

Combining Low-Rank and Deep Plug-and-Play Priors for Snapshot Compressive Imaging

Yong Chen¹, Xinfeng Gui¹, Jinshan Zeng¹, Xi-Le Zhao¹, and Wei He¹, *Senior Member, IEEE*

Abstract—Snapshot compressive imaging (SCI) is a promising technique that captures a 3-D hyperspectral image (HSI) by a 2-D detector in a compressed manner. The ill-posed inverse process of reconstructing the HSI from their corresponding 2-D measurements is challenging. However, current approaches either neglect the underlying characteristics, such as high spectral correlation, or demand abundant training datasets, resulting in an inadequate balance among performance, generalizability, and interpretability. To address these challenges, in this article, we propose a novel approach called LR2DP that integrates the model-driven low-rank prior and data-driven deep priors for SCI reconstruction. This approach not only captures the spectral correlation and deep spatial features of HSI but also takes advantage of both model-based and learning-based methods without requiring any extra training datasets. Specifically, to preserve the strong spectral correlation of the HSI effectively, we propose that the HSI lies in a low-rank subspace, thereby transforming the problem of reconstructing the HSI into estimating the spectral basis and spatial representation coefficient. Inspired by the mutual promotion of unsupervised deep image prior (DIP) and trained deep denoising prior (DDP), we integrate the unsupervised network and pre-trained deep denoiser into the plug-and-play (PnP) regime to estimate the representation coefficient together, aiming to explore the internal target image prior (learned by DIP) and the external training image prior (depicted by pre-trained DDP) of the HSI. An effective half-quadratic splitting (HQS) technique is employed to optimize the proposed HSI reconstruction model. Extensive experiments on both simulated and real datasets demonstrate the superiority of the proposed method over the state-of-the-art approaches.

Index Terms—Deep image prior (DIP), hyperspectral compressive imaging, low-rank representation, plug-and-play (PnP) prior.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) are 3-D data with dozens to hundreds of spectral bands which contain 2-D spatial information and 1-D-specific spectral information on the target scene. Due to their abundant spectral information, HSIs have been widely applied in various areas, such as remote sensing [1], [2], medical imaging [3], military [4], [5].

To capture 3-D HSI, common imaging systems scan the scene along spatial and spectral dimensions with 1-D or 2-D detectors which are time-consuming and memory-consuming. Recently, compressive sensing theory-based imaging systems called snapshot-hyperspectral compressive imaging (SCI) systems are popular. Compared to scanning-based technologies, SCI has the advantages of low memory cost, low time consumption, and low power dissipation. Among those SCI systems [6], [7], [8], the coded aperture snapshot spectral imaging (CASSI) system [9] is the representative one. The first phase of CASSI encodes the 3-D HSI into a single 2-D snapshot measurement, and the second phase reconstructs desirable HSI from the snapshot measurement via reconstruction algorithms. This article focuses on developing an efficient reconstruction method for the second phase.

To reconstruct high-quality HSIs from the corresponding snapshot measurements, various reconstruction algorithms have been proposed. These approaches can be broadly divided into three categories: model-based, learning-based, and hybrid-based approaches. Model-based approaches utilize handcrafted priors, such as sparsity [10], [11], [12], total variation (TV) [13], [14], low-rank [15], [16], [17], and non-local self-similarity (NSS) [18], [19], [20], and then design an iterative optimization algorithm for SCI reconstruction. Overall, model-based approaches enjoy good interpretability and generalization since they consider the physical prior characteristic and imaging model of desired HSI. However, handcrafted priors lack an adaptive ability to highly complex features. To remedy this weakness, learning-based methods attempt to learn the mapping function from 2-D measurement to 3-D HSI by various parameterized neural networks, such as convolutional neural network (CNN) [21], [22], spatial-spectral attention network [23], [24], and transformer network [25], [26]. Although promising results have been achieved, learning-based methods demand sufficient training data, which

Manuscript received 4 December 2022; revised 5 May 2023 and 7 June 2023; accepted 1 July 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62101222, Grant 61977038, Grant 62131005, Grant 12171072 and Grant 42271370; in part by the National Science Foundation of Jiangxi, China, under Grant 20232ACB212001, Grant 20224BAB212001, and Grant 20224ACB212004; in part by Thousand Talents Plan of Jiangxi Province under Grant jxsq2019201124; in part by the National Key Research and Development Program of China under Grant 2020YFA0714001; in part by the Open Research Fund Program of Data Recovery Key Laboratory of Sichuan Province under Grant DRN2302; in part by the Fundamental Research Funds for the Central Universities under Grant 2042022rc0027; and in part by the Postgraduate Innovation Fund of the Jiangxi Provincial Department of Education under Grant YC2022-s328. (Corresponding author: Jinshan Zeng.)

Yong Chen, Xinfeng Gui, and Jinshan Zeng are with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang 330022, China (e-mail: chenrong1872008@163.com; xiaofeng@jxnu.edu.cn; jinshanzeng@jxnu.edu.cn).

Xi-Le Zhao is with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 610051, China (e-mail: xlzhao122003@163.com).

Wei He is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: weihe1990@whu.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TNNLS.2023.3294262>.

Digital Object Identifier 10.1109/TNNLS.2023.3294262

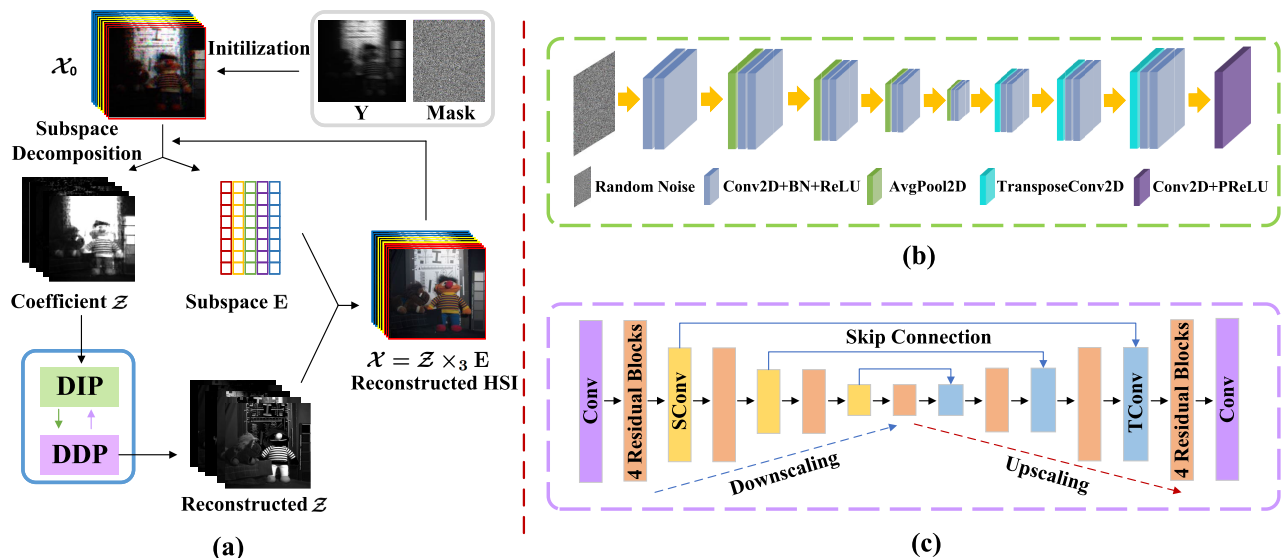


Fig. 1. Overall architecture of LR2DP. (a) It consists of three main stages: 1) low-rank subspace decomposition; 2) spatial representation coefficient reconstruction by two deep priors; and 3) HSI reconstruction from the subspace and spatial representation coefficient. (b) Network architecture of DIP. (c) Network architecture of DDP.

is expensive to acquire for network pretraining. Moreover, the network trained under specific training datasets cannot be well applied to other imaging systems that are not included in the training data.

Recently, to address the issue of HSI training data, hybrid-based methods integrate deep priors depicted by pre-trained deep network or untrained deep network and handcrafted priors into the plug-and-play (PnP) framework for SCI reconstruction [27], [28], [29]. Since these methods do not require extra training data and incorporate an imaging model, thus it is a good trade-off for reconstruction performance and generalization. However, most of the hybrid-based methods usually directly design the deep priors for HSI itself, while ignoring the discriminatively intrinsic spectral low-rank structure of the HSI, which can promote SCI reconstruction with accurate spectral information. Moreover, the rough combination of pre-trained or untrained deep priors with handcrafted priors may ignore the potential cooperation between deep priors which depict different target features.

To address these issues, we formulate the SCI reconstruction into a popular PnP framework (see Fig. 1), which aims at taking the hybrid advantages of model-based and learning-based methods and making these strengths promote each other. **First**, to preserve the strong spectral correlation of the HSI effectively, we design model-based low-rank subspace representation (LRSR) to project the high-dimensional HSI into a low-dimensional spectral subspace. As a result, the HSI reconstruction problem is transformed into the estimation of spectral basis and spatial representation coefficient, which will usually be more precise than reconstructing the original HSI directly since the number of unknown variables is significantly reduced. Moreover, the model-driven low-rank prior can inherit the established advantages of model-based approaches, such as high interpretability and excellent generalization ability. **Second**, it is different from existing methods to regularize the original HSI directly by employing handcrafted

prior or designing an unsupervised deep image prior (DIP) or pre-trained deep denoising prior (DDP) alone, we integrate the unsupervised DIP and the pre-trained deep denoiser to explore the internal and external structure of the representation coefficient, respectively, which is more stable and reasonable since the subspace of representation coefficient is the same as the one where spanned by original HSI. Especially, the internal structural features are learned from the data itself via the DIP network, while the external features are obtained by pre-trained deep denoiser from diverse additional datasets. The advantage of data-driven deep priors we designed is that they can not only greatly reduce the overall error induced by the model-driven handcrafted priors but also enjoy the power of deep neural networks and do not require any extra training data. **Finally**, considering that model-driven and data-driven priors could promote each other, we propose a novel method (called LR2DP) by combining the low-rank and deep priors to enhance SCI reconstruction with the crucial of strong interpretability, superior generalization ability, and high reconstruction capacity.

