DNU: Deep Non-local Unrolling for Computational Spectral Imaging

Lizhi Wang Chen Sun Maoqing Zhang Ying Fu Hua Huang
Beijing Institute of Technology
{lzwang, sunchen, zmq, fuying, huahuang}@bit.edu.cn

Abstract

Computational spectral imaging has been striving to capture the spectral information of the dynamic world in the last few decades. In this paper, we propose an interpretable neural network for computational spectral imaging. First, we introduce a novel data-driven prior that can adaptively exploit both the local and non-local correlations among the spectral image. Our data-driven prior is integrated as a regularizer into the reconstruction problem. Then, we propose to unroll the reconstruction problem into an optimization-inspired deep neural network. The architecture of the network has high interpretability by explicitly characterizing the image correlation and the system imaging model. Finally, we learn the complete parameters in the network through end-to-end training, enabling robust performance with high spatial-spectral fidelity. Extensive simulation and hardware experiments validate the superior performance of our method over state-of-the-art methods.

1. Introduction

The spectral image delineates a detailed scene representation about the scene, which is beneficial to a diverse range of field, from the fundamental research areas, *e.g.*, medical diagnosis, health care, and remote sensing [4, 6, 32], to the computer vision applications, *e.g.*, face recognition, appearance modeling and object tracking [40, 47, 23]. Conventional spectrometers generally scan the scene along either the spatial dimension or the spectral dimension, requiring multiple exposures to capture a full spectral image. Thus, these systems are unsuitable for measuring dynamic scenes. To this end, researchers have developed quite a few computational spectral imaging prototypes [9, 29, 18, 8, 58]. Based on the foundations of the compressive sensing (CS) theory [14], coded aperture snapshot spectral imaging (CASSI) stands out as a promising solution [3, 48, 50]. However, the bottleneck of CASSI lies in the limited reconstruction quality. The core problem of spectral image reconstruction is how to derive the underlying 3D spectral image from the under-sampled 2D measurement. Theoretically, the reconstruction quality is affected by two aspects: prior regularization and optimization method.

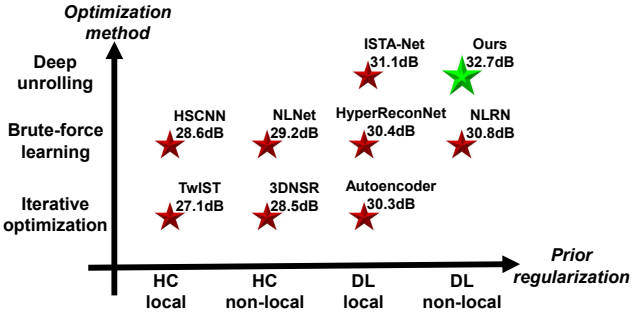


Figure 1. Development trends of the optimization method and the prior regularization. Our method integrates the power of deep unrolling method and non-local prior into an interpretable neural network, which is customized to solve the reconstruction problem of computational spectral imaging, achieving the best performance according to the PSNR. Note that HC and DL are abbreviations for hand-crafted and deep-learning, respectively. Our code is open sourced at [1]

Since the reconstruction problem is under-determined, regularizations based on image priors are required to delineate the structure characteristic in spectral images. Previous approaches model image priors within a local window, like Total Variation (TV) [55], Markov Random Field (MRF) [46] and sparsity [15]. The recently proposed deep denoising prior [61] is also a local prior, as it gradually processes the information from a local neighborhood [54]. Complementary with the local prior, the non-local prior is an alternative in spectral image reconstruction to exploit the long-range dependence [52, 31]. However, existing non-local prior-based regularizations, like non-local mean [7], collaborative filtering [12] and joint sparsity [35], require manually tweaking parameters to handle the various characteristics of the scenes.

Besides the prior regularization, the optimization method is also crucial for the reconstruction quality. Early development of CASSI employs model-based optimization techniques in general [5, 16]. Later, some works propose to learn a brute-force mapping between the compressive im-

ages and the underlying spectral images [53, 38]. However, these methods ignore the system imaging model. They thus lack flexibilities when being used in real systems, as the system imaging models among different hardware implementations differ from each other at a large extent. Recently, researchers have developed some deep unrolling-based optimization methods in the field of natural image CS [20, 45, 60]. These substitute the iterations in model-based optimization with a neural network. However, they still inherit the local prior by explicitly enforcing the feature maps to be sparse.

In this paper, we propose the deep non-local unrolling, an interpretable neural network, for computational spectral imaging. First, we introduce a novel data-driven prior that regularizes the optimization problem to boost the spatial-spectral fidelity. Our data-driven prior explicitly learns both the local prior and the non-local prior. We further delve into the balance of the local prior and the non-local prior by adaptively learning the contribution weights. Then, we incorporate our regularizer into the reconstruction problem and unroll the reconstruction problem into an optimization-inspired deep neural network (DNN). The architecture of the network is intuitively interpretable by explicitly characterizing the image priors and the system imaging model. Finally, we learn the complete parameters in the reconstruction network through end-to-end training to achieve robust performance with high accuracy. Extensive simulation and hardware experiments validate the superior performance of our method over state-of-the-art methods.

2. Related work

2.1. Spectral Image Prior

Regularization based on image priors is a fundamental technique to solve the under-determined optimization problem and is essential to computational spectral imaging. As it is hard to model the prior of a high-dimensional spectral image, most of the previous approaches focus on image local prior characterizing the spectral image structure within a local area. By regularizing the image gradients, the TV model imposes the first-order smoothness prior on the spectral image [55]. Sparse representation methods build a dictionary to model the sparsity prior for image patches [29]. However, the hand-crafted image priors are insufficient to capture the characteristics in various spectral images. Recently, by leveraging the large datasets, the concept of deep denoising prior has been proposed by learning an implicit but more accurate prior based on deep learning [61, 43, 37]. A similar idea has also been exploited in the autoencoder-based computational spectral imaging [11]. As the network extracts the information from local neighborhoods with convolutions, the deep denoising prior and autoencoder prior also belong to local priors [21].

Aiming to exploit the long-range dependence, NLS-based prior has been extensively studied in the literature. The off-the-shelf method to exploit the NLS is based on the Euclidean distance between similar patches. Under the frameworks of sparse representation and low-rank approximation, NLS-based prior has achieved impressive results in computational spectral imaging [52, 17, 31, 62], which demonstrates the effectiveness of NLS in spectral image reconstruction. NLS-based prior is further studied in the framework of deep learning where the block matching operation is treated as a pre-processing step before feeding the similar patches into a neural network [27]. To improve the accuracy of block matching, a non-local neural network is proposed for video classification by exploiting the NLS in an implicit transform domain [54]. Then the non-local neural network module is embedded into a recurrent neural network for image restoration [30]. This paper follows the development trend of the NLS exploitation method. We further propose an interpretable neural network specifically for computational spectral imaging.

