# Hyperspectral Image Reconstruction Using a Deep Spatial-Spectral Prior

Lizhi Wang[1]   Chen Sun[1]   Ying Fu[1]   Min H. Kim[2]   Hua Huang[1]

[1]Beijing Institute of Technology   [2]Korea Advanced Institute of Science and Technology

{lzwang, sunchen, fuying, huahuang}@bit.edu.cn, minhkim@kaist.ac.kr

## Abstract

*Regularization is a fundamental technique to solve an ill-posed optimization problem robustly and is essential to reconstruct compressive hyperspectral images. Various hand-crafted priors have been employed as a regularizer but are often insufficient to handle the wide variety of spectra of natural hyperspectral images, resulting in poor reconstruction quality. Moreover, the prior-regularized optimization requires manual tweaking of its weight parameters to achieve a balance between the spatial and spectral fidelity of result images. In this paper, we present a novel hyperspectral image reconstruction algorithm that substitutes the traditional hand-crafted prior with a data-driven prior, based on an optimization-inspired network. Our method consists of two main parts: First, we learn a novel data-driven prior that regularizes the optimization problem with a goal to boost the spatial-spectral fidelity. Our data-driven prior learns both local coherence and dynamic characteristics of natural hyperspectral images. Second, we combine our regularizer with an optimization-inspired network to overcome the heavy computation problem in the traditional iterative optimization methods. We learn the complete parameters in the network through end-to-end training, enabling robust performance with high accuracy. Extensive simulation and hardware experiments validate the superior performance of our method over the state-of-the-art methods.*

## 1. Introduction

Hyperspectral imaging captures the spectral power distributions of a scene or an object as a three-dimensional (3D) tensor, which describes spectral intensity per wavelength at each pixel location. By leveraging the rich spectral information in hyperspectral images, various imaging and vision applications have been benefited extending the current dimensions of the recognition and classification techniques to higher dimensions of novel applications, *e.g.*, re-mote sensing, medical imaging, vision inspection, digital forensics, etc. [1, 2, 3, 4, 5, 6].

Since a hyperspectral image is a 3D tensor, the scenes have to be measured with multiple exposures with one 1D or 2D sensor. This imaging process is time-consuming and limited to static objects [7]. With an objective to capture dynamic objects and scenes, quite a few snapshot hyperspectral imaging system have been proposed [8, 9, 10, 11, 12]. Based on compressive sensing [13], coded aperture snapshot spectral imaging (CASSI) [14] introduces a promising solution. CASSI encodes the incident light into a snapshot-imaging sensor and employs an optimization algorithm to reconstruct the hyperspectral image as a 3D tensor.

To solve the under-determined reconstruction problem, regularization technique is used to introduce the image priors, for instance, the total variation (TV) [15, 16], sparsity [17, 18] and non-local similarity (NLS) [19, 20]. They have been formulated analytically to restrict the solution space when solving the data term. However, these prior are designed empirically and often insufficient to model the wide variety of spectra of the real world. Also, the optimization based on a hand-crafted prior requires manual tweaking of its weight parameters to handle the complicated characteristics of the target scenes.

Moreover, the optimization of reconstructing hyperspectral images cannot be solved with a closed form. Therefore, iterative optimization techniques have been used in general. Recently, a novel research direction is to substitute the iteration-based optimization in compressive sensing with a deep neural network, *e.g.*, LISTA [21] ADMM-Net [22] and ISTA-Net [23]. These networks learn unrolled solutions of an iterative optimization based on natural image statistics by means of deep learning. However, they still inherit the sparsity prior by explicitly limiting features to be sparse in a few layers. The optimization using the sparsity prior suffers from the prior deviation problem as addressed by Zhang et al. [24]. Besides, these neural network-based approaches only account for CS reconstruction within the spatial dimension, but in ignorance of the spectral dimension, as they do not learn any prior of spectral information

of natural images. Choi et al. proposed to train an autoencoder to learn the spectral prior [25]. Nevertheless, the solution is still based on iterative optimization using the sparsity regularizer.

In this paper, we propose a novel hyperspectral image reconstruction algorithm that substitutes the traditional hand-crafted prior with a data-driven prior, based on an optimization-inspired network. The proposed method combines the merits from two aspects: the structure insight of the optimization and the prior modeling capacity of the deep neural network. First, we learn a novel data-driven prior that regularizes the optimization problem to exploit the spatial and spectral correlation. Our data-driven prior learns both local coherence and dynamic characteristics of natural hyperspectral images. Second, we combine our regularizer with an optimization-inspired network to overcome heavy computation of the traditional iterative optimization methods. We learn the entire parameters in the complete network by end-to-end training, enabling robust performance with high accuracy. Extensive simulations and hardware experiments verify that the proposed method can achieve a significant improvement over the state-of-the-art methods according to both comprehensive quantitative metrics and perceptual quality.

In a nutshell, this work integrates the power of deep learning and the optimization framework into the reconstruction problem of compressive hyperspectral imaging to take a substantial step forward to make CASSI into practice.

## 2. Related work

### 2.1. Prior Modeling in Hyperspectral Image

Solving an inverse optimization problem stands at the core of hyperspectral image reconstruction, *i.e.*, the subsequent reconstruction problem is how to derive the underlying 3D hyperspectral image from a 2D compressive image. In general, a hyperspectral image prior is used to regularize the inverse problem, since the prior can identify the most feasible solution from the infinite set of solutions by enforcing specific feature to the solution. Thereby, designing an appropriate prior plays a key role in finding a solution for the reconstruction problem in compressive hyperspectral imaging.

