Unsupervised Sparse Dirichlet-Net for Hyperspectral Image Super-Resolution

Ying Qu[1]*, Hairong Qi[1], Chiman Kwan[2]
[1]The University of Tennessee, Knoxville, TN [2] Applied Research LLC, Rockville, MD
yqu3@vols.utk.edu hqi@utk.edu chiman.kwan@arllc.net

Abstract

In many computer vision applications, obtaining images of high resolution in both the spatial and spectral domains are equally important. However, due to hardware limitations, one can only expect to acquire images of high resolution in either the spatial or spectral domains. This paper focuses on hyperspectral image super-resolution (HSI-SR), where a hyperspectral image (HSI) with low spatial resolution (LR) but high spectral resolution is fused with a multispectral image (MSI) with high spatial resolution (HR) but low spectral resolution to obtain HR HSI. Existing deep learning-based solutions are all supervised that would need a large training set and the availability of HR HSI, which is unrealistic. Here, we make the first attempt to solving the HSI-SR problem using an unsupervised encoder-decoder architecture that carries the following uniquenesses. First, it is composed of two encoder-decoder networks, coupled through a shared decoder, in order to preserve the rich spectral information from the HSI network. Second, the network encourages the representations from both modalities to follow a sparse Dirichlet distribution which naturally incorporates the two physical constraints of HSI and MSI. Third, the angular difference between representations are minimized in order to reduce the spectral distortion. We refer to the proposed architecture as unsupervised Sparse Dirichlet-Net, or uSDN. Extensive experimental results demonstrate the superior performance of uSDN as compared to the state-of-the-art.

1. Introduction

Hyperspectral image (HSI) analysis has become a thriving and active research area in computer vision with a wide range of applications [7, 5], including, for example, object recognition and classification [24, 12, 53, 31], tracking [44, 13, 42, 43], environmental monitoring [40, 35], and change detection [25, 6]. Compared to multispectral images (MSI with around 10 spectral bands) or conventional color images (RGB with 3 bands), HSI collects hundreds of contiguous bands which provide finer details of spectral signature of different materials. However, its spatial resolution becomes significantly lower than MSI or RGB due to hardware limitations [20, 3]. On the contrary, although MSI or RGB has high spatial resolution, their spectral resolution is relatively low. Very often, to yield better recognition and analysis results, images with both high spectral and spatial resolution are desired [46]. A natural way to generate such images is to fuse hyperspectral images with multispectral images or conventional color images. This procedure is referred to as *hyperspectral image super-resolution (HSI-SR)* [3, 27, 8] as shown in Fig. 1.

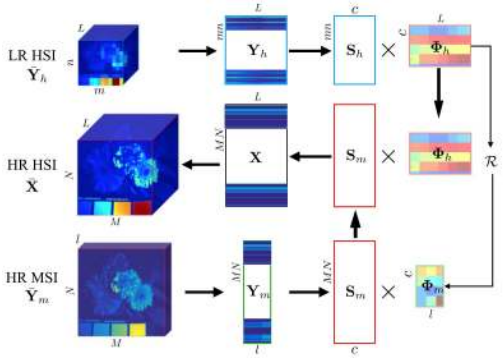

Figure 1. General procedure of HSI-SR.

The problem of HSI-SR originates from *multispectral pan-sharpening (MSI-PAN)* in the remote sensing field, where the spatial resolution of MSI is further improved by a high-resolution panchromatic image (PAN). Note that, in general, resolution refers to the spatial resolution. Usually, MSI has much higher resolution than HSI, but PAN has even higher resolution than MSI. We use LR to denote low spatial resolution and HR for high spatial resolution. There are roughly two groups of MSI-PAN methods, namely, the component substitution (CS) [41, 38, 2] and the multi-resolution analysis (MRA) based approaches [1]. Although MSI-PAN has been well developed through decades of innovations [41, 29, 54], they cannot be readily adopted to solve the HSI-SR problems. On one hand, the amount of spectral information to be preserved for HSI-SR is much higher than that of MSI-PAN, thus it is easier to introduce spectral distortion, *i.e.*, the output image does not preserve

the accurate spectral information [29, 51, 3, 8]. On the other hand, HSI possesses much lower resolution than that of MSI, making it more challenging to improve the spatial resolution.

There have been few methods specifically designed for HSI-SR, including mainly Bayesian based and matrix factorization based approaches [29, 51, 10]. The unique framework of Bayesian offers a convenient way to regularize the solution space of HR HSI by employing a proper prior distribution such as Gaussian [47]. Simoes *et al*. proposed HySure [39], which applied a total variation regularization to smooth the image. Akhtar *et al*. [3] introduced a non-parametric Bayesian strategy to extract spectral dictionary and spatial coefficients from LR HSI and HR MSI, respectively. Matrix factorization based approaches have been actively studied recently [20, 52, 10, 27, 45]. Yokoya *et al*. [52] decomposed both the LR HSI and HR MSI alternatively to achieve the optimal non-negative bases and coefficients that used to generate HR HSI. Lanaras *et al*. [27] further improved the fusion results by introducing a sparse constraint. However, most existing HSI-SR approaches generally assume that the downsampling function between the spatial coefficients of HR HSI and LR HSI are known beforehand. This assumption is not always true due to the distortions caused by both the sensors and complex environmental conditions [3].

HSI-SR is also closely related to the natural image super-resolution (SR) problem, which has been extensively studied and achieved excellent performance through the state-of-the-art *deep learning* [9, 30, 37, 21, 22, 28, 26, 16]. The main principle of SR is to learn a mapping function between LR images and HR images in a supervised fashion. Natural image SR methods usually work on up to $4\times$ upscaling. There have been three attempts to address the MSI-PAN problem with deep learning where the mapping function is learned using different frameworks including tied-weights denoising/ autoencoder [19], SRCNN [32], and deep residual network [16, 48]. These deep learning based methods, including natural image SR and MSI-PAN are all supervised, making their adoption on HSI-SR a challenge due to two reasons. First, they are designed to find an end-to-end mapping function between the LR images and HR images under the assumption that the mapping function is the same for different images. However, the mapping function may not be the same for images acquired with different sensors. Even for the data collected from the same sensor, the mapping function for different spectral bands may not be the same. Thus the assumption may cause severe spectral distortion. Second, training a mapping function is a supervised solution which requires a large dataset, the down-sampling function, and the availability of the HR HSI, that are not realistic for HSI.