2.2. Spectral Image Reconstruction

Besides the prior regularization, the computational optimization methods play an essential role in computational spectral imaging for a faithful reconstruction. Previously, model-based iterative optimization methods are used to incorporate with the hand-crafted priors [5, 16, 29]. However, these methods have to iteratively solve the optimization problems, and thus suffer from parameter tuning and high computational complexity. Later, learning-based methods have been developed for CS image reconstruction [39, 26, 33, 36]. The first category of the learning-based methods for computational spectral imaging is to fit a brute-force mapping from the compressive image to the desirable image [53, 38, 63]. Nevertheless, these methods can solely work on the specific system imaging model that is used during the training. Since the pixel-to-pixel correspondence between the coded aperture pixel and the sensor pixel is fragile to be changed, the system imaging model would vary and differ from the one used during the training. Thus these methods would lack flexibilities in real hardware systems.

To this end, the deep unrolling-based optimization methods have been exploited for image CS [60, 45, 34, 49]. These methods unroll the iterations in the optimization into a DNN and learn the optimization parameters and network parameters simultaneously. Although the deep unrolling-based methods have achieved state-of-the-art results, they still inherit the local prior to regularize the optimization. As we discussed above, the local prior is less potent than the non-local prior.

The motivation of this paper originates from the success of the NLS-based regularization with the model-based op-

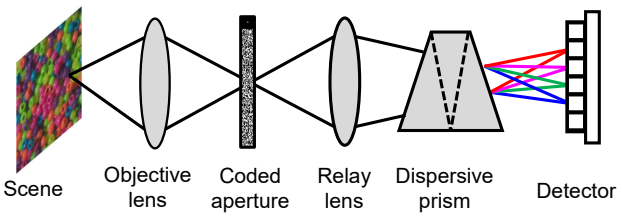


Figure 2. The schematic of CASSI system.

timization method. In this paper, we delve into an interpretable DNN to integrate the NLS prior and the mathematical optimization for compressive spectral imaging.

3. Methodology

3.1. System Imaging Model

Computational spectral imaging is an increasing trend by optically encoding the spectral information and then recovering it through computational reconstruction. Let us start with an in-depth analysis about the CASSI imaging model. Figure 2 illustrates a schematic of CASSI. The spectral information is first spatially modulated by a coded aperture with a fixed pattern and then spectrally dispersed by the dispersive prism before being detected by the detector. Mathematically, considering that a spectral image patch with Λ bands $\{\boldsymbol{F}_\lambda\}_1^\Lambda \in \mathbb{R}^{M\times N}$ is modulated by a coded aperture with pattern $\boldsymbol{C} \in \mathbb{R}^{M\times N}$, the measurement $\boldsymbol{G} \in \mathbb{R}^{M\times N}$ is formulated as

$$\boldsymbol{G} = \sum_{\lambda=1}^{\Lambda} \boldsymbol{C}_\lambda \circ \boldsymbol{F}_\lambda, \tag{1}$$

where $\circ$ means point-wise product. $\boldsymbol{C}_\lambda$ represents the band-wise modulation which is derived by shifting the coded aperture according to the dispersive function $J(\lambda)$ as

$$\boldsymbol{C}_\lambda(m,n) = \boldsymbol{C}(m - J(\lambda), n), \tag{2}$$

where m and n index the spatial coordinates. Note Eq. (2) assumes a dispersion along the vertical dimension, and the inference hereafter is also applicable for horizontal dispersion. The CASSI imaging model in Eq. (1) can be rewritten in the matrix-vector form as

$$\boldsymbol{g} = \boldsymbol{\Phi}\boldsymbol{f}, \tag{3}$$

where $\boldsymbol{g} \in \mathbb{R}^{MN}$ and $\boldsymbol{f} \in \mathbb{R}^{MN\Lambda}$ are the vectorized representation of the compressive image and the underlying spectral image, and $\boldsymbol{\Phi} \in \mathbb{R}^{MN\times MN\Lambda}$ is the sensing matrix that describes the system imaging model. Our observation is that the sensing matrix $\boldsymbol{\Phi}$ is a block diagonal matrix and can be written as

$$\boldsymbol{\Phi} = [d(\boldsymbol{C}_1), \cdots, d(\boldsymbol{C}_\lambda), \cdots, d(\boldsymbol{C}_\Lambda)], \tag{4}$$

where $d(\cdot)$ means an operation that builds a diagonal matrix with the operator. Recall Eq. (2), we can see that the sensing matrix $\boldsymbol{\Phi}$ only depends on the coded aperture $\boldsymbol{C}$. Figure 3 shows the transition from the coded aperture to the sensing matrix, where a spectral image patch with a size of $4\times 4\times 3$ is assumed. The sensing matrix contains as many diagonal patterns as the spectral bands. Each diagonal pattern corresponds to the vectorized coded aperture, and adjacent diagonal patterns are with a uniform shift. It is worth noting that the special structure of the sensing matrix $\boldsymbol{\Phi}$ will lead to $\boldsymbol{\Phi}\boldsymbol{\Phi}^\intercal$ being a diagonal matrix, which determines our development of the network architecture.

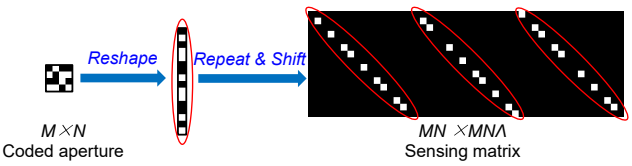


Figure 3. The transition from the coded aperture to the sensing matrix. To synthesize the block diagonal sensing matrix, the coded aperture is first reshaped as a vector, then repeated in the horizontal direction, each time with a uniform shift in the vertical direction, as many times as the number of spectral band.

