IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 2, 2019, accepted March 11, 2019, date of publication March 18, 2019, date of current version April 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2905511

Convolutional Autoencoder-Based Multispectral Image Fusion

ARIAN AZARANG, (Student Member, IEEE), HAFEZ E. MANOOCHEHRI, (Student Member, IEEE), AND NASSER KEHTARNAVAZ, (Fellow, IEEE)
Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA
Corresponding author: Arian Azarang (azarang@utdallas.edu)

ABSTRACT This paper presents a deep learning-based pansharpening method for fusion of panchromatic and multispectral images in remote sensing applications. This method can be categorized as a component substitution method in which a convolutional autoencoder network is trained to generate original panchromatic images from their spatially degraded versions. Low resolution multispectral images are then fed into the trained convolutional autoencoder network to generate estimated high resolution multispectral images. The fusion is achieved by injecting the detail map of each spectral band into the corresponding estimated high resolution multispectral bands. Full reference and no-reference metrics are computed for the images of three satellite datasets. These measures are compared with the existing fusion methods whose codes are publicly available. The results obtained indicate the effectiveness of the developed deep learning-based method for multispectral image fusion.

INDEX TERMS Multispectral image fusion by convolutional autoencoder, fusion of panchromatic and multispectral images in remote sensing, convolutional autoencoder-based pansharpening.

I. INTRODUCTION

Pansharpening [1]–[9] or fusion of spectral and spatial image attributes is commonly used in remote sensing. The fusion process is done in order to address the limitation of bandwidth in capturing image data by satellites at a certain Signal-to-Noise Ratio (SNR). Two types of data are captured by satellites: PANchromatic (PAN) image data providing spatial information and MultiSpectal (MS) image data providing spectral information. After the preprocessing steps of registration and noise reduction, the fusion process is applied to merge the spectral content of MS image data with the spatial content of PAN image data. The effectiveness of fusion is often assessed by the level of artifacts and inconsistencies in the fused outcome [10].

A number of pansharpening methods have been introduced in the literature that can be grouped into three categories: (1) Component Substitution (CS), e.g. [11]–[15], (2) Multi-Resolution Analysis (MRA), e.g. [16]–[20], and (3) Model-Based (MB), e.g. [21]–[23]. In the first two categories, the Low Resolution MS (LRMS) image data are used while different procedures are used to extract the detail map. Most CS-based methods are based on the difference between PAN image data and a linear combination of LRMS bands. Depending on the weights in the linear combination, different methods are resulted such as Adaptive Intensity Hue Saturation (AIHS) and its improved version (IAIHS) [24], [25], Weighted Least Square (WLS) [26] and Nonlinear Intensity Hue Saturation (NIHS) [27]. Due to the fact that different spectral bands do not share the same detail map, adaptive injection gains were developed in [28]. In IAIHS, an edge detector was derived from both PAN and MS image data which was then applied to the detail map extracted by AIHS. In [29], a fractional-order edge detector on the detail map was used to obtain a refined detail map. In MRA-based methods, the detail map is derived from PAN image data and by applying high frequency filters. In MB methods, pansharpening is formulated as an optimization problem based on a prior model which is normally considered to be Bayesian [30].

The past few years have witnessed an increasing use of deep neural networks in remote sensing. As far as pansharpening is concerned, the convolutional neural network (CNN) architecture was utilized in [31]. In [32], the Denoising AutoEncoder (DAE) architecture was used as part of a

The associate editor coordinating the review of this manuscript and approving it for publication was Choon Ki Ahn.

MRA-based method leading to enhanced fusion outcomes. In [33], the Deep Residual Network (DRN) architecture was used to achieve multispectral image fusion. In [34], the DAE architecture was considered as part of a CS-based method where LRMS images were computed via a Gaussian filter. More recently, in [35], the Generative Adversarial Network (GAN) architecture was used for pansharpening. A comprehensive study of deep learning-based pansharpening methods is provided in [36].

In this paper, a new deep learning-based method for multispectral image fusion is introduced based on the convolutional autoencoder architecture. The fusion outcome of the developed method is then compared with a representative set of existing methods using both full reference and no-reference objective measures.

The rest of the paper is organized as follows: section II describes the proposed fusion method in detail. In section III, the datasets and the performance measures used are mentioned. The experimental results and their discussion are also covered in section III. Finally, the conclusion is stated in section IV.

II. DEVELOPED CONVOLUTIONAL AUTOENCODER FUSION METHOD

A. FUSION FRAMEWORK

The general framework of CS-based methods can be stated as the following equation [6]

$$\mathbf{M}_i^{\mathrm{H}} = \mathbf{M}_i^{\mathrm{L}} + \mathrm{g_i}\,(\mathbf{P} - \mathbf{I}) \tag{1}$$

where $\mathbf{M}_i^{\mathrm{H}}$ and $\mathbf{M}_i^{\mathrm{L}}$ denote the high and low resolution MS image data in the i^{th} band, respectively, $\mathrm{g_i}$ the i^{th} injection gain of the detail map which is represented by the difference between the PAN image $\mathbf{P}$ and a linear combination of LRMS bands $\mathbf{I}$ expressed as:

$$\mathbf{I} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{w_i}\mathbf{M}_i^{\mathrm{L}} \tag{2}$$

with $\mathrm{w_i}$s as weights and N as the number of spectral bands covering the spectral range of the PAN image. Optimal weights are found via solving the following optimization problem

$$w_i^* = \arg\min_{\mathrm{w_i}} \left\| \mathbf{P} - \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{w_i}\mathbf{M}_i^{\mathrm{L}} \right\|^2 \tag{3}$$

where $\mathrm{w_i^*}$s denote the optimal weights. The detail map $\mathbf{D}$ is then obtained as follows:

$$\mathbf{D} = \mathbf{P} - \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{w_i^*}\mathbf{M}_i^{\mathrm{L}} \tag{4}$$

Considering that different MS bands do not share the same detail map, the injection gains are adjusted according to this equation [26]

$$\mathrm{g_i} = \frac{cov\left(\mathbf{M}_i^{\mathrm{L}}, \mathbf{I}\right)}{var(\mathbf{I})} \tag{5}$$

B. CONVOLUTIONAL AUTOENCODER

Convolutional AutoEncoder (CAE) is a type of CNN that reconstructs the input at the output and can be used as a feature extractor with 2D images as its input. The architecture of a CAE consists of two main parts: encoding part for feature representation or providing a compressed version of the input and a decoding part for reconstructing the input from the compressed version. The encoding part includes convolution and max pooling layers, while the decoding part consists of deconvolution and upscaling layers. Fig. 1 exhibits an illustration of the CAE architecture.

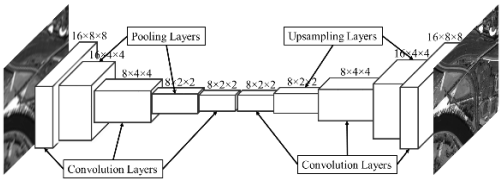

FIGURE 1. Architecture of the convolutional autoencoder network used.

