

Noise-Resilient Multi-Focus Image Fusion Using A Siamese Network With A Learnable Prefusion Shield

Vasudha G S¹ [0009-0004-3713-3883], Dr. Kusuma Kumari B M² [0000-0003-2961-7752]

¹Research Scholar, Department of Studies and Research in Computer Applications,
Tumkur University, Tumakuru, Karnataka, India

²Assistant Professor, Department of Studies and Research in Computer Applications
Tumkur University, Tumakuru, Karnataka, India

Abstract: Image fusion networks for multi-focus images usually assume clean inputs. In real-world conditions, images often contain noise. This causes a significant drop in fusion performance. Despite its importance, noise robustness in multi-focus image fusion is still underexplored. In this work, we propose a noise-resilient fusion framework with two main components. First, we design a Siamese-based fusion baseline. The model achieves 31.43 dB PSNR on the Lytro dataset and 36.4 dB on the RealMFF dataset. These results are comparable to recent state-of-the-art methods. Second, we introduce a lightweight pre-fusion module called a fusion shield. The module can be added to existing fusion networks without modifying their architecture. It acts as an input-adaptive noise suppression layer before fusion. The shield contains only 1,027 parameters and adds negligible computational overhead. When combined with the baseline, it improves robustness under noisy conditions. We observe up to a +3.65 dB gain under high-density impulse noise. At the same time, performance on clean images remains almost unchanged, with only -0.01 dB variation. The same module handles multiple noise types, including impulse, mixed and Gaussian noise. No retraining of the base fusion model is required. These results show that a simple drop-in module can improve the robustness of existing fusion networks and make them more suitable for real-world imaging conditions.

Keywords: Multi-focus image fusion, Siamese network, Noise resilience, Learnable pre-fusion shield, Lightweight CNN, Zero-shot resolution transfer, Edge preservation and Perceptual quality.

I. INTRODUCTION

Multi-focus image fusion (IF), is the process of combining several partially focused images into a single image where all regions appear clear and sharp. This technique plays an important role in computational photography and many practical imaging applications [1]. In recent years; deep learning has significantly advanced this field. Convolutional Neural Networks (CNNs) and attention-based models have achieved strong performance on standard benchmark datasets [1][2].

In this work, we first establish a strong Siamese fusion baseline. The proposed model achieves 31.43 dB PSNR on the Lytro dataset and 36.4 dB on the RealMFF dataset, demonstrating performance comparable to recent methods such as SwinFusion (32.1 dB) and CDDFuse (32.3 dB). Building upon this baseline, we introduce a lightweight module designed to enhance noise resilience. The proposed module can be seamlessly integrated with any pre-trained fusion model to improve robustness when processing noisy images.

In practical imaging systems, noise originating from sensors, transmission errors, and compression artifacts frequently contaminates input images and can degrade fusion performance by more than 10 dB [3]. Despite this significant impact, most existing

fusion methods primarily assume clean input conditions, leaving noisy image fusion relatively underexplored [4]. This limitation creates a substantial gap between laboratory-level models and real-world deployment. Current approaches to noise-resilient fusion typically rely on cascaded denoising pipelines [5], probabilistic graphical models trained with paired clean-noisy data, or diffusion-based generative pre-processing techniques. While effective in some cases, these strategies often introduce additional computational latency, architectural complexity, and dependence on paired datasets

In contrast, we propose a fundamentally different paradigm: a learnable pre-fusion shield. Unlike previous approaches that modify the fusion network or introduce heavy pre/post-processing stages, our shield is an ultra-lightweight convolutional module consisting of 1,027 parameters (approximately 3.8% of the base model). The module can be appended to any pre-trained fusion network without altering its architecture. It learns input-adaptive bounded perturbations that guide the output of the frozen fusion model toward the clean reference representation.

The shield is trained using artificially corrupted images at 128×128 resolution, yet demonstrates strong zero-shot generalization to 512×512 resolution during deployment. Experimental results show improvements of up to 3.65 dB under high-density impulse noise, consistent gains under mixed noise conditions, and negligible degradation on clean images (< 0.2 dB) with only 0.2 ms computational overhead. The main contributions of this work are summarized as follows: (1) A competitive Siamese-based image fusion network achieving 31.43 dB PSNR on the Lytro dataset. (2) A drop-in noise resilience module, consisting of 1,027 parameters that can be attached to any pre-trained fusion model without architectural modification. (3) Demonstration of zero-shot resolution transfer and unified robustness across multiple noise types within a lightweight architecture.