3.2. Interpretable Unrolling-based Reconstruction

Given the compressive image $\boldsymbol{g}$ and the sensing matrix $\boldsymbol{\Phi}$, the subsequent task is to estimate the underlying spectral image. Since the reconstruction problem is severely under-determined, it need to solve the following minimization problem [44]:

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} ||\boldsymbol{g} - \boldsymbol{\Phi}\boldsymbol{f}||^2 + \tau R(\boldsymbol{f}), \tag{5}$$

where $R(\cdot)$ is the regularization term that enforces some image prior on the solution, and τ is a parameter that tweaks the weights of the data term and the regularization term. The optimization problem in Eq. (5) cannot be directly solved and a general strategy is to decouple it into two subproblems. By introducing an auxiliary valuable, Eq. (5) can be written as a constrained optimization problem:

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} ||\boldsymbol{g} - \boldsymbol{\Phi}\boldsymbol{f}||^2 + \tau R(\boldsymbol{h}), \quad s.t. \quad \boldsymbol{h} = \boldsymbol{f}. \tag{6}$$

Then, we adopt the half quadratic splitting (HQS) method to convert the above constrained optimization problem to a non-constrained optimization problem

$$(\hat{\boldsymbol{f}}, \hat{\boldsymbol{h}}) = \arg\min_{\boldsymbol{f},\boldsymbol{h}} ||\boldsymbol{g} - \boldsymbol{\Phi}\boldsymbol{f}||^2 + \eta||\boldsymbol{h} - \boldsymbol{f}||^2 + \tau R(\boldsymbol{h}), \tag{7}$$

where η is a penalty parameter. Eq. (7) can be split into two subproblems:

$$\hat{\boldsymbol{f}}^{(k+1)} = \arg\min_{\boldsymbol{f}} ||\boldsymbol{g} - \boldsymbol{\Phi}\boldsymbol{f}||^2 + \eta||\boldsymbol{h}^{(k)} - \boldsymbol{f}||^2, \tag{8}$$

$$\hat{\boldsymbol{h}}^{(k+1)} = \arg\min_{\boldsymbol{h}} \eta||\boldsymbol{h} - \boldsymbol{f}^{(k+1)}||^2 + \tau R(\boldsymbol{h}). \tag{9}$$

In this viewpoint, the HQS algorithm separates the sensing matrix $\mathbf{\Phi}$ and the regularization $R(\cdot)$, and these two subproblems can be solved alternatively.

In this section, we focus on finding a plausible solution for the $f-$subproblem in Eq. (8). Note we will depict that the $h-$subproblem in Eq. (9) can be solved with a spectral image prior network in Sec. 3.3. Here we would like to skip the details about the $h-$subproblem and just give a general solver to enable the subsequent deduction:

$$\boldsymbol{h}^{(k+1)} = S(\boldsymbol{f}^{(k+1)}). \tag{10}$$

The f-subproblem in Eq. (8) is a quadratic regularized least-squares problem. A closed form is given as

$$\boldsymbol{f}^{(k+1)} = (\mathbf{\Phi}^\intercal\mathbf{\Phi} + \eta\boldsymbol{I})^{-1}(\mathbf{\Phi}^\intercal\boldsymbol{g} + \eta\boldsymbol{h}^{(k)}), \tag{11}$$

where $\boldsymbol{I}$ is an identity matrix with desired dimensions. Previous methods in the field of image restorations extensively admit that since the matrix $\mathbf{\Phi}^\intercal\mathbf{\Phi} + \eta\boldsymbol{I}$ is very large, it is impossible to directly compute the inverse matrix [13, 42, 41]. Instead, iterative conjugate gradient (CG) algorithm is employed, which, however, requires many iterations and cannot guarantee find the exact solution. In this paper, owning to the specific structure of the sensing matrix $\mathbf{\Phi}$ as shown in Figure 3, we instead adapts the recent advances [59, 31] on computing Eq. (11) to obtain the exact solution directly. Specifically, given a block diagonal sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{MN\times MN\Lambda}$, we can simply calculate a diagonal matrix $\mathbf{\Phi_i}\mathbf{\Phi_i}^\intercal$ as

$$\mathbf{\Phi_i}\mathbf{\Phi_i}^\intercal = \text{diag}\{\phi_1, ...\phi_i, ..., \phi_{MN}\}, \tag{12}$$

where ϕ_i can be pre-calculated according to the coded aperture pattern $\boldsymbol{C}$. By following the matrix inverse lemma, the matrix inversion in Eq. (11) can be written as

$$(\mathbf{\Phi}^\intercal\mathbf{\Phi}+\eta\boldsymbol{I})^{-1} = \eta^{-1}\boldsymbol{I}-\eta^{-1}\mathbf{\Phi}^\intercal(\boldsymbol{I}+\mathbf{\Phi}\eta^{-1}\mathbf{\Phi}^\intercal)^{-1}\mathbf{\Phi}\eta^{-1}. \tag{13}$$

According to Eq.(12), we know

$$(\boldsymbol{I} + \mathbf{\Phi}\eta^{-1}\mathbf{\Phi}^\intercal)^{-1} = \text{diag}\{\frac{\eta}{\eta+\phi_1}, ...\frac{\eta}{\eta+\phi_i}, ...\frac{\eta}{\eta+\phi_n}\} \tag{14}$$

By plugging Eq. (12), Eq. (13) and Eq. (14) into Eq. (11) and simplifying the formula, we have

$$\boldsymbol{f}^{(k+1)} = \boldsymbol{h}^{(k)} + \mathbf{\Phi}^\intercal[(\boldsymbol{g} - \mathbf{\Phi}\boldsymbol{h}^{(k)})./(\eta + \mathbf{\Phi}\mathbf{\Phi}^\intercal)]. \tag{15}$$

In this manner, the $f-$subproblem can be solved to obtain an accurate solution. Further, the calculation of Eq. (15) only needs linear operations with much fewer computational cost compared with the iterative CG-based method.

We then unify the two subproblems as a whole by substituting Eq. (10) into Eq. (15)

$$\boldsymbol{f}^{(k+1)} = S(\boldsymbol{f}^{(k)}) + \mathbf{\Phi}^\intercal[(\boldsymbol{g} - \mathbf{\Phi}S(\boldsymbol{f}^{(k)}))./(\eta + \mathbf{\Phi}\mathbf{\Phi}^\intercal)]. \tag{16}$$

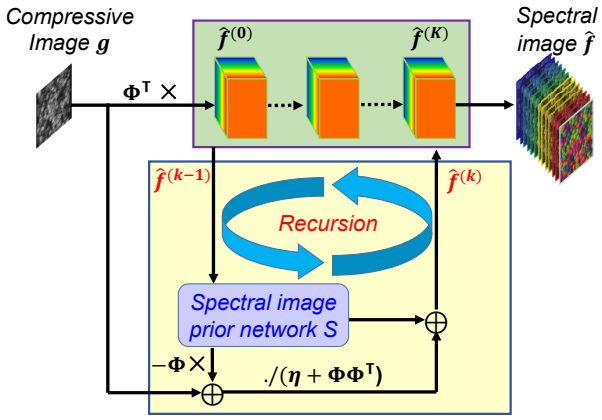


Figure 4. Illustration of the proposed neural network. The network integrates the insight of the optimization method and exploits the specific structure of the sensing matrix. It is composed of multiple recursion, and each recursion includes one spectral image prior network concatenated with some linear connections that accords with the imaging model.

