Multimedia Tools and Applications
https://doi.org/10.1007/s11042-018-6850-3

FusionCNN: a remote sensing image fusion algorithm based on deep convolutional neural networks

Fajie Ye[1,2] · Xiongfei Li[1,2] · Xiaoli Zhang[1,2]

Received: 22 May 2018 / Revised: 31 October 2018 / Accepted: 6 November 2018
Published online: 13 November 2018

Abstract
In remote sensing image fusion field, traditional algorithms based on the human-made fusion rules are severely sensitive to the source images. In this paper, we proposed an image fusion algorithm using convolutional neural networks (FusionCNN). The fusion model implicitly represents a fusion rule whose inputs are a pair of source images and the output is a fused image with end-to-end property. As no datasets can be used to train FusionCNN in remote sensing field, we constructed a new dataset from a natural image set to approximate MS and Pan images. In order to obtain higher fusion quality, low frequency information of MS is used to enhance the Pan image in the pre-processing step. The method proposed in this paper overcomes the shortcomings of the traditional fusion methods in which the fusion rules are artificially formulated, because it learns an adaptive strong robust fusion function through a large amount of training data. In this paper, Landsat and Quickbird satellite data are used to verify the effectiveness of the proposed method. Experimental results show that the proposed fusion algorithm is superior to the comparative algorithms in terms of both subjective and objective evaluation.

Keywords Remote sensing image fusion · Convolutional neural networks · Deep learning · Image enhancement

1 Introduction

In recent years, remote sensing images have been widely used in various applications such as environmental management and detection, geological hazard prevention, precision agriculture, and national defense security. Due to the limitations of satellite sensors, we can only obtain

✉ Xiaoli Zhang
zhangxiaoli@jlu.edu.cn

[1] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China

[2] College of Computer Science and Technology, Jilin University, Changchun 130012, China

multispectral images (MS) with high spectral and low spatial resolution or panchromatic images (Pan) with high spatial and low spectral resolution. However, in practical applications, we need to use both high-spectrum and high-spatial resolution information. High spectral resolution is used for accurate object classification, and high spatial resolution is used for accurate descriptions of the texture and shapes [8, 11]. The purpose of remote sensing image fusion is to produce images with both high spatial and high spectral resolution by combing the MS and Pan images [15, 34].

In many remote sensing image fusion algorithms, fusion rules are designed based on single pixel. However, the neighboring pixels in images are spatially related. If neighborhood information is ignored, spatial distortion may be generated in the fused image. In addition, the calculation of the contribution degree is based on artificially created fusion rules, and the final fusion quality is also changed by the fusion rules. These fusion rules have strong dependence on the fused image itself, and the same method has different fusion qualities on different remote sensing images. The fusion rule can be regarded as a fusion function whose inputs are Pan and MS images and output is a fused image. Finding a suitable fusion function is the key for remote sensing image fusion because it directly determines the quality of fused images.

Since it has been introduced in the 1940s, neural networks have experienced two ups and downs for various reasons. Until 2006, Hindon et al. used a “greedy layer-wise pretraining” strategy to successfully train a neural network, which is called a deep belief network [14]. Since then, neural networks have returned in the name of deep learning. Due to the significant progress in theoretical research, the dramatic increase of hardware computing capabilities, and the ever-increasing amount of data, deep learning has obtained significant achievements in computer vision [18, 21], natural language processing [6, 23], machine translation [29], et al.

In this paper, we propose a method that uses the properties of remote sensing image fusion and the merits of convolutional neural networks to create a multi-scale convolutional neural network model, which is then used to fuse Pan and MS images. We first construct a Convolutional Neural Networks for image fusion (FusionCNN), which is a typical regression model. As no dataset can be used to train FusionCNN in remote sensing field, we construct a new dataset from a natural image set by simulating PAN and MS images. In order to obtain better fusion quality, we use the low frequency information of MS to enhance the Pan image, which is denoted as EPAN. The MS and EPAN images are input into the trained FusionCNN, so that fused images can be obtained as the output of the network. The main advantages of the proposed method are as follows: (1) The convolutional neural network is used to regressively learn a complex fusion rule with high robustness and good fusion quality. The final fused image can well preserve both the spectral information of MS images and the spatial information of Pan images; (2) The fusion procedure is simple: the first step is to convert PAN to EPAN, and then input EPAN and MS image to the fusion model to get the final fused image.

The remainder of the paper is organized as follows: the related works are shown in Section 2; in Section 3, we describe the CNNs-based remote sensing image fusion method used in this paper; the construction of the training set, the details of the model, and model training is introduced in Section 4; in Section 5, we present and analyze the experimental results of FusionCNN and other five comparative algorithms; concluding remarks are given in Section 6.

2 Related work

The existing remote sensing image fusion methods are generally classified into three categories: component substitution (CS), multi-resolution analysis (MRA) and hybrid methods.

The CS methods are based on a reversible transform, by which the MS images are projected into another color space. The transform can make the spectral structure of the MS image separated from the spatial structure. The separated spatial components are replaced with the Pan image after histogram matching. Finally, inverse transform is used to convert the data to the original space to get final fused images. The IHS (Intensity hue saturation) method can keep spectral information in components H and S while keeping most of the spatial information in component I [5]. Tu et al. proposed a fast IHS fusion method, which can extend the traditional third-order transform to any order with high computational efficiency [31]. Similar to the IHS method, the fast IHS algorithm is also liable to spectral distortions. To overcome these problems, González-Audícana et al. used the minimization algorithm to improve the fast IHS method [13]. However, this method is not efficient enough for a large amount of remote sensing data. PCA (Principal component analysis) is a statistical technique which converts cross-correlated multivariate data into uncorrelated variables, and then replaces the first principal component image with the Pan image. Shahdoosti et al. proposed a hybrid algorithm combining spectral PCA and spatial PCA [27]. The Brovey transform normalizes the three bands first, and then multiplies the result by the expected data to obtain fused images [30].

In MRA based methods, source images are decomposed by multi-scale decomposition algorithm, then fusion rules are applied to different scale levels of the original images, and finally the fused image is synthesized by the inverse transform [35]. The multi-scale analysis algorithms used in image fusion include Laplacian pyramid decomposition [2], wavelet transforms [28], contourlet transforms [7, 25], curvelet transforms [4, 10, 16], etc. The adaptive high-pass filter (HPFA) method gets fused images by injecting the structure and texture details of high-resolution image into low-resolution images. Gangkofner et al. improved this method to be an adjustable and standardized image fusion method [9]. Shahdoosti et al. proposed an optimization filter that can extract relevant and non-redundant information from Pan images [26]. Compared with other methods, the optimal filter coefficients extracted from the statistical properties of images are more consistent with the types and textures of remote sensing images.

The hybrid methods combine the advantages of CS and MRA [22]. Valizadeh et al. proposed a remote sensing image fusion method using IHS and curvelet transform [32]. To overcome the spectral distortion of PCA and the spatial resolution reduction of wavelet transforms, Luo et al. proposed a fusion method based on the PCA method and the wavelet transform [20]. Cheng et al. proposed a fusion framework combining wavelet transform and sparse representation, and obtained relatively satisfactory experimental results [3].

