Anuário do Instituto de Geociências
Universidade Federal do Rio de Janeiro
https://revistas.ufrj.br/index.php/aigeo/

ISSN 0101-9759
e-ISSN 1982-3908

An Improved Pan-Sharpening Method Based on Convolutional Autoencoder and Guided Filter (CAE-GF) on Quickbird Images

Um Método Aprimorado de Pan-Sharpening Baseado em Autoencoder Convolucional e Filtro Guiado (CAE-FG) em Imagens Quickbird

Jessica da Silva Costa, Hideo Araki

Universidade Federal do Paraná, Programa de Pós-graduação em Ciências Geodésicas, Curitiba, PR, Brasil
E-mails: jessica.dsilvacosta@gmail.com, haraki@ufpr.br

Abstract

Pan-sharpening is the process of combining high-resolution panchromatic and low-resolution multispectral images to generate a high-resolution multispectral image. However, some pan-sharpening methods have poor preservation of spectral and spatial information. In this context, the aim of this study is to improve the pan-sharpening method based on a convolutional autoencoder and a guided filter (CAE-GF), by developing a new Convolutional Autoencoder (CAE) network. A CAE network was designed and trained to generate original panchromatic images from their spatially degraded versions. The trained network was used to improve the spatial details of the intensity component of the multispectral image obtained through the adaptive intensity-hue-saturation (AIHS) method. The pan-sharpening process is achieved by applying the multi-scale guided filter to improve the original PAN image using the enhanced intensity component and by injecting the detail map into the multispectral image. To analyze the preservation of the spectral and spatial information in the proposed method, full-reference and no-reference quality indices were calculated, and a visual analysis was performed. These analyses were compared with traditional pan-sharpening methods of component substitution. The results showed the developed method's efficacy in generating Pan-sharpened QuickBird images.

Keywords: Deep learning; Image fusion; Remote sensing

Resumo

Pan-sharpening é o processo de combinação de imagens pancromáticas de alta resolução e multiespectral de baixa resolução para gerar uma imagem multiespectral de alta resolução. Entretanto, alguns métodos de pan-sharpening apresentam pouca preservação da informação espectral e espacial. Neste contexto, o objetivo deste estudo é realizar uma melhoria no método pan-sharpening baseado em um Autoencoder Convolucional e um filtro guiado (CAE-FG) por meio do desenvolvimento de uma nova rede Autoencoder Convolucional (CAE). Uma rede CAE foi elaborada e treinada para gerar imagens pancromáticas originais a partir de suas versões espacialmente degradadas. A rede treinada foi usada para aprimorar os detalhes espaciais do componente de intensidade da imagem multiespectral, obtido pelo método adaptativo intensidade-matiz-saturação (AIHS). O processo de pan-sharpening é alcançado com a aplicação do filtro guiado multi-escala, para aperfeiçoar a imagem PAN original usando o componente de intensidade aprimorado, e com a injeção do mapa de detalhes na imagem multiespectral. Para analisar a preservação das informações espectrais e espaciais do método proposto foram calculados índices de qualidade de referência completa (full-reference) e sem referência (no-reference), e realizou-se uma análise visual. Estas análises foram comparadas com métodos tradicionais de pan-sharpening de substituição de componentes. Os resultados mostraram a eficácia do método desenvolvido na geração de imagens QuickBird pan-sharpened.

Palavras-chave: Deep learning; Fusão de imagens; Sensoriamento remoto

Received: 08 May 2023; Accepted: 02 July 2024

1 Introduction

Many Earth observation remote sensing platforms acquire multispectral (MS) and panchromatic (PAN) images, with MS images that include rich spectral content but have lower spatial resolution than PAN bands (Huang et al. 2022; Liu et al. 2023). The application of only MS or PAN images means that parts of the information content are discarded. Image fusion methods are frequently used to combine two or more images to produce enhanced images (Beene et al. 2022; Meng et al. 2020). The fusion of PAN and MS images, which is commonly called "Pan-sharpening", can be used to integrate the geometric details of the high spatial resolution (HR) PAN image and the spectral information of the low spatial resolution (LR) MS image to obtain an HR-MS image (Kaur et al. 2021; Vivone et al. 2021).

Several pan-sharpening methods have been developed, such as those presented by Dadrass Javan et al. (2021). These methods can be classified into two main groups based on the manner of extracting spatial details from a PAN image: (1) Component Substitution (CS) and (2) Multi-Resolution Analysis (MRA). However, some methods are not a part of these two categories, as is the case with Variational Optimization (VO) methods (Meng et al. 2019; Vivone et al. 2015). CS-based methods extract spatial details through the difference between the PAN image and the linear combination of the MS image, such as the Intensity-Hue-Saturation (IHS) method (Tu et al. 2001), Principal Component Analysis (PCA) method (Chavez & Kwarteng 1989), Gram-Schmidt (GS) method (Klonus & Ehlers 2009), Brovey Transformation (BT) method (Jiang et al. 2011), Sample Mean (SM) (Alcaras, Parente & Vallario 2021), Esri (Esri 2023) and Adaptive Intensity-Hue-Saturation (AIHS) (Rahmani et al. 2010). In contrast, the MRA-based methods extract spatial details through the difference between the PAN image and its low resolution, such as the Laplacian Pyramids (Aiazzi et al. 2006) and Wavelets (Otazu et al. 2005) methods.

Many pan-sharpening techniques have been traditionally used. However, some studies hold the viewpoint that traditional methods have poor preservation of spectral and spatial information (Mahyari & Yazdi 2011; Meng et al. 2019; Rahaman, Hassan & Ahmed, 2017). For example, CS-based methods are simple and regularly provide better spatial quality but exhibit a spectral distortion in the pan-sharpened image. On the other hand, MRA-based techniques generally show good spectral performance but include a spatial distortion (Dadrass Javan et al. 2021; Kaur et al. 2021; Meng et al. 2019).

Recently, Deep Learning (DL) based pan-sharpening methods using Convolutional Autoencoders (CAE) have been developed and have shown a great ability to simultaneously preserve the spatial details and spectral features of the pan-sharpened images. In the study by Azarang, Manoochehri & Kehtarnavaz (2019), a CAE architecture was used to develop a CS-based pan-sharpening method. Al Smadi et al. (2021) developed a pan-sharpening method based on a CAE and a guided filter (CAE-GF). The multi-scale guided filter was used to enhance the semantic detail map by utilizing the enhanced intensity image as a guidance image obtained by CAE. Al Smadi et al. (2022) present a pan-sharpening technique based on the non-subsampled contourlet transform (NSCT) and a CAE on QuickBird images. However, these studies perform patch-based CAE network training, in which the original image is divided into small patches, each patch is processed individually, and then the image is reconstructed from the individual processing. However, this approach causes the appearance of block artifacts in the reconstructed image, which influences the quality of the pan-sharpened image generated by the method.

