

Coupling Model- and Data-Driven Methods for Remote Sensing Image Restoration and Fusion

*Improving
physical
interpretability*

IMAGE LICENSED BY INGRAM PUBLISHING

**HUANFENG SHEN, MENGHUI JIANG, JIE LI, CHENXIA ZHOU,
QIANGQIANG YUAN, AND LIANGPEI ZHANG**

In the fields of image restoration and image fusion, model- and data-driven methods are the two representative frameworks. However, both approaches have their respective advantages and disadvantages. Model-driven techniques consider the imaging mechanism, which is deterministic and theoretically reasonable; however, they cannot easily model complicated nonlinear problems. Data-driven schemes have a stronger prior-knowledge learning capability for huge data, especially for nonlinear statistical features; however, the interpretability of the networks is poor, and they are overdependent on training data. In this article, we systematically investigate the coupling of model- and data-driven methods, which has rarely been considered in the remote sensing image restoration and fusion communities. We are the first to summarize the coupling approaches into the following three categories: 1) data- and model-driven cascading methods, 2) variational models with embedded learning, and 3) model-constrained network learning methods. The typical existing and potential coupling techniques for remote sensing

image restoration and fusion are introduced with application examples. This article also gives some new insights into potential future directions, in terms of both methods and applications.

INTRODUCTION

Remote sensing data are carriers of spatial information and geographical knowledge, and the data quality directly affects the application, both intensively and extensively. However, the imaging process is influenced by many factors, such as the observation capability of the remote sensing satellite sensors, land-cover type, atmospheric conditions, and lighting conditions, resulting in complex and diverse data quality problems [1].

The quality of remote sensing data is closely related to spatial, spectral, and temporal indicators. Specifically, spatial resolution is the ability to discriminate the spatial detail information, which is the ground range corresponding to a pixel in the actual satellite observation image. *Spectral resolution* refers to the minimum wavelength interval that the sensor can resolve when receiving electromagnetic wave information radiated by a ground object. Temporal resolution is the revisit time of the sensor.

.....
Digital Object Identifier 10.1109/MGRS.2021.3135954
Date of current version: 20 June 2022

Due to hardware limitations of the sensors, the energy that a sensor can receive is limited. A single remote sensing image needs to be a tradeoff between the spatial, spectral, and temporal resolutions, leading to a low expression capability for the land surface. What is more, due to the interaction of sensor imaging, atmosphere transmission, and surface reflection, noise often degrades the spatial texture of remote sensing images. The two most typical noise types are speckle noise in synthetic aperture radar (SAR) images and hybrid noise in hyperspectral images (HSIs). Moreover, the thick and thin cloud/haze appearing in poor atmospheric conditions can obscure the land, leading to missing spatial information and distorted spatial-spectral features.

Therefore, to overcome these degradation problems, researchers have proposed many image processing schemes, including image denoising, cloud removal, and image fusion methods. The traditional remote sensing image processing approaches are based mostly on filtering, regression, fitting, Fourier transforms, and wavelet transforms. However, these methods rarely consider the image degenerative process or the image priors existing in statistical and structural information. Therefore, variational procedures have been proposed, which regard image processing as an ill-posed inverse problem and construct the energy function according to degradation models between the ideal image and the degraded observations. These techniques are considered model-based methods in which the energy function is usually constructed based on Bayesian maximum a posteriori (MAP) estimation or sparse representation. These methods generally include two parts: a data fidelity and a prior model. The data fidelity model constrains the relationship between the ideal image and the degraded observations. The prior model employs the structure and statistical characteristics of the image itself to optimize the solution. Common image priors are the total variation (TV) [2], Laplacian [3], nonlocal [4], and low-rank priors [5]. Because of rigorous theory, the accuracy of the variational model-based methods is often higher than that of traditional schemes, but they cannot handle complicated nonlinear problems accurately.

Recently, deep learning has been applied to various remote sensing problems due to its promising performance in describing the nonlinear relationships among different data. A variety of networks with good performances have been developed, such as residual neural networks [6], U-Net [7], encoder-decoder networks [8], DenseNet [9], and generative adversarial networks [10]. However, although deep learning has powerful feature extraction and expression capabilities, it lacks a theoretical foundation and relies heavily on massive data. The model- (i.e., variational model-based methods) and data-driven methods (i.e., the deep learning-based procedures) are complementary, to a large extent. On the one hand, the combination of the two approaches can improve the interpretability of the network in deep learning and reduce the network's dependence on massive data. On the

other hand, it can reduce the pressure of accurate modeling nonlinear problems of model-driven methods.

The coupling of model- and data-driven schemes has been utilized in various computer vision tasks and has been found to be effective in image denoising [11], image super-resolution [12], image fusion [13], and image dehazing [14]. Along with the successful coupling of model- and data-driven techniques in natural image processing [11]–[17], the coupling of these methods has also become a popular and promising trend in remote sensing. Different from natural images, on the one hand, the noise of remote sensing images is more complex, including Gaussian noise, multiplicative noise, and even mixed noise. On the other hand, remote sensing images have rich spectral characteristics. Therefore, it is necessary to consider the distribution of different noises and impose constraints on spatial and spectral information when constructing the model. In this article, for remote sensing image restoration and fusion, we divide the coupling systems into three categories: 1) data- and model-driven cascading procedures, 2) variational models with embedded learning, and 3) model-constrained network learning methods.

In the following sections, we first introduce the main idea behind each coupling approach and then introduce some specific applications of each coupling approach in the field of remote sensing image restoration and fusion. Although some attempts at the coupling of model- and data-driven methods have been made in remote sensing image restoration and fusion, there are still many problems to be solved. We also discuss the future developments of data- and model-driven combinations in remote sensing.

THE MODEL-DRIVEN FRAMEWORK

PROBLEM DEFINITIONS AND OBJECTIVE FUNCTIONS IN REMOTE SENSING

The degradation of remote sensing images often caused by noise, haze, cloud, and the lack of spatial-spectral-temporal resolution. The different degradation models in remote sensing can be uniformly expressed in the following framework:

$$Y = MTDHX + N, \quad (1)$$

where Y is the observed degraded image, and X is the ideal clean image. T , D , H , and M represent the different degradation processes, as shown in Figure 1. Among these processes, the downsampling matrix D and blur matrix H are usually jointed to describe degradation of the imaging spatial scale. T denotes an uneven variation of the intensity distribution, and it describes the portion of the electromagnetic radiation that reaches the sensor under the influence of complicated atmospheric scattering (due to thin clouds and haze) and the obstruction of incident light. M reflects the missing information, such as objects covered by thick clouds or dead pixels caused by sensor failure. N denotes generalized noise, such as the classic Gaussian noise; Poisson, impulse, and stripe noises, even clouds and fog.

To summarize T , D , H , and M to be a matrix A , (1) can be simplified into

$$Y = AX + N \quad (2)$$

with the use of an MAP estimation, the optimization problem can be formulated as

$$X = \operatorname{argmin}_X \|Y - AX\|_p + \lambda g(X), \quad (3)$$

where the first and second terms are the data fidelity and regularization ones, respectively. $g(X)$ in the second term is the prior operator, such as the TV operator, Laplacian operator, and so on.

Based on sparse representation theory, the image restoration optimization problem can be rewritten as

$$X = \operatorname{argmin}_X \|Y - A\psi\alpha\|_2 + \lambda \|\alpha\|_0 \text{ with } X = \psi\alpha, \quad (4)$$

where the original clean image can be represented by the dictionary ψ and the sparse coefficient α based on spatial or spectral redundancy. The regularization term $g(\cdot)$ is also rewritten as some constraints associated to α as the sparsity prior of l_0 norm.

Furthermore, when improving the image quality with auxiliary data, such as image fusion with two observation images, the energy function model can be given as

$$X = \operatorname{argmin}_X \|Y - AX\|_p + \lambda g(X) + \gamma u(X, Z), \quad (5)$$

where the third term is the data fidelity term between ideal image X and the other complementary observation Z .

However, for SAR intensity images with speckle noise, different from (1) for optical data, the degradation model is usually described as

$$Y = XN, \quad (6)$$

where Y and X are the observed speckled and speckle-free images, respectively; and N follows a Gamma law with one mean, which is signal-dependent multiplicative noise. Its density function is defined as

$$p(N) = \frac{L^L}{\Gamma(L)} N^{L-1} e^{-LN} \mathbf{1}_{\{N \geq 0\}}, \quad (7)$$

where L is the number of looks. According to Bayesian MAP theory, SAR image despeckling can be regarded as an optimization problem [18], for which the model is as follows:

$$X = \operatorname{argmin}_X \lambda \left(\log X + \frac{Y}{X} \right) + g(X). \quad (8)$$

In summary, the energy function of the image restoration problem can be expressed as

$$X = \operatorname{argmin}_X f(X, Y) + \lambda g(X), \quad (9)$$

where $f(\cdot)$ is the data fidelity term that maintains consistency between the observed degraded images and the ideal clean image. The regularization term $g(\cdot)$ promotes solutions with an optimum performance, and parameter λ controls the balance of consistency and performance. For the different tasks, $f(\cdot)$ and $g(\cdot)$ can be used to describe different data relationships and prior constraints, respectively, in (3), (5), and (8), which is discussed in detail in the ‘‘Application Examples’’ section.

BASIC OPTIMIZATION METHODS

For the solution of the aforementioned objective functions, two typical optimization strategies are often used. The simplest strategy is to solve the objective function directly using the gradient descent method, Newton’s method, and so forth. Taking the case of gradient descent, (9) can be rewritten as

$$X^{t+1} = X^t - \delta \frac{\partial}{\partial (X^t)} (f(X^t, Y) + \lambda g(X^t)), \quad (10)$$

where t represents the iteration number, and parameter δ denotes the step size.

Another efficient strategy is the use of a variable splitting algorithm, such as half-quadratic splitting (HQS), the alternating direction method of multipliers (ADMM), or proximal gradient descent (PGD). In this strategy, an auxiliary variable V is first introduced to split the data fidelity and regularization terms. Equation (9) can then be rewritten as

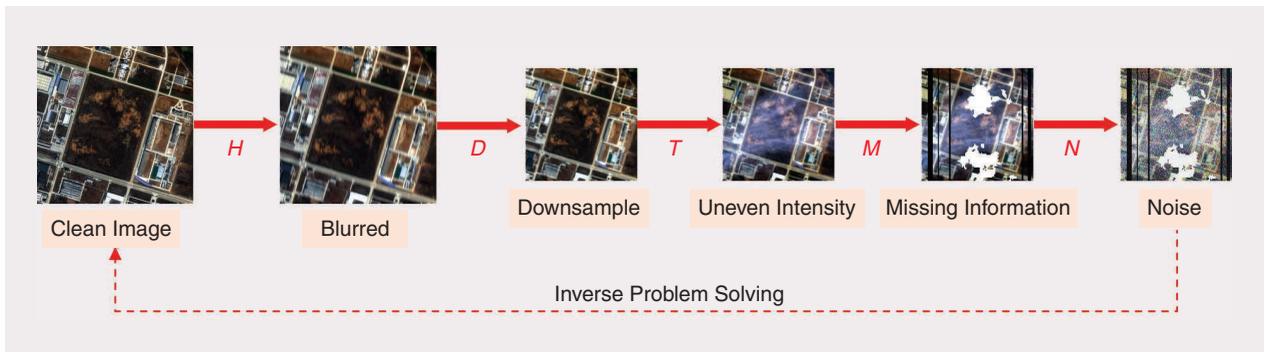


FIGURE 1. The degradation processes of remote sensing images.

$$\begin{aligned} X, V &= \underset{X, V}{\operatorname{argmin}} f(X, Y) + \lambda g(V), \\ \text{s.t. } & AX + BV = C. \end{aligned} \quad (11)$$

The augmented Lagrangian of (11) is then

$$X, V = \underset{X, V}{\operatorname{argmin}} f(X, Y) + \lambda g(V) + \frac{\rho}{2} \|AX + BV - C + W\|_2^2, \quad (12)$$

where W is the Lagrange multiplier, and ρ is the penalty parameter.