We would like to highlight that it is the first time to deduce the recursion formula in Eq. (16) for the regularization-based optimization, owning to the specific structure of the sensing matrix in computational spectral imaging.

To faithfully solve Eq. (16), we propose to unroll the recursion via a DNN, as shown in Figure 4. The network is composed of multiple recursions, each of which includes one spectral image prior network (as introduced in Sec. 3.3.) concatenated with linear connections that accords with Eq. (16). In the proposed network, the recursions run in a feed-forward manner. The network is trained end-to-end to obey the imaging model and exploit the image priors simultaneously, which is advantageous over the separative solvers in previous methods.

Specifically, the input compressive patch $\boldsymbol{g}$ is first fed into a linear layer parameterized by the transpose of the sensing matrix $\mathbf{\Phi}^\intercal$. The output vector is treated as the initialization: $\boldsymbol{f}^{(0)} = \mathbf{\Phi}^\intercal\boldsymbol{g}$. For the k_{th} recursion, the input $\boldsymbol{f}^{(k-1)}$ is successively fed into the spectral image prior network $S(\cdot)$ and a residual network block [22]. In the residual network block, the identical connection and the residual connection mimic the first part and second part in Eq. (16), respectively. In the residual connection, the input is fed into a linear layer parameterized by $\mathbf{\Phi}$, summed up with the compressive image $\boldsymbol{g}$ and followed by linear connections parameterized by $\eta + \mathbf{\Phi}\mathbf{\Phi}^\intercal$ and $\mathbf{\Phi}^\intercal$, respectively. Such recursion is run K times. According to the recent evidence in [60, 11], we set the recursion number $K = 11$ in the following simulations and experiments, to obtain a balance between accuracy and memory.

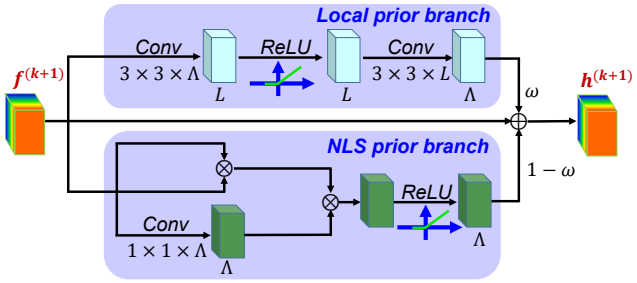
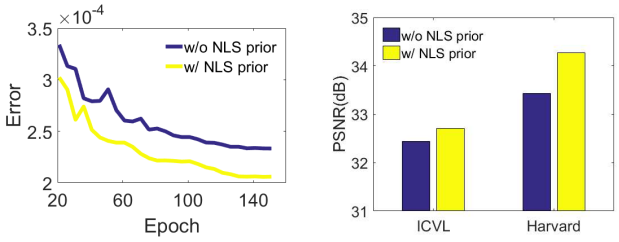


Figure 5. Architecture of the spectral image prior network. It contains a local prior branch and a NLS prior branch, which are adaptively integrated with a learnable weight.



(a) (b)

Figure 6. The performance comparison between the method with (w/) and without (w/o) the NLS prior. (a) Training error. (b) PSNR.

3.3. Neural Local and Non-local Priors

Now we turn to discuss the h-subproblem in Eq. (9), which is actually a proximal operator of $R(\boldsymbol{h})$ computed at point $\boldsymbol{f}^{(k+1)}$. Moreover, it has been demonstrated that both local and NLS prior exist in spectral images [52, 62]. Thus, we propose to explicitly incorporate these two types of priors and rewrite the general form of the proximal operator in Eq. (9) as

$$\hat{\boldsymbol{h}}^{(k+1)} = \arg\min_{\boldsymbol{h}} \eta||\boldsymbol{h} - \boldsymbol{f}^{(k+1)}||^2 + \tau_l R_l(\boldsymbol{h}) + \tau_n R_n(\boldsymbol{h}), \tag{17}$$

where $R_l(\boldsymbol{h})$ and $R_l(\boldsymbol{h})$ represent regularizations based on local prior and non-local prior, respectively. τ_l and τ_n are regularization parameters.

Instead of explicitly modeling the regularizations and solving the proximal operator, we propose to directly learn a solver $S(\cdot)$ for the proximal operator with a customized neural network. In this manner, the spectral image priors are not explicitly modeled but learned with the neural network, which introduces nonlinearity in prior modeling and improves the accuracy of the hand-crafted image priors.

The spectral image prior network is illustrated in Figure 5. Two intuitions guide the design of the spectral image prior network. First, it should enable to exploit local prior and NLS prior simultaneously. Second, it should be as simple as possible to facilitate the training. Following these intuitions, we propose a spectral image prior network that consists of two branches: the local prior branch and the NLS prior branch. The local prior branch is simple and contains only two linear convolutional layers interleaved by one rectified linear unit (ReLU) layer. This design is motivated by the excellent work on image super-resolution that removes the unnecessary layers (such as batch normalization) in the neural networks [28]. In the local prior branch, the first convolutional layer uses $3 \times 3 \times \Lambda$ filters and produces $L = 64$ features, while the second convolutional layer uses $3 \times 3 \times L$ filters and produces Λ features.