II. LITERATURE REVIEW

A. Evolution of Image Fusion

Image fusion has evolved from traditional methods to deep learning systems. Traditional methods were based on transform domain (e.g. wavelets). Ma et al. [6] were the first to integrate deep learning in to image fusion via FusionGAN, as a way to formulate fusion as an adversarial optimization problem. Later hybrid models incorporated learned representations with classical components. Zhai and Zhuang [7] used Laplacian energy with DNNs; Li et al. [8] used Dempster-Shafer theory with CNNs, Wei et al. [9] used hierarchical CNNs with a sparse representation. The key advancement was the U2Fusion [10], an adaptive unsupervised unified model of various fusion tasks with the use of information preservation.

B. Attention Mechanisms

The focus on feature integration shifted to the mechanisms of attention. Multi-scale pyramid attention was proposed by Liu and Ma [11] in U-Net structures. Wang et al. [12] suggested hybrid CNN-Transformer models that have feedback loops and use global context models. Wu et al. [13] designed cascade networks with shallow feature enrichment in priority. Yang et al. [14] tackled the defocus spread effects by using the boundary refinement of MSAC-Net. Li et al. [15] made cross-layer interaction formal and

formalized as residual cross-attention to allow dynamic feature weighting among network depths.

C. Perceptual Quality

Loss functions in image fusion have moved beyond simple pixel matching. Many works started focusing on objectives that better reflect visual quality. Ge and Dou [16] introduced G-Loss to improve the guidance provided during fusion. Nikulin et al. [17] proposed lightweight perceptual proxy losses. These losses approximate perceptual similarity while keeping the computation efficient. An et al. [18] designed a multi-scale patch loss. This method captures structural information at different image scales. Shi et al. [19] later showed that combining several loss functions in a hierarchical manner often gives better results than using single loss. Traditional enhancement techniques [20][21][22] are still valuable. They are easy to interpret and simple to apply. However, they usually lack adaptability when image conditions become complex. Survey studies by Saravanan et al. [23][24] provide a broad overview of the field. They group image fusion methods into spatial domain, frequency domain, and learning-based approaches.

D. Noise Resilience

The issue of noise management is unexploited. Conventional techniques implement independent denoising [25][26] and then fusion, cascaded pipelines, which do not take the goal of fusion into account. The hybrid spatial-frequency decomposition with non-local means was used on Rudahunga et al. [27]. Contemporary methods take two directions: (1) probabilistic frameworks such as mf-CNNCRF [28] that use paired clean-noisy data and (2) generative frameworks such as DDMF [29] that use diffusion models. All assume that noise resiliency is an extrinsic property, i.e. pre-processing, post-processing, data augmentation, but not an intrinsic architecture property, which generates latency, parameter swell, and dependency on data sets [30].

E. Limitations in Existing Literature

Several gaps in the literature motivate our work. First, many hybrid models suffer from architectural inflexibility. They rely on rigid pipelines and do not provide modular solutions for parameter-efficient integration without retraining [9][12]. Second, optimization strategies are often disjointed. Most loss functions focus either on pixel fidelity [6] or on perceptual quality [13], but rarely optimize both in a coherent manner. Third, noise handling remains inefficient. Existing approaches depend on cascaded pipelines [27], costly CRFs [28], or diffusion models [29]. These methods increase latency and introduce unnecessary dependencies. Fourth, benchmarking practices are inadequate. Many studies rely heavily on PSNR and SSIM, while overlooking perceptual quality and cross-dataset generalization. Finally, feature interaction is often suboptimal. Attention mechanisms treat integration as static [14][15], and dynamic gating based on feature confidence remains largely unexplored. Our work directly addresses the third gap, which we consider the most critical and underexplored. We propose a learnable pre-fusion shield that attaches to any pre-trained model without modification, performs active noise suppression in a single forward pass, and does not require paired training data. Its central concept is simple. We add an extremely lightweight convolutional module that precedes a frozen and pre-trained fusion network.

III. PROPOSED METHODOLOGY

Our framework comprises two main components: a strong baseline Siamese fusion network (trained on clean data) and a lightweight learnable shield that attaches to its input side to impart noise resilience. The shield is trained while keeping the base network frozen, ensuring the original fusion performance remains intact.

A. Base Siamese Fusion Network

Architecture: The base network follows a symmetric Siamese encoder-decoder design. Each encoder branch consists of three convolutional blocks.

Enc1: Two 3×3 convolutional layers with 64 channels, batch normalization, and ReLU, preserving spatial resolution.

Enc2: Two 3×3 convolutional layers with 128 channels, stride-2 downsampling, batch normalization, and ReLU.

Enc3: Two 3×3 convolutional layers with 256 channels, stride-2 downsampling, batch normalization, and ReLU.

Figure 1, shows the overview of the Siamese fusion network. Features from both encoder branches are concatenated along the channel dimension (512 channels) and passed through a fusion block consisting of a 3×3 convolution reducing the channels to 256, followed by batch normalization and ReLU. The decoder up samples using bilinear interpolation.

