

Learning Tensor Low-Rank Prior for Hyperspectral Image Reconstruction

Shipeng Zhang¹ Lizhi Wang² Lei Zhang² Hua Huang^{3*}

¹Xi'an Jiaotong University ²Beijing Institute of Technology ³Beijing Normal University
zsp6869123@stu.xjtu.edu.cn, {wanglizhi, leizhang}@bit.edu.cn, huahuang@bnu.edu.cn

Abstract

Snapshot hyperspectral imaging has been developed to capture the spectral information of dynamic scenes. In this paper, we propose a deep neural network by learning the tensor low-rank prior of hyperspectral images (HSI) in the feature domain to promote the reconstruction quality. Our method is inspired by the canonical-polyadic (CP) decomposition theory, where a low-rank tensor can be expressed as a weight summation of several rank-1 component tensors. Specifically, we first learn the tensor low-rank prior of the image features with two steps: (a) we generate rank-1 tensors with discriminative components to collect the contextual information from both spatial and channel dimensions of the image features; (b) we aggregate those rank-1 tensors into a low-rank tensor as a 3D attention map to exploit the global correlation and refine the image features. Then, we integrate the learned tensor low-rank prior into an iterative optimization algorithm to obtain an end-to-end HSI reconstruction. Experiments on both synthetic and real data demonstrate the superiority of our method.

1. Introduction

Hyperspectral imaging systems capture the spectral information of the scene across tens to hundreds of discrete bands. The rich spectral details are beneficial to various computer vision tasks, such as face recognition [28], object tracking [16] and appearance modeling [26]. To obtain the 3D hyperspectral image (HSI), conventional hyperspectral imaging systems [4, 30, 33, 46] scan the scene with multiple exposures, which makes those systems cannot be used in the dynamic scenes. To this end, numerous snapshot hyperspectral imaging systems have been developed [7, 13, 27, 49] in the last few decades. Based on the compressive sensing theory [3, 6, 22], coded aperture snapshot spectral imaging (CASSI) [2, 36] draws an increasing attention due to the promising ability in capturing the dynamic target. However, the bottleneck of CASSI is the limited quality of reconstructing the 3D HSI from the 2D

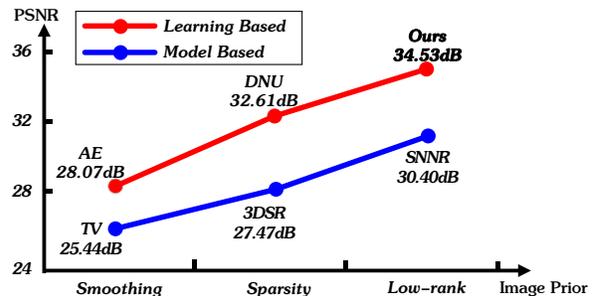


Figure 1: Development trends of the HSI priors. Model based methods and learning based methods show a similar development trend in both image priors and reconstruction accuracy (PSNR). We exploit the tensor low-rank prior with deep learning and obtain the best reconstruction result.

compressive measurement. Since this problem is under-determined, exploiting the image intrinsic prior is essential to regularize the reconstruction solution.

So far, two kinds of methods, i.e., model based methods and learning based methods, have been developed to exploit the image prior for HSI reconstruction and show a similar development trend, as shown in Figure 1.

In model based methods, numerous priors, from piecewise smoothing prior [42, 50], sparsity prior [35, 41] to low-rank prior [15, 24], are proposed to finish HSI reconstruction with conventional iterative optimization algorithms. Besides, different data structures, from 2D matrix to 3D tensor, have been employed to characterize those priors. Among these priors, the tensor low-rank prior stands out as a convincing way to model the high contextual correlation in the high-dimensional structure [52]. However, these priors equipped with conventional optimization algorithms suffer from low computational efficiency in practice.

In deep learning, image priors, including piecewise smoothing [12, 45] and sparsity prior [39, 47], are characterized by neural networks for HSI reconstruction. They usually exploit those priors in the feature domain. Methods in this venues have obtained promising reconstruction accuracy and efficiency. However, they do not take the low-rank prior into consideration, especially in tensor structure, which limits the improvement of reconstruction quality.

*Corresponding author

In nature images, structures in different frequencies carry discriminative structures. Actually, the feature maps in convolution networks can also be treated as a mixture of structures at different frequencies. Meanwhile, the features possess rich redundancy and high contextual correlation like the original image [11, 32]. An interesting and meaningful question arises: can the low-rank prior in tensor structure be explicitly characterized with the image features, like the prior regularization in model-based methods? This new prior, once realized, could promote HSI reconstruction with both accuracy and efficiency.

In this paper, we propose a novel deep neural network by exploiting the tensor low-rank prior of HSI in the feature domain to boost the reconstruction performance for snapshot hyperspectral imaging. Our method is inspired by the canonical-polyadic (CP) decomposition theory, where a low-rank tensor can be expressed as the weight summation of several rank-1 individual component tensors with discriminative structures. Specifically, we first characterize the tensor low-rank prior of the image features with two steps: (a) we generate rank-1 tensors with discriminative components to collect the contextual information from both spatial and channel dimensions of the image features; (b) we aggregate those rank-1 tensors into a low-rank tensor as a 3D attention map to exploit the global correlation and refine the image features. Then, we integrate the learned tensor low-rank prior into the half quadratic splitting (HQS) based optimization algorithm to obtain an end-to-end HSI reconstruction. Experiments on both synthetic and real data demonstrate the superiority of our method.

In summary, we make the following contributions:

- (1). We propose a deep neural network to characterize the tensor low-rank prior of HSI in the feature domain.
- (2). We characterize the tensor low-rank prior by exploiting the contextual correlation of the image features with CP decomposition.
- (3). We obtain an end-to-end HSI reconstruction by integrating the image prior into the HQS algorithm.

