EURASIP Journal on Image and Video Processing

RESEARCH Open Access

Pansharpening based on convolutional autoencoder and multi-scale guided filter

Ahmad AL Smadi[1], Shuyuan Yang[1*], Zhang Kai[2], Atif Mehmood[1], Min Wang[3] and Ala Alsanabani[1]

*Correspondence:
syyang2009@gmail.com
[1]School of Artificial Intelligence, Xidian University, No. 2 South Taibai Road, 710071 Xian, China
Full list of author information is available at the end of the article

Abstract

In this paper, we propose a pansharpening method based on a convolutional autoencoder. The convolutional autoencoder is a sort of convolutional neural network (CNN) and objective to scale down the input dimension and typify image features with high exactness. First, the autoencoder network is trained to reduce the difference between the degraded panchromatic image patches and reconstruction output original panchromatic image patches. The intensity component, which is developed by adaptive intensity-hue-saturation (AIHS), is then delivered into the trained convolutional autoencoder network to generate an enhanced intensity component of the multi-spectral image. The pansharpening is accomplished by improving the panchromatic image from the enhanced intensity component using a multi-scale guided filter; then, the semantic detail is injected into the upsampled multi-spectral image. Real and degraded datasets are utilized for the experiments, which exhibit that the proposed technique has the ability to preserve the high spatial details and high spectral characteristics simultaneously. Furthermore, experimental results demonstrated that the proposed study performs state-of-the-art results in terms of subjective and objective assessments on remote sensing data.

Keywords: Pansharpening, Convolutional autoencoder, Guided image filtering, Adaptive intensity-hue-saturation AIHS

1 Introduction

There are many applications based on remote sensing satellites that require observation of the alterations of the earth, such as image fusion [1–3] and mapping land cover [4]. Given that, pansharpening is one of the essential interests of many scientists. It is difficult that the remote sensing satellites can obtain a panchromatic image (PAN) and a multi-spectral image (MS) with the qualities of both high spatial resolution and high spectral resolution at the same time due to data transmission impediment. However, the main objective of pansharpening is fusing the high spatial resolution PAN image with the corresponding high spectral resolution MS image to acquire high spatial and spectral resolutions for MS image [5].

As indicated by [6–8], a wide assortment of image fusion techniques can be classified into two classes based on the way of extracting a spatial detail from a PAN image: (1) component substitution (CS) and (2) multi-resolution analysis (MRA). And some methods do

not belong to these two categories, such as model-based pansharpening method [9, 10]. Among the conventional component substitution-based methods include intensity-hue-saturation (IHS) [11], principal component analysis (PCA) [12], Gram-Schimidt [13], and Brovey transform [14], etc. in which the detail information is extracted by the difference between the PAN image and linear combination of the upsampled MS image; therefore, the component substitution-based methods have a spectral distortion in the fused image. In contrast, the multi-resolution analysis-based methods, such as Smoothing Filter-based Intensity Modulation (SFIM) [15], generalized Laplacian pyramid (MTF-GLP) [16], and indusion [17], extract the detail information by the difference between the PAN image and its low resolution. These methods offer an outstanding spectral resolution, but they suffer from spatial distortion in the fused image. The edge-preserving filtering techniques have drawn an important role in pansharpening. Guided image filter [18] is one of the well-known techniques. Yang et al. [19] introduced multi-scale guided filer based on adaptive intensity-hue-saturation (MSGF); they used the intensity image as a guidance image to enhance the PAN image. In our work, the multi-scale guided filter is used to enhance the semantic detail map by utilizing the enhanced intensity image as a guidance image that is obtained by CAE.

Recently, the use of deep neural networks has been a hot topic in many fields [20–25]. Researchers have started investigating this topic for pansharpening. Scarpa et al. [21] proposed the convolutional neural network-based pansharpening method.

Residual convolutional neural network (RCNN) was utilized to achieve pansharpening [26]. Huang et al. [27] introduced a pansharpening model using deep neural networks (DNN), which utilized the relationship between PAN image patches and MS image patches for training the neural network. More recently, in [28], convolutional autoencoder (CAE)-based multi-spectral image fusion was introduced in which the low-resolution MS images is fed into the trained CAE to generate estimated high-resolution MS images; then, the fusion process is achieved by injecting the detailed map of each image into the corresponding estimated high-resolution MS bands. Inspired by this, we propose a pansharpening technique based on a convolutional autoencoder. First, the convolutional autoencoder is trained from the degraded PAN image patches to generate the original PAN image patches; the AIHS component is then tested on the trained network to obtain enhanced intensity components. Further, the guided filter is employed to enhance the PAN image using the enhanced intensity component. Finally, the experiments are conducted on both real and degraded datasets. We showed that the fusion process of the convolutional autoencoder with a guided filter is capable of preserving the high spatial details and high spectral characteristics simultaneously, which is a start-of-the-art approach on multiple tasks. And our method is also more robust against spectral and spatial distortions.

1.1 Convolutional autoencoder

Autoencoder belongs to unsupervised learning that considers an input image and attempts to reconstruct it back. The convolutional autoencoder is a sort of convolutional neural network that reproduces the input image patches at the output. However, the design of a convolutional autoencoder comprises two fundamental phases, which are the encoding phase and the decoding phase. The encoding phase represents half of the network, and it incorporates convolution and max-pooling layers. In contrast,

the decoding phase for recreating the input image patches from the degraded pieces comprises deconvolution and upscaling layers [29].

1.1.1 Encoding phase

A convolution among an input volume $I = \{I_1, \cdots, I_D\}$ with D dimension and every convolutional layer is composed of n convolutional filters $F^{(1)} = \left\{F_1^{(1)}, \ldots, F_{\mathrm{n}}^{(1)}\right\}$ which is considered to produce m features.

$$O_m = a\left(I * F_m^{(1)} + b_m^{(1)}\right) \quad m = 1, 2, \cdots, n \tag{1}$$

O_m represents the feature maps of the input I, b_m represents the bias, and a denotes an activation function.

1.1.2 Decoding phase

The produced m feature maps considered to be used as input to the decoder, to reconstruct the input image, which is obtained by the consequence of the convolution between $O = \{O_{i=1}\}^n$ with convolutional filters $F^{(2)} = \left\{F_1^{(2)}, \ldots, F_{\mathrm{n}}^{(2)}\right\}$ that estimated as follows:

$$\tilde{I} = a\left(O * F_{\mathrm{m}}^{(2)} + b_{\mathrm{m}}^{(2)}\right) \tag{2}$$

Considering that both the output image patches and its input have the same dimension, therefore, it is conceivable to relate I and $\tilde{I}$ using a loss function to update the weights during training, for example, mean square error (MSE).

$$\mathcal{L}(I, \tilde{I}) = \frac{1}{2}\|I - \tilde{I}\|_2^2 \tag{3}$$

1.2 Adaptive intensity-hue-saturation

The IHS technique belongs to CS-based methods that introduced [30], and it is just appropriate for MS images with three bands [11]. Even though the IHS strategy displays extraordinary spatial quality, it severely experiences spectral distortion. The general formula for generating an intensity component is as follows:

$$I = \sum_{i=1}^{n} \alpha_i M_{i'th} \tag{4}$$

where α_i denotes the weight coefficients, and n represents the number of spectral bands. M_i indicates the i_{th} band of the upsampled MS band. Therefore, Rahmani et al. [31] AIHS was introduced, in which the optimal weights are obtained by solving the following optimization problem:

$$\alpha_i^* = \arg\min_{\alpha_{\mathrm{i}}} \left\| PAN - \sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_{\mathrm{i}} \mathbf{M}_{i'th} \right\|^2 \tag{5}$$

where PAN denotes panchromatic image.

1.3 Guided filter

The guided filter **GF** was introduced by He et al. [32]. The uses of guided filter have been widely utilized in image processing fields such as detail enhancement and image fusion. The guided filter can maintain a strategic distance from ringing artifacts. The **GF** depends on a local linear model that is using the guided image **gui** to filter the input image **inp**. Therefore, the output image **Out** can conserve the essential data of the **inp** and obtain the

variation trend of **gui** at the same time [19]. Mathematically, the guided filter is employed to find a pair of scalar values a_i and b_i that solves the following problem [33]:

$$\underset{a_i,b_i}{\operatorname{argmin}} \frac{1}{n} \left\| \mathbf{inp}_i - \left(a_i \mathbf{gui}_i + b_i\right) \right\|_2^2 + \zeta \left|a_i\right|_2^2 \tag{6}$$

Here, n denotes to the number of pixels in a squared window $\mathbf{w}$ with size (2r+1)×(2r+ 1), and ζ is a small regularization constant that prevents large a_i.

