GEOMATICS AND ENVIRONMENTAL ENGINEERING • Volume 17 • Number 5 • 2023

https://doi.org/10.7494/geom.2023.17.5.145

Mahmoud M. Hammad[1], Tarek A. Mahmoud[2],
Ahmed Saleh Amein[3], Tarek S. Ghoniemy[4]

Satellite Image Fusion Using a Hybrid Traditional and Deep Learning Method

Abstract: Due to growing demand for ground-truth in deep learning-based remote sensing satellite image fusion, numerous approaches have been presented. Of these approaches, Wald's protocol is the most commonly used. In this paper, a new workflow is proposed consisting of two main parts. The first part targets obtaining the ground-truth images using the results of a pre-designed and well-tested hybrid traditional fusion method. This method combines the Gram–Schmidt and curvelet transform techniques to generate accurate and reliable fusion results. The second part focuses on the training of a proposed deep learning model using rich and informative data provided by the first stage to improve the fusion performance. The demonstrated deep learning model relies on a series of residual dense blocks to enhance network depth and facilitate the effective feature learning process. These blocks are designed to capture both low-level and high-level information, enabling the model to extract intricate details and meaningful features from the input data. The performance evaluation of the proposed model is carried out using seven metrics such as peak-signal-to-noise-ratio and quality without reference. The experimental results demonstrate that the proposed approach outperforms state-of-the-art methods in terms of image quality. It also exhibits the robustness and powerful nature of the proposed approach which has the potential to be applied to many remote sensing applications in agriculture, environmental monitoring, and change detection.

Keywords: deep learning image fusion, remote sensing image fusion, remote sensing optical image, pan-sharpening, remote sensing image

Received: 15 February 2023; accepted: 12 July 2023

[1] Military Technical College, Cairo, Egypt, email: radok2003@hotmail.com
[2] Military Technical College, Cairo, Egypt, email: tarek.mahmoud.mtc@gmail.com, https://orcid.org/0000-0002-5714-4596
[3] October 6 University, Cairo, Egypt, email: ahmed.saleh.csis@o6u.edu.eg, https://orcid.org/0000-0002-7675-2140
[4] Military Technical College, Cairo, Egypt, email: ghoniemy_t@mtc.edu.eg, https://orcid.org/0000-0003-4919-4232

1. Introduction

Pansharpening is the process of combining both spatial and spectral information into one image [1]. Recently, the significance of creating accurate pansharpening algorithms has become evident, leading to the proposal of multiple methods to solve the pansharpening problem. Image fusion methods can be categorized into traditional approaches and deep learning techniques. The traditional methods involve three main processes: image transformation, activity level measurement, and fusion rule design. Being manually designed and based on theoretical assumptions, the selection and design of activity level measurements and fusion rules are challenging. Additionally, their performance can vary with imaging sensors, land cover characteristics, acquisition geometry, and complex transformations. Traditional methods are classified into three categories: component substitution (CS), multiresolution analysis (MRA), and hybrid methods. CS-based methods separate spatial and spectral information and replace the spatial details with a panchromatic (PAN) image. However, these methods require a high correlation among image components to minimize spectral distortion. In contrast, MRA-based methods address spectral deformation by employing spatial detail extraction methods while preserving spectral accuracy [2]. Hybrid methods, using both CS-based and MRA-based categories, combine the benefits of both methods. Deep learning methods, particularly convolutional neural networks (CNNs), provide an automatic solution to the pansharpening problem as they are shift, scale, and distortion invariant [3]. Using neural networks (NN) targets the preservation of both spatial and spectral information in the fused images. The exponential growth of datasets and improvements in computational power have led to the remarkable success of deep learning in extracting image features for use in image processing applications. As a result, many deep learning networks have been developed specifically for image fusion tasks. Researchers have explored various techniques to enhance fusion performance by increasing network depth, transfer learning, and using multiscale and multi-depth CNNs, as discussed in the following section.

In the paper, a novel approach called traditional deep learning image fusion (TDIF) is introduced as a solution for satellite remote sensing image fusion. Unlike many other deep learning-based image fusion methods, TDIF takes a different approach using a well-established and accurate traditional image fusion method as the ground truth during the training process. Typically, deep learning-based image fusion methods rely on low-resolution images as the ground truth for training their models. However, TDIF deviates from this approach and instead leverages the expertise and accuracy of traditional image fusion methods to generate high-quality fused images. By using traditional image fusion methods as the ground truth, TDIF benefits from their established performance and accuracy to enhance the training process. This unique characteristic of TDIF makes it distinguishable from other deep learning-based image fusion approaches and also contributes to improving the fusion results. By incorporating the power of traditional fusion methods and deep

learning capabilities, TDIF aims to provide a robust and accurate solution for remote sensing satellite image fusion tasks.

To optimize this approach, the following steps can be taken:

1. Experiment with different traditional image fusion methods to find the one that produces the best results as the ground-truth for training the deep learning model.
2. Explore different deep learning architectures and training strategies to optimize the performance of the TDIF method.
3. Use a larger and more diverse dataset for training and testing the TDIF method to increase its robustness and generalizability.
4. Evaluate the performance of the TDIF method using a variety of quality metrics, both qualitatively and quantitatively, to ensure that it performs well in different scenarios and applications.
5. Compare the TDIF method to other state-of-the-art image fusion techniques to demonstrate its superiority.
6. Fine-tune the parameters of the traditional method and the deep learning model to improve the results and make the method more robust.

The organization of the remaining parts of the paper is as follows. In Section 2, previous related researches are discussed. Section 3 offers a summary of the framework of the proposed approach. Section 4 details the data and outlines the experimental setup. This section also contains the experimental results and presents a discussion of them. Finally, Section 5 presents some conclusions and the future work of the study.