The CS methods can usually well preserve spatial information of Pan images in fused images, and they are simple to implement. However, they do not consider the local differences between Pan and MS images, resulting in significant spectral distortion in the final fused images. In multi-scale analysis, the number of image decomposition stages and the filters used will have a great impact on fusion results. Typical multi-scale analysis methods such as wavelet transforms have significant spatial information distortion. Although the hybrid methods combine CS and MRA, the final fused images still have different degrees of spectral distortion and spatial structure distortion. There is a great correlation between the fusion quality and the specific fusion methods used.

3 Remote sensing image fusion algorithm

In this paper, we propose a remote sensing image fusion algorithm based on convolutional neural network. The key issue in this algorithm is the network construction. In this section, the CNN for image fusion is presented first, and then we detail the fusion rule, after which the framework of the proposed image fusion algorithm is summarized.

In the algorithm, the input contains a PAN image and a MS image, and they are denoted as PAN and MS. In addition, we assume that the two images are well denoised and registered.

3.1 Convolutional neural network for fusion (FusionCNN)

In remote sensing image fusion, suppose $F(i,j)$ is the pixel of the fused image at (i,j), $PAN(i,j)$ and $MS(i,j)$ are the corresponding pixels of the Pan and MS images, respectively. To calculate $F(i,j)$ reasonably, we must first consider the neighboring pixels of $PAN(i,j)$ and $MS(i,j)$, which are denoted as $NP(i,j)$ and $NM(i,j)$. Then, the fused pixel $F(i,j)$ is calculated by $NP(i,j)$ and $NM(i,j)$. In the algorithm, we propose to use the convolution operation of CNN to calculate the NP and NM, and further get the fused image F. The pixels in NP and NM are obtained by convolving image PAN and MS with kernels in size of $N \times N$, and the fused image F is obtained by convolving NP and NM with a kernel with the size of 1×1. The pure convolution operation is a linear operation. In order to improve the quality of the fused images, a nonlinear operation is performed after the convolution. The final convolution operation is defined as

$$F = \mathrm{ReLU}(X \circ w) \tag{1}$$

where X is the convolution input, w is the convolution kernel, and ReLU is a nonlinear activation function:

$$\mathrm{ReLU}(x) = \max\{0, x\} \tag{2}$$

The FusionCNN model is shown in Fig. 1. The inputs of the FusionCNN are a pair of source images with 3 channels, and the output of the FusionCNN is a color fused image. One attractive feature of the network is that it can handle arbitrary-size images. The FusionCNN only contains 10 convolution layers, among which 6 ones are in size of $N \times N(N > 1)$ and the rest is 1×1. We use this network to implicitly represent the fusion function $F(MS, PAN) \rightarrow F$. Let $MS_k(k = 1,2,3)$ and $PAN_k(k = 1,2,3)$ represent the output of the kth convolution layer of images MS and PAN

$$MS^k(i,j,c) = \mathrm{ReLU}\left(MS^{k-1}(i,j,c) \circ w^k_{MS(c)}\right) \tag{3}$$

$$PAN^k(i,j,c) = \mathrm{ReLU}\left(PAN^{k-1}(i,j,c) \circ w^k_{PAN(c)}\right) \tag{4}$$

where c is the index of channels, $w^k_{MS(c)}$ and $w^k_{PAN(c)}$ are the convolution kernels. Here, we set $MS^0 = MS$, $PAN^0 = PAN$.

In FusionCNN, let $PANMS_1$ denote the fused image of images PAN and MS by weighted average, and it is calculated as follows: the images PAN and MS are first combined to construct

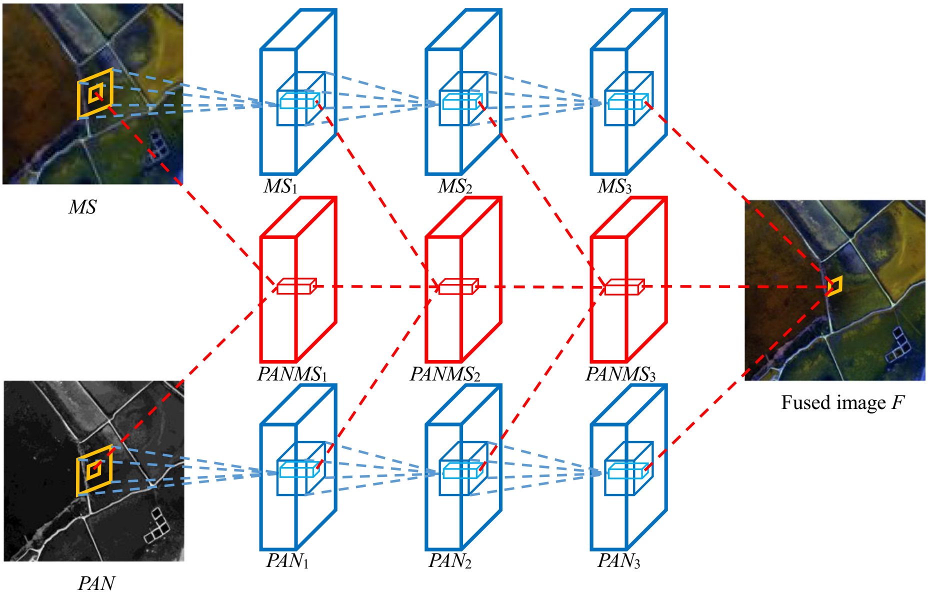


Fig. 1 Fusion model FusionCNN

a new image *PANMS* with the channel of six; then a 1 × 1 convolution calculation is conducted.

$$PANMS_1(i,j,c) = w(PAN,c)PANMS(i,j,c) + w(MS,c)PANMS(i,j,c+3) \tag{5}$$

$PANMS_2$,$PANMS_3$ and the fused image F can be obtained in a similar way of $PANMS_1$. From Fig. 1, it can be observed that the inputs contain three rather than two images when calculating $PANMS_2$, $PANMS_3$ and F. Taking $PANMS_2$ as an example, the images *PAN, MS* and $PANMS_1$ are first combined to construct a new image *PANMS*, and then a 1 × 1 convolution calculation is conducted.

$$PANMS_1(i,j,c) = \sum_{l=0}^{2} w_{PAN(3l+c)} PANMS(i,j,(3l+c)) \tag{6}$$

Each level of fusion takes the result of the upper level fusion into account, and we can regard it as a special kind of multi-scale fusion method. The final fused image F is the result of merging fused images at different scales.

For each pixel in the final fused image, in order to increase its corresponding pixel area in the original image (increase the fusion scale), we can choose to increase the size of the convolution kernels or deepen the network model. Compared with increasing the size of convolution kernels, deepening the network can effectively improve the fusion ability of the fusion model.

3.2 The input of FusionCNN

The core idea of the algorithm is to use a convolutional neural network to train a regression model. Using the trained regression model, we fuse the panchromatic image

PAN and the multi-spectral image MS together to obtain final fused image. The fused image combines the spatial information of the Pan image with the spectral information of the MS image.

