

Article

A Multi-Scale Wavelet 3D-CNN for Hyperspectral Image Super-Resolution

Jingxiang Yang ^{1,2}, Yong-Qiang Zhao ^{2,*}, Jonathan Cheung-Wai Chan ³ and Liang Xiao ¹

¹ School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

² Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, China

³ Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussel 1050, Belgium

* Correspondence: zhaoyq@nwpu.edu.cn

Received: 21 May 2019; Accepted: 28 June 2019; Published: 30 June 2019



Abstract: Super-resolution (SR) is significant for hyperspectral image (HSI) applications. In single-frame HSI SR, how to reconstruct detailed image structures in high resolution (HR) HSI is challenging since there is no auxiliary image (e.g., HR multispectral image) providing structural information. Wavelet could capture image structures in different orientations, and emphasis on predicting high-frequency wavelet sub-bands is helpful for recovering the detailed structures in HSI SR. In this study, we propose a multi-scale wavelet 3D convolutional neural network (MW-3D-CNN) for HSI SR, which predicts the wavelet coefficients of HR HSI rather than directly reconstructing the HR HSI. To exploit the correlation in the spectral and spatial domains, the MW-3D-CNN is built with 3D convolutional layers. An embedding subnet and a predicting subnet constitute the MW-3D-CNN, the embedding subnet extracts deep spatial-spectral features from the low resolution (LR) HSI and represents the LR HSI as a set of feature cubes. The feature cubes are then fed to the predicting subnet. There are multiple output branches in the predicting subnet, each of which corresponds to one wavelet sub-band and predicts the wavelet coefficients of HR HSI. The HR HSI can be obtained by applying inverse wavelet transform to the predicted wavelet coefficients. In the training stage, we propose to train the MW-3D-CNN with L1 norm loss, which is more suitable than the conventional L2 norm loss for penalizing the errors in different wavelet sub-bands. Experiments on both simulated and real spaceborne HSI demonstrate that the proposed algorithm is competitive with other state-of-the-art HSI SR methods.

Keywords: super-resolution; hyperspectral; wavelet; multi-scale; CNN

1. Introduction

Hyperspectral image (HSI) is collected in contiguous bands over a certain electromagnetic spectrum, and the spectral and spatial information in HSI is helpful for identifying and discriminating different materials in the scene. HSI has been applied to many fields, including target detection [1], environment monitoring [2], and land-cover classification [3]. However, the spatial resolution of HSI is often limited due to the trade-off between the spatial and spectral resolutions. Some Earth Observation applications, such as urban mapping [4] and fine mineral mapping [5], require high resolution (HR) HSI. Therefore, enhancing the spatial resolution of HSI is of significance for the application of HSI.

There are several ways to enhance the spatial resolution of HSI. Some auxiliary images, e.g., panchromatic image and multispectral image (MSI), often have higher spatial resolution [6]. Hyperspectral pan-sharpening reconstructs HR HSI by fusing the low resolution (LR) HSI with a HR panchromatic image taken over the same area at the same time (or a similar period). Pan-sharpening could be implemented by

different methods, such as component substitution methods [7], multi-resolution analysis methods [8], and variational methods [9]. As an effective model in extracting features and representing mapping function, deep learning, particularly convolutional neural network (CNN), has attracted more and more interest in pan-sharpening [10,11]. In [12], a pan-sharpening CNN network (PNN) was proposed to learn the mapping between the LR MSI, HR panchromatic image and HR MSI. Combined with residual learning, the performance of PNN could be further boosted [13]. In order to preserve detailed structures, the pan-sharpening network could also be learned in the high-pass filtering domain rather than image domain [14]. In [15], Yuan et al. proposed a multi-scale and multi-depth CNN (MSDCNN) for pan-sharpening, in which multi-scale features in HSI can be exploited.

HSI-MSI fusion, which fuses the LR HSI with a HR MSI taken over the same area, is another option to enhance HSI resolution [16]. HSI-MSI fusion methods can be classified into four categories: unmixing based method, dictionary learning based method, variational method, and deep learning method. HR HSI can be reconstructed with the endmembers of LR HSI and the abundance of MSI. Utilizing this idea, several unmixing based fusion methods have been proposed. For example, in [17], the LR HSI and the MSI were alternatively unmixed via nonnegative matrix factorization, the HR HSI was reconstructed with the endmembers of LR HSI and the abundance of MSI. HR HSI can also be reconstructed using a dictionary. In [18], a spatial dictionary was learned from HR MSI, and the HR HSI was then reconstructed via joint sparse coding. In [19], a spectral dictionary was learned from LR HSI, then it was used to reconstruct the HR HSI based on the abundance of MSI. The HSI-MSI fusion problem could also be solved in a variational framework [20–22]. Sparsity [20], vector-total-variation [21], and low rank [22] could be utilized as regularizers for fusion in the variational framework. CNN has also shown its potential in HSI-MSI fusion. In [23], a CNN with two branch architecture was proposed for HSI-MSI fusion, deep features of LR HSI and HR MSI can be extracted and fused by the two CNN branches. In [24], the LR HSI and the MSI was fused in a deep learning model with low rank used as prior information.

Single-frame HSI super-resolution (SR) tries to reconstruct HR HSI using only one LR HSI [25]. Compared with pan-sharpening and HSI-MSI fusion, it does not require any auxiliary data and is more flexible. A basic single-frame HSI SR involves interpolating the LR HSI band-by-band (e.g., bicubic interpolation). Such methods are simple and fast, but the image details in HR HSI are prone to being blurred. In [26], a sparse representation based HSI SR method was proposed, and sparsity and non-local similarity regularizers were exploited. In [27], in order to exploit the self-similarity in spatial and spectral domains, a group sparse representation method was proposed for HSI SR. HSI can be represented as a tensor, and a tensor-based HSI SR method was proposed in [28] via non-local low-rank tensor approximation. In [29], a CNN was used to initially super-resolve the LR HSI band-by-band, then the HR HSI was refined via collaborative matrix factorization. The authors in [30] proposed a spectral difference convolutional network (SDCNN) to learn the mapping of spectral differences between the LR and HR HSIs, and the SDCNN could be further integrated with a spatial error correction model to rectify the artifacts of HR HSI [31]. 3D convolution could exploit the spectral-spatial correlation in HSI, a 3D CNN based HSI SR method was proposed in [32], where the mapping between the LR and HR HSIs was represented by 3D CNN.

Despite the above progress, the deep learning based single-frame HSI SR technology still faces the challenge in reconstructing detailed structures of HR HSI due to the fact that there is no HR auxiliary data providing structural information. In order to accurately reconstruct detailed structures in HSI without HR auxiliary data, the information in the spectral and spatial domains of LR HSI should be fully exploited for SR. Extracting deep spectral-spatial features from the HSI is an effective way to exploit the information in HSI. On the other hand, in wavelet domain, global topology and local textural information of different scales and orientations can be captured by different wavelet sub-bands. Training a deep learning network that predicts the wavelet coefficients, particularly the high-frequency wavelet sub-bands, would encourage the network to produce more structural details in the image SR problem [33–36].

In this study, we propose a single-frame HSI SR method based on multi-scale wavelet 3D CNN (MW-3D-CNN) that predicts the wavelet package coefficients of the latent HR HSI, rather than directly inferring the HR HSI. The network is built on 3D convolutional layers, which could extract hierarchical features from both spectral and spatial domains of HSI. Specifically, the MW-3D-CNN consists of two subnets, one is an embedding subnet and another one is a predicting subnet. The embedding subnet projects the LR HSI into a feature space and represents it with deep spectral-spatial feature cubes. The feature cubes are then fed to the predicting subnet. The predicting subnet is composed of multiple output branches, each of which corresponds to one wavelet sub-band and predicts the wavelet coefficients of the latent HR HSI. By applying inverse wavelet package transformation to the predicted wavelet coefficients, the HR HSI can be obtained. It should be noted that the wavelet coefficients have larger values in the low-frequency sub-band but smaller values in the high-frequency sub-bands. Previous L2 norm loss will over-penalize the larger errors in the low-frequency sub-band while neglect the smaller errors in the high-frequency sub-bands [37]. Therefore, we propose to train the MW-3D-CNN with L1 norm loss, which is more suitable to equally penalize the errors in both low- and high- frequency sub-bands. Furthermore, the L1 norm loss could lead to SR result with sharper and clearer structures [38,39].

We consider four characteristics of the proposed HSI SR method:

- Unlike the previous deep learning models that reconstruct HR HSI directly [29–32], the proposed network predicts the wavelet coefficients of the latent HR HSI, which is beneficial for reconstructing detailed textures in HSI.
- In the predicting subnet, different branches corresponding to different wavelet sub-bands are trained jointly in a unified network, and the inter sub-band correlation can be utilized.
- The network is built based on 3D convolutional layers, which could exploit the correlation in both spectral and spatial domains of HSI.
- Instead of the conventional L2 norm, we propose to train the network with the L1 norm loss, which is fit for both low- and high- frequency wavelet sub-bands.

The remainder of the paper is organized as follows. In Section 2, we introduce some related works. In Section 3, we present the proposed HSI SR method, including the architecture and training of the network. The experimental results are given in Section 4. In Section 5, we present some analyses and discussion on the experiment. In Section 6, we conclude with observations specific to the potential of our approach to single-frame HSI SR.

2. Related Works

2.1. CNN Based Single Image SR

CNN could extract features from the local neighborhood of image by convolving with trainable kernels, which makes it easy to exploit spatial correlation in an image. CNN has become the most popular deep learning model in many image processing tasks, particularly in image SR [40–46].

In [38], Dong et al. proposed to learn the mapping between the LR and HR images using a CNN. The HR image can be inferred from its LR version with the trained network. Inspired by this idea, several CNN based single image SR methods have been proposed [41–46]. In [41], a very deep CNN for SR was proposed and trained with a residual learning strategy. Trainable parameters would drastically increase in very deep CNN, and a recursive CNN was proposed to address this issue by sharing the parameters of different layers in [42]. Most CNN SR methods employed the high-level features for reconstruction and neglected the low- and mid- level features. In [43,44], the authors proposed a residual dense network for SR, in which layers were densely connected to make full use of the hierarchical features. To address the challenge of super-resolving an image by large factors, the authors in [45] proposed progressive deep learning models to upscale the image gradually. Similarly, a Laplacian Pyramid SR CNN (LapSRN) was proposed in [46], which could progressively reconstruct the high-frequency details of different sub-bands of the latent HR image.

