Hyperspectral Image Super-Resolution with Optimized RGB Guidance

Ying Fu[1] Tao Zhang[1] Yinqiang Zheng[2] Debing Zhang[3] Hua Huang[1]
[1]Beijing Institute of Technology [2]National Institute of Informatics [3]DeepGlint
{fuying,tzhang,huahuang}@bit.edu.cn yqzheng@nii.ac.jp debingzhang@deepglint.com

Abstract

To overcome the limitations of existing hyperspectral cameras on spatial/temporal resolution, fusing a low resolution hyperspectral image (HSI) with a high resolution RGB (or multispectral) image into a high resolution HSI has been prevalent. Previous methods for this fusion task usually employ hand-crafted priors to model the underlying structure of the latent high resolution HSI, and the effect of the camera spectral response (CSR) of the RGB camera on super-resolution accuracy has rarely been investigated. In this paper, we first present a simple and efficient convolutional neural network (CNN) based method for HSI super-resolution in an unsupervised way, without any prior training. Later, we append a CSR optimization layer onto the HSI super-resolution network, either to automatically select the best CSR in a given CSR dataset, or to design the optimal CSR under some physical restrictions. Experimental results show our method outperforms the state-of-the-arts, and the CSR optimization can further boost the accuracy of HSI super-resolution.

1. Introduction

Hyperspectral imaging systems capture detailed spectral distribution of a scene, and have found numerous applications in remote sensing [7, 34], object classification [8], anomaly detection [41], fluorescent analysis [15, 16], and so on.

For these applications, images with sufficient spectral and spatial resolution are usually desired [20, 37]. However, to achieve high spectral resolution, traditional hyperspectral imaging systems [6, 18, 36] often sacrifice the temporal and spatial resolution. In contrast, conventional RGB (or multispectral) cameras integrate the radiance across a wide wavelength range, and can easily capture high spatial resolution images in real time. Therefore, a natural way to obtain high resolution HSI is to fuse the low resolution HSI with its corresponding high resolution RGB image, which acts as spatial guidance.

Most existing approaches for HSI super-resolution of a hybrid camera system [1, 2, 3, 12, 14, 24, 28, 30, 31, 45] employ various hand-crafted priors to model the underlying structure of the latent high resolution HSI. Nevertheless, to hammer out proper priors for a specific scene remains to be an art.

Recent alternative approaches [13, 35] leverage on deep learning to alleviate the dependence on hand-crafted priors, and show that the CNN scheme can effectively exploit the intrinsic characteristics of HSIs. Nevertheless, these methods either use the CNN scheme to refine the initialized results in a supervised way [13], or resort to step-by-step alternating optimization [35]. In this work, we present a simple and efficient CNN-based end-to-end method for HSI super-resolution with RGB guidance, which can effectively approximate the spectral nonlinear mapping between the RGB and the spectral space, and utilize the spatial consistency. Neither delicate hand-crafted priors nor training data are needed in our method. This allows our method to handle various scenes more easily.

In addition, all these methods mainly focus on RGB-guided HSI super-resolution under a given CSR function of the RGB camera. Recent researches on HSI super-resolution from a single RGB image [5, 17, 33] have shown that the CSR is critical in improving super-resolution accuracy. This motivates to explore the effect of CSR in our hybrid super-resolution task. Through experiments, we have found that the performance of HSI super-resolution methods for the hybrid camera system is obviously dependent on the CSR. To optimize the RGB guidance, in this paper, we present a CSR optimization layer to automatically determine the best CSR in a given CSR dataset, or even to design an unprecedented CSR function that is optimal for the task of RGB-guided HSI super-resolution. To the best of our knowledge, this work is the first to evaluate the effect of CSR in a hybrid hyperspectral imaging system, and optimize the RGB guidance for hybrid HSI super-resolution. In summary, our main contributions are that

1. we present an unsupervised CNN-based method for HSI super-resolution, which can effectively exploit the underlying characteristics of the HSI and adapt itself to variant scenes more easily;

2. we evaluate the effect of RGB CSR functions for HSI super-resolution, and develop a CSR selection layer to retrieve the best CSR of a given CSR dataset;
3. Beyond CSR selection, we simulate the CSR as a convolution layer to learn the optimal CSR for RGB-guided HSI super-resolution.

2. Related Work

HSI super-resolution is closely related to multispectral image (MSI) pan-sharping, in which a low resolution MSI is usually fused with a high resolution panchromatic image [29]. The upscaling ratio of existing methods [10, 42] is usually no more than four times.

To lift the HSI super-resolution ratio, the recent trend is to use a RGB (or MSI) image to guide the HSI super-resolution. Most existing approaches for RGB-guided HSI super-resolution are based on matrix factorization [1, 3, 14, 24, 28, 30]. Kawakami *et al.* [24] used matrix factorization method to learn a spectral dictionary representation in the low resolution HSI under the sparsity assumption, and restored the high resolution HSI using shared sparse coefficients estimated from the RGB observation. Wycoff *et al.* [40] took into account the non-negative physical property of the materials in the scene to improve performance. Yokoya *et al.* [44] used a coupled non-negative matrix factorization approach without any sparse constraints, while Lanaras *et al.* [30] employed similar approach with sparse constraints. Later, Dong *et al.* [14] further employed non-negative structured sparse coding model for HSI super-resolution after the matrix factorization.

Besides, Bayesian representation [2, 3] is also used for HSI super-resolution. Akhtar *et al.* [2] learned the spectral dictionary by using non-parametric Bayesian model and constructed the high resolution HSI with the learned dictionary under sparse constraints. Later, Akhtar *et al.* [3] employed a Bayesian representation model and used the hierarchical Beta process with Gaussian process prior for HSI super-resolution.

Another class of HSI super-resolution methods are based on tensor factorization [12, 31, 45]. Dian *et al.* [12] proposed to use non-local sparse tensor factorization for this super-resolution task. Li *et al.* [31] formulated the estimation of the dictionaries and the core tensor as a coupled tensor factorization of the low resolution HSI and the high resolution multispectral image. Zhang *et al.* [45] presented a clustering manifold structure based HSI super-resolution method under tensor representation.