However, we find that MS images also contain some texture information. If a pair of MS and PAN images is input into the FusionCNN, the fused images tend to result in reduction of contrast in fused images. Hence, we enhance the PAN image with the texture information from MS image before inputting them into the FusionCNN. Considering that the texture of the MS image is used, MS and PAN are first transformed into HLS color space, and their L channel is denoted as L_{MS} and L_{MS}. Then, L_{MS} is up sampled to the same size with PAN. Finally, L_{MS} is injected into the image PAN. The algorithm of the last step is as follows:

a) L_{MS} and L_{PAN} are decomposed into N sub-bands using Non-Subsampled Laplacian Pyramid (NSLP) decomposition:

$$L_{MS} \overset{NSLP}{\rightarrow} \left\{L_{MS}^{0}, L_{MS}^{1}, \ldots, L_{MS}^{N-1}\right\} \tag{7}$$

$$L_{PAN} \overset{NSLP}{\rightarrow} \left\{L_{PAN}^{0}, L_{PAN}^{1}, \ldots, L_{PAN}^{N-1}\right\} \tag{8}$$

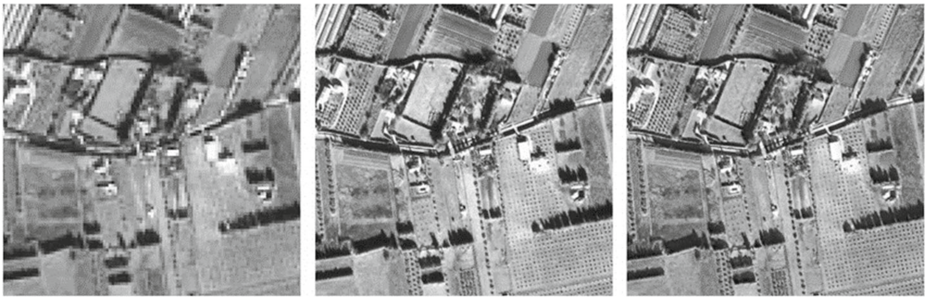
where L_{MS}^{0} and L_{PAN}^{0} are the lowpass sub-bands of L_{MS} and L_{PAN}; $L_{MS}^{i}(1 < i < N)$ and L_{PAN}^{i} $(1 < i < N)$ represent the ith highpass sub-bands of L_{MS} and L_{PAN}.

b) Construct the sub-bands of the L channel of the enhanced PAN (EPAN) in Non-Subsampled Laplacian Pyramid space by combining the lowpass sub-band of L_{MS} with the highpass sub-bands of L_{PAN}:$\left\{L_{MS}^{0}, L_{PAN}^{1}, \ldots, L_{PAN}^{N-1}\right\}$.

c) Reconstruct L channel of EPAN by performing the inverse NSLP transform

$$L_{PAN} \overset{iNSLP}{\leftarrow} \left\{L_{MS}^{0}, L_{PAN}^{1}, \ldots, L_{PAN}^{N-1}\right\} \tag{9}$$

Once we get the L channel of EPAN, we transform it together with the H and S channel of PAN into RGB space to obtain the enhanced PAN image EPAN. Figure 2 shows an example of PAN image enhancement.



(a) L_{MS} (b) PAN (c) EPAN

Fig. 2 An example of PAN image enhancement

3.3 Framework of the fusion algorithm

The framework of the proposed algorithm based on convolutional neural network is shown in Fig. 3. The detail steps of the algorithm can be presented as follows:

Step1 Train fusion model *FusionCNN* on the selected training dataset;
Step2 Obtain the enhanced Pan image *EPAN*;
Step3 The MS is up sampled to the *EPAN* resolution using bicubic interpolation. The upsampled *MS* and *EPAN* are input into the trained fusion model to get the final fused image *F*.

4 Training dataset, model details and model training

4.1 Training dataset

The starting point of this paper is to establish a fusion function $F(MS, EPan) = Fusion$ using a regression model. But solving this regression function requires a large amount of training data. For this reason, we need to construct a suitable training data set.

In actual satellite remote sensing images, there are no groundtruth images, for which regression targets for the fusion model do not exist. We can take the original MS images as groundtruth images, and down-sampled MS and PAN images as the two inputs of the network. However, such operation can hardly train a satisfactory network because of the following three facts: (1) The color of remote sensing images is relatively simple. The trained model has poor fusion quality when dealing with new remote sensing images. In other words, the model does not have strong generalization ability; (2) Remote sensing image resolution is very low, and there is little difference among each pixel and its neighborhoods, that is, there is less texture information. Hence, it is difficult to train the fusion model; (3) Remote sensing image acquisition is relatively expensive. The fusion model can’t be trained very well with a small amount of training data. In order to tackle this problem, we try to construct a reasonable training data set from natural image set. For any image *I*, it is regarded as a groundtruth image *F*, while its corresponding low-resolution image is used as an MS image. We regard the *L* component of the image *I* in the HLS space as *EPAN*, the other input of FusionCNN.

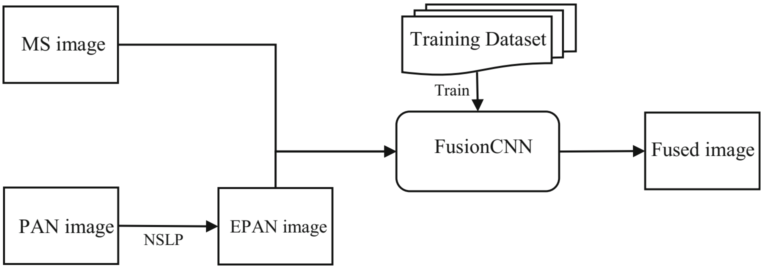


Fig. 3 Framework of the proposed algorithm

In this paper, the natural images come from the CIFAR data set. CIFAR includes CIFAR10 and CIFAR100, each containing 60,000 real images with the size of 32 × 32. It is a standard dataset for image recognition tasks in the field of computer vision. This dataset is selected due to the following reasons: (1) There are many categories in the CIFAR dataset, in which images are diversified and the amount of data is relatively large; (2) The image size is 32 × 32, which is exactly suitable for the input image size of the fusion model used in this paper; (3) The images in the CIFAR dataset are captured from real world and then reduced to 32 × 32. Though the images are very blurry, they contain as much information as possible of the original images, especially strong texture features. Hence, they are suitable for simulating MS and PAN images. Figure 4 shows some of the constructed training samples. In CIFAR dataset, 1000 images are extracted for model test, and all the remaining images are used for model training. In order to avoid losing more spectral information, we only down-sample the original image (bilinear interpolation) to half of the original size, and then up-sample (bicubic interpolation) it to the original size of the MS image.

4.2 Model details

Considering that great kernel size and deep network will weaken the fusion performance and increase the computational costs, we the convolution kernel size N to 3, and the depth of the network to 4 in the FusionCNN model. That is, the pixels in the fused image are calculated according to the 1, 3, 5, and 7 neighborhood pixels of the corresponding original image pixels. The number of feature maps of MS_1, MS_2, and MS_3 are 32, 64, and 128, respectively. PAN_1, PAN_2, and PAN_3 are the same as above. The number of fused feature maps, $PANMS_1$, $PANMS_2$, and $PANMS_3$, are 32, 64, and 128, respectively. The final fused image F has three feature maps (R, G, B channels). It should be noted that three 1 × 1 convolution layers are used, with the same number of feature maps, when generating fused feature maps. Similarly, two convolution layers (3 × 3 and 1 × 1) are used when generating MS_i and PAN_i.

