Hyperspectral Compressive Snapshot Reconstruction via Coupled Low-Rank Subspace Representation and Self-Supervised Deep Network

Yong Chen, Wenzhen Lai, Wei He, *Senior Member, IEEE*, Xi-Le Zhao, *Member, IEEE*, and Jinshan Zeng, *Member, IEEE*

***Abstract*— Coded aperture snapshot spectral imaging (CASSI) is an important technique for capturing three-dimensional (3D) hyperspectral images (HSIs), and involves an inverse problem of reconstructing the 3D HSI from its corresponding coded 2D measurements. Existing model-based and learning-based methods either could not explore the implicit feature of different HSIs or require a large amount of paired data for training, resulting in low reconstruction accuracy or poor generalization performance as well as interpretability. To remedy these deficiencies, this paper proposes a novel HSI reconstruction method, which exploits the global spectral correlation from the HSI itself through a formulation of model-driven low-rank subspace representation and learns the deep prior by a data-driven self-supervised deep learning scheme. Specifically, we firstly develop a model-driven low-rank subspace representation to decompose the HSI as the product of an orthogonal basis and a spatial representation coefficient, then propose a data-driven deep guided spatial-attention network (called *DGSAN*) to adaptively reconstruct the implicit spatial feature of HSI by learning the deep coefficient prior (DCP), and finally embed these implicit priors into an iterative optimization framework through a self-supervised training way without requiring any training data. Thus, the proposed method shall enhance the reconstruction accuracy, generalization ability, and interpretability. Extensive experiments on several datasets and imaging systems validate the superiority of our method. The source code and data of this article will be made publicly available at https://github.com/ChenYong1993/LRSDN.**

Manuscript received 19 November 2022; revised 28 October 2023; accepted 7 January 2024. Date of publication 22 January 2024; date of current version 25 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62101222, Grant 42271370, Grant 62376110, Grant 12371456, Grant 12171072, Grant 62131005, and Grant 61977038; in part by the Natural Science Foundation of Jiangxi, China, under Grant 20232ACB212001, Grant 20224BAB212001, and Grant 20224ACB212004; in part by the Young Elite Scientists Sponsorship Program by Jiangxi Association for Science and Technology (JXAST) under Grant 2023QT12; in part by the Thousand Talents Plan of Jiangxi Province under Grant jxsq2019201124; in part by the Sichuan Science and Technology Program under Grant 2023ZYD0007; in part by the Fund of Hubei Key Laboratory of Inland Shipping Technology under Grant NHHY2023003; and in part by the National Key Research and Development Program of China under Grant 2020YFA0714001. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hairong Qi. *(Corresponding authors: Wei He; Jinshan Zeng.)*

Yong Chen, Wenzhen Lai, and Jinshan Zeng are with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang 330022, China (e-mail: chenyong1872008@163.com; 202141600128@jxnu.edu.cn; jinshanzeng@jxnu.edu.cn).

Wei He is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: weihe1990@whu.edu.cn).

Xi-Le Zhao is with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: xlzhao122003@163.com).

Digital Object Identifier 10.1109/TIP.2024.3354127

***Index Terms*— Hyperspectral imaging, hyperspectral image reconstruction, low-rank subspace representation, self-supervised deep network.**

I. Introduction

HYPERSPECTRAL imaging systems capture the spectral signature of a spatial scene as three-dimensional (3D) cubic data over tens to hundreds of discrete bands. The abundant spectral information in hyperspectral image (HSI) has been widely used in many fields, including computer vision [1], remote sensing [2], and medical image processing [3], [4] and so on.

To obtain the 3D HSI, conventional hyperspectral imaging systems scan the scene with multiple exposures along the spatial or spectral dimension, which is time-consuming for the imaging procedure and cannot be used to capture dynamic scenes or video with high-speed rates [5], [6]. Recently, motivated by the mature compressive sensing theory, snapshot compressive imaging (SCI) systems [7], [8], [9], [10], [11] have attracted much attention due to their advantages on capturing dynamic scenes and balancing the temporal and spatial resolution. Among existing SCI systems, the coded aperture snapshot spectral imaging (CASSI) system [12], [13] is a representative one, which samples snapshots along the spectral dimension by the coded aperture in each spectral band, and then compresses the sampled images along the spectrum into a single 2D measurement, as shown in Fig. 2. CASSI systems are generally divided into two phases: the exposure measure phase for encoding the 3D HSI into a single 2D compressive image, and the computational reconstruction phase for recovering the underlying HSI from the snapshot measurement, where the reconstruction of high-quality HSI from the coded 2D measurements is one key phase in this imaging system.

In the past decade, numerous approaches have been proposed for HSI reconstruction from 2D compressed measurements [14], [15], [16], [17]. Since the reconstruction process is an ill-posed inverse problem, model-based approaches design hand-crafted priors, such as total variation (TV) [18], [19], sparsity [20], [21], low-rank [22], [23], and nonlocal self-similarity [24], [25], for HSI reconstruction. Although the kind of model-based methods generally have good interpretability and sometimes achieve satisfactory performance,

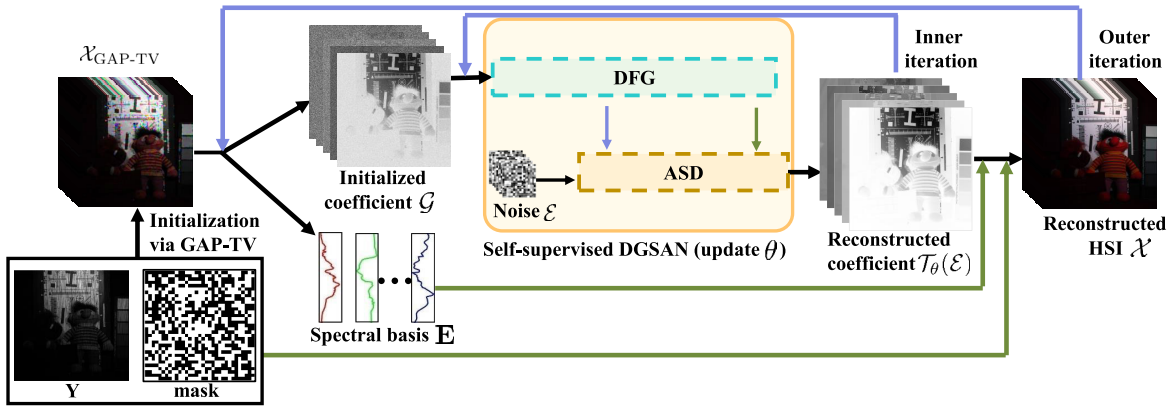


Fig. 1. Illustration of the proposed method. The reconstructed HSI is first initialized by GAP-TV. Then, the global spectral correlation of HSI is depicted by model-driven low-rank subspace representation, and the spatial feature of HSI is reconstructed by data-driven self-supervised DGSAN. Finally, the model-driven and data-driven priors are embedded in an iterative optimization algorithm to promote each other.

the hand-designed priors may not well depict the discriminatively intrinsic structures of HSIs, resulting in unsatisfactory results. Alternatively, the kind of deep learning-based methods have attracted extensive attention recently. Different from model-based approaches designing the prior manually, deep learning-based approaches [26], [27], [28], [29], [30] learn the prior through an end-to-end reconstruction manner from the coded 2D measurement, and often achieve impressive performances. However, deep learning-based methods typically require a large amount of paired training samples, which can be expensive to collect, and their generalization ability may be limited when applied to other systems. These factors restrict the widespread adoption of deep learning-based methods. To alleviate the data collection burden and motivated by the universal prior between the natural images and HSIs, several model-based optimization with deep priors approaches [31], [32], [33] have been recently introduced for HSI reconstruction, through incorporating pre-trained or untrained deep networks into a traditional optimization framework. This type of approach enhances the interpretability and generalization ability compared to the model-based and supervised learning-based approaches. However, the performance of these approaches heavily depends on the design of the deep denoiser or untrained networks.

In this paper, we focus on the development of efficient HSI reconstruction methods by combining the advantages of both model-driven and data-driven approaches to deal with the existing challenges described before. Specifically, we propose a novel HSI reconstruction method called *LRSDN* via coupled low-rank subspace representation and self-supervised deep network to respectively capture the global spectral low-rank prior and deep coefficient prior (DCP). First, instead of completely using data-driven deep learning approaches, we develop the model-driven low-rank prior to explore the global spectral correlation of HSI based on a hypothesis that spectral vectors of the HSI lie in a low-dimensional subspace and can be represented by the product of orthogonal spectral basis and representation coefficient [23], [34], [35], [36]. Then, the data-driven DCP is learned from the proposed self-supervised deep guided spatial-attention network (DGSAN) without any external training data to reconstruct the complex nonlinearity spatial feature of HSI. Finally, the model-driven low-rank prior and data-driven DCP could promote each other in an iterative optimization algorithm. In each iteration, the optimization variables and network parameters are updated by closed-form solutions and learned DGSAN, respectively. The illustration of the proposed reconstruction method is shown in Fig. 1. Our contributions can be summarized as follows:

- We formulate the HSI reconstruction into a self-supervised model-driven and data-driven framework, and propose an iterative optimization method that couples model-driven low-rank prior and data-driven DCP. In the iterative process, two priors can promote each other to improve the reconstructed interpretability and accuracy.
- Instead of directly employing existing unsupervised neural networks or pre-trained deep denoising networks, we design a novel self-supervised neural network called DGSAN to learn the DCP. DGSAN reconstructs the representation coefficient from the guidance data and snapshot measurement without any external data for pre-training. Thus, the proposed method can guarantee generalization ability and is suitable for different imaging systems and datasets.
- An efficient alternating direction method of multipliers (ADMM) algorithm is designed to develop an iterative optimization algorithm for HSI reconstruction. Extensive experimental results on both DD-CASSI and SD-CASSI systems illustrate that the proposed method outperforms model-based optimization with hand-crafted and deep priors methods and achieves competitive results with supervised learning-based methods.

The rest of this paper is organized as follows. Section II reviews some related works used for HSI reconstruction. In Section III, we present some notations and the problem formulation of two different CASSI systems. Section IV gives the proposed HSI reconstruction model, DGSAN, and optimization algorithm. Experimental results on several datasets and the discussion are illustrated in Section V. Finally, a summary is presented in Section VI.

II. Related Work

In this section, we briefly review the related HSI reconstruction methods, which can be roughly divided into three categories: model-based approaches, deep learning-based approaches, and model-based optimization with deep priors approaches.

A. Model-Based HSI Reconstruction Approaches

Due to the ill-posed problem of HSI reconstruction from a snapshot measurement, model-based methods usually exploit various hand-crafted priors to model the intrinsic properties of HSI and then design corresponding regularizers to reconstruct the desired HSIs by solving optimization problems. The sparse prior with different bases (such as wavelet basis or overcomplete dictionary) were utilized to explore the spatial-spectral sparsity of HSI [12], [20], [37], [38]. To characterize the local spatial piece-wise smoothness of HSI, the TV regularization has been employed for HSI reconstruction [18], [19], [39]. Furthermore, Low-rank matrix/tensor approximation were widely designed to explore the global spatial-spectral correlation [22], [23], [40] and nonlocal self-similarity [24], [25], [41]. The interpretability and generalization of model-based methods can be guaranteed, but these hand-crafted priors lack an adaptive ability to capture the characteristics of different HSIs, resulting in unsatisfactory reconstruction quality.

B. Deep Learning-Based HSI Reconstruction Approaches

Recently, benefiting from the ability to learn complex structural features, deep learning-based methods have been demonstrated to achieve promising results in HSI reconstruction tasks. These methods can be roughly categorized into three classes: end-to-end (E2E) [26], [27], [30], [42], [43], [44], [45], [46], [47], [48], deep unfolding [28], [29], [45], [49], [50], [51], and single-sample generative models [52], [53], [54], [55]. The principle of E2E reconstruction is to implicitly learn the image priors from sufficient training data and then construct an E2E mapping function between the observed measurement and the original HSI. For example, a unified convolutional neural network (CNN) framework is proposed in [43] to reconstruct the HSI from spectrally undersampled projections. To utilize the attention mechanisms, λ-net [26] introduced spatial attention blocks in U-Net for HSI reconstruction. Furthermore, HDNet [30] designed the spatial-spectral attention module to provide fine pixel-level features. Inspired by the transformer being more effective than CNN in many tasks, MST-L [46] used a transformer framework to capture the remote dependence of HSI and further explore the spectral structure using spectral self-attention.