The aforementioned methods from these three categories all formulate the fusion problem as the optimization problem constrained by various hand-crafted priors, like low-rankness and sparsity. More recently, CNN-based approaches have been presented for HSI super-resolution [13, 35], in which hand-crafted prior modeling is no longer necessary. Dian *et al.* [13] initially restored the high resolution HSI from the fusion framework via solving a Sylvester equation, and then employed CNN-based method to enhance the initialized results with prior training on a HSI dataset in a supervised way. Qu *et al.* [35] attempted to solve the HSI super-resolution problem using an unsupervised encoder-decoder architecture without training in a HSI dataset. Although it restored the high resolution HSI by using a CNN-based end-to-end network in an unsupervised way, this method needs to be carefully optimized step-by-step in an alternating way. In this work, we present a CNN-based end-to-end method for HSI super-resolution, which is easy to optimize and independent of prior training on a HSI dataset.

In parallel to hybrid fusion, HSI super-resolution from a single RGB image has attracted attention, and the very recent trend is to optimize the CSR of the RGB image so as to maximize the reconstruction accuracy. Arad and Ben-Shahar [5] first recognized that the quality of HSI recovery from a single RGB image is sensitive to the CSR selection. Fu *et al.* [17] presented a CNN-based method to select the optimal CSR in a CSR dataset, which has much lower time complexity. Nie *et al.* [33] went beyond selection and automatically designed optimal CSR using CNN.

Inspired by these methods, we investigate the effect of CSR functions for HSI super-resolution with RGB guidance, and develop a CSR optimization layer to select or design the CSR to boost the accuracy of hybrid HSI super-resolution.

3. RGB-Guided HSI Super-resolution

In this section, we first formulate the problem for HSI super-resolution with RGB guidance and describe the motivation for our method. Then, we introduce our CNN-based method for the HSI super-resolution, which can effectively learn the internal recurrence of spectral information and guarantee spatial consistency.

3.1. Formulation and Motivation

The aim is to restore a high resolution HSI $\mathbf{X} \in \mathbb{R}^{B\times MN}$ by fusing a low resolution HSI $\mathbf{X}_l \in \mathbb{R}^{B\times mn}$ and a high resolution RGB image $\mathbf{Y} \in \mathbb{R}^{b\times MN}$. M, N, and B are the number of rows, columns, and bands for the restored high resolution HSI $\mathbf{X}$. Correspondingly, m and n denote the number of rows and columns for the low resolution image $\mathbf{X}_l$. The input low resolution HSI $\mathbf{X}_l$ is the downsampled version of $\mathbf{X}$ in the spatial dimension and the high resolution RGB $\mathbf{Y}$ can be obtained by downsampling $\mathbf{X}$ across spectra. The relationship among these three images is generally linear and can be described as

$$\mathbf{X}_l = \mathbf{X}\mathbf{H}, \quad \text{and} \quad \mathbf{Y} = \mathbf{C}\mathbf{X}, \tag{1}$$

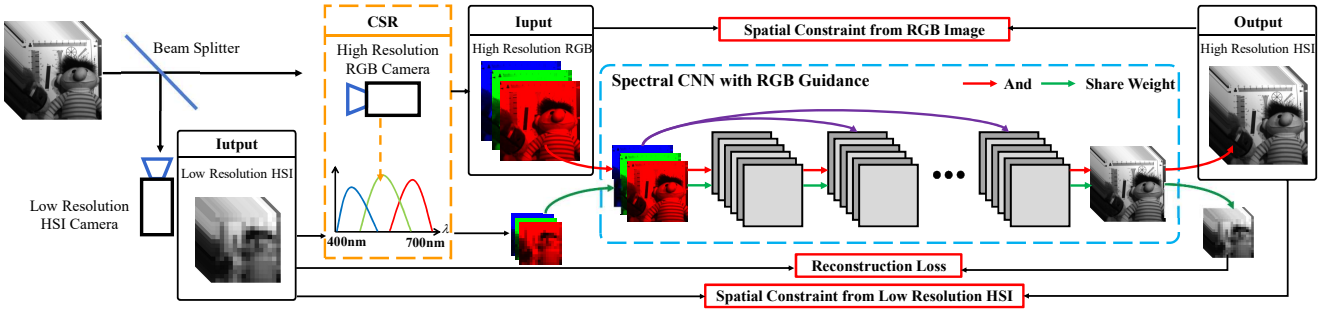


Figure 1. Overview of our CNN-based HSI super-resolution with RGB guidance.

where $\mathbf{H} \in \mathbb{R}^{MN \times mn}$ denotes the downsampling along spatial dimension, and $\mathbf{C} \in \mathbb{R}^{b \times B}$ is the CSR function to integrate the spectra into R, G, and B channels.

Most state-of-the-art methods model HSI super-resolution from a low resolution HSI and a high resolution RGB image as

$$E(\mathbf{X}) = E_d(\mathbf{X}, \mathbf{X}_l, \mathbf{Y}) + \lambda E_s(\mathbf{X}), \tag{2}$$

where λ is a predefined parameter. The first term $E_d(\mathbf{X}, \mathbf{X}_l, \mathbf{Y})$ is the data term such that the recovered $\mathbf{X}$ should be projected to $\mathbf{X}_l$ and $\mathbf{Y}$ under constraints in Equation (1). The second term $E_s(\mathbf{X})$ is the prior regularization for $\mathbf{X}$. To model this term, previous works [1, 3, 14, 24, 30] often assume that the scene contains a small number of distinct materials and can be described in a linear way under sparsity assumption.

Recently, [22] shows that the nonlinear mapping between RGB values and spectra for each spatial point can effectively assist HSI recovery from a single RGB image. [11] reveals that the nonlinear spectral representation of the HSI can significantly improve the accuracy of restored HSI instead of the linear representation.

Accordingly, we present a CNN-based method for HSI super-resolution with RGB guidance, which can effectively learn the nonlinear spectral representation and add the spatial constraints in an unsupervised way. Concretely, we utilize multiple CNN layers in spectral CNN to deeply learn the nonlinear mapping between spectra and RGB space, and employ spatial constraint for the high resolution HSI to guarantee the spatial consistency. Besides, our method uses the input RGB image to further guide the HSI super-resolution with the spectral CNN. Figure 1 shows the HSI super-resolution network for a hybrid camera system.