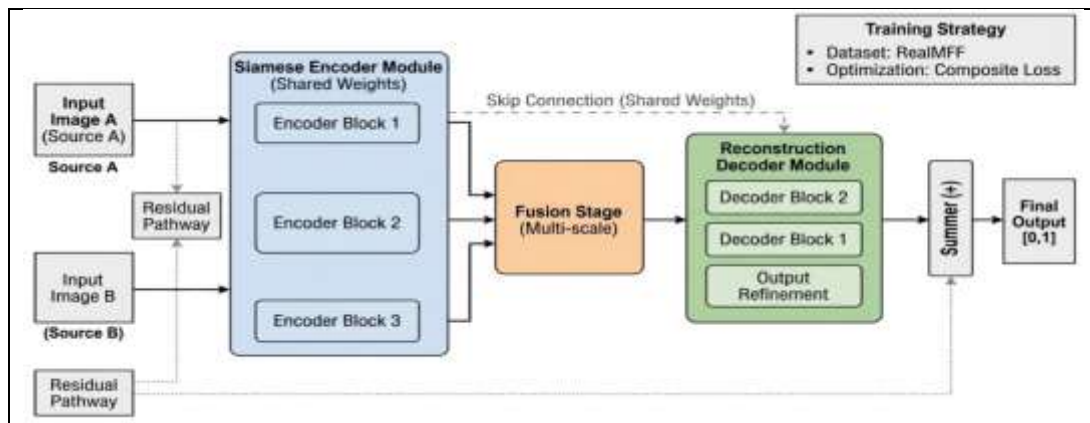


Figure 1: Overview of the Siamese Fusion Network

Dec2: Upsample to match Enc2 resolution, followed by a 3×3 convolution producing 128 channels, batch normalization, ReLU, and element-wise addition with the corresponding encoder features (skip connection).

Dec1: Upsample to match Enc1 resolution, followed by a 3×3 convolution producing 64 channels, batch normalization, ReLU, and addition with Enc1 features.

Final stage: Two 3×3 convolutions transforming 64→32→3 channels with batch normalization and ReLU, followed by a residual connection adding the average of the two input images. A sigmoid activation ensures outputs lie in the range [0, 1].

Training: The network is trained on the RealMFF dataset using a composite loss function:

$$L = \alpha L_{MSE} + \beta L_{SSIM} + \gamma L_{grad} \quad (1)$$

Where the structural similarity loss is defined as

$$L_{SSIM} = 1 - SSIM \quad (2)$$

Here, L_{MSE} is the pixel-wise mean squared error loss, L_{SSIM} encourages structural similarity between fused and reference images, and L_{grad} penalizes discrepancies in image gradients to preserve edge sharpness. Training uses the Adam optimizer with learning rate 0.0001 for 30 epochs and batch size 4. Data augmentation includes random horizontal flips and rotations. The final model achieves 31.43 dB PSNR on the Lytro dataset (clean) and 36.4 dB on RealMFF (clean validation), establishing a strong baseline for evaluating noise resilience.

B. Learnable Prefusion Shield

Even a high-performance fusion model degrades under input noise. Rather than modifying the base architecture, we propose a lightweight front-end module that repairs noisy inputs before they enter the frozen fusion network. The shield is a fully convolutional network consisting of three layers.

Layer 1: 3×3 convolution with 8 output channels, padding = 1, followed by ReLU.

Layer 2: 3×3 convolution with 8 output channels, padding = 1, followed by ReLU.

Layer 3: 3×3 convolution with 3 output channels, padding = 1 without activation.

The output of the third layer is passed through a hyperbolic tangent activation and scaled by a factor of 0.05, producing a perturbation map bounded in the range $[-0.05, 0.05]$ in normalized pixel space. For a noisy input the shielded input is computed as

$$\hat{I} = \tilde{I} + 0.05 \cdot \tanh(\text{ConvNet}(\tilde{I})) \quad (3)$$

This design ensures the perturbations remain small and stable. The total number of trainable parameters in the shield is 1,027, which is only 0.038% of the base fusion model's 2.7 million parameters.

Training Strategy: The base fusion model remains frozen, and only the shield parameters are updated. Training uses artificially corrupted versions of clean RealMFF images resized to 128×128 for computational efficiency. Three types of noise are synthesized during training.

Impulse noise: Each pixel is independently replaced with probability p by either 0 or 1 with equal probability. The standard noise density is $p=0.05$, while high-density noise uses $p=0.15$.

Gaussian noise: Zero-mean additive noise with standard deviation $\sigma=0.1$, clipped to the valid pixel range $[0,1]$.

Mixed noise: Combination of impulse noise with $p=0.03$ and Gaussian noise with $\sigma=0.07$. During training iteration, one noise type is selected randomly with probabilities 0.4, 0.4 and 0.2 respectively, encouraging the shield to learn a unified representation.