Differing from E2E methods, the deep unfolding methods combine the prior knowledge of the observation model and unfold the reconstruction process based on iterative optimization into a multi-stage network, with each stage corresponding to one iteration in the optimization algorithm. Zhang et al. [49] learned a tensor low-rank prior for HSI in the feature domain and integrated it into an iterative optimization algorithm. JR2net [51] proposed a joint non-linear representation and recovery unfolding network for compressive spectral imaging. DAUHST [50] designed a half-shuffle transformer and incorporated it into the degradation-aware unfolding framework. With sufficient training data and time, E2E and deep unfolding methods yield impressive performance. However, the available HSIs for training are very limited due to the high cost of collecting the HSI. Single-sample generative models design an untrained network for HSI reconstruction, which does not require training data. Bacca et al. [54] utilized Tucker representation to analyze the structure of HSI and modeled it within a deep neural network, enhancing its representational capability in high-dimensional structural information. Gelvez et al. [55], [56] decomposed the HSI into the product of basis matrices and coefficient matrices, which are individually learned as the weights and features of the deep neural network. Although single-sample generative models can solve the problem of training data, there is a lack of model-driven optimization algorithms to guide network learning.

C. Model-Based Optimization With Deep Priors Approaches

Model-based optimization with deep priors approaches employ the pre-trained or untrained deep networks as regularization and then combine it with traditional optimization algorithms for HSI reconstruction [31], [32], [33], [57], [58]. PnP-GAP [33] introduced pre-trained deep denoiser into a generalized alternating projection framework for HSI reconstruction. To make the deep prior and hand-crafted prior promote each other, TV-FFDNet [32] merged the FFDNet with TV prior into the PnP framework. However, pre-trained deep denoisers still require sufficient training data and struggle to represent complex spatial information in different HSIs adaptively. Recently, PnP-DIP-HSI [31] adopted untrained deep image priors (DIP) [52] within the PnP mechanism, significantly enhancing the generalization capability. The introduction of untrained deep neural networks into the iterative optimization algorithm has shown great potential. Unfortunately, due to the lack of additional data, untrained networks introduce uncertainty during the parameter optimization stage, limiting their performance.

III. Notations and Problem Formulation

A. Notations

In this paper, lowercase and uppercase (b, $B \in \mathbb{R}$), boldface lowercase ($\mathbf{x} \in \mathbb{R}^b$), boldface capital letter ($\mathbf{X} \in \mathbb{R}^{I \times J}$), and calligraphic letter ($\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$) are employed to represent scalars, vectors, matrices, and tensors respectively. The element value of $\mathcal{X}$ in location $(i_1, i_2, \cdots, i_n)$ is represented by $\mathcal{X}(i_1, i_2, \cdots, i_n)$. The mode-k unfolding of tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ is represented as $\mathbf{X}_{(k)} \in \mathbb{R}^{I_k \times I_1 I_{k-1} \cdots I_{k+1} I_n}$. In contrast, matrix $\mathbf{X}_{(k)}$ along the k-mode folds to a tensor is denoted as $\mathcal{X} = \text{fold}_k(\mathbf{X}_{(k)})$. $\mathcal{X}(:, :, i_3)$, $\mathcal{X}(:, i_2, :)$, and $\mathcal{X}(i_1, :, :)$ are the frontal, lateral, and horizontal slices of a 3D tensor $\mathcal{X}$, respectively. The symbol of mode-k tensor-matrix product is $\times_k$, and the operator is defined as $(\mathcal{X} \times_k \mathbf{U})_{i_1, \cdots, i_{k-1} j i_{k+1}, \cdots, i_n} = \sum_{i_k} x_{i_1, i_2, \cdots, i_n} \cdot u_{j, i_k}$. For the Frobenius norm of $\mathcal{X}$, it is denoted as $||\mathcal{X}||_F = (\Sigma_{i_1, i_2, \cdots, i_n} x^2_{i_1 i_2 \cdots i_n})^{\frac{1}{2}}$.

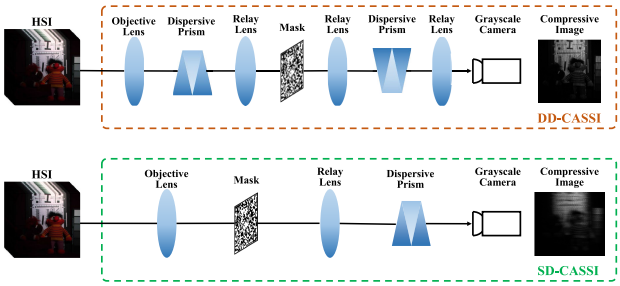


Fig. 2. Illustration of the optical coding principles of two representative CASSI imaging systems [42]. (top) Double dispersers CASSI measurement process. (Bottom) Single disperser CASSI measurement process.

B. Problem Formulation

CASSI is capable of capturing 3D HSI into 2D measurements, and we introduce two representative coded aperture snapshot imaging systems, i.e., double dispersers (DD-CASSI) and single disperser (SD-CASSI).

The first row of Fig. 2 presents the coding principle of the DD-CASSI imaging system [13]. The imaging systems first disperse the incident light field, then create a coded field by the coded aperture mask, and design another optics to unshear this coding. Mathematically, $\mathcal{X} \in \mathbb{R}^{M\times N\times C}$ denotes the 3D HSI, M and N are the spatial size, and C is the number of spectral channels. The intensity of final snapshot measurement $\mathbf{Y}$ in location (m, n) can be expressed as follows:

$$\mathbf{Y}(m,n)=\sum_{c=1}^{C}\varphi(m-\sigma^{dd}(c),n)\mathcal{X}(m,n,c), \tag{1}$$

where $1 \le c \le C$ is the spectral coordinate, $\varphi(m, n)$ represents the transfer function of the coded aperture mask, and $\sigma^{dd}(c)$ represents the wavelength-dependent dispersion function based on the dispersion prism in DD-CASSI.

In the SD-CASSI system [12], as shown in the second row of Fig. 2. The imaging systems first establish a coding of the incident light field and then shear the coded field by a dispersive element. The intensity of final snapshot measurement $\mathbf{Y}$ in location (m, n) can be expressed as follows:

$$\mathbf{Y}(m,n)=\sum_{c=1}^{C}\varphi(m-\sigma^{sd}(c),n)\mathcal{X}(m-\sigma^{sd}(c),n,c), \tag{2}$$

where $\sigma^{sd}(c)$ represents the wavelength-dependent dispersion function based on the dispersion prism in SD-CASSI.

For convenience, the degradation model of two CASSI imaging processes can be described as follows:

$$\mathbf{Y}=\Phi(\mathcal{X})+\mathbf{Z}, \tag{3}$$

where $\mathbf{Y} \in \mathbb{R}^{M\times N'}$ represents the captured measurement, whose size is dependent on the CASSI systems. For the SD-CASSI system, due to the effect of individual dispersion, the size of compression measurement is $M \times (N + C - 1)$. Since the second dispersion in DD-CASSI can offset the effect of the first dispersion, the compression measurement captured by the DD-CASSI system is the same as the original spatial dimensions $M \times N$. The operator $\Phi(\cdot) : \mathbb{R}^{M\times N\times C} \to \mathbb{R}^{M\times N'}$ contains all operations of the whole imaging process, and $\mathbf{Z}$ is the error or additive noise.

IV. Proposed Coupled HSI Reconstructed Method

The fundamental task of snapshot compression reconstruction is to reconstruct the HSI $\mathcal{X}$ from the measurement $\mathbf{Y}$ and imaging operator Φ. Due to the reconstruction problem being an ill-posed inverse problem, it is difficult to recover $\mathcal{X}$ directly from the degradation model (3). Therefore, it is necessary to constrain the solution space using regularization methods, and the reconstruction model can be formulated as:

$$\arg\min_{\mathcal{X}}\frac{1}{2}\|\mathbf{Y}-\Phi(\mathcal{X})\|_F^2+\lambda R(\mathcal{X}), \tag{4}$$

where $R(\mathcal{X})$ is the regularization term, characterizing the prior information of desirable HSI $\mathcal{X}$, and λ is the positive regularization parameter.

A. Low-Rank Subspace Representation of HSI

From a linear mixture model, each spectral signature can be represented by a linear combination of a small number of endmembers [59], which means a high spectral correlation in HSI. The high correlation is naturally captured by low-rank subspace representation that is demonstrated as a powerful tool for HSI processing tasks, such as denoising [34], [35] and superresolution [60]. Therefore, to capture the spectral correlation of HSI, we design model-driven low-rank subspace representation to approximate it:

$$\mathcal{X}=\mathcal{W}\times_3\mathbf{E}, \tag{5}$$

where $\mathbf{E} \in \mathbb{R}^{C\times K}$ $(K \ll C)$ is the spectral basis with the orthogonality of columns, i.e., $\mathbf{E}^T\mathbf{E} = \mathbf{I}$. $\mathcal{W} \in \mathbb{R}^{M\times N\times K}$ is the spatial representation coefficient.

The spectral basis $\mathbf{E}$ may be approximately learned from the HSI data itself using the singular value decomposition (SVD) or HySime algorithm [61]. In our work, we employ the SVD of $\mathcal{X}$ from the last iteration to learn the approximation solution of spectral basis $\mathbf{E}$ [62]. Given the estimated result $\mathcal{X}$, the spectral basis is updated as follows:

$$\mathbf{E}=\mathbf{U}(:,1:K), \tag{6}$$

where $\mathbf{U}$ is the left singular vector of $\mathbf{X}_{(3)}$. Given an estimation of spectral basis $\mathbf{E}$, the HSI reconstruction problem can be transformed into a reconstruction of the representation coefficient, so as to achieve the desirable HSI $\mathcal{X}$.

B. Proposed Coupled Model

With the spectral basis E known, and by introducing the low-rank subspace representation of HSI in (4), the reconstruction model can be formulated as:

$$\arg\min_{\mathcal{X},\mathcal{W}}\frac{1}{2}\|\mathbf{Y}-\Phi(\mathcal{X})\|_F^2,\quad s.t.\quad \mathcal{X}=\mathcal{W}\times_3\mathbf{E}. \tag{7}$$

Although the prior knowledge of spectral correlation is effectively explored by model (7), the spatial prior which promotes each other with spectral prior is ignored. With the orthogonal constraint of $\mathbf{E}$, the spatial prior of original HSI $\mathcal{X}$ can be

reflected on the reduced-dimensionality representation coefficient $\mathcal{W}$. Therefore, the reconstruction of spatial information of original $\mathcal{X}$ can be transformed into the estimation of $\mathcal{W}$. Thus, we can formulate the reconstruction model as:

$$\begin{aligned} &\arg\min_{\mathcal{X},\mathcal{W}} \frac{1}{2}\|\mathbf{Y}-\Phi(\mathcal{X})\|_F^2+\lambda R(\mathcal{W}),\\ &s.t.\ \ \mathcal{X}=\mathcal{W}\times_3\mathbf{E}, \end{aligned} \tag{8}$$

where $R(\mathcal{W})$ is the regularization term related to the spatial prior depiction of representation coefficient $\mathcal{W}$.

Since the representation coefficient $\mathcal{W}$ inherits the spatial characteristic of original HSI $\mathcal{X}$, the model-driven hand-crafted spatial prior regularization can also be employed to depict $\mathcal{W}$, such as TV regularization [36] and nonlocal self-similarity [34], [35]. However, hand-crafted priors may not suit different HSIs. Therefore, to adaptively represent the complex spatial features of different HSIs, the data-driven DCP that is learned by deep networks is used to excavate the implicit features. In general, by absorbing the regularization term $R(\mathcal{W})$, the coupled model can be generalized as:

$$\arg\min_{\mathcal{X},\theta}\frac{1}{2}\|\mathbf{Y}-\Phi(\mathcal{X})\|_F^2,\quad s.t.\ \ \mathcal{X}=\mathcal{T}_\theta(\mathcal{E})\times_3\mathbf{E}. \tag{9}$$

DCP assumes that the desired representation coefficient $\mathcal{W}$ is the output of deep neural network $\mathcal{T}_\theta(\mathcal{E})$, where $\mathcal{E}$ is a random tensor whose size is the same as $\mathcal{W}$, and θ is the network parameters to be learned.