3.2. Spectral CNN with RGB Guidance

To better model the spectral relationship between the HSI and RGB image, the low resolution HSI is first downsampled across spectra to a RGB image by a pre-determined CSR, which is denoted as $\mathbf{Y}_l = \mathbf{C}\mathbf{X}_l$. A spectral CNN is designed to learn the spectral nonlinear mapping between

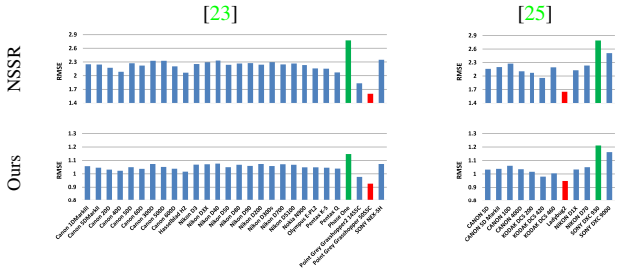


Figure 2. The RMSE results of NSSR and our HSI recovery network on ICVL and two CSR datasets. The red and green bars indicate the best and worst CSR functions for the HSI recovery in a brute force way, respectively.

the RGB values and the corresponding spectra from the downsampled RGB image and the low resolution HSI. Besides, the input RGB image is also used to guide the spatial information reconstruction, which is modeled by stacking the input RGB image and the output for each layer. Thus, the spectral CNN consists of L layers and the output of the l-th layer can be expressed as

$$\mathbf{F}_l = \text{LeakyReLU}(\mathbf{W}_l * \text{stack}(\mathbf{F}_{l-1}, \mathbf{Y}) + \boldsymbol{b}_l), \tag{3}$$

where $\mathbf{F}_0 = 0$ and $\text{LeakyReLU}(x) = \max\{x, \alpha x\}$, denoting a leaky rectified linear unit [32], and we empirically set $\alpha = 0.05$. $\mathbf{W}_l$ and $\boldsymbol{b}_l$ represent the convolutional kernels and biases for the l-th layer, respectively. In the experiments, we empirically set $L = 5$. To learn the intrinsic recurrence of spectral information, the size of the convolutional kernels is set to be 1×1.

The parameters for the HSI super-resolution network are denoted as $\mathbf{\Theta} = \{\mathbf{W}, \mathbf{b}\}$, and can be achieved by minimizing the Mean Squared Error (MSE) between the reconstructed HSI $\hat{\mathbf{X}}_l$ and the corresponding ground truth image

$$\mathcal{L}_s(\mathbf{\Theta}) = \|f(\mathbf{Y}_l, \mathbf{\Theta}) - \mathbf{X}_l\|^2. \tag{4}$$

3.3. Spatial Constraints

The learned parameters in Equation (4) are also shared for the high resolution RGB image and the latent high resolution HSI in the learning process. Thus, the relationship

between high resolution RGB and latent high resolution HSI can be expressed as

$$\mathbf{X} = f(\mathbf{Y}, \mathbf{\Theta}). \tag{5}$$

According to Equation (1), the latent high resolution HSI should be consistent with low resolution HSI and high resolution RGB image after linear mappings. Given the relationship in Equation (5), the spatial constraints from inputs, which are also used to learn the model parameters, can be described as

$$\mathcal{L}_d(\mathbf{\Theta}) = \|\mathbf{Y} - \mathbf{C}f(\mathbf{Y}, \mathbf{\Theta})\|^2 + \tau_1 \|\mathbf{X}_l - f(\mathbf{Y}, \mathbf{\Theta})\mathbf{H}\|^2. \tag{6}$$

3.4. Learning Details

In our method, the spectral nonlinear mapping and spatial consistency are involved into a unified framework to learn the model, and high resolution HSI is restored by minimizing the loss

$$\mathcal{L}_{sd} = \mathcal{L}_d(\mathbf{\Theta}) + \tau_2 \mathcal{L}_s(\mathbf{\Theta}) + \eta_1 \|\mathbf{\Theta}\|_2^2. \tag{7}$$

The underlying characteristics of the latent high resolution HSI in Equation (7) is modeled in deep prior instead of hand-crafted priors and learned only on the input images with no need for the training set in an unsupervised way.

The loss is minimized with the adaptive moment estimation method [26] and τ_1, τ_2, η_1 are set to 50, 50, 10^{-4}, respectively. The learning rate is initially set to be 0.01, which will be divided by 10 every 2000 iterations. All convolution layer's weights are initialized by the method in [19].

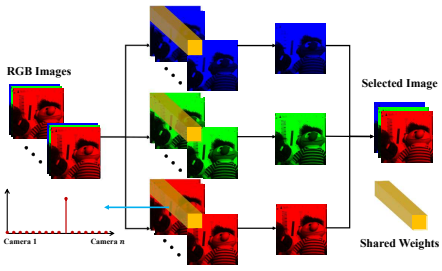
4. CSR Optimization

In this section, we first describe the motivation for the CSR optimization, then introduce two approaches for CSR optimization layer, which is appended in front of the spectra reconstruction network as shown in the orange box in Figure 1, and they are jointly learned when optimizing the CSR layer.

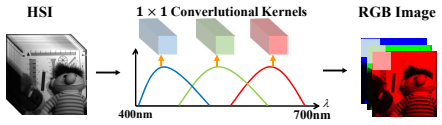
4.1. Motivation

Previous researches on HSI recovery from a single RGB image [5, 17, 33] have shown that the CSR significantly affects the quality of HSI recovery. However, this effect has never been investigated in existing literature on RGB-guided HSI super-resolution. To investigate the effect from the CSR, we perform HSI super-resolution methods by using the synthetic RGB images from a HSI dataset under different CSR functions in a brute force way.