In this paper, we propose an *unsupervised* network structure to address the challenges of HSI-SR. To the best of our knowledge, this is the first effort to solving the HSI-SR problem with deep learning in an unsupervised fashion. The novelty of this work is three-fold. First, the network extracts both the spectral and spatial information from LR HSI and HR MSI with two deep learning networks which share the same decoder weights, as illustrated in Fig. 2. Second, in order to incorporate the two physical constraints of HSI and MSI data representation, i.e., sum-to-one and sparsity, the network encourages the representations from both modalities to follow a Dirichlet distribution which naturally incorporates the sum-to-one property. Since each pixel of the image only consists of a few spectral bases, the sparsity of the representations is guaranteed by minimizing their entropy function. Third, to address the challenge of spectral distortion, instead of adopting the down-sampling function (as an estimated mapping function) to relate the representations of the two modalities, we minimize the angular difference of these representations such that they have similar patterns. In this way, the spectral distortion is largely reduced. The proposed method is referred to as uSDN.

2. Problem Formulation

Given the LR HSI, $\bar{\mathbf{Y}}_h \in \mathbb{R}^{m\times n\times L}$, where m, n and L denote the width, height and number of spectral bands of the HSI, respectively, and the corresponding HR MSI, $\bar{\mathbf{Y}}_m \in \mathbb{R}^{M\times N\times l}$, where M, N and l denote the width, height and number of spectral bands of the MSI, respectively, the goal is to estimate the HR HSI, $\bar{\mathbf{X}} \in \mathbb{R}^{M\times N\times L}$, with both high spatial and spectral resolution. In general, MSI has much higher spatial resolution than HSI, *i.e*., $M \gg m$, $N \gg n$, and HSI has much higher spectral resolution than MSI, *i.e*., $L \gg l$. To facilitate the subsequent processing, we unfold the 3D images into 2D matrices, *i.e*., each row of the 2D matrix denotes the spectral reflectance of a given pixel. The unfolded matrices are written as $\mathbf{Y}_h \in \mathbb{R}^{mn\times L}$, $\mathbf{Y}_m \in \mathbb{R}^{MN\times l}$ and $\mathbf{X} \in \mathbb{R}^{MN\times L}$. This is illustrated in Fig. 1.

Assuming that each row of $\mathbf{Y}_h$ is a linear combination of c basis vectors (or spectral signatures), as expressed in Eq. (1), where $\mathbf{\Phi}_h \in \mathbb{R}^{c\times L}$ and each row of which denotes the spectral basis that preserves the spectral information and $\mathbf{S}_h \in \mathbb{R}^{mn\times c}$ is the corresponding proportional coefficients (referred to as *representations* in deep learning). Since the coefficients indicate how the spectral bases are mixed at specific spatial locations, they preserve the spatial structure of HSI.

Similarly, $\mathbf{Y}_m$ can be expressed as Eq. (2), where $\mathbf{\Phi}_m \in \mathbb{R}^{c\times l}$ and each row of which indicates the spectral basis of MSI. $\mathcal{R} \in \mathbb{R}^{L\times l}$ is the transformation matrix given as a prior from the sensor [20, 52, 47, 29, 39, 46, 27, 8], which describes the relationship between HSI and MSI bases. With $\mathbf{\Phi}_h \in \mathbb{R}^{c\times L}$ carrying the high spectral infor-

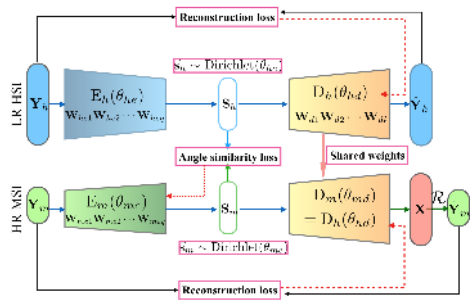

Figure 2. Simplified architecture of uSDN.

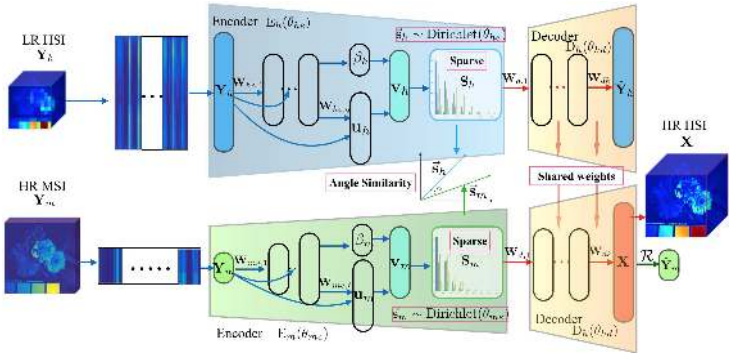

Figure 3. Details of the encoder nets.

mation and $\mathbf{S}_m \in \mathbb{R}^{MN\times c}$ carrying the high spatial information, the desired HR HSI, $\mathbf{X}$, is generated by Eq. (3). See Fig. 1.

$$\mathbf{Y}_h = \mathbf{S}_h\mathbf{\Phi}_h, \tag{1}$$

$$\mathbf{Y}_m = \mathbf{S}_m\mathbf{\Phi}_m, \quad \mathbf{\Phi}_m = \mathbf{\Phi}_h\mathcal{R} \tag{2}$$

$$\mathbf{X} = \mathbf{S}_m\mathbf{\Phi}_h. \tag{3}$$

The problem of HSI-SR can be described mathematically as $P(\mathbf{X}|\mathbf{Y}_h, \mathbf{Y}_m)$. Since the ground truth $\mathbf{X}$ is not available, the problem should be solved in an unsupervised fashion. The key to addressing this problem is to take advantage of the shared information, *i.e.*, $\mathbf{\Phi}_h \in \mathbb{R}^{c\times L}$, to extract desired high spectral bases $\mathbf{\Phi}_h$ and spatial representations $\mathbf{S}_m$ from two different modalities.

In addition, three unique requirements of HSI-SR need to be given special consideration. First, in representing HSI or MSI as a linear combination of spectral signatures, the representation vectors should be non-negative and sum-to-one. That is, $\sum_{j=1}^{c} s_{ij} = 1$, where $\mathbf{s}_i$ is the row vector of either $\mathbf{S}_h$ or $\mathbf{S}_m$ [20, 52, 10, 27, 45]. Second, due to the fact that each pixel of image only consists of a few spectral bases, the representations should be sparse. Third, spectral distortion should be largely reduced in the process in order to preserve the spectral information of HR HSI while gaining spatial resolution.