In the conventional hyperspectral image reconstruction methods, most of the hyperspectral image priors are hand-crafted based on empirical observation. Given the fact that natural hyperspectral images are usually sparse after being transformed in the frequency or gradient domain, priors that enforce the sparsity of transformation coefficients are widely used in CASSI reconstruction [15, 18, 26]. To improve the diversity of the transformation, blind compressive sensing has been proposed to solve the CASSI reconstruction problem [27]. This technique makes efforts to jointly

infer the underlying hyperspectral image and the transformation basis from the compressive image. Further, by making use of the repetitive structures in hyperspectral images, NLS-based regularizers have been adopted by combining with the sparsity representation [19] or low-rank approximation [20, 28], which have shown improvement over the local regularizers. However, these hand-crafted priors are often to generic, in that many non-image signals can also satisfy the constraint.

Different from the carefully designed prior, the learning-based methods can learn implicit image priors from a large dataset. Choi et al. [25] proposed an autoencoder-based method for CASSI reconstruction, where an autoencoder is pre-trained as a deep image prior and integrated in the optimization as a regularizer. A similar approach has been exploited for natural image restoration, where deep image priors are adopted as regularizers in the optimization [24]. Compared with the hand-crafted image priors, the data-driven priors can characterize nonlinear correlation in hyperspectral images and thus lead to superior performance. But it has to iteratively solve the optimization problems, which suffers from parameter tuning.

### 2.2. Deep Learning for Image Reconstruction

Compared with the prevalence of deep learning in the field of high-level visions [29, 30], few works focus on involving deep learning to solve the hyperspectral image reconstruction problem, especially concerning the deployment on real hardware systems.

Pioneering works on natural image reconstruction try to learn a brute-force mapping function from the compressive image to the underlying image, for instance [31, 32], which, however, lacks flexibility. The learned mapping function would be ineffective and need to be retrained, even if the observation model deviates very slightly from that one used during the training. Actually, this is a common phenomenon in hardware implementation. For the first time, Xiong et al. introduced a convolutional neural network to solve the hyperspectral image reconstruction problem in CASSI [33]. It transfers the hyperspectral image reconstruction problem to an hyperspectral image enhancement problem and learns a mapping function from a low-quality reconstruction to the desired hyperspectral images. In contrast, the proposed network-based method is designed with inspiration from the optimization framework, which fully integrates the observation model in the neural network. Thus the proposed method can generalize well to an untrained observation model.

The motivation of this work originates from the recently proposed deep neural network for natural image compressive sensing, including the LISTA [21] ADMM-Net [22] and ISTA-Net [23]. These methods mimic the structure of the prior-regularized optimization and unroll the itera-
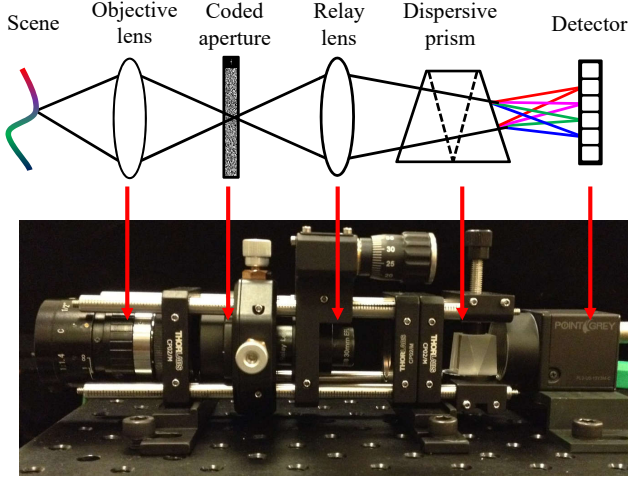
Figure 1. The optical principle in CASSI and the prototype we build for verifying the proposed method with real hardware experiment.
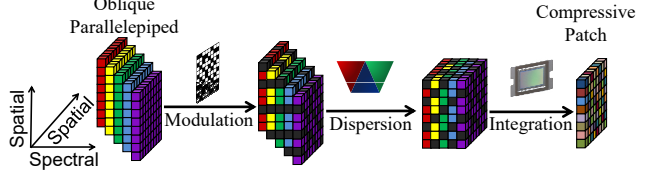


Figure 2. Patch-based modeling for CASSI. One patch in the compressive image exactly corresponds to one oblique parallelepiped cube in the underlying hyperspectral image

tions in optimization into a deep network. Consequently, they can learn the transformations and parameters by end-to-end training. But they still utilize the sparsity prior in the neural network by explicitly limiting the features in a few layer to be sparse, which suffers from the same drawback as the hand-crafted image priors. Furthermore, these methods are specifically designed and developed for natural image compressive sensing and cannot be trivially extended to hyperspectral image reconstruction, since hyperspectral image lies in high dimension, thereby spectral prior should be exploited to facilitate the reconstruction.

## 3. The Proposed Framework

### 3.1. CASSI Observation Model

A schematic diagram of CASSI is shown in Figure 1. CASSI encodes the 3D hyperspectral information into a 2D compressive image. Let $\boldsymbol{F}(m, n, \lambda)$ indicate the intensity of incident light where $1 \leq m \leq M$ and $1 \leq n \leq N$ index the spatial dimension and $1 \leq \lambda \leq \Lambda$ indexes the spectral dimension. A coded aperture creates spatial modulation by its transmission function $\boldsymbol{C}(m, n)$, while a dispersive prism creates spectral shear along a spatial dimension, according to the wavelength-dependent dispersion function $\psi(\lambda)$. By following the observation model of CASSI, the 2D compressive image can be represented as an integral over the spectral wavelength $\lambda$ as

$$\boldsymbol{G}(m, n) = \sum_{\lambda=1}^{\Lambda} \boldsymbol{C}(m - \psi(\lambda), n)\boldsymbol{F}(m - \boldsymbol{\psi}(\lambda), n, \lambda). \quad (1)$$

Note the shear in Eq. (1) is along the vertical dimension and the inference hereafter is also applicable to the horizontal shear. The CASSI observation model can be rewritten in matrix-vector form as

$$\boldsymbol{g} = \boldsymbol{\Phi}\boldsymbol{f}, \quad (2)$$

where $\boldsymbol{g} \in \mathbb{R}^{(M+\Lambda-1)N}$ and $\boldsymbol{f} \in \mathbb{R}^{MN\Lambda}$ are the vectorized representation of the compressive image and the underlying hyperspectral image, and $\boldsymbol{\Phi} \in \mathbb{R}^{(M+\Lambda-1)N \times MN\Lambda}$ represents the measurement matrix of CASSI.