$$a_i = \frac{\frac{1}{n}\left(\mathbf{inp}_i - \bar{\mathbf{inp}}_i\right)^{\mathrm{T}}\left(\mathbf{gui}_i - \bar{\mathbf{gui}}_i\right)}{\frac{1}{n}\left(\mathbf{gui}_i - \bar{\mathbf{gui}}_i\right)^{\mathrm{T}}\left(\mathbf{gui}_i - \bar{\mathbf{gui}}_i\right) + \zeta} \tag{7}$$

$$= \frac{\operatorname{cov}\left(\mathbf{inp}_i - \bar{\mathbf{inp}}_i, \mathbf{gui}_i - \bar{\mathbf{gui}}_i\right)}{\operatorname{var}\left(\mathbf{gui}_i - \bar{\mathbf{gui}}_i\right) + \zeta} \tag{8}$$

$$b_i = \bar{\mathbf{inp}}_i - a_i \bar{\mathbf{gui}}_i \tag{9}$$

Here, $\bar{\mathbf{inp}}_i$ and $\bar{\mathbf{gui}}_i$ represent the input image mean and the guidance image mean, respectively. Thus, after computing a_i; b_i for all windows in the image, the filtering output is computed as follows:

$$\mathbf{Out}_i = \bar{a}_i \mathbf{inp}_i + \bar{b}_i \tag{10}$$

The following equation represented the guided filter operation in this paper:

$$\mathbf{Out} = \mathbf{GF}(\mathbf{gui}, \mathbf{inp}) \tag{11}$$

2 Methodology

In this paper, we propose a pansharpening technique based on a convolutional autoencoder and CS-based method. First, we highlight the steps for building our technology are:

- Utilize the convolutional autoencoder to enhance to enhance the intensity component which is obtained by AIHS from MS and PAN images. And the spatial resolution enhancement of the degraded PAN image is used the to train the model.
- Generate the intensity component of the MS image by utilizing AIHS-based method, which is then fed to trained convolutional autoencoder considering this as a testing step.
- Utilize the estimated intensity component to enhance the PAN image by using the guided filter.
- The fusion step represents the last phase of the proposed technique. However, it will be explained in detail later.

Figure 1 illustrates the schematic of the proposed method.

2.1 Enhancing the spatial detail

To enhance the spatial detail of the intensity component, we utilize the convolutional autoencoder network in which the relationship between PAN image patches and its degraded form is learned. Note that the degraded PAN image is generated using bi-cubic interpolation. The convolutional autoencoder is used to minimize the difference between input image patches and reconstruction output original image patches. Figure 2 illustrates the applied structure of the convolutional autoencoder.

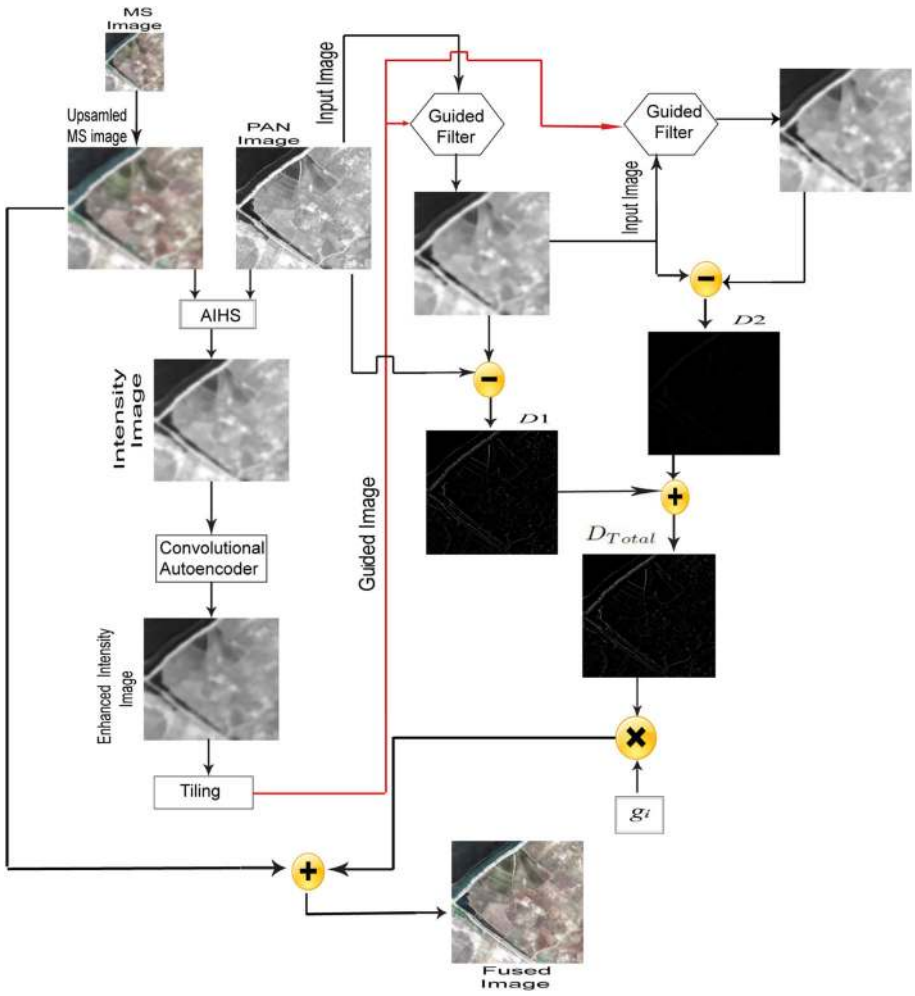

Fig. 1 The proposed methodology schematic. **D** represents the detailed map, and **g** represents the injection gain

According to [28], the same description of the training network would apply here: the PAN image and its spatially degraded image are partitioned into 8×8 patches with 5 overlapping pixels that include 500,000 patch pairs, 30 epochs for training, considering that the relationship between PAN image patches and its degraded image patches is learned

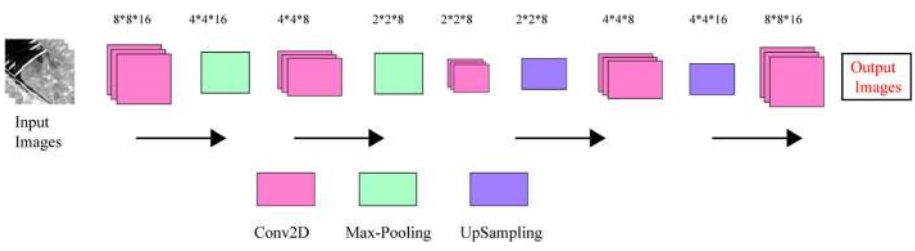

Fig. 2 The structure of the convolutional autoencoder used

by the training network. The following equation illustrates the output patches of the convolutional autoencoder network at each iteration:

$$\left\{\tilde{P}_{\mathrm{i}}\right\}_{\mathrm{i}=1}^{\mathrm{n}} = \mathrm{Dec}\left(\mathrm{Enc}\left(\left\{P_{\mathrm{i}}^{\mathrm{L}}\right\}_{\mathrm{i}=1}^{\mathrm{n}}\right)\right) \tag{12}$$

where $\left\{\tilde{P}_{\mathrm{i}}\right\}_{\mathrm{i}=1}^{\mathrm{n}}$, $\left\{P_{\mathrm{i}}^{\mathrm{L}}\right\}_{\mathrm{i}=1}^{\mathrm{n}}$ represent the output and input patches, respectively. Enc and Dec indicated the encoding and decoding processes, respectively. The encoding process involves several layers starting with (1) the input image patch 8×8; (2) the Conv2D layer that indicates a 2D convolutional layer with 16 filters 3×3 kernel size, activation "ReLU" and padding "same"; the "ReLU" activation is used due to its simplicity and computation efficiency compared to other activation functions [34]. (3) MAX-Pooling layer that indicates a 2D max-pooling 2×2 region with padding "same"; (4) Conv2D layer with 8 filters 3×3 kernel size, activation "ReLU" and padding "same"; (5) Max-Pooling 2×2 region with padding "same"; and (6) Conv2D layer with 8 filters 3×3 kernel size, activation "ReLU" and padding "same". The CAEs are fully convolutional networks; thus, the decoding process is including a convolution. The decoding process involves several layers starting with (1) the Conv2D layer that indicates a 2D convolutional layer with 8 filters 3×3 kernel size, activation "ReLU" and padding 'same'; (2) the UpSampling layer that indicates a 2D UpSampling 2×2 region; (3) the Conv2D layer with 8 filters 3×3 kernel size, activation "ReLU" and padding "same"; (4) UpSampling 2×2 region; (5) the Conv2D layer with 16 filters 3×3 kernel size, activation "ReLU" and padding "same"; and (6) the Conv2D layer with 1 filter 3×3 kernel size, activation "linear" and padding "same". Thus, Adadelta optimization is used throughout training, and the MSE between the reconstructed output patches and the target patches $\left\{P_{\mathrm{i}}^{\mathrm{H}}\right\}_{\mathrm{i}=1}^{\mathrm{n}}$ is used for updating the weights as follows:

$$\mathcal{L}\left(\left\{\tilde{P}_{\mathrm{i}}\right\}_{\mathrm{i}=1}^{\mathrm{n}}, \left\{P_{\mathrm{i}}^{\mathrm{H}}\right\}_{\mathrm{i}=1}^{\mathrm{n}}\right) = \frac{1}{2}\sum_{i=1}^{\mathrm{n}}\left\|\tilde{P}_{\mathrm{i}} - P_{\mathrm{i}}^{\mathrm{H}}\right\|_2^2 \tag{13}$$

After updating the weights, the back-propagation algorithm is utilized for training the convolutional autoencoder network. In the stage of testing, because of similar characteristics between the PAN and the corresponding intensity component of the MS image, the trained network is relied upon to improve the intensity component of MS image; firstly, the intensity component I which is generated by Eq. (5) is partitioned $\{I_{\mathrm{i}}\}_{\mathrm{i}=1}^{\mathrm{n}}$ and is then fed to the trained network for generating an estimated intensity component $\left\{E_{I_i}\right\}_{\mathrm{i}=1}^{\mathrm{n}}$. Thus, the $\left\{E_{I_i}\right\}_{\mathrm{i}=1}^{\mathrm{n}}$ is being tiled.