Fig. 2 presents an example to illustrate that the model-driven low-rank prior and data-driven deep priors are indeed complementary to each other. We can observe that the overall structure of the image can be reconstructed when only DIP is used, but additional noise emerges and spectral information is lost [see Fig. 2(b)]. When DDP is introduced into DIP, the performance is improved since the noise has disappeared. However, the result reconstructed by these two deep priors is over-smoothed [see Fig. 2(c)]. Similarly, the low-rank prior can promote DIP to preserve the spectral information but the artifacts exist in the image [see Fig. 2(d)]. In contrast, the proposed method, which incorporates low-rank prior and two deep priors, reconstructs the overall structure without additional noise or blur and effectively preserves the spectral information [see Fig. 2(e)]. The major contributions of this article can be summarized as follows.

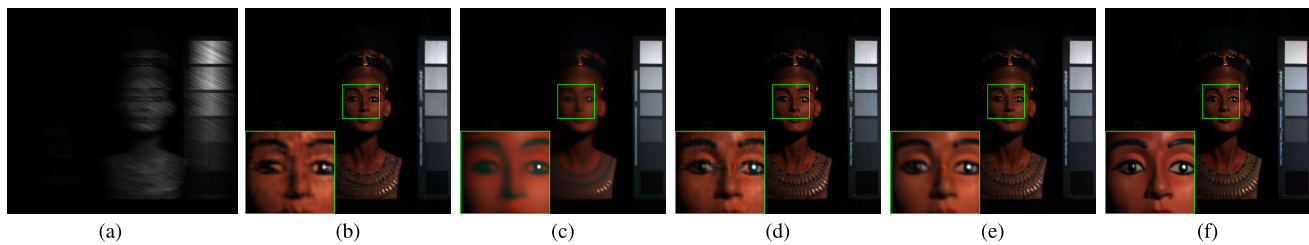


Fig. 2. Illustration of SCI reconstruction results in which model-driven low-rank prior and data-driven deep prior promote each other. (a) Two-dimensional measurement. (b) Using DIP alone (PSNR: 37.64 dB). (c) Combining DIP with DDP (PSNR: 37.30 dB). (d) Combining DIP with low-rank prior (PSNR: 40.10 dB). (e) Proposed method (PSNR: 40.40 dB). (f) Ground truth.

- 1) We propose a novel method that combines the advantages of model-based and learning-based methods for SCI reconstruction so that the strong spectral correlation and spatial structure of the HSI are captured by model-driven low-rank prior and data-driven deep priors, respectively.
- 2) The unsupervised DIP and pre-trained DDP are designed as deep priors to regularize the reduced-dimensionality representation coefficient, not HSI itself, which can effectively explore deep internal and external features of HSI without extra training datasets.
- 3) An efficient optimization framework half-quadratic splitting (HQS) is designed to solve the proposed coefficient reconstruction model. A series of experiments on both simulated and real datasets demonstrate that the proposed method achieves state-of-the-art performance.

The rest of this article is organized as follows. We introduce the related works of SCI reconstruction in Section II. Section III introduces the notations and the problem formulation of CASSI. The proposed SCI reconstruction method is presented in Section IV. Section V shows the experimental results and detailed discussion of the proposed method, and the conclusion is drawn in Section VI.

II. RELATED WORK

Recently, the SCI reconstruction methods have been widely researched, which are mainly divided into three directions: model-based, learning-based, and hybrid-based methods. In this section, we briefly review the related works.

A. Model-Based Methods

Due to the ill-posed inverse problem of SCI reconstruction from a snapshot measurement, model-based methods design different handcrafted priors to exploit the spatial-spectral structure of the desired HSI and regularize the inverse problem for a stable solution. The TV regularization has been employed to preserve the spatial piecewise smoothness of HSI in SCI reconstruction [13], [14]. The sparsity prior achieves a better result than TV regularization since it describes the sparse characteristic of HSI in a redundancy dictionary or fixed transform domain [10], [11]. To explore the spatial, spectral, and nonlocal correlations of HSI, low-rank matrix/tensor approximations have been designed for SCI reconstruction [16], [17], [18], [19]. Overall, these model-based optimization algorithms are handcrafted based on prior knowledge, so they are usually

interpretable but they are not always sufficient to capture the characteristics in various spectral images.

B. Learning-Based Methods

Learning-based methods have been employed for SCI reconstruction via learning the end-to-end function from the observed measurement to the desired HSI [24], [30], [31], [32], [33], [34]. The sophisticated CNN is employed to explore spatial-spectral correlations of HSI from sufficient external datasets for SCI reconstruction [30]. To address the deficiencies in long-range dependency and content independence between images and convolution kernels, the attention mechanism is introduced in CNN for SCI reconstruction. TSA-Net [24] designed a CNN with spatial-spectral self-attention to tackle each dimension sequentially for SCI reconstruction. Recently, with the relevancy between computer vision and natural language processing, the transformer and recurrent neural networks have been introduced into SCI reconstruction tasks. Cai et al. [33] proposed a mask-guided spectral-wise transformer for SCI reconstruction, which can capture the long-range inter-spectra dependencies. Even though these learning-based methods achieve excellent reconstruction performance, they all demand sufficient training samples for supervised learning, which is difficult in the research fields related to HSI. Moreover, each trained network is usually dedicated to one specific imaging device, as the imaging device changes, they need to retrain a new network, which constrains the generalization of this kind of method.

C. Hybrid-Based Methods

To cope with the problem of insufficient HSI training data, hybrid-based methods [27], [28], [35], [36] try to combine conventional physical model and PnP priors (pre-trained or untrained deep denoiser networks) for SCI reconstruction, which inherits the advantages of model-based and learning-based methods. To balance the speed, accuracy, and generalization for SCI reconstruction, Yuan et al. [36] presented an efficient PnP-GAP algorithm by using various denoisers, such as learning-based FFDNet. Since using a single pre-trained denoiser alone cannot obtain satisfactory results, FFDNet-TV [28] combined the FFDNet and TV denoisers to improve the existing hybrid-based algorithms for SCI reconstruction. As the pre-trained deep denoiser FFDNet is learned from nature images, it may not be able to represent the HSI.

To address this challenge, PnP-DIP [27] developed a self-supervised neural network that integrates DIP into the PnP regime for SCI reconstruction. Overall, hybrid-based methods integrate the implicit deep denoisers or explicit handcraft denoisers in the regularization model, which can improve the interpretability and generalization of SCI reconstruction. Despite the effectiveness of this kind of method, they usually directly design the PnP priors to learn HSI itself but ignore the discriminatively intrinsic low-rank structure of the HSI, which limited the improvement of spectral reconstruction quality. Moreover, the usual combination of handcrafted denoisers or deep denoisers lacks further exploration of the spatial information of HSI.

Inspired by the effectiveness and generalization of PnP priors, we have attempted to involve the advantages of hybrid-based methods in the coded SCI reconstruction and remedy the deficiencies to effectively boost the reconstruction quality.

III. NOTATIONS AND PROBLEM FORMULATION

A. Notations

In this article, scalars, vectors, matrices, and tensors are represented by 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 ($\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n}$), respectively. The element value of \mathcal{X} in location (i_1, i_2, \dots, i_n) is represented by $\mathcal{X}(i_1, i_2, \dots, i_n)$. The mode- k unfolding of tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n}$ is denoted as $\mathbf{X}_{(k)} \in \mathbb{R}^{I_k \times I_1 I_2 \dots I_{k-1} I_{k+1} \dots I_n}$. The frontal, lateral, and horizontal slices of a 3-D tensor \mathcal{X} are denoted as $\mathcal{X}(:, :, i_3)$, $\mathcal{X}(:, i_2, :)$, and $\mathcal{X}(i_1, :, :)$, respectively. The mode- k product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n}$ and a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_k}$ is defined as $\mathcal{Y} = \mathcal{X} \times_k \mathbf{U}$, where $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_{k-1} \times J \times I_{k+1} \times \dots \times I_n}$ and $\mathcal{Y}_{i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_n} = \sum_{i_k=1}^{I_k} \mathcal{X}_{i_1, i_2, \dots, i_n} \cdot \mathbf{U}_{j, i_k}$. For the Frobenius norm of \mathcal{X} , it is denoted as $\|\mathcal{X}\|_F = \sum_{i_1, i_2, \dots, i_n} (\mathcal{X}_{i_1, i_2, \dots, i_n})^{(1/2)}$.

