ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

A COMPARATIVE STUDY OF CNN-BASED AND TRADITIONAL MULTI-FOCUS IMAGE FUSION TECHNIQUES WITH THE INTRODUCTION OF A NOVEL HYBRID MODEL

Muhammad Abdullah Irfan Khan[1], Muhammad Ahmad [2], Asim Amin [2], Syed Hammad [3], Dr Amna Khan [4], Noor Fatima [5]

[1,2,5] Department of computer science and information technology, Superior University
su92-mspmw-f23-003@superior.edu.pk
ahmadkahloon@superior.edu.pk
su92-mspmw-004@superior.edu.pk
[2,3]Department of computer science and information technology, The university of Chenab
asim@cs.uchenab.edu.pk
hammad@cs.uchenab.edu.pk
[4]Department of computer science and information technology, The university of Hafr AlBatin
amna.khan@uhb.edu.sa

Abstract-
This paper provides an extensive review of multi-focus image fusion techniques, highlighting recent advancements in transforming domain methods, deep learning approaches and hybrid strategies. The proposed framework combines classical multiscale decompositions such as Discrete Wavelet Transform (DWT) and Laplacian Pyramid with convolutional neural networks (CNNs) for enhanced refinement, aiming to capture intricate structural and perceptual details for superior fusion quality. Experimental evaluations reveal that hybrid models consistently outperform traditional methods in both visual fidelity and objective metrics. Despite these improvements, challenges remain, including real-time processing limitations, robustness under diverse imaging conditions and the absence of standardized benchmarks. Future research should focus on optimizing computational efficiency and developing adaptive fusion frameworks capable of addressing varied real-world scenarios.

Keywords: multi-focus image fusion, hybrid fusion, DWT, Laplacian Pyramid, CNN, residual learning PSNR, SSIM, PIQE.

1. Introduction

Image fusion is a vital area of image processing that integrates complementary information from multiple inputs into a single, more informative representation. Within this field, multi-focus image fusion has gained prominence due to the inherent depth-of-field limitations of imaging systems: a single exposure cannot keep all regions of a scene simultaneously sharp. By combining several images of the same scene each focused on different regions multi-focus fusion produces a comprehensive result in which all relevant objects appear crisp and clear. This capability is crucial across real-world domains such as medical diagnostics, remote sensing, surveillance, industrial quality control, and consumer
photography, where visual clarity directly influences decision-making. Over time, a broad spectrum of fusion strategies has been explored to enhance quality. Early methods centered on transform-based techniques including wavelet, contourlet, and shear let transforms which decompose images into multiple scales and orientations to capture both fine and coarse details. While these approaches proved effective in enhancing edges and structural content, they often struggled with robustness under noise and were limited in modelling complex patterns. To mitigate these issues, sparse representation frameworks were introduced, representing images

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

as linear combinations of learned dictionary atoms; this improved detail preservation but incurred substantial computational costs. The advent of deep learning has reshaped the landscape. Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and more recently attention-based models have enabled data-driven learning of fusion rules from large corpora. In contrast to handcrafted strategies, these models adaptively capture spatial and contextual relationships, yielding fused images with superior sharpness, structural integrity, and contrast. In parallel, hybrid methods that integrate classical signal processing with deep learning have emerged, aiming to balance interpretability, efficiency, and performance. Despite these advances, several challenges persist. Achieving high fusion accuracy in real time remains difficult, especially in time-critical settings such as medical imaging and autonomous systems where latency is unacceptable. Moreover, variations in illumination, noise, and motion can degrade fusion effectiveness. A further barrier is the absence of universally accepted benchmark datasets and evaluation metrics, which complicate fair and reproducible comparisons among methods. Overall, multi-focus image fusion is a dynamic, practically significant research area. The field has progressed from early transformation-domain strategies to sophisticated learning-based models that markedly improve image clarity and structural preservation. Looking ahead, tighter integration of domain knowledge with data-driven techniques is expected to further enhance the quality, reliability, and adaptability of multi-focus fusion, broadening its suitability for diverse and complex real-world scenarios.

