Generative Dual-Adversarial Network With Spectral Fidelity and Spatial Enhancement for Hyperspectral Pansharpening

Wenqian Dong, Member, IEEE, Shaoxiong Hou, Song Xiao, Jiahui Qu, Member, IEEE, Qian Du, Fellow, IEEE, and Yunsong Li, Member, IEEE

Abstract—Hyperspectral (HS) pansharpening is of great importance in improving the spatial resolution of HS images for remote sensing tasks. HS image comprises abundant spectral contents, whereas panchromatic (PAN) image provides spatial information. HS pansharpening constitutes the possibility for providing the pansharpened image with both high spatial and spectral resolution. This article develops a specific pansharpening framework based on a generative dual-adversarial network (called PS-GDANet). Specifically, the pansharpening problem is formulated as a dual task that can be solved by a generative adversarial network (GAN) with two discriminators. The spatial discriminator forces the intensity component of the pansharpened image to be as consistent as possible with the PAN image, and the spectral discriminator helps to preserve spectral information of the original HS image. Instead of designing a deep network, PS-GDANet extends GANs to two discriminators and provides a high-resolution pansharpened image in a fraction of iterations. The experimental results demonstrate that PS-GDANet outperforms several widely accepted state-of-the-art pansharpening methods in terms of qualitative and quantitative assessment.

Index Terms—Generative dual-adversarial network, hyperspectral (HS) pansharpening, spatial discriminator, spectral discriminator.

I. INTRODUCTION

HYPERSPECTRAL (HS) pansharpening aims to combine an HS image and a panchromatic (PAN) image acquired for the same scene to produce a pansharpened image featuring both high spectral and spatial resolution, which cannot be achieved by a single sensor because of physical constraints. Pansharpening is widely used for enhancing images in some remote sensing applications, such as change detection [1], [2], snow mapping [3], and spatial feature extraction [4]. Furthermore, the demand for high-resolution images in some commercial products such as Bing maps and Google Earth continues to grow. Pansharpening has aroused wide interest from many investigators.

Over the past decades, various pansharpening methods have been proposed and they can be roughly divided into two main families: classical simple model-based methods and recent complex model-based approaches [5], [6]. The latter techniques are superior to the classical methods at the expense of computational complexity. This family includes the coupled nonnegative matrix factorization (CNMF) [7], HySure [8], Bayesian sparsity-promoted Gaussian prior (Bayesian sparse) [9], and Bayesian Naïve Gaussian prior (Bayesian Naïve) [10] using the sparse representations and total variation penalization terms to restore the ideal HS image from its degraded version. However, it is difficult and time consuming to solve the ill-posed inverse problem.

The component substitution (CS)-based pansharpening techniques are widely accepted because their realization is easy and fast. Specifically, the HS image is decomposed into different principal components or color spaces. Then, the separated spatial component is substituted by the PAN image. The high-resolution HS (HR-HS) images can be finally obtained by inverse spectral transformation. Among the CS-based methods, intensity-hue-saturation (IHS) transform [11], principal component substitution (PCA) [12], Gram–Schmidt (GS) [13], adaptive GS (GSA) [14], guided filter PCA (GFPCA) [15], and improved detail extraction-based method [16] are representative approaches. Generally, the CS-based pansharpening methods can be efficiently implemented. However, these methods approximate PAN images by linear combination of HS bands, and the pansharpening performance depends heavily on the definition of weights. Therefore, the performance of these methods is usually unstable.

The multiresolution analysis (MRA)-based methods obtain the pansharpened result by injecting the high-pass component

Manuscript received 20 November 2020; revised 7 February 2021 and 7 April 2021; accepted 20 May 2021. Date of publication 10 June 2021; date of current version 1 December 2022. This work was supported in part by the National Defense Pre-Research Foundation; in part by the Yangtze Grant 2021JQ-194 and Grant 2021JQ-197; in part by the Fundamental Natural Science Basic Research Plan in Shaanxi Province of China under Grant 61501346, Grant 61502367, and Grant 61701360; in part by the National Defense Pre-Research Foundation; in part by the 111 Project of current version 1 December 2022. This work was supported in part by the Ten Thousand Talent Program. (Corresponding authors: Song Xiao; Jiahui Qu.)

Wenqian Dong, Shaoxiong Hou, Jiahui Qu, and Yunsong Li are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi’an 710071, China (e-mail: wqdong@xidian.edu.cn; xhou1997@163.com; jhq@xidian.edu.cn; ysl@mail.xidian.edu.cn).

Song Xiao is with the Beijing Electronic Science and Technology Institute, Beijing 100070, China, and also with the State Key Laboratory of Integrated Service Network, Xidian University, Xi’an 710071, China (e-mail: xiao.song@mail.xidian.edu.cn).

Qian Du is with the Department of Electronic and Computer Engineering, Mississippi State University, Starkville, MS 39759 USA (e-mail: du@ece.msstate.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2021.3084745.

of the PAN image into the interpolated HS image [17], [18].
The details are extracted by different modalities of MRA: curvelets transform [19], wavelet transform [20], and Laplacian pyramid [21]. Because the MRA methods keep the structure of the original HS image unchanged during pansharpening, they usually provide better spectral fidelity performance than the CS-based methods, but may cause serious spatial distortion because of the aliasing effects.

Another family of pansharpening approaches are based on tensor factorization. They define the HS image as a 3-D tensor, and formulate the fusion problem as the estimation of three dictionaries and core tensor. The work [22] formulates the estimation of a core tensor and dictionaries of the three modes as a coupled sparse tensor factorization of the input images. A nonlocal sparse tensor factorization (NLSTF) was proposed by Dian et al. [23]. Based on this, they present a low tensor-train rank (LTTR)-based fusion method, where the spectral, spatial, and nonlocal modes of the similar HS cubes can be effectively learned by an LTTR prior [24]. Compared with the traditional methods, tensor factorization-based methods usually create better results. However, the input HS image in these methods is not used for estimating the core tensor.

In recent years, neural networks, especially convolutional neural networks (CNNs), have shown great potential in many image processing [25]–[27] and visual tasks [28], [29]. Since Huang et al. [30] first proposed a CNN-based pansharpening approach, various CNN-based approaches have been proposed and achieved state-of-the-art pansharpening performance because of their nonlinearity [31]–[34]. Masi et al. [33] incorporated the idea of super-resolution (SR) network and proposed a lightweight architecture for pansharpening (PNN). Inspired by the success of PNN, Wei et al. [35] proposed a very deep CNN architecture with residual learning for pansharpening (DRPNN). Burt and Adelson [36] have successfully designed a multidepth and multiscale CNN structure for pansharpening that adopts a multiscale block to allow the CNN to go deeper for a better performance.

