A generative model method for unsupervised multispectral image fusion in remote sensing

Arian Azarang1 · Nasser Kehtarnavaz1

Received: 3 December 2020 / Revised: 10 April 2021 / Accepted: 28 May 2021 / Published online: 19 June 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract
This paper presents a generative model method for multispectral image fusion in remote sensing which involves training without supervision. This method eases the supervision of learning and also uses a multi-objective loss function to achieve image fusion. The loss function used incorporates both spectral and spatial distortions. Two discriminators are designed to minimize the spectral and spatial distortions of the generative output. Extensive experimentations are conducted using three public domain datasets. The comparison results across four reduced-resolution and three full-resolution objective metrics show the superiority of the developed method over several recently developed methods.

Keywords Generative model · Image fusion in remote sensing · Deep learning

1 Introduction
The Earth observation satellites capture the earth’s surface information in different modalities including spectral, spatial, and temporal. In the spectral modality, MultiSpectral (MS) data or images are captured at different wavelengths with low spatial resolutions. In the spatial modality, PAN-chromatic (PAN) data or images are captured over a long range of wavelengths with high spatial resolutions. The fusion of MS and PAN data, named pansharpening in the remote sensing literature [1–8], involves combining the spatial and spectral modalities. In multispectral image fusion, the objective is to recover higher spatial resolutions for multispectral images by reducing the degradation processes that occur during data collection. This recovery can be viewed as an inverse image processing problem. This inverse problem is ill-posed, meaning that there exist a large number of high-resolution images that can get mapped to low-resolution input images. Image fusion techniques attempt to narrow down the search towards obtaining high-resolution images.

A number of review articles [9, 10] have categorized pansharpening or multispectral image fusion methods into three main groups consisting of (1) Component Substitution (CS) methods, e.g., [11–13], (2) Multi-Resolution Analysis (MRA) methods, e.g., [14–16], and (3) Model-Based (MB) methods, e.g., [17–19]. The main difference between the first two groups is in the way detail map computation is done. In CS methods, a detail map is acquired by subtracting a PAN image from a linear/nonlinear combination of Low-Resolution MS (LRMS) images, whereas in MRA methods, this map is obtained by subtracting a Low-Resolution PAN (LRPAN) image from the PAN image. A LRPAN can be computed by applying decomposition methods such as wavelet transform of the PAN image. The third group of pansharpening methods involves using a Bayesian model and posing the fusion as an optimization problem.

In the last few years, deep learning models have been applied to multispectral image fusion generating better outcomes than conventional methods. An initial attempt was made in [20] by solving the pansharpening problem via a deep neural network framework where the nonlinear relationship between the low-resolution and high-resolution images was formulated as a denoising autoencoder. In [21], a three-layer convolutional neural network was designed to turn the MS image fusion problem into a super-resolution problem. The concept of residual learning in MS image fusion was first introduced in [22], where a deep convolutional neural network was used. In [23], a deep denoising
The methods mentioned above primarily use a single objective learning to optimize network parameters and generalize its capability. However, other loss functions that can represent both modalities (spatial and spectral) have recently gained more attention. For instance, in [32], based on the correlation maps between MS target images and PAN input images, a loss function was designed to minimize the artifacts of fused images. Also, in [33], it was shown that although a linear combination of MS bands could be estimated from the PAN image, a rather large difference in luminance was resulted. Thus, certain objects could not be differentiated properly. To address this issue, a color-aware perceptual (CAP) loss function was designed to obtain the features of a pre-trained VGG network that were more sensitive to spatial details and less sensitive to color differences. The aforementioned methods rely on the availability of GT data for regularizing the network parameters. However, in practice, such data is not available [27].

The work presented in this paper eases the above two limitations of the existing deep learning models for remote sensing image fusion. The first limitation involves dependency of GT data towards training a network and the second limitation involves the use of a generic loss function. The first limitation is eased by an unsupervised learning strategy based on generative adversarial networks. What is meant by unsupervised learning here is that the label/reference target is not available for training the deep model. A key point for unsupervised learning is that while the data passed through the deep model are abundant, the targets and labels are quite sparse or even non-existent. The second limitation is eased by designing a multi-objective loss function to reflect both spatial and spectral attributes at the same time.

