Hyperspectral Image Reconstruction using Deep External and Internal Learning

Tao Zhang Ying Fu Lizhi Wang Hua Huang
School of Computer Science and Technology, Beijing Institute of Technology
{tzhang,fuying,lzwang,huahuang}@bit.edu.cn

Abstract

To solve the low spatial and/or temporal resolution problem which the conventional hypelrspectral cameras often suffer from, coded snapshot hyperspectral imaging systems have attracted more attention recently. Recovering a hyperspectral image (HSI) from its corresponding coded image is an ill-posed inverse problem, and learning accurate prior of HSI is essential to solve this inverse problem. In this paper, we present an effective convolutional neural network (CNN) based method for coded HSI reconstruction, which learns the deep prior from the external dataset as well as the internal information of input coded image with spatial-spectral constraint. Our method can effectively exploit spatial-spectral correlation and sufficiently represent the variety nature of HSIs. Experimental results show our method outperforms the state-of-the-art methods under both comprehensive quantitative metrics and perceptive quality.

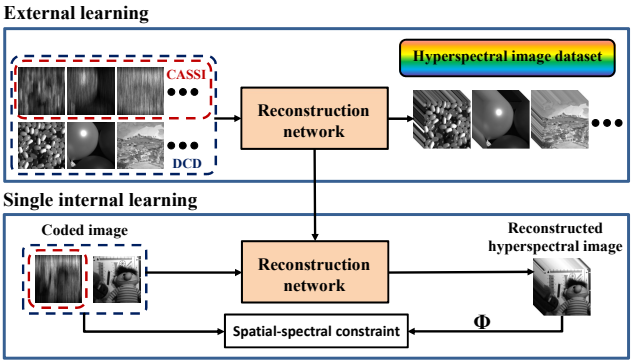


Figure 1. Overview of our CNN-based coded HSI reconstruction method. The reconstruction network is first learned from an external dataset, and then customized with spatial-spectral constraint in term of the internal information of the input coded image for each target scene.

1. Introduction

Hyperspectral imaging systems capture the spectral signature of each spatial location in a scene with much more than three bands. The rich spectral details in HSI can show deterministic information about the lighting and material, which is beneficial to various fields, including remote sensing [5, 28], computer vision [7] and medical diagnosis [4, 27].

To obtain a full 3D HSI, the conventional hyperspectral imagers needs multiple exposures to scan the scene [2, 30, 31], which is time-consuming and cannot capture dynamic scenes. To improve the temporal resolution, various snapshot hyperspectral imaging systems [6, 10, 29] have been proposed by multiplexing the 3D HSI into a 2D spatial sensor, which, however, sacrifice the spatial resolution. Recently, by leveraging the compressive sensing (CS) theory, coding-based hyperspectral imaging techniques [16, 22, 25, 26, 35, 37] have attracted increasing attention due to the potential to overcome the trade-off between temporal and spatial resolution.

With elaborate optical design, these coding-based techniques encode the 3D HSI into a 2D compressive measurement. Now the bottleneck lies in how to faithfully reconstruct the desirable HSI. Since the reconstruction problem is under-determined, prior knowledge of the unknown HSI is required to regularize the reconstruction. To this end, the well-known model-based methods employ various handcrafted priors, such as the total variation [22, 37], sparsity [35, 38, 39] and low-rankness [13]. However, these handcrafted priors are insufficient to represent the variety of the real-world spectral data.

Different from the model-based methods that rely on carefully designed priors, the learning-based methods [9, 40, 43] can implicitly learn the prior by leveraging the external dataset. However, the learning-based methods [40] often attempt to fit a brute-force mapping between the compressive image and the desirable image, which ignores the internal imaging model. Thus, the learned mapping function would be ineffective even if the observation model deviates very slightly from that one used to synthesize the training data. Recently, an autoencoder-based method [9] is proposed by pre-training an autoencoder to exploit the

image prior and integrating it to the model-based reconstruction framework. However, the autoencoder cannot be jointly optimized with the other parameters of the optimization algorithm, leaving a headache of parameters tuning. Moreover, all the learning-based methods rely on an assumption of strong prior similarity between the training and testing data. In fact, different kinds of hyperspectral imager would produce heterogeneous HSIs with totally different centric wavelength and spectral response. So the learning-based methods usually suffer from the problem of over-fitting and lacking generalization ability.

In this paper, we present a CNN-based coded HSI reconstruction method by jointly exploiting deep external and internal learning (Figure 1). First, we develop a CNN-based channel attention reconstruction network to effectively exploit the spatial-spectral correlation of the HSI. Then, we train the reconstruction network by leveraging an arbitrary external dataset to exploit the general spatial-spectral correlation. Finally, we customize the network by internal learning with spatial-spectral constraint by the coded image, which makes use of the internal imaging model to learn specific prior for current desirable image. Our method are verified on two representative snapshot hyperspectral imaging systems, i.e., CASSI [35] and DCD[37].

In summary, our main contributions are that we

1. present a CNN-based method for coded HSI reconstruction, which can effectively combine deep external and internal learning;
2. exploit the spatial-spectral correlation of the HSI by external learning;
3. guarantee generalization ability and adapt itself to variant scenes by internal learning.

2. Related Work

In this section, we review the most relevant studies on hyperspectral imaging system and coded HSI reconstruction methods.

2.1. Hyperspectral Imaging Systems

Conventional hyperspectral cameras usually make trade-off between temporal/spatial and spectral resolution. Scanning-based acquisition systems [2, 30, 31, 44] captured the full scene pointwisely/linewisely in spatial domain or bandwisely in spectral domain. All these systems sacrifice the temporal resolution. To capture dynamic scene, several snapshot spectral imagers have been proposed, which multiple the 3D HSI into a 2D spatial sensor but trade the spatial resolution for spectral resolution [6, 11, 14, 15].