2.2. Application of Wavelet in SR

Wavelet describes image structures in different orientations. Employing wavelet in image SR, particularly the high-frequency wavelet sub-bands, is beneficial for preserving the detailed image structures. Many wavelet based SR methods have been proposed [47–50]. In [47], the LR image was decomposed into different wavelet sub-bands, the high-frequency sub-bands were interpolated and then combined with the LR image to generate HR image via inverse wavelet transformation. Similarly, the LR image was decomposed by two types of wavelets, and the high-frequency sub-bands of the two wavelets were then combined and followed by inverse wavelet transformation [48]. In [49,50], edge prior was utilized in the high-frequency sub-bands estimation to make the SR result sharper. Wavelet could also be used in CNN to better infer image details and enhance the sparsity of the network. For example, in [34,35], the mapping between the LR and HR images was learned by a CNN in wavelet domain for single image SR. However, these SR methods were designed for a single image, therefore applying these methods to HSI in band-by-band fashion would neglect the spectral correlation in HSI and lead to high spectral distortion.

3. Multi-Scale Wavelet 3D CNN For HSI SR

In this study, we transform the HSI SR problem into predicting the wavelet coefficients of HSI. In this section, we first introduce some basics on wavelet package analysis and 3D CNN, then we propose the MW-3D-CNN for HSI SR, including the architecture and the loss function.

3.1. Wavelet Package Analysis

Wavelet package transformation (WPT) could transform an image into a serial of wavelet coefficients sub-bands with the same size. An example of WPT with Haar wavelet function is given in Figure 1. The one-level decomposition is shown in Figure 1b. It can be found that the low-frequency sub-band (i.e., the top-left patch) describes the global topology. The detailed structures in vertical, horizontal, and diagonal orientations can be captured by different high-frequency sub-bands (i.e., the rest patches). By repeating the decomposition to each sub-band recursively, we can obtain higher-level WPT results, such as the two-level decomposition in Figure 1c. It is noted that the decomposition is applied to both the low- and high-frequency sub-bands, so the sub-bands of higher-level decomposition are of the same size. The original image can be reconstructed from these sub-bands via inverse WPT.

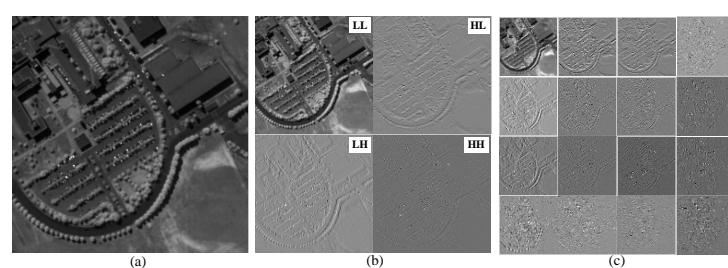


Figure 1. An example of wavelet package transformation (WPT) of different levels: (a) the original image, cropped from the 100th band of *Pavia University* data, (b) one-level decomposition, (c) two-level decomposition.

3.2. 3D CNN

For HSI, both spatial and spectral domains should be exploited in feature extraction. By convolving with 3D kernels, 3D CNN could extract features from different domains of volumetric data. The activity of the k -th feature cube in the d -th layer following formulation in [51] can be written as

$$F_{(x,y,z)}^{d,k} = g(b^{d,k} + \sum_c \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} \sum_{w=0}^{W-1} \omega_{(u,v,w)}^{d,k,c} F_{(x+u,y+v,z+w)}^{d-1,c}), \quad (1)$$

where, c means the set of feature cubes in the $(d-1)$ -th layer connected to the k -th feature cube in the d -th layer, $\omega_{(u,v,w)}^{d,k,c}$ is the value at position (u, v, w) of the 3D kernel associated with the k -th feature cube. The size of the 3D kernel is $U \times V \times W$. $F_{(x,y,z)}^{d,k}$ is the value at position (x, y, z) of the k -th feature cube in the d -th layer. $g(\cdot)$ is a non-linear activation function such as Rectified Linear Unit (ReLU) and Sigmoid function, etc. By convolving with different kernels, several 3D feature cubes can be extracted in each layer of 3D CNN, as shown in Figure 2b. Pixels of spatial neighborhood and adjacent bands are involved in 3D convolution, and the spectral-spatial correlation in HSI can be jointly exploited in feature extraction [52,53].

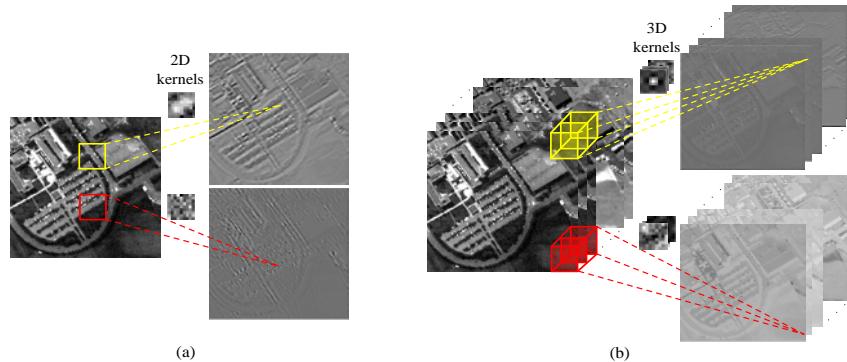


Figure 2. The illustration of (a) 2D and (b) 3D convolutional operations, feature maps and feature cubes are generated in each layer of 2D convolutional neural network (CNN) and 3D CNN respectively.

3.3. Network Architecture of MW-3D-CNN

The correlation exists not only in the spatial and spectral domains, but also among the wavelet package sub-bands of HSI. Considering the inter wavelet package sub-bands correlation, an embedding subnet is designed to learn shared features for different wavelet package sub-bands. These shared features are then fed to a predicting subnet to infer the wavelet package coefficients. Both of the embedding and predicting subnets are built based on 3D convolutional layers, which could naturally exploit the spectral-spatial correlation in HSI. The overall architecture of MW-3D-CNN is shown in Figure 3.

3.3.1. Embedding Subnet

The embedding subnet projects the LR HSI into deep feature space and represents it as a set of feature cubes that are shared by different wavelet package sub-bands. 3D convolutional layers and non-linear activation layers are alternatively stacked in the embedding subnet. The embedding subnet extracts feature cubes from the LR HSI $\mathbf{X} \in \mathbb{R}^{m \times n \times L}$, where m , n , and L are the number of rows, columns, and spectral bands, respectively. Both spectral and spatial information of HSI can be encoded by 3D convolution during the feature extraction, after several 3D convolutional layers, the LR HSI \mathbf{X} could be represented by a serial of spectral-spatial feature cubes, which are expressed as $\psi(\mathbf{X}) \in \mathbb{R}^{m \times n \times L \times S}$, where S is the number of feature cubes, $\psi : \mathbb{R}^{m \times n \times L} \rightarrow \mathbb{R}^{m \times n \times L \times S}$ denotes the function of embedding subnet. It is noted that zero padding is adopted in each convolutional layer to make the feature cubes the same size with the LR HSI.

3.3.2. Predicting Subnet

The embedding subnet is followed by a predicting subnet, which infers wavelet package coefficients. There are multiple output branches in the predicting subnet, each of which corresponds to one wavelet package sub-band. The predicting subnet takes the feature cubes extracted by the embedding subnet as input, each branch of the predicting subnet is trained to infer the wavelet coefficients at each sub-band. Similar to the embedding subnet, each branch in the predicting subnet is also stacked by

3D convolutional layers and non-linear activation layers with zero padding strategy adopted, and the predicted wavelet coefficients have the same spatial size with the LR HSI. The desired HR HSI is obtained by applying inverse WPT to the predicted wavelet coefficients, so the upscaling factor of SR depends on the number of WPT levels. Specifically, suppose the number of WPT levels is l , there would be $N_w = 4^l$ wavelet package sub-bands, and the number of output branches in the predicting subnet is also 4^l . Taking the shared feature cubes $\psi(\mathbf{X})$ as input, the i -th branch φ_i predicts the i -th wavelet package sub-band as $\varphi_i(\psi(\mathbf{X})) \in \mathbb{R}^{m \times n \times L}$, where $\varphi_i : \mathbb{R}^{m \times n \times L \times S} \rightarrow \mathbb{R}^{m \times n \times L}$, $i = 1, 2, \dots, N_w$ denotes the function of the i -th branch. The output of MW-3D-CNN can be denoted as a set of wavelet package coefficients:

$$\{\varphi_1(\psi(\mathbf{X})), \varphi_2(\psi(\mathbf{X})), \dots, \varphi_i(\psi(\mathbf{X})), \dots, \varphi_{N_w}(\psi(\mathbf{X}))\}, i = 1, 2, \dots, N_w. \quad (2)$$

In the training stage, the MW-3D-CNN learns the mapping between the LR HSI and the wavelet package coefficients of the latent HR HSI. In the testing stage, given the LR HSI, the MW-3D-CNN would infer the wavelet package coefficients at each sub-band. Applying inverse WPT to the predicted wavelet package coefficients, the HR HSI can be obtained:

$$\hat{\mathbf{Y}} = \phi\{\varphi_1(\psi(\mathbf{X})), \varphi_2(\psi(\mathbf{X})), \dots, \varphi_i(\psi(\mathbf{X})), \dots, \varphi_{N_w}(\psi(\mathbf{X}))\}, \quad (3)$$

where, ϕ denotes inverse WPT, $\hat{\mathbf{Y}} \in \mathbb{R}^{(r \times m) \times (r \times n) \times L}$ is the estimated HR HSI, $r = 2^l$ is the upscaling factor of SR.

Different wavelet sub-bands share the common deep layers in the embedding subnet due to the inter wavelet sub-bands correlation. The embedding subnet learns the shared feature cubes and the predicting subnet optimizes with respect to each wavelet package sub-band. The embedding subnet connects different branches into a unified predicting subnet and allows them to be jointly optimized. Specifically, the errors in each wavelet package sub-band can be jointly back-propagated to the embedding subnet to learn the shared features, and the embedding subnet will refine different branches in the predicting subnet. Compared with training each branch independently, such joint training could make different branches facilitate each other and implicitly capture the correlation among different wavelet sub-bands.