By analyzing the imaging mechanism within CASSI, we decompose the observation model from full image-based modeling to patch-based one to relieve the computational complexity and then facilitate the network training. As shown in Figure 2, we now consider a $P \times P$ patch of the compressive image, and trace the energy in the patch back through the system. The source hyperspectral image is no longer a standard cube but an oblique parallelepiped cube with $\Lambda$ shifted spectral bands. Each band has a one-pixel shift relative to its neighboring bands in the shear direction. In this manner, we can isolate the patch-based mapping between the compressive image and the underlying hyperspectral image, and avoid cross-talk between different mapping pairs. Then, the observation model of such oblique parallelepiped cube can be expressed in the matrix-vector form as

$$\boldsymbol{g}_i = \boldsymbol{\Phi}_i\boldsymbol{f}_i, \quad (3)$$

where $i$ is the index number of the selected patch. Note that Eq. (3) is the patch-based observation model of Eq. (2). For simplicity, we remove the subscripts in Eq. (3).

### 3.2. Hyperspectral Image Prior Network

Physically, the hyperspectral image reconstruction is under-determined, thus image priors are adopted as the regularization to constrain the solution space [34]. From a Bayesian perspective, the underlying hyperspectral image can be obtained by solving a minimization problem as

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}}||\boldsymbol{g} - \boldsymbol{\Phi}\boldsymbol{f}||^2 + \tau R(\boldsymbol{f}), \quad (4)$$

where $\tau$ is a balancing parameter. The data term guarantees that the solution accords with the observation model and the regularization term enforces the output with desired hyperspectral image prior $R(\cdot)$.

When the regularization term is not differentiable, it often employs the variable splitting technique to decouple the
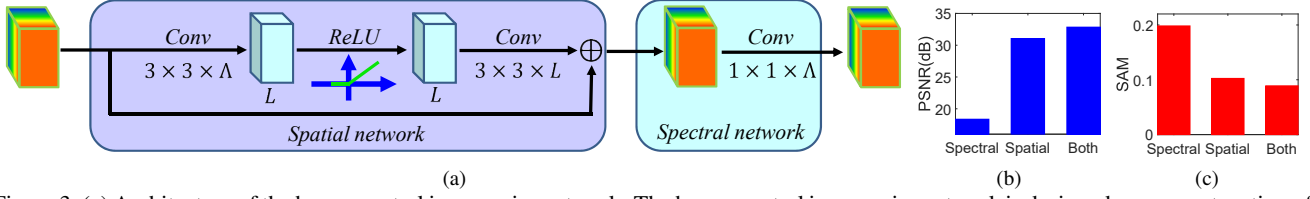
Figure 3. (a) Architecture of the hyperspectral image prior network. The hyperspectral image prior network is designed as a concatenation of spatial network part and spectral network part, which focus on exploit spatial and spectral correlation, respectively. (b) and (c) Performance test with only one part being used.

data term and the regularization term in Eq. (4). Specifically, by introducing an auxiliary variable $\boldsymbol{h}$, Eq. (4) can be reformulated as a constrained optimization problem which is given by

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} ||\boldsymbol{g} - \boldsymbol{\Phi}\boldsymbol{f}||^2 + \tau R(\boldsymbol{h}), \quad s.t. \quad \boldsymbol{h} = \boldsymbol{f}. \quad (5)$$

Then, by adopting the half quadratic splitting (HQS) method, the above constrained optimization problem can be converted as a non-constrained optimization problem

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

where $\eta$ is a penalty parameter. Eq. (6) can be split into two subproblems as

$$\hat{\boldsymbol{f}}^{(k+1)} = \arg\min_{\boldsymbol{f}} ||\boldsymbol{g} - \boldsymbol{\Phi}\boldsymbol{f}||^2 + \eta ||\boldsymbol{h}^{(k)} - \boldsymbol{f}||^2, \quad (7)$$

$$\hat{\boldsymbol{h}}^{(k+1)} = \arg\min_{\boldsymbol{h}} ||\boldsymbol{h} - \boldsymbol{f}^{(k+1)}||^2 + \frac{\tau}{\eta} R(\boldsymbol{h}). \quad (8)$$

We can see that the $h$-subproblem in Eq. (8) is a proximal operator of the hyperspectral image prior $R(\cdot)$ with penalty $\tau$. When the hyperspectral image prior uses the $l_1$ sparsity, the proximal operator implies simply soft-thresholding on $\boldsymbol{f}^{(k+1)}$. From this perspective, the HQS method separates the observation model $\boldsymbol{\Phi}$ and the hyperspectral image prior $R(\cdot)$. The hyperspectral image prior solely appears in the form of the proximal operator. Therefore, instead of explicitly learning a hyperspectral image prior $R(\cdot)$ and solving the proximal operator with the constraint of the hyperspectral image prior, we propose to directly learn a solver $S(\cdot)$ for the proximal operator with a hyperspectral image prior network as follows:

$$\boldsymbol{h}^{(k+1)} = S(\boldsymbol{f}^{(k+1)}). \quad (9)$$

In this manner, the hyperspectral image prior is not explicitly modeled, but is learned with the hyperspectral image prior network, which introduces nonlinearity in prior modeling and avoids the inaccuracy of the explicit hand-crafted image priors, such as a TV prior or sparsity prior.