Section 2 presents the formulation of the developed generative model method as well as its architecture. Section 3
covers the datasets and objective metrics used in this paper. The experimental results covering both objective metrics and visual comparisons are then provided in Sect. 4. Finally, the paper is concluded in Sect. 5.

2 Generative model method

This section provides a description of the developed generative model method. To set the stage, let us begin with the general framework of CS methods. The CS framework can be mathematically expressed by the following equation:

\[
\tilde{M}_k = M_k + g_k (P - I_k)
\]  (1)

where \(\tilde{M}_k\) and \(M_k\) denote the high-resolution and upsampled low-resolution MS images, respectively, \(g_k\)'s are injection gains for spectral bands, \(P\) denotes the PAN image, and \(I_k\) is the \(k\)-th intensity component defined as

\[
I_k = F(\tilde{M}_k)
\]  (2)

where \(F(.)\) denotes a linear/nonlinear combination of spectral bands [1–5].

2.1 Generative adversarial networks

The use of Generative Adversarial Networks (GANs) has been steadily growing. Furthermore, these networks have facilitated the recognition of new categories of learning schemes leading to the synthesis of realistic data [34]. In the setting of the GAN structure, rather than a single deep neural network (DNN), training encompasses two DNNs, a “generator” and a “discriminator” architecture, where the former synthesizes realistic data given an input, and the latter classifies inputs as real or synthetic.

In the original form of the GAN architecture [34], the generator is initialized with randomized input noise yielding several realizations depending on the noise statistics. For image enhancement problems, a specific type of GANs, called conditional GANs (cGANs), is developed since the input to the generator is the image itself, while it could be dissimilar from the output such as an edge map [35]. A noteworthy paper that shows the abilities of GAN in inverse image processing problems is the super-resolution GAN (SRGAN) architecture [35].

The GAN architecture for the inverse image processing problem here involves an iterative training or learning process that alternates between synthesizing of a high-quality image \(M^S\) given a low-quality input image \(M^\text{IN}\) performed by the generator \(G\), and the classification of the high-quality image as real \(M^R\) or synthetic \(M^S\) performed by the discriminator \(D\). Thus, training a GAN translates into optimizing a min–max problem where the aim is to estimate the network parameters (weights and biases) of the generator \(\theta_G\) and the discriminator \(\theta_D\) based on the following equation:

\[
\min_{\theta_G} \max_{\theta_D} \mathbb{E} \left[ \log D_{\theta_D}(M^R) \right] + \mathbb{E} \left[ \log \left( 1 - D_{\theta_D} \left( G_{\theta_G}(M^\text{IN}) \right) \right) \right]
\]  (3)

Fig. 2 Sample PAN-MS image pairs for three datasets: Pleiades-1A; a MS, b PAN, WorldView-2; c MS, d PAN, and GeoEye-1; e MS, f PAN.
2.2 Fusion framework

The main contribution in this work is to formulate the fusion problem as a multi-objective loss function represented by a deep generative network in which both the spectral and spatial distortions are minimized simultaneously.

2.2.1 Spectral preservation learning process

For minimizing the spectral distortion in the fused image, a spectral metric is used to deal with spectral consistency. For this purpose, a discriminator for the learning process is considered, named spectral discriminator here. The MS image data at the original resolution is used as the input of this discriminator. Initially, the output of the generator is inputted to the spectral discriminator. The following objective function is then used to minimize the spectral distortion of the fused image:

\[ \mathcal{L}_1 = UIQI(\hat{M}_k^E, M_k) \]  

where \( UIQI(\ldots) \) denotes the Universal Image Quality Index (UIQI) as described in [36], and \( \hat{M}_k^E \) and \( M_k \) are the estimated high-resolution MS image at the output of the generator and the MS input image at the original resolution, respectively.