Then, taking the ADMM as an example, to solve the aforementioned optimization problem, the iteration equations are

$$X^{t+1} = \underset{X}{\operatorname{argmin}} f(X, Y) + \frac{\rho}{2} \|AX + BV^t - C + W^t\|_2^2 \quad (13)$$

$$V^{t+1} = \underset{V}{\operatorname{argmin}} \lambda g(V) + \frac{\rho}{2} \|AX^{t+1} + BV - C + W^t\|_2^2 \quad (14)$$

$$W^{t+1} = W^t + AX^{t+1} + BV^{t+1} - C. \quad (15)$$

Finally, (13) and (14) can be solved in different ways according to the specific tasks. When combined with deep learning, (14) can be implicitly replaced by a deep network prior with discriminative information for a specific problem.

METHODS BASED ON THE COUPLING OF MODEL AND DATA DRIVEN FRAMEWORKS

DATA- AND MODEL-DRIVEN CASCADING

The coupling techniques can be categorized into 1) data- and model-driven cascading methods; 2) variational models with embedded learning; and 3) model-constrained network learning approaches. Model-driven methods have difficulty in achieving accurate modeling, while data-driven ones rely on massive data with representative features. *Data- and model-driven cascading* refers to the sequential use of data- and model-driven methods. This can be further categorized into two approaches (see also Figure 2): 1) first model and then data driven, i.e., first build a rough model to obtain an initialization result, and then use deep learning to generate a more accurate result; and 2) first data and then model driven, i.e., first use deep learning to mine the

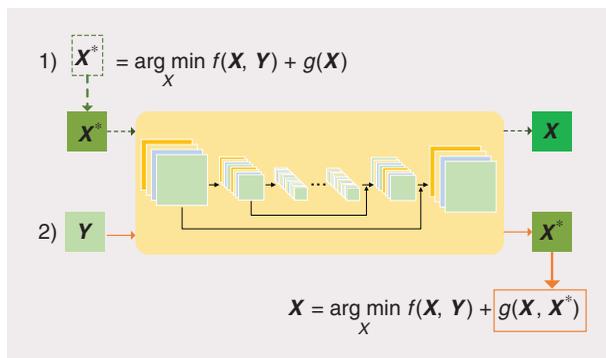


FIGURE 2. The framework of data- and model-driven cascading.

nonlinear features of the images, and then use the learned prior information to build an accurate model.

As depicted in the upper-left part of Figure 2, the dashed green arrow indicates that the rough, variational model, which utilizes the task-specific degradation domain knowledge, is first constructed to obtain a coarse reconstruction result, i.e., X^* . The coarse reconstruction result is then fed into the convolutional neural network (CNN) to allow the network to generate a more accurate reconstruction result, i.e., X .

In the lower-right part of Figure 2, the solid orange arrow indicates that the CNN is first used to mine the deep prior information in the image, i.e., X^* . The prior is then used to construct an accurate energy function, which avoids complicated prior assumptions or incorrect linear assumptions. Equation (9) can then be rewritten as

$$X = \underset{X}{\operatorname{argmin}} f(X, Y) + \lambda g(X, X^*), \quad (16)$$

where $g(X, X^*)$ represents the relationship function between X and X^* , such as the difference between the two or the difference between the gradients of the two. It should be noted that the learned prior image X^* is not only used to construct the regularization term [13], [19], i.e., $g(X, X^*)$, but can also be used in the data fidelity term [20], i.e., $f(X, Y)$,

VARIATIONAL MODELS WITH EMBEDDED LEARNING

Both the handcrafted priors in the traditional variational models and the deep priors in data- and model-driven cascading approaches explicitly define the regularization term. However, the degradation types of remote sensing images are complex and diverse, and an explicit prior cannot handle the various latent degradation types. Thus, the plug-and-play prior strategy [21], which is known for its flexible and effective handling of various inverse problems, is used. The main idea of this approach is to unfold the energy function into subproblems by the use of a variable splitting algorithm, and then to embed the pretrained CNN to solve the prior term-related subproblem.

As shown in Figure 3, the auxiliary variables V introduced by the ADMM algorithm decouple the data fidelity and regularization terms of the model into individual subproblems. The variable X of the data fidelity term-related subproblem is solved using a conventional solution, such as gradient descent or the least-squares algorithm. The variable V of the regularization term-related subproblem is solved through the plug-and-play pretrained network. Most of the existing plug-and-play prior-based image restoration methods treat the CNN Gaussian denoiser as the prior [11], [22], and some treat the CNN superresolver as the prior [12].

MODEL-CONSTRAINED NETWORK LEARNING

The previous two approaches are aimed at capturing a more accurate prior by the data-driven method, and then the pretrained prior is used to construct the energy function or solve the prior-related subproblem. In these two

approaches, network training and model solving are separated. In contrast, the model-constrained network learning approach integrates model- and data-driven methods into an end-to-end network, simultaneously performing network parameter optimization and model solving. According to the different alternatives, the model-constrained network learning approach can be further divided into two subcategories: 1) a model-constrained network structure and 2) model-constrained loss function. The former uses the model to make the network structure interpretable, and the latter uses the model to constrain the optimization space of the network parameters.

THE MODEL-CONSTRAINED NETWORK STRUCTURE

As is known, variable splitting and alternate iterative optimization algorithms are typically used for optimizing the solutions of model-driven methods. However, processing of the optimization has to manually adjust some tedious parameters and only obtain the shallow feature priors. Hence, to avoid these problems, this approach introduces a recursive network structure to realize automatic updating parameters and deep feature mining. Each iteration in the original model-driven methods will be unfolded into one subnetwork in recursion. The multicascaded subnetwork form a recursive network structure in which subnetworks can share or not share the same parameters [15], [16], [23]–[25].

Similar to the variational models with embedded learning, a variable splitting algorithm, taking the ADMM algorithm as an example, first decouples the energy function (9) into individual subproblems, such as (13)–(15). The network is then designed according to the iterative update processes of the subproblems, as shown in Figure 4. A deep convolutional structure (DCS) is used to solve at least one

subproblem. Figure 4 takes the solution of a subproblem as an example [15], where the whole network contains T stages, corresponding to T-iterative processes. At each stage, the subvariables are updated in turn, the DCS solves the regularization term, and the data fidelity term is solved using the traditional method, which is given by

$$\begin{cases} \mathbf{X}^{t+1} = \mathbf{X}^t - \delta \frac{\partial}{\partial (\mathbf{X}^t)} \left(f(\mathbf{X}^t, \mathbf{Y}) + \frac{\rho}{2} \|\mathbf{A}\mathbf{X}^t + \mathbf{B}\mathbf{V}^t - \mathbf{C} + \mathbf{W}^t\|_2^2 \right) \\ \mathbf{V}^{t+1} = \text{DCS}(\mathbf{X}^{t+1}, \mathbf{W}^t) \\ \mathbf{W}^{t+1} = \mathbf{W}^t + \mathbf{A}\mathbf{X}^{t+1} + \mathbf{B}\mathbf{V}^{t+1} - \mathbf{C}. \end{cases} \quad (17)$$

All the parameters are jointly optimized by an end-to-end training scheme. Naturally, other optimization methods [16], [23]–[25] can also be applied to solve this category. For example, the iterative shrinkage-thresholding algorithm based deep

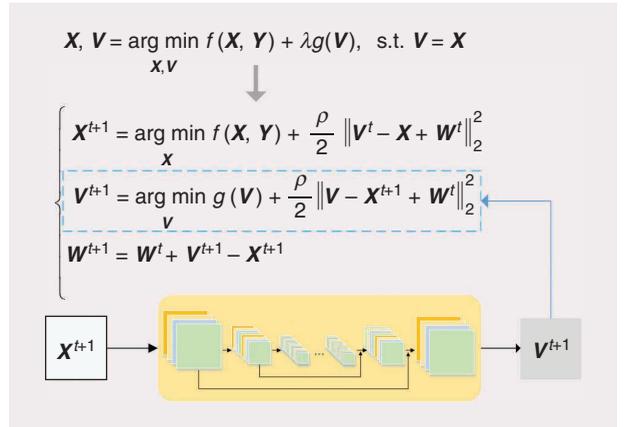


FIGURE 3. The framework of variational models with embedded learning.

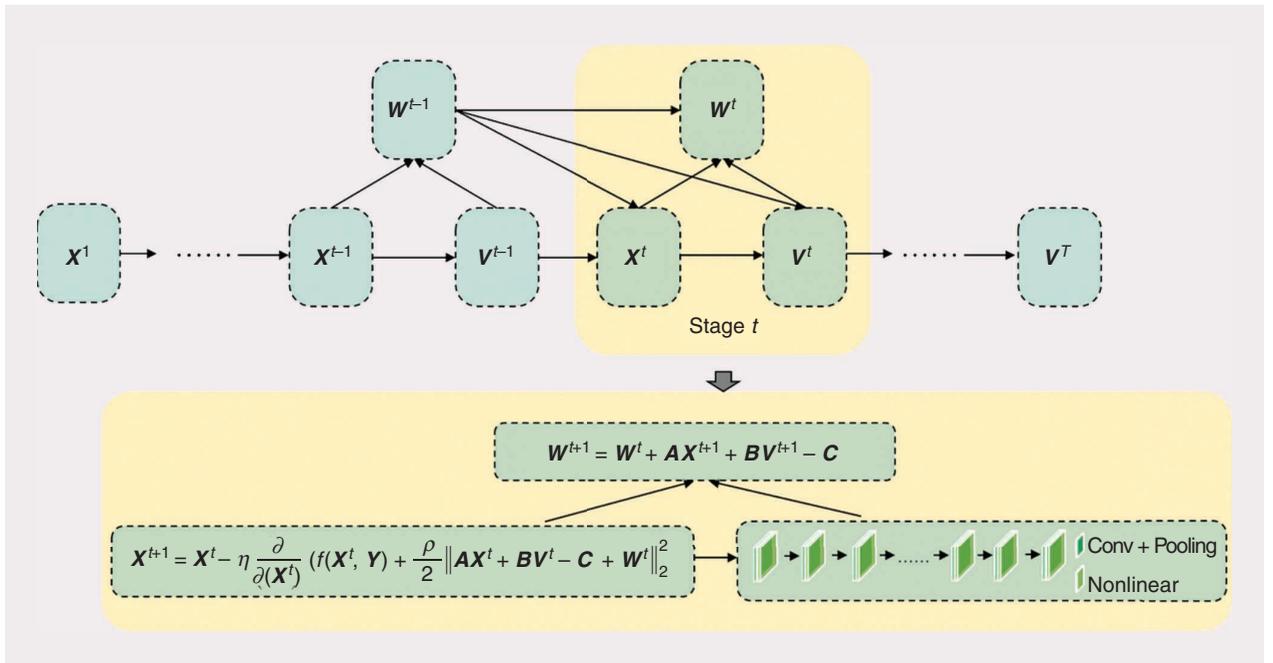


FIGURE 4. The framework of a model-constrained network structure.

network [16] and the dual-path deep neural network [23] were inspired by the iterative shrinkage-thresholding algorithm and HQS, respectively. Most of these practices employed a DCS for optimization of the regularization term while using the gradient descent or Newton's method for optimization of the data fidelity term, which was embedded in the network with an equation. However, in [24], Kobler et al. proposed a variational network that employed the PGD method and used a rectified linear unit instead of proximal mapping.