In order to eliminate the influence of boundary, we do not pad zeros around the feature maps at each convolution layer. Zeros padding is performed directly around the Pan image and MS image, and the original size 32 × 32 is increased to 38 × 38. After being acted upon by

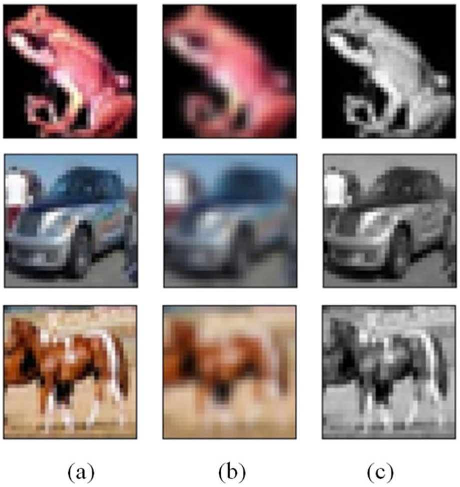


Fig. 4 Constructed training samples. **a** original image, **b** Low resolution version of (**a**), and **c** L channel of (**a**) in HLS space

three 3 × 3 convolution layers, the size of the fused image is 32 × 32. With this design, the model can adapt any size ($M \times N$) remote sensing image. The last 1 × 1 convolutional layer, which generates the final fused image F, does not use an activation function. The other convolutional layers in the network use the ReLU activation function in Eq. (2).

4.3 Model training

Since this paper uses a regression model to train an implicit fusion function $Fusion = F(EPan, MS)$, we use the mean squared error as the loss function, which is defined as follows:

$$L(\theta) = \frac{1}{n} \sum_{\mathrm{i}=1}^{n} \left\| I - F\left(\theta; EPAN, MS\right) \right\|^2 \tag{10}$$

where I is the grouthtruth image (the original image in the training set); $EPAN$ is enhanced panchromatic image; MS is multispectral image (low resolution image); $F(\theta; EPAN, MS)$ is the model output fused image and n is the number of training samples. We solve the fusion function F by minimizing L. In addition, the pixel value range of the image is 0–255, which is normalized to interval of [0, 1] before being input into the model.

The model is trained on Nvidia GTX1080 Ti GPU and the optimization algorithm is Adam algorithm, which is an adaptive learning rate optimization algorithm based on stochastic gradient descent. We train the model for 12 epochs with a mini-batch size of 128. The initial learning rate is set to 0.001, and is divided by 10 at 50% and75% of the total number of training epochs. The total training time is 40 min, while the model's final training mean square error is 0.00012. Figure 5 shows some fused results of the trained model on the Ciafr test set, (a) Original images, (b) MS images (low resolution images), (c) APan images (L component in HLS space) and (d) model fused images. Both the test mean square error and the training mean square error belong to the same magnitude. It can be seen that the model has a good generalization ability.

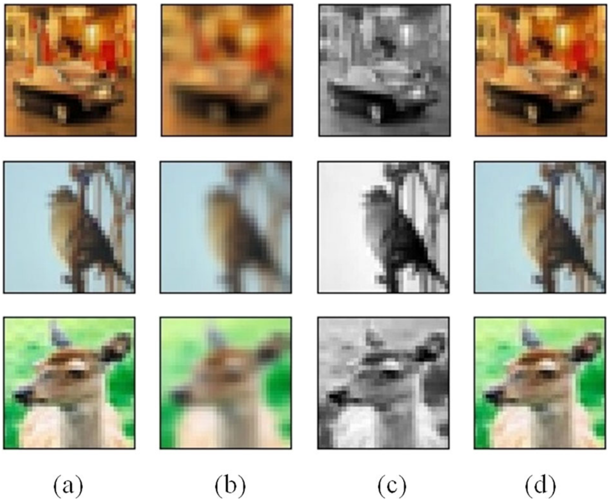


Fig. 5 Fused results of the trained model on the CIFAR test set: **a** original images, **b** low resolution images (for stimulating MS images), **c** L component of original images (for stimulating EPAN images), **d** fused images

5 Experimental results and analysis

In order to fully demonstrate the effectiveness and robustness of the proposed method, we performed experiments on the GLCF public dataset which contains remote sensing images from Landsat and Quickbird satellites [12]. For Landsat images, different band combinations are selected for experiments. For Quickbird images, RGB bands are chosen for the fusion experiment. Five fusion algorithms, High Pass Filter (HPF) [9], Brovey [24], Wavelet Transform and Sparse Representation (WT-SR) [19], Hyperspherical Color Space (HCS) [33] and Nonsubsampled Contourlet Transform (NSCT) [17], are selected as comparative fusion methods. In order to compare the performance of these fusion algorithms, subjective and objective evaluation are conducted separately.

5.1 Objective evaluation metrics

In objective evaluation, six metrics, including mean square error (RMSE), correlation coefficient (CC), universal quality index (UQI), global relative spectral loss (ERGAS), structural similarity index (SSIM) and peak signal to noise ratio (PSNR), are chosen to quantitatively assess the quality of fused images obtained by the five fusion algorithms.

(1) RMSE

$$RMSE = \sqrt{\frac{\sum_{p=1}^{M\times N}\sum_{i=1}^{d}\left(X_i(p)-Y_i(p)\right)^2}{M\times N\times d}} \tag{11}$$

where X is the reference image, Y is fused image, M and N are the length and width of the image respectively, and d is the number of bands. The closer the RMSE is to 0, the more similar the fused image is to the reference image, and the better the fusion quality is.

(2) CC

$$CC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(X_{i,j}-\overline{X}\right)\left(Y_{i,j}-\overline{Y}\right)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(X_{i,j}-\overline{X}\right)^2\sum_{i=1}^{M}\sum_{j=1}^{N}\left(Y_{i,j}-\overline{Y}\right)^2}} \tag{12}$$

where X is the reference image, and Y is fused image. It is used to compute the similarity of spectral features between the reference and the fused images. The value of CC should be close to +1, if the reference and fused images are the same.

(3) UQI

$$UQI = \frac{4\sigma_{XY}(\mu_X+\mu_Y)}{(\mu_X^2+\mu_Y^2)(\sigma_X^2+\sigma_Y^2)} \tag{13}$$

where μ_X and μ_Y represent the means of reference image X and fused image Y; σ_X and $\sigma_Y$denote the standard deviation of X and Y; σ_{XY} is the covariance of X and Y. UQI is used to calculate the amount of similar information in reference image X and fused image Y. The range of this measure is from −1 to 1. The value 1 indicates that the reference and fused images are similar.

(4) ERGAS

$$ERGAS = 100\frac{h}{l}\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{RMSE(n)}{\mu(n)}\right)^2} \tag{14}$$

where h/l represents the ratio of the Pan image and MS image in terms of resolution, and N is the number of bands. It is used to compute the quality of fused image according to normalized average error of each band. Higher ERGAS indicates larger distortion in the fused image.