Recently, to overcome the trade-off between temporal and spatial resolution, several coding-based systems have been prevalent by relying on CS theory. CASSI utilized two dispersers [16] or one disperser [35] with a coded aperture to optically encode the spectral signal along spatial dimension. To improve the performance of CASSI, multiple shots with varying coded apertures [22] and DCD with an aligned panchromatic camera [37] have been proposed. Besides, a dual-coded compressive spectral imager [26] was proposed as an advancement of CASSI, which separately encoded spatial and spectral dimensions using a digital micromirror device and a liquid crystal on silicon, respectively. Later, [26] was upgraded to spatial-spectral encoded compressive spectral imager [25], which jointly considers the spatial-spectral coding mechanism but only employs one coded aperture. In this paper, we mainly vertify the effectiveness of our method on CASSI [35] and DCD[37]

2.2. Coded HSI Reconstruction Methods

Model-based methods often formulate coded HSI reconstruction as a Maximum a Posterior problem. Various hand-crafted priors are served as regularizers in these methods and the HSI is reconstructed by solving an optimization problem. Kittle *et al.* [22] and Wang *et al.* [37] employed two-step iterative shrinkage/thresholding algorithm [3] with the total variation prior. The total variation prior [22, 37] can effectively model the piecewise smooth spatial structure, but the recovered HSI trend to be over-smooth and lack details. The sparsity prior [33, 35, 38, 39] has shown better results than the total variation prior. Wagadarikar *et al.* [35] employed the gradient projection for sparse reconstruction algorithm. Tan *et al.* [33] utilized approximate message passing by integrating Winner filter as a denoiser in each iteration. Wang *et al.* [38] utilized an adaptive dictionary based algorithm to reconstruct 4D hyperspectral video and the sparse basis is learned from panchromatic image of DCD. Wang *et al.* [39] further integrated non-local similarity with sparse representation to improve performance. Besides, the low-rankness prior [13] is also used for coded HSI reconstruction. Fu *et al.* [13] exploited the spectral-spatial correlation with low-rank approximation and non-local similarity for this reconstruction task.

Recently, by leveraging the power of deep learning, learning-based approaches have been presented for coded HSI reconstruction [9, 40, 43]. Xiong *et al.* [43] first initially reconstructed the HSI via an existing model-based method [22], and then employed a CNN-based method to enhance the initialized result with prior trained on the training set. Choi *et al.* [9] trained the autoencoder to learn nonlinear spectral representation as a deep prior, and then combined it with the total variation prior in the optimization as regularizers to reconstruct HSI. Wang *et al.* [40] employed an end-to-end CNN-based method for coded HSI reconstruction, considering the spatial correlation between neighboring spatial locations and spectral correlation between neighboring bands.

The hand-crafted priors only model the linear character-

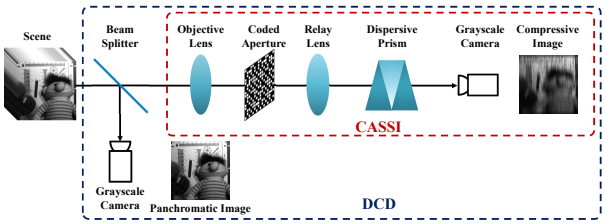


Figure 2. Illustration of two representative imaging systems.

istic in HSI, therefore, are insufficient to exploit the nonlinearity in HSI. The deep prior only learned from external dataset lacks similarity with which the test image desire, thus often fails to work for unknown data. In this work, we present an efficient CNN-based method for coded HSI reconstruction to learn the deep prior from external dataset and internal input image, combining deep external and internal learning. Neither delicate hand-crafted priors are required nor strong prior similarity is relied on in our method.

3. Coded HSI Reconstruction

In this section, we first formulate the problem for coded HSI reconstruction. Then, we introduce our CNN-based method for the coded HSI reconstruction, which can effectively learn spatial-spectral correlation in the HSI for each target scene by using external and internal learning.

3.1. Observation Model

Here, we mainly show the observation model for two representative snapshot hyperspectral imaging systems, i.e., CASSI [35] and DCD [37].

In the CASSI system, as shown in Figure 2, the incident light is first projected into the coded aperture through the objective lens, which plays a spatial modulation. Then, the modulated incident light goes through the relay lens and is spectrally dispersed by the prism. Finally, the spectral dispersed information is captured by a panchromatic camera. Let $\mathbf{X}(m,n,\lambda)$ indicate the intensity of incident light where $1 \leq m \leq M$ and $1 \leq n \leq N$ index the spatial coordinates and $1 \leq \lambda \leq \Lambda$ indexes the spectral coordinate. The (m,n)-th pixel of the compressive measurement can be represented as

$$y^c(m,n) = \sum_{\lambda=1}^{\Lambda} \varphi(m-\psi(\lambda),n)x(m-\psi(\lambda),n,\lambda)\omega(\lambda), \tag{1}$$

where $\varphi(m,n)$ denotes the transmission function of the coded aperture, $\psi(\lambda)$ denotes the wavelength-dependent dispersion function for the prism, x is the spectral distribution of the (m,n,λ)-th pixel of the HSI, $\omega(\lambda)$ represents the response function of the detector. The CASSI observation model can be rewritten in matrix form as

$$\mathbf{Y}^c = \mathbf{\Phi}^c\mathbf{X}, \tag{2}$$

where $\mathbf{\Phi}^c$ denotes the projection matrix of CASSI and jointly determined by $\varphi(m,n)$, $\psi(\lambda)$ and $\omega(\lambda)$, $\mathbf{Y}^c$ denotes the vectorized representation of the compressive image $y^c(m,n)$, $\mathbf{X}$ is the underlying HSI.

In the DCD system, as shown in Figure 2, the incident light is first divided into two direction by the beam splitter. The light in one direction is captured by CASSI, while the light in another direction is captured by the same kind of panchromatic detector. The (m,n)-th pixel of the panchromatic measurement can be represented as

$$y^p(m,n) = \sum_{\lambda=1}^{\Lambda} x(m,n,\lambda)\omega(\lambda). \tag{3}$$

Similar with the CASSI formulation in Equation (2), Equation (3) can also be rewritten in matrix form as

$$\mathbf{Y}^p = \mathbf{\Phi}^p\mathbf{X}, \tag{4}$$

where $\mathbf{\Phi}^p$ denotes the projection matrix of panchromatic camera and determined by $\omega(\lambda)$, $\mathbf{Y}^p$ is the vectorized representation of the panchromatic image.

The DCD sensing process can be generally expressed as

$$\mathbf{Y}^d = \mathbf{\Phi}^d\mathbf{X}, \tag{5}$$

where $\mathbf{Y}^d = [\mathbf{Y}^c; \mathbf{Y}^p]$ and $\mathbf{\Phi}^d = [\mathbf{\Phi}^c; \mathbf{\Phi}^p]$.