MODEL-CONSTRAINED LOSS FUNCTION

The commonly used loss functions in image restoration tasks, such as the L1 and L2 norms, focus on minimizing the difference between the network output and labeled data. However, they do not consider the physical model behind the degradations and thus cannot be categorized as model driven. In this article, model-constrained loss function is established by integrating some degenerate relationships between the observed and estimated images and can then be regarded as a network parameter optimization problem with degradation model constraints, as shown in (18). The framework of this approach is shown in Figure 5.

$$\begin{aligned} \hat{X} &= \operatorname{argmin}_X f(X, Y) + \lambda g(X) + \beta R(X, X') \\ &= \operatorname{argmin}_X \|Y - AX\|_F^2 + \lambda g(X) + \beta \|X - X'\|_F^2, \end{aligned} \quad (18)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. λ and β are predefined balance scalars. The energy function $f(X, Y)$ can be defined according to different application tasks, indicating the multiplicative and additive degenerations in the spectral and spatial domains. For example, the noise that obeys a Gamma distribution in SAR imagery, downscaling in the superresolution problem, the blur kernel in deblurring, thin clouds/haze in optical images, and the noise that follows a Bayes model. As displayed in Figure 5, $f(X, Y)$ is specified by A and N , where A represents the comprehensive degradation process spatially and spectrally, and N represents interference, such as clouds, fog, and noise. X is the output of the network, X' is the labeled data of X , and Y indicates the observed image with one or multiple degradation factors. $g(X)$ is the handcrafted prior term, such as the TV prior, nonlocal prior, metric learning prior [26], and so on. Traditionally, $R(X, X')$ is the mean square error (MSE), which acts as a loss function to constrain the relationship between the labeled data and the estimated solution. Designing a loss function of the network according to A and the handcrafted prior $g(X)$ can effectively reduce the optimization space of the network parameters and improve the accuracy of the network.

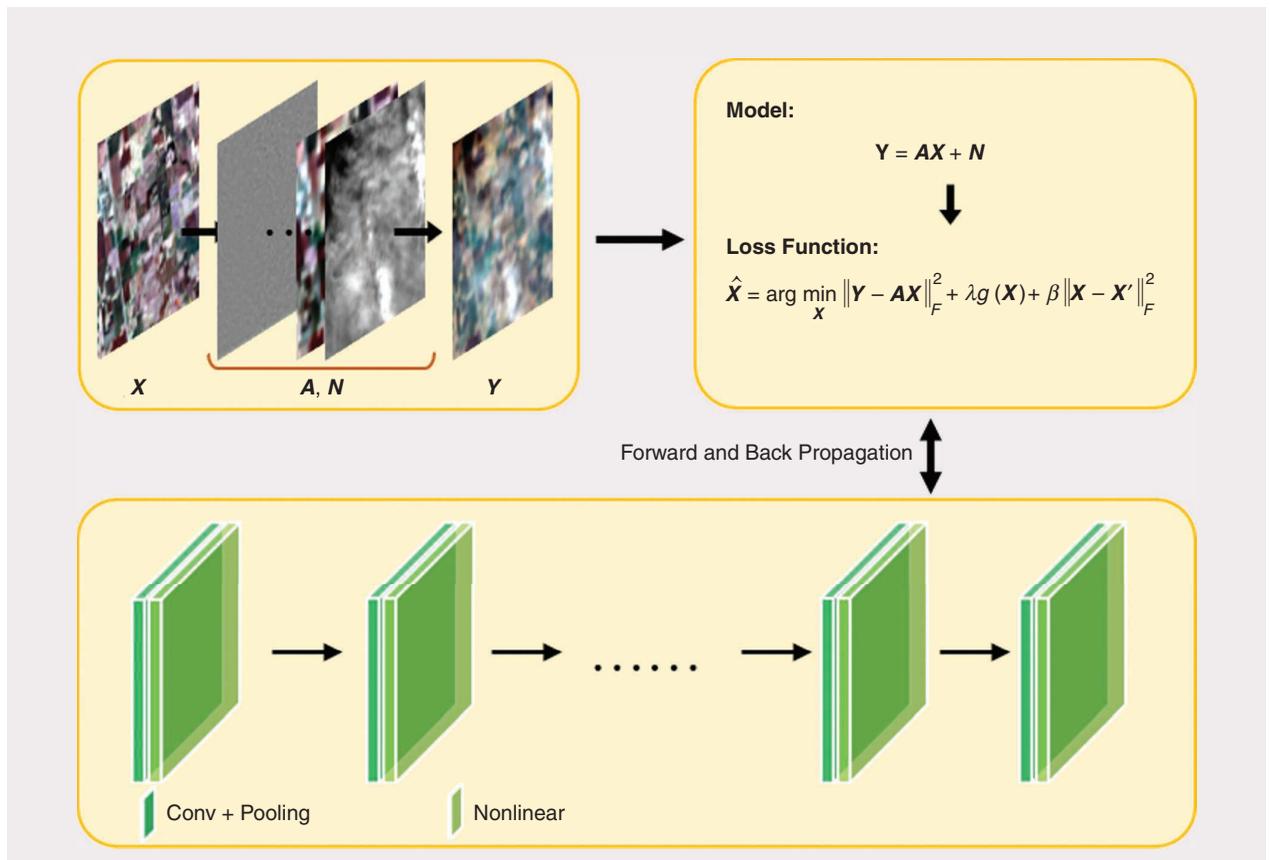


FIGURE 5. The framework of the model-constrained loss function.

APPLICATION EXAMPLES

SPECKLE NOISE REDUCTION IN SAR IMAGES

SAR systems are active, remote sensing systems that play an important role in Earth surface monitoring, irrespective of weather conditions. However, in SAR images, speckle noise has the characteristics of multiplicative Gamma noise and inherently affect image quality, making it hard to classify, interpret, or segment due to the coherent interference of radar waves reflected from the many basic scatterers. Therefore, the reduction of speckle noise in SAR imagery is important for many applications. Generally speaking, various despeckling approaches can be categorized as model- or data-driven methods. The coupling model- and data-driven techniques are few in number. In the following sections, we introduce some examples for the two categories [25], [27]–[29], i.e., variational models with embedded learning and model-constrained network learning approaches (see Table 1).

DESPECKLING USING VARIATIONAL MODELS WITH EMBEDDED LEARNING

In the class of variational models with embedded learning, there are mainly three combined modes for single-polarization SAR image despeckling, i.e., in the original-intensity, log-intensity, and original complex domains. In 2017, Deledalle et al. [27] proposed a plug-and-play SAR image reconstruction framework in the log-intensity domain in which the multiplicative noise is transformed into additive noise by a log transform. And then, in 2019, Alver et al. [28] proposed a different framework in the original complex domain in which the authors assume a Fourier transform-based forward model for constructing the additive noise observation mode.

Currently, there is no relevant paper in the original intensity domain based on the original multiplicative observational model. Therefore, for the first time, we propose a novel method based on the original multiplicative degradation model of (7), called *SAR despeckling based on a plug-and-play (SAR-PNP)*. Here we use the AA variational model proposed by Aubert and Aujol [18], which is presented in (8). Then, taking the variable splitting technique and HQS algorithm, the iteration equations are as follows:

$$X^{t+1} = \underset{X}{\operatorname{argmin}} \lambda \left(\log X + \frac{Y}{X} \right) + \frac{\rho}{2} \|X - V^t\|_2^2 \quad (19)$$

$$V^{t+1} = \underset{V}{\operatorname{argmin}} g(X) + \frac{\rho}{2} \|V - X^{t+1}\|_2^2, \quad (20)$$

where subproblem V is solved by the deep CNN network $V^{t+1} = DCNN(X^t)$, which is a pretrained network, and the subproblem X is solved by the gradient descent method.

DESPECKLING USING THE MODEL-CONSTRAINED NETWORK STRUCTURE APPROACH

As an example of a model-constrained network structure approach for SAR despeckling, Shen et al. [25] proposed a recursive deep CNN (DCNN) prior model (SAR-RDCP),

which jointly optimizes the data fidelity and deep CNN-based regularization terms. In this method, Shen et al. apply the same framework presented in (19) and (20). The difference is the strategy of the network learning. Specifically, the two subproblems, X and V , are optimized and updated jointly in a recursive network structure instead of only updating the subproblem V in network.

Real-data experiments were undertaken to validate the effectiveness of variational models with the embedded learning and model-constrained network learning approaches. The pure network approach, called *SAR-DCNN* [25], the same as the structure of the DCNN in the SAR-PNP and SAR-RDCP, is selected for comparison with the SAR-PNP and SAR-RDCP. (The source codes for SAR-DCNN, SAR-PNP, and SAR-RDCP are available at <https://github.com/SGGJerryLi/WHU-SGG-RS-Pro-Group.git>.) Here the University of California, Merced's land-use data set [30] was used as the training set for the deep learning-based methods. The experiment selected a *Sentinel-1* single-look complex format image of the city of Wuhan in China with single polarized channel, horizontal transmit and vertical receive (VH), as presented in Figure 6. The image was cropped to 500×500 pixels for the experiment. From the results, it can be seen that the detail preservation and

TABLE 1. THE APPLICATIONS OF COUPLED MODEL- AND DATA-DRIVEN METHODS.

REFERENCE	CATEGORY	APPLICATION
Shen et al. [25]	Model-constrained network structure	SAR despeckling
Deledalle et al. [27]	Variational models with embedded learning	SAR despeckling
Alver et al. [28]	Variational models with embedded learning	SAR despeckling
Zeng et al. [31]	Variational models with embedded learning	HSI denoising
Lin et al. [32]	Variational models with embedded learning	HSI denoising
Choi et al. [35]	Variational models with embedded learning	HSI reconstruction
Wang et al. [33]	Model-constrained network structure	HSI reconstruction
Sidorov and Hardeberg [34]	Model-constrained loss function	HSI denoising
Zhang et al. [36]	Model-constrained loss function	HSI denoising
Dian et al. [13]	Data- and model-driven cascading	MSI/HSI fusion
Xie et al. [19]	Data- and model-driven cascading	Pan/HSI fusion
Shen et al. [20]	Data- and model-driven cascading	Pan/MSI fusion
Dian et al. [22]	Variational models with embedded learning	MSI/HSI fusion
Xie et al. [47]	Model-constrained network structure	MSI/HSI fusion
Zhang et al. [48]	Model-constrained loss function	MSI/HSI fusion

speckle reduction of the SAR-RDCP and SAR-PNP schemes are superior to that of SAR-DCNNs. For the SAR-PNP and SAR-RDCP procedures, from the two enlarged images of Figure 6(c) and (d), the place selected in the red box indicates that the SAR-PNP retains the details of the water-body area better, but from the place selected in the blue box, the SAR-RDCP has better retention of strong scattering points.

From the results described previously, the significant improvement of the image quality obtained using the SAR-RDCP and SAR-PNP schemes shows that use of the model-constrained network learning structure with optimizing guidance and use of a pretrained individual network may have greater potential than the pure network. In addition, the two techniques show different advantages in different scenes. Thus, for different scenes, it is also meaningful to choose appropriate coupling methods to obtain better solutions.

Recently, Molini et al. [29] applied a Bayesian framework relying on blind-spot CNNs to the self-supervised SAR image despeckling task. The main idea of this work is to minimize the negative logarithm-likelihood distribution in the training phase. Motivated by the method, it can be observed that model-constrained loss function with an appropriate likelihood distribution also makes a

breakthrough that takes into account the characteristics of the speckle in network optimization.

Overall, coupling methods are being more widely used in SAR image despeckling and obtain satisfactory performances. Up to now, these processes are mainly for single-polarization SAR regardless of fully polarimetric SAR (PolSAR). The extreme lack of real, clean PolSAR images makes us put more attention on unsupervised strategies in which a variational model can be used to better constrain both the network structure and loss function.