3. Proposed Approach

We propose an unsupervised architecture as shown in Fig. 2. We highlight the three structural uniquenesses here. First, the architecture consists of two deep networks, for the representation learning of the LR HSI and HR MSI, respectively. These two networks share the same decoder weights, enabling the extraction of both spectral and spatial information from multi-modalities in an unsupervised fashion. Second, in order to satisfy the sum-to-one constraint of the representations, both $\mathbf{S}_h$ and $\mathbf{S}_m$ are encouraged to follow a Dirichlet distribution where the sum-to-one property is naturally incorporated in the network with a further sparsity constraint. Third, to address the challenge of spectral distortion, the representations of two modalities are encouraged to have similar patterns by minimizing their angular difference.

3.1. Network Architecture

As shown in Fig. 2, the network reconstructs both the LR HSI $\mathbf{Y}_h$ and HR MSI $\mathbf{Y}_m$ in a coupled fashion. Taking the LR HSI network (the top network) as an example. The network consists of an encoder $\mathrm{E}_h(\theta_{he})$, which maps the input data to low-dimensional representations (latent variables on the Bottleneck hidden layer), *i.e.*, $p_{\theta_{he}}(\mathbf{S}_h|\mathbf{Y}_h)$, and a decoder $\mathrm{D}_h(\theta_{hd})$ which reconstructs the data from the representations, *i.e.*, $p_{\theta_{hd}}(\hat{\mathbf{Y}}_h|\mathbf{S}_h)$. Both the encoder and decoder are constructed with multiple fully-connected layers. Note that the bottleneck hidden layer $\mathbf{S}_h$ behaves as the representation layer that reflect the spatial information and the weights θ_{hd} of the decoder $\mathrm{D}_h(\theta_{hd})$ serve as $\mathbf{\Phi}_h$ in Eq. (1), respectively. This correspondence is further elaborated below.

The HSI is reconstructed by $\hat{\mathbf{Y}}_h = f_k(\mathbf{W}_{dk}f_{k-1}(...(f_1(\mathbf{S}_h\mathbf{W}_{d1} + b_1)...) + b_{k-1}) + b_k)$, where $\mathbf{W}_{dk}$ denotes the weights in the kth layer. To extract the spectral basis from LR HSI, the latent variables of the representation layer $\mathbf{S}_h$ act as the proportional coefficients, where $\mathbf{S}_h$ follows a Dirichlet distribution with the sum-to-one property naturally incorporated. Suppose the activation function is an identity function and there is no bias in the decoder, we have $\theta_{hd} = \mathbf{W}_1\mathbf{W}_2...\mathbf{W}_k$. That is, the weights θ_{hd} of the decoder correspond to the spectral basis $\mathbf{\Phi}_h$ in Eq. (1) and $\mathbf{\Phi}_h = \theta_{hd}$. In this way, $\mathbf{\Phi}_h$ preserves the spectral information of LR HSI, and the latent variables $\mathbf{S}_h$ preserves the spatial information effectively.

Equivalently, the bottom network reconstructs the HR MSI in a similar way with encoder $\mathrm{E}_m(\theta_{me})$ and decoder $\mathrm{D}_m(\theta_{md})$. However, since $l \leq c \leq L$, *i.e.*, the number of latent variables, L, is much larger than the number of input nodes, l, the MSI network is very unstable and hard to train. On the other hand, the spectral basis of HR MSI

can be transformed from those of LR HSI which possesses more spectral information, the decoder of the MSI is designed to share the weights with that of HSI in terms of $\theta_{md} = \mathbf{\Phi}_m = \theta_{hd}\mathcal{R} = \mathbf{\Phi}_h\mathcal{R}$. Then the reconstructed HR MSI can be obtained by $\hat{\mathbf{Y}}_m = \mathbf{S}_m\mathbf{\Phi}_h\mathcal{R}$. In this way, only the encoder $\mathrm{E}_m(\theta_{me})$ of the MSI is updated during the optimization, where the HR spatial information $\mathbf{S}_m$ is extracted from MSI. Eventually, the desired HR HSI is generated directly by $\mathbf{X} = \mathbf{S}_m\mathbf{\Phi}_h$. Note that the dashed lines in the image show the path of backpropagation which will be elaborated in Sec. 3.4.

3.2. Sparse Dirichlet-Net with Dense Connectivity

To extract stable spectral information, we need to enforce the proportional coefficients $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_i, \cdots, \mathbf{s}_p)^T$ of each pixel to sum-to-one [52, 49, 27, 27], *i.e.*, $\sum_{j=1}^{c} s_{ij} = 1$. Without loss of generality, $\mathbf{S}$ represents either $\mathbf{S}_h$ with $p = mn$ or $\mathbf{S}_m$ with $p = MN$. In addition, due to the fact that only a few spectral bases actually contribute in the linear combination of the spectral reflectance of each pixel, the coefficients should also be sparse. In the proposed architecture, the latent variables (or representations) of the hidden layer $\mathbf{S}_h$ or $\mathbf{S}_m$ correspond to the proportional coefficients in Eqs. (1) and (2). To naturally incorporate the sum-to-one property, the representations are encouraged to follow a Dirichlet distribution which is accomplished with stick-breaking process as illustrated in Fig. 3. Furthermore, entropy function is adopted to reinforce the sparsity of the representations.

The stick-breaking process was first proposed by Sethuranman [36] back in 1994. It is used to generate random vectors $\mathbf{s}$ with Dirichlet distribution. The process can be illustrated as breaking a unit-length stick into c pieces, the length of which follows a Dirichlet distribution. Assuming that the generated vector is denoted as $\mathbf{s} = (s_1, \cdots, s_j, \cdots, s_c)$, we have $0 \leq s_j \leq 1$, and the variables in the vector are sum to one, *i.e.*, $\sum_{j=1}^{c} s_j = 1$. Mathematically [36], a single variable s_j is defined as

$$s_j = \begin{cases} v_1 & \text{for} \quad j = 1 \\ v_j \prod_{o<j}(1 - v_o) & \text{for} \quad j > 1, \end{cases} \tag{4}$$

where v_o is drawn from a Beta distribution, *i.e.*, $v_o \sim \mathrm{Beta}(u, \alpha, \beta)$. Nalisnick and Smyth successfully coupled the expressiveness of generative networks with Bayesian nonparametric model through stick-breaking process [33]. The network uses a Kumaraswamy distribution [23] as an approximate posterior which takes in the samples from a randomly generated uniform distribution during the training procedure.

