International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 15 Issue 2 (2026) | Pages: 136-143

# LG Fusion++: Adaptive Nonlinear Frequency-Aware and Task-Driven Multimodal Image Fusion

J. Aajila[1], E. P. Kannan[2]

[1,2] Department of Electronics and Communication Engineering, Arunachala College of Engineering for Women
Manavilai, Kanyakumari district

**Peer Review Information**

***Type**: Article*
***Received**: 29 March 2026*
***Revised**: 25 April 2026*
***Accepted**: 02 June 2026*
***Published**: 23 June 2026*

**Abstract**

Multimodal image fusion aims to combine the complementary information from infrared, visible and medical imaging modalities to generate a single enhanced image. Existing methods based on wavelet transform and convolutional neural network are often not able to simultaneously preserve the global structural information and the local texture details well. In this paper, we propose LG Fusion++, an adaptive nonlinear frequency-aware fusion framework that can effectively separate low-frequency structural information and high-frequency texture details. The framework incorporates lightweight attention mechanisms and Mamba based context modeling for efficient global feature learning. A contrast-aware perceptual loss and multi-task optimization strategy are adopted to achieve better edge preservation, texture enhancement and downstream task performance. The experimental results indicate that the proposed framework outperforms existing approaches in terms of entropy, edge preservation, and structural similarity, making it suitable for surveillance, medical imaging, and autonomous systems.

**Keywords:** Multimodal Image Fusion; Infrared Imaging; Frequency-Aware Fusion; Mamba Modeling; Attention Mechanism; Medical Imaging

**How to Cite This Article**

Aajila, J., & Kannan, E. P. (2026). LG Fusion++: Adaptive nonlinear frequency-aware and task-driven multimodal image fusion. *International Journal on Advanced Computer Engineering and Communication Technology*, 15(2), 136–143.

## Introduction

Multimodal image fusion is an important area of research in image processing and computer vision, because it can combine complementary information from different imaging modalities into a single enhanced image. In various practical applications including surveillance, medical imaging, remote sensing, autonomous navigation and defense systems, a single imaging modality is typically not sufficient to provide complete scene information. For instance, visible images have rich color, texture, and structural details, but in bad lighting or bad weather, they function far worse. On the other hand, infrared photographs typically lack fine texture and comprehensive scene representation, but they do capture thermal radiation and offer great target visibility even in low-light conditions. Thus, fusing visible and infrared images can result in a fused image that maintains both detailed visual content and thermal information, improving interpretability and decision-making.

Conventional image fusion techniques mostly rely on transform-domain methods like Non-Subsampled Contourlet Transform (NSCT), Laplacian Pyramid, and Discrete Wavelet Transform (DWT). These techniques use pre-established fusion rules to merge images after breaking them down into frequency components. Even though these methods can maintain certain structural characteristics, they frequently have low robustness and little flexibility in challenging imaging scenarios.

Problems including blurring edges, decreased contrast, and loss of crucial information result from fixed fusion techniques' inability to accommodate fluctuations in illumination, contrast, noise, or modality-specific properties. Convolutional Neural Networks (CNNs) have become popular for multimodal picture fusion due to the quick development of deep learning. By automatically deriving hierarchical representations from data, CNN-based techniques enhance feature extraction. Compared to conventional methods, these techniques are better at maintaining edge information and local textures. However, CNNs frequently have trouble capturing long-range dependencies and global contextual linkages within the image since they mostly concentrate on local receptive fields. Because of this, the fused outputs might still have poor global feature representation and insufficient structural coherence.

Transformer-based fusion techniques have recently been developed to get around these restrictions. Transformers use self-attention processes to effectively simulate global contextual information and long-range dependencies. Transformer-based techniques are less appropriate for real-time and resource-constrained applications since they often need significant computing resources and large memory consumption, despite their great feature modeling capacity. Additionally, a lot of current deep learning fusion models are less able to strike a compromise between local texture preservation and global consistency since they do not explicitly distinguish between modality-specific details and shared structural information.