2. Literature Review

Advanced multi-focus image fusion method based on a convolutional neural network (CNN) that leverages a residual atrous spatial pyramid pooling (RASPP) module a disparities attention mechanism (DAM) and residual blocks to enhance feature extraction and correspondence. The network employs supervised training on a specially constructed dataset derived from VOC201 enabling it to effectively fuse images with different focal points by capturing multi-scale multi-level features and maintaining feature consistency. Experimental results including ablation studies and comparisons with other methods demonstrate that this approach achieves superior subjective and objective performance metrics such as PSNR SSIM and information entropy while also reducing computational complexity. Overall, the method effectively preserves details and boundaries in fused images contributing to improved image clarity and fidelity in multi-focus fusion tasks.[1] Deep Learning has recently achieved great success in Multi-Focus Image Fusion (MFIF).Conventional two-class focus classification methods often fail causing artifacts and sensitivity to misregistration. To address this, we propose MCMS CNN, a multi-classification and multi-scale decomposition-based MFIF method. A CNN classifier first generates focus probability maps which are then fused using refined multi-class rules. Experiments show that our method delivers superior fusion quality and stronger robustness compared to existing approaches.[2] A discrete wavelet transform (DWT)-based method for multi-focus image fusion. Different fusion rules are applied to low and high frequency sub-bands. Low-frequency coefficients are selected using a maximum sharpness focus measure. High-frequency coefficients are fused using a maximum neighboring energy scheme with consistency verification. Experimental results show the proposed method outperforms conventional fusion techniques.[3] Multi-focus image fusion is a process that combines several images of the same scene, each focused on different areas into one clear and complete image. In this work, a new method is introduced that separates each image into two parts: the smooth shapes and the detailed pattern. This is done using an improved decomposition technique that works quickly and accurately captures the natural structure of the image. Different fusion rules are then applied to both parts and the results are merged to create a single sharp image. This

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

approach not only keeps the original structures intact but also reduces unwanted artifacts and noise. Experiments show that this method provides better results both visually and through numerical measurements compared to existing techniques.[4] This paper introduces a new method for multi-focus image fusion. The process starts by building a joint dictionary from multiple smaller dictionaries which are learned directly from the source images using the K-SVD algorithm. Unlike other methods, this approach does not require prior knowledge or external training datasets. Next sparse coefficients are calculated using the batch-OMP algorithm which speeds up the coding process. The fused image is then reconstructed by applying a maximum weighted multi-norm rule ensuring that it captures the most important details from the source images. To test its effectiveness experiments were carried out on different multi-focus and artificially blurred images. The results show that this method performs better than many existing techniques, both visually and in terms of numerical evaluation.[5] Multi-focus image fusion helps overcome the depth-of-field problem in cameras by combining multiple partially focused images into one clear fully focused image. This paper introduces boundary segmentation as one category of fusion methods and proposes a new way to classify fusion algorithms into four groups transform domain method, boundary segmentation methods, deep learning methods, and hybrid methods. It also outlines both subjective and objective evaluation standards explaining eight commonly used objective indicators in detail. Based on a wide review of existing studies, the paper compares different representative methods, highlights current challenges, and discusses possible directions for future research in multi-focus image fusion. This discusses different multi-focus image fusion techniques used in image processing. It explains that blurred images often occur because cameras have a limited depth of field meaning only certain areas are in focus while others appear unclear. Multi-focus image fusion is designed to solve this issue by combining multiple focused images into one sharp, all-in-focus image. [6] Capturing both foreground and background sharply in DSLR photography is difficult due to limited depth of field often leading to blurred images. Existing multi-focus fusion methods face challenges with image quality and varying input angles. To overcome this, we propose a new fusion method combining Convolutional Neural Networks (CNNs) with Gaussian Pyramid techniques. The process involves four steps: visual search, segmentation, compression analysis, and synthesis. The Gaussian Pyramid enhances edge detection and object identification, while CNNs improve feature learning. Evaluations using PIQE, PSNR, SSIM, and connectivity show our model achieves clearer images, with a 4.70% improvement in PIQE over previous CNN-based methods.[7] Image fusion combines information from multiple images into one clearer and more complete image. It helps improve image quality and widens the scope of applications. different image fusion techniques used in research. Basic methods include averaging, selecting maximum and selecting minimum values. Advanced techniques include Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA). A comparison of these methods highlights the most effective approaches and suggests directions for future research. The goal of image fusion is to combine important details from multiple images into one reliable, clear image. A main challenge in multi-focus image fusion is identifying which regions are in focus. This paper introduces a new method that uses the Mean shift algorithm to locate focused regions followed by edge detection and morphological techniques to find their boundaries. For boundary fusion, a combination of pulse coupled neural networks (PCNN) and Gaussian fuzzy methods is applied in the NSCT domain. Finally, focused regions and fused boundaries are merged to form the final image. Experimental results show that this approach more accurately identifies focused regions and produces higher-quality fused images compared to traditional methods, both visually and through objective measures.[8] Traditional multi-focus image fusion methods often fail to make full use of spatial context information. To overcome this, a