Different from the generative network, we aim to find shared representations that better reconstruct the data. Therefore, the weights of the network should be changed according to the input data instead of randomly generated distribution. It has been proved that when $v_o \sim \mathrm{Beta}(u, 1, \beta)$, $\mathbf{s}$ follows a Dirichlet distribution. Since it is difficult to draw samples directly from Beta distribution, we draw samples from the inverse transform of Kumaraswamy distribution, as shown in Eq. (5), which is equivalent to Beta distribution when $\alpha = 1$ or $\beta = 1$,

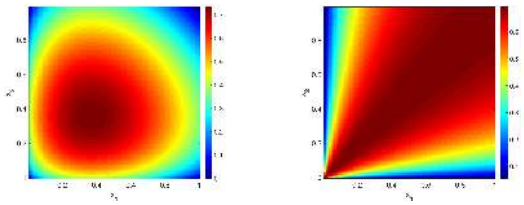

Figure 4. Shannon entropy (L) and Shannon entropy function (R).

$$\mathrm{kuma}(u, \alpha, \beta) = \alpha\beta u^{\alpha-1}(1 - u^{\alpha})^{\beta-1} \tag{5}$$

where $\alpha > 0$, $\beta > 0$ and $u \in (0, 1)$. The benefit of Kumaraswamy distribution is that it has a closed-form CDF, where the inverse transform is defined as

$$v_o \sim (1 - (1 - u^{\frac{1}{\beta}})^{\frac{1}{\alpha}}). \tag{6}$$

Let $\alpha = 1$, parameters u and β are learned through the network as illustrated in Fig. 3. Because $\beta > 0$, a softplus is adopted as the activation function [11] at the β layer. Similarly, a sigmoid [15] is used to map u into $(0, 1)$ range at the $\mathbf{u}$ layer. To avoid gradient vanishing and increase the representation power of the proposed method, the encoder of the network is densely connected, *i.e.*, each layer is fully connected with all its subsequent layers [17].

To further increase the variability of u and β (theoretically, we want the learned u and β to be any number within their range), instead of concatenating all the preceding layers, the input of the kth layer is the summation of all the preceding layers x_0, x_1, x_{k-1} with their own weights, *i.e.*, $\mathbf{W}_0x_0 + \mathbf{W}_1x_1 + ... + \mathbf{W}_{k-1}x_{k-1}$. In this way, fewer number of layers is required to learn the optimal representations.

Although the stick-breaking structure encourages the representations to follow a Dirichlet distribution, it does not guarantee the sparsity of the representations. In addition, the widely used l_1 regularization or Kullback-Leibler divergence [14] will not encourage the representation layer to be sparse either, because they guarantee the sparsity by reducing the mean of active value, *i.e.*, mean of the representation layer. However, due to the stick-breaking structure, the mean of $\mathbf{S}_h$ or $\mathbf{S}_m$ is almost one. Therefore, we introduce a generalized Shannon entropy function [18] to reinforce the sparsity of the representation layer which works effectively even with the sum-to-one constraint.

The entropy function was first proposed in compressive sensing field to solve the signal recovery problem. It is de-

fined as

$$\mathcal{H}_p(\mathbf{s}) = -\sum_{j=1}^{N} \frac{|s_j|^p}{\|\mathbf{s}\|_p^p} \log \frac{|s_j|^p}{\|\mathbf{s}\|_p^p}. \tag{7}$$

Compared to the more popular Shannon entropy, the entropy function Eq. (7) decreases monotonically when the data become sparse. To illustrate the effect, we show the phenomena with 2D variables in Fig. 4. Shannon entropy is small when both x_1 and x_2 are small or large. But for Shannon entropy function, the local minimum only occurs at the boundaries of the quadrants. This nice property guarantees the sparsity of arbitrary data even the data are with the sum-to-one constraint. Due to the stick-breaking structure, the latent variables at the representation layer are positive. We choose $p = 1$ which is more efficient and will encourage the variables to be sparse.

3.3. Angle Similarity

Extracting spatial information from HR MSI is quite challenging and easy to introduce spectral distortion in the subsequent HR HSI results. The main cause to this problem is that the number of the representations c (number of nodes in the representation layer) is much larger than the dimension of the MSI, *i.e.*, $c \gg l$. Previous researchers assume the down-sampling function is available *a-priori* to build a relationship between the representations of HSI and MSI. However, the down-sampling function is usually unknown for real applications.

Therefore, instead of taking the down-sampling function as a prior, we encourage the representations $\mathbf{S}_h$ and $\mathbf{S}_m$ of the two networks following a similar pattern to prevent spectral distortion. And such similarity is measured by the angular difference between the two representations. Spectral angle mapper (SAM) is employed to measure this angular difference. SAM is a spectral evaluation method in remote sensing [29, 51, 34], which measures the angular difference between the estimated image and the ground truth image. The lower the SAM score, the smaller the spectral angle difference, and the more similar the two representations.

Since the HSI and MSI networks share the same decoder weights, the representations should have similar angle in order to generate high quality image with less spectral distortion. Besides encouraging the representation layer to follow a sparse Dirichlet distribution, we further reduce the angular difference of the representations of HSI and MSI during the optimization procedure.