1) ENCODER

To capture structural attributes of the input image data across Q dimensions $K = \{K_1, K_2, \ldots, K_{\mathrm{Q}}\}$, n convolution filters $F^{(1)} = \{F_1^{(1)}, F_2^{(1)}, \ldots, F_{\mathrm{n}}^{(1)}\}$ are considered to generate n intermediate features expressed by the following equation

$$T_{\mathrm{m}} = s\left(K * F_{\mathrm{m}}^{(1)} + b_{\mathrm{m}}^{(1)}\right) \quad \mathrm{m} = 1, 2, \ldots, \mathrm{n} \tag{6}$$

where s denotes the sigmoid activation function and $b_{\mathrm{m}}^{(1)}$ is the bias for the m^{th} feature map. The intermediate feature maps correspond to the compressed version of the input data. Several convolution layers are normally concatenated.

2) DECODER

Decoding involves reconstructing the input image K from its n intermediate feature maps. The reconstructed input image $\tilde{K}$ is obtained via the convolution of the intermediate feature maps $T = \{T_{\mathrm{m}}\}_{\mathrm{m}=1}^{\mathrm{n}}$ with the convolutional filters $F^{(2)} = \{F_1^{(2)}, F_2^{(2)}, \ldots, F_{\mathrm{n}}^{(2)}\}$ expressed by the following equation

$$\tilde{K} = s(T * F_{\mathrm{m}}^{(2)} + b_{\mathrm{m}}^{(2)}) \tag{7}$$

With the output having the same dimension as the input, any loss function $\mathcal{L}(.)$ such as Mean Square Error (MSE)

$$\mathcal{L}\left(K, \tilde{K}\right) = \frac{1}{2}\left\|K - \tilde{K}\right\|_2^2 \tag{8}$$

can be used to update the network weights during the training phase.

C. DEVELOPED CAE-BASED METHOD

A CAE network is used in this paper to improve the spatial information of the LRMS bands by learning the nonlinear relationship between a PAN image and its spatially degraded version. The steps involved in this process are described next.

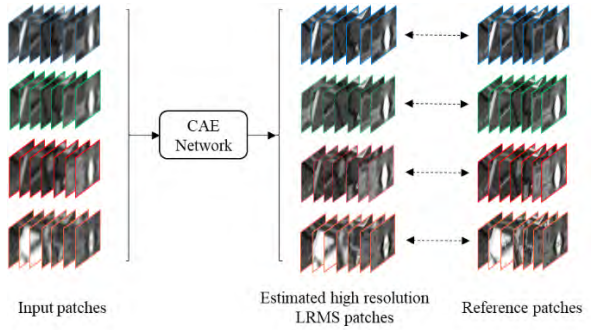

FIGURE 2. Input and output of the convolutional autoencoder network.

TABLE 1. Geographical areas of the datasets examined.

Sensor	Captured Area
QuickBird	Sundarbans, Bangladesh
Pleiades-1A	Melbourne, Australia
GeoEye-1	Washington, USA

A PAN image is considered to be the reference spatial image. A spatially degraded version of the PAN image is first generated using an interpolation filter. The original and degraded PAN images are partitioned into $p \times p$ patches with r overlapping pixels. A CAE network is used to learn the nonlinear relationship between the original PAN patches and corresponding degraded PAN patches as its output and input, respectively. After training, the CAE network generates estimated high resolution LRMS patches as its output in response to LRMS patches as its input.

Mathematically, let the original PAN patches, denoted by $\left\{X_{\mathrm{i}}^{\mathrm{H}}\right\}_{\mathrm{i}=1}^{\mathrm{C}}$, and spatially degraded PAN patches, denoted by $\left\{X_{\mathrm{i}}^{\mathrm{L}}\right\}_{\mathrm{i}=1}^{\mathrm{C}}$, form the target and the input of the CAE network, respectively, where C indicates the number of patches. The network is effectively designed to learn how to inject the spatial information into the degraded image. At each iteration, the output patches of the CAE network are computed as follows:

$$\left\{Y_{\mathrm{i}}^{\mathrm{L}}\right\}_{\mathrm{i}=1}^{\mathrm{C}} = F_{Decode}(F_{Encode}\left(\left\{X_{\mathrm{i}}^{\mathrm{L}}\right\}_{\mathrm{i}=1}^{\mathrm{C}}\right)) \tag{9}$$

where $\left\{Y_{\mathrm{i}}^{\mathrm{L}}\right\}_{\mathrm{i}=1}^{\mathrm{C}}$ denotes the output and F_{Encode} and F_{Decode} correspond to the encoding and decoding operations. The weights at each iteration are updated based on the MSE between the original PAN patches and their reconstructed versions

$$\mathcal{L}\left(\left\{Y_{\mathrm{i}}^{\mathrm{L}}\right\}_{\mathrm{i}=1}^{\mathrm{C}}, \left\{X_{\mathrm{i}}^{\mathrm{H}}\right\}_{\mathrm{i}=1}^{\mathrm{C}}\right) = \frac{1}{2}\sum_{i=1}^{\mathrm{C}}\left\|Y_{\mathrm{i}}^{\mathrm{L}} - X_{\mathrm{i}}^{\mathrm{H}}\right\|_2^2 \tag{10}$$

The backpropagation algorithm is then employed for training the network. To test the effectiveness of the trained

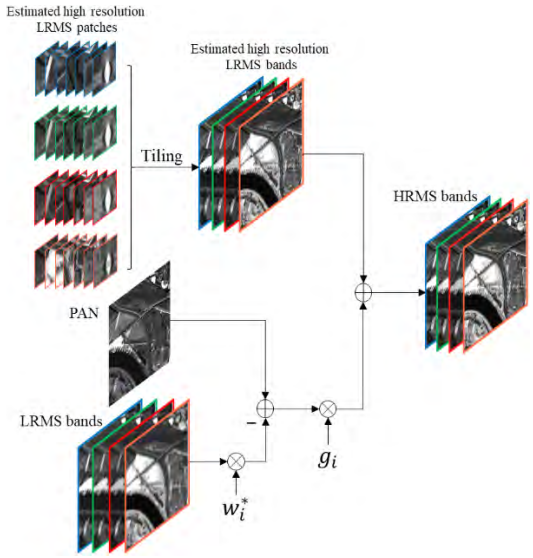

FIGURE 3. Steps involved in the developed fusion method.

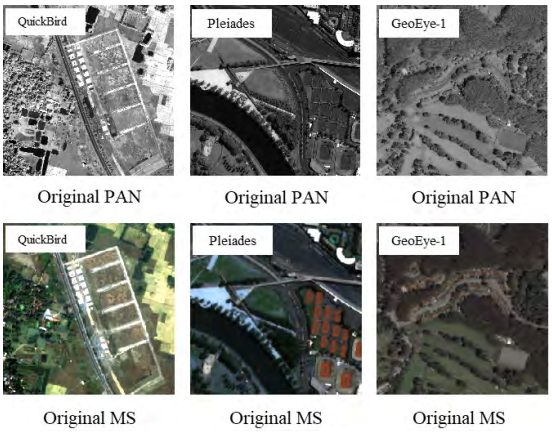

FIGURE 4. Sample images of the datasets examined.

network, the LRMS image is partitioned into patches across N bands, that is $\left\{Z_{\mathrm{i,j}}^{\mathrm{L}}\right\}_{\mathrm{i}=1}^{\mathrm{C}}$ for $\mathrm{j} = 1 : \mathrm{N}$.