The aim is to reconstruct high quality HSI $\mathbf{X}$ from coded image $\mathbf{Y}^c$ for CASSI and $\mathbf{Y}^d$ for DCD.

3.2. Coded HSI Reconstruction Network

Previous works have shown that effectively exploiting the latent intrinsic properties of the HSI — spatial correlation [40] and spectral correlation [9] — can reconstruct high quality HSI from coded image. To better explore the spatial-spectral correlation in the HSI, we conduct a deep CNN to model the spatial-spectral correlation trough multiple layers of nonlinear transformations with dense connection and channel attention. As shown in Figure 3, the CASSI reconstruction network consists of L Dense Blocks [12, 32] between two convolutional layers. Let C^c_{in} denotes the first convolutional layer and C^c_{out} denotes the last convolutional layer in the reconstruction network. For the l-th Dense Block, the inputs are B^c_0 to B^c_{l-1} and the output can be expressed as

$$B^c_l = \mathcal{D}^c_l(B^c_0, \cdots, B^c_{l-1}), \tag{6}$$

where $B^c_0 = C^c_{in}(\mathbf{Y}^c)$ and $\mathcal{D}^c_l$ denotes the l-th Dense Block function in CASSI reconstruction network, respectively.

In each Dense Block, there are K residual channel attention (RCA) Modules. The k-th RCA Module can be expressed as

$$H^c_{l,k} = \mathcal{A}^c_{l,k}(\mathcal{R}^c_{l,k}(H^c_{l,k-1})) + H^c_{l,k-1}, \tag{7}$$

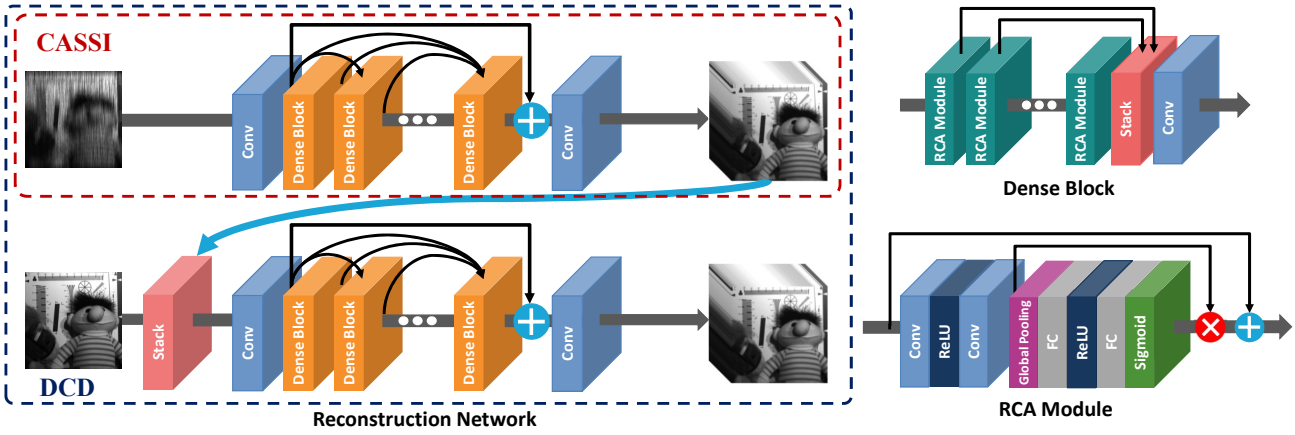


Figure 3. The architecture of coded HSI reconstruction network.

where $H^c_{l,0} = B^c_{l-1}$, $\mathcal{R}^c_{l,k}$ and $\mathcal{A}^c_{l,k}$ denote the residual component [18] function and channel attention component [19, 46] function in k-th RCA Module of l-th Dense Block. Residual component is adaptively rescaled by the channel attention component, which is beneficial to exploit the spectral correlation.

The final output can be expressed as

$$\widehat{\mathbf{X}}^c = C^c_{out}(B^c_L + C^c_{in}(\mathbf{Y}^c)) = f^c(\mathbf{Y}^c, \mathbf{\Theta}^c), \quad (8)$$

where B^c_L denotes the output of L-th Dense Block in the CASSI reconstruction network, f^c is the mapping function of CASSI reconstruction network and $\mathbf{\Theta}^c$ represents parameters in the CASSI reconstruction network, respectively.

For coded HSI reconstruction from DCD, to further fuse the CASSI reconstructed result $\widehat{\mathbf{X}}^c$ and panchromatic image $\mathbf{Y}^p$, we append an identical layout with CASSI reconstruction network, as shown in Figure 3. The final output can be expressed as

$$\begin{aligned}\widehat{\mathbf{X}}^d &= C^p_{out}(B^p_L + C^p_{in}(stack(\widehat{\mathbf{X}}^c, \mathbf{Y}^p))) \\ &= f^d(\mathbf{Y}^c, \mathbf{Y}^p, \mathbf{\Theta}^d),\end{aligned} \quad (9)$$

where B^p_L denotes the output of $2L$-th Dense Block in DCD reconstruction network and f^d is the mapping function of DCD reconstruction network and $\mathbf{\Theta}^d$ represents parameters in the DCD reconstruction network, respectively.

In our coded HSI reconstruction network, we empirically set L, K and feature maps to be 4, 4, 64, respectively. For kernel size, the last convolutional layers in Dense Blocks are set to be 1×1 and the others are set to be 3×3.

3.3. External Learning

To explore latent intrinsic characteristics of the HSI, the deep prior is first learned from external dataset. We uniformly extract from each HSI with size of 256×256 and its corresponding coded image under projection matrix $\mathbf{\Phi}^c$ or $\mathbf{\Phi}^d$ to constitute the patch pairs in the dataset.

For CASSI reconstruction, the model is optimized by minimizing the loss

$$\mathcal{L}^c_{ex}(\mathbf{\Theta}^c_{ex}) = \frac{1}{T}\sum_{t=1}^{T} \|f^c(\mathbf{Y}^c_{ex,t}, \mathbf{\Theta}^c_{ex}) - \mathbf{X}_{ex,t}\|^2, \quad (10)$$

where $\mathbf{Y}^c_{ex,t}$ denotes the t-th external compressive image, $\mathbf{X}_{ex,t}$ represents the t-th corresponding ground truth from external dataset, and $\mathbf{\Theta}^c_{ex}$ is the parameters of the external trained CASSI reconstruction network, T is the number of training samples in the external dataset, respectively.