In the network, representations $\mathbf{S}_h \in \mathbb{R}^{mn\times c}$ and $\mathbf{S}_m \in \mathbb{R}^{MN\times c}$, from two different modalities have different dimensions. To minimize the angular difference, we increase the size of the low-dimensional $\mathbf{S}_h$ by duplicating its values at each pixel to its nearest neighborhood. Then the duplicated representations $\tilde{\mathbf{S}_h} \in \mathbb{R}^{MN\times c}$ have the same dimension as $\mathbf{S}_m$. With vectors of equal size, the angular difference is defined as

$$\mathcal{A}(\tilde{\mathbf{S}_h}, \mathbf{S}_m) = \frac{1}{MN}\sum_{i=1}^{MN} \arccos\left(\frac{\tilde{\mathbf{s}}_h^{\,i} \cdot \mathbf{s}_m^{\,i}}{\|\tilde{\mathbf{s}}_h^{\,i}\|_2 \|\mathbf{s}_m^{\,i}\|_2}\right) \tag{8}$$

To map the range of the angle within $(0, 1)$, Eq. (8) is divided by the circular constant π.

$$\mathcal{J}(\tilde{\mathbf{S}_h}, \mathbf{S}_m) = \frac{\mathcal{A}(\tilde{\mathbf{S}_h}, \mathbf{S}_m)}{\pi} \tag{9}$$

3.4. Optimization and Implementation Details

To prevent over-fitting, we applied an l_2 norm on the decoder weights. The objective functions of the proposed network architecture can then be expressed as:

$$\begin{aligned}\mathcal{L}(\theta_{he}, \theta_{hd}) &= \frac{1}{2}\|\mathbf{Y}_h(\theta_{he}, \theta_{hd}) - \hat{\mathbf{Y}}_h(\theta_{he}, \theta_{hd})\|_F^2 \\ &+ \lambda\mathcal{H}_1(\mathbf{S}_h(\theta_{he})) + \mu\|\theta_{hd}\|_F^2,\end{aligned} \tag{10}$$

$$\begin{aligned}\mathcal{L}(\theta_{me}) &= \frac{1}{2}\|\mathbf{Y}_m(\theta_{me}, \theta_{hd}) - \hat{\mathbf{Y}}_m(\theta_{me}, \theta_{hd})\|_F^2 \\ &+ \lambda\mathcal{H}_1(\mathbf{S}_m(\theta_{me})),\end{aligned} \tag{11}$$

$$\mathcal{L}(\theta_{me}) = \mathcal{J}(\tilde{\mathbf{S}}_h(\theta_{he}), \mathbf{S}_m(\theta_{me})), \tag{12}$$

where λ and μ are parameters that balance the trade-off between the reconstruction error and the sparsity and weights loss, respectively.

The proposed architecture consists of two sparse Dirichlet-Nets which extract the spectral information Φ_h from HSI and spatial information $\mathbf{S}_m$ from MSI. The network is optimized with back-propagation following the procedure described below, also illustrated in Fig. 2 with the dashed line.

Step 1: Since the decoder weights θ_{hd} of the HSI network preserves the spectral information Φ_h, we first update the HSI network, given the objective function in Eq. (10), to find the optimal θ_{hd}. To prevent over-fitting, an l_2 norm is applied on the decoder of the HSI network.

Step 2: The estimated decoder weights θ_{hd} are fixed and shared with the decoder of the MSI network. Update the encoder weights θ_{me} of the MSI network given the objective function in Eq. (11).

Step 3: To reduce spectral distortion, every 10 iterations, we minimize the angular difference between the representations of two modalities given the objective function in Eq. (12). Since we already have θ_{he} from the first step, only the encoder θ_{me} of the MSI network is updated during the optimization.

For all the experiments, both the input and output of the HSI network have 31 nodes, representing the number of spectral bands in the data. The numbers of densely-connected layers and nodes of the encoder are shown in

Table 1. There are 3 layers in the HSI network and each layer contains 10 nodes. The MSI network has 5 layers with the number of nodes increases from 4 to 10. The $\mathbf{v}_h$/$\mathbf{v}_m$ are drawn with Eq. (6) given $\mathbf{u}_h$/$\mathbf{u}_m$ and β_h/β_m, which are learned by back-propagation. Both β_h and β_m have only one node, denoting the distribution parameter of each pixel. The representation layers, $\mathbf{S}_h$ and $\mathbf{S}_m$ with 10 nodes are constructed with $\mathbf{v}_h$ and $\mathbf{v}_m$, respectively, according to Eq. (4). The network shares the decoder with 2 layers and each layer has 10 nodes. Since different images have different spectral bases and representations, the network is trained on each pair of LR HSI and HR MSI to reconstruct each image accurately.

Table 1. The number of layers and nodes in the network.

Dirichlet-Net	Encoder			
	#layers and #nodes	$\mathbf{u}$	β	$\mathbf{v}$
HSI	3 / [10,10,10]	10	1	10
MSI	5 / [4,5,7,910]	10	1	10

4. Experiments and Results

4.1. Datesets and Experimental Setup

The proposed uSDN has been thoroughly evaluated with two widely used benchmark datasets, CAVE [50] and Harvard [7]. The CAVE dataset consists of 32 HR HSI images and each of which has a dimension of 512×512 with 31 spectral bands. These spectral images are taken within the wavelength range $400 \sim 700$nm with an interval of 10 nm. The Harvard dataset includes 50 HR HSI images with both indoor and outdoor scenes. The dimension of the images in this dataset is 1392×1040, with 31 bands taken at an interval of 10nm within the wavelength range of $420 \sim 720$nm. Note that for this dataset, the top left corner of size $1024 \times 1024 \times 31$ is cropped as the HR HSI.

For the two benchmark datasets, the LR HSI $\mathbf{Y}_h$ is obtained by averaging the HR HSI over 32×32 disjoint blocks. The HR MSI images with 3 bands are generated by multiplying the HR HSI with the given spectral response matrix $\mathcal{R}$ of Nikon D700. All the images are normalized between 0 and 1. Note that the CAVE dataset is in general considered a more challenging set than Harvard since images in Harvard usually contain more smooth reflections; and since the images have higher spatial resolution, pixels within close vicinity usually have similar spectral reflectance. Hence, even the images are down-sampled by the 32×32 kernel, most spectral information is still preserved in the LR HSI.

The results of the proposed method on individual images are compared with seven state-of-the-art methods, *i.e.*, CS based [2], MRA based [1], CNMF [52], Bayesian Sparse (BS) [47], HySure [39], Lanaras's 15 (CSU) [27], and Akhtar's 15 (BSR) [3] that belong to different categories of approaches described in Sec. 1. These methods also reported the best performance [29, 3, 27], with the original code made available by the authors. We also directly list results [4] from Akhtar's 16 (HBPG) since the code is not available. The average results on the complete dataset is also reported to evaluate the robustness of the proposed method. For quantitative comparison, the root mean squared error (RMSE) and spectral angle mapper (SAM) are applied to evaluate the reconstruction error and the amount of spectral distortion, respectively.