2. Related Works

The development of deep learning algorithms for image fusion, which can be seen as a subset of image super resolution, began in 2016 with the creation of the simplest super resolution deep learning algorithm, super-resolution convolutional neural network SRCNN, which utilized three convolutional neural networks (CNNs) [4]. This led to the development of the first deep learning image fusion algorithm, pan-sharpening by convolutional neural networks (PNN) [5]. PNN effectively preserves the spectral information at the cost of losing the spatial one. Many subsequent works continued to build on this foundation; for example, the author in [6] created a two-stage model relying on a SRCNN to increase the resolution of the intensity component of multispectral (MUL) image. The final fused image was produced by combining PAN image and the result from the first stage using the Gram–Schmidt method. In [7], the author proposed the pan-sharpening network (PanNet) that improved upon previous methods by better preserving both spatial and spectral information. However, this came at the cost of some blurring in the final fused image. Over time, researchers have sought to improve the results of these algorithms by making the networks

deeper using, for example, transfer learning, changing the loss function, and utilizing multiscale and multi-depth CNNs [8]. In 2017, the deep residual pan-sharpening neural network (DRPNN) [9] method was proposed using a deeper image super-resolution network to produce better results. In 2018, the use of two NN branches was introduced to extract more features from both PAN and MUL images [10, 11]. Dual NN paths approach was also proposed [12] consisting of a local and global NN. More recently, there has been a focus on the training data itself, such as in [13], where the authors attempted to find suitable training data relationships by using a dynamic blurring kernel and residual deep learning model. In [14], the authors explored the relationship between the PAN and MUL input images in their loss function without the use of label data. Also, Generative Adversarial Networks (GANs) have been used for satellite image fusion, but they require large amounts of training data, are computationally intensive, and can require significant processing power to train [15].

To date, there are three approaches for training the NN for PAN and MUL image fusion. The first method involves using Wald's protocol to make MUL images as ground truth images [5, 6], but this simple blur and interpolation process can lead to the loss of spatial information. The second method trains the NN on other images that do not have the same characteristics as remote sensing images [16]. The final method uses the relationship between the PAN and MUL images in the loss function without using labelling data [14].

3. Methodology

Typically, remote sensing image fusion methods utilize one of the following deep learning architectures: autoencoder (AE)-based architecture, CNN-based architecture, or GAN-based architecture. This paper particularly utilizes the second scheme and implements a CNN-based architecture for the image fusion task. A visual representation of the general workflow diagram is presented in Figure 1.

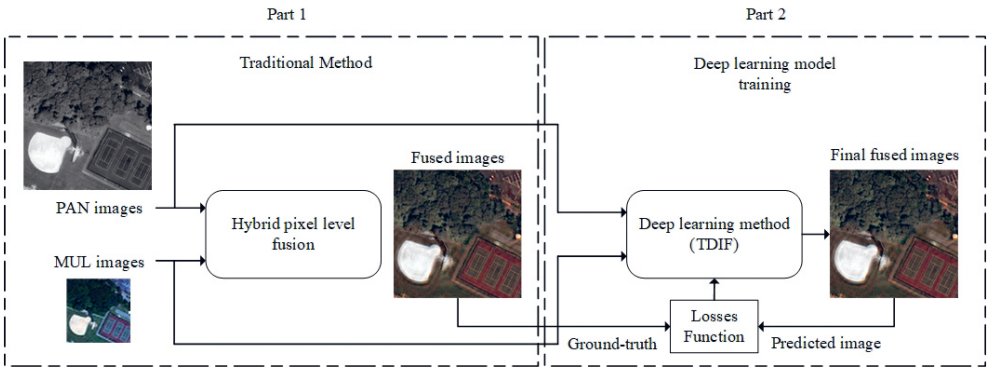

Fig. 1. General workflow diagram of the proposed TDIF method

The figure illustrates the overall structure of the network and how data flows through it, aiming to provide a clear and concise explanation of the proposed approach making it simple for others to follow and implement.

A new workflow is suggested to tackle the difficulty of obtaining ground-truth high-resolution multispectral images for deep learning-based remote sensing satellite image fusion. This workflow is comprised of two primary components. The initial stage involves acquiring ground-truth images through a hybrid traditional fusion approach that combines the Gram–Schmidt (GS) and curvelet transform techniques (CVT). This well-established fusion method producing precise and dependable fusion outcomes which are then utilized as the reference data for training the deep learning model. The proposed hybrid traditional fusion method overcomes the limitations of traditional fusion methods by combining two fusion techniques, to take advantage of both methods [17]. The resulting traditional fused images (TMUL) are evaluated and compared to numerous CS-based and MRA-based methods, as well as recently published methods, using various datasets.

The subsequent phase is dedicated to training the proposed deep learning model using the obtained ground-truth images. By capitalizing on the high-quality fusion results derived from the traditional method, the deep learning model can effectively learn from the valuable and information-rich dataset, leading to enhanced fusion performance. The proposed TDIF model is based on the enhanced super-resolution generative adversarial network (ESRGAN) generator [18]. The overall framework of the proposed method is summarized in Algorithm 1.

By incorporating the strengths of the traditional fusion method and deep learning, this workflow addresses the challenge of acquiring ground-truth high-resolution multispectral images. It utilizes the reliable fusion results from the traditional method to provide robust training data for the deep learning model, enabling it to learn and generalize from this information, ultimately enhancing fusion performance. The main layers of the model are illustrated in Figure 2.

The first convolutional layer takes the upsampled MUL image as input and performs a 3 × 3 kernel operation with a stride of 1 and padding of 1. The second convolutional layer takes the PAN image as input and performs the same operation.

A sequence of eight residual in residual dense blocks (RRDB) is applied to enhance its feature learning capabilities. Each of these blocks, consisting of three dense residual blocks and including five convolutional layers, each performing a 3 × 3 kernel operation with a stride of 1 and padding of 1. A non-linear activation function (LeakyReLU) is applied after each convolutional layer except for the last one to further refine the output and introduce non-linearity to the network. These dense residual blocks allow the network to capture both low-level and high-level information, making it more effective in preserving spatial and spectral details during the fusion process. Another convolutional layer is applied to the output of the residual blocks. It performs a 3 × 3 kernel operation with a stride of 1 and padding of 1.

Algorithm 1 The proposed TDIF method

Input: PAN images; MUL images; Target images $\hat{M}$; TMUL images M_T
Output: Fused image $\tilde{M}$
// Prepare the dataset for training

1. Downsampling PAN Images p and MUL images then interpolate MUL images.
2. Use [19] to create the reduced resolution MUL based on Wald's protocol.
3. Divide the dataset for training and validating and testing steps.
4. Get the trusted image by fusing PAN and MUL images using the traditional method M_T.

