumieu, and C. Germain, "Generative adversarial network for pansharpening with spectral and spatial discriminators," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4401611.

[37] Q. Liu, H. Zhou, Q. Xu, X. Liu, and Y. Wang, "PSGAN: A generative adversarial network for remote sensing image," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10 227–10 242, Dec. 2021.

[38] W. G. C. Bandara and V. M. Patel, "HyperTransformer: A textural and spectral feature fusion transformer for pansharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1757–1767.

[39] K. Zhang, Z. Li, F. Zhang, W. Wan, and J. Sun, "Pan-sharpening based on transformer with redundancy reduction," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5513205.

[40] F. Zhang, K. Zhang, and J. Sun, "Multiscale spatial–spectral interaction transformer for pan-sharpening," *Remote Sens.*, vol. 14, no. 7, Apr. 2022, Art. no. 1736.

[41] M. Zhou et al., "Spatial-frequency domain information integration for pan-sharpening," in *Proc. Eur. Conf. Comput. Vis.*, pp. 274–291, 2022.

[42] M. Zhou, K. Yan, J. Huang, Z. Yang, X. Fu, and F. Zhao, "Mutual information-driven pan-sharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1788–1798.

[43] M. Zhou, X. Fu, J. Huang, F. Zhao, A. Liu, and R. Wang, "Effective pan-sharpening with transformer and invertible neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5406815.

[44] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[45] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2021, pp. 1–21.

[46] Y. Yan, J. Liu, S. Xu, Y. Wang, and X. Cao, "MD^{3}Net: Integrating model-driven and data-driven approaches for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5411116.

[47] Z. Wu, T. Huang, L. Deng, J. 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, 2022, Art. no. 5401016.

[48] K. Zhang, A. Wang, F. Zhang, W. Wan, J. Sun, and L. Bruzzone, "Spatial–spectral dual back-projection network for pansharpening," *IEEE Trans. Geos. Remote Sens.*, vol. 61, 2023, Art. no. 5402216.

[49] Z. Xiang, L. Xiao, J. Yang, W. Liao, and W. Philips, "Detail-injection-model-inspired deep fusion network for pansharpening," *IEEE Trans. Geos. Remote Sens.*, vol. 60, 2022, Art. no. 5411315.

[50] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, Jun. 1997.

[51] K. Zheng et al., "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2487–2502, Mar. 2021.

[52] J. Liu, Z. Wu, L. Xiao, and X. -J. Wu, "Model inspired autoencoder for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5522412.

[53] J. -Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.

[54] B. Sim, G. Oh, and J. C. Ye, "Optimal transport structure of CycleGAN for unsupervised learning for inverse problems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 8644–8647.

[55] J. Peng et al., "PSMD-Net: A novel pan-sharpening method based on a multiscale dense network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4957–4971, Jun. 2021.

[56] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3880–3891, Nov. 2009.

[57] D. Ingrid, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909–996, 1988.

[58] D. Li et al., "Involution: Inverting the inherence of convolution for visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021 pp. 12 316–12 325.
[59] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
[60] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 228–236, Jan. 2008.
[61] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for p XS image fusion," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 43–58, 2006.
[62] L. Wald, "Quality of high resolution synthesized images: Is there a simple criterion?," in *Proc. 3rd Conf. Fusion Earth Data*, 2000, pp. 99–105.
[63] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.
[64] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 313–317, Oct. 2004.
[65] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[66] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogrammetric Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.

Wenxiu Diao received the B.S. and M.S. degrees in computer science and technology from Shandong Normal University, Jinan, China, in 2019 and 2022, respectively. She is currently working toward the Ph.D. degree in computer science and technology with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China.

Her research interests include image processing and deep learning.

Mengying Jin received the B.S. degree in mathematics and applied mathematics from Yancheng Teachers University, Yancheng, China, in 2014, and the M.Sc. degree in maths from the Nanjing University of Information Science and Technology, Nanjing, China, in 2020. She is currently working toward the Ph.D. degree with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing.

Her main current research interests include image processing and deep learning.

Kai Zhang (Member, IEEE) was born in Shanxi, China, in 1992. He received the B.S. degree in electrical engineering and automation from the North University of China, Taiyuan, China, in 2013, and the Ph.D. degree in circuits and systems from Xidian University, Xi'an, China, in 2018.

From November 2021 to November 2022, he was a Postdoctoral Fellow with the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento, Trento, Italy. He is currently an Associate Professor with the School of Information Science and Engineering, Shandong Normal University, Jinan, China. His research interests include multisource remote sensing image fusion, change detection, and deep learning.

Liang Xiao (Member, IEEE) received the B.S. degree in applied mathematics and the Ph.D. degree in computer science from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 1999 and 2004, respectively.

From 2006 to 2008, he was a Postdoctoral Research Fellow with the Pattern Recognition Laboratory, NJUST. From 2009 to 2010, he was a Postdoctoral Fellow with Rensselaer Polytechnic Institute, Troy, NY, USA. Since 2013, he has been the Deputy Director with the Jiangsu Key Laboratory of Spectral Imaging Intelligent Perception, NJUST. Since 2014, he has been the second Director with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, NJUST, where he is a Professor with the School of Computer Science and Engineering. His research interests include remote sensing image processing, image modeling, computer vision, machine learning, and pattern recognition.