The reconstruction model (9) simultaneously captures the spatial and spectral information of HSI by model-driven low-rank prior and data-driven DCP, respectively. However, we find that the two priors in model (9) could not be fused effectively. The reason is that we only constrain the results of subspace representation, thus the $\mathcal{T}_\theta(\mathcal{E})\times_3\mathbf{E}$ close to the observed measurement $\mathbf{Y}$, which is the only given input to the algorithm. In the process of learning DCP, the optimization of network parameters θ is not directly related to the available input $\mathbf{Y}$, which results in the difficulty for the network to learn spatial information. To fuse these two priors more effectively, we add a fidelity term into the model (9). Therefore, the final proposed coupled model is rewritten as follows:

$$\begin{aligned} &\arg\min_{\mathcal{X},\theta}\frac{1}{2}\|\mathbf{Y}-\Phi(\mathcal{X})\|_F^2+\frac{\lambda}{2}\|\mathbf{Y}-\Phi(\mathcal{T}_\theta(\mathcal{E})\times_3\mathbf{E})\|_F^2,\\ &s.t.\ \ \mathcal{X}=\mathcal{T}_\theta(\mathcal{E})\times_3\mathbf{E}. \end{aligned} \tag{10}$$

The proposed model is inherently interpretable and achieves better generalizability than fully model-based and supervised learning-based approaches. On the one hand, fully model-based approaches design hand-crafted priors that may not fit every data. In contrast, our model explores the DCP that can adaptively learn the implicit features of different data. On the other hand, supervised learning-based approaches are heavily dependent on sufficient training data and difficult to preserve the spectral correlation. The proposed model (10) applies the interpretable model-driven low-rank subspace representation to explore the global spectral correlation of HSI since the property exists on almost all HSIs. Moreover, although the deep network is used in our model, we do not need any training data. Therefore, we can expect the proposed model to bring robust and better HSI reconstruction results.

Previous single-sample generative models also combine low-rank and deep priors for HSI reconstruction [54], [55], [56]. However, they are very different from our work. First, they fused the low-rank and deep priors in an unsupervised deep generative network, while the proposed method couples low-rank and deep priors in an iterative optimization algorithm. Second, these methods employ the low-rank prior to guide the network to learn the low-dimensional structure of the image, i.e., these low-rank factors are learned as the weights and the features of the proposed network. On the contrary, the proposed method explores the low-rank prior of HSIs by low-rank subspace representation, where the spectral basis is learned by model optimization and the subspace representation coefficient is obtained by the proposed unsupervised deep generative network. Third, they design U-Net-based, AutoencoderNet-based, and ResNet-based as the unsupervised deep network. However, we propose a novel self-supervised neural network called DGSAN to learn the DCP.

C. Optimization Algorithm

The proposed model (10) is a constrained minimization problem, and the well-known ADMM [63], [64] is an efficient algorithm to solve it. Following the framework of ADMM, the augmented Lagrangian function of (10) is:

$$\begin{aligned} \mathcal{L}\{\mathcal{X},\theta,\mathcal{B}\}=\ &\frac{1}{2}\|\mathbf{Y}-\Phi(\mathcal{X})\|_F^2+\frac{\lambda}{2}\|\mathbf{Y}-\Phi(\mathcal{T}_\theta(\mathcal{E})\\ &\times_3\mathbf{E})\|_F^2+\frac{\mu}{2}\|\mathcal{X}-\mathcal{T}_\theta(\mathcal{E})\times_3\mathbf{E}-\mathcal{B}\|_F^2, \end{aligned} \tag{11}$$

where μ represents the positive penalty parameter, and $\mathcal{B}$ is the Lagrange multiplier. We then present how to solve each subproblem according to the ADMM framework.

1) θ-subproblem: Fixing other variables except θ, the θ-subproblem is formulated as:

$$\begin{aligned} \arg\min_{\theta}\frac{\lambda}{2}\|\mathbf{Y}-\Phi(\mathcal{T}_\theta(\mathcal{E})\times_3\mathbf{E})\|_F^2+\frac{\mu}{2}\|\mathcal{X}-\mathcal{T}_\theta(\mathcal{E})\times_3\mathbf{E}\\ -\mathcal{B}\|_F^2. \end{aligned} \tag{12}$$

It is worth noting that this subproblem is a regression problem using neural network models, where $\mathcal{T}_\theta$ is the network parameterized by θ, and $\mathcal{E}$ is the network input. Hence, existing off-the-shelf neural network optimizers, such as the network of DIP [52] and guided deep decoder [65], can be employed to update θ. However, the network architecture of DIP does not fully exploit the semantic features of an image. In our work, the self-supervised DGSAN (see Section IV-D) is designed to explore the DCP and update the parameter θ, which has been demonstrated to be more effective than DIP (see ablation study). To solve the θ subproblem, the gradient descent algorithm is a popular tool in complex network learning, and the gradient to θ can be computed by the standard backpropagation algorithm. Moreover, the Adam algorithm [66] is adopted as the optimizer, and the minimization problem (12) is chosen as the loss function of the DGSAN.

2) $\mathcal{X}$-subproblem: The subproblem of $\mathcal{X}$ is formulated as

$$\arg\min_{\mathcal{X}} \frac{1}{2}\|\mathbf{Y}-\Phi(\mathcal{X})\|_F^2+\frac{\mu}{2}\|\mathcal{X}-\mathcal{T}_\theta(\mathcal{E})\times_3\mathbf{E}-\mathcal{B}\|_F^2. \tag{13}$$

To facilitate the solution, we rewrite the $\mathcal{X}$-subproblem in the following equivalent form:

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_F^2+\frac{\mu}{2}\|\mathbf{x}-\mathbf{r}-\mathbf{b}\|_F^2, \tag{14}$$

where $\mathbf{y}$, $\mathbf{x}$, $\mathbf{r}$, and $\mathbf{b}$ are the vectorization forms of $\mathcal{Y}$, $\mathcal{X}$, $\mathcal{T}_\theta(\mathcal{E})\times_3\mathbf{E}$, and $\mathcal{B}$, respectively. $\mathbf{H}$ represents the sensing matrix, whose special structure can be referred to [31]. For the SD-CASSI system, the sensing matrix $\mathbf{H}$ is defined as:

$$\mathbf{H}=[\mathbf{D}_1,\ldots,\mathbf{D}_C]\in\mathbb{R}^{M(N+C-1)\times MNC},$$

where $\mathbf{D}_k=\begin{bmatrix}\mathbf{0}^{(1)}\\ \mathbf{A}_k\\ \mathbf{0}^{(2)}\end{bmatrix}\in\mathbb{R}^{M(N+C-1)\times MNC}$ with $\mathbf{A}_k=\text{diag}(\text{vec}(\mathbf{M}))\in\mathbb{R}^{MN\times MN}$ being a diagonal matrix with $\text{vec}(\mathbf{M})$ as its diagonal elements, where $\mathbf{M}$ is the sensing matrix. $\mathbf{0}^{(1)}\in\mathbb{R}^{M(k-1)\times MN}$ and $\mathbf{0}^{(2)}\in\mathbb{R}^{M(C-k)\times MN}$ are zero matrices. Similarly, the sensing matrix $\mathbf{H}$ in DD-CASSI system is defined as

$$\mathbf{H}=[\mathbf{D}_1,\ldots,\mathbf{D}_C]\in\mathbb{R}^{MN\times MNC},$$

where $\mathbf{D}_k=\text{diag}(\text{vec}(\mathbf{M}))\in\mathbb{R}^{MN\times MN}$ being a diagonal matrix with $\text{vec}(\mathbf{M})$ as its diagonal elements.

Problem (14) is a least-square problem, which is equivalent to solving the following linear problem:

$$\mathbf{x}=(\mathbf{H}^T\mathbf{H}+\mu\mathbf{I})^{-1}[\mathbf{H}^T\mathbf{y}+\mu(\mathbf{r}+\mathbf{b})]. \tag{15}$$

Because $\mathbf{H}$ is a fat matrix, the computation cost of $(\mathbf{H}^T\mathbf{H}+\mu\mathbf{I})^{-1}$ is high. Therefore, we simplify the inverse problem $(\mathbf{H}^T\mathbf{H}+\mu\mathbf{I})^{-1}$ using the matrix inversion formula,

$$(\mathbf{H}^T\mathbf{H}+\mu\mathbf{I})^{-1}=\mu^{-1}-\mu^{-1}\mathbf{H}^T(\mathbf{I}+\mu^{-1}\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H}\mu^{-1}. \tag{16}$$

Therefore, Eq. (15) can be rewritten as follows:

$$\mathbf{x}=\frac{\mathbf{H}^T\mathbf{y}+\mu(\mathbf{r}+\mathbf{b})}{\mu}-\frac{\mathbf{H}^T(\mathbf{I}+\mu^{-1}\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{H}^T\mathbf{y}}{\mu^2}-\frac{\mathbf{H}^T(\mathbf{I}+\mu^{-1}\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H}(\mathbf{r}+\mathbf{b})}{\mu}. \tag{17}$$

Since $\mathbf{H}\mathbf{H}^T$ is a diagonal matrix, we define it as $\mathbf{H}\mathbf{H}^T=\text{diag}\{\sigma_1,\ldots,\sigma_n\}$. Consequently, $(\mathbf{I}+\mu^{-1}\mathbf{H}\mathbf{H}^T)^{-1}$ and $(\mathbf{I}+\mu^{-1}\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{H}^T$ can be expressed as follows:

$$\begin{aligned}(\mathbf{I}+\mu^{-1}\mathbf{H}\mathbf{H}^T)^{-1}&=\text{diag}\{\frac{\mu}{\mu+\sigma_1},\ldots,\frac{\mu}{\mu+\sigma_n}\},\\ (\mathbf{I}+\mu^{-1}\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{H}^T&=\text{diag}\{\frac{\mu\sigma_1}{\mu+\sigma_1},\ldots,\frac{\mu\sigma_n}{\mu+\sigma_n}\}.\end{aligned} \tag{18}$$

Let y_i and $[\mathbf{H}(\mathbf{r}+\mathbf{b})]_i$ denotes the i-th element of $\mathbf{y}$ and $\mathbf{H}(\mathbf{r}+\mathbf{b})$, respectively. We plug Eq. (18) into Eq. (17) as:

$$\begin{aligned}\mathbf{x}=&\frac{\mathbf{H}^T\mathbf{y}}{\mu}+(\mathbf{r}+\mathbf{b})-\frac{1}{\mu}\mathbf{H}^T\\ &\left[\frac{y_1\sigma_1+\mu[\mathbf{H}(\mathbf{r}+\mathbf{b})]_1}{\mu+\sigma_1},\ldots,\frac{y_n\sigma_n+\mu[\mathbf{H}(\mathbf{r}+\mathbf{b})]_n}{\mu+\sigma_n}\right]^\top\\ =&(\mathbf{r}+\mathbf{b})\\ &+\mathbf{H}^T\left[\frac{y_1-[\mathbf{H}(\mathbf{r}+\mathbf{b})]_1}{\mu+\sigma_1},\ldots,\frac{y_n-[\mathbf{H}(\mathbf{r}+\mathbf{b})]_n}{\mu+\sigma_n}\right]^\top\\ =&\mathbf{H}^T[\mathbf{y}-\mathbf{H}(\mathbf{r}+\mathbf{b})]\oslash(\text{diag}(\mathbf{H}\mathbf{H}^T)+\mu)+(\mathbf{r}+\mathbf{b}).\end{aligned} \tag{19}$$

When the solution of $\mathbf{x}$ is obtained, then the original variable $\mathcal{X}$ can be achieved by reshaping the vector $\mathbf{x}$ as a tensor form.

3) multiplier $\mathcal{B}$ update: Based on the ADMM, the multiplier is further calculated by the following formulation:

$$\mathcal{B}\leftarrow\mathcal{B}-(\mathcal{X}-\mathcal{T}_\theta(\mathcal{E})\times_3\mathbf{E}). \tag{20}$$

Summarizing the optimization procedure of the whole process, we can obtain the pseudocode of the proposed LRSDN for HSI reconstruction in Algorithm 1. Before the reconstruction, the spectral basis should be learned in advance from the HSI. Since the original HSI is not available, we follow the previous HSI reconstruction works [24], [25] that initialize the HSI by GAP-TV [19] since its high efficiency. Moreover, the trade-off between reconstruction ability and image preservation ability is reflected in the subspace dimension of spectral basis. At the beginning of reconstruction, the initialized image is of low quality, leading to a correspondingly low-quality initial guidance coefficient. The initialized lower rank can achieve the satisfying reconstruction result without noise, but it leads to missing details. After the iteration, the reconstruction result is substantially improved, and we need a larger rank value to preserve more details of the image. Therefore, the iteration refining of the rank and spectral basis is designed to improve the reconstruction results. However, due to the dimension mismatch between the two iterations, we cannot update $\mathbf{E}$ from model (10) by the ADMM algorithm. Thus, we lose the closed-form solution of optimizing $\mathbf{E}$ via (6), which can be efficiently computed. The optimization of $\mathbf{E}$ is simply related to the reconstructed HSI of the last iteration and suitable for any rank K.