4.2. Experimental Results

Tables 2 and 3 show the experimental results of 7 groups of images from the CAVE and Harvard datasets, which are commonly benchmarked by existing literature [20, 3, 4]. We observe that traditional CS-based and MRA-based methods suffer from spectral distortion, thus could not achieve competitive performance. The Bayesian based approach, BS [47], fails due to the fact that it assumes the representation $\mathbf{S}_m$ follows a Gaussian distribution, which is not always true. However, the Bayesian non-parametric based method BSR [3] outperforms BS because it estimates the spectra through non-parametric learning. The matrix-based approaches,CNMF [52] and CSU [27], are not as competitive on the CAVE dataset due to their predefined down-sampling function, although they perform much better on the Harvard dataset. We also observe that some methods like Hysure can achieve better RMSE, but worse SAM scores, that is because they cannot preserve the spectral information properly which has caused large spectral distortion. Based on the experiments, the proposed uSDN powered by the unique sparse Dirichlet-net outperforms all of the other approaches in terms of both RMSE and SAM, and it is quite stable for different types of input images.

Table 2. Benchmarked results in terms of RMSE.

Methods	CAVE					Harvard	
	balloon	CD	cloth	photo	spool	img1	imgb5
CS	25.4	19.4	22.0	18.2	25.8	16.7	17.8
MRA	12.5	14.2	15.4	4.8	11.3	4.7	8.9
BS	14.2	15.3	17.6	11.3	15.2	10.9	14.7
Hysure	14.9	20.3	14.8	4.6	12.5	4.4	5.4
BSR	2.6	7.9	4.3	2.1	6.2	2.3	2.5
CNMF	9.0	11.9	10.1	5.2	12.2	3.2	4.5
CSU	13.3	10	6.7	3.1	7.9	2.2	2.6
uSDN	**1.8**	**4.8**	**3.7**	**2.0**	**5.3**	**2.0**	**0.7**
HBPG	1.9	5.3	3.7	–	–	2.2	0.8

To further demonstrate the robustness of the proposed uSDN, we report the mean of RMSE and SAM over the complete CAVE and Harvard dataset in Table 3. We only list the performance of matrix factorization based CSU and Bayesian based BSR, since they demonstrated better performance as shown in Tables 1 and 2. We observe that since

Table 3. Benchmarked results in terms of SAM.

Methods	CAVE					Harvard	
	balloon	CD	cloth	photo	spool	img1	imgb5
CS	19	17	17	82	48	15	14
MRA	12	9	11	14	15	13	15
BS	11	16	10	18	24	17	18
Hysure	18	24	18	19	38	18	19
BSR	11.9	17.9	6	14	16	1.9	3.4
CNMF	10	9	7	11	20	10	13
CSU	8.9	25	12.6	10	17	1.8	2.8
uSDN	**4.7**	**10**	**4.8**	**5.4**	**13**	**1.6**	**1.7**
HBPG	7.6	10.6	5.0	–	–	2.5	2.1

Table 4. The average RMSE and SAM scores over complete benchmarked datasets.

Methods	CAVE		Harvard	
	RMSE	SAM	RMSE	SAM
CSU[27]	9.96	15.63	3.37	5.35
BSR[3]	5.29	13.63	2.61	4.46
uSDN	**4.09**	**6.95**	**1.78**	**4.05**

BSR estimates the representations separately from the spectral bases, although it can achieve good RMSE scores, its SAM scores are not promising. While CSU relates the representations with a predefined down-sampling function, and thus achieves better results on the Harvard dataset, it generates worse results on the CAVE dataset. Both methods may cause spectral distortion in different scenarios. The proposed approach consistently outperforms the other methods in terms of both RMSE and SAM as reported in Table 4.

We also make two further observations. First, since the Harvard dataset is less challenging than the CAVE dataset, the improvement on the former is not as apparent as that on the latter. This, on the other hand, demonstrates that uSDN can handle challenging scenarios much better than state-of-the-art. Second, uSDN is very effective in preserving the spectral signature of the reconstructed HR HSI, showing much improved performance especially on SAM on CAVE. The main reason that contributes to the success of the proposed approach is that it relates the representations $\mathbf{S}_h$ and $\mathbf{S}_m$ with statistics and angular difference, *i.e.*, both representations are encouraged to follow a Dirichlet distribution, and their angular difference is enforced to be small. In this way, both the reconstruction error and spectral distortion are effectively reduced. Since the representation is enforced to be sparse Dirichlet over each pixel, not the entire image, the proposed structure is capable of recovering different pixels individually. And the total number of the recovered samples, that equals the number of pixels, is large. This demonstrates the representation capacity of the proposed structure.

To visualize the results, we show the reconstructed samples from CAVE and Harvard taken at wavelengths 460, 540, and 670 nm in Fig. 5. The first through fourth columns show the LR images, reconstructed images from our method, ground truth images, and the absolute difference between the images at the second and third columns, respectively. We also compare the proposed method with CSU and BSR on the challenge dataset CAVE and show the results in Fig. 6. The effectiveness of the proposed method can be readily observed from the difference images, where the proposed approach is able to preserve both the spectral and spatial information.

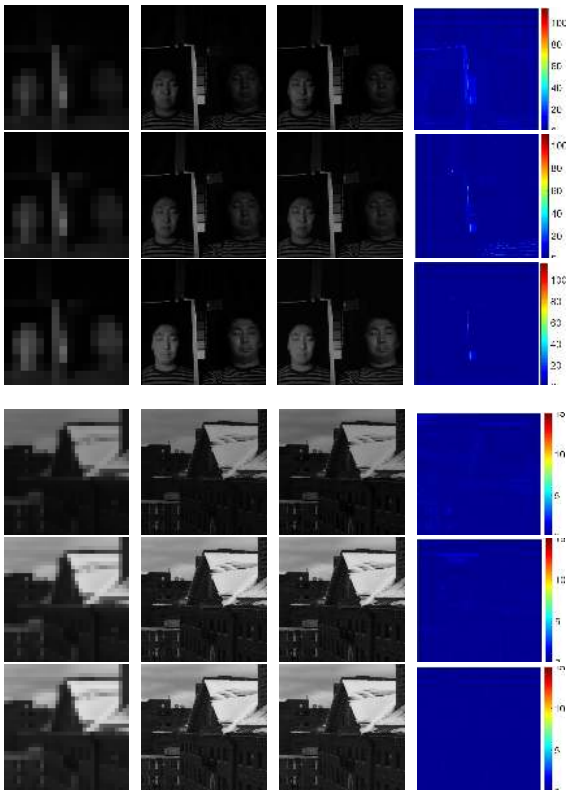

Figure 5. Reconstructed images from the CAVE (top) and Harvard dataset (bottom) at wavelength 460, 540 and 620 nm. First column: LR images (16×16). Second: estimated images (512×512). Third: ground truth images. Fourth: absolute difference.