Thus, this study aims to improve the CAE-GF method by developing a new CAE network architecture. In the proposed method, a CAE network is developed and trained to generate HR-PAN images from their spatially degraded versions. Next, the intensity component, obtained by the adaptive intensity-hue-saturation (AIHS) method, is tested on the trained network to obtain enhanced intensity components. Also, the guided filter is employed to enhance the PAN image using the intensity component estimated by the network. The pan-sharpening process is achieved by injecting the detail map into the LR-MS image. The result of the developed method is compared with the traditional pan-sharpening techniques using full-reference and non-reference metrics and visual analysis.

2 Methodology

To develop the proposed method, a CAE network was developed to enhance the spatial details of the intensity component of the upsampled MS-LR image. Subsequently, spatial details were extracted from the PAN image using a multiscale guided filter method, in which the estimated intensity component was used as the guiding image. Then, the spatial details of the PAN image and an injection gain matrix were injected into the upsampled MS-LR image, resulting in the pan-sharpened HR-MS image. Figure 1 presents the methodological flowchart of the proposed pan-sharpening method.

2.1 Convolutional Autoencoder (CAE)

In Autoencoder, unsupervised learning is performed using a neural network that is trained to attempt to copy its

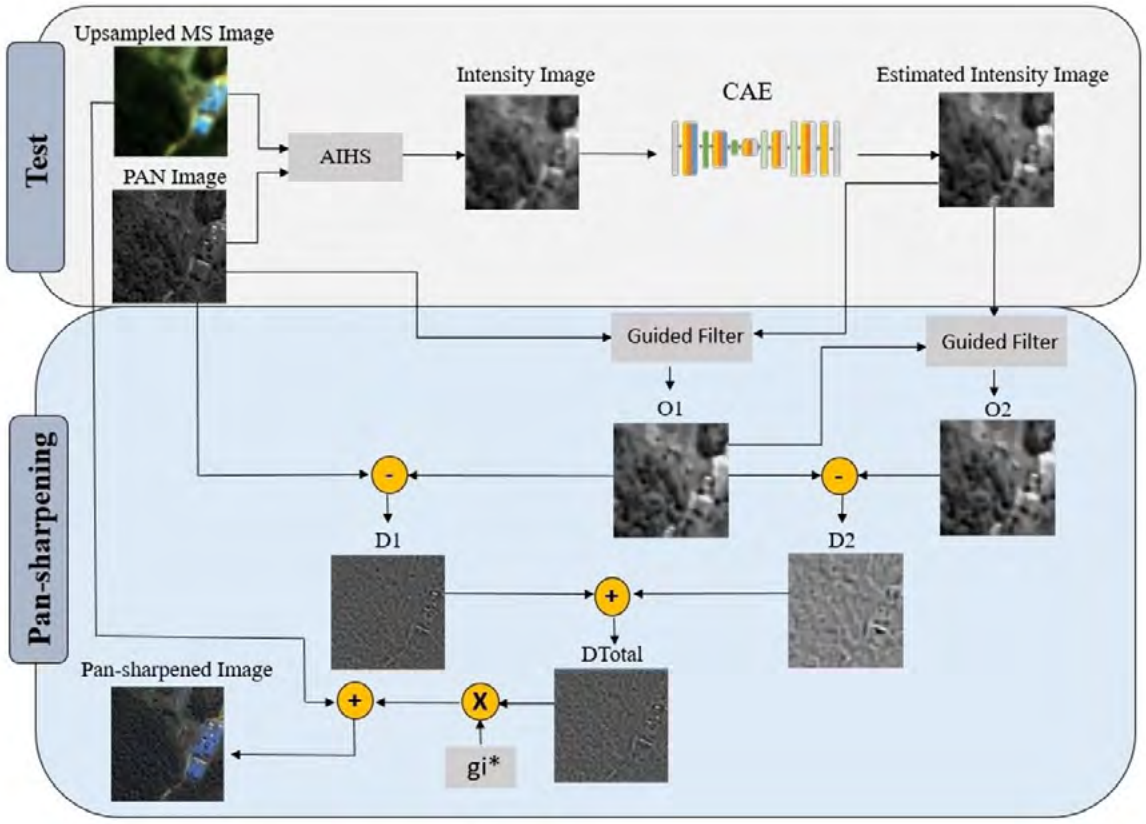

Figure 1 Methodological flowchart of the proposed method.

input to its output. Internally, it has a hidden layer (h) that describes a code used to represent the input. The network consists of two parts: an encoding function h=f(x) and a decoding function that produces a reconstruction r=g(h) (Goodfellow, Bengio & Courville 2016). The Convolutional Autoencoder is a type of convolutional neural network that can be used to generate an output image based on the input image. The encoding part includes convolution and max-pooling layers, while the decoding part consists of deconvolution and upscaling layers (Azarang, Manoochehri & Kehtarnavaz 2019).

2.2 Improved CAE-GF Method

The method proposed in this study was developed based on the pan-sharpening CAE-GF method developed by Al Smadi et al. (2021) but using a new Convolutional Autoencoder (CAE) architecture.

In the CAE network employed in the CAE-GF method, the input images must be 8x8 in size, so the PAN and MS images are partitioned into 8x8 patches with 5 overlapping pixels and then applied to the network. After that, the images are reconstructed and applied in the pan-sharpening process. Deep learning model training with patches allows for increased model performance and reduced computational time. However, processing each patch individually can generate block artifacts in the reconstructed image, causing a loss in image quality (Tarchouli et al. 2022).

In this regard, the method proposed in this study uses a CAE network with size 128x128 input images and presents more layers in its structure. This makes it possible to use larger-sized images, without the need to perform patch partitioning of the original images, which influences the quality of the generated pan-sharpened image. Also, the proposed method is simpler to apply because it does not need to perform this patch-based step.

2.3 Developed Convolutional Autoencoder

The developed Convolutional Autoencoder (CAE) network is composed of an encoding part, responsible for compressing the input data into a latent space representation,

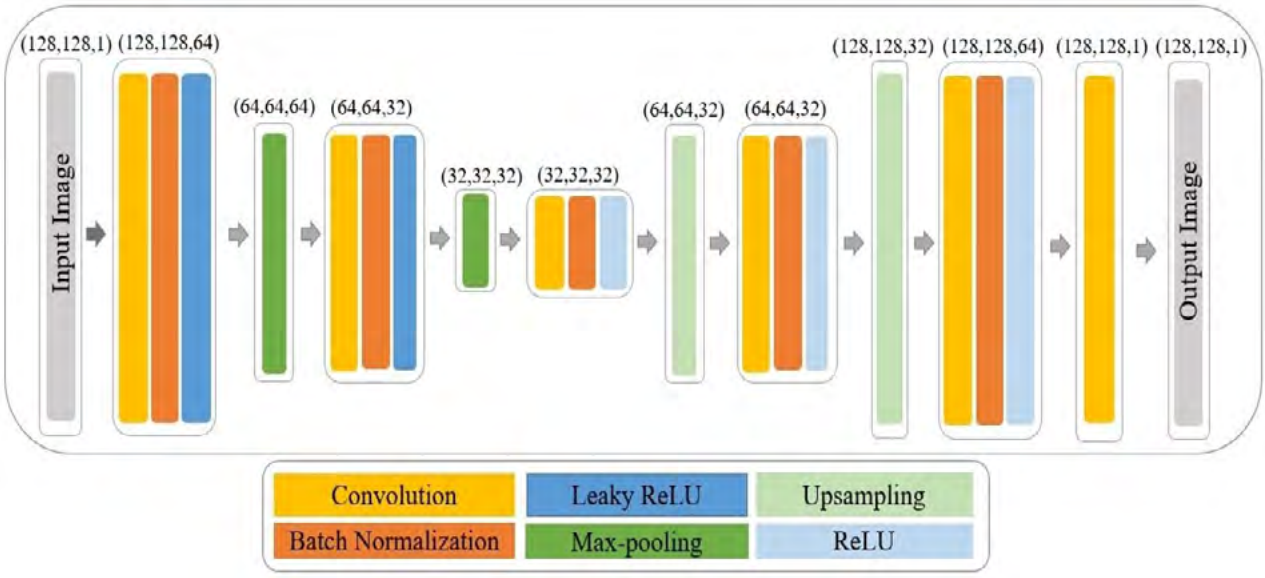

Figure 2 Developed CAE architecture.