Next, the patch-wised LRMS bands are fed into the CAE trained network. Due to the similarity in the spectral characteristics of PAN and MS images, the trained network is expected to improve the spatial information of the LRMS bands. The input and output of the CAE network are illustrated in Fig. 2. From this figure, it can be seen that an estimated high resolution LRMS is obtained at the output of the network for each band. In fact, not only the reconstructed version preserves the spectral information of LRMS bands but also it carries more spatial information in comparison with the input patches. What makes the proposed approach different than the previous approaches is that the estimated high resolution LRMS is used instead of LRMS in Eq. (1). After tiling the estimated high resolution LRMS bands, the fusion process is carried out via the following equation:

$$\mathbf{M}_{\mathrm{i}}^{\mathrm{H}} = \hat{\mathbf{M}}_{\mathrm{i}}^{\mathrm{H}} + \mathrm{g_i}(\mathbf{P} - \mathbf{I}) \tag{11}$$

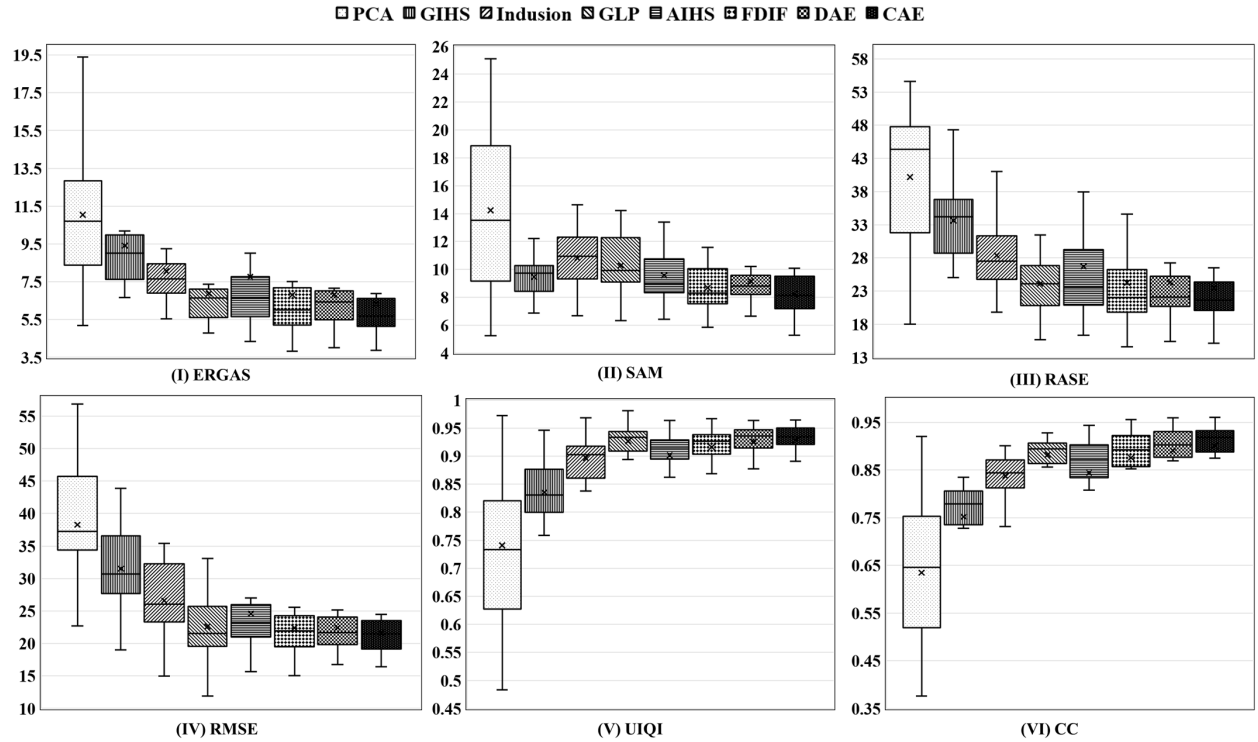

FIGURE 5. Average and percentile performance measures of PAN-MS pairs for QuickBird dataset images correspond to the full reference protocol.

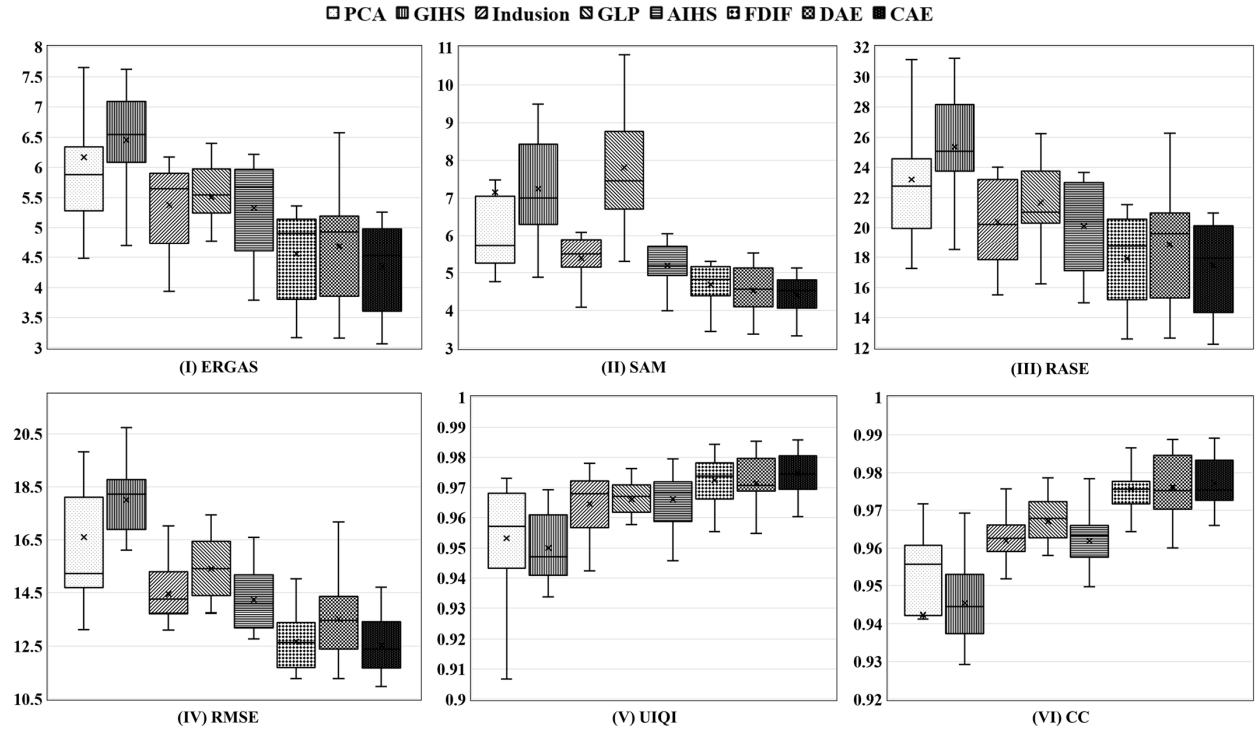

FIGURE 6. Average and percentile performance measures of PAN-MS pairs for Pleiades-1A dataset images corresponding to the full reference protocol.

where $\hat{\mathbf{M}}_i^H$ denotes the i^{th} estimated high resolution LRMS band obtained from the trained CAE network. Fig. 3 provides an illustration of the steps involved in the developed method. As illustrated in Fig. 3, the optimal weights derived from Eq. (3) are applied to the LRMS bands in order to obtain an estimation of the low resolution PAN image. Then, the