2. Related Works

2.1. Model Based HSI Prior

According to the compressive sensing theory, the desired signal can be reconstructed by solving a prior regularized convex optimization problem. Researchers attempted to model the intrinsic property of HSI with various prior regularization forms. The TV regularization was first introduced to keep the piecewise smoothing of HSI, but the reconstructed results tend to be over-smooth [40, 48, 50, 51]. The sparsity prior maintains the sparse property of HSI in a fixed transform domain or an over-completed dictionary

and derives a more robust performance than TV [14, 36]. Tan et al. [35] used approximate message passing to pursue the sparsity of HSI with the wavelet transform. Wang et al. [41] proposed to learn the 3D adaptive nonlocal sparse representation (3DNSR) of HSI. Then low-rank matrix approximation (LRMA) was proposed in [15, 24] to exploit the non-local correlation. However, these methods always expressed the 3D HSI as a 1D vector or a 2D matrix, which inevitably breaks the high-dimensional nature of HSI. Recently, Zhang et al. [52] proposed the low-rank tensor recovery based on the sum of nuclear norm regularization (SNNR) and obtained impressive accuracy. However, solving this problem with convex optimization is not sufficient to fit the data diversity and suffers from low efficiency. In this paper, we design a deep neural network to exploit the tensor low-rank prior and improve the reconstruction performance of snapshot hyperspectral imaging.

2.2. Learning Based HSI Prior

By leveraging the advantage of deep learning, researchers characterized the image priors with the deep neural networks to finish HSI reconstruction. Xiong et al. [45] proposed HSCNN to treat the HSI reconstruction as an image enhancement task and learn a brute-force mapping from the initialized image. Choi et al. [12] utilized auto-encoder (AE) to learn the non-linear sparsity prior and the TV based smoothing prior of HSI. Zhang and Ghanem [47] exploited the deep sparsity prior with an interpretable optimization-inspired deep network (ISTA) for natural image compressive sensing. Recently, attention mechanism based methods were proposed to explore the non-local similarity prior of HSI. Miao et al. [25] proposed λ -Net for CASSI reconstruction by combining the spatial attention block and U-net. Wang et al. [39] developed a deep non-local unrolling (DNU) method by integrating a spatial attention block with a local sparsity block to further boost the reconstruction accuracy. However, they failed to consider the low-rank prior of HSI, especially in tensor structure, which limited the improvement of reconstruction quality. In this work, we characterize the tensor low-rank prior of HSI in the feature domain to promote the reconstruction accuracy for snapshot hyperspectral imaging.

2.3. Low-rank Tensor Recovery

Low-rank tensor recovery tries to estimate the desired tensor with the lower-rank constraint from the degraded input using various tensor decomposition models. The most widely used decomposition models are Tucker decomposition and CP decomposition [19]. Tucker decomposition expresses the input tensor as the product of several matrices and a core tensor. CP decomposition expresses the input tensor as a weighted combination of few rank-1 tensors. The rank-1 tensors contain discriminative

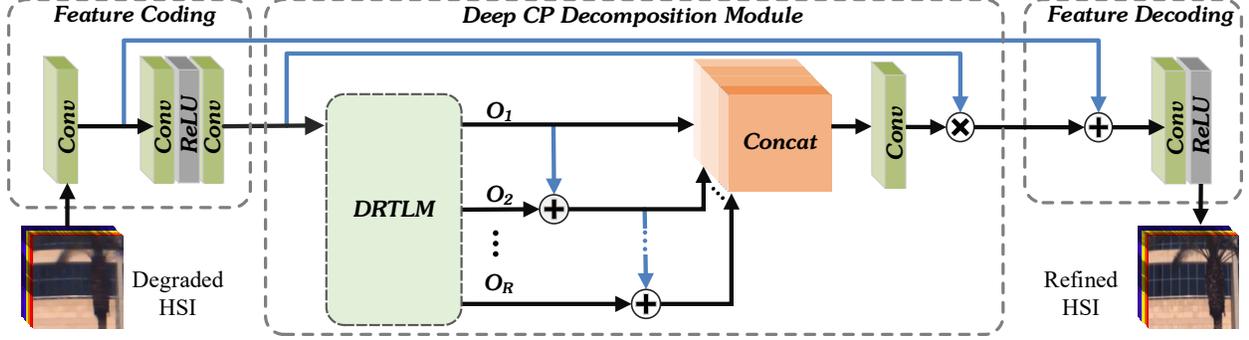


Figure 2: The framework of the tensor low-rank prior learning network (TLPLN). We input the degraded image and output the refined image after implementing three parts, i.e., feature coding, the deep CP decomposition module and feature decoding.

components with different importance. CP decomposition is an efficient tool for delivering the low-rank nature of data and has been widely used in various computer vision tasks [9, 10, 21, 29, 44]. In this paper, we leverage CP decomposition to model the tensor low-rank prior of HSI in the feature domain with deep learning.

3. Tensor Low-rank Prior Learning

The feature maps in convolution networks can be treated as a linear projection of the original image and delivers different information of various frequencies. For example, the features in higher frequencies carry structure details and those in lower frequencies carry smoothing information. Meanwhile, rich redundancy and high contextual correlation also exist in the image features like the original image [11, 32]. This key observation motivates us to model the tensor low-rank prior of HSI in the feature domain.

To this end, we design a deep neural network to learn the tensor low-rank prior. Meanwhile, a deep CP decomposition module is developed to collect the contextual information and refine the image features.

3.1. Prior Learning Network

Our tensor low-rank prior learning network, as shown in Figure 2, consists of three steps: feature coding, the deep CP decomposition module, and feature decoding.

Before implementing the deep CP decomposition module, we first employ several neural layers to code the image features of the input degraded HSI. Specifically, we use convolution (Conv)-Conv-ReLU-Conv to generate the features of the input image. During the feature coding, we fix the spatial kernel size as 3×3 and the channel number as 64. Then, we deliver the coded image features into the proposed deep CP decomposition module to restore the refined features. Finally, we employ Conv-ReLU with the spatial kernel size of 3×3 to decode the refined features into the improved HSI. Next, we introduce the design of the deep CP decomposition module in detail.

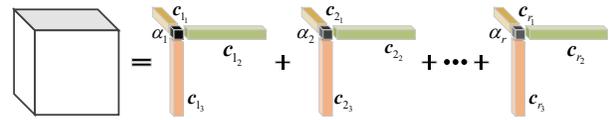


Figure 3: A visual display of 3D CP decomposition.

3.2. Deep CP Decomposition Module

We choose CP decomposition as the pipeline to learn the tensor low-rank prior. The definition of CP decomposition is given as the following theorem, and a visual display of the CP decomposition for 3D tensor is shown in Figure 3.