For DCD reconstruction, to improve the final reconstruction accuracy, we further add constraint for CASSI reconstruction and the loss can be expressed as

$$\begin{aligned}\mathcal{L}^d_{ex}(\mathbf{\Theta}^d_{ex}) = \frac{1}{T}\sum_{t=1}^{T} (&\|f^d(\mathbf{Y}^c_{ex,t}, \mathbf{Y}^p_{ex,t}, \mathbf{\Theta}^d_{ex}) - \mathbf{X}_{ex,t}\|^2 \\ &+ \eta\|f^c(\mathbf{Y}^c_{ex,t}, \mathbf{\Theta}^c_{ex}) - \mathbf{X}_{ex,t}\|^2),\end{aligned} \quad (11)$$

where $\mathbf{Y}^p_{ex,t}$ denotes the t-th panchromatic image from external training dataset, $\mathbf{\Theta}^d_{ex}$ represents the parameters of the external trained DCD reconstruction network, and η is a predefined parameter which we empirically set to be 0.5.

These losses are minimized with the adaptive moment estimation method [21]. We set the mini-batch size, momentum parameter and weight decay to be 1, 0.9 and 10^{-4}, respectively. The learning rate is initially set to be 0.0001, which will be divided by 10 every 30 epochs. All learnable layer's weights are initialized by the method in [17]. The networks are trained with the deep learning tool Caffe [20] on NVIDIA Titan X Pascal GPU.

3.4. Internal Learning

Existing HSI datasets are still kind of small and the distribution variation between training and testing data cannot

be avoided. These may make the external learned deep prior lack similarity with the prior desired by the testing image and less practical for unknown data. Inspired by [34], we further learn the deep prior on a single image using internal learning with spatial-spectral constraint.

For CASSI reconstruction, given the relationship between coded image and corresponding HSI in Equation (2), the latent HSI in Equation (8) should be consistent with the input coded image after the linear mapping $\mathbf{\Phi}^c$. To reduce the effect of distribution variation, the reconstruction network can be updated with spatial-spectral constraint from the input coded image for each scene, which can be represented as

$$\mathcal{L}_{in}^c(\mathbf{\Theta}_{in}^c) = \|\mathbf{\Phi}^c f^c(\mathbf{Y}_{in}^c, \mathbf{\Theta}_{in}^c) - \mathbf{Y}_{in}^c\|^2, \quad (12)$$

where $\mathbf{Y}_{in}^c$ denotes the input coded image, and $\mathbf{\Theta}_{in}^c$ is the parameters of CASSI reconstruction network in internal learning and is initialized by $\mathbf{\Theta}_{ex}^c$.

For DCD reconstruction, given the relationship between the capture image and corresponding HSI in Equation (5), the latent HSI in Equation (9) should be consistent with the input image after the linear mapping $\mathbf{\Phi}^d$. We add the spatial-spectral constraint in Equation (11) for DCD reconstruction and the internal learning loss can be expressed as

$$\begin{aligned}\mathcal{L}_{in}^d(\mathbf{\Theta}_{in}^d) = \|\mathbf{\Phi}^d f^d(\mathbf{Y}_{in}^c, \mathbf{Y}_{in}^p, \mathbf{\Theta}_{in}^d) - \mathbf{Y}_{in}^d\|^2 \\ +\eta\|\mathbf{\Phi}^d f^c(\mathbf{Y}_{in}^c, \mathbf{\Theta}_{in}^c) - \mathbf{Y}_{in}^d\|^2,\end{aligned} \quad (13)$$

where $\mathbf{Y}_{in}^p$ denotes the internal panchromatic image, and $\mathbf{\Theta}_{in}^d$ is the parameters of internal trained DCD reconstruction network and is initialized by $\mathbf{\Theta}_{ex}^d$.

In internal learning, the underlying characteristics of the latent HSI in Equation (12) and Equation (13) is modeled in deep prior instead of hand-crafted priors compared with model-based methods and learned on the input image compared with learning-based methods. Our method effectively combines internal and external information in the deep learning architecture.

The internal learning details are similar with external learning, except we fix the learning rate to be 0.0001 and all learnable layer's weights are initialized by the external learned model.

4. Experimental Results

In this section, we first introduce the datasets and metrics for quantitative evaluation in our experiments. Then, our method is compared with several state-of-the-art methods on synthetic data. In addition, the generalization ability of our method is discussed. Finally, we implement our coded HSI reconstruction method on the real images.

4.1. Datasets and Metrics

Our method is evaluated on three public HSI datasets, including the the CAVE dataset [45], the Harvard dataset [8], and ICVL dataset [1]. The CAVE dataset consists of 32 HSIs and the spatial resolution is 512×512. The Harvard dataset consists of 50 outdoor images captured under daylight illumination, whose spatial resolution is 1024×1392. Due to large-area high-light pixels, we remove 6 deteriorated HSIs. The ICVL dataset consists of 201 images, which is by far the most comprehensive natural HSI dataset. The spatial resolution of HSIs is 1300×1392. The spectral range of CAVE and ICVL datasets are from 400 nm to 700 nm, the spectral range of Harvard dataset is from 420 nm to 720 nm, and the spectral range of all three HSI datasets are divided into 31 spectral bands with 10 nm interval. We random select 10 images in CAVE dataset, 9 images in Harvard dataset and 50 images in ICVL dataset for testing and the rest for training, respectively.

Three quantitative image quality metrics are utilized to evaluate the performance of all methods, including peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [41], spectral angle mapping (SAM) [23] and relative dimensionless global error in synthesis (ERGAS) [36]. Larger values of PSNR and SSIM suggest better performance, while a smaller value of SAM and ERGAS implies a better reconstruction.

4.2. Evaluation on Synthetic Data

We compare our method with six state-of-the-art HSI reconstruction methods on synthetic data, including three model-based methods i.e., total variation based method (TV) [3], sparse representation based method (NSR) [39], and low-rank matrix approximation based method (LRMA) [13] and three learning-based methods, i.e., the HSCNN method[43], the Autoencoder method [9] and the HyperReconNet method [40]. We make great effort to reproduce the best results for competitive methods with the codes that are released publicly or provided privately by the authors.