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

new segmentation-based approach is introduced. The method works in two main stages: first, PSPNet is used to identify the focused regions in the source images; second, ConvCRFs refine the segmentation map for greater accuracy. The final fusion is carried out using these improved maps. Experiments on 20 pairs of color multi-focus images show that this method delivers clearer and more visually appealing results than many existing state-of-the-art techniques, both in subjective observation and objective evaluation.[9] Multi-focus image fusion is widely used in image processing and computer vision to combine important details from multiple images. This paper introduces a new fusion method that detects focused regions using a combination of mean filter and guided filter. First, rough focus maps are generated using a mean filter and difference operator, then refined with a guided filter. An initial decision map is created with the pixel-wise maximum rule and further improved using the guided filter. Finally, the fused image is produced through a pixel wise weighted averaging approach. Experiments show that this method is more robust against noise, faster in computation and delivers better visual quality and objective results compared to many existing techniques.[10] Multi-focus image fusion helps solve the depth-of-field limitation in photography by combining several partially focused images into one clear all-in-focus result. In recent years, progress in areas such as multi-scale analysis sparse representation and deep learning has pushed this field forward. This survey provides an overview of existing fusion methods and introduces a new way of classifying them into four groups transform domain spatial domain, hybrid approaches and deep learning-based methods. Each category is explained with representative techniques, and a comparative study of 18 methods is carried out using 30 pairs of multi-focus images and 8 evaluation metrics. The authors also share datasets, metrics, and results online to serve as benchmarks for future studies. The paper closes by highlighting key challenges that remain and suggesting possible directions for further research.[11] Coupled Neural P (CNP) systems are a new type of distributed and parallel computing model inspired by coupled and spiking neurons. Unlike traditional spiking neural P (SNP) systems, they work with three types of data units and use a coupled firing mechanism with dynamic thresholds. This paper applies CNP systems to the problem of multi-focus image fusion and introduces a new method in the non-subsampled contourlet transform (NSCT) domain. Two CNP systems with local topology are used to manage the fusion of low-frequency coefficients. The method is tested on a dataset of 19 multi-focus images using five evaluation measures and compared against 11 advanced fusion techniques. Both visual and numerical results show that the proposed method produces superior image quality and fusion performance.[12] A new multi-focus image fusion method using the non-subsampled shear let transform (NSST). An initial fused image is created with a standard multi-resolution fusion approach. Then by comparing pixel errors between the source images and the initial fusion the focused regions are identified. Morphological opening and closing are applied to refine these regions. The focused areas and their borders are then used to guide the fusion process in the NSST domain and the final fused image is reconstructed using inverse NSST. Experiments show that this method captures more important details while reducing unwanted artifacts, performing better than DWT-based, NSCT-based, and earlier NSST-based methods in both visual quality and objective evaluations.[13] Measuring pixel sharpness is crucial for effective multi-focus image fusion. In this work, a gray image is treated as a two-dimensional surface and pixel sharpness is evaluated using a neighbor distance measure derived from differential geometry. A smooth image surface is reconstructed through kernel regression and the neighbor distance filter is then applied within a multi-scale analysis framework. Based on this approach, a new multi-focus fusion method is proposed. Experimental results show that it outperforms traditional fusion techniques, achieving better results on evaluation metrics such as spatial frequency, standard deviation and average gradient.[14] The image is treated as a two-dimensional surface and sharpness is calculated using a neighbor distance measure derived