2.2 Fusion process

The estimated intensity component E_I is employed to enhance the PAN image by using the two-scale guided filter. Firstly, the E_I is being used as the guidance image and the PAN image as the input image.

$$O_1 = \mathbf{GF}(E_I, PAN) \tag{14}$$

The difference between the approximation image O_1 and the input image E_I is represented by the spatial detail D_1. Hence, D_1 will blend with low-frequency component and may cause serious spectral distortion [35]; therefore, D_1 is then utilized as the input image for the second scale of guided filter O_2.

$$D_1 = PAN - O_1 \tag{15}$$

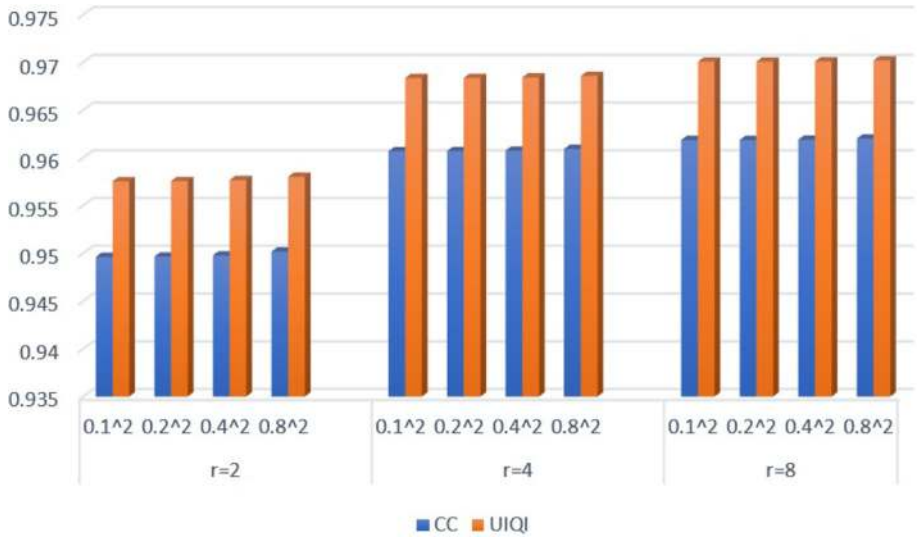

Fig. 3 The influence of the parameters r and ζ on the fusing result concerning CC and UIQI indexes. Higher CC and UIQI indicate the better-fused effect

$$O_2 = \mathbf{GF}(E_I, O_1) \tag{16}$$

The difference between O_1 and O_2 is represented by the spatial detail D_2.

$$D_2 = O_1 - O_2 \tag{17}$$

The total semantic map D_{Total} is injected into the upsampled MS image through injection gains g_i which are adjusted by (19).

$$D_{Total} = D_1 + D_2 \tag{18}$$

$$g_i = \frac{\mathrm{cov}\,(MS_i, E_I)}{\mathrm{var}(E_I)} \tag{19}$$

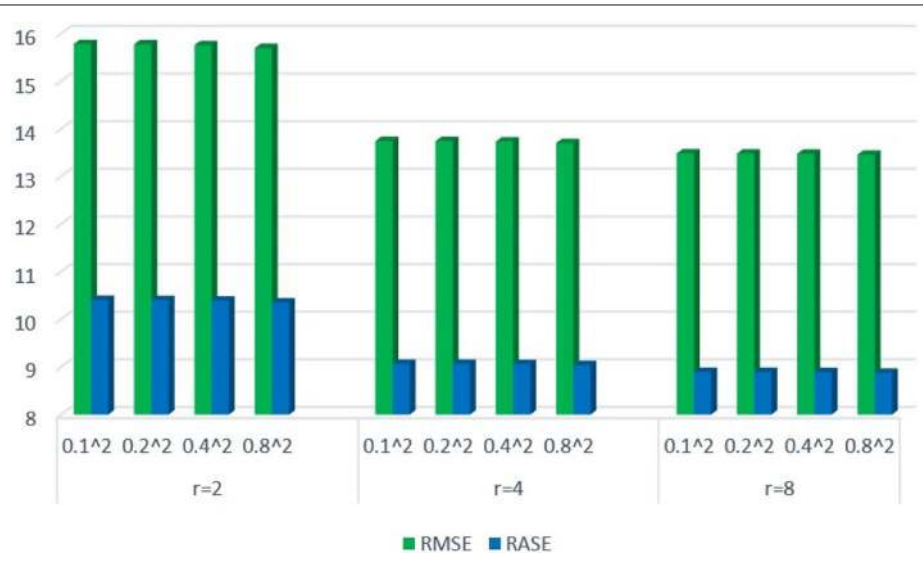

Fig. 4 The influence of the parameters r and ζ on the fusing result concerning RMSE and RASE indexes. Lower RMSE and RASE indicate the better-fused effect

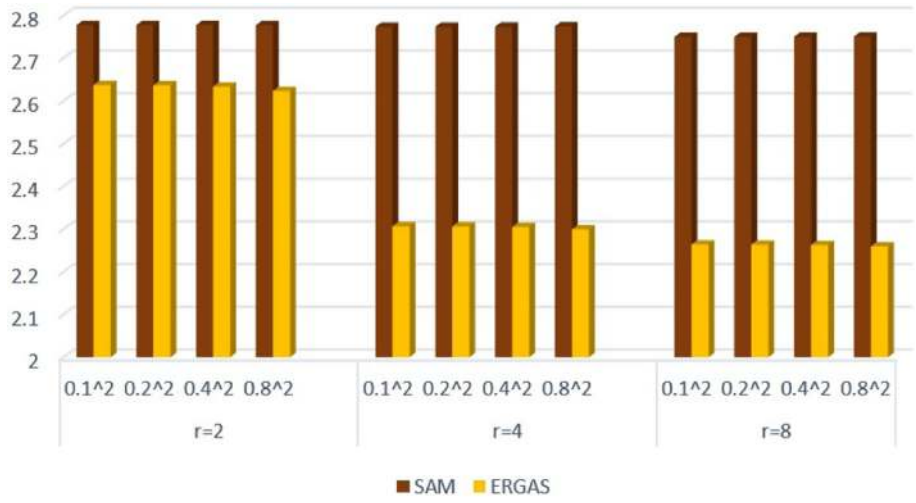

Fig. 5 The influence of the parameters r and ζ on the fusing result concerning SAM and ERGAS indexes. Lower SAM and ERGAS indicate the better-fused effect

The high-resolution multi-spectral (**HRMS**) fused image is conducted by the following equation:

$$\mathbf{HRMS} = MS_{\mathrm{i}} + \mathrm{g_i} D_{Total} \tag{20}$$

3 Results and discussion

In this section, several experiments were performed on different datasets to evaluate the performance of the model based on some quality metrics. Here, 8×8 patches with 5 overlapping pixels of the degraded PAN and the original PAN images that include 500,000 patch pairs were utilized for training the network. In total, six datasets have been selected for implementation purposes. Three degraded datasets (full reference), which means the reference image is available, and three real datasets (no reference image), namely QuickBird and GeoEye.

Therefore, we compared our technique with several conventional efficient pansharpening methods, such as IHS [11], PCA [12], BDSD [36], PRACS [37], and AIHS [31], and several state-of-the-art methods such as SFIM [15], MTF-GLP [16], Indusion [17], MSGF

Table 1 Descriptions of the experimental datasets

Datasets	Size	Spatial resolution	Region and date
Degraded QuickBird Figs. 6 and 7	PAN, 256×256 MS, 64×64	0.7 m/2.8 m	Sundarbans, India on 21 November 2002
Degraded GeoEye Fig. 8	PAN, 256×256 MS, 64×64	0.5 m/2 m	Hobart, Australia on 24 February 2009
Real QuickBird Fig. 9	PAN, 1024×1024 MS, 256×256	0.7 m/2.8 m	Xi'an, China on 30 September 2008
Real QuickBird Fig. 10	PAN, 1024×1024 MS, 256×256	0.7 m/2.8 m	Sundarbans, India on 21 November 2002
Real GeoEye Fig. 11	PAN, 1024×1024 MS, 256×256	0.5 m/2 m	Hobart, Australia on 24 February 2009

