National Journal of Signal and Image Processing, ISSN: 3107-5231 Vol. 1, No. 2, 2025 (pp. 66 -72)

RESEARCH ARTICLE

ECEJOURNALS.IN

Fusion of Multispectral and Panchromatic Images for Enhanced Remote Sensing Resolution

Kasil Teyene[1]*, M.P. Bates[2]

[1]*Electrical and Computer Engineering Addis Ababa University Addis Ababa, Ethiopia*

[2]*Dept. of Electrical Engineering, University of Zimbabwe, Harare, Zimbabwe*

KEYWORDS:
Remote Sensing,
Pansharpening,
Multispectral Image,
Panchromatic Image,
Image Fusion,
Spatial Resolution Enhancement,
Spectral Preservation,
Wavelet Transform,
Principal Component Analysis (PCA),
Vision Transformer,
Deep Learning,
CNN-Based Fusion,
Quality Metrics,
SAM,
ERGAS,
Q-index

ARTICLE HISTORY:
Submitted : 16.01.2025
Revised : 15.02.2025
Accepted : 19.04.2025

https://doi.org/10.17051/NJSIP/01.02.09

Abstract

This combination of multispectral (MS) and panchromatic (PAN) data commonly known as pansharpening has become a ubiquitous operation in all remotely determined research and development and it enables the creation of images that produce the saturation of the multispectral datasets combined with the fine-resolution capacity of the panchromatic data. This paper gives an extensive comparative study as well as practical use of new high-quality pansharpening methods, with a view to overcome the current insatiable needs in high-quality remote sensing information as well as rigorously reserving the spectral fidelity. Such state-of-the-art methods as convolutional neural networks (CNNs), Vision Transformer and deep learning-based methods which are Component Substitution (CS) and Multiresolution Analysis (MRA) are also included. Techniques are then tested on quantifiable scales with firm measures like Spectral Angle Mapper (SAM), Correlation Coefficient (CC), Error Relative Globules Adimensionnelle de Synthaje (ERGAS) and Quality Index (Q-index) so that there is the multi-faceted analysis of respective spectral and spatial work. Experimental validations were conducted with high-resolution WorldView-3 satellite images of large variety of types of land cover and scene complexities. The results reveal that deep-learning-based-fusion techniques, mainly the Transformer-architectures-based, have achieved significant bettering over conventional approaches in terms of providing enhanced spatial sharpness and superior spectral preservation with enhancements in all forms of assessment measures. The results illustrate the extent of the Transformer models in learning the complex cross-modality relationships and present the possibility of usage in real-time, large-scale, application-specific remote sensing applications like land cover classification, urban mapping, and precision agriculture. Not only has this work created a critical benchmarking of both present and developing fusion methods but it has also suggested practical lessons on possible future research on unsupervised fusion, model generalizability, and finally edge deployment in the service of in-situ applications of remote sensing.

Author's e-mail: teyene.kasil@aait.edu.et, bates.mles.pn@gmail.com

How to cite this article: Teyene K, Bates M P. Fusion of Multispectral and Panchromatic Images for Enhanced Remote Sensing Resolution. National Journal of Signal and Image Processing, Vol. 1, No. 2, 2025 (pp. 66-72).

Introduction

Remote sensing technology has climbed so high as to be incorporated into so many uses and branches, such as precision agriculture, monitoring the environment, urban planning, disaster mapping, and military surveillance among others. At the heart of these applications lies data collection and analysis of high-resolution satellite images which may be used to show minute features on the surface of Earth and the atmosphere. The performance of the remote sensing data itself is however limited by already restrictive performance of the imaging sensors, which usually entails compromise between those in the spatial and spectral resolutions. High resolution panchromatic (PAN) images that measure broad-spectrum light as a single grayscale band give an outstanding spatial resolution but do not allow distinguishing between various materials on the basis of their spectral properties. Multispectral (MS) images are in contrast heavily spectrally rich in a number of distinct bands, e.g., visible to near-infrared electromagnetic wavelength, at coarser spatial resolutions.

The limitations of this so-called spatial-spectral trade-off that is restricted by the sensor design and constraints posed due to the on-board hardware, complicates things in applications that require both the high spatial resolution as well as the true spectral information.