Figure 2 shows the results from NSSR [14] and our method on ICVL dataset (more details will be provided in Section 5). It can be observed that HSI super-resolution with RGB guidance is also obviously dependent on the used



(a) The optimal CSR selection



(b) The optimal CSR design

Figure 3. The illustration of the CSR optimization.

CSR. Therefore, to boost the accuracy of hybrid HSI super-resolution, it is essential to optimize the CSR.

Here, we present two approaches to obtain the optimal CSR. For existing RGB cameras, we design a selection convolution layer to retrieve the optimal CSR. Beyond the CSR selection, the optimal CSR is further learned under some physical restrictions by simulating the CSR as a convolution layer. The optimized CSR is selected or designed to boost HSI super-resolution, so the CSR optimization layer should be appended onto the HSI super-resolution network and learned together with high resolution HSI restoration.

Note that, our CNN-based HSI super-resolution method itself works on the input image pair only, without requiring any training data. Therefore, it is possible in principle to optimize the CSR for the given input image pair. However, we have found that the optimal CSR varies a lot according to the input scene, and it is not meaningful to customize a RGB camera for every scene. Therefore, we optimize the CSR by using a HSI dataset, such that the selected/designed CSR is generally applicable.

4.2. Optimal CSR Selection

To select the optimal CSR function from existing cameras, RGB images for each HSI are first synthesized with all CSR functions in a candidate dataset, described as $\mathbf{Y}_{j,t} = \mathbf{C}_j \mathbf{X}_t$ for the j-th CSR function and the t-th HSI in the training dataset. For each scene, the input RGB images are obtained by stacking all RGB images under different CSR functions, and denoted as $\mathcal{Y}_t = \text{stack}(\mathbf{Y}_{1,t}, \cdots, \mathbf{Y}_{j,t}, \cdots, \mathbf{Y}_{J,t})$.

The optimal CSR selection is equivalent to the RGB image selection in $\mathcal{Y}_t$, which is synthesized with the selected optimal CSR. As shown in Figure 3a, we first separate RGB bands into three branches, which share the same 1×1 convolution kernel $\boldsymbol{V}$. Thus, the output of this optimal CSR

selection network can be expressed as

$$\hat{\mathbf{Y}}_t = \text{stack}(\boldsymbol{V} * \mathcal{Y}_t(R), \boldsymbol{V} * \mathcal{Y}_t(G), \boldsymbol{V} * \mathcal{Y}_t(B)), \quad (8)$$

where $\mathcal{Y}_t(R)$, $\mathcal{Y}_t(G)$, and $\mathcal{Y}_t(B)$ denote all the red, green, and blue channels in $\mathcal{Y}_t$, respectively.

As the simulated RGB image should be positive, the weight for CSR selection should be nonnegative. To correctly identify the best CSR, a sparsity constraint is introduced. The $\boldsymbol{V}$ can be determined by minimizing the MSE under the nonnegative sparse constraint between the selected RGB image $\hat{\mathbf{Y}}_t$ and the corresponding ground truth image

$$\mathcal{L}_{cs}(\boldsymbol{V}) = \frac{1}{T}\sum_{t=1}^{T} \|\hat{\mathbf{Y}}_t(\boldsymbol{V}) - \mathbf{Y}_t\|^2 + \eta_2 \|\boldsymbol{V}\|_1, \quad s.t. \ \ \boldsymbol{V} \geq 0, \quad (9)$$

where $\hat{\mathbf{Y}}_t$ is the t-th output, $\mathbf{Y}_t$ is the t-th ground truth image, and T is the number of training samples. A larger value in $\boldsymbol{V}$ implies that its corresponding CSR is better for HSI recovery. Consequently, the largest value in $\boldsymbol{V}$ reveals the optimal CSR.

4.3. Optimal CSR Learning

The linear relationship between RGB and the latent high resolution HSI is $\mathbf{Y} = \mathbf{C}\mathbf{X}$. Each row of the CSR function $\mathbf{C}$ can be regarded as the weight in the 1×1 convolution layer with three kernels. As shown in Figure 3b, each spatial point could be interpreted as the output activation map of a convolution. Thus, the layer for optimal CSR learning can be expressed as

$$\hat{\mathbf{Y}} = \boldsymbol{U} * \mathbf{X}, \quad (10)$$

where $\boldsymbol{U}$ is the convolutional representation of $\mathbf{C}$.

As the simulated RGB image should be positive, all values in CSR are non-negative. Besides, according to [33], the CSR function should be smooth to facilitate filter realization. Thus, the values in $\boldsymbol{U}$ can be determined by minimizing the MSE under the nonnegative smooth constraint between the synthesized RGB image $\hat{\mathbf{Y}}$ and the corresponding ground truth image

$$\begin{aligned}\mathcal{L}_{co}(\boldsymbol{U}) = &\frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{U} * \mathbf{X}_t - \mathbf{Y}_t\|^2 + \eta_3 \|\boldsymbol{U}\|_2^2 \\ &+ \eta_4 \|G\boldsymbol{U}\|_2^2, \quad s.t. \ \ \boldsymbol{U} \geq 0,\end{aligned} \quad (11)$$

where G is the first derivative matrix to account for smoothness, and $\mathbf{Y}_t$ is the t-th input RGB image corresponding learned CSR. The optimal CSR can be obtained from the learned weight of three 1×1 convolution kernels $\boldsymbol{U}$.

4.4. Optimization Details

The CSR optimization layer is appended onto the HSI super-resolution network to optimize CSR and the parameters for HSI super-resolution are learned together. To select the best CSR in a candidate dataset, the entire network is trained by minimizing the loss

$$\mathcal{L} = \mathcal{L}_{cs}(\boldsymbol{V}) + \tau_3 \mathcal{L}_{sd}(\boldsymbol{\Theta}). \quad (12)$$

The CSR corresponding to the largest value in $\boldsymbol{V}$ is selected as the optimal CSR.

To optimize CSR via design, the entire network is trained by minimizing the loss

$$\mathcal{L} = \mathcal{L}_{co}(\boldsymbol{U}) + \tau_4 \mathcal{L}_{sd}(\boldsymbol{\Theta}). \quad (13)$$

The optimal CSR can be obtained from the learned weight of the 1×1 convolution kernels $\boldsymbol{U}$.

