*Article*

# Spectral Super-Resolution with Optimized Bands

**Utsav B. Gewali** [1,]*[iD] **, Sildomar T. Monteiro** [2][iD] **and Eli Saber** [1,3]

1    Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY 14623, USA
2    Boeing Research and Technology, Huntsville, AL 35824, USA
3    Department of Electrical & Microelectronic Engineering, Rochester Institute of Technology, Rochester, NY 14623, USA
*    Correspondence: ubg9540@rit.edu

**Abstract:** Hyperspectral (HS) sensors sample reflectance spectrum in very high resolution, which allows us to examine material properties in very fine details. However, their widespread adoption has been hindered because they are very expensive. Reflectance spectra of real materials are high dimensional but sparse signals. By utilizing prior information about the statistics of real HS spectra, many previous studies have reconstructed HS spectra from multispectral (MS) signals (which can be obtained from cheaper, lower spectral resolution sensors). However, most of these techniques assume that the MS bands are known apriori and do not optimize the MS bands to produce more accurate reconstructions. In this paper, we propose a new end-to-end fully convolutional residual neural network architecture that simultaneously learns both the MS bands and the transformation to reconstruct HS spectra from MS signals by analyzing large quantity of HS data. The learned band can be implemented in hardware to obtain an MS sensor that collects data that is best to reconstruct HS spectra using the learned transformation. Using a diverse set of real-world datasets, we show how the proposed approach of optimizing MS bands along with the transformation can drastically increase the reconstruction accuracy. Additionally, we also investigate the prospects of using reconstructed HS spectra for land cover classification.

**Keywords:** hyperspectral spectra; multispectral signals; spectral super-resolution; optimized bands; convolutional neural networks

## 1. Introduction

The reconstruction of hyperspectral (HS) spectrum from multispectral (MS) signal is an ill-posed problem. This is because MS bands are broader and fewer in number, so they do not capture fine details in reflectance spectra that are captured by HS bands. Therefore, it becomes necessary to use prior knowledge about the statistics of the real HS spectra in order to accurately reconstruct HS spectra from MS signals. Past studies have successfully utilised machine learning techniques, such as, sparse dictionary learning [1] and deep learning [2], for this purpose. However, they have only concentrated on learning the transformation from MS signal to HS spectrum for fixed MS bands. They do not optimize the location or the resolution of the MS bands. In this paper, we propose a deep learning based approach that can jointly optimize the MS bands and the transformation to reconstruct HS spectra by analyzing large quantity of HS data. Our method is based on a novel end-to-end fully convolutional residual neural network architecture that utilizes a new band re-sampling layer and a new loss layer.

Hyperspectral sensors capture reflected energy at very high spectral resolution over hundreds of narrow, contiguous bands over visible and infrared spectral range, while on the other hand, multispectral sensors captures energy over the same spectral range with a small number of broad bands (typically less that 15) [3]. While the higher spectral resolution is essential for analyzing material

properties in fine detail and at sub-pixel scales, high spectral resolution sensors are very expensive. It is a well-known fact that HS signals are oversampled signal and lie in a low-dimensional subspace within a high dimensional space [4]. This means that there is a possibility that HS spectrum could be reconstructed from low dimensional signal, such as MS signal. Due to the potential of obtained high resolution spectra from cheaper MS sensor, recently the topic of utilizing super-resolution techniques to reconstruct HS spectrum from MS signal captured by MS sensors has gained a lot of attention [2].

In the past, super-resolution techniques have been primarily applied to hyperspectral remote sensing for improving the spatial resolution of the HS images by fusing a low spatial resolution HS image with a high spatial resolution panchromatic or MS image of the same scene [5]. However, there have been growing interests in methods, called single image super-resolution, that map low spatial resolution HS image to high spatial resolution image without the use of high spatial resolution panchromatic or MS imagery [6–8]. These methods utilize prior knowledge about spatial distribution of HS images, which is learned by training models on large dataset of imagery. Similarly, models can be trained to perform single image spectral super-resolution, i.e., reconstruct hyperspectral imagery from multispectral imagery without the aid of any additional imagery or data. In these methods, the spatial resolution of the input image and the reconstructed image is the same but the number of bands is different–the input image contains few broad bands while the output image contains numerous contiguous narrow bands.

Conventional approaches for single image spectral super-resolution have utilized models such as PCA [9,10], kernel ridge regression [11], and radial basis network [12]. Most of the modern methods utilize dictionary learning or deep networks, however there has been on-going research on application of other machine learning algorithms, such as Gaussian processes, for spectral super-resolution [13]. Arad and Ben-Shahar [1] proposed a method that jointly learned a RGB (red-green-blue) dictionary that mapped RGB values to sparse code and a HS dictionary that mapped the sparse code to HS spectrum in the visible spectral range. Similarly, many studies have recently utilized deep neural networks to map an RGB image to a visible spectral range hyperspectral image [2]. Galliani et al. [14] trained a 2 dimensional fully-convolutional deep network, inspired from Tiramisu network [15], to predict hyperspectral image from RGB image. They showed that their method can produce superior result compared to the dictionary learning approach of Arad and Ben-Shahar [1]. Apart from using a very deep architecture with 56 layers, their method is different from [1] in one important aspect. Their method uses spatial context while performing spectral resolution as it uses 2 dimensional convolutional network trained on patches of HS images. Following this finding, Aeschbacher et al. [16] proposed a dictionary learning method, based on A+ method [17], that outperformed the method by Galliani et al. [14] and showed that a very deep network is not necessarily needed for spectral super-resolution. One of the improvements in their method was that they utilized spatial context by training on spectra of spatially down-sampled images or training on $3 \times 3$ neighborhood of pixels. More recently, studies [18,19] have found that moderately deep residual networks [20] can provide better performance than very deep model, such as [14], and dictionary learning approaches (shallow models), such as [16]. Gwn Lore et al. [21] investigated generative adversarial networks (GANs) for spectral super-resolution, however their method did not outperform previous methods.

The proposed method is different from all of methods discussed so far. The above mentioned methods only optimize the mapping from MS signal to HS spectrum, however our method jointly optimizes the MS bands and the transformation from MS signal to HS spectra. Our method is based on the hypothesis that better reconstructions of HS spectra could be possible if we were to optimize the centers and the resolutions of the MS bands along with the model to reconstruct HS spectra. The MS bands and the super-resolution method can be optimized by analyzing a large quantity of HS data and later the optimized bands can be implemented in hardware to build a MS sensor whose signal can be used to reconstruct HS signal using learned super-resolution model. This is similar to the idea of HS compressed sensing, where the goal is to develop hardware and data processing techniques to

reconstruct high dimensional HS spectra by making measurements much fewer than the dimension of the HS spectra [22–24].

Many optimization techniques have been applied to select best MS bands for spectra super-resolution. A classical approach utilized simulated annealing to optimize the centers and the full-width-half-maximums (FWHMs) of Gaussian MS bands such that those bands produced best reconstruction [25]. Some recent studies have investigated the feasibility of finding optimum MS bands for spectral super-resolution using discrete optimization, where the goal in this approach is to find a subset from a large collection of MS bands which produces the best reconstruction accuracy [26–29]. Unfortunately, these approaches are not scalable when the number of extracted MS bands becomes much larger than 3 and the spectral range covered is not just visible range, as used in these studies. It is because in these cases the number of possible combination of bands becomes very large. Nie et al. [30] used a convolutional neural network to learn the response value of the MS bands at every wavelength measured by a HS sensor over the visible spectral range. However, their method becomes ineffective when learning MS bands to cover a larger spectral region, for example, the entire visible to shortwave spectral region, because their method optimizes bands freely without bounds to the width of the bands. Hence, a single extracted band could extend very large spectral ranges (have very large FWHM), possibly the entire visible to shortwave spectral region, which is not realizable in real sensor hardware.

The proposed method assumes the MS bands are Gaussian in shape (similar to [25]) and allows for optimization of the band center and the band FWHMs using back-propagation along with the optimization of super-resolution neural network. Our method allows constraints to be applied on the centers and the FWHMs of the optimized bands to handle design constraints that may arise when building sensor hardware. Another important aspect of our study is that unlike almost all of the previous studies which have concentrated on terrestrial HS images with spectral range only in visible region, we experiment with terrestrial, airborne, and space-based imagery with some of them spanning a spectral range covering the entirety of visible to shortwave infrared (350 nm to 2500 nm). The main contributions of the paper are as follows. We present a neural network architecture for optimizing MS bands and a super-resolution network that provides high fidelity reconstruction of HS spectra. The network utilizes a novel layer called tunable spectral sub-sampling layer which contains tunable MS bands. We also propose a new loss function for reconstruction of HS spectra. Experiments are performed on diverse set of images. The performance of the optimized bands is evaluated for fine-level land cover classification, one of the important applications of HS imaging [31]. The paper is organized as follows. Section 2 introduces the proposes method, Section 3 provides experimental evaluation of the proposed method on various datasets and Section 4 concludes the paper.

## 2. Proposed Method

We propose an end-to-end convolutional residual neural network, with a novel tunable spectral sub-sampling layer and a new spectral loss function, to jointly optimize the multispectral (MS) bands and the transformation to reconstruct hyperspectral (HS) spectra from MS signal. This method can optimize both the center and the resolution of the MS bands. Only the number of MS bands to be optimized should be fixed a priori. Additionally, constraints on the MS band centers and resolutions, that could occur in real sensor design situations, can be incorporated in our method.

### 2.1. Tunable Spectral Sub-Sampling Layer

Let radiance or reflectance, $r(\lambda)$, be measured within a wavelength range $\lambda_{\min}$ to $\lambda_{\max}$ by M consecutive hyperspectral (HS) bands with band centers $[\lambda_1, \cdots, \lambda_M]$ to obtain a HS signal, $\mathbf{h} = [h_1, \cdots, h_M]$. Here, $h_i$ is the radiance/reflectance measured by the $i$-th HS band. If we are to assume that the number of HS bands is very large with the bands being narrow and contiguous (i.e., $\lambda_{i+1} - \lambda_i \to 0$, $\forall i \in \{1, \cdots, M-1\}$), their responsivities can be approximated by delta functions, e.g., $\delta(\lambda - \lambda_i)$ for $i$-th band. Then, $r(\lambda)$ can be approximated as,

$$r(\lambda) \approx \sum_{i=1}^{M} h_i \delta\left(\lambda - \lambda_i\right). \tag{1}$$

Let $r(\lambda)$ be measured again by N multispectral bands to obtain a MS signal, $\mathbf{m} = [m_1, \cdots, m_N]$, such that $m_j$ is the radiance/reflectance measured by the $j$-th MS band. The MS bands are much fewer in number than the HS bands ($N \ll M$). We assume the MS bands to be Gaussian in shape with $[\mu_1, \cdots, \mu_N]$ being the centers and $[\sigma_1, \cdots, \sigma_N]$ being the standard deviations. $[\mu_1, \cdots, \mu_N]$ are the locations (band centers) of the bands and $[\sigma_1, \cdots, \sigma_N]$ control the resolution of the bands (i.e., FWHM of $i$-th band is approximately $2.3548\,\sigma_i$).

The signal measured by each MS band is given by

$$m_j = \int_{\lambda_{\min}}^{\lambda_{\max}} r(\lambda)\, g(\lambda, \mu_j, \sigma_j)\, d\lambda, \ \forall j \in \{1, \cdots, N\}, \tag{2}$$

where, $g(\lambda, \mu_j, \sigma_j) = \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\lambda - \mu_j}{\sigma_j}\right)^2}$.

On combining (1) and (2),

$$m_j \approx \sum_{i=1}^{M} h_i\, g(\lambda_i, \mu_j, \sigma_j), \ \forall j \in \{1, \cdots, N\}. \tag{3}$$

This result has been used by numerous past studies, e.g., [10,25,32]. It should be noted that (3) is only accurate when the spectral resolution and the band spacing of the HS bands are very small compared to spectral resolution of MS bands as assumed earlier. Equation (3) can be written in matrix algebra form as,

$$\mathbf{m}_{1 \times N} = \mathbf{h}_{1 \times M}\, \mathbf{R}_{M \times N}, \tag{4}$$

where, $R_{i,j} = g(\lambda_i, \mu_j, \sigma_j)$ is the $i$-th row and $j$-th column element of the responsivity matrix $\mathbf{R}$. The dimensions of the MS signal vector $\mathbf{m}$, the HS signal vector $\mathbf{h}$, and the responsivity matrix are shown in Equation (4) for clarity.