Theorem 1 For an arbitrary N -order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, it can be decomposed into the linear combination of a series of Kronecker bases, written as:

$$\mathcal{X} = \sum_{i=1}^r \alpha_i c_{i_1} \circ c_{i_2} \circ \dots \circ c_{i_N}, \quad (1)$$

where $c_{i_n} \in \mathbb{R}^{I_n}$ is a rank-1 Kronecker basis vector, α_i is the scalar weight, \circ denotes Kronecker product, and r is a predefined number called CP rank. Generally, $c_{i_1} \circ c_{i_2} \circ \dots \circ c_{i_N}$ produces a rank-1 tensor.

According to Theorem 1, CP decomposition can be regarded as a high-dimensional extension of image singular value decomposition, where the component bases with different importance contain distinctive structures of different frequencies. Therefore, CP decomposition can characterize discriminative information of different frequencies in the image features effectively.

The key of CP decomposition is to learn the discriminative rank-1 tensor components and aggregate them into a low-rank tensor [19]. Specifically, we first design a sub block to generate a single rank-1 tensor. Given the input image features, we apply global average pooling (GAP)-Conv-Sigmoid on the channel, height, and width dimension [10] in parallel to generate the rank-1 vector c_{r_1} , c_{r_2} and c_{r_3} . Specifically, GAP is used to deliver the contextual information across the corresponding dimension and Conv-Sigmoid

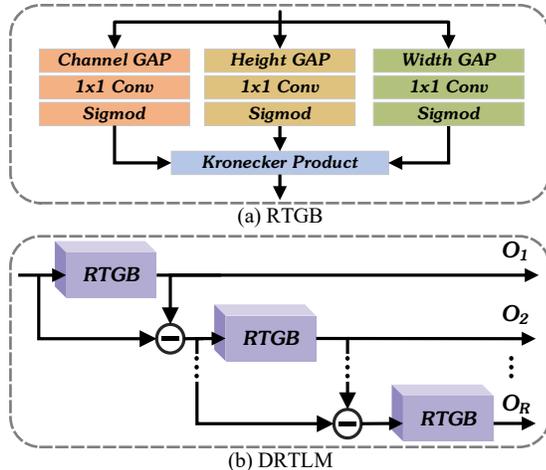


Figure 4: Architecture of the rank-1 tensor generating block (RTGB) and the discriminative rank-1 tensor learning module (DRTLML). The DRTLML embed the RTGBs to produce discriminative rank-1 component tensors.

is employed for generating the non-linear projection from the result of GAP to the desired vector. Once the rank-1 vectors are obtained, we impose the Kronecker product on the rank-1 vector c_{r_1} , c_{r_2} and c_{r_3} to generate the rank-1 tensor \mathcal{O}_r , which delivers the contextual information in spatial dimension and channel dimension jointly. We denote the above process as the rank-1 tensor generating block, as shown in Figure 4 (a).

Then, we introduce residual learning to learn the discriminative rank-1 tensors of different frequencies. Residual learning is a powerful tool to learn the image information of different frequencies [31]. Specifically, we first employ a rank-1 tensor generating block on the input features to generate the rank-1 tensor \mathcal{O}_1 . Then we extract the residual part, which can be regarded as the information with higher frequency that \mathcal{O}_1 fails to restore. The residual part is consequently processed by a rank-1 tensor generating block again to generate \mathcal{O}_2 . Finally, given the predefined CP rank r , we embed the rank-1 generating blocks for r times to generate r rank-1 tensors $\{\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_r\}$. Each rank-1 generating block is responsible for producing a tensor with the individual contextual information and passing the residual part into the next rank-1 generating block. We denote this process as the discriminative rank-1 tensor learning module, as shown in Figure 4 (b).

Once the rank-1 component tensors are available, we aggregate them into a low-rank tensor. Specifically, we first concatenate r rank-1 tensors into a stacked tensor. Since the contextual information in a higher frequency is more difficult to restore, we add the skip connection from the lower rank-1 tensor before concatenation. The skip connection can faithfully deliver valuable information that is easier to be restored, which benefits avoiding the over-fitting and

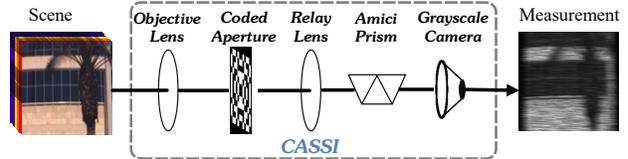


Figure 5: Diagram of the CASSI system.

boosting the back-propagation. Then, we employ a Conv layer with the spatial size of 3×3 to aggregate the stacked tensor into the desired low-rank tensor, which can be equivalent to learn the weights α_r in CP decomposition. Accordingly, the restored low-rank tensor \mathcal{X}_{lr} based on CP decomposition can be expressed as:

$$\mathcal{X}_{lr} = \text{Conv}(\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_r). \quad (2)$$

It is also worth noting that aggregating rank-1 tensors with weighted summation in Eq. 1 gives the same weights along different dimensions within a rank-1 tensor, which ignores the non-local correlation between different dimensions. Instead, we employ a Conv layer to learn the distinctive weights along different dimensions and produce a more robust aggregation result. Since the aggregated tensor contains rich contextual information, it can be regarded as a 3D attention map of modeling the global correlation from different dimensions and different frequencies. Therefore, we finally refine the image features using the Hadamard product on the aggregated tensor and the input features.

In the following, we leverage the learned prior to finish reconstruction of snapshot hyperspectral imaging.

4. HSI Reconstruction

4.1. Snapshot Imaging System

We give a brief introduction on the representative snapshot hyperspectral imaging system, i.e., CASSI. Other imaging systems, such as multiple snapshot hyperspectral system [18] and spatial-spectral encoded imaging system [23], are also applicable for our method.