The architecture of the hyperspectral image prior network is illustrated in Figure 3a. There are two technical insights that guide the design of the hyperspectral image prior network. First, it should enable to exploit spatial and spectral correlation simultaneously. Second, it should be as simple as possible to facilitate the training. Following the first insight, the proposed prior network consists of two main parts, *i.e.*, the spatial network part and the spectral network part, which focuses on exploiting the spatial correlation and spectral correlation, respectively. Following the second insight, the spatial network part employs residual network structure, since residual learning enables fast and stable training which relieves the computational burden [30, 35]. There are two linear convolutional layers interleaved by one rectified linear unit (ReLU) layer. This spatial network design is motivated by the recent work on image spatial super-resolution with excellent performance by removing the unnecessary layers (such as batch normalization) in neural networks [35]. The spectral network contains one convolutional layer to reach a simple architecture and focus on exploiting the spectral correlation. Figure 3b and Figure 3c show the performance test when only one part in the prior network being used. It can been seen that both the spatial network part and the spectral network part have an impact on the final performance, which validates the technical insights for the network design. As introduced in Section 3.1, the input of the hyperspectral image prior network is an oblique parallelepiped cube, which corresponds to a patch on the compressive image. The first convolutional layer uses $3 \times 3 \times \Lambda$ filters and produces $L$ features, while the second convolutional layer uses $3 \times 3 \times L$ filters and produces $\Lambda$ features. The spectral network uses $1 \times 1 \times \Lambda$ filters.

### 3.3. Optimization-inspired Reconstruction Method

We propose to solve Eqs. (7) and (8) in a united framework. Compared with the splitting and iterative way, the proposed framework re-bridge the observation model and the image prior as a whole. Recall that $f$-subproblem in Eq. (7) is a quadratic regularized least-squares problem which ensures the data fidelity. A direct solution is given in the closed form as

$$\boldsymbol{f}^{(k+1)} = (\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi} + \eta\boldsymbol{I})^{-1}(\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{g} + \eta\boldsymbol{h}^{(k)}). \quad (10)$$
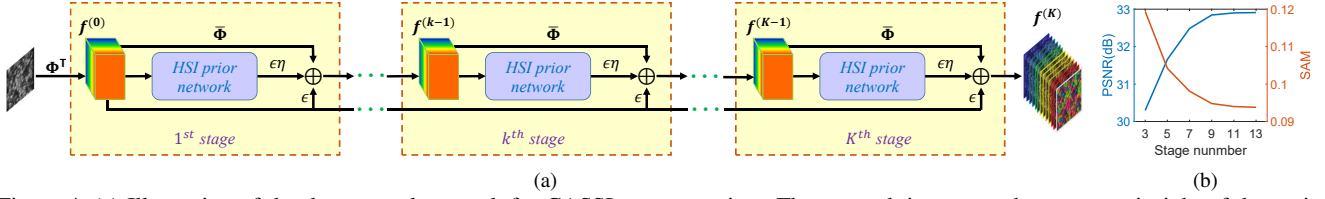
Figure 4. (a) Illustration of the deep neural network for CASSI reconstruction. The network integrates the structure insight of the optimization. It is composed of multiple stages, and each stage includes one hyperspectral image prior network paralleled with two linear connections that accords with the observation model. (b) Impact of the stage number on the reconstruction accuracy. For more than 9 stages, there is no significant increase in PSNR and decrease in SAM.

where $\boldsymbol{I}$ is an identity matrix. Since $\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi} + \eta\boldsymbol{I}$ is very large, it is computationally expensive to invert that matrix. Instead, we employ the conjugate gradient (CG) algorithm to solve the $f$-subproblem. Thus, the solution of Eq. (7) can be expressed as

$$\begin{aligned}
\boldsymbol{f}^{(k+1)} &= \boldsymbol{f}^{(k)} - \epsilon[\boldsymbol{\Phi}^\mathsf{T}(\boldsymbol{\Phi}\boldsymbol{f}^{(k)} - \boldsymbol{g}) + \eta(\boldsymbol{f}^{(k)} - \boldsymbol{h}^{(k)})] \\
&= \bar{\boldsymbol{\Phi}}\boldsymbol{f}^{(k)} + \epsilon\boldsymbol{f}^{(0)} + \epsilon\eta\boldsymbol{h}^{(k)},
\end{aligned}$$
(11)

where $\epsilon$ is the step size in gradient descent, $\boldsymbol{f}^{(0)} = \boldsymbol{\Phi}^\mathsf{T}\boldsymbol{g}$ represents the initialization and $\bar{\boldsymbol{\Phi}} = (1 - \epsilon\eta)\boldsymbol{I} - \epsilon\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}$. Then, we can unify the two subproblems as a single problem by substituting Eq. (9) into Eq. (11)

$$\boldsymbol{f}^{(k+1)} = \bar{\boldsymbol{\Phi}}\boldsymbol{f}^{(k)} + \epsilon\boldsymbol{f}^{(0)} + \epsilon\eta S(\boldsymbol{f}^{(k)}). \quad (12)$$

