*remote sensing*

*Article*

# Mixed 2D/3D Convolutional Network for Hyperspectral Image Super-Resolution

**Qiang Li, Qi Wang * and Xuelong Li**

School of Computer Science and the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China; liqbest@mail.nwpu.edu.cn (Q.L.); xuelong_li@nwpu.edu.cn (X.L.)

* Correspondence: crabwq@nwpu.edu.cn

Received: 2 April 2020; Accepted: 20 May 2020; Published: 21 May 2020

**Abstract:** Deep learning-based hyperspectral image super-resolution (SR) methods have achieved great success recently. However, there are two main problems in the previous works. One is to use the typical three-dimensional convolution analysis, resulting in more parameters of the network. The other is not to pay more attention to the mining of hyperspectral image spatial information, when the spectral information can be extracted. To address these issues, in this paper, we propose a mixed convolutional network (MCNet) for hyperspectral image super-resolution. We design a novel mixed convolutional module (MCM) to extract the potential features by 2D/3D convolution instead of one convolution, which enables the network to more mine spatial features of hyperspectral image. To explore the effective features from 2D unit, we design the local feature fusion to adaptively analyze from all the hierarchical features in 2D units. In 3D unit, we employ spatial and spectral separable 3D convolution to extract spatial and spectral information, which reduces unaffordable memory usage and training time. Extensive evaluations and comparisons on three benchmark datasets demonstrate that the proposed approach achieves superior performance in comparison to existing state-of-the-art methods.

**Keywords:** hyperspectral image; super-resolution (SR); convolutional neural networks (CNNs); mixed convolution; local feature fusion

## 1. Introduction

Hyperspectal imaging system collects surface information in tens to hundreds of continuous spectral bands to acquire hyperspectral image. Compared with multispectral image or natural image, hyperspectral image has more abundant spectral information of ground objects, which can reflect the subtle spectral properties of the measured objects in detail [1]. As a result, it is widely used in various fields, such as mineral exploration [2], medical diagnosis [3], plant detection [4], etc. However, the obtained hyperspectral image is often low-resolution because of the interference of environment and other factors. It limits the performance of high-level tasks, including change detection [5], image classification [6], etc.

To better and accurately describe the ground objects, the hyperspectral image super-resolution (SR) is proposed [7–9]. It aims to restore high-resolution hyperspectral image from degraded low-resolution hyperspectral image. In practical applications, the objects in the image are often detected or recognized according to the spectral reflectance of the object. Therefore, the change of spectral curve should be taken into account in reconstruction, which is different from natural image SR in computer vision [10].

Since the spatial resolution of hyperspectral images is lower than that of RGB image [11], existing methods mainly fuse high-resolution RGB image with low-resolution hyperspectral image [12,13]. For instance, Kwon et al. [12] use the RGB image corresponding to high-resolution hyperspectral image to obtain poorly reconstructed image. Then, the image in local is refined by sparse coding to obtain better SR image. Under the prior knowledge of spectral and spatial transform responses, Wycoff et al. [14] formulate the SR problem into non-negative sparse factorization. The problem is effectively addressed by alternating direction method of multipliers [15]. These methods realize hyperspectral image SR under the guidance of RGB images generated by the same camera spectral response (CSR) (http://www.maxmax.com/aXRayIRCameras.htm Access date: 29 April 2020), ignoring the differences of CSR between datasets or scenes. Suppose that the same CSR value is used in the process of reconstruction, which will obviously lead to the poor robustness of the algorithm. To address this issue, Fu et al. [16] design the CSR function selection layer, which can automatically select the optimal CSR according to a particular scene. In addition to the CSR function selection mechanism, the method simulates CSR as the convolutional layer to learn the optimal CSR function. It significantly improves the performance of hyperspectral image SR. However, such a scheme requires the pair of images to be well registered, which is usually difficult to follow in practice. Moreover, the scholars claim that these algorithms are unsupervised, but they are not actually unsupervised in that the ground-truth for RGB image is adopted during reconstruction. Therefore, in our paper, we focus on analyzing image super resolution without using RGB image.

The research of natural image SR has achieved great success in recent years due to the powerful representational ability of convolution neural networks (CNNs) [17,18]. Its main principle is to learn the mapping function between low-resolution and high-resolution image in a supervised way. The typical methods include SRCNN [19], EDSR [20], and SRGAN [21], etc. Due to the satisfying performance in natural image SR, the scholars apply these methods for hyperspectral image SR [22–25]. A remarkable characteristic of hyperspectral image is that the adjacent bands have strong correlation. Therefore, there is much recent literature on hyperspectral image SR using 3D convolution [26–29]. In terms of typical 3D convolution, these designed networks simultaneously extract spectral and spatial features so that it significantly improves performance. For example, Mei et al. [26] first present 3D full convolution neural network (3D-FCNN) to conduct hyperspectral image SR task. Yang et al. [27] design multi-scale wavelet 3D convolutional neural network (MW-3D-CNN). However, using typical 3D convolution leads to increasing the number of parameters. Later, Li et al. [30] propose dual 1D-2D spatial-spectral convolutional neural network. It uses 1D and 2D convolution to extract spectral and spatial features. Although this method effectively reduce the number of parameters, it lacks more exploration of the spatial information of image. The above works either use typical 3D convolution to analyze, resulting in more parameters of the network, or does not more focus on the mining of spatial information for hyperspectral image. Therefore, when the spectral information can be extracted, how to increase the spatial exploration of image and reduce the parameters of the model still needs more research efforts.

Considering the issues depicted above, in this paper, we propose a mixed 2D/3D convolutional network (MCNet) for hyperspectral image super-resolution. Our method learns the mapping function in a supervised way without using RGB image. The whole network uses 2D/3D convolution to extract hyperspectral image features instead of only one convolution. In each mixed convolutional module (MCM), the network can adaptively learn more effective spatial and spectral features from all the hierarchical 2D units. To reduce unaffordable memory usage and training time, in 3D unit, we employ separable 3D convolution to extract spatial and spectral information. Through three evaluation indexes, we demonstrate on three datasets that the performance of MCNet is superior to the state-of-the-art hyperspectral image SR approaches based on deep learning. In summary, our main contributions are follows:

- The novel mixed convolutional module (MCM) is proposed to mine the potential features. Using the correlation between 3D and 2D feature maps, 3D and 2D convolution share spatial information by reshaping. Compared with using only 3D convolution, it not only reduce the parameters of the network, but also makes the network learning relatively easy.
- Spatial and spectral separable 3D convolution is employed to extract spatial and spectral features in each 3D unit. It can effectively reduce unaffordable memory usage and training time.

- The local feature fusion is designed to adaptively preserve the accumulated features for 2D unit. It makes full use of all the hierarchical features in each 2D unit after changing the size of feature maps.
- Extensive experiments on three benchmark datasets demonstrate that the proposed approach achieves superior performance in comparison to existing state-of-the-art methods.

## 2. Related Work

There exists an extensive body of literature on hyperspectral image SR. Here we first outline several deep learning-based hyperspectral image SR methods. To better understand the proposed method, we then give a brief introduction to 3D convolution.

### *2.1. Deep Learning-Based Methods*

Recently, deep learning-based methods [31] have achieved remarkable advantages in the field of hyperspectral image SR. Here, we will briefly introduce several methods with CNNs. Li et al. [25] propose a deep spectral difference convolutional neural network (SDCNN) by using five convolutional layers to improve spatial resolution. Under spatial constraint strategy, it makes the reconstructed hyperspectral image preserve spectral information through post-processing. Jia et al. [24] present spectral-spatial network (SSN), including spatial and spectral sections. It tries to learn the mapping function between low-resolution and high-resolution images and fine-tune spectrum. Yuan et al. [23] use the knowledge from natural image to restore high-resolution hyperspectral image by transfer learning, and collaborative non-negative matrix factorization is proposed to enforce collaborations between low-resolution and high-resolution hyperspectral image. All of these methods need two steps to achieve image reconstruction, i.e., the algorithm first improves the spatial resolution. To avoid spectral distortion, some constraint criteria are then employed to retain the spectral information. It is clear that the spatial resolution may be changed while maintaining the spectral information. Inspired by deep recursive residual network [32], Li et al. [22] propose grouped deep recursive residual network (GDRRN) to execute hyperspectral image SR task. When 2D convolution is employed, the above networks can only extract the spatial information of hyperspectral images (see Figure 1a). They do not use the information of spectral dimension, thus achieving poor performance.