which contains three two-dimensional convolution layers, two max-pooling layers, and a decoding part, which performs the reconstruction of the input data from the latent space representation, containing three two-dimensional convolution layers and two upsampling layers. Batch Normalization (BN) (Ioffe & Szegedy 2015) is applied to all layers except the output layer. This can avoid model instability and help the flow of gradients in the networks (Zhu et al. 2020). The Leaky Rectified Linear Unit (Leaky ReLU) activation function (Xu et al. 2015) is used after the encoding convolutions, and the Rectified Linear Unit (ReLU) activation function (Nair & Hinton 2010) is used after the decoding convolutions. The Sigmoid activation function is used for the output layer. The Mean Square Error (MSE) between the reconstructed output data and the input data is used to update the weights. After updating the weights, the backpropagation algorithm is used to train the network. The architecture of the developed CAE network can be seen in Figure 2.

To train the network, PAN-HR images were reduced to PAN-LR images generated by applying the bilinear interpolation filter. The network receives the spatially reduced PAN-LR image as input and produces an estimated PAN-HR image as output. In the testing step, the upsampled MS-LR image intensity component is introduced into the trained network to generate an estimated intensity component. Due to the similar characteristics between the PAN and the corresponding intensity component of the MS image, the trained network is reliable for enhancing the MS image's intensity component (Al Smadi et al. 2021). To calculate the intensity component, the Adaptive Intensity-Hue-Saturation (AIHS) method (Rahmani et al. 2010) was used, as per Equations 1 and 2, in which denotes the weight coefficients, and n represents the number of spectral bands. indicates the *ith* band of the MS image and PAN corresponds to panchromatic image.

$$I = \sum_{i=1}^{n} \alpha i M_{i'th} \quad (1)$$

$$\alpha_i^* = \arg \min_{\alpha_i} \left| PAN - \sum_{i=1}^{n} \alpha_i \; M_{i'th} \right|^2 \quad (2)$$

2.4 Pan-sharpening Process

In the pan-sharpening process development, the intensity component () estimated by the CAE network is employed to enhance the original PAN image using a guided filter (GF) (Ochotorena & Yamashita 2020) in two scales. In the first step, the is used as the guidance image and the PAN image as the input image, as observed in Equation 3. The difference between the approximate image and the

input PAN image is represented by the spatial detail , as seen in Equation 4.

$$O_1=GF(E_I,PAN) \quad (3)$$

$$D_1=PAN-O_1 \quad (4)$$

Then, in Equation 5, is used as the input image for the second scale of the guided filter . The difference between and is represented by the spatial detail , as shown in Equation 6.

$$O_2=GF(E_I,O_1) \quad (5)$$

$$D_2=O_1-O_2 \quad (6)$$

The total semantic map), shown in Equation 7, is injected into the upsampled MS image through injection gains which are calculated using Equation 8.

$$D_{Total}=D_1+D_2 \quad (7)$$

$$g_i = \frac{cov(MS_i,E_i)}{var(E_i)} \quad (8)$$

Thus, the high resolution multispectral (HR-MS) pan-sharpened image is obtained by Equation 9.

$$HR\text{-}MS = MSi+ giDtotal \quad (9)$$

2.5 Datasets

To perform the experiment, PAN images of 70cm spatial resolution and MS images of 2.8m spatial resolution, containing 4 bands (Blue, Green, Red & NIR), from the QuickBird sensor, located in the urban region of Curitiba, Paraná State, were used. The QuickBird images that were used exhibit a perspective effect because they were acquired from an inclined view, and not a nadiral one.

To train the network, 150 spatially degraded LR-PAN images at a scale of 4 and 150 original HR-PAN images of 128x128 pixels were used as input and validation, respectively, using 100 training images and 50 test images, and 100 epochs were performed. The CAE network training was performed using the Nvidia Tesla T4 GPU (Graphics Processing Units) provided by Google Colab (Google Colab 2023), with a processing time of 100.11s. Figure 3 presents the curve of changing loss value during training, showing a good fit, with the training and test loss values decreasing to a point of stability, but with the test data showing instability in the first 20 epochs.

Before applying the pan-sharpening methods used in this study, the MS and PAN images were cut so that they presented the same coverage region. After that, the MS image was geometrically compatible with the PAN image, and bilinear interpolation was performed.

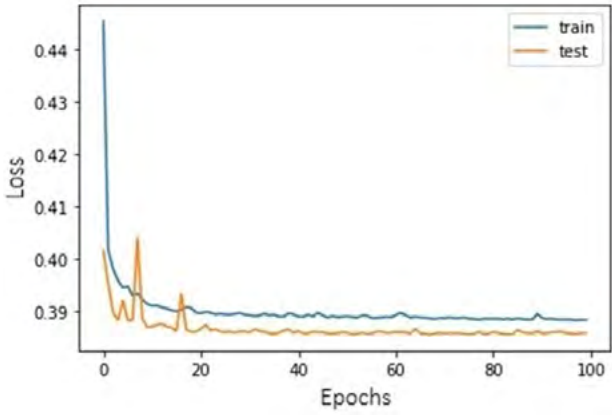

Figure 3 Change curve of the loss value on training.

2.6 Quality Analysis

To perform the spectral and spatial quality analysis of the proposed method, we performed a visual analysis and used quality indices that are commonly used in the evaluation of pan-sharpening methods.

In the quality analysis, two protocols that are commonly used in the validation of pan-sharpening methods were considered. The first protocol, known as Wald's protocol (Wald 2002), is considered full reference, in which the MS and PAN images are spatially degraded, on a scale of 4, so that the original MS image is used as a reference for the evaluation of the results. For this protocol, the Root Mean Square Error (RMSE) (Yang et al. 2015), Relative Average Spectral Error (RASE) (Ranchin & Wald 2000), Erreur Relative Globale Adimensionnelle de Synthese (ERGAS) (Palsson et al. 2016), Spectral Angle Mapper (SAM) (Rashmi et al. 2014), Universal Image Quality Index (UIQI) (Wang & Bovik 2002) and Correlation Coefficient (CC) (Ouahab & Belbachir 2020) spectral analysis indices were used. The second protocol is called no-reference

(Zhou, Civco & Silander 1998), in which the reference MS image is upsampled to the size of the PAN image. The analysis according to this protocol used the Quality Without Reference (QNR) index and its spectral distortion Dλ and spatial distortion Ds components (Alparone et al. 2008).