Indeed, these methods provided a novel mean to solve the pansharpening problem by using deep network architectures. However, they were originally developed for multispectral (MS) pansharpening. As one may expect, HS pansharpening can be regarded as a special case of MS pansharpening, but HS images cover a much wider spectral range than MS images. Therefore, how to achieve good pansharpening performance in the spectral range not covered by PAN image remains a great challenge. He et al. [42] attempted to address the HS pansharpening problem by combining the spectral predictive structure into an end-to-end CNN network. Following the idea of MRA, Li et al. [32] developed a detail-based deep Laplacian pansharpening method and provided a remarkable performance. Xie et al. [31] formulated HS pansharpening as a constraint optimization problem and solved it by learning the priors through CNN. He et al. [42] designed a CNN structure with spectral fidelity, which also achieves satisfying pansharpening performance.

Thanks to the nonlinearity of various deep network architectures, the recently proposed CNN-based methods have achieved a remarkable HS pansharpening performance. However, similar to MS pansharpening, most of these CNN architectures are based on SR. Low-resolution-HS (LR-HS) constitutes the sole possibility for creating HR-HS in SR, whereas both HS and PAN images have important contributions to the pansharpened product. Existing CNN-based methods take PAN image as an additional band of HS image for network training. CNNs operate as “black boxes” and process each band indiscriminately, so the contribution of PAN image to the pansharpened image cannot be well achieved. To tackle these problems, we propose a powerful pansharpening method based on generative adversarial network (GAN) with a clear physical interpretation (called PS-GDANet) in this article. The proposed pansharpening framework based on a generative dual-adversarial network (PS-GDANet) intends to preserve the spectral information of the original HS image well, and tries to make the spatial distribution of pansharpened product consistent with that of the PAN image. To achieve this goal, the PS-GDANet is designed with a feature extraction network (FENet), a generator, a spatial discriminator, and a spectral discriminator. On the one hand, the spatial discriminator establishes the adversarial game with the generator to make the spatial distribution of the pansharpened product close as possible to that of the PAN image acquired by the sensor with such a high resolution. On the other hand, the spectral discriminator establishes the adversarial game with the generator to force the spectral fidelity of the pansharpened product to be as consistent as possible with that of the original HS image. The PS-GDANet is superior to some widely used state-of-the-art approaches because of the following advantages:

1) We formulate the pansharpening as a dual-task problem and propose a generative dual-adversarial pansharpening network with two discriminators, in which the spatial discriminator aims to improve spatial resolution and the spectral discriminator aims to maintain spectral information.

2) To effectively extract spatial features, we develop a new unsupervised FENet, which is based on the fact that the spatial component of the HS image acquired by an ideal high-resolution sensor can be approximated by the PAN image and focuses on extracting multilevel details from the HS image by discovering the relationship between HS and PAN images in the spatial domain.

3) We incorporate guided filtering with deep CNN to design a generator specifically for pansharpening to ensure the enhancement of spatial-spectral information.

The rest of this article is organized as follows: the proposed PS-GDANet method is introduced in detail in Section II. In Section III, the experimental results are described. Section IV draws the conclusions, and Section V gives the discussions.

II. PROPOSED METHOD

The flowchart of the proposed PS-GDANet is illustrated in Fig. 1. Unlike traditional GANs, which establish a two-player game between a generator and a discriminator, our proposed PS-GDANet builds a minmax three-player game between one generator and two discriminators according to two different tasks to be solved. Overall, our proposed PS-GDANet model consists of a FENet, a generator, a spatial discriminator, and a spectral discriminator.
In the training phase of the proposed PS-GDANet, the feature maps of HS images at different scales learned from FENet, the PAN image, and the HS image constitute the input samples of the generator. The generator builds a two-player adversarial game with the spatial discriminator, where the generator aims to create a pansharpened image whose intensity component is similar to the PAN image, and the spatial discriminator intends to force the pansharpened image to learn more spatial details from the PAN image. The generator builds another two-player adversarial game with the spectral discriminator. The essence of the adversarial learning between the generator and the spectral discriminator is to train the generator to capture the spectral information of the original HS image. Therefore, through adversarial learning, the generator preserves the spectral content of the original HS image while continuously fitting the spatial distribution of the PAN image into the pansharpened image. When the spatial discriminator cannot distinguish the intensity component of the pansharpened image from the PAN image and the spectral discriminator cannot distinguish the blurred down-sampled version of the pansharpened image from the original HS image, it is considered that the pansharpened image has injected as much spatial details of PAN image as possible, and its spectral distribution is consistent with that of the original image. In other words, LR-HS can be obtained by spatially blurring and then down-sampling of HR-HS, i.e.,

$$\mathbf{H} = \mathbf{H} \mathbf{B} \mathbf{S}$$  \hspace{1cm} (2)

where $\mathbf{S}$ represents the down-sampling matrix, and $\mathbf{B}$ denotes the blur matrix, which can be expressed as

$$\mathbf{B} = \mathbf{F} \mathbf{D} \mathbf{F}^{H}$$  \hspace{1cm} (3)

where $\mathbf{F}$ and $\mathbf{F}^{H}$ represent discrete-Fourier transform and inverse fast Fourier transform matrices, respectively, and $\mathbf{D}$ denotes a diagonal matrix, which contains the eigenvalues of $\mathbf{B}$. 

Authorized licensed use limited to: Mississippi State University Libraries. Downloaded on December 08,2022 at 05:25:11 UTC from IEEE Xplore. Restrictions apply.
Fig. 2. Overall flowchart of the proposed feature extraction network.

Based on the analysis described earlier, HS pansharpening can be defined as the following optimization problem:

$$\min_{\tilde{H}} \| \text{PAN} - R\tilde{H} \|^2_F + \| H - \tilde{H}\text{BS} \|^2_F$$  \hspace{1cm} (4)

where $\|\cdot\|^2_F$ is a Frobenius norm. Obviously, it is an ill-posed inverse problem to recover ideal HR-HS images from LR-HS images.

Formula (4) reveals that the spectral information of the pansharpened image should be as close as possible to that of LR-HS image, while its spatial component should be consistent with that of the PAN image. In the following, pansharpening is formulated as a multiobjective programming problem with the objective functions of minimum spectral loss and maximum spatial quality improvement. In the proposed PS-GDANet, the spatial discriminator establishes the adversarial game with the generator to approximate the formula (1), and the spectral discriminator establishes the adversarial game with the generator to approximate the formula (2).