An example is land cover classification and the vegetation health assessment that would not only need the spectral richness that MS images have but also the fine spatial structure that PAN images can offer. Combination of both MS and PAN imagery, often termed pansharpening, has since become one of the most hot-researched topics in the field of remote sensing with the hope, to achieve a realization of the so-called best-of-both-words.

In the last decades, a wide variety of pansharpening techniques have been proposed, including classical mathematical transforms (Intensity-Hue-Saturation (IHS), Principal Component Analysis (PCA), and Bovey Transform), more complex in multiresolution analysis methods, such as the ones based in wavelets and Laplacian pyramids. Such methods have proven very popular because of their ease of concept whereas they are very computationally efficient. They have however the tendency to produce undesirable spectral distortion particularly in scenes where the content is complex and heterogeneous. Recent breakthrough in deep learning, most notably convolutional neural networks (CNNs) and, more recently, transformers-based networks have been able to provide significant potential in modelling the complex non-linear dependence between MS and PAN images to produce pansharpened image with high spatial fidelity and sound spectral integrity Figure 1.

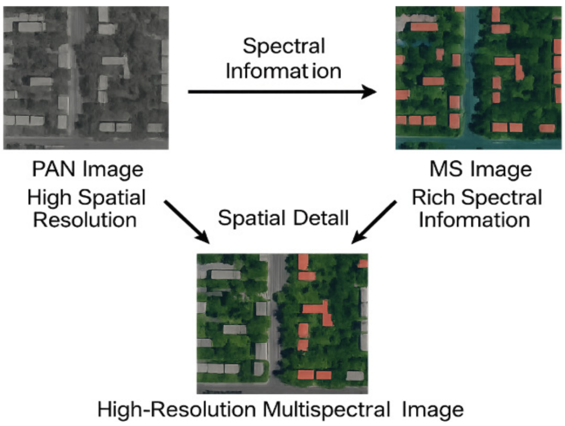

Fig. 1: Spatial and Spectral Trade-Off in PAN and MS Remote Sensing Images

Even with substantial advancement, the area still requires addressing issues pertinent to spectral regularity, maintenance of smaller spatial features, applicability of the model to distinct types of sensors and scenes, and performance with regard to real-time or on-board implementation. In the current paper, the author attempted to overcome these issues by carrying out an extensive survey of state-of-the-art methods in pansharpening both classical and deep learning-based models. Among the most important contributions, there is the strict comparison of these methods based on quantitative quality indicators, the thorough discussion of the advantages and weaknesses of the methods, and the emergence of a new framework of transformer-loaded fusion architecture distinctive to high levels of remote sensing.

Overall, this paper has not only contributed to the body of published knowledge regarding MS-PAN image fusion but it has actually been able to render practical guidance in this regard of selecting as well as applying fusion methods in actual remote sensing processes. The products will have the potential to advance the interests of many diverse scientific and operational applications users who require the improvement of the spatial and spectral accuracy of the satellite images to make decisions of utmost importance.

Related Work

The combination of multispectral (MS) and panchromatic (PAN) images in Pansharpening has undergone tremendous methodological developments, covering the foundation of classical mathematical models to the latest paradigms of the learning style. In the earlier methods, simple transformations with efficient computation efforts were of concern. One of the earliest methods that combined spatial and spectral information was the Intensity-Hue-Saturation (IHS) method, Principal Component Analysis (PCA), and Brovey Transform.[1, 2] These approaches gave more spatial resolution but could introduce spectral distortion, in particular, in heterogeneous or urban areas.[3]

In order to overcome such shortcomings, multiresolution analysis (MRA) techniques like Laplacian pyramid and wavelet transform intrafractals were developed. MRA methods replace spectrum of MS bands by injecting high-frequency spatial information of the PAN into the MS bands on a range of scales hence providing better spectral fidelity but these methods are computationally expensive and are prone to artifacts in complicated scenes.[4, 5] A model based solution to fusion, sparse representation techniques and adaptive Bayesian inference techniques have been proposed to fuse, as a model based approach uses image statistics to generate spatially and spectrally more accurate images.[6, 7]