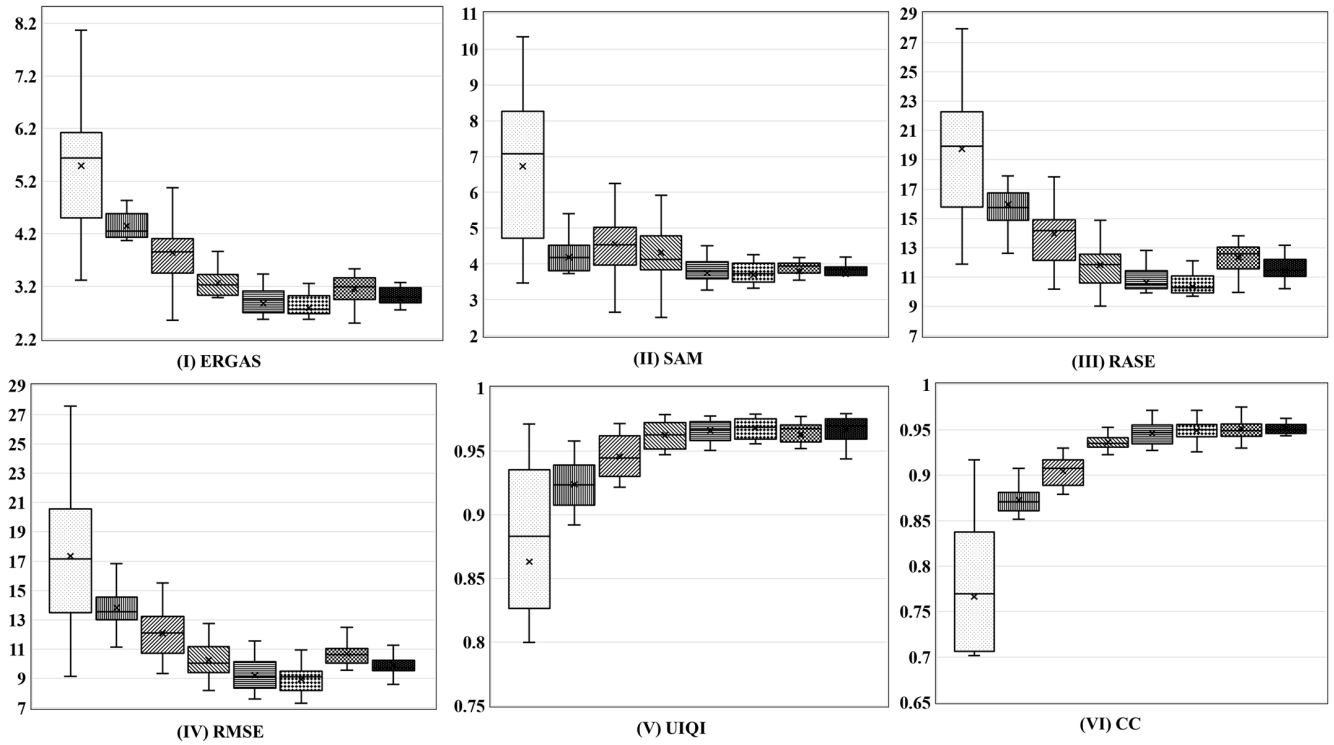

FIGURE 7. Average and percentile performance measures of PAN-MS pairs for GeoEye-1 dataset images corresponding to the full reference protocol.

TABLE 2. Average performance measures for QuickBird dataset images; the last row shows the ideal value for each measure and the best values are bolded with the second best values underlined.

Method	ERGAS	SAM	RASE	RMSE	UIQI	CC
PCA	11.03	14.23	40.16	38.26	0.74	0.63
GIHS	9.41	9.45	33.61	31.49	0.84	0.75
Indusion	8.06	10.80	28.37	26.56	0.90	0.84
GLP	6.87	10.26	<u>24.09</u>	22.55	0.93	0.88
AIHS	7.75	9.56	26.69	24.54	0.90	0.84
FDIF	6.80	<u>8.68</u>	24.26	<u>22.35</u>	0.92	0.88
DAE	<u>6.80</u>	9.14	24.27	22.39	<u>0.93</u>	<u>0.89</u>
CAE	**6.36**	**8.21**	**23.47**	**21.66**	**0.93**	**0.90**
IDEAL	0	0	0	0	1	1

primitive detail map is obtained via Eq. (4). Next, the injection gains noted in Eq. (5) are used to acquire the refined detail map. After obtaining and tiling the estimated high resolution LRMS patches, the first term of Eq. (11) is computed. The fusion is achieved by combining its two components, see Fig. 3.

III. EXPERIMENTAL RESULTS

In this section, the datasets used are first described followed by the performance measures examined. A quantitative as well as a visual comparison with a representative set of existing methods are then reported for the conventional full reference (reduce-resolution) and no-reference (full resolution) protocols.

A. DATASETS

Three datasets are examined in this paper: (a) The QuickBird dataset which includes LRMS images at 2.44-2.88m resolution as well as PAN images at 0.61-0.72m resolution taken from a rural area. (b) The Pleiades-1A dataset which includes 0.5m resolution images that are taken from an urban area. (c) The GeoEye-1 dataset which includes 0.46m resolution images taken from a forest area. In all of these datasets, 4 bands (blue, green, red, and near infrared) are available,

TABLE 3. Average performance measures for Pleiades-1A dataset images; the last row shows the ideal value for each measure and the best values are bolded with the second best values underlined.

Method	ERGAS	SAM	RASE	RMSE	UIQI	CC
PCA	6.17	7.15	23.19	16.59	0.95	0.94
GIHS	6.45	7.24	25.35	17.99	0.95	0.95
Indusion	5.38	5.38	20.37	14.44	0.96	0.96
GLP	5.51	7.81	21.65	15.40	0.97	0.97
AIHS	5.32	5.18	20.08	14.23	0.97	0.96
FDIF	4.55	4.68	17.91	12.66	0.97	0.98
DAE	4.68	4.52	18.85	13.52	0.97	0.98
CAE	**4.34**	**4.40**	**17.46**	**12.53**	**0.98**	**0.98**
IDEAL	0	0	0	0	1	1

TABLE 4. Average performance measures for GeoEye-1 dataset images; the last row shows the ideal value for each measure and the best values are bolded with the second best values underlined.

