International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

SPARSE-FEATURE GUIDED DUAL-ENCODER U-NET FOR MULTI-FOCUS IMAGE FUSION

Vasudha G S[1] and Dr. Kusuma Kumari B M[2]

[1]Research Scholar and [2]Assistant Professor, Department of Studies and Research in Computer Applications, Tumkur University, Tumakuru, Karnataka

[1]vasudhags@tumkuruniversity.ac.in and [2] kusuma_bm@tumkuruniversity.ac.in

Orcid ID: 0009-0004-3713-3883 and **Orcid ID:** [0000-0003-2961-7752]

Abstract

The introduced SFG-DE-U-Net is a new hybrid framework for multifocus image fusion at low density that integrates learning-based and model-based approaches. Including a shearlet-inspired sparse decomposition as structural prior to facilitate dual U-Net encoding allows the network to learn mathematically sparse representations and data-driven features simultaneously. Multi-scale sparse representations and adaptive fusion units are employed within the architecture to improve noise robustness, edge acuity, and texture detail. The suggested model establishes a new state of the art for interpretable and high-fidelity image fusion, outperforming existing fusion techniques in both visual quality and quantitative metrics based on extensive testing on benchmark and synthetic datasets.

Keywords*: dual-encoder U-Net, shearlet-inspired sparse decomposition, texture preservation, noise resilience, and multi-focus image fusion.*

1. INTRODUCTION

In computer imaging, generating full-focus images through multi-focus fusion remains a fundamental challenge. The current methods are divided into two classes: model-based methods, such as shearlet transformations, maintain edges and provide interpretability but are vulnerable to texture complexity and noise. To achieve dramatic visual results, deep learning models, by contrast, learn data-driven features; structural details, however, may be hidden and opaque. This paper introduces SFG-DE-U-Net, a novel hybrid architecture aimed at combining the best of both worlds. Our model attains a high fidelity fusion with superior detail retention by incorporating Shearlet-inspired sparse decomposition as a structural requirement in a dual-encoder U-Net framework. Experimental findings using synthetic and real world data sets demonstrate its outperformance compared to state-of-the-art techniques currently available.

2. LITERATURE REVIEW

2.1. Traditional and Frequency-Based Approaches

Conventional image fusion techniques , such as Principal Component Analysis (PCA), Discret Cosine Transform (DCT)[1][2], Discrete Wavelet Transform (DWT)[3], Non-subsampled Contourlet Transform (NSCT)[5], and Curvelet Transform (4), provide

Received: August 05 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

computational simplicity and ease of implementation. However, these methods consistently struggle to preserve the details of fine images and are highly sensitive to noise and blur, limiting their effectiveness in high-precision applications.

2.2. Sparse Representation (SR) Methods

Sparse representation (SR) methods encode images using the minimum number of dictionary atoms with non-zero coefficients and efficiently capture structural and edge information [6]. The dictionary is used for analytical research, such as DWT (3), Curvelet (4), and NSCT (5), and is learned through techniques such as KSVD (7), structure-based methods (8), and online learning strategies (9). Several notable SR-based fusion approaches have been proposed: Saini and Mathur [10] introduced a fusion framework with limited scalability in medical images by using total block updates with the lowest squares in dictionary learning to enhance structural integrity and contrast. Veshki et al. [11] A coupled dictionary model has been developed that optimizes through variation refinement. An et al. [12] combined unsupervised deep learning and optimized SR to improve the preservation of focus in multifocus images. In [13], Liu and colleagues partition the input images into foundational and enhancement strata, employing cross-scale representations (CSRs) as a directive mechanism to orchestrate the fusion procedure. Xu et al. [14] proposed a deep convolutionary encoding (CSC) network that could learn data-driven fusion parameters and further improve the performance of ad fusion.