B. Problem Formulation of CASSI

There are two popular types of imaging systems in CASSI, i.e., SD-CASSI employing a single disperser encoded in the spatial domain and DD-CASSI encoded in both spatial and spectral domains [37]. The imaging principle of CASSI is presented in Fig. 3, which encodes the 3-D hyperspectral cube onto a 2-D measurement. The key imaging processes of SD-CASSI include the following steps: 1) the 3-D cube is modulated by a physical mask; 2) it is dispersed by the prism; and 3) the coded cube is integrated into a 2-D measurement along the spectral dimension. Mathematically, $\mathcal{X} \in \mathbb{R}^{h \times w \times B}$ denotes the target HSI to be captured, which is modulated by the mask $\mathbf{M} \in \mathbb{R}^{h \times w}$, where h , w , and B represent the height, width, and number of bands, respectively, then we can obtain

$$\mathcal{X}'(:, :, b) = \mathcal{X}(:, :, b) \odot \mathbf{M}, \quad (b = 1, 2, \dots, B) \quad (1)$$

where $\mathcal{X}'(:, :, b)$ represents the band b of modulated target HSI, \odot represents the element-wise multiplication. \mathcal{X}' becomes tilted along the spatial vertical direction after it passes the disperser. We use $\mathcal{X}'' \in \mathbb{R}^{(h+d(B-1)) \times w \times B}$ to denote the tilted HSI, where d denotes the shifting step, and we have

$$\mathcal{X}''(u, v, b) = \mathcal{X}'(h + d(\lambda_b - \lambda_a), w, b) \quad (2)$$

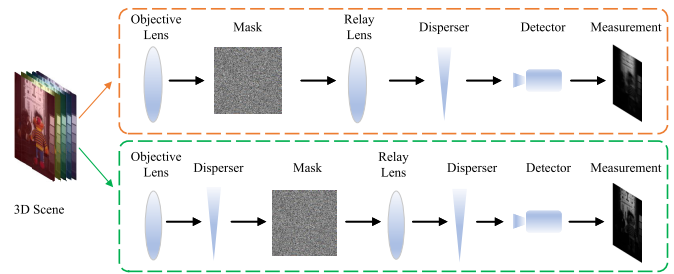


Fig. 3. Illustration of two coded imaging principles in CASSI. (Top) SD-CASSI. (Bottom) DD-CASSI.

where (u, v) represents the coordinate system on the detector plane, λ_b and λ_a represent the wavelength of the b th band and the reference wavelength, respectively. Therefore, we can obtain the final 2-D measurement $\mathbf{Y} \in \mathbb{R}^{(h+d(B-1)) \times w}$ by

$$\mathbf{Y} = \sum_{b=1}^B \mathcal{X}''(:, :, b) + \mathbf{N}'' \quad (3)$$

where $\mathbf{N}'' \in \mathbb{R}^{(h+d(B-1)) \times w}$ denotes measurement noise generated in the imaging process.

For the DD-CASSI, the target 3-D HSI is first spectrally dispersed by the prism and modulated by the physical mask. Then the modulated HSI goes through the relay lens and is dispersed by another inverse prism for spectral gathering. Similar to (3), The final captured 2-D compressive measurement in DD-CASSI can be represented as follows:

$$\mathbf{Y} = \sum_{b=1}^B \mathcal{X}'(:, :, b) + \mathbf{N}' \quad (4)$$

where $\mathbf{Y} \in \mathbb{R}^{h \times w}$ and $\mathbf{N}' \in \mathbb{R}^{h \times w}$ represent 2-D compressive measurement and noise, respectively.

For convenience, we denote the whole imaging process of two CASSI imaging processes as follows:

$$\mathbf{Y} = \mathcal{H}(\mathcal{X}) + \mathbf{N} \quad (5)$$

where $\mathcal{H}(\cdot) : \mathbb{R}^{h \times w \times B} \rightarrow \mathbb{R}^{(h+d(B-1)) \times w}$ or $\mathbb{R}^{h \times w}$ is an operator that contains whole imaging process.

IV. PROPOSED METHOD

The estimation of HSI \mathcal{X} from the measurement \mathbf{Y} is an ill-posed problem. Model-based methods reconstruct the HSI by exploring the prior knowledge of desired \mathcal{X} . The reconstruction model can be formulated as follows:

$$\arg \min_{\mathcal{X}} \frac{1}{2} \|\mathbf{Y} - \mathcal{H}(\mathcal{X})\|_F^2 + \lambda R(\mathcal{X}) \quad (6)$$

where the first term is the data-fidelity term, $R(\mathcal{X})$ is the regularization term specifying the prior knowledge of HSI, and λ is a positive regularization parameter. The handcrafted priors, such as TV [13], low-rank [16], and NSS [18], [38], [39], can be employed to regularize \mathcal{X} . However, these handcrafted priors may not be able to explore the deep features of HSI. Recently, deep neural networks can capture the inherent characteristics of HSI by a large number of training datasets or degraded measurement itself. Therefore, the deep priors

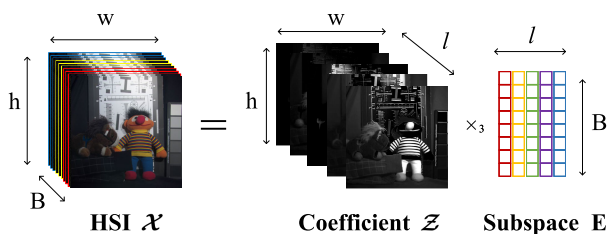


Fig. 4. LRSR of the HSI.

are introduced to the reconstruction model in (6) for SCI reconstruction [27], [28]. Although these methods perform satisfactory results, few researchers consider both model-driven low-rank prior and data-driven deep priors for SCI reconstruction.

A. Low-Rank Prior of HSI

The global spectral correlation is an intrinsic prior that exists in HSI, which has been widely used in different applications, such as compressive hyperspectral imaging [17], super-resolution [40], [41], [42], and restoration [43], [44], [45]. To capture the global spectral correlation of HSI, the LRSR is introduced to approximate it as follows:

$$\mathcal{X} = \mathcal{Z} \times_3 \mathbf{E} \quad (7)$$

where $\mathbf{E} \in \mathbb{R}^{B \times l}$ ($l \ll B$) denotes the orthogonal basis matrix capturing the common subspace of different spectral signatures, and $\mathcal{Z} \in \mathbb{R}^{h \times w \times l}$ is the subspace (spatial) representation coefficient. Fig. 4 presents the global spectral correlation of HSI via LRSR.

Using the low-rank subspace decomposition, the HSI reconstruction can be transformed into the spectral basis and subspace coefficient estimation. Since the ground-truth \mathcal{X} is unknown, it is difficult to learn the spectral basis directly from (7). To obtain the spectral basis, we use the singular value decomposition (SVD) of $\hat{\mathcal{X}}$ from the previous iteration to learn the approximation solution [46]. Given the result of $\hat{\mathcal{X}}$, we obtain the spectral basis as follows:

$$\mathbf{E} = \mathbf{U}(:, 1:l) \quad (8)$$

where \mathbf{U} is the left singular matrix of $\hat{\mathbf{X}}_{(3)}$.

B. Combining Low-Rank and Deep PnP Priors

With the spectral basis \mathbf{E} known, we need to further estimate the spatial coefficient \mathcal{Z} . Based on the maximum a posteriori (MAP) estimation, we can obtain the following coefficient estimation model by combining (6) and (7):

$$\arg \min_{\mathcal{Z}} \frac{1}{2} \|\mathbf{Y} - \mathcal{H}(\mathcal{Z} \times_3 \mathbf{E})\|_F^2 + \lambda R(\mathcal{Z}) \quad (9)$$

where $R(\mathcal{Z})$ is the regularization term describing the prior of \mathcal{Z} . With the column orthogonal constraint of subspace basis \mathbf{E} , the spatial characteristic of \mathcal{X} in the image domain can be transformed to the subspace coefficient domain. Thus, the regularization term $R(\mathcal{X})$ in (6) is transformed as $R(\mathcal{Z})$ in (9). There are many handcrafted priors can be used to constrain

\mathcal{Z} , e.g., BM3D [46], TV [47], NSS [48]. However, they lack an adaptive ability to capture different spatial features in HSI.

Recently, deep neural networks devote to adaptively learning the image prior. To take full advantage of the ability of neural networks to learn priors, we introduce the unsupervised DIP [49], [50], [51] to regularize the coefficient \mathcal{Z} . The major advantage of DIP is that it does not require ground truth during the learning and can exploit the internal target image prior. By incorporating the DIP prior and absorbing the regularization $R(\mathcal{Z})$, the coefficient reconstruction model is suggested as follows:

$$\arg \min_{\Theta} \frac{1}{2} \|\mathbf{Y} - \mathcal{H}(\mathcal{F}_{\Theta}(\mathcal{E}) \times_3 \mathbf{E})\|_F^2 \quad (10)$$

where $\mathcal{F}_{\Theta}(\cdot)$ denotes the network of DIP, Θ is the network parameters, and \mathcal{E} is a random noise input into the network whose size is the same as \mathcal{Z} . It is worth noting that the optimization of \mathcal{Z} is transformed to learn the network parameters Θ . Then, an approximation solution of \mathcal{Z} is computed as $\mathcal{Z} = \mathcal{F}_{\Theta}(\mathcal{E})$.

It seems that model in (10) can achieve satisfactory reconstruction results since it simultaneously utilizes the spectral low-rank prior and spatial deep prior. However, the performance of model in (10) still has room for improvement since the result of DIP is limited for learning multidimensional data [52]. Moreover, as the iterative termination criteria of DIP is blind, it will introduce extra noises or blurred features. To improve the performance of model in (10), extra regularization can be introduced to promote the performance of DIP, such as handcrafted TV prior [28]. Instead of using the handcrafted prior, we employ the prior learned from extensive images to regularize the coefficient. Therefore, motivated by the effectiveness of the PnP framework, the well-trained DDP is employed to represent the external feature of the coefficient. This leads to the final reconstruction model as follows:

$$\arg \min_{\Theta} \frac{1}{2} \|\mathbf{Y} - \mathcal{H}(\mathcal{F}_{\Theta}(\mathcal{E}) \times_3 \mathbf{E})\|_F^2 + \lambda \phi(\mathcal{F}_{\Theta}(\mathcal{E})) \quad (11)$$

where $\phi(\cdot)$ is the regularization term describing the external deep features of the representation coefficient.