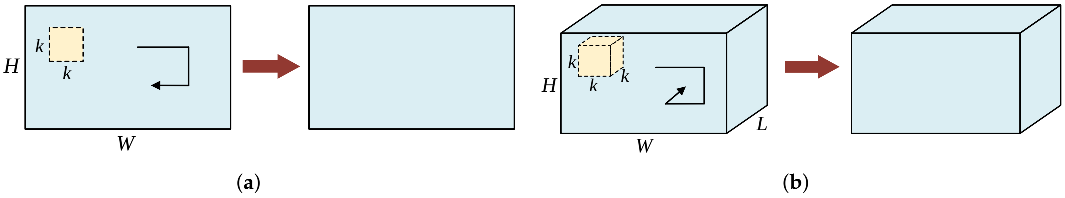


**Figure 1.** The illustration of 2D and 3D convolution operation adapted from [33]. (**a**) employing 2D convolution on an image. (**b**) employing 3D convolution on an image cube.

Since 3D convolution can extract spectral and spatial information at the same time (see Figure 1b), Mei et al. [26] present 3D full convolution neural network (3D-FCNN) that contains five layers. It explores the relationship of the spatial information and adjacent pixels between spectra. However, the method changes the size of the estimated hyperspectral image, which is not suitable for the purpose of image reconstruction. Yang et al. [27] design multi-scale wavelet 3D convolutional neural network (MW-3D-CNN). The network includes pre-processing and post-processing. Inspired by generative adversarial network (GAN), many hyperspectral image SR algorithms using GAN are proposed. Li et al. [28] design 3D-GAN-based hyperspectral image SR. Jiang et al. [34] propose a GAN contains spectral and spatial feature extraction section. Usually, GAN-based on SR is not easy to train. Furthermore, these networks either have many parameters or do not extract spatial and spectral features at the same time. Then, Li et al. [30] propose dual 1D-2D spatial-spectral convolutional neural network. It uses 1D and 2D convolution to extract spectral and spatial features, respectively, and fuses them by reshape operation. Although this method effectively solve the above issues, it lacks of more exploration of the spatial information of image.

### *2.2. 3D Convolution*

For natural image SR, the scholars usually employ 2D convolution to extract the features and obtain good performance [35,36]. As we introduced earlier, the hyperspectral image contains many continuous bands, which results in a significant characteristic that there is a great correlation between adjacent bands [37]. If we directly use 2D convolution to conduct hyperspectral image SR task, it will make it impossible to effectively exploit potential features between bands. Therefore, in order to make full use of this characteristic, we design network by using 3D convolution to analyze the spatial and spectral features of hyperspectral image in our paper.

Since 3D convolution takes into account the inter-frame motion information in the time dimension, it is widely used in video classification [38], action recognition [39] and other fields. Unlike 2D convolution, the 3D convolution operation is implemented by convolving a 3D kernel with feature maps. Intuitively, the number of parameters of the training network using 3D convolution is an order of magnitude more than that of the 2D convolution. To address this problem, Xie et al. [40] develop typical separable 3D CNNs (S3D) model to accelerate video classification. In this model, the standard 3D convolution is replaced by spatial and temporal separable 3D convolution (see Figure 2), which demonstrates that this way can effectively reduce the number of parameters while still maintain good performance.

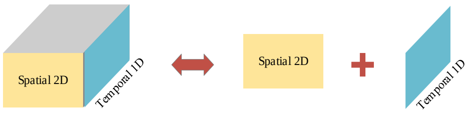


**Figure 2.** The illustration that standard 3D convolution can be separated into two parts: spatial convolution and temporal convolution.

## 3. Proposed Method

### *3.1. Network Structure*

In this section, we will detail the overall architecture of our MCNet, whose flowchart is shown in Figure 3. As can be seen from this figure, our method mainly consists of three parts: initial feature extraction (IFE) sub-network, deep feature extraction (DFE) sub-network, and image reconstruction (IR) sub-network. Let $I_{LR} \in R^{L \times W \times H}$ and $I_{SR}$ represent the input low-resolution hyperspectral image and the output reconstructed hyperspectral image, where $W$ and $H$ are the width and height of each band, and $L$ represents the total number of the bands in hyperspectral image. As we said earlier, 3D convolution can analyze information other than spatial dimensions. Therefore, in this paper, we use 3D convolution to extract spatial and spectral information from hyperspectral image. Since the size of the input low-resolution image is $L \times W \times H$, in order to employ 3D convolution, we need to reshape $I_{LR}$ into four dimensions ($1 \times L \times W \times H$) at the beginning of the network. Then, a standard 3D convolution is applied to extract shallow features about $I_{LR}$, i.e.,

$$F_0 = f_c(Reshape(I_{LR})), \tag{1}$$

where $Reshape(\cdot)$ is the function that changes the size of feature maps, and $f_c(.)$ denotes 3D convolution operation. The initial features of $F_0$ is fed into spatial-spectral residual module, which is described in detail in Section 3.2. After $D$ residual modules and global skip connection, the deep feature maps $F_D$ are denoted as

$$F_D = F_0 + M_D(M_{D-1}(...M_1(F_0) + F_0...) + F_0), \tag{2}$$

where $M_d(\cdot)$ denotes the operation of the $d$-th residual module. With respect to the impact of the number of residual module $D$ in our network, we will analyze it in Section 4.4.1. For IR sub-network, we use transposed convolution layer to upsample these feature maps to the desired scale via scale factor $r$, which is followed by a convolution layer. After reshaping, the output size becomes $L \times W \times H$. Finally, the output of MCNet can be obtained by

$$I_{SR} = Reshape(f_c(f_{up}(F_D, r))), \tag{3}$$

where $f_{up}(\cdot)$ is the function for upsampling.

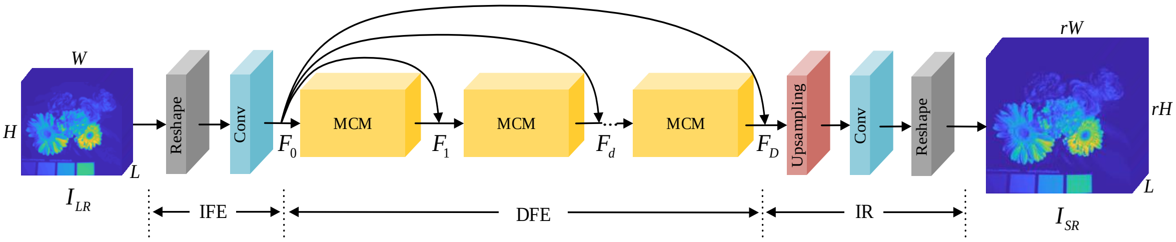


**Figure 3.** Overall architecture of our proposed MCNet. The whole network mainly contains three parts. First, the low-resolution hyperspectral image $I_{LR}$ is fed into initial feature extraction (IFE) sub-network. Then, these shallow features go through several mixed convolution modules (MCMs) and skip connections to obtain the depth features of the hyperspectral image in deep feature extraction (DFE) sub-network. Finally, these feature maps are upsampled according to scale factor $r$ in image reconstruction (IR) sub-network. The reconstructed hyperspectral image $I_{SR}$ is obtained by convolution and reshape.

### *3.2. Mixed Convolutional Module*

The architecture of mixed convolutional module (MCM) is illustrated in Figure 4. As provided in this figure, the module mainly contains four 3D units, three 2D units, and local feature fusion. In the $d$-th MRM, suppose $F_{d-1}$ and $F_d$ are the input and output feature maps, respectively. Under the local residual connection, the output $F_d$ of the $d$-th MCM can be defined as

$$F_d = f_{3D}(F_{d,f} + F_{d-1}), \tag{4}$$

where $f_{3D}(\cdot)$ is the function of 3D unit. Next, we will present the details about the proposed two units and local feature fusion.

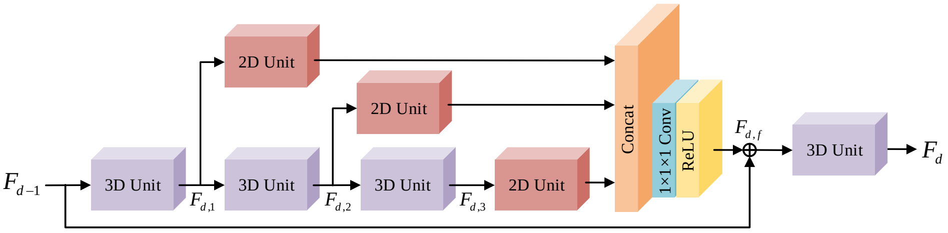


**Figure 4.** Architecture of the $d$-th mixed convolutional module (MCM). The module mainly contains four 3D units, three 2D units, and local feature fusion. The feature maps from $F_{d-1}$ are first fed into the first 3D unit. After two 3D units, the output feature maps of each 3D unit is reshaped. These feature maps are fed into 2D unit, respectively. Then, the feature output from 2D units of different depths are concatenated together. More effective features are attached to 3D unit after local residual learning. Finally, the output of the module $F_d$ is obtained.