(5) SSIM

$$SSIM = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \tag{15}$$

SSIM is used to represent the structural similarity between the reference image and the fused image.μ_X and μ_{Y} represent the mean of the reference image and fused image respectively.$\sigma_{X\mathrm{Y}}$ represents the covariance of the reference image and the fused image. σ_X and σ_{Y} represent the

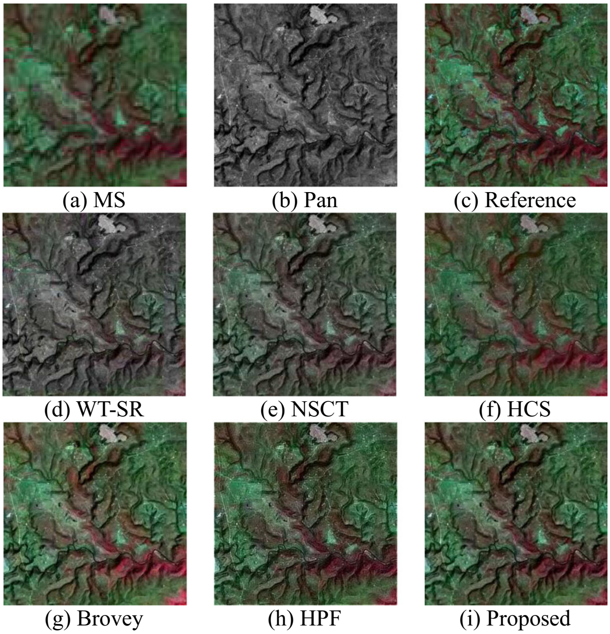

(a) MS (b) Pan (c) Reference
(d) WT-SR (e) NSCT (f) HCS
(g) Brovey (h) HPF (i) Proposed

Fig. 6 Fusion results of the 1st group of Landsat images

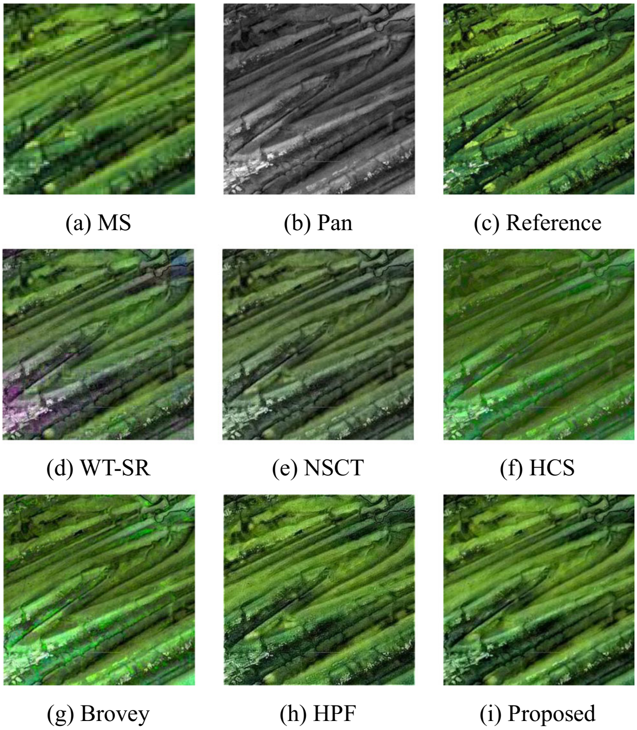

(a) MS
(b) Pan
(c) Reference
(d) WT-SR
(e) NSCT
(f) HCS
(g) Brovey
(h) HPF
(i) Proposed

Fig. 7 Fusion results of the 2nd group of Landsat images

standard deviation of the reference image and fused image respectively. The range is from −1 to 1. The closer it is to 1, the closer the fused image is to the reference image.

(6) PSNR

$$PSNR = 20\log_{10}\left\{\frac{L}{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(X(i,j)-Y(i,j)\right)^2}\right\} \tag{16}$$

Here, L is the number of gray levels. The larger the PSNR is, the more similar the fused image is to the reference image, and the better the fusion quality is.

Table 1 Evaluation results of the fused images in Fig. 6

	RMSE	CC	UQI	ERGAS	SSIM	PSNR
WT-SR	20.3770	0.8061	0.7864	6.9213	0.8036	24.5064
NSCT	16.7573	0.8589	0.8110	5.9400	0.8275	24.6050
HCS	**14.4165**	0.8705	0.8224	5.5550	0.8223	25.5915
Brovey	20.0427	0.8799	0.8105	6.5433	0.8097	22.4086
HPF	21.0655	0.8087	0.7961	8.0899	0.7505	22.1699
Proposed	14.4870	**0.8976**	**0.8268**	**5.3729**	**0.8314**	**25.6976**

Table 2 Evaluation results of the fused images in Fig. 7

	RMSE	CC	UQI	ERGAS	SSIM	PSNR
WT-SR	15.4933	0.8512	0.8766	7.2431	0.8109	24.0231
NSCT	14.9028	0.8892	0.8813	6.5432	0.8231	24.8839
HCS	13.4550	0.8889	0.8932	5.2262	0.8492	25.9448
Brovey	19.5596	0.8923	0.8719	6.4110	0.8239	22.5330
HPF	13.9074	0.9070	0.8902	5.4103	**0.8638**	25.9249
FusionCNN	**12.3618**	**0.9155**	**0.8934**	**4.8591**	0.8607	**26.9076**

5.2 Experimental results on Landsat images

Bands 1, 2, 3, 4, 5, and 7 in the Landsat TM images are multi-spectral bands with resolution of 30 m, and band 8 is a panchromatic one with resolution of 15 m. Figures 7 and 8 show two results which contain MS images, PAN images, reference images, and the fused images obtained by six fusion algorithms. In Fig. 6, the original MS image is constructed with bands 4, 3, 2; in Fig. 7, the original MS image is constructed with bands 7, 5, 2. For the sake of comparison, we use the original MS image as reference image [1, 22] (with a resolution of 30 m). The Pan image is down-sampled to the reference image resolution as a new Pan image (with a resolution of 30 m), and the original MS image is down-sampled with the same scale as a new MS image (with a resolution of 60 m). The resolution of MS and Pan images used in this paper is 200 × 200 and 400 × 400, respectively, and the reference image is 400 × 400.

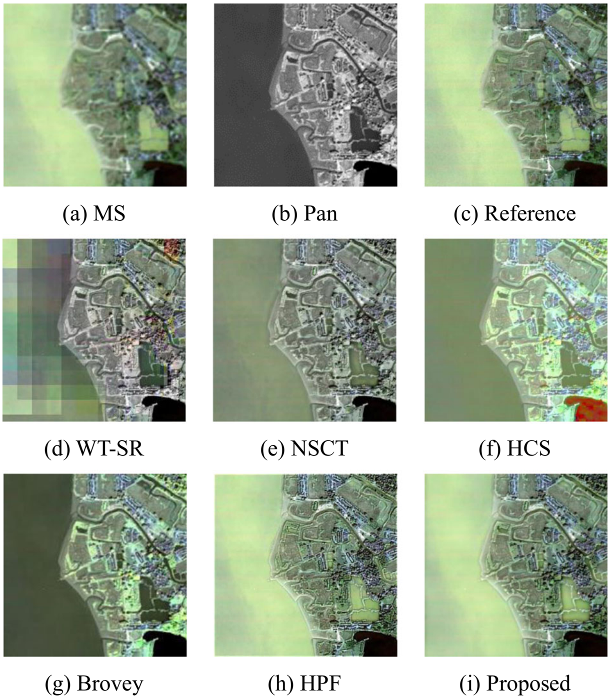
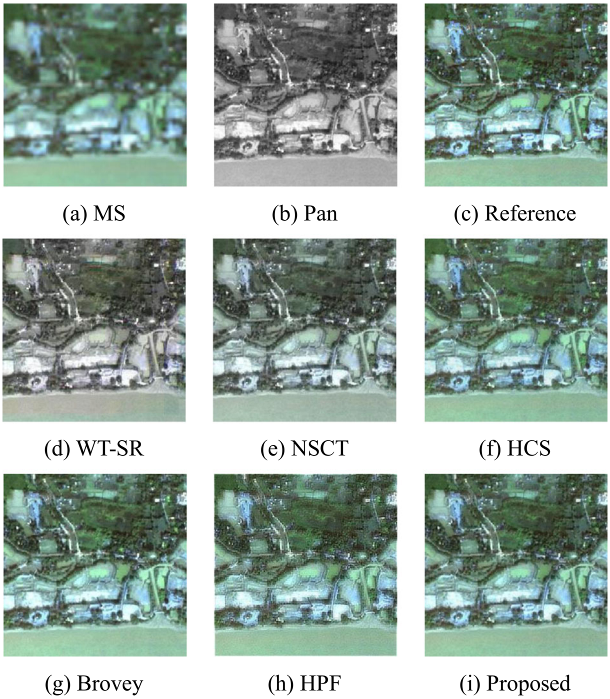