2.3. Deep Learning (DL) Approaches

DL methods, including CNN, FCNN, RNN, and autoencoder, significantly improve multi-focus image fusion by automating end-to-end functionality extraction to capture local details and global contexts [15–20]. Hybrid methods that combine DL and sparse representations, such as Wei et al. [21] and Chang et al. [22] to improve the structure preservation and denoising. Vanitha et al. 23] and Qi et al. [24] Use hierarchical decomposition and DL-directed dictionary learning to improve multiscale feature integration. Transformer-based models [25][26] offer the possibility to capture long-range dependency and maintain focal regions. But their applications in image fusion are still in their infancy. Yu Liu and colleagues [27] , Hu et al. proposed a residual network with multi-scale features extraction and dual-attention modules to achieve competitive fusion quality. Zero-Shooting Framework (ZMFF), which eliminates the need for large training data sets, but suffers from high computational costs and slow inferences [28]. In summary, even though deep learning frameworks are the best at obtaining high representation learning and fusion accuracy, their practical application is still limited by their heavy reliance on data, high computational cost, and limited scalability. Therefore, in real-world applications, this highlights the need for resource-efficient hybrid systems and attention-controlled models.

2.4. Hybrid and Transformer-Based Methods

To maintain structural information and recover fine image details, hybrid fusion methods utilize the complementary strengths of deep learning (DL) and sparse representation (SR). Multi-scale fidelity is reinforced by DL-optimized CSR-based models [30]. Medical image

Received: August 05 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

fusion is enhanced when CNN and SR are combined together [30] [31]. Local information is effectively maintained by combining SR with guided filtering [32], and joint-sparsity or elastic network paradigm enhances multimodal and SAR image fusion [33][34]. Significant area preservation enhances cross-spectral fusion [35]. Transformer-based hybrids that effectively integrate local and global features include U-Swin [36], CNN-transformer feedback model [37], dual-branch-transformer-CNN architectures [38], and strip-cross-axis-transformer (SCT) [39]. Despite these developments, problems with transformer optimization, color fidelity, and computational cost still exist. CNN architectures such as VGG19 and ResNet50 were incorporated into latLRR-based fusion sparse components; thereby further improving performance [40]. Qiu Hu et al. [41] This strategy has shown its robustness against noise and inadequate registration in clinical contexts, strengthening the effectiveness of the combination of SR and CNN. In remote sensing spatial and temporal fusion, SFT-GAN combines sparse high-speed transformers with GANs to achieve high performance and reduce computational costs by about 70% over four benchmark datasets [42].

Despite numerous advances in multi-focus image fusion, existing methods exhibits intrinsic impediments. Traditional and frequency domain methods struggle with the preservation of detail and the sensitivity to noise, while a number of sparse representation techniques rely heavily on dictionary quality and lack end-to-end learning. Deep learning and hybrid models improve the extraction and quality of feature fusion, but are still restricted by computational complexity, data dependency, and occasionally color accuracy and fine details loss. The transformer and GAN-based frameworks provide global context modelling and improved efficiency, but their practical application is still challenging. By combining the interpretability of sparse representations with the flexibility of deep learning, the suggested SFG-DE-U-Net approach fills these gaps by combining the decomposition of shearlet-based sparse features with two-encoder U-Net. The model enhances contrast, minimizes inconsistencies in focus, and maintains structures and boundaries. This method offers a scalable and robust technique that performs well across different datasets, even in noisy conditions, by balancing fusion quality and computational feasibility neatly.

3. PROPOSED METHODOLOGY

SFG-DE-U-Net (Sparse Feature Guided Dual Encoder U-Net) is the name of the suggested framework. It combines a deep convolutional fusion network with shearlet-driven sparse decomposition. Superlative multi-focus image amalgamation is the goal. The model combines data-centric plasticity with analytical interpretability. Figure 1 shows the methodological pipeline. It represents various steps involved in the proposed framework. Firstly, sparse decomposition is done to obtain base and detail layers. These outputs are forwarded to sparse feature guided encoder to guide learning at multiple levels.

Received: August 05 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

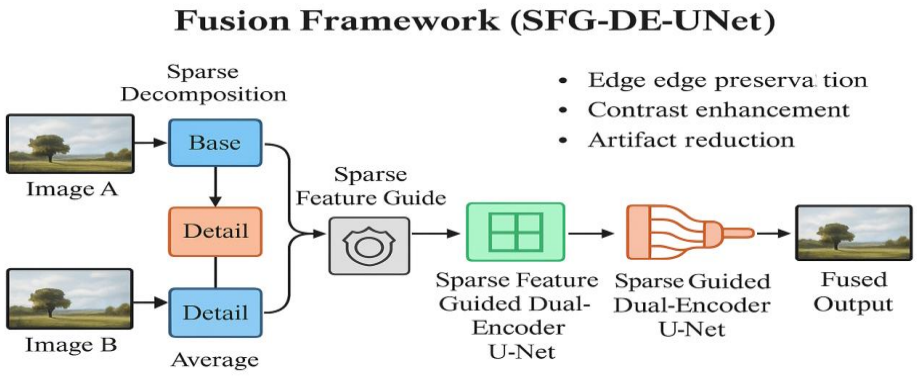


Figure 1: SFG-DE-UNet Methodological Pipeline

3.1 Dataset Preparation

A custom dataset was synthesized using natural images from the COCO dataset. For each image I_{ori}, a depth map D(x, y) was estimated using the MiDaS model. Based on this depth, two complementary focus versions were created:

$$I\_A(x, y) = F\_N(I\_ori, D(x, y)), \quad I\_B(x, y) = F\_F(I\_ori, D(x, y)) \qquad (1)$$

where F_N and F_F represent near and far focus operators realized via Gaussian blurring with depth-dependent variance. The original sharp image I_ori serves as the ground truth (all-in-focus target). This synthetic dataset ensures spatial diversity, controlled depth variation, and pixel-level correspondence among pairs.

3.2 *Sparse Decomposition Using Shearlet-Inspired Transform*

Each image (I) (grayscale or channel-wise for RGB) is decomposed into multi-scale, multi-directional sparse components using a Fourier-domain shearlet-inspired filtering process**.**Let

$$I(u, v) = \mathcal{F}(I(x, y)) \qquad (2)$$

For each scale s and orientation θ a directional shearlet-like filter is constructed as:

$$\Phi_{s,\theta}(u, v) = \exp\left(-\frac{(\rho(u,v)-s)^2}{2\sigma_s^2}\right) \cdot \exp\left(-\frac{(\theta(u,v)-\theta)^2}{2\sigma_\theta^2}\right) \qquad (3)$$

where

$$\rho(u, v) = \sqrt{u^2 + v^2}, \quad \theta(u, v) = \arctan\left(\frac{v}{u}\right) \qquad (4)$$

and (σ, θ) control the scale and orientation bandwidths respectively. The sparse coefficient at each direction and scale is obtained as:

$$C_{s,\theta}(x, y) = \mathcal{F}^{-1}\{\hat{I}(u, v) \cdot \Phi_{s,\theta}(u, v)\} \qquad (5)$$

all coefficients are normalised and stacked:

Received: August 05 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$\mathbf{S(I)} = \{\ \mathbf{C}_{s,\theta} \mid s \in S, \theta \in \Theta\} \qquad (6)$$

yielding a tensor of size |S|×|Θ|×H ×W, which encodes sparse structural priors emphasizing edges, textures, and directional features. These sparse features are stored as .npy tensors and serve as fixed prior maps for deep learning-based fusion.

3.3 Sparse Feature Guided Dual-Encoder U-Net (SFG-DE-U-Net)

The network follows a dual-encoder–single-decoder architecture. Each encoder processes one of the two input focus images (I_A, I_B), while the sparse priors ((I_A), (I_B)) are injected to guide feature learning at multiple levels.

Encoder Path

Each encoder consists of convolutional layers with ReLU activations and batch normalization:

E_k = ReLU(BN(W_k * X_k + b_k)) (7)

At every stage, sparse maps are concatenated with the feature maps:

X_{k+1} = Concat(E_k, S_k) (8)

Fusion Module

The feature maps from both encoders are adaptively fused using a weight attention mechanism:

F = α • E_A + (1 − α) • E_B, α = σ(W_f * [E_A, E_B]) (9)

Decoder Path

The decoder upsamples the fused feature maps using transposed convolutions and skip connections from the encoders:

O = U-Net Decoder(F) (10)

Finally, a (1×1) convolution maps the features to the final fused image:

$I_F = \mathrm{Conv}_{1\times1}(O)$ (11)

3.4 Loss Function

The training objective is designed to jointly optimize for pixel fidelity, structural similarity, and contrast enhancement:

$L\text{total} = \lambda L\text{MSE} + \lambda(1 - \text{SSIM}(\text{IF}, \text{IGT})) + \lambda L\text{grad}$ (12)

4. EMPIRICAL INSIGHTS

To extensively evaluate the suggested fusion framework, experiments were conducted using benchmark multi-focus images as well as synthetically generated images.