### 3.2.1. 3D Unit

As we said in Section 2, the previous works use spatial and temporal separable 3D convolution to represent the standard 3D convolution for video classification, i.e, the size of the filter $k \times k \times k$ is modified as $k \times 1 \times 1$ and $1 \times k \times k$, which has been proven to perform better [40]. Therefore, to reduce unaffordable memory usage and training time, in our paper, we use this method to replace the standard 3D convolution in 3D unit. Please note that the temporal information refers to spectral information for hyperspectral image.

With respect to 3D unit (see Figure 5a), the filter $1 \times k \times k$ is adopted to first extract the spatial features of each band, and the filter $k \times 1 \times 1$ is used to extract the features between spectra. After each convolution operation, we add the rectified linear unit (ReLU). Through the local skip connection, the output of $n$-th 3D unit can be formulated as

$$f_{3D}(z) = \sigma(f_c(\sigma(f_c(z))) + z, \tag{5}$$

where $\sigma$ denotes the ReLU activation function. In this way, it does not just effectively mine the potential information between spectra, but also speeds up the implementation of the algorithm.

### 3.2.2. 2D Unit

If the network more focus on the analysis of spectral and spatial information by using 3D convolution, it would significantly increase the parameters of the network. This makes it impossible to design a deeper network. The main purpose of using spectral information is to improve spatial information. From this point of view, it is not necessary for all layers of the designed network to analyze the information between spectra. We need to more focus on the spatial features of image and reduce the feature analysis between spectra. Therefore, we hope that the designed network not only explore more spatial information but also reduce the parameters. Based on this motivation, we add 2D unit after each 3D unit to mine spatial features.

Now we present the proposed 2D unit, which is shown in Figure 5b. Specifically, in order to use 2D convolution, the output feature maps of 3D unit are first reshaped from $N \times C \times L \times W \times H$ to $(N * L) \times C \times W \times H$, where $C$ is the number of channel, and $N$ denotes batch size. Then, two 2D convolutions and ReLU activation function are added in this unit. Finally, the feature maps after these operations are reshaped into its original size. The output of $n$-th 2D unit can be obtained by

$$f_{2D}(z) = Reshape(f_c(\sigma(f_c(Reshape(z))))). \tag{6}$$

By making use of the correlation between 3D and 2D feature maps, 3D and 2D convolution can effectively share spatial information. In addition, 2D spatial features are relatively easy to learn. By doing so, there are two main benefits. On the one hand, it can promote the learning of 3D features. One the other hand, compared with using only 3D convolution, the 2D unit can greatly reduce the parameters of the network. Furthermore, it also enables the network to more mine spatial features of hyperspectral image, while the spectral information can be extracted.

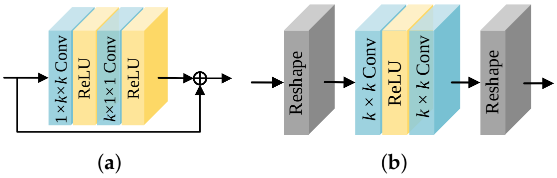


**Figure 5.** The detailed architecture of the 2D/3D unit. (**a**) 3D unit. (**b**) 2D unit.

### 3.2.3. Local Feature Fusion

To make the network learn more useful information, we design local feature fusion strategy (see Figure 4) to adaptively retain the cumulative features from 2D unit. It enables the network can fully extract hyperspectral image features. Specifically, the features from different 2D units are first concatenated to learn fusion information. To do a local residual learning between the fused result and input $F_{d-1}$, it is necessary to reduce the number of feature maps. Thus, we add a convolution layer whose filter size is $1 \times 1 \times 1$ to adaptively retain valid information. Besides, we also set the ReLU activation function after convolution. As a result, the output of local feature fusion $F_{d,f}$ is formulated as

$$F_{d,f} = \sigma(f_c(Concat[f_{2D}(F_{d,1}), f_{2D}(F_{d,2}), f_{2D}(F_{d,3})])), \tag{7}$$

where *Concat* denotes the concatenation operation.

### *3.3. Skip Connections*

As the depth of the network increases, the weakening of information flow and the disappearance of gradient hinder the training of the network. Recently, there are many ways to solve these problems. For instance, He et al. [41] first use skip connection between layers so as to improve the information flow and make it easier to train. To fully explore the advantages of skip connection, Huang et al. [42] propose DenseNet. The network has the advantages of strengthening feature propagation, supporting feature reuse, and reducing the number of parameters.

For SR task, the input low-resolution image is greatly similar to the output high-resolution image, i.e., the low-frequency information carried by the low-resolution image is similar to that of the high-resolution image [43]. According to this characteristic, the researchers use dense connections to enhance the information flow of the whole network and alleviate the disappearance of the gradient for natural image SR, thus effectively improving the performance of the algorithm. Therefore, we add several global residual connections in our network. Since the shallow network can retain more edge or texture information of hyperspectral image, the feature maps from IFE are fed into the the back of each module, which can enhance the performance of the entire network.

### *3.4. Network Learning*

For network training, the MCNet is optimized by minimizing the difference between reconstructed hyperspectral image $I_{SR}$ and corresponding ground-truth hyperspectral image $I_{HR}$. Mean square error (MSE) is often used as loss function to study the parameters of the network for hyperspectral image SR algorithms based on deep learning [25]. Additionally, some methods design two terms in loss function to minimize the difference, including MSE and spectral angle mapping (SAM) [22,44]. In fact, these loss functions do not make the network converge better and obtain poor results, which is proved in the experiment section. For natural image SR, as far as we know, many networks in recent years usually use $L1$ as loss function, and the experiments also demonstrate that the $L1$ can obtain more powerful performance and convergence [17]. Therefore, in this paper, we refer to the natural image SR method and adopt $L1$ as the loss function of our designed network. The loss function of MCNet is

$$\mathcal{L}(I_{SR}, I_{HR}; \theta) = \frac{1}{M} \sum_{m=1}^{M} ||{I_{HR}}^{(m)} - {I_{SR}}^{(m)}||_1, \tag{8}$$

where $M$ is the number of training patches and $\theta$ denotes the parameter set of the MCNet network.

## 4. Results

To verify the effectiveness of the proposed MCNet, in this section, we first introduce three public datasets. Then, the implementation details and evaluation indexes are described. Finally, we assess the performance of our MCNet by comparisons to the state-of-the-art methods.

### *4.1. Datasets*

(a) CAVE dataset: The CAVE dataset (http://www1.cs.columbia.edu/CAVE/databases/multispectral/ Access date: 29 April 2020) is gathered by cooled CCD camera at a 10 nm step from 400 nm to 700 nm (31 bands) [45]. The dataset contains 31 scenes, divided into 5 sections: real and fake, skin and hair, paints, food and drinks, and stuff. The size of all hyperspectral image is 512 × 512 × 31 in this dataset. Each band is stored as a 16-bit grayscale PNG image.

(b) Harvard dataset: The Harvard dataset (http://vision.seas.harvard.edu/hyperspec/explore.html Access date: 29 April 2020) is obtained by Nuance FX, CRI Inc. camera in the wavelength range of 400 nm to 700 nm. [46]. The dataset consists of 77 hyperspectral images of real-world indoor or outdoor scenes under daylight illumination. The size of each hyperspectral image is 1040 × 1392 × 31 in this dataset. Unlike CAVE dataset, this dataset is stored as .mat file.

(c) Foster dataset: The Foster dataset (https://personalpages.manchester.ac.uk/staff/d.h.foster/Local_Illumination_HSIs/Local_Illumination_HSIs_2015.html Access date: 29 April 2020) is collected using a low-noise Peltier-cooled digital camera (Hamamatsu, model C4742-95-12ER) [47]. The dataset includes 30 images from the Minho region of Portugal during late spring and summer of 2002 and 2003. Each hyperspectral image has 33 bands with the size of 1204 × 1344 pixels. Similarly, the dataset is also stored as .mat file. Some RGB images corresponding to hyperspectral images are shown in Figure 6.

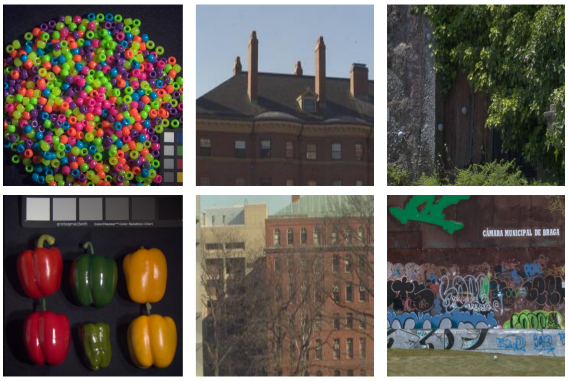


**Figure 6.** Some RGB images corresponding to hyperspectral images on three datasets.

### *4.2. Implementation Details*

As mentioned earlier, different datasets are gathered by different hyperspectral cameras, so we need to train and test each dataset individually, which is different from the natural image SR. In our work, 80% of the samples are randomly selected as training set, and the rest are used for testing.