The NLS prior branch is designed per the principle of non-local mean operation [7]. Specifically, the NLS prior branch inputs an intermediate spectral image $\boldsymbol{f}$ and generates a refined output $\hat{\boldsymbol{f}}$. A generic formula of the non-local operation is

$$\hat{\boldsymbol{f}}_{\boldsymbol{i}} = \text{ReLU}(\sum_{\boldsymbol{j}} d(\boldsymbol{f}_{\boldsymbol{i}}, \boldsymbol{f}_{\boldsymbol{j}})\psi(\boldsymbol{f}_{\boldsymbol{j}})) \tag{18}$$

where $\boldsymbol{i}$ and $\boldsymbol{j}$ are the indexes of the spatial locations. A pairwise function $d(\cdot, \cdot)$ computes a scaler that represents the distance (*i.e.*, similarity) between input at two locations. The function $\psi(\cdot)$ computes an embedding representation of the input. For simplicity, we set $\psi(\cdot)$ as a linear convolutional embedding and set $d(\cdot, \cdot)$ as a dot-product similarity: $d(\boldsymbol{f}_{\boldsymbol{i}}, \boldsymbol{f}_{\boldsymbol{j}}) = \boldsymbol{f}_{\boldsymbol{i}}^{\mathsf{T}} \boldsymbol{f}_{\boldsymbol{j}}$. The NLS prior branch that accords with Eq. (18) is shown in Figure 5. The convolution in the embedding uses $1 \times 1 \times \Lambda$ kernels and generates Λ filters.

Finally, we propose to integrate the local prior branch and the NLS prior branch with a weight parameter ω. We further employ the residual network design [22] in the spectral image prior network, since residual learning enables fast and stable training and relieves the computational burden. Figure 6 shows the training loss and the final PSNR benefiting from the NLS prior branch, which validates the intuitions for the network design.

3.4. Adaptive Parameters Learning

Once the network is built, we train it by end-to-end training to learn the network parameters and the optimization parameters simultaneously. In our implementation, all the parameters are set to be different among each recursion, as with the recursion increasing, the reconstruction quality is improved; thus the network parameters and the optimization parameters should be changed accordingly. We would like to highlight that instead of manually setting the parameter ω, we propose to learn it to adaptively balance the contribution of the local prior and the NLS prior.

Given a set of spectral image patches $\boldsymbol{F}^{(l)}$ and its corresponding compressive patch $\boldsymbol{g}^{(l)}$ as the training samples, the network is trained according to the MSE-based lose func-

Table 1. Performance comparisons on ICVL and Harvard datasets (3% compressive ratio). The best performance is labeled in bold.

Dataset	Metric	TwIST	GPSR	BPDN	3DNSR	SSLR	HSCNN	ISTA-Net	Autoencoder	HyperReconNet	Ours
ICVL	PSNR	26.15	24.56	26.77	27.95	29.16	29.48	31.73	30.44	32.36	**34.27**
	SSIM	0.936	0.909	0.947	0.958	0.964	0.973	0.984	0.970	0.986	**0.991**
	SAM	0.053	0.09	0.052	0.051	0.046	0.043	0.042	0.036	0.037	**0.034**
Harvard	PSNR	27.16	24.96	26.67	28.51	29.68	28.55	31.13	30.30	30.34	**32.71**
	SSIM	0.924	0.907	0.935	0.94	0.952	0.944	0.967	0.952	0.964	**0.978**
	SAM	0.119	0.196	0.155	0.132	0.101	0.118	0.114	0.098	0.115	**0.091**
Time(s)		555	302	705	8648	6986	3.11	1.15	521	30	**0.98**

tion, which can be expressed as

$$(\ddot{\boldsymbol{\Theta}}, \ddot{\eta}) = \arg\min_{\boldsymbol{\Theta}, \boldsymbol{\eta}} \frac{1}{L} \sum_{l=1}^{L} ||\widetilde{\boldsymbol{F}}(\boldsymbol{g}^{(l)}; \boldsymbol{\Theta}, \eta) - \boldsymbol{F}^{(l)}||^2, \quad (19)$$

where $\widetilde{\boldsymbol{F}}(\cdot)$ denotes the output of the network given the input and the parameters.

We employ TensorFlow to implement the network, minimize the loss function using a stochastic gradient descent method, and train it up to 150 epochs. Figure 6a plots the testing loss along with the training epochs, which verifies the convergence of the proposed method. The mini-batch size and momentum are set as 64 and 0.9, respectively. The learning rate is initially set as 0.001 and exponentially decays to 90% for every ten epochs. The network parameters are initialized with the method in [19]. We use a machine equipped with an Intel Core i7-6800K CPU with 64GB memory and an NVIDIA Titan X PASCAL GPU.

4. Simulations on Synthetic Data

4.1. Configurations

For a comprehensive evaluation, we conduct simulations on two public spectral datasets, *i.e.*, the ICVL dataset [2] and the Harvard dataset [10]. The basic configurations of the spectral cameras for obtaining these datasets are different, so the property of the images from different datasets are diverse and heterogeneous. Specifically, the spectral images in the ICVL dataset are acquired using a Specim PS Kappa DX4 spectral camera and a rotary stage for spatial scanning. The spectral range is from 400nm to 700nm, which is divided into 31 spectral bands with approximate 10nm bandwidth for each band. There are 201 spectral images in ICVL dataset. To avoid over-fitting, we exclude 31 spectral images with similar backgrounds and 20 spectral images with similar contents. Then we randomly select 100 spectral images for training and 50 spectral images for testing. The spectral images in the Harvard dataset are acquired using a CRI Nuance FX spectral camera with a liquid crystal tunable filter for spectral scanning. The spectral range is from 420nm to 720nm with 31 spectral bands. The Harvard dataset consists of 50 spectral images with distinct natural scenes. We remove 6 deteriorated spectral images with large-area saturated pixels, and randomly select 35 spectral images for training and 9 spectral images for testing. We set the patch size as 48×48, and randomly select 80% patches for training and the rest patches for validation. The coded aperture patterns are constructed by following random matrix in Bernoulli distributions with $p = 0.5$. However, the coded aperture patterns in testing and training are totally different, as this configuration represents the real case in hardware systems where the system imaging model is fragile to varies.

We compare our methods with nine main stream methods, including five hand-crafted prior based methods: TV based TwIST method [24], sparsity based GPSR method [48] and BPDN method [29], and NLS based 3DNSR method [52] and SSLR method [17], and four learning based methods: HSCNN [57], ISTA-Net [60], Autoencoder [11] and HyperReconNet [53]. All the codes for competitive methods are released publicly or provided privately to us by the authors, and we make great efforts to produce their best results.

Three quantitative image quality metrics are employed to evaluate the performance of these methods, including peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [56], and spectral angle mapping (SAM) [25]. PSNR and SSIM are calculated on each 2D spatial image and averaged over all spectral bands. A larger value of PSNR and SSIM indicates a higher accuracy in the spatial domain. SAM is calculated on each 1D spectral vector and averaged over all spatial points. A smaller value of SAM suggests a smaller error in the spectral domain.