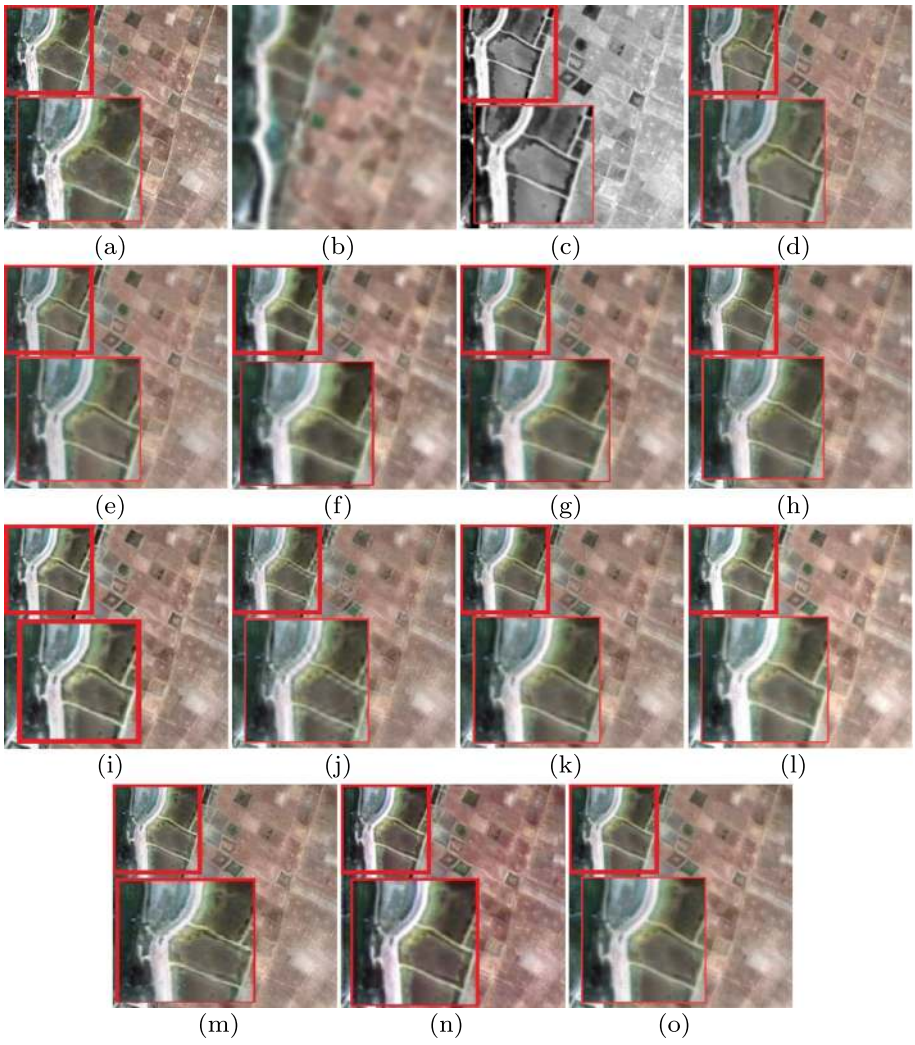

Fig. 6 Fusion results of the full reference QickBird-1 dataset. **a** Reference image (256×256). **b** Degraded MS image (64×64). **c** PAN image (256×256). **d** IHS method. **e** AIHS method. **f** PCA method. **g** BDSD method. **h** PRACS method. **i** SFIM method. **j** Indusion method. **k** MTF-GLP method. **l** CAE method. **m** MSGF method. **n** PNN method. **o** Proposed method

[19], CAE [28], and PNN [38]. Moreover, seven image quality indexes are broadly utilized, to assess the quality of the fused image, which are:

1 Correlation coefficient (CC) [39]
2 Universal Image Quality Index (UIQI) [40]
3 Quaternion Theory-based Quality Index (Q4) [40]
4 Root mean square error (RMSE) [41]
5 Relative average spectral error (RASE) [42]
6 Spectral Angle Mapper (SAM) [43]
7 Erreur Relative Globale Adimensionnelle de Synthese (ERGAS) [44]

To assess the quality of the fused images concerning real datasets, D_s, D_λ, and QNR [45] were employed. The ideal value of each quality index is shown in parentheses in the tables.

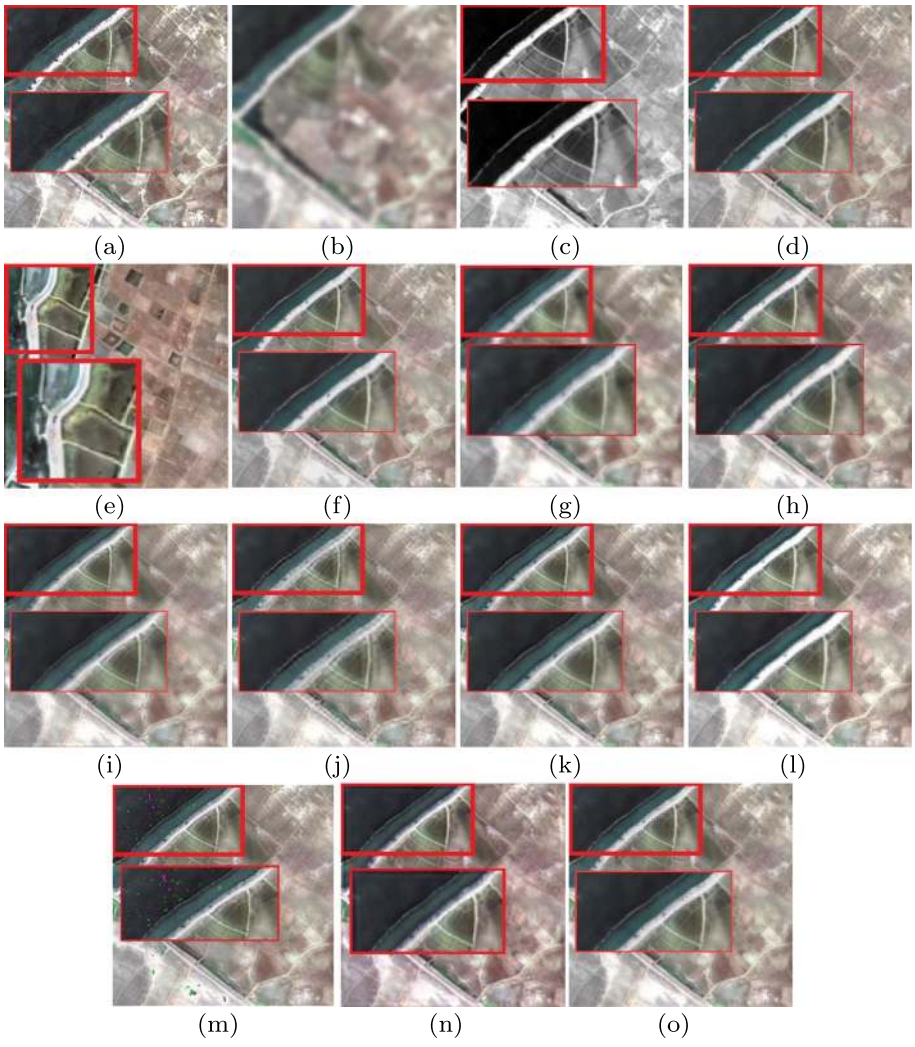

Fig. 7 Fusion results of the full reference QickBird-2 dataset. **a** Reference image (256×256). **b** Degraded MS image (64×64). **c** PAN image (256×256). **d** IHS method. **e** AIHS method. **f** PCA method. **g** BDSD method. **h** PRACS method. **i** SFIM method. **j** Indusion method. **k** MTF-GLP method. **l** CAE method. **m** MSGF method. **n** PNN method. **o** Proposed method

3.1 Parameter investigation

Here, we study the influence of parameter setting in the guided filter on the fusion simulation of degraded QuickBird-1 dataset, namely, window size r and the regularization parameter ζ. Figures 3, 4, and 5 illustrate the influence of these parameters, where the horizontal axis is the regularization parameter ζ concerning three cases of window size r and the vertical axis is quality index results. Therefore, as can be seen, the best performance results originated from setting the parameters r and ζ at 8 and 0.8^2, respectively.

3.2 Fusion results of degraded datasets (full reference)

In this section, the simulations were carried out on degraded datasets that have the reference image to evaluate our proposed method according to Wald's protocol [46]. Regarding the degraded datasets (QuickBird, GeoEye), the sizes of the MS image and the PAN image

Table 2 Numerical results of the full reference QuickBird-1 dataset

Method	CC(1)	UIQI(1)	RMES(0)	RASE(0)	SAM(0)	ERGAS(0)	Q4(1)
IHS	0.91446	0.93471	18.509	12.673	3.837	3.2713	0.8211
AIHS	0.93518	0.95778	15.045	10.301	3.7497	2.7063	0.88383
PCA	0.93743	0.94871	16.408	11.234	3.6063	2.9025	0.84805
BDSD	0.92229	0.95204	16.435	11.253	3.7728	2.9444	0.86477
PRACS	0.93038	0.95991	14.992	10.265	3.8317	2.7466	0.86389
SFIM	0.89958	0.92403	20.297	13.897	3.6792	3.5142	0.81325
Indusion	0.91806	0.93911	17.967	12.302	3.5524	3.1483	0.81981
MTF-GLP	0.91144	0.93984	18.306	12.534	3.6399	3.2217	0.84308
CAE	0.94467	0.95735	23.955	16.402	3.3501	4.1479	**0.8968**
MSGF	0.9264	0.93907	18.165	12.437	3.3836	3.1604	0.88853
PNN	0.89955	0.92229	21.857	14.965	4.8681	3.7866	0.8183
Proposed	**0.94563**	**0.96355**	**14.096**	**9.6514**	**3.3264**	**2.5357**	0.88685

are 64×64 and 256×256, respectively. The descriptions of the experimental datasets are shown in Table 1.