The schematic diagram of CASSI is shown in Figure 5. The incident spectral information is first projected on a coded aperture through an objective lens. Then the light is spatially modulated by the pre-given pattern of the coded aperture and spectrally dispersed by a dispersive prism. Finally, the light arrives on the focal plane of a grayscale camera and forms a 2D compressive measurement. We denote the original HSI as $\mathcal{X} \in \mathbb{R}^{M \times N \times \Lambda}$ and its element as $\mathcal{X}(m, n, \lambda)$, where m, n index the spatial coordinate and λ indexes the spectral coordinate. Accordingly, the intensity $Y(m, n)$ of the CASSI measurement $\mathbf{Y} \in \mathbb{R}^{(M+\Lambda-1) \times N}$ at position (m, n) is

$$Y(m, n) = \sum_{\lambda=1}^{\Lambda} \rho(\lambda) \varphi(m - \lambda, n) \mathcal{X}(m - \lambda, n, \lambda), \quad (3)$$

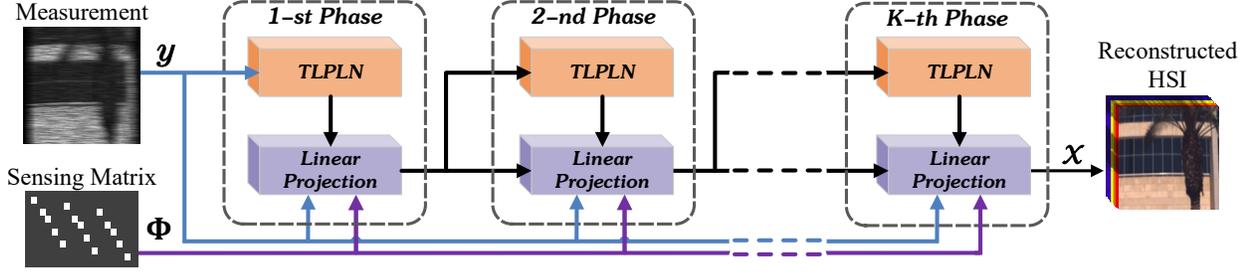


Figure 6: Flowsheet of our proposed reconstruction network. We first initialize HSI from the input sensing matrix and measurement. Then the TLPLN and linear projection work alternatively in each phase. We iterate the procedure for K times and finally output the reconstructed HSI.

where $\rho(\lambda)$ is the spectral response of the camera and $\varphi(m, n)$ is the pattern of the coded aperture. For brevity, the imaging process can be represented as the following mathematical formulation:

$$\mathbf{y} = \Phi \mathbf{x}, \quad (4)$$

where $\Phi \in \mathbb{R}^{((M+\Lambda-1) \times N) \times MNA}$ is the sensing matrix [2], $\mathbf{x} = \text{Vec}(\text{Vec}(\mathcal{X}(:, :, 1)), \text{Vec}(\mathcal{X}(:, :, 2)), \dots, \text{Vec}(\mathcal{X}(:, :, \Lambda)))$ and $\mathbf{y} = \text{Vec}(\mathbf{Y})$.

4.2. Reconstruction Network

We introduce the deep interpretable optimization framework [34, 47] to reconstruct HSI by integrating the tensor low-rank prior learning network with the HQS algorithm.

The fundamental task of HSI reconstruction is to recover the underlying image \mathbf{x} with the given sensing matrix Φ and the captured measurement \mathbf{y} . According to the compressive sensing theory, the problem of HSI reconstruction with the tensor-low rank prior based regularization can be described as the following convex optimization formulation:

$$\arg \min_{\mathcal{X}} R(\mathcal{X}), \quad \text{s.t. } \mathbf{y} = \Phi \mathbf{x}, \quad (5)$$

where $R(\cdot)$ denotes the learned tensor low-rank prior based regularization. The above constrained minimization problem is equivalent to its Lagrangian form:

$$\arg \min_{\mathcal{X}} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \beta R(\mathcal{X}), \quad (6)$$

Here we utilize the HQS algorithm to optimize the reconstruction problem in Eq. 6. Specifically, after introducing an auxiliary variable $\mathcal{S} = \mathcal{X}$, Eq. 6 can be reformulated into a constrained optimization problem as:

$$\arg \min_{\mathcal{X}, \mathcal{S}} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \beta R(\mathcal{S}), \quad \text{s.t. } \mathcal{S} = \mathcal{X}. \quad (7)$$

The above problem can be converted to a non-constrained optimization problem:

$$\arg \min_{\mathcal{X}, \mathcal{S}} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \beta R(\mathcal{S}) + \tau \|\mathcal{S} - \mathcal{X}\|_F^2, \quad (8)$$

where τ is the penalty factor. Then, \mathcal{X} and \mathcal{S} can be estimated by solving the two sub-problems alternatively:

$$\mathcal{S}^{t+1} = \arg \min_{\mathcal{S}} R(\mathcal{S}) + \frac{\tau}{\beta} \|\mathcal{S} - \mathcal{X}^t\|_F^2, \quad (9)$$

$$\mathcal{X}^{t+1} = \arg \min_{\mathcal{X}} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \tau \|\mathcal{S}^{t+1} - \mathcal{X}\|_F^2. \quad (10)$$

\mathcal{S} sub-problem. Eq. 9 turns to optimize the learned tensor low-rank prior based regularized problem. We denote the tensor low-rank prior learning network as a proximal operator $\Upsilon(\cdot)$, so the solution can be expressed as:

$$\mathcal{S}^{t+1} = \Upsilon(\mathcal{X}^t). \quad (11)$$

\mathcal{X} sub-problem. Eq. 10 is a quadratic optimization problem and admits a straightforward least-square solution:

$$\mathcal{X}^{t+1} = (\Phi^T \Phi + \tau \mathbf{I})^{-1} (\Phi^T \mathbf{y} + \tau \mathcal{S}^{t+1}). \quad (12)$$

Here we employ the gradient descent algorithm to obtain an approximate solution with linear projection given as:

$$\mathcal{X}^{t+1} = \hat{\Phi}(\mathcal{X}^t) + \varepsilon \Phi^T \mathbf{y} + \varepsilon \tau \mathcal{S}^{t+1}, \quad (13)$$

where $\hat{\Phi} = (1 - \varepsilon \tau) \mathbf{I} - \varepsilon \Phi^T \Phi$.

We integrate the two updating steps, i.e., Eq. 11 and Eq. 13, into an end-to-end manner and obtain the proposed reconstruction network, as shown in Figure 6. Specifically, we first initial $\mathcal{X}^0 = \Phi^T \mathbf{y}$. Then, the tensor low-rank prior learning network and the linear projection are performed alternatively in each phase for K times. Finally, the network outputs the reconstructed HSI.