We define a new layer, called tunable spectral sub-sampling layer, to perform the operation in (3), i.e., convert a HS signal to MS signal. This layer is similar to a fully connected layer that connects M input nodes to N output nodes. However, unlike a fully connected layer whose weight matrix is unconstrained (i.e., each element can take any real value), the weights matrix of the tunable spectral sub-sampling layer is constrained to have Gaussian column vectors. This is shown in Figure 1. The values of elements of $\mathbf{R}$ are parameterized by $[\mu_1, \cdots, \mu_N]$ and $[\sigma_1, \cdots, \sigma_N]$. Therefore, $[\mu_1, \cdots, \mu_N]$ and $[\sigma_1, \cdots, \sigma_N]$ are the parameters of the tunable spectral sub-sampling layer. In other words, the parameters of the sub-sampling layers are the locations and the resolutions of the N MS bands, all of which are independently tunable.

We can optimize $[\mu_1, \cdots, \mu_N]$ and $[\sigma_1, \cdots, \sigma_N]$ via backpropagation. The partial derivatives of the layer output with respect to the parameters, needed for backpropagation, are given by,

$$\frac{\partial m_j}{\partial \mu_j} = \sum_{i=1}^{M} h_i \frac{\partial g(\lambda_i, \mu_j, \sigma_j)}{\partial \mu_j}$$

$$= \sum_{i=1}^{M} h_i \frac{\lambda_i - \mu_j}{\sigma_j^2} g(\lambda_i, \mu_j, \sigma_j) \tag{5}$$

$$\frac{\partial m_j}{\partial \sigma_j} = \sum_{i=1}^{M} h_i \frac{\partial g(\lambda_i, \mu_j, \sigma_j)}{\partial \sigma_j}$$

$$= \sum_{i=1}^{M} h_i \left[ \frac{(\lambda_i - \mu_j)^2}{\sigma_j^3} - \frac{1}{\sigma_j} \right] g(\lambda_i, \mu_j, \sigma_j). \tag{6}$$

To summarize, (4) gives expression for the forward pass through the tunable spectral sub-sampling layer and (5) and (6) provide partial derivatives for backward pass through the layer.
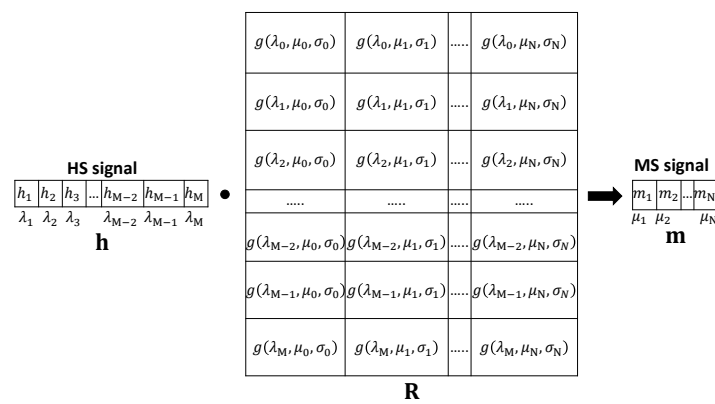


**Figure 1.** Tunable spectral sub-sampling layer.

## 2.2. Network Architecture

Figure 2 shows the proposed network architecture. The input to the network is a hyperspectral spectrum and the output is a reconstructed version of the input spectrum. There are two major parts in the network–(i) tunable spectral sub-sampler layer and (ii) spectral super-resolution sub-network. The tunable spectral sub-sampler portion takes as input HS spectrum (**h**) and resamples the HS spectrum to produce a MS signal (**m**). The spectral super-resolution portion of the network takes a MS signal (**m**) as input and reconstructs the HS signal ($\hat{\mathbf{h}}$).
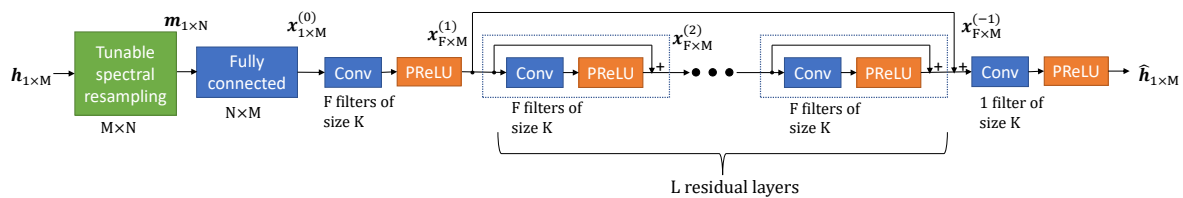


**Figure 2.** Proposed network architecture.

The spectral super-resolution sub-network consists of a fully connected layer, followed by a one-dimensional convolutional layer, a non-linear layer, a series of residual blocks, a one-dimensional convolutional layer, and a non-linear layer in a cascade. The input of the spectral super-resolution sub-network is a MS signal and the output is the reconstructed HS signal. The fully connected layer converts the N dimensional MS signal into M dimensional signal, M being the dimension of HS signal. Hypothetically, had $M \geq N$, we would require only a fully connected layer to reconstruct HS signal from MS signal. When $M = N$, the fully connected layer could learn weight matrix $\mathbf{R}^{-1}$ and when

M > N, the fully connected layer could learn weight matrix $\mathbf{R}^{\dagger}$ (pseudoinverse of $\mathbf{R}$) to retrieve $\mathbf{h}$ from $\mathbf{m}$. However, in real world scenario, M is always much smaller than N.

When M $\ll$ N, we cannot recover HS signal by just using a fully connected layer because the problem becomes ill-posed. We need to use prior knowledge about real spectra to perform the conversion. Had real hyperspectral spectra spanned the entire M dimensional space, $\mathbb{R}^M$, (i.e., no structure), it would be impossible to recover them. But, real spectra have structure and they lie in a subspace within high dimensional space [33]. So we can utilize information about statistics about real spectra to reconstruct HS spectra from MS signal. In our network, we utilize a residual network architecture for this purpose. After the fully connected layer, the proposed network consists of a convolution layer with F 1-D filters of size K and parametric rectified linear unit (PrReLU) layer. The purpose of this is to obtain F feature maps of length M, $\mathbf{x}^{(1)}$. We use PrReLU as non-linearity through out the network and all of filters used in all convolutional layers are of length K. Following the PrReLU layer, we have a series of L residual blocks. Each residual block consists of a convolutional layer with F filters of length K and a PrReLU layer. There is also a skip connection from the input of the first residual block to the output of the final residual block. After the final residual block, we convolve with one filter of length K and follow with a PrReLU layer to obtain the final prediction ($\hat{\mathbf{h}}$).

Another way to view our network is in the terms of encoder-decoder network. The tunable spectral sub-sampler is the encoder and the spectral super-resolution sub-network is the decoder. The encoder projects the high-dimensional HS signal onto a linear (not non-linear) subspace encoding (MS signal) and the super-resolution portion of the network projects the low dimensional encoding onto the original high dimensional space.

The network is trained on a large dataset of unlabeled HS data to find best MS bands and super-resolution network by jointly optimizing parameters of the tunable spectral sub-sampling layer and the super-resolution sub-network by back-propagation. Once the network is trained, we can separate the optimal band information and the super-resolution network. Then, a new MS sensor can be designed matching the specification of the optimal bands. Any MS image acquired by this sensor can be projected into HS image using the super-resolution network.

While training the network, we do not directly optimize $[\mu_1, \cdots, \mu_N]$ and $[\sigma_1, \cdots, \sigma_N]$. It is because we want to make sure that the bands lie within the measured spectral range. Failing to constrain band center could cause network to get stuck in local minima (because MS signal is zero if the band is outside of the range) or a portion of the band falling outside of the measured spectral wavelength range, $[\lambda_{min}, \lambda_{max}]$. $[\sigma_1, \cdots, \sigma_N]$ should be also constrained such that FWHM of the MS bands are significantly larger than the FWHM of the HS bands, so that assumptions made for (3) are true. Also, in a real sensor design problem we have limitations on the range of values of band centers and FWHM. To accommodate all these, we further parameterize $[\mu_1, \cdots, \mu_N]$ and $[\sigma_1, \cdots, \sigma_N]$ as functions of $[\mu'_1, \cdots, \mu'_N]$ and $[\sigma'_1, \cdots, \sigma'_N]$:

$$\sigma_j = \sigma_{j,\min} - (\sigma_{j,\max} - \sigma_{j,\min}) \, S(\sigma'_j), \tag{7}$$

$$\mu_j = (\lambda_{j,\min} - b\sigma_j) - (\lambda_{j,\max} - \lambda_{j,\min} - 2b\sigma_j) \, S(\mu'_j), \tag{8}$$

$\forall j \in \{1, \cdots, N\}$, where $S(x) = \frac{1}{1+\exp(-x)}$ is the sigmoid function. $\lambda_{j,\min}$ and $\lambda_{j,\max}$ are the minimum and maximum values of the wavelengths that the $j$-th band can span. In other words, this means that the $j$-th band can have non-zero responsivity only between the wavelength range $[\lambda_{j,\min}, \lambda_{j,\max}]$. Similarly, $\sigma_{j,\min}$ and $\sigma_{j,\max}$ are the minimum and maximum value limits of $\sigma_j$, respectively, for all $j \in \{1, \ldots, N\}$. The values of $\lambda_{j,\min}, \lambda_{j,\max}, \sigma_{j,\min}$, and $\sigma_{j,\max}$ for all $j \in \{1, \ldots, N\}$ are fixed by the sensor designer before training the model. In (7), the sigmoid function maps the value of $\sigma'_j$ onto $[0, 1]$ and $\sigma_j$ gets a value in the range $[\sigma_{j,\min}, \sigma_{j,\max}]$. Similarly, $\mu_j$ is mapped within the range $[\lambda_{j,\min} - b\sigma_j, \lambda_{j,\max} - b\sigma_j]$ in (8). The guard gap of $b\sigma_j$ guarantees that the width of the band lie within the limits, $[\lambda_{j,\min}, \lambda_{j,\max}]$. If there were no guard gap, the model could possibly put $\mu_j$ too close to $\lambda_{j,\min}$ or $\lambda_{j,\max}$, and a significant portion of the band may fall outside of $[\lambda_{j,\min}, \lambda_{j,\max}]$ range. $b$ controls the width of the guard gap and is set by sensor designer. A value of $\geq 3$ is a good choice for

$b$, as a value of Gaussian function is negligible past three times of the standard deviation. In this way, our method gives flexibility to the designer of the sensor, to set specifications for the extracted bands by fixing limits for individual band centers and band resolution. In the tunable spectral sub-sampling layer, $[\mu'_1, \cdots, \mu'_N]$ and $[\sigma'_1, \cdots, \sigma'_N]$ are the actual parameters that are optimized via back-propagation. Their value can be any real number. For back-propagation, we can easily calculate partial derivatives needed to optimize $\mu'_j$ and $\sigma'_j$ by calculating partial derivatives of (7) and (8) and combining with (5) and (6) using the chain rule.

In design situations, where there is no specification for limits of $[\mu_1, \cdots, \mu_N]$ and $[\sigma_1, \cdots, \sigma_N]$, $\lambda_{j,\min}$ and $\lambda_{j,\max}$ has to be set to $\lambda_{\min}$ and $\lambda_{\max}$ so that the extent (or width) of the extracted bands do not cross the limit $[\lambda_{\min}, \lambda_{\max}]$. $\sigma_{j,\min}$ has to be set to a value much greater than resolution of the HS bands, but in our study we found that $\sigma_{j,\max}$ is unnecessary, so that a simpler form, such as $\sigma_j = \sigma_{j,\min} + |\sigma'_j|$, can be used instead of (7).