B. Feature Extraction Network (FENet)

FENet focuses on extracting spatial-spectral information in an unsupervised manner by discovering the relationship between HS image and PAN image. In general, the deeper the network, the more parameters and storage space are required. FENet transforms a difficult problem that requires a deep network into several subproblems of shallower networks cooperatively working. It follows the general idea of MRA and extracts multilevel features from the HS image. More specifically, to reduce reconstruction artifacts, we upsample the HS image level by level. Then the convolution and deconvolution blocks are adopted to construct the feature extraction subnetwork for each level. The features extracted from the level $c$ are input into the level $c+1$. Overall, as shown in Fig. 2, FENet is composed of a set of cascading subnetworks.

The main task of the FENet is to fuse the features learned by each level. The features learned in the first intermediate layer of each level can be formulated as

$$F_1^i = f_{\text{conv}}[f_{\text{conv}}(H^i)], \quad i = 0, 1, 2$$  \hspace{1cm} (5)

where $H^i$ and $F_1^i$ represent the input of the $i$th level and the output of the first intermediate layer in this level, respectively, $f_{\text{conv}}$ is a convolution block comprising the convolutional layer, the batch normalization (BN) layer, and the rectified linear unit (ReLU) layer. The features learned in the second intermediate layer of each level can be expressed as

$$F_2^i = \begin{cases} f_{\text{conv}}[g_{\text{deconv}}(F_1^i)], & \text{if } i = 0 \\ f_{\text{conv}}[f_{\text{conv}}(F_1^i)], & \text{if } i = 1, 2 \end{cases}$$  \hspace{1cm} (6)

where $g_{\text{deconv}}$ is a function of deconvolution block containing the deconvolution layer, BN layer, and ReLU layer. The deconvolution layers are adopted to upsample the feature maps in the subnetworks to reduce the memory and calculation time required by FENet. Similarly, the features learned in the third intermediate layer of each level can be expressed as

$$F_3^i = \begin{cases} f_{\text{conv}}[g_{\text{deconv}}(F_2^i)], & \text{if } i = 0 \\ f_{\text{conv}}[f_{\text{conv}}(F_2^i)], & \text{if } i = 1, 2 \end{cases}$$  \hspace{1cm} (7)

Therefore, the features $F_1$, $F_2$, and $F_3$ learned in three intermediate layers of FENet are

$$\begin{align*}
F_1 &= \text{concat}[F_1^0, \uparrow \text{concat}[\uparrow F_1^0, F_1^1]] \\
F_2 &= \text{concat}[F_2^0, \uparrow \text{concat}[\uparrow F_2^0, F_2^1]] \\
F_3 &= \text{concat}[F_3^0, F_3^1, F_3^2]
\end{align*}$$  \hspace{1cm} (8)

where symbol $\uparrow$ refers to the upsampling operation, and $\text{concat}(A, B)$ represents the cascading operation of $A$ and $B$. 

Authorized licensed use limited to: Mississippi State University Libraries. Downloaded on December 08,2022 at 05:25:11 UTC from IEEE Xplore. Restrictions apply.
 Recently, GAN is considered to be a very promising technique for SR and various GAN-based SR methods have been explored. Among them, super-resolution generative adversarial network (SRGAN) has attracted wide attention for its advantages of remarkable performance, fast convergence rate, and so on [39]. Inspired by the success of SRGAN, Wang et al. [40] explored enhanced SRGAN (ESRGAN) for SR, which introduced residual-residual dense block (RRDB), and removed the BN layers to reduce computational complexity and improve SR performance. The experimental results show that ESRGAN can provide more realistic texture structure than SRGAN. This is because when mean and variance of the training dataset differ significantly from those of the testing dataset, BN layers tend to limit the generalization ability and cause some artifacts. Moreover, the BN layer is more likely to bring artifacts when it is trained in a deeper network [41].

Because the pansharpened HS image is highly similar to the input HS image, most of the values in residual image generated by the residual network (ResNet) are very small or even zero. Therefore, the ResNet has great potential in solving the pansharpening problem. DenseNet facilitates to back-propagate gradients and extracts more representative features. RRDB is a perfect combination of ResNet and DenseNet, which can significantly improve gradient flow and increase the learning representation.

Based on these observations, we propose a novel generator framework, in which the BN layers are removed and the RRDBs are adopted, for a stable and consistent performance. As shown in Fig. 3, the generator keeps the high-level structure design of ESRGAN. Our proposed generator consists of a feature extraction module (FEM), a two-stage SR module (TSRM), a feature reconstruction module (FRM), and a feature injection module (FIM) as shown in Fig. 3.

1) Feature Extraction Module (FEM): The generator takes the input LR-HS image \( H \), the PAN image \( PAN \) and the extracted features \( F_1 \), \( F_2 \), and \( F_3 \) as inputs. The generator starts with a FEM in which the first convolution layer is adopted to extract the initial feature representation from the input HS image

\[
F_I = f_{\text{conv}}(H) \tag{9}
\]

where \( f_{\text{conv}}(\cdot) \) denotes the first convolution operation in FEM, and \( F_I \) indicates the initial feature.

Following the first convolution layer are RRDBs. Based on the key principle in deep learning that “the deeper the better,” the proposed PS-GDANet adopts six RRDBs with the same structure in the generator. As depicted in Fig. 3, residual connections occur not only between the pairs of RRDBs, but also between the dense blocks in each RRDB and the convolutional blocks in each dense block. Therefore, RRDBs allow for a deeper CNN to increase the learning representation and solve the problem of vanishing/exploding gradients. This process can be formulated as

\[
F_N = f_{\text{RRDB}_N}(F_I) = f_{R,N-1}(\cdots f_{R,1}(F_I) \cdots) \tag{10}
\]

where \( f_{\text{RRDB}_N}(\cdot) \) indicates the function of the RRDBs, and \( f_{R,N}(\cdot) \) denotes the operation of the \( N \)th RRDB.