This research presents LG Fusion++, an adaptive nonlinear frequency-aware multimodal image fusion framework intended for medical and infrared-visible image fusion applications, as a solution to these problems. The suggested framework divides input images into high-frequency texture components and low-frequency structural components using a nonlinear frequency decomposition technique. The system adaptively learns how to integrate modality-specific data and shared structural information using learnable fusion weights, as opposed to predefined fusion rules. This allows the framework to effectively maintain both global and local image properties while constantly adapting to changing image conditions.

The suggested method also uses Mamba-based sequence modeling and lightweight attention techniques to effectively capture both local feature relevance and global contextual links. Additionally, a contrast-aware perceptual loss algorithm is added to enhance the fused image's overall perceptual quality, edge preservation, and texture visibility. Additionally, picture fusion and downstream tasks like segmentation and classification are concurrently optimized using a multi-task learning technique, increasing the fused output's semantic value.

### Related Work

Using deep learning, frequency-domain processing, and attention-based architectures, a number of academics have made substantial contributions to the fields of multimodal picture fusion and low-light image enhancement.

Cha et al. proposed a low-light image enhancement method based on Retinex theory with improved illumination map estimation [1].Their framework enhanced image brightness and preserved scene visibility by refining illumination correction and reducing uneven lighting effects. Although the method improved low-light performance, maintaining texture details under highly complex environments remained challenging.

Feng and Qi introduced a motion deblurring and image color enhancement framework using Wasserstein Generative Adversarial Networks (WGAN) [2]. Their method effectively restored blurred images while improving color representation and visual quality.

Guo et al. developed an adaptive partition-based quality evaluation method for night vision anti-halation fusion images [3]. Their study focused on designing an effective evaluation framework for assessing image fusion quality under nighttime conditions.

Gurbina et al. proposed a low-light image enhancement framework based on the absorption light scattering model [4]. Their approach improved image visibility by modeling environmental scattering effects and restoring image brightness.

Li et al. proposed an infrared and visible image fusion framework based on sparse representation and guided filtering in the Laplacian pyramid domain [5]. Their method decomposes source images into low-frequency and high-frequency components, where sparse representation is applied to preserve global structural information and guided filtering is used to enhance edge continuity and texture details

Parihar proposed a single-image dehazing framework using non-local atmospheric light estimation [6]. The model improved haze removal capability and enhanced image clarity by estimating atmospheric components more accurately. However, the method showed limitations in preserving edge details and minimizing color distortion.

Ren et al. developed a multi-scale convolutional neural network for single-image dehazing [7]. Their framework extracted multiscale features to restore visibility and improve image quality. Although the model achieved better restoration performance than traditional methods, it introduced increased computational complexity and limited long-range dependency modeling.

Hua et al. presented an image dehazing technique using near-infrared information integrated with the dark channel prior approach [8]. The proposed method effectively utilized complementary infrared information to improve visibility and remove haze.

Yan et al. proposed TGLFusion, a temperature-guided lightweight fusion framework for infrared and visible image fusion[9]. The method utilizes temperature-aware feature extraction to guide the fusion process and effectively preserve thermal targets while maintaining visible image texture information.

Recent multimodal fusion methods have also learned to integrate attention mechanisms and Transformer architectures for improved global context learning. Attention-based medical image fusion networks, e.g. AMMNet, use economical convolutional architectures with channel and spatial attention modules to effectively preserve diagnostic features. In a similar fashion, adaptive spatial-frequency expert fusion networks explicitly distinguish spatial and frequency information to improve texture preservation and structural consistency.

Despite promising results obtained by existing approaches, some challenges remain:

- Lack of balance between global structure and local texture preservation.
- Loss of contrast and edge information in difficult imaging condition
- Computationally expensive Transformer-based models.
- Limited robustness and generalization to multiple multimodal datasets.

In order to address these problems, LG Fusion++, our proposed framework, makes use of adaptive nonlinear frequency decomposition, light attention schemes, Mamba-based contextual modeling, and contrast perception learning methods.