The remote sensing area has been revolutionized in the past ten years by using deep learning. Convolutional Neural Networks (CNNs) have proven to learn complex nonlinear relationships between low-resolution MS and high-resolution PAN data, and have been shown to be extremely successful in eliminating the spectral and spatial artifacts.[8, 9] Other implementations of Deep learning models as FPGA-based accelerators have also been evaluated in recent studies regarding reconfigurable computing to aid in efficient pansharpening and other remote sensing as a ground category.[10] Besides, transformer-based architectures have been demonstrated to further increase pansharpening performance by learning long-range dependence relationships among image data,[11] and Generative Adversarial Networks (GANs) further explore the envelope of perceptual quality.[12]

Moreover, data science and predictive analytics have been quickly picked up in related engineering fields which have been enhancing the resilience of the models and their applicability,[13] and other antenna and hardware design innovations, like the combinations used in 5G MIMO and GNSS signal processing, demonstrate the significance of validation based on spatial dimension and real-time processing that can be developed into the context of remote sensing.[14, 15] The creation of frequency-attentive CNN to detect environmental sounds also indicates potential variability and cross-dimensional applicability of extended deep learning models.[16]

Nevertheless, despite the recent developments, the fundamental problems are still the same: spectral consistency, the preservation of sharp spatial features, and high computational efficiency to implement in the field. Future work that still tries to satisfy these needs is on new architectures, domain adaptation, and edge-based processing.

METHODOLOGY

Dataset Description

Source: Satellite Imagery WorldView-3

The experiments and assessments in this paper employ the use of imagery obtained by a modern state of the commercial Earth-observation satellite known as WorldView-3 that was created by Maxar Technologies. WorldView-3 is well known as the satellite capable of simultaneous high-resolution panchromatic (PAN) and multispectral (MS) imagery of the same geographic area, which is why it is one of the popular sources of remote sensing fusion studies. The high spatial and generous spectral coverage of the satellite enables a wide variety of applications such as urban mapping and agricultural monitoring, and environmental analysis.

Spatial Resolution

WorldView-3 offers various spatial resolutions of images with specific modalities of the sensor. The panchromatic sensor provides images with remarkable ground sample distance (GSD) of 0.31 meters, which means that it is possible to see finer details of surfaces and patterns. The multispectral sensor in contrast has lower GSD of 1.24 meters. Such intentional design compromise enables one to gain a spectrally rich information in wider swaths but ensuring high spatial acuity in the PAN channel as well. Insofar as pansharpening is concerned, the PAN image would carry the spatial information whilst the MS image would carry the spectral richness to be maintained in the product of the fusion.

Spectral bands:

The multispectral payload on the WorldView-3 has a series of eight different spectral bands, which would include coastal blue to near infra-red (NIR). These are: coastal blue (400450nm), blue (450510nm), green (510580nm), yellow (585625nm), red (630690nm), red edge (705745nm), and near-infrared 1 (770895nm), and near-infrared 2 (8601040nm). This mass spectral coverage is very important in the precise separation of the surface materials, vegetation health, and presence of water bodies amongst other advanced remote sensing assignments. In contrast the panchromatic channel sums the reflected radiance over a wide range of the spectrum (450800 nm) and hence the single wavelength grayscale image has a very high spatial resolution Figure 2.

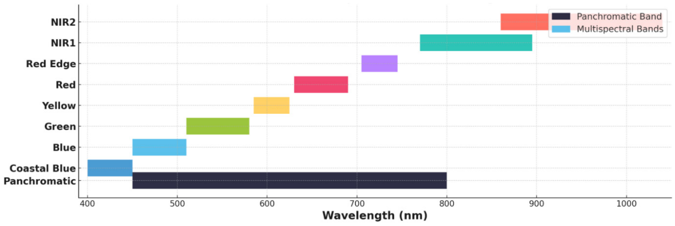

Fig. 2: Spectral Band Coverage of WorldView-3 Panchromatic and Multispectral Sensors

Classical Fusion Techniques

IHS Transform Based Fusion

One of the most popular methods of signal substitution (component substitution- CS) in the scope of pan-sharpening is the Intensity-Hue-Saturation (IHS) transformation, which are usually used to process RGB or three-band multispectral pictures. It starts with the transformation of the actual MS image in the RGB color space into the IHS color space and virtually de-mplexes the intensity (spatial detail) part, the hue and the saturation (spectral content) part of the image. The intensity channel of the IHS representation is then replaced with the high-resolution PAN image that includes better spatial information. This replacement is done so that the spatial detail available in the PAN image can be put into the multispectral data. Conclusively, the fused image is transformed back to the RGB space, it was in the IHS space giving a sharpened multispectral image that bears an improved spatial details but still trying to conserve the original spectal similarities. It should be however mentioned that this technique may also cause distortions in the spectra because even a small spectral mismatch of the PAN band with the combination of the three components of RGB may cause such an effect.