4.2. Evaluation

Numerical Results. Table 1 summarizes the numerical results on ICVL and Harvard datasets where the compressive ratio is 3%. The proposed method outperforms all the existing methods according to the metrics in both spatial and spectral domains, which demonstrates the superiority of the integration of NLS and deep unrolling. Specifically, our method exhibits remarkably higher accuracy compared with the methods based on hand-crafted priors. This indicates that our method manages to capture more accurate priors by the spectral image prior network. Further, the gains of our method over the brute-force learning methods,

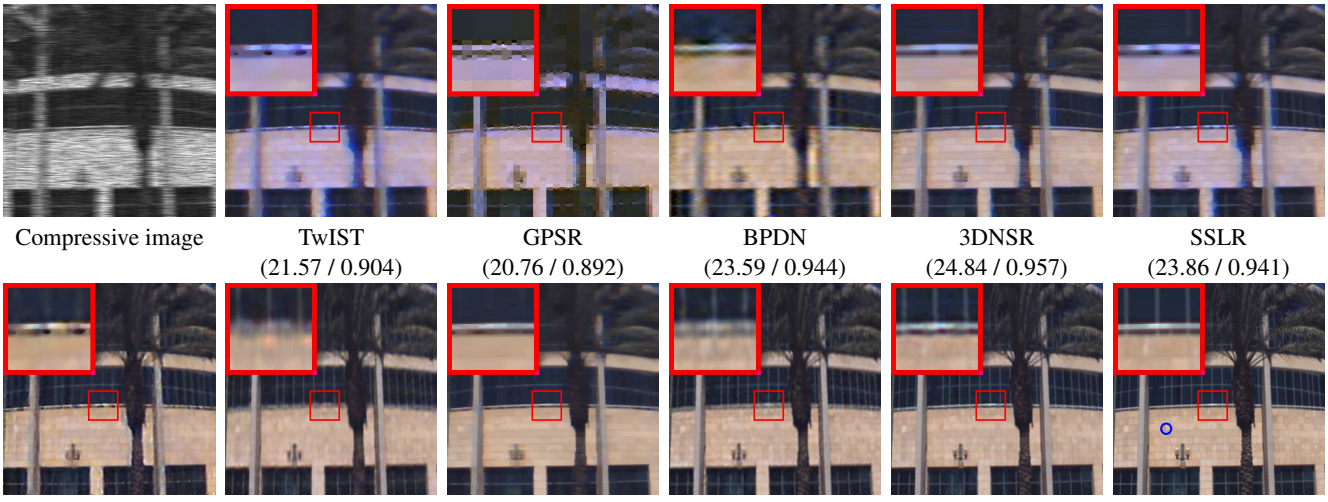
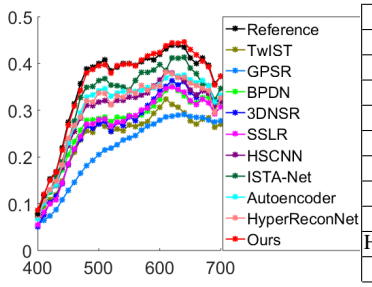


Figure 7. Visual quality comparison. The PSNR and SSIM for the result images are shown in the parenthesis. Our methods outperforms all the competitive methods in terms of both perceptual quality and quantitative metrics.



	RMSE	SAM
TwIST	0.116	0.046
GPSR	0.144	0.137
BPDN	0.091	0.060
3DNSR	0.093	0.097
SSLR	0.095	0.065
HSCNN	0.063	0.033
ISTA-Net	0.034	0.032
Autoencoder	0.052	0.021
HyperReconNet	0.057	0.027
Ours	**0.018**	**0.011**

Figure 8. Comparison of spectral accuracy. The point is indicated in Figure 7. The spectra reconstructed by the our method is closer to the reference compared with the other methods. The RMSE and SAM numbers further demonstrate the superiority of our methods on spectral fidelity.

i.e., HSCNN and HyperReconNet, demonstrate the effectiveness of our derivation of the specific architecture for deep unrolling. Last but not least, our method is considerably better than ISTA-Net and Autoencoder, which are learning-based methods but solely employ local priors. We attribute this to the exploitation of the NLS-based prior in our method.

Perceptual Quality. We show the reconstruction results of one representative image from ICVL dataset in Figure 7. To simultaneously present the results of all spectral bands, we convert the spectral images to sRGB via the CIE color matching function. We also provide the PSNR and SSIM values for each result image. Clearly, the proposed method can produce visually pleasant results with less artifact and sharper edges compared with other methods, which is consistent with the numerical metrics.

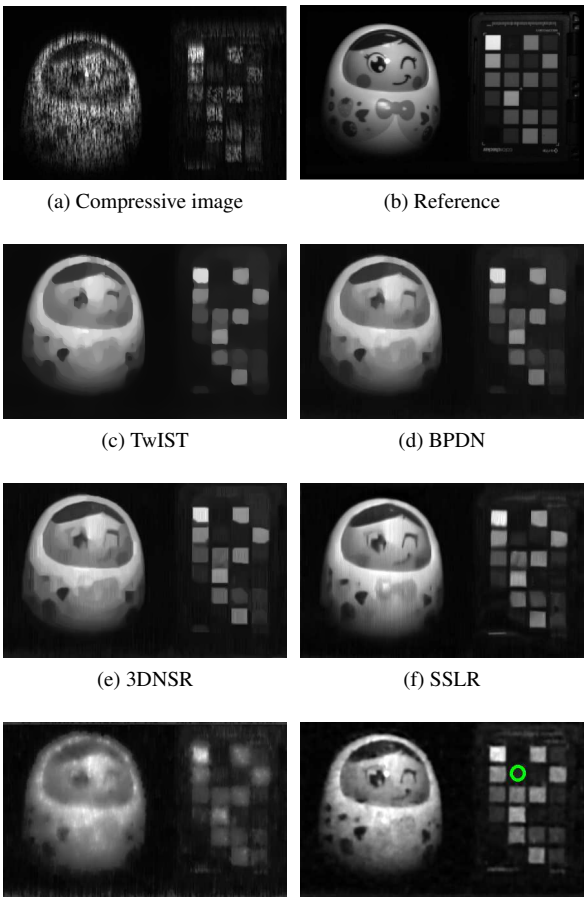
Table 2. Performance comparisons on natural image CS.