For the training phase, since there are too few images in these datasets for deep learning algorithm, we augment the training data by randomly selecting 24 patches. Each patch is flipped horizontally, rotated (90°, 180°, and 270°), and scaled (1, 0.75, and 0.5). According to scale factor $r$, these patches are downsampled as low-resolution hyperspectral images with the size of $32 \times 32 \times L$ by bicubic interpolation [48]. Before feeding the mini-batch into our network, we subtract the average value of the entire training images for patches. In our work, we set the size of filter in 3D unit as $3 \times 1 \times 1$ and $1 \times 3 \times 3$ in each convolution layer expect those for initial feature extraction and image reconstruction (the size of filter is set to $3 \times 3 \times 3$). The size of the filter in 2D unit is set to $3 \times 3$. The number of filter for all layer in our network is 64. We initialize each convolutional filter using [49]. The ADAM optimizer with $\beta_1 = 0.9, \beta_2 = 0.999$ is employed to train our network. The learning rate is initialized as $10^{-4}$ for all layers, which decreases by a half at every 35 epochs.

For the test phase, in order to improve the efficiency of the test, we only use the top left 512 × 512 region of each test image for evaluation. Our method is conducted using the PyTorch framework with NVIDIA GeForce GTX 1080 GPU.

### *4.3. Evaluation Metrics*

To qualitatively measure the proposed MCNet, three evaluation methods are employed to verify the effectiveness of the algorithm, including peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and spectral angle mapping (SAM). They are defined as

$$PSNR = \frac{1}{L}\sum_{l=1}^{L} 10log_{10}\left(\frac{MAX_l^2}{MSE_l}\right) \tag{9}$$

$$MSE_l = \frac{1}{WH}\sum_{w=1}^{W}\sum_{h=1}^{H}\left(I_{SR}(w,h,l) - I_{HR}(w,h,l)\right)^2 \tag{10}$$

$$SSIM = \frac{1}{L}\sum_{l=1}^{L}\frac{\left(2\mu_{I_{SR}}^l\mu_{I_{HR}}^l + c_1\right)\left(2\sigma_{I_{SR}I_{HR}}^l + c_2\right)}{\left((\mu_{I_{SR}}^l)^2 + (\mu_{I_{HR}}^l)^2 + c_1\right)\left((\sigma_{I_{SR}}^l)^2 + ((\sigma_{I_{HR}}^l)^2 + c_2\right)} \tag{11}$$

$$SAM = arccos\left(\frac{< I_{SR}, I_{HR} >}{||I_{SR}||_2||I_{HR}||_2}\right) \tag{12}$$

where $MAX_l$ is the maximal pixel value for $l$-th band, $\mu_{I_{SR}}^l$ and $\mu_{I_{HR}}^l$ denote the mean of $I_{SR}$ and $I_{HR}$ for $l$-th band, respectively, $\sigma_{I_{SR}}^l$ and $\sigma_{I_{HR}}^l$ are the variance of $I_{SR}$ and $I_{HR}$ for $l$-th band, $\sigma_{I_{SR}I_{HR}}^l$ is the covariance of $I_{SR}$ and $I_{HR}$ for $l$-th band, $c_1$ and $c_2$ are two constants, $< \cdot, \cdot >$ represents the dot product operation, and $|| \cdot ||_2$ is l2 norm operation.

In general, the larger the PSNR and SSIM is and the smaller the SAM is, the better the performance of the reconstructed hyperspectral image is.

### *4.4. Model Analysis*

In this section, we conduct sufficient experiments, including study of $D$ module and ablation study analysis. To make a simple and fair comparison, we analyze the results for scale factor ×2 on CAVE dataset.

#### 4.4.1. Study of D Module

The structure of our proposed MCNet is determined by the number of the mixed convolutional module $D$. Thus, we set the range of $D$ from 2 to 5 to analyze the effect of the parameter $D$ on the performance using three evaluation metrics. The results are displayed in Table 1. It can be seen that when D is set with different values, all three indicators have a certain degree of change. Specifically, the values of SAM and SSIM remain basically the same. Compared with these two results, the values of PSNR increase significantly when $D < 5$. However, the value of each evaluation index has been decreased when $D$ is set to 5, especially for PSNR. In our view, there are two main reasons for this phenomenon. The one is the increase of network parameters caused by the use of more 3D convolution in the network. The other is that the network becomes deeper. These are not easy to train the network. In summary, we empirically set the parameter $D$ to 4 in our paper.

**Table 1.** Analysis of the influence of the number of the mixed convolutional module $D$ on the performance.

| **Evaluation Metrics** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|
| PSNR | 45.013 | 45.043 | 45.102 | 45.051 |
| SSIM | 0.9734 | 0.9736 | 0.9738 | 0.9735 |
| SAM | 2.247 | 2.244 | 2.241 | 2.243 |

### 4.4.2. Ablation Study Analysis

Table 2 shows the ablation study on the impacts of 2D unit (2U), local feature fusion (LFF), and global residual learning (GRL). We set the different combinations of components to analyze the performance of the proposed MCNet. To simply do fair comparison, our network with 4 modules is adopted to implement ablation investigation.

**Table 2.** Ablation study about the components.

| **Components** | **Different Combinations of Components** | | | | | |
|---|---|---|---|---|---|---|
| 2D unit (2U) | × | × | × | √ | × | √ |
| Local feature fusion (LFF) | × | √ | × | √ | √ | √ |
| Global residual learning (GRL) | × | × | √ | × | √ | √ |
| PSNR | 44.068 | 44.395 | 44.533 | 44.617 | 45.014 | 45.101 |
| SSIM | 0.9727 | 0.9729 | 0.9730 | 0.9734 | 0.9735 | 0.9738 |
| SAM | 2.451 | 2.325 | 2.318 | 2.312 | 2.283 | 2.241 |

First, there are no 2U, GRL, and LFF components, only 3D units are included in deep feature extraction (DFE) sub-network (the network is defined as baseline). It yields the worst performance. It mainly lacks adequate learning of effective features, which also shows that spectral and spatial features cannot be extracted well without these components. Thus, these components are required in our network. Then, we add one of these components to the baseline. The performance of the network is improved in PSNR and SAM. Accordingly, two of these components are added to the baseline. Evaluation indexes attain relatively better results than in previous evaluations. In short, the experiments demonstrate that each component can clearly enhance the performance of the network. This indicates that each component plays a key role in making the network easier to train. Finally, three components are attached to the baseline. The table shows that the results of three components are significantly better than the performance of only one or two, which reveals the effectiveness and benefits of the proposed components.

### *4.5. Comparisons with the State-of-the-Art Methods*

In this section, we adopt three public hyperspectral image datasets to evaluate the effectiveness of our MCNet with five existing SR methods. They are Bicubic [48], GDRRN [22], 3D-FCNN [26], EDSR [20], SSRNet [29]. Table 3 depicts the quantitative evaluation of state-of-the-art SR algorithms by average PSNR/SSIM/SAM for different scale factors.

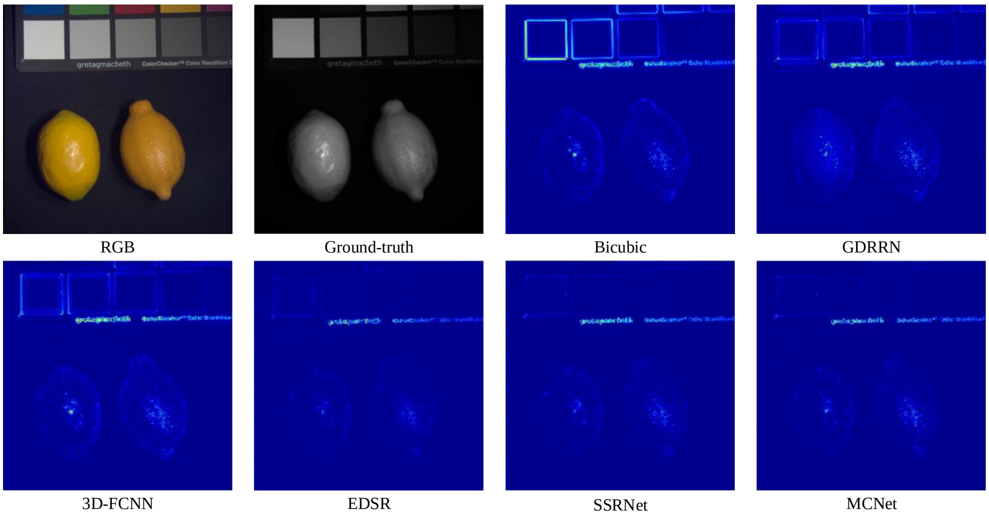
As shown in the table, our method can achieve better results than other algorithms using the CAVE dataset. Specifically, Bicubic produces the worst performance among these competitors. For the GDRRN algorithm, all the results are slightly higher than the worst Bicubic but lower than other methods. It is caused by the addition of a SAM item in the loss function. As a result, the network cannot optimize the difference between reconstructed and high-resolution image. Furthermore, the results of 3D-FCNN in PSNR and SSIM are lower than that of EDSR, but the performance in SAM of 3D-FCNN is obviously higher than that of EDSR, which is due to the fact that 3D-FCNN uses 3D convolution to extract the spectral features of hyperspectral image. Thus, this algorithm can void the spectral distortion of the reconstructed hyperspectral image well. However, the image obtained by 3D-FCNN lose part of the bands (the algorithm only obtains 23 bands on hyperspectral image with 31 bands), which is not suitable for image SR. For SSRNet algorithm, its results are better than that of the previous four methods. Compared with the existing SR approaches, our method obtains excellence performance. The proposed method is significantly superior to the scale factor ×4 of algorithm with the second performance (SSRNet) in terms of three evaluation metrics (+0.082 dB, +0.0007, and −0.005).