C. Efficient Optimization

The objective in (11) is hard to directly optimize, due to the network parameters Θ coupled in the implicit regularization term. To obtain the solution for the model in (11), the HQS algorithm [53] is employed for its simplicity and fast convergence. By introducing an auxiliary variable \mathcal{U} , the problem of (11) is approximately equivalent to the following problem:

$$\arg \min_{\Theta, \mathcal{U}} \frac{1}{2} \|\mathbf{Y} - \mathcal{H}(\mathcal{F}_{\Theta}(\mathcal{E}) \times_3 \mathbf{E})\|_F^2 + \lambda \phi(\mathcal{U}) + \frac{\beta}{2} \|\mathcal{U} - \mathcal{F}_{\Theta}(\mathcal{E})\|_F^2 \quad (12)$$

where β is the penalty parameter. Based on the HQS algorithm, the problem in (12) can be solved by alternately iterating two subproblems Θ and \mathcal{U} .

1) Θ -Subproblem: The subproblem of (12) with respect to Θ is formulated as follows:

$$\arg \min_{\Theta} \frac{1}{2} \|\mathbf{Y} - \mathcal{H}(\mathcal{F}_{\Theta}(\mathcal{E}) \times_3 \mathbf{E})\|_F^2 + \frac{\beta}{2} \|\mathcal{U} - \mathcal{F}_{\Theta}(\mathcal{E})\|_F^2. \quad (13)$$

Obviously, the Θ subproblem is a quadratic minimization problem, which can be regarded as a generalized form of the DIP framework [49]. In our implementation, we adopt a similar UNet structure for optimizing the parameter Θ . The repeated convolution modules are used to design the sampling layers, where each downsampling layer consists of 3×3 strided convolution, and the upsampling layer consists of 3×3 transposed convolution. Moreover, we employ P-Relu instead of Relu as an activation function in the last upsample layer because the values in the coefficient images are not necessarily all positive numbers. To reduce the network parameters, we discard the skip connections and decrease the number of feature channels. Moreover, the loss function of the classic DIP framework usually only has the fidelity term. In this particular case, the proximity regularization that forces $\mathcal{F}_{\Theta}(\mathcal{E})$ to be close \mathcal{U} is provided. The whole formula in (13) is employed as the loss function instead of just the fidelity term, which provides a stable and robust effect to the network optimization.

2) \mathcal{U} -Subproblem: The optimization of \mathcal{U} subproblem can be solved by the following minimization problem:

$$\arg \min_{\mathcal{U}} \lambda \phi(\mathcal{U}) + \frac{\beta}{2} \|\mathcal{U} - \mathcal{F}_{\Theta}(\mathcal{E})\|_F^2. \quad (14)$$

Inspired by the spirit of the PnP framework, the problem of (14) can be regarded as denoising of $\mathcal{F}_{\Theta}(\mathcal{E})$ with known Gaussian noise level of $\sigma^2 = (\lambda/\beta)$. The unsupervised DIP network is employed to learn the internal feature of the subspace coefficient. To adequately explore the external characteristics, we introduce DDP, depicted by the pre-trained deep denoiser which is trained from sufficient images, to effectively capture the externally shared image structures and remove the noise. Many popular pre-trained deep denoisers can be selected to solve the denoising problem, such as DnCNN [54] and FFDNet [55]. In this article, we choose a well-trained deep denoiser named DRUNet [56] to depict the DDP as the denoising engine [see Fig. 1(c)], which is demonstrated as more efficient than current deep denoisers like FFDNet. Especially, the generalization of DRUNet is better than other deep denoisers since it is suitable for a wide range of seen and unseen noise levels. Thus, although it is difficult to select proper parameters λ and β which ensures satisfactory results and proper noise level σ , DRUNet can handle this issue. DRUNet employs the UNet with skip connections as its backbone which contains four stages of upscaling and downscaling. In each scale, the layers are composed of four successive residual blocks and one scaling layer, in which the downscaling layer consists of 2×2 strided convolution and the upscaling layer consists of 2×2 transposed convolution. Then, the solution of (14) can be formulated as follows:

$$\hat{\mathcal{U}}(:, :, p) = \text{DRUNet}(\mathcal{Z}(:, :, p), \sigma_{c_p}), \quad (p = 1, 2, \dots, l) \quad (15)$$

where $\mathcal{Z}(:, :, p)$ is the p th slice of $\mathcal{F}_{\Theta}(\mathcal{E})$, and σ_{c_p} is the calculated noise level of p th slice in the coefficient.

Since each slice of the coefficient is not a natural image, we first need to scale the values of each slice into the same range. Each slice of coefficient into $[0, 1]$ is scaled by the min-max scaling, and the noise level is scaled as well

$$\begin{aligned} \hat{\mathcal{U}}(:, :, p) &= \text{DRUNet}(\hat{\mathcal{Z}}(:, :, p), \hat{\sigma}_{c_p}), \quad (p = 1, 2, \dots, l) \\ \hat{\mathcal{Z}}(:, :, p) &= s_i \mathcal{Z}(:, :, p) + d_i, \quad \hat{\sigma}_{c_p} = s_i \sigma_{c_p} \end{aligned} \quad (16)$$

where $\hat{\mathcal{Z}}$ is the normalized $\mathcal{F}_{\Theta}(\mathcal{E})$, $\hat{\sigma}_{c_p}$ is the normalized coefficient, and $s_i = 1/(\max(\mathcal{Z}(:, :, p)) - \min(\mathcal{Z}(:, :, p)))$, $d_i = -s_i \times \min(\mathcal{Z}(:, :, p))$. Finally, we need to scale the updated results back.

We summarize the proposed LR2DP method for SCI reconstruction in Algorithm 1. To perform Algorithm 1, we need to learn the subspace basis from \mathcal{X} in advance. Following the initialization setting of DeSCI, we also employ the GAP-TV to initialize the variables \mathcal{X} because of its high efficiency, and the variable \mathcal{U} is initialized as $\mathcal{X} \times_3 \mathbf{E}^T$.

Algorithm 1 Proposed LR2DP Method for SCI Reconstruction

Input: Measurement \mathbf{Y} , operator \mathcal{H} .

Parameter: Parameters λ and μ , subspace l .

Output: Reconstructed \mathcal{X} .

- 1: Initialize: β , Θ , \mathcal{X}_0 , \mathcal{U} , $k = 0$, and $t = 0$.
 - 2: **while** not converged **do**
 - 3: $k = k + 1$.
 - 4: Learn the subspace \mathbf{E} by (8).
 - 5: **while** $t \leq t_{\max}$ **do**
 - 6: $t = t + 1$.
 - 7: Update network parameters Θ via DIP by (13).
 - 8: Update \mathcal{U} via DRUNet denoiser by (16).
 - 9: Update $\beta = \mu * \beta$.
 - 10: **end while**
 - 11: Update desired HSI via $\mathcal{X} = \mathcal{F}_{\Theta}(\mathcal{E}) \times_3 \mathbf{E}$.
 - 12: Check the convergence condition $\frac{\|\mathcal{X}^k - \mathcal{X}^{k-1}\|_F}{\|\mathcal{X}^{k-1}\|_F} \leq \epsilon$ and $k \leq k_{\max}$.
 - 13: **end while**
-

V. EXPERIMENTS

In this section, we conduct experiments on simulated data and real data for evaluating the performance of the proposed LR2DP. Several state-of-the-art methods are employed for comparison, including two model-based methods: GAP-TV [13] and DeSCI¹ [18], and two hybrid-based methods: TV-FFDNet² [28] and PnP-DIP³ [27]. For their parameter selection of these methods, we carefully follow the author's suggestions in their papers to guarantee optimal results.

A. Simulated Experiments

For a comprehensive evaluation, two coded imaging systems SD-CASSI and DD-CASSI are conducted to verify the effectiveness of our method.