Method	ERGAS	SAM	RASE	RMSE	UIQI	CC
PCA	5.49	6.72	19.74	17.34	0.86	0.77
GIHS	4.35	4.18	15.93	13.81	0.92	0.87
Indusion	3.84	4.56	13.94	12.09	0.95	0.90
GLP	3.26	4.32	11.823	10.25	0.96	0.94
AIHS	2.87	3.74	10.63	9.20	0.97	0.94
FDIF	**2.78**	3.70	**10.34**	**8.94**	0.97	0.94
DAE	3.15	3.79	12.33	10.68	0.96	0.95
CAE	2.97	**3.70**	11.45	9.91	**0.97**	**0.96**
IDEAL	0	0	0	0	1	1

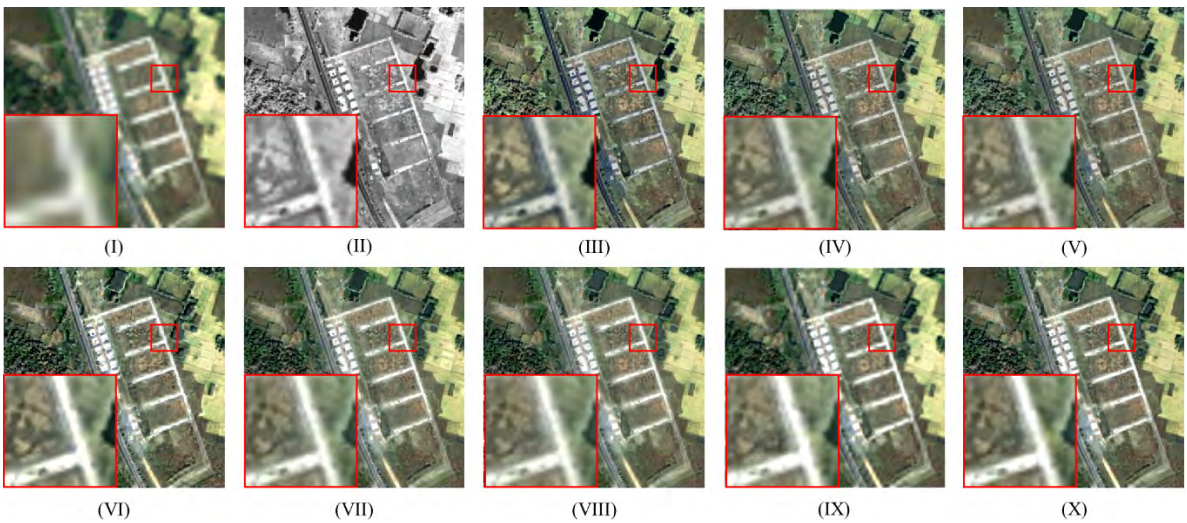

FIGURE 8. Fusion outcomes for a sample QuickBird dataset image: (I) LRMS, (II) PAN, (III) PCA, (IV) GIHS, (V) Indusion, (VI) GLP, (VII) AIHS, (VIII) FDIF, (IX) DAE, (X) CAE.

covering the visible and near-infrared regions of the spectrum. The geographical areas associated with the datasets are listed in Table 1. For the experimentations, 8 × 8 patches of the spatially degraded version of PAN and the original PAN images consisting of 500,000 patch pairs were used to serve as the input and target of the CAE network, respectively.

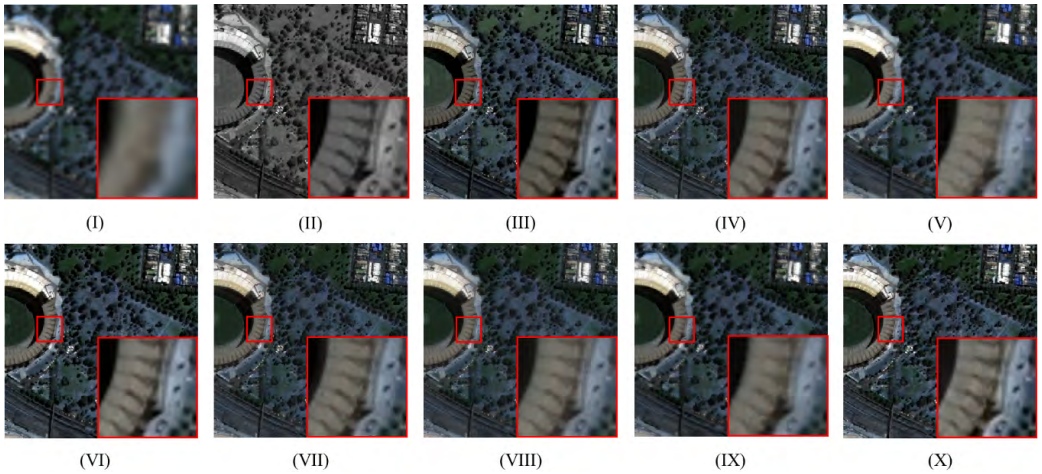

FIGURE 9. Fusion outcomes for a sample Pleiades-1A dataset image: (I) LRMS, (II) PAN, (III) PCA, (IV) GIHS, (V) Indusion, (VI) GLP, (VII) AIHS, (VIII) FDIF, (IX) DAE, (X) CAE.

TABLE 5. Average no-reference performance measures for QuickBird dataset images; the last row shows the ideal value for each measure and the best values are bolded with the second best values underlined.

Method	D_s	D_λ	QNR
PCA	0.33	0.25	0.51
GIHS	0.29	0.18	0.58
Indusion	0.14	0.19	0.70
GLP	0.14	0.16	0.73
AIHS	0.12	0.15	0.75
FDIF	0.09	0.09	0.83
DAE	<u>0.06</u>	<u>0.06</u>	<u>0.88</u>
CAE	**0.05**	**0.04**	**0.91**
IDEAL	0	0	1

TABLE 6. Average no-reference performance measures for Pleiades-1A dataset images; the last row shows the ideal value for each measure and the best values are bolded with the second best values underlined.

Method	D_s	D_λ	QNR
PCA	0.14	0.10	0.77
GIHS	0.15	0.10	0.77
Indusion	0.11	0.09	0.81
GLP	0.09	0.09	0.83
AIHS	0.09	0.09	0.83
FDIF	0.09	0.06	0.86
DAE	<u>0.06</u>	<u>0.05</u>	<u>0.89</u>
CAE	**0.05**	**0.04**	**0.91**
IDEAL	0	0	1

To get the estimated high resolution LRMS, 8×8 patches were used separately on each band on the trained CAE network.

B. PERFORMANCE MEASURES

An issue of concern in pansharpening involves the unavailability of High Resolution MS (HRMS) images. That is why two protocols are often considered in remote sensing. In the first protocol, which is called full reference or reduced-resolution [37], the intact MS image is considered to be the reference. In the second protocol, which is called no-reference or full resolution [38], the reference MS image is upsampled to the size of the PAN image.

Here, the outcome of our method is compared to a representative set of pansharpening methods including PCA [12], GIHS [39], Indusion [16], GLP [17], AIHS [24], FDIF [29] and DAE [32]. Both the full reference and no-reference protocols are considered. The results reported next correspond to the averages and percentiles of the entire QuickBird and Pleiades-1A dataset images and 50% of the GeoEye-1 dataset images (randomly selected).

1) FULL REFERENCE

Several measures have been utilized in the literature for evaluating the spatial and spectral distortions of fused images with respect to an available reference image. According to the first or Wald's protocol, both scalar (i.e., measurements on a single spectral band) and vector (i.e., jointly considering all the spectral bands) (dis)similarity indices are computed. The ones that are widely used are Erreur Relative Globale Adimensionnelle de Synthese (ERGAS) [39], Spectral Angle Mapper (SAM) [40], Relative Average Spectral Error (RASE) [41], Root Mean Square Error (RMSE) [42], Universal Image Quality Index (UIQI) [43] and Correlation Coefficient (CC) [44]. RMSE denotes the square mean root error between the fused and original MS images. ERGAS is a normalized version of RMSE indicating a measure of distortion. SAM provides the angle between the corresponding pixels in