4.3. Learning Strategy

Learning the parameters from the external datasets is essential for deep neural networks. In the reconstruction algorithm, the hyper-parameters τ and ε leave an important effect on accuracy. However, tuning hyper-parameters is always challenging and nontrivial for different scenes. Toward this end, we set the two hyper-parameters $\Theta_h = (\tau, \varepsilon)$

Table 1: Evaluation results and reconstruction time (in second) on two datasets. The best results are highlighted in bold.

Dataset	Index	TV	3DNSR	LRMA	SNNR	HSCNN	AE	ISTA	λ -Net	DNU	Ours
ICVL	PSNR	25.442	27.472	29.793	30.400	28.446	28.068	30.498	29.006	32.610	34.530
	SSIM	0.906	0.914	0.940	0.943	0.934	0.923	0.947	0.946	0.966	0.977
	SAM	0.051	0.047	0.032	0.032	0.043	0.038	0.044	0.053	0.037	0.030
	ERGAS	89.950	70.665	56.328	50.605	61.557	68.668	49.381	56.501	37.531	30.065
Harvard	PSNR	26.228	28.638	30.499	31.136	27.603	29.205	29.870	29.373	31.111	32.433
	SSIM	0.889	0.903	0.930	0.932	0.895	0.912	0.913	0.909	0.929	0.941
	SAM	0.113	0.108	0.078	0.080	0.109	0.091	0.110	0.133	0.101	0.090
	ERGAS	127.482	97.269	84.133	74.915	105.897	91.625	85.211	89.015	73.533	62.510
Running Time(s)		150	3240	3862	6942	0.78	130	0.29	0.07	0.25	0.36

also as learnable variables to balance the contribution between the data fidelity and the deep prior based regularization adaptively. It should be noted that our tensor low-rank prior learning network is trained directly during the end-to-end reconstruction process. The prior learning network in different phases will handle degraded images with various noise levels and obtain adaptive parameters.

We introduce the MSE loss as the evaluation criteria. Consequently, our training loss can be expressed as:

$$\mathcal{L}(\Theta_n, \Theta_h) = \frac{1}{L} \sum_{l=1}^L \left\| F(\mathbf{y}^l, \Phi, \Theta_n, \Theta_h) - \mathbf{x}^l \right\|^2, \quad (14)$$

where Θ_n is the set of all network parameters, $F(\cdot)$ denotes the reconstructed output, \mathbf{x}^l and \mathbf{y}^l denote the paired HSI and the corresponding measurement, respectively.

In our work, we employ TensorFlow as the running framework and use Adam [17] to minimize the loss function. The learning rate, the penalty τ and the decent step ε are initially set as 0.0001, 0.8 and 0.04, respectively. Meanwhile, the learning rate exponentially decays to 95% for every ten epochs. We set the CP rank as 4 and the max phase number K as 11 empirically. We execute our implementation on a platform with Inter i7 6700 and NVIDIA 1080TI.

5. Experiments on Synthetic Data

5.1. Implementation Details

Datasets. We evaluate our method on two benchmark datasets, including the ICVL dataset [1] and the Harvard dataset [8]. The ICVL dataset consists of 201 real-world objects, each with 1300×1392 spatial resolution and 31 spectral bands collected from 400nm to 700 nm in a 10nm step. The Harvard dataset consists of 50 outdoor scenes, each with 512×512 spatial resolution and 31 spectral bands collected from 420nm to 720nm in a 10nm step. We set the patch size as 48×48 across full spectral bands. Meanwhile, we randomly collect 50 images in the ICVL dataset and 9 images in the Harvard dataset for testing and the rest for training. In our experiment, the resolution of all tested images is cropped into 256×256 .

Comparison Methods. We compare our method on synthetic data with the state-of-the-art reconstruction methods, including 4 model based methods, i.e., TV [5], 3DNSR [41], LRMA [15] and SNNR [52], and 5 learning based methods, i.e., HSCNN [12], AE [12], ISTA [47], λ -Net [25] and DNU[38]. We make great efforts to produce their best results according to their codes released publicly or provided by the authors privately.

Evaluation Indexes. We adopt four image quality indexes, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM) [43], spectral angle mapping (SAM) [20] and relative dimensionless global error in synthesis (ERGAS) [37], for quantitative evaluation. Specifically, PSNR measures the visual quality, SSIM measures the structure similarity, while SAM and ERGAS measure the spectral fidelity. Generally, bigger values of PSNR and SSIM, smaller results of SAM and ERGAS suggest a better reconstruction accuracy.

5.2. Evaluation Results

Table 1 shows the average numerical results on the ICVL data and the Harvard dataset. The best results for each index are highlighted in bold. We can see that our method outperforms all the existing methods on both spatial accuracy and spectral fidelity, which demonstrates the superiority by learning the tensor low-rank prior with the deep neural network. Specifically, compared with the model based methods, the proposed deep prior can better capture the intrinsic characteristic of HSI. In particular, the improvement upon SNNR indicates that using the deep neural network is more sufficient than the model based method to deliver the tensor low-rank property of HSI and fit the data diversity. Our method can also produce a remarkable improvement upon the learning based methods. The boost upon HSCNN AE and ISTA states clearly that integrating the tensor low-rank prior is more conducive for HSI reconstruction than other learning based priors. The promotion upon λ -Net and DNU demonstrates that the proposed tensor low-rank prior is more powerful to exploit the contextual correlation