While training models, we use different learning rates for the spectral sub-sampling layer and the super-resolution sub-network, say, $\alpha_{\text{sub}}$ and $\alpha_{\text{sup}}$ respectively. All of the weights of convolutional filters in the super-resolution sub-network have $L_2$-regularization with a weight decay factor $\beta$.

## 2.3. Spectral Loss Function

Mean squared error (MSE) is a poor choice for loss function to reconstruct HS spectra because it tend to produce blurred/smooth reconstruction. This is problematic because in reflectance spectra the information about material properties is manifested as sharp valleys and peaks in reflectance curve, called spectral features [34]. Past studies have shown that the first and the second derivative of the spectra with respect to wavelength can capture these spectral features [35]. This can be seen in the example in Figure 3, where $\mathbf{h}$, $\hat{\mathbf{h}}_1$, and $\hat{\mathbf{h}}_2$ are the original spectrum and two reconstructions, respectively. Both of the reconstructions have equal squared Euclidean distance from the original, but the second, $\hat{\mathbf{h}}_2$, is much smoother and does not capture the shape (features) of the original spectrum while the first, $\hat{\mathbf{h}}_2$, does. That means if we just look at squared error, both reconstructions are equally good. However, if we are to compare the first and the second derivatives of the reconstructions to the original, the difference is clear. The first reconstruction has much smaller squared Euclidean distances between its and original spectrum's first and second derivatives.
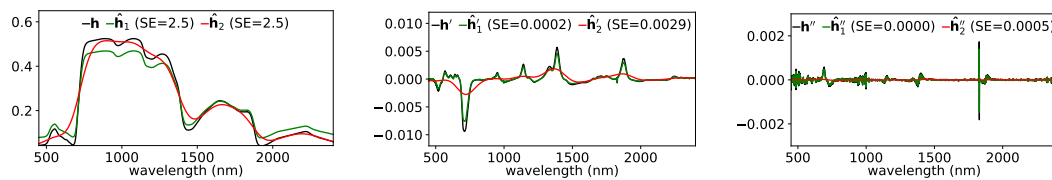


**Figure 3.** Comparison of magnitude, first derivative, and second derivative of the original spectrum ($\mathbf{h}$) and two reconstructions ($\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$). SE is the squared Euclidean distance between the original and reconstructions.

Based on this observation, we propose the following error metric $d_{\text{spectral}}(\mathbf{h}, \hat{\mathbf{h}})$ to compute the error between the original and the reconstructed spectra,

$$d_{\text{spectral}}(\mathbf{h}, \hat{\mathbf{h}}) = \left\| \mathbf{h} - \hat{\mathbf{h}} \right\|^2 + w_1 \left\| \frac{d\mathbf{h}}{d\lambda} - \frac{d\hat{\mathbf{h}}}{d\lambda} \right\|^2 + w_2 \left\| \frac{d^2\mathbf{h}}{d^2\lambda} - \frac{d^2\hat{\mathbf{h}}}{d^2\lambda} \right\|^2.$$

It is a weighted sum of the Euclidean distances between the original and reconstruction and their first and second derivatives. $w_1$ and $w_2$ are the hyperparameters. The derivatives can be implemented using finite difference approximations. The loss function for optimization is given by the average of the spectral error metrics computed over a batch.

### 2.4. Hyperparameter Tuning Using Bayesian Optimization

We tune the hyperparameters of the proposed model (F, K, L, $\beta$, $\alpha_{\text{sub}}$, $\alpha_{\text{sup}}$, $w_1$, and $w_2$) using Bayesian optimization. Bayesian optimization [36] is a derivative-free black-box optimization algorithm that can optimize an objective function by sequentially querying its value for different input values. It is best suited for optimizing function whose value is difficult or expensive to evaluate and whose input space is low dimensional (<20) [37]. Common black-box optimization schemes, such as grid-search and random search, are not applicable in this scenario as they require a large number of function evaluations to get accurate results.

Let us assume we want to minimize an objective function, $f(\mathbf{x})$, whose input is vector $\mathbf{x}$. Elements of $\mathbf{x}$ can be either continuous or discrete. The only operation that can be performed on $f(\mathbf{x})$ is that it can be evaluated for arbitrary $\mathbf{x}$ in its input domain. As stated earlier, it is very costly to evaluate $f(\mathbf{x})$, so we want to intelligently search over the input space in order to minimize the number of function evaluations. Bayesian optimization iteratively queries $f(\mathbf{x})$, while internally modeling $f(\mathbf{x})$ using a function, $g(\mathbf{x})$, learned on the queried input and output. The $g(\mathbf{x})$ is called the surrogate model. Popular choices for surrogate models are Gaussian processes [38] and random forests [36]. It captures the beliefs about $f(\mathbf{x})$ and is updated in each iteration with the input-output values of the query. Acquisition function, $\gamma(g(\mathbf{x}))$, defined over $g(\mathbf{x})$ picks the next query point based on some objective that balances searching unexplored regions in the input space and searching around good solutions to refine them. Some of the common acquisition functions are probability of improvement, expected improvement, and entropy search [36]. It is important to properly choose $g(\mathbf{x})$ and $\gamma(g(\mathbf{x}))$ because the form $g(\mathbf{x})$ controls how accurately $f(\mathbf{x})$ is modeled and $\gamma(g(\mathbf{x}))$ defines the search strategy. Algorithm 1 shows the steps involved in Bayesian optimization. Readers who are interested in learning more about Bayesian optimization are encouraged to study the review by Shahriari et al. [36].

Bayesian optimization is commonly used to tune continuous as well as discrete hyperparameters of machine learning models. When tuning hyperparameters of models, the validation loss can be used as the objective function, $f(\mathbf{x})$, with the elements of $\mathbf{x}$ being the hyperparameters of the model [38].

---

**Algorithm 1:** Bayesian Optimization

**Let** : Objective function: $f(\mathbf{x})$, Surrogate model: $g(\mathbf{x})$, Acquisition function: $\gamma(g(\mathbf{x}))$
**Initialize:** $n \leftarrow 0$, $\mathcal{D}_0 \leftarrow \{(\mathbf{x}, f(\mathbf{x})) : \text{random } \mathbf{x}\}$, $y_{min} \leftarrow \infty$

1 **while** *not stopping criteria* **do**
2     Select a new query point, $\mathbf{x}_{n+1}$, by maximizing $\gamma$:   $\mathbf{x}_{n+1} \leftarrow \underset{\mathbf{x}}{\arg\max}\, \gamma\left(g\left(\mathbf{x}; \mathcal{D}_n\right)\right)$
3     Query objective function: $y_{n+1} \leftarrow f(\mathbf{x}_{n+1})$
4     **if** $y_{n+1} < y_{min}$ **then**
5        $(\mathbf{x}_{min}, y_{min}) \leftarrow (\mathbf{x}_{n+1}, y_{n+1})$
6     Augment the data: $\mathcal{D}_{n+1} \leftarrow \{\mathcal{D}_n, (\mathbf{x}_{n+1}, y_{n+1})\}$
7     Update the surrogate model: learn $g(\mathbf{x}; \mathcal{D}_{n+1})$
8     $n \leftarrow n + 1$
9 **return** $(\mathbf{x}_{min}, y_{min})$

---

### 2.5. Validation Procedure and Datasets

We experiment with six diverse real world datasets. Each dataset was divided into separate training, validation, and testing sets. Models were trained on the training set, and the performance of those models were evaluated on the testing set and reported in the results. The validation set was used for early stopping the training of the network and also to tune the hyperparameters of the model. For training the neural network, Adam [39] was used as optimizer and a mini-batch size of 128 was used. The weights of the model were initialized using the Xavier method [40]. Early stopping with validation after every 1000 Adam iteration and a patience of 25 was used as stopping criteria.

The validation was performed on 64,000 randomly selected spectra from the validation set and the average spectral angle (SAM) between the input and the reconstruction was used as the measure for validation loss. The reason average SAM, not root mean square error (RMSE), was used to measure validation performance is that spectral angle is a better metric to compare the shapes of the spectra and also because if we are to use RMSE to measure the validation loss, it would be easier to minimize the loss by setting $w_1$ and $w_2$ to equal to zero and this would suppress the benefits of our unique loss function.

The hyperparameters of the model were tuned using Bayesian optimization (BO) by selecting values that minimize the validation loss. There are 8 hyperparameters of the proposed model–F, K, L, $\beta$, $\alpha_{sub}$, $\alpha_{sup}$, $w_1$, and $w_2$. Some of them are continuous and some are discrete. Limits were set for possible values of each hyperparameter. The values of the learning rates ($\alpha_{sub}$ and $\alpha_{sup}$) and the weight decay factor ($\beta$) are continuous, and were bounded to $[10^{-5}, 10^{-1}]$. The limit values of $w_1$ and $w_2$ was set to $[0, 1]$. The number of features (F), the filter size (K), and the number of residual layers (L) are discrete hyperparameters whose possible values were limited to fixed sets of integer values. L was limited to a integer a value in the interval $[1, 16]$, K was limited to $\{3, 5, 7, 9, 11\}$, and F was limited to $\{4, 8, 16, 32, 64, 128\}$. Gaussian process was used as the surrogate function ($g(\mathbf{x})$) and expected improvement was used as the acquisition function ($\gamma(g(\mathbf{x}))$). The cost minimized by the BO is the validation loss of the trained neural network. BO was run for 50 function evaluations (BO not finished by 2.5 days was prematurely stopped). In each evaluation, a model was trained with certain setting of hyperparmeters (selected by the acquisition function) and the validation loss (as measured by average spectral angle between 64,000 randomly selected spectra from the validation set and their reconstructions was computed. The model with lowest validation loss was returned as final model. The models were trained on a computer with NVIDIA Titan X Pascal GPU, Intel Core i7-980 at 3.3 GHz processor, and 12 GB of RAM.

Datasets

**Outdoor scenes dataset:** The first and the second datasets were acquired by terrestrial cameras. The first dataset, which we call "Outdoor scenes dataset", contains images of outdoor scenes measured in raw sensor readings over visible wavelengths. It was collected by Arad and Ben-Shahar [1] and contains Specim PS Kappa DX4 hyperspectral camera acquired images with 31 HS bands in the visible wavelengths–400 nm to 700 nm at 10 nm increments. We use "rural" and "park" subsets of the dataset in the experiment. The rural subset contains 5 images and the park subset contain 9 images, each $1392 \times 1300$ pixels in size. We train the model on one subset and test on another. From the subset that is used for training, 70% of spectra is used for training and the remaining 30% is used for validation. We use this dataset to demonstrate that our method with optimized bands can outperform baseline methods that use fixed bands and conventional spectral estimation technique.

**CAVE dataset:** The second dataset, called CAVE dataset [41], contains indoor images of natural and artificial materials captured over visible wavelengths in raw sensor readings. The images contain 31 bands covering 400 nm to 700 nm at 10 nm increments as well. We use this dataset to compare our method with state-of-the-art methods for HS reconstructions, as many of them have published results on this dataset. We follow the same experimental procedure as used by previous studies, e.g., [14,16], and measure the reconstruction accuracy by performing 4-fold cross-validation by dividing the 32 images into 4 groups, each consisting of 8 images.

**Aerial images dataset:** The remaining four datasets contain remote sensing images. The third dataset, which will be referred to as "Aerial images dataset", contains NASA Goddard's LiDAR, Hyperspectral & Thermal Imager (G-LiHT) [42] airborne HS reflectance images collected over different regions in the United States. The images have 114 bands covering a spectral range in visible and near infrared (400 nm to 1000 nm) and pixel size on the ground is around 1 m. We construct training, validation, and testing sets with spectra from images captured at separate geographical locations in the United States. The training set contains spectra from images collected at Bowie in Maryland, Rhode

Island, and Hanover in Connecticut. The validation set contains spectra from images captured at San Marcos in California and University of Tennessee-Knoxville. The testing set contains spectra from images collected at Cape Cod in Massachusetts, White Mountains in Maine, and Rochester in New York. This is the largest of all datasets in this paper, containing about $189 \times 10^6$ training spectra, $9 \times 10^6$ validation spectra and $53 \times 10^6$ testing spectra (Cape Cod: $33 \times 10^6$, White Mountain: $13 \times 10^6$, and Rochester: $7 \times 10^6$). We use this dataset to show how we can constrained the optimized bands in our method to match real sensor design situations.