**System Analysis**

*Existing System*

Classic methods like DWT (Discrete Wavelet Transform), LP (Laplacian Pyramid), Curvelet Transform, and NSCT (Non-Subsampled Contourlet Transform) decompose an image into several sub-bands based on a set of predetermined fusion strategies. Such approaches are effective from both computational complexity and implementation perspectives; nevertheless, there is a drawback that such algorithms do not adjust to different image features, leading to poor contrast and lost information within the resulting image. Recently, there is a great deal of interest in CNN based fusion schemes as they offer an effective framework for automatic feature learning from massive datasets. CNN based methods enhance the process of texture preservation and extraction of local features as compared to conventional techniques. Despite this, CNN frameworks emphasize local receptive field processing and are generally poor at modeling long range dependencies and capturing global context Transformer-based fusion techniques of images have been proposed to mitigate some of the issues with CNNs.

These techniques make use of self-attention schemes to capture global dependency between different regions of an image and deliver better fusion outcomes. Despite being effective in offering more context-awareness, Transformer models are computationally expensive, require high memory space, and consume a lot of time to train In addition to this, current approaches for image fusion use either static or semi-static techniques which are incapable of balancing both low frequency structural information and high frequency textual details. Due to this, there is an increased possibility that the resultant image will have blurred edges and missing fine details.

*Limitations Of Existing System*

- Utilize predetermined fusion operations
- Fail to efficiently retain structure and texture information at the same time.
- CNN approaches lack efficient understanding of global contextual information.

- Transformer-based approaches need excessive computing capabilities.
- Lower efficiency under difficult and low-light scenarios.
- High consumption of memory space and processing time..

*Proposed System*

The proposed LG Fusion++ framework tries to solve the problems associated with traditional multimodal image fusion techniques with its adaptive, frequency-sensitive, and efficient fusion algorithm. The LG Fusion++ framework is comprised of the following stages:

- Input Image Acquisition and Preprocessing
- Nonlinear Frequency Decomposition
- Adaptive Frequency Fusion
- Attention and Mamba-Based Context Modeling
- Contrast-Aware Perceptual Learning
- Multi-Task Learning Optimization
- Fused Image Reconstruction

The framework processes infrared and visible images to generate high-quality fused outputs with enhanced contrast and texture preservation.

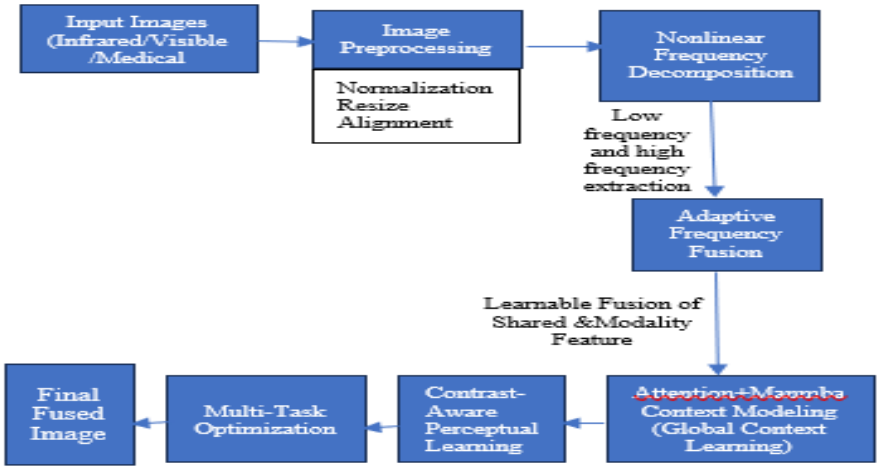


***Fig. 1.*** *Block Diagram of Proposed System*

Input Image Preprocessing

The preprocessing phase will enable the multimodal images to undergo proper fusion. The two input images, that is, the infrared image and the visible image, are resized to be of the same size. Normalization is performed on the images to make the intensity values more similar. Registration will align the two images to ensure proper fusion.The preprocessed images are represented as:

where $I_{IR}$ denotes the infrared image and $I_{VIS}$ denotes the visible image

Nonlinear Frequency Decomposition

The new approach breaks down the input images into Low frequency that carries overall structure of image data and High frequency that carries texture and details in images. In contrast to conventional frequency decomposition approaches that have fixed frequencies, LG Fusion++ incorporates nonlinearly convolution layers where frequency components are separated depending upon image features..