3.2.1 Experiments on degraded QuickBird datasets

In this section, two pairs of QuickBird satellite datasets were examined; Fig. 6 illustrates the fusion results of the degraded QuickBird-1 dataset. For better comparison, the red square area is enlarged and then displayed at the bottom left of the fusion image. As can be observed, Fig. 6d–j methods have more inferior pansharpening results than CAE and proposed methods.

Figure 6i–j suffer from spatial distortion. Figure 6m suffers from spatial and spectral distortions. The fusion result of the PNN method is depicted in Fig. 6n, which produces some unnatural color compared with the reference image. Furthermore, Fig. 6l CAE and proposed method Fig. 6o look most similar to the reference image Fig. 6a, but the proposed method performs better in terms of spectral and spatial fidelity. Similar observations can be made regarding the experimental results from the QuickBird-2 dataset. Figure 7 displays the fusion results of the degraded QuickBird-2 dataset. For better visual comparison, the red rectangle area is enlarged and then displayed at the bottom of the selected area; thus, the proposed and CAE methods have performed better visual effects.

Table 3 Numerical results of the full reference QuickBird-2 dataset

Method	CC(1)	UIQI(1)	RMES(0)	RASE(0)	SAM(0)	ERGAS(0)	Q4(1)
IHS	0.94819	0.95489	16.013	10.565	2.9752	2.6898	0.85877
AIHS	0.96171	0.96337	13.465	9.8843	2.8699	2.2927	0.8984
PCA	0.95507	0.95906	15.256	10.066	2.9418	2.557	0.84805
BDSD	0.9425	0.9547	16.41	10.828	3.0727	2.7548	0.85536
PRACS	0.96067	0.97082	13.29	8.7685	3.0473	2.2505	0.89358
SFIM	0.92309	0.9339	19.612	12.94	3.0571	3.2638	0.82182
Indusion	0.93623	0.94457	17.754	11.714	2.9083	2.9657	0.82554
MTF-GLP	0.93661	0.94967	17.309	11.42	2.9938	2.9018	0.85643
CAE	0.96082	0.96583	18.999	12.535	2.871	3.1639	0.8957
MSGF	0.9372	0.94047	18.635	12.295	2.9182	3.1012	0.90274
PNN	0.92756	0.93768	19.765	13.041	3.8949	3.2934	0.84088
Proposed	**0.96366**	**0.97157**	**13.286**	**8.7657**	**2.7686**	**2.2271**	**0.90617**

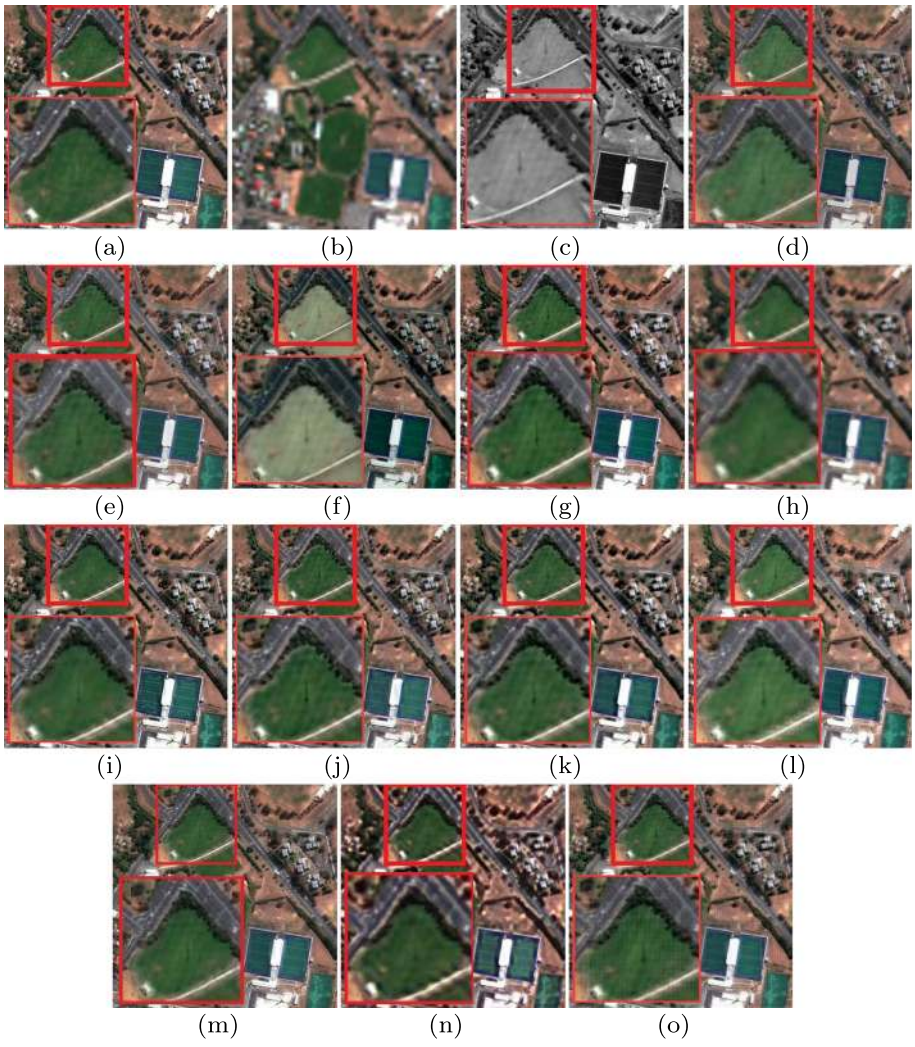

Fig. 8 Fusion results of the full reference GeoEye-1 dataset. **a** Reference image (256×256). **b** Degraded MS image (64×64). **c** PAN image (256×256). **d** IHS method. **e** AIHS method. **f** PCA method. **g** BDSD method. **h** PRACS method. **i** SFIM method. **j** Indusion method. **k** MTF-GLP method. **l** CAE method. **m** MSGF method. **n** PNN method. **o** Proposed method

Table 4 Numerical results of the full reference GeoEye-1 dataset

Method	CC(1)	UIQI(1)	RMES(0)	RASE(0)	SAM(0)	ERGAS(0)	Q4(1)
IHS	0.94344	0.94327	19.622	18.015	5.6283	4.6538	0.8837
AIHS	0.95117	0.95566	17.432	16.004	5.2561	4.0669	0.90041
PCA	0.71449	0.80704	36.19	33.227	14.476	8.9614	0.68996
BDSD	0.95674	0.96131	17.53	16.095	6.2016	4.0897	0.91719
PRACS	0.94091	0.9545	17.857	16.395	4.6631	4.2619	0.89865
SFIM	0.7315	0.6906	57.368	52.671	6.0017	14.559	0.89337
Indusion	0.94438	0.94489	20.182	18.529	5.8298	4.4905	0.88997
MTF-GLP	0.95403	0.95517	18.87	17.325	6.6894	4.248	0.90944
CAE	0.95082	0.94401	26.323	24.168	5.389	5.9969	0.8993
MSGF	0.9575	0.9571	16.217	14.791	**4.6017**	3.7496	0.92439
PNN	0.93825	0.91003	27.623	25.361	9.9312	5.6757	0.86618
Proposed	**0.95794**	**0.96376**	**16.103**	**14.784**	4.8976	**3.7145**	**0.91728**