**Table 3.** Quantitative evaluation of state-of-the-art SR algorithms by average PSNR/SSIM/SAM for different scale factors. The red and blue indicate the best and second performance, respectively.

| **Scale Factor** | **Methods** | **CAVE** | **Harvard** | **Foster** |
|---|---|---|---|---|
| | | **PSNR/SSIM/SAM** | **PSNR/SSIM /SAM** | **PSNR/SSIM/SAM** |
| ×2 | Bicubic | 40.762/0.9623/2.665 | 42.833/0.9711/2.023 | 55.155/0.9981/4.391 |
| | GDRRN | 41.667/0.9651/3.842 | 44.213/0.9775/2.278 | 53.527/0.9963/5.634 |
| | 3D-FCNN | 43.154/0.9686/2.305 | 44.454/0.9778/1.894 | 60.242/0.9987/5.271 |
| | EDSR | 43.869/0.9734/2.636 | 45.480/0.9824/1.921 | 57.371/0.9978/5.753 |
| | SSRNet | 44.991/0.9737/2.261 | 46.247/0.9825/1.884 | 58.852/0.9987/4.064 |
| | MCNet | 45.102/0.9738/2.241 | 46.263/0.9827/1.883 | 58.878/0.9988/4.061 |
| ×3 | Bicubic | 37.532/0.9325/3.522 | 39.441/0.9411/2.325 | 50.964/0.9943/5.357 |
| | GDRRN | 38.834/0.9401/4.537 | 40.912/0.9523/2.623 | 50.464/0.9926/6.833 |
| | 3D-FCNN | 40.219/0.9453/2.930 | 40.585/0.9480/2.239 | 55.551/0.9958/6.303 |
| | EDSR | 40.533/0.9512/3.175 | 41.674/0.9592/2.380 | 52.983/0.9956/7.716 |
| | SSRNet | 40.896/0.9524/2.814 | 42.650/0.9626/2.209 | 54.937/0.9967/5.134 |
| | MCNet | 41.031/0.9526/2.809 | 42.681/0.9627/2.214 | 55.017/0.9970/5.126 |
| ×4 | Bicubic | 35.755/0.9071/3.944 | 37.227/0.9122/2.531 | 48.281/0.9880/5.993 |
| | GDRRN | 36.959/0.9166/5.168 | 38.596/0.9259/2.794 | 47.836/0.9877/7.696 |
| | 3D-FCNN | 37.626/0.9195/3.360 | 38.143/0.9188/2.363 | 52.188/0.9918/7.798 |
| | EDSR | 38.587/0.9292/3.804 | 39.175/0.9324/2.560 | 50.362/0.9915/7.103 |
| | SSRNet | 38.944/0.9312/3.297 | 40.001/0.9365/2.412 | 52.210/0.9939/5.702 |
| | MCNet | 39.026/0.9319/3.292 | 40.081/0.9367/2.410 | 52.225/0.9941/5.685 |

Similarly, the MCNet outperforms other competitors on the Harvard dataset, except for SAM. Concretely, unlike on CAVE dataset, GDRRN and 3D-FCNN achieved approximately the same results, because the number of hyperspectral images on augmented Harvard dataset is more than that on CAVE dataset. This is more beneficial to network training with many parameters, such as EDSR and SSRNet. Moreover, in most cases, it also enables our approach to achieve higher performance on this dataset for scale factor ×4. Likewise, the proposed approach achieves good performance in comparison to existing state-of-the-art methods on Foster dataset, particularly in SSIM and SAM.

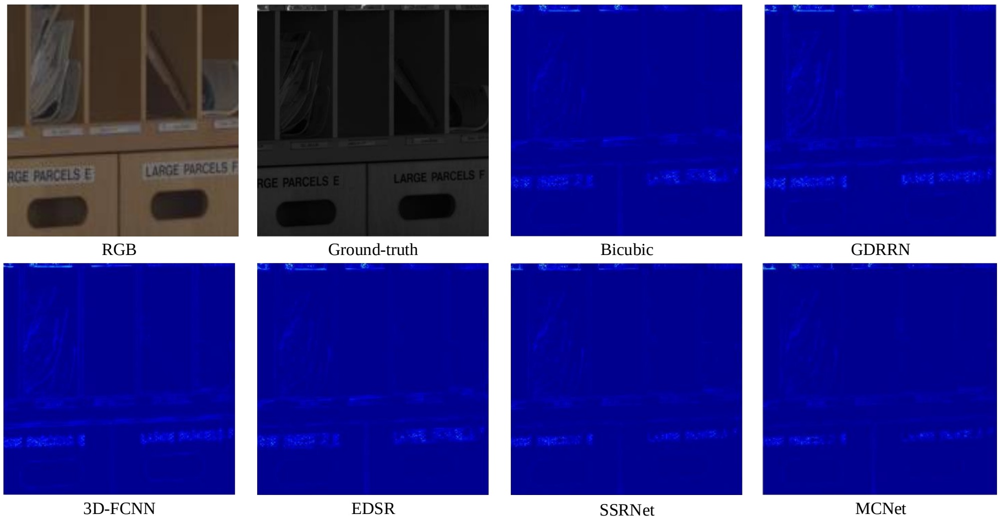
In Figures 7–9, we show visual comparisons with different algorithms for scale factor ×4 on three datasets. The figures only provide visual results of the 27-th band of three typical scenes. As revealed in the figure, the ground-truth is grey. So in order to observe the difference between reconstructed hyperspectral image and ground-truth clearly, the absolute error map between them is presented. In general, the bluer the absolute error map is, the better the reconstructed image is. Please note that each hyperspectral image is normalized. From these figures, we can see that the proposed MCNet obtains very low absolute error results. In some regions, especially for the edges of the image, our method generates shallow edge information or no edge information. It means our proposed MCNet generates more realistic visual results compared with other methods, which is consistent with our analysis in Table 3.



RGB Ground-truth Bicubic GDRRN

3D-FCNN EDSR SSRNet MCNet

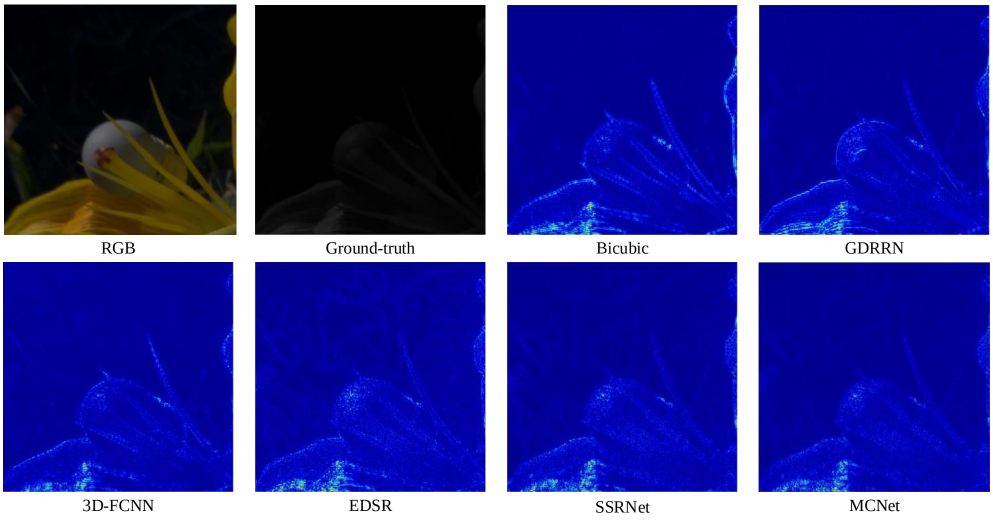
**Figure 7.** Absolute error map comparisons for image *fake_and_real_lemons* on CAVE dataset.