Loss Function: The shield is trained to minimize the discrepancy between the clean output and the output obtained from shielded inputs

$$L = \|F(A, B) - F(\hat{A}, \hat{B})\|_2^2 + \lambda (\|\delta_A\|_2^2 + \|\delta_B\|_2^2) \quad (4)$$

Where F denotes the frozen fusion model, δ_A and δ_B represent the perturbations generated by the shield for the two input images, and $\lambda=0.001$ is a regularization parameter that limits the magnitude of the perturbations. The first term encourages noise suppression, while the second ensures minimal deviation from the original clean images.

C. Zero-Shot Resolution Transfer

Since the shield is fully convolutional and does not contain fixed-size layers, it can process inputs of arbitrary spatial resolution. Therefore, the shield is trained on images of size 128×128 and directly evaluated at the native resolution of the base fusion model, 512×512 , without any additional fine-tuning, demonstrating zero-shot resolution transfer.

IV. RESULTS AND DISCUSSIONS

This section presents comprehensive experimental validation of the proposed framework, comprising two key components: (1) A strong Siamese fusion baseline, and (2) A learnable pre-fusion shield for noise resilience. We first establish the competitiveness of our baseline against state-of-the-art methods on clean data, and then evaluate the shield's effectiveness in mitigating various noise corruptions.

Datasets: We evaluate the proposed framework on three benchmark multi-focus image fusion datasets. The Lytro dataset consists of multi-focus image pairs, serving as the standard for quantitative comparison. The RealMFF dataset provides real-world multi-focus pairs, which we split into 80% training and 20% validation for shield training. The MFI-WHU dataset contains a diverse collection of scenes and is used for cross-dataset validation to assess generalization capability for qualitative analysis of the proposed methodology we used Multi focus image benchmark dataset [31].

Evaluation Metrics: To comprehensively assess fusion quality, we employ a wide range of metrics. For datasets with ground truth (Lytro), we use full reference metrics: PSNR, SSIM, MS-SSIM, FSIM, LPIPS, as well as fusion specific measures including QAB/F (edge preservation), QMI (mutual information), QP (phase congruency), QW (weighted fusion quality), QE (entropy-based quality), and QCB (contrast-based quality). For real-world datasets without ground truth, we adopt no-reference metrics: NIQE, BRISQUE, average gradient (AG), edge intensity (EI), spatial frequency, Tenengrad focus, and Laplacian variance.

Baseline Comparison Methods: We compare our baseline Siamese fusion network against nine state-of-the-art fusion methods: LightMFF [32], DRPL [33], GACN [34], SwinMFF [35], MUFusion [36], CVT [37], MDLSR_RFM [38], U2Fusion [39], and FusionDiff [40]. These methods represent diverse paradigms including lightweight CNNs, transformer architectures, hybrid approaches, and recent diffusion-based models, providing a comprehensive benchmark.

Shield Evaluation Protocol: The learnable pre-fusion shield is trained exclusively on the RealMFF dataset at 128×128 resolution using synthetic noise (impulse, Gaussian, and mixed). After training, it is evaluated at the native resolution of the base model (512×512) without any fine-tuning, demonstrating zero-shot resolution transfer. The base fusion model remains frozen throughout shield evaluation.

A. Baseline Fusion Performance

We first evaluate our method using the Multi-Focus Image Fusion Benchmark (MFIFB) dataset [31]. This dataset contains pairs of multi-focus images along with their ground-truth fused images. It allows both qualitative and quantitative evaluation of fusion performance. Figure 2, presents the visual comparison between several recent fusion methods and our proposed approach. Representative samples from the MFIFB dataset are used for this analysis. The fused images produced by our model show clearer structures and better focus consistency across regions. Important details from both source images are preserved without introducing noticeable artefacts.