Two datasets of PAN and MS images of sizes 512x512 and 128x128, respectively, were selected to be evaluated according to the full-reference protocol. In addition, two datasets were used for evaluation in the no-reference protocol, with PAN images of size 128x128 and MS images of size 32x32. In all the images presented in this study, a 1% linear contrast was applied, in which thresholds are positioned to discriminate 1% of the pixels on each side of the histogram, to improve the visual aspect of the images.

The results of the proposed method were compared with a set of traditional pan-sharpening methods of component substitution, including the Brovey Transformation (BT), Sample mean (SM), Esri, and Intensity-Hue-Saturation (IHS) methods obtained through ArcGIS software, the Gram-Schmidt (GS) method obtained through ENVI software, and the Adaptive Intensity-Hue-Saturation (AIHS) method developed in Python language.

3 Results and Discussion

3.1 Pan-Sharpened Results of the Full Reference Datasets

Tables 1 and 2 present the results of the quality indices of the Pan-sharpened images generated by the full reference datasets. For the two analyzed datasets, the proposed method presented the best results for all calculated indices compared with the results obtained by the other methods. However, it obtained values close to the AIHS and GS methods. The AIHS method presented the second-best performance for the indexes in the first dataset (Table 1) and for the ERGAS, SAM, UIQI, and CC indexes in the second dataset (Table 2). The GS method presented the second-best performance for the RMSE and RASE indexes for the second dataset (Table 2). Overall, the results indicate that the proposed method showed greater preservation of the spectral information compared to the reference MS image than the other methods.

In the visual analysis, it is possible to observe that in Figures 4F and 5F, the proposed method presents a good preservation of spatial information, presenting spatial characteristics similar to those of the reference MS image (Figures 4A and 5A). Also, it is possible to see that the proposed method (Figures 4F and 5F) has a good preservation of spectral information, maintaining a coloration similar to that of the reference MS image (Figures 4A and 5A). The proposed method (Figures 4F and 5F) and the AIHS (Figures 4K and 5K) and GS (Figures 4L and 5L) methods presented a better performance in color preservation, especially in the coloration of vegetation areas, differently from that obtained by the BT (Figures 4G and 5G), SM (Figures 4H and 5H), Esri (Figures I4 and I5) and IHS (Figures 4J and 5J) methods, which exhibited a different coloration from the reference.

However, the proposed method (Figures 4F and 5F) presented a slight spectral and spatial distortion, with a small loss of spectral information and spatial details, exhibiting a blurry appearance in some areas, such as near the edges of buildings and shaded locations. Similarly, this was also observed in the images generated by AIHS (Figures 4K and 5K) and GS (Figures 4L and 5L). Since the Quickbird image used in this study was obtained from an inclined view in an urban region, there is a large presence of shadow and occlusion areas, which may have caused a loss of spectral and spatial information in the processing of the proposed pan-sharpening method, resulting in small distortions.

Table 1 Quality indices for dataset 1 full reference. The last row shows the ideal value for each index, in bold are the best values and in underlined are the second-best values.

Method	RMSE	RASE	ERGAS	SAM	UIQI	CC
Proposed	**39,943**	**17,316**	**4,484**	**4,234**	**0,933**	**0,852**
BT	79,001	34,247	10,027	13,491	0,840	0,675
SM	85,972	37,269	13,077	12,483	0,786	0,803
Esri	90,699	39,318	12,681	10,935	0,772	0,780
IHS	98,500	42,700	12,456	11,967	0,747	0,710
AIHS	40,907	17,733	4,707	4,398	0,931	0,843
GS	41,997	18,206	4,953	4,706	0,923	0,800
Ideal	0	0	0	0	1	1

Table 2 Quality indices for dataset 2 full reference. The last row shows the ideal value for each index, in bold are the best values and in underlined are the second-best values.

Method	RMSE	RASE	ERGAS	SAM	UIQI	CC
Proposed	**57,090**	**22,732**	**5,706**	**3,916**	**0,844**	**0,751**
BT	106,478	42,398	11,597	16,799	0,701	0,556
SM	94,313	37,554	11,306	10,937	0,685	0,702
Esri	124,392	49,531	14,458	10,418	0,594	0,667
IHS	108,692	43,279	11,773	10,195	0,631	0,573
AIHS	60,121	23,939	6,004	4,005	0,832	0,733
GS	58,967	23,480	6,129	4,429	0,819	0,687
Ideal	0	0	0	0	1	1