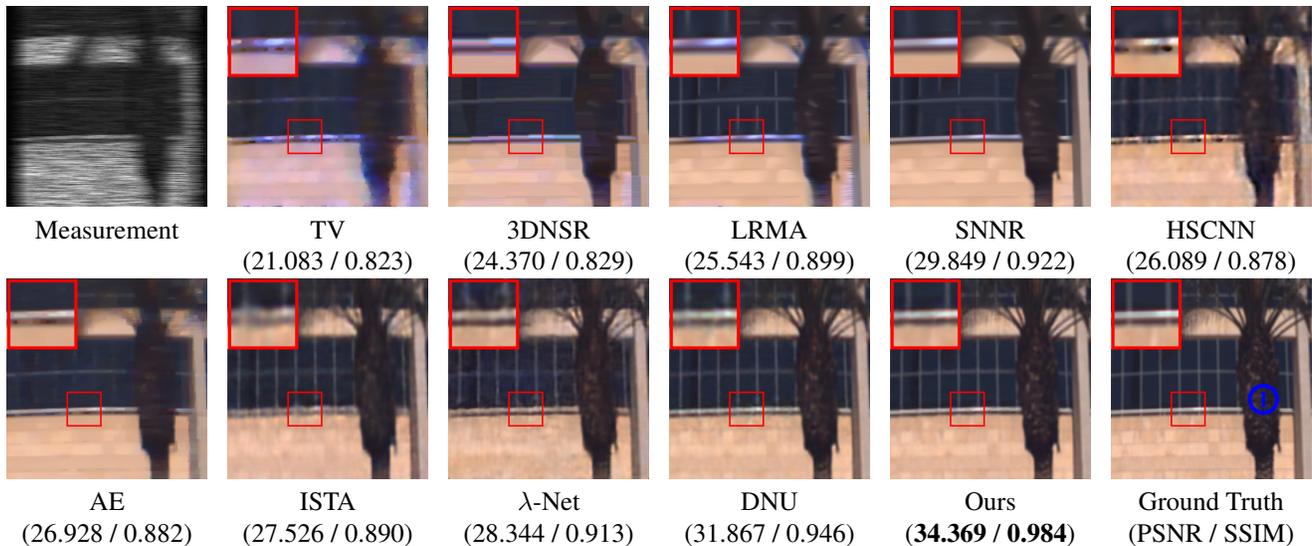


Figure 7: Visual quality comparison of one example scene from the ICVL dataset. The PSNR and SSIM results for the reconstructed images are shown in the parenthesis. Our method obtains better spatial contents and textures.

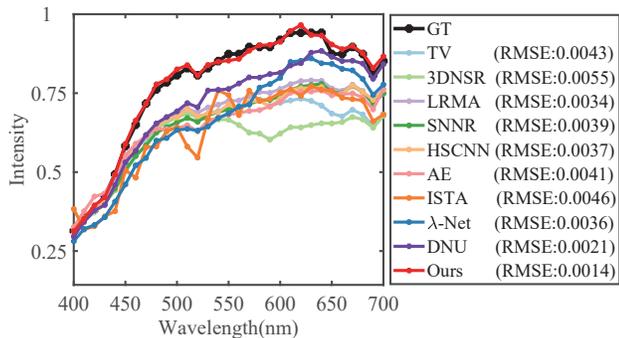


Figure 8: Spectral curves and their reconstructed RMSE results of the region labeled in Figure 7.

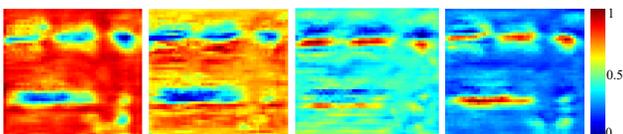


Figure 9: Visualization of four sub-attention maps.

of HSI. The last row in Table 1 lists the average running time (in seconds). We can see that the proposed method can obtain a comparable efficiency with other learning based methods, such as HSCNN and ISTA.

To give a visual evaluation, we display the reconstructed images of one representative scene from the ICVL dataset in Figure 7. For better vision, we convert the reconstructed HSI into RGB image using the CIE color mapping function. By comparing the full images and the local magnification maps, we can see that more spatial details and sharper textures are produced in our method. We display the spectral curves of the tree, which is labeled in Figure 7, and

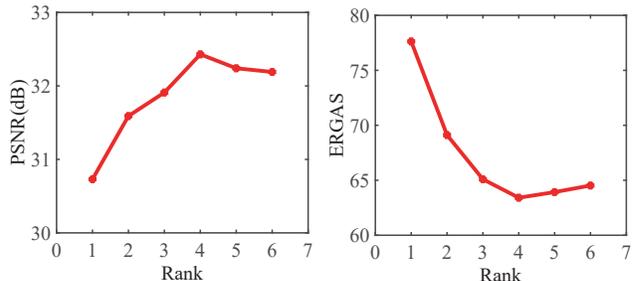


Figure 10: The reconstructed spatial accuracy (PSNR) and spectral fidelity (ERGAS) under different CP ranks.

the reconstructed RMSE results in Figure 8. It shows that our method provides a lower spectral error than other methods. Therefore, our method outperforms all the competitive methods in terms of both spatial and spectral accuracy.

We further visualize the heat maps of the learned attention map of the image in Figure 7 to check the part of features they activate. As shown in Figure 9, different sub-attention maps activate discriminative structures. For example, the first map activates most area of the feature, which is the smoothing information with lower frequencies. While the last map activates few small areas, which is the structure details. Therefore, the deep CP decomposition module can capture contextual information of different frequencies.

5.3. Ablation Study

We conduct extended experiments to investigate the effects of parameters and validate the proposed motivation.

CP rank. The CP rank determines reconstruction quality and plays an important role in our model. Figure 10 shows the average results on the Harvard dataset with different CP

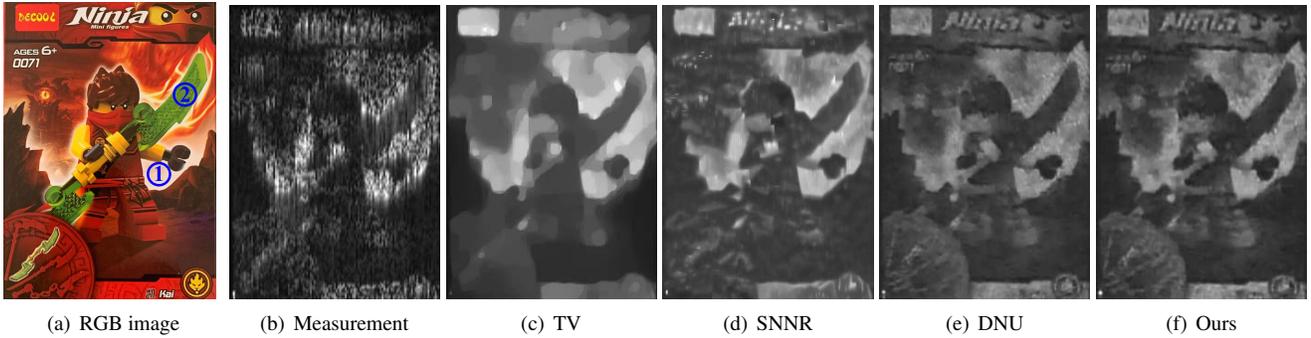


Figure 11: Performance comparison on real data.