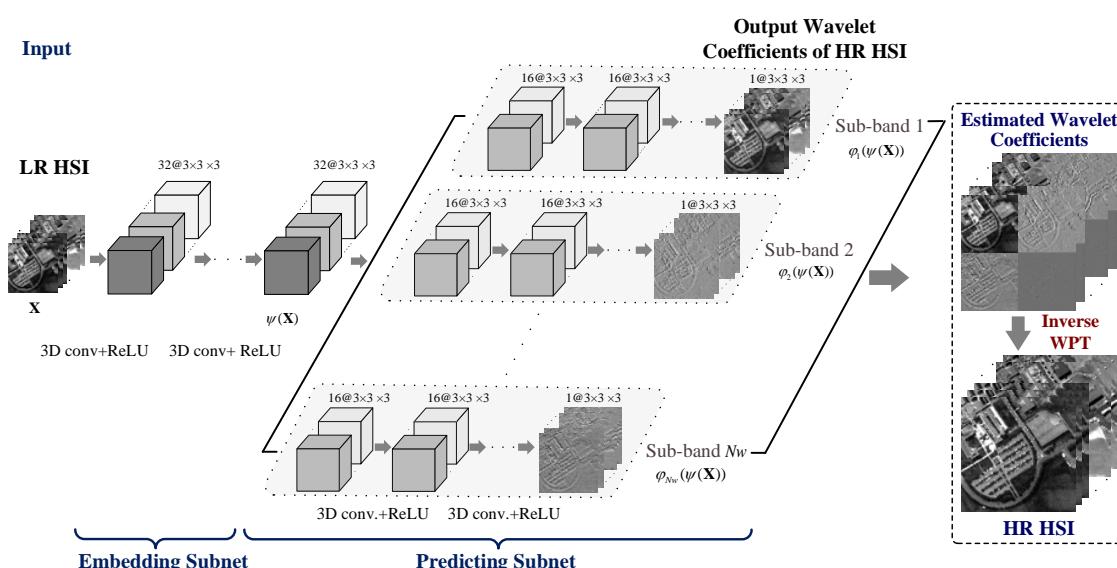


Figure 3. The architecture of the proposed multi-scale wavelet (MW)-3D-CNN, the number and the size of convolutional kernels are denoted at each layer, and the embedding subnet and predicting subnet have three and four layers respectively.

Our MW-3D-CNN focuses on predicting the wavelet package coefficients of HR HSI, compared with predicting the HR HSI directly, we consider three advantages. Firstly, wavelet coefficients describe the detailed textural information in HSI. Training the MW-3D-CNN to predict the wavelet coefficients is beneficial for recovering the detailed structures in HSI [33,36]. Secondly, a network with sparse activations is easier to be trained [34,35]. Wavelet coefficients have sparsity characteristics in the high-frequency sub-bands, and predicting wavelet coefficients promotes the sparsity of the MW-3D-CNN and makes the training easier and the trained network more robust. Finally, the MW-3D-CNN extracts features from the LR HSI directly. Compared with extracting features from the interpolated LR HSI, such as in [40,41], information in larger receptive field can be exploited.

3.4. Training of MW-3D-CNN

All the convolutional kernels and bias in the embedding and predicting subnets are trained in an end-to-end manner. L2 norm, which measures mean square error, is often used in loss function in the conventional CNN based image SR methods. However, the output of our network is the wavelet coefficients, which have larger values in the low-frequency sub-band and smaller values in the high-frequency sub-bands, as shown in the histograms in Figure 4. The L2 norm loss penalizes heavily on larger errors and is less sensitive to smaller errors [37]. On the contrary, the L1 norm loss penalizes equally on both larger and smaller errors, and it is more suitable than the L2 norm loss for wavelet coefficients prediction. In addition, compared with the L2 norm loss, the L1 norm loss is helpful for recovering sharper image structures with faster convergence [38]. Therefore, we propose to train the MW-3D-CNN with the L1 norm loss, the loss function is written as

$$L = \frac{1}{NN_w} \sum_{j=1}^N \sum_{i=1}^{N_w} \lambda_i \|C_j^i - \hat{C}_j^i\|_1, \quad (4)$$

where, C_j^i and $\hat{C}_j^i = \varphi_i(\psi(X_j))$ are the ground truth and the predicted wavelet package coefficients of the i -th sub-band respectively, $j = 1, 2, \dots, N$, N is the number of training samples, $i = 1, 2, \dots, N_w$, $N_w = 4^l$ is the number of sub-bands. X_j is the LR HSI of the j -th training sample. λ_i is the weight balancing the trade-off between different wavelet sub-bands, which is set to 1 for simplicity in the experiment. The loss function is optimized using the adaptive moment estimation (ADAM) method with standard back propagation. The trainable convolutional kernels and bias are updated according to the following rule [54]:

$$\theta_{(t)} = \theta_{(t-1)} - \alpha \cdot \tilde{m}_{(t)} / (\sqrt{\tilde{v}_{(t)}} + \epsilon) \quad (5)$$

where, $\theta_{(t)}$ denotes the trainable parameters (i.e., convolutional kernels and bias) at the t -th iteration, α is learning rate, ϵ is a constant to stabilize the updating, which is set to 10^{-6} . $\tilde{m}_{(t)}$ and $\tilde{v}_{(t)}$ are bias-corrected first and second moment estimates respectively:

$$\tilde{m}_{(t)} = m_{(t)} / (1 - \beta_1^t), \quad (6)$$

$$\tilde{v}_{(t)} = v_{(t)} / (1 - \beta_2^t), \quad (7)$$

$$m_{(t)} = \beta_1 \cdot m_{(t-1)} + (1 - \beta_1) \cdot \frac{\partial L_{(t-1)}}{\partial \theta}, \quad (8)$$

$$v_{(t)} = \beta_2 \cdot v_{(t-1)} + (1 - \beta_2) \cdot \left(\frac{\partial L_{(t-1)}}{\partial \theta} \right)^2, \quad (9)$$

where $\frac{\partial L_{(t-1)}}{\partial \theta}$ is the gradient with respect to the trainable parameters θ . β_1 and β_2 are two exponential decay rates for the moment estimation. In our implementation, the learning rate α is initially set to 0.001 and decreased by half for every 50 training epochs. The exponential decay rates β_1 and β_2 are set to 0.9 and 0.999 respectively. The batch size is set to 64. The number of training epochs is 200.

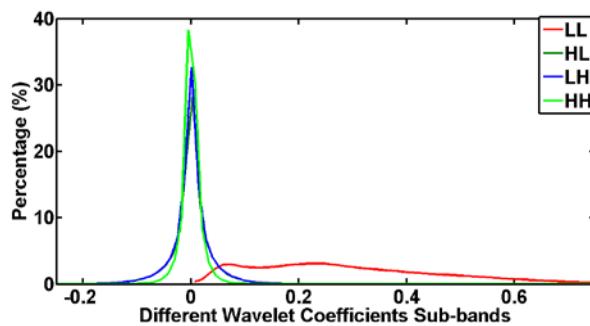


Figure 4. The histograms of different wavelet sub-bands of one-level WPT applied to each band of *Pavia University*, LL is the low-frequency sub-band, HL, LH, and HH are high-frequency sub-bands.

4. Experimental Results

In this section, we compare the MW-3D-CNN with other state-of-the-art HSI SR methods on several simulated HSI datasets. In order to demonstrate the applicability of MW-3D-CNN, we also validate it on real spaceborne Hyperion HSI. Since there is no reference HSI for SR assessment in real data case, we use the no-reference HSI assessment method in [55] to evaluate the SR performance.

4.1. Experiment Setting

Three datasets were used in the experiment. The first one is the Reflective Optics System Imaging Spectrometer (ROSIS) dataset, which contains two images taken over Pavia University and Pavia Center with sizes of 610×340 and 1096×715 , respectively. The spatial resolution is 1.3 m. After discarding the noisy bands, there are 100 bands remained in the spectral range 430~860 nm. The second dataset was collected by Headwall Hyperspec-VNIR-C imaging sensor over Chikusei, Japan, on July 29, 2014 [56]. The size is 2517×2335 with spatial resolution 2.5 m. There are 128 bands in the spectral range of 363~1018 nm. The third dataset is 2018 IEEE GRSS Data Fusion Contest data (denoted as “grss_dfc_2018”), which was acquired by the National Center for Airborne Laser Mapping (NCALM) over Houston University, on February 16, 2017 [57]. The size of this data is 1202×4172 . The spatial resolution is 1 m. It has 48 bands in the spectral range of 380~1050 nm.

The above data was treated as original HR HSI, the LR HSI was simulated via Gaussian down-sampling, which is a process of simulating LR HSI via applying a Gaussian filter to HR HSI and then down-sampling it in both vertical and horizontal directions. The Gaussian down-sampling was implemented using the “Hyperspectral and Multispectral Data Fusion Toolbox” [16]. For down-sampling by a factor of two, the Gaussian filter was of size 2×2 with zero mean and standard deviation 0.8493; for down-sampling by a factor of four, the Gaussian filter was of size 4×4 with zero mean and standard deviation 1.6986. All these parameters in down-sampling are suggested in [16,17].

We cropped three sub-images with rich textures from the original HSI as testing data, and the remainder was used as training data. About 100,000 LR-HR pairs were extracted as training samples to train the MW-3D-CNN. Each LR HSI sample was of size $16 \times 16 \times 16$. For training the MW-3D-CNN by an upscaling factor of two, there were four branches in the predicting subnet, the output wavelet coefficients in each branch were of size $16 \times 16 \times 16$, and the corresponding HR HSI sample was of size $32 \times 32 \times 16$. For training the MW-3D-CNN by an upscaling factor of four, there were 16 branches in the predicting subnet, the output wavelet coefficients in each branch were also of size $16 \times 16 \times 16$, and the corresponding HR HSI sample was of size $64 \times 64 \times 16$. It is noted that there was no overlapping between the training and testing regions. The network parameters of MW-3D-CNN were set according to network parameters in Figure 3. Haar wavelet function was used in WPT.

4.2. Comparison with State-of-the-Art SR Methods

In this sub-section, we compare the proposed method with other state-of-the-art HSI SR methods. Spectral-spatial group sparse representation HSI SR method (denoted as SSG) [27], and two CNN based SR algorithms, i.e., SRCNN [40] and 3D-CNN [32], were used for comparison. As an often used benchmark, bicubic interpolation was also compared. All the parameters of SSG, SRCNN, and 3D-CNN followed the default setting as described in [27,40], and [32]. The training samples and training epochs of SRCNN and 3D-CNN were the same with that of MW-3D-CNN, which guarantees the fairness of comparison.

The SR performance was assessed using peak-signal-noise-ratio (PSNR, dB), structural similarity index measurement (SSIM) [58], feature similarity index measurement (FSIM) [59], and spectral angle mean (SAM). We compute the PSNR, SSIM, and FSIM indices on each band, and then calculated the mean values over all the spectral bands.