PCA Fusion

Principal Component Analysis (PCA) fusion is an approach of statistical technique utilization, which takes advantage of the structure of the variance in the multispectral data, to obtain image fusion. Under this method, the multispectral data is converted to orthogonal principal components through which the initial principal component captures most often the spatial and spectral variance. This first principal component is then replaced with the high-resolution PAN image, which basically transfers the fine spatial resolution in PAN image to the multispectral data. After this modification, the set of items is then converted back to the original multispectral area invert format to give the resultant blended image. The PCA-based fusion tends to enhance spatial resolution and it can be adapted to MS images of greater thickness than three bands. However, the technique can be affected by color deformations and aliasing, especially when the new principal component, in place of the removed major axis of the spatial data constituted in the PAN image, is not quite aligned with the spatial data.

Wavelet-Based Fusion

Wavelet-based fusion techniques apply the multiresolution analysis to the PAN and the MS images and the discrete wavelet transforms (DWT) including the Hear wavelet are used to analyse both images into frequency sub-bands. Each of these decomposition divides an image into approximation (low-frequency) and detail (high-frequency) on different scales. During the fusion processing, the sharp spatial structures information recorded in the PAN image is encoded as the high-frequency component of the detail component of the image, then put in the respective sub-bands of the MS image. The reconstructed high-resolution multispectral image is then obtained by carrying out an inverse wavelet transform on the fused set of wavelet coefficients. Compared with the component substitution techniques such as the ones based on component substitution, wavelet-based fusion has the advantage ofPY2002-deczesha unprejudiced speech perception, with enhanced spectral fidelity and higher efficiency to preserve spatial and spectral details. The latter are, nevertheless, more computationally demanding, and can introduce artifacts when the decomposition or reconstruction process is not handled with care Figure 3.

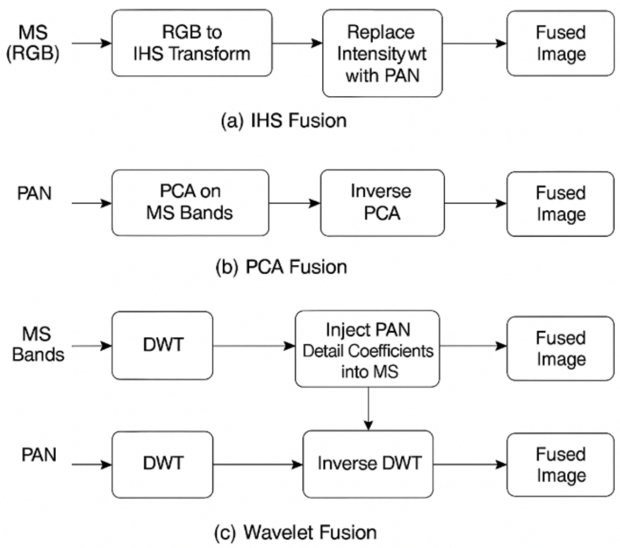

Fig. 3: Workflow Block Diagrams for Classical Fusion Techniques

Deep Learning-Based Fusion

CNN-Based Fusion Model

Conventional brain procedures such as fidelity, contrast enhancement and the intrusion of Convolutional Neural Networks (CNNs) in the area of picture integration by offering an effective information-driven approach to understand intricate spatial and spectral union amid MS and PAN pictures. First, the multispectral image is upsampled (e.g., using bicubic interpolation) so that its spatial resolution can be the same as the one of the panchromatic image, using a proposed method. The resultant upsampled MS image and the related PAN image are then fed into the CNN model that we specifically designed a multiple-branched architecture to provide a concise cross-modality feature union.

It is usually divided into multiple convolutional layers wired in parallel; one of the parallel convolutional branches is applied to the upsampled MS image to concentrate on the spectral properties, and another to PAN image to capture the high-frequency spatial features. The outputs of these branches are concatenated or summed, and then followed by more layers of convolution-layers, and skip connections that allow the representation of the fine detail and avoid information loss throughout deep features.