**Satellite images dataset:** The fourth dataset, which we call "Satellite images dataset", contains satellite imagery obtained from Hyperion sensor [43]. Spectra in these images are measured as radiance in $W\,m^{-2}\,sr^{-1}\,nm^{-1}$ and each pixel is around 30 m on the ground. Hyperion has two separate sensors for visible-near infrared and shortwave spectral ranges with some overlapping bands [43]. We extracted 196 contiguous band ranging from 400 nm to 2500 nm for our experiments. Training, validation, and testing sets comprises images of different cities in the USA. The training set contains HS spectra from images collected from Baltimore, Boston, New York, Philadelphia, Pittsburgh, and Syracuse. The validation set contains HS spectra from image captured from New Heaven. The test set contains HS spectra from images collected from Atlanta, DC, and Rochester. In total, there are about $8.9 \times 10^6$ training spectra, $1.8 \times 10^6$ validation spectra, and $2.5 \times 10^6$ testing spectra (Atlanta: $8 \times 10^5$, DC: $8.9 \times 10^5$, Rochester: $8.3 \times 10^5$). We use this dataset to show how learned bands can be better for reconstruction than bands of LANDSAT-8 [44] MS sensor. Using the third and the fourth datasets, we also demonstrate how models trained on images collected in one geographical region can be applied for HS reconstruction in completely different geographical region.

**Land cover classification dataset:** The fifth dataset called "Land cover classification dataset", contains radiance images, spanning visible to shortwave infrared wavelengths, collected by NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor [45]. The imagery contains 220 bands and was collected over an agricultural site in northwestern Indiana called Indian pines test site 3 [46]. The spectral range covered is 400 nm to 2500 nm and the ground sampling distance is 20 m. This dataset contains data from two flight lines. A small portion of one of the flight lines, spanning an area of about $2 \times 2$ miles, has a corresponding ground truth land cover map available. This labeled area is $145 \times 145$ pixels in size and a land cover label is available for most of those pixels. We divide unlabeled pixels ($2.7 \times 10^6$ spectra) into training (70%) and validation (30%). The testing set consists of spectra in the $145 \times 145$ area with ground cover class labels. Instead of calculating reconstruction metrics between the reconstruction and the original signal in the testing subset, we compare the classification performance of the models trained on the original signal and the models trained on reconstruction. The classification model used throughout the experiments is a support vector machine (SVM) with radial basis function kernel. A 3-fold cross-validation is used to tune the cost parameter (C parameter) and the kernel scale of the SVM. Also, of all the land cover classes in the labeled set, we only experiment with 13 classes which cover more than 50 pixels in the image.

**Transfer test dataset:** The last dataset, which we call "Transfer test dataset", also contains imagery collected by AVIRIS sensor [47]. We use this dataset to investigate whether the optimized bands and the super-resolution sub-network trained on data from one sensor can be use applied to data obtained from a different sensor. For this purpose, we evaluate models trained on G-LiHT sensor collected data (Aerial images dataset) on this dataset. In total, $28.6 \times 10^6$ spectra are present in this dataset and were extracted from imagery collected by AVIRIS sensor during 3 flight lines over Bay Area in California.

## 3. Experimental Results

### 3.1. Natural Images

3.1.1. Outdoor Scenes

We now evaluate the proposed method on the outdoor scenes dataset. In order to compare with results published in Arad and Ben-Shahar [1], we optimize three MS bands with the proposed approach in this experiment. In their work, Arad and Ben-Shahar used sparse dictionary learning to reconstruct 31 bands from a red-green-blue (RGB) image. They used CIE standard color matching functions to simulate RGB from 31 bands. We present results of our model with tunable spectral sub-sampling layer (with no constraints on band centers and FWHMs) and with spectral sub-sampling layer replaced with CIE color matching function (i.e., no optimization of the MS bands). Table 1 shows the comparison of reconstruction performance of the proposed method against [1]. Four metrics are reported. Two metrics (RRMSE and RMSE) are same as the ones reported in Arad and Ben-Shahar [1]. RRMSE and RMSE are relative root mean square error and root mean square error between the test image and reconstruction, averaged over the test images. RRMSE is root mean square of the difference between input and reconstruction normalized by the input and RMSE is root mean square of difference between input and reconstruction normalized to 8-bit (0-255) value. SAM (Spectral Angle Mapper) and $R$ are the angle and the Pearson's correlation coefficient between the input spectrum and the reconstruction averaged over all test spectra in all test images.

Additionally, we have included the results from a traditional PCA-based spectral reconstruction [9] and a conventional approach to optimize MS bands [25] in Table 1. The PCA-based spectral reconstruction was performed on MS images obtained with the RGB bands, the learned bands acquired from our method and the bands optimized by conventional approach. The conventional approach to band optimization [25] used simulated annealing [48] to find the bands that produces best reconstruction with PCA-based spectral super-resolution approach [9]. The original approach used a colorimetric cost function to measure the reconstruction accuracy but we chose to use RMSE, as the main focus of this paper is remote sensing applications where the spectral range is typically not just constrained to visible range.

**Table 1.** Comparison of reconstruction accuracy on Outdoor scenes dataset.

| Dataset | | Method | RRMSE | RMSE | SAM | R |
|---|---|---|---|---|---|---|
| Train | Test | | | | | |
| Rural | Park | Ours (Learned bands) | **0.0317** | **1.433** | **0.0269** | **0.987** |
| | | Ours (RGB bands) | 0.0443 | 2.082 | 0.0307 | 0.980 |
| | | PCA [9] (RGB bands) | 0.0465 | 1.964 | 0.0379 | 0.948 |
| | | PCA [9] (Our learned bands) | 0.0492 | 1.609 | 0.0342 | 0.943 |
| | | PCA [9] (bands optimized by simulated annealing [25]) | 0.0469 | 1.579 | 0.0341 | 0.946 |
| | | Arad & Ben-Shahar [1] | 0.0801 | 2.693 | - | - |
| Park | Rural | Ours (Learned bands) | **0.0264** | **1.591** | **0.0216** | **0.995** |
| | | Ours (RGB bands) | 0.0303 | 2.455 | 0.0265 | 0.993 |
| | | PCA [9] (RGB bands) | 0.0337 | 2.333 | 0.029 | 0.991 |
| | | PCA [9] (Our Learned bands) | 0.0283 | 1.64 | 0.0235 | 0.994 |
| | | PCA [9] (bands optimized by simulated annealing [25]) | 0.0342 | 1.769 | 0.0246 | 0.994 |
| | | Arad & Ben-Shahar [1] | 0.0592 | 3.121 | - | - |

The proposed method showed superior performance to all compared approaches. We find that optimized MS bands provide better reconstruction accuracy compared to RGB bands. Furthermore, it is also seen that convolutional neural network based methods can provide better reconstruction than sparse dictionary learning and traditional PCA-based approach. It is interesting to observe that the PCA-based approach performed better when reconstructing MS images obtained from our

learned bands than MS images obtained from the standard RGB bands. In one case (testing on Park subset), PCA-based method using our learned bands outperformed PCA-based method with traditional band optimization. This reconfirms the superiority of the bands acquired from our method for spectral reconstructions.

Figure 4 shows our learned bands along with mean per band RRMSE (with standard deviation) computed over the test images. Interestingly, the three optimized bands seem to roughly cover blue, green, and red spectral range. For all examples in Figure 4, we find that per band error is minimum around centers of the MS bands and increase as the separation from the center increases. The error is particularly high towards the extreme ends of the spectral range of the HS signal. It should be because in the edge the network has information from only one side of the HS band being predicted. It is also seen that error is generally higher in areas where the responsivity of all of the MS bands are low. This could be happening because in those areas the neural network gets less information from the MS bands to predict HS reflectance accurately.
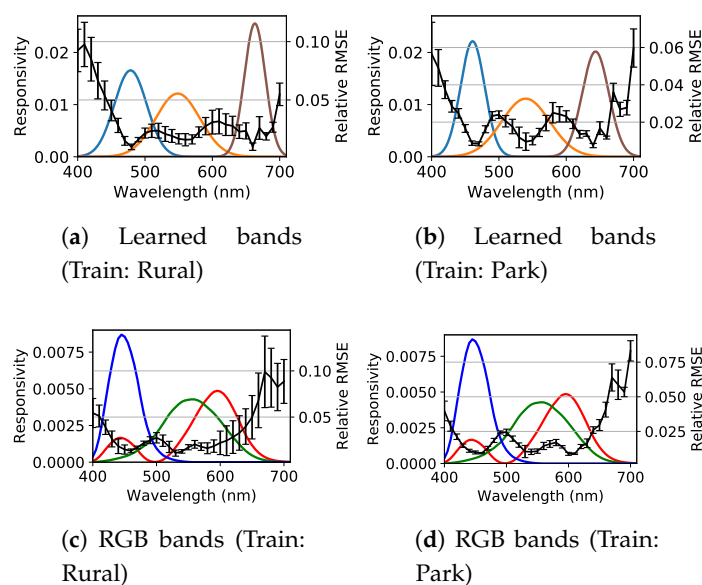


(**a**) Learned bands (Train: Rural)

(**b**) Learned bands (Train: Park)

(**c**) RGB bands (Train: Rural)

(**d**) RGB bands (Train: Park)

**Figure 4.** Responsivity of our learned bands and average relative root mean square error (RMSE) across reconstructed bands (black line) for Outdoor scenes dataset.

For the model trained on Rural subset, our learned MS bands had centers 478.48 nm, 548.58 nm, and 663.14 nm with FWHMs 56.74 nm, 84.34 nm, and 46.58 nm, respectively. Similarly, for the model trained on Park subset, our learned MS bands had centers 460.89 nm, 539.79 nm, and 643.52 nm with FWHMs 42.45 nm, 84.34 nm, and 46.58 nm, respectively. We see that there is slight difference in the band locations and resolutions when the model was trained in different subsets. It could have stem from the fact that the size of these subset are relatively small and substantially different from each other. Another important fact to note is that there is no guarantee that the solution obtained by optimizing a neural network is a global minimum. Different solutions could be obtained with similar dataset if the models find different local minima solutions.

For comparison, the MS bands obtained from the traditional approach had centers 473.09 nm, 538.09 nm, and 643.16 nm with FWHMs 57.16 nm, 50.56 nm, and 45.17 nm, respectively, for Rural subset and centers 478.73 nm, 554.12 nm, and 625.19 nm with FWHMs 57.16 nm, 61.45 nm, 51.5 nm, and 58.11 nm, respectively, for Park subset. It is interesting to observe the similarity between the bands found by ours and the traditional approach.