Algorithm 1 LRSDN-Based HSI Reconstruction Method

Input: Compressive measurement $\mathbf{Y}$, imaging operator Φ, parameters λ and μ.

Initialize: $\mathcal{B}=\mathcal{O}$, $\mathcal{X}=\mathcal{X}_{\text{GAP-TV}}$, subspace dimension $K=K^0$, and initialize $\mathcal{E}$ by random tensor.
for $t=1:T$ **do**
  Update subspace dimension $K^{(t)}=K^{(t-1)}+\gamma*(t-1)$.
  Learn orthogonal spectral basis $\mathbf{E}$ via (6) and initialize guidance coefficient via $\text{fold}_3(\mathbf{E}^T\mathbf{X}_{(3)})$.
  for $p=1:P$ **do**
    Update θ via (12).
    Update $\mathcal{X}$ via (19).
    Update multiplier $\mathcal{B}$ via (20).
    Update guidance coefficient $\mathcal{G}$ via $\mathcal{T}_\theta(\mathcal{E})$.
    Update $\mu=\eta*\mu$.
  end for
end for

Output: reconstructed HSI $\mathcal{X}$.

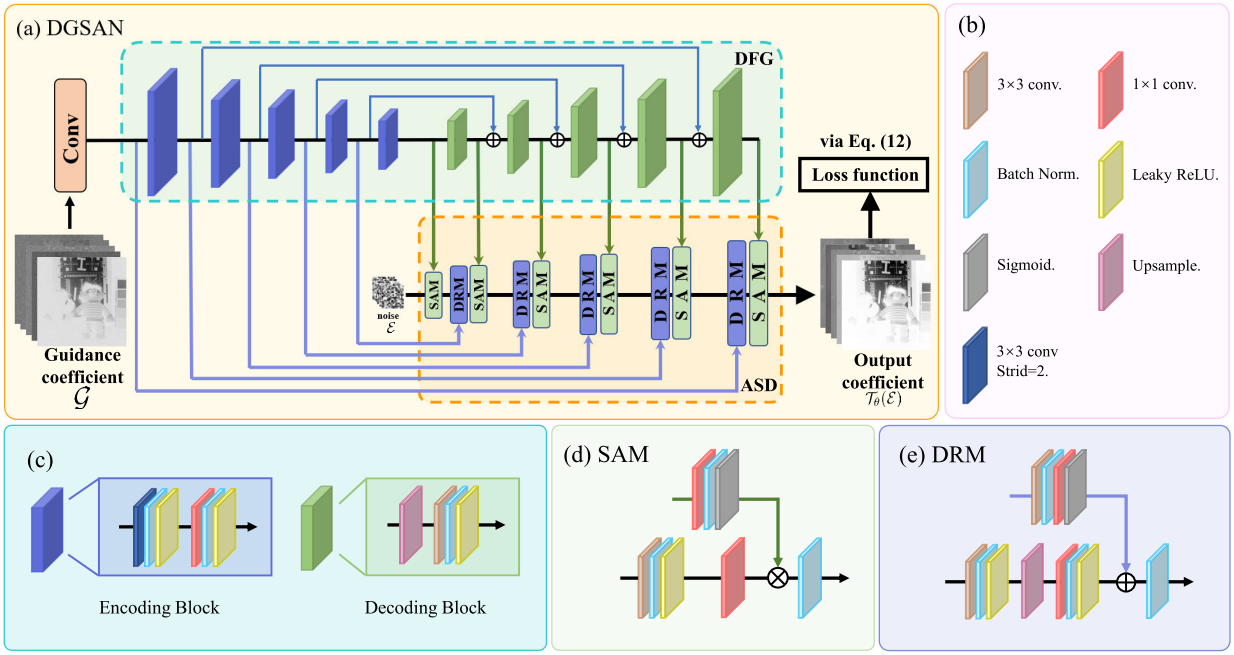


Fig. 3. The architecture of the proposed DGSAN for representation coefficient reconstruction. (a) The overall network, (b) operators used in the network, (c) the detail of encoding and decoding block in DFG, (d) the detail of SAM module in ASD, (e) the detail of DRM module in ASD.

D. Self-Supervised Deep Guided Spatial-Attention Network

In this part, we present in detail how the proposed self-supervised DGSAN explores the DCP and reconstructs the representation coefficient. Fig. 3 shows the schematic concept of the DGSAN, which mainly consists of two parts: deep feature generator (DFG) and attention-based spatial decoder (ASD). The idea of our DGSAN is to extract the multi-scale features of the input guidance coefficient $\mathcal{G}$ from low to high levels in the DFG module, and then the multi-scale features are employed to guide the parameter optimization in the deep decoder of the ASD module. In the ASD module, the random noise is generated as the initial data to reconstruct the representation coefficient. To gradually improve the quality of the guidance information, we employ the output of the last iteration of the network as the guidance coefficient. Since the guidance coefficient is obtained during the iterative process and the noise is randomly generated, DGSAN does not need any training data. Moreover, DGSAN leverages the representation coefficient from different stages. Thus, the proposed DGSAN effectively explores the data-driven prior of the representation coefficient. In the following, we describe the different modules of the proposed network in detail.

1) Deep Feature Generator: The DFG is an encoder-decoder network with skip connections, and the whole layout is similar to U-net, which is employed to extract the multi-scale features of the guidance coefficient. First, the initial encoded feature F_{en}^0 is obtained by a convolution operation on the guidance coefficient. Subsequently, the multi-scale features F_{en}^k are obtained by inputting F_{en}^0 into multiple encoding blocks. The whole encoding blocks are formulated as:

$$F_{en}^k = f_{en}^k(F_{en}^{k-1}), \tag{21}$$

where $f_{en}^k(\cdot)$ $(k = 1, 2, \cdots, K)$ denotes the operations contained by the k-th encoding block, and K denotes the number of encoding blocks.

The function of the decoding block enables to obtain decoded features F_{dc}^k that are paired with the encoded features in the spatial dimension. Especially, different from the traditional U-net with a skip connection that directly concatenates the encoding and decoded features, we utilize the idea of the residual network to add up these two features, which not only can alleviate the information missing in the encoding-decoding process but also further can promote the semantic alignment of paired encoded-decoded features. Therefore, the decoded features F_{dc}^K can be expressed as:

$$\begin{aligned} F_{dc}^1 &= f_{dc}^1(F_{en}^K), \\ F_{dc}^k &= f_{dc}^k(F_{dc}^{k-1} + f_{skip}(F_{en}^{K-k+1})), \end{aligned} \tag{22}$$

where $f_{dc}^k(\cdot)$ denotes the k-th decoding operation, and $f_{skip}(\cdot)$ is the information extraction of encoded features. In summary, the multi-scale encoded-decoded features extracted by DFG provide important information to guide the parameter optimization in ASD.

2) Attention-Based Spatial Decoder: The ASD module is the core deep decoder for the reconstruction of the representation coefficient, and it contains a consecutive semantic attention module (SAM) and a detail refinement module (DRM). Among them, the SAM module first extracts the attention weights from the decoded features F_{dc}^k of guidance coefficient, and then guides the semantic alignment of coefficient generator features. As the decoding features are achieved by merging the pairing features F_{dc}^{k-1} and F_{en}^{K-k-1}, thus they contain more abundant semantic information. In summary, the operation of the SAM module can be formulated as:

$$F_{sam}^0 = f_{sam}^0(\mathcal{E}, F_{en}^K),$$

TABLE I
QUANTITATIVE RESULTS OF COMPARISON METHODS ON THE CAVE DATASET

Index	Method	balloons	beads	beers	cd	face	feathers	flowers	food	slices	toy	Average
PSNR	GAP-TV	28.34	20.38	27.420	27.74	29.81	24.87	26.87	27.14	28.85	25.08	26.65
	DeSCI	27.64	21.46	34.228	26.69	33.21	27.66	27.13	29.03	30.34	26.33	28.37
	TV-FFDNet	40.00	24.16	37.857	34.97	37.59	30.83	34.18	35.35	34.21	29.61	33.87
	PnP-DIP-HSI	31.02	20.30	31.229	28.39	35.38	29.63	32.90	32.41	30.62	30.18	30.21
	NLTT-TV	38.40	22.45	36.201	28.55	38.88	31.36	31.98	34.06	36.14	33.75	33.18
	DPLR	38.75	27.82	37.25	31.65	39.82	35.21	39.19	36.21	36.22	35.36	35.75
	LRSDN	**42.03**	**31.26**	**40.48**	**35.37**	**41.79**	**37.54**	**40.36**	**39.76**	**39.71**	**39.21**	**38.75**
SSIM	GAP-TV	0.886	0.652	0.879	0.830	0.919	0.802	0.834	0.826	0.847	0.859	0.833
	DeSCI	0.957	0.739	0.966	0.928	0.946	0.912	0.898	0.909	0.930	0.910	0.909
	TV-FFDNet	0.984	0.796	0.979	0.945	0.966	0.937	0.937	0.950	0.950	0.927	0.937
	PnP-DIP-HSI	0.848	0.654	0.835	0.890	0.921	0.788	0.876	0.850	0.902	0.867	0.843
	NLTT-TV	0.979	0.752	0.974	0.926	0.976	0.917	0.923	0.926	0.945	0.946	0.926
	DPLR	0.956	0.755	0.960	0.815	0.952	0.880	0.925	0.889	0.897	0.894	0.892
	LRSDN	**0.985**	**0.921**	**0.983**	**0.960**	**0.983**	**0.969**	**0.972**	**0.970**	**0.971**	**0.977**	**0.969**
FSIM	GAP-TV	0.892	0.800	0.869	0.873	0.922	0.870	0.887	0.885	0.891	0.883	0.877
	DeSCI	0.954	0.834	0.960	0.928	0.952	0.927	0.922	0.926	0.935	0.926	0.926
	TV-FFDNet	0.989	0.877	0.982	0.955	0.974	0.952	0.955	0.964	0.964	0.948	0.956
	PnP-DIP-HSI	0.956	0.801	0.968	0.931	0.974	0.924	0.957	0.949	0.936	0.951	0.934
	NLTT-TV	0.985	0.860	0.982	0.949	0.984	0.954	0.957	0.965	0.972	0.977	0.959
	DPLR	0.967	0.928	0.974	0.898	0.979	0.958	0.977	0.948	0.965	0.968	0.956
	LRSDN	**0.992**	**0.966**	**0.988**	**0.964**	**0.989**	**0.983**	**0.985**	**0.983**	**0.986**	**0.990**	**0.982**
ERGAS	GAP-TV	166.01	462.87	132.940	252.69	194.32	263.73	271.43	294.05	226.15	232.42	249.66
	DeSCI	176.70	404.44	57.076	300.40	127.41	189.91	268.98	237.33	190.31	195.77	214.83
	TV-FFDNet	42.74	295.01	37.663	110.25	76.10	132.18	114.86	112.89	121.92	134.14	117.77
	PnP-DIP-HSI	121.01	480.07	83.443	243.80	98.80	153.28	135.07	159.05	185.85	128.74	178.91
	NLTT-TV	52.28	384.89	47.651	231.04	66.43	131.35	152.36	131.17	104.15	100.96	140.23
	DPLR	49.50	194.60	40.90	176.59	60.64	80.99	65.32	142.91	100.40	70.20	98.20
	LRSDN	**33.91**	**132.30**	**28.325**	**105.30**	**47.30**	**63.34**	**56.71**	**67.37**	**66.24**	**45.58**	**64.64**
SA	GAP-TV	15.14	27.85	7.778	21.96	17.30	19.56	20.54	20.62	23.79	17.62	19.22
	DeSCI	6.39	19.09	2.788	9.40	13.23	9.83	11.88	10.31	11.16	12.15	10.62
	TV-FFDNet	**4.23**	17.06	1.982	8.44	11.40	7.95	9.78	7.24	9.92	11.44	8.94
	PnP-DIP-HSI	15.09	28.60	5.736	11.47	17.10	18.32	19.09	16.35	15.50	20.75	16.80
	NLTT-TV	6.43	25.48	2.982	14.11	10.26	14.43	16.56	16.16	19.12	15.93	14.15
	DPLR	7.50	16.48	2.62	15.12	14.38	10.70	13.68	13.63	15.98	16.98	12.71
	LRSDN	4.91	**10.84**	**1.880**	**6.67**	**9.00**	**5.70**	**8.13**	**6.74**	**8.39**	**8.68**	**7.09**

$$F_{sam}^{k} = f_{sam}^{k}(F_{drm}^{k}, F_{dc}^{k}), \tag{23}$$

where $f_{sam}^{k}(\cdot)$ denotes the SAM operation in k-layer.