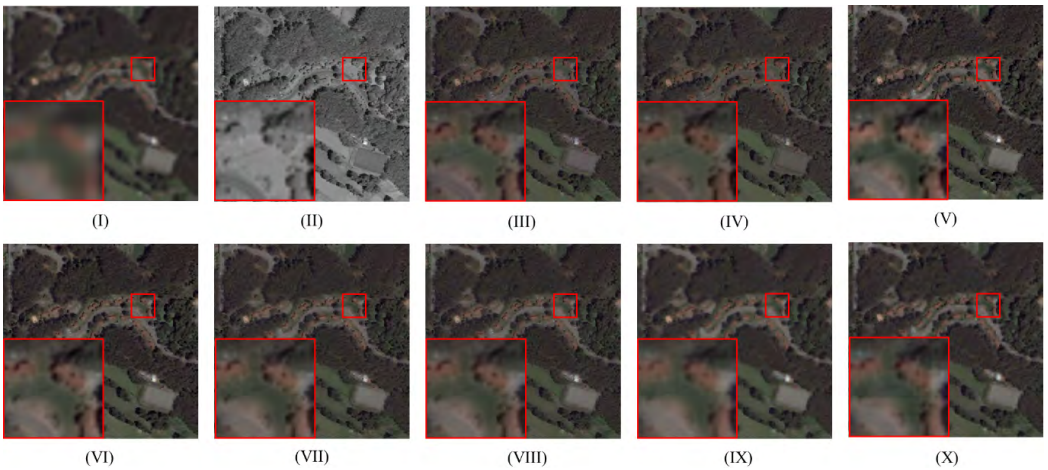

FIGURE 10. Fusion outcomes for a sample GeoEye-1 dataset image: (I) LRMS, (II) PAN, (III) PCA, (IV) GIHS, (V) Indusion, (VI) GLP, (VII) AIHS, (VIII) FDIF, (IX) DAE, (X) CAE.

the original MS and the pansharpened images. RASE captures the average performance of different methods in terms of spectral bands. CC reflects spectral distortion based on the correlation coefficient between the original MS bands and the fused image bands. UIQI is another measure of spectral distortion obtained by the product of correlation, luminance distortion, and contrast distortion.

2) NO-REFERENCE

In order to perform the quality evaluation at the original image or data resolution, the Quality of No Reference (QNR) index was proposed in [38]. The QNR index is defined as

$$\mathrm{QNR} = (1 - \mathrm{D}_\lambda)^\alpha \, (1 - \mathrm{D}_\mathrm{S})^\beta \tag{12}$$

where D_λ and D_s denote quantifications of the spectral and spatial distortions, respectively, with α and β as weight factors.

C. COMPARISON RESULTS

1) FULL REFERENCE

a: QUANTITATIVE COMPARISON

The performance measures of our method and the existing methods for the three datasets of QuickBird, Pleiades-1A, and GeoEye-1 are provided in Figs. 5 through 7, respectively. To make the quantitative comparison easy to see, the average values of the measures are also reported in Tables 2 through 4. The ideal values of each measure are shown in the last row of these tables with the best values bolded. From the above figures and tables, it can be observed that when the performance measures are considered collectively, on average, our method generated the best performance measures with the least amounts of variance.

b: VISUAL COMPARISON

Figs. 8 through 10 illustrate sample LRMS and PAN images as well as corresponding pansharpened results. From these figures, it can be seen that all the pansharpened images did better than LRMS in terms of the spatial content. The PCA, GIHS and Indusion methods suffered from spectral intensity distortions and colors also got slightly distorted in LRMS. On the other hand, the AIHS and FDIF methods did not capture the spatial details more than the other methods, in particular near the edges. It can be observed that the strength of our method lies in the preservation of colors.

For better visualization, a small area of each dataset is zoomed and shown on the bottom side of the images. As seen in the figures, the PCA, GIHS and Indusion methods suffered from spectral distortions. The GLP and AIHS methods led to some loss of the spectral information. Although the FDIF and DAE methods preserved the spectral and spatial information in the fused outcome, some areas, in particular near the edges, got blurred. Our method did not generate local artifacts and provided spectral and spatial consistency.

The measures reported reflect the average distortion for an entire fused image and not local area distortions. The error

TABLE 7. Average no-reference performance measures for GeoEye-1 dataset images; the last row shows the ideal value for each measure and the best values are bolded with the second best values underlined.

Method	D_s	D_λ	QNR
PCA	0.13	0.09	0.79
GIHS	0.18	0.13	0.71
Indusion	0.13	0.19	0.71
GLP	0.09	0.12	0.80
AIHS	0.12	0.14	0.75
FDIF	0.08	0.10	0.83
DAE	<u>0.07</u>	<u>0.09</u>	<u>0.85</u>
CAE	**0.05**	**0.07**	**0.88**
IDEAL	0	0	1

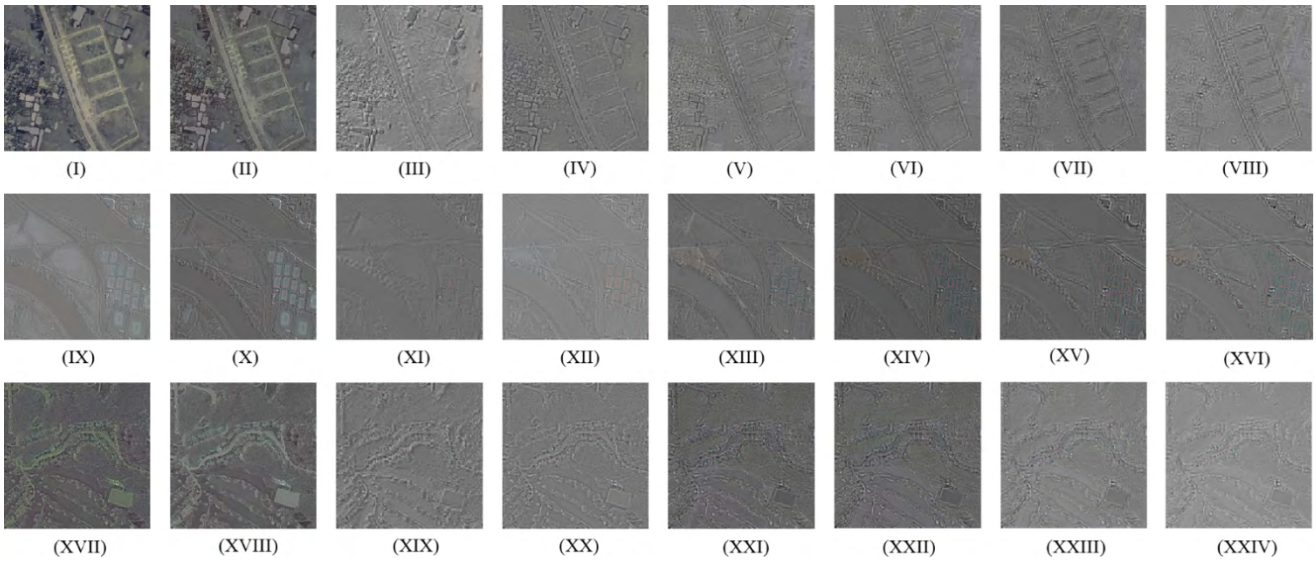

FIGURE 11. Error maps - for visualization purposes, the original error maps are linearly mapped to [0,255]: (I)-(VIII) QuickBird, (IX)-(XVI) Pleiades-1A, (XVII)-(XXIV) GeoEye-1.