HSI DENOISING AND RECONSTRUCTION

Hyperspectral imaging is a technique used to acquire the radiation characteristics of observed objects with a fine spectral resolution. On account of the rich spectral information, HSIs are utilized in many applications. However, the increase in spectral channels of the sensors generates spectra with low signal-to-noise ratios (SNRs). Due to observation conditions and sensors, HSIs are always degraded by multiple types of noise, such as Gaussian, stripe, and impulse noises. Conversely, HSIs are also compressed to avoid the pressures related to data storage, transmission, and processing of airborne or spaceborne remote sensing

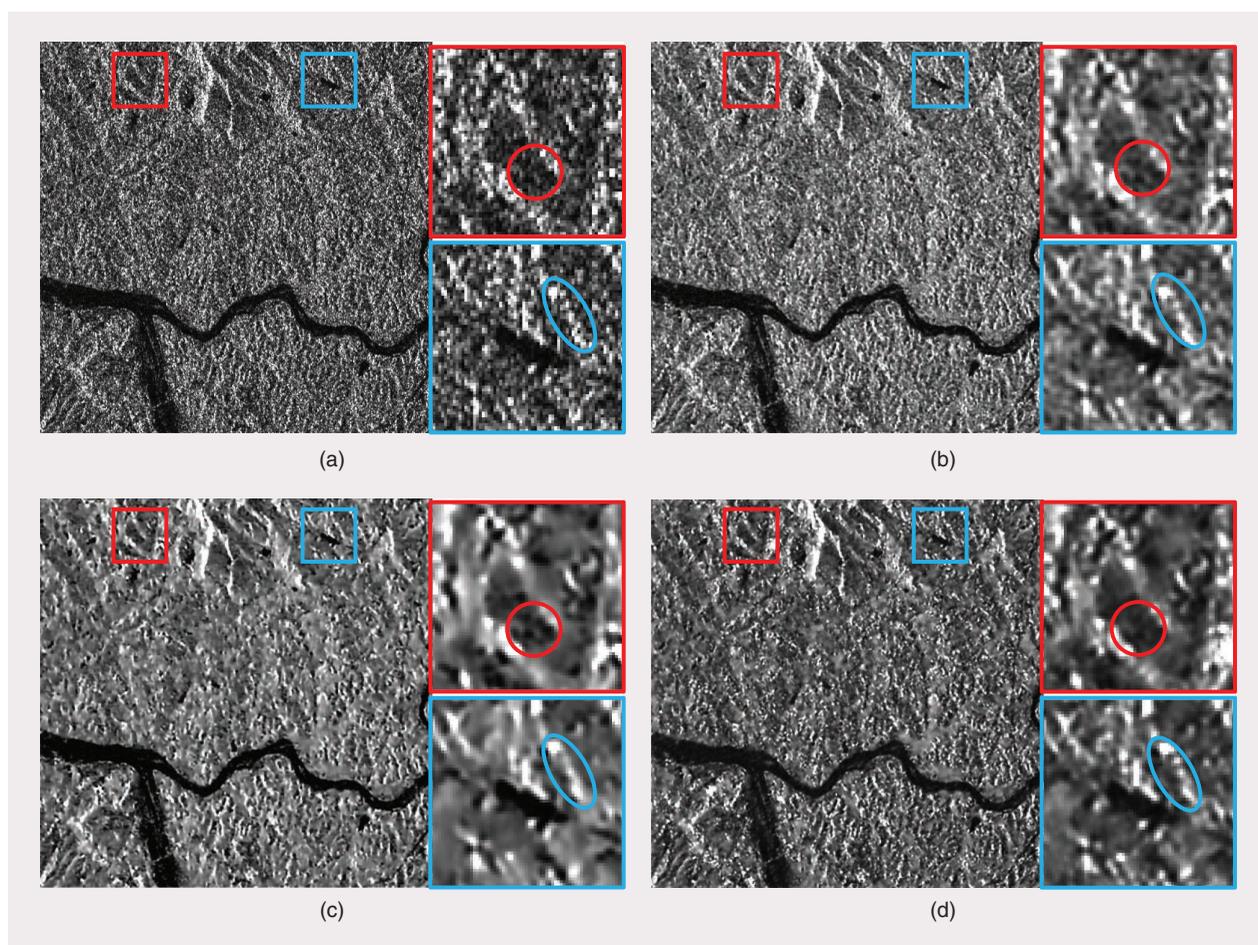


FIGURE 6. The despeckling results for the *Sentinel-1* VH single-look complex image of the city of Wuhan. (a) An original speckled image, (b) SAR-DCNN, (c) SAR-PNP, and (d) SAR-RDCP.

imaging system. Therefore, HSI denoising and reconstruction are common tasks of recovering clean HSI from its noisy and compressed versions.

As is known, data-driven methods have shown better performances than conventional model-optimization-based approaches due to their powerful representation capability; however, they lack flexibility. Hence, few practices combining data- and model-driven approaches have been attempted for HSI denoising and reconstruction and have shown advantages in exploiting large training data sets and introducing the explicable structure of the prior-regularized optimization. These coupling schemes are mainly based on the variational model with an embedded learning approach and model-constrained network learning.

HSI DENOISING AND RECONSTRUCTION USING VARIATIONAL MODELS WITH EMBEDDED LEARNING

For variational models with embedded learning, the plug-and-play framework is the main approach used in HSI denoising. The framework provides us with the possibility of integrating the capabilities of multiple priors, including deep learning priors, in one restoration model. For example, Zeng et al. [31] embedded both a low-rank and deep learning prior into a plug-and-play framework. This method uses a submodel of a Tucker decomposition-based, low-rank tensor approximation to remove the sparse noise and a part of the Gaussian noise, and leaves the residual by the dilated, deep residual network. Otherwise, a spectral mixing model can also be integrated into the framework, and a deep learning prior can be used to excavate the implicit features of the abundance matrix [32]. In general, this type of approach can be generalized as

$$\begin{aligned} \underset{X}{\operatorname{argmin}} & \|Y - X + S + N\|_F^2 + \tau g(Z) + \lambda \|S\|_1 + \beta \|N\|_F^2 \\ \text{s.t. } & X = Z, \end{aligned} \quad (21)$$

where $\|Y - X + S + N\|_F^2$ is equivalent to $f(\cdot)$ in (9), and $g(\cdot)$ denotes the implicit CNN-based function. N and S denote the Gaussian and sparse noises, respectively, which consist of impulse noise, stripes, deadlines, and so forth. As in [32], HSIs can also be transformed into signal subspaces for mining deeper, characteristic relationships, e.g., sparse representation in the spectral dimension, with $X = E\alpha$, where E and α are the endmember and abundance matrices, respectively. λ and β are positive-regularization parameters.

In HSI reconstruction, a deep autoencoder can also act as a deep operator in the process of iterative optimization. Specifically, to yield good, reconstructed images in spectral accuracy, the convolutional autoencoder in [35] is leveraged to first train an encoder network E to learn a nonlinear representation α of HSIs in a nonlinear space, then uses the decoder network D to reconstruct the final image from coded sensor data. The fidelity of nonlinear representations, therefore, can act as an HSI prior in the following:

$$\begin{aligned} \hat{\alpha} &= \underset{\alpha}{\operatorname{argmin}} \|Y - \Phi D(\alpha)\|_F^2 + \tau \|\alpha - E(D(\alpha))\|_F^2 + R(V), \\ V &= Q\alpha \\ \Rightarrow & \begin{cases} \alpha^{k+1} = \underset{\alpha}{\operatorname{argmin}} \|Y - \Phi D(\alpha)\|_F^2 + \tau \|\alpha - E(D(\alpha))\|_F^2 \\ + \frac{\rho}{2} \|V - Q\alpha + W\|_F^2 \\ V^{k+1} = \underset{V}{\operatorname{argmin}} R(V) + \frac{\rho}{2} \|V - Q\alpha + W\|_F^2, \end{cases} \end{aligned} \quad (22)$$

where Φ represents the measurement matrix, the second term regularizes the fidelity using the encoder-decoder pair, and the third term can be defined to favor sparsity of representation with operator Q . Compared with the linear representations based on sparse coding, the second term will allow us to cover a wider range of real-world spectral features, resulting in good spatial resolution and spectral accuracy. Different from a plug-and-play deep prior for subproblem V , (22) uses the trained encoder-decoder network in each iteration for subproblem α . Although the strategy is currently only used in natural HSI compressive reconstruction, it can be extended to remote sensing superresolution and potentially achieves better results.

HSI RECONSTRUCTION USING THE MODEL-CONSTRAINED NETWORK LEARNING APPROACH

The model-constrained network learning approach is currently used in HSI reconstruction. Wang et al. [33] first introduced iterative optimization into a deep convolutional network. This method focuses on exploiting the spatial and spectral correlation, which can lead to a superior performance. First, the optimization problem, which guarantees the relationship between the desired and original HSIs, is unfolded into an iteration-based optimization problem, as shown in (17), by the use of the conjugate gradient algorithm. Second, the structural insight of the iterative processing is integrated into the network and forms a data-driven prior, which can also be called an *optimization-inspired network*. The data-driven prior can regularize the optimization problem to exploit the spatial and spectral correlation, thus removing the influence of the noises and avoiding the heavy computational load of the traditional iterative optimization methods. The optimization inspired algorithm can be given as

$$\begin{aligned} \hat{X} &= \underset{X}{\operatorname{argmin}} \|Y - \Phi X\|_F^2 + \tau g(V), \text{ s.t. } V = X \\ \Rightarrow & \begin{cases} X^{k+1} = X^k - \varepsilon [\Phi^T (\Phi X^k - Y) + \eta (X^k - V^k)] \\ V^{k+1} = \underset{V}{\operatorname{argmin}} \frac{\tau}{\eta} g(V) + \|V - X^{k+1}\|_2^2 = \text{DCS}(X^{k+1}), \end{cases} \end{aligned} \quad (23)$$

where auxiliary variable V is introduced to convert the optimization problem into a subproblem about V related to the HSI prior $g(\cdot)$. Compared to the reconstruction problem, HSI denoising is a more basic degradation problem. Naturally, when Φ is an identity matrix, (23) can be converted to deal with noises in HSIs.

HSI DENOISING USING THE MODEL-CONSTRAINED LOSS FUNCTION APPROACH

In the model-constrained, loss function approach, the image prior can be found directly in the space of the network's parameters through the optimization process with loss function considering the degeneration relationship between input and output images, which can be called *degeneration loss function* in this article. This function, called *deep hyperspectral prior (DHSP)*, [34] can describe the fidelity between the estimated and corrupted images in denoising, inpainting, and superresolution and can be written as

$$\hat{\theta} = \underset{x}{\operatorname{argmin}} F(\operatorname{Net}_{\theta}(\mathbf{Y}), \mathbf{Y}), \text{ s.t. } \mathbf{X} = \operatorname{Net}_{\theta}(\mathbf{Y}). \quad (24)$$

Function $F(\cdot)$ addresses the degeneration relationship using simple linear relationships, and network $\operatorname{Net}_{\theta}(\mathbf{Y})$ is used to obtain an ideal clean image. Considering the non-independent identically distributed noise of HSIs, a denoising framework called the *deep spatio-spectral Bayesian posterior method* [36] was designed to simultaneously describe signal-independent and signal-dependent noises across different bands in HSI [37] and can obtain a good performance.

To analyze the effectiveness of the variational model with the embedded learning and model-constrained network learning approaches, some representative methods are chosen in the following experiments. These approaches are the deep learning model fast and flexible discriminative

CNN denoiser (FFDNet) (<https://github.com/NavyZeng/DPLRTA>) [38], deep plug-and-play model with the combined FFDNet and low-rank tensor approximation (DPLRTA) (<https://github.com/NavyZeng/DPLRTA>) [31], and DHSP (<https://github.com/acecreamu/deep-hs-prior>) [34], which is the model-constrained, loss function-based procedure. The Hyperspectral Digital Imagery Collection Experiment Urban data set, with a cropped size of $256 \times 256 \times 191$, was adopted in the simulated HSI data-denoising experiments to interpret the performance of different approaches. As the DHSP-based system is mainly proposed to remove Gaussian noise, $\sigma_b = 25$ is only added to each band, and these images before and after restoration are displayed as false color with the 57th, 27th, and 17th bands in Figure 7. From the results, it is apparent that all are able to remove the Gaussian noise, but the DHSP can better maintain the spectral characteristic than others, while the image restored by the DPLRTA shows good structure and texture information.