We propose to implement the deduction in Eq. (12) via a deep neural network as shown in Figure 4a. The network is composed of multiple stages, each of which includes one hyperspectral image prior network as introduced in Section 3.2 paralleled with two linear connections that accords with Eq. (12). In the proposed network, given the initial hyperspectral image estimation $\boldsymbol{f}^{(0)}$, the stages are concatenated in a feed-forward manner. Different from the iteration-based optimization, the network is trained in an end-to-end manner to obey the observation model and exploit the image prior 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 modeled by the transpose of the measurement matrix $\boldsymbol{\Phi}^\mathsf{T}$. The output vector is reshaped to an oblique parallelepiped cube which is treated as the initialization as $\boldsymbol{f}^{(0)} = \boldsymbol{\Phi}^\mathsf{T}\boldsymbol{g}$. For the $k$th stage, its updated result $\boldsymbol{f}^{(k)}$ comes from three parts as shown in Eq. (12). The first part is derived from the previous result $\boldsymbol{f}^{(k-1)}$ which is fed into the hyperspectral image prior network and then weighted by the parameter $\epsilon\eta$; The second part is also from the previous result $\boldsymbol{f}^{(k-1)}$, which is fed into a linear layer parameterized by $\bar{\boldsymbol{\Phi}}$; The third part is a skip connection to the initialization $\boldsymbol{f}^{(0)}$ weighted by parameter $\epsilon$. This stage is repeated $K$ times. Figure 4b shows the impact of the stage number on

the reconstruction accuracy. It can be seen that after 9 stages there is no significant improvement in accuracy. Given the tradeoff between accuracy and memory, we set the stage number to 9 in the following simulations and experiments.

### 3.4. Adaptive Parameters Learning

We propose to train the network by end-to-end training to learn the network parameters $\boldsymbol{\Theta}$ and the optimization parameters $\epsilon$ and $\eta$ simultaneously. In our implementation, all the parameters are set to be different among each stage, as with the stage increasing, the reconstruction quality is improved, thus the network parameters and the optimization parameters should be adaptively changed.

Given a set of oblique parallelepiped cubes $\boldsymbol{f}_{(l)}$ as the training samples and its corresponding compressive patch $\boldsymbol{g}_{(l)}$, the network is trained according to the MSE-based loss function, which can be expressed as

$$(\hat{\boldsymbol{\Theta}}, \hat{\epsilon}, \hat{\eta}) = \arg\min_{\boldsymbol{\Theta}, \epsilon, \eta} \frac{1}{L} \sum_{l=1}^{L} ||\widetilde{\boldsymbol{f}}(\boldsymbol{g}_{(l)}; \boldsymbol{\Theta}, \epsilon, \eta) - \boldsymbol{f}_{(l)}||^2,$$
(13)

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

We employ MatConvNet to implement the network, minimize the loss function in Eq. (13) using the stochastic gradient descent method, and train it up to 150 epochs. The mini-batch size and momentum parameter are set to 64 and 0.9, respectively. The learning rate is set to $10^{-3}$. The network parameters for each layers are initialized with the method in [36]. The optimization parameters are initialized with all zeros. We use a machine equipped with an Intel Core i7-6800K CPU with 64GB memory and an NVIDIA Titan X PASCAL GPU with 12GB memory.

## 4. Simulation Results on Synthetic Data

### 4.1. Configurations

For comprehensive evaluation, we conduct simulations on the public ICVL [37], Harvard [38] and KAIST [25] hyperspectral image datasets. We follow the principles in [39, 25, 33] to partition the training and testing sets. We

Table 1. Performance comparisons on the ICVL and Harvard datasets. The best performance is labeled in bold and the second best performance is underlined.

| Dataset | Metric | TwIST | GPSR | AMP | 3DNSR | SSLR | HSCNN | ISTA-Net | Autoencoder | Proposed$_D$ | Proposed$_I$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ICVL | PSNR | 26.15 | 24.56 | 26.77 | 27.95 | 29.16 | 29.48 | 31.73 | 30.44 | <u>33.43</u> | **34.13** |
| | SSIM | 0.936 | 0.909 | 0.947 | 0.958 | 0.964 | 0.973 | 0.984 | 0.970 | <u>0.990</u> | **0.992** |
| | SAM | 0.053 | 0.09 | 0.052 | 0.051 | 0.046 | 0.043 | 0.042 | 0.036 | <u>0.030</u> | **0.028** |
| Harvard | PSNR | 27.16 | 24.96 | 26.67 | 28.51 | 29.68 | 28.55 | 31.13 | 30.30 | <u>32.44</u> | **32.84** |
| | SSIM | 0.924 | 0.907 | 0.935 | 0.94 | 0.952 | 0.944 | 0.967 | 0.952 | <u>0.976</u> | **0.979** |
| | SAM | 0.119 | 0.196 | 0.155 | 0.132 | 0.101 | 0.118 | 0.114 | 0.098 | <u>0.093</u> | **0.089** |
| Time(s) | | 555 | 302 | 705 | 8648 | 6986 | 3.11 | <u>1.15</u> | 521 | **1.11** | **1.11** |

set the patch size and the number of feature map in prior network to $64 \times 64$ and 64, respectively. The corresponding coded aperture $C(m, n)$ is constructed by generating a binary matrix in Bernoulli distribution with $p = 0.5$.

We compare our methods with several state-of-the-art methods, including five hand-crafted prior-based methods, *i.e.*, TwIST with TV prior [15], GPSR and AMP with sparisty prior [18, 26], and 3DNSR and SSLR with NLS prior [19, 20], and three learning-based methods, *i.e.*, HSCNN [33], ISTA-Net[1] [23] and Autoencoder [25]. All the codes for the competitive methods are released publicly or provided privately to us by the authors, and we make great effort to produce their best results. For the proposed method, there are two kinds of variation, *i.e.*, the coded apertures used in training and testing are identical (indicated by Proposed$_I$) or different (indicated by Proposed$_D$).