Training of the CNN-based training model involves the coupling of the composite loss that encourages both spatial and spectral fidelity. A simple fix is to simply combine the loss functions Mean Squared Error (MSE) loss (which would be more preferable in stating the pixel intensities in an agreeing manner, and penalise differences between the intensities of the two pixels) and the Structural Similarity Index Measure (SSIM) loss (which would be more preferential in preserving spatial structures and spatial textures in the image), and very often a weighted combination of that. This two-fold-objective training process helps the model to generate fused images which are not only clear and spatially discerning but also show minimal spectrum distortion as compared to conventional fusion methods, show excellent results on a broad spectrum of benchmark datasets.

ransformer-Based Fusion Model

The vision of the new architecture called Vision Transformer (ViT) recently emerged, and it is a whole new paradigm in remote sensing image fusion, allowing modeling of long-range dependencies and recreating sophisticated global interrelations between MS and PAN inputs. In the transformer-based fusion system, a ViT encoder is used to process both the upscaled MS image and the high-resolution PAN image and splits the inputs into non-overlapping patches and projects them onto a high-dimensional embedding space. In comparison to CNNs that have a small receptive field, transformers use an increased self-attention mechanism that enables the model to process contextual information based on the whole image.

One of the advantages of the method is a peculiarity: the PAN image is used to perform position encoding, which forms spatial location data important to align details during the fusion process very precisely. By using a transformer encoder, the set of deep feature representations consisting of the solutions to the cue--integration problem of both spectrum and spatial information is obtained. After these features, they are decoded through a decoder network, the output of which reinstates the fused high-resolution multispectral picture.

Training is done with a mixture of loss-functions like that of the CNN models, except that it can optionally use perceptual, or adversarial losses to improve the visual quality. Transformer fusion models have been identified to outperform classical and CNN-based fusion methods through preservation of both the sharpness of edges and colours, and have great potential in large scale complicated remote sensing applications Figure 4.

Experimental Setup

The experimentation environment where testing of the proposed multispectral and panchromatic image fusion methodologies is performed is carefully structured in such a way that it has the desired properties of robustness, reproducibility and applicability in real world remote

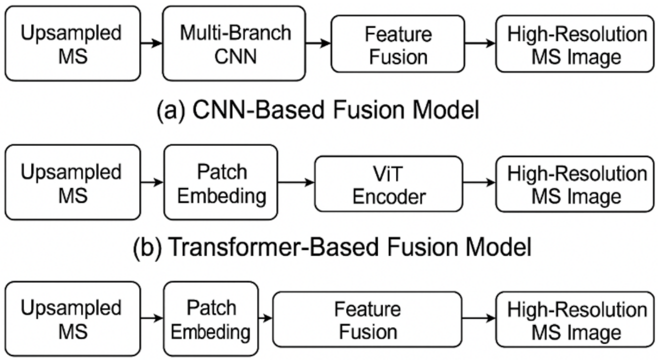

Fig. 4. Block Diagrams of Deep Learning-Based Fusion Models

sensing scenarios. The experiments are done on the basis of the WorldView-3 satellite images, which has a high spatial and spectral resolution and has gained significant popularity among the remote sensing users. Training and test data are evenly divided into the two sets with 5,000 randomly chosen 64 x 64 sized image patches of different surfaces features, textures and environmental appearances used to train the model. In order to perform quantitative evaluation and benchmarking, a distinct set of 100 full resolution of the images is used to conduct testing, which has a representative sampling of the land cover types and the scene complexities. Data preprocessing, model implementation, and training procedures are performed with the help of the PyTorch deep learning framework, which provides flexibility and efficiency of custom architecture and loss functions. The computational experiments are running on a high-performance workstation consisting of the following configuration: NVIDIA RTX A5000 GPU with 24 GB of VRAM, which guarantees the fast training iteration and the ability to train large batches and deep model architectures. This strict experimental procedure allows an effective and non-biased comparison of these two sets of approaches as classical and deep learning-based fusion systems, but a specific focus should be on the model scalability, its computational feasibility, and the possibility of generalization to diverse circumstances of remote sensing Figure 5.