With the mean peak SNR (PSNR), the mean structural similarity (SSIM) index and mean spectral angle (MSA) mapper served as evaluation indices, the quantitative assessment also indicates that DPLRTA obtained the best PSNR and SSIM, and the DHSP gives the best MSA. In addition, the authors in [31] have proven that DPLRTA can remove the large-scale sparse noise compared with the FFDNet. However, as an unsupervised method, the DHSP

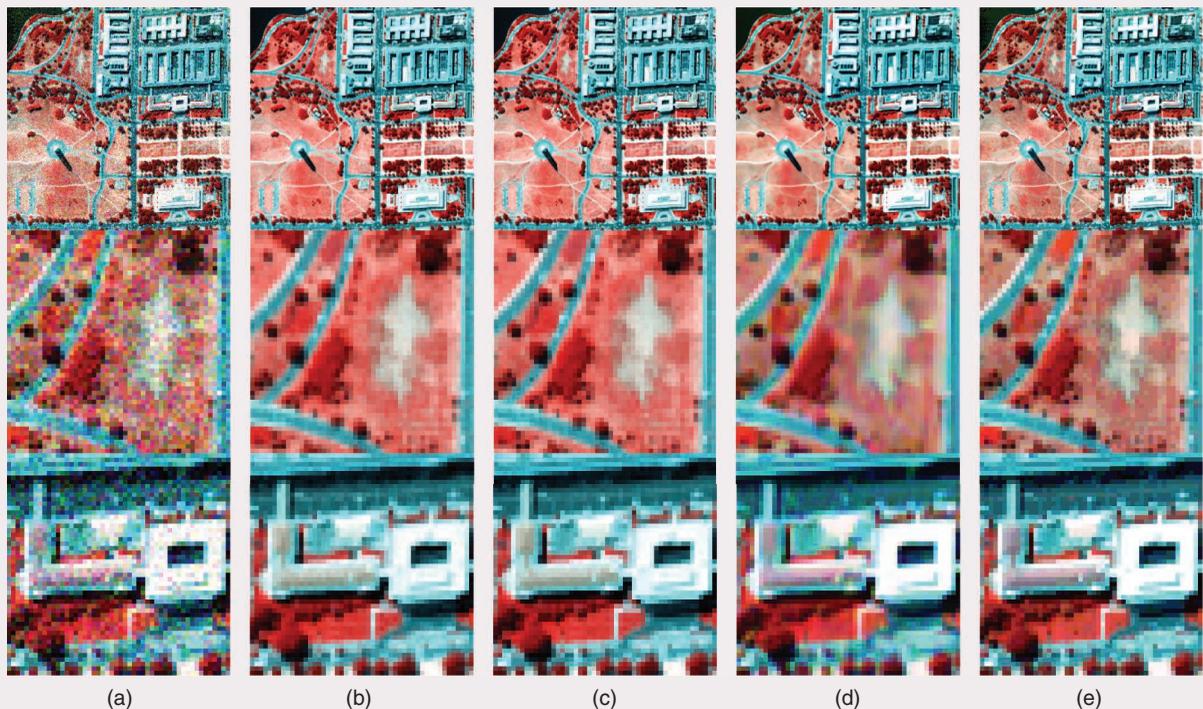


FIGURE 7. A visual comparison and quantitative evaluation with peak SNR (dB), mean structural similarity, and mean spectral angle values of the denoised results in the Washington D.C. Mall data set. (a) A noisy image (57, 27, 17) (20.613 dB, 0.741, 12), (b) FFDNet (25.775 dB, 0.931, 6.602), (c) DPLRTA (28.320dB, 0.953, 5.978), (d) DHSP (27.786dB, 0.924, 4.777), and (e) an original image.

can obtain an image prior within a CNN itself and flexibly adapt for situations without an significant amount of training data available. Certainly, to overcome the hybrid noise as DPLRTA, the loss function should be adjusted to describe the properties of different noises, rather than only the MSE function for Gaussian noise.

Overall, compared with variational models with embedded learning, the model-constrained network learning approach can adaptively manage hyperparameter learning in the network training process, instead of manually adjusting regularization parameters, and can obtain better convergence results for nonlinear problems. Conversely, variational models with embedded learning can be flexibly adjusted for different types of noise by adding different prior models. However, the accuracy and training speed of HSI denoising using coupled model- and data-driven methods still need to be considered.

REMOTE SENSING IMAGE FUSION

Due to the limitations of hardware, remote sensing images are tradeoffs between spatial and spectral resolution. *Remote sensing image fusion* refers to the fusion of multiple images with complementary information to obtain higher-resolution remote sensing images. Here we refer to spatial-spectral fusion, which aims at obtaining a fused image with both high-spatial and spectral resolutions by fusing high-spatial-resolution (HR)/low-spectral-resolution images and low-spatial resolution (LR)/high-spectral-resolution images. This includes panchromatic (Pan)/multispectral image (MSI) fusion, Pan/HSI fusion, and MSI/HSI fusion. Pan/HSI fusion can be regarded as a special case of MSI/HSI fusion.

For remote sensing image fusion, the energy function can be generally represented as (5), where $X \in \mathbb{R}^{WH \times S}$ is the ideal high-spatial-spectral resolution image. W , H , and S are the width, height, and band number, respectively, of the ideal image. $Y \in \mathbb{R}^{wh \times s}$ denotes the observed LR/high-spectral-resolution image. W/w is the spatial resolution ratio of the corresponding X and Y . $Z \in \mathbb{R}^{WH \times s}$ is the observed HR/low-spectral-resolution image, with $s \ll S$. The first term is the spectral fidelity term, where $H \in \mathbb{R}^{wh \times HW}$ is the downsampling and blurring matrix. The third term is the spatial fidelity term.

In remote sensing fusion, scholars have proposed methods belonging to each coupling class. In the following sections, we introduce representative fusion techniques for each coupling class in detail.

FUSION USING THE DATA- AND MODEL-DRIVEN CASCADING APPROACH

As an example of the data- and model-driven cascading approach, Shen et al. [20] proposed the the combination of deep learning and variational model (DL-VM) technique, which utilizes the gradient prior obtained from a CNN to construct the spatial fidelity term for Pan/MSI fusion, and it is the first scheme combining data- and model-driven methods for remote sensing image fusion. (The source

codes for DL-VM are available at <https://github.com/SGGJeryLi/WHU-SGG-RS-Pro-Group.git>.) The energy function of [20] is written as

$$X = \underset{x}{\operatorname{argmin}} \frac{1}{2} \|Y - HX\|_2^2 + \frac{\gamma}{2} \sum_{j=1}^2 \|\nabla_j X - G_j\|_2^2 + \frac{\lambda}{2} \|QX\|_2^2 \quad (25)$$

The second term is the spatial fidelity term, where $G_j \in \mathbb{R}^{WH \times S}$ with $j = 1, 2$ denotes gradient images of the high-resolution MSI in the horizontal and vertical directions learned through the network, corresponding to X^* in Figure 2. $\nabla_j \in \mathbb{R}^{WH \times WH}$ with $j = 1, 2$ means the global first-order, finite-difference matrix in the horizontal and vertical directions, respectively. Constructing the spatial fidelity term with the learned gradient priors, G_j avoids the inaccurate linear assumption of the relationship between the Pan and high-resolution MSI. $Q \in \mathbb{R}^{MN \times MN}$ in the third term indicates the Laplacian matrix, which is a common Laplacian prior term.

In this class, in addition to pansharpening, Dian et al. [13] proposed a deep HSI sharpening method (DHSIS) for MSI/HSI fusion. Xie et al. [19] proposed a HS pansharpening method with deep priors for Pan/HSI fusion. The data- and model-driven cascading strategy uses the prior learned from a pretrained network to construct the energy function, which avoids the linear assumption in the spatial fidelity or regularization term of the classic model-driven method. However, some limitations exist in this strategy, such as 1) it requires retraining of the network when fusing data of different sensors, and 2) it takes time to iteratively solve the energy function.

To show the effectiveness of the data- and model-driven cascading approach in Pan/MSI fusion, a low-resolution, QuickBird MSI with the size of $248 \times 248 \times 4$ and a QuickBird Pan image with the size of $992 \times 992 \times 1$, with resolutions of 2.44 and 0.61 m, respectively, were adopted in the simulated experiment according to Wald's protocol. The traditional component, substitution-based approach of adaptive IHS pan-sharpening (AIHS) [39], model-based two-step sparse coding method (TSSC) method [40], and deep learning-based technique of deep residual pan-sharpening neural network (DRPNN) [41] were applied to compare with the cascading DL-VM scheme [20]. Figure 8 shows the red, green, blue-band combinations of the various fusion systems, where the lower-right corner is a magnified display of the image inside the red rectangle. By comparing the results, it can be observed that, for AIHS and TSSC, sharpened spatial features are achieved, but with severe spectral distortion, as depicted in the vegetation area in Figure 8(c) and (d). DRPNN shows good performance in spectral fidelity but poor performance in spatial texture information enhancement, as illustrated in the zoomed-in area in Figure 8(e). The fusion result obtained by DL-VM is the closest to the reference image, both in the fusion of the spatial details and in the preservation of the spectral fidelity.

FUSION USING VARIATIONAL MODELS WITH THE EMBEDDED LEARNING APPROACH

In the variational models using the embedded learning approach, Dian et al. [22] proposed the CNN denoiser based HSI and MSI fusion (CNN-Fus) method, which combines subspace representation and a CNN denoiser for MSI/HSI fusion. The augmented Lagrangian function of the constructed model can be written as

$$\begin{aligned} \alpha, V, W = \operatorname{argmin}_{\alpha, V, W} & \|Y - H\alpha\psi\|_F^2 + \gamma \|Z - \alpha\psi P\|_F^2 \\ & + \rho \|V - \alpha + W\|_F^2 + \lambda g(V) \\ \text{s.t. } & V = \alpha; \quad \text{with } X = \alpha\psi, \end{aligned} \quad (26)$$

where $\psi \in \mathbb{R}^{L \times S}$ is the subspace according to the subspace estimation [42], which can be estimated from the observed low-resolution MSI. $\alpha \in \mathbb{R}^{WH \times L}$ represents the coefficients. ψ and α are similar to the overcomplete dictionary and coefficient in sparse representation, respectively, as presented in (4). As shown in (12), V is the auxiliary variable introduced by the ADMM algorithm.

As depicted in Figure 3, variables V , α , and W are iteratively solved. Among these variables, variable α is updated by a fast method based on solving a Sylvester equation

[43]. Variable V is updated by the pretrained CNN denoiser (FFDnet [38] is utilized) as

$$V = \mathcal{F}\left(\left(\alpha^{t+1} - W^t\right), \frac{\lambda}{2\rho^{t+1}}; \Theta\right), \quad (27)$$

where \mathcal{F} is the CNN denoiser, Θ represents the parameters of \mathcal{F} , $\alpha^{t+1} - W^t$ is the input of \mathcal{F} , and $(\lambda/2\rho^{t+1})$ is noise level σ^2 in the Gaussian denoiser.

The CNN denoiser helps to suppress noise and the potential artifacts generated in the iteration. Meanwhile, the flexibility of the CNN denoiser guarantee that a well-trained denoising network can be plugged into various image restoration problems without retraining. However, when the noise of remote sensing images is not significant, a denoising prior may not be the optimal option.

