Multi-Focus Image Fusion Using Soft Decision Maps and Dual Fusion Rules

Braulio Lopez-Morales , Luis M. Ledesma-Carrillo * , Sebastián Salazar-Colores , Misael Lopez-Ramirez , Carlos Rodriguez-Donate , Eduardo Cabal-Yepez

Posted Date: 10 June 2026

doi: 10.20944/preprints202606.0829.v1

Keywords: multi-focus image fusion; local focus measurement; soft decision map; dual fusion rules

 doi:10.20944/preprints202606.0829.v1

Article

Multi-Focus Image Fusion Using Soft Decision Maps and Dual Fusion Rules

Braulio Lopez-Morales [1], Luis M. Ledesma-Carrillo [1,*], Sebastian Salazar-Colores [2], Misael Lopez-Ramirez [1], Carlos Rodriguez-Donate [1] and Eduardo Cabal-Yepez [1]

[1] Departamento de Estudios Multidisciplinarios, División de Ingenierías, Campus Irapuato-Salamanca, Universidad de Guanajuato, Yuriria, Guanajuato, 38944, México

[2] Centro de Investigaciones en Óptica A.C. (CIO), León, Guanajuato, 37150, México

* Correspondence: lm.ledesma@ugto.mx

Abstract

Multi-Focus Image Fusion is a technique that combines the in-focus regions from multiple images of the same scene, captured at different focal planes, to generate a single image with an extended Depth of Field compared to the source images. Despite improved fusion quality, recent approaches are constrained by high computational cost and the need for large training datasets, reducing their feasibility in resource-limited scenarios. In this work, a multi-focus fusion method based on a soft decision map is proposed, where the map is generated from local focus measurements in each scene. Subsequently, two Fusion Rules are employed: the Additive Fusion Rule, commonly used in the state of the art, and a newly proposed Multiplicative Fusion Rule. The method is evaluated using the public MFI-WHU and Lytro datasets through both reference-based and non-reference metrics. The results demonstrate that the proposed method produces visually coherent fused images and achieves competitive performance in terms of structural similarity, edge preservation, and information content, while maintaining low computational complexity.

Keywords: multi-focus image fusion; local focus measurement; soft decision map; dual fusion rules

1. Introduction

In recent years, conventional optical systems used to capture images of a scene for subsequent analysis have improved significantly. However, one of their main limitations remains the difficulty of achieving an extended Depth of Field without compromising resolution [1–3]. Among the most widely used techniques to address this issue is Multi-Focus Image Fusion (MFIF), which aims to integrate multiple images of the same scene captured at different focal planes [4]. In this context, the ability to combine high-resolution information from different regions is essential, as it helps mitigate the loss of relevant information in subsequent analysis or recognition tasks, particularly in applications such as microscopy, industrial inspection, robotic vision, and automated monitoring systems [5–8].

Traditionally, low-computational-cost and easy-to-implement MFIF methods have relied on the use of Focus Measure Operators (FMO) to estimate the Local Focus Measurement (LFM) in the input images. These measurements, combined through pixel-wise decision rules, determine the fusion process [4,9]. Among the most commonly used operators for estimating local focus are those based on the spatial domain, such as gradients, variance, and energy, as well as those in the frequency domain, including the Fourier transform, discrete cosine transform, wavelet transform, and other multiscale transforms [10–12]. These methods are well suited for real-time applications due to their efficiency; however, when Fusion Rule (FR) rely on binary decisions, i.e., selecting entirely the pixel from the source image with the higher focus measure, they may produce abrupt transitions and artifacts in regions where both images exhibit similar levels of sharpness [13]. Subsequently, methods based on multiscale transforms, guided image filtering, cross bilateral filtering, and Gaussian

 doi:10.20944/preprints202606.0829.v1

curvature filtering, among others, have been developed to improve spatial consistency and edge preservation. In this context, Ming et al. propose an MFIF method in the Nonsubsampled Contourlet Transform domain (NSCT) domain[14], where a distance-weighted regional energy measure is used for fusing low-frequency components, whereas a tensor-based structure operator is employed to process high-frequency components, enabling better preservation of details and edges. Additionally, a consistency verification mechanism is incorporated to enhance the robustness of focused region selection. Similarly, Wan et al. develop a multiscale decomposition-based approach that combines energy measures for low-frequency components with a parameter-adaptive pulse-coupled neural network (PA-PCNN) model for high-frequency components [15], effectively reducing artifacts in transition regions. Chen et al. [16], introduce an MFIF algorithm, noteworthy because of its computational efficiency and suitability for real-time processing, based on multiscale feature aggregation, where the input images are first downsampled to estimate local defocus levels using difference-of-Gaussians and Laplacian filters; then, they are upsampled before applying the final decision map. Although these methods represent significant improvements in mitigating abrupt transitions and artifacts in regions with similar sharpness levels, the use of multiscale transforms in most of these approaches considerably increases their computational cost.