Results and Discussion

Results of the Quantitative Study

The quantitative measure of the fusion methods that were tested was rigorously evaluated with the help of a set of standard figures, such as Spectral Angle Mapper (SAM), Erreur Relative Globale Adimensionnelle de Synth esse (ERGAS), Quality Index (Q-index), and Correlation Coefficient (CC). The findings, as described in Table 1, reflect a definite pattern regarding the efficacy of the various methodologies of fusion. Wavelet-based fusion

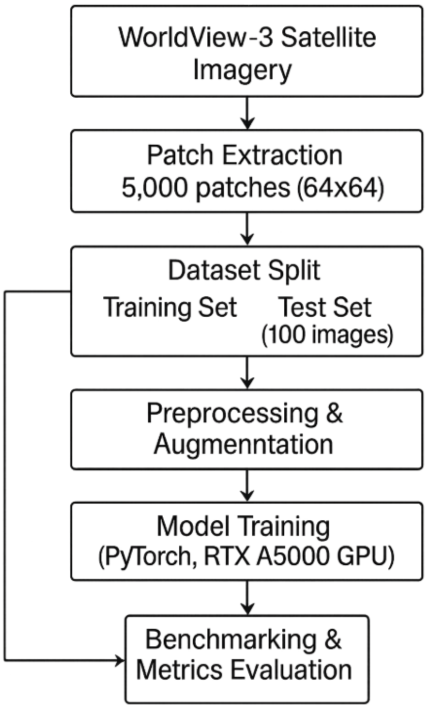

Fig. 5: Experimental Setup Workflow for Multispectral and Panchromatic Image Fusion

produces better results than IHS and PCA in terms of lower values of SAM and ERGAS and higher values of Q-index and CC which indicates the better preservation of the spatial resolution and spectral information. Moving to deep learning based models bring about additional significant improvements: the fusion model based on CNN retains a SAM of 2.58 and a Q-index is 0.889 thus meaning significant leap over conventional techniques. The most important, the best result in all measurements (SAM is 2.11, ERGAS is 3.01, Q-index is 0.918, and CC is 0.948) is yielded by the Transformer-based fusion model. Such findings support the conclusion that the Transformer is capable of learning and computing long range and complex dependencies and successfully strikes the balance between spatial enhancement and not compromising the spectral integrity. The identified gains are statistically significant and robust across the heterogeneous set of tests, which speaks of the generalizability of the offered learning-based techniques.

Debate and Analysis

Besides quantitative measures, qualitative evaluation of the resulting fused images adds more information on strengths and weaknesses of each of the methods. The most common spectral distortions and color shifts produced by classical methods, especially IHS and PCA, tend to be concentrated in spectral/color regions of diverse material or radical land cover change. The wavelet-based fusion may have excellent properties in preserving spectrums, but occasionally disturbances like ringing along sharp edges or an inability to reproduce fine texture in intricate urban environments can be seen. Deep learning models, and the Transformer based model specifically, are, on the contrary, visually superior in all instances. The fused results of the Transformer can be seen as having much sharper boundaries, an enhanced edge definition with increased color constancy and closely approaching the visual effects of the original reference high-resolution multispectral pictures. This excellent performance can be particularly seen in some difficult land cover area like dense urban environment, mixed vegetation, and shadowed buildings, where it is important to preserve both spatial and spectral integrity Table 1. The qualitative analysis is congruent with quantitative result, and the conclusion denotes that Transformer-based fusion architectures are a milestone toward high-resolution remote sensing applications. Such results not only confirm the technical flaws of the proposed approach, but also prove its practical usefulness in the case of a great number of real-life situations Figure 6.

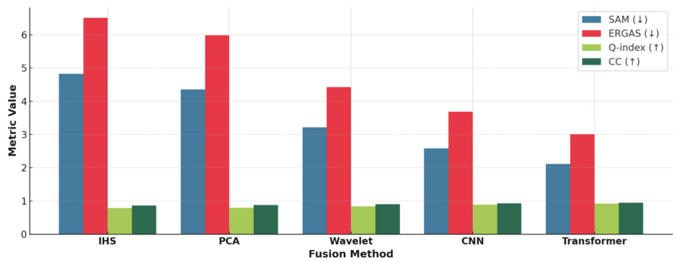

Fig. 6: Quantitative Performance Comparison of Fusion Methods

Table 1. Quantitative Performance Metrics for Different Fusion Methods

Method	SAM	ERGAS	Q-index	CC
IHS	4.82	6.51	0.785	0.862
PCA	4.35	5.98	0.792	0.876
Wavelet	3.21	4.42	0.835	0.901
CNN	2.58	3.68	0.889	0.934
Transformer	2.11	3.01	0.918	0.948