// Training and validating the network

5. Use $L1$ losses for the training dataset from Wald's protocol $L1_1$ and add $L1_2$ losses for the TMUL and MUL images:

$$L1_1 = 1/N \cdot \sum_{i=1}^{N} \left| \hat{M} - \tilde{M} \right|$$

$$L1_2 = 1/N \cdot \sum_{i=1}^{N} \left| M_T - \tilde{M} \right|$$

$$L = L1_1 + \alpha L1_2$$

// Testing the network

6. Use the test dataset to test the network.
7. Evaluate the resulted fused image.

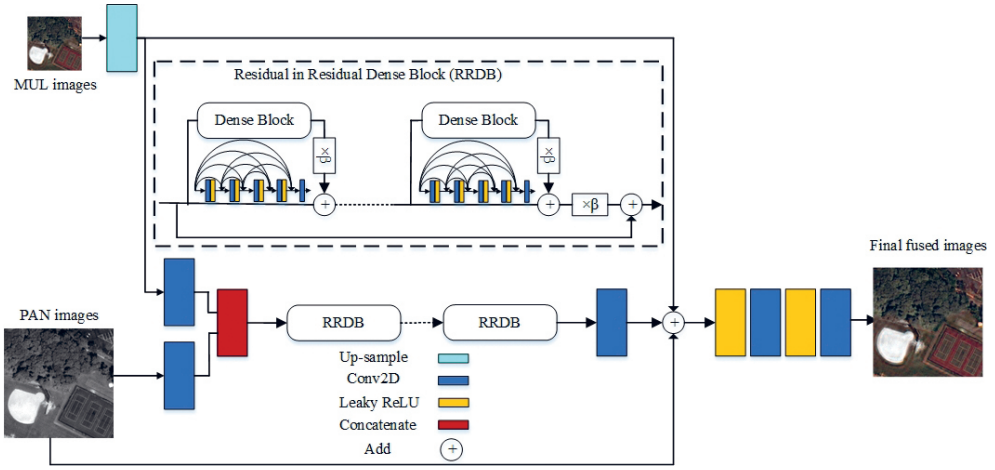

Fig. 2. Architecture of the proposed model

The output from the convolution layer is added to PAN, MUL images to maintain the spatial and spectral information from both PAN and MUL images then apply the result to the final stage. The final block consists of a leaky ReLU activation function followed by two more convolutional layers. The first convolutional layer has a kernel size of 3 × 3 and the second one has a kernel size of 1 × 1. The number of output channels in the last convolutional layer matches the number of channels in the MUL image. Additionally, all the convolutional layers are initialized using Kaiming normal initialization for weight initialization.

3.1. Network Training

The TDIF model is optimized using the Adam optimizer because of its advantageous features that enhance the accuracy and speed of the proposed models. By utilizing an adaptive learning rate and a momentum-based approach, the Adam optimizer facilitates faster learning and quicker convergence towards the optimal parameter values that minimize the cost or loss function. The TDIF model is optimized using the Adam optimizer with specific configuration settings. It starts with an initial learning rate of 0.0001 and does not incorporate weight decay. The momentum parameters, β1 and β2, are set to 0.9 and 0.999, respectively. To introduce non-linearity in the activation layers, a Leaky ReLU activation function is applied with a slope of 0.2. This activation function allows for the propagation of small negative values, preventing the complete saturation of neurons. The weights of the model are initialized using the Kaiming initialization method. This method takes into account the specific activation function and aims to initialize the weights in a way that prevents the signal from vanishing or exploding during forward and backward propagation. By incorporating these settings, the TDIF model optimizes its performance and ensures effective learning and convergence during the training process.

The model undergoes training for a total of 100 epochs. Throughout the training process, the performance of the model is evaluated using two metrics: peak signal to noise ratio (PSNR) and losses. In each iteration or epoch, the model's output is compared to the ground truth or target data using the PSNR metric. PSNR measures the quality of the model's output by quantifying the ratio of the peak signal power to the noise power. Higher PSNR values indicate better image quality. Additionally, the losses incurred during each iteration are calculated to quantify the discrepancy between the model's predicted output and the ground truth as outlined in Algorithm 2.

The training process involves minimizing the L_1 loss, which is defined as the sum of all the absolute differences between the ground truth value and the predicted value $\tilde{M}$. The overall loss is calculated by taking into account two ground truth values with different priorities. The training data is represented by a set of pairs of down-sampled MUL and PAN images, with the MUL image $\hat{M}$ being the label. The training data is represented as $\left[p^{(i)}, M^{(i)}, \hat{M}^{(i)}, M_T^{(i)} \right]_{i=1}^{N}$ for $i = 1$ to N, where N is the number of training samples in each iteration.

Algorithm 2 The proposed training algorithm

Require: train data set, batch size, epoch number, number of channel output of conv2D layers, number of RRDB, and number of dense block, define the optimizer, β1, β2, learning rate, learning rate steps, define the losses function.
Initialize model weights
Split the dataset into training and validation sets
For epoch = 1:100 ***do***
    For each batch size = 4 ***do***
        Perform a forward pass on the current batch of data to get the predicted output
        Compute L_1 losses
        Compute gradient
        Update weights
        Calculate PSNR for each epoch (training/validation) dataset
        Save lowest losses value
        If the model is stable then add L_1 losses between MUL and TMUL
        Calculate PSNR for each epoch (training/validation) dataset
        Save lowest losses value
    End For
End For

4. Dataset, Results, and Analysis

The use of Wald's protocol for creating MUL images as ground truth is a common approach in the field of image fusion for remote sensing images. This method is based on the idea of using a simple blur and interpolation process to create a low-resolution image from a high-resolution image. The NN is then trained to produce a high-resolution image from the low-resolution image and the corresponding high-resolution PAN image. Wald's protocol has been widely used in the remote sensing community due to its simplicity and ease of implementation. However, as already mentioned, this method can lead to the loss of important spatial information, and it may not always produce the best results. Nevertheless, it remains a widely used approach and has been applied in various studies and applications.

The model was implemented using the PyTorch framework and trained on an Intel(R) Xeon(R) W-2125 CPU @ 4.00GHz with a NVIDIA Quadro P500. The dataset was divided into a training set (80%), validation set (10%), and testing set (10%).

4.1. Dataset

In this study, verified pansharpened images are used as additional ground truth images in the training process of the deep learning model. These images are

produced through a well-established pansharpening method (GS-CVT based) and are used to further improve the results of the model. The preparation of the training dataset is depicted in Figure 3.