Table 1 and Table 2 provide the averaged reconstructed results for CASSI and DCD over all test images on three datasets, to quantitatively compare our method with TV, NSR, LRMA, HSCNN, Autoencoder and HyperReconNet. Note that HyperReconNet is specially designed for CASSI reconstruction and difficultly extended for DCD reconstruction, so we only compare with it on CASSI reconstruction. The best results are in bold on each dataset. It can be seen that DCD always has much better performance than CASSI, which demonstrates the advantage of DCD. Comparing the results with different methods in the same system, our method outperforms the existing methods in most case according to the metrics in spatial and spectral domains. This reveals the advantages of deeply exploiting the intrinsic properties of HSIs and verifies the effectiveness of our deep external and internal learning method.

To visualize the experimental results, three representative restored results on three datasets are shown in Figures

Table 1. Evaluation CASSI reconstructed results of different methods on three HSI datasets.

Methods	CAVE				Harvard				ICVL			
	PSNR	SSIM	SAM	ERGAS	PSNR	SSIM	SAM	ERGAS	PSNR	SSIM	SAM	ERGAS
TV	24.099	0.8917	8.928	41.132	27.163	0.9242	6.800	37.141	26.155	0.9358	3.020	15.971
NSR	26.644	0.9247	**7.440**	32.085	28.508	0.9400	7.572	32.722	27.949	0.9576	2.939	12.468
LRMA	25.871	0.9198	8.473	34.545	30.113	0.9572	**5.086**	26.506	29.975	0.9720	1.819	9.921
HSCNN	25.017	0.9151	11.911	38.836	28.548	0.9442	6.759	30.758	29.475	0.9733	2.469	10.034
Autoencoder	25.741	0.9186	8.506	35.542	30.300	0.9520	5.615	26.441	30.440	0.9700	2.063	10.485
HyperReconNet	24.444	0.9043	12.996	40.913	30.341	0.9644	6.609	25.358	32.363	0.9861	2.121	7.295
Ours	**29.055**	**0.9570**	8.260	**24.786**	**33.585**	**0.9824**	5.362	**17.138**	**35.884**	**0.9937**	**1.462**	**4.847**

Table 2. Evaluation DCD reconstructed results of different methods on three HSI datasets.

Methods	CAVE				Harvard				ICVL			
	PSNR	SSIM	SAM	ERGAS	PSNR	SSIM	SAM	ERGAS	PSNR	SSIM	SAM	ERGAS
TV	34.090	0.9849	5.817	14.445	36.294	0.9892	5.114	15.937	36.902	0.9914	1.807	5.612
NSR	35.890	0.9864	5.105	13.087	38.564	0.9934	4.553	12.801	39.915	0.9960	1.404	3.774
LRMA	38.370	0.9933	5.019	9.188	40.429	0.9959	4.110	9.737	41.124	0.9972	1.322	3.090
HSCNN	33.834	0.9852	7.880	14.899	37.492	0.9924	4.845	12.652	38.797	0.9966	1.516	3.551
Autoencoder	33.863	0.9874	7.294	14.816	39.052	0.9952	4.679	10.096	41.499	0.9983	1.225	2.624
Ours	**38.458**	**0.9954**	**4.179**	**8.454**	**41.825**	**0.9975**	**3.972**	**7.360**	**44.355**	**0.9991**	**0.981**	**1.847**

Table 3. Evaluation generalization ability of distribution variation between training and testing data and all models are externally trained on ICVL dataset.

Methods	CAVE				Harvard				ICVL			
	PSNR	SSIM	SAM	ERGAS	PSNR	SSIM	SAM	ERGAS	PSNR	SSIM	SAM	ERGAS
CASSI												
HSCNN	22.219	0.8824	17.945	54.624	26.973	0.9197	10.715	42.144	29.475	0.9733	2.469	10.034
Autoencoder	22.200	0.8348	18.142	56.088	25.234	0.8785	16.258	58.358	30.441	0.9702	2.063	10.485
HyperReconNet	21.171	0.7866	23.320	69.475	24.765	0.8636	14.510	71.778	32.363	0.9861	2.121	7.295
Ours-external	19.974	0.7201	24.825	77.604	24.807	0.8873	17.745	68.265	32.012	0.9862	2.629	7.768
Ours	**26.888**	**0.9320**	**10.882**	**32.614**	**32.767**	**0.9792**	**6.054**	**19.186**	**35.884**	**0.9937**	**1.462**	**4.847**
DCD												
HSCNN	28.965	0.9320	11.615	30.666	34.625	0.9744	7.439	27.728	38.797	0.9966	1.516	3.551
Autoencoder	26.849	0.9309	18.747	38.914	29.148	0.9402	17.480	45.519	41.499	0.9983	1.225	2.624
Ours-external	22.199	0.9757	24.751	68.439	27.138	0.9162	24.153	62.007	37.282	0.9954	2.006	4.477
Ours	**38.377**	**0.9952**	**4.283**	**8.840**	**41.465**	**0.9972**	**4.044**	**7.770**	**44.355**	**0.9991**	**0.981**	**1.847**

4 and 5. The error maps are the average absolution errors between ground truth and restored results across spectra. To show the scenes, we convert the original HSIs to RGBs via the CIE color mapping function, as shown in the last column of Figure 5. The recovered results from our method are consistently more accurate for all scenes, which demonstrates our method can provide higher spatial accuracy. The absolute error between ground truth and reconstructed results of scenes in Figures 4 and 5 along spectra for all methods are shown in Figure 6. It can be seen that the results of our method are much closer to the ground truth, which verifies that our method obtain higher spectral fidelity.

4.3. Ablation Study

Due to the space limitation, we only show the results of the model externally trained on ICVL dataset.

To evaluate the generalization ability on distribution variation between training and testing data, we compare our method with other learning-based methods on different testing set, and all reconstruction networks are trained on ICVL dataset. Our method with only external learning on ICVL dataset is denoted as 'Ours-external'. According to the Table 3, we can seen that the performance of learning-based methods and our external learned model decrease dramatically on CAVE and Harvard testing sets, while our method combining external and internal learning can obviously improve the spatial accuracy and spectral fidelity. It verifies our method has better generalization ability on data distribution variation.