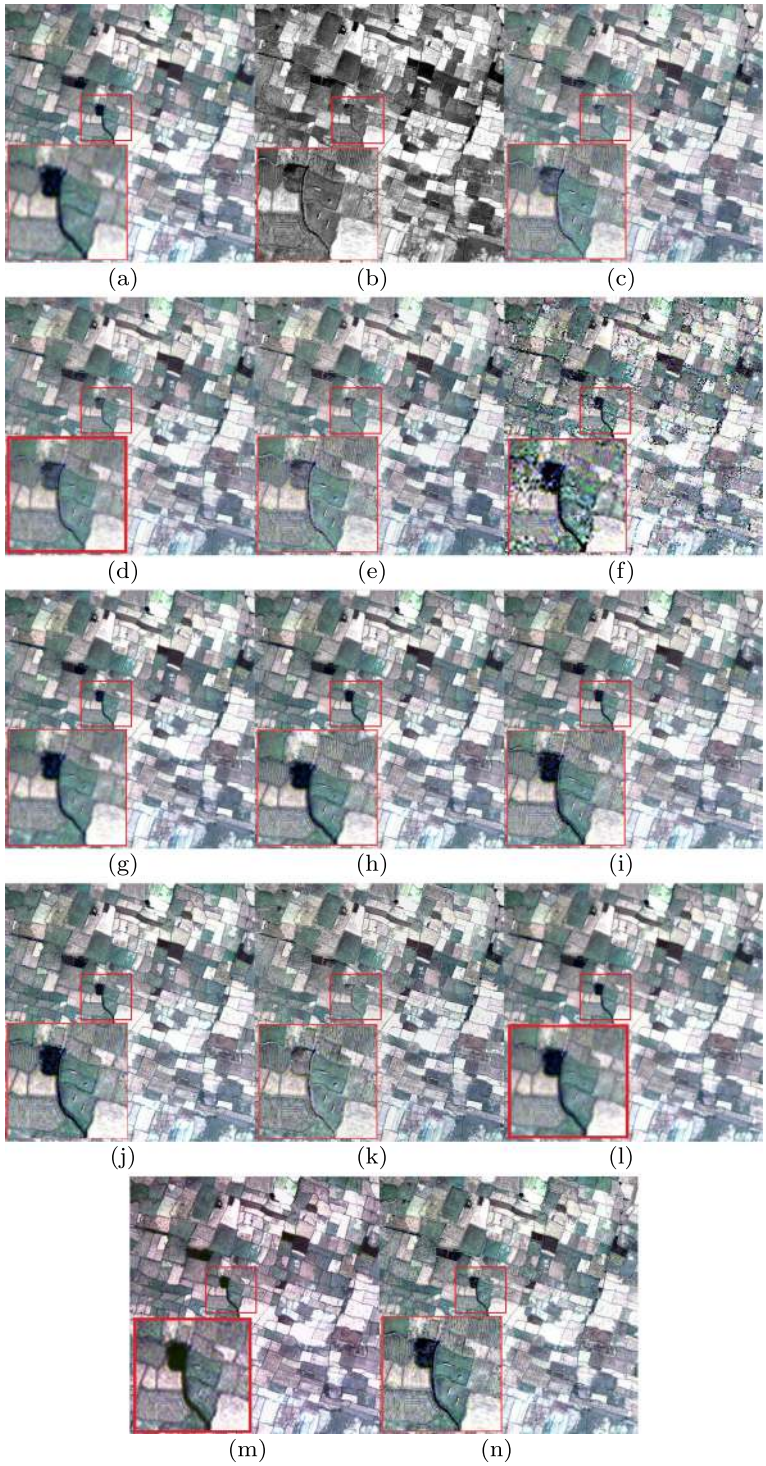

Fig. 9 Fusion results of the real QuickBird-1 dataset. **a** Upsampled MS image (1024×1024). **b** PAN image (1024×1024). **c** IHS method. **d** AIHS method. **e** PCA method. **f** BDSD method. **g** PRACS method. **h** SFIM method. **i** Indusion method. **j** MTF-GLP method. **k** CAE method. **l** MSGF method. **m** PNN method. **n** Proposed method

In terms of objective evaluation, the numerical indexes of fused images for Figs. 6 and 7 are computed and reported in Tables 2 and 3, respectively. From both tables, it is clear that our method can contribute to the best values in terms of quality indexes.

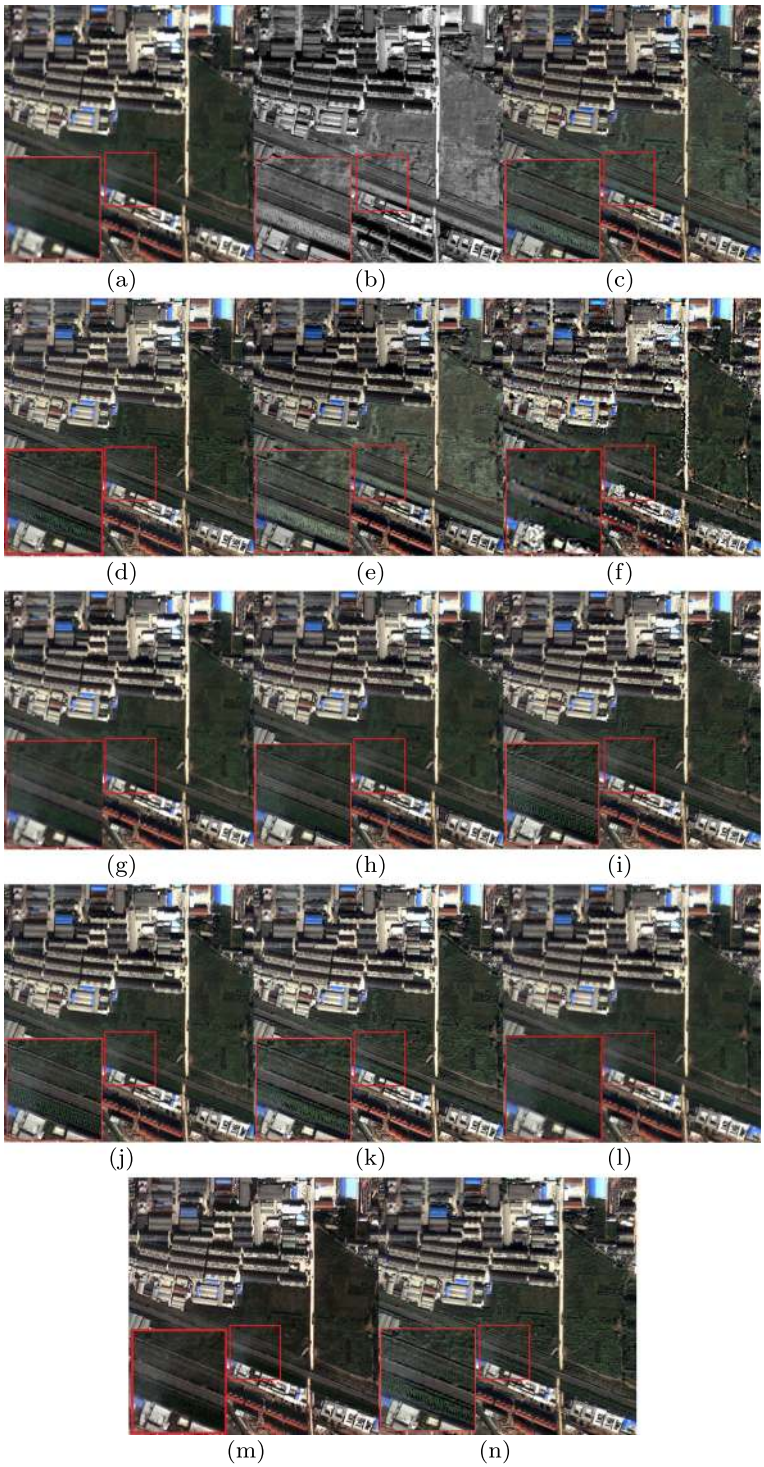

Fig. 10 Fusion results of the real QuickBird-2 dataset. **a** Upsampled MS image (1024×1024). **b** PAN image (1024×1024). **c** IHS method. **d** AIHS method. **e** PCA method. **f** BDSD method. **g** PRACS method. **h** SFIM method. **i** Indusion method. **j** MTF-GLP method. **k** CAE method. **l** MSGF method. **m** PNN method. **n** Proposed method

Table 5 Numerical results of the real QuickBird-1 dataset

Method	$D_s(0)$	$D_\lambda(0)$	QNR(1)
Proposed	**0.1289**	**0.0008**	**0.9699**
PNN	0.1957	0.2305	0.8976
MSGF	0.151	0.019	0.9603
CAE	0.16027	0.001	0.9620
MTF-GLP	0.15097	0.068214	0.94468
Indusion	0.13001	0.042951	0.95876
SFIM	0.13109	0.052531	0.95375
PRACS	0.1834	0.0382	0.9469
BDSD	0.1998	0.1235	0.92189
IHS	0.24386	0.06756	0.91959
AIHS	0.22026	0.043312	0.93374
PCA	0.24261	0.05516	0.92611

3.2.2 Experiment on degraded GeoEye dataset

Figure 8 displays the fusion results of the degraded GeoEye-1 dataset. The red square area is enlarged and then displayed at the bottom left of the fusion image. As shown in Fig. 8f, PCA produced seedy color in the fused image, and Fig. 8f–h suffer from the spectral distortion. Here, it can be seen that the SFIM, Indusion, and MTF-GLP methods perform well, as shown in Fig. 8i–k. We can also observe from Fig. 8l that the result of the CAE method has a color problem at the vegetation area compared with the reference image. The colors of the fusion image for MSGF and PNN methods have remarkable distortion, as shown in Fig. 8m, n. Overall, the proposed method created the fused image, with appropriate spectral and spatial resolution, as shown in Fig. 8o compared with others.

The numerical indexes of fused images for Fig. 8 are computed and reported in Table 4. From the table, it is clear that our method can contribute to the best values in the most quality indexes.

3.3 Fusion results of real datasets (no reference)

Regarding real datasets, two kinds of real datasets (QuickBird, GeoEye) were implemented, and the sizes of the MS image and the PAN image are 256×256 and 1024×1024, respectively.