Figures 5 and 6 show examples of reconstructed spectra and examples of bands in the reconstructed image obtained from our model with learned bands, respectively. These plots indicate that there is variability in the reconstruction error observed for pixels belonging to different material

classes. This is expected as reconstruction accuracy for particular class of material is assumed to be function of how many training pixels belonging to that class were present in the training set and how much inter-class spectral variability is shown by that material.
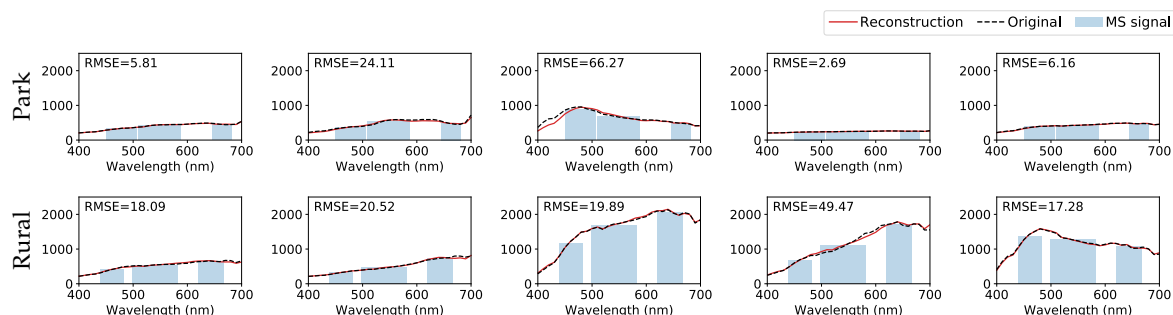


**Figure 5.** Examples of spectra reconstructions from our method with learned bands. Spectra randomly selected from the test set of Outdoor scenes dataset. The RMSE shown is unnormalized.
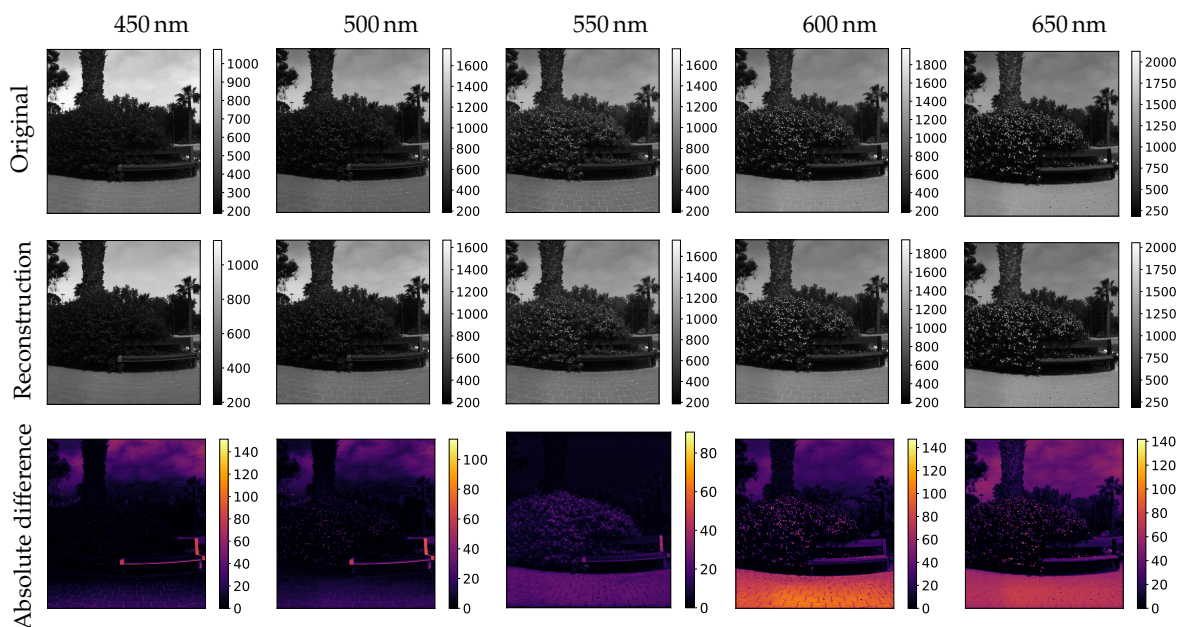


**Figure 6.** Five bands of an image in Outdoor scenes dataset reconstructed from learned bands by our method.

### 3.1.2. Comparison with State-of-the-Art Approaches on CAVE Dataset

We compare our method with the state-of-the-art methods on CAVE [41] dataset. Our approach is different than other methods in that our method optimizes three MS bands along with the super-resolution network while the other approaches use fixed three MS bands (RGB bands) and only optimize the transformation from RGB to HS. Table 2 lists the reconstruction error of our method and the state-of-the-art approaches in terms of per-pixel mean root mean square error (RMSE). RMSE is only shown because this is the common metric published by most of the state-of-the-art methods.

We find that our method outperforms state-of-the-art methods. The result is congruent to the results in the last experiment and demonstrate that optimizing the MS bands along with the transformation from MS signal to HS spectra can vastly improve the reconstruction accuracy. It should be noted that most of the state-of-the-art methods utilize spatial contextual information in their prediction, which our method does not. The reconstruction performance of our method could be further improve if we are to replace the current spectral super-resolution sub-network with a two

dimensional network and train on image patches. However, that is currently out of the scope of this paper and will be explored in future studies.

**Table 2.** Comparison with state-of-the-art approaches on CAVE dataset.

| Method | RMSE |
|---|---|
| Ours | **3.47** |
| Galliani et al. [14] | 4.76 |
| Arad & Ben-Shahar [1] | 5.4 |
| A+ [16] | 6.70 |
| Can & Timofte [18] | 3.49 |
| Han et al. [19] | 4.78 |
| Gwn Lore et al. [21] | 8.06 |

*3.2. Remote Sensing Imagery*

3.2.1. Aerial Images

We will now use Aerial images (G-LiHT sensor) dataset to demonstrate the feature of the proposed model that allows constraining the center and the resolution (FWHM) of the optimized bands. Such constraints may arise in real sensor design scenario due to restriction in sensing hardware and technology. Table 3 shows a comparison of reconstruction accuracy of models trained with different constraints on MS bands. In the table, there are 7 different models. "Free" is a model trained to optimize 4 MS bands without any constraints. Similarly, "Free (8 bands)" is a model trained to optimize 8 MS bands without any constraints. Except for "Free (8 bands)", all models are trained to optimize 4 MS bands. In "Fixed centers", only FWHM is optimized and the centers of the 4 MS bands are fixed to 500 nm, 600 nm, 700 nm, and 800 nm respectively. Similarly, in "Fixed FWHMs", the FWHM of all four bands are set fixed to 100 nm and only the band centers are optimized. In "Bounded centers", the extent of 4 bands are bounded to 450 nm to 550 nm, 550 nm to 650 nm, 650 nm to 750 nm, and 750 nm to 850 nm, respectively. In "Bounded FWHMs", the FWHM of the bands are limited to a range 13.55 nm to 22.59 nm while the band center is freely optimized. Finally, in "Bounded centers & FWHMs" both the extent and the FWHM of each of the bands are bounded by limits during optimization. The extent and the FWHM of the first band were bounded to ranges 482 nm to 520 nm and 50 nm to 100 nm, respectively. The extent and the FWHM of the second band were bounded to ranges 580 nm to 620 nm and 50 nm to 100 nm, respectively. The extent and the FWHM of the third band were bounded to ranges 680 nm to 720 nm and 50 nm to 100 nm, respectively. Finally, the extent and the FWHM of the fourth band were bounded to ranges 780 nm to 824 nm and 50 nm to 150 nm, respectively.

These kind of constraints are important in real sensor design scenarios. For examples, if we know the location of a certain important spectral feature, we may want to set the band center of one of the MS band at that location and only optimize FWHM. If there is some uncertainty in the location of the spectral feature, we could bound the center of the MS bands to an interval around the location of the feature. Similarly, for many filters that are used to build MS sensor, there are limits to FWHM values, so we may need to bound the FWHM value of the optimized bands. For MS hardware that capture large spectral span, they might contain different kind of sensors. For example, Silicon sensor can be used to capture visible (400 nm to 1000 nm) radiations but different technology is needed to sense spectral range beyond 1000 nm. In this case, we do not want to have a MS band that spans over spectral range covered by two different sensor technologies. This can be handled by setting appropriate limits on the values of band centers and band FWHM in our model.

In Table 3, we find that increasing the number of MS bands from 4 to 8, significantly reduces the reconstruction error. This is expected as larger number of MS bands can capture more information about the reflectance spectra. Another interesting finding is that as constraints are added to the optimized bands, the reconstruction error increases, which is also expected. This happens because as constraints are added the set of all possible solutions decreases in size.

Figure 7 shows the optimized MS bands along with per band RMSE. It is seen that the error is high in areas where the responsivity of all MS bands are low and towards the ends of the spectral range of the HS spectra, as it was observed in previous experiment with Outdoor scenes dataset. Figure 8 shows reconstructed bands of a portion of test imagery obtained using the model trained on 4 bands without any restriction. We see smaller objects with high reconstruction error in the images. They could be material classes that were readily not present in the training set.

**Table 3.** Comparison of our model with different constraints on multispectral (MS) bands using Aerial images dataset.

| Test Dataset | Bands | RMSE | SAM | R |
|---|---|---|---|---|
| Cape Cod | Free | $0.0027 \pm 0.0022$ | $0.0252 \pm 0.0243$ | $0.9917 \pm 0.0433$ |
| | Free (8 bands) | $0.0016 \pm 0.0009$ | $0.0157 \pm 0.0174$ | $0.9973 \pm 0.0149$ |
| | Fixed centers | $0.0047 \pm 0.0029$ | $0.0382 \pm 0.0319$ | $0.9866 \pm 0.0664$ |
| | Fixed FWHMs | $0.0036 \pm 0.0021$ | $0.0309 \pm 0.0294$ | $0.9918 \pm 0.0396$ |
| | Bounded centers | $0.0025 \pm 0.0017$ | $0.0231 \pm 0.0147$ | $0.9977 \pm 0.0186$ |
| | Bounded FWHMs | $0.0029 \pm 0.0019$ | $0.0261 \pm 0.0229$ | $0.9937 \pm 0.0300$ |
| | Bounded centers & FWHMs | $0.0033 \pm 0.0019$ | $0.0303 \pm 0.0265$ | $0.9924 \pm 0.0346$ |
| White Mountains | Free | $0.0024 \pm 0.0011$ | $0.0296 \pm 0.0183$ | $0.9958 \pm 0.0428$ |
| | Free (8 bands) | $0.0018 \pm 0.0007$ | $0.0228 \pm 0.0176$ | $0.9963 \pm 0.0449$ |
| | Fixed centers | $0.0040 \pm 0.0023$ | $0.0471 \pm 0.0279$ | $0.9939 \pm 0.0481$ |
| | Fixed FWHMs | $0.0040 \pm 0.0019$ | $0.0479 \pm 0.0273$ | $0.9920 \pm 0.0705$ |
| | Bounded centers | $0.0033 \pm 0.0016$ | $0.0404 \pm 0.0229$ | $0.9940 \pm 0.0549$ |
| | Bounded FWHMs | $0.0029 \pm 0.0016$ | $0.0372 \pm 0.0244$ | $0.9948 \pm 0.0447$ |
| | Bounded centers & FWHMs | $0.0026 \pm 0.0015$ | $0.0314 \pm 0.0176$ | $0.9958 \pm 0.0433$ |
| Rochester | Free | $0.0040 \pm 0.0014$ | $0.0291 \pm 0.0157$ | $0.9927 \pm 0.0356$ |
| | Free (8 bands) | $0.0030 \pm 0.0009$ | $0.0216 \pm 0.0114$ | $0.9965 \pm 0.0173$ |
| | Fixed centers | $0.0054 \pm 0.0026$ | $0.0366 \pm 0.0214$ | $0.9895 \pm 0.0479$ |
| | Fixed FWHMs | $0.0055 \pm 0.0029$ | $0.0355 \pm 0.0186$ | $0.9920 \pm 0.0357$ |
| | Bounded centers | $0.0048 \pm 0.0019$ | $0.0333 \pm 0.0164$ | $0.9913 \pm 0.0390$ |
| | Bounded FWHMs | $0.0042 \pm 0.0016$ | $0.0300 \pm 0.0172$ | $0.9930 \pm 0.0318$ |
| | Bounded centers & FWHMs | $0.0046 \pm 0.0022$ | $0.0322 \pm 0.0175$ | $0.9927 \pm 0.0323$ |



(**a**) Free



(**b**) Free (8 bands)