(a) MS
(b) Pan
(c) Reference
(d) WT-SR
(e) NSCT
(f) HCS
(g) Brovey
(h) HPF
(i) Proposed

Fig. 8 Fusion results of the 1st group of Quickbird images

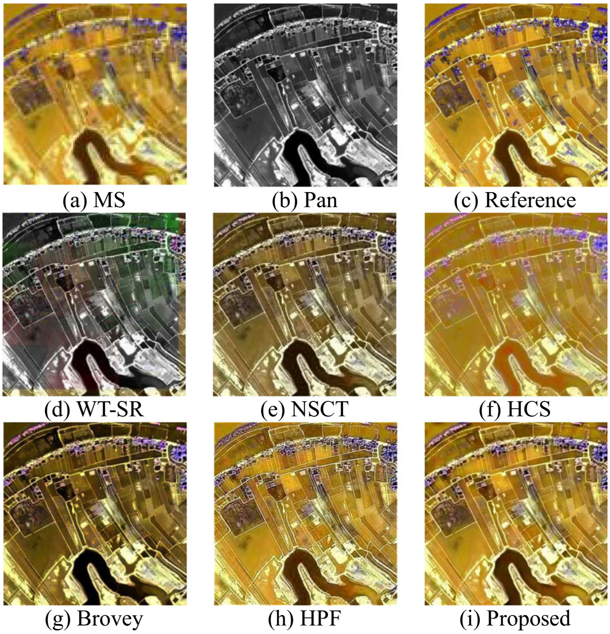


(a) MS (b) Pan (c) Reference

(d) WT-SR (e) NSCT (f) HCS

(g) Brovey (h) HPF (i) Proposed

Fig. 9 Fusion results of the 2nd group of Quickbird images



(a) MS (b) Pan (c) Reference

(d) WT-SR (e) NSCT (f) HCS

(g) Brovey (h) HPF (i) Proposed

Fig. 10 Fusion results of the 3th group of Quickbird images

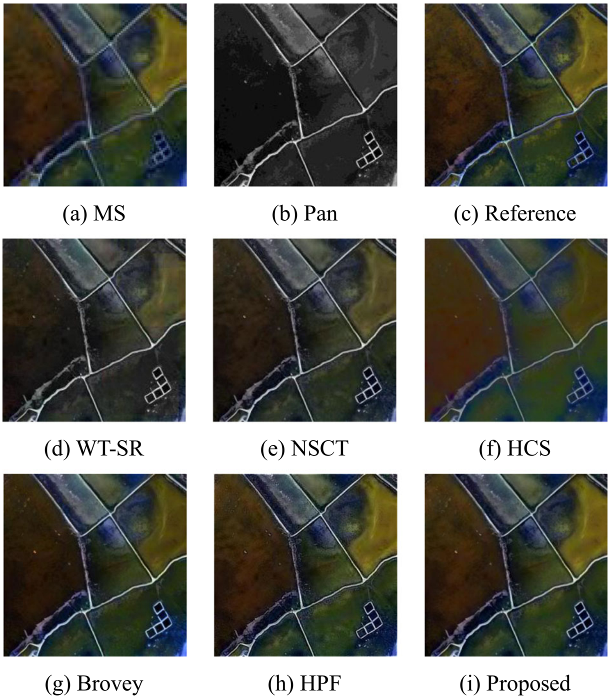


Fig. 11 Fusion results of the 4th group of Quickbird images

It can be seen from Fig. 6 that the fused image obtained by WT-SR is heavily corrupted by spectral distortion. In the fused image by NSCT, the fused image is darker and the spectrum is distorted. The texture in the fused image by HCS is distorted. The spatial structure of the Brovey remains better, but the spectral distortion is serious. The HPF method has better spectral retention, but the spatial structure is distorted. Compared with these algorithms, the proposed method can preserve spectral information of MS image and spatial information of Pan image well. Table 1 shows the evaluation results of the fused images in Fig. 6, where the best results corresponding to each metric are marked in bold. It can be seen that the fusion method proposed in this paper is superior to other algorithms on the five metrics (except RMESE), confirming the subjective evaluation from an objective perspective.

Table 3 Evaluation results of the fused images in Fig. 8

	RMSE	CC	UQI	ERGAS	SSIM	PSNR
WT-SR	62.3001	0.3172	0.7241	12.6698	0.4844	11.9162
NSCT	50.7761	0.5452	0.6495	10.1216	0.6040	13.6348
HCS	71.2143	−0.0825	0.7887	12.0615	0.4881	10.7184
Brovey	89.7570	−0.1407	0.6436	21.4769	0.3987	8.6026
HPF	29.3566	0.8838	0.9391	5.0770	0.6789	19.1467
FusionCNN	**23.6326**	**0.9128**	**0.9592**	**4.1491**	**0.7283**	**20.9645**

Table 4 Evaluation results of the fused images with in Fig. 9

	RMSE	CC	UQI	ERGAS	SSIM	PSNR
WT-SR	25.3392	0.9032	0.9014	10.3740	0.7544	19.9021
NSCT	50.4384	0.8897	0.6432	20.4801	0.6038	14.9845
HCS	29.8448	0.9053	0.9021	6.5009	0.7162	19.0340
Brovey	54.6062	0.8307	0.6714	13.7904	0.6521	13.3559
HPF	35.2090	0.8863	0.8879	8.9135	0.7142	17.6623
FusionCNN	**26.3109**	**0.9334**	**0.9190**	**6.2391**	**0.7744**	**20.1174**

Figure 7 shows the fusion results of the second group Landsat images. The WT-SR method results in serious distortion both in spectral and spatial structure. The NSCT method leads to more serious spectral distortion. The HCS method causes spectral distortion and spatial structure distortion. The Brovey method produces brighter spectrum than the reference image does. Although the overall spectral information in the HPF method remains better, the spatial distortion exits. The fusion result of the proposed method can preserve spectral information and spatial information well. Table 2 shows the evaluation results of various fused images in Fig. 7. Except SSIM, the method proposed in this paper has achieved the best results on other metrics, which is also consistent with our subjective evaluation.

5.3 Experimental results on Quickbird images

Since the resolution of the Landsat satellite is relatively low, the fusion effect cannot be well represented. We only selected two groups of images for experiments. In order to better examine the advantages of the method proposed in this paper, we performed experiments on the Quickbird satellite dataset with higher resolution. Bands 1, 2, 3 and 4 in the QuickBird image are multispectral bands with resolution of 2.8 m, while the resolution of PAN image is 0.7 m. Similar to the experimental setting in subsection 5.2, the original MS image is used as reference image (with a resolution of 2.8 m), and the Pan image is down-sampled to the resolution of the reference image. The MS image is down-sampled on the same scale as new MS image. The MS and Pan images used in this paper are 200 × 200 and 400 × 400, respectively, and the reference image is 400 × 400. Figs. 8, 9, 10, and 11 display four group fusion results of Quickbird images.