RGB Ground-truth Bicubic GDRRN

3D-FCNN EDSR SSRNet MCNet

**Figure 8.** Absolute error map comparisons for image *imgd5* on Harvard dataset.

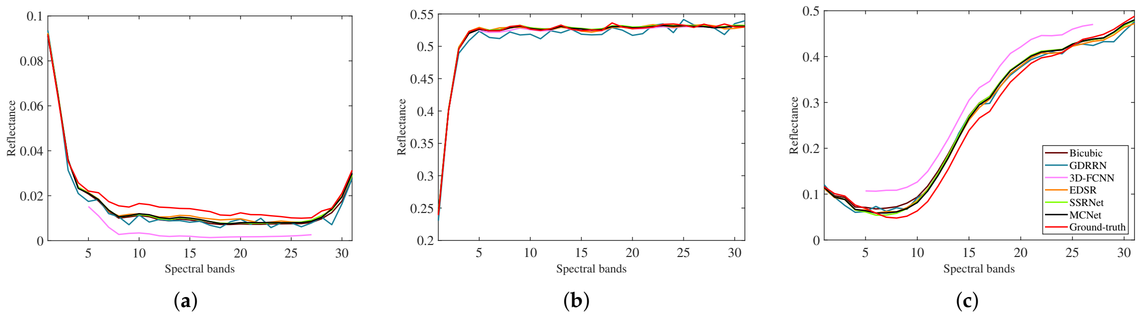


RGB Ground-truth Bicubic GDRRN

3D-FCNN EDSR SSRNet MCNet

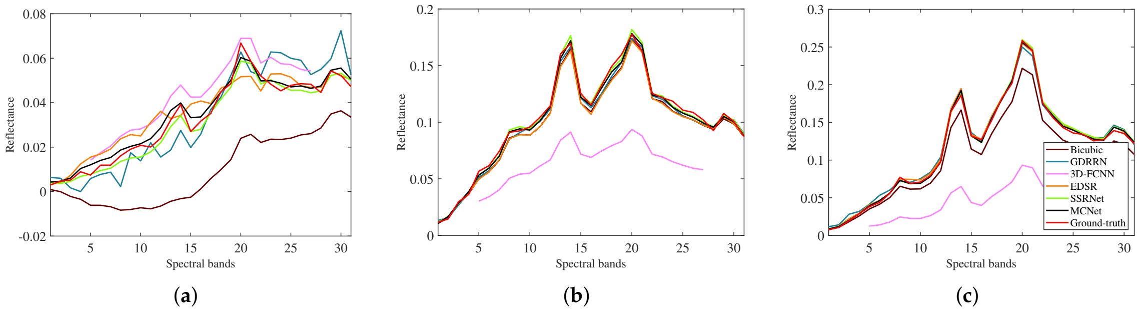
**Figure 9.** Absolute error map comparisons for image *Bom_Jesus_Bush* on Foster dataset.

We also visualize the spectral distortion of the reconstructed image by drawing spectral curves for three scenes, which are presented in Figures 10–12. Since all convolutions are not padding during reconstruction for 3D-FCNN, the actual output of the network is smaller than the input. We only show some of bands for this algorithm. To alleviate the problem caused by random selection, we selected three pixel positions ((20, 20), (100, 100), and (340, 340)) to analyze the distortion of the spectrum. As shown in Figure 10, the spectral curves of all competitors are basically consistent with that of ground-truth for image (*fake_and_real_lemons*). With respect to two images (*imgd5* and *Bom_Jesus_Bush*) in Figures 11 and 12, it can be seen from these figures that the distortion for 3D-FCNN is the most severe. The distortion of the spectral curve obtained by Bicubic is relatively small compared with 3D-FCNN. Moreover, the curves of other methods have certain deviation from the corresponding ground-truth. However, the results of our method are much closer to the ground-truth in most cases, which proves that our algorithm attains higher spectral fidelity. Of course, in order to show more clearly the spectral degree of three pixel positions, we also show the spectral distortion comparisons in three scenes by calculating SAM (see Table 4). As displayed in this table, the values of SAM in our method are better than that of other algorithms in most cases. In summary, MCNet does not just outperform state-of-the-art SR algorithms through quantitative evaluation, but also yields more realistic visual results.



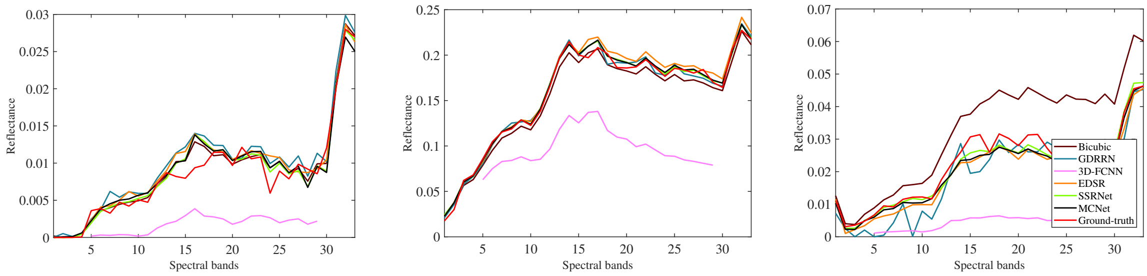
(**a**) (**b**) (**c**)

**Figure 10.** Visual comparison of spectral distortion for image *fake_and_real_lemons* on CAVE dataset. (**a**–**c**) Results of the spectral curve in pixel position (20, 20), (100, 100), and (340, 340).



(**a**) (**b**) (**c**)

**Figure 11.** Visual comparison of spectral distortion for image *imgd5* on Harvard dataset. (**a**–**c**) Results of the spectral curve in pixel position (20, 20), (100, 100), and (340, 340).



(**a**) (**b**) (**c**)

**Figure 12.** Visual comparison of spectral distortion for image *Bom_Jesus_Bush* on Foster dataset. (**a**–**c**) Results of the spectral curve in pixel position (20, 20), (100, 100), and (340, 340).

**Table 4.** Spectral distortion comparisons in three scenes by SAM. The red and blue indicate the best and second performance, respectively.

| Image | Pixel Position | Bicubic | GDRRN | 3D-FCNN | EDSR | SSRNet | MCNet |
|---|---|---|---|---|---|---|---|
| fake_and_real_lemons | (20,20) | 8.250 | 9.922 | 34.660 | 5.996 | 7.047 | 6.753 |
| | (100,100) | 0.220 | 0.756 | 0.246 | 0.322 | 0.227 | 0.222 |
| | (340,340) | 3.537 | 3.530 | 6.204 | 2.942 | 3.126 | 2.777 |
| imgd5 | (20,20) | 0.547 | 1.120 | 0.659 | 0.517 | 0.531 | 0.545 |
| | (100,100) | 0.829 | 1.230 | 0.814 | 0.912 | 0.832 | 0.813 |
| | (340,340) | 0.983 | 2.003 | 3.153 | 1.198 | 1.042 | 0.971 |
| Bom_Jesus_Bush | (20,20) | 6.810 | 8.137 | 18.503 | 8.334 | 7.846 | 8.390 |
| | (100,100) | 1.584 | 1.740 | 8.209 | 1.685 | 1.394 | 1.328 |
| | (340,340) | 6.590 | 12.134 | 5.632 | 5.288 | 4.376 | 4.339 |

### *4.6. Application on Real Hyperspectral Image*

In this section, we apply the MCNet to a real hyperspectral image dataset to demonstrate its applicability. The real hyperspectral images was collected by a progressive-scanning monochrome digital camera. This dataset (https://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral_images_of_natural_scenes_02.html Access date: 29 April 2020) has 30 scenes, such as rocks and trees [50]. The size of each scene is different, but there are still 31 bands in each scene. In our work, the images of eight representative scenes that are proved in [50] are used to demonstrate its applicability. Due to the limitation of hardware, we only use the top left 260 × 260 region of each hyperspectral image for evaluation.

Because there is no reference image for evaluation, some traditional evaluation metrics (such as, PSNR and SSIM) cannot be used here. Thus, the universal non-reference hyperspectral image quality evaluation methods (i.e., NIQE [51]) are adopted to evaluate the performance of the reconstruction. Generally, the higher values of NIQE mean a better visual quality. Table 5 shows the no-reference image quality assessment of existing SR methods. It can seen from the table that our method also achieves good results in real hyperspectral image dataset. This is consistent with our results in Table 3. It also demonstrates that the proposed algorithm has strong applicability. Since there is no reference image, absolute error map cannot be displayed. Therefore, we only provide visual results of the 27-th band in Figure 13. One can observe that our method generates better sharper edges and clearer structures than other algorithms.