Table 2: Accuracy comparison under different methods of two essential operations in our model.

Residual Learning	Aggregation Manner	PSNR	ERGAS
×	Convolution Learning	32.160	64.370
✓	Weighted Summation	31.320	71.494
✓	Convolution Learning	32.433	62.510

ranks. We can see that the best result is obtained when rank = 4, which validates the effectiveness of aggregating the contextual information with a low-rank constraint.

Essential Operations. Generating discriminative rank-1 tensors with residual learning and aggregating rank-1 tensors into a low-rank tensor with convolution learning are two essential operations in our model. We conduct the experiments when we generate rank-1 tensors without residual learning, and when we implement aggregation with weighted summation to validate the effectiveness of our design. The results in Table 2 demonstrate that generating discriminative rank-1 tensors with residual learning and aggregating the rank-1 tensors with convolution learning can produce more accurate results in CP decomposition.

6. Experiments on Real Data

We further evaluate the performance of the proposed method on the real captured data. The captured scene is a cartoon cover under the laboratory ambient light condition. The RGB image and the CASSI compressive measurement are shown in Figure 11 (a) and (b), respectively. To generate the training data, we utilize the spectral interpolation on the synthetic dataset according to the corresponding wavelength. We display the reconstructed images at 648nm of TV, SNNR, DNU, and our method in Figure 11 (c)-(f). It shows that our method obtains better results with clearer contents and textures compared with other methods.

The reconstructed spectral curves of the representative regions, which are labeled as ① and ② in Figure 11, are shown in Figure 12, and the corresponding RMSE results are also attached in parentheses. We use a scanning-based commercial spectrometer (SOC 710) to capture the refer-

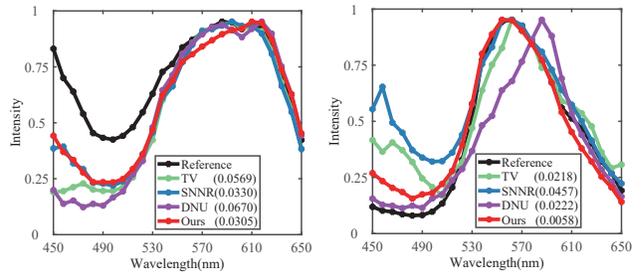


Figure 12: Spectra results (from left to right) of the regions labeled as ① and ② in Figure 11, respectively.

ence spectrum. It shows that the spectral curve reconstructed by our method is closer to the reference compared with other methods. Meanwhile, our method obtained the least error according to the RMSE results, which demonstrates the superiority of our method in real data.

7. Conclusion

In this paper, we proposed a deep neural network by characterizing the tensor low-rank prior of HSI in the feature domain to boost the reconstruction performance for snapshot hyperspectral imaging. Specifically, we first characterized the tensor low-rank prior of the image features using CP decomposition with two steps: (a) we generated rank-1 tensors with discriminative components to collect the contextual information from both spatial and channel dimensions; (b) we aggregated those rank-1 tensors into a low-rank tensor as a 3D attention map to exploit the global correlation and refine the image features. Then, we integrated the learned tensor low-rank prior into an iterative optimization algorithm to obtain an end-to-end HSI reconstruction. Experiments on both synthetic and real data demonstrated the superiority of our method.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant 62072038 and Grant 61922014.

References

- [1] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34, 2016. [6](#)
- [2] Gonzalo R. Arce, David J. Brady, Lawrence Carin, Henry Arguello, and David S. Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2014. [1](#), [5](#)
- [3] Richard G. Baraniuk. Compressive sensing [lecture notes]. *IEEE Signal Processing Magazine*, 24(4):118–121, 2007. [1](#)
- [4] Robert W. Basedow, Dwayne C. Carmer, and Mark E Anderson. Hydice system: Implementation and performance. In *Imaging Spectrometry*, volume 2480, pages 258–268, 1995. [1](#)
- [5] J. M. Bioucas-Dias and M. A. T. Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004, 2007. [6](#)
- [6] Emmanuel J. Candes. Compressive sampling. In *Proceedings of the International Congress of Mathematicians*, volume 3, pages 1433–1452, 2006. [1](#)
- [7] Xun Cao, Hao Du, Xin Tong, Qionghai Dai, and Stephen Lin. A prism-mask system for multispectral video acquisition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2423–2435, 2011. [1](#)
- [8] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–200, 2011. [6](#)
- [9] Yi Chang, Luxin Yan, and Sheng Zhong. Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5901–5909, 2017. [3](#)
- [10] Wanli Chen, Xinge Zhu, Ruoqi Sun, Junjun He, Ruiyu Li, Xiaoyong Shen, and Bei Yu. Tensor low-rank reconstruction for semantic segmentation. In *European Conference on Computer Vision*, pages 52–69, 2020. [3](#)
- [11] Yunpeng Chen, Haoqi Fan, Bing Xu, Zhicheng Yan, Yan-nis Kalantidis, Marcus Rohrbach, Shuicheng Yan, and Jiashi Feng. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In *IEEE International Conference on Computer Vision*, pages 3435–3444, 2019. [2](#), [3](#)
- [12] Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics*, 36(6):218, 2017. [1](#), [2](#), [6](#)
- [13] Michael Descour and Eustace Dereniak. Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. *Applied Optics*, 34(22):4817–26, 1995. [1](#)
- [14] Mário AT Figueiredo, Robert D. Nowak, and Stephen J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):586–597, 2007. [2](#)
- [15] Ying Fu, Yinqiang Zheng, Imari Sato, and Yoichi Sato. Exploiting spectral-spatial correlation for coded hyperspectral image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3727–3736, 2016. [1](#), [2](#), [6](#)
- [16] Kate S. He, Duccio Rocchini, Markus Neteler, and Harini Nagendra. Benefits of hyperspectral remote sensing for tracking plant invasions. *Diversity and Distributions*, 17(3):381–392, 2011. [1](#)
- [17] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. [6](#)
- [18] D. Kittle, K Choi, A Wagadarikar, and D. J. Brady. Multi-frame image estimation for coded aperture snapshot spectral imagers. *Applied Optics*, 49(36):6824, 2010. [4](#)
- [19] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009. [2](#), [3](#)
- [20] Fred A. Kruse, AB Lefkoff, JW Boardman, KB Heidebrecht, AT Shapiro, PJ Barloon, and AFH Goetz. The spectral image processing system (sips)-interactive visualization and analysis of imaging spectrometer data. In *AIP Conference Proceedings*, volume 283, pages 192–201, 1993. [6](#)
- [21] Chang Li, Yong Ma, Jun Huang, Xiaoguang Mei, and Jiayi Ma. Hyperspectral image denoising using the robust low-rank tensor recovery. *Journal of the Optical Society of America A*, 32(9):1604–1612, 2015. [3](#)
- [22] Zhetao Li, Jingxiong Xie, Gengming Zhu, Xin Peng, Yanrong Xie, and Youngjune Choi. Block-based projection matrix design for compressed sensing. *Chinese Journal of Electronics*, 25(3):551–555, 2016. [1](#)
- [23] Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics*, 33(6):1–11, 2014. [4](#)
- [24] Yang Liu, Xin Yuan, Jinli Suo, David Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2990–3006, 2018. [1](#), [2](#)
- [25] Xin Miao, Xin Yun, Yunchen Pu, Athitsos, and Vassilis Athitsos. lambda-net: Reconstruct hyperspectral images from a snapshot measurement. In *IEEE International Conference on Computer Vision*, pages 4058–4068, 2019. [2](#), [6](#)
- [26] H. Kim Min, Todd Alan Harvey, David S. Kittle, Holly Rushmeier, Julie Dorsey, Richard O. Prum, and David J. Brady. 3d imaging spectroscopy for measuring hyperspectral patterns on solid objects. *ACM Transactions on Graphics*, 31(4):1–11, 2012. [1](#)
- [27] Takayuki Okamoto and Ichirou Yamaguchi. Simultaneous acquisition of spectral image information. *Optics Letters*, 16(16):1277–1279, 1991. [1](#)
- [28] Zhihong Pan, G. Healey, M. Prasad, and B. Tromberg. Face recognition in hyperspectral images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1552–1560, 2003. [1](#)
- [29] Yi Peng, Deyu Meng, Zongben Xu, Chenqiang Gao, Yi Yang, and Biao Zhang. Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2949–2956, 2014. [3](#)
- [30] Wallace M. Porter and Harry T. Enmark. A system overview of the airborne visible/infrared imaging spectrometer (aviris). In *Imaging Spectroscopy II*, volume 834, pages 22–32, 1987. [1](#)
- [31] Yajun Qiu, Ruxin Wang, Dapeng Tao, and Jun Cheng.