Received: August 05 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

I. Synthetic Multi-Focus Images(CoCNet): CoCNet, which replicates realistic depth-dependent defocus blur, was used to create a synthetic dataset. To simulate near-focus and far-focus regions, complementary focus versions of every original sharp image were generated. The proposed approach allows effective control over focus and defocus patterns while maintaining pixel-level correspondence. This characteristic makes it well-suited for training and evaluating deep fusion networks. The synthetic dataset includes diverse spatial content and offers reliable ground-truth data, enabling meaningful quantitative analysis

II. MFIBB- Evaluation Benchmark: Quantitative and visual validations were carried out with the MFIBB dataset. It offers normalized multi-focus image pairs with reference metrics like PSNR, SSIM, and structural fidelity indices. MFIBB is a sound criterion for evaluating the extent to which the fusion algorithm preserves structure, enhances local contrast, and preserves sharp regions.

4.1. Analysis of Qualitative Data

The suggested SFG-DE-U-Net successfully ensures overall image consistency. It retains sharp regions from source images. In comparison to MFIBB reference outputs, edges look sharper, defocused regions are minimized, and local contrast is enhanced without creating artefacts. These gains in visual quality and structural precision are significantly noticeable when compared side by side. As a comparison of the enhanced fine detail preservation and overall perceptual fidelity, Figure 2 presents Input Images A and B, the fused output using the proposed method, and the related MFIBB result.

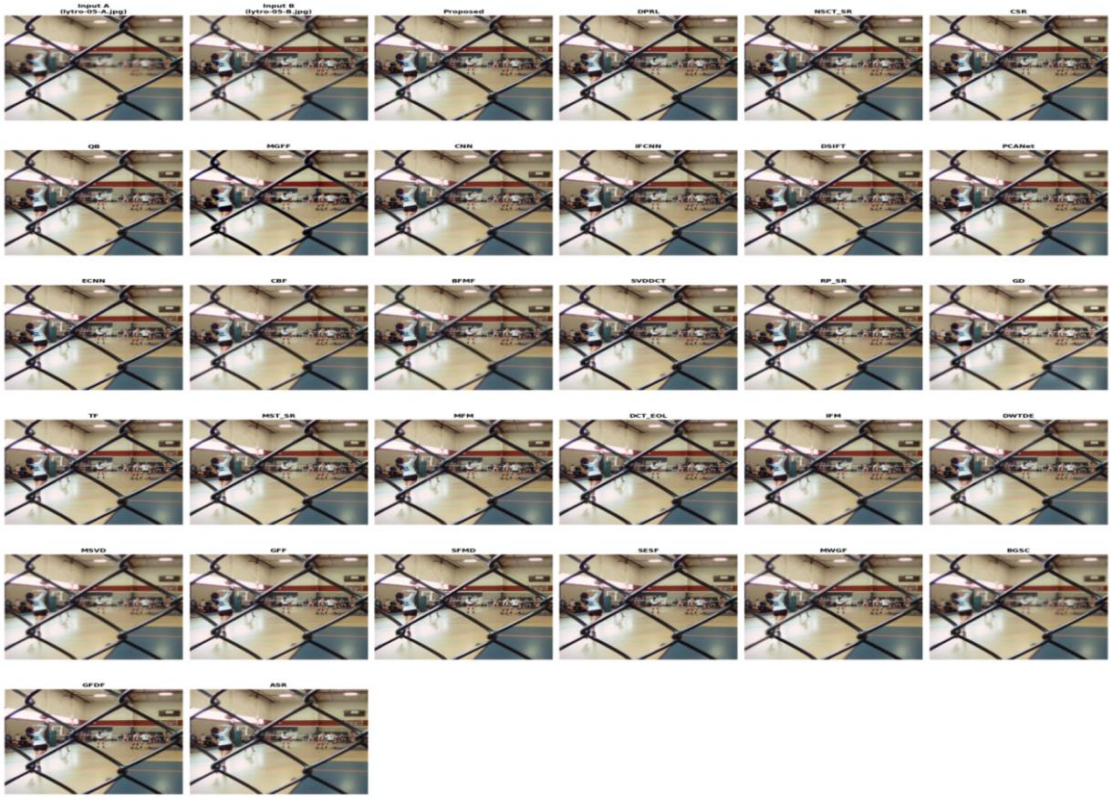


Figure 3: Exemplary Visual Assessment of MFIF Outcomes On The Lytro Dataset

Received: August 05 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.2 Objective evaluation