These losses are minimized with the adaptive moment estimation method [26], τ_3, τ_4, η_2, η_3 and η_4 are set to 1, 1, 0.8, 0.01 and 0.1, respectively. We set the learning rates for the selection layer, design layer and HSI super-resolution to be 0.01, 0.001 and 0.001, respectively. To fit the nonnegative constraint for the optimal CSR selection and optimal CSR design, its convolution layer's weights are initialized as $\frac{1}{J}$ and $\frac{1}{B}$, where J is the number of CSRs and B is the number of HSI bands, respectively. All negative weights for CSR optimization layer are set to zero during the forward and back propagation. The network has been trained with the deep learning tool Caffe [21] on a NVIDIA Titan X GPU.

5. Experimental Results

In this section, we first introduce the datasets and settings in our experiments. Then, we compare our method with several state-of-the-art methods under a typical CSR. In addition, the effectiveness of our CSR optimization method is evaluated. Finally, we implement our HSI super-resolution method on the real images.

5.1. Dataset and Setup

Our method is evaluated on three public hyperspectral datasets, including the ICVL dataset [4], the CAVE dataset [43], and the Harvard dataset [9]. The ICVL dataset consists of 201 images, which is by far the most comprehensive natural hyperspectral dataset. The spatial resolution of HSIs is 1300×1392.

The Harvard dataset consists of 50 outdoor images captured under daylight illumination, whose spatial resolution is 1024×1392. The CAVE dataset has 32 HSIs and the spatial resolution is 512×512. All HSIs in these datasets have 31 bands. To fairly compare with results provided in [35], we follow it to use the top left 1024×1024 image region in Harvard dataset and the whole image in CAVE dataset to perform the comparison on these datasets.

We randomly select 151 images in ICVL dataset to train the optimal CSR and use the rest for testing. To compare

Table 1. Evaluation results of different unsupervised HSI super-resolution methods on three HSI datasets.

Methods	Metrics	ICVL	Harvard	CAVE
NSSR	RMSE	1.6004	1.8524	2.5288
	SSIM	0.9901	**0.9814**	**0.9859**
	ERGAS	0.1168	0.3045	**0.3143**
	SAM	1.3766	3.2738	5.2171
BSR	RMSE	3.7453	2.5854	5.8107
	SSIM	0.9673	0.9775	0.9508
	ERGAS	0.3012	0.3390	0.9508
	SAM	2.8155	4.0001	12.4576
CSTF	RMSE	1.8455	2.5793	3.0002
	SSIM	0.9841	0.9598	0.9669
	ERGAS	0.1204	0.3210	0.3830
	SAM	1.5182	4.1837	7.5866
UDL	RMSE	/	**1.78**	4.09
	SAM	/	4.05	6.95
Ours	RMSE	**0.9673**	1.9347	**2.5055**
	SSIM	**0.9932**	0.9800	0.9846
	ERGAS	**0.0505**	**0.2554**	0.3175
	SAM	**0.8103**	**3.1582**	**4.5232**

with [13], we follow it to separate the training and testing sets for CAVE and Harvard datasets.

Two CSR datasets are used to evaluate the optimal CSR selection. The first dataset [23] contains 28 CSR curves and the second dataset [25] contains 12 CSR curves. Both datasets cover different camera types and brands.

The original HSIs in datasets serve as ground truth for HSI super-resolution. The low-resolution HSI is obtained by downsampling the original HSI with a scaling factor of 32, which means that we average over 32×32 spatial patch to produce a spatial pixel in low resolution HSI. This procedure has been widely used in existing works for HSI super-resolution with RGB guidance [1, 2, 3, 12, 14, 24, 28, 30, 31, 45]. The RGB image is simulated by integrating the original HSI along the spectral dimension by the CSR function.

Four image quality metrics are utilized to evaluate the performance of all methods, including root-mean-square error (RMSE), structural similarity (SSIM) [39], relative dimensionless global error in synthesis (ERGAS) [38], and spectral angle mapping (SAM) [27]. Smaller values of RMSE, ERGAS, and SAM suggest better performance, while a larger value of SSIM implies better fidelity.

5.2. Evaluation on HSI Super-resolution

Here, we first compare our method with state-of-the-arts from four categories, including matrix factorization, Bayesian representation, tensor factorization, and deep learning. Five typical methods for comparison are selected from these categories. Non-negative structured sparse representation based method (NSSR) [14], Bayesian sparse representation based method (BSR) [2], and coupled sparse tensor factorization based method (CSTF) [31] belong to the first three categories, respectively. We also evaluate two deep learning based methods, including supervised CNN-based method (SDL) [13] and unsupervised CNN-based method (UDL) [35]. We use the CSR function of Nikon D700[1], which has been used in [2, 13, 14, 31, 35], to synthesize RGB values.

Table 2. Comparison with supervised HSI super-resolution method SDL on three HSI datasets.

Methods	Metrics	ICVL	Harvard	CAVE
SDL	RMSE	1.2788	2.0426	**2.5092**
	SSIM	0.9929	**0.9829**	0.9812
	ERGAS	0.0647	0.3448	0.2652
	SAM	1.0218	4.1778	6.0237
Ours	RMSE	**0.9673**	**2.0058**	2.6612
	SSIM	**0.9932**	0.9815	**0.9817**
	ERGAS	**0.0505**	**0.3166**	**0.2649**
	SAM	**0.8103**	**3.6369**	**4.7417**

Table 1 provides the averaged results over all HSIs on three HSI datasets, to quantitatively compare our method with NSSR, BSR, CSTF, and UDL. The best results are highlighted in bold. UDL performs better on the Harvard dataset and NSSR shows relatively larger advantage over the other methods on CAVE. The NSSR shows that the hand-crafted prior in NSSR is effective for the CAVE dataset. Since SDL needs training, we individually compared with it on its testing set instead of the full dataset and provide the results in Table 2.

According to Tables 1 and 2, our method provides better results in most cases for all error metrics and the improvement on the ICVL dataset is in general more significant. This reveals the advantages of deeply exploiting the intrinsic properties of HSIs and verifies the effectiveness of our HSI super-resolution network.