The assessment indices of different SR methods are given in Tables 1 and 2. The scores of our method are better than the compared methods in most cases. The 3D-CNN in [32] could extract spectral-spatial features from HSI and jointly reconstruct different spectral bands, so it could lead to less spectral distortion than the SRCNN, as shown in Tables 1 and 2. Both 3D-CNN and MW-3D-CNN are in the framework of 3D CNN, and the MW-3D-CNN predicts the wavelet coefficients of the HR HSI, rather than directly predicting the HR HSI. Focusing on the wavelet coefficients makes the MW-3D-CNN more effective in preserving structures in HR HSI, so the results of MW-3D-CNN have higher PSNR values. In order to test the robustness of MW-3D-CNN over larger upscaling factor, we also implemented the SR by a factor of four and report the indices in Table 2. It can be found that the MW-3D-CNN can also achieve competitive results in most cases by an upscaling factor of four. In Figure 5, we plot the PSNR indices of different SR methods on each band. It is clear that the proposed method outperforms other methods on most spectral bands.

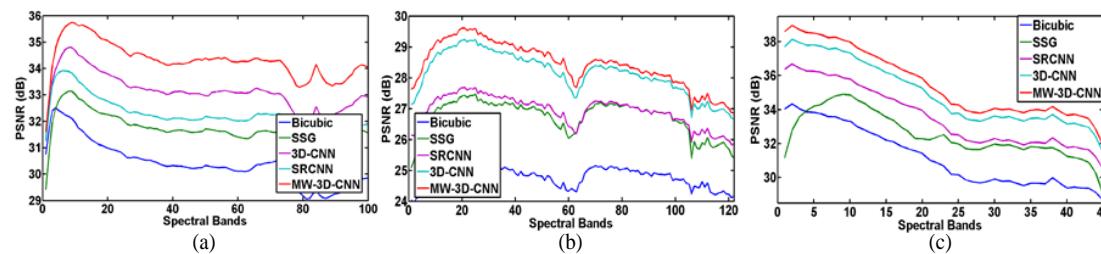


Figure 5. Peak-signal-noise-ratio (PSNR) indices of each bands of different hyperspectral image super-resolution (HSI SR) methods by an upscaling factor of two, (a) on *Pavia University*, (b) on *Chikusei* (c) on *Houston University* (grss_dfc_2018).

Table 1. The assessment indices of different HSI SR methods by an upscaling factor of two.

Data	Indices	Bicubic	SSG [27]	SRCNN [40]	3D-CNN [32]	MW-3D-CNN
<i>Pavia University</i>	PSNR (dB)	30.4032	31.7092	32.1961	33.1397	34.9394
	SSIM	0.8867	0.9132	0.9234	0.9398	0.9537
	FSIM	0.9191	0.9460	0.9517	0.9643	0.9754
	SAM	4.0979°	4.6845°	3.7519°	3.5470°	3.3302°
<i>Chikusei</i>	PSNR (dB)	24.7892	26.7419	26.9271	28.0397	28.4288
	SSIM	0.8596	0.9148	0.9301	0.9344	0.9396
	FSIM	0.8889	0.9313	0.9408	0.9483	0.9544
	SAM	4.2283°	3.7700°	3.0919°	2.9650°	2.9248°
<i>Houston University</i> (grss_dfc_2018)	PSNR (dB)	31.2005	32.5020	33.5990	34.9816	35.5552
	SSIM	0.9280	0.9480	0.9596	0.9669	0.9710
	FSIM	0.9878	0.9953	0.9991	0.9993	0.9997
	SAM	2.5757°	3.4858°	2.4268°	2.1029°	1.9252°

Table 2. The assessment indices of different HSI SR methods by an upscaling factor of four.

Data	Indices	Bicubic	SSG [27]	SRCNN [40]	3D-CNN [32]	MW-3D-CNN
<i>Pavia University</i>	PSNR (dB)	27.5136	27.6828	27.8132	28.7122	29.1069
	SSIM	0.7187	0.7328	0.7327	0.7745	0.7928
	FSIM	0.7905	0.8186	0.8058	0.8450	0.8620
	SAM	6.1537°	7.7461°	5.9707°	5.6644°	5.8828°
<i>Chikusei</i>	PSNR (dB)	19.8308	20.3108	21.0739	21.1284	20.6069
	SSIM	0.5623	0.6280	0.6723	0.6741	0.6853
	FSIM	0.7039	0.7646	0.7985	0.7979	0.7934
	SAM	7.8073°	7.9160°	6.5647°	6.5458°	7.2638°
<i>Houston University</i> (grss_dfc_2018)	PSNR (dB)	25.3139	26.0628	26.7927	27.8006	28.4968
	SSIM	0.7410	0.7703	0.7971	0.8259	0.8514
	FSIM	0.8988	0.9233	0.9372	0.9528	0.9653
	SAM	4.6611°	6.9780°	4.2034°	4.0398°	3.6881°

The SR results are presented in Figures 6–10. Some selected bands of the reconstructed HR HSIs are shown in Figures 6, 8 and 10. To compare the difference of the SR methods, in Figures 7, 9 and 11, we also give the residual maps of SR results, in which reconstruction error at each pixel can be reflected. In Figure 6, it is clear that the result of MW-3D-CNN is closer to the reference image, and the results of other compared methods are much brighter than the original HR image, which means that the spectral distortion is heavier. We also display some small areas by enlarging them to highlight the details of the SR results. In Figures 6 and 10, both SSG and SRCNN results suffer from artifacts with stripe-like patterns. By comparing the details in Figure 10, it can be found that our MW-3D-CNN SR results are sharper than the 3D-CNN results.

In the residual maps, it can be observed that all the SR results contain errors in the edges and details. Compared with other methods, our MW-3D-CNN method generates less errors. For example, in Figure 11, the error values in the MW-3D-CNN residual map are much sparser, which also demonstrates that predicting the wavelet coefficients is helpful for recovering the edges and detailed structures in the HR HSI.

We also present running time comparison of different SR methods in Tables 3 and 4. Most of the SR methods could infer HR HSI quickly. In the SSG method, dictionary learning and sparse coding is time consuming, so SSG takes the longest time to reconstruct HR HSI. The running time of MW-3D-CNN is comparable to 3D-CNN, as both of them could super-resolve HSI within 2 s. The running time comparison in Tables 3 and 4 indicates that our proposed method could achieve competitive performance in both SR accuracy and running time.

Table 3. The running time of different SR methods by an upscaling factor of two.

Data	Bicubic	SSG [27]	SRCNN [40]	3D-CNN [32]	MW-3D-CNN
<i>Pavia University</i>	0.42 s	2.37 h	233.45 s	0.96 s	1.18 s
<i>Chikusei</i>	0.44 s	2.86 h	241.84 s	1.14 s	1.30 s
<i>Houston University</i>	0.97 s	4.33 h	402.71 s	1.76 s	1.92 s

Table 4. The running time of different SR methods by an upscaling factor of four.

Data	Bicubic	SSG [27]	SRCNN [40]	3D-CNN [32]	MW-3D-CNN
<i>Pavia University</i>	0.24 s	2.28 h	237.58 s	1.12 s	1.16 s
<i>Chikusei</i>	0.28 s	2.77 h	247.75 s	1.20 s	1.42 s
<i>Houston University</i>	0.49 s	4.21 h	409.54 s	1.76 s	1.87 s

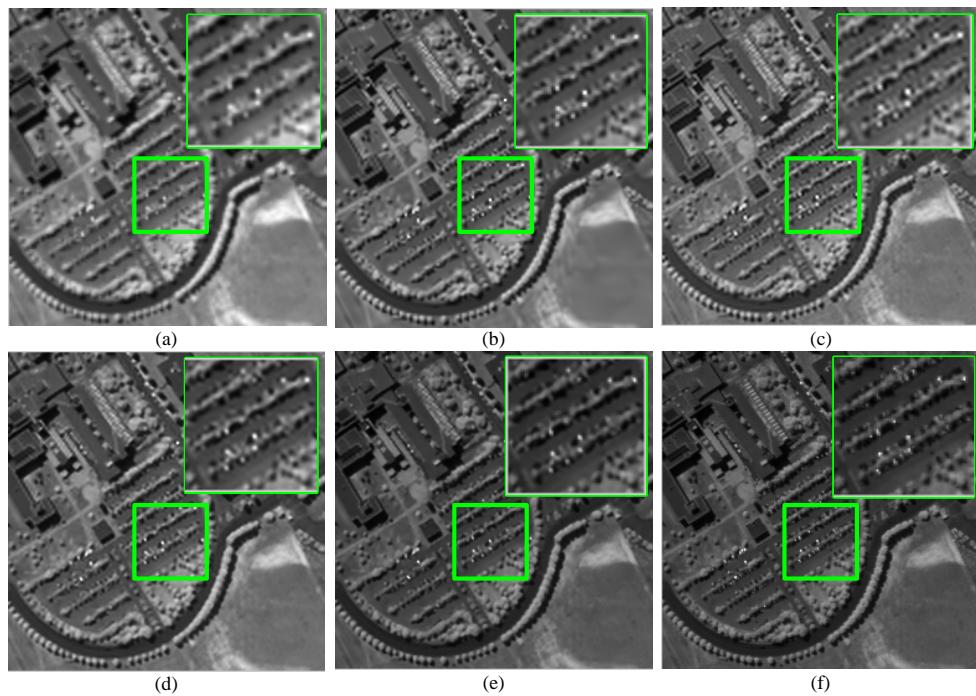


Figure 6. The SR results (band 85) of different methods by an upscaling factor of two, the testing data is cropped from *Pavia University* with size 256×256 . (a) Result of Bicubic, (b) result of spectral-spatial group sparse representation HSI SR method (SSG) [27], (c) result of super resolution CNN (SRCNN) [40], (d) result of 3D-CNN [32], (e) result of the proposed MW-3D-CNN, and (f) original HR image.