However, the spatial information of generator features is inevitably lost by bilinear interpolation. DRM is different from SAM in that the deep spatial feature is weighted by the high-level decoding semantic features of the guidance coefficient. As the low-level encoded features of the guidance coefficient contain abundant spatial information, the DRM is designed to extract the spatial information used to compensate for the spatial detail of deep features. Therefore, we transfuse the obtained spatial information into the generator features, and the operation of DRM can be expressed as:

$$F_{drm}^{k} = f_{drm}^{k}(F_{sam}^{k-1}, F_{en}^{K-k}), \tag{24}$$

where $f_{drm}^{k}(\cdot)$ denotes the DRM operation in k-layer.

In summary, the features of the deep decoder ASD are weighted in SAM and compensated in DRM, which leads to the DCP that can more explicitly exploit the semantic features and spatial details of the guidance coefficient, resulting in high-precision reconstruction. The loss function of DGSAN is defined as the minimization problem (12), which has three advantages: 1) denoising $(\mathcal{X}-\mathcal{B})\times_3\mathbf{E}^T$, 2) minimizing the loss between the observed measurement and the low-rank approximation of original HSI, and 3) algorithm interpretability.

V. EXPERIMENT

We validate the performance of our LRSDN on two representative CASSI systems DD-CASSI and SD-CASSI and compare it with several state-of-the-art (SOTA) CASSI reconstruction algorithms. We implement DGSAN in the Pytorch framework and minimize the loss function using the ADAM optimizer ($\beta 1 = 0.9$ and $\beta 2 = 0.999$). For the DD-CASSI and SD-CASSI systems, we set the learning rate as 0.1 and 0.002, respectively. All experiments are run on a platform with Inter i9-12900K and NVIDIA GeForce RTX 3090.

A. Comparison Methods and Evaluation Metrics

We compare our proposed method with fourteen SOTA HSI reconstruction methods, including three model-based methods GAP-TV [19], DeSCI [24], and NLTT-TV [25], six E2E methods λ-Net [26], TSA-Net [27], HDNet [30], MST-L [46], MST++ [47], and CST-L [48], two deep unfolding methods DGSMP [29] and DAUHST [50], one single-sample generative models DPLR [54], and two model-based optimization with deep priors methods TV-FFDNet [32] and PnP-DIP-HSI [31].

Specifically, we compare the proposed LRSDN with GAP-TV, DeSCI, TV-FFDNet, PnP-DIP-HSI, NLTT-TV, and DPLR in the DD-CASSI system. In the SD-CASSI system, our method is compared with GAP-TV, DeSCI, λ-Net, PnP-DIP-HSI, TSA-Net, DGSMP, HDNet, MST-L, MST++, CST-L and DAUHST. The implementation codes for all the compared methods are available from the authors' websites, and the hyper-parameters for different experiments are set following the authors' code or suggestions from the reference papers to achieve the best possible results. Regarding the parameter selection for our method, we set the regularization parameter $\lambda = 0.1$ and penalty parameter $\mu = 0.03$ in the DD-CASSI experiments, while $\lambda = 1$ and $\mu = 0.003$ are used in the SD-CASSI experiments. It is worth noting that the codes and results of almost all deep learning-based methods can be downloaded from the MST homepage.[1]

Five image quality metrics are employed to quantitatively evaluate the reconstruction results, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM), feature similarity (FSIM), relative dimensionless global error in synthesis (ERGAS), and spectral angle (SA). PSNR, SSIM, and FSIM are used to evaluate the reconstruction quality of spatial features, while ERGAS and SA evaluate the preservation ability of spectral signatures. The better reconstruction results are achieved by the larger PSNR, SSIM, and FSIM values and the smaller ERGAS and SA values.

B. Simulated Experiments on DD-CASSI System

The imaging system of DD-CASSI is presented in the first row of Fig. 2. For a comprehensive evaluation, two simulated datasets are chosen as benchmark hyperspectral datasets, including CAVE dataset[2] and Harvard dataset.[3] The CAVE dataset contains 32 indoor HSIs with the size $512 \times 512 \times 31$. To avoid randomness and the influence of specific data, ten representative scenes in the CAVE dataset are selected for the experiments. The Harvard dataset has 50 HSIs of indoor and outdoor scenes under daylight illumination and an additional 25 HSIs under artificial and mixed illumination. The spatial resolution and spectral number of each HSI in Harvard datasets are 1040×1392 and 31, respectively. We resize the spatial resolution of the Harvard dataset as 512×512 and randomly select five images for testing. For simulating the 2D observed snapshot image, the real coded mask used here is available from DeSCI[4] [24]. The coded aperture is a random code, and the code's transmittance is 50%.

Tables I and II list the quantitative values of different methods on the CAVE and Harvard datasets, respectively. The best results for each index are marked in bold. Specifically, GAP-TV and DeSCI design the hand-crafted prior on the HSI, thus it obtains relatively unsatisfactory results compared with other methods. NLTT-TV method integrates the nonlocal similarity and tensor low-rank prior, which further improves the reconstruction quality than GAP-TV and DeSCI. Additionally,

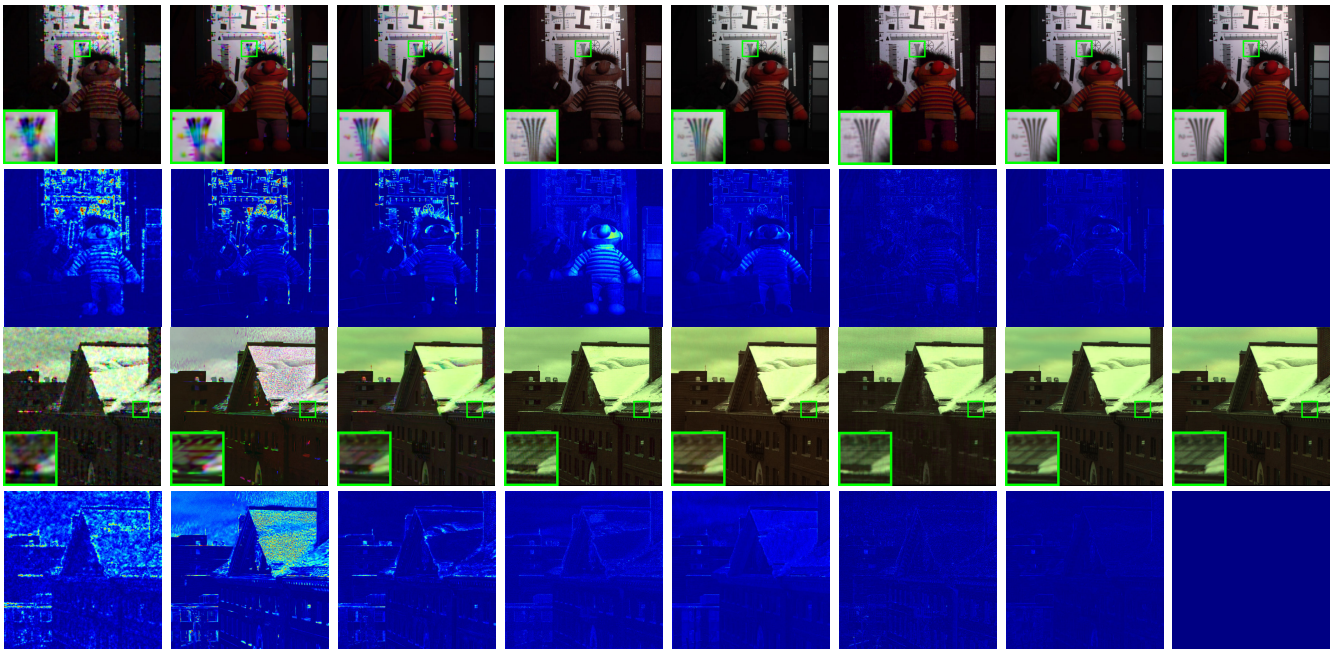
TABLE II
QUANTITATIVE RESULTS OF COMPARISON METHODS ON THE HARVARD DATASET

Index	Method	img 1	img b8	img c4	img d3	img h0	Average
PSNR	GAP-TV	23.35	21.20	26.25	27.30	22.64	24.15
	DeSCI	19.30	21.27	27.89	28.66	23.42	24.11
	TV-FFDNet	31.50	24.95	32.34	33.01	30.50	30.46
	PnP-DIP-HSI	33.25	29.38	35.14	35.08	30.16	32.60
	NLTT-TV	34.92	28.11	36.16	37.42	31.30	33.58
	DPLR	35.41	29.03	35.74	38.60	32.91	34.34
	LRSDN	**39.67**	**31.77**	**40.52**	**42.15**	**35.73**	**37.97**
SSIM	GAP-TV	0.696	0.583	0.763	0.802	0.714	0.712
	DeSCI	0.708	0.612	0.837	0.893	0.794	0.769
	TV-FFDNet	0.890	0.754	0.903	0.940	0.880	0.873
	PnP-DIP-HSI	0.835	0.854	0.925	0.911	0.872	0.879
	NLTT-TV	0.932	0.857	0.940	0.950	0.913	0.918
	DPLR	0.870	0.739	0.872	0.929	0.859	0.854
	LRSDN	**0.952**	**0.892**	**0.970**	**0.979**	**0.956**	**0.950**
FSIM	GAP-TV	0.782	0.774	0.850	0.879	0.813	0.819
	DeSCI	0.810	0.794	0.872	0.918	0.877	0.854
	TV-FFDNet	0.946	0.861	0.934	0.958	0.919	0.924
	PnP-DIP-HSI	0.966	**0.955**	0.973	0.975	0.958	0.965
	NLTT-TV	0.977	0.932	0.973	0.983	0.955	0.964
	DPLR	0.962	0.928	0.966	0.978	0.958	0.959
	LRSDN	**0.990**	0.954	**0.989**	**0.992**	**0.977**	**0.980**
ERGAS	GAP-TV	382.76	521.57	373.78	347.01	437.59	412.54
	DeSCI	599.88	439.67	226.84	264.77	370.85	380.40
	TV-FFDNet	104.98	290.59	151.55	149.02	148.42	168.91
	PnP-DIP-HSI	84.88	165.17	96.53	113.02	119.25	115.77
	NLTT-TV	87.65	281.30	120.13	112.97	141.57	148.72
	DPLR	59.72	156.39	97.64	72.46	100.93	97.43
	LRSDN	**38.75**	**138.63**	**52.82**	**47.89**	**69.55**	**69.53**
SA	GAP-TV	17.11	23.28	17.38	19.40	15.58	18.55
	DeSCI	15.52	22.02	11.87	11.15	11.66	14.45
	TV-FFDNet	7.08	13.26	7.97	7.85	5.68	8.37
	PnP-DIP-HSI	8.01	11.31	7.48	11.55	7.90	9.25
	NLTT-TV	5.58	9.15	6.70	6.82	5.62	6.77
	DPLR	4.15	7.57	6.92	5.92	5.39	5.99
	LRSDN	**3.72**	**6.91**	**4.25**	**4.44**	**3.35**	**4.53**

TV-FFDNet, PnP-DIP-HSI, and DPLR explore deep priors, enabling them to achieve better results than hand-crafted prior-based methods GAP-TV and DeSCI. This demonstrates the superior performance of deep neural networks in capturing spatial features. DPLR outperforms TV-FFDNet and PnP-DIP-HSI on both datasets, thanks to its excellent generalization performance across different datasets, which is attributed to the combination of low-rank and deep priors. In contrast, the proposed LRSDN exhibits significant superiority over other comparison methods in terms of all metrics, highlighting the effectiveness of modeling the global spectral correlation of HSI through low-rank subspace representation and exploring the DCP of representation coefficient using a self-supervised deep neural network.