Recently, deep-learning-based approaches have shown competitive results by modeling fusion as either a supervised or unsupervised learning problem[17,18]. In this context, Zhang et al. propose an end-to-end image fusion framework based on convolutional neural network (IFCNN) [19], in which features extracted from source images are fused using element-wise rules and subsequently reconstructed to generate the final image. This approach eliminates the need for post-processing and enables joint model optimization, whereas leveraging multi-focus datasets with reference images to improve performance and generalization capability. On the other hand, Xu et al. develop a unified unsupervised image fusion network (U2Fusion) [20], which employs a loss function focused on preserving structural information, thereby removing the need for fully focused reference images during training. Tian et al. [17], propose a two-stage MFIF method that combines a DenseNet-based encoder–decoder network with a polarized self-attention module for image reconstruction, along with a fusion strategy based on edge maps derived from encoded features. These maps enable pixel-level discrimination of focused regions and generate a decision map that guides a weighted fusion process, improving detail preservation and smooth transitions between regions. Liu et al. [18], introduce a multi-focus fusion method based on an encoder–decoder architecture with multiscale residual blocks and hybrid attention, trained in an unsupervised manner. The model incorporates a up-and-down sampling projection (UDP) module to enhance edge information and generates a decision map from feature analysis in the spatial–frequency domain, which guides the final fusion. This approach achieves improved detail preservation and greater robustness in fused image quality. However, despite their strong performance, many of these approaches require training processes that involve large volumes of data and substantial computational resources, which may limit their applicability in resource-constrained environments or in scenarios where sufficient data are not available [21].

Considering the previously described scenario, in this work, a MFIF methodology is developed based on a spatial-domain FMO, capable of maintaining low computational cost whereas achieving competitive performance. To this end, particular attention is given to one of the most critical aspects of the fusion process: the transition from binary decision maps (BDM) to soft decision maps (SDM), which preserve spatial coherence without introducing abrupt transitions or significantly increasing algorithmic complexity. This issue is addressed in this work by formulating a fusion scheme based on a Local Focus Measure (LFM) derived from the Tenengrad criterion [12]. Specifically, energy maps are generated for each source image, from which a soft decision map (SDM) is constructed to guide the fusion process through a newly proposed FR. This introduced approach effectively reduces the presence of artifacts. Additionally, two FR are analyzed: the widely used linear additive interpolation and a newly proposed Multiplicative FR. Experimental validation, as well as an exhaustive execution time analysis, are conducted on the well-known Lytro and MFI-WHU datasets [22,23], using both

 doi:10.20944/preprints202606.0829.v1

reference-based and non-reference metrics. The results demonstrate that the proposed framework achieves spatial consistency and stable quantitative performance with low computational cost, supporting its viability as a lightweight and reproducible alternative.

2. Proposed Method

The proposed method is structured into four main stages: local-focus measurement, soft decision map construction, FR application, and quantitative evaluation. Figure 1 illustrates the overall workflow of the proposed method.

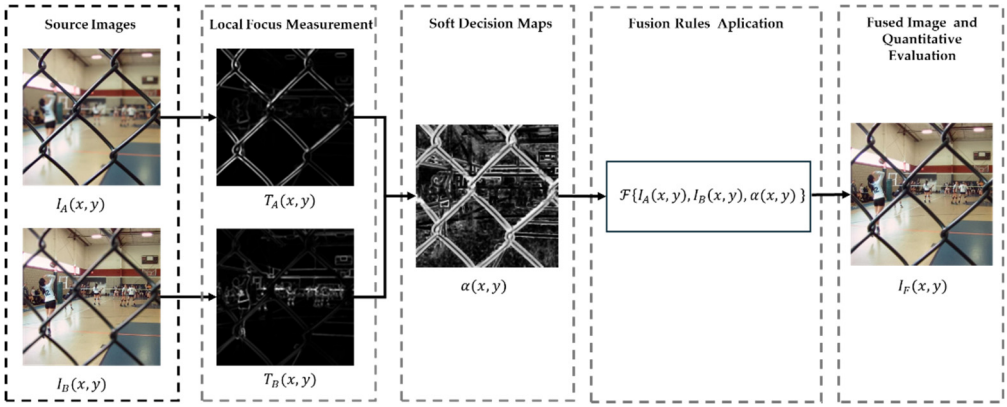

Figure 1. Proposed-method stages.

Given a pair of multi-focus input images $I_A(x,y)$ and $I_B(x,y)$, the focus measures $T_A(x,y)$ and $T_B(x,y)$ are first computed. Subsequently, a soft decision map $\alpha(x,y)$ is constructed, containing values within the interval $[0, 1]$. Finally, the FR $F\{\blacksquare\}$ is applied to obtain the fused image $I_F(x,y)$, which is then evaluated using qualitative metrics.