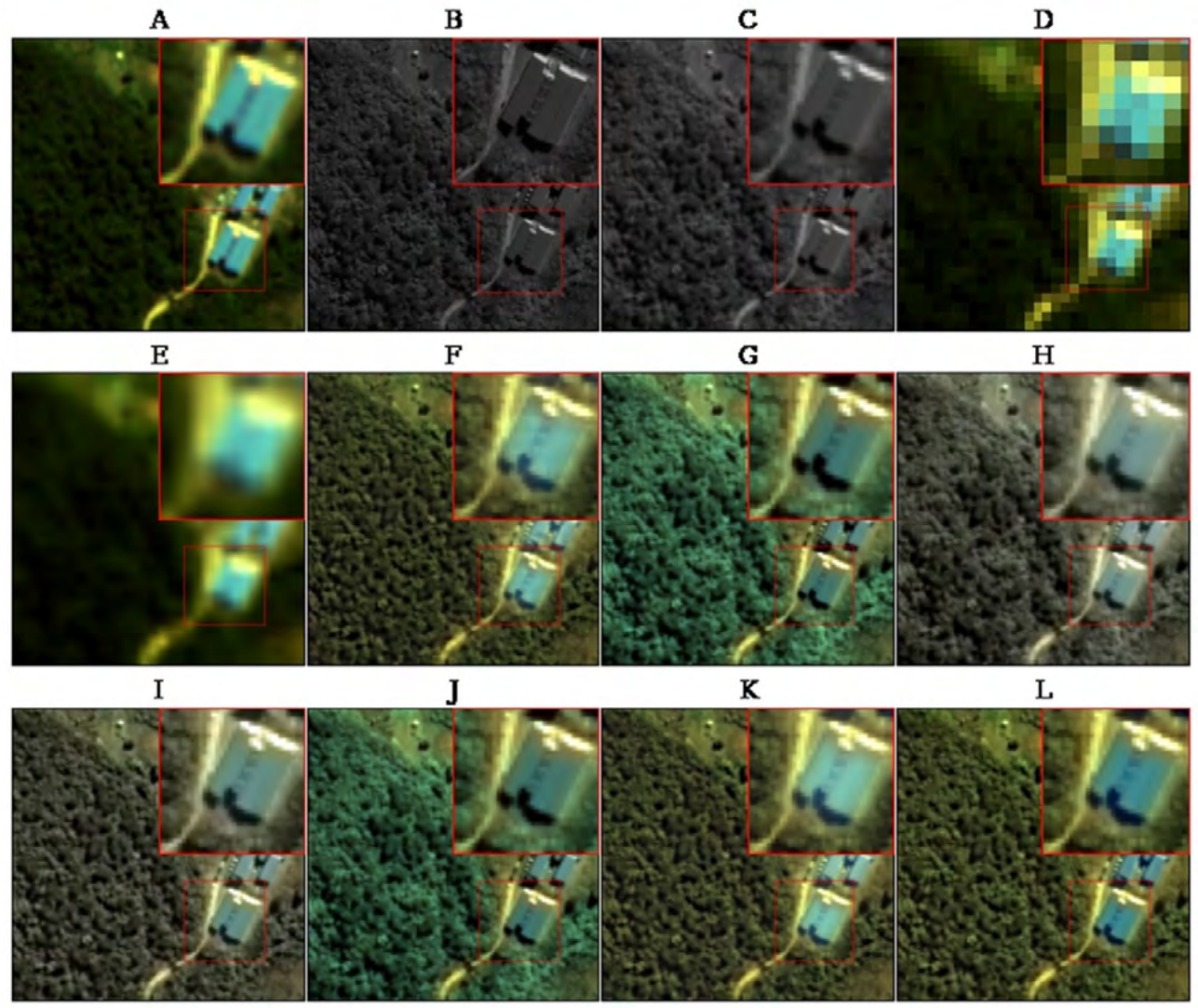

Figure 4 Pan-sharpened results of dataset 1 full reference: A. Reference MS image (128×128); B. PAN image (512×512); C. Degraded PAN image (128×128); D. Degraded MS image (32×32); E. Degraded MS image Upsampled (128×128); F. Proposed method; G. BT method; H. SM method; I. Esri method; J. IHS method; K. AIHS method; L. GS method.

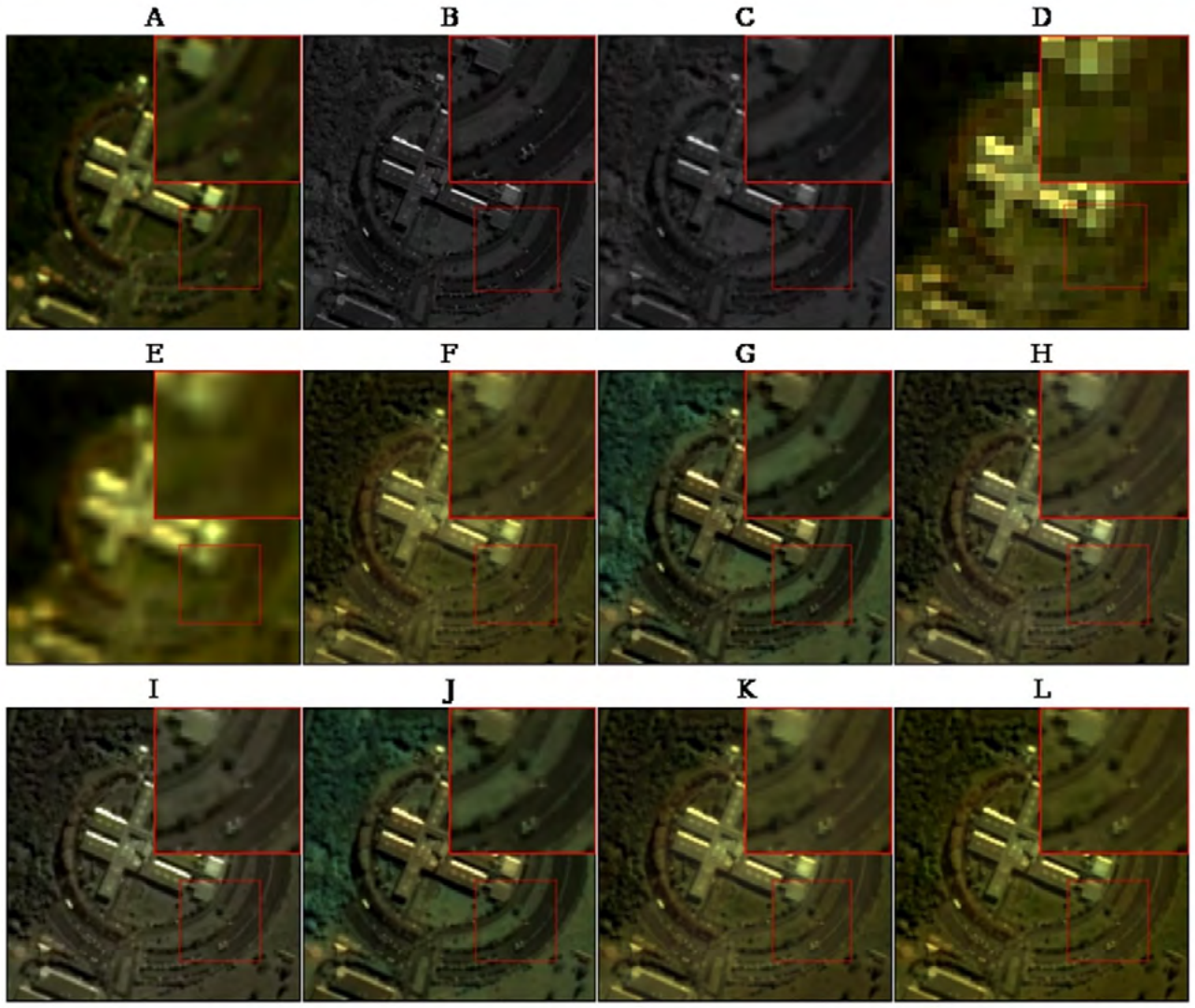

Figure 5 Pan-sharpened results of dataset 2 full reference: A. Reference MS image (128×128); B. PAN image (512×512); C. Degraded PAN image (128×128); D. Degraded MS image (32×32); E. Degraded MS image Upsampled (128×128); F. Proposed method; G. BT method; H. SM method; I. Esri method; J. IHS method; K. AIHS method; L. GS method.

3.2 Pan-Sharpened Results of the No-Reference Datasets

The results of the quality indices of the Pan-sharpened images generated for the no-reference datasets can be seen in Tables 3 and 4. For the two datasets that were analyzed, the GS method presented a better performance for the dλ component and the QNR index, indicating less spectral and global distortion, respectively. However, the proposed method showed, for both datasets, the second-best result for the dλ component and the QNR index. This shows the method's great potential for obtaining low spectral and global distortion. However, the proposed method showed better performance for the ds index in both datasets, indicating less spatial distortion of the generated Pan-sharpened image compared to the other methods. Furthermore, Tables 3 and 4 demonstrate that the proposed method showed results close to the required optimal values for the ds, dλ, and QNR indices for both datasets.

In Figures 6D and 7D, it is possible to see that the proposed method was better than the MS-LR image (Figures 6B and 7B) in terms of spatial content, indicating the pan-sharpened image's spatial information preservation potential (Figures 6A and 7A). Also, we can easily analyze the great color preservation ability of the Pan-sharpened image generated by the proposed method (Figures 6D and 7D) compared with the MS image (Figures 6C and 7C), compared with the traditional methods analyzed in this

Table 3 Quality indices for dataset 1 no-reference. The last row shows the ideal value for each index, in bold are the best values and in underlined are the second-best values.