The receptive field of the generator is determined by the number of RRDBs. For pansharpening, the local spatial details are more important than the global structural information. Based on experimental analysis, six RRDBs are adopted in the generator, and each RRDB contains three residual in dense blocks (RDBs). The residual scaling is set to 0.2.
The FEM ends with an additional layer, which is connected to the last RRDB to realign the channel of the feature maps to the same number as that of the input features and reconstruct the residual feature maps:

$$\tilde{F}_{\text{res}} = f_{\text{conv}}(F_N) = f_{\text{conv}}[f_{\text{RRDB}}(F_I)]$$

where $f_{\text{conv}}(\cdot)$ denotes the last convolution operation in FEM, and $\tilde{F}_{\text{res}}$ represents the estimated 128-channel residual features. We add a skip connection between the first and the last convolution layer to alleviate the vanishing gradient problem. The output of FEM can be formulated as

$$F_{\text{FEM}} = F_I + \tilde{F}_{\text{res}}$$

(12)

where $F_{\text{FEM}}$ is the feature map estimated by FEM.

2) Two-Stage SR Module (TSRM): The main task of TSRM is to upsample the feature maps estimated by FEM to the same size as the PAN image without distorting the spatial-spectral characteristics of the input HS image.

TSRM consists of two SR blocks (SRBs). For the first SRB, the 128-channel feature map is fed into a convolution layer and a 512-channel feature map is obtained. Followed the first convolution layer is the leaky rectified linear unit (LeakyReLU) operation. The 512-channel feature map is then fed into a subpixel convolution layer to be reshaped to an upsampled feature map, which contains 128 channels but is doubled in size. The first SRB of TSRM can be formulated as

$$F_{\text{SR}}^1 = \uparrow \left[ f_{\text{conv}}(F_{\text{FEM}}) \right]$$

(13)

where $F_{\text{SR}}^1$ represents the output of the first SRB, $f_{\text{conv}}(\cdot)$ is the convolutional operation with LeakyReLU served as the activation function, $\uparrow$ is upsampling operation in subpixel convolution. The second SRB has the same architecture design of the first SRB. The second SRB of TSRM can be formulated as

$$F_{\text{SR}}^2 = \uparrow \left[ f_{\text{conv}}(F_{\text{SR}}^1) \right]$$

(14)

where $F_{\text{SR}}$ denotes the output of TSRM, which is an upsampled feature map with twice the resolution of the input.

3) Feature Reconstruction Module (FRM): The objective of FRM is to reconstruct the upsampled feature map estimated by TSRM to an upsampled HS image. As illustrated in Fig. 3, FRM consists of two convolution blocks, each of which adopts a Conv layer followed by LeakyReLU operation. The Conv layer in the first convolution block containing 128 kernels, whereas the Conv layer in the second convolution block contains $C$ kernels, where $C$ is the channel of HS image. With the generated $F_{\text{SR}}$, we applied FRM to reconstruct the upsampled HS image $\overline{H}$

$$\overline{H} = f_{\text{conv}}(f_{\text{conv}}(F_{\text{SR}}))$$

(15)

4) Feature Injection Module (FIM): As illustrated in Fig. 3, FIM aims to inject the feature maps extracted by FENet and the spatial details of the PAN image into the upsampled HS image $\overline{H}$. Based on the analysis described earlier, $F_1$ and $F_2$ mainly concentrate on learning the spectral features of HS image, while $F_3$ contains most of the spatial information of HS image. The network will inevitably lose some features in the process of learning the spatial-spectral information of HS image. To solve this problem, instead of designing a deep network with PAN image as the only or primary input, the FIM cascades all the features into a shallow feature enhancement network to estimate the missing spatial-spectral features of the upsampled HS image. The guided filter [42] is applicable for structure-transferring and exhibits the nice property of edge-preserving smoothing. FIM uses guided filtering to transfer the structural information of the PAN image to $F_3$, which can not only enable a shallower network to fully extract the spatial information of the PAN image, but also effectively solves the problem of pansharpened image blur caused by the low-frequency component of the PAN image. The output image $F_F$ is a linear transform of the guidance image PAN in a window $v_t$, and the linear coefficients are computed by minimizing the difference between $F_F$ and the input image $F_3$

$$F_F^h = a_1 F_N^h + b_1 \quad \forall h \in v_t$$

(16)

$$\min \sum_{h \in v_t} (a_1 F_N^h + b_1 - (F_3^h)^2 + \zeta a_1^2)$$

(17)

where $(F_F^h)$, $(F_1^h)$, and $F_N^h$ are the $h$th pixel of $F_F$, $F_3$, and PAN, $\zeta$ is a regularization parameter. Here, $a_1$ and $b_1$ are the linear coefficients that are computed by solving (17). $a_1$ and $b_1$ for all the overlapping windows $v_t$ that covers $t$ are used to compute the final filtering output

$$F_F^h = \frac{1}{|v_t|} \sum_{h \in v_t} (a_1 F_N^h + b_1)$$

(18)

where $|v_t|$ denotes the number of pixels in $v_t$, $(F_3^h)$ denotes the mean of $F_3$ in $v_t$, $\tau_1$, and $\sigma_1^2$ denotes the mean and variance of PAN in $v_t$. The features learned by FIM can be expressed by

$$F = f_{\text{conv}}(f_{\text{conv}}(F_1; F_2; F_F))$$

(19)

where $f_{\text{conv}}(\cdot)$ denotes the convolution operation, and $[F_1; F_2; F_F]$ represents the concatenation of three features.

With the upsampled HS image $\overline{H}$ and its missing features estimated, the final pansharpened image $H_P$ can be obtained by injecting the estimated features into the upsampled HS image

$$H_P = \overline{H} + G \cdot F$$

(20)

where G is an injection gain defined as many widely accepted pansharpening methods

$$G_k = \frac{\overline{H}_k}{\sum_{k=1}^{C} H_k}$$

(21)

where the subscript $k$ denotes the $k$th band in HS image. The injection gain G helps to keep the spectral ratio of HS unchanged in the pansharpened image.
Algorithm 1 Generative Dual-Adversarial Networks With Spatial Enhancement and Spectral Fidelity for Hyperspectral Pansharpening Algorithm