To show the effectiveness of variational models using the embedded learning approach in MSI/HSI fusion, in [22], a low-resolution Hyperion MSI with the size of $100 \times 100 \times 89$ and a high-resolution *Sentinel-2* MSI with the size of $300 \times 300 \times 4$ with resolutions of 30 and 10 m, respectively, were adopted in the real-data experiment. The original low-resolution Hyperion HSI has 220 spectral bands in the spectral range of 400–2,500 nm, of which 89 bands were retained after removing the bands with a low SNR. The original high-resolution

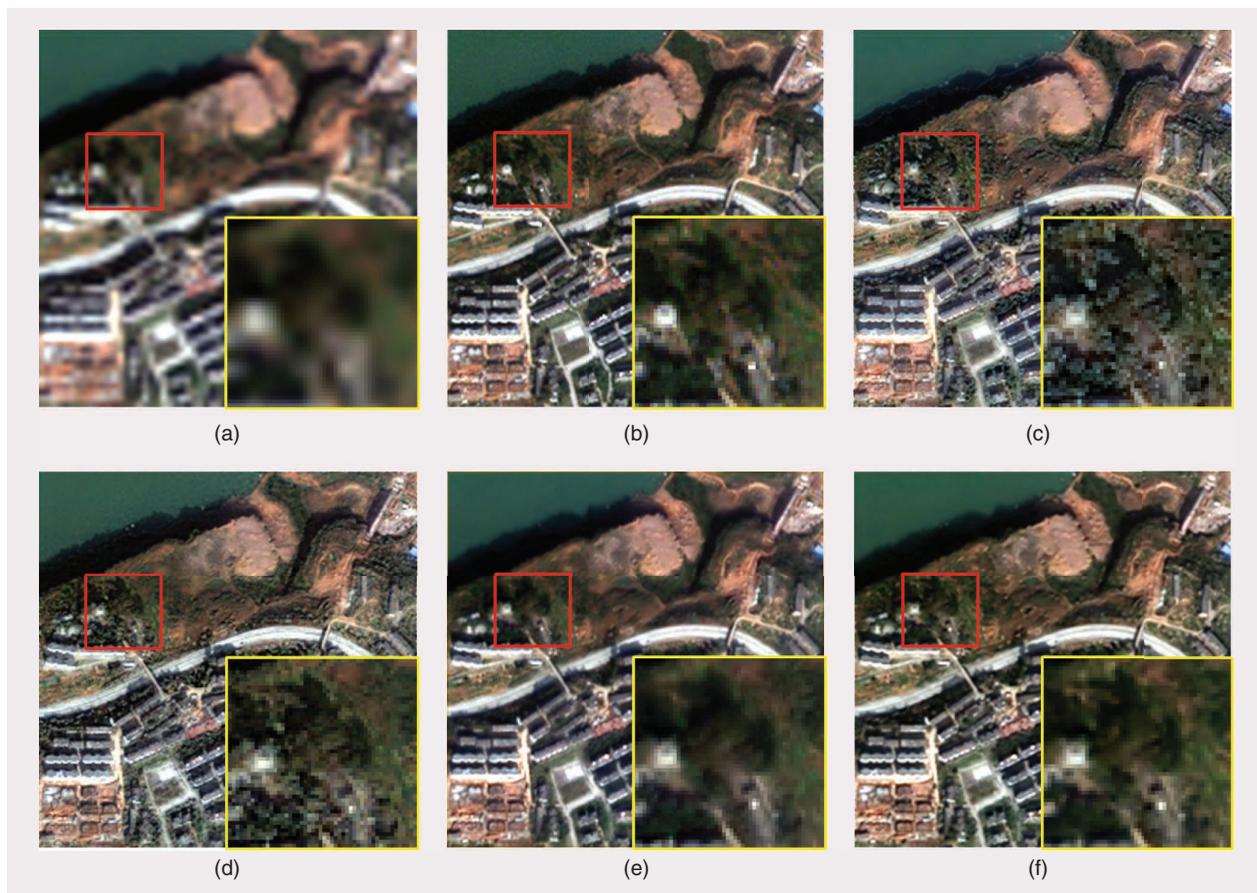


FIGURE 8. The fusion results for the QuickBird MSI and Pan images. (a) A low-resolution MSI image, (b) reference image, (c) AIHS, (d) TSSC, (e) DRPNN, and (f) DL-VM.

Sentinel-2 MSI has 13 spectral bands. Four bands with central wavelengths, i.e., 490, 560, 665, and 842 nm, were utilized in this experiment. Three state-of-the-art fusion approaches were used for comparison with the CNN-Fus coupling method, which includes a nonnegative structured sparse representation (NSSR) [44], coupled spectral unmixing (CSU) [45], and coupled sparse tensor factorization (CSTF) [46]. Figure 9 shows the false-color images consisting of the 16th, fifth, and second bands of the fused, high-resolution HSIs. As presented in Figure 9, all of the fusion approaches can improve the spatial resolution of the observable low-resolution HSI. The CSU method performs poorly in the spectral fidelity term, and obvious artifacts can be seen in the fusion results of the NSSR and CSTF. The fusion results of CNN-Fus have much fewer flaws, which shows the superiority of the CNN denoiser compared to the traditional handcrafted regularization prior.

FUSION USING THE MODEL-CONSTRAINED NETWORK STRUCTURE APPROACH

In the model-constrained network structure approach, Xie et al. [47] proposed the MS/HS Fusion Net method, which unfolds the algorithm into an optimization-inspired deep network for MSI/HSI fusion. The energy function can be written as

$$\hat{Z} = \min_Z \|H(ZA + \hat{Z}B) - Y\|_F^2 + \lambda f(\hat{Z})$$

with $X = ZA + \hat{Z}B$. (28)

A novel, relational formulation between the observed high-resolution MSI Z and ideal high-resolution HSI X is introduced as $X = ZA + \hat{Z}B$, where $\hat{Z} \in \mathbb{R}^{r \times s}$ is the unknown variable to be sought after, and $A \in \mathbb{R}^{s \times s}$ and $B \in \mathbb{R}^{r \times s}$ are the corresponding known coefficient matrices. Then the spectral fidelity term between the LR HSI Y and X can be expressed as $Y = H(ZA + \hat{Z}B)$. Additional details of this method can be found in [47].

Equation (28) is first decomposed using a proximal gradient algorithm then unfolded into an end-to-end deep network, which is similar to Figure 4. Different approaches of this class can adopt different optimization algorithms. Figure 4 takes the ADMM algorithm as an example, while in [47] it is the proximal gradient algorithm.

Similar to remote sensing image denoising, the fusion network structure in [47] is designed based on the optimization processing of the objective function to make the network interpretable, i.e., (28). Each block of the network represents an iterative solution. Hence, for this approach, the selection of a good network design and a robust optimization strategy can contribute to the admirable performance of this coupling method.

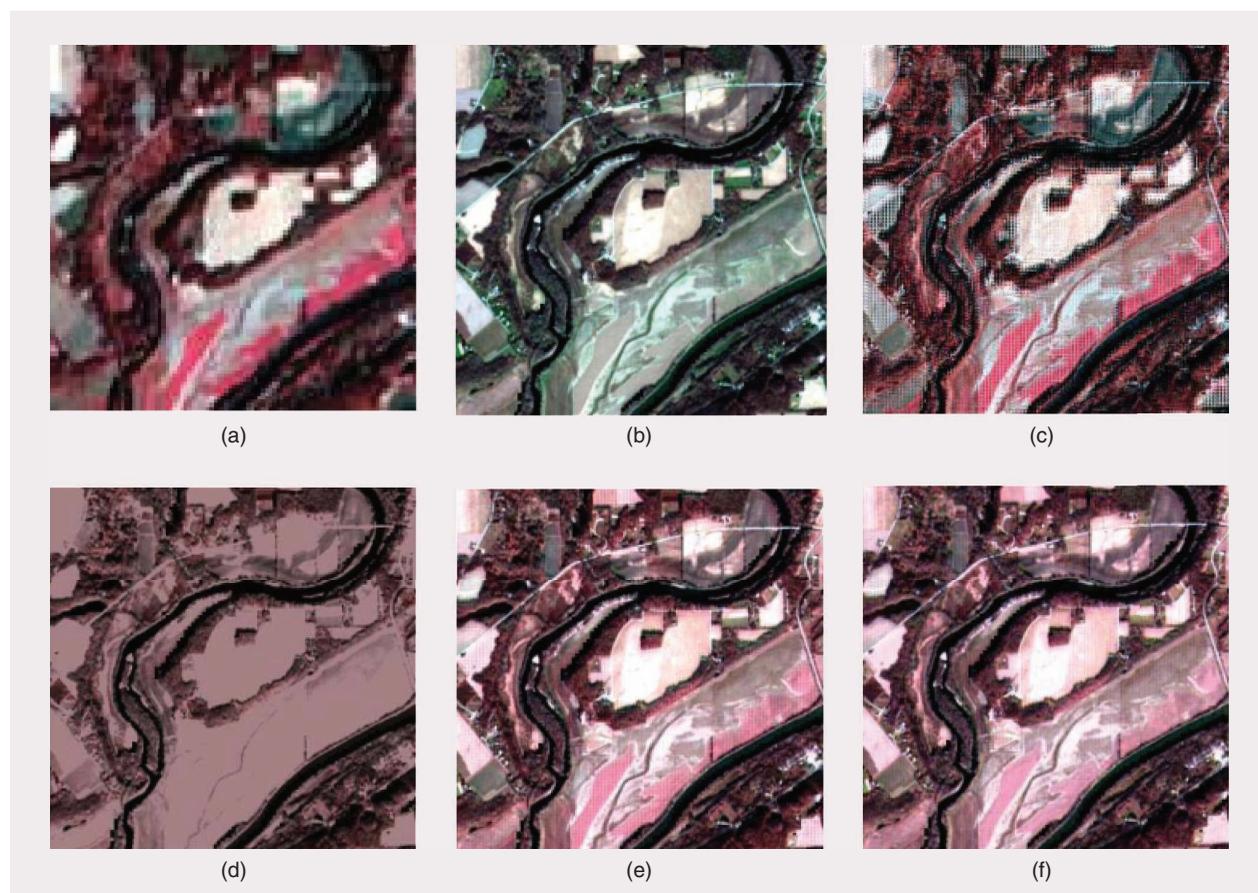


FIGURE 9. The fusion results for the Hyperion MSI and *Sentinel-2A* MSI images (a) A Hyperion low-resolution MSI, (b) *Sentinel-2A* high-resolution MSI, (c) NSSR, (d) CSU, (e) CSTF, and (f) CNN-Fus [22].

FUSION USING THE MODEL-CONSTRAINED LOSS FUNCTION APPROACH

In the model-constrained, loss function approach, Zhang et al. [48] proposed an unsupervised deep framework for blind HSI superresolution. The loss function in [48] is

$$\begin{aligned} \min_{\Theta, P, k} & \|Z - XP\|^2 + \|Y - (X * k)_{\downarrow r}\|^2 + \lambda (\|k\|_2^2 + \|P\|_F^2) \\ \text{s.t. } & X = f_{\Theta}(E), \end{aligned} \quad (29)$$

where $P \in \mathbb{R}^{s \times s}$ denotes the degeneration in the spectral domain, k denotes the unknown blur kernel, \downarrow_r indicates the downsampling operation in the spatial domain with scaling factor r , and $*$ indicates the blurring operation in the spatial domain. The first two terms correspond to the spatial fidelity and spectral fidelity terms, respectively. The third term is the regularization term that imposes constraints on the degradation processes, that is, k and P , which guides the network to estimate the degradations, and improves the ability of the fusion network to comply with the unknown degenerations in real HSI superresolution applications. Θ -parameterized $f_{\Theta}(\cdot)$ represents the image generator network for the latent X , and E is a precomputed code that contains the image-specific statistics of X . Integrating the observation model into the loss

function construction reduces the need for ideal images as labeled data for network training, which is very practical and suitable for situations where limited sample data are available.

In general, in remote sensing image fusion, the data- and model-driven cascading approaches and the variational model with embedded learning approach are simple and easy to implement. The model-constrained network structure approach shows good performance due to its inter-

pretable structure, but it needs careful mathematical derivation and network structure design. Therefore, there is still not much related work in remote sensing image fusion. Due to the low dependence on ideal, high-quality images as labeled data, the model-constrained, loss function approach has been used in some unsupervised fusion schemes recently, and it has the potential to be joined with the other three combined approaches.

FURTHER DEVELOPMENTS

Although the coupled model- and data-driven methods have shown their potential in learning the nonlinear correlations in remote sensing imagery and characterizing the structural this insight of optimization based on the image degradation process, challenges still exist due to the

complexity of physical information, availability of samples, and universality of the models. In the following sections, we suggest a few future directions.