Method	ds	dλ	QNR
Proposed	**0,093**	0,039	0,872
BT	0,286	0,176	0,588
SM	0,578	0,339	0,279
Esri	0,497	0,203	0,401
IHS	0,640	0,260	0,267
AIHS	0,118	0,070	0,820
GS	0,106	**0,017**	**0,878**
Ideal	0	0	1

Table 4 Quality indices for dataset 2 no-reference. The last row shows the ideal value for each index, in bold are the best values and in underlined are the second-best values.

Method	ds	dλ	QNR
Proposed	**0,082**	0,038	0,883
BT	0,225	0,197	0,623
SM	0,534	0,332	0,311
Esri	0,442	0,105	0,499
IHS	0,602	0,331	0,266
AIHS	0,097	0,048	0,860
GS	0,086	**0,010**	**0,905**
Ideal	0	0	1

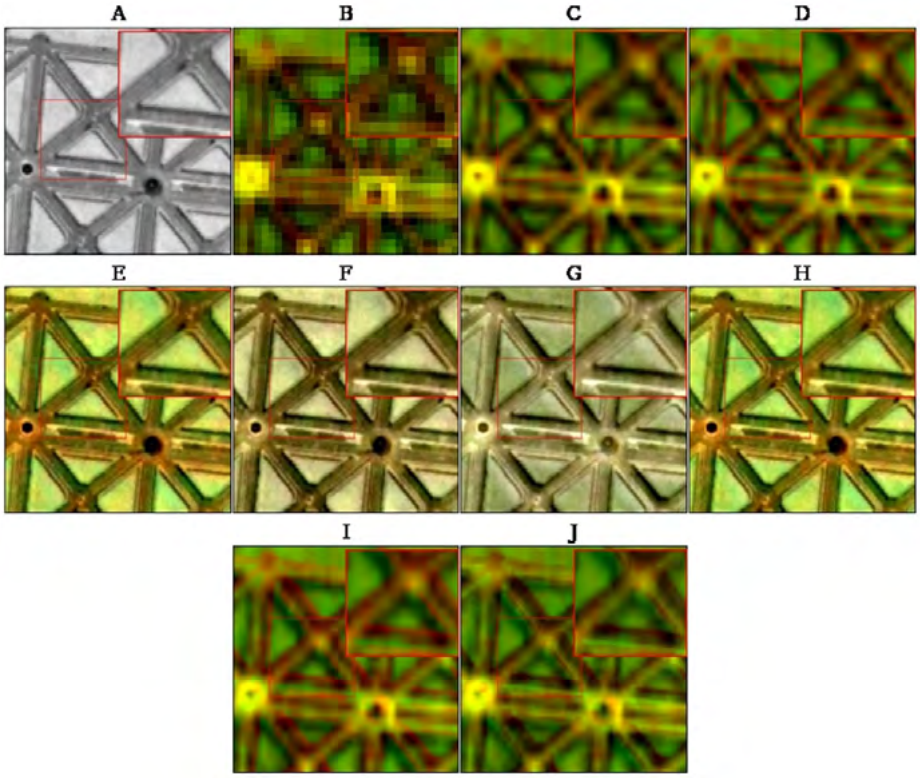

Figure 6 Pan-sharpened results of dataset 1 no-reference: A. PAN image (128×128); B. MS image (32×32); C. MS image Upsampled (128×128); D. Proposed method; E. BT method; F. SM method; G. Esri method; H. IHS method; I. AIHS method; J. GS method.

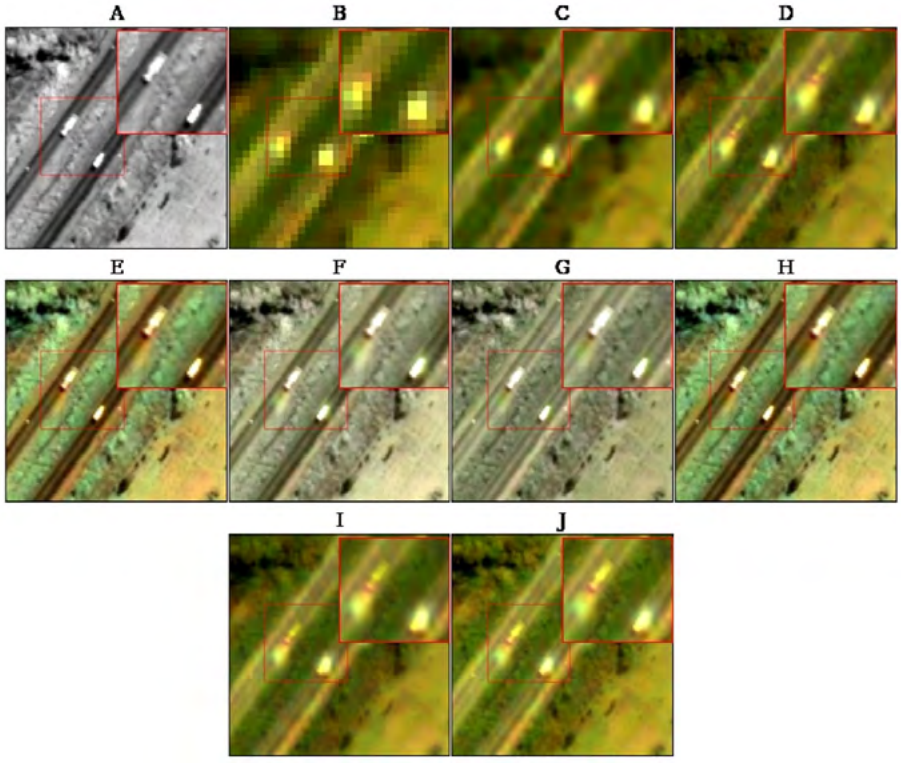

Figure 7 Pan-sharpened results of dataset 2 no-reference: A. PAN image (128×128); B. MS image (32×32); C. MS image Upsampled (128×128); D. Proposed method; E. BT method; F. SM method; G. Esri method; H. IHS method; I. AIHS method; J. GS method.

study. The study conducted by Azarang, Manoochehri & Kehtarnavaz (2019) using a CAE architecture to develop a pan-sharpening method also showed improved color preservation performance. As observed in the previous section, is the method exhibits a greater potential for preserving the color in vegetation areas than that observed in the BT (Figures 6E and 7E), SM (Figures 6F and 7F), Esri (Figures 6G and 7G) and IHS (Figures 6H and 7H) methods. It is apparent in Figures 6 and 7 that the proposed method (Figures 6D and 7D) exhibited a coloration close to that obtained by the AIHS (Figures 6I and 7I) and GS (Figures 6J and 7J) methods for both datasets. However, the AIHS and GS methods have lighter colors compared with the MS image (Figures 6C and 7C).