4.4. Evaluation on Real Data

We further evaluate the effectiveness of our method on the real data. A cartoon cover under the laboratory ambient light condition with complex texture is captured by CASSI and DCD imaging systems. The compressive image of CASSI is shown in Figure 7(a) and the panchro-

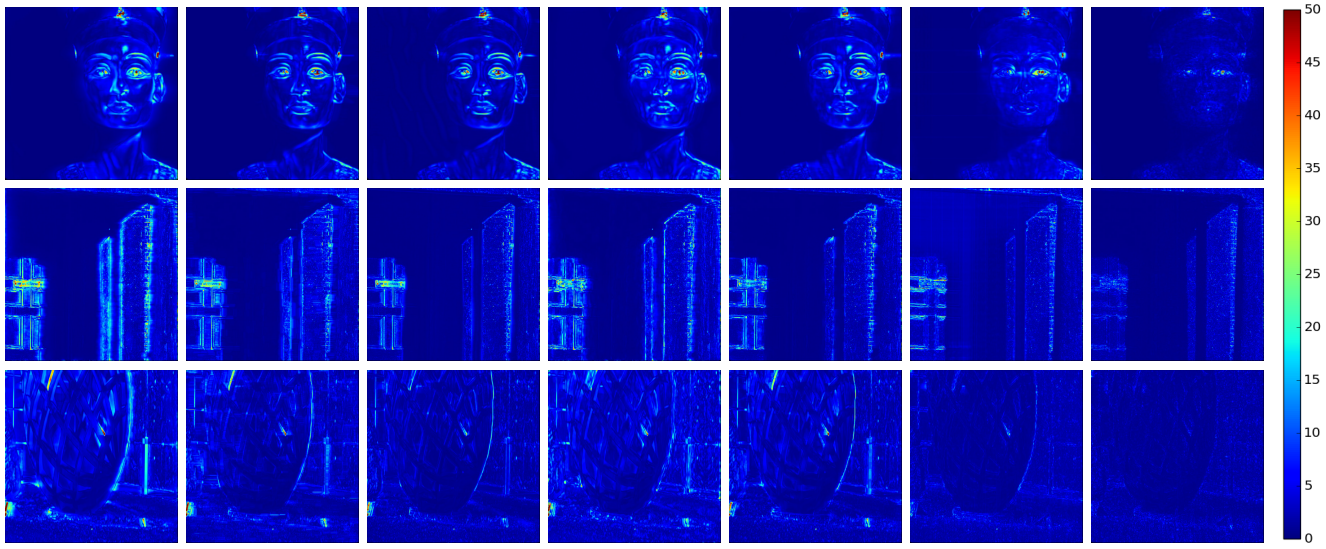


Figure 4. Visual quality comparison on three typical scenes in HSI datasets for CASSI. The error maps for TV/NSR/LRMA/HSCNN/Autoencoder/HyperReconNet/our CASSI recovered results are shown from left to right and the corresponding scenes are shown in Figure 5.

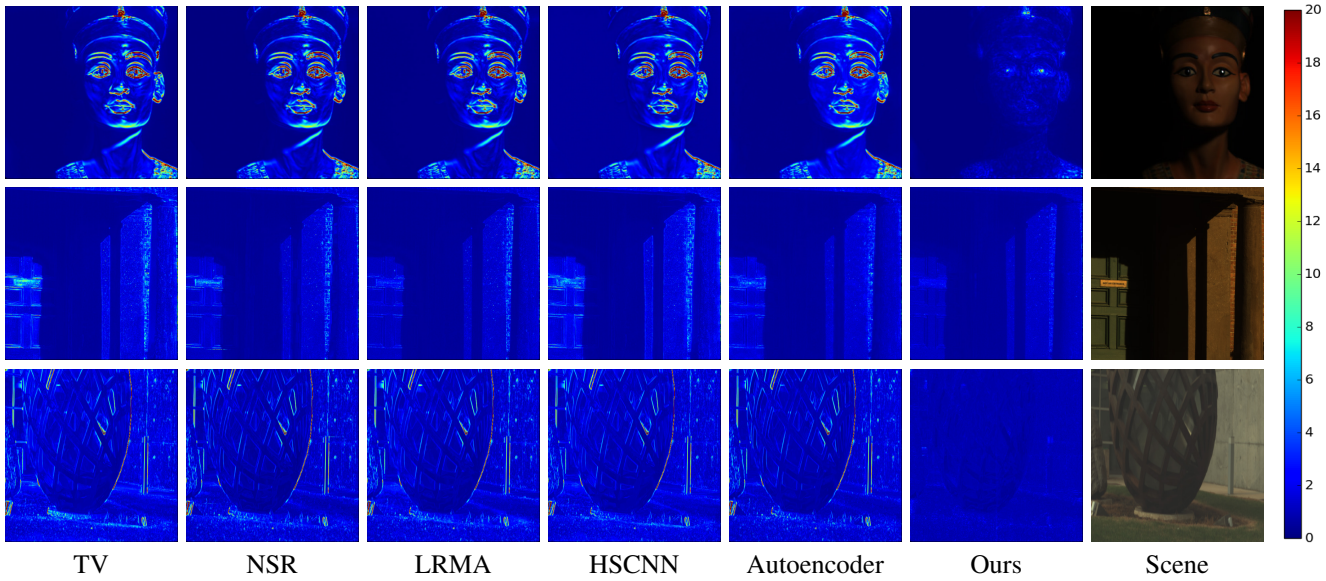
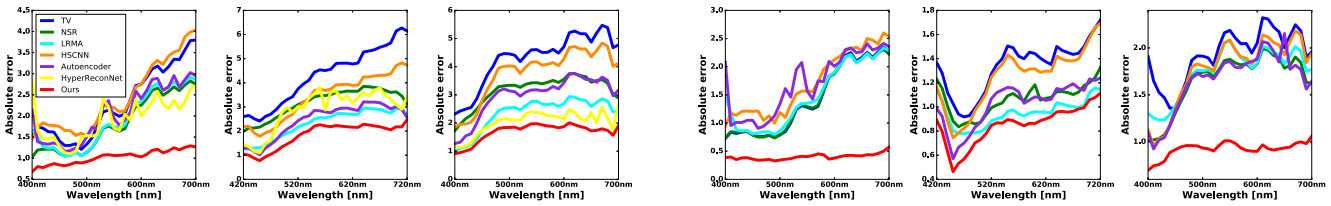


Figure 5. Visual quality comparison on three typical scenes in HSI datasets for DCD. The error maps for TV/NSR/LRMA/HSCNN/Autoencoder/our DCD recovered results and the scenes are shown from left to right.