Input: The PAN image PAN, the HS image H, and the reference HR-HS image H
Output: The pansharpened HR-HS image HP
1: Obtain the feature 1 F1, feature 2 F2, and feature 3 F3 based on feature extraction network (FENet)
2: Initialize the parameters of generative network θG, the parameters of spatial discriminator network θD1, and the parameters of spectral discriminator θD2
3: Set the hyperparameters. Epoch is set to 500 (max_epoch = 500). The learning rate is 10^(-4) and decreases by a factor of 10 per 350 epochs (Lr = 10^(-4), Lr_de = 350)
4: while epoch < max_epoch do
5: for s = 1, 2, ..., N do
6: if s % D1-critic == 0 then
7: HP ← G(PAN, H, F; θG)_1,2,3
8: Calculate \( \hat{g}_1 \) according to equation (27)
9: Calculate \( L_{D_1}(\theta_{D_1}) \) according to equation (25)
10: \( \theta_{D_1} \leftarrow \text{Adam}(\nabla_{\theta_{D_1}} L_{D_1}(\theta_{D_1}), \theta_{D_1}, Lr) \)
11: end if
12: if s % D2-critic == 0 then
13: HP ← G(PAN, H, F; θG)_1,2,3
14: Calculate \( \hat{g}_2 \) according to equation (28)
15: Calculate \( L_{D_2}(\theta_{D_2}) \) according to equation (26)
16: \( \theta_{D_2} \leftarrow \text{Adam}(\nabla_{\theta_{D_2}} L_{D_2}(\theta_{D_2}), \theta_{D_2}, Lr) \)
17: end if
18: HP ← G(PAN, H, F; θG)_1,2,3
19: Calculate \( L_G(\theta_G) \) according to equation (24)
20: \( \theta_G \leftarrow \text{Adam}(\nabla_{\theta_{G}} L_G(\theta_{G}), \theta_G, Lr) \)
21: end for
22: if epoch % Lr_de == 0 then
23: \( Lr \leftarrow Lr/10 \)
24: end if
25: epoch ++
26: end while
27: HP ← G(PAN, H, F; θG)_1,2,3
28: return HP

D. Discriminator Networks

As mentioned earlier, the generator performs adversarial learning with two discriminators, which have the same structures with different inputs. The spatial discriminator takes the spatial component of the generated HS image or the PAN image as the input. The spectral discriminator takes the original HS image or the degraded version of the generated HS image as the input. Through the adversarial games, the spatial discriminator forces the spatial component of the HS image generated by the generator to be closer to the PAN image, and the spectral discriminator forces the spectral distribution of the generated HS image to be more consistent with that of the original HS image. While in SRGAN and ESRGAN, the discriminator is learned to distinguish the reconstructed image from the real image. Both the spatial and spectral discriminators are simple convolution neural networks. For the first four convolution blocks, they are all composed of a 3 × 3 convolutional layer followed by an activation layer. In these four convolution blocks, the stride of the first layer is set to 1, and the second one is set to 2. The feature maps are all set to 3. We borrow the idea from ESRGAN [39] to remove the BN layers and use the leaky ReLU as activation function. Following the four convolution blocks are two 3 × 3 convolution layers, in which the stride is set to 2 without padding. Essentially, the discriminator is a classifier. The last two layers are fully connected layers.

E. Loss Function

1) FENet Loss Function: FENet is an unsupervised network that is used to extract the multilevel features. Learned features F3 are fed into two convolution layers to reconstruct the PAN image, which can be expressed as

\[
\bar{P} = f'_{conv}(f_{conv}(F_3))
\]

where, \( \bar{P} \) is the output of FENet representing the reconstructed PAN image, \( f'_{conv} \) is the last convolutional operation to reconstruct the feature map to generate the PAN image, and \( f_{conv} \) is a function for the convolution block, in which the convolutional layer is followed by the BN layer and the ReLU layer. The L1-norm is chosen as the loss criterion of FENet, which can be expressed as

\[
L_{FE}(\Theta) = \frac{1}{Q} \sum_{q=1}^{Q} \left\| \bar{P}^q - \text{PAN}^q \right\|_1
\]

where \( \Theta \) denotes the parameters of FENet, \( Q \) is the number of training samples, and \( \text{PAN}^q \) and \( \bar{P}^q \) are the qth sample of PAN and \( \bar{P} \), respectively.

2) Generator Loss Function: The generator selects the L1-norm distance between the pansharpened HR-HS image \( G(PAN, H, F; \theta_G)_1,2,3 \), and the reference HR-HS image \( \hat{H} \) as the content loss. The generative losses that are defined based on the probabilities of the spatial discriminator and spectral discriminator are also adopted. In addition to the content loss and the generative losses, we also impose root mean squared error (RMSE) as a global constraint on spatial and spectral distortion to further minimize the error between the generated HR-HS image and the reference HR-HS image. To sum up, the proposed generator can be formulated to minimize the overall loss defined as

\[
L_G(\theta_G) = \left\| G(PAN, H, F; \theta_G) - \hat{H} \right\|_1 + \alpha_1 \left[ -D_1(R \cdot G(PAN, H, F; \theta_G); \theta_{D_1}) \right] + \alpha_2 \left[ -D_2(G(PAN, H, F; \theta_G) \cdot BS; \theta_{D_2}) \right] + \alpha_3 \left\| G(PAN, H, F; \theta_G) - \hat{H} \right\|_F/\sqrt{MNS}
\]
where \(\|\|_F\) is the Frobenius norm and \(a_1, a_1, \) and \(a_3\) are the coefficients to balance different loss terms and are empirically set to \(5 \times 10^{-3}, 5 \times 10^{-4},\) and \(10^{-3}\), respectively.

3) Discriminators Loss Function: The proposed PS-GDANet has two discriminators, which are the spatial discriminator and the spectral discriminator. The spatial discriminator aims at distinguishing the spatial component of the generated HR-HS image from real PAN image, and the spectral discriminator is learned to differentiate the degraded HS version of the generated HS image from real LR-HS image. The traditional GANs have mode collapse and vanishing gradient problems, and the Wasserstein GAN (WGAN) [43] uses weight clipping to make the discriminator lie within the space of Lipschitz functions so as to improve the stability of learning procedure. However, weight clipping in WGAN leads to optimization difficulties and pushes weights toward two values, which are the extremes of the clipping range. WGAN-GP [44] utilizes gradient penalty for enforcing Lipschitz constraint, which can avoid the same problems. The spatial and spectral discriminators in PS-GDANet are trained by adopting WGAN-GP, and their loss functions are defined as

\[
L_{D_1}(\theta_{D_1}) = D_1(R \cdot G(PAN; H, F_i; \theta_G); \theta_{D_1}) - D_1(PAN; \theta_{D_1}) + \lambda_1(\|\nabla_{\theta_D} D_1(\hat{g}_1; \theta_{D_1})\|_2 - 1)^2
\]

(25)

\[
L_{D_2}(\theta_{D_1}) = D_2(G(PAN, H, F_i; \theta_G) \cdot BS; \theta_{D_2}) - D_2(H; \theta_{D_2}) + \lambda_2(\|\nabla_{\theta_D} D_2(\hat{g}_2; \theta_{D_2})\|_2 - 1)^2
\]