In Figs. 8 and 9, the MS image is constructed with bands 1, 2 and 3; in Figs. 10 and 11, the MS image is constructed with bands 2, 3 and 4. Tables 3, 4, 5, and 6 show the objective evaluation results of the fused images in Figs 8, 9, 10, and 11, respectively. Figure 12 shows 9

Table 5 Evaluation results of the fused images in Fig. 10

	RMSE	CC	UQI	ERGAS	SSIM	PSNR
WT-SR	26.9740	0.9034	0.8384	3.6483	0.9001	22.4932
NSCT	25.7403	0.9649	0.8992	3.7421	0.9028	23.0324
HCS	12.3505	0.9781	0.9175	2.4870	0.9204	27.0992
Brovey	12.1475	0.9794	0.9189	**2.3723**	0.9260	27.1448
HPF	21.5072	0.9260	0.9023	4.3108	0.8108	21.9461
FusionCNN	**11.8945**	**0.9870**	**0.9199**	2.4253	**0.9288**	**27.6808**

Table 6 Evaluation results of the fused images in Fig. 11

	RMSE	CC	UQI	ERGAS	SSIM	PSNR
WT-SR	18.0341	0.8495	0.8114	9.2559	0.7821	24.0965
NSCT	15.3007	0.8902	0.7352	7.7850	0.8193	25.3899
HCS	16.1732	0.8713	0.7926	7.8649	0.7665	24.3812
Brovey	16.2029	0.8970	0.8012	8.4801	0.8156	24.1564
HPF	16.3949	0.9048	0.8215	8.3057	0.8001	24.5242
FusionCNN	**12.1526**	**0.9398**	**0.8390**	**6.1008**	**0.8393**	**27.0405**

groups of test MS images. Figure 13 shows the averaged evaluation results of the 9 groups of fusion results.

Subjective evaluation It can be observed from the four groups of fusion results that the WT-SR produces distortions both in spectral information and spatial structure, and the algorithm is sensitive to the type of objects. The NSCT preserves the spatial information of the MS image well in all four groups of images, but the spectral distortion is serious. HCS method results are similar to PCA in terms of spectral distortion and spatial distortion. Brovey method has great difference in fusion quality for different fusion bands and object types. The HPF method can retain spectral information of MS images well, but the spatial structure is obviously distorted. Compared with the comparative algorithms, the fusion method proposed in this paper can preserve spectral information of MS images and spatial information of Pan images for maximum degree.

Objective evaluation Tables 3, 4, 5, and 6 are the evaluation results of the fused images in Figs 8, 9, 10, and 11, respectively, where the best results for each metric are marked in bold. From these tables, we can see that the proposed method achieved the best results in all metrics except the ERGAS in Fig. 10. Figure 13 shows the average evaluation results of the 9 groups of fused images. The method proposed in this paper achieved the best results on all six metrics. This also quantitatively shows the rationality of the subjective evaluations above. The most

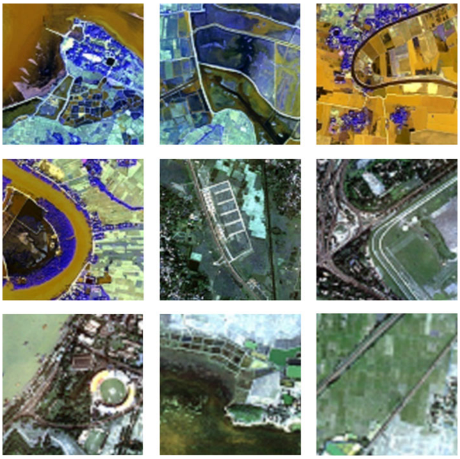
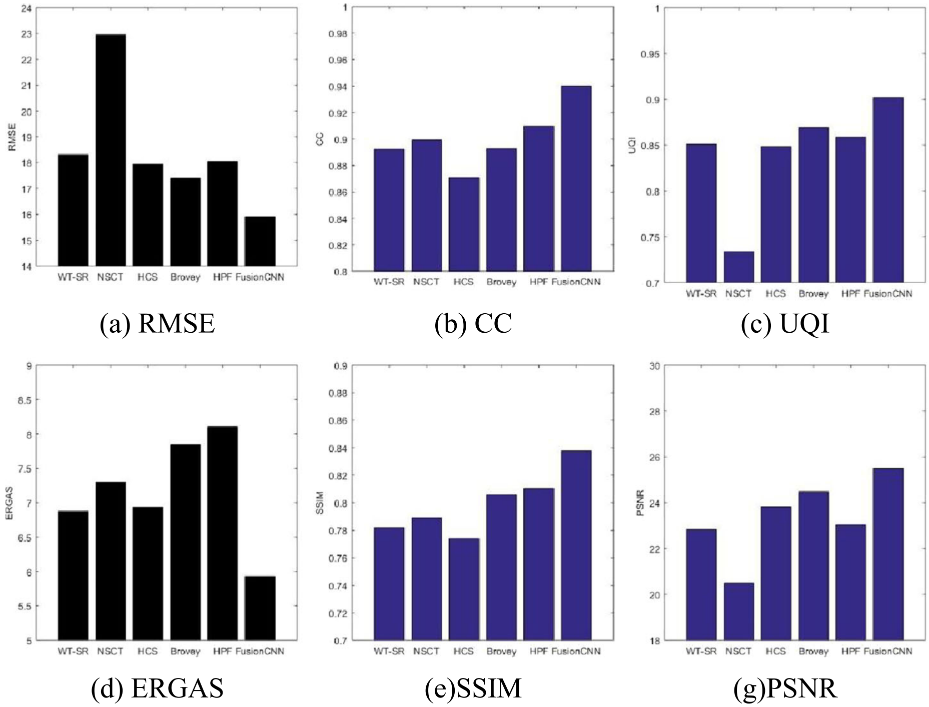


Fig. 12 9 groups of Quickbird MS images.



(a) RMSE (b) CC (c) UQI

(d) ERGAS (e)SSIM (g)PSNR

Fig. 13 The averaged evaluation results of the 9 groups of fusion results

important thing is that our method can obtain good fusion results for different bands of MS images and different types of ground objects. Spectral and spatial information can be well preserved with strong robustness.

6 Conclusion

In this paper, we utilized the characteristics of remote sensing image fusion and the advantages of convolutional neural networks to invent an algorithm for image fusion. By constructing a new data set from natural image set CIFAR, we successfully train the fusion network, which enables end-to-end fusion of the MS image and Pan image. In the algorithm, we first enhance the Pan image by injecting the low frequency information of MS image using Non-Subsampled Laplacian Pyramid decomposition. And then, the EPAN image and MS image are input into the trained fusion model to generate the final fused image. Comparing with five classical methods, the experiments show that the proposed method can well process different bands of MS images and different types of objects. It can achieve satisfactory fusion quality: the fused image can maintain the MS image spectral information and Pan image spatial information well with a strong robustness.