2.2.2 Spatial preservation learning process

Another discriminator is considered for the minimization of spatial distortion, named spatial discriminator here. To inject spatial details into the fused image and by noting that the PAN image denotes the reference spatial information, the PAN image at the original resolution is used as the input to the discriminator. The following loss function is then used during the training phase of the generative model:

\[ \mathcal{L}_2 = UIQI(\hat{I}_k^E, \hat{P}_k) \]  

where \( \hat{I}_k^E \) is the linear combination of estimated high-resolution MS images at the output of the generator and \( \hat{P}_k \) is the histogram matched PAN image with respect to the \( k \)-th spectral band. The learning process of the developed method is illustrated in Fig. 1.

3 Experimental studies and discussion

For our experimental studies, the following three public domain datasets are used: Pleiades-1A, WorldView-2, and GeoEye-1. The geographical areas for the Pleiades-1, WorldView-2, and GeoEye-1 datasets correspond to Melbourne-Australia, Paris-France, and Washington-USA, respectively. The MS data for each dataset has four different bands including Blue (B), Green (G), Red (R), and Near InfraRed (NIR). Since the original datasets are quite large, they are divided into \( 1024 \times 1024 \) and \( 256 \times 256 \) subimages for PAN and MS, respectively. For Pleiades-1A, WorldView-2, and GeoEye-1, 400, 300, 500 pairs of MS-PAN images were used for the

| Table 1 Average performance of reduced-resolution mode of pleiades-1A dataset |
|-----------------------------|------------------|---------|---------|--------|
| SAM | ERGAS | CC | Q4 |
| BDSD | 3.66 | 4.12 | 0.96 | 0.86 |
| AIHS | 3.24 | 3.86 | 0.96 | 0.87 |
| DNN | 3.15 | 3.82 | 0.96 | 0.92 |
| CAE | 3.05 | 3.75 | 0.96 | 0.92 |
| MTF-GLP | 3.63 | 3.95 | 0.95 | 0.87 |
| FDIF | 3.43 | 4.01 | 0.95 | 0.86 |
| Multi-Objective | 3.03 | 2.79 | 0.97 | 0.93 |
| Developed | **2.53** | **2.36** | **0.97** | **0.94** |
| Ideal | 0 | 0 | 1 | 1 |

Best values are indicated in bold face

| Table 2 Average Performance of Reduced-Resolution Mode of WorldView-2 Dataset |
|-----------------------------|------------------|---------|---------|--------|
| SAM | ERGAS | CC | Q4 |
| BDSD | 4.32 | 2.12 | 0.95 | 0.87 |
| AIHS | 4.14 | 1.86 | 0.95 | 0.88 |
| DNN | 3.83 | 1.73 | 0.97 | 0.90 |
| CAE | 3.55 | 1.70 | 0.97 | 0.91 |
| MTF-GLP | 4.60 | 2.34 | 0.94 | 0.86 |
| FDIF | 4.43 | 2.30 | 0.93 | 0.85 |
| Multi-Objective | 3.30 | 1.64 | 0.97 | 0.92 |
| Developed | **3.16** | **1.41** | **0.97** | **0.92** |
| Ideal | 0 | 0 | 1 | 1 |

Best values are indicated in bold face

| Table 3 Average performance of reduced-resolution mode of geoeye-1 dataset |
|-----------------------------|------------------|---------|---------|--------|
| SAM | ERGAS | CC | Q4 |
| BDSD | 2.76 | 1.88 | 0.93 | 0.86 |
| AIHS | 2.34 | 1.78 | 0.93 | 0.88 |
| DNN | 2.25 | 1.64 | 0.94 | 0.90 |
| CAE | 2.15 | 1.55 | 0.94 | 0.90 |
| MTF-GLP | 2.60 | 2.02 | 0.93 | 0.86 |
| FDIF | 2.43 | 2.12 | 0.93 | 0.88 |
| Multi-Objective | 2.03 | 1.55 | 0.97 | 0.93 |
| Developed | **1.84** | **1.30** | **0.97** | **0.95** |
| Ideal | 0 | 0 | 1 | 1 |

Best values are indicated in bold face
In the experiments reported here, respectively. Sample images of each dataset are shown in Fig. 2. Note that the regions are selected from different surface indices, e.g., coastal, urban, and jungle areas.