Dataset	CS Ratio	SDA	ReconNet	ISTA-Net	Ours
Set11	10%	22.65	24.28	25.80	**25.96**
	4%	20.12	20.63	21.23	**22.11**
BSD68	10%	23.12	24.15	25.02	**25.46**
	4%	21.32	21.66	22.12	**22.94**

Spectral Fidelity. Figure 8 plots the recovered spectra of one point indicated in Figure 7. The spectrum reconstructed by the proposed method is closest to the reference. The RMSE and SAM further demonstrate the superior performance of the proposed method on spectral fidelity.

Computational Complexity. Table 1 lists the running time for reconstructing one spectral image with size of $512 \times 512 \times 31$. All the codes are implemented on an Intel Core i7-6800K CPU. As can be seen, the proposed method is comparable with ISTA-Net and HSCNN (in seconds), and much faster than the other methods. In our method, the total number of multiplication for spectral image prior network is approximately $M \times N \times 10^5$. In contrast, for example, the GPSR method adopts the sparse coding technique to solve the proximal operation and its multiplication is approximately $M \times N \times 10^7$ when a $2\times$ over-complete dictionary is used.

Generality on Natural Image CS. The proposed deep nonlocal unrolling model is a general model for image restoration. To test its generality beyond spectral imaging, we conduct further simulations on natural image CS by following the configurations in [26, 60]. We compare our method with three state-of-the-art optimization-based methods: SDA [39], ReconNet [26] and ISTA-Net [60]. Ta-



(a) Compressive image (b) Reference

(c) TwIST (d) BPDN

(e) 3DNSR (f) SSLR

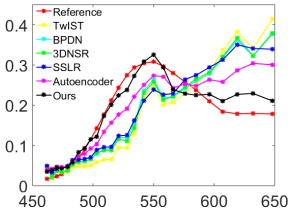
(g) Autoencoder (h) Ours

Figure 9. Comparison on real captured data. Together with the reconstrcution results, we also show the compressive image and the panchromatic image of the target for reference. Our method can produce the results with clearer visual information compared with the other methods. The center wavelength for the selected band is 632nm.

ble 2 lists the average PSNR results on Set11 and BSD68 with two CS ratios. It can be observed that the proposed method outperforms all the competing methods, which further demonstrates the merits of deep non-local unrolling.

5. Experiments on a hardware System

We conduct experiments on a real hardware system to demonstrate the practicability of our method [48, 51]. To handle the real-world scenes, we retrain the network by combining the spectral images in two datasets. Figure 9 shows the reconstructed images by our method together with TwIST, BPDN, 3DNSR, SSLR, and Autoencoder. We also capture a panchromatic image of the target for reference. It can be seen that the proposed method can produce better results with fewer artifacts and clearer content compared with the other methods. The qualitative results with spectral plots are also provided in Figure 10. The reference spectrum is obtained with a commercial spectrometer (Ocean Optics). The spectrum reconstructed by our method is closer to the reference compared with the other methods. The RMSE and SAM numbers further show the superiority of our methods on spectral fidelity, which demonstrates the effectiveness of our method in the real hardware system.



	RMSE	SAM
TwIST	0.109	0.563
BPDN	0.093	0.476
3DNSR	0.095	0.491
SSLR	0.090	0.461
Autoencoder	0.056	0.288
Ours	**0.023**	**0.120**

Figure 10. Comparison of spectral accuracy. The point is indicated in Figure 9. The spectrum reconstructed by the our method is closer to the reference compared with the other methods. The RMSE and SAM numbers further demonstrate the superiority of our methods on spectral fidelity.

6. Conclusion

In this paper, we have presented an interpretable neural network for computational spectral imaging. The proposed method integrates the merits of deep unrolling and NLS as follows: (1) By exploiting the block diagonal nature of the sensing matrix, we derive a novel recursion formula, based on which we develop an interpretable neural network for solving the reconstruction problem in computational spectral imaging. (2) By recognizing that previous deep learning methods solely consider local prior via local convolution, we propose a mechanism to adaptively incorporate the non-local prior into the spectral image prior network. Our method is free of optimization parameter tuning and reduces the computational complexity. We have also validated the effectiveness of the proposed method on real hardware prototype. One future direction of interest is to extend the proposed method for more spectral image processing problems, *e.g.*, spectral interpolation, and demosaicing. The other direction is to accelerate the proposed method to reach a video-rate reconstruction, thus enabling the real-time acquisition of hyperspectral video.

Acknowledgments

This work is supported in part by National Natural Science Foundation of China under Grant 61701025 and Grant 61672096, in part by Beijing Municipal Science and Technology Commission under Grant Z181100003018003 and in part by Beijing Institute of Technology Research Fund Program for Young Scholars.

References

[1] https://github.com/wang-lizhi/DeepNonlocalUnrolling. 1

[2] B. Arad and O. Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34, 2016. 6

[3] G. Arce, D. Brady, L. Carin, H. Arguello, and D. Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2014. 1

[4] V. Backman, M. B. Wallace, L. Perelman, J. Arendt, R. Gurjar, M. Müller, Q. Zhang, G. Zonios, E. Kline, T. McGillican, et al. Detection of preinvasive cancer cells. *Nature*, 406(6791):35, 2000. 1

[5] J. M. Bioucas-Dias and M. A. Figueiredo. A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004, 2007. 1, 2

[6] M. Borengasser, W. S. Hungate, and R. Watkins. *Hyperspectral remote sensing: principles and applications*. CRC press, 2007. 1

[7] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 60–65, 2005. 1, 5

[8] X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin. A prism-mask system for multispectral video acquisition. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 33(12):2423–2435, 2011. 1

[9] X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. J. Brady. Computational snapshot multispectral cameras: toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5):95–108, 2016. 1

[10] A. Chakrabarti and T. Zickler. Statistics of real-world hyperspectral images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–200, 2011. 6

[11] I. Choi, D. S. Jeon, G. Nam, D. Gutierrez, and M. H. Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on on Graphics (SIGGRAPH Asia)*, 36(6):218, 2017. 2, 4, 6

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 1

[13] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu. Denoising prior driven deep neural network for image restoration. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 41(10):2305–2318, 2018. 4

[14] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. 1

[15] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 1

[16] M. A. Figueiredo, R. D. Nowak, and S. 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. 1, 2

[17] Y. Fu, Y. Zheng, I. Sato, and Y. Sato. Exploiting spectral-spatial correlation for coded hyperspectral image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3727–3736, 2016. 2, 6

[18] L. Gao, R. T. Kester, N. Hagen, and T. S. Tkaczyk. Snapshot image mapping spectrometer (ims) with high sampling density for hyperspectral microscopy. *OSA Optics Express*, 18(14):14330–14344, 2010. 1