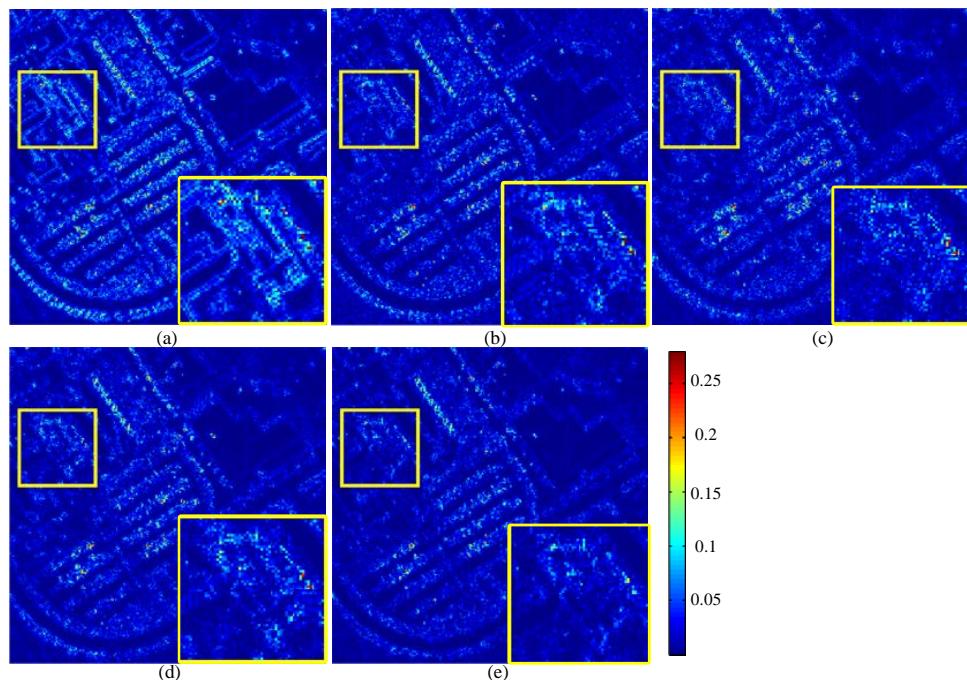


Figure 7. The residual maps on *Pavia University* by an upscaling factor of two. (a) Bicubic result, (b) SSG result [27], (c) SRCNN result [40], (d) 3D-CNN result [32], and (e) the proposed MW-3D-CNN result. The residual maps are displayed by scaling to the minimum and maximum errors.

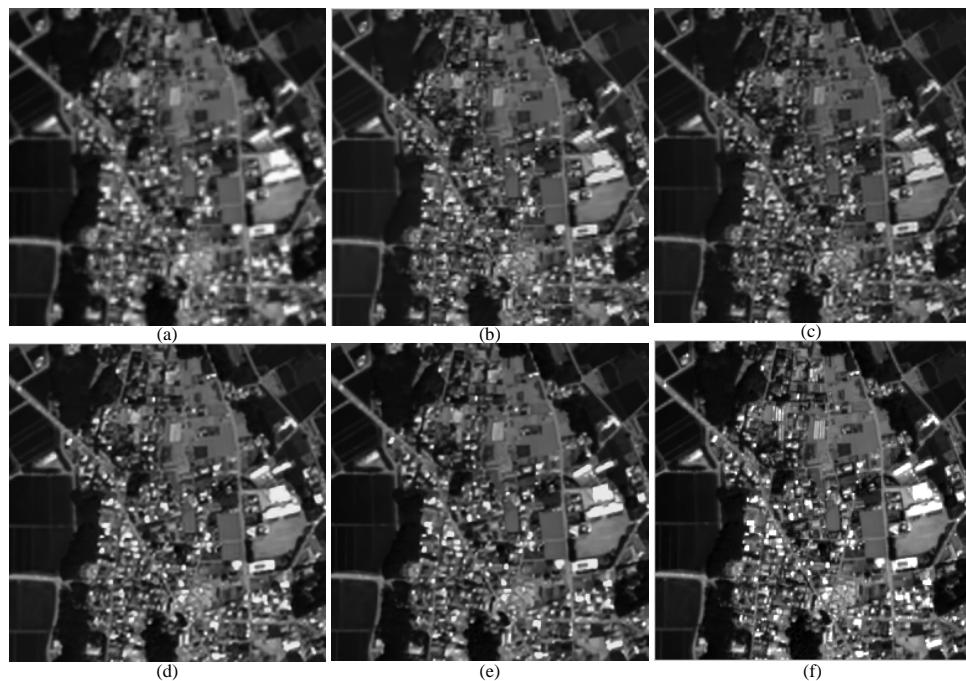


Figure 8. The SR results (band 20) of different methods by an upscaling factor of two, the testing data is cropped from *Chikusei* with size 256×256 . (a) Result of Bicubic, (b) result of SSG [27], (c) result of SRCNN [40], (d) result of 3D-CNN [32], (e) result of the proposed MW-3D-CNN, and (f) original HR image.

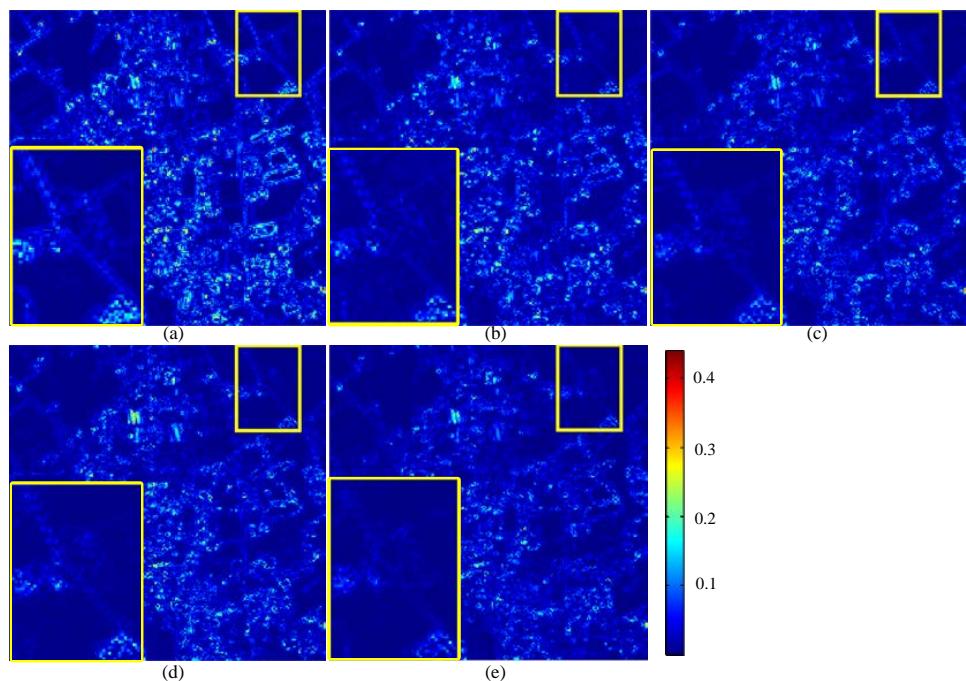


Figure 9. The residual maps on *Chikusei* by an upscaling factor of two. (a) Bicubic result, (b) SSG result [27], (c) SRCNN result [40], (d) 3D-CNN result [32], and (e) the proposed MW-3D-CNN result. The residual maps are displayed by scaling to the minimum and maximum errors.

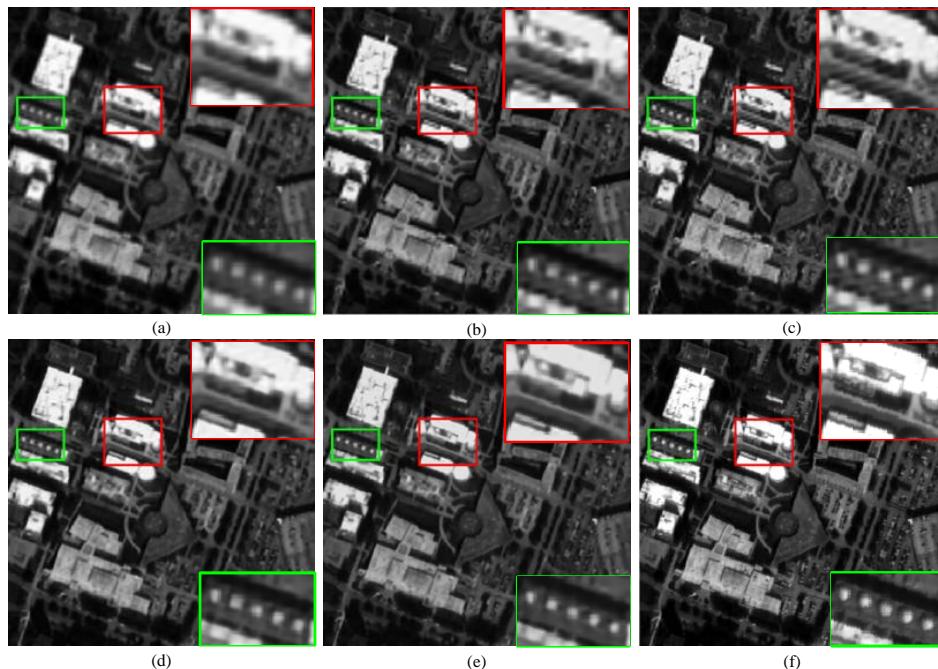


Figure 10. The SR results (band 5) of different methods by an upscaling factor of four, the testing data is cropped from *Houston University* (grss_dfc_2018) with size 512×512 . (a) Result of Bicubic, (b) result of SSG [27], (c) result of SRCNN [40], (d) result of 3D-CNN [32], (e) result of the proposed MW-3D-CNN, and (f) original HR image.

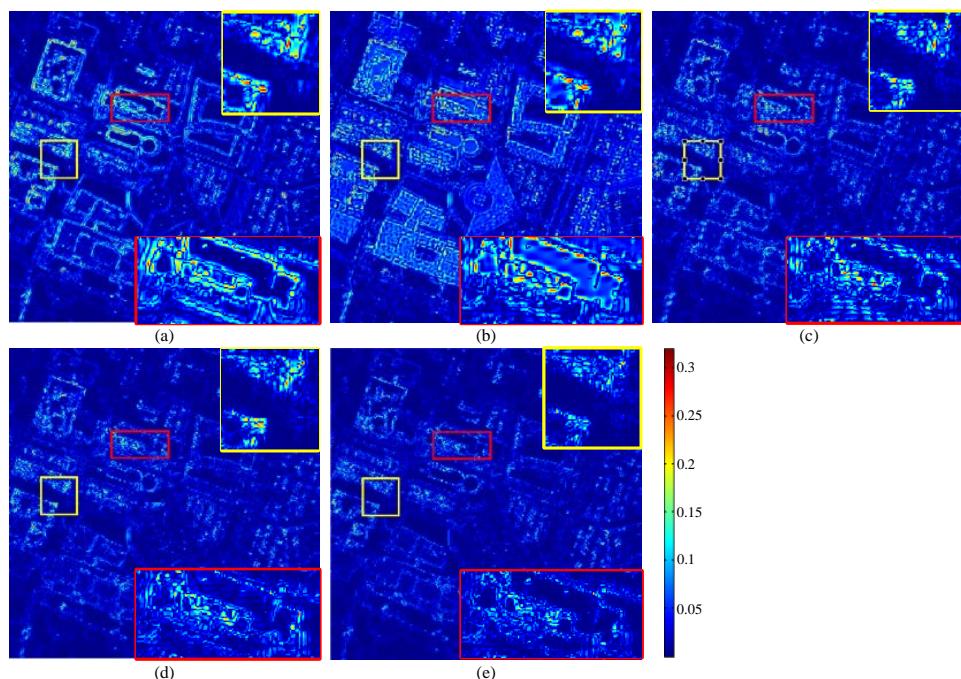


Figure 11. The residual maps on *Houston University* by an upscaling factor of four. (a) Bicubic result, (b) SSG result [27], (c) SRCNN result [40], (d) 3D-CNN result [32], and (e) the proposed MW-3D-CNN result. The residual maps are displayed by scaling to the minimum and maximum errors.