¹<https://github.com/XiaoYangLiu-FinRL/DeSCI>

²<https://github.com/ucker/SCI-TV-FFDNet>

³<https://github.com/mengziyi64/CASSI-Self-Supervised>

TABLE I
SD-CASSI AND DD-CASSI QUANTITATIVE RESULTS OF ALL COMPARISON METHODS ON THE CAVE, HARVARD, AND ICVL DATASETS

Index	Methods	CAVE					Harvard					ICVL				
		Face	Toy	Clay	Egyptian	Cloth	Img0	Img1	Img3	Img4	Img8	Eve	Bgu	Lab02	Lab10	Prk
SD-CASSI																
PSNR	GAP-TV	28.78	23.22	25.85	31.52	22.13	22.82	23.66	27.66	26.29	22.85	25.77	25.30	29.81	33.39	29.61
	DeSCI	35.82	26.38	39.37	35.73	23.45	26.14	29.90	31.83	30.29	25.51	29.52	27.61	32.04	34.96	31.61
	TV-FFDNet	35.44	25.78	37.49	35.21	23.92	26.85	29.25	30.61	29.61	25.48	29.68	27.64	33.00	36.42	32.74
	PnP-DIP	33.89	27.83	33.58	35.40	26.89	29.54	30.61	33.88	31.33	27.56	30.12	30.41	32.33	35.86	30.20
	LR2DP	39.24	33.25	42.53	40.40	29.92	35.46	35.21	39.95	39.17	31.70	36.12	36.51	38.97	42.11	38.10
SSIM	GAP-TV	0.857	0.745	0.753	0.872	0.439	0.581	0.679	0.760	0.698	0.503	0.737	0.694	0.836	0.905	0.821
	DeSCI	0.941	0.852	0.956	0.934	0.506	0.678	0.828	0.870	0.813	0.599	0.826	0.734	0.867	0.920	0.850
	TV-FFDNet	0.941	0.827	0.955	0.931	0.523	0.698	0.815	0.860	0.796	0.592	0.820	0.735	0.875	0.933	0.865
	PnP-DIP	0.876	0.774	0.815	0.899	0.741	0.800	0.772	0.869	0.818	0.684	0.796	0.826	0.867	0.919	0.819
	LR2DP	0.967	0.933	0.968	0.969	0.848	0.934	0.896	0.961	0.954	0.852	0.934	0.943	0.949	0.972	0.943
FSIM	GAP-TV	0.867	0.791	0.774	0.906	0.740	0.773	0.772	0.839	0.804	0.786	0.817	0.809	0.840	0.898	0.864
	DeSCI	0.949	0.892	0.958	0.956	0.728	0.801	0.906	0.911	0.863	0.800	0.858	0.810	0.859	0.899	0.864
	TV-FFDNet	0.950	0.871	0.961	0.957	0.755	0.819	0.898	0.905	0.857	0.812	0.862	0.824	0.883	0.928	0.897
	PnP-DIP	0.940	0.889	0.927	0.946	0.914	0.925	0.916	0.946	0.915	0.921	0.911	0.933	0.905	0.932	0.874
	LR2DP	0.978	0.965	0.963	0.980	0.948	0.977	0.969	0.989	0.985	0.966	0.971	0.980	0.973	0.983	0.974
ERGAS	GAP-TV	211.09	280.44	372.82	214.72	331.49	338.76	311.65	276.24	297.27	367.13	183.41	197.13	126.12	114.32	149.38
	DeSCI	93.97	195.44	76.38	132.60	291.39	193.86	114.02	151.80	155.20	231.69	108.71	153.10	100.73	98.30	117.34
	TV-FFDNet	97.77	210.26	93.36	140.55	275.64	171.30	121.96	176.42	170.17	231.36	106.75	150.43	90.76	82.73	102.89
	PnP-DIP	117.60	164.82	154.31	138.24	191.10	124.58	103.94	119.09	138.20	182.28	101.54	108.08	101.00	88.88	139.65
	LR2DP	63.32	88.41	54.91	77.14	132.89	63.30	61.99	59.22	55.83	115.54	50.83	53.66	45.40	42.41	55.39
SAM	GAP-TV	18.361	18.127	24.920	30.255	16.829	11.170	12.270	15.757	13.951	14.956	7.711	7.730	5.482	6.414	7.962
	DeSCI	11.614	12.118	11.039	20.445	11.444	5.850	6.140	8.851	7.999	9.802	3.529	5.259	3.515	4.513	5.350
	TV-FFDNet	10.803	11.533	11.178	22.974	10.151	4.750	5.307	7.852	6.974	8.285	3.003	4.524	3.227	3.498	4.540
	PnP-DIP	23.466	23.136	26.636	40.008	8.966	6.471	6.400	10.121	7.633	9.132	4.259	4.182	3.189	3.889	6.653
	LR2DP	9.977	10.981	12.053	22.058	6.094	3.079	3.767	4.240	3.845	4.287	2.053	1.999	1.671	1.910	2.859
DD-CASSI																
PSNR	GAP-TV	29.81	25.08	24.49	32.71	21.15	22.64	23.35	27.30	26.25	21.20	25.58	25.36	29.57	33.54	29.82
	DeSCI	34.43	26.33	32.88	33.48	20.57	27.28	26.98	29.07	29.04	20.93	28.70	26.61	30.56	34.23	30.65
	TV-FFDNet	37.59	29.61	39.19	36.97	24.28	30.50	31.50	33.01	32.34	24.95	31.84	29.44	34.37	38.24	33.40
	PnP-DIP	35.53	28.76	36.38	36.78	26.87	30.91	32.05	35.43	34.72	29.31	32.55	33.06	35.95	36.99	33.07
	LR2DP	38.51	33.58	39.23	40.65	30.01	32.03	36.95	40.04	38.53	32.03	34.19	35.65	39.26	43.46	37.19
SSIM	GAP-TV	0.919	0.859	0.757	0.939	0.591	0.714	0.696	0.802	0.764	0.583	0.808	0.812	0.884	0.934	0.871
	DeSCI	0.957	0.910	0.921	0.956	0.576	0.853	0.830	0.905	0.857	0.610	0.874	0.831	0.907	0.949	0.890
	TV-FFDNet	0.966	0.927	0.962	0.965	0.754	0.880	0.890	0.940	0.903	0.754	0.918	0.885	0.942	0.967	0.916
	PnP-DIP	0.932	0.821	0.886	0.939	0.812	0.853	0.808	0.931	0.908	0.829	0.918	0.909	0.939	0.950	0.895
	LR2DP	0.971	0.943	0.956	0.973	0.889	0.935	0.943	0.976	0.971	0.932	0.964	0.975	0.982	0.992	0.966
FSIM	GAP-TV	0.922	0.883	0.802	0.953	0.796	0.813	0.782	0.879	0.850	0.774	0.851	0.852	0.871	0.926	0.894
	DeSCI	0.959	0.926	0.957	0.966	0.733	0.905	0.891	0.926	0.889	0.763	0.886	0.870	0.887	0.930	0.894
	TV-FFDNet	0.974	0.948	0.974	0.976	0.853	0.919	0.946	0.959	0.934	0.861	0.924	0.901	0.933	0.960	0.928
	PnP-DIP	0.975	0.939	0.957	0.979	0.931	0.959	0.948	0.977	0.971	0.954	0.969	0.974	0.970	0.970	0.947
	LR2DP	0.981	0.970	0.949	0.982	0.947	0.971	0.988	0.990	0.985	0.968	0.979	0.982	0.980	0.991	0.974
ERGAS	GAP-TV	194.32	232.42	453.57	188.23	369.38	437.59	382.76	347.01	373.78	521.57	217.61	221.14	135.06	120.45	159.73
	DeSCI	109.84	195.77	173.15	171.33	405.25	191.22	175.91	233.47	188.11	446.10	118.31	165.17	116.30	103.97	130.32
	TV-FFDNet	76.10	134.14	80.77	114.48	260.45	148.42	104.98	149.02	151.55	290.59	83.77	121.01	74.72	66.14	94.82
	PnP-DIP	97.38	149.90	115.45	117.19	191.23	112.31	92.21	108.22	103.36	161.37	78.77	80.37	63.17	82.18	99.98
	LR2DP	69.59	88.37	87.80	76.23	131.08	98.17	62.75	61.24	63.84	122.32	66.23	61.99	44.23	37.47	62.01
SAM	GAP-TV	17.305	17.617	28.303	25.264	23.866	15.583	17.115	19.402	17.380	23.279	10.532	11.122	7.310	8.489	10.480
	DeSCI	11.917	12.151	10.589	15.798	19.137	8.060	10.409	10.558	10.995	20.856	5.672	8.983	4.810	6.005	7.937
	TV-FFDNet	11.404	11.443	10.500	20.502	12.636	5.679	7.081	7.855	7.969	13.260	4.011	6.313	3.410	4.365	6.178
	PnP-DIP	16.324	22.576	19.717	32.075	12.200	7.701	6.605	9.355	8.562	12.002	4.447	5.040	3.599	4.773	6.480
	LR2DP	11.203	14.539	13.572	24.593	8.052	6.601	3.760	5.325	5.034	7.873	3.530	3.331	2.414	2.784	4.464

1) *Datasets*: To show the flexibility of the proposed method, we test the SD-CASSI and DD-CASSI experiments on five public datasets, including three benchmark datasets from a computer vision society, i.e., CAVE,⁴ Harvard,⁵

and ICVL,⁶ and two popular datasets from remote sensing HSI, i.e., Washington DC Mall (WDC) and Pavia University (PaviaU). The CAVE dataset consists of 32 hyperspectral scenes and 31 wavelength bands from 400 to 700 nm with a step of 10 nm whose band sizes are each 512×512 .