Three quantitative image quality metrics are employed to evaluate the performance of these methods, including peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [40] and spectral angle mapping (SAM) [41]. PSNR and SSIM show the spatial fidelity which are calculated on each 2D spatial image, and averaged over all spectral bands. a larger value of PSNR and SSIM indicates better performance. SAM shows the spectral fidelity, which is calculated on each 1D spectral vector and averaged over all spatial points. A smaller value of SAM suggests a better reconstruction.

### 4.2. Evaluation

**Numerical Results.** Table 1 summarizes the numerical results on the ICVL and Harvard datasets. It can be seen that the proposed methods outperform all the existing methods by a large margin according to the metrics in both spatial and spectral domains. For simulations here, the ratio between the pixel numbers of compressive image and the underlying hyperspectral image is extremely high (3.2% for 31 spectral bands), so the reconstruction is challenging for

---

[1] Since it was originally designed for natural imaging, we modified it with great efforts to be applicable to hyperspectral imaging.

Table 2. Performance comparisons on the KAIST dataset. For the case with noise, the noise variance is 0.05.

| | Metric | HSCNN | ISTA-Net | Autoencoder | Proposed$_D$ | Proposed$_I$ |
|---|---|---|---|---|---|---|
| without noise | PSNR | 25.18 | 29.39 | 27.90 | <u>29.77</u> | **30.03** |
| | SSIM | 0.947 | 0.979 | 0.971 | <u>0.986</u> | **0.988** |
| | SAM | 0.153 | 0.140 | 0.136 | <u>0.122</u> | **0.114** |
| with noise | PSNR | 24.90 | <u>29.28</u> | 27.65 | 29.17 | **29.86** |
| | SSIM | 0.931 | 0.968 | 0.970 | <u>0.976</u> | **0.978** |
| | SAM | 0.203 | 0.143 | 0.143 | <u>0.141</u> | **0.136** |

the methods with hand-crafted image priors. This demonstrates the superiority of the prior modeling capability of the hyperspectral image network. Compared with the learning-based methods, the proposed network drives from the inspiration of the optimization, sets the parameters adaptively by end-to-end training, and thus achieves better performance. Even when the coded aperture used in the testing is different with that in the training, the proposed method still produces superior result compared with all the previous methods.

**Noise Test.** We conduct an additional simulation experiment on the KAIST dataset with consideration of imaging noise. Here, we focus on comparing the performance of the learning-based methods. For the case with noise, the noise variance is 0.05. The results are listed in Table 2. As can be seen, our method outperforms the other methods under both the noise-free and the noisy cases, which promotes CASSI to work at a high frame rate.

**Perceptual Quality.** To visualize the reconstruction results, two representative result images of all methods are shown in Figure 5. To simultaneously show the results of all spectral bands, we convert the spectral images to sRGB via the CIE color matching function. The PSNR and SSIM values are provided for each result. Clearly, the proposed method can produce the visually pleasant results with less artifacts and sharper edges compared with other methods, which is consistent with the numerical metrics.

**Spectral Fidelity.** Figure 6 shows the recovered spectrum of two points in the selected image. It can be seen
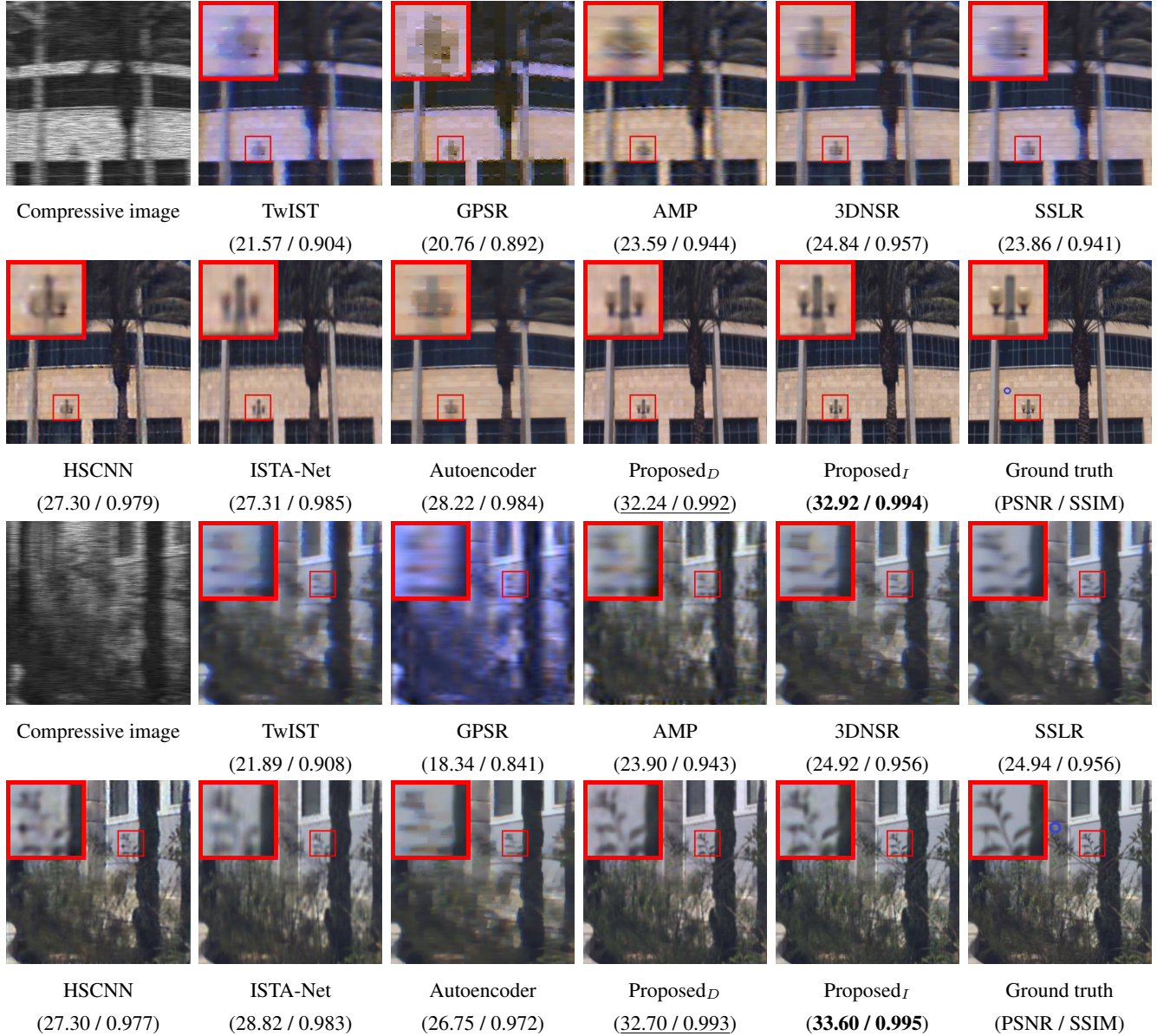
Figure 5. Visual quality comparison. The PSNR and SSIM for the result images are shown in the parenthesis. Our methods outperforms all the competitive methods in terms of spatial and spectral accuracy.
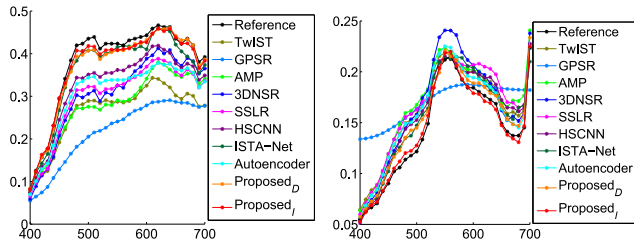
that, among all the results, the spectra reconstructed by the proposed methods are closer to the reference. The overlaid SAM values of the reconstructed spectra further demonstrate the superior performance of the proposed methods on spectral recovery.