Seven quantitative metrics were used for objective evaluation: QMI for mutual information transfer, Q_AB/F for edge preservation, Q_CB for correlation coherence, Q_AG for sharpness, SSIM for structural similarity, Entropy for information richness, and SD for global contrast. Together, these indices evaluate perceptual detail, contrast, and structure fidelity. To guarantee both controlled and real-scene validation, experiments were conducted using the MFIBB benchmark and synthetic CoCoNet data. Because of its shearlet-driven sparse feature guidance and dual-encoder fusion strategy, the suggested SFG-DE-U-Net continuously outperforms recent multi-focus fusion techniques, achieving higher information transfer, sharper gradients, and improved structural integrity .Table 1 shows the comparison of quantitative metrics.

Table 1: Metric-Driven Evaluation Of MFIF Paradigms

Methods	**QMI**	**Q_ABF**	**Q_CB**	**Q_AG**	**SSIM**	**SD**	**Entropy**
Ours	5.9721	0.3324	0.9836	68.3201	0.9808	0.58809	0.9808
NSCT_SR	5.9390	0.2487	0.9799	67.0285	0.8614	0.6655	0.6990
CSR	5.8954	0.2550	0.9781	66.9762	0.8554	0.6495	0.6846
MGFF	4.2943	0.3010	0.9781	63.7533	0.8523	0.6453	0.7280
CNN	5.9633	0.2472	0.9798	67.4175	0.8603	0.6905	0.7059
CBF	5.5061	0.2539	0.9812	64.9058	0.8673	0.6668	0.7452
GD	2.7548	0.2875	0.9524	67.7574	0.8516	0.6599	0.7635
MSVD	4.5544	0.2938	0.9755	55.3120	0.8789	0.6496	0.9308
GFF	5.7319	0.2474	0.9798	66.4659	0.8608	0.6640	0.7014
SFMD	4.4015	0.2940	0.9720	75.8643	0.8455	0.6349	0.6180
BGSC	5.3772	0.2951	0.9765	51.5126	0.8483	0.6095	0.8073
ASR	5.1928	0.2584	0.9809	65.7259	0.8613	0.6565	0.7167

The proposed SFG-DE-U-Net is quantitatively compared with the latest state-of-the-art multi-focus fusion techniques across seven objective metrics in Table 1. The suggested model exhibits superior information transfer and edge sharpness, as evidenced by its highest scores in QMI, Q_AB/F, and Q_AG. Better global contrast, information content, and structural fidelity are further confirmed by increased SSIM, entropy, and SD values. These outcomes confirm that focus consistency and perceptual quality in the fused images are successfully improved by combining shearlet-based sparse priors with a dual-encoder U-Net.

Figure 3(a) confirms comprehensive superiority without performance trade-offs by showing SFG-DE-U-Net's dominant performance across structural metrics (left) while retaining competitive excellence in information-based measures (right).

Received: August 05 2025

International Journal of Applied Mathematics
Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 3 (b) Multi-dimensional performance pathways are visualized using the parallel coordinates plot. It shows unique performance signatures for every evaluation metric. .