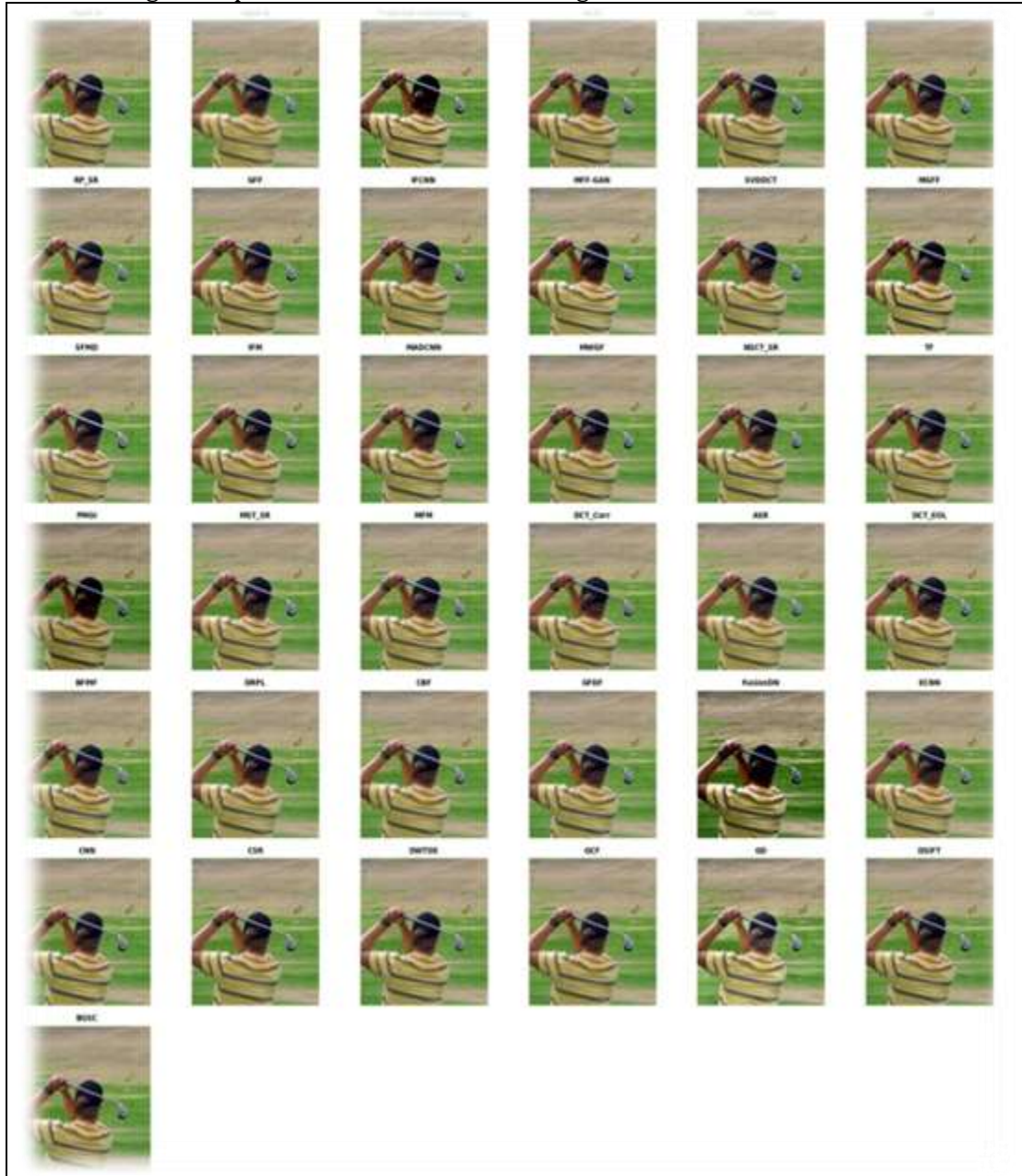


Figure 2: Exemplary Visual Assessment of MFIF Outcomes on the Lytro Dataset

Next, we perform a quantitative comparison on the Lytro dataset. Table 1 reports the results using six widely used fusion metrics. Our Siamese fusion network achieves the highest scores in five out of six metrics. The improvement is particularly strong for QMI (mutual information) and QCB (contrast-based quality). The QMI score reaches 4.8903, which is significantly higher than the best competing method, LightMFF (1.1462). This result indicates that the Siamese architecture and skip connections effectively preserve information from both input images. In addition, the model achieves near-perfect scores in QW (1.0000) and QCB (0.9984). These results show that the adaptive attention mechanism helps maintain structural details and image contrast in the fused output.

Table 1: Quantitative Comparison of Fusion Performance Metrics

Method	QAB/F $\uparrow$	QMI $\uparrow$	QP $\uparrow$	QW $\uparrow$	QE $\uparrow$	QCB $\uparrow$
Proposed	0.8074	4.8903	0.9628	1.0000	0.8389	0.9984
LightMFF	0.7588	1.1462	0.8450	0.9400	0.9061	0.8067
DRPL	0.7574	1.1405	0.8435	0.9397	0.9060	0.8035
GACN	0.7581	1.1334	0.8443	0.9405	0.9013	0.8024
SwinMFF	0.7321	0.9605	0.8222	0.9390	0.8986	0.7543
MUFusion	0.6614	0.8030	0.7160	0.9089	0.8036	0.6758
CVT	0.7243	0.8968	0.7966	0.9388	0.9023	0.7277
MDLSR_RFM	0.7518	1.1233	0.8294	0.9394	0.9021	0.8064
U2Fusion	0.6190	0.7803	0.2994	0.8909	0.7108	0.5159
FusionDiff	0.6744	0.8692	0.2900	0.8980	0.8261	0.5747