To visualize the experimental results, several representative restored HSIs and the corresponding error images on three datasets are shown in Figure 4. The ground truth, error images for NSSR/BSR/CSTF/SDL/our methods, and RMSE results along spectra for all methods are shown from top to bottom. The 16-th band for all scenes is shown. The error images are the average absolution errors between the ground truth and the recovered results across spectra. The results of SDL is much close to our method in terms of Table 2, but the accuracy of SDL relies on the similarity between training and testing set. For example, the *book scene* in the Harvard dataset is restored much worse compared with all other methods, for there is no similar scene in the training set. The HSI super-resolution results from our method are consistently more accurate for all scenes, which verifies that our method can provide higher spatial accuracy. The RMSE results along spectra for all methods show that the results

[1] http://www.maxmax.com/spectral response.htm

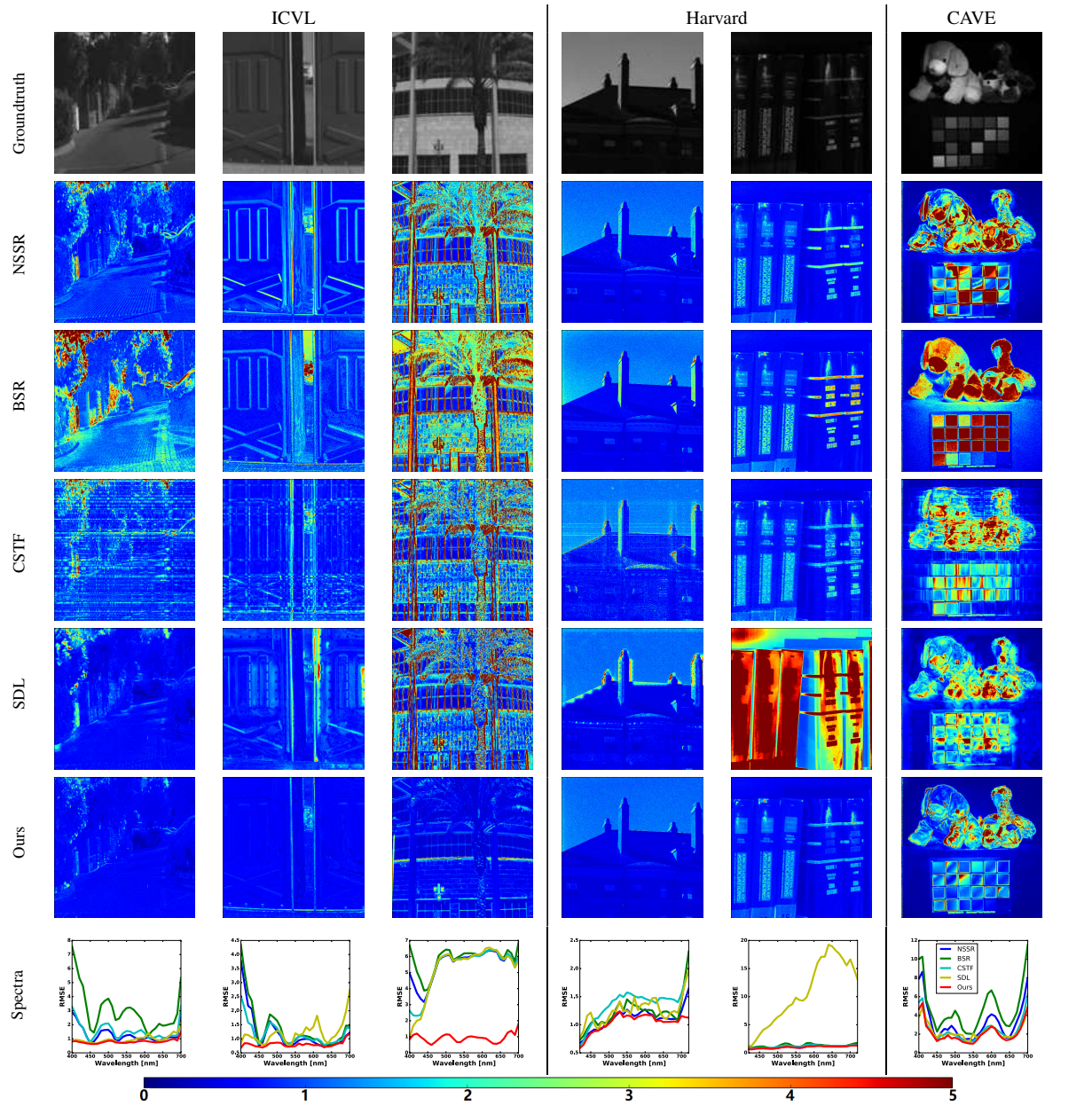
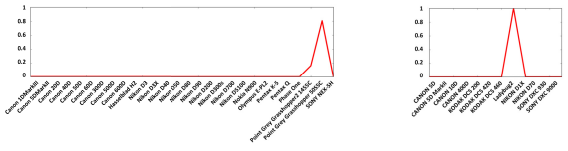


Figure 4. Visual quality comparison on six typical scenes in HSI datasets. The ground truth, the error map for NSSR/BSR/CSTF/SDL/our results, and RMSE results along spectra for all methods are shown from top to bottom.

of our method are much closer to the ground truth in most cases, which demonstrates that our approach obtains higher spectral fidelity.

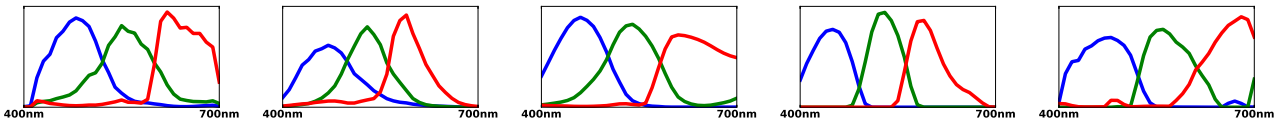
5.3. Evaluation on CSR Optimization

Due to the space limitation, we only show the CSR optimization results on the ICVL dataset. To obtain an optimal CSR for improved HSI super-resolution, we first use a convolutional layer under nonnegative sparsity constraint to select the optimal CSR in the training set of ICVL. As



(a) [23] (b) [25]

Figure 5. The selected optimal CSR by our methods on three HSI datasets. The largest value means the optimal CSR in the CSR dataset which is consistent with the best one determined by exhaustive search in Figure 2.