To visually compare the performance of different methods for HSI reconstruction, we choose a representative CAVE-Toy scene and Harvard-Img1 to present the reconstruction results in Fig. 4, respectively. The first row and third row present the reconstructed false-color image (composed of bands 31, 11, and 6) obtained by different methods, and the second row and fourth row illustrate corresponding residual images, which is achieved by averaging the absolute error between the reconstruction results and ground truth. To enable a comprehensive comparison of the reconstructed results, we have

[1] https://github.com/caiyuanhao1998/MST
[2] http://www1.cs.columbia.edu/CAVE/databases/multispectral
[3] http://vision.seas.harvard.edu/hyperspec/download.html
[4] https://github.com/liuyang12/DeSCI



(a) GAP-TV (b) DeSCI (c) TV-FFDNet (d) PnP-DIP-HSI (e) NLTT-TV (f) DPLR (g) LRSDN (h) Ground Truth

Fig. 4. Reconstructed results of different methods on CAVE-Toy image and Harvard-Img1 image. The first and third rows show the false color images which are composed of bands (R: 31, G: 11, and B: 6). The second and fourth rows show the corresponding absolute error map between the original and reconstructed images.

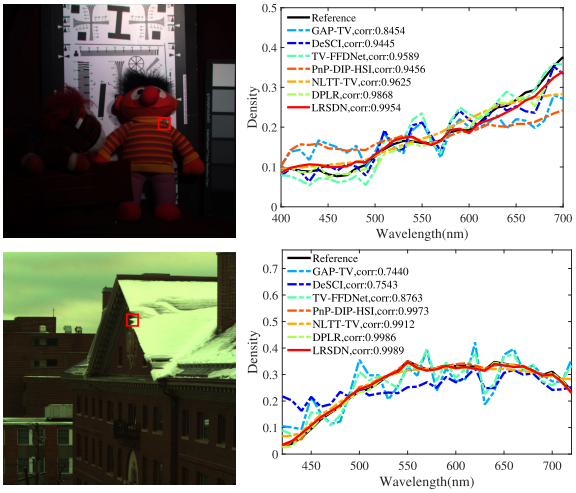


Fig. 5. The reconstructed spectral curves of different methods on CAVE-Toy image (top) and Harvard-Img1 image (down).

included zoomed-in views of specific image areas outlined by rectangles. As we can see from the result, all methods can reconstruct the HSI where the spatial information is visible. However, GAP-TV, DeSCI, and TV-FFDNet destroy the image detail as shown in the enlarged box. Since PnP-DIP-HSI ignores the global spectral correlation of HSI, the reconstructed spectral information is distorted. The results of NLTT-TV and DPLR are better than other methods because they both leverage the low-rank property of HSI. However, NLTT-TV still exhibits artifacts, whereas DPLR eliminates these artifacts, thanks to the effectiveness of deep priors. In contrast, the proposed method can effectively reconstruct the spatial structures and spectral signatures of HSI, thus demonstrating the capability of LRSDN to utilize the spectral low-rank characteristics of HSI, and verifying the effectiveness of self-supervised learning. From residual map results, the reconstructed HSI produced by our method has fewer errors than that of other comparison methods, which further validates the superior performance of the proposed method for HSI reconstruction.

To illustrate the spectral preservation ability of different methods, we also select a patch to present the reconstructed spectral curves in Fig. 5. Moreover, we give the correlation coefficient of the reconstructed spectral and the ground truth in the legend. By comparing the degree of similarity with the real curve and the correlation coefficient, we can observe that our LRSDN is closer to the reference curve and obtains a higher correlation value, indicating the significant advantage of our proposed method to preserve the spectral signature.

C. Simulated Experiments on SD-CASSI System

To further illustrate the generalization ability of the proposed method to different CASSI systems, the SD-CASSI imaging system is employed to test. The benchmark data that contains 10 scenes from the KAIST dataset [44] are adopted for testing. The detailed information of coded apertures can be referred to TSA-Net[5] [27]. For a fair comparison, the experimental settings including the ground truth and real mask keep the same as that of previous works [27], [30], [31], [46], [47], [48], [50]. Consistent with the comparison methods, the PSNR and SSIM are employed to evaluate the HSI reconstruction performance.

[5]https://github.com/mengziyi64/TSA-Net

TABLE III
QUANTITATIVE RESULTS OF COMPARISON METHODS ON THE KAIST DATASET

Method	Scene1	Scene2	Scene3	Scene4	Scene5	Scene6	Scene7	Scene8	Scene9	Scene10	Average
GAP-TV	26.82 0.754	22.89 0.610	26.31 0.802	30.65 0.852	23.64 0.703	21.85 0.663	23.76 0.688	21.98 0.655	22.63 0.682	23.10 0.584	24.36 0.669
DeSCI	27.13 0.748	23.04 0.620	26.62 0.818	34.96 0.897	23.94 0.706	22.38 0.683	24.45 0.743	22.03 0.673	24.56 0.732	23.59 0.587	25.27 0.721
λ-Net	30.10 0.849	28.46 0.805	27.73 0.870	37.01 0.934	26.19 0.817	28.64 0.853	26.47 0.806	26.09 0.831	27.50 0.826	27.13 0.816	28.53 0.841
PnP-DIP-HSI	32.68 0.890	27.26 0.833	31.30 0.914	40.54 0.962	29.79 0.900	30.39 0.877	28.18 0.913	29.44 0.874	34.51 0.927	28.51 0.851	31.26 0.894
TSA-Net	32.03 0.892	31.00 0.858	32.25 0.915	39.19 0.953	29.39 0.884	31.44 0.908	30.32 0.878	29.35 0.888	30.01 0.890	29.59 0.874	31.46 0.894
DGSMP	33.26 0.915	32.09 0.898	33.06 0.925	40.54 0.964	28.86 0.882	33.08 0.937	30.74 0.886	31.55 0.923	31.66 0.911	31.44 0.925	32.63 0.917
HDNet	34.95 0.948	32.52 0.953	34.52 0.957	43.00 0.981	32.49 0.957	35.96 0.965	29.18 0.937	34.00 0.961	34.56 0.958	32.22 0.950	34.34 0.957
MST-L	35.40 0.941	35.87 0.944	36.51 0.953	42.27 0.973	32.77 0.947	34.80 0.955	33.66 0.925	32.67 0.948	35.39 0.949	32.50 0.941	35.18 0.948
MST++	35.80 0.943	36.23 0.947	37.34 0.957	42.63 0.973	33.38 0.952	35.38 0.957	34.35 0.934	33.71 0.953	36.67 0.953	33.38 0.945	35.99 0.951
CST-L	35.96 0.949	36.84 0.955	38.16 0.962	42.44 0.975	33.25 0.955	35.72 0.963	34.86 0.944	34.34 0.961	36.51 0.957	33.09 0.945	36.12 0.957
DAUHST-9stg	**37.25** **0.958**	**39.02** **0.967**	**41.05** **0.971**	**46.15** **0.983**	**35.80** **0.969**	**37.08** **0.970**	37.57 0.963	**35.10** **0.966**	**40.02** **0.970**	**34.59** **0.956**	**38.36** **0.967**
LRSDN	35.44 0.923	34.89 0.909	38.90 0.961	45.29 0.985	34.71 0.949	33.18 0.930	**37.76** **0.964**	30.57 0.901	39.49 0.963	30.62 0.889	36.08 0.938

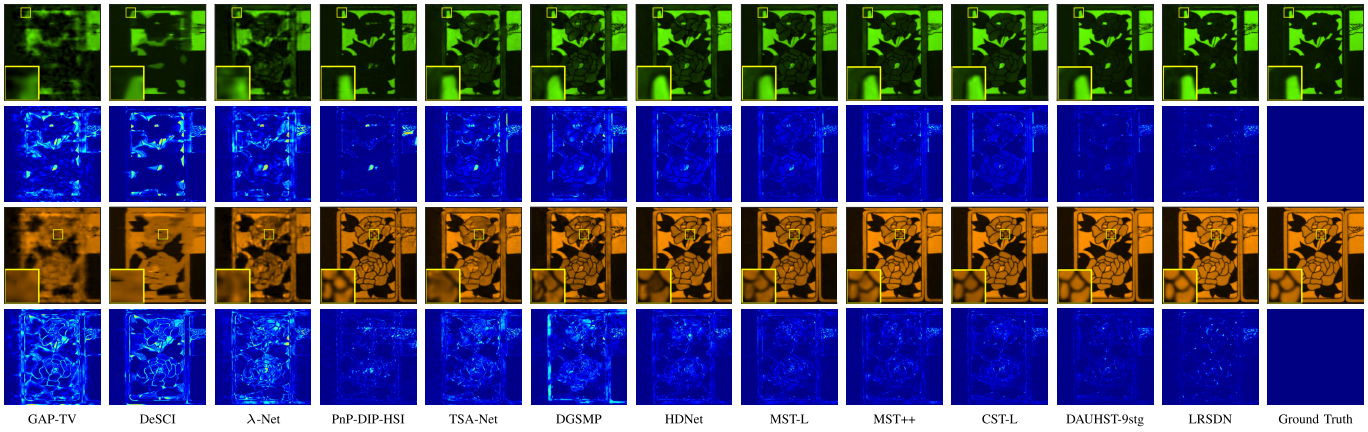


Fig. 6. Reconstructed results of different methods on KAIST-Scene7 image, including 2 (out of 28) spectral channels and the corresponding error maps.

Table III lists the reconstruction results of different methods on the KAIST dataset. Since the experimental setting is the same as the previous works, the results of comparison methods are derived from [46]. It can be seen that the supervised learning-based methods usually outperform the model-based optimization with hand-crafted or deep priors methods. Our method performs a significant improvement over model-based GAP-TV and DeSCI, indicating the advantage of the proposed deep prior. PnP-DIP-HSI method directly employs the deep priors to HSI itself, while the proposed method explores the global spectra correlation of HSI and designs a deep prior for representation coefficient, thus our method is superior to PnP-DIP-HSI. Furthermore, our method achieves competitive results with recently proposed supervised learning methods. Although DAUHST-9stg and CST-L respectively outperform the proposed LRSDM by 2.28 and 0.04 dB in PSNR, our method still outperforms other supervised learning methods λ-Net, TSA-Net, and DGSMP. Moreover, compared with supervised learning methods, our LRSDN method has the virtue of strong representation ability, superior generalization ability, and high reconstruction flexibility.

The visual reconstruction results and the corresponding error maps for different methods on scene 7 with 2 channels are presented in Fig. 6. As can be seen from the reconstructed results and the zoom-in patches of the selected regions, model-based optimization with hand-crafted or deep priors methods

TABLE IV
QUANTITATIVE RESULTS OF COMPARISON METHODS ON THE BIRD DATASET

Dataset	Index	GAP-TV	DeSCI	TV-FFDNet	PnP-DIP-HSI	NLTT-TV	DPLR	LRSDN
Bird	PSNR	24.59	35.72	39.74	34.00	32.59	37.26	**42.17**
	SSIM	0.736	0.938	0.958	0.875	0.918	0.912	**0.967**
	FSIM	0.823	0.957	0.972	0.960	0.965	0.974	**0.986**
	ERGAS	417.74	90.19	57.56	99.51	132.76	76.24	**38.00**
	SA	12.78	3.65	2.66	5.88	5.59	3.07	**2.37**

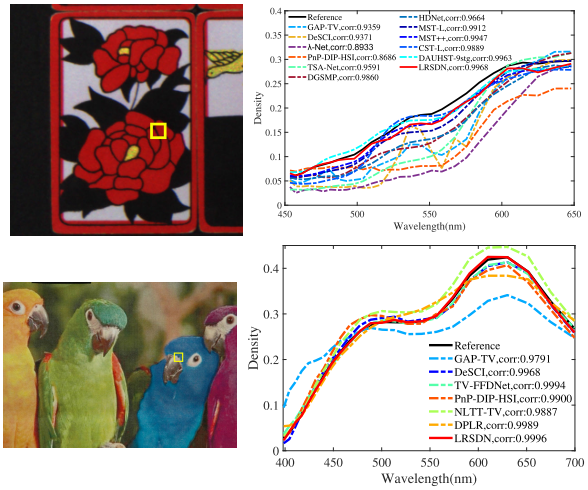


Fig. 7. The reconstructed spectral curves of different methods on KAIST-Scene7 (top) and real Bird image (down).

and most learning-based methods blur the details and produce artifacts to some extent. In contrast, DAUHST-9stg and our LRSDN are more capable of reconstructing perceptually pleasing images and preserving the spatial details and sharp edges. From the results of error maps, DAUHST-9stg and our LRSDN also perform the smaller error result, indicating that we can preserve more image information. This is mainly because our LRSDN organically couples the advantages of model-driven low-rank prior and data-driven deep prior.