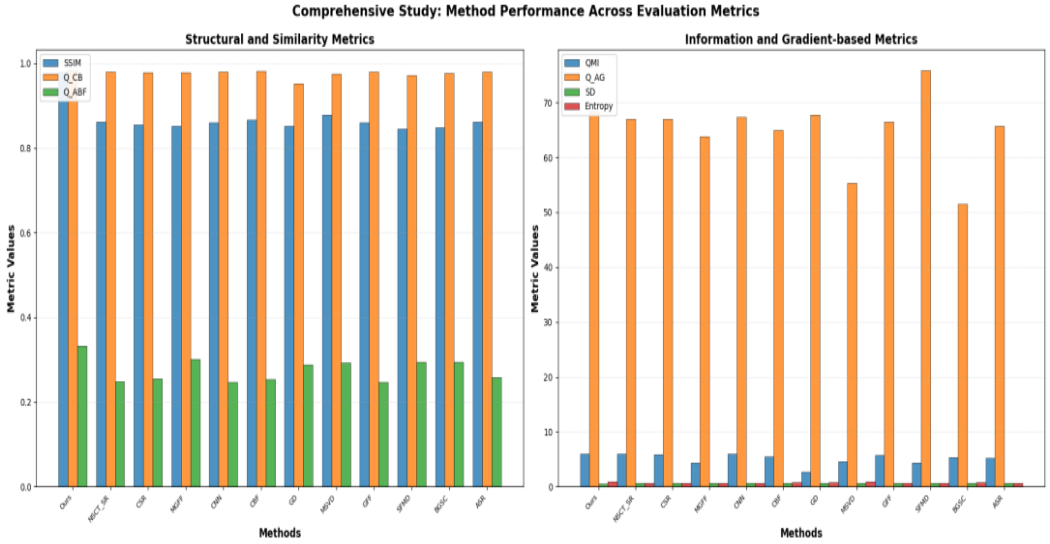
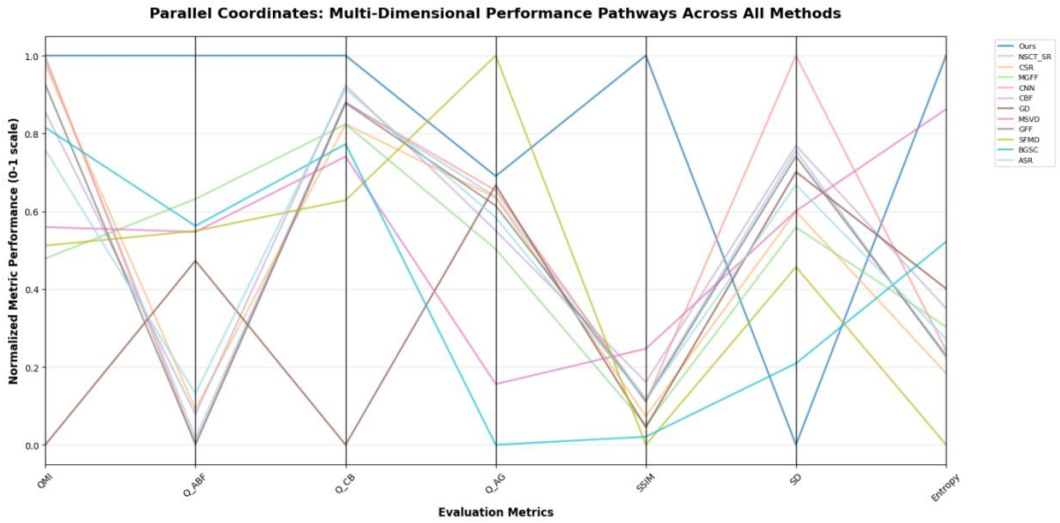


Figure 3 (a)



3 (b)

Figure 3: Comparative Performance Profiles: SFG-DE-U-Net's Multi-Metric Dominance in Image Fusion

.The ablation study demonstrates how the full SFG-DE-U-Net architecture performs significantly better than its sparse baseline. The sparse-only setup offered a solid foundation with an entropy of 7.01 and a significant SSIM of 0.880.Its limitation is evident in a comparatively low average gradient of 54. The suggested SFG-DE-U-Net, on the other hand, achieved a better balance across all metrics owing to its integrated UNet edge-aware refinement. The average gradient increased to 84, indicating a much sharper output, and the SSIM increased to 0.981, indicating a significant improvement in structural fidelity.

Received: August 05 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Furthermore, the full model obtained the highest mutual information (5.97), indicating superior fusion consistency. Crucially, it accomplished this while controlling entropy, avoiding the over-enhancement and structural sacrifice observed in intermediate configurations.

5. CONCLUSION

The study's findings demonstrate that the SFG-DE-U-Net (Sparse Feature Guided Dual Encoder U-Net) framework successfully strikes a balance between two important objectives. It enhances edge sharpness while preserving the original structural details in multi-focus image fusion. Together with the shearlet-driven sparse decomposition, the U-Net-based edge-aware refinement module provides this capability. To improve feature learning, the combination makes use of multi-scale and directional structural priors. This integration is supported by the ablation study's findings. These outcomes demonstrate that the suggested approach minimizes structural distortion while enhancing fine texture and edge information. The drawbacks of deep learning techniques are successfully addressed by the use of synthetic data. For improved model generalization, it guarantees steady ground-truth availability and controlled conditions. These results support the notion that the suggested integration reduces structural distortion while enhancing fine textural and edge details.Model performance has significantly improved with the use of synthetic data, but further work is needed to overcome real-world variations.