### 3.1 Reduced resolution metrics

One of the widely used metrics at full-reference mode is Spectral Angle Mapper (SAM) \[9\]. The color differences between the fused and MS images are characterized by this metric. The following equation is used to compute local SAM values:

\[
\text{SAM}(x, y) = \frac{\langle F, M \rangle}{\|F\|_2 \|M\|_2} \tag{6}
\]

where \(F\) and \(M\) denote the pixels of the fused image and the original MS image, respectively. The global value of SAM is computed by taking the average of all the pixels in the SAM map. It is represented in degree (°) or radian. The optimal value for global SAM is zero which means no color distortion in the fused image. Note that SAM is regarded as a spectral distortion metric.

Another objective metric that is widely used is Correlation Coefficient (CC). The range for CC is \([-1, 1]\), where 1 means the highest correlation between images. This metric is computed as follows:

\[
\text{CC} = \frac{\sum_{x=1}^{N} \sum_{y=1}^{N} (F(x, y) - \mu_F)(M(x, y) - \mu_M)}{\sqrt{\sum_{x=1}^{N} \sum_{y=1}^{N} (F(x, y) - \mu_F)^2 (M(x, y) - \mu_M)^2}} \tag{7}
\]

UIQI in \[36\] is also a widely used metric. This metric denotes a similarity index which characterizes spectral and spatial distortions, and it is computed using the following equation:

\[
\text{UIQI} = \frac{\sigma_F \sigma_M}{\sigma_F^2 + \mu_F^2 + \sigma_M^2 + \sigma_F^2 + \sigma_M^2} \tag{8}
\]

in which the term \(\sigma(.)\) represents the standard deviation and \(\mu(.)\) denotes the average. The Q4 metric is a vectorized version of the UIQI metric.

The last metric used here in the reduced-resolution mode is Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS), which is an improvement of the Mean Squared Error (MSE) by taking into consideration the scale ratio of the PAN and MS images. It reflects the global distortion in the fused image according to the following equation:

\[
\text{ERGAS} = 100 \frac{d_h}{d_l} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{\text{RMSE}(F, M)}{\mu(i)}} \tag{9}
\]
where \( d_h \) denotes the ratio between pixel sizes of PAN and MS, e.g., \( \frac{1}{4} \) for Pleiades-1A, WorldView-2, and GeoEye-1.

### 3.2 Full resolution metrics

The metric \( D_s \) is computed between the low-resolution MS image and the fused image at the PAN scale. The UIQI metric between the MS bands, e.g., \( M_1 \) and \( M_2 \), is first computed and then subtracted from the corresponding multiplication at high-resolution (fused image, i.e., \( F_1 \) and \( F_2 \)). This metric is computed using the following equation:

\[
D_s = \left( \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} \left| \text{UIQI}(M_i, M_j) - \text{UIQI}(F_1, F_2) \right| \right)^q 
\]

The exponent \( p \) is set to one by default but can be chosen to show larger differences between the two terms. Low \( D_s \) metric values indicate less spectral distortion and the ideal value is zero.

The metric \( D_s \) represents spatial distortion where is computed by the following equation:

\[
D_s = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left| \text{UIQI}(M_i, P_i) - \text{UIQI}(F_i, P) \right|} 
\]

The exponent \( q \) is set to one by default. The ideal value for \( D_s \) is zero which denotes no spatial distortion.

The results of the extensive experimentations carried out to examine the developed generative model are reported. The two commonly used protocols of reduced-resolution and high-resolution were considered. In the first protocol known as Wald’s protocol (reduced resolution protocol), the fusion process is performed at the lower resolution and the original MS image remains intact for full-reference analysis. The second protocol involves the full resolution data and in which the no-reference quality metrics are used.