2.1. Datasets

The proposed method is evaluated using two widely used datasets in the MFIF literature: Lytro (comprising 20 pairs of images captured at different focal planes) and MFI-WHU (consisting of 120 pairs of multi-focus images). In both cases, each image pair corresponds to the same scene with variations in the focused regions, enabling the evaluation of the algorithms ability to properly integrate sharp areas from different source images.

In the case of the MFI-WHU dataset, where defocused regions are synthetically generated, a fully focused image (Ground Truth, GT) is also provided for each scene. This feature enables the use of reference-based metrics for a more accurate quantitative evaluation. In contrast, the Lytro dataset contains only multi-focus image pairs without GT, making it suitable for assessing the method performance in non-reference scenarios, which are closer to real-world applications where a fully focused image is not available.

2.2. Local Focus Measure Estimation

To estimate the level of local focus in the source images, the classical Tenengrad focus measure [14] is employed. This is a low computational cost approach based on gradient magnitude, capable of estimating local sharpness. The measure relies on the premise that focused regions exhibit higher high-frequency content and, consequently, larger gradient magnitudes.

Thus, for a grayscale version of the source images $I(x,y)$, the horizontal and vertical gradient components, $G_x(x,y)$ and $G_y(x,y)$, are computed along the x and y directions, as defined in (1) and (2), respectively.

$$G_x(x,y) = I(x,y) * K_x(x,y)) \tag{1}$$

 doi:10.20944/preprints202606.0829.v1

$$G_y(x,y) = I(x,y) * K_y(x,y) \tag{2}$$

where, the symbol $*$ denotes the convolution operator, and K_x and K_y represent the Sobel masks in the horizontal and vertical directions, respectively. Subsequently, the Tenengrad focus measure is applied to obtain the local gradient energy.

$$T(x,y) = \frac{1}{2w+1} \sum_{i=-w}^{w} \sum_{j=-w}^{w} \{G_x(x+i,y+j)^2 + G_y(x+i,y+j)^2\} \tag{3}$$

In (3), w denotes the size of the neighborhood centered at (x, y). This procedure is applied independently to each source image using a neighborhood size of $w = 3$, generating the energy maps $T_A(x,y)$ and $T_B(x,y)$, which represent a measure of the focus level across the entire image.

2.3. Soft Decision Map

A key aspect of the proposed methodology is the construction of a decision map that serves as a reference to determine the contribution of the focus level from each source image, thereby improving the performance of the FR. To this end, the approach is based on the premise that, in most classical MFIF methods, the FR relies on binary decision maps, where each pixel is entirely selected from the image with the higher focus measure. Although this approach is simple, as previously mentioned, it introduces abrupt transitions and may generate artifacts in regions where both images exhibit similar levels of focus. Therefore, to address this limitation, a non-binary decision map is adopted, namely a soft decision map $\alpha(x,y) \in [0,1]$, whose values are bounded within this interval. To construct it, a local-energy ratio is computed as:

$$\alpha(x,y) = \frac{T_A(x,y)}{T_A(x,y) + T_B(x,y) + \varepsilon} \tag{4}$$

where $T_A(x,y)$ and $T_B(x,y)$ denote the previously computed local focus measures for each input image, and ε is a constant introduced to ensure numerical stability and prevent indeterminate values in the function.

From (4), it can be observed that $\alpha(x,y) \approx 1$when $T_A(x,y) \gg T_B(x,y)$, i.e., when the local energy of image $I_A(x,y)$ is significantly higher than that of $I_B(x,y)$. Conversely, $\alpha(x,y) \approx 0$ when $T_A(x,y) \ll T_B(x,y)$. Finally, when both local energy measures are similar, $\alpha(x,y) \approx 0.5$. This soft decision map improves regions where the input images exhibit similar local focus measures, resulting in smoother transitions.

Figure 2 presents the numerical simulation of the proposed procedure. The first and second columns show the input images, Image A and Image B, respectively, whereas the third and fourth columns display the computation of the local focus measure for each image, LSM A and LSMB, respectively. Finally, the last column shows the soft decision map, where non-binary variations are represented in grayscale according to the previously computed local focus measures.