In addition, we plot the spectral signature curves of the selected region and give the correlation values with ground truth in Fig. 7. It can be seen that the signature curve of LRSDN is closest to ground truth, and our LRSDN achieves the highest correlation coefficient, which further illustrates the superior performance of LRSDN for spectral signature preservation. In summary, the proposed method can achieve competitive results with supervised learning methods without any training data.

D. Experiments With Real Data

We further validate the performance of the proposed LRSDN on two real datasets: the Bird dataset from DD-CASSI system [13] and five real scenes from SD-CASSI system [27]. The Bird dataset consists of 24 spectral bands with the spatial resolution 1021×703. The real CASSI measurement is shown in Fig. 8 (a). We list the quantitative results of all comparison methods for the Bird dataset in Table IV. It can be seen clearly that our LRSDN provides optimal reconstructed indices. The reconstruction results of all comparison methods are shown in Figs. 8 (b)-(h). From the results, we can observe that LRSDN outperforms the comparison methods by reconstructing the most spatial details as shown in the enlarged box. Fig. 7 shows the spectral signature profiles and corresponding spectral correlation coefficients of a selected region by different methods. Compared with other methods, our method can obtain spectral signature profiles closer to the reference and obtain higher spectral correlations.

We further conduct the experiments on five real scenes from the real SD-CASSI system [27]. Each measurement has a spatial size of 660×714, and the HSI to be recovered covers a wavelength range from 450nm to 650nm with 28 bands and a spatial size of 660×660. Due to limited space, we have shown a visual comparison of two scenes. Fig. 9 presents the visual comparisons of Scene 1 and Scene 5 between the proposed LRSDN and eight supervised learning methods. It can be observed that λ-Net, TSA-Net, DGSMP, HDNet, and MST++ cannot reconstruct local detail or appear edge blurring in some regions. The proposed LRSDN achieves visual reconstruction results that are competitive with the supervised learning-based SOTA methods MSL-L, CST-L, and DAUHST-9stg. This is attributed to the ability of our method to capture complex nonlinear information in HSI using deep priors in the low-rank subspace framework, further confirming the generalization capability of the proposed method.

E. Discussion

In this section, we perform the ablation study to illustrate the effectiveness of two priors (low-rank prior and deep prior) promoting each other and the performance of different modules in DGSAN. To make the ablation study convincing, we test ten scenes from the CAVE dataset and ten scenes from the KAIST dataset in the DD-CASSI system and SD-CASSI system, respectively. Moreover, we provide the sensitivity analysis of the regularization parameter λ involved in the proposed model (10). Finally, we discussed the numerical convergence of the proposed method.

1) Ablation Study on Low-Rank Prior and Deep Prior: To verify the validity of these two priors in the proposed LRSDN, we conducted an ablation study by disabling each prior individually. The low-rank prior and deep prior (DGSAN) are disabled, which are referred to as w/o LR and w/o DGSAN, respectively. The complete disabling of DGSAN will lead to the instability of the model, thus the result of w/o DGSAN is achieved by replacing DGSAN with DIP. Table V lists the reconstructed PSNR and SSIM of two different CASSI systems by averaging ten scenes from the CAVE dataset and KAIST dataset, respectively. It is clear that LRSDN

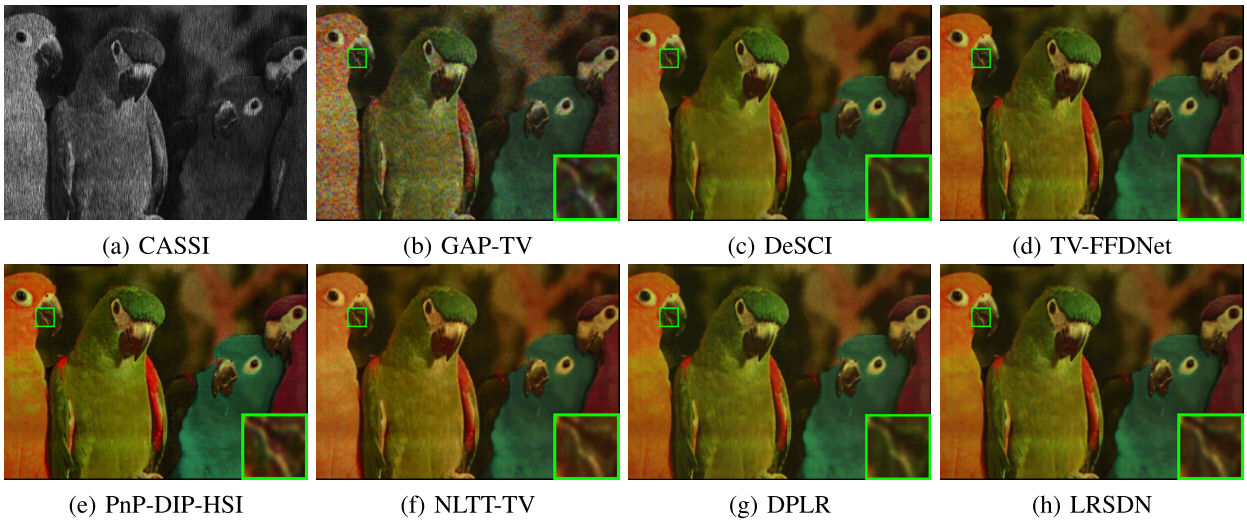


Fig. 8. Reconstructed results of different methods on real Bird image. The false color image is composed of bands (R: 24, G: 12, and B: 6).

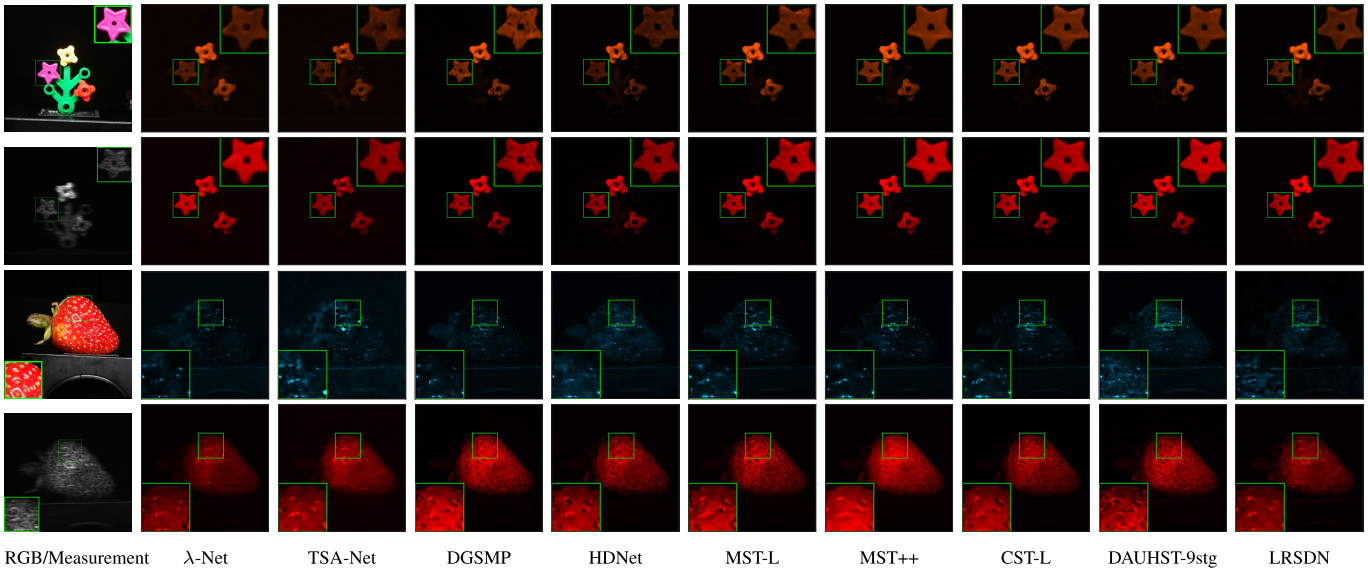


Fig. 9. Real HSI reconstruction results of LRSDN and eight supervised learning-based methods on Scene 1 (first two rows) and Scene 5 (Last two rows) with 2 (out of 28) spectral channels.

TABLE V

QUANTITATIVE RESULTS OF ABLATION STUDY ON LOW-RANK PRIOR AND DEEP PRIOR IN LRSDN

System	Index	w/o LR	w/o DGSAN	LRSDN
DD-CASSI	PSNR	37.44	35.22	**38.84**
	SSIM	0.962	0.948	**0.969**
SD-CASSI	PSNR	35.14	32.32	**36.08**
	SSIM	0.925	0.877	**0.938**

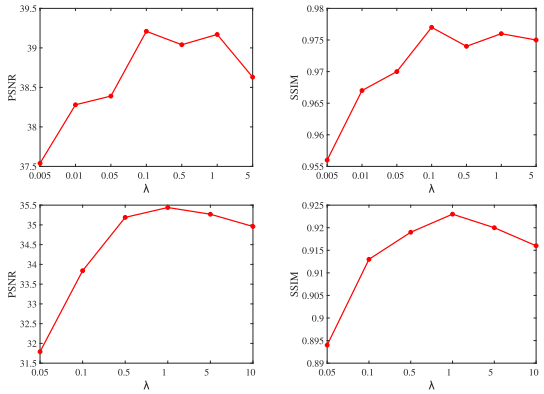
TABLE VI

QUANTITATIVE RESULTS OF ABLATION STUDY ON DIFFERENT MODULES IN DGSAN

Method		w/o DFG	w/o SAM	w/o DRM	DGSAN
DD-CASSI	PSNR	36.12	36.49	37.60	**38.84**
	SSIM	0.942	0.945	0.955	**0.969**
SD-CASSI	PSNR	30.89	33.72	34.66	**36.08**
	SSIM	0.840	0.901	0.921	**0.938**

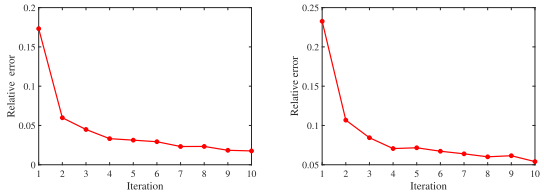
outperforms its variants, and the two components contribute significantly to the success of the proposed LRSDN method. Especially, LRSDN outperforms the w/o LR by 1.4 dB and 0.94 dB in average PSNR, indicating that low-rank prior promotes the deep prior. Moreover, the proposed DGSAN surpasses the recent sophisticated DIP by 3.62 dB and 3.76 dB in average PSNR on two different systems, demonstrating the effectiveness of DGSAN over DIP.

2) Ablation Study on Different Modules in DGSAN: The proposed DGSAN mainly includes an encoding-decoding DFG module and a deep decoding ASD module, in which the ASD module also contains SAM and DRM modules. We further conduct an ablation study to validate the performance of different modules in DGSAN. The results are summarized in Table VI. It is clear that DGSAN outperforms its variants, and all of the three modules contribute significantly to the success of the proposed DGSAN.



(a) (b)

Fig. 10. Sensitivity analysis of regularization parameter λ on two different systems. Top row: CAVE-Toy scene on DD-CASSI system. Bottom row: KAIST-Scene1 on SD-CASSI system.



(a) CAVE-Toy (b) KAIST-Scene1

Fig. 11. Relative change values of reconstructed HSI versus the iteration number of the proposed algorithm.

3) Parameter Analysis of λ: The proposed model (10) involves a regularization parameter λ, which is used to balance the two fidelity terms. The CAVE-Toy data on the DD-CASSI system and KAIST-Scene1 on the SD-CASSI system are employed for the parameter analysis. Fig. 10 presents the PSNR and SSIM values of the proposed method with different λ values. It can be seen that with the increase of λ value, the performances of PSNR and SSIM increase gradually, indicating the effectiveness of two fidelity terms. As we continue to increase the value of λ, the performance tends to decrease. Thus, we fix $\lambda = 0.1$ and $\lambda = 1$ for DD-CASSI and SD-CASSI systems in all experiments, respectively.

4) Numerical Convergence: Since a deep network is embedded in the proposed model (10), it is difficult to present a theoretical convergence guarantee via Algorithm 1. To demonstrate the convergence of the proposed method, we present the numerical result. Fig. 11 presents the relative change values of reconstructed HSI versus the iteration number of the Algorithm 1. It can be seen that, as the number of iterations increases, the relative changes converge to a stable value, which indicates the convergence of the proposed method.