(26)

where \(\lambda_1\) and \(\lambda_2\) are set to 10, and \(\nabla\) denotes the operation of partial derivative. The first two terms in \(L_{D_1}(\theta_{D_1})\) and \(L_{D_2}(\theta_{D_2})\) are the loss functions of WGAN, and the last term is the gradient penalty. \(\hat{g}_1\) and \(\hat{g}_2\) are sampled from the distributions, which sample uniformly between pairs of points sampled from the real data distribution and the generated distribution, and can be expressed as

\[
\hat{g}_1 = \gamma_1 \cdot PAN + (1 - \gamma_1) \cdot [R \cdot G(PAN, H, F_i; \theta_G)]
\]

(27)

\[
\hat{g}_2 = \gamma_2 \cdot H + (1 - \gamma_2) \cdot [(G(PAN, H, F_i; \theta_G) \cdot BS]
\]

(28)

where \(\gamma_1\) and \(\gamma_2\) are the random numbers uniformly distributed between 0 and 1.

The algorithm of the proposed PS-GDANet is shown in Algorithm 1.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Datasets and Experimental Setup

The CAVE dataset was acquired by the generalized assorted pixel camera. This dataset contains 32 HS images with controlled illumination under lab conditions. Each HS image is characterized by 31 bands in the wavelength range 400–700 nm. The CAVE dataset is with the spatial resolution of \(512 \times 512\). Correspondingly, the spatial resolution of PAN images is \(128 \times 128\). In the experiment, we randomly selected 22 HS images to train the model, and the remaining 10 HS images were used as test samples. In this article, the training of PS-GDANet model is based on patches because of the insufficient samples. The reference HS image, PAN image, and LR-HS image are divided into 16 nonoverlapping patches with sizes of \(128 \times 128 \times 31\), \(128 \times 128 \times 1\), and \(32 \times 32 \times 31\), respectively.

The Pavia Center dataset was taken over the Pavia, Northern Italy. It was collected by the reflective optics system imaging spectrometer (ROSIS), which is characterized by 115 bands in the visible and infrared range of 430–860 nm. The noise and water vapor absorption bands have been removed from the dataset, leaving 102 bands for experiments. The Pavia Center dataset contains 1096 scan lines, each of which contains 1096 pixels. Because some samples do not contain any information, for efficiency, we discarded the noninformation fragment and intercepted the image of 960 × 640 for experiment. The experimental Pavia Center dataset is with the spatial resolution of \(40 \times 40\). Correspondingly, the spatial resolution of PAN images is \(160 \times 160\). Because of the limitations of HS image samples and memory, we trained the model based on patch. Specifically, the entire image is first divided into 12 patches of the same size without overlapping, each of which is \(160 \times 320\). Nine patches were randomly selected from 12 patches as training samples, and the remaining three patches were used to test the performance of the model. Furthermore, we slide each patch into 21 small patches with a size of \(160 \times 160\).

The CAVE and Pavia Center datasets are all semisynthetic datasets. Given the high-resolution reference images, the semisynthetic PAN and HS images can be generated by Wald’s protocol [46]. According to the Wald’s protocol, the HR-HS image is down-sampled using 9*9 filtering to obtain a LR-HS image, and the spectral response of the HR-HS image in the visible range is averaged to obtain the PAN image.

The proposed method is conducted with the PyTorch framework and trained on the NVIDIA GTX 2080Ti GPU. In the FENet, the Adam optimization is chosen for training with the batch size parameterized by 16, and the epoch is 300. The learning rate is initialized to 0.006 and decreases by a factor of 10 per 64 epochs. In both the generator and discriminator, the Adam optimization is employed, and the optimization is terminated at the 500th epoch. The batch size is set to 8. The learning rate is \(10^{-4}\) and decreases by a factor of 10 per 350 epochs. We train the generative dual-adversarial network based on WGAN-GP. During the training, the parameters of the spatial discriminator and the spatial discriminator are updated alternately by the optimizer, to ensure the fast and stable convergence of the network. When the parameters of spatial discriminator are updated every two iterations and the parameters of spectral discriminator are updated every three iterations, the generative dual-adversarial network achieves the best performance of HS pansharpening.

B. Compared Methods and Quality Assessment

For comprehensive validation, ten pansharpening methods are utilized for comparison, including SFIM, MGH, GSA, PCA, GFPCA, CNMF, Hysure, BayesSparse, LTTR, and DRPNN. Among these competing methods, the first two belong to the MRA family. GSA and PCA are the most representatives of the CS family. GFPCA is a hybrid
pan sharpening method that combines the advantages of CS- and MRA-based methods. As demonstrated in some publications, CNMF is a matrix factorization-based method and usually performs well in terms of objective evaluation. BayesSparse and Hysure are recently proposed model-based methods that usually outperform the CS- and MRA-based pansharpening methods. LTTR is a recently proposed tensor-based method. Moreover, we also compare the proposed method with the DL-based DRPNN method. The best parameters for all the competing methods are adopted in the experiment.

It is difficult to distinguish the slight difference between the pasharpening results of the competing methods only by visual inspection, especially the two methods with comparable performance. Four widely accepted quantitative indexes [46], [47] are adopted, including cross correlation (CC), spectral angle mapper (SAM), RMSE, and erreur relative globale adimensionnelle de synthèse (ERGAS). Specifically, CC is a spatial quality metric to measure the spatial similarity between the pansharpened result and the reference image. The closer the CC is to 1, the smaller the spatial distortion of the fused image. SAM compares the similarity of the spectral vectors of the fused image and the reference image at the corresponding pixels. Generally speaking, the smaller the SAM value, the better the performance. As for the metrics RMSE and ERGAS, they are both the global quality indexes to evaluate the spatial and spectral distortion. RMSE is adopted to assess the difference of values of the pixels between the pansharpened result and the reference image. ERGAS is used to evaluate the dynamic range change. The closer RMSE and ERGAS are to the ideal value 0, the better the spatial-spectral performance of the pansharpened image is.

C. Ablation Study

To verify the effectiveness of the PS-GDANNet proposed in this article, we compared and studied the impact of different configurations on performance in this section.