 doi:10.20944/preprints202606.0829.v1

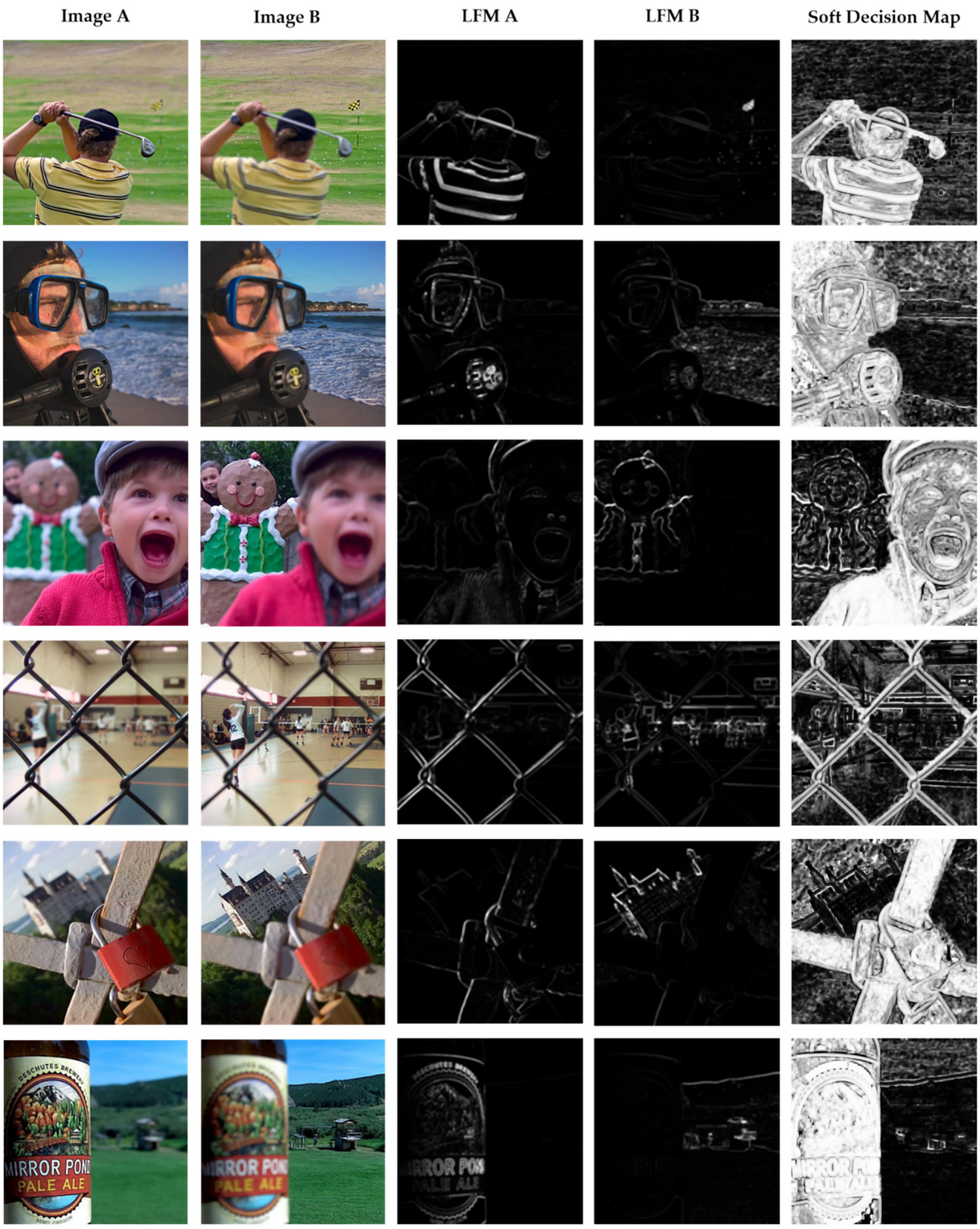

Figure 2. Numerical simulation of the local focus measure and the soft decision map computation.

2.4. Fusion Rule

For the FR described in (5), two types of fusion operators $F\{\blacksquare\}$ are analyzed to study the quality of the fused image based on the two input images and the soft decision map. First, the well-known weighted linear (additive) fusion is employed; additionally, a new fusion strategy, based on a weighted geometric mean (multiplicative), is proposed, which, to the best of our knowledge, has not been previously reported for MFIF.

$$I_F(x,y) = \{I_A(x,y), I_B(x,y), \alpha(a,y)\} \tag{5}$$

 doi:10.20944/preprints202606.0829.v1

2.4.1 Additive Fusion

The first FR corresponds to the well-known linear interpolation, widely used in the state of the art [24,25], and is defined as:

$$I_F(x,y) = \alpha(x,y)I_A(x,y) + \big(1-\alpha(x,y)\big)I_B(x,y) \quad (6)$$

From (6), it is clear that in regions where $\alpha = 1$, the fused output image is determined by image I_A. Conversely, when $\alpha = 0$, it is determined by image I_B. For intermediate values of α, the fused image is obtained as a weighted average of both images. This expression can be generalized to N input images as the weighted sum of all images:

$$I_F(x,y) = \sum_{k=1}^{N} I_k(x,y)\alpha_k \quad ; \quad \sum_{k=1}^{N} \alpha_k = 1 \quad (7)$$