4.3. Application on Real Spaceborne HSI

In this sub-section, we also apply the MW-3D-CNN to real spaceborne HSI SR to demonstrate its applicability. Earth Observing-1 (EO-1)/Hyperion HSI was used as testing data. The spatial

resolution of Hyperion HSI is 30 m. There are 242 spectral bands in the spectral range of 400~2500 nm. The Hyperion HSI suffers from noise, and after removing the noisy bands and water absorption bands, 83 bands remain. The Hyperion HSI in this experiment was taken over Lafayette, LA, USA on October 2015. We cropped a sub-image with size 341×365 from it as the study area.

As there is no HR HSI in real application, we used the Wald protocol to train the networks [24]. The original 30 m HSI was regarded as HR HSI, and LR HSI with resolution 60 m was simulated via down-sampling. The LR-HR HSI pairs were used to train the MW-3D-CNN that could super-resolve HSI by a factor of two. The trained MW-3D-CNN was then applied to the 30 m Hyperion HSI, and HR HSI with 15 m resolution could be obtained. The super-resolved Hyperion HSIs are shown in Figure 12. In Figures 13 and 14, we show, in zoom, the results of the compared methods. The resolution of Hyperion HSI is enhanced significantly through SR. Compared with other methods, the proposed MW-3D-CNN generates HSI with sharper edges and clearer structures, as indicated by the area highlighted in the dashed boxes.

Since there is no reference image for assessment, the traditional evaluation indices such as PSNR cannot be used here. We used the no-reference HSI quality assessment method in [55], which measures the deviation of reconstructed HSI from pristine HSI, to evaluate the super-resolved Hyperion HSIs. The original Hyperion images were first screened for noisy bands and water absorption bands. The remaining bands were used as training data, quality-sensitive features were extracted from the training data and a benchmark multivariate Gaussian model was learned for the no-reference HSI assessment. The no-reference HSI quality scores after SR are listed in Table 5. It shows that by an upscaling factor of two where the SR image is at 15 m resolution, the proposed MW-3D-CNN performs better than other methods with a lower score, which means that the SR result deviates less from the pristine HSI than other SR results.

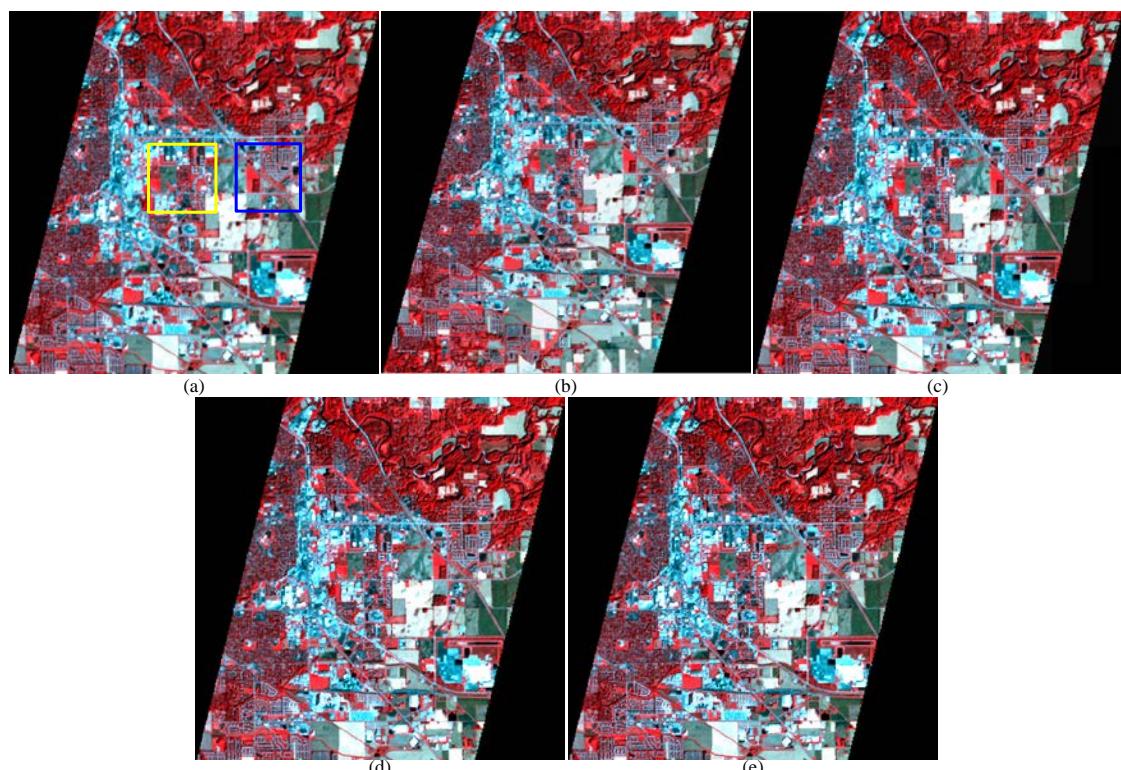


Figure 12. False color composite (band 45, 21, 14) of different Hyperion SR results. The upscaling factor is two, the size of the enhanced image is 682×780 with 15 m resolution. (a) result of Bicubic, (b) result of SSGS [27], (c) result of SRCNN [40], (d) result of 3D-CNN [32], and (e) result of MW-3D-CNN.

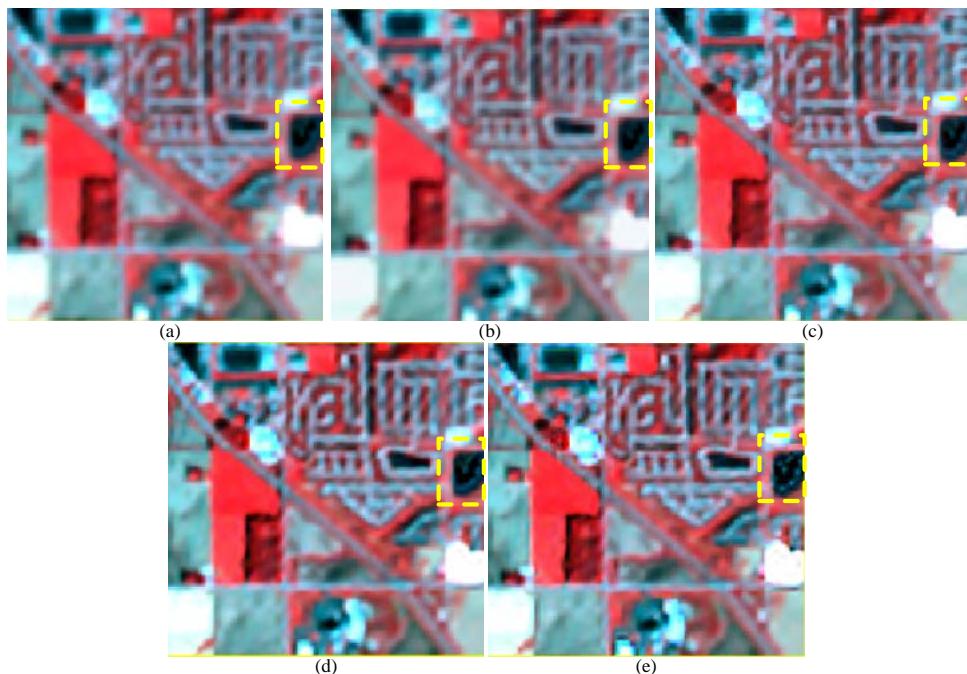


Figure 13. False color composite (band 45, 21, 14) of the enlarged area in the blue box of Figure 12. The size of the area is 200×200 . (a) Result of Bicubic, (b) result of SSGS [27], (c) result of SRCNN [40], (d) result of 3D-CNN [32], and (e) result of MW-3D-CNN.

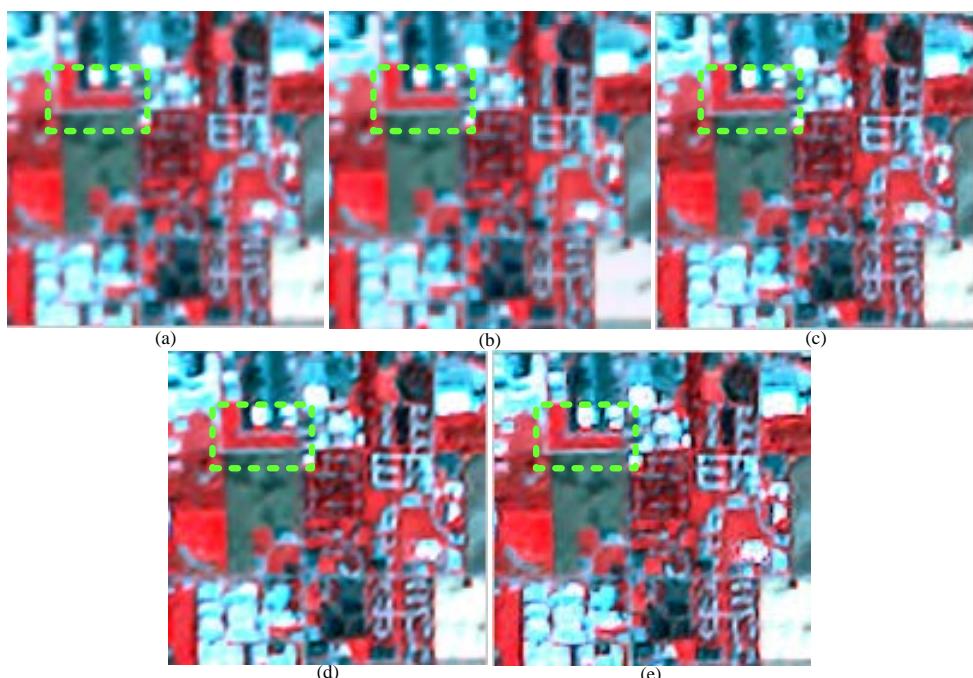


Figure 14. False color composite (band 45, 21, 14) of the enlarged area in the yellow box of Figure 12. The size of the area is 200×200 . (a) Result of Bicubic, (b) result of SSGS [27], (c) result of SRCNN [40], (d) result of 3D-CNN [32], and (e) result of MW-3D-CNN.