Ablation Study: Taking the 'pompom' image from the CAVE dataset as an example, we further evaluate 1) the necessity of enforcing the representation to follow sparse Dirichlet and 2) the usage of angle similarity loss. Fig. 7 illustrates the RMSE of the reconstructed HR HSI using 4 different network structures, *i.e.*, autoencoder (AE) without any constraints, AE with the sparsity constraint (SAE), a simple Dirichlet-Net without any constraints, and the proposed sparse Dirichlet-Net. We observe that the adoption of Dirichlet-Net significantly reduces RMSE as compared to AE and SAE; and the proposed sparse Dirichlet-Net reduces RMSE even further, especially as the number of iterations increases. Fig. 9 shows the summation of elements in $\mathbf{s}_j$ averaged over all pixels in the image, where

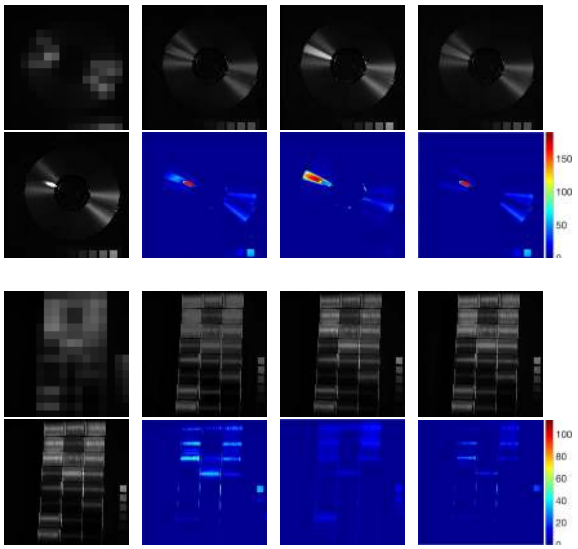

Figure 6. Reconstructed images of two examples (top two rows and bottom two rows) from the CAVE dataset at wavelength 670 nm. The first column shows the LR image (top) and the ground truth image (bottom). The second, third and fourth columns are the reconstructed results (top) and the absolute difference (bottom) from CSU, BSR and uSDN, respectively.

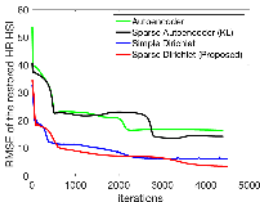

Figure 7. Sparse Dirichlet constraint.

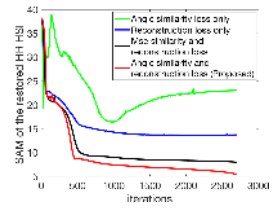

Figure 8. Spectral angle constraint.

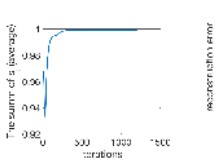

Figure 9. Summation.

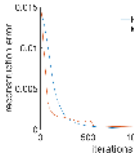

Figure 10. Learning curves.

Figure 11. The RMSE curve.

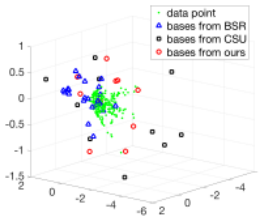

Figure 12. Spectral basis.

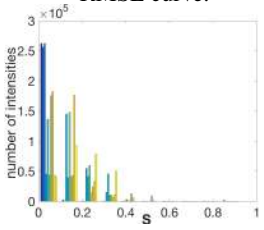

Figure 13. Histogram of $\mathbf{S}_m$.

we observe that representations $\mathbf{s}_j$ are sum-to-one almost surely after around 300 iterations with Dirichlet-Net. Fig. 8 demonstrates the spectral angle mapper (SAM) of the reconstructed HR HSI using 4 different loss functions when the MSI network is updated, *i.e.*, only with angle similarity loss, only with reconstruction loss, reconstruction loss with MSE similarity, and the proposed reconstruction loss with angle similarity, respectively. We observe that reconstruction loss significantly stabilizes/regulates the convergence process; and reconstruction loss with angle similarity presents the lowest SAM and fastest convergence speed.

Convergence Study: During the optimization, both the HSI and MSI networks converge smoothly as shown in Fig. 10. The MSI network has a little bit fluctuation caused by the angular difference which is minimized every 10 iterations between the representations of two modalities.

Effect of Free Parameters: There are two free parameters in the algorithm design, i.e., μ for the decoder weight loss and λ for the sparsity control, as shown in Eq. (10). We keep $\mu = 1e^{-6}$ during the experiments. Fig. 11 shows how RMSE is decreasing when we increase λ from 2×10^{-7} to 1×10^{-6}. We set $\lambda = 1 \times 10^{-6}$ in the experiments.

Visualizing $\mathbf{S}_m$ and $\mathbf{\Phi}_h$: The proposed structure is based on the assumption that the LR HSI, HR MSI, and HR HSI can be formulated as a linear combination of their corresponding spectral bases. Here, we would like to provide visualization results of the spatial representation, $\mathbf{S}_m$, its sparsity property, and the spectral bases, $\mathbf{\Phi}_h$. We use the pompom image from the CAVE dataset as the testing image to generate all the visualization. In order to visually see if the linear combination assumption is valid or not, we project the estimated bases, $\mathbf{\Phi}_m$ into a 3D space using singular value decomposition. In Fig. 12, we observe that the learned bases from CSU is a little bit far away from the data, while the bases from BSR cluster with each other and do not cover all the data. The bases from our method circumscribe the entire data, indicating a more effective representation of the data. We also study if $\mathbf{S}_m$ is indeed sparse or not. The histogram of the learned representations $\mathbf{S}_m$ is shown in Fig. 13, where the sparsity is clearly evident.