VI. Conclusion

In this paper, we propose a hyperspectral compressive snapshot reconstruction method by combining model-driven-based low-rank prior and data-driven-based deep prior, which explores the global spectral correlation of the HSI and deep feature of the corresponding spatial representation coefficient, respectively. Especially, a self-supervised deep network DGSAN is proposed to learn the DCP directly from the compressed measurement and guidance coefficient, which can better exploit semantic features and spatial details of the guidance coefficient to reconstruct the spatial representation coefficient. Therefore, we integrate this self-supervised deep network-based DCP into the low-rank subspace representation framework of HSI and solve it by the ADMM algorithm. The proposed method has been tested on two representative coded hyperspectral imaging systems, including SD-CASSI and DD-CASSI. Experimental results demonstrate that our method outperforms current model-based and Model-based optimization with deep priors state-of-the-art methods on several benchmark datasets. Moreover, we have achieved competitive results when compared with supervised deep learning-based approaches, which need sufficient training data. In the future, we believe that our proposed self-supervised framework can extensively be used for other advanced CASSI-type architecture, such as colored CASSI system [67].

References

[1] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1552–1560, Dec. 2003.

[2] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[3] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, Jan. 2014, Art. no. 010901.

[4] Z. Meng, M. Qiao, J. Ma, Z. Yu, K. Xu, and X. Yuan, "Snapshot multispectral endomicroscopy," *Opt. Lett.*, vol. 45, no. 14, p. 3897, 2020.

[5] R. W. Basedow, D. C. Carmer, and M. E. Anderson, "HYDICE system: Implementation and performance," *Proc. SPIE*, vol. 2480, pp. 258–267, Jun. 1995.

[6] Y. Y. Schechner and S. K. Nayar, "Generalized mosaicing: Wide field of view multispectral imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 10, pp. 1334–1348, Oct. 2002.

[7] X. Cao et al., "Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world," *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 95–108, Sep. 2016.

[8] P. Llull et al., "Coded aperture compressive temporal imaging," *Opt. Exp.*, vol. 21, no. 9, p. 10526, 2013.

[9] A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," *Opt. Exp.*, vol. 17, no. 8, p. 6368, 2009.

[10] X. Yuan, D. J. Brady, and A. K. Katsaggelos, "Snapshot compressive imaging: Theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 65–88, Mar. 2021.

[11] H. Arguello, H. Rueda, Y. Wu, D. W. Prather, and G. R. Arce, "Higher-order computational model for coded aperture spectral imaging," *Appl. Opt.*, vol. 52, no. 10, p. D12, 2013.

[12] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, p. B44, 2008.

[13] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Exp.*, vol. 15, no. 21, p. 14013, 2007.

[14] L. Huang, R. Luo, X. Liu, and X. Hao, "Spectral imaging with deep learning," *Light, Sci. Appl.*, vol. 11, no. 1, Mar. 2022.

[15] J. Zhang, R. Su, Q. Fu, W. Ren, F. Heide, and Y. Nie, "A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging," *Sci. Rep.*, vol. 12, no. 1, p. 11905, Jul. 2022.

[16] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 105–115, Jan. 2014.

[17] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, "Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 116–126, Jan. 2014.

[18] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 54, no. 4, p. 848, 2015.

[19] X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2539–2543.

[20] L. Wang, Z. Xiong, G. Shi, F. Wu, and W. Zeng, "Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 10, pp. 2104–2111, Oct. 2017.

[21] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial–spectral encoded compressive hyperspectral imaging," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 1–11, Nov. 2014.

[22] S. Zhang, L. Wang, Y. Fu, X. Zhong, and H. Huang, "Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10182–10191.

[23] W. He, N. Yokoya, and X. Yuan, "Fast hyperspectral image recovery of dual-camera compressive hyperspectral imaging via noniterative subspace-based fusion," *IEEE Trans. Image Process.*, vol. 30, pp. 7170–7183, 2021.

[24] Y. Liu, X. Yuan, J. Suo, D. J. Brady, and Q. Dai, "Rank minimization for snapshot compressive imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 2990–3006, Dec. 2019.

[25] Y. Wang, Y. Han, K. Wang, and X.-L. Zhao, "Total variation regularized nonlocal low-rank tensor train for spectral compressive imaging," *Signal Process.*, vol. 195, Jun. 2022, Art. no. 108464.

[26] X. Miao, X. Yuan, Y. Pu, and V. Athitsos, "Lambda-Net: Reconstruct hyperspectral images from a snapshot measurement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4058–4068.

[27] Z. Meng, J. Ma, and X. Yuan, "End-to-end low cost compressive spectral imaging with spatial–spectral self-attention," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 187–204.

[28] Z. Meng, S. Jalali, and X. Yuan, "GAP-Net for snapshot compressive imaging," 2020, *arXiv:2012.08364*.

[29] T. Huang, W. Dong, X. Yuan, J. Wu, and G. Shi, "Deep Gaussian scale mixture prior for spectral compressive imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16211–16220.

[30] X. Hu et al., "HDNet: High-resolution dual-domain learning for spectral compressive imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17521–17530.

[31] Z. Meng, Z. Yu, K. Xu, and X. Yuan, "Self-supervised neural networks for spectral snapshot compressive imaging," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2602–2611.

[32] H. Qiu, Y. Wang, and D. Meng, "Effective snapshot compressive-spectral imaging via deep denoising and total variation priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9123–9132.

[33] X. Yuan, Y. Liu, J. Suo, and Q. Dai, "Plug-and-play algorithms for large-scale snapshot compressive imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1444–1454.

[34] W. He et al., "Non-local meets global: An integrated paradigm for hyperspectral image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2089–2107, Apr. 2022.

[35] X. Chen, W. He, X.-L. Zhao, T.-Z. Huang, J. Zeng, and H. Lin, "Exploring nonlocal group sparsity under transform learning for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2022, Art. no. 5537518.

[36] Y. Chen, T.-Z. Huang, W. He, X.-L. Zhao, H. Zhang, and J. Zeng, "Hyperspectral image denoising using factor group sparsity-regularized nonconvex low-rank approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022, Art. no. 5515916.

[37] L. Zhang, W. Wei, Y. Zhang, C. Shen, A. van den Hengel, and Q. Shi, "Dictionary learning for promoting structured sparsity in hyperspectral compressive sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7223–7235, Dec. 2016.

[38] J. Tan, Y. Ma, H. Rueda, D. Baron, and G. R. Arce, "Compressive hyperspectral imaging via approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 2, pp. 389–401, Mar. 2016.

[39] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Appl. Opt.*, vol. 49, no. 36, p. 6824, 2010.

[40] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3556–3570, Aug. 2020.

[41] Y. Fu, Y. Zheng, I. Sato, and Y. Sato, "Exploiting spectral–spatial correlation for coded hyperspectral image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3727–3736.

[42] Y. Fu, T. Zhang, L. Wang, and H. Huang, "Coded hyperspectral image reconstruction using deep external and internal learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3404–3420, Jul. 2022.

[43] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, and F. Wu, "HSCNN: CNN-based hyperspectral image recovery from spectrally undersampled projections," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 518–525.

[44] I. Choi, D. S. Jeon, G. Nam, D. Gutierrez, and M. H. Kim, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–13, Dec. 2017.

[45] L. Wang, C. Sun, M. Zhang, Y. Fu, and H. Huang, "DNU: Deep non-local unrolling for computational spectral imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1658–1668.

[46] Y. Cai et al., "Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17481–17490.

[47] Y. Cai et al., "MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 744–754.

[48] Y. Cai et al., "Coarse-to-fine sparse transformer for hyperspectral image reconstruction," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 686–704.

[49] S. Zhang, L. Wang, L. Zhang, and H. Huang, "Learning tensor low-rank prior for hyperspectral image reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12001–12010.

[50] Y. Cai et al., "Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging," in *Proc. Adv. Neural. Inf. Process. Syst.*, vol. 35, 2022, pp. 37749–37761.

[51] B. Monroy, J. Bacca, and H. Arguello, "JR2Net: A joint non-linear representation and recovery network for compressive spectral imaging," *Appl. Opt.*, vol. 61, no. 26, p. 7757, 2022.

[52] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[53] Z. Sun, Y. Yang, Q. Liu, and M. Kankanhalli, "Unsupervised spatial-spectral network learning for hyperspectral compressive snapshot reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022, Art. no. 5514314.

[54] J. Bacca, Y. Fonseca, and H. Arguello, "Compressive spectral image reconstruction using deep prior and low-rank tensor representation," *Appl. Opt.*, vol. 60, no. 14, p. 4197, 2021.

[55] T. Gelvez, J. Bacca, and H. Arguello, "Interpretable deep image prior method inspired in linear mixture model for compressed spectral image recovery," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 1934–1938.

[56] T. Gelvez-Barrera, J. Bacca, and H. Arguello, "Mixture-Net: Low-rank deep image prior inspired by mixture models for spectral image recovery," *Signal Process.*, vol. 216, Mar. 2024, Art. no. 109296.

[57] S. Zheng et al., "Deep plug-and-play priors for spectral snapshot compressive imaging," *Photon. Res.*, vol. 9, no. 2, p. B18, 2021.

[58] Y. Chen, X. Gui, J. Zeng, X.-L. Zhao, and W. He, "Combining low-rank and deep plug-and-play priors for snapshot compressive imaging," *IEEE Trans. Neural Netw. Learn. Syst.*, p. 1, 2023, doi: 10.1109/TNNLS.2023.3294262.

[59] J. M. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[60] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, Oct. 2019.

[61] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[62] Y. Chen, J. Zhang, J. Zeng, W. Lai, X. Gui, and T.-X. Jiang, "A guidable nonlocal low-rank approximation model for hyperspectral image denoising," *Signal Process.*, vol. 215, Feb. 2024, Art. no. 109266.

[63] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, vol. 78, no. 1, pp. 29–63, Jan. 2019.

[64] J. Zeng, S.-B. Lin, Y. Yao, and D.-X. Zhou, "On ADMM in deep learning: Convergence and saturation-avoidance," *J. Mach. Learn. Res.*, vol. 22, no. 199, pp. 1–67, 2021.

[65] T. Uezato, D. Hong, N. Yokoya, and W. He, "Guided deep decoder: Unsupervised image pair fusion," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 87–102.
[66] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[67] H. Arguello and G. R. Arce, "Colored coded aperture design by concentration of measure in compressive spectral imaging," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1896–1908, Apr. 2014.

Yong Chen received the B.S. degree from the School of Science, East China University of Technology, Nanchang, China, in 2015, and the Ph.D. degree from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2020. From 2018 to 2019, he was a Research Intern with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Japan. He is currently with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang. His research interests include hyperspectral image processing, low-rank matrix/tensor representation, and model-driven deep learning. More information can be found on his homepage at: https://chenyong1993.github.io/yongchen.github.io/

Wenzhen Lai received the B.S. degree from the Nanchang Institute of Technology, Nanchang, China, in 2021. He is currently pursuing the M.S. degree with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang. His research interests include hyperspectral image reconstruction and deep learning.

Wei He (Senior Member, IEEE) received the B.S. degree from the School of Mathematics and Statistics and the Ph.D. degree in surveying, mapping and remote sensing from LIESMARS, Wuhan University, Wuhan, China, in 2012 and 2017, respectively. From 2018 to 2020, he was a Researcher with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan, where he was a Research Scientist, from 2020 to 2021. He is currently a Full Professor with LIESMARS, Wuhan University. His research interests include image quality improvement, remote sensing image processing and low-rank representation, and deep learning.

Xi-Le Zhao (Member, IEEE) received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009 and 2012, respectively. He was a Postdoctoral Researcher along with Prof. Michael K. Ng at Hong Kong Baptist University from 2013 to 2014 and a Visiting Scholar along with Prof. Jose Bioucas Dias at the University of Lisbon from 2016 to 2017. He is currently a Professor with the School of Mathematical Sciences, UESTC. His research interests include image processing, machine learning, and scientific computing. More information can be found on his homepage at: https://zhaoxile.github.io/

Jinshan Zeng (Member, IEEE) received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 2015. Since July 2022, he has been the Associate Dean of the School of Computer and Information Engineering. He is currently a Professor with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, China. He has published over 60 papers in high-impact journals and conferences, such as *Journal of Machine Learning Research*, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE/IEEE TRANSACTIONS ON SIGNAL PROCESSING/IEEE TRANSACTIONS ON IMAGE PROCESSING/IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS/IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, ICML, and AAAI. His research interests include nonconvex optimization, machine learning, remote sensing, and computer vision. He has had two papers coauthored with collaborators that received the International Consortium of Chinese Mathematicians (ICCM) Best Paper Award in 2018 and 2020.