The decomposition process can be represented as:

$$F_L = D_L(I)$$

$$F_H = D_H(I)$$

where:

$F_L$ represents low-frequency structural information

$F_H$ represents high-frequency texture information

$D_L$ and $D_H$ are learnable decomposition functions

Adaptive Frequency Fusion

Adaptive fusion involves the dynamic integration of the features at low and high frequency of infrared and visual modalities. Fusion weights learned from the images decide the influence of each modality. Fusion features are enhanced by applying convolutions and non-linear activations. The adaptive fusion approach ensures that the fusion process captures the following characteristics of images: Shared global structure of low frequency Modality-specific textures and edges of high frequency.Adaptive fusion weights are learned automatically during training

$$F_{LF} = \alpha F_L^{IR} + (1-\alpha)F_L^{VIS}$$

$$F_{HF} = \beta F_H^{IR} + (1-\beta)F_H^{VIS}$$

where:

$\alpha$= structural fusion weight

$\beta$= texture fusion weight

Attention and Mamba-Based Context Modeling

Lightweight attention mechanisms focus on significant regions within the image like edges and salient objects by downplaying irrelevant data. Sequence modeling using Mamba is highly efficient in capturing long-distance dependencies and global relationships in context, unlike Transformers.

The combination of attention and Mamba modelling ensures:

- Improved global consistency.
- Enhanced contextual understanding.
- Efficient feature representation.

$$h_t = Ah_{t-1} + Bx_t$$

$$y_t = Ch_t + Dx_t$$

where:

$x_t$is the input feature sequence

$h_t$represents hidden states

$y_t$denotes output features

$A, B, C, D$are learnable parameters

Contrast-Aware Perceptual Learning

A contrast-aware perceptual loss function is used to preserve brightness consistency, texture details, and edge sharpness. High-frequency information is emphasized during training to maintain visual clarity. The perceptual learning module enhances edge preservation ,improves texture visibility and maintains structural consistency. Deep feature similarity is computed using pretrained convolutional networks to ensure perceptual quality beyond pixel-level similarity.

The total loss consists of three components:

1. Reconstruction Loss

Measures pixel-level similarity by:

$$L_{rec} = ||I_F - I_{GT}||_1$$

2. Perceptual Loss

Extracts semantic similarity using deep feature maps:

$$L_{perc} = ||\phi(I_F) - \phi(I_{GT})||_2$$

where $\phi$ denotes pretrained feature extraction layers.

3. Contrast Enhancement Loss

Preserves image contrast and edge sharpness:

$$L_{con} = 1 - \frac{\sigma(I_F)}{\sigma(I_{GT})}$$

The final objective function becomes:

$$L_{total} = \lambda_1 L_{rec} + \lambda_2 L_{perc} + \lambda_3 L_{con}$$

where:
$\lambda_1, \lambda_2, \lambda_3$are balancing coefficients.

This loss formulation encourages the model to generate visually appealing fused images with enhanced contrast and texture preservation.

Multi-Task Learning Optimization
The proposed framework jointly optimizes image fusion and downstream vision tasks such as:

- Object Detection
- Image Segmentation
- Image Classification

Shared feature representations improve generalization and robustness across different datasets. Multiple loss functions are combined into a unified optimization objective to balance fusion quality and semantic understanding.

Image Reconstruction
The final fused image is generated THROUGH a reconstruction network consisting of convolutional layers and residual blocks.

The reconstruction process combines:

- Adaptively fused low-frequency features
- Adaptively fused high-frequency features
- Global contextual information learned by Mamba blocks

The reconstructed fused image is represented as:

$$I_F = R(F_{LF}, F_{HF})$$

where $R(\cdot)$denotes the reconstruction module.