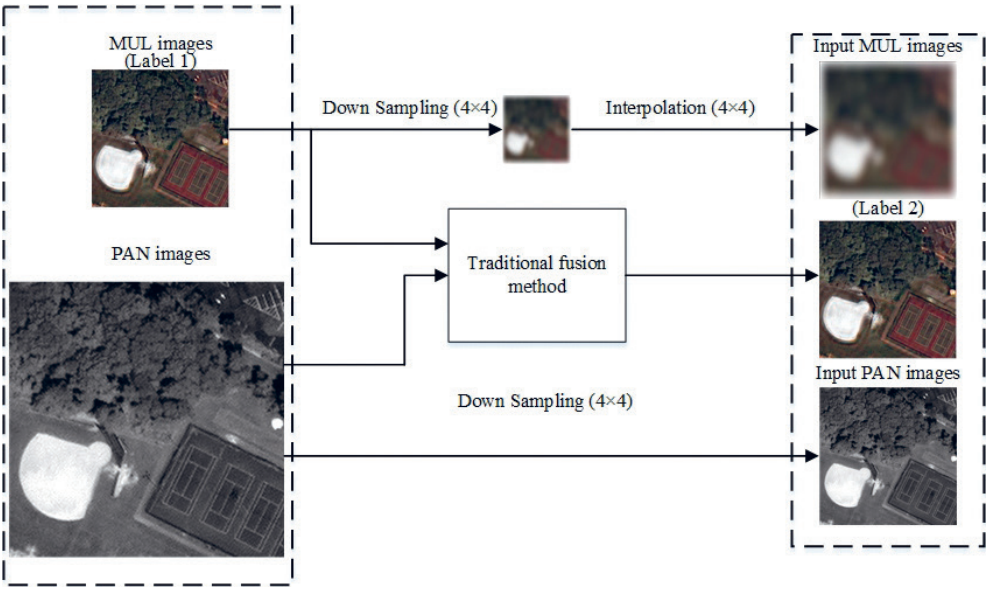

Fig. 3. Dataset preparation

The study uses two different satellite images, WorldView2 and GeoEye1, for its datasets. These images capture different types of regions including both man-made and natural areas. The images are divided into patches for training and testing purposes. For WorldView2 dataset, the PAN images are 512 × 512 in size and the MUL images are 128 × 128. For GeoEye1 dataset, the PAN images are 128 × 128 in size while the MUL images are 32 × 32. These differences in scales allow the model to be tested with varying data sizes. The results of the traditional method are produced using the same dataset as the one used to train and test the model. The satellite datasets details can be seen in Table 1.

Table 1. Satellites specifications

Satellite	Resolution [m]		Date	City	Downloaded sites
	PAN	MUL			
WorldView2	0.4	1.6	26.09.2016	Washington DC	https://resources.maxar.com/product-samples
GeoEye1	0.41	1.64	4.12.2020	Vientiane	

The proposed ground truth dataset is produced using a combination of GS and CVT methods, which are based on local energy and maximum fusion rules [17]. This combination of methods aims to reduce the limitations of the individual method and improve the preservation of both spectral and spatial information in the output. It is important to note that the traditional method used can be changed based on the specific design environment, requirements, or dataset being used.

The effectiveness of the proposed traditional method is evaluated using 3600 pairs of WorldView2 and 14,400 pairs of GeoEye1 satellite images satellite images with varying patch sizes. Both qualitative and quantitative evaluations are performed using seven image quality evaluation metrics: peak signal to noise ratio (PSNR), quality with no reference (QNR) index, spectral correlation coefficient (SCC), spectral angular mapper (SAM), structural similarity index measure (SSIM), error relative global dimensionless synthesis (ERGAS), quality index (Q_{index}). The results are shown in Figures 4, 5 and Tables 2, 3. The results obtained from the comparison of the proposed traditional method with eight other traditional methods demonstrate the superiority and robustness of the proposed approach. The performance of the proposed method outperforms the other traditional methods in various evaluation metrics. These findings highlight the significance and reliability of the proposed traditional method as a preferred choice for the specific application or problem domain.

In Figure 4, a sample of the quantitative evaluation of WorldView2 images are displayed using eight different traditional fusion methods (GS-CVT, CVT, Brovey, GS, intensity hue saturation (IHS), principal component analysis (PCA) and Ehler transform (EL)). The GS-CVT fusion method resulted in the best image in the qualitative evaluation.

The other methods either had high spectral or spatial distortion. For example, in the GS-based method, the fused image had a red hue, while the DWT and CVT fusion methods have clear details but poor coloration.

Table 2 shows the evaluation of traditional fusion methods based on different seven evaluation metrics reveals varying levels of performance over WorldView2 images. Brovey, IHS, PCA, EL, CVT, and DWT exhibit lower performance in metrics such as SSIM, SCC, QNR, Q_{index}, PSNR, ERGAS, and SAM compared to GS-CVT. GS shows comparable performance to GS-CVT in most metrics, indicating its effectiveness as a fusion method. GS-CVT serves as a reliable baseline, demonstrating its superior performance compared to other traditional fusion methods. These findings emphasize the significance of GS-CVT as a benchmark method and highlight its potential for achieving improved fusion results.

In Figure 5, a sample of the quantitative evaluation of GeoEye1 satellite image fusion results using eight different traditional fusion methods is displayed. The output of the GS-CVT method is closest to the MUL images, which are used as the ground truth images.