An analysis of Figures 6 and 7 demonstrates that the proposed method presents spectral and spatial distortions in some locations, for example, at the edges of more detailed objects (Figure 6D) and at locations with moving objects (Figure 7D), a phenomenon that was also observed in the images generated by the AIHS (Figures 6I and 7I) and GS (Figures 6J and 7J) methods. In Figure 6, in the BT (Figure 6E), SM (Figure 6F), Esri (Figure 6G) and IHS (Figure 6H) methods, it is possible to observe the image details more easily, but they exhibited a loss of spectral information. However, the proposed method (Figure 6D) was able to preserve the spectral information and get a coloration closer to the MS image (Figure 6C) while still obtaining an image with better spatial resolution. In Figure 7, it is possible to observe that the proposed method (Figure 7D) and the AIHS (Figure 7I) and GS (Figure 7J) methods showed a blur in areas with moving objects. Due to the different sensor characteristics, physical locations, and capture time, PAN and MS images inevitably have a large pixel misalignment that is even worse for locally moving objects, such as cars on a highway. This usually leads to the presence of various artifacts in the generated pan-sharpened images, such as double edges and blurred artifacts, especially near moving objects, with extreme misalignment (Lee, Seo & Kim 2021).

4 Conclusion and Future Studies

In this study, a pan-sharpening method based on the CAE-GF method was proposed, with the development of a new Convolutional Autoencoder network. The network was trained to learn the relationship between the original PAN images and their degraded versions to produce HR-PAN images. The trained network is used to enhance the spatial details of the intensity component. Then, the multi-scale guided filter is applied to enhance the original PAN image. The results showed that the proposed method performed

better for all quality indices tested in the full reference protocol, and for the ds component of the QNR index for the no-reference protocol, compared to the traditionally used pan-sharpening methods of component substitution. In the visual analysis, one can observe that the proposed method presented small spectral and spatial distortions in some places in the generated image. It also showed blurred artifacts in areas with moving objects. However, the proposed method showed a great ability to maintain the colors of the MS image when compared with the comparative methods. Thus, we can conclude that the proposed method has great potential in preserving the spatial information of the HR-PAN image and the spectral information of the LR-MS image during the pan-sharpening process. Also, it is possible to conclude that the CAE-FG method can be improved with satisfactory results.

For future studies, it is recommended to apply the proposed methodology in other images with higher spatial resolution, such as Pleiades, GeoEye, and WorldView images. It is also advisable to use images that present more spectral bands, such as Landsat images, to analyze the method's performance in images that present only part of the spectral bands covered by the PAN image's spectral range. Furthermore, it is recommended to analyze the potential of the images generated by this method in the calculation of vegetation indices, such as NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index), and SAVI (Soil Vegetation Index), analyzing the method's effectiveness in generating an HR-MS image with a distribution of vegetation indices similar to the LR-MS image. Also, we suggest the development of pan-sharpening methods that minimize the effects of blurring in the images caused by the misalignment of the PAN and MS images, which happens due to the different acquisition times for moving objects.

5 References

Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A. & Selva, M. 2006, 'MTF-tailored multiscale fusion of high-resolution MS and Pan imagery', *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 5, pp. 591-6, DOI:10.14358/PERS.72.5.591.

Al Smadi, A., Yang, S., Abugabah, A. Alzubi, A.A. & Sanzogni, L. 2022, 'A Pansharpening Based on the Non-Subsampled Contourlet Transform and Convolutional Autoencoder: Application to QuickBird Imagery', *IEEE Access*, vol. 10, pp. 44778-88, DOI:10.1109/ACCESS.2022.3169698.

Al Smadi, A., Yang, S., Kai, Z., Mehmood, A., Wang, M. & Alsanabani, A. 2021, 'Pansharpening based on convolutional autoencoder and multi-scale guided filter', *J Image Video Proc,* vol. 25, DOI:10.1186/s13640-021-00565-3.

Alcaras, E., Parente, C. & Vallario, A. 2021, 'Automation of Pan-Sharpening Methods for Pléiades Images Using GIS Basic Functions', *Remote Sens.,* vol. 13*,* no. 8, 1550, DOI:10.3390/rs13081550.

Alparone, L., Aiazzi, B., Baronti, S., Garzelli, A., Nencini, F. & Selva, M. 2008, 'Multispectral and panchromatic data fusion assessment without reference', *Photogrammetric Engineering & Remote Sensing*, vol. 74, no. 2, pp.193-200, DOI:10.14358/PERS.74.2.193.

Azarang, A, Manoochehri, H.E. & Kehtarnavaz, N. 2019, 'Convolutional Autoencoder-Based Multispectral Image Fusion', *IEEE Access*, vol. 7, pp. 35673-83, DOI:10.1109/ACCESS.2019.2905511.

Beene, D., Zhang, S., Lippitt, C.D. & Bogus, S.M. 2022, 'Performance Evaluation of Multiple Pan-Sharpening Techniques on NDVI: A Statistical Framework', *Geographies*, vol. 2, pp.435-452, DOI:10.3390/geographies2030027.

Chavez, A. & Kwarteng, P. 1989, 'Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis', *Photogrammetric Engineering & Remote Sensing,* vol. 55, no. 1, pp. 339-48.

Esri 2023. *Fundamentals of Panchromatic Sharpening. ArcGIS for Desktop*, viewed 27 March 2023, <https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/fundamentals-of-panchromatic-sharpening.htm>.

Goodfellow, I., Bengio, Y. & Courville, A. 2016, *Deep Learning*, MITPress.

Google Colab 2023. *Google Colab*, viewed 27 March 2023, <https://colab.research.google.com/notebooks/intro.ipynb>.

Huang, M., Liu, S., Li, Z., Feng, S., Wu, D., Wu, Y. & Shu, F. 2022, 'Remote Sensing Image Fusion Algorithm Based on Two-Stream Fusion Network and Residual Channel Attention Mechanism', *Wireless Communications and Mobile Computing,* vol. 2022, 14, DOI:10.1155/2022/8476000.

Ioffe, S. & Szegedy, C. 2015, 'Batch normalization: Accelerating deep network training by reducing internal covariate shift', *International conference on machine learning*, vol. 37, pp. 448-56, DOI:10.48550/arXiv.1502.03167.

Javan, F.D., Samadzadegan, F., Mehravar, S., Toosi, A., Khatami, R. & Stein, A. 2021, 'A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery', *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.171, pp. 101-17, DOI:10.1016/j.isprsjprs.2020.11.001.

Jiang, D., Zhuang, D., Huang, Y. & Fu, J. 2011, 'Survey of multispectral image fusion techniques in remote sensing applications', *Image fusion its Appl*., pp.1-23, DOI:10.5772/10548.

Kaur, G., Saini, K.S., Singh, D. & Kaur, M. 2021, 'A Comprehensive Study on Computational Pansharpening Techniques for Remote Sensing Images', *Archives of Computational Methods in Engineering*, vol. 28, pp. 4961-978, DOI:10.1007/s11831-021-09565-y.

Klonus, S. & Ehlers, M. 2009, 'Performance of evaluation methods in image fusion', *12th International Conference on Information Fusion*, IEEE, Seattle, WA, pp. 1409-16.