(a) Best in [23] (b) Worst in [23] (c) Best in [25] (d) Worst in [25] (e) Optimal

Figure 6. The best and worst CSRs from both CSR datasets and the designed CSR under ICVL training set.

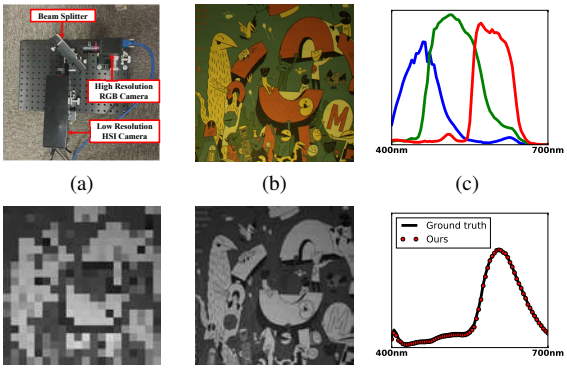
Table 3. RMSE, SSIM, and SAM results for different CSR on ICVL dataset.

CSRs	RMSE	SSIM	ERGAS	SAM
Best in [23]	0.9267	0.9933	0.0511	0.7801
Worst in [23]	1.1472	0.9918	0.0565	0.8808
Best in [25]	0.9481	0.9933	0.0495	0.7954
Worst in [25]	1.2117	0.9912	0.0578	0.9218
Optimal	0.9212	0.9934	0.0500	0.7802

shown in Figure 5, the largest value means the optimal CSR in the CSR dataset and our method can effectively select the optimal CSR, which is consistent with the best one determined by exhaustive search in Figure 2. The first to fourth rows of Table 3 are the best and worst results under different CSRs in [23, 25], respectively. Besides, we also design the CSR as three convolutional kernels to optimize the CSR in the CNN-based architecture. The corresponding results by training and testing in the ICVL dataset are shown in the fifth row of Table 3. The corresponding curves of these CSRs are provided in Figure 6. It can be seen that HSI super-resolution results with selected best and designed optimal CSRs significantly better than that with worst CSRs. The performance under the designed CSR is better than the selected best one. Besides, the selected best and designed optimal CSRs have the similar appearance (e.g. the higher sensitivity on the longer wavelength compared with worst one) and are much different from worst ones. It implies the effectiveness of the CSR optimization and provides a guidance for the CSR design of RGB cameras used for the HSI super-resolution.

5.4. Real Images

To further evaluate the effectiveness of our method, we set up a real hybrid camera system to capture the real images. It has a hyperspectral camera (EBA Japan NH-7) and a high resolution RGB camera (Nikon D5), as shown in Figure 7a. A cartoon scene is used for test. The captured RGB image is shown in Figure 7b, and its CSR is shown in Figure 7c. Figure 7d shows the low resolution HSI at 600 nm, and its corresponding restored high resolution result by our method is shown in Figure 7e. Figure 7f provides the recovered spectra for a randomly selected red area in the real scene. It can approximate the ground truth well. These convincingly demonstrate the effectiveness of the proposed method, especially in the real capture system.



(a) (b) (c) (d) (e) (f)

Figure 7. Results on real data. (a) The coaxial hybrid camera system. (b) The captured RGB image. (c) The CSR of the RGB camera. (d) The captured low resolution HSI at 600 nm. (e) The corresponding restored high resolution HSI by our method. (f) The recovered spectra for a randomly selected red area.

6. Conclusion

In this paper, we have proposed an unsupervised CNN-based end-to-end method for HSI super-resolution with optimal RGB guidance. Our network is easier to train, and does not rely on hand-crafted priors. We have also recognized that the CSR of the RGB camera is critical in maximizing the restoration accuracy, and proposed to optimize the RGB response via automatic optimal selection or design in an unified CNN framework. Experimental results showed that our HSI super-resolution method can provide substantial improvements over the current state-of-the-arts, and the CSR optimization can further boost the HSI restoration fidelity.

Our HSI super-resolution method employed a full three-channel RGB image and did not take into account the mosaic effect of a CCD/CMOS sensor. In addition, the designed optimal CSR has not been realized by optical filters. We will leave them as our future work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants No. 61425013 and No. 61672096, the Beijing Municipal Science and Technology Project under Grant No. Z181100003018003, and the JSPS KAKENHI under Grant No. 19K20307.

References

[1] N. Akhtar, F. Shafait, and A. Mian. Sparse Spatio-spectral Representation for Hyperspectral Image Super-resolution. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 63–78, Sept. 2014. 1, 2, 3, 6

[2] N. Akhtar, F. Shafait, and A. Mian. Bayesian sparse representation for hyperspectral image super resolution. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 1, 2, 6

[3] N. Akhtar, F. Shafait, and A. Mian. Hierarchical Beta process with Gaussian process prior for hyperspectral image super resolution. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 103–120, Oct. 2016. 1, 2, 3, 6

[4] B. Arad and O. Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *Proc. of European Conference on Computer Vision (ECCV)*, pages 19–34, Oct. 2016. 5

[5] B. Arad and O. Ben-Shahar. Filter selection for hyperspectral estimation. In *Proc. of International Conference on Computer Vision (ICCV)*, Oct. 2017. 1, 2, 4

[6] R. W. Basedow, D. C. Carmer, and M. E. Anderson. Hydice system: Implementation and performance. In *SPIE's Symposium on OE/Aerospace Sensing and Dual Use Photonics*, pages 258–267, 1995. 1

[7] M. Borengasser, W. S. Hungate, and R. Watkins. *Hyperspectral Remote Sensing: Principles and Applications*. Remote Sensing Applications Series. CRC Press, Dec. 2007. 1