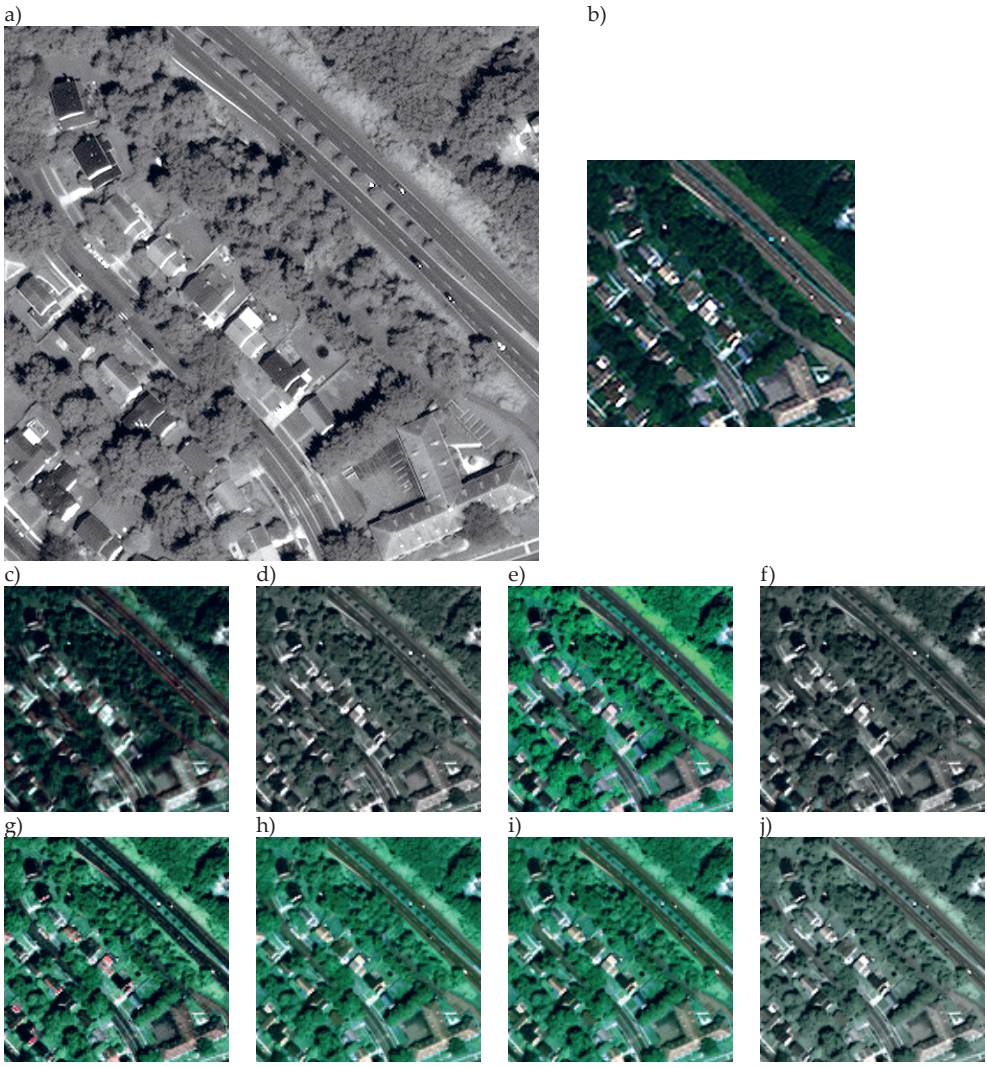

Fig. 4. Sample of WorldView2 images by different pansharpening algorithms:
a) PAN; b) MUL; c) GS-CVT; d) CVT; e) DWT; f) Brovey; g) GS; h) IHS; i) PCA; j) EL

Table 2. Objective evaluation of traditional fusion results for WorldView2 images

	Brovey	IHS	GS	PCA	EL	CVT	DWT	GS-CVT
SSIM↑	0.8421	0.9012	0.8773	0.9017	0.9099	0.923	0.857	0.919
SCC↑	0.8035	0.8099	0.7731	0.8101	0.7949	0.859	0.851	0.862
QNR↑	0.8876	0.6773	0.6944	0.6758	0.6387	0.640	0.694	0.887
Q_{index}↑	0.6998	0.7038	0.6660	0.7037	0.6598	0.746	0.774	0.735
PSNR↑	25.897	23.590	24.278	23.543	20.024	24.09	26.48	31.49
ERGAS↓	64.884	67.444	73.658	67.392	88.319	66.70	68.54	63.82
SAM↓	12.659	14.868	15.048	14.871	15.955	15.11	15.62	13.52

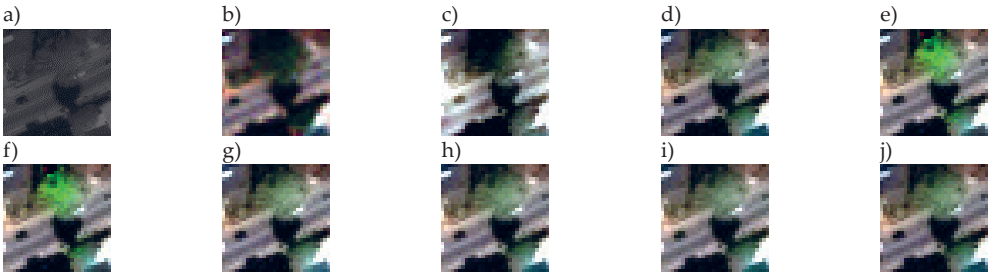

Fig. 5. Sample of GeoEye1 images by different pansharpening algorithms:
a) PAN; b) MUL; c) GS-CVT; d) CVT; e) DWT; f) Brovey; g) GS; h) IHS; i) PCA; j) EL

In Table 3, the performance of the fusion methods is evaluated over a different dataset (GeoEye1 images). It shows that the GS-CVT method has the highest QNR, SCC and PSNR values and DWT method has highest Q_{index} compared to the other methods. However, the ERGAS value is the best for the GS-CVT method compared to the other methods.

Table 3. Objective evaluation of traditional fusion results for GeoEye1 images

	Brovey	IHS	GS	PCA	EL	CVT	DWT	GS-CVT
SSIM↑	0.838	0.9146	0.870	0.9133	0.9142	0.768	0.8441	0.8528
SCC↑	0.831	0.8364	0.827	0.8356	0.8352	0.813	0.8673	0.8997
QNR↑	0.813	0.8259	0.852	0.8250	0.8136	0.821	0.8363	0.8414
Q_{index}↑	0.675	0.6854	0.685	0.6863	0.6923	0.585	0.6965	0.6879
PSNR↑	27.788	25.010	26.25	24.945	23.860	28.67	27.128	32.060
ERGAS↓	68.363	66.132	66.00	65.918	65.703	78.57	60.408	60.249
SAM↓	10.663	12.293	12.52	12.286	13.335	12.43	13.558	11.979

It is important to note that different metrics may prioritize different aspects of image quality, such as sharpness, colour preservation, or noise reduction. For example, SSIM values closer to one indicate a higher level of similarity between the fused image and the reference image, while PSNR values closer to infinity indicate a lower level of noise in the image. On the other hand, the ERGAS metric measures the quality of the detail representation and how well the global and local structures are preserved in the fused image.

Overall, the performance of these traditional fusion methods varies in different metrics compared to GS-CVT. While some methods show comparable or slightly better performance in certain metrics, others may have trade-offs in different aspects. The selection of the fusion method should be based on the specific requirements and priorities of the application.

4.2. Results

The evaluation was performed on the fusion result quantitatively and qualitatively over the eight different image fusion methods (GS-CVT, PCA, CVT, GS, IHS, Brovey, Ehler, DWT) and two deep learning methods that trained from scratch on

the same datasets (PNN and PanNet) and finally the proposed method without using traditional results and with using traditional results. The same seven performance metrics that were used for dataset evaluation were also used for comparing the results (SSIM, SCC, QNR, Q_{index}, PSNR, ERGAS and SAM).