STRUCTURE DEVELOPMENT

INTRODUCTION OF GRAPH NEURAL NETWORKS

Two aspects limit the development of combined strategies in remote sensing applications. First, the convolution operation underpinning all CNN architectures is unable to capture nonlocal, self-similarity patterns because of the locality of convolution kernels. Furthermore, the objects or pixels in multimodal data, including optical data, infrared data, SAR data, lidar data, and socioeconomic data, contain complex relationships and interdependencies. The complexity of the data has imposed significant challenges on existing CNN architectures. Graph neural networks (GCNs) are able to describe the complex relationships between data by the nodes and edges in the feature space of the network [49]–[51] and have shown great potential for capturing self-similarity information and coping with heterogeneous multimodal data. Very recently, a graph convolutional denoiser network [52] was proposed to aggregate k -nearest neighboring patches for image denoising. In addition, by further mining the multiscale recurrence property of a natural image, cross-scale internal GCN [53] was proposed to construct a graph on different-resolution patches and successfully recover more detailed textures. However, due to the big difference between CNNs and GCNs, when the data are structured as a graph, the combination of iterative optimization and GCNs still faces big challenges. First, graph-based models are less efficient than CNN-based ones. The graph-based models usually need to handle the whole image, with complex topology structures containing a large number of nodes. For each iteration in the optimization, the graph should be updated with changes of an optimal solution. Hence, graph convolutions need to adapt to the dynamicity of the graphs. The change of adjacency relations for each node, however, may also pose a burden to computational efficiency and introduce turbulence into network optimization.

UTILIZATION OF UNSUPERVISED LEARNING

Most of the existing data-driven image restoration methods are based on supervised learning, i.e., they utilize the ideal clean images as labeled data and train the network through a large number of samples. However, in actual situations, the following two problems may be encountered. One is that there may be no ideal clean images as labeled data of networks, such as clean HSIs or PolSAR images in the denoising. The other is the insufficient number of sample data pairs, for example, due to the difficulty in obtaining pairwise hyperspectral and multispectral images at the same time. Therefore, scholars have to degrade the observations and use the original observations as

ALTHOUGH THE COUPLED MODEL- AND DATA-DRIVEN METHODS HAVE SHOWN THEIR POTENTIAL IN LEARNING THE NONLINEAR CORRELATIONS IN REMOTE SENSING IMAGERY AND CHARACTERIZING THE STRUCTURAL THIS INSIGHT OF OPTIMIZATION.

ideal clean images. In this case, unsupervised learning that does not require ideal clean images is required. Some scholars have used unsupervised learning for remote sensing image restoration [17], [32], [54], [55]. They have attempted to integrate the observation model in 1) into an unsupervised network, but all of them have mostly adopted the combined form of the model-constrained, loss function approach. Thus, the network structure of the unsupervised network still lacks interpretability. In the future, we should consider other model- and data-driven combinations in unsupervised learning, especially in the form of a model-constrained network structure approach. Furthermore, in unsupervised learning, it will be valuable to combine the model-driven loss function approach with other approaches.

COMBINED WITH TENSOR THEORY

Due to the inherent 3D characteristics of remote sensing images, the previous vector/matrix-based combined methods have a limited ability to fully exploit the multidimensional structural correlation, in comparison with directly working on the high-order tensor format image. Tensor-based procedures have received increased attention for preserving the intrinsic structural correlation, obtaining better restoration results, especially the low-rank, tensor-based methods in the last two years. Currently, the tensor rank is mostly applied to accelerate the deep learning-based optimization process [56]. Tensor-based models have also been combined with deep learning, but mainly in hyperspectral denoising [31], [57], and they have rarely been used in other remote sensing applications. This combination has been achieved with variational models using embedded learning in which a global variational framework is established based on tensor representation, and a deep convolutional network is employed to obtain a plug-and-play prior of ideal images. In the future, on the one hand, tensor models could be embedded into deep learning using other combined processes and then provide more physical constraints for data-driven methods. On the other hand, tensor decomposition could be used to obtain more essential feature components of the data, which could be extracted or updated using network training.

APPLICATION CHALLENGES

THE COMPLEX DEGRADATION PROBLEM

Thin clouds, haze, and shadow often cause unevenness in remote sensing images, which can influence human interpretation. Haze/thin cloud degradation is triggered by the scattering of radiance from turbid particles in the atmosphere, which makes the ground information distorted. Shadow leads to an intensity decrease due to the obstruction of the incident light. Therefore, it is essential to correct the unevenness. Many deep learning-based cloud-/shadow-removal methods have been proposed, but only a few are combined with variational models [14],

[58] to handle natural images. For example, Yang and Sun [14] unfold the iterative algorithm with the transmission and dark channel priors to be a deep network, and Liu et al. [58] formulated image dehazing as the minimization of a variational model with favorable data fidelity terms and prior terms, and then solved the variational model based on the classical gradient descent method with built-in, deep CNNs. However, no model-driven deep learning approaches have been proposed to remove thin clouds, haze, and shadow from remote sensing images. Compared to natural images that are covered with uniform haze or shadow, the actual intensity in remote sensing images is usually nonuniform and thus more complex in spatial and spectral domains. Two possible alternative ideas are put forward here to solve the problem. One possibility is to simulate the nonuniform and complex degenerations through a convolutional layer rather than simple image statistics. Second, although the unevenness caused by haze, thin clouds, or shadow has no obvious distribution law in the spatial domain, a strict physical scattering law related to the intensity of spectral bands can be incorporated into deep learning to improve the solution [59]–[61].

LARGE AREA MISSING INFORMATION

Due to sensor malfunction and adverse atmospheric conditions, there can often be a great deal of missing information in optical remote sensing data. This makes missing information reconstruction technology important. Recently, some researchers have devoted efforts to develop inpainting based on coupled model- and data-driven methods [34], [62]. For example, Sidorov and Hardeberg [34] used the intrinsic properties of a CNN without any training to obtain the regularization terms in a variational model, and Lahiri et al. [62] first trained a generative model to map a latent prior distribution to the natural image manifold and search for the best-matching prior to reconstruct the signal. However, in remote sensing images, the land-cover types are complex, especially for high-resolution images. Moreover, thick clouds are usually accompanied with cloud shadows, which can be further divided into umbra and penumbra. This makes the inpainting for remote sensing images face serious challenges.

HETEROGENEOUS IMAGE FUSION

The existing model- and data-driven combinations are mostly adopted for single-type image restoration, such as SAR image denoising or optical image fusion. The fusion of heterogeneous images with different statistical properties can, in theory, improve the performance of remote sensing images, such as missing information reconstruction and resolution enhancement by merging HR SAR and high-spectral-resolution optical images. In fact, there have already been some studies of heterogeneous image fusion [63]–[67]. Most of the early heterogeneous image

fusion methods were based on a simple linear assumption [64], [65]. Since then, some heterogeneous fusion techniques based on the model-driven approach [66], [67] have gradually been developed. However, due to the difference in imaging mechanisms between heterogeneous images, it is difficult to construct an explicit and linear relationship. In recent years, with the rapid development of deep learning, some scholars have turned to the use of deep learning for heterogeneous image fusion [68]–[70]. In heterogeneous image fusion, the coupling of model- and data-driven methods has not yet been attempted, and there is much research space.

CONCLUSIONS

In this article, we systematically reviewed the combinations of model- and data-driven methods for remote sensing image restoration and fusion. The combined approaches solve the black-box problem of deep learning methods by introducing the model-driven scheme to direct parameter learning and improve physical interpretability of the model. The coupling approaches can be further divided into three types: 1) data- and model-driven cascading, 2) variational models with embedded learning, and 3) model-constrained network learning. These techniques have been widely applied in remote sensing image restoration and fusion, especially in SAR image despeckling, HSI hybrid noise reduction, and remote sensing image fusion. The results described in this article confirm that the use of these approaches can result in a significant improvement for remote sensing image restoration and fusion.

However, research into the coupling of model- and data-driven methods is still young. We believe that some new insights into the potential improvement for remote sensing applications have been provided in this article. From the perspective of the model's structure, GCNs and deep tensor decomposition [71] can be integrated into a combined framework for better utilization of the spatial nonlocal self-similarity patterns and high spectral-correlation property. The essential tasks in the future will be determining how best to remove dependency on the large numbers of training samples, the best way to achieve transfer learning among data sets, and unsupervised learning without labeled HR images. Conversely, more diverse application directions are worth studying, such as dehazing, cloud removal, and heterogeneous image fusion. Exploring proper data-driven priors based on an optimization-inspired variational mode for these more complex and specific remote sensing problems also remains a big challenge.

ACKNOWLEDGMENTS

This work was supported, in part, by the National Natural Science Foundation of China under grants 62071341 and 41971303. Jie Li is the corresponding author.

AUTHOR INFORMATION

Huanfeng Shen (shenhf@whu.edu.cn) received his B.S. degree in surveying and mapping engineering and his Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 2002 and 2007, respectively. In 2007, he joined the School of Resource and Environmental Sciences (SRES), Wuhan University, Wuhan, 430079, China, where he is currently a Luojia Distinguished Professor and dean of the SRES. He was or is the principal investigator of two projects supported by the National Key Research and Development Program of China and six projects supported by the National Natural Science Foundation of China. He has authored more than 100 research papers in peer-reviewed international journals. He is currently a member of the editorial board of *Journal of Applied Remote Sensing* and *Geography and Geo-Information Science*. His research interests include remote sensing image processing, multisource data fusion, and intelligent environmental sensing. He is a Senior Member of IEEE, council member of the China Association of Remote Sensing Application, an Education Committee member of the Chinese Society for Geodesy Photogrammetry and Cartography, and Theory Committee member of the Chinese Society for Geospatial Information Society.

Menghui Jiang (jiangmenghui@whu.edu.cn) received her B.S. degree in geographical science from Wuhan University, China, in 2017. She is currently pursuing her Ph.D. degree in the School of Resource and Environmental Sciences, Wuhan University, Wuhan, 430079, China. Her research interests include image data fusion, quality improvement, remote sensing image processing, and deep learning.

Jie Li (aaronleecool@whu.edu.cn) received his B.S. degree in sciences and techniques of remote sensing and his Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 2011 and 2016, respectively. He is currently an associate professor with the School of Geodesy and Geomatics, Wuhan University, Wuhan, 430079, China. His research interests include image quality improvement, image superresolution reconstruction, data fusion, remote sensing image processing, sparse representation, and deep learning. He is a Member of IEEE.

Chenxia Zhou (zhoucx31@whu.edu.cn) received her B.S. degree in geographical information science from Anhui University, Hefei, China, in 2018 and her M.S. degree from the School of Resource and Environmental Sciences, Wuhan University, China, in 2021. Her research interests include synthetic aperture radar image despeckling, quality improvement, remote sensing image processing, and deep learning.

Qiangqiang Yuan (yqiang86@gmail.com) received his B.S. degree in surveying and mapping engineering and his Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 2006 and 2012,

respectively. In 2012, he joined the School of Geodesy and Geomatics, Wuhan University, where he is a Professor. He published more than 90 research papers, including more than 70 peer-reviewed articles in international journals, such as *Remote Sensing of Environment*, *ISPRS Journal of Photogrammetry and Remote Sensing*, *IEEE Transactions Image Processing*, and *IEEE Transactions on Geoscience and Remote Sensing*. He is an associate editor of five international journals and has frequently served as a referee for more than 40 international journals for remote sensing and image processing. His research interests include image reconstruction, remote sensing image processing and application, and data fusion. He is a Member of IEEE.

Liangpei Zhang (zlp62@whu.edu.cn) received his B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, his M.S. degree in optics from Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and his Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998. Currently, he is a Chang-Jiang Schola Chair Professor, appointed by the Ministry of Education of China, in the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, 430072, China. Previously, he was a principal scientist for the China State Key Basic Research Project (2011–2016) and was appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has published more than 700 research papers and five books. He is an Institute for Scientific Information highly cited author and a Clarivate highly cited researcher. He holds 30 patents. He was a recipient of the 2010 Best Paper Boeing Award, the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing, the 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics, the 2020 TGRS award, and the 2020 GRSS David Landgrebe Award from IEEE. He also serves as an associate editor or editor for more than 10 international journals. He is currently serving as an associate editor for *IEEE Transactions on Geoscience and Remote Sensing*. He is the founding chair of the IEEE Geoscience and Remote Sensing Society Wuhan Chapter. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence. He is a fellow of the Institution of Engineering and Technology. He is a Fellow of IEEE.