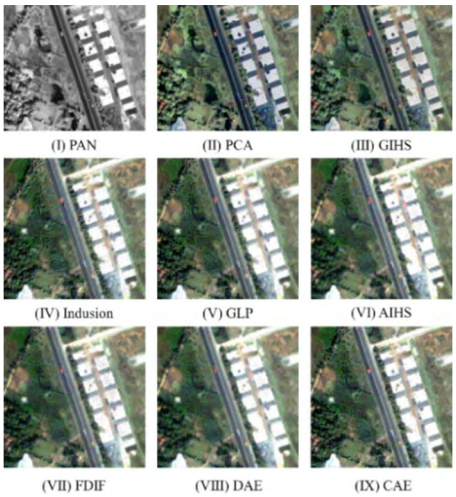

FIGURE 12. Full resolution fusion outcomes for a sample QuickBird image (for better visual comparison, a cropped part of the entire image is shown).

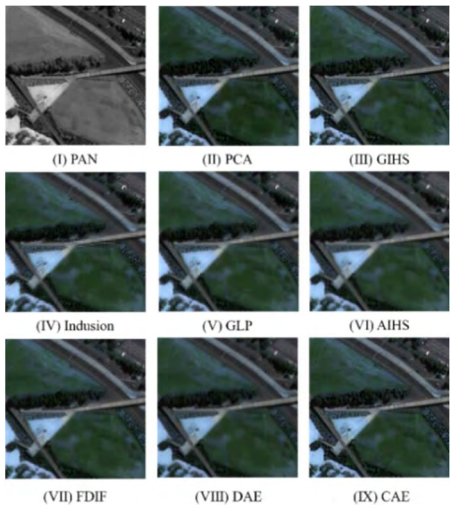

FIGURE 13. Full resolution fusion outcomes for a sample Pleiades-1A image (for better visual comparison, a cropped part of the entire image is shown).

maps, i.e. the difference between the fused image and the reference image (HRMS), can be examined in order to see the structural uniformity and edges. As shown in the error maps in Fig. 11, our method achieved better performance in most areas of the image compared to the other methods.

2) NO-REFERENCE

a: QUANTITATIVE COMPARISON

Table 5 compares the result of our method (CAE) with the other methods for the QuickBird dataset in terms of D_S, D_λ and QNR which represent the spatial, spectral and overall distortions, respectively. As seen in this table, our CAE method achieved better performance in all the cases. Similarly, Tables 6 and 7 show our CAE method outperformed the other methods for the Pleiades-1A and GeoEye-1 datasets, respectively.

b: VISUAL COMPARISON

The no-reference fused sample images are shown in Figs. 12 through 14 for the three datasets QuickBird, Pleiades-1A and GeoEye-1, respectively. The PCA and GIHS methods performed poorly, mostly because of the spectral distortion, while the Indusion and AIHS methods generated blurriness. The FDIF and DAE methods produced good spectral information but suffered from a lack of spatial information in some areas. Our CAE method performed similar to the GLP method, but it produced more spatial information while preserving the spectral information. As an example, in Fig. 12,

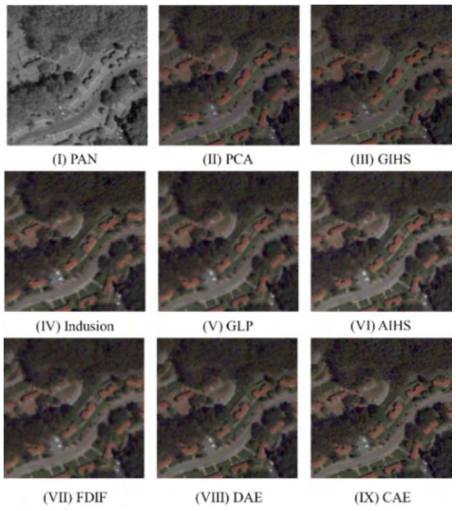

FIGURE 14. Full resolution fusion outcomes for a sample Pleiades-1A image (for better visual comparison, a cropped part of the entire image is shown).

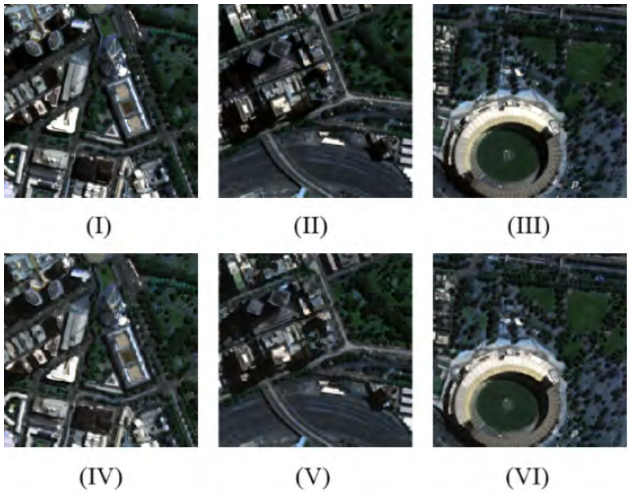

FIGURE 15. Fusion outcome corresponding to an unseen area: (I)-(III) the reference data and (IV) – (VI) the processed data.

the white squares in the PCA and GIHS methods became gray although the spatial information was injected into the fused image. The edges of the white squares in our method can be seen more clearly here in comparison with the other methods.

D. NETWORK GENERALIZATION

In this section, the generalization capability of the trained network on the Pleiades-1A dataset is considered. This was done by training the network on one area of the image and testing it for other areas. For this purpose, three non-overlapping areas of the Pleiades-1A dataset were chosen to serve as unseen data corresponding to vegetation and urban areas. The LRMS bands of these areas were then fed into the trained network after being patched by the same size as used in the training. The output patches were reconstructed and fed into the fusion framework with the results appearing in Fig. 15. As can be seen from this figure, the network was still able to inject the spatial information into the LRMS patches although they were chosen from different areas of the image. As an example, for the white area in the middle of the image in Fig. 15(IV), the spatial information was effectively injected into the fusion outcome without distorting any color information.

IV. CONCLUSION

In this paper, a pansharpening method has been developed based on a convolutional autoencoder network. A spatially degraded PAN image is first obtained via an interpolation filter. The original PAN images and their degraded versions are used to train a convolutional autoencoder network in a patch-wise manner. The LRMS bands are partitioned and fed into the trained network to obtain an estimation of high resolution LRMS images. It has been shown that this network can effectively inject the spatial information by learning the nonlinear mapping of degraded and clean PAN patches. The above approach is then used as part of the component substitution fusion framework to generate fused images. Objective performance measures were computed for three public domain datasets and the results obtained have demonstrated the effectiveness of the introduced fusion method relative to a representative set of existing fusion methods.

ACKNOWLEDGMENT

The authors would like to thank Prof. H. Ghassemian who provided valuable discussions during this work.

REFERENCES

[1] G. Khademi and H. Ghassemian, "A multi-objective component-substitution-based pansharpening," in *Proc. IEEE 3rd Int. Conf. Pattern Recog. Image Anal. (IPRIA)*, Apr. 2017, pp. 248–252.

[2] H. R. Shahdoosti and H. Ghassemian, "Combining the spectral PCA and spatial PCA fusion methods by an optimal filter," *Inf. Fusion*, vol. 27, pp. 150–160, Jan. 2016.

[3] M. Ghamchili and H. Ghassemian, "Panchromatic and multispectral images fusion using sparse representation," in *Proc. IEEE Artif. Intell. Sig. Process. Conf. (AISP)*, Oct. 2017, pp. 80–84.