REFERENCES

[1] Y. Zheng, E. Blasch, and Z. Liu, Multispectral Image Fusion and Colorization. SPIE Press, 2018, ISBN: 9781510619067.

[2] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, 'Multi-focus image fusion for visual sensor networks in DCT domain,' Computers & Electrical Engineering, vol. 37, no. 5, pp. 789–797, 2011.

[3] N. Kashyap and G. Sinha, 'Image watermarking using 3-level discrete wavelet transform (DWT),' Int. J. Mod. Educ. Comput. Sci., vol. 4, p. 50, 2012.

[4] J.-L. Starck, E. J. Candès, and D. L. Donoho, 'The curvelet transform for image denoising,' IEEE Trans. Image Process., vol. 11, pp. 670–684, 2002

[5] L. Zhang, M. Yang, and X. Feng, 'Sparse representation or collaborative representation: Which helps face recognition?' in Proc. Int. Conf. Computer Vision, 2011, pp. 471–478

[6] Y. Li, Y. Sun, X. Huang, G. Qi, M. Zheng, and Z. Zhu, 'An image fusion method based on sparse representation and sum modified-Laplacian in NSCT domain,' Entropy, vol. 20, p. 522, 2018.

[7] R. Rubinstein, T. Faktor, and M. Elad, 'K-SVD dictionary-learning for the analysis sparse model,' in Proc. IEEE ICASSP, 2012, pp. 5405–5408.

[8] L. Shen, S. Wang, G. Sun, S. Jiang, and Q. Huang, 'Multi-level discriminative dictionary learning towards hierarchical visual categorization,' in Proc. IEEE CVPR, 2013, pp. 383–390.

Received: August 05 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[9] H. Wang, P. Wang, L. Song, B. Ren, and L. Cui, 'A novel feature enhancement method based on improved constraint model of online dictionary learning,' IEEE Access, vol. 7, pp. 17599–17607, 2019.

[10] Saini, Lalit Kumar, and Pratistha Mathur. “Medical image fusion by sparse-based modified fusion framework using block total least-square update dictionary learning algorithm.” Journal of medical imaging (Bellingham, Wash.) vol. 9,5 (2022): 052403. doi:10.1117/1.JMI.9.5.052403

[11] F. G. Veshki and S. A. Vorobyov, 'Coupled feature learning via structured convolutional sparse coding for multimodal image fusion,' in Proc. IEEE ICASSP, 2022.

[12] F.-P. An, X.-M. Ma, and L. Bai, 'Image fusion algorithm based on unsupervised deep learning-optimized sparse representation,' Biomed. Signal Process. Control, vol. 71, p. 103140, 2022

[13] Y. Liu et al., 'Image fusion with convolutional sparse representation,' IEEE Signal Process. Lett., vol. 23, no. 12, pp. 1882–1886, 2016.

[14] S. Xu et al., 'Deep convolutional sparse coding networks for image fusion,' arXiv:2005.08448, 2020.

[15] Y. Bengio, 'Learning deep architectures for AI,' Found. Trends Mach. Learn., vol. 2, no. 1, pp. 1–127.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, 'ImageNet classification with deep convolutional neural networks,' Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017.

[17] S. Hochreiter and J. Schmidhuber, 'Long short-term memory,' Neural Comput., vol. 9, no. 8, pp. 1735–1780.

[18] N. Ahmed, Z. A. Aghbari, and S. Girija, 'A systematic survey on multimodal emotion recognition using learning algorithms,' Intell. Syst. Appl., vol. 17, p. 200171, 2023.

[19] D. Ahmedt-Aristizabal, M. A. Armin, S. Denman, C. Fookes, and L. Petersson, 'A survey on graph-based deep learning for computational histopathology,' Computerized Med. Imag. Graph., vol. 95, p. 102027, 2022

[20] M. Tschannen, O. Bachem, and M. Lucic, 'Recent advances in autoencoder-based representation learning,' arXiv:1812.05069, 2018.

[21] B. Wei, X. Feng, K. Wang, and B. Gao, 'The multi-focus-image-fusion method based on convolutional neural network and sparse representation,' Entropy, vol. 23, p. 827, 2021.

[22] L. Chang et al., 'Image decomposition fusion method based on sparse representation and neural network,' Appl. Opt., vol. 56, no. 28, pp. 7969–7977, 2017.