(**c**) Fixed centers



(**d**) Fixed FWHMs



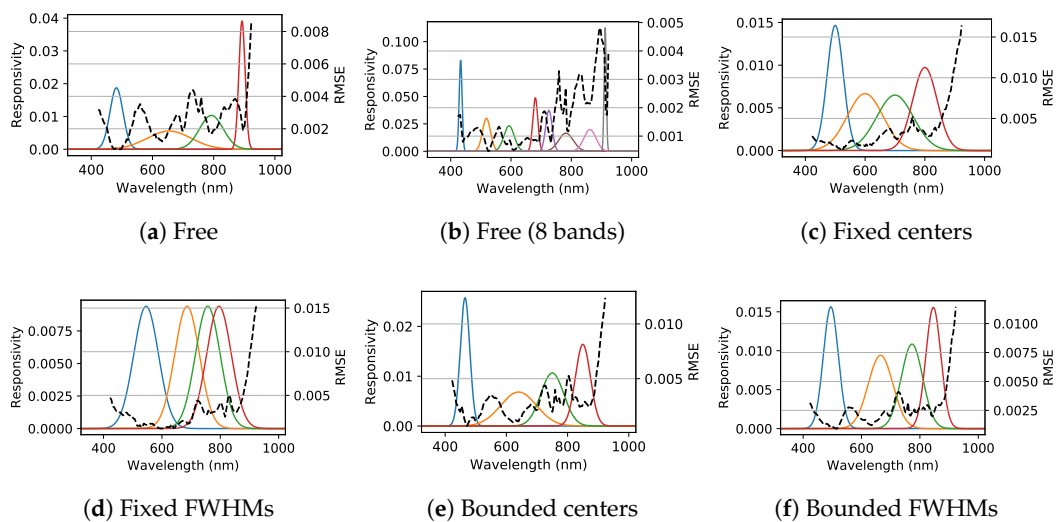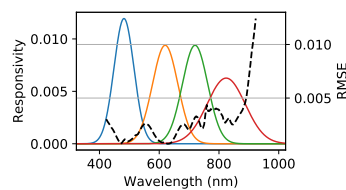(**e**) Bounded centers



(**f**) Bounded FWHMs

**Figure 7.** *Cont.*

(**g**) Bounded centers &
FWHMs

**Figure 7.** Responsivity of multispectral bands obtained using Aerial images dataset and RMSE across reconstructed bands (dotted line).
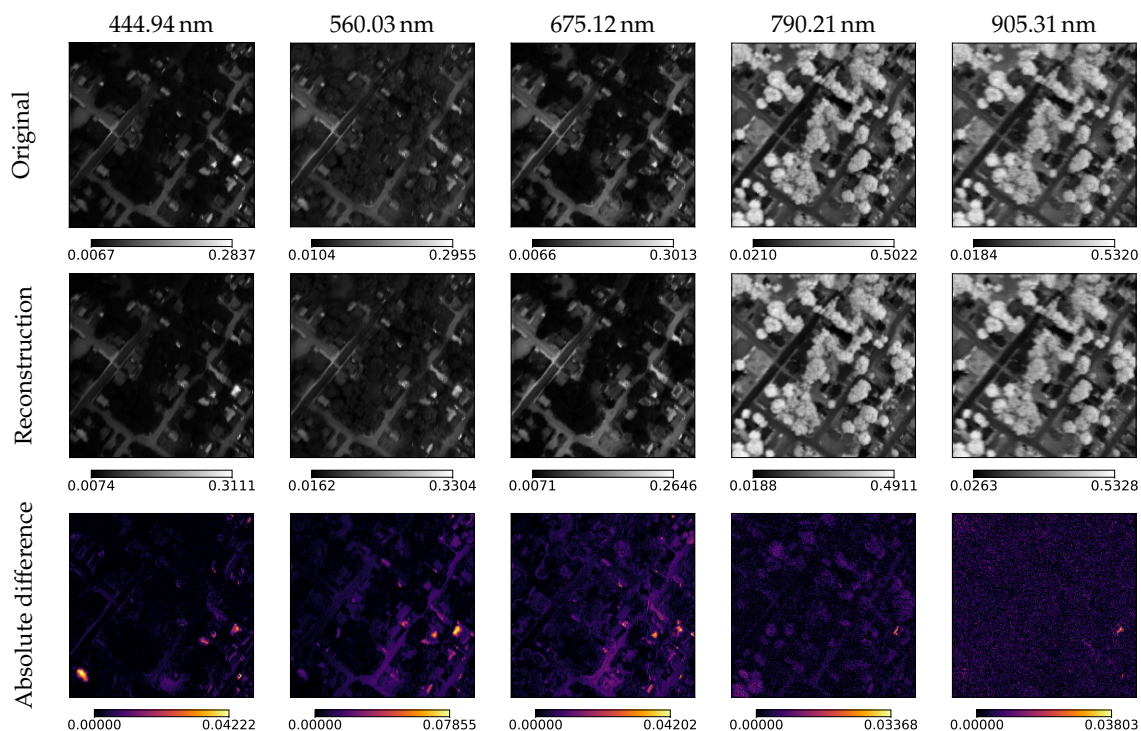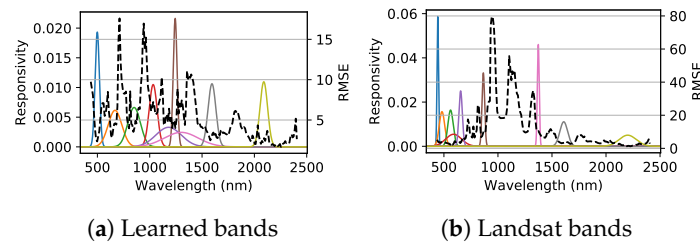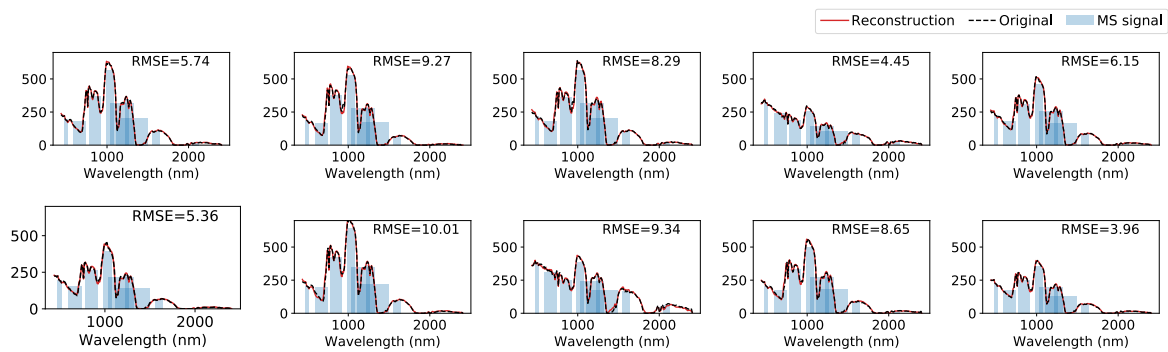


**Figure 8.** Examples of reconstructed bands for a portion of test imagery predicted using model with 4 free bands trained on Aerial images dataset.

### 3.2.2. Satellite Images

In this subsection, we compare the reconstruction performance of learned bands and Landsat-8 MS sensor bands on satellite images (Hyperion sensor) dataset. We extract 9 MS bands without any constraints. The reconstruction accuracy of a model using those 9 bands is compared with the one of a model that uses fixed Landsat-8 MS bands (Gaussian bands with Landsat-8 centers and FWHMs) in Table 4. The metrics are the same as the last (aerial images) experiments. Figure 9 shows the extracted bands with per band RMSE computed over the entire testing set and Figure 10 shows few sample reconstructions. We again find that learned MS bands are more accurate that fixed MS bands for reconstruction. The per band error was high in areas where all of the MS bands had low responsivity as observed before. It is interesting to see that the model did not place any band in the spectral range 1800 nm to 2000 nm because this region contains atmospheric water absorption bands and the HS signal in that region highly attenuated [49].

**Table 4.** Reconstruction accuracy for satellite images. RMSE is measured in $\mathrm{W\,m^{-2}\,sr^{-1}\,nm^{-1}}$.

| Test Image | Bands | RMSE | SAM | R |
|---|---|---|---|---|
| Atlanta | Learned | **6.751** ± 2.081 | **0.033** ± 0.007 | **0.999** ± 0.001 |
| | Landsat | 26.460 ± 10.869 | 0.119 ± 0.042 | 0.986 ± 0.009 |
| DC | Learned | **4.978** ± 3.039 | **0.047** ± 0.016 | **0.998** ± 0.004 |
| | Landsat | 11.046 ± 4.952 | 0.096 ± 0.024 | 0.990 ± 0.007 |
| Rochester | Learned | **4.797** ± 1.701 | **0.078** ± 0.038 | **0.995** ± 0.007 |
| | Landsat | 7.369 ± 5.359 | 0.094 ± 0.034 | 0.994 ± 0.008 |



(**a**) Learned bands          (**b**) Landsat bands

**Figure 9.** Responsivity of multispectral bands obtained using the satellite images dataset and RMSE across reconstructed bands (dotted line).



**Figure 10.** Examples of Hyperion spectra (satellite images dataset) reconstructions using learned bands (randomly selected from testing set). RMSE between the original and the reconstructed spectrum is measured in $\mathrm{W\,m^{-2}\,sr^{-1}\,nm^{-1}}$.

### 3.2.3. Land Cover Classification

In previous experiments, we evaluated the reconstruction accuracy of the optimized MS bands, now we will examine the performance of the optimized bands for the task of land cover classification. We optimize separate sets of 5 MS bands and 10 MS bands using the proposed network and the land cover classification dataset. The optimized bands are shown in Figure 11. It is interesting to see that the model has placed most of the MS bands in visible and near infrared region (esp. in case of 10 bands) because for this dataset those regions are the important ones as majority of the spectra are vegetation spectra. Table 5 shows comparison of classification performance of various models measured in terms of overall accuracy (OA), average accuracy (AA), and kappa coefficient for different training sizes. We randomly sampled a fixed number of pixels belonging to each land cover class to train classification model and evaluated the model on the remaining pixels. This process was repeated 30 times to obtain the mean and the standard deviation of the metrics which have been listed in the table. We observe that when training set is large, models trained on real (original) data and tested on real data performs the best. This is expected because MS bands and reconstructions from MS bands are likely to have less information contents than the original signal. However, surprisingly, when the training set was very small (equal to 5 per class), models tested on reconstruction from 10 MS bands performed the best. This

could be because the reconstruction is less noisy than the original signal because many details in the original signal that are not relevant to overall shape of the spectra are removed during reconstruction. This is similar to how dimensionality reduction improves classification performance when the dataset is small [50]. As expected, we also observe that models trained and tested on reconstructions from 10 bands performed better than the ones trained and tested on reconstructions from 5 bands.
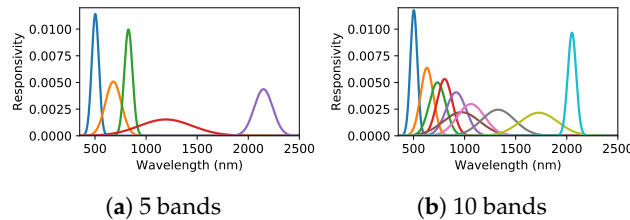


(**a**) 5 bands        (**b**) 10 bands

**Figure 11.** Bands extracted from Land cover classification dataset.

The importance of spectral super-resolution sub-network is also seen in the results. Models trained and tested on reconstructed HS signal performed better than the ones that utilized MS signal for training and testing. This means if we are to use our approach to optimize bands for a MS sensor and design a MS sensor in real-world, it is better to convert the MS signal observed by the new sensor to HS signal using the super-resolution sub-network before analysis. This is probably due to the fact that reconstructed HS signal, which utilizes statistics of real HS signal, in a way, acts like a feature extractor for MS signal. Finally, we observe interesting results regarding models that were trained on real HS data but tested on reconstructed HS data. We find that such models can provide sufficiently accurate classification performance. This is important because this means that if we build a real MS sensor using our method, the data obtained from the new sensor could be analyzed by pre-existing methods/models which were developed for real HS spectra by projecting the captured data to HS signal space using the spectral super-resolution sub-network. This could save time and costs associated with developing new models for analysing the data obtained from the new sensor.