The results from Figures 6 and 7 indicate that the proposed method provides the best outcome in terms of both spatial and spectral information preservation. The CVT and DWT methods do not effectively preserve the spatial details, while other methods preserve the spectral details but not the spatial details. However, the proposed method is successful in preserving both aspects of the information. Additionally, the use of traditional results as additional ground truth improves the results of the proposed method, as demonstrated by the figures.

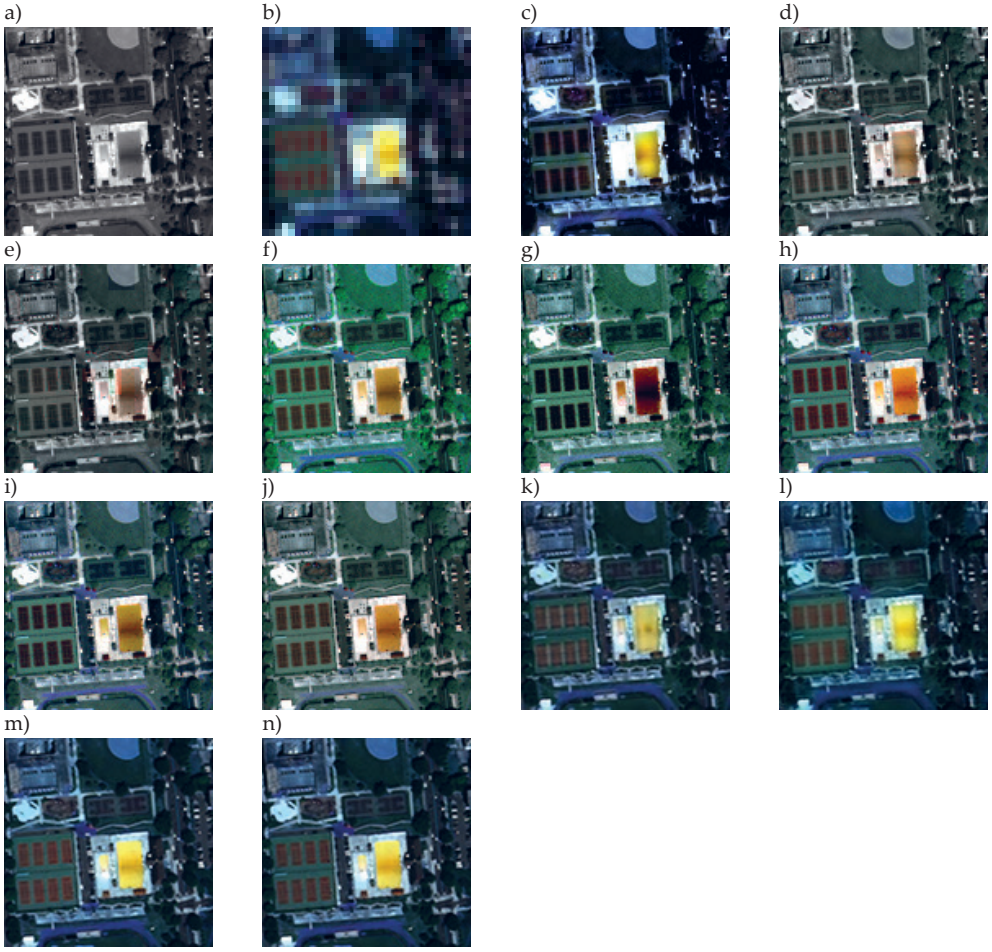

Fig. 6. Sample of WorldView2 images by different pansharpening algorithms:
a) PAN; b) MUL; c) GS-CVT; d) CVT; e) DWT; f) Brovey; g) GS; h) IHS; i) PCA; j) EL;
k) PNN; l) PanNet; m) TDIFW; n) TDIF

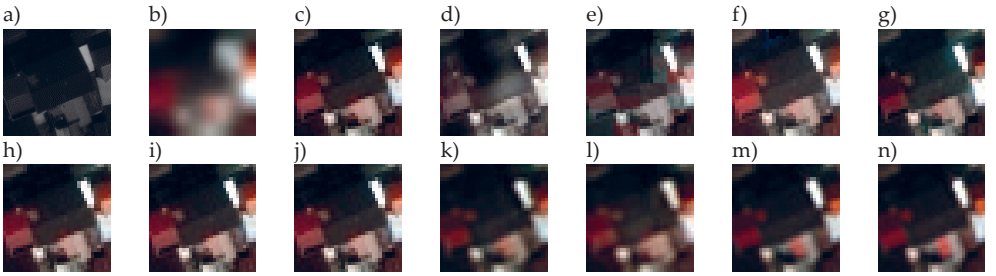

Fig. 7. Sample of GeoEye1 images by different pansharpening algorithms: a) PAN; b) MUL; c) GS-CVT; d) CVT; e) DWT; f) Brovey; g) GS; h) IHS; i) PCA; j) EL; k) PNN; l) PanNet; m) the proposed method without using traditional results (TDIFW); n) the proposed method (TDIF)

4.3. Analysis

As the subjective evaluation depends on the vision of the interpreter, there is a need for objective analysis. The evaluation of different fusion methods for WorldView2 and GeoEye1 images by seven metrics reveals that TDIF demonstrates competitive performance compared to the other methods. While some methods show improvements or slight decreases in performance compared to TDIF, others exhibit more significant drops. TDIFW stands out as a method that shows comparable performance to TDIF, with only slight decreases in various evaluation metrics. Overall, the proposed fusion method (TDIF) proves to be effective in improving the performance of the fusion process, outperforming several other methods in terms of evaluation metrics.

According to the results from Figures 8 and 9, TDIFW exhibits slightly lower similarity, correlation, and quality scores (SSIM, SCC, and QNR) compared to TDIF. PNN demonstrates significantly lower structural similarity (SSIM) and lower quality in terms of noise reduction (QNR) than TDIF. Similarly, PANNET shows lower similarity (SSIM) and correlation (SCC) scores compared to TDIF. Brovey, as well as IHS, GS, PCA, CVT, DWT, and GS-CVT, exhibit lower similarity, correlation, and quality scores compared to TDIF, with varying degrees. Overall, TDIF performs better than TDIFW, PNN, and PANNET in terms of image fusion quality. The percentage differences in evaluation metrics range from −29.2% to −2.9% for WorldView2 images and from 29.5% to 99.8% for GeoEye1 images. However, TDIFW demonstrates comparable performance to TDIF as shown in Table 4, with slight decreases ranging from 1.6% to 6.9% for WorldView2 images and from 0.2% to 7.9% for GeoEye1 images across different evaluation metrics.