The results of the developed method were compared with seven recently developed methods including Band Dependent Spatial Detail (BDSD) [9], Adaptive Intensity-Hue-Saturation (AIHS) [10], CS-based deep learning model (abbreviated as DNN) [20], Convolutional AutoEncoder-based pansharpening (abbreviated as CAE) [6], Modulation Transfer Function Generalized Laplacian Pyramids (MTF-GLP) [9], Fractional-order Differentiation in Image Fusion (FDIF) [1], our previously developed (Multi-Objective)

| Method       | t(s)    |
|--------------|---------|
| BDSD         | 0.20 (CPU) |
| AIHS         | 0.25 (CPU) |
| DNN          | 0.3 (GPU)  |
| CAE          | 0.3 (GPU)  |
| MTF-GLP      | 0.4 (CPU)  |
| FDIF         | 0.15 (CPU) |
| Multi-objective | 0.7 (GPU)  |
| Developed    | 0.5 (GPU)  |

The exponent \( p \) is set to one by default but can be chosen to show larger differences between the two terms. Low \( D_s \) metric values indicate less spectral distortion and the ideal value is zero.

The results of the developed method were compared with seven recently developed methods including Band Dependent Spatial Detail (BDSD) [9], Adaptive Intensity-Hue-Saturation (AIHS) [10], CS-based deep learning model (abbreviated as DNN) [20], Convolutional AutoEncoder-based pansharpening (abbreviated as CAE) [6], Modulation Transfer Function Generalized Laplacian Pyramids (MTF-GLP) [9], Fractional-order Differentiation in Image Fusion (FDIF) [1], our previously developed (Multi-Objective)
method [37]. All the experiments were done in both reduced and full-scale modes.

Tables 1 through 4 exhibit the results of the developed method as well as the above representative existing methods. As can be seen from these tables, the developed method provided improved metric values. In particular, the SAM and ERGAS metric values were considerably better (Tables 2 and 3). These metrics denote the spectral and overall distortions of the fused image, respectively (Tables 5 and 6). For the other metrics, the developed method also provided improved values in comparison with the other methods. Sample fused images for the three datasets examined are shown in Figs. 3 through 5 at full-resolution. The processing time for each method is listed in Table 7.

Fig. 4 True color representation of sample fusion results for WorldView-2 dataset at full-resolution: a MS, b PAN, c BDSD, d AIHS, e DNN, f CAE, g MTF-GLP, h FDIF, i Multi-Objective, j Developed

Fig. 5 True color representation of sample fusion results for GeoEye-1 dataset at full-resolution: a MS, b PAN, c BDSD, d AIHS, e DNN, f CAE, g MTF-GLP, h FDIF, i Multi-Objective, j Developed
From a visual inspection perspective, one can see that the developed method performed better in terms of both the spatial and spectral contextual information. For example, as can be seen from Fig. 3, the color information of the green area was better preserved. Moreover, as seen from Fig. 3, the spatial details of the PAN image were more effectively injected into the fused image. In Fig. 3c, e, and f, the fusion results got blurred. It can be clearly seen that Fig. 3d and g suffer from color distortion in some regions. Moreover, Fig. 3h is over-sharpened and the level of spatial details is high. The result of Fig. 3i and j is competitive; however, in most regions, the developed method performs better. To make the visual inspection easier, the harbor area in the center of Fig. 4 is magnified. It can be seen that the AIHS method generated color distortion especially in the green area. The MTF-GLP method generated a blurry outcome at the harbor edges. The green area in the FDIF image turned into dark green in comparison to the green color of the LRMS image. The BDSD, DNN, and CAE methods oversharpened the fused image with a slight color distortion. The Multi-Objective method visually produced similar outcomes but the developed method preserved the spectral information better. Another example is shown in Fig. 5. In this figure, one can see that the method AIHS suffered from the spectral distortion in some regions. The building edges when using the DNN, CAE, and FDIF methods appeared blurred. The color information in the MTF-GLP and BDSD methods were lost in some regions. The GLP-HRI method suffered from over-sharpening. The colors associated with the developed generative model appeared better preserved across different datasets in comparison to our previous (Multi-Objective) method.