[8] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley. Hyperspectral image classification with markov random fields and a convolutional neural network. *IEEE Trans. Image Processing*, 27(5):2354–2367, 2018. 1

[9] A. Chakrabarti and T. Zickler. Statistics of real-world hyperspectral images. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 193–200, June 2011. 5

[10] C. Chen, Y. Li, W. Liu, and J. Huang. Image fusion with local spectral consistency and dynamic gradient sparsity. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2760–2765, June 2014. 2

[11] I. Choi, D. S. Jeon, G. Nam, D. Gutierrez, and M. 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. 3

[12] R. Dian, L. Fang, and S. Li. Hyperspectral image super-resolution via non-local sparse tensor factorization. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5344–5353, June 2017. 1, 2, 6

[13] R. Dian, S. Li, A. Guo, and L. Fang. Deep hyperspectral image sharpening. *IEEE Trans. Neural Networks and Learning Systems*, 29(11):5345–5355, Nov. 2018. 1, 2, 6

[14] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Trans. Image Processing*, 25(5):2337–2352, May 2016. 1, 2, 3, 4, 6

[15] Y. Fu, A. Lam, Y. Kobashi, I. Sato, T. Okabe, and Y. Sato. Reflectance and fluorescent spectra recovery based on fluorescent chromaticity invariance under varying illumination. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2171–2178, June 2014. 1

[16] Y. Fu, A. Lam, I. Sato, T. Okabe, and Y. Sato. Separating reflective and fluorescent components using high frequency illumination in the spectral domain. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 457–464, 2013. 1

[17] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang. Joint camera spectral sensitivity selection and hyperspectral image recovery. In *Proc. of European Conference on Computer Vision (ECCV)*, Sept. 2018. 1, 2, 4

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

[19] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proc. of International Conference on Artificial Intelligence and Statistics*, pages 249–256, May 2010. 4

[20] S. Hao, W. Wang, Y. Ye, E. Li, and L. Bruzzone. A deep network architecture for super-resolution-aided hyperspectral image classification with classwise loss. *IEEE Trans. Geoscience and Remote Sensing*, 56(8):4650–4663, 2018. 1

[21] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proc. of international conference on Multimedia(MM)*, pages 675–678, Nov. 2014. 5

[22] Y. Jia, Y. Zheng, L. Gu, A. Subpa-Asa, A. Lam, Y. Sato, and I. Sato. From rgb to spectrum for natural scenes via manifold-based mapping. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 4715–4723, Oct. 2017. 3

[23] J. Jiang, D. Liu, J. Gu, and S. Ssstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 168–179, 2013. 3, 6, 7, 8

[24] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2329–2336, June 2011. 1, 2, 3, 6

[25] R. Kawakami, H. Zhao, R. T. Tan, and K. Ikeuchi. Camera spectral sensitivity and white balance estimation from sky images. *International Journal of Computer Vision*, 105(3):187–204, 2013. 3, 6, 7, 8

[26] D. P. Kingma and J. L. Ba. Adam: a method for stochastic optimization. In *Proc. of International Conference on Learning representations(ICLR)*, May 2015. 4, 5

[27] 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. 6

[28] H. Kwon and Y.-W. Tai. RGB-guided hyperspectral image upsampling. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 307–315, 2015. 1, 2, 6

[29] L. Laetitia, L. B. d. Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, J.-Y. Tourneret, M. A. Veganzones, G. Vivone, Q. Wei, and N. Yokoya. Hyperspectral pansharp-

ening: A review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):27–46, Sept 2015. 2

[30] C. Lanaras, E. Baltsavias, and K. Schindler. Hyperspectral super-resolution by coupled spectral unmixing. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 3586–3594, 2015. 1, 2, 3, 6

[31] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias. Fusing hyperspectral and multispectral images via coupled sparse tensor factorization. *IEEE Trans. Image Processing*, 27(8):4118–4130, 2018. 1, 2, 6

[32] H. A. Y. Maas, Andrew L and A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. of International Conference on Machine Learning (ICML)*, volume 30, June 2013. 3

[33] S. Nie, L. Gu, Y. Zheng, A. Lam, O. Nobutaka, and I. Sato. Deeply learned filter response functions for hyperspectral reconstruction. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1, 2, 4, 5

[34] L. Ojha, M. B. Wilhelm, S. L. Murchie, A. S. McEwen, J. J. Wray, J. Hanley, M. Mass, and M. Chojnacki. Spectral evidence for hydrated salts in recurring slope lineae on Mars. *Nature Geoscience*, 8(11):829–832, Nov. 2015. 1

[35] Y. Qu, H. Qi, and C. Kwan. Unsupervised sparse dirichlet-net for hyperspectral image super-resolution. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2862–2869, June 2018. 1, 2, 5, 6

[36] Y. Schechner and S. Nayar. Generalized mosaicing: wide field of view multispectral imaging. *IEEE Trans. Pattern Analysis and Machine Intelligence* , 24(10):1334–1348, Oct. 2002. 1

[37] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. R. R. Licciardi, and L. Wald. A critical comparison among pansharpening algorithms. *IEEE Trans. Geoscience and Remote Sensing*, 53(5):2565–2586, 2015. 1

[38] L. Wald. Quality of high resolution synthesised images: Is there a simple criterion ? In *Proc. of Conference on Fusion Earth Data*, pages 99–103, 2000. 6

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

[40] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1409–1413, May 2013. 2

[41] X. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei. Anomaly detection in hyperspectral images based on low-rank and sparse representation. *IEEE Trans. Geoscience and Remote Sensing*, 54(4):1990–2000, 2016. 1

[42] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley. Pannet: A deep network architecture for pan-sharpening. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 1753–1761, Oct 2017. 2

[43] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range and spectrum. *IEEE Trans. Image Processing*, 19(9):2241–2253, 2010. 5

[44] N. Yokoya, T. Yairi, and A. Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geoscience and Remote Sensing*, 50(2):528–537, Feb. 2012. 2

[45] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang. Exploiting clustering manifold structure for hyperspectral imagery super-resolution. *IEEE Trans. Image Processing*, 27(12):5969–5982, 2018. 1, 2, 6