⁴<https://www1.cs.columbia.edu/CAVE/databases/multispectral/>

⁵<http://vision.seas.harvard.edu/hyperspec/>

⁶<https://icvl.cs.bgu.ac.il/hyperspectral/>

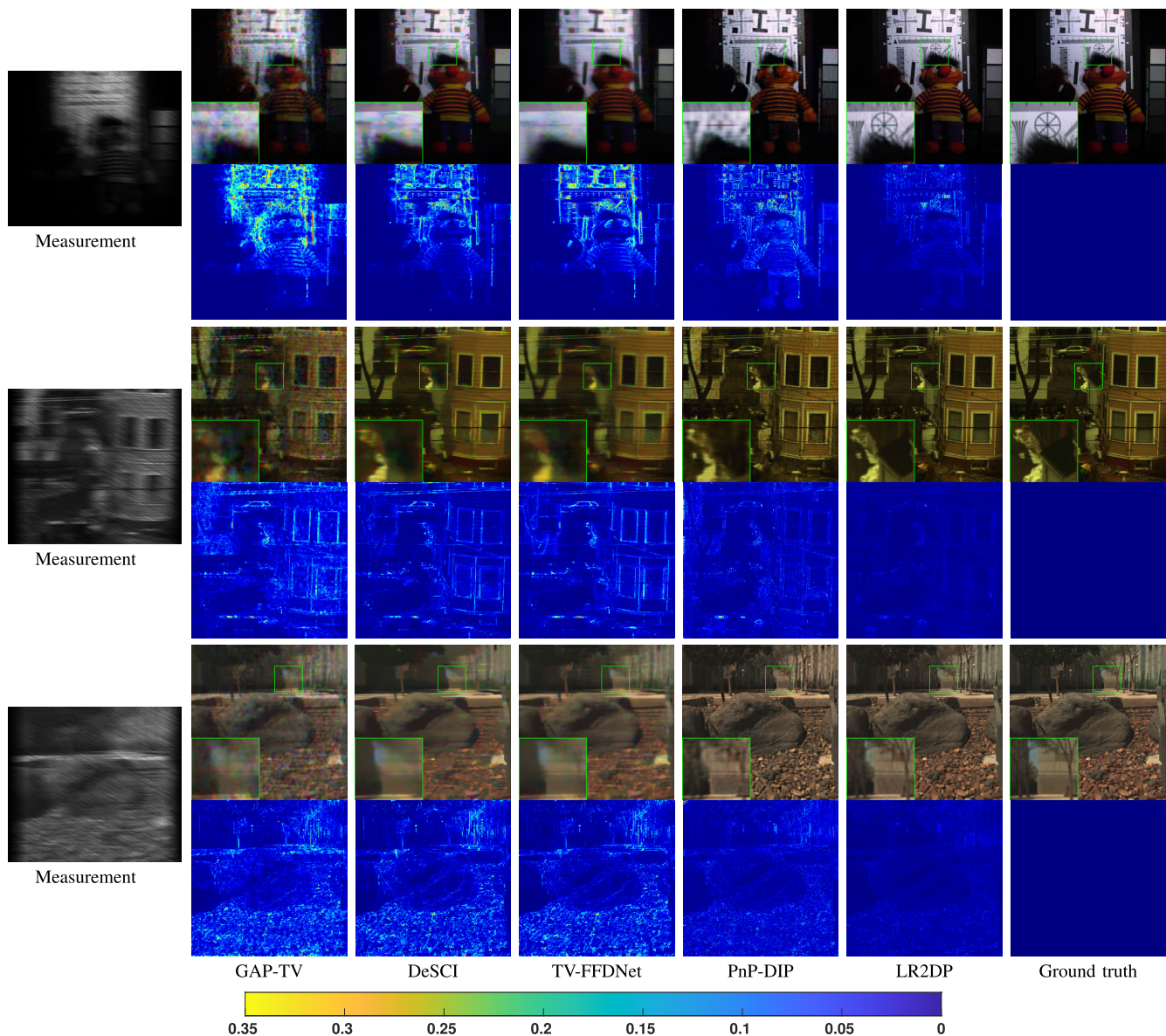


Fig. 5. Reconstruction results of all comparison methods on CVAE-Toy, Harvard-Imgc4, and ICVL-Bgu image for SD-CASSI, respectively. The first, third, and fifth rows are the visual results, and the second, fourth, and sixth rows are the corresponding absolute error maps between the ground truth and reconstructed images. The false color images are composed of bands (R: 28, G: 15, B: 4), (R: 26, G: 12, B: 5), and (R: 29, G: 14, B: 7), respectively.

The Harvard dataset consists of 50 hyperspectral images (HSIs) of indoor and outdoor scenes with 31 bands from 420 to 720 nm at a step of 10 nm, and the spatial size of this dataset is 1040×1392 . Meanwhile, the ICVL dataset contains 201 real-world scenes, and the spatial and spectral sizes of each scene are 1300×1392 and 31 spectral bands collected from 400 nm to 700 nm in a 10 nm step, respectively. We severally choose five scenes from these three datasets as the ground-truth data and crop their spatial resolution into 512×512 . The original WDC contains 1208×307 pixels and 191 spectral bands acquired by the HYDICE sensor, and the PaviaU contains 610×340 spatial size with 103 spectral bands acquired by the ROSIS-03 optical sensor over the area of Pavia University. For convenience, we cropped the corresponding subimage with $256 \times 256 \times 31$ to conduct the experiments, where 31 bands

are extracted from the continuous band 11 to band 41 in WDC and PaviaU.

2) *Quantitative Metrics*: Five quantitative image quality metrics are employed to thoroughly evaluate the performance of all the competing methods [43], including peak signal-to-noise ratio (PSNR), relative dimensionless global error in synthesis (ERGAS), structure similarity (SSIM), feature similarity (FSIM), and spectral angle mapper (SAM). PSNR, SSIM, and FSIM are employed to measure spatial structure reconstruction, while ERGAS and SAM measure spectral fidelity. Generally, the higher values of PSNR, SSIM, and FSIM and smaller results of ERGAS and SAM suggest a better reconstruction accuracy.

3) *Experimental Results on CAVE, Harvard, and ICVL Datasets*: Table I lists the quantitative results for SD-CASSI

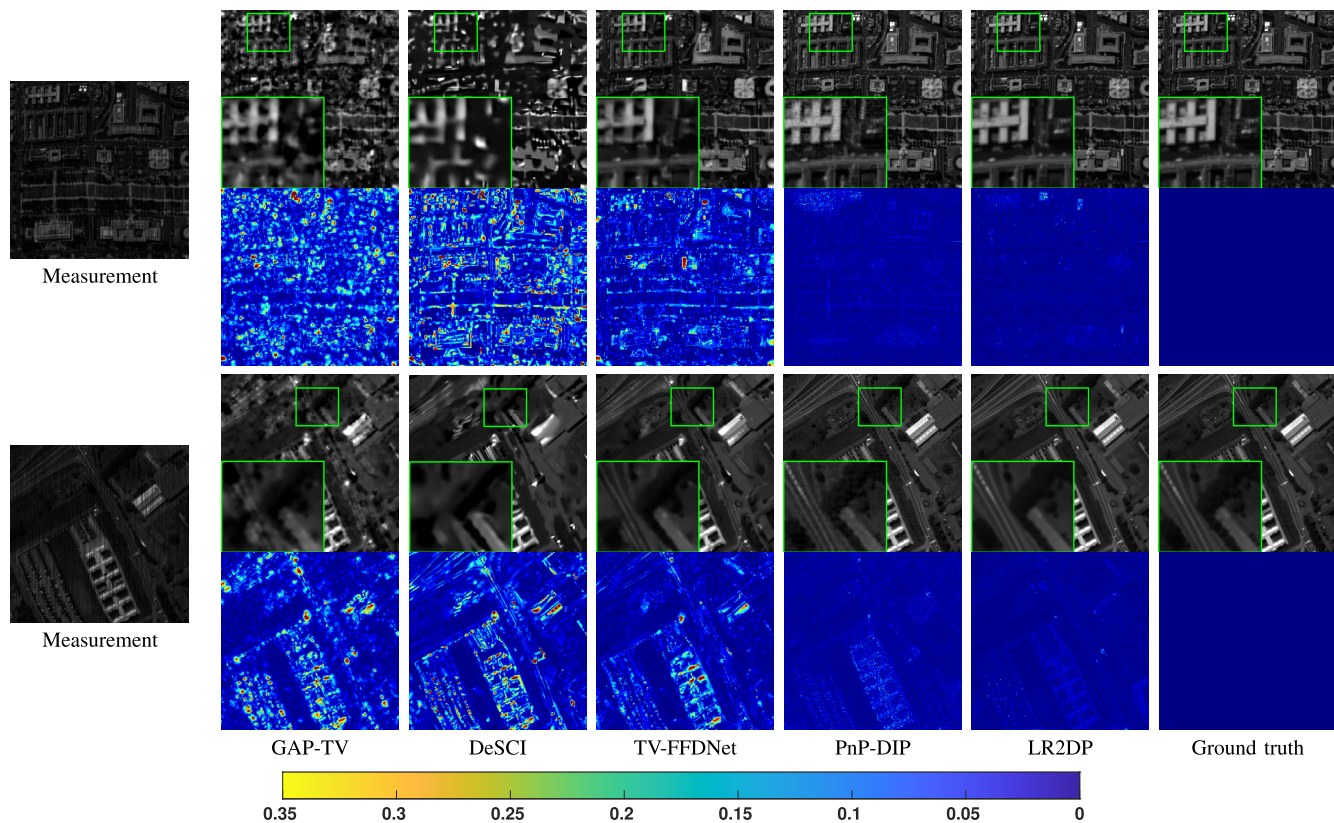


Fig. 6. Reconstruction results of all comparison methods on WDC and PaviaU datasets for DD-CASSI, respectively. The first and third rows are the visual results, and the second and fourth rows are the corresponding absolute error maps between the ground truth and reconstructed images.

and DD-CASSI over five scenes on the CAVE, Harvard, and ICVL datasets. We highlight the best results for each index in bold. From the Table I, we can find that hybrid-based methods are usually superior to model-based methods. Although TV-FFDNet and PnP-DIP employ the PnP denoiser as the prior for reconstruction, none of them take into account the spectral correlation of HSI. Therefore, the results of the SAM index in most cases are significantly inferior to the proposed LR2DP method. Moreover, the competing methods design the regularization to the HSI itself, while the proposed method introduces the spatial deep priors to the reduced-dimensionality representation coefficient, which leads to a better reconstruction of the spatial information. In summary, comparing the quantitative results of different methods on the same imaging system, our method outperforms the existing methods in most cases and metrics. This reveals that our method can effectively preserve the spatial structures and spectral signatures of the reconstructed HSI, demonstrating the capability of LR2DP to utilize the low-rank characteristics of HSI and verifying the effectiveness of mutual promotion between unsupervised deep prior and pre-trained deep prior.