Table 4. Objective evaluation TDIF and TDIFW

Satellite image	Method	SSIM↑	SCC↑	QNR↑	Q_{index}↑	PSNR↑	ERGAS↓	SAM↓
WorldView2 images	TDIF	0.92	0.97	0.93	0.94	31.52	22.2	7.58
	TDIFW	0.91	0.96	0.91	0.92	30.47	23.2	8.11
GeoEye1 images	TDIF	0.92	0.98	0.97	0.96	32.10	13.87	6.41
	TDIFW	0.91	0.98	0.96	0.96	29.94	14.99	6.4

Ranking Tables 2, 3 and the values of Figures 8, 9, so that all the metrics values of the methods are being compared are in order. When these methods are numbered from 1 to 12, with the best metric value taking the highest number and the worst metric value taking the smallest, as shown in Figure 10, the proposed fusion method is the best that can improve the performance of the fusion process because it has the highest value.

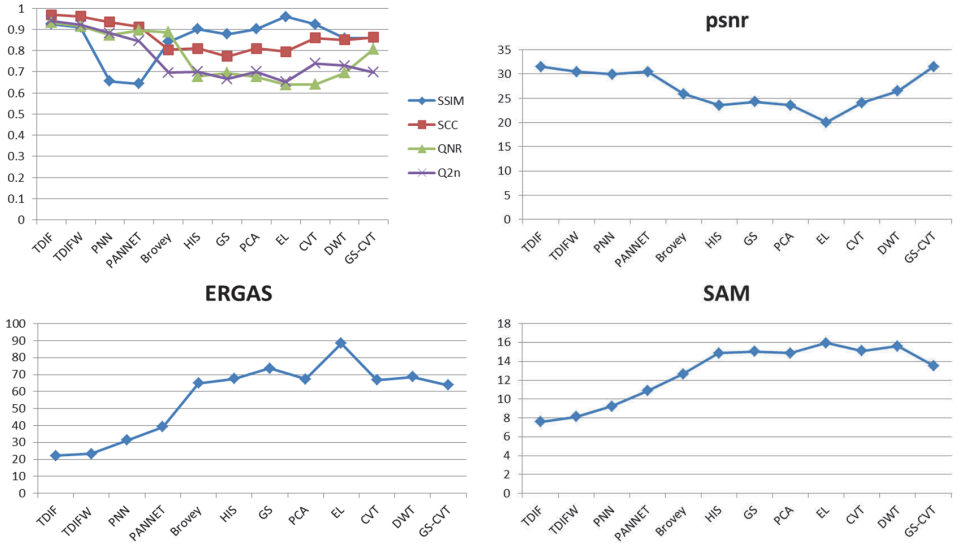

Fig. 8. Objective evaluation of fusion results for WorldView2 images

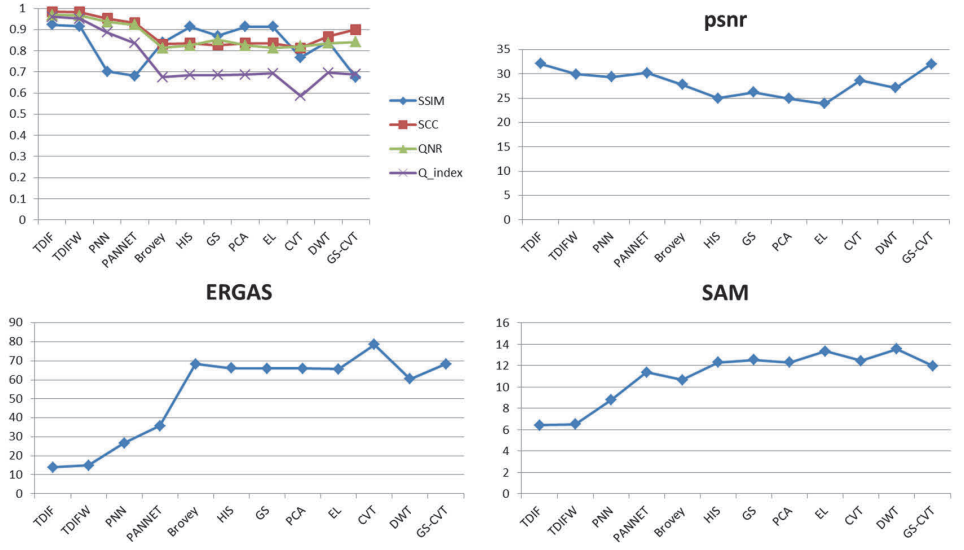

Fig. 9. Objective evaluation of fusion results for GeoEye1 images

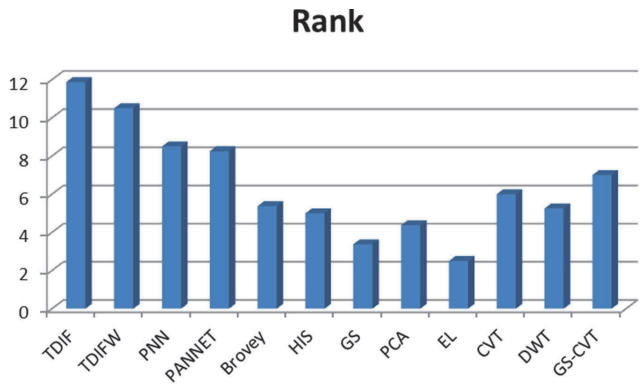

Fig. 10. Ranking of fusion methods

5. Conclusion and Future Work

The main challenge addressed in this study is the training phase as it requires significant processing power and memory resources. Furthermore, designing lightweight architectures and optimizing the inference process are important considerations. Two contributions are introduced within the scope of the proposed approach. The first proposes a deep learning model for the fusion of PAN and MUL images before testing it to ensure its superiority. Second, it proposes a new perspective for training the deep learning network. The proposed (TDIF) approach is based on the results obtained from pre-designed and well-tested hybrid traditional methods providing a novel perspective on solving the image fusion problem compared to existing deep learning methods.

By comparing the performance of different fusion methods to the proposed TDIF, significant variations in the evaluation metrics can be observed. For WorldView2 images, the percentage differences range from −29.2% to −2.9%, indicating lower performance compared to TDIF. On the other hand, for GeoEye1 images, the percentage differences range from 29.5% to 99.8%, highlighting even greater disparities in performance. These variations emphasize the impact of different fusion methods on image quality and the need to carefully consider the choice of method based on the specific dataset or application. Overall, the proposed TDIF approach demonstrates competitive performance compared to the other fusion methods, with some methods showing improvements or slight decreases, while others exhibit more significant performance drops.