**Table 5.** No-reference image quality assessment of state-of-the-art SR algorithms by average NIQE for different scale factors. The red and blue indicate the best and second performance, respectively.

| Scale Factor | Bicubic | GDRRN | 3D-FCNN | EDSR | SSRNet | MCNet |
|---|---|---|---|---|---|---|
| ×2 | 20.3403 | 20.5709 | 20.4427 | 20.8163 | 20.9145 | 20.9800 |
| ×3 | 20.1674 | 20.3097 | 20.3595 | 20.4087 | 20.3576 | 20.3797 |
| ×4 | 20.4322 | 20.5553 | 20.5113 | 20.5321 | 20.5714 | 20.5856 |

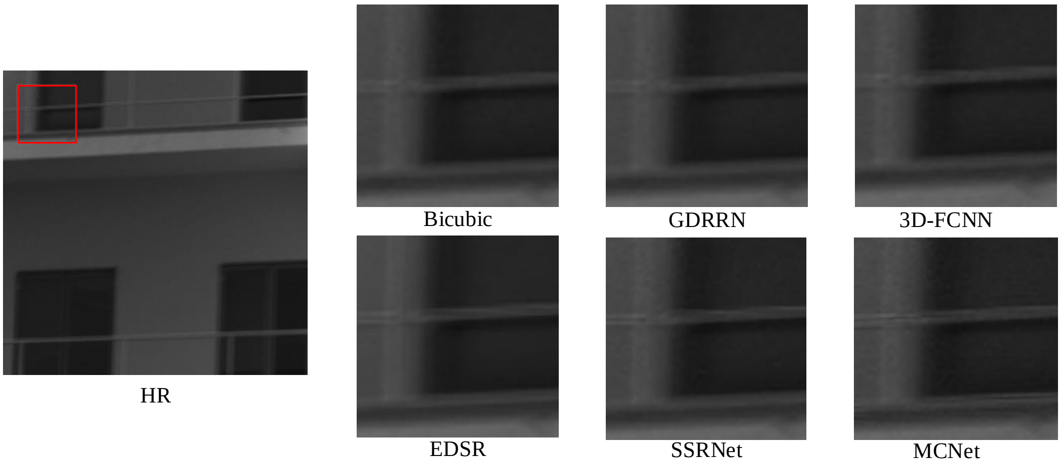


**Figure 13.** Visual comparison on real hyperspectral image dataset.

## 5. Discussions

In this section, we discuss the impact on the performance of the algorithm from the following two parts: loss function and study of 3D unit. Similarly, the experiments is implemented on CAVE dataset for scale factor ×2 to illustrate the influence.

### 5.1. Loss Function Analysis

To demonstrate the effect of different loss functions, the loss functions of [25,44], and L1 in our work are employed to train MCNet. The evaluation results are shown in Table 6. When adding SAM in loss function, it is clear that the spatial resolution has changed, and the spectral distortion has become more serious. Moreover, the loss function containing MSE and SAM gets a lower PSNR value, which is mainly due to the fact that the loss function weakens the performance of spatial resolution. As seen from this table, L1 in our paper can achieve the best performance than other loss functions for three indexes. It verifies our method can effectively optimize the difference between $I_{SR}$ and $I_{HR}$ using L1.

**Table 6.** Analysis of the influence for three loss functions.

| Loss Function | PSNR | SSIM | SAM |
|---|---|---|---|
| MSE | 44.980 | 0.9731 | 2.284 |
| L1 | 45.101 | 0.9738 | 2.241 |
| 0.5*MSE+0.5*SAM | 43.763 | 0.9704 | 2.346 |

### 5.2. Efficiency Study of 3D Unit

In this section, we study the efficiency of the proposed 3D unit using different types in module, including standard 3D convolution and separable 3D convolution. The one is that we use 3D unit with separable 3D convolution, the other is standard 3D convolution that has removed ReLU activation function. Please note that the convolution operations in initial feature extraction and image reconstruction are not replaced by separable 3D convolution in our network. The comparison results are shown in Table 7. Obviously, our proposed 3D unit can greatly reduce parameters, which can effectively reduce memory footprint. With respect to the results of PSNR, using standard 3D convolution is lower than that of separable 3D convolution. We think that there are too many parameters of the network, which makes the network more difficult to train, resulting in a decline in performance. Moreover, the training time of separable 3D convolution is lower than the standard 3D convolution, which mainly benefits from the reduction of the number of parameters. Generally speaking, the two methods are adopted to perform SR task, and the results of the proposed algorithms are approximately the same, expect for training time. This also verifies the effectiveness of the proposed algorithm when a small number of parameters are used.

**Table 7.** Comparison of the performance of standard 3D convolution and separable 3D convolution.

| Evaluation Metrics | Standard 3D Convolution | Separable 3D Convolution |
|---|---|---|
| Parameter | 3.16M | 1.93M |
| PSNR | 45.083 | 45.102 |
| SSIM | 0.9738 | 0.9738 |
| SAM | 2.240 | 2.241 |
| Training Time | 28h | 20h |

## 6. Conclusions

When the spectral information can be extracted, most existing models do not pay much attention to the mining of spatial information of hyperspectral images. To deal with this issue, in our paper, we develop a mixed 2D/3D convolutional network (MCNet) to reconstruct hyperspectral image, claiming the following contributions: (1) we propose a novel mixed convolutional module (MCM) to mine the potential features by 2D/3D convolution instead of one convolution; (2) To reduce the parameters for the designed network, we employ separable 3D convolution to extract spatial and spectral features respectively, thus reducing unaffordable memory usage; and (3) we design local feature fusion strategy to make full use of all the hierarchical features in each 2D unit after changing the size of feature maps. Extensive benchmark evaluations well demonstrate that our MCNet does not just outperform state-of-the-art SR algorithms, but also yields more realistic visual results.

In the future, we plan to improve in two aspects. First, in the mixed convolution module (MCM), the network does not effectively use the results of 2D unit, but only concatenates this information for analysis. Therefore, this can make full use of each 2D unit to optimize the structure of the network. Second, comparing with 2D convolution, the use of 3D convolution still results in a significant increase the number of parameters. From this point of view, the network can increase more 2D units and reduce 3D units, thus effectively reducing the number of parameters.

**Author Contributions:** Methodology, Q.L.; Writing and original draft preparation, Q.L. and Q.W.; Writing, review, and editing, X.L. and Q.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Key R&D Program of China under Grant 2018YFB1107403, National Natural Science Foundation of China under Grant U1864204, 61773316, U1801262, and 61871470.

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

## References

1. Li, Q.; Wang, Q.; Li, X. An Efficient Clustering Method for Hyperspectral Optimal Band selection via Shared Nearest Neighbor. *Remote Sens.* **2019**, *11*, 350. doi:10.3390/rs11030350.
2. Sabins, F.F. Remote Sensing for Mineral Exploration. *Ore Geol. Rev.* **1999**, *14*, 157–183. doi:10.1016/S0169-1368(99)00007-4.