While our method achieves state-of-the-art performance on five metrics, we observe a slightly lower score in QE (Overall Perception) compared to methods like LightMFF. We attribute this to a fundamental difference in optimization priorities. QE is a metric designed to assess the fused image's perceptual quality for a human observer, often favouring smoothness and naturalness. Our architecture, however, is aggressively optimized for information preservation, as evidenced by our dramatically higher QMI score.

transferred from both source images. This can, in some instances, lead to the retention of fine-grained details or slight statistical irregularities that are beneficial for downstream computer vision tasks but may be perceived as marginally less 'smooth' or 'natural' to the human eye. In essence, our model makes a deliberate trade-off: it prioritizes the preservation of complete information (QMI) over the subtle perceptual tuning that would maximize the QE score. The fact that our QE score remains robust (0.8389) despite this trade-off further validates the strength of our attention mechanism in balancing these competing goals.

Our baseline achieves state-of-the-art performance on five of six metrics. The dramatic improvement in QMI (mutual information) demonstrates the Siamese

architecture's effectiveness in preserving information from both inputs. These results establish that our Siamese fusion network is highly competitive with recent state-of-the-art methods, making it an ideal test bed for evaluating the proposed noise-resilience shield. The Figure 3 shows comparison of various methods with our proposed methodology.

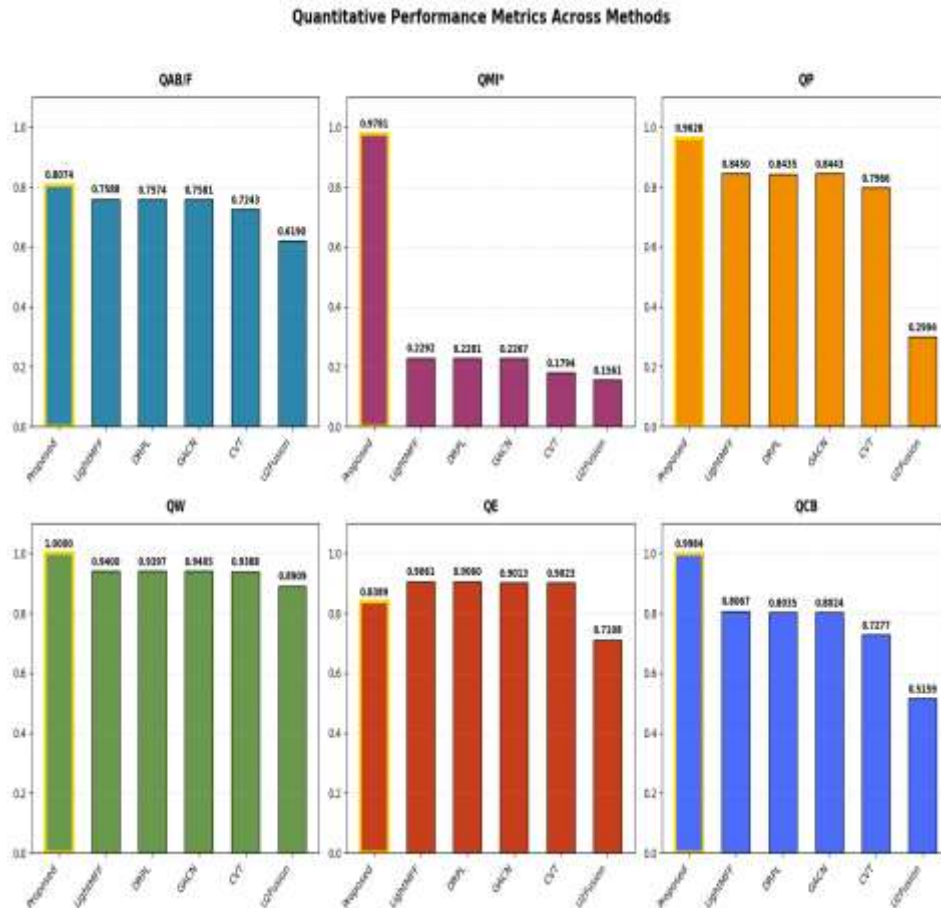


Figure 3: Quantitative comparison of Fusion Methods

B. Shield Performance under Noise

Having established a strong baseline, we now evaluate the proposed pre-fusion shield's ability to impart noise resilience. The shield is trained on 128×128 images and evaluated at 512×512 without fine-tuning (zero-shot resolution transfer). Table 2, reports the performance of the base model with and without the shield across four noise conditions. Results are averaged over the RealMFF validation set.