[23] K. Vanitha, D. Satyanarayana, and M. N. G. Prasad, 'Medical image fusion based on deep decomposition and sparse representation,' in MIND 2020, CCIS, vol. 1240, Springer, 2020. https://doi.org/10.1007/978-981-15-6315-7_22.

Received: August 05 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[24] X. Qin et al., 'Improved image fusion method based on sparse decomposition,' Electronics, vol. 11, p. 2321, 2022.

[25] J. Zhang, A. Liu, D. Wang, Y. Liu, Z. J. Wang, and X. Chen, "Transformer-Based End-to-End Anatomical and Functional Image Fusion," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022, Art. no. 5019711, doi: 10.1109/TIM.2022.3200426.

[26] V. Vs, J. M. Jose Valanarasu, P. Oza, and V. M. Patel, "Image Fusion Transformer," in *2022 IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France, 2022, pp. 3566–3570, doi: 10.1109/ICIP46576.2022.9897280.

[27] Liu, Yu, et al. "Multi-focus image fusion with deep residual learning and focus property detection." Information Fusion 86 (2022): 1-16

[28] Hu, Xingyu, et al. "ZMFF: Zero-shot multi-focus image fusion." Information Fusion 92 (2023): 127-138.

[29] Wang, Zeyu, et al. "A self-supervised residual feature learning model for multifocus image fusion." IEEE Transactions on Image Processing 31 (2022): 4527-4542.

[30] S. Nirmalraj and G. Nagarajan, 'Fusion of visible and infrared image via compressive sensing using convolutional sparse representation,' ICT Express, vol. 7, no. 3, pp. 350–354, 2021.

[31] A. S. Yousif, Z. Omar, and U. U. Sheikh, 'An improved approach for medical image fusion using sparse representation and Siamese CNN,' Biomed. Signal Process. Control, vol. 72, Part B, p. 103357, 2022.

[32] P. Guo, G. Xie, R. Li, and H. Hu, 'Multi-modal image fusion via convolutional morphological component analysis and guided filter,' J. Circuits, Syst. Comput., vol. 30, no. 02, p. 2130003, 2021.

[33] M. Abavisani and V. M. Patel, "Deep multimodal sparse representation-based classification," in Proc. IEEE Int. Conf. Image Process. (ICIP), 2020, pp. 773–777, doi: 10.1109/ICIP40778.2020.9191317.

[34] O. Kechagias-Stamatis and N. Aouf, 'Fusing deep learning and sparse coding for SAR ATR,' IEEE Trans. Aerosp. Electron. Syst., vol. 55, no. 2, pp. 785–797, 2018. https://doi.org/10.1109/TAES.2018.2873804

[35] Qi, Biao, et al. "A novel saliency-based decomposition strategy for infrared and visible image fusion." *Remote Sensing* 15.10 (2023): 2624..

[36] H. H. Xia et al., 'Multi-focus microscopy image fusion based on Swin Transformer architecture,' Applied Sciences, vol. 13, no. 23, p. 12798, 2023

[37] X. Wang, Z. Hua, and J. Li, "Multi-focus image fusion framework based on transformer and feedback mechanism," Ain Shams Eng. J., vol. 14, no. 5, p. 101978, 2023.

[38] Z. Duan et al., 'Combining transformers with CNN for multi-focus image fusion,' Expert Syst. Appl., vol. 235, p. 121156, 2023.

Received: August 05 2025

Volume 38 No. 1s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[39] G. Zhang and J. Li, 'Multi-focus image fusion under strip-cross axis transformer,' in 2024 5th Int. Symp. Comput. Eng. Intell. Commun. (ISCEIC), IEEE, 2024.

[40] Hao, Chen-Yu, et al. "Using Sparse Parts in Fused Information to Enhance Performance in Latent Low-Rank Representation-Based Fusion of Visible and Infrared Images." Sensors 24.5 (2024): 1514.

[41] Hu, Qiu, et al. "Adaptive convolutional sparsity with sub-band correlation in the NSCT domain for MRI image fusion." Physics in Medicine & Biology 69.5 (2024): 055022.

[42] Ma, Zhaoxu, et al. "SFT-GAN: Sparse Fast Transformer Fusion Method Based on GAN for Remote Sensing Spatiotemporal Fusion." Remote Sensing 17.13 (2025): 2315.

Received: August 05 2025