2.4.2 Multiplicative Fusion

To explore a nonlinear behavior, a FR based on a multiplicative-type weighted geometric mean is proposed, defined as:

$$I_F x,y) = I_A(x,y)^{\alpha(x,y)} \cdot I_B(x,y)^{(1-\alpha(x,y))} \quad (8)$$

From (8), similarly to the linear FR, for $\alpha = 1$ and $\alpha = 0$, the fused image is determined by I_A or I_B, respectively. For intermediate values of α, the fused image corresponds to a weighted geometric mean of both images. This expression can be generalized to N input images as the product of the weighted images:

$$I_F(x,y) = \prod_{k=1}^{N} I_k(x,y)^{\alpha_k} \quad ; \quad \sum_{k=1}^{N} \alpha_k = 1 \quad (9)$$

The proposed FR exhibits behavior analogous to the Cobb–Douglas production model, widely used in economics, where multiple factors contribute multiplicatively to the output through exponents that represent their influence [26]. This strategy tends to emphasize the input-image contributions in terms of their structural information, promoting greater discrimination between regions with similar focus levels while preserving fine details.

2.5. Quantitative Evaluation

To evaluate the performance of the proposed methodology, both reference-based and non-reference metrics are employed. Reference-based metrics assess the similarity between the fused image and a ground-truth reference image, whereas non-reference metrics evaluate the quality of the fused image based on perceptual properties.

2.5.1. Reference-Based Metrics

For the MFI-WHU dataset, the following metrics are computed:

- PSNR (Peak Signal-to-Noise Ratio): quantifies the mean squared error with respect to the ground-truth (GT) image. Higher values indicate lower distortion.
- SSIM (Structural Similarity Index): evaluates structural similarity by considering luminance, contrast, and local structure. Values close to 1 indicate high structural similarity.

These metrics quantify the global and structural similarity between the fused image and the fully focused ground-truth reference image.

2.5.2. Non-Reference Metrics

For both datasets, widely used non-reference metrics from the image fusion literature are employed:

 doi:10.20944/preprints202606.0829.v1

- NMI (Normalized Mutual Information): measures the amount of information shared between the fused image and the source images.
- Q_{ABF} (Edge-Based Fusion Quality Index): evaluates the preservation of edges and gradient orientations in the fused image.
- Q_{CB} (Chen–Blum Contrast Fusion Quality Index): quantifies the consistency of local contrast with respect to the source images.
- SF (Spatial Frequency): measures the overall SF of the image; higher values indicate greater high-frequency content and, consequently, a higher level of detail.

These metrics assess the quality of the fused image without requiring a reference image, based on changes in statistical and perceptual properties.

2.6. Computational-Cost Analysis

In addition to the quality assessment, the execution time corresponding to each stage of the proposed methodology are recorded. The total execution time of the algorithm is defined as the sum of the time required by the following stages: (a) local focus measurement, (b) soft decision map construction, and (c) fusion-rule application.

3. Experimental Evaluation

In this section, the MFI-WHU dataset, consisting of 120 pairs of multi-focus images and their corresponding *ground-truth* images, is used to evaluate the performance of the proposed method using the reference-based metrics PSNR and SSIM. Additionally, the non-reference metrics NMI, Q_{ABF}, Q_{CB}, and SF are computed on the same dataset. Similarly, the Lytro dataset, which contains 20 pairs of multi-focus images, is employed to assess the performance of the proposed method using the aforementioned non-reference metrics. The obtained results were compared against those of eight representative state-of-the-art fusion algorithms: Multi-Focus Image Fusion based on Residual Removal (MFIF-RR) [27], Multimodal Medical Image Fusion based on Joint Bilateral Filter (MMIF-JBF) [28], Generative Adversarial Network for Multi-Focus Image Fusion (MFIF-GAN) [29], A General Unsupervised Image Fusion Network based on Memory Unit (MUFusion) [30], Small-Area-Aware Multi-Focus Image Fusion (SAMF) [31], Adaptive Region Division Multi-Focus Image Fusion (RDMF)[32], Bridging the Gap between Multi-Focus and Multi-Modal Image Fusion (MFIF-MMIF)[33], and Gradient-Based Multi-Focus Image Fusion (Gradient-MFIF) [34]. In the proposed method, a 7 × 7 neighborhood window is used for local focus measure computation. Furthermore, two FR are evaluated: the conventional Additive FR and the Multiplicative FR introduced in this work. The reported results correspond to the average values obtained across all scenes in each dataset.

In addition to fusion quality, the computational cost of the proposed method is evaluated, measured in seconds, to assess its suitability for real-time applications. The experiments are conducted on a workstation running Pop!_OS 22.04 LTS (64-bit), equipped with an Intel Core i7-10750H processor (12 threads, up to 5.0 GHz) and 16 GB of RAM. Although the system includes an NVIDIA GeForce GTX 1650 Ti mobile Graphics Processing Unit (GPU), all computations are performed exclusively on the Central Processing Unit (CPU).