5. Conclusion

We proposed an unsupervised sparse Dirichelet-Net (uSDN) to solve the problem of hyperspectral image super-resolution (HSI-SR). To the best of our knowledge, this is the first effort to solving the problem of HSI-SR in an unsupervised fashion. The network extracts the spectral basis from LR HSI with rich spectral information and spatial representations from HR MSI with high spatial information through a shared decoder. The representations from two modalities are encouraged to follow a sparse Dirichlet distribution. In addition, the angular difference of two representations is minimized during the optimization to reduce spectral distortion. Extensive experiments on two benchmark datasets demonstrate the superiority of the proposed approach over state-of-the-art.

Acknowledgement: This work was supported in part by NASA NNX12CB05C and NNX16CP38P.

References

[1] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva. Mtf-tailored multiscale fusion of high-resolution ms and pan imagery. *Photogrammetric Engineering & Remote Sensing*, 72(5):591–596, 2006.

[2] B. Aiazzi, S. Baronti, and M. Selva. Improving component substitution pansharpening through multivariate regression of ms+ pan data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10), 2007.

[3] N. Akhtar, F. Shafait, and A. Mian. Bayesian sparse representation for hyperspectral image super resolution. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3631–3640, 2015.

[4] N. Akhtar, F. Shafait, and A. Mian. Hierarchical beta process with gaussian process prior for hyperspectral image super resolution. *European Conference on Computer Vision*, pages 103–120, 2016.

[5] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(2), 2012.

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

[7] A. Chakrabarti and T. Zickler. Statistics of real-world hyperspectral images. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 193–200, 2011.

[8] R. Dian, L. Fang, and S. Li. Hyperspectral image super-resolution via non-local sparse tensor factorization. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5344–5353, 2017.

[9] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[10] 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 Transactions on Image Processing*, 25(5):2337–2352, 2016.

[11] C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia. Incorporating second-order functional knowledge for better option pricing. *Advances in neural information processing systems*, pages 472–478, 2001.

[12] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2013.

[13] Y. Fu, Y. Zheng, I. Sato, and Y. Sato. Exploiting spectral-spatial correlation for coded hyperspectral image restoration. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[14] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[15] J. Han and C. Moraga. The influence of the sigmoid function parameters on the speed of backpropagation learning. *From Natural to Artificial Neural Computation*, pages 195–201, 1995.

[16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[17] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*, 2016.

[18] S. Huang and T. D. Tran. Sparse signal recovery via generalized entropy functions minimization. *arXiv preprint arXiv:1703.10556*, 2017.

[19] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang. A new pan-sharpening method with deep neural networks. *IEEE Geoscience and Remote Sensing Letters*, 12(5):1037–1041, 2015.

[20] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2329–2336, 2011.

[21] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[22] J. Kim, J. Kwon Lee, and K. Mu Lee. Deeply-recursive convolutional network for image super-resolution. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[23] P. Kumaraswamy. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1-2):79–88, 1980.

[24] C. Kwan, B. Ayhan, G. Chen, J. Wang, B. Ji, and C.-I. Chang. A novel approach for spectral unmixing, classification, and concentration estimation of chemical and biological agents. *IEEE Transactions on Geoscience and Remote Sensing*, 44(2):409–419, 2006.

[25] H. Kwon and N. M. Nasrabadi. Kernel matched signal detectors for hyperspectral target detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 6–6, 2005.

[26] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2017.

[27] C. Lanaras, E. Baltsavias, and K. Schindler. Hyperspectral super-resolution by coupled spectral unmixing. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3586–3594, 2015.

[28] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016.

[29] L. Loncan, L. B. de Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, et al. Hyperspectral pansharpening:

a review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3), 2015.

[30] J. Lu and D. Forsyth. Sparse depth super resolution. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[31] E. Maggiori, G. Charpiat, Y. Tarabalka, and P. Alliez. Recurrent neural networks to correct satellite image classification maps. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9):4962–4971, Sept 2017.

[32] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7):594, 2016.

[33] E. Nalisnick and P. Smyth. Deep generative models with stick-breaking priors. *ICML*, 2017.

[34] S. Ozkan, B. Kaya, E. Esen, and G. B. Akar. Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing. *arXiv preprint arXiv:1708.01894*, 2017.

[35] A. Plaza, Q. Du, J. M. Bioucas-Dias, X. Jia, and F. A. Kruse. Foreword to the special issue on spectral unmixing of remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4103–4110, 2011.

[36] J. Sethuraman. A constructive definition of dirichlet priors. *Statistica sinica*, pages 639–650, 1994.

[37] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient subpixel convolutional neural network. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[38] S. C. Sides, J. A. Anderson, et al. Comparison of three different methods to merge multiresolution and multispectral data- landsat tm and spot panchromatic. *Photogrammetric Engineering and remote sensing*, 57(3):295–303, 1991.

[39] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3373–3388, 2015.

[40] L. H. Spangler, L. M. Dobeck, K. S. Repasky, A. R. Nehrir, S. D. Humphries, J. L. Barr, C. J. Keith, J. A. Shaw, J. H. Rouse, A. B. Cunningham, et al. A shallow subsurface controlled release facility in bozeman, montana, usa, for testing near surface co2 detection techniques and transport models. *Environmental Earth Sciences*, 60(2):227–239, 2010.

[41] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot. Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1301–1312, 2008.

[42] B. Uzkent, M. J. Hoffman, and A. Vodacek. Real-time vehicle tracking in aerial video using hyperspectral features. *The IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2016.

[43] B. Uzkent, A. Rangnekar, and M. Hoffman. Aerial vehicle tracking by adaptive fusion of hyperspectral likelihood maps. *The IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017.

[44] H. Van Nguyen, A. Banerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 44–51, 2010.

[45] M. A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot. Hyperspectral superresolution of locally low rank images from complementary multisource data. *IEEE Transactions on Image Processing*, 25(1):274–288, 2016.

[46] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5), 2015.

[47] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7), 2015.

[48] Y. Wei, Q. Yuan, H. Shen, and L. Zhang. Boosting the accuracy of multi-spectral image pan-sharpening by learning a deep residual network. *arXiv preprint arXiv:1705.07556*, 2017.

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

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

[51] N. Yokoya, C. Grohnfeldt, and J. Chanussot. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 5(2):29–56, June 2017.

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

[53] F. Zhang, B. Du, and L. Zhang. Scene classification via a gradient boosting random convolutional network framework. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3):1793–1802, 2016.

[54] J. Zhou, C. Kwan, and B. Budavari. Hyperspectral image super-resolution: a hybrid color mapping approach. *Journal of Applied Remote Sensing*, 10(3):035024–035024, 2016.