REFERENCES

- [1] H. Shen *et al.*, "Missing information reconstruction of remote sensing data: A technical review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 61–85, Sep. 2015, doi: 10.1109/MGRS.2015.2441912.
- [2] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 318–322, Jan. 2014, doi: 10.1109/LGRS.2013.2257669.
- [3] X. Lu, L. Dong, and Y. Yuan, "Subspace clustering constrained sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3007–3019, May 2020, doi: 10.1109/TGRS.2019.2946751.
- [4] J. Duran, A. Buades, B. Coll, and C. Sbert, "A nonlocal variational model for pansharpening image fusion," *SIAM J. Imaging Sci.*, vol. 7, no. 2, pp. 761–796, 2014, doi: 10.1137/130928625.
- [5] P. Liu, L. Xiao, and T. Li, "A variational pan-sharpening method based on spatial fractional-order geometry and spectral-spatial low-rank priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1788–1802, Dec. 2017, doi: 10.1109/TGRS.2017.2768386.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.
- [8] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, arXiv:1406.1078.
- [9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.
- [10] X. Lu, T. Gong, and X. Zheng, "Multisource compensation network for remote sensing cross-domain scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2504–2515, Apr. 2020, doi: 10.1109/TGRS.2019.2951779.
- [11] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 2808–2817, doi: 10.1109/CVPR.2017.300.
- [12] K. Zhang, W. Zuo, and L. Zhang, "Deep plug-and-play super-resolution for arbitrary blur kernels," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 1671–1681, doi: 10.1109/CVPR.2019.00177.
- [13] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 1–11, Nov. 2018, doi: 10.1109/TNNLS.2018.2798162.
- [14] D. Yang and J. Sun, "Proximal dehaze-net: A prior learning-based deep network for single image dehazing," in *Proc. Eur. Conf. Comput. Vision (ECCV)*, 2018, pp. 702–717.
- [15] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 10–18.
- [16] J. Zhang, and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 1828–1837, doi: 10.1109/CVPR.2018.00196.

- [17] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 9446–9454, doi: 10.1109/CVPR.2018.00984.
- [18] G. Aubert and J.-F. Aujol, "A variational approach to removing multiplicative noise," *SIAM J. Appl. Math.*, vol. 68, no. 4, pp. 925–946, 2008, doi: 10.1137/060671814.
- [19] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1529–1543, May 2020, doi: 10.1109/TNNLS.2019.2920857.
- [20] H. Shen, M. Jiang, J. Li, Q. Yuan, Y. Wei, and L. Zhang, "Spatial-spectral fusion by combining deep learning and variational model," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6169–6181, Apr. 2019, doi: 10.1109/TGRS.2019.2904659.
- [21] S. Venkatakrisnan, C. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2013, pp. 945–948, doi: 10.1109/GlobalSIP.2013.6737048.
- [22] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by CNN denoiser," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1124–1135, Apr. 2020, doi: 10.1109/TNNLS.2020.2980398.
- [23] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2305–2318, 2018, doi: 10.1109/TPAMI.2018.2873610.
- [24] E. Kobler, T. Klatzer, K. Hammernik, and T. Pock, "Variational networks: connecting variational methods and deep learning," in *Proc. 39th German Conf. Pattern Recognit. (GCPR)*, 2017, pp. 281–293.
- [25] H. Shen, C. Zhou, J. Li, and Q. Yuan, "SAR image despeckling employing a recursive deep CNN prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 273–286, May 2021, doi: 10.1109/TGRS.2020.2993319.
- [26] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018, doi: 10.1109/TGRS.2017.2783902.
- [27] C. Deledalle, L. Denis, S. Tabti, and F. Tupin, "MuLoG, or how to apply Gaussian denoisers to multi-channel SAR speckle reduction?" *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4389–4403, 2017, doi: 10.1109/TIP.2017.2713946.
- [28] M. B. Alver, A. Saleem, and M. Çetin, "A novel plug-and-play SAR reconstruction framework using deep priors," in *Proc. IEEE Radar Conf. (RadarConf)*, 2019, pp. 1–6, doi: 10.1109/RA-DAR.2019.8835598.
- [29] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "Speckle2Void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3065461.
- [30] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (GIS)*, 2010, pp. 270–279, doi: 10.1145/1869790.1869829.
- [31] H. Zeng, X. Xie, H. Cui, Y. Zhao, and J. Ning, "Hyperspectral image restoration via CNN denoiser prior regularized low-rank tensor recovery," *Comput. Vision Image Underst.*, vol. 197, p. 103004, Aug. 2020.
- [32] B. Lin, X. Tao, and J. Lu, "Hyperspectral image denoising via matrix factorization and deep prior regularization," *IEEE Trans. Image Process.*, vol. 29, pp. 565–578, Jun. 2019, doi: 10.1109/TIP.2019.2928627.
- [33] L. Wang, C. Sun, Y. Fu, M.H. Kim, and H. Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 8032–8041, doi: 10.1109/CVPR.2019.00822.
- [34] O. Sidorov and J. Yngve Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vision Workshop*, 2019, pp. 3844–3851, doi: 10.1109/ICCVW.2019.00477.
- [35] I. Choi, M. H. Kim, D. Gutierrez, D. S. Jeon, and G. Nam, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–13, 2017, doi: 10.1145/3130800.3130810.
- [36] Q. Zhang, Q. Yuan, J. Li, F. Sun, and L. Zhang, "Deep spatio-spectral Bayesian posterior for hyperspectral image non-IID noise removal," *ISPRS J. Photogramm. Remote Sens.*, vol. 164, pp. 125–137, Jun. 2020, doi: 10.1016/j.isprsjrs.2020.04.010.
- [37] Z. Yue, H. Yong, Q. Zhao, L. Zhang, and D. Meng, "Variational denoising network: Toward blind noise modeling and removal," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 1688–1699.
- [38] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, May 2018, doi: 10.1109/TIP.2018.2839891.
- [39] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 746–750, May 2010, doi: 10.1109/LGRS.2010.2046715.
- [40] C. Jiang, H. Zhang, H. Shen, and L. Zhang, "Two-step sparse coding for the pan-sharpening of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 5, pp. 1792–1805, May 2014, doi: 10.1109/JSTARS.2013.2283236.
- [41] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017, doi: 10.1109/LGRS.2017.2736020.
- [42] M. A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 274–288, Jan. 2016, doi: 10.1109/TIP.2015.2496263.
- [43] Q. Wei, N. Dobigeon, J.-Y. Tourneret, J. Bioucas-Dias, and S. Godsill, "R-FUSE: Robust fast fusion of multiband images based on solving a Sylvester equation," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1632–1636, Nov. 2016, doi: 10.1109/LSP.2016.2608858.

- [44] W. Dong *et al.*, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016, doi: 10.1109/TIP.2016.2542360.
- [45] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, 2015, pp. 3586–3594, doi: 10.1109/ICCV.2015.409.
- [46] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018, doi: 10.1109/TIP.2018.2836307.
- [47] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multi-spectral and hyperspectral image fusion by MS/HS fusion net," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 1585–1594, doi: 10.1109/CVPR.2019.00168.
- [48] L. Zhang, J. Nie, W. Wei, Y. Li, and Y. Zhang, "Deep blind hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2388–2400, Jun. 2021, doi: 10.1109/TNNLS.2020.3005234.
- [49] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw.*, vol. 32, no. 1, pp. 4–24, Jan. 2021, doi: 10.1109/TNNLS.2020.2978386.
- [50] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, and Y. Y. Tang, "Spectral-spatial graph convolutional networks for semisupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 241–245, Feb. 2019, doi: 10.1109/LGRS.2018.2869563.
- [51] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang, "Multiscale dynamic graph convolutional network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3162–3177, May 2020, doi: 10.1109/TGRS.2019.2949180.
- [52] D. Valsesia, G. Fracastoro, and E. Magli, "Deep graph-convolutional image denoising," *IEEE Trans. Image Process.*, vol. 29, pp. 8226–8237, Aug. 2020, doi: 10.1109/TIP.2020.3013166.
- [53] S. Zhou, J. Zhang, W. Zuo, and C. Loy, "Cross-scale internal graph neural network for image super-resolution," Oct. 2020, arXiv:2006.16673.
- [54] A. Gokaslan, V. Ramanujan, D. Ritchie, K. I. Kim, and J. Tompkin, "Improving shape deformation in unsupervised image-to-image translation," in *Proc. Eur. Conf. Comput. Vision (ECCV)*, 2018, pp. 649–665.
- [55] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inf. Fusion*, vol. 62, pp. 110–120, Oct. 2020, doi: 10.1016/j.inffus.2020.04.006.
- [56] S. Gao and X. Zhuang, "Rank-one network: An effective framework for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2020.3046476.
- [57] Y. Liu, X. Zhao, Y. Zheng, T. Ma, and H. Zhang, "Hyperspectral image restoration by tensor fibered rank constrained optimization and plug-and-play regularization," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3045169.
- [58] Y. Liu, J. Pan, J. Ren, and Z. Su, "Learning deep priors for image dehazing," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, 2019, pp. 2492–2500, doi: 10.1109/ICCV.2019.00258.
- [59] H. Shen, C. Zhang, H. Li, Q. Yuan, and L. Zhang, "A spatial-spectral adaptive haze removal method for visible remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6168–6180, Sep. 2020, doi: 10.1109/TGRS.2020.2974807.
- [60] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, 2010, doi: 10.1109/TPAMI.2010.168.
- [61] P. Chavez Jr., "An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data," *Remote Sens. Environ.*, vol. 24, no. 3, pp. 459–479, 1988, doi: 10.1016/0034-4257(88)90019-3.
- [62] A. Lahiri, A. K. Jain, S. Agrawal, P. Mitra, and P. K. Biswas, "Prior guided GAN based semantic inpainting," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2020, pp. 13,696–13,705, doi: 10.1109/CVPR42600.2020.01371.
- [63] S. C. Kulkarni and P. P. Rege, "Pixel level fusion techniques for SAR and optical images: A review," *Inf. Fusion*, vol. 59, pp. 13–29, Jul. 2020, doi: 10.1016/j.inffus.2020.01.003.
- [64] J. R. Harris, R. Murray, and T. Hirose, "IHS Transform for the integration of radar imagery with other remotely sensed data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 12, pp. 1631–1641, 1990.
- [65] C. A. Dupas, "SAR And LANDSAT TM image fusion for land cover classification in the Brazilian Atlantic forest domain," *Int. Arch. Photogram. Remote Sens.*, vol. 33, pp. 96–103, Jan. 2000.
- [66] W. Zhang and L. Yu, "SAR and Landsat ETM+ image fusion using variational model," in *Proc. Int. Conf. Comput. Commun. Technol. Agriculture Eng.*, 2010, pp. 205–207, doi: 10.1109/CCTAE.2010.5544210.
- [67] M. Ghahremani and H. Ghassemian, "Compressed-sensing-based pan-sharpening method for spectral distortion reduction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2194–2206, 2016, doi: 10.1109/TGRS.2015.2497309.
- [68] C. Grohnfeldt, M. Schmitt, and X. Zhu, "A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from sentinel-2 images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2018, pp. 1726–1729, doi: 10.1109/IGARSS.2018.8519215.
- [69] J. Gao, Q. Yuan, J. Li, H. Zhang, and X. Su, "Cloud removal with fusion of high resolution optical and SAR images using generative adversarial networks," *Remote Sens.*, vol. 12, no. 1, p. 191, 2020, doi: 10.3390/rs12010191.
- [70] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 333–346, Aug. 2020, doi: 10.1016/j.isprsjprs.2020.05.013.
- [71] S. Oymak and M. Soltanolkotabi, "End-to-end learning of a convolutional neural network via deep tensor decomposition," May 2018, arXiv: 1805.06523.