Table 6 Numerical results of the real QuickBird-2 dataset

Method	$D_s(0)$	$D_\lambda(0)$	QNR(1)
Proposed	0.13935	**0.081319**	**0.96598**
PNN	0.181486	0.095215	0.96407
MSGF	0.082709	0.102	0.96379
CAE	0.19834	0.093565	0.94344
MTF-GLP	0.13929	0.092786	0.93708
Indusion	**0.030669**	0.087048	0.96348
SFIM	0.066346	0.14364	0.93699
PRACS	0.124	0.09274	0.9646
BDSD	0.3088	0.0974	0.9593
IHS	0.2534	0.10823	0.91008
AIHS	0.16242	0.095591	0.93691
PCA	0.28645	0.13137	0.90127

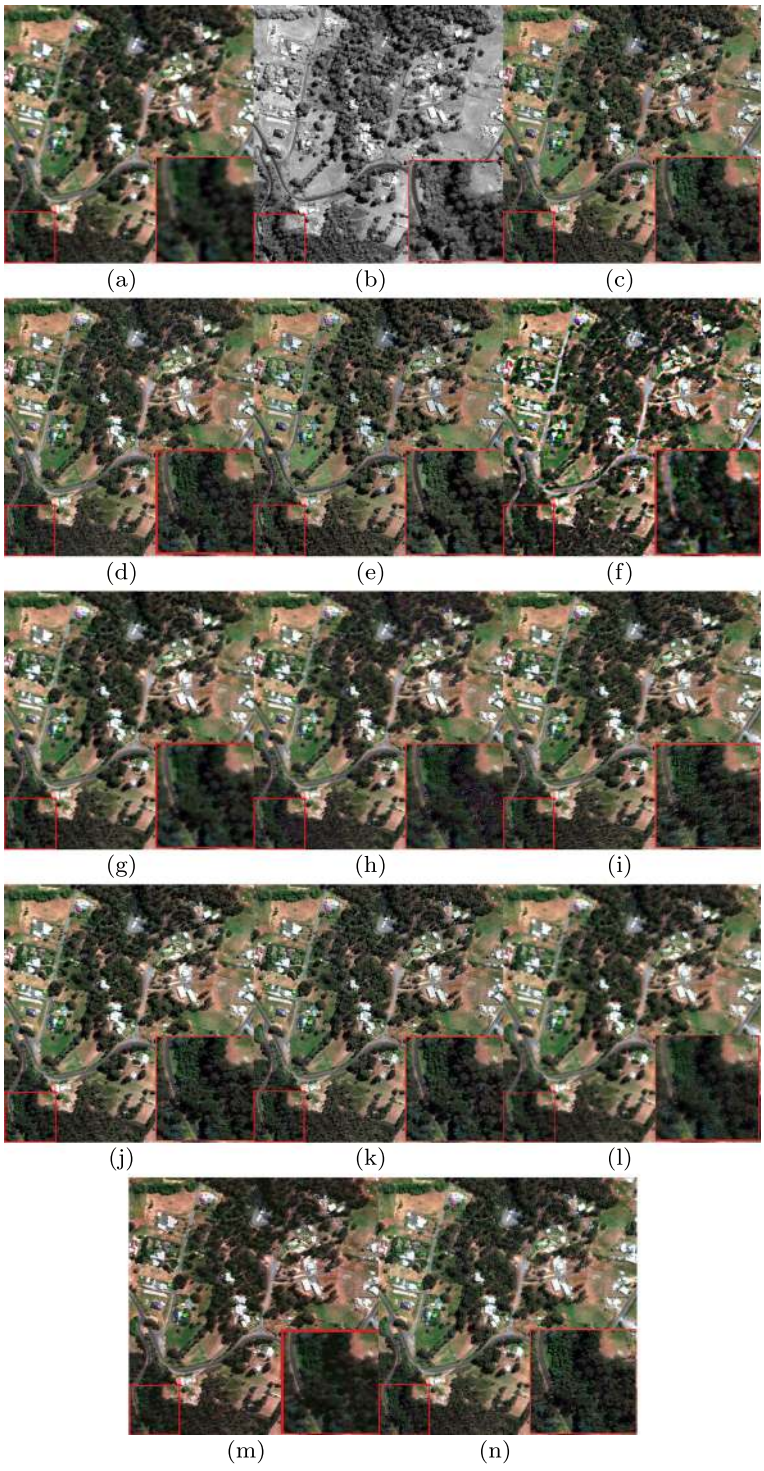

Fig. 11 Fusion results of the real GeoEye-1 dataset. **a** Upsampled MS image (1024×1024). **b** PAN image (1024×1024). **c** IHS method. **d** AIHS method. **e** PCA method. **f** BDSD method. **g** PRACS method. **h** SFIM method. **i** Indusion method. **j** MTF-GLP method. **k** CAE method. **l** MSGF method. **m** PNN method. **n** Proposed method

3.3.1 Experiments on real QuickBird datasets

Two pairs of real QuickBird satellite datasets were examined; for better visual comparison, the red square area is enlarged and then displayed at the bottom left of the fusion image. Figure 9 displays the fusion results of real QuickBird-1 dataset.

The fusion results of all methods improved, but the CS-based method and CAE method suffer from spectral distortion, as shown in Fig. 9c, e, and k. The BDSD fusion method has remarkable distortions. For SFIM, Indusion, and MTF-GLP methods, they can achieve relatively better results regarding spectral resolution than others, as shown in Fig. 9h–j. The MSGF method suffers from spatial distortion, as shown in Fig. 9l, and the colors of the fusion image for the PNN method have remarkable distortions. However, the fusion result of the proposed method can perform better than others, as shown in Fig. 9o. Similarly, the observations can be done regarding the experimental results from the real QuickBird-2 dataset. Figure 10 displays the fusion results of the real QuickBird-2 dataset. The CS-based methods suffer from spectral distortion, as shown in Fig. 10c, e. The BDSD fusion method has remarkable distortions as shown in Fig. 10e. The CAE method can achieve well concerning the spatial aspect but still has a lighter color in the vegetation area compared with the upsampled MS image, as shown in Fig. 10k.

The fusion results of SFIM, Indusion, MTF-GLP, MSGF, PNN, and proposed methods improved in both aspects of spectral and spatial.

The numerical measurements of real data fused images for Figs. 9 and 10 are computed and listed in Tables 5 and 6, respectively.

Table 5 illustrates the proposed method performed the best value in terms of D_λ and D_s. Thus, our method showed the best value in terms of D_λ and QNR, as reported in Table 6.

3.3.2 Experiment on real GeoEye dataset

Figure 11 displays the fusion results of the real GeoEye-1 dataset. The selected red square area is enlarged and then displayed at the bottom right of the fusion image for better visual comparison. As shown in Fig. 11c–e, these methods can perform well regarding spatial aspect but suffer from spectral distortion, and Fig. 11f–i and l, suffer from notable spectral and spatial distortion. Here, it can be seen that the MTF-GLP, CAE, and proposed methods perform well, as shown in Fig. 11j, k, and o.

Overall, the proposed method created the fused image, with appropriate spectral and spatial resolution.

Table 7 Numerical results of the real GeoEye-1 dataset

Method	D_s (0)	D_λ (0)	QNR(1)
Proposed	0.1083	0.0806	**0.95706**
PNN	0.094048	**0.060553**	0.95616
MSGF	0.1102	0.095	0.95434
CAE	0.13473	0.14395	0.93975
MTF-GLP	0.12752	0.13915	0.92936
Indusion	0.050782	0.11372	0.9568
SFIM	0.13797	0.20061	0.91727
PRACS	**0.0464**	0.0936	0.9566
BDSD	0.1242	0.1269	0.9446
IHS	0.1593	0.1277	0.92838
AIHS	0.123	0.123	0.9452
PCA	0.1869	0.1555	0.92199

The numerical indexes of fused images for Fig. 11 are computed and reported in Table 7. From Table 7, the PNN method can perform the best value in terms of D_λ, followed by our method. Overall, our method can still contribute to the best values concerning quality indexes.

4 Conclusion

In this paper, we have proposed a pansharpening technique based on a convolutional autoencoder with AIHS and a multi-scale guided filter. The proposed method first trained the convolutional autoencoder to learn the relationship between the panchromatic image and its degraded version. The trained network is used to enhance the intensity component. Furthermore, the multi-scale guided filter is used to enhance the original panchromatic image. Several experiments were conducted, and the article has put in place the results of the experiment. The outcomes of this research are, first, in terms of visual aspect, the proposed method includes more of the spectral detail of the MS image and spatial detail of the panchromatic image than existing fusion methods. Second, the quality indexes of our method show significant enhancements compared with comparative methods. Overall, the model developed in this research was able to preserve appropriate spatial and spectral aspects of fusion image compared with comparative methods in both aspects, subjective and objective evaluations.

Abbreviations

PAN: Panchromatic image; MS: Multi-spectral image; CNN: Convolutional neural network; AIHS: Adaptive intensity-hue-saturation; CS: Component substitution; MRA: Multi-resolution analysis; PCA: Principal component analysis; GS: Gram-Schimidt; BT: Brovey transform; SFIM: Smoothing Filter-based Intensity Modulation; MTF-GLP: Generalized Laplacian pyramid; MSGF: Multi-scale guided filer; RCNN: Residual convolutional neural network; CAE: Convolutional autoencoder; GF: Guided filter; BDSD: Band-dependent spatial-detail; PRACS: Partial replacement adaptive CS; PNN: Pansharpening by convolutional neural networks. CC: Correlation coefficient; UIQI: Universal Image Quality Index; RMSE: Root mean square error; RASE: Relative average spectral error; SAM: Spectral Angle Mapper; ERGAS: Erreur Relative Globale Adimensionnelle de Synthese. D_s: Spatial distortion ; D_λ: Spectral distortion; QNR: Quality with no reference

Acknowledgements

No other acknowledgments.

Authors' contributions

AAL and YS conceptualized and carried out the implementation; AAL, KZ, and AM wrote and reviewed the paper; AS and MW were in charge of the overall research and contributed to the paper writing; YS contributed to funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 61771380, 61906145, U1730109, 91438103, 61771376, 61703328, 91438201, U1701267, 61703328), the Equipment pre-research project of the 13th Five-Years Plan (Nos. 6140137050206, 414120101026, 6140312010103, 6141A020223, 6141B06160301, 6141B07090102), the Major Research Plan in Shaanxi Province of China (Nos. 2017ZDXM-GY-103,017ZDCXL-GY-03-02), the Foundation of the State Key Laboratory of CEMEE (Nos. 2017K0202B, 2018K0101B 2019K0203B, 2019Z0101B), and the Science Basis Research Program in Shaanxi Province of China (Nos. 16JK1823, 2017JM6086, 2019JQ-663).