3. Lin, J.; Clancy, N.T.; Qi, J.; Hu, Y.; Tatla, T.; Stoyanov, D.; Maier-Hein, L.; Elson, D.S. Dual-modality Endoscopic Probe for Tissue Surface Shape Reconstruction and Hyperspectral Imaging Enabled by Deep Neural Networks. *Med. Image Anal.* **2018**, *48*, 162–176. doi:10.1016/j.media.2018.06.004.
4. Lowe, A.; Harrison, N.; French, A.P. Hyperspectral Image Analysis Techniques for the Detection and Classification of the Early Onset of Plant Disease and Stress. *Plant Methods* **2017**, *13*, 80. doi:10.1186/s13007-017-0233-z.
5. Wang, Q.; Yuan, Z.; Li, X. GETNET: A General End-to-end Two-dimensional CNN Framework for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3–13.
6. Wang, Q.; He, X.; Li, X. Locality and Structure Regularized Low Rank Representation for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 911–923. doi:10.1109/TGRS.2018.2862899.
7. Xie, W.; Jia, X.; Li, Y.; Lei, J. Hyperspectral Image Super-Resolution Using Deep Feature Matrix Factorizationk. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6055–6067. doi:10.1109/TGRS.2019.2904108.
8. Dong, W.; Fu, F.; Shi, G.; Gao, X.; Wu, J.; Li, G.; Li, X. Hyperspectral Image Super-Resolution via Non-Negative Structured Sparse Representation. *IEEE Trans. Image Process.* **2016**, *25*, 2337–2352. doi:10.1109/TIP.2016.2542360.
9. Akgun, T.; Altunbasak, Y.; Mersereau, R.M. Super-resolution Reconstruction of Hyperspectral Images. *IEEE Trans. Image Process.* **2005**, *14*, 1860–1875. doi:10.1109/TIP.2005.854479.
10. Hu, Y.; Li, J.; Huang, Y.; Gao, X. Channel-wise and Spatial Feature Modulation Network for Single Image Super-Resolution. *IEEE Trans. Circuits Syst. Video Technol.* **2019**. doi:10.1109/tcsvt.2019.2915238.
11. Qu, Y.; Qi, H.; Kwan, C. Unsupervised Sparse Dirichlet-Net for Hyperspectral Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 2511–2520. doi:10.1109/cvpr.2018.00266.
12. Kwon, H.; Tai, Y. RGB-guided Hyperspectral Image Upsampling. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 307–315. doi:10.1109/ICCV.2015.43.
13. Akhtar, N.; Shafait, F.; Mian, A.S. Hierarchical Beta Process with Gaussian Process Prior for Hyperspectral Image Super Resolution. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 103–120. doi:10.1007/978-3-319-46487-9_7.
14. Wycoff, E.; Chan, T.; Jia, K.; Ma, W.; Ma, Y. A Non-negative Sparse Promoting Algorithm for High Resolution Hyperspectral Imaging. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 1409–1413. doi:10.1109/ICASSP.2013.6637883.
15. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed Optimization and Statistical Learning via Alternating Direction Method of Multipliers. *Found. Trends® Mach. Learn.* **2011**, *3*, 1–122.
16. Fu, Y.; Zhang, T.; Zheng, Y.; Zhang, D.; Huang, H. Hyperspectral Image Super-Resolution with Optimized RGB Guidance. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. doi:10.1109/CVPR.2019.01193.
17. Anwar, S.; Khan, S.; Barnes, N. A Deep Journey into Super-Resolution: A survey. *arXiv* **2019**, arXiv:1904.07523.
18. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Super-Resolution. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481. doi:10.1109/cvpr.2018.00262.
19. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning A Deep Convolutional Network for Image Super-Resolution. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 184–199. doi:10.1007/978-3-319-10593-2_13.
20. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140. doi:10.1109/CVPRW.2017.151.
21. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z. Photo-realistic Single Image Super-Resolution Using A Generative Adversarial Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.

22. Li, Y.; Zhang, L.; Ding, C.; Wei, W.; Zhang, Y. Single Hyperspectral Image Super-Resolution with Grouped Deep Recursive Residual Network. In Proceedings of the 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), Xi'an, China, 13–16 September 2018; pp. 1–4. doi:10.1109/bigmm.2018.8499097.
23. Yuan, Y.; Zheng, X.; Lu, X. Hyperspectral Image Superresolution by Transfer Learning. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *10*, 1963–1974. doi:10.1109/JSTARS.2017.2655112.
24. Jia, J.; Ji, L.; Zhao, Y.; Geng, X. Hyperspectral Image Super-Resolution with Spectral–Spatial Network. *Proc. Int. J. Remote Sens.* **2018**, *39*, 7806–7829. doi:10.1080/01431161.2018.1471546.
25. Li, R.; Hu, J.; Zhao, X.; Xie, W.; Li, J. Hyperspectral Image Super-Resolution Using Deep Convolutional Neural Network. *Neurocomputing* **2017**, *266*, 29–41. doi:10.1016/j.neucom.2017.05.024.
26. Mei, S.; X. Yuan, J.J.; Zhang, Y.; Wan, S.; Du, Q. Hyperspectral Image Spatial Super-Resolution via 3D Full Convolutional Neural Network. *Remote Sens.* **2017**, *9*, 1139. doi:10.3390/rs9111139.
27. Yang, J.; Zhao, Y.; Chan, J.C.; Xiao, L. A Multi-Scale Wavelet 3D-CNN for Hyperspectral Image Super-Resolution. *Remote Sens.* **2019**, *11*, 1557. doi:10.3390/rs11131557.
28. Li, J.; Cui, R.; Li, Y.; Li, B.; Du, Q.; Ge, C. Multitemporal Hyperspectral Image Super-Resolution through 3D Generative Adversarial Network. In Proceedings of the 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Shanghai, China, 5–7 August 2019; pp. 1–4. doi:10.1109/Multi-Temp.2019.8866956.
29. Wang, Q.; Li, Q.; Li, X. Spatial-Spectral Residual Network for Hyperspectral Image Super-Resolution. *arXiv* **2020**, arXiv:2001.04609.
30. Li, J.; Cui, R.; Li, B.; Li, Y.; Du, S.M.Q. Dual 1D-2D Spatial-Spectral CNN for Hyperspectral Image Super-Resolution. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2019), Yokohama, Japan, 28 July–2 August 2019; pp. 3113–3116. doi:10.1109/IGARSS.2019.8898352.
31. He, Z.; Lin, L. Hyperspectral Image Super-Resolution Inspired by Deep Laplacian Pyramid Network. *Remote Sens.* **2018**, *10*, 1939. doi:10.3390/rs10121939.
32. Tai, Y.; Yang, J.; Liu, X. Image Super-Resolution via Deep Recursive Residual Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3147–3155. doi:10.1109/CVPR.2017.298.
33. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning Spatiotemporal Features with 3D Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4489–4497.
34. Jiang, R.; Li, X.; Gao, A.; Li, L.; Meng, H.; Yue, S.; Zhang, L. Learning Spectral and Spatial Features Based on Generative Adversarial Network for Hyperspectral Image Super-Resolution. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2019), Yokohama, Japan, 28 July–2 August 2019; pp. 3161–3164. doi:10.1109/IGARSS.2019.8900228.
35. 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. doi:10.1109/TPAMI.2015.2439281.
36. Kappeler, A.; Yoo, S.; Dai, Q.; Katsaggelos, A.K. Video Super-Resolution With Convolutional Neural Networks. *IEEE Trans. Comput. Imaging* **2016**, *2*, 109–122. doi:10.1109/TCI.2016.2532323.
37. Wang, Q.; Li, Q.; Li, X. Hyperspectral Band Selection via Adaptive Subspace Partition Strategy. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2019**. doi:10.1109/JSTARS.2019.2941454.
38. Tran, D.; Wang, H.; Torresani, L.; Feiszli, M. Video Classification with Channel-Separated Convolutional Networks. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 5552–5561. doi:10.1109/ICCV.2019.00565.
39. 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. doi:10.1109/TPAMI.2012.59.
40. Xie, S.; Sun, C.; Huang, J.; Tu, Z.; Murphy, K. Rethinking Spatiotemporal Feature Learning: Speed-accuracy Trade-offs in Video Classification. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2018; pp. 318–335. doi:10.1007/978-3-030-01267-0_19.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. doi:10.1109/CVPR.2016.90.

42. Huang, G.; Liu, Z.; Maaten, L.V.D.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708. doi:10.1109/CVPR.2017.243.
43. Kim, J.; Lee, J.K.; Lee, K.M. Accurate Image Super-Resolution Using very Deep Convolutional Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654. doi:10.1109/CVPR.2016.182.
44. Wang, C.; Liu, Y.; Bai, X.; Tnag, W.; Lei, P.; Zhou, J. Deep Residual Convolutional Neural Network for Hyperspectral Image Super-Resolution. In *International Conference on Image and Graphics*; Springer: Cham, Switzerland, 2017; pp. 370–380. doi:10.1007/978-3-319-71598-8_33.
45. Yasuma, F.; Mitsunaga, T.; Iso, D.; Nayar, S.K. Generalized Assorted Pixel Camera: Postcapture Control of Resolution, Dynamic Range, and Spectrum. *IEEE Trans. Image Process.* **2010**, *19*, 2241–2253. doi:10.1109/TIP.2010.2046811.
46. Chakrabarti, A.; Zickler, T. Statistics of Real-world Hyperspectral Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011), Providence, RI, USA, 20–25 June 2011; pp. 193–200. doi:10.1109/CVPR.2011.5995660.
47. Nascimento, S.M.C.; Amano, K.; Foster, D.H. Spatial Distributions of Local Illumination Color in Natural Scenes. *Vis. Res.* **2016**, *120*, 39–44. doi:10.1016/j.visres.2015.07.005.
48. Miller, F.P.; Vandome, A.F.; Mcbrewster, J. *Bicubic Interpolation*; Alphascript Publishing: Riga, Latvia, 2010.
49. Yu, J.; Fan, Y.; Yang, J.; Xu, N.; Wang, Z.; Wang, X.; Huang, T. Wide Activation for Efficient and Accurate Image Super-Resolution. *arXiv* **2018**, arXiv:1808.08718v2.
50. Nascimento, S.M.C.; Ferreira, F.P.; Foster, D.H. Statistics of Spatial Cone-Excitation Ratios in Natural Scenes. *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* **2002**, *19*, 1484–1490.
51. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. doi:10.1109/LSP.2012.2227726.

© 2020 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/).