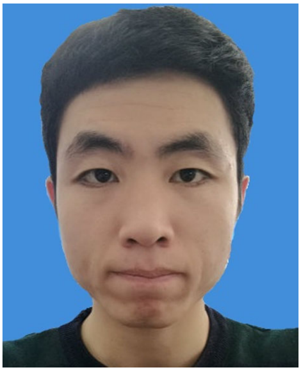
Acknowledgements The work was supported by National Science & Technology Pillar Program of China (Grant No. 2012BAH48F02), National Natural Science Foundation of China (Grant No. 61272209, 61801190), Natural Science Foundation of Jilin Province (Grant No. 20180101055JC), Outstanding Young Talent Foundation of Jilin Province (Grant No. 20180520029JH) and China Postdoctoral Science Foundation (Grant No. 2017M611323). The authors would like to thank Dr. Shuang Yu for her help on technical editing of the manuscript, and Prof. Xiaoying Sun for scientific advices.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Amro I, Mateos J, Vega M et al (2011) A survey of classical methods and new trends in pansharpening of multispectral images. EURASIP J Adv Signal Process 79:1–22. https://doi.org/10.1186/1687-6180-2011-79
2. Burt PJ, Adelson EH (1983) The Laplacian pyramid as a compact image code. IEEE Trans Commun COM-31(4):532–540
3. Cheng J, Liu H, Liu T et al (2015) Remote sensing image fusion via wavelet transform and sparse representation. ISPRS J Photogramm Remote Sens 104:158–173
4. Choi M, Kim RY, Nam MR, Kim HO (2005) Fusion of multispectral and panchromatic satellite images using the curvelet transform. IEEE Geosci Remote Sens Lett 2(2):136–140
5. Chu H, Zhu W (2008) Fusion of IKONOS satellite imagery using IHS transform and local variation. IEEE Geosci Remote Sens Lett 5(4):653–657
6. Collobert R (2011) Natural language processing (almost) from scratch. J Mach Learn Res 12:2493–2537
7. Do MN, Vetterli M (2005) The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans Image Process 14(12):2091–2106
8. Fan C, Wang L, Liu P et al (2016) Compressed sensing based remote sensing image reconstruction via employing similarities of reference images. Multimed Tools Appl 75(19):12201–12225
9. Gangkofner UG, Pradhan PS, Holcomb DW (2008) Optimizing the high pass filter addition technique for image fusion. Photogramm Eng Remote Sens 74(9):1107–1118
10. Ghahremani M, Ghassemian H (2015) Remote-sensing image fusion based on Curvelets and ICA. Int J Remote Sens 36(16):4131–4143
11. Ghassemian H (2016) A review of remote sensing image fusion methods. Inform Fusion 32(PA):75–89
12. Global Land Cover Facility. http://www.landcover.org/. Accessed 11 Nov 2018
13. González-Audícana M, Saleta JL, Catalán RG et al (2004) Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition. IEEE Trans Geosci Remote Sens 42(6):1291–1299
14. Hinton GE, OsinderoS TYW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7): 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527
15. Hnatushenko VV, Vasyliev VV (2016) Remote sensing image fusion using Ica and optimized wavelet transform. Int Arch Photogramm Remote Sens Spat Inf Sci XLI-B7:653–659
16. Ji X, Zhang G (2017) Image fusion method of SAR and infrared image based on Curvelet transform with adaptive weighting. Multimed Tools Appl 76(17):17633–17649
17. Kong WW, Lei YJ, Lei Y et al (2011) Image fusion technique based on non-subsampled contourlet transform and adaptive unit-fast-linking pulse-coupled neural network. IET Image Process 5(2):113–121
18. Krizhevsky A, Sutskever I, Hinton G (2012) ImageNet classification with deep convolutional neural networks. In proc. Adv Neural Inf Proces Syst 25:1090–1098
19. Liu Y, Wang Z (2014) A practical pan-sharpening method with wavelet transform and sparse representation. IEEE international conference on imaging systems and techniques, 288–293
20. Luo Y, Liu R, Zhu YF (2011) Fusion of remote sensing image base on the PCA + ATROUS wavelet transform. Appl Mech Mater 353-356:172–176
21. Nakazawa T, Kulkarni DV (2018) Wafer map defect pattern classification and image retrieval using convolutional neural network. IEEE Trans Semicond Manuf 31(2):309–314
22. Paramanandham N, Rajendiran K (2017) Multi sensor image fusion for surveillance applications using hybrid image fusion algorithm. Multimed Tools Appl. https://doi.org/10.1007/s11042-017-4895-3
23. Park CC, Kim Y, Kim G (2018) Retrieval of sentence sequences for an image stream via coherence recurrent convolutional networks. IEEE Trans Pattern Anal Mach Intell 40(4):945–957
24. PohlC V (1998) Multisensor image fusion in remote sensing: concepts, methods and applications. Int J Remote Sens 19(5):823–854

25. Shah VP, Younan NH, King RL (2008) An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. IEEE Trans Geosci Remote Sens 46(5):1323–1335
26. Shahdoosti HR, Ghassemian H (2015) Fusion of MS and PAN images preserving spectral quality. IEEE Geosci Remote Sens Lett 12(3):611–615
27. Shahdoosti HR, Ghassemian H (2016) Combining the spectral PCA and spatial PCA fusion methods by an optimal filter. Information Fusion 27:150–160
28. Shensa MJ (1992) The discrete wavelet transform: wedding the àtrous and Mallat algorithm. IEEE Trans Signal Process 40(10):2464–2482
29. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. Adv Neural Inf Proces Syst 27:3104–3112
30. Tu TM, Su SC, Shyu HC et al (2001) A new look at ihs-like image fusion methods. Inform Fusion 2(3): 177–186
31. Tu TM, Huang PS, Hung CL et al (2004) A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery. IEEE Geosci Remote Sens Lett 1(4):309–312
32. Valizadeh SA, Ghassemian H (2012) Remote Sensing image fusion using combining HIS and Curvelet transform. In: International symposium on telecommunications, 1184–1189
33. Wu B, Fu Q, Sun L et al (2015) Enhanced hyperspherical color space fusion technique preserving spectral and spatial content. J Appl Remote Sens 9(1):097291
34. Yang Y, Wan W, Huang S et al (2017) Remote sensing image fusion based on adaptive IHS and multiscale guided filter. IEEE Access 4:4573–4582
35. Zhang X, Li X, Feng Y (2017) Image fusion based on simultaneous empirical wavelet transform. Multimed Tools Appl 76(6):8175–8193



Fajie Ye received his B.S. degree in geo-exploration science and technology from Jilin University in 2016. He is currently a master candidate in School of computer science and technology in Jilin University. His research interests include multimedia processing and information fusion.

Xiongfei Li received the BS degree in computer software in 1985 from Nanjing University, the MSc degree in computer software in 1988 from the Chinese academy of sciences, the PhD degree in communication and information system in 2002 from Jilin University. Since 1988, he has been a member of the faculty of the computer science and technology at Jilin University, Changchun, China. He is a professor of computer software and theory at Jilin University. He has authored more than 100 research papers. His research interests include data mining, intelligent network, image processing and analysis. Prof. Li is a member of the IEEE.

Xiaoli Zhang received his M.S. and Ph.D. degree in computer science and technology from Jilin University in 2012 and 2016, respectively. He is currently a faculty member with the School of computer science and technology in Jilin University. His research interests include multimedia processing, information fusion, and data mining. He has published more than 30 papers in journals and conferences.