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Artificial Intelligence, Xidian University, No. 2 South Taibai Road, 710071 Xian, China. [2]School of Information Science and Engineering, Shandong Normal University, Jinan 250358 Shandong, China. [3]Key Laboratory of Radar Signal Processing, Xidian University, No. 2 South Taibai Road, Xian, China.

Received: 20 October 2020 Accepted: 10 June 2021
Published online: 19 July 2021

References

1. K. Zhang, M. Wang, S. Yang, L. Jiao, Spatial–spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **11**(4), 1030–1040 (2018)
2. A. Al Smadi, A. Abugabah, in *Proceedings of the 2018 the 2nd International Conference on Video and Image Processing*, Intelligent information systems and image processing: a novel pan-sharpening technique based on multiscale decomposition, (2018), pp. 208–212
3. F. Zhang, K. Zhang, Superpixel guided structure sparsity for multispectral and hyperspectral image fusion over couple dictionary. Multimedia Tools Appl. **79**(7), 4949–4964 (2020)
4. J. Xu, H. Zhao, P. Yin, D. Jia, G. Li, Remote sensing classification method of vegetation dynamics based on time series Landsat image: a case of opencast mining area in China. EURASIP J. Image Video Process. **2018**(1), 113 (2018)
5. A. Alsmadi, S. Yang, K. Zhang, Pansharpening via deep guided filtering network. Int. J. Image Process. Vis. Commun. **5**, 1–8 (2018)
6. G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, L. Wald, A critical comparison among pansharpening algorithms. IEEE Trans. Geosci. Remote Sens. **53**(5), 2565–2586 (2014)
7. L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, L. M. Bruce, Comparison of pansharpening algorithms: outcome of the 2006 GRS-S data-fusion contest. IEEE Trans. Geosci. Remote Sens. **45**(10), 3012–3021 (2007)
8. A. Mookambiga, V. Gomathi, Comprehensive review on fusion techniques for spatial information enhancement in hyperspectral imagery. Multidim. Syst. Sign. Process. **27**(4), 863–889 (2016)
9. F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, J. A. Benediktsson, in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Model based pansharpening method based on TV and MTF deblurring (IEEE, 2015), pp. 33–36
10. W. Li, Y. Li, Q. Hu, L. Zhang, Model-based variational pansharpening method with fast generalized intensity–hue–saturation. J. Appl. Remote. Sens. **13**(3), 036513 (2019)
11. T.-M. Tu, S.-C. Su, H.-C. Shyu, P. S. Huang, A new look at IHS-like image fusion methods. Inf. Fusion. **2**(3), 177–186 (2001)
12. P. Kwarteng, A. Chavez, Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. Photogramm. Eng. Remote Sens. **55**(1), 339–348 (1989)
13. B. Aiazzi, S. Baronti, M. Selva, Improving component substitution pansharpening through multivariate regression of ms + pan data. IEEE Trans. Geosci. Remote Sens. **45**(10), 3230–3239 (2007)
14. A. R. Gillespie, A. B. Kahle, R. E. Walker, Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques. Remote Sens. Environ. **22**(3), 343–365 (1987)
15. J. Liu, Smoothing filter-based intensity modulation: a spectral preserve image fusion technique for improving spatial details. Int. J. Remote Sens. **21**(18), 3461–3472 (2000)
16. B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, M. Selva, MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. Photogramm. Eng. Remote Sens. **72**(5), 591–596 (2006)
17. M. M. Khan, J. Chanussot, L. Condat, A. Montanvert, Indusion: fusion of multispectral and panchromatic images using the induction scaling technique. IEEE Geosci. Remote Sens. Lett. **5**(1), 98–102 (2008)
18. K. He, J. Sun, X. Tang, Guided image filtering. IEEE Trans. Pattern. Anal. Mach. Intell. **35**(6), 1397–1409 (2012)
19. Y. Yang, W. Wan, S. Huang, F. Yuan, S. Yang, Y. Que, Remote sensing image fusion based on adaptive IHS and multiscale guided filter. IEEE Access. **4**, 4573–4582 (2016)
20. W. Shi, S. Liu, F. Jiang, D. Zhao, Z. Tian, Anchored neighborhood deep network for single-image super-resolution. EURASIP J. Image Video Process. **2018**(1), 34 (2018)
21. G. Scarpa, S. Vitale, D. Cozzolino, Target-adaptive CNN-based pansharpening. IEEE Trans. Geosci. Remote Sens. **56**(9), 5443–5457 (2018)
22. S. Huang, J. Wu, Y. Yang, P. Lin, Multi-frame image super-resolution reconstruction based on spatial information weighted fields of experts. Multidim. Syst. Sign. Process. **31**(1), 1–20 (2020)
23. S. Baghersalimi, B. Bozorgtabar, P. Schmid-Saugeon, H. K. Ekenel, J.-P. Thiran, Dermonet: densely linked convolutional neural network for efficient skin lesion segmentation. EURASIP J. Image Video Process. **2019**(1), 71 (2019)
24. A. Mehmood, M. Maqsood, M. Bashir, Y. Shuyuan, A deep Siamese convolution neural network for multi-class classification of Alzheimer disease. Brain Sci. **10**(2), 84 (2020)
25. Y. Wang, H. Bai, L. Zhao, Y. Zhao, Cascaded reconstruction network for compressive image sensing. EURASIP J. Image Video Process. **2018**(1), 77 (2018)
26. Y. Rao, L. He, J. Zhu, in *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)*, A residual convolutional neural network for pan-shaprening (IEEE, 2017), pp. 1–4
27. W. Huang, L. Xiao, Z. Wei, H. Liu, S. Tang, A new pan-sharpening method with deep neural networks. IEEE Geosci. Remote Sens. Lett. **12**(5), 1037–1041 (2015)
28. A. Azarang, H. E. Manoochehri, N. Kehtarnavaz, Convolutional autoencoder-based multispectral image fusion. IEEE Access. **7**, 35673–35683 (2019)
29. S. Dolgikh, Spontaneous concept learning with deep autoencoder. Int. J. Comput. Intell. Syst. **12**(1), 1–12 (2018)
30. W. CARPER, T. LILLESAND, R. KIEFER, The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data. Photogramm. Eng. Remote Sens. **56**(4), 459–467 (1990)
31. S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, T. Wittman, An adaptive IHS pan-sharpening method. IEEE Geosci. Remote Sens. Lett. **7**(4), 746–750 (2010)
32. K. He, J. Sun, X. Tang, in *European Conference on Computer Vision*, Guided image filtering (Springer, 2010), pp. 1–14
33. C. N. Ochotorena, Y. Yamashita, Anisotropic guided filtering. IEEE Trans. Image Process. **29**, 1397–1412 (2019)
34. Y. Bengio, I. Goodfellow, A. Courville, *Deep Learning*, vol. 1. (MIT Press, Massachusetts, USA, 2017)
35. Y. Song, W. Wu, Z. Liu, X. Yang, K. Liu, W. Lu, An adaptive pansharpening method by using weighted least squares filter. IEEE Geosci. Remote Sens. Lett. **13**(1), 18–22 (2015)
36. A. Garzelli, F. Nencini, L. Capobianco, Optimal MMSE pan sharpening of very high resolution multispectral images. IEEE Trans. Geosci. Remote Sens. **46**(1), 228–236 (2007)

37. J. Choi, K. Yu, Y. Kim, A new adaptive component-substitution-based satellite image fusion by using partial replacement. IEEE Trans. Geosci. Remote Sens. **49**(1), 295–309 (2010)
38. G. Masi, D. Cozzolino, L. Verdoliva, G. Scarpa, Pansharpening by convolutional neural networks. Remote Sens. **8**(7), 594 (2016)
39. M. Imani, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **11**(12), 4994–5004 (2018)
40. Z. Wang, A. C. Bovik, A universal image quality index. IEEE Signal Process. Lett. **9**(3), 81–84 (2002)
41. P. Jagalingam, A. V. Hegde, A review of quality metrics for fused image. Aquat. Procedia. **4**, 133–142 (2015)
42. P. Mhangara, W. Mapurisa, N. Mudau, Comparison of image fusion techniques using satellite pour l'Observation de la Terre (SPOT) 6 satellite imagery. Appl. Sci. **10**(5), 1881 (2020)
43. G. P. Petropoulos, K. P. Vadrevu, C. Kalaitzidis, Spectral angle mapper and object-based classification combined with hyperspectral remote sensing imagery for obtaining land use/cover mapping in a Mediterranean region. Geocarto Int. **28**(2), 114–129 (2013)
44. F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, J. A. Benediktsson, Quantitative quality evaluation of pansharpened imagery: consistency versus synthesis. IEEE Trans. Geosci. Remote Sens. **54**(3), 1247–1259 (2015)
45. L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, M. Selva, Multispectral and panchromatic data fusion assessment without reference. Photogramm. Eng. Remote Sens. **74**(2), 193–200 (2008)
46. T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti, L. Wald, Image fusion–the arsis concept and some successful implementation schemes. ISPRS J. Photogramm. Remote. Sens. **58**(1-2), 4–18 (2003)

Publisher's Note