### 3.3 Ablation study

In this study, the effect of adding new terms to the conventional loss function was examined. This study was conducted for all the three datasets. To show the effectiveness of the developed method, the error maps of the fused products were computed with respect to the reference image and specific locations in each image were cropped and selected. To make the error maps more visible, the images were adjusted in the range of [0, 255] via linear transformation. The results corresponding to the Pleiades-1A, WorldView-2, and GeoEye-1 datasets are shown in part I, II, and III of Fig. 6, respectively. As can be seen from part I of Fig. 6 (Pleiades-1A dataset results), the developed method exhibited less spectral and spatial distortions. The same outcome was seen for the WorldView-2 and GeoEye-1 datasets.

### 4 Conclusion

In this paper, a new generative model method for unsupervised learning process of multispectral image fusion has been developed. The developed method addresses the ill-posed pansharpening problem in a more comprehensive manner. The model consists of two separate discriminators for learning the spectral and spatial information. The former uses the MS data at the original scale as the input of the spectral discriminator. The latter uses the PAN image as the input. A comprehensive comparison was conducted with seven recent pansharpening methods and the results obtained show fused images obtained by the developed method generated less distortion compared to these methods.
References

1. Azarang, A., Ghassemian, H.: Application of fractional-order differentiation in multispectral image fusion. Remote Sens. Lett. 10(1), 91–100 (2018)

2. Azarang, A., and Ghassemian, H.: An adaptive multispectral image fusion using particle swarm optimization. In: 2017 Iranian Conference on Electrical Engineering (ICEE) (pp. 1708–1712). IEEE (2017)

3. Azarang, A., Kehtarnavaz, N.: Multispectral image fusion based on map estimation with improved detail. Remote Sens. Lett. 11(8), 797–806 (2020)

4. Aiazzi, B., Baronti, S., Selva, M.: Improving component substitution pansharpening through multivariate regression of MS S+ S Pan data. IEEE Trans. Geosci. Remote Sens. 45(10), 3230–3239 (2007)

5. Zhu, X.X., Bamler, R.: A sparse image fusion algorithm with application to pan-sharpening. IEEE Trans. Geosci. Remote Sens. 51(5), 2827–2836 (2012)

6. Azarang, A., Manoochehri, H.E., Kehtarnavaz, N.: Convolutional autoencoder-based multispectral image fusion. IEEE access 7, 35673–35683 (2019)

7. Kang, X., Li, S., Benediktsson, J.A.: Pansharpening with mabling model. IEEE Trans. Geosci. Remote Sens. 52(8), 5088–5099 (2013)

8. Shah, V.P., Younan, N.H., King, R.L.: An efficient pansharpening method via a combined adaptive PCA approach and contourlets. IEEE Trans. Geosci. Remote Sens. 46(5), 1323–1335 (2008)

9. Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G., Restaino, R. and Wald, L.: A critical comparison of pansharpening algorithms. In: 2014 IEEE Geoscience and Remote Sensing Symposium (pp. 191–194). IEEE (2014)

10. Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A.: Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J. Photogramm. Remote. Sens. 152, 166–177 (2019)

11. Zhong, S., Zhang, Y., Chen, Y., Wu, D.: Combining component substitution and multiresolution analysis: a novel generalized BDS pan-sharpening algorithm. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10(6), 2867–2875 (2017)

12. Wang, W., Jiao, L., Yang, S.: Novel adaptive component-substitution-based pan-sharpening using particle swarm optimization. IEEE Geosci. Remote Sens. Lett. 12(4), 781–785 (2014)

13. Choi, J., Yu, K., Kim, Y.: A new adaptive component-substitution-based satellite image fusion by using partial replacement. IEEE Trans. Geosci. Remote Sens. 49(1), 295–309 (2010)

14. Licciardi, G., Vivone, G., Dalla Mura, M., Restaino, R., Chanussot, J.: Multi-resolution analysis techniques and nonlinear PCA for hybrid pansharpening applications. Multidimension. Syst. Signal Process. 27(4), 807–830 (2016)