*Estimated Comparative Analysis*

***Table. 1.*** *Estimated Comparative Analysis*

| **Feature** | **Traditional Methods** | **CNN-Based Methods** | **Transformer Methods** | **Proposed LGFusion++** |
|---|---|---|---|---|
| Structural Preservation | Medium | Good | Excellent | Excellent |
| Texture Preservation | Medium | Good | Excellent | Excellent |
| Global Context Learning | Poor | Limited | Excellent | Excellent |
| Computational Cost | Low | Medium | Very High | Low-Medium |
| Adaptive Fusion | No | Partial | Yes | Yes |
| Real-Time Capability | High | Moderate | Low | High |
| Edge Preservation | Moderate | Good | Very Good | Excellent |

The comparison demonstrates that the proposed framework outperforms existing approaches in terms of fusion quality while maintaining lower computational complexity.

*Advantages of Proposed System*
The proposed LGFusion++ framework offers several advantages:

- Adaptive Nonlinear Frequency Decomposition Learns optimal frequency representations automatically and eliminates dependence on fixed decomposition filters.
- Provides improved structural and texture preservation.
- Mamba architecture provides efficient global context modeling by capturing long-range dependencies efficiently.
- Reduced memory requirements compared to Transformer-based methods.
- Contrast-aware perceptual optimization improves image appearance.
- Applicable to surveillance, medical imaging, autonomous vehicles, and remote sensing.

**Results And Discussion**

*Fusion Result*

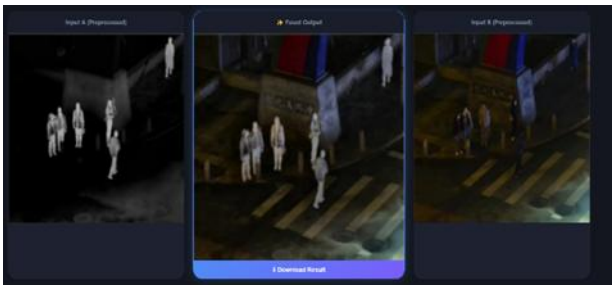

***Fig. 2.*** *Fusion result*

The fusion result shows the output of the LG Fusion++ model, where the preprocessed infrared image (left) and visible image (right) are combined to produce the fused output (center).

The fused image successfully preserves the thermal structural details from the infrared source while incorporating the color and texture information from the visible image, resulting in an enhanced representation of the road scene.

*Intermediate Feature Maps*


***Fig. 3.*** *Intermediate feature maps*

The intermediate feature maps visualize the internal processing stages of the LGFusion++ model, showing the low-frequency structural components and high-frequency texture/edge details extracted separately from both the infrared (A) and visible (B) input images.

The fused features and context features represent the progressively refined representations produced by the adaptive frequency fusion and Mamba-based sequence modeling modules respectively.

*Quality Metrics*

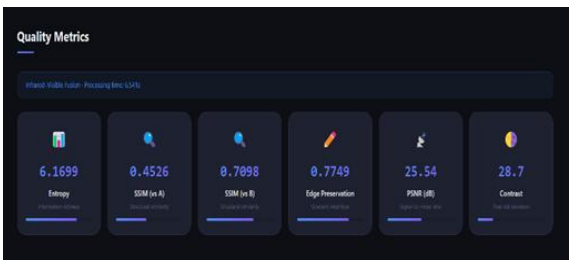


***Fig. 4.*** *Quality metrics*

The quality metrics evaluate the performance of the LGFusion++ fusion output using six standard measures — Entropy (6.97) confirms high information richness, SSIM values (0.64 and 0.40) indicate structural similarity with both source images, Edge Preservation (0.88) demonstrates strong retention of gradient details, PSNR (23.23 dB) reflects acceptable signal quality, and Contrast (31.22) shows adequate pixel variation in the fused result.