1) Residual-Residual Dense Blocks (RRDBs): We first compare the performance of the generator using different numbers of RRDBs. Intuitively, the more RRDBs in the generator, the deeper the network, and the better the performance. We trained three generator models by using different RRDB levels: 4, 6, and 8. The training curves of the average CC, SAM, RMSE, and ERGAS results on CAVE dataset are shown in Fig. 4. As shown in Fig. 4, the performance of a generator using six RRDBs is significantly better than that of a generator using four RRDBs, especially after 350 iterations. However, from the experiment, we found that when RRDB increased from 6 to 8, the performance would become worse. This is because the nonlinear mapping of six RRDBs can well decompose the input HS image and simplify the learning problem. Continuously increasing the number of RRDBs may lead to overfitting of the model. Therefore, in the generator network, six RRDBs are adopted.

2) Discriminators: To improve the spatial and spectral performance, the proposed PS-GDANet generates the pasharpened HS image by establishing the adversarial games between the generator and two discriminators, where one makes the spectral content of the pansharpened HS image to be close to that of the LRMS image, and the other tries to force the spatial information to be consistent with that of PAN image. To verify the effectiveness of the design, the performance of different discriminator combinations is compared. We train three model structures on CAVE dataset, i.e., Spatial D using only one spatial discriminator, Spectral D using only one spectral discriminator, and the proposed PS-GDANet using both spatial and spectral discriminator. In Table I, each of these models shows their best performance. It can be seen that the values of CC, SAM, RMSE, and ERGAS obtained by PS-GdANet are significantly better than those obtained by the model using only spatial D or spectral D. This indicates that both spatial and spectral discriminators play an important role in performance improvement in adversarial learning.

3) Upsampling Techniques: In pansharpening, it is necessary to upsample the HS image to the same size as the PAN image. In our experiments, the spatial resolution of the PAN image is four times that of the available HS image. Bilinear interpolation is one of the most widely used interpolation techniques in the classical pansharpening methods. As shown in Fig. 5, we have also conducted an experiment on the CAVE dataset with different interpolation techniques. It can be seen that the two-time subpixel convolution
technology adopted in this article achieves the best performance on quantitative indicators, followed by a four-time subpixel convolution. This is because the subpixel convolution has the potential to effectively suppress reconstruction artifacts caused by bilinear interpolation. Obviously, the performance obtained by directly using the four-time bilinear interpolation technique is the worst. This may be because directly sampling the HS image by a factor of four would also cause severe reconstruction artifacts. Therefore, it is difficult for bilinear interpolation to balance the spectral fidelity and spatial enhancement.

4) FENet: After confirming the structures of generator and discriminator, we analyzed the impact of FENet on pansharpening. Fig. 6 verifies the effectiveness of FENet, which is illustrated by the curves of indexes. We adopt FENet for two major reasons. First, FENet can fully extract the spatial features of the HS image. Second, some spectral features may be lost in the convolution operation; using FENet helps to preserve the spectral information and reduce the spectral distortion of the pansharpened image. All the indexes show relatively consistent results that using FENet can bring a better performance and a faster convergence rate.

5) Guided Filtering: In the proposed generator, the guided filter is introduced to transfer the structures of the PAN image to feature 3. We test the performance of guided filtering on feature 1, feature 2, and feature 3. As shown in Fig. 7, the performance can be significantly improved by applying guided filtering to feature 3. However, from the comparison experiment, we can see that the guided filtering of feature 1 and feature 2 brings limited improvement, and the performance is the worst when the guided filtering is performed on all three feature maps. This is because feature 1 and feature 2 mainly contain the spectral information of HS images, whereas the spatial information of HS image is mainly represented by feature 3. Thus, we use guided filtering to transfer the spatial structure of the PAN image to feature 3.

D. Experimental Results

The color displays, composed of bands 3-12-25, of the fusion results obtained by different methods on the CAVE dataset are shown in Fig. 8. Fig. 9 shows error images of the competing approaches for CAVE dataset at 12th band. The available reference HS images are shown in Figs. 8(a) and 9(a), respectively. By visual comparison of the pansharpened results and the reference images, we found that the remarkable performance in the spectral preservation and the spatial quality of the pansharpened results with respect to the reference image have been achieved by these competing methods. Both SFIM and MGH created an HR-HS image with similar color to the reference HS image. However, they produce halo artifacts around the edges of objects, although such artifacts make the structures of objects in the image look sharper [see Fig. 8(b) and (c) and Fig. 9(b) and (c)]. In other words, despite good spectral fidelity, the SFIM and MGH methods inject too much details into the pansharpened results, such as the redundant structures as shown in marginal areas of Fig. 9(b) and (c). As for GSA, its pansharpened results exhibit remarkable improvement in both spectral and spatial qualities with respect to LR-HS images. However, it produced somewhat blurred results on account of insufficient details of injection. For example, the pansharpened image produced by GSA in Fig. 10(d) lacks some details in square area at the top of the image. As reported in some studies, the PCA method usually causes brightness or spectral distortion. Based on the comparison of Figs. 8(e) and 9(e) with their corresponding reference images, it can be concluded that the PCA method may decrease the contrast and brightness of the original image. It is clear that the pansharpened images provided by GFPCA suffer from severe spectral and spatial distortion [see Figs. 8(f) and 9(f)]. The pansharpened results obtained by GFPCA are significantly blurred. The CNMF method can usually create pansharpened results, which are slight spectral distortions and spatial artifacts, and it may lose some important spatial details. According to visual comparison of the pansharpened results obtained by Bayesian model-based approaches, it can be further confirmed that the HySure and BayeSparse exhibit a slight spatial distortion. In contrast, the spatial information injected by BayeSparse appears to be inadequate. The fused result provided by LTTR method is too smooth or even blurred because of insufficient spatial detail injection [see Figs. 8(j) and 9(j)]. Overall, the DRPNN and the proposed PS-GDANet pansharpening methods show considerable potential for maintaining the spectral fidelity and improving the spatial resolution.

Table II summarizes the quantitative results of the different methods on the CAVE dataset. It can be seen that, for the average of the ten test samples in the CAVE dataset, the
proposed PS-GDANet yields the largest quality index values for CC, and the smallest values for SAM, RMSE, and ERGAS. In fact, the CC value provided by the proposed PS-GDANet is always the largest for each image. Moreover, the values of SAM, RMSE, and ERGAS of the proposed PS-GDANet are always the smallest for each image of the CAVE dataset. This means that our proposed method performs well in preserving the spectral content of the original HS image and injecting the right amount of detail information. As for SFIM, MGH, and DRPNN, their SAM and ERGAS are much higher than those of the PS-GDANet, although their CC and RMSE are also satisfactory. This means that they cannot offer a very stable performance on the CAVE dataset.