**Table 5.** Comparison of classification performance on real and reconstructed hyperspectral images.

| Image | | Number of Training Pixels per Class | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 5 | | | 25 | | | 50 | | |
| Train | Test | OA (%) | AA (%) | kappa (%) | OA (%) | AA (%) | kappa (%) | OA (%) | AA (%) | kappa (%) |
| Real | Real | $51.4 \pm 6.8$ | $59.6 \pm 5.9$ | $45.5 \pm 7.3$ | **73.6** $\pm$ **1.6** | **81.6** $\pm$ **1.3** | **70.2** $\pm$ **1.7** | **78.8** $\pm$ **1.3** | **85.9** $\pm$ **0.6** | **75.9** $\pm$ **1.4** |
| Real | HSI-5 | $47.6 \pm 4.2$ | $55.3 \pm 3.0$ | $41.6 \pm 4.1$ | $53.8 \pm 4.0$ | $58.8 \pm 3.6$ | $48.2 \pm 4.2$ | $53.3 \pm 4.8$ | $57.6 \pm 3.9$ | $47.5 \pm 4.9$ |
| HSI-5 | HSI-5 | $48.8 \pm 5.8$ | $57.2 \pm 4.6$ | $43.0 \pm 6.1$ | $66.7 \pm 2.6$ | $73.9 \pm 1.8$ | $62.5 \pm 2.8$ | $69.5 \pm 3.3$ | $76.7 \pm 2.1$ | $65.5 \pm 3.6$ |
| MSI-5 | MSI-5 | $46.8 \pm 4.4$ | $54.2 \pm 4.9$ | $40.5 \pm 4.9$ | $65.2 \pm 2.5$ | $72.3 \pm 2.0$ | $60.8 \pm 2.7$ | $65.5 \pm 3.3$ | $73.8 \pm 1.8$ | $61.0 \pm 3.5$ |
| Real | HSI-10 | **52.6** $\pm$ **5.3** | **60.7** $\pm$ **4.7** | **47.1** $\pm$ **5.7** | $68.1 \pm 2.4$ | $73.7 \pm 2.5$ | $63.9 \pm 2.6$ | $70.0 \pm 2.3$ | $74.7 \pm 2.7$ | $66.0 \pm 2.5$ |
| HSI-10 | HSI-10 | **54.6** $\pm$ **5.0** | **62.5** $\pm$ **3.7** | **49.2** $\pm$ **5.3** | $72.3 \pm 2.7$ | $79.4 \pm 1.7$ | $68.7 \pm 2.9$ | $74.5 \pm 3.1$ | $81.6 \pm 2.1$ | $71.1 \pm 3.4$ |
| MSI-10 | MSI-10 | $48.0 \pm 5.6$ | $55.8 \pm 4.8$ | $41.8 \pm 6.0$ | $68.2 \pm 2.3$ | $75.8 \pm 1.7$ | $64.1 \pm 2.5$ | $69.9 \pm 4.5$ | $77.7 \pm 3.7$ | $65.9 \pm 4.9$ |

MSI-5 and MSI-10 are derived 5-band and 10-band multispectral images. HSI-5 and HSI-10 are synthetic hyperspectral images recontructed from MSI-5 and MSI-10 with RMSE of $0.0685 \pm 0.018$ and $0.0589 \pm 0.014$ $W \, cm^{-2} \, sr^{-1} \, nm^{-1}$ respectively.

Figure 12 plots average accuracy (the mean and the standard deviation of 30 random trial) as a function of training set size, as the training set size is increased from 15 per class to 150 per class. The experimental procedure used to obtain the result is same as the one used in Table 5, except that classes which covered more than 200 labeled pixels in the labeled image were only considered to accommodate larger training set sizes. We find that the performance of all methods initially increases as the training set size is increased but saturates after reaching a certain training set size. The performance of the methods that were trained on real data and tested on reconstructions saturated earlier at around a training set size of 25, while the same for methods that were trained and tested on either real data or reconstructions started saturating at around a training set size of 100. In general, models that were trained and tested on same data source (real or reconstructions) had much better performance than

models trained on real data and tested on reconstructions when the training set was large. This could have happened because the classifier trained on real spectra (which has never seen reconstructed spectra during training) cannot find subtle features in reconstructions which it deemed as important while training looking at the real data. So adding more training data does not help much because the subtle features are not present in the reconstructed data that is used for testing.
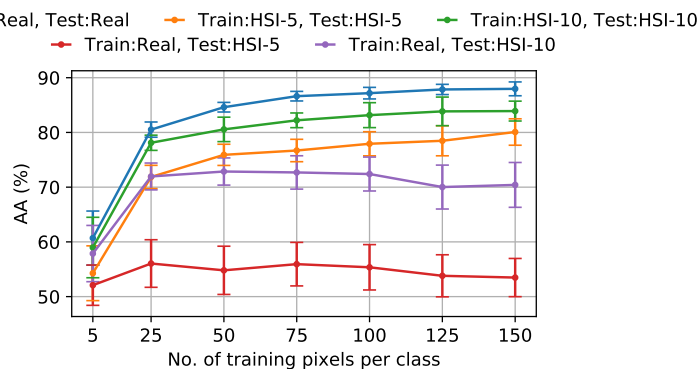


**Figure 12.** Average accuracy as a function of training set size for different models trained on land cover classification dataset.

Figure 13 shows normalized confusion matrices for models trained on real HS data and tested on real data and reconstructions when training set was set to 25 pixels per class. It is observed that when the classification model is tested on real data, reconstructions from 10 MS bands and reconstructions from 5 MS bands get more confused about similar classes (such as different types of soybeans, corn and grasses) in increasing order. This is because less fine details about the reflectance spectra are present in reconstructions from 10 MS bands than in real and in reconstructions from 5 MS bands than in reconstructions from 10 MS bands. So depending on the application, the number of optimized MS band should be adjusted to obtain desired level of performance.
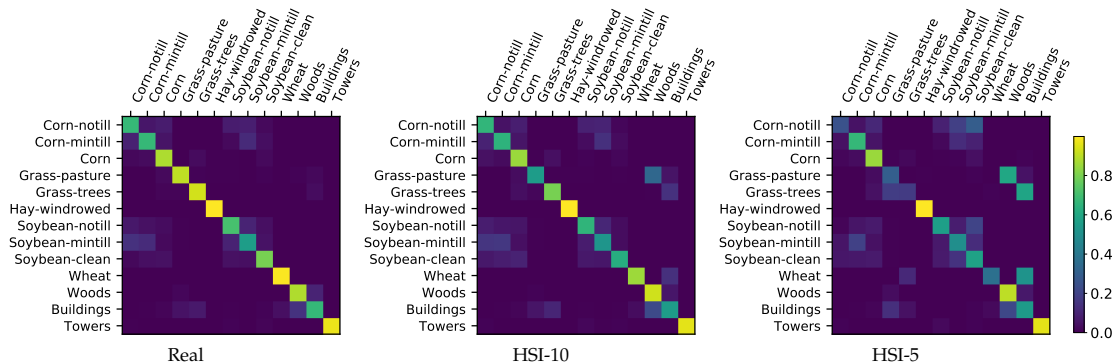


**Figure 13.** Normalized confusion matrices (25 training samples per class) averaged over 30 trials for classification model trained on real data and tested on real, HSI-10 (reconstruction from 10 MS bands), and HSI-5 (reconstruction from 5 MS bands).

### 3.2.4. Transfer of Learned Bands

Up to this point, we trained and evaluated our method on data from the same sensor, albeit different geographical locations. In this section, we will evaluate the reconstruction accuracy of the learned bands and the spectral super-resolution sub-network on data from a separate sensor. Figure 14 shows a block diagram explaining the design of the experiment.
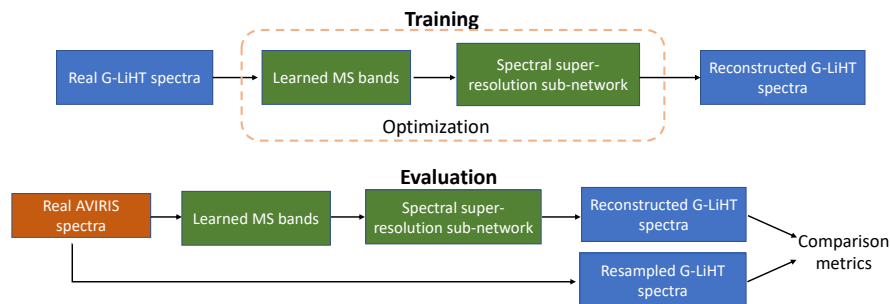
**Figure 14.** Evaluation of MS bands learned using G-LiHT data on AVIRIS imagery.

We utilize two models trained on G-LiHT spectra (Aerial images dataset) in Section 3.2.1 for the experiments. The two models used are the networks to extract 4 MS bands and 8 MS bands (without bounds on the band centers and the FWHMs). We evaluate these models on data obtained from AVIRIS sensor. For evaluation, we first convert AVIRIS spectra to MS signal using the learned bands' responsivity functions. Then, this MS signal is used to reconstruct HS spectra using the spectral super-resolution sub-network. The reconstructed HS spectra has the same spectral range and band spacing as G-LiHT spectra. It cannot be directly compared with the original AVIRIS spectra.

The AVIRIS sensor measures a spectral range of 400 nm to 2500 nm at approximately 10 nm increments, while the G-LiHT sensor measures a spectral range of 400 nm to 1000 nm at approximately 4.43 nm increments. Hence, the visible and near infrared portion of the original AVIRIS spectra was resampled to match the bands of the G-LiHT spectra. Then, performance metrics were computed to compare the predicted spectra and the resampled version of the original spectra. The results are shown in Table 6. For comparison, we also trained the conventional method (PCA + simulated annealing) on G-LiHT spectra and tested on resampled AVIRIS data. Since, the Aerial images dataset is very huge, we used Incremental PCA [51] to find the principal components needed for the conventional approach.

**Table 6.** Performance of MS bands learned using G-LiHT data on AVIRIS imagery.

| Methods | RMSE | SAM | R |
|---|---|---|---|
| Ours (4 bands) | $0.0148 \pm 0.01$ | $0.0729 \pm 0.0363$ | $0.9520 \pm 0.1492$ |
| Ours (8 bands) | $\mathbf{0.0046} \pm 0.0028$ | $\mathbf{0.0227} \pm 0.0087$ | $\mathbf{0.9843} \pm 0.0589$ |
| PCA with simulated annealing [9,25] (4 bands) | $0.4436 \pm 0.02$ | $0.0772 \pm 0.0116$ | $0.95081 \pm 0.1324$ |
| PCA with simulated annealing [9,25] (8 bands) | $0.46055 \pm 0.01808$ | $0.10655 \pm 0.00381$ | $0.95729 \pm 0.09921$ |

The results show that our method performed much better than the traditional approach. In fact, the model learned using the traditional approach showed very high error when tested on the AVIRIS spectra, even though it had better reconstruction error for the G-LiHT test set (e.g., $0.006 \pm 0.0003$ RMSE for Rochester subset using 4 bands). We hypothesize that the traditional approach is more vulnerable to noise and variability, which led to poor performance in this experiment.

The reconstruction performance of our method is also weaker compared to the results in Table 3. This is expected as the training data and the testing data come from two different HS sensors with different parameters, such as ground sampling distance, flying altitude, and signal-to-noise ratio. However, the error of the reconstruction from 8 bands is significantly smaller than the reconstruction from 4 bands and comparable to reconstruction error for Rochester subset (4 bands) in Table 3. This indicates more bands are better when there is large variability in the properties and distribution of the training set and the testing set. Figure 15 shows the sample reconstructions.
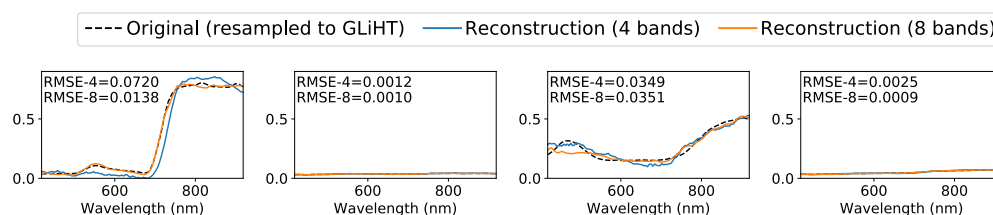
**Figure 15.** Reconstructions of AVIRIS spectra using G-LiHT trained models. (Left to right) the worst reconstruction with 4 bands, the best reconstruction with 4 bands, the worst reconstruction with 8 bands, and the best reconstruction with 8 bands. RMSE-4 and RMSE-8 are RMSE between original spectra and reconstructions from 4 bands and 8 bands respectively.