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

from geometry refined with kernel regression. A multi-scale framework is then built using this sharpness measure to guide the fusion process. Experiments show that the method produces clearer fused images than conventional techniques with better results on measures like spatial frequency standard deviation and average gradient[15] Authors proposes a deep learning-based method for multi-focus image fusion. Instead of using complex filters and rules, a CNN is trained with sharp and blurred image patches to directly generate the focus map. This approach simplifies the process, improves fusion quality and runs fast enough for practical use. It also shows potential for other image fusion tasks.[16].

3. Methodology

3.1 Data and Preprocessing

We consider pairs of multi-focus images (I_1, I_2). Images are converted to a common color space (YCbCr) and fused on the luminance (Y) channel; chroma channels are carried from the source patch with the larger local focus measure. Prior to fusion, images are resized to 256×256, intensity-normalized to [0,1], and aligned (when necessary) using feature-based registration to reduce ghosting. For training the CNN refiner, we use a mix of real pairs (e.g., Lytro-style scenes) and synthetic pairs generated from high-quality all-in-focus images by applying spatially varying blur masks. Data augmentation includes rotations (±90°), horizontal/vertical flips, random contrast scaling, and mild Gaussian noise.

3.2 Multiscale Decomposition and Initial Fusion

Each source image undergoes a two-branch multi-scale analysis: (a) Discrete Wavelet Transform **(DWT)** producing sub-bands **{LL, LH, HL, HH}** and (b) a Laplacian Pyramid (LP) capturing band-pass details across levels. For the low-frequency band (LL), we adopt a max-selection rule to retain global contrast from the more informative source: $\mathbf{LL\_f(i,j) = max(LL_1(i,j), LL_2(i,j))}$. For high-frequency bands $\mathbf{C \in \{LH, HL, HH\}}$, coefficients are selected by

magnitude: $\mathbf{C\_f(i,j) = argmax\_\{k\in\{1,2\}\} |C\_k(i,j)}$ with a small consistency check (3×3 majority) to suppress isolated artifacts. Reconstruction via inverse transforms yields an initial fused image **I_init** that already exhibits improved sharpness but may contain ringing or subtle seam artifacts.

3.3 CNN Residual Refiner

To correct residual artifacts, we train a lightweight CNN R to predict a residual map Δ such that $\mathbf{I\_out = I\_init + \Delta}$. The network accepts a three-channel tensor formed by concatenating $\mathbf{[I_1, I_2, I\_init]}$ along the channel dimension and comprises five convolutional layers: three 3×3 layers (64 filters) followed by two 5×5 layers (32 then 1 filter). Each hidden layer uses ReLU activations and batch normalization; the last layer uses sigmoid to keep outputs in [0,1]. We optimize Mean Squared Error (MSE) against reference all-in-focus images where available (synthetic data) and a self-supervised proxy loss combining MSE on gradient maps and structural similarity terms for real pairs. Training uses Adam (learning rate 1e-3), batch size 16, 50 epochs with early stopping (patience 8).

3.4 Evaluation Protocol

We report PSNR and SSIM on pairs with available ground truth (synthetic test set) and PIQE on real pairs lacking references. Metrics are computed on the Y channel; PIQE scores (lower is better) are reported on the fused RGB image.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

We compare three settings: (i) DWT-only, (ii) CNN-only (direct mapping from [I_1, I_2] to fused image without transforms) and (iii) the proposed Hybrid (DWT/LP + CNN refiner).

$$PSNR = 10 \cdot \log_{10} \frac{MAX_I^2}{MSE}$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

4. Experiments and Results

Training curves show rapid loss reduction in the first 10 epochs followed by gradual convergence, indicating stable learning. Quantitatively, the hybrid model yields consistent gains over both baselines.