15. Palsson, F., Sveinsson, J.R. and Ulfarsson, M.O.: Optimal component substitution and multi-resolution analysis pansharpening methods using a convolutional neural network. In: IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 3177–3180). IEEE (2019)

16. Mitianoudis, N., Tzimiropoulos, G. and Stathaki, T.: Fast wavelet-based pansharpening of multi-spectral images. In: 2010 IEEE International Conference on Imaging Systems and Techniques (pp. 11–16). IEEE (2010)

17. Ait, H.A., Sharma, G.: A regularized model-based optimization framework for pan-sharpening. IEEE Trans. Image Process. 23(6), 2596–2608 (2014)

18. Wang, W., Liu, H., Liang, L., Liu, Q., Xie, G.: A regularised model-based pansharpening method for remote sensing images with local dissimilarities. Int. J. Remote Sens. 40(8), 3029–3054 (2019)

19. Dong, W., Liang, J., Xiao, S.: Saliency analysis and Gaussian mixture model-based detail extraction algorithm for hyperspectral pansharpening. IEEE Trans. Geosci. Remote Sens. 58(8), 5462–5476 (2020)

20. Huang, W., Xiao, L., Wei, Z., Liu, H., Tang, S.: A new pan-sharpening method with deep neural networks. IEEE Geosci. Remote Sens. Lett. 12(5), 1037–1041 (2015)

21. Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G.: Pansharpening by convolutional neural networks. Remote Sens. 8(7), 594 (2016)

22. Wei, Y., Yuan, Q., Shen, H., Zhang, L.: Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. IEEE Geosci. Remote Sens. Lett. 14(10), 1795–1799 (2017)

23. Azarang, A., and Ghassemian, H.: A new pansharpening method using multi resolution analysis framework and deep neural networks. In: 2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA) (pp. 1–6). IEEE (2017)

24. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 38, 295–307 (2015)

25. Yuan, Q., Wei, Y., Meng, X., Shen, H., Zhang, L.: A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11(3), 978–989 (2018)

26. Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X. and Paisley, J.: PanNet: A deep network architecture for pan-sharpening. In Proceedings of the IEEE International Conference on Computer Vision (pp. 5449–5457) (2017)

27. Liu, X., Wang, Y. and Liu, Q.: PSGAN: A generative adversarial network for remote sensing image pan-sharpening. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 873–877). IEEE (2018)

28. Vitale, S.: A cnn-based pansharpening method with perceptual loss. In: IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 3105–3108). IEEE (2019)

29. Scarpa, G., Vitale, S., Cozzolino, D.: Target-adaptive CNN-based pansharpening. IEEE Trans. Geosci. Remote Sens. 56(9), 5443–5457 (2018)

30. Vitale, S., Scarpia, G.: A detail-preserving cross-scale learning strategy for CNN-based pansharpening. Remote Sens. 12(3), 348 (2020)

31. Wang, D., Li, Y., Ma, L., Bai, Z., Chan, J.C.W.: Going deeper with densely connected convolutional neural networks for multispectral pansharpening. Remote Sens. 11(22), 2608 (2019)

32. Choi, J.S., Kim, Y., Kim, M.: S3: A spectral-spatial structure loss for pan-sharpening networks. IEEE Geosci. Remote Sens. Lett. 17(5), 829–833 (2019)

33. Bello, J.L.G., Seo, S., and Kim, M.: Pan-sharpening with color-aware perceptual loss and guided re-colorization. In: 2020 IEEE International Conference on Image Processing (ICIP) (pp. 908–912). IEEE (2020)

34. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (pp. 2672–2680) (2014)

35. Mirza, M. and Osindero, S.: Conditional generative adversarial nets. arXiv preprint (2014)

36. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Process. Lett. 9(3), 81–84 (2002)

37. Azarang, A., Kehtarnavaz, N.: Image fusion in remote sensing by multi-objective deep learning. Int. J. Remote Sens. 41(10), 9507–9524 (2020)

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