## 4. Conclusions

We showed the efficacy of the proposed method on six real world HS dataset. The results of the experiments showed that optimizing MS bands along with the transformation from MS signal to HS spectrum can greatly improve the reconstruction accuracy compared to only optimizing the transformation for fixed MS bands. The proposed method allows for constraining the location and the resolution of the optimized bands, which could be very helpful to handle design constraints of sensor hardware. We also showed that the reconstructed HS spectra can be used for land cover classification.

In future, we want to investigate if more accurate reconstructions can be obtained using generative adversarial networks (GANs) with the super-resolution sub-network. Currently, we are using the mean squared error between the first and the second derivatives of the original signal and the reconstruction to capture fine details in HS spectra. However, leveraging GAN as prior for HS reconstruction could provide better reconstruction of fine details. Another promising avenue for improving the reconstruction accuracy is to utilize spatial contexts in the super-resolution sub-network by using a two dimensional convolutional neural network architecture, which is trained end-to-end to predict HS image from MS image. It would also be interesting to evaluate the performance of reconstructed HS image for other image analysis applications, such as, physical/chemical parameter estimation, target detection and unmixing. We would also like to investigate tunable sub-sampling layer with non-Gaussian bands. One of the limitations of this study is that we have only validated the proposed method in software. Future goal would be to implement the MS bands acquired from our method in hardware and evaluate the reconstruction performance of a framework containing the hardware implementation of the MS bands followed by a data processing pipeline containing the spectral super-resolution sub-network.

**Author Contributions:** U.B.G. wrote the manuscript. U.B.G., S.T.M., and E.S. conceptualized and contributed to the algorithmic methods.

## References

1. Arad, B.; Ben-Shahar, O. Sparse recovery of hyperspectral signal from natural RGB images. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin, Germany, 2016; pp. 19–34.

2. Arad, B.; Ben-Shahar, O.; Timofte, R.; Van Gool, L.; Zhang, L.; Yang, M.H. NTIRE 2018 challenge on spectral reconstruction from RGB images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18 June 2018; pp. 1042–104209.

3. Schott, J.R. *Remote Sensing: The Image Chain Approach*; Oxford University Press on Demand: Oxford, UK, 2007.

4.   Bioucas-Dias, J.M.; Nascimento, J.M. Hyperspectral subspace identification. *IEEE Trans. Geosci. Remote. Sens.* **2008**, *46*, 2435–2445. [CrossRef]

5.   Loncan, L.; De Almeida, L.B.; Bioucas-Dias, J.M.; Briottet, X.; Chanussot, J.; Dobigeon, N.; Fabre, S.; Liao, W.; Licciardi, G.A.; Simoes, M.; et al. Hyperspectral pansharpening: A review. *IEEE Geosci. Remote. Sens. Mag.* **2015**, *3*, 27–46. [CrossRef]

6.   Yuan, Y.; Zheng, X.; Lu, X. Hyperspectral image superresolution by transfer learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1963–1974. [CrossRef]

7.   Li, Y.; Hu, J.; Zhao, X.; Xie, W.; Li, J. Hyperspectral image super-resolution using deep convolutional neural network. *Neurocomputing* **2017**, *266*, 29–41. [CrossRef]

8.   Lin, G.; Wu, Q.; Qiu, L.; Huang, X. Image super-resolution using a dilated convolutional neural network. *Neurocomputing* **2018**, *275*, 1219–1230. [CrossRef]

9.   Fairman, H.S.; Brill, M.H. The principal components of reflectances. *Color Res. Appl.* **2004**, *29*, 104–110. [CrossRef]

10.  Agahian, F.; Amirshahi, S.A.; Amirshahi, S.H. Reconstruction of reflectance spectra using weighted principal component analysis. *Color Res. Appl.* **2008**, *33*, 360–371. [CrossRef]

11.  Heikkinen, V.; Jetsu, T.; Parkkinen, J.; Hauta-Kasari, M.; Jaaskelainen, T.; Lee, S.D. Regularized learning framework in the estimation of reflectance spectra from camera responses. *JOSA A* **2007**, *24*, 2673–2683. [CrossRef]

12.  Nguyen, R.M.; Prasad, D.K.; Brown, M.S. Training-based spectral reconstruction from a single RGB image. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin, Germany, 2014; pp. 186–201.

13.  Heikkinen, V. Spectral Reflectance Estimation Using Gaussian Processes and Combination Kernels. *IEEE Trans. Image Process.* **2018**, *27*, 3358–3373, doi:10.1109/TIP.2018.2820839. [CrossRef]

14.  Galliani, S.; Lanaras, C.; Marmanis, D.; Baltsavias, E.; Schindler, K. Learned spectral super-resolution. *arXiv* **2017**, arXiv:1703.09470.

15.  Jégou, S.; Drozdzal, M.; Vazquez, D.; Romero, A.; Bengio, Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.

16.  Aeschbacher, J.; Wu, J.; Timofte, R. In defense of shallow learned spectral reconstruction from RGB images. In Proceedings of the IEEE International Conference on Computer Vision Honolulu, HI, USA, 21–26 July 2017; pp. 471–479.

17.  Timofte, R.; De Smet, V.; Van Gool, L. A+: Adjusted anchored neighborhood regression for fast super-resolution. In Proceedings of the Asian Conference on Computer Vision, Singapore, 1–5 Novembe 2014; Springer: Berlin, Germany, 2014; pp. 111–126.

18.  Can, Y.B.; Timofte, R. An efficient CNN for spectral reconstruction from RGB images. *arXiv* **2018**, arXiv:1804.04647.

19.  Han, X.H.; Shi, B.; Zheng, Y. Residual HSRCNN: Residual Hyper-Spectral Reconstruction CNN from an RGB Image. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 2664–2669.

20.  He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June–1 July 2016; pp. 770–778.

21.  Gwn Lore, K.; Reddy, K.K.; Giering, M.; Bernal, E.A. Generative Adversarial Networks for Spectral Super-Resolution and Bidirectional RGB-To-Multispectral Mapping. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 15–21 June 2019.

22.  Sun, T.; Kelly, K. Compressive sensing hyperspectral imager. In *Computational Optical Sensing and Imaging*; Optical Society of America: Washington, DC, USA, 2009; p. CTuA5.

23.  Wang, L.; Feng, Y.; Gao, Y.; Wang, Z.; He, M. Compressed sensing reconstruction of hyperspectral images based on spectral unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2018**, *11*, 1266–1284. [CrossRef]

24.  Martín, G.; Bioucas-Dias, J.M.; Plaza, A. HYCA: A new technique for hyperspectral compressive sensing. *IEEE Trans. Geosci. Remote. Sens.* **2015**, *53*, 2819–2831. [CrossRef]

25.  López-Álvarez, M.A.; Hernández-Andrés, J.; Valero, E.M.; Romero, J. Selecting algorithms, sensors, and linear bases for optimum spectral recovery of skylight. *JOSA A* **2007**, *24*, 942–956. [CrossRef] [PubMed]

26. Shen, H.L.; Yao, J.F.; Li, C.; Du, X.; Shao, S.J.; Xin, J.H. Channel selection for multispectral color imaging using binary differential evolution. *Appl. Opt.* **2014**, *53*, 634–642. [CrossRef] [PubMed]

27. Arad, B.; Ben-Shahar, O. Filter selection for hyperspectral estimation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3153–3161.

28. Li, Y.; Majumder, A.; Zhang, H.; Gopi, M. Optimized Multi-Spectral Filter Array Based Imaging of Natural Scenes. *Sensors* **2018**, *18*, 1172. [CrossRef] [PubMed]

29. Fu, Y.; Zhang, T.; Zheng, Y.; Zhang, D.; Huang, H. Joint Camera Spectral Sensitivity Selection and Hyperspectral Image Recovery. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 788–804.

30. Nie, S.; Gu, L.; Zheng, Y.; Lam, A.; Ono, N.; Sato, I. Deeply Learned Filter Response Functions for Hyperspectral Reconstruction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4767–4776.

31. Gewali, U.B.; Monteiro, S.T.; Saber, E. Machine learning based hyperspectral image analysis: A survey. *arXiv* **2018**, arXiv:1802.08701.

32. Lanaras, C.; Baltsavias, E.; Schindler, K. Hyperspectral super-resolution by coupled spectral unmixing. In Proceedings of the IEEE International Conference on Computer Vision, Araucano Park, Las Condes, Chile, 11–18 December 2015; pp. 3586–3594.

33. Chang, C.I.; Du, Q. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Trans. Geosci. Remote. Sens.* **2004**, *42*, 608–619. [CrossRef]

34. Van Der Meer, F. Analysis of spectral absorption features in hyperspectral imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2004**, *5*, 55–68. [CrossRef]

35. Tsai, F.; Philpot, W. Derivative analysis of hyperspectral data. *Remote. Sens. Environ.* **1998**, *66*, 41–51. [CrossRef]

36. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* **2016**, *104*, 148–175. [CrossRef]

37. Frazier, P.I. A tutorial on Bayesian optimization. *arXiv* **2018**, arXiv:1807.02811.

38. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian optimization of machine learning algorithms. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Lake Tahoe, NT, USA, 3–6 December 2012; pp. 2951–2959.

39. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

40. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.

41. Yasuma, F.; Mitsunaga, T.; Iso, D.; Nayar, S.K. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. *IEEE Trans. Image Process.* **2008**, *9*, 2241–2253. [CrossRef] [PubMed]

42. Cook, B.D.; Corp, L.A.; Nelson, R.F.; Middleton, E.M.; Morton, D.C.; McCorkel, J.T.; Masek, J.G.; Ranson, K.J.; Ly, V.; Montesano, P.M. NASA Goddard's LiDAR, hyperspectral and thermal (G-LiHT) airborne imager. *Remote. Sens.* **2013**, *5*, 4045–4066. [CrossRef]

43. Pearlman, J.; Carman, S.; Segal, C.; Jarecke, P.; Clancy, P.; Browne, W. Overview of the Hyperion imaging spectrometer for the NASA EO-1 mission. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Sydney, Australia, 9–13 July 2001; Volume 7, pp. 3036–3038.

44. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote. Sens. Environ.* **2014**, *145*, 154–172. [CrossRef]

45. Vane, G.; Green, R.O.; Chrien, T.G.; Enmark, H.T.; Hansen, E.G.; Porter, W.M. The airborne visible/infrared imaging spectrometer (AVIRIS). *Remote. Sens. Environ.* **1993**, *44*, 127–143. [CrossRef]

46. Baumgardner, M.F.; Biehl, L.L.; Landgrebe, D.A. *220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3*; Purdue University: West Lafayette, IN, USA, 2015; doi:10.4231/R7RX991C.

47. Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Data Products Portal. Available online: https://aviris.jpl.nasa.gov/alt_locator/ (accessed on 8 April 2019).

48. Aarts, E.; Korst, J. *Simulated Annealing and Boltzmann Machines*; John Wiley and Sons Inc.: New York, NY, USA, 1988.

49. Vermote, E.F.; Tanré, D.; Deuze, J.L.; Herman, M.; Morcette, J.J. Second simulation of the satellite signal in the solar spectrum, 6S: An overview. *IEEE Trans. Geosci. Remote. Sens.* **1997**, *35*, 675–686. [CrossRef]

50. Su, J.; Yi, D.; Liu, C.; Guo, L.; Chen, W.H. Dimension reduction aided hyperspectral image classification with a small-sized training dataset: Experimental comparisons. *Sensors* **2017**, *17*, 2726. [CrossRef]

51. Zhao, H.; Yuen, P.C.; Kwok, J.T. A novel incremental principal component analysis and its application for face recognition. *IEEE Trans. Syst. Man, Cybern. Part B (Cybern.)* **2006**, *36*, 873–886. [CrossRef]