(a) CASSI (b) DCD

Figure 6. The absolute error between ground truth and recovered results of scenes in Figure 4 and Figure 5 along spectra for all methods.

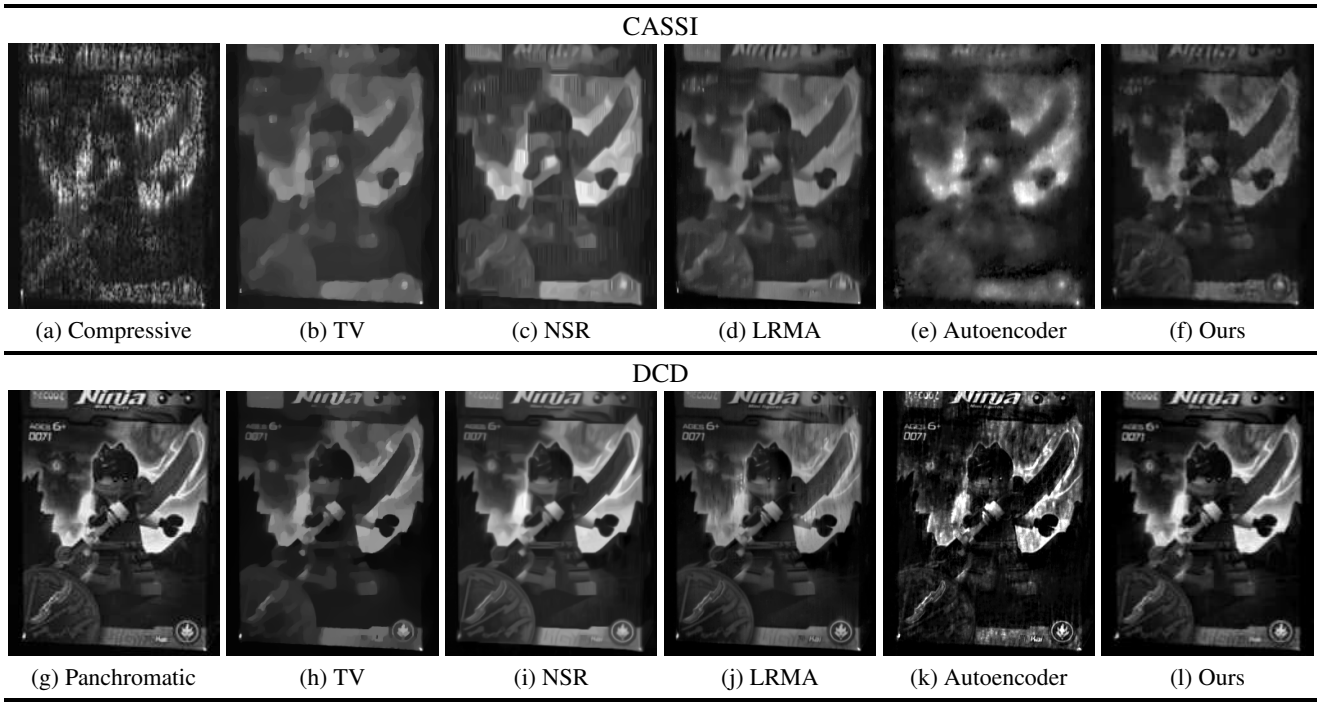


Figure 7. Reconstruction results of the spectral band in 607 nm. (a) Compressive image. (g) Panchromatic image. (b)-(f) TV/NSR/LRMA/Autoencoder/our CASSI reconstruction results. (h)-(l) TV/NSR/LRMA/Autoencoder/our DCD reconstruction results.

matic image of DCD is shown in Figure 7(g), respectively. We show the reconstruction results at 607 nm for all methods. By comparing the CASSI reconstruction results in Figure 7(b,c,d,e,f) and DCD reconstruction results in Figure 7(h,i,j,k,l), we can see that the DCD outperforms CASSI with better vision quality. It suggests that the panchromatic image can obviously assist the HSI reconstruction. According to the reconstruction results of all methods, our method outperforms other methods in both systems. We can see that TV, NSR, LRMA and Autoencoder suffer from oversmoothing or lack structure information, and our method produces better results with more details and less artifacts compared with other methods.

5. Conclusion

In this paper, we propose a novel CNN-based coded HSI reconstruction method via combining deep external and internal learning, which learns the deep prior from both the external dataset and internal information of input coded image. The proposed method can effectively exploit the spectral-spatial correlation and adapt itself to variant scenes. Experimental results show that the proposed method outperforms current state-of-the-art methods both on synthetic data and real data and has better generalization ability compared with the learning-based methods.

Our method is mainly evaluated on CASSI and DCD, which could be directly implemented on other coded HSI reconstruction, e.g., spatial-spectral encoded imaging system [25], multiple snapshot imaging system [22, 42] and so on. Besides, it is worth investigating the effect from the combination of deep external and internal learning for nature image/video compressive sensing reconstruction [24], image restoration [32], and so on.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants No. 61425013 and No. 61672096 and the Beijing Municipal Science and Technology Commission under Grant No. Z181100003018003.

References

[1] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *Proc. of European Conference on Computer Vision (ECCV)*, 2016.

[2] Robert W. Basedow, Dwayne C. Carmer, and Mark E. Anderson. Hydice system: Implementation and performance. In *SPIE's Symposium on OE/Aerospace Sensing and Dual Use Photonics*, pages 258–267, 1995.

[3] Jose M. Bioucas-Dias and Mario A. Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Processing*, 16(12):2992–3004, Dec. 2007.

[4] Asgeir Bjorgan and Lise Lyngsnes Randeberg. Towards real-time medical diagnostics using hyperspectral imaging tech-

nology. In *Clinical and Biomedical Spectroscopy and Imaging IV*, page 953712, 2015.

[5] Marcus Borengasser, William S. Hungate, and Russell Watkins. *Hyperspectral Remote Sensing: Principles and Applications*. Remote Sensing Applications Series. CRC Press, Dec. 2007.

[6] Xun Cao, Hao Du, Xin Tong, Qionghai Dai, and Stephen Lin. A prismmask system for multispectral video acquisition. *IEEE Trans. Pattern Analysis and Machine Intelligence* , 33(12):2423–2435, Dec. 2011.