- Embedded block residual network: A recursive restoration model for single-image super-resolution. In *IEEE International Conference on Computer Vision*, pages 4180–4189, 2019. 4
- [32] Joanna Rownicka, Peter Bell, and Steve Renals. Multi-scale octave convolutions for robust speech recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7019–7023, 2020. 2, 3
- [33] Yoav Y. Schechner and Shree K. Nayar. Generalized mosaicing: Wide field of view multispectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1334–1348, 2002. 1
- [34] Jian Sun, Huibin Li, Zongben Xu, et al. Deep admn-net for compressive sensing mri. In *Advances in Neural Information Processing Systems*, pages 10–18, 2016. 5
- [35] Jin Tan, Yanting Ma, Hoover Rueda, Dror Baron, and Gonzalo R Arce. Compressive hyperspectral imaging via approximate message passing. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):389–401, 2015. 1, 2
- [36] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, 47(10):B44–B51, 2008. 1, 2
- [37] Lucien Wald. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002. 6
- [38] Lizhi Wang, Chen Sun, Ying Fu, Min H Kim, and Hua Huang. Hyperspectral image reconstruction using a deep spatial-spectral prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8032–8041, 2019. 6
- [39] Lizhi Wang, Chen Sun, Maoqing Zhang, Ying Fu, and Hua Huang. Dnu: Deep non-local unrolling for computational spectral imaging. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1661–1671, 2020. 1, 2
- [40] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, and Feng Wu. Dual-camera design for coded aperture snapshot spectral imaging. *Applied Optics*, 54(4):848–858, 2015. 2
- [41] Lizhi Wang, Zhiwei Xiong, Guangming Shi, Feng Wu, and Wenjun Zeng. Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(10):2104–2111, 2017. 1, 2, 6
- [42] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008. 1
- [43] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6
- [44] Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(8):1888–1902, 2018. 3
- [45] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *IEEE International Conference on Computer Vision Workshops*, pages 518–525, 2017. 1, 2
- [46] Masahiro Yamaguchi, Hideaki Haneishi, Hiroyuki Fukuda, Junko Kishimoto, Hiroshi Kanazawa, Masaru Tsuchida, Ryo Iwama, and Nagaaki Ohyama. High-fidelity video and still-image communication based on spectral information: Natural vision system and its applications. In *Spectral Imaging: Eighth International Symposium on Multispectral Color Science*, volume 6062, page 60620G, 2006. 1
- [47] Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1828–1837, 2018. 1, 2, 5, 6
- [48] Maoqing Zhang, Lizhi Wang, Lei Zhang, and Hua Huang. Compressive hyperspectral imaging with non-zero mean noise. *Optics Express*, 27(13):17449–17462, 2019. 2
- [49] Maoqing Zhang, Lizhi Wang, Lei Zhang, and Hua Huang. High light efficiency snapshot spectral imaging via spatial multiplexing and spectral mixing. *Optics Express*, 28(14):19837–19850, 2020. 1
- [50] Shipeng Zhang, Hua Huang, and Ying Fu. Fast parallel implementation of dual-camera compressive hyperspectral imaging system. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(11):3404–3414, 2018. 1, 2
- [51] Shipeng Zhang, Lizhi Wang, Ying Fu, and Hua Huang. Gpu assisted towards real-time reconstruction for dual-camera compressive hyperspectral imaging. In *Pacific Rim Conference on Multimedia*, pages 711–720, 2018. 2
- [52] Shipeng Zhang, Lizhi Wang, Ying Fu, Xiaoming Zhong, and Hua Huang. Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery. In *IEEE International Conference on Computer Vision*, pages 10183–10192, 2019. 1, 2, 6