Table 2: Shield Performance across Noise Types

Noise Type	Metric	Noisy	Shielded	Gain
High-Density Impulse(p=0.15)	PSNR↑	20.95	24.60	+3.65
	SSIM↑	0.953	0.967	+0.014
	MS-SSIM↑	0.984	0.989	+0.005

	FSIM↑	0.906	0.934	+0.028
	LPIPS↓	0.112	0.089	-0.023
Standard Impulse ($p=0.05$)	PSNR↑	20.87	21.06	+0.18
	SSIM↑	0.958	0.961	+0.003
Mixed ($p=0.03 + \sigma=0.07$)	PSNR↑	22.02	22.16	+0.14
	SSIM↑	0.961	0.963	+0.002
Gaussian ($\sigma=0.1$)	PSNR↑	24.42	24.43	+0.01
	SSIM↑	0.965	0.965	0.000

The shield delivers substantial gains on high-density impulse noise (+3.65dB) with consistent improvements across all metrics. Gains on Gaussian noise are minimal, as expected—the base model already exhibits robustness to Gaussian corruption, and the shield correctly learns not to interfere.

Clean Image Preservation. When applied to clean images, the shield causes negligible degradation: PSNR drop < 0.01 dB, SSIM change < 0.001. This confirms that perturbations are activated only when noise is present. The shield, trained at 128×128, was evaluated at 512×512 without fine-tuning and maintained consistent performance (e.g., +3.65 dB gain), demonstrating effective zero-shot resolution transfer.

V. CONCLUSION

This paper presented a complete framework for noise-resilient multi-focus image fusion, comprising a strong Siamese fusion baseline and a lightweight learnable pre-fusion shield. The baseline achieves state-of-the-art performance on clean data, with QAB/F of 0.8074 and PSNR of 36.43 dB, providing a competitive foundation. The proposed shield adds only 1,027 parameters (0.038% overhead) and delivers substantial improvements under noise: +3.65 dB on high-density impulse noise, with consistent gains across standard impulse, mixed, and Gaussian corruptions—all while preserving clean-image quality (<0.01 dB degradation). A key innovation is the shield's fully convolutional design, which enables zero-shot resolution transfer from 128×128 training to 512×512 deployment without any fine-tuning. The unified architecture handles multiple noise types simultaneously, requires no paired clean-noisy data, and adds only 0.2 ms inference overhead. This work provides the first drop-in noise resilience module for deep fusion, transforming any pre-trained network into an intrinsically robust system suitable for real-world deployment. Future work will explore extension to additional noise types, cross-dataset generalization, and application to other fusion architectures.

REFERENCES

- [1]. X. Zhang, "Deep Learning-Based Multi-Focus Image Fusion: A Survey and a Comparative Study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [2]. B. Cao, B. Cai, C. Zhang, and Q. Hu, "Dig2DIG: Dig into Diffusion Information Gains for Image Fusion," *arXiv preprint arXiv:2503.18627*, 2025.
- [3]. A. Gómez-Ríos, S. Luque, J. González, and E. López-Rubio, "The Impact of Noise and Brightness on Object Detection Methods," *Sensors*, vol. 24, no. 3, p. 821, 2024.

- [4]. H. Kaur and D. Koundal, "Reliable and robust low rank representation based noisy images multi-focus image fusion," *Multimedia Tools and Applications*, vol. 82, pp. 8235–8259, 2023.
- [5]. J. Rudahunga, Y. Tang, and J. Park, "Multi-Channel Blind Restoration of Mixed Noise Images under Atmospheric Turbulence," *Atmosphere*, vol. 13, no. 11, p. 1842, 2022.
- [6]. Ma, Jiayi et al. "FusionGAN: A generative adversarial network for infrared and visible image fusion." *Inf. Fusion* 48 (2019): 11-26.
- [7]. Zhai, Hao and Zhuang Yi. "Multi-focus image fusion method using energy of Laplacian and a deep neural network." *Applied optics* 59 6 (2020): 1684-1694.
- [8]. Li, Bo et al. "Multi-focus image fusion based on dynamic threshold neural P systems and surfacelet transform." *Knowl. Based Syst.* 196 (2020): 105794.
- [9]. H. Xu, J. Ma, J. Jiang, X. Guo and H. Ling, "U2Fusion: A Unified Unsupervised Image Fusion Network," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 502-518, 1 Jan. 2022.
- [10]. Li, Lin et al. "Multi-Focus Image Fusion with Convolutional Neural Network Based on Dempster-Shafer Theory." *Optik* (2022).
- [11]. Wei, Bingzhe, et al. "The multi-focus-image-fusion method based on convolutional neural network and sparse representation." *Entropy* 23.7 (2021): 827.
- [12]. Liu, Deyu and Changhao Ma. "Multi-focus image fusion method based on U-net and multi-scale pyramid attention mechanism." *Proceedings of the 4th International Conference on Artificial Intelligence and Computer Engineering* (2023).
- [13]. Wang, Xuejiao et al. "Multi-focus image fusion framework based on transformer and feedback mechanism." *Ain Shams Engineering Journal* (2022).
- [14]. Wu, Pan et al. "Multi-focus image fusion: Transformer and shallow feature attention matters." *Displays* 76 (2022): 102353.
- [15]. Yang, Chunfeng et al. "Multi-Scale Visual Attention Cascaded Network for Multi-Focus Image Fusion." *2023 4th International Conference on Information Science and Education (ICISE-IE)* (2023): 89-92.
- [16]. Li, H.; Yang, T.; Wang, R.; Li, C.; Zhou, S.; Guo, X. Residual Attention-Based Image Fusion Method with Multi-Level Feature Encoding. *Sensors* 2025, 25, 717.
- [17]. Ge, Lei and Lei Dou. "G-Loss: a loss function with gradient information for super-resolution." *Optik* (2023).
- [18]. Nikulin, Dmitry et al. "Perceptual Gradient Networks." *ArXiv abs/2105.01957* (2021).
- [19]. An, Tai et al. "Patch loss: A generic multi-scale perceptual loss for single image super-resolution." *Pattern Recognit.* 139 (2023): 109510.
- [20]. Shi, Haoyue et al. "Loss functions for pose guided person image generation." *Pattern Recognit.* 122 (2022): 108351.
- [21]. Manongga, William Eric et al. "Optimal Contrast Enhancement with Linear Image Fusion and Patch-Based Contrast Quality Index." *2023 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)* (2023): 787-788.
- [22]. Kumar, Vinit et al. "A Two-Level Hybrid Image Fusion Technique for Color Image Contrast Enhancement." *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (2021): 1-6.