Table 5. The no-reference assessment scores of super-resolved 15 m Hyperion HSI.

SR Methods	Bicubic	SSGS [27]	SRCNN [40]	3D-CNN [32]	MW-3D-CNN
Scores	31.3888	28.3041	26.9271	25.6205	25.4930

5. Analysis and Discussions

5.1. Sensitivity Analysis on Network Parameters

It is theoretically hard to estimate the optimal network parameters of a deep learning architecture. We empirically tuned the network parameters and presented them in Figure 3. In this sub-section, we give the sensitivity analysis of MW-3D-CNN over the network parameters. We vary one network parameter and fix others, then observe the SR performance.

The sensitivity analysis over the size of 3D convolutional kernel is in Table 6. Proper large convolutional kernel size is necessary for collecting spatial and spectral information for HSI SR. It is clear that the best performance is achieved with convolutional kernel size $3 \times 3 \times 3$. The performance decreases when the convolutional kernel size is set to $5 \times 5 \times 5$. More spatial and spectral information can be exploited by larger convolutional kernel, but higher complexity of the network will be caused, and more parameters need to be trained. This may explain why the performance drops with the increase of kernel size.

The number of 3D convolutional kernels determines the number of feature cubes extracted by each layer. In our MW-3D-CNN, we set 32 convolutional kernels for each layer of the embedding subnet and 16 convolutional kernels for each layer of the predicting subnet, which leads to the best performance in most cases, as shown in Table 7. With the increase of convolutional kernel number, more feature cubes could be extracted, but the complexity of the network would be increased.

Usually, the deeper the network, the better the performance. With deeper architecture, the network would have larger capacity. In Table 8, it is shown that the best performance can be obtained in most cases when the number of convolutional layers in the embedding subnet and predicting subnet is set to three and four.

Table 6. PNSR (dB) indices of the sensitivity analysis of MW-3D-CNN over the size of 3D convolutional kernels. The upscaling factor is two.

Size of 3D Conv. Kernel	Pavia University	Chikusei	Houston University
$1 \times 1 \times 1$	30.3859	23.3061	31.0492
$3 \times 3 \times 3$	34.9394	28.4288	35.5552
$5 \times 5 \times 5$	34.5399	27.9122	35.4294

Table 7. PSNR (dB) indices of the sensitivity analysis of MW-3D-CNN over the number of 3D convolutional kernels in embedding and predicting subnets. The upscaling factor is two.

Number of 3D Conv. Kernels	Pavia University	Chikusei	Houston University
16 (embedding subnet), 8 (predicting subnet)	34.8725	28.3497	35.6839
32 (embedding subnet), 16 (predicting subnet)	34.9394	28.4288	35.5552
64 (embedding subnet), 32 (predicting subnet)	34.8568	28.2704	35.3547

Table 8. PSNR (dB) indices of the sensitivity analysis of MW-3D-CNN over the number of 3D convolutional layers in embedding and predicting subnets. The upscaling factor is two.

Number of 3D Conv. Layers	Pavia University	Chikusei	Houston University
2 (embedding subnet), 3 (predicting subnet)	34.9282	28.3663	35.4573
3 (embedding subnet), 4 (predicting subnet)	34.9394	28.4288	35.5552
4 (embedding subnet), 5 (predicting subnet)	35.1095	28.3744	35.4720

5.2. The Rationality Analysis of L1 Norm Loss

In order to verify the rationality of L1 norm loss, we trained the MW-3D-CNN using the L2 norm loss written as

$$\text{loss} = \frac{1}{NN_w} \sum_{j=1}^N \sum_{i=1}^{N_w} \lambda_i \|C_j^i - \hat{C}_j^i\|_2' \quad (10)$$

then compared it with the one trained using the L1 norm loss in Equation (4). The comparison is presented in Table 9. The L1 norm loss could mitigate the unbalance in penalizing low- and high-frequency wavelet package sub-bands caused by the L2 norm loss, so the MW-3D-CNN trained with the L1 norm loss performs better than the L2 norm loss on the testing data, as shown in Table 9.

In the training stage, the errors of the i -th wavelet package sub-band predicted by the MW-3D-CNN can be expressed as $(C_j^i - \hat{C}_j^i)$, where $j = 1, 2, \dots, N$, N is the number of training sample. We present the histograms of the errors after 200 training epochs in Figure 15. It is clear that the errors of different wavelet package sub-bands have similar statistics, as most of the errors are close to zero and tend to follow Laplacian distributions. Compared with the L2 norm, the L1 norm is more suitable for penalizing the Laplacian-like errors, which demonstrates the rationality of the L1 norm loss as well.

Table 9. PSNR (dB) of MW-3D-CNN trained with different losses. The upscaling factor is two.

Loss Functions	Pavia University	Chikusei	Houston University
L1 Norm Loss	34.9394	28.4288	35.5552
L2 Norm Loss	34.6417	28.3176	35.2615

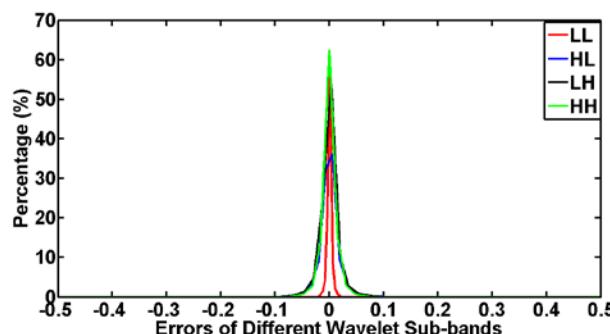


Figure 15. The histograms of errors in different wavelet sub-bands after 200 training epochs. The training data is extracted from *Pavia University*, the MW-3D-CNN is trained with the L1 norm loss, and the upscaling factor is two.

5.3. The Rationality Analysis of 3D Convolution

In this sub-section, in order to analyze the advantage of 3D convolution over 2D convolution for HSI SR, we replaced all the 3D convolutional layers in the MW-3D-CNN with 2D convolutional layers. In this case, it reduces to the architecture as the wavelet-SRNet method in [36]. Then we compared the MW-3D-CNN with the wavelet-SRNet. The loss function of wavelet-SRNet was originally designed with L2 norm in [36]. Here, we also trained the wavelet-SRNet with L1 norm as loss function, and the corresponding results are denoted as wavelet-SRNet-L2 and wavelet-SRNet-L1. The comparison between the MW-3D-CNN and the wavelet-SRNet is presented in Table 10.

In Table 10, it can be found that the MW-3D-CNN performs better than the wavelet-SRNet on the three datasets. The MW-3D-CNN is based on 3D convolutional layers, which could naturally exploit the spectral correlation and reduce the spectral distortion in HSI SR. We could also find that when the L1 norm is used as loss function for the wavelet-SRNet, the SR performance is slightly better than the L2 norm, which also demonstrates the effectiveness of L1 norm.

Table 10. PSNR (dB) of MW-3D-CNN and wavelet-SRNet with upscaling factor two.

Methods	Pavia University	Chikusei	Houston University
MW-3D-CNN	34.9394	28.4288	35.5552
Wavelet-SRNet-L2	32.2569	27.0149	34.1717
Wavelet-SRNet-L1	32.3658	27.0903	34.1537

5.4. Robustness over Wavelet Functions

In the experiment, we used Haar wavelet function in WPT. In this sub-section, we also perform the MW-3D-CNN with other two wavelet functions, Daubechies-2 and biorthogonal wavelet functions, to evaluate the robustness of MW-3D-CNN over the wavelet function. In Table 11, it can be found that the SR performance with different wavelet functions is close to each other. The SR performance changes slightly with different wavelet functions, which demonstrates the robustness of MW-3D-CNN over the wavelet functions.

Table 11. PSNR (dB) indices of MW-3D-CNN with different wavelet functions in WPT. The upscaling factor is two.

Wavelets	Pavia University	Chikusei	Houston University
Haar wavelet	34.9394	28.4288	35.5552
Daubechies-2 wavelet	35.0468	28.6751	35.5202
Biorthogonal wavelet	34.9695	28.4213	35.5594

The MW-3D-CNN is implemented on Tensorflow [60], with a NVIDIA GTX 1080Ti graphic card. It takes about 7 h and 20 h to train the MW-3D-CNN with upscaling factor two and four respectively. In the testing stage, inferring a HR HSI only takes less than two seconds, it is fast because there is only feed forward operation involved.

6. Conclusions

In this study, a MW-3D-CNN for HSI SR was proposed. Instead of predicting the HR HSI directly, we predicted the wavelet package coefficients of the latent HR HSI, and then reconstructed the HR HSI via inverse WPT. The MW-3D-CNN is constituted by an embedding subnet and a predicting subnet, both of which are built on 3D convolutional layers. The embedding subnet projects the input LR HSI into feature space and represents it with a set of feature cubes. These feature cubes are then fed to the predicting subnet, which consists of several output branches. Each branch corresponds to a wavelet package sub-band and predicts the wavelet package coefficients of each sub-band. The HR HSI can be reconstructed via inverse WPT. The experiment results on both simulated and real spaceborne HSI demonstrate that the proposed MW-3D-CNN could achieve competitive performance. The MW-3D-CNN learns the knowledge from the external training data for HSI SR. HSI has its prior information in both spectral and spatial domains, such as the structural self-similarity [26] and low rank prior [61–63]. Exploiting these prior information helps regularize the ill-posed HSI SR problem. How to combine such internal prior with external learned knowledge in the deep learning will need to be examined in future work. Furthermore, integrating adversarial loss [64] in training the network is another direction to boost the SR performance.

Author Contributions: Conceptualization, methodology, and writing—original draft preparation, J.Y.; writing—review and editing, J.C., Y.Z., and L.X.