3.1. Quantitative Comparison

Table 1 and Table 2 present the quantitative comparison between the proposed method and eight representative methods from the literature, using both reference-based and non-reference metrics on the two evaluated datasets. In Table 1, corresponding to the MFI-WHU dataset, it can be observed that none of the compared methods report reference-based metrics such as PSNR and SSIM, which limits the assessment of fusion quality with respect to a fully focused reference image. In contrast, the proposed method, in its two variants referred to as Additive Fusion and Multiplicative Fusion, achieves PSNR values of 33.84 dB and 33.75 dB, respectively, as well as an SSIM value of 0.985

 doi:10.20944/preprints202606.0829.v1

for both variants, indicating a high degree of structural similarity with the ground-truth image. Regarding the non-reference metrics, the Multiplicative Fusion variant achieved the highest NMI value (0.953) and the best Q_{CB} performance (0.912), indicating a greater amount of shared information and improved local contrast preservation. Furthermore, both proposed variants achieve the highest Q_{ABF} score (0.817), reflecting superior preservation of edges and structural details in the fused image. It is worth noting that both proposed variants obtain slightly lower SF values. However, since this metric measures high-frequency content without distinguishing between meaningful details and noise, these results suggest that the proposed method preserves relevant structures better while simultaneously reducing the amplification of unwanted noise components.

Table 1. Performance comparison of state-of-the-art methods on the MFI-WHU dataset.

Method	PSNR (dB)	SSIM	NMI	Q_{ABF}	Q_{CB}	SF
MFIF-RR	-	-	1.184	0.728	0.822	26.760
MMIF-BJF	-	-	1.038	0.723	0.778	26.696
MFIF-GAN	-	-	1.184	0.732	0.822	26.850
MUFusion	-	-	0.728	0.599	0.647	22.585
SAMF	-	-	1.188	0.726	0.820	26.667
RDMF	-	-	1.174	0.728	0.823	26.591
MFIF-MMIF	-	-	1.038	0.717	0.791	25.846
Gradient MFIF	-	-	1.185	0.733	0.825	26.783
Proposed (Additive Fusion)	**33.84**	**0.985**	**0.949**	**0.817**	**0.907**	**23.711**
Proposed (Multiplicative Fusion)	**33.75**	**0.985**	**0.953**	**0.817**	**0.912**	**23.886**

Table 2 presents the results obtained on the Lytro dataset using only non-reference metrics. Consistent with the results observed for the MFI-WHU dataset, the proposed variants achieved competitive NMI values of 0.974 and 0.982, indicating effective preservation of the shared information between the fused image and the source images. Regarding the Q_{ABF} and Q_{CB} metrics, the proposed methods achieve the best results, reaching values of up to 0.841 and 0.924, respectively. These results demonstrate excellent edge preservation and high contrast quality in the fused images. On the other hand, the obtained SF values are slightly lower than those reported by some of the competing methods. However, as previously discussed, this metric does not distinguish between meaningful details and noise. Therefore, the results suggest that the proposed method favors the preservation of relevant structures while limiting noise amplification, ultimately contributing to improved visual quality in the fused images.

Table 2. Performance results on the Lytro dataset using non-reference metrics.

Method	NMI	Q_{ABF}	Q_{CB}	SF
MFIF-RR	1.1261	0.7530	0.8018	19.4040
MMIF-BJF	0.8931	0.7140	0.6690	18.6682
MFIF-GAN	1.1313	0.7529	0.8005	19.4271
MUFusion	0.7983	0.6624	0.6770	18.9482
SAMF	1.1191	0.7511	0.7951	19.3820
RDMF	1.1221	0.7518	0.8010	19.3444
MFIF-MMIF	0.9329	0.7318	0.7338	18.5671
Gradient MFIF	-	0.7531	0.8024	19.4308
Ours (Additive Fusion)	**0.974**	**0.834**	**0.915**	**18.00**
Ours (Multiplicative Fusion)	**0.982**	**0.841**	**0.924**	**18.32**

Based on the results obtained across both datasets, the proposed method demonstrates a more balanced performance compared with state-of-the-art approaches when all evaluation metrics are

 doi:10.20944/preprints202606.0829.v1

considered jointly. It particularly stands out in terms of Q_{ABF} and Q_{CB}, indicating superior edge preservation and better local contrast consistency in the fused images. Moreover, it maintains competitive NMI and SSIM values, reflecting effective information retention and high structural fidelity. Overall, these results demonstrate that the proposed method provides a favorable balance among structural quality, information content, and perceptual image quality.

3.1. Qualitative Results

To visually compare the fusion results produced by the proposed method, Figure 3 and Figure 4 present the output obtained by using the two FR considered in this work on both datasets. Unlike the quantitative comparison, this section focuses exclusively on the proposed variants, as they achieve superior performance compared with the state-of-the-art methods according to the evaluated objective metrics.