**Computational Complexity.** The computational complexity is proportional to the number of multiplications in solving Eq. (8) [25]. In our method, the total number of multiplication for the hyperspectral 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 Eq. (8) and its multiplication is approximately $M \times N \times 10^7$.

with a $2\times$ over-complete dictionary. Further, we record the running time for reconstructing one hyperspectral image with size of $512 \times 512 \times 31$ shown in Table 1. All the codes are implemented on an Intel Core i7-6800K CPU and no specific parallel operation and code optimization are conducted. The proposed method is comparable with ISTA-Net and much faster than the other methods.

## 5. Experiments on a Real CASSI System

In this section, we conduct experiments on real hardware system to demonstrate the practicability of our system. To this end, we build a prototype of a CASSI system,

SAM values of the reconstructed spectra

|     | TwIST | GPSR | AMP | 3DNSR | SSLR | HSCNN | ISTA-Net | Autoencoder | Proposed$_D$ | Proposed$_I$ |
|-----|-------|------|-----|-------|------|-------|----------|-------------|----------|----------|
| (a) | 0.040 | 0.144 | 0.106 | 0.101 | 0.052 | 0.035 | 0.027 | 0.027 | 0.026 | 0.016 |
| (b) | 0.077 | 0.226 | 0.090 | 0.054 | 0.093 | 0.069 | 0.063 | 0.071 | 0.051 | 0.026 |

Figure 6. Comparison of spectral accuracy. The points are indicated in Figure 5. The spectra reconstructed by the our methods are closer to the reference compared with the other methods. The SAM numbers further demonstrate the superiority of our methods on spectral reconstruction.

as shown in Figure 1. The system is made of a 16mm objective lens (AZURE 1614), a coded aperture, a relay lens (Edmund 45762), a dispersive prism and a detector. The detector is a Point Grey FL3-U3-13Y3M with a pixel pitch of 4.9um and $1280 \times 1080$ pixels in total. The coded aperture includes random binary patterns made through lithographic chrome etching on a quartz plate and the pixel pitch is 9.8um. A pixel on the coded aperture corresponds to two-by-two pixel on the detector. The dispersive prism is manufactured by Shanghai Optics, producing a 26-pixel dispersion from 450nm to 650nm. We calibrate the prototype by following the principle in [18] to obtain the optical properties of the system.

To handle real-world scenes from our prototype, we retrain the network by combining the hyperspectral images from these three datasets. The training data is further augmented with scale invariance [42]. We use the configuration of the Proposed$_D$ to train the network for real experiment. Figure 7 shows the reconstructed images of one channel by our method together with TwIST, 3DNSR, and Autoencoder. It can be seen that the proposed methods can produce better results with less artifacts and clearer contents compared with the other methods. Further, we select one patch in the color checker in Figure 7 (indicated by ⊗) and plot the spectral signatures in Figure 8. The reference spectrum is obtained with a commercial spectrometer. The spectrum of our method are closest to the reference and the corresponding RMSE also verify the superior performance of our method.