Method	PSNR (dB)	SSIM	PIQE ↓
DWT-only	28.4	0.89	25.6
CNN-only	30.1	0.93	21.3
Hybrid (Proposed)	32.8	0.96	18.2

The hybrid approach benefits from strong initial priors provided by coefficient-selection and learns to suppress ringing and boundary inconsistencies via residual correction. Improvements in PIQE indicate perceptual gains even when full-reference metrics saturate. Visual inspection (edges, textures, and fine structures) confirms fewer halos and better focus continuity across transition regions.

5. Conclusion

This study highlights the progression of multi-focus image fusion from traditional transform-based techniques to advanced deep learning approaches, emphasizing the superior performance of hybrid models that integrate classical signal processing with CNN-driven refinement. These innovations significantly improve image clarity, structural consistency, and perceptual quality. However, challenges such as high computational requirements, sensitivity to noise, and variability in input conditions continue to hinder large-scale practical adoption. Overcoming these limitations through optimized algorithms, standardized benchmarking, and adaptive frameworks will be essential for advancing multi-focus fusion to meet diverse real-world application needs. Future work will explore attention-augmented refinement, deformable registration, and quantization-aware training for edge deployment.

References

[1] L. Jiang, H. Fan, and J. Li, 'A multi-focus image fusion method based on attention mechanism and supervised learning,' Appl Intell, 52(1), 339–357, 2022.
[2] L. Ma, Y. Hu, B. Zhang, J. Li, Z. Chen, and W. Sun, 'A new multi-focus image fusion method based on multi-classification focus learning and multi-scale decomposition,' Appl Intell, 53(2), 1452–1468, 2023.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

[3] Y. Yang, 'A Novel DWT Based Multi-focus Image Fusion Method,' Procedia Engineering, 24, 177–181, 2011.
[4] Z. Liu, Y. Chai, H. Yin, J. Zhou, and Z. Zhu, 'A novel multi-focus image fusion approach based on image decomposition,' Information Fusion, 35, 102–116, 2017.
[5] H. Yin, Y. Li, Y. Chai, Z. Liu, and Z. Zhu, 'A novel sparse-representation-based multi-focus image fusion approach,' Neurocomputing, 216, 216–229, 2016.
[6] A. Anish and T. J. Jebaseeli, 'A Survey on Multi-Focus Image Fusion Methods,' 1(8).
[7] K. Ayub, M. Ahmad, F. Nasim, S. Noor, and K. Pervaiz, 'CNN and Gaussian Pyramid-Based Approach For Enhance Multi-Focus Image Fusion,' 07(02).
[8] K. He, D. Zhou, X. Zhang, and R. Nie, 'Multi-focus: Focused region finding and multi-scale transform for image fusion,' Neurocomputing, 320, 157–170, 2018.
[9] J. Zhou, M. Hao, D. Zhang, P. Zou, and W. Zhang, 'Fusion PSPnet Image Segmentation Based Method for Multi-Focus Image Fusion,' IEEE Photonics J., 11(6), 2019.
[10] X. Qiu, M. Li, L. Zhang, and X. Yuan, 'Guided filter-based multi-focus image fusion through focus region detection,' Signal Processing: Image Communication, 72, 35–46, 2019.
[11] Y. Liu, L. Wang, J. Cheng, C. Li, and X. Chen, 'Multi-focus image fusion: A Survey of the state of the art,' Information Fusion, 64, 71–91, 2020.
[12] H. Peng, B. Li, Q. Yang, and J. Wang, 'Multi-focus image fusion approach based on CNP systems in NSCT domain,' Computer Vision and Image Understanding, 210, 103228, 2021.
[13] G. Guorong, X. Luping, and F. Dongzhu, 'Multi-focus image fusion based on non-subsampled shearlet transform,' IET Image Processing, 7(6), 633–639, 2013.
[14] H. Zhao, Z. Shang, Y. Y. Tang, and B. Fang, 'Multi-focus image fusion based on the neighbor distance,' Pattern Recognition, 46(3), 1002–1011, 2013.
[15] S. Bhat and D. Koundal, 'Multi-focus image fusion techniques: a survey,' Artif Intell Rev, 54(8), 5735–5787, 2021.
[16] D. Gai, X. Shen, H. Chen, and P. Su, 'Multi-focus image fusion method based on two stage of convolutional neural network,' Signal Processing, 176, 107681, 2020.