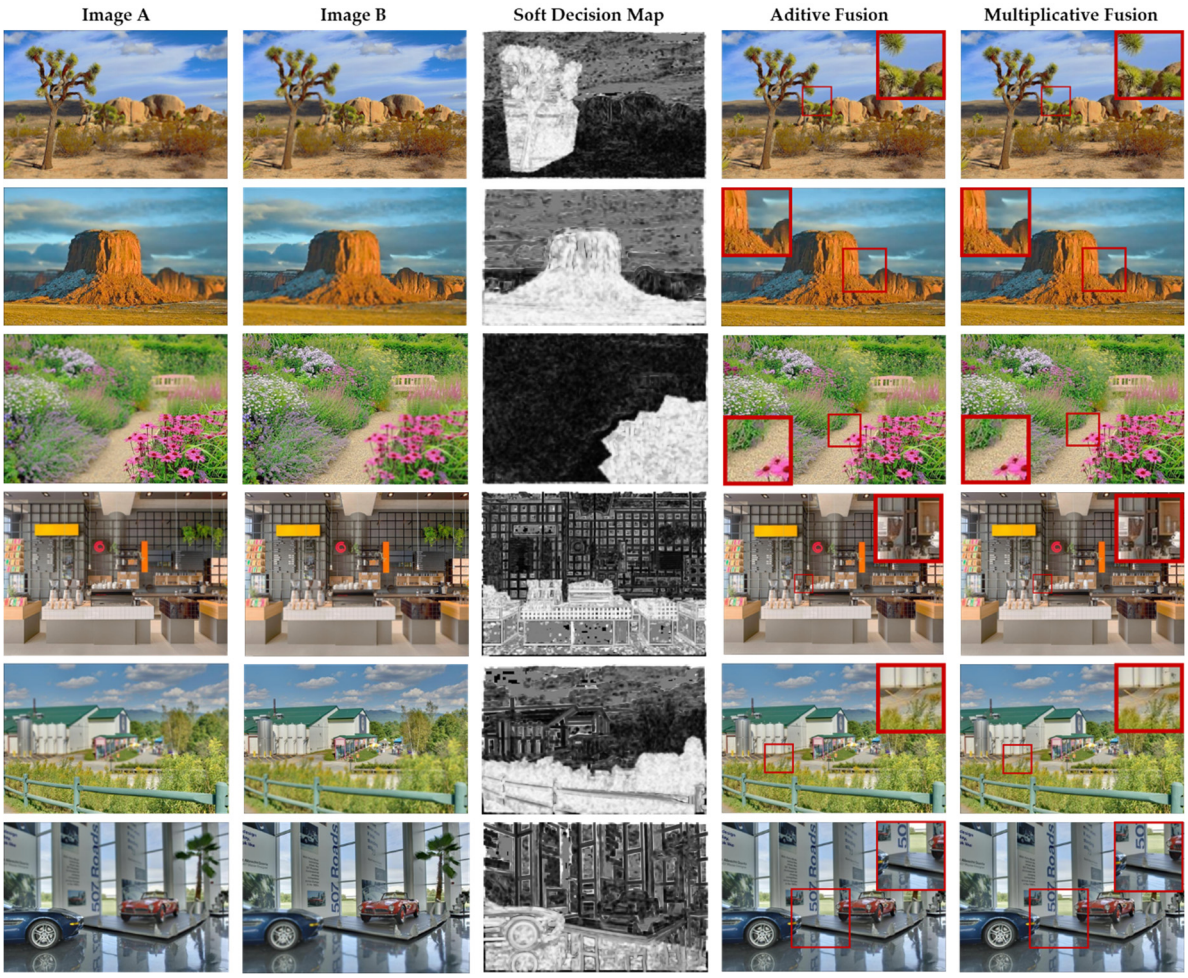

Figure 3. Source Images, Image A and Image B, Soft Decision Maps, and Fused Images Generated by the Proposed Method on the MFI-WHU Dataset.

Figure 3 presents six image pairs from the MFI-WHU dataset, showing the source images (Image A and Image B), the Soft Decision Map generated by the proposed methodology, and the fused images obtained using the Additive and Multiplicative FR. As can be observed, the soft decision map effectively identifies the regions with the highest local focus level in each image, thereby facilitating the fusion process. In the resulting fused images, the Additive FR produces smoother transitions between regions, helping to reduce the occurrence of visual artifacts. In contrast, the Multiplicative FR provides sharper edge and texture definition, enhancing the preservation of fine details in the fused image.

 doi:10.20944/preprints202606.0829.v1

Figure 4 presents six image pairs from the Lytro dataset. Similar to the results observed on the MFI-WHU dataset, the Additive FR produces visually homogeneous images with smooth transitions between regions and without abrupt changes. In contrast, the Multiplicative FR stands out for its ability to preserve fine details and enhance the visual perception of the scene, particularly in regions with greater structural complexity and depth.