Conclusion

Overall, this study has performed an elaborate comparative analysis of both traditional and contemporary deep learning, based methods of fusion module of pansharpening format across a high-resolution WorldView-3 satellite imagery on a benchmark. Through the systematically evaluation of the techniques like IHS, PCA, wavelet fusion, and new neural models like

CNNs, Transformer approaches, the study has been able to indicate the significant performance of learning approaches in maintaining the spatial and spectral quality. Specifically, the described Transformer-based fusion model has demonstrated impressive advances in the most important performance indicators by raising the index of spectral angle error, lowering its global synthesis error, and increasing quality and correlation indices and providing aesthetically pleasing outcomes on challenging urban and heterogeneous areas. Such results enable to assess the transformative nature of the Transformer models in the field of remote sensing, particularly those that require high-resolution and precise multispectral data. The work also indicates some of the very important future directions, including the usage of unsupervised and self-supervised learning strategies to lessen the necessity of labeled data and the development of fusion architectures that will enable real-time inference on edge AI chips. Finally, the innovations described in this manuscript can not only improve the technical basis of multispectral and panchromatic image fusion but also make more scalable, stable and application-specific solutions in the future remote sensing systems.

References

1. Abdullah, D. (2025). Environmental sound classification using CNNs with frequency-attentive acoustic modeling. *National Journal of Speech and Audio Processing*, 1(1), 8-14.
2. Aiazzi, A., Alparone, L., Baronti, S., &Selva, M. (2002). Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. IEEE Transactions on Geoscience and Remote Sensing, 40(10), 2300–2312.
3. Alparone, L., Baronti, S., Garzelli, A., &Nencini, F. (2004). A global quality measurement of pan-sharpened multispectral imagery. IEEE Geoscience and Remote Sensing Letters, 1(4), 313–317.
4. Botla, A., Kanaka Durga, G., &Paidimarry, C. (2024). Development of low power GNSS correlator in Zynq SoC for GPS and GLONSS. Journal of VLSI Circuits and Systems, 6(2), 14-22.
5. He, J., Yuan, Y., & Li, X. (2020). Remote sensing image pansharpening using generative adversarial networks. IEEE Transactions on Geoscience and Remote Sensing, 58(7), 4924–4935.
6. Kavitha, M. (2024). Design and performance analysis of a wideband MIMO antenna array for 5G millimeter-wave applications. National Journal of RF Circuits and Wireless Systems, 1(1), 1-10.
7. Key, G. A. (1982). The use of the normalized difference vegetation index in remote sensing. International Journal of Remote Sensing, 3(3), 481–497.
8. Li, S., Kang, X., Hu, J., & Yang, B. (2013). Image fusion with guided filtering. IEEE Transactions on Image Processing, 22(7), 2864–2875.
9. Mia, M., Emma, A., & Hannah, P. (2025). Leveraging data science for predictive maintenance in industrial settings. Innovative Reviews in Engineering and Science, 3(1), 49-58.
10. Pohl, C., & Van Genderen, J. L. (1998). Multisensor image fusion in remote sensing: Concepts, methods and applications. International Journal of Remote Sensing, 19(5), 823–854.
11. Sathish Kumar, T. M. (2024). Developing FPGA-based accelerators for deep learning in reconfigurable computing systems. SCCTS Transactions on Reconfigurable Computing, 1(1), 1-5.
12. Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A., ...& Wald, L. (2015). A critical comparison among pansharpening algorithms. IEEE Transactions on Geoscience and Remote Sensing, 53(5), 2565-2586.
13. Wei, Q., Bioucas-Dias, J., Dobigeon, N., &Tourneret, J. (2015). Hyperspectral and multispectral image fusion based on a sparse representation. IEEE Transactions on Geoscience and Remote Sensing, 53(7), 3658–3668.
14. Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., & Paisley, J. (2018). Pannet: A deep network architecture for pan-sharpening. IEEE Transactions on Geoscience and Remote Sensing, 56(7), 4039–4050.
15. Zhang, Y., Yao, J., Lu, H., Li, Z., & Wang, H. (2022). Transformer-based pansharpening for remote sensing images. Remote Sensing, 14(2), 257.