For future work, the TDIF approach can be further refined and optimized by exploring different architectures and techniques within the deep learning model. This may involve investigating the use of different types of layers, loss functions, or incorporating attention mechanisms to enhance the model performance. The TDIF could also be used as a generator, adding a discriminator to form GAN architectures

and testing the results. Additionally, the TDIF approach can be applied and evaluated in a wider range of remote sensing applications, such as agriculture, environmental monitoring, and change detection. This will allow for a better assessment of its effectiveness in various real-world scenarios and provide insights into its potential practical applications.

References

[1] Xiao G., Bavirisetti D. P., Liu G., Zhang X.: *Image Fusion*. Springer, Singapore 2020.

[2] Kaur H., Koundal D., Kadyan V.: *Image fusion techniques: A survey*. Archives of Computational Methods in Engineering, vol. 28, 2021, pp. 4425–4447. https://doi.org/10.1007/s11831-021-09540-7.

[3] Tsagkatakis G., Aidini A., Fotiadou K., Giannopoulos M., Pentari A., Tsakalides P.: *Survey of deep-learning approaches for remote sensing observation enhancement*. Sensors, vol. 19(18), 2019, 3929. https://doi.org/10.3390/s19183929.

[4] Dong C., Loy C. C., He K., Tang X.: *Image super-resolution using deep convolutional networks*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38(2), 2016, pp. 295–307. https://doi.org/10.1109/TPAMI.2015.2439281.

[5] Masi G., Cozzolino D., Verdoliva L., Scarpa G.: *Pansharpening by convolutional neural networks*. Remote Sensing, vol. 8(7), 2016, 594. https://doi.org/10.3390/rs8070594.

[6] Zhong J., Yang B., Huang G., Zhong F., Chen Z.: *Remote sensing image fusion with convolutional neural network*. Sensing and Imaging, vol. 17, 2016, 10. https://doi.org/10.1007/s11220-016-0135-6.

[7] Yang J., Fu X., Hu Y., Huang Y., Ding X., Paisley J.: *PanNet: A Deep Network Architecture for Pan-Sharpening*. [in:] *2017 IEEE International Conference on Computer Vision ICCV 2017: Proceedings: 22–29 October 2017, Venice, Italy 2017*, IEEE, Piscataway 2017, pp. 1753–1761. https://doi.org/10.1109/ICCV.2017.193.

[8] Nguyen H.V., Ulfarsson M.O., Sveinsson J.R., Mura M.D.: *Deep SURE for unsupervised remote sensing image fusion*. IEEE Transactions on Geoscience and Remote Sensing, vol. 60, 2022, pp. 1–13. https://doi.org/10.1109/TGRS.2022.3215902.

[9] Wei Y., Yuan Q.: *Deep residual learning for remote sensed imagery pansharpening*. [in:] *RSIP 2017: International Workshop on Remote Sensing with Intelligent Processing: Proceedings: May 19–21, Shanghai, China*, IEEE, Piscataway 2017, pp. 1–4. https://doi.org/10.1109/RSIP.2017.7958794.

[10] Shao Z., Cai J.: *Remote sensing image fusion with deep convolutional neural network*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11(5), 2018, pp. 1656–1669. https://doi.org/10.1109/JSTARS.2018.2805923.

[11] Liu X., Liu Q., Wang Y.: *Remote sensing image fusion based on two-stream fusion network*. Information Fusion, vol. 55, 2020, pp. 1–15. https://doi.org/10.1016/j.inffus.2019.07.010.

[12] Wang J., Shao Z., Huang X., Lu T., Zhang R.: *A dual-path fusion network for pan-sharpening*. IEEE Transactions on Geoscience and Remote Sensing, vol. 60, 2022, pp. 1–14. https://doi.org/10.1109/TGRS.2021.3090585.

[13] Guo A., Dian R., Li S.: *Unsupervised blur kernel learning for pansharpening*. [in:] *IGARSS 2020 – 2020 IEEE: International Geoscience and Remote Sensing Symposium: International Geoscience and Remote Sensing Symposium: September 26 – October 2, 2020: Virtual Symposium*, IEEE, Piscataway 2020, pp. 633–636. https://doi.org/10.1109/IGARSS39084.2020.9324543.

[14] Luo S., Zhou S., Feng Y., Xie J.: *Pansharpening via unsupervised convolutional neural networks*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, 2020, pp. 4295–4310. https://doi.org/10.1109/JSTARS.2020.3008047.

[15] Benzenati T., Kessentini Y., Kallel A.: *Pansharpening approach via two-stream detail injection based on relativistic generative adversarial networks*. Expert Systems with Applications, vol. 188, 2022, 115996. https://doi.org/10.1016/j.eswa.2021.115996.

[16] Ye F., Li X., Zhang X.: *FusionCNN: a remote sensing image fusion algorithm based on deep convolutional neural networks*. Multimedia Tools and Applications, vol. 78, 2019, pp. 14683–14703. https://doi.org/10.1007/s11042-018-6850-3.

[17] Hammad M., Ghoniemy T., Mahmoud T., Amein A.: *Hybrid fusion using Gram Schmidt and Curvelet transforms for satellite images*. IOP Conference Series: Materials Science and Engineering, vol. 1172, 2021, 012016. https://doi.org/10.1088/1757-899X/1172/1/012016.

[18] Wang X., Yu K., Wu S., Gu J., Liu Y., Dong C., Qiao Y. et al.: *Esrgan: Enhanced super-resolution generative adversarial networks*. [in:] Leal-Taixé L., Roth S. (eds.), *Computer Vision – ECCV 2018 Workshops: Munich, Germany, September 8–14, 2018: Proceedings, Part V*, Lecture Notes in Computer Science, vol. 11133, Springer, Cham 2019, pp. 63–79. https://doi.org/10.1007/978-3-030-11021-5_5.

[19] Vivone G., Alparone L., Chanussot J., Dalla Mura M., Garzelli A., Licciardi G.A., Restaino R. et al.: *A critical comparison among pansharpening algorithms*. IEEE Transactions on Geoscience and Remote Sensing, vol. 53(5), 2014, pp. 2565–2586. https://doi.org/10.1109/TGRS.2014.2361734.