To provide the qualitative comparison, we respectively choose one representative scene from CAVE, Harvard, and ICVL datasets in SD-CASSI to show the reconstructed results. Fig. 5 shows the reconstructed false color images and corresponding residual maps of three scenes before and after reconstruction. To compare the local details of the reconstructed images, we present zoomed-in views of an image area

represented by green rectangles. From the results, we can find that DeSCI reconstructs the global image structures roughly, but the local details are lost. TV-FFDNet and PnP-DIP ignore the spectral correlation of HSI, thus the spectral signature of the reconstructed image is distorted. Moreover, TV-FFDNet blurs the image by introducing handcrafted TV regularization. In contrast, the proposed LR2DP effectively reconstructs the spatial details as well as preserves the spectral information of the original image. Although the proposed method employs the GAP-TV as initialization, we can find that the reconstruction results of GAP-TV deviate greatly from the ground truth image. Therefore, the better reconstruction results are because of our method rather than the effectiveness of initialization. The residual images show the average absolute errors between the reconstruction result and the ground truth. From the error maps, we can find that LR2DP achieves the minimum reconstruction errors, indicating better reconstruction results.

4) *Experimental Results on WDC and PaviaU Datasets:* To further validate that the proposed method can be applied to the reconstruction of remote sensing HSIs, two popular remote sensing HSIs are chosen for the SD-CASSI and DD-CASSI experiments. Table II presents the quantitative results of different reconstruction methods on the two datasets using the SD-CASSI and DD-CASSI measurements. From the results, hybrid-based methods still achieve better results than model-based ones in most cases, indicating the effectiveness of deep PnP priors over handcrafted priors. PnP-DIP obtains better results than TV-FFDNet, demonstrating the advantage

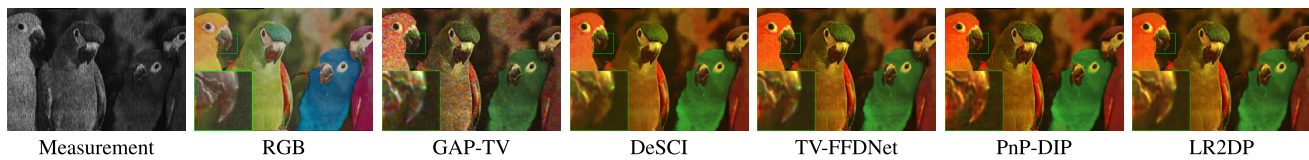


Fig. 7. Reconstruction results of all comparison methods on the real Bird dataset, the false color image is composed by bands (R: 24, G: 11, B: 4).

TABLE II
SD-CASSI AND DD-CASSI QUANTITATIVE RESULTS OF ALL
COMPARISON METHODS ON THE WDC AND
PAVIAU DATASETS

Dataset	Index	GAP-TV	DeSCI	TV-FFDNet	PnP-DIP	LR2DP
SD-CASSI						
WDC	PSNR	23.43	23.33	23.65	30.60	32.88
	SSIM	0.487	0.483	0.501	0.864	0.909
	FSIM	0.740	0.693	0.716	0.930	0.950
	ERGAS	314.01	317.83	306.42	137.79	106.07
	SAM	18.085	17.881	13.204	10.639	5.123
PaviaU	PSNR	23.49	24.59	24.18	31.34	34.49
	SSIM	0.602	0.624	0.636	0.860	0.922
	FSIM	0.750	0.729	0.754	0.917	0.950
	ERGAS	287.13	252.76	265.55	116.05	80.71
	SAM	12.208	10.392	10.651	7.685	3.819
DD-CASSI						
WDC	PSNR	21.81	20.08	25.29	33.88	35.10
	SSIM	0.578	0.502	0.784	0.951	0.967
	FSIM	0.817	0.704	0.865	0.973	0.978
	ERGAS	379.05	462.77	253.84	97.57	85.34
	SAM	34.035	36.006	22.150	13.283	7.614
PaviaU	PSNR	24.05	23.33	27.41	33.39	36.03
	SSIM	0.765	0.734	0.877	0.922	0.962
	FSIM	0.857	0.802	0.906	0.959	0.975
	ERGAS	268.76	292.54	182.86	93.60	70.31
	SAM	18.626	18.615	11.945	9.668	5.772

of the unsupervised DIP prior over pre-trained deep prior. In contrast, we can easily find that our LR2DP achieves the best performance among all the competing methods in all evaluation metrics, again illustrating the superiority of exploiting the model-driven low-rank property of HSIs and data-driven deep priors.

We also provide some visualization results of different methods in Fig. 6, which depicts the SCI reconstruction results of the concerned methods and corresponding error maps on two remote sensing HSIs for DD-CASSI. It can be observed from Fig. 6 that model-based methods fail to reconstruct the serious spatial structure in the WDC dataset. TV-FFDNet and PnP-DIP improve the results, but they are still blurred. The reconstructed results from our method provide the best, with much more detailed information compared to the ground truth. The absolute error maps between ground truth and reconstructed results of scenes are also much closer to the ground truth, verifying that our method obtains higher spatial accuracy. From the SAM index in Table II, the proposed method still achieves optimal values, demonstrating that our method can provide a higher spectral fidelity.

B. Real Experiments

To further show the effectiveness of the proposed method, we test the experiment on real snapshot-hyperspectral

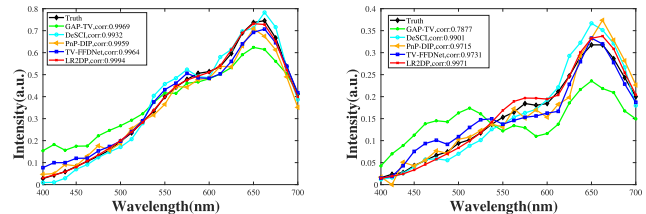


Fig. 8. Spectral curves of the real Bird data. (Left) Spatial position (118 396) and (Right) spatial position (394 750).

compressive imaging data, i.e., the Bird data.⁷ The Bird data consist of 24 spectral channels with a spatial size of 1021×703 . The reconstructed results of all comparison methods are shown in Fig. 7. We present the RGB reference image of the Bird dataset to compare the spatial information of these reconstructed results. GAP-TV introduces some noises in the reconstruction result, and DeSCI produces an artifact, resulting in the loss of spatial details. PnP-DIP and TV-FFDNet obtain satisfactory results, but some details are blurred. In contrast, the proposed method can reconstruct accurate spatial structures, leading to better reconstruction quality.

To compare the spectral information of reconstructed results, we plot the spectral signatures of two spatial positions by different methods in Fig. 8. It can be observed that our method can reconstruct more accurate spectral signatures than other methods, which illustrates better spectral information preservation ability. In summary, our method has the potential to be applied to real SCI reconstruction.

C. Ablation Study and Parameter Analysis

1) *Ablation Study*: To demonstrate the effectiveness of mutual promotion between the model-driven LRSR and two data-driven deep PnP priors (i.e., unsupervised DIP [49] and pre-trained DDP [56]), the ablation study is presented. We conduct the ablation experiments for the SD-CASSI system on five scenes from the CAVE dataset. Table III lists the average quantitative results over the CAVE dataset on five scenes. The result of DIP is employed as the baseline, then we incorporate the DDP, LRSR, and both of them (proposed LR2DP method) to the baseline, respectively. From Table III, we can find that DDP and LRSR can significantly promote the performance of DIP. LR2DP combines the benefits of the model-driven low-rank prior, data-driven DIP, and pre-trained DDP to achieve the best result, demonstrating that our method can enable these priors to promote each other, and thereby achieve much better

⁷<https://github.com/XiaoYangLiu-FinRL/DeSCI>

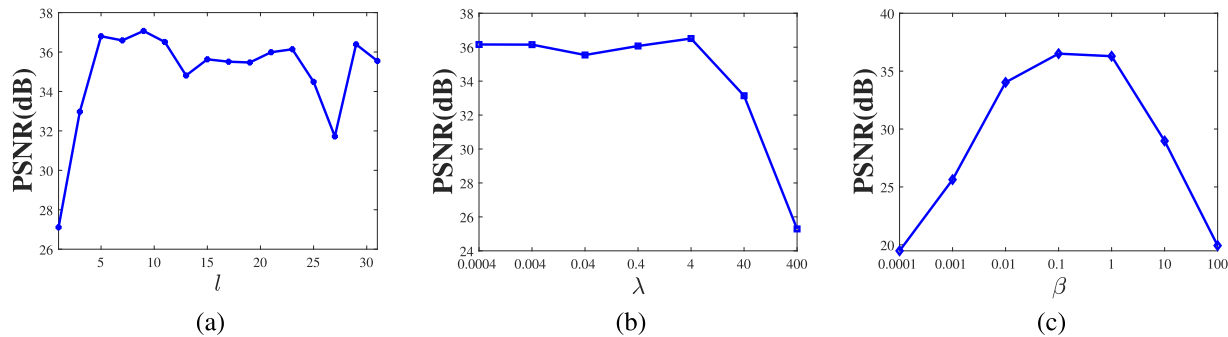


Fig. 9. PSNR curves versus the different parameters. (a) Subspace dimension l . (b) Regularization parameter λ . (c) Penalty parameter β .