[4] G. Khademi and H. Ghassemian, "Incorporating an adaptive image prior model into Bayesian fusion of multispectral and panchromatic images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 6, pp. 917–921, Jun. 2018.

[5] A. Azarang and H. Ghassemian, "An adaptive multispectral image fusion using particle swarm optimization," in *Proc. Iranian Conf. Elec. Eng. (ICEE)*, May 2017, pp. 1708–1712.

[6] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.

[7] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.

[8] S. Vitale, G. Ferraioli, and G. Scarpa, "A CNN-based model for pansharpening of worldview-3 images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 5108–5111.

[9] M. Gargiulo, A. Mazza, R. Gaetano, G. Ruello, and G. Scarpa, "A CNN-based fusion method for super-resolution of sentinel-2 data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 4713–4716.

[10] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, Nov. 2016.

[11] Y. Yang, L. Wu, S. Huang, J. Sun, W. Wan, and J. Wu, "Compensation details-based injection model for remote sensing image fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 734–738, May 2018.

[12] X. Li, Y. Pan, A. Gao, L. Li, S. Mei, and S. Yue, "Pansharpening based on joint Gaussian guided upsampling," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jun. 2018, pp. 7184–7187.

[13] P. Liu, L. Xiao, and T. Li, "A variational pan-sharpening method based on spatial fractional-order geometry and spectral-spatial low-rank priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1788–1802, Mar. 2018.

[14] H. R. Shahdoosti and N. Javaheri, "Pansharpening of clustered MS and Pan images considering mixed pixels," *IEEE Trans. Geosci. Remote. Lett.*, vol. 14, no. 6, pp. 826–830, Jun. 2017.

[15] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Geosci. Remote Sens. Lett.*, vol. 49, no. 1, pp. 295–309, Jan. 2011.

[16] M. M. Khan, J. Chanussot, L. Condat, and A. Montanvert, "Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 98–102, Jan. 2008.

[17] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.

[18] H. Li and L. Jing, "Improvement of MRA-based pansharpening methods through the considerasion of mixed pixels," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 5112–5115.

[19] Y. Xing, M. Wang, S. Yang, and K. Zhang, "Pansharpening with multiscale geometric support tensor machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2503–2517, May 2018.

[20] X. Lu, J. Zhang, and Y. Zhang, "An improved non-subsampled contourlet transform-based hybrid pan-sharpening algorithm," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 3393–3396.

[21] H. AanÆs, J. R. Sveinsson, A. A. Nielsen, T. Bovith, and J. A. Benediktsson, "Model-based satellite image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1336–1346, May 2008.

[22] X. Kang, S. Li, and J. A. Benediktsson, "Pansharpening with matting model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5088–5099, Aug. 2014.

[23] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model based pansharpening method based on TV and MTF deblurring," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 33–36.

[24] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS Pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 746–750, Oct. 2010.

[25] Y. Leung, J. Liu, and J. Zhang, "An improved adaptive intensity-hue-saturation method for the fusion of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 985–989, May 2014.

[26] Y. Song, W. Wu, Z. Liu, X. Yang, and K. Liu, and W. Lu, "An adaptive pan-sharpening method by using weighted least squares filter," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 18–22, Jan. 2016.

[27] M. Ghahremani and H. Ghassemian, "Nonlinear IHS: A promising method for pan-sharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1606–1610, Nov. 2016.

[28] M. R. Vicinanza, R. Restaino, G. Vivone, M. Dalla Mura, and J. Chanussot, "A pansharpening method based on the sparse representation of injected details," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 180–184, Jan. 2015.

[29] A. Azarang and H. Ghassemian, "Application of fractional-order differentiation in multispectral image fusion," *Remote Sens. Lett.*, vol. 9, no. 1, pp. 91–100, Jan. 2018.

[30] T. Wang, F. Fang, F. Li, and G. Zhang, "High-quality Bayesian pansharpening," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 227–239, Jan. 2019.

[31] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.

[32] A. Azarang and H. Ghassemian, "A new pansharpening method using multi resolution analysis framework and deep neural networks," in *Proc. IEEE 3rd Int. Conf. Pattern Recog. Image Anal. (IPRIA)*, Apr. 2017, pp. 1–6.

[33] Y. Rao, L. He, and J. Zhu, "A residual convolutional neural network for pan-shaprening," in *Proc. IEEE Int. Workshop Remote Sens. Intell. Process. (RSIP)*, Sep. 2017, pp. 1–4.

[34] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pan-sharpening method with deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1037–1041, May 2015.

[35] X. Liu, Y. Wang, and Q. Liu. (2018). "PSGAN: A generative adversarial network for remote sensing image pan-sharpening." [Online]. Available: https://arxiv.org/abs/1805.03371

[36] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018.

[37] L. Wald, *Data Fusion: Definitions and Architectures Fusion of Images of Different Spatial Resolutions*. Paris, France: Presses des MINES, 2002.

[38] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, 1998.

[39] H. Hallabia, A. Kallel, A. Ben Hamida, and S. Le Hégarat-Mascle, "High spectral quality pansharpening approach based on MTF-matched filter banks," *Multidimensional Syst. Signal Process.*, vol. 27, no. 4, pp. 831–861, Oct. 2016.

[40] C. B. Hsu, J. C. Lee, and T. M. Tu, "Generalized IHS-BT framework for the pansharpening of high-resolution satellite imagery," *J. Appl. Remote Sens.*, vol. 12, no. 4, Oct. 2018, Art. no. 046008.

[41] H. Li, W. Li, G. Han, and F. Liu, "Coupled tensor decomposition for hyperspectral pansharpening," *IEEE Access*, vol. 6, pp. 34206–34213, Jun. 2018.

[42] Y. Yang, L. Wu, S. Huang, W. Wan, and Y. Que, "Remote sensing image fusion based on adaptively weighted joint detail injection," *IEEE Access*, vol. 6, pp. 6849–6864, 2018.

[43] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[44] M. Imani, "Band dependent spatial details injection based on collaborative representation for pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4994–5004, Jul. 2018.

ARIAN AZARANG (S'15) received the B.S. degree from Shiraz University, Iran, in 2015, and the M.S. degree from Tarbiat Modares University, Iran, in 2017. He is currently pursuing the Ph.D. degree with The University of Texas at Dallas, all in electrical engineering. His research interests include signal and image processing, deep learning, remote sensing, and chaos theory. He received the first rank award for the B.S. degree.

HAFEZ E. MANOOCHEHRI (S'15) received the M.S. degree in computer science from The University of Texas at Dallas, in 2015, where he is currently pursuing the Ph.D. degree. Since 2018, he has been with the Bioinformatics Department, UT Southwestern Medical Center, Dallas, as a Research Collaborator. His research interests include image analysis and machine learning in biological applications.

NASSER KEHTARNAVAZ (S'82–M'86–SM'92–F'12) is currently an Erik Jonsson Distinguished Professor with the Department of Electrical and Computer Engineering and the Director of the Signal and Image Processing Laboratory, The University of Texas at Dallas. His research interests include signal and image processing, machine learning, and real-time implementation on embedded processors. He has authored or co-authored ten books and more than 380 journal papers, conference papers, patents, manuals, and editorials in these areas. He is a Fellow of SPIE and a Licensed Professional Engineer. He is currently serving as an Editor-in-Chief of the *Journal of Real-Time Image Processing*.