The color displays, composed by bands 30-40-50, of the fusion results on two sets of the Pavia Center datasets obtained by different methods are shown in Figs. 10 and 11. The available reference HS images are shown in Figs. 10(a) and 11(a). The Pavia Center dataset is more challenging, because it contains complex urban scenes with more structure information. It is clear that the SFIM method causes severe spatial distortion on edge and characteristic of texture, although it performs well with less spectral distortion. The SFIM creates a pansharpened image whose color display is similar to the reference HS image, but whose edges appear too smooth because of insufficient spatial information extracted from the PAN image. The buildings and vegetation areas in Figs. 10(b) and 11(b)
are blurred. The color of the pansharpened results created by the MGH and GSA methods is similar to that of the reference one as shown in Figs. 10(a) and 11(a). In terms of spatial quality, they inject insufficient spatial details, such as the car shown on the street. Comparing Fig. 10(c) and (d) and Fig. 11(c) and (d), we can see that the GSA algorithm causes a little more blurring than the MGH algorithm, especially in vegetation areas. As for PCA and GFPCA, they yield less clear results than other competing methods. Compared with GFPCA, PCA provides better spectral fidelity performance, but because of insufficient spatial information, its fused result is obviously more blurred. Because of the spectral range mismatch between HS and PAN images, the GFPCA method results in serious spectral distortion. It can be seen from Fig. 10(f) that the color display of the pansharpening result of the GFPCA method was significantly different from that of the reference image. Similar to GFPCA, the CNMF method can also cause spectral distortion. The difference is that the GFPCA method on the Pavia Center dataset produces a darker pansharpening result, whereas CNMF reduces the contrast and brightness of the original HS image. Figs. 10(i) and 11(i) indicate that the BayesSparse method may introduce different levels of spectral distortion, because it produces a result that looks significantly brighter than the reference image. Obviously, the LTTR method produces severe spatial distortion on the Pavia Center dataset. The edges and structures of the objects in Figs. 10(j) and 11(j) are too smooth. By contrast, the HySure, DRPNN, and the proposed GANet work well spatially and spectrally.

In Table III, the four quality metrics are considered together to evaluate the fusion performance. Similar to Table II, all four metrics obtained by PCA and GFPCA are also relatively poor. The global quality metrics RMSE and ERGAS of GSA, PCA, GFPCA, CNMF, LTTR, and BayesSparse are quite large when
compared with that of the proposed method. This means that these pansharpening methods cannot work well in both spatial and spectral aspects. By contrast, the SFIM, MGH, HySure, and DRPNN offer much better pansharpening performance in terms of CC, SAM, RMSE, and ERGAS on the Pavia Center dataset. Consistent with color display, the proposed PS-GDANet provides the best performance for all four quality indexes, i.e., CC, RMSE, SAM, and ERGAS. This further confirms that the proposed PS-GDANet performs well in both improving the spatial resolution and maintaining the spectral information of the original HS image.

IV. DISCUSSION

From the experimental results, we can see that the dual-adversarial network proposed in this article achieves a good trade-off between maintaining the spatial information of PAN image and preserving the spectral information of HS image. Fig. 7 shows that the best performance can be achieved by applying guided filtering to feature 3. As shown in Fig. 6, it can be seen that the proposed method without FENet can well preserve the spectral content of the original HS image, but fail in fully estimating the missing details of HS image. FENet is designed to seek an accurate representation from HS to PAN image, aiming at learning spatial information in an unsupervised manner. FENet plays an important role in further improving spatial performance, which is demonstrated in Figs. 6 and 7.

V. CONCLUSION

In this article, we have developed a novel pansharpening framework, PS-GDANet, based on generative dual-adversarial network. The proposed PS-GDANet pays special attention to spatial enhancement and spectral fidelity, which is essential for many applications of HS images. On the one hand, PS-GDANet automatically estimates multilevel spatial details of HS image in an unsupervised manner by exploring the spatial relationship between HS and PAN images. On the other hand, PS-GDANet extends GAN to two discriminators, in which the spatial discriminator improves the spatial resolution by constraining the spatial distribution of the generated image and that of PAN image, and the spectral discriminator ensures the spectral fidelity by constraining the spectral similarity between the generated image and the original
HS image. Compared to several widely accepted pansharpening approaches, PS-GDA-Net shows comparable or better performance in terms of four reliable quantitative indexes and creates a pansharpened HS image much closer to the reference HS image. Limited by the structure, the proposed method can only be applied to the specific case where the spatial resolution of the PAN image is 2 times that of the HS image. In the future, we will study on extending the network to pansharpening with any level of scaling factors.

REFERENCES


[48] Jiahui Qu (Member, IEEE) received the B.S. degree in communication engineering from Yantai University, Yantai, China, in 2014, and the Ph.D. degree in communication and information systems of Xidian University, Xi’an, China, in June 2020. She was an Exchange Ph.D. Student of Mississippi State University, Starkville, MS, USA, from 2018 to 2019. She is currently a Lecturer with the State Key Laboratory of Integrated Services Networks, Xidian University. She has published over 20 articles in known academic journals and conferences, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the Remote Sensing. Her research interests include hyperspectral image detection, image fusion, neural networks, and deep learning.

[49] Song Xiao received the M.S. degree in communication and information system and the Ph.D. degree in signal and information processing from Xidian University, Xi’an, China, in 2001 and 2004, respectively. From 2006 to 2007, she was with the Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA, as a Post-Doctoral Researcher. She is currently a Professor and the Ph.D. Director of Communication and Information System with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi’an, China. She has authored over 80 international journal articles and conference papers. Her research interests include image compression and coding, joint source channel coding, multimedia transmission systems over wired/wireless network, and compressed sensing.

Dr. Xiao is the Secretary-General of Image Application in Military and Civil Integration (IMCI) of Professional Committee of China Society of Image and Graphics. She is a Council Member of Shaanxi Society of Image and Graphics and a member of the IEEE Multimedia Communication Technology Committee and the IEEE Signal Processing Society.

[50] Yunsong Li (Member, IEEE) received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi’an, China, in 1999 and 2002, respectively. In 1999, he joined the School of Telecommunications Engineering, Xidian University, where he is currently a Professor. He is also the Director of the Image Coding and Processing Center, State Key Laboratory of Integrated Services Networks, Xidian University. His research interests include image and video processing, hyperspectral image processing, and high-performance computing.

Authorized licensed use limited to: Mississippi State University Libraries. Downloaded on December 08,2022 at 05:25:11 UTC from IEEE Xplore. Restrictions apply.