## Conclusion

The proposed LGFusion++ framework enhances multimodal image fusion by integrating complementary information from infrared–visible and medical images. The introduction of adaptive nonlinear frequency fusion allows the model to intelligently balance shared structural information and modality-specific details, resulting in more accurate and informative fused outputs. The incorporation of a contrast-aware perceptual loss helps to improve the quality of the fused images by enhancing texture visibility, preserving fine details, and maintaining essential structural information. In addition, the use of lightweight attention mechanisms combined with Mamba-based sequence modeling ensures computational efficiency while still capturing both local and global contextual information. This balance between performance and efficiency makes the proposed framework suitable for real-time and resource-constrained applications. Overall, LGFusion++ achieves superior fusion quality, better generalization, and improved robustness compared to existing methods, making it highly effective for applications in surveillance, medical imaging, and computer vision applications.

## Future Work

Future enhancements of the LG Fusion++ framework can be made by extending its capabilities to support a wider range of multimodal imaging types, such as hyperspectral images, satellite imagery, and other emerging sensing modalities. This will extend the applicability of the model to fields like remote sensing, environmental monitoring, and geospatial analysis. Further improvements can also be made to optimize the model for real-time image fusion applications. This involves reducing computational complexity, improving inference speed, and enhancing hardware efficiency to enable deployment on edge devices and embedded systems.Other direction for future research is the integration of more advanced deep learning architectures to further improve feature extraction and fusion performance. Incorporating novel architectures may help the model to capture more complex patterns and improve the quality of the fused output. Additionally, self-supervised and unsupervised learning approaches can reduce the dependency on labeled datasets, which are often limited in multimodal fusion tasks. These learning strategies can help the model learn more generalized and robust representations from unlabeled data, improving scalability and adaptability across diverse datasets. Overall, these future directions aim to further improve the performance, efficiency, and applicability of the LGFusion++ framework, making it more versatile and effective for a wide range of real-world applications.

## References


1. Cha W., Kim J., Kim Y. and Lee S. (2022), ‘Low-Light Image Enhancement Method Based on Retinex Theory by Improving Illumination Map’, Applied Sciences, Vol. 12, No. 10, Art. No. 5257.
2. Feng J. and Qi S. (2020), ‘Motion Deblurring in Image Color Enhancement by WGAN’, International Journal of Optics, Vol. 2020, Art. No. 1295028.
3. Guo Q., Chai G. and Li H. (2020), ‘Quality Evaluation of Night Vision Anti-halation Fusion Image Based on Adaptive Partition’, Journal of Electronics & Information Technology, Vol. 42, No. 7, pp. 1750–1757. doi:10.11999/JEIT190453.
4. Gurbina M., et al. (2021), ‘Low-Light Image Enhancement via the Absorption Light Scattering Model’, Pattern Recognition Letters, Vol. 150, pp. 37–43.
5. Li L., Shi Y., Lv M., Jia Z., Liu M., Zhao X., Zhang X. and Ma H. (2024), ‘Infrared and Visible Image Fusion via Sparse Representation and Guided Filtering in Laplacian Pyramid Domain’, *Remote Sensing*, Vol. 16, No. 20, Art. No. 3804.
6. Parihar A.S. (2021), ‘Single Image Dehazing Using Non Local Atmospheric Light Estimation’, Proceedings of ICICNIS 2020 (SSRN Electronic Journal), Art. No. 3769839.
7. Ren W., Liu S., Zhang H., Pan J., Cao X. and Yang M.H. (2016), ‘Single Image Dehazing via Multi-scale Convolutional Neural Networks’, In: Leibe B., Matas J., Sebe N. and Welling M. (eds), Computer Vision – ECCV 2016, Lecture Notes in Computer Science, Vol. 9906.
8. Hua Z., Ding Y. and Li J. (2021), ‘Image Dehazing Using Near-Infrared Information Based on Dark Channel Prior’, Procedia Computer Science, Vol. 187, pp. 18–23.
9. Yan B., Zhao L., Miao K., Wang S., Li Q. and Luo D. (2024), ‘TGLFusion: A Temperature-Guided Lightweight Fusion Method for Infrared and Visible Images’, Sensors, Vol. 24, No. 6, p. 1735.
10. Zhang H., Cheng Y. and Huang W. (2025), ‘SMFusion: Semantic-Preserving Fusion of Multimodal Medical Images for Enhanced Clinical Diagnosis’, IEEE Journal of Biomedical and Health Informatics, pp. 1–12.