[7] Xiangyong Cao, Feng Zhou, Lin Xu, Deyu Meng, Zongben Xu, and John Paisley. Hyperspectral image classification with markov random fields and a convolutional neural network. *IEEE Trans. Image Processing*, 27(5):2354–2367, 2018.

[8] Ayan Chakrabarti and Todd E. Zickler. Statistics of real-world hyperspectral images. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 193–200, 2011.

[9] Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Trans. on Graphics (Proc. of SIGGRAPH Asia)*, 36(6):218:1–218:13, Nov. 2017.

[10] Michael Descour and Eustace Dereniak. Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. *Applied Optics*, 34(22):4817–26, 1995.

[11] Bridget Ford, Michael Descour, and Ronald Lynch. Large-image-format computed tomography imaging spectrometer for fluorescence microscopy. *Optics Express*, 9(9):444–53, 2001.

[12] Ying Fu, Tao Zhang, Yinqiang Zheng, Debing Zhang, and Hua Huang. Joint camera spectral sensitivity selection and hyperspectral image recovery. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018.

[13] Ying Fu, Yinqiang Zheng, Imari Sato, and Yoichi Sato. Exploiting spectral-spatial correlation for coded hyperspectral image restoration. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3727–3736, 2016.

[14] Liang Gao, Robert T. Kester, Nathan Hagen, and Tomasz S. Tkaczyk. Snapshot Image Mapping Spectrometer (IMS) with high sampling density for hyperspectral microscopy. *Optics Express*, 18(14):14330–14344, 2010.

[15] Nahum Gat, Gordon Scriven, John Garman, Ming De Li, and Jingyi Zhang. Development of four-dimensional imaging spectrometers (4d-IS). In *Proc. of SPIE Optics + Photonics*, volume 6302, pages 63020M–63020M–11, 2006.

[16] Michael E. Gehm, Renu John, David J. Brady, Rebecca Willett, and Timothy J. Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics Express*, 15(21):14013–27, 2007.

[17] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proc. of International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[19] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[20] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proc. of international conference on Multimedia(MM)*, pages 675–678, Nov. 2014.

[21] Diederik P. Kingma and Jimmy Lei Ba. Adam: a method for stochastic optimization. In *Proc. of International Conference on Learning Representations(ICLR)*, 2015.

[22] David Kittle, Kerjil Choi, Ashwin Wagadarikar, and David J. Brady. Multiframe image estimation for coded aperture snapshot spectral imagers. *Applied Optics*, 49(36):6824–6833, 2010.

[23] 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, May 1993.

[24] Kuldeep Kulkarni, Suhas Lohit, Pavan Turaga, Ronan Kerviche, and Amit Ashok. Reconnet: Non-iterative reconstruction of images from compressively sensed random measurements. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 449–458, 2016.

[25] Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. Spatial-spectral Encoded Compressive Hyperspectral Imaging. *ACM Trans. on Graphics (Proc. of SIGGRAPH Asia)*, 33(6):233:1–233:11, Nov. 2014.

[26] Xing Lin, Gordon Wetzstein, Yebin Liu, and Qionghai Dai. Dual-coded compressive hyperspectral imaging. *Optics Letters*, 39(7):2044–2047, 2014.

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

[28] Lujendra Ojha, Mary Beth Wilhelm, Scott L. Murchie, Alfred S. McEwen, James J. Wray, Jennifer Hanley, Marion Mass, and Matt Chojnacki. Spectral evidence for hydrated salts in recurring slope lineae on Mars. *Nature Geoscience*, 8(11):829–832, Nov. 2015.

[29] Takayuki Okamoto and Ichirou Yamaguchi. Simultaneous acquisition of spectral image information. *Optics Letters*, 16(16):1277–1279, 1991.

[30] Wallace M. Porter and Harry T. Enmark. A system overview of the airborne visible/infrared imaging spectrometer (aviris). In *Annual Technical Symposium*, pages 22–31, 1987.

[31] Yoav Y. Schechner and Shree K. Nayar. Generalized mosaicing: wide field of view multispectral imaging. *IEEE Trans. Pattern Analysis and Machine Intelligence* , 24(10):1334–1348, Oct. 2002.

[32] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[33] Jin Tan, Yanting Ma, Hoover Rueda, Dror Baron, and Gonzalo R. Arce. Compressive hyperspectral imaging via approximate message passing. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):389–401, 2016.

[34] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[35] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, 47(10):44–51, Apr. 2008.

[36] Lucien Wald. Quality of high resolution synthesised images: Is there a simple criterion ? In *Proc. of Conference on Fusion Earth Data*, pages 99–103, 2000.

[37] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, and Feng Wu. Dual-camera design for coded aperture snapshot spectral imaging. *Applied Optics*, 54(4):848–58, Feb. 2015.

[38] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, Wenjun Zeng, and Feng Wu. High-speed hyperspectral video acquisition with a dual-camera architecture. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4942–4950, 2015.

[39] Lizhi Wang, Zhiwei Xiong, Guangming Shi, Feng Wu, and Wenjun Zeng. Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. *IEEE Trans. Pattern Analysis and Machine Intelligence* , 39(10):2104–2111, Oct. 2017.

[40] Lizhi Wang, Tao Zhang, Ying Fu, and Hua Huang. Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. *IEEE Trans. Image Processing*, 28(5):2257–2270, Nov. 2018.

[41] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, Apr. 2004.

[42] Yuehao Wu, Iftekhar O. Mirza, Gonzalo R. Arce, and Dennis W. Prather. Development of a digital-micromirror-devicebased multishot snapshot spectral imaging system. *Applied Optics*, 36(14):2692–2694, 2011.

[43] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proc. of International Conference on Computer Vision - Workshops*, 2017.

[44] Masahiro Yamaguchi, Hideaki Haneishi, Hiroyuki Fukuda, Junko Kishimoto, Hiroshi Kanazawa, Masaru Tsuchida, Ryo Iwama, and Nagaaki Ohyama. High-fidelity video and still-image communication based on spectral information: natural vision system and its applications. In *Proc. of SPIE*, 2006.

[45] Fumihito Yasuma, Tommo Mitsunaga, Daisuke Iso, and Shree K. Nayar. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range and spectrum. *IEEE Trans. Image Processing*, 19(9):2241–2253, 2010.

[46] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018.