Funding: This work is supported by the National Natural Science Foundation of China (61771391, 61371152), the National Natural Science Foundation of China and South Korean National Research Foundation Joint Funded Cooperation Program (61511140292), the Fundamental Research Funds for the Central Universities (3102015ZY045), the Jiangsu Provincial Social Developing Project (BE 2018727), the China Scholarship Council for joint PhD students (201506290120), the Science, Technology and Innovation Commission of Shenzhen Municipality (JCYJ20170815162956949), and the Innovation Foundation of Doctor Dissertation of Northwestern Polytechnical University (CX201621).

Acknowledgments: The authors gratefully acknowledge Space Application Laboratory, Department of Advanced Interdisciplinary Studies, University of Tokyo for providing the Hyperspec-VC Chikusei data. The authors would also like to thank the National Center for Airborne Laser Mapping and the Hyperspectral Image Analysis Laboratory at the University of Houston and the IEEE GRSS Image Analysis and Data Fusion Technical Committee for acquiring and providing the “grss_dfc_2018” data used in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nasrabadi, N.M. Hyperspectral target detection: An overview of current and future challenges. *IEEE Signal Process. Mag.* **2014**, *31*, 34–44. [[CrossRef](#)]
2. Clark, M.L.; Buck-Diaz, J.; Evens, J. Mapping of forest alliances with simulated multi-seasonal hyperspectral satellite imagery. *Remote Sens. Environ.* **2018**, *210*, 490–507. [[CrossRef](#)]
3. Yang, J.; Zhao, Y.Q.; Chan, J.C.W. Learning and transferring deep joint spectral–spatial features for hyperspectral classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4729–4742. [[CrossRef](#)]
4. Chen, F.; Wang, K.; Van de Voorde, T.; Tang, T.F. Mapping urban land cover from high spatial resolution hyperspectral data: An approach based on simultaneously unmixing similar pixels with jointly sparse spectral mixture analysis. *Remote Sens. Environ.* **2017**, *196*, 324–342. [[CrossRef](#)]
5. Yokoya, N.; Chan, J.C.W.; Segl, K. Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated EnMAP and Sentinel-2 images. *Remote Sens.* **2016**, *8*, 172. [[CrossRef](#)]
6. 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](#)]
7. Dalla Mura, M.; Vivone, G.; Restaino, R.; Addesso, P.; Chanussot, J. Global and local Gram-Schmidt methods for hyperspectral pansharpening. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Milan, Italy, 26–31 July 2015; pp. 37–40.
8. Shahdoosti, H.R.; Ghassemian, H. Combining the spectral PCA and spatial PCA fusion methods by an optimal filter. *Inf. Fusion* **2016**, *27*, 150–160. [[CrossRef](#)]
9. Yang, S.; Zhang, K.; Wang, M. Learning low-rank decomposition for pan-sharpening with spatial-spectral offsets. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *20*, 3647–3657.
10. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [[CrossRef](#)]
11. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [[CrossRef](#)]
12. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. Pansharpening by convolutional neural networks. *Remote Sens.* **2016**, *8*, 594. [[CrossRef](#)]
13. Wei, Y.; Yuan, Q.; Shen, H.; Zhang, L. Boosting the Accuracy of multispectral image pansharpening by learning a deep residual network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1795–1799. [[CrossRef](#)]
14. Yang, J.; Fu, X.; Hu, Y.; Huang, Y.; Ding, X.; Paisley, J. PanNet: A deep network architecture for pan-sharpening. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5449–5457.
15. Yuan, Q.; Wei, Y.; Meng, X.; Shen, H.; Zhang, L. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 978–989. [[CrossRef](#)]
16. Yokoya, N.; Grohnfeldt, C.; Chanussot, J. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 29–56. [[CrossRef](#)]
17. Yokoya, N.; Yairi, T.; Iwasaki, A. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 528–537. [[CrossRef](#)]

18. Zhu, X.X.; Grohnfeldt, C.; Bamler, R. Exploiting joint sparsity for pansharpening: The J-SparseFI algorithm. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2664–2681. [[CrossRef](#)]
19. Akhtar, N.; Shafait, F.; Mian, A. Sparse spatio-spectral representation for hyperspectral image super-resolution. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 63–78.
20. Wei, Q.; Bioucas-Dias, J.; Dobigeon, N.; Tourneret, J.Y. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3658–3668. [[CrossRef](#)]
21. Simões, M.; Bioucas-Dias, J.; Almeida, L.B.; Chanussot, J. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3373–3388. [[CrossRef](#)]
22. Zhang, L.; Wei, W.; Bai, C.; Gao, Y.; Zhang, Y. Exploiting clustering manifold structure for hyperspectral imagery super-resolution. *IEEE Trans. Image Process.* **2018**, *27*, 5969–5982. [[CrossRef](#)]
23. Yang, J.; Zhao, Y.Q.; Chan, J.C.W. Hyperspectral and Multispectral Image Fusion via Deep Two-Branches Convolutional Neural Network. *Remote Sens.* **2018**, *10*, 800. [[CrossRef](#)]
24. Xie, Q.; Zhou, M.; Zhao, Q.; Meng, D.; Zuo, W.; Xu, Z. Multispectral and Hyperspectral Image Fusion by MS/HS Fusion Net. *arXiv* **2019**, arXiv:1901.03281.
25. Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J.; Rasti, B.; Plaza, A. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 37–78. [[CrossRef](#)]
26. Zhao, Y.Q.; Yang, J.; Chan, J.C.-W. Hyperspectral imagery super-resolution by spatial–spectral joint nonlocal similarity. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2671–2679. [[CrossRef](#)]
27. Li, J.; Yuan, Q.; Shen, H.; Meng, X.; Zhang, L. Hyperspectral Image Super-Resolution by Spectral Mixture Analysis and Spatial-Spectral Group Sparsity. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1250–1254. [[CrossRef](#)]
28. Wang, Y.; Chen, X.A.; Han, Z.; He, S. Hyperspectral image super-resolution via nonlocal low-rank tensor approximation and total variation regularization. *Remote Sens.* **2017**, *9*, 1286. [[CrossRef](#)]
29. Yuan, Y.; Zheng, X.; Lu, X. Hyperspectral image super-resolution by transfer learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1963–1974. [[CrossRef](#)]
30. Li, Y.; Hu, J.; Zhao, X.; Xie, W.; Li, J.J. Hyperspectral image super-resolution using deep convolutional neural network. *Neurocomputing* **2017**, *266*, 29–41. [[CrossRef](#)]
31. Hu, J.; Li, Y.; Xie, W. Hyperspectral image super-resolution by spectral difference learning and spatial error correction. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1825–1829. [[CrossRef](#)]
32. Mei, S.; Yuan, X.; Ji, J.; Zhang, Y.; Wan, S.; Du, Q. Hyperspectral image spatial super-resolution via 3D full convolutional neural network. *Remote Sens.* **2017**, *9*, 1139. [[CrossRef](#)]
33. Liu, P.; Zhang, H.; Zhang, K.; Lin, L.; Zuo, W. Multi-level wavelet-CNN for image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 773–782.
34. Guo, T.; Mousavi, H.S.; Vu, T.H.; Monga, V. Deep wavelet prediction for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 104–113.
35. Bae, W.; Yoo, J.J.; Ye, J.C. Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 1141–1149.
36. Huang, H.; He, R.; Sun, Z.; Tan, T. Wavelet-SRNet: A wavelet-based CNN for multi-scale face super resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1689–1697.
37. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **2017**, *3*, 47–57. [[CrossRef](#)]
38. Cai, J.; Gu, S.; Zhang, L. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Trans. Image Process.* **2018**, *27*, 2049–2062. [[CrossRef](#)] [[PubMed](#)]
39. Scarpa, G.; Vitale, S.; Cozzolino, D. Target-adaptive CNN-based pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5443–5457. [[CrossRef](#)]
40. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [[CrossRef](#)] [[PubMed](#)]

41. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1646–1654.
42. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1637–1645.
43. Tai, Y.; Yang, J.; Liu, X.; Xu, C. Memnet: A persistent memory network for image restoration. In Proceedings of the IEEE Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4539–4547.
44. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
45. Wang, Y.; Perazzi, F.; McWilliams, B. A Fully Progressive Approach to Single-Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, Salt Lake City, UT, USA, 18–22 June 2018.
46. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep Laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 624–632.
47. Anbarjafari, G.; Demirel, H. Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image. *ETRI J.* **2010**, *32*, 390–394. [[CrossRef](#)]
48. Demirel, H.; Anbarjafari, G. Image resolution enhancement by using discrete and stationary wavelet decomposition. *IEEE Trans. Image Process.* **2011**, *20*, 1458–1460. [[CrossRef](#)]
49. Chavez-Roman, H.; Ponomaryov, V. Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1777–1781. [[CrossRef](#)]
50. Demirel, H.; Anbarjafari, G. Discrete wavelet transform-based satellite image resolution enhancement. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1997–2004. [[CrossRef](#)]
51. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 221–231. [[CrossRef](#)]
52. Li, Y.; Zhang, H.; Shen, Q. Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sens.* **2017**, *9*, 67. [[CrossRef](#)]
53. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [[CrossRef](#)]
54. Kingma, D.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.
55. Yang, J.; Zhao, Y.; Yi, C.; Chan, J.C.W. No-reference hyperspectral image quality assessment via quality-sensitive features learning. *Remote Sens.* **2017**, *9*, 305. [[CrossRef](#)]
56. Yokoya, N.; Iwasaki, A. *Airborne Hyperspectral Data over Chikusei*; Technical Report; SAL-2016-05-27; Space Appl. Lab., University of Tokyo: Tokyo, Japan, 2016.
57. 2018 IEEE GRSS Data Fusion Contest. Available online: <http://www.grss-ieee.org/community/technical-committees/data-fusion> (accessed on 10 June 2018).
58. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
59. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [[CrossRef](#)] [[PubMed](#)]
60. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
61. Xue, J.; Zhao, Y.; Liao, W.; Chan, J.C.W. Nonlocal Low-Rank Regularized Tensor Decomposition for Hyperspectral Image Denoising. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5174–5189. [[CrossRef](#)]
62. Yi, C.; Zhao, Y.Q.; Chan, J.C.-W. Spectral super-resolution for multispectral image based on spectral improvement strategy and spatial preservation strategy. *IEEE Trans. Geosci. Remote Sens.* **2019**.

63. Pan, L.; Hartley, R.; Liu, M.; Dai, Y. Phase-only Image Based Kernel Estimation for Single-image Blind Deblurring. *arXiv* **2018**, arXiv:1811.10185.
64. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).