## 6. Conclusion

We have presented a novel hyperspectral image reconstruction method, which outperforms current state-of-the art methods. There are the two key steps in our method: (1) we
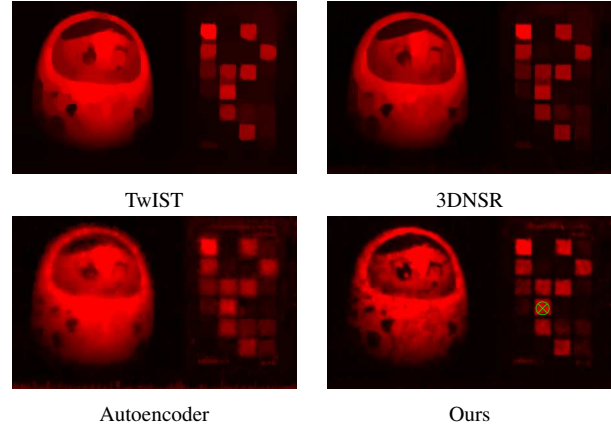


Figure 7. Experiments results on a real CASSI system. The center wavelength for the selected band is 632nm. Our method can achieve results with clearer spatial details.
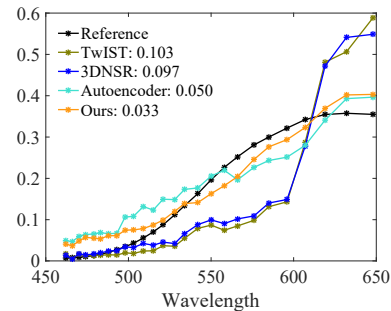


Figure 8. The reconstructed spectra on the selected patch in Figure 7 (indicated by ⊗ ) and the corresponding RMSE. Our method can achieve results with smaller spectral errors.

learn a novel data-driven prior that regularizes the optimization problem to exploit the spatial-spectral correlation; (2) we combine our regularizer with an optimization-inspired network to enable end-to-end training. Our reconstruction method also reduces the computational cost. We have also built a prototype system to validate the effectiveness of the proposed methods. One future direction of interest is to extend the proposed method for other hyperspectral image processing problem, *e.g.*, hyperspectral interpolation and demosaicing. The other direction is to further accelerate the proposed method to reach a real-time reconstruction, thus enabling to acquire hyperspectral images at a video frame rate.

## Acknowledgments

# References

[1] D. J. Brady, *Optical imaging and spectroscopy*. John Wiley & Sons, 2009. 1

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

[3] J. Solomon and B. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1152, 1985. 1

[4] 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*, vol. 31, no. 4, pp. 38:1–38:11, 2012. 1

[5] 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*, 2010, pp. 44–51. 1

[6] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1552–1560, 2003. 1

[7] J. James, *Spectrograph design fundamentals*. Cambridge University Press, 2007. 1

[8] 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*, vol. 33, no. 5, pp. 95–108, 2016. 1

[9] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Transactions on on Graphics*, vol. 33, no. 6, p. 233, 2014. 1

[10] H. Du, X. Tong, X. Cao, and S. Lin, "A prism-based system for multispectral video acquisition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 175–182. 1

[11] Y. Y. Schechner and S. K. Nayar, "Generalized mosaicing: Wide field of view multispectral imaging," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 24, no. 10, pp. 1334–1348, 2002. 1

[12] S.-H. Baek, I. Kim, D. Gutierrez, and M. H. Kim, "Compact single-shot hyperspectral imaging using a prism," *ACM Transactions on on Graphics*, vol. 36, no. 6, p. 217, 2017. 1

[13] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006. 1

[14] G. Arce, D. Brady, L. Carin, H. Arguello, and D. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 105–115, 2014. 1

[15] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *OSA Applied Optics*, vol. 49, no. 36, pp. 6824–6833, 2010. 1, 2, 6

[16] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," *OSA Applied Optics*, vol. 54, no. 4, pp. 848–858, 2015. 1

[17] 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*, 2015, pp. 4942–4950. 1

[18] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *OSA Applied Optics*, vol. 47, no. 10, pp. B44–B51, 2008. 1, 2, 6, 8

[19] 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*, vol. 39, no. 10, pp. 2104–2111, 2017. 1, 2, 6

[20] 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*, 2016, pp. 3727–3736. 1, 2, 6

[21] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *International Conference on Machine Learning*. Omnipress, 2010, pp. 399–406. 1, 2

[22] J. Sun, H. Li, Z. Xu *et al.*, "Deep admm-net for compressive sensing mri," in *Advances in Neural Information Processing Systems*, 2016, pp. 10–18. 1, 2

[23] 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*, June 2018, pp. 1828–1837. 1, 2, 6

[24] 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*, 2017, pp. 2808–2817. 1, 2

[25] 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)*, vol. 36, no. 6, p. 218, 2017. 2, 5, 6, 7

[26] J. Tan, Y. Ma, H. Rueda, D. Baron, and G. R. Arce, "Compressive hyperspectral imaging via approximate message passing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 389–401, 2016. 2, 6

[27] A. Rajwade, D. Kittle, T.-H. Tsai, D. Brady, and L. Carin, "Coded hyperspectral imaging and blind compressive sensing," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 782–812, 2013. 2

[28] Y. Liu, X. Yuan, J. Suo, D. Brady, and Q. Dai, "Rank minimization for snapshot compressive imaging," *IEEE Transactions Pattern Analysis and Machine Intelligence*, 2018. 2

[29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015. 2

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. 2, 4

[31] 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*, 2016, pp. 449–458. 2

[32] M. Iliadis, L. Spinoulas, and A. K. Katsaggelos, "Deep fully-connected networks for video compressive sensing," *Digital Signal Processing*, vol. 72, pp. 9–18, 2018. 2

[33] 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*, vol. 2, 2017. 2, 5, 6

[34] S. Roth and M. J. Black, "Fields of experts," *International Journal of Computer Vision*, vol. 82, no. 2, p. 205, 2009. 3

[35] 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*, vol. 1, no. 2, 2017, p. 4. 4

[36] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256. 5

[37] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *European Conference on Computer Vision*, 2016, pp. 19–34. 5

[38] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 193–200. 5

[39] 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*, vol. 28, no. 5, pp. 2257–2270, 2019. 5

[40] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. 6

[41] 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*, vol. 44, no. 2-3, pp. 145–163, 1993. 6

[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015. 8