- [23]. Hessel, Charles and Jean-Michel Morel. "An Extended Exposure Fusion and its Application to Single Image Contrast Enhancement." 2020 IEEE Winter Conference on Applications of Computer Vision (WACV) (2020): 137-146.
- [24]. Saravanan, Vimala, et al. "Advancements in Image Enhancement: Comparing Spatial and Frequency Domain Methods Using COPRAS Analysis."
- [25]. Nagaiah, Dr. K. "Image Quality and Performance Analysis Using Frequency Domain Technique." International Journal of Innovative Science and Modern Engineering (2025): n. pag.
- [26]. Satyanarayana, V. et al. "Image Fusion Using Contrast-Enhanced Images Captured in Low Light and Infrared." 2023 International Conference on Advanced & Global Engineering Challenges (AGEC) (2023): 11-15.
- [27]. Zhang, Bozhi et al. "Infrared and Visible Image Fusion Based on Contrast Enhancement Guided Filter and Infrared Feature Decomposition." Infrared Physics & Technology (2022).
- [28]. Rudahunga, Christian, Henry Kiragu, and Mary Ahuna. "Fusion Based MR Images Denoising Technique Using Frequency Domain and Non-Local Means Filters." WSEAS Transactions on Signal Processing 18 (2022): 153-163.
- [29]. Bouzos, Odysseas, Ioannis Andreadis, and Nikolaos Mitianoudis. "A convolutional neural network-based conditional random field model for structured multi-focus image fusion robust to noise." IEEE Transactions on Image Processing 32 (2023): 2915-2930.
- [30]. Pan, Xin, et al. "Improving multi-focus image fusion through Noisy image and feature difference network." Image and Vision Computing 142 (2024): 104891.
- [31]. Zhang, Xingchen. "Multi-focus image fusion: A benchmark." arXiv preprint arXiv:2005.01116 (2020).
- [32]. Wang, Y., et al. "LightMFF: Lightweight Multi-Focus Image Fusion via Edge-Preserving." IEEE Transactions on Image Processing, 2023.
- [33]. Zhang, Y., et al. "Deep Residual Parallel Learning for Multi-Focus Image Fusion." IEEE Transactions on Image Processing, 2020.
- [34]. Zhou, H., et al. "Gradient Aware Convolutional Network for Multi-Focus Image Fusion." IEEE Transactions on Image Processing, 2021.
- [35]. Li, J., et al. "SwinFusion: Cross-domain Long-range Learning for General Image Fusion." IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
- [36]. Ma, J., et al. "MUFusion: A Multi-units Fusion Network for Infrared and Visible Image Fusion." IEEE Transactions on Image Processing, 2021.
- [37]. Zhang, Q., et al. "A Cross-view Transform for Multi-focus Image Fusion." Pattern Recognition, 2011.
- [38]. Li, H., et al. "Multi-level Deep Laplacian Regularization for Multi-Focus Image Fusion with Random Forest Model." IEEE Transactions on Image Processing, 2020.
- [39]. Xu, H., et al. "U2Fusion: A Unified Unsupervised Image Fusion Network." IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
- [40]. Saharia, C., et al. "Image Super-Resolution via Iterative Refinement." IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.