[19] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010. 6

[20] K. Gregor and Y. LeCun. Learning fast approximations of sparse coding. In *International Conference on Machine Learning*, pages 399–406, 2010. 2

[21] S. Gu, R. Timofte, and L. Van Gool. Integrating local and non-local denoiser priors for image restoration. In *International Conference on Pattern Recognition*. 2

[22] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 4, 5

[23] M. H. Kim, T. A. Harvey, D. S. Kittle, H. Rushmeier, J. Dorsey, R. O. Prum, and D. J. Brady. 3d imaging spectroscopy for measuring hyperspectral patterns on solid objects. *ACM Transactions on on Graphics*, 31(4):38:1–38:11, 2012. 1

[24] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady. Multiframe image estimation for coded aperture snapshot spectral imagers. *OSA Applied Optics*, 49(36):6824–6833, 2010. 6

[25] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz. The spectral image processing system (SIPS)–interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2-3):145–163, 1993. 6

[26] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 449–458, 2016. 2, 7

[27] S. Lefkimmiatis. Non-local color image denoising with convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3587–3596, 2017. 2

[28] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1132–1140, 2017. 5

[29] X. Lin, Y. Liu, J. Wu, and Q. Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on on Graphics*, 33(6):233, 2014. 1, 2, 6

[30] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang. Non-local recurrent network for image restoration. In *Advances in Neural Information Processing Systems*, pages 1680–1689, 2018. 2

[31] Y. Liu, X. Yuan, J. Suo, D. J. Brady, and Q. Dai. Rank minimization for snapshot compressive imaging. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 41(12):2990–3006, 2018. 1, 2, 4

[32] G. Lu and B. Fei. Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, 19(1):010901, 2014. 1

[33] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos. Using deep neural networks for inverse problems in imaging: beyond analytical methods. *IEEE Signal Processing Magazine*, 35(1):20–36, 2018. 2

[34] J. Ma, X.-Y. Liu, Z. Shou, and X. Yuan. Deep tensor admm-net for snapshot compressive imaging. In *International Conference on Computer Vision*, pages 10223–10232, 2019. 2

[35] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2272–2279, 2009. 1

[36] M. T. McCann, K. H. Jin, and M. Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95, 2017. 2

[37] T. Meinhardt, M. Moller, C. Hazirbas, and D. Cremers. Learning proximal operators: Using denoising networks for regularizing inverse imaging problems. In *International Conference on Computer Vision*, pages 1781–1790, 2017. 2

[38] X. Miao, X. Yuan, Y. Pu, and V. Athitsos. λ-net: Reconstruct hyperspectral images from a snapshot measurement. In *International Conference on Computer Vision*, pages 4059–4069, 2019. 2

[39] A. Mousavi, A. B. Patel, and R. G. Baraniuk. A deep learning approach to structured signal recovery. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing*, pages 1336–1343, 2015. 2, 7

[40] Z. Pan, G. Healey, M. Prasad, and B. Tromberg. Face recognition in hyperspectral images. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 25(12):1552–1560, 2003. 1

[41] N. Qi, Y. Shi, X. Sun, J. Wang, B. Yin, and J. Gao. Multi-dimensional sparse models. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 40(1):163–178, 2018. 4

[42] J. Ren, J. Liu, and Z. Guo. Context-aware sparse decomposition for image denoising and super-resolution. *IEEE Transactions on Image Processing*, 22(4):1456–1469, 2013. 4

[43] J. Rick Chang, C.-L. Li, B. Poczos, B. Vijaya Kumar, and A. C. Sankaranarayanan. One network to solve them all–solving linear inverse problems using deep projection models. In *International Conference on Computer Vision*, pages 5888–5897, 2017. 2

[44] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 860–867, 2005. 3

[45] J. Sun, H. Li, Z. Xu, et al. Deep admm-net for compressive sensing mri. In *Advances in Neural Information Processing Systems*, pages 10–18, 2016. 2

[46] M. F. Tappen. Utilizing variational optimization to learn markov random fields. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. 1

[47] H. Van Nguyen, A. Banerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *IEEE Computer Vision and Pattern Recognition Workshops*, pages 44–51, 2010. 1

[48] A. Wagadarikar, R. John, R. Willett, and D. Brady. Single disperser design for coded aperture snapshot spectral imaging. *OSA Applied Optics*, 47(10):B44–B51, 2008. 1, 6, 8

[49] L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang. Hyperspectral image reconstruction using a deep spatial-spectral prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8032–8041, 2019. 2

[50] L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu. High-speed hyperspectral video acquisition with a dual-camera architecture. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4942–4950, 2015. 1

[51] L. Wang, Z. Xiong, H. Huang, G. Shi, F. Wu, and W. Zeng. High-speed hyperspectral video acquisition by combining nyquist and compressive sampling. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 41(4):857–870, 2018. 8

[52] L. Wang, Z. Xiong, G. Shi, F. Wu, and W. Zeng. Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 39(10):2104–2111, 2017. 1, 2, 5, 6

[53] L. Wang, T. Zhang, Y. Fu, and H. Huang. Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. *IEEE Transactions on Image Processing*, 28(5):2257–2270, 2019. 2, 6

[54] X. Wang, R. Girshick, A. Gupta, and K. He. Non-local neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7794–7803, 2018. 1, 2

[55] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008. 1, 2

[56] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6

[57] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, and F. Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *IEEE International Conference on Computer Vision Workshops*, volume 2, 2017. 6

[58] Z. Xiong, L. Wang, H. Li, D. Liu, and F. Wu. Snapshot hyperspectral light field imaging. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3270–3278, 2017. 1

[59] X. Yuan. Generalized alternating projection based total variation minimization for compressive sensing. In *IEEE International Conference on Image Processing*, pages 2539–2543. IEEE, 2016. 4

[60] J. Zhang and B. 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. 2, 4, 6, 7

[61] K. Zhang, W. Zuo, S. Gu, and L. Zhang. Learning deep cnn denoiser prior for image restoration. In *IEEE Conference*

on Computer Vision and Pattern Recognition, pages 2808–2817, 2017. 1, 2

[62] S. Zhang, L. Wang, Y. Fu, X. Zhong, and H. Huang. Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery. In *International Conference on Computer Vision*, pages 10183–10192, 2019. 2, 5

[63] T. Zhang, Y. Fu, L. Wang, and H. Huang. Hyperspectral image reconstruction using deep external and internal learning. In *International Conference on Computer Vision*, pages 8559–8568, 2019. 2