TABLE III

QUANTITATIVE RESULTS OF ABLATION STUDY ON CAVE DATASET

Method	PSNR	SSIM	FSIM	ERGAS	SAM
DIP	33.72	0.880	0.945	120.11	16.02
DIP + DDP	34.20	0.861	0.905	121.90	12.82
DIP + LRSR	35.56	0.882	0.956	103.86	15.45
LR2DP	37.07	0.937	0.967	83.33	12.23

results. Visually, Fig. 2 shows the ablation study conducted on CAVE-Egyptian, which illustrates the reconstruction effect benefits from the above priors.

2) *Parameter Analysis*: Subsequently, we analyze the parameter selection of the proposed LR2DP model. There are three key parameters, including subspace dimension l , regularization parameter λ , and penalty parameter β , in our method. To illustrate the influence of each parameter, we plot the PSNR change curves with different parameters shown in Fig. 9. From Fig. 9(a), we can find that satisfactory results of subspace dimension are achieved in the range of [9, 11]. Thus, the subspace dimension l is set as [9, 11] in all experiments, which can represent a low dimensional spectral subspace and preserve sufficient spatial information of HSI. For the parameter λ , it is relatively robust because satisfactory results can be achieved by choosing different λ in a wide range, and we set $\lambda = 4$ in all experiments. Fig. 9(c) presents the PSNR results obtained by the proposed methods with different β . It can be observed that the optimal result is obtained by $\beta = 0.1$, thus we set $\beta = 0.1$ in all experiments. Moreover, the growth rate μ of the penalty parameter is set to 1.1 in all experiments. In summary, the relative optimal reconstruction results can be achieved across a wide range of parameter settings, which shows the robustness advantage of the proposed LR2DP method.

VI. CONCLUSION

In this article, we provide a novel hybrid model-based and learning-based method, named LR2DP, for SCI reconstruction, which can effectively exploit the strong spectral correlation and deep spatial structure of HSI, and adapt itself to diverse scenes. The model-driven low-rank representation motivates LR2DP to maintain the generalization ability and interpretability of model-based methods for the underlying structure of

the global spectral correlation of HSI. The data-driven DIP and DDP enable LR2DP to absorb the superior representation capabilities of learning-based approaches for discriminative exploitation of the internal and external structure of HSI. We plug two data-driven deep priors into the HQS iteration to estimate the spatial representation coefficient. Experimental comparisons on simulated and real datasets show that our LR2DP outperforms current state-of-the-art methods. Moreover, two different imaging systems and six different HSI scenes also demonstrate superior generalization ability and high reconstruction capacity of the proposed method. In the future, we believe that our proposed framework can also be used in other high-dimension data imaging applications, such as video SCI reconstruction problems [57].

REFERENCES

- [1] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, May 2017.
- [2] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.
- [3] Z. Meng, M. Qiao, J. Ma, Z. Yu, K. Xu, and X. Yuan, "Snapshot multispectral endomicroscopy," *Opt. Lett.*, vol. 45, no. 14, pp. 3897–3900, 2020.
- [4] J. Yang, Y. Zhao, J. C. Chan, and S. G. Kong, "Coupled sparse denoising and unmixing with low-rank constraint for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1818–1833, Mar. 2016.
- [5] D. Hong et al., "Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6518–6531, Nov. 2022.
- [6] 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.
- [7] P. Llull et al., "Coded aperture compressive temporal imaging," *Opt. Exp.*, vol. 21, no. 9, pp. 10526–10545, May 2013.
- [8] 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.
- [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, pp. 6368–6388, 2009.
- [10] X. Yuan et al., "Low-cost compressive sensing for color video and depth," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3318–3325.
- [11] 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.
- [12] Z. Zha, B. Wen, X. Yuan, J. Zhou, C. Zhu, and A. C. Kot, "A hybrid structural sparsification error model for image restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4451–4465, Sep. 2022.

- [13] 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.
- [14] 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, pp. 848–858, 2015.
- [15] W. He, Y. Chen, N. Yokoya, C. Li, and Q. Zhao, "Hyperspectral super-resolution via coupled tensor ring factorization," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108280.
- [16] 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.
- [17] W. He, N. Yokoya, and X. Yuan, "Fast hyperspectral image recovery of dual-camera compressive hyperspectral imaging via non-iterative subspace-based fusion," *IEEE Trans. Image Process.*, vol. 30, pp. 7170–7183, 2021.
- [18] 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.
- [19] 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.
- [20] L. Zhang, L. Song, B. Du, and Y. Zhang, "Nonlocal low-rank tensor completion for visual data," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 673–685, Feb. 2021.
- [21] T. Zhang, Y. Fu, L. Wang, and H. Huang, "Hyperspectral image reconstruction using deep external and internal learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8558–8567.
- [22] Y. Sun, J. Chen, Q. Liu, and G. Liu, "Learning image compressed sensing with sub-pixel convolutional generative adversarial network," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107051.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] Z. Wt, J. Zhangt, and C. Mou, "Dense deep unfolding network with 3D-CNN prior for snapshot compressive imaging," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4872–4881.
- [30] 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.
- [31] 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. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8024–8033.
- [32] Z. Cheng et al., "Recurrent neural networks for snapshot compressive imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2264–2281, Feb. 2023.
- [33] 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.
- [34] Y. Quan, X. Qin, M. Chen, and Y. Huang, "High-quality self-supervised snapshot hyperspectral imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 1526–1530.
- [35] J. Ma, X. Liu, Z. Shou, and X. Yuan, "Deep tensor ADMM-net for snapshot compressive imaging," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10222–10231.
- [36] 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.
- [37] Y. 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, 2022, Art. no. 5514314.
- [38] Z. Zha, B. Wen, X. Yuan, J. T. Zhou, J. Zhou, and C. Zhu, "Triply complementary priors for image restoration," *IEEE Trans. Image Process.*, vol. 30, pp. 5819–5834, 2021.
- [39] Z. Zha, B. Wen, X. Yuan, J. Zhou, C. Zhu, and A. C. Kot, "Low-rankness guided group sparse representation for image restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 2, 2022, doi: 10.1109/TNNLS.2022.3144630.
- [40] 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.
- [41] Y. Chen, J. Zeng, W. He, X. Zhao, and T. Huang, "Hyperspectral and multispectral image fusion using factor smoothed tensor ring decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5515417.
- [42] S. Yang, K. Zhang, and M. Wang, "Learning low-rank decomposition for pan-sharpening with spatial-spectral offsets," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3647–3657, Aug. 2018.
- [43] Y. Chen, T. Huang, W. He, X. 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, 2022, Art. no. 5515916.
- [44] W. He et al., "Non-local meets global: An iterative paradigm for hyperspectral image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2089–2107, Apr. 2022.
- [45] Y. Chen, W. He, N. Yokoya, and T. 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.
- [46] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.
- [47] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.
- [48] L. Zhuang, X. Fu, M. K. Ng, and J. M. Bioucas-Dias, "Hyperspectral image denoising based on global and nonlocal low-rank factorizations," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10438–10454, Dec. 2021.
- [49] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.
- [50] Y. Miao, X. Zhao, X. Fu, J. Wang, and Y. Zheng, "Hyperspectral denoising using unsupervised disentangled spatio-spectral deep priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5513916.
- [51] T. Yokota, H. Hontani, Q. Zhao, and A. Cichocki, "Manifold modeling in embedded space: An interpretable alternative to deep image prior," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1022–1036, Mar. 2022.
- [52] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3844–3851.
- [53] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, 2009.
- [54] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [55] 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, Sep. 2018.
- [56] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022.
- [57] Z. Cheng et al., "BIRNAT: Bidirectional recurrent neural networks with adversarial training for video snapshot compressive imaging," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer*, 2020, pp. 258–275.



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.

He is currently working with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, China. From 2018 to 2019, he was a Research Intern with the Geoinformatics unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. His current 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/>.



Xinfeng Gui received the B.S. degree from the Jiangxi Agricultural University, 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 current research interests include hyperspectral image processing and low-rank optimization.



Jinshan Zeng received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 2015.

He is currently a Distinguished Professor with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, China, and serves as the Director of the Department of Data Science and Big Data. He has authored more than 40 papers in high-impact journals and conferences such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), *Journal of Machine Learning Research*, IEEE TRANSACTIONS ON SIGNAL PROCESSING (TSP), International Conference on Machine Learning, and Association for the Advancement of Artificial Intelligence. He has coauthored two papers with collaborators that received the International Consortium of Chinese Mathematicians (ICCM) Best Paper Award in 2018 and 2020). His current research interests include nonconvex optimization, machine learning (in particular deep learning), and remote sensing.



Xi-Le Zhao 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 worked as a Post-Doctoral Researcher with Prof. Michael K. Ng at Hong Kong Baptist University, Hong Kong, from 2013 to 2014. He worked as a Visiting Scholar with Prof. Jose Bioucas Dias at the University of Lisbon, Lisbon, Portugal from 2016 to 2017. He is currently a Professor with the School of Mathematical Sciences, UESTC, Shenzhen, China. His current research interests include image processing, machine learning, and scientific computing. More information can be found on his homepage at: <https://zhaoxile.github.io/>



Wei He (Senior Member, IEEE) received the B.S. degree from the School of Mathematics and Statistics, Wuhan University, Wuhan, China, in 2012, and the Ph.D. degree in surveying, mapping, and remote sensing (LIESMARS) from Wuhan University, in 2017.

From 2018 to 2020, he was a Researcher at 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 at LIESMARS, Wuhan University. His current research interests include image quality improvement, remote sensing image processing and low-rank representation, and deep learning.