Lee J., Seo S. & Kim, M. 2021, 'SIPSA-Net: Shift-Invariant Pan Sharpening with Moving Object Alignment for Satellite Imagery', *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* Nashville, TN, USA, pp. 10161-69. DOI:10.1109/CVPR46437.2021.01003.

Liu H., Deng L., Dou Y., Zhong X. & Qian Y. 2023, 'Pansharpening Model of Transferable Remote Sensing Images Based on Feature Fusion and Attention Modules', *Sensors*, vol. 23, no. 6, 3275, DOI:10.3390/s23063275.

Mahyari, A.G. & Yazdi. M. 2011, 'Panchromatic and multispectral image fusion based on maximization of both spectral and spatial similarities'*, IEEE Transactions on Geoscience and Remote Sensing*. vol. 49, no. 6, pp. 1976-85, DOI:10.1109/TGRS.2010.2103944.

Meng, X., Shen, H., S., Li, H., Zhang, L. & Randi Fu, R. 2019, 'Review of the Pansharpening Methods for Remote Sensing Images Based on the Idea of Meta-analysis: Practical Discussion and Challenges', *Information Fusion*, vol. 46, pp. 102-13, DOI:10.1016/j.inffus.2018.05.006.

Meng, X., Xiong, Y., Shao, F., Shen, H., Sun, W., Yang, G., Yuan, Q., Fu, R., & Zhang, H. 2020, 'A Large-Scale Benchmark Data Set for Evaluating Pansharpening Performance: Overview and Implementation', *IEEE Geoscience and Remote Sensing Magazine***,** vol. 9, no. 1, pp. 18-52, DOI:10.1109/MGRS.2020.2976696.

Nair, V. & Hinton, G.E. 2010, 'Rectified Linear Units Improve Restricted Boltzmann Machines', *International Conference on Machine Learning*, Haifa, Israel, pp. 807-14.

Ochotorena, C.N. & Yamashita, Y. 2020, 'Anisotropic guided filtering', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 29, pp. 1397-412, DOI:10.1109/TIP.2019.2941326.

Otazu, X., Gonzalez-Audícana, M., Fors, O. & Núnez, J. 2005, 'Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2376-85, DOI:10.1109/TGRS.2005.856106.

Ouahab, A. & Belbachir, M.F. 2020, 'A comparison analysis of pan-sharpening methods on Alsat-2A images', *2020 2nd International Conference on Mathematics and Information Technology (ICMIT)*, Adrar, Algeria, pp. 138-41, DOI:10.1109/ICMIT47780.2020.9046990.

Palsson, F., Sveinsson, J.R., Ulfarsson, M.O. & Benediktsson, J.A. 2016, 'Quantitative quality evaluation of pansharpened imagery: Consistency versus synthesis', *IEEE Transactions on Geoscience and Remote Sensing,* vol. 54, no. 3, pp. 1247-59, DOI:10.1109/TGRS.2015.2476513.

Rahaman, K.R., Hassan, Q.K. & Ahmed, M.R. 2017, 'Pan-sharpening of Landsat-8 images and its application in calculating vegetation greenness and canopy water contents', *ISPRS International Journal of Geo-Information,* vol. 6*,* no.6, 168.

Rahmani, S., Strait, M., Merkurjev, D., Moeller, M. & Wittman, T. 2010, 'An adaptive IHS pan-sharpening method', *IEEE Geoscience and Remote Sensing Letters,* vol.7, no. 4, pp.746-50, DOI:10.1109/LGRS.2010.2046715.

Ranchin T. & Wald, L. 2000, 'Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation'*, Photogrammetric Engineering & Remote Sensing*, vol. 66, no. 1, pp. 49-61.

Rashmi, S., Addamani, S., Venkat, & Ravikiran, S. 2014, 'Spectal Angle Mapper Algorithm for remote Sensing Image Classification', *International Journal of Innovative Science, Engineering & Technology*, vol. 1, no. 4, pp. 201-5.

Tarchouli, M., Pelurson, S., Guionnet, T., Hamidouche, W., Outtas, M. & Deforges, O. 2022, 'Patch-based image coding with end-to-end learned codec using overlapping', *Computer Science & Information Technology*, vol. 12, no. 23, pp. 53-63, DOI:10.5121/csit.2022.122305.

Tu, T.M., Su, S.C., Shyu, H.C. & Huang, P.S. 2001, 'A new look at IHS-like image fusion methods', *Information Fusion,* vol.2, no.3, pp.177-86, DOI:10.1016/S1566-2535(01)00036-7.

Vivone, G.*,* Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G.A., Restaino, R. & Wald, L. 2015, 'A Critical Comparison Among Pansharpening Algorithms', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565-86, DOI:10.1109/TGRS.2014.2361734.

Vivone, G., Dalla Mura, M., Garzelli, A., Restaino, R., Scarpa, G., Ulfarsson, M.O., Alparone, L., & Chanussot, J. 2021*,* 'A New Benchmark Based on Recent Advances in Multispectral Pansharpening: Revisiting Pansharpening With Classical and Emerging Pansharpening Methods', *IEEE Geoscience and Remote Sensing Magazine,* vol. 9, no. 1, pp. 53-81, DOI:10.1109/MGRS.2020.3019315.

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

Wang, Z. & Bovik, A. C. 2002, 'A universal image quality index', *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–4. https://doi.org/10.1109/97.995823

Xu, B., Wang, N., Chen, T. & Li, M. 2015, 'Empirical Evaluation of Rectified Activations in Convolutional Network', *ArXiv e-prints*, DOI:10.48550/arXiv.1505.00853.

Yang, Y., Tong, S., Huang, S. & Lin, P. 2015, 'Multifocus image fusion based on NSCT and focused area detection', *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2824-38, DOI:10.1109/JSEN.2014.2380153.

Zhou, J., Civco, D.L. & Silander, J.A. 1998, 'A wavelet transform method to merge Landsat TM and SPOT panchromatic data', *International Journal of Remote Sensing*, vol. 19, no. 4, pp.743-57, DOI:10.1080/014311698215973.

Zhu, D., Cheng, X., Zhang, F. Yao, X., Gao, Y. & Liu, Y. 2020, 'Spatial interpolation using conditional generative adversarial neural networks', *International Journal of Geographical Information Science*, vol. 34, no. 4, pp.735-58, DOI:10.1080/13658816.2019.1599122.

Author contributions

Jessica da Silva Costa: conceptualization; formal analysis; methodology; validation; writing-original draft; writing – review and editing; visualization. **Hideo Araki:** conceptualization; formal analysis; methodology; validation; writing-original draft; writing – review and editing; supervision; visualization.

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

Scripts and code are available on request.

Funding information

Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Finance Code 001.

Editor-in-chief

Dr. Claudine Dereczynski

Associate Editor

Dr. Jader de Oliveira Santos

How to cite:

Costa, J. S. & Araki, H. 2024, 'An Improved Pan-Sharpening Method Based on Convolutional Autoencoder and Guided Filter (CAE-GF) on Quickbird Images', *Anuário do Instituto de Geociências,* 47:58601. https://doi.org/10.11137/1982-3908_2024_47_58601