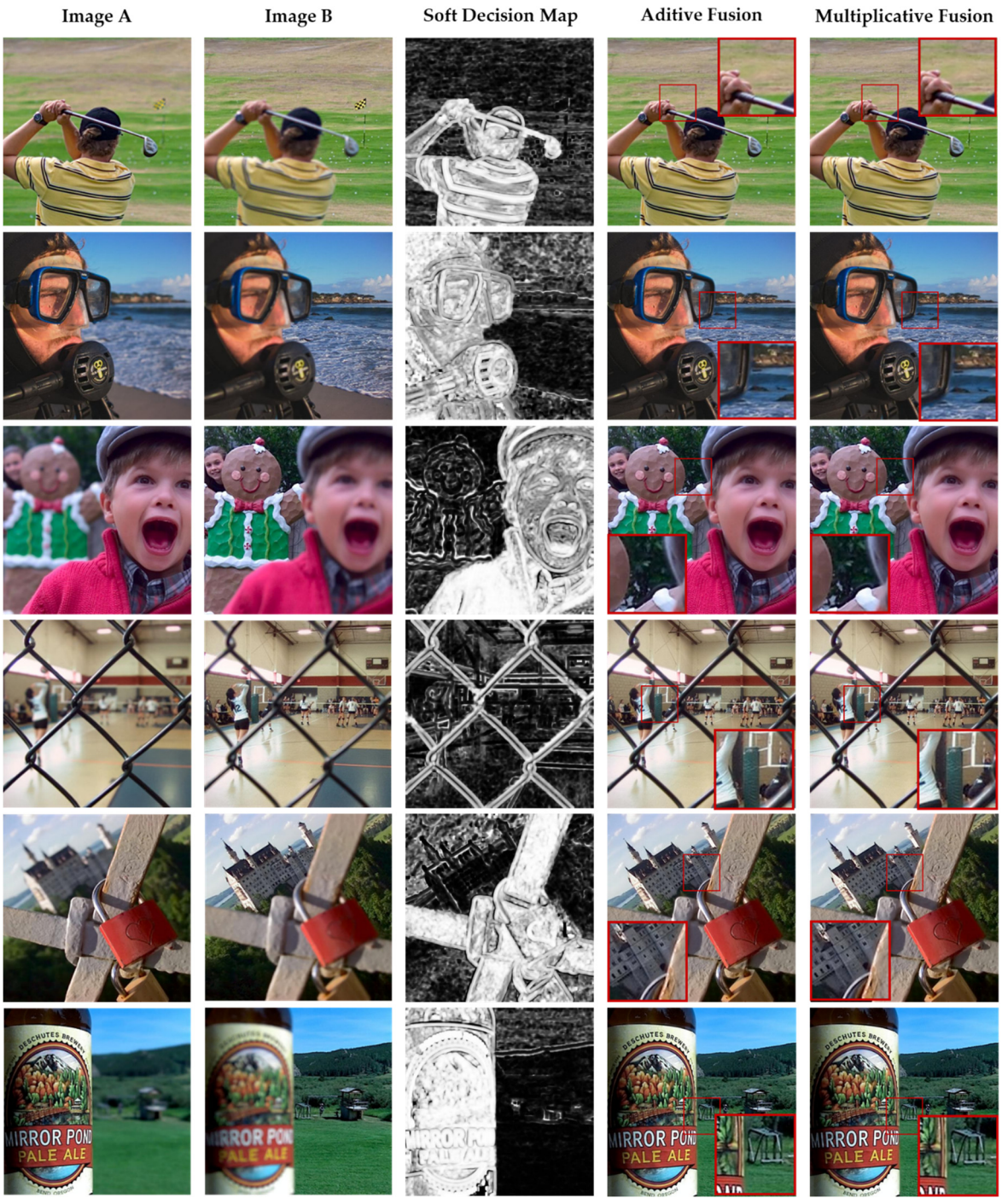

Figure 4. Visual Comparison of Fusion Results Obtained Using the Additive and Multiplicative FR on the Lytro Dataset.

3.4. Computational Cost

In addition to the quality of the fused images, the computational cost of the proposed method is evaluated to assess its suitability for real-time applications. The algorithm is implemented in Python

 doi:10.20944/preprints202606.0829.v1

using standard numerical and image-processing libraries. Since the methodology relies primarily on local gradient operations and simple arithmetic computations, it does not require GPU acceleration, thereby maintaining low computational complexity and facilitating reproducibility across different hardware platforms.

Table 3 presents the results of the computational cost analysis. The total execution time of the algorithm is defined as the sum of the processing times required by its three main stages: **(a)** local focus measurement, **(b)** soft decision map construction, and **(c)** application of the FR.

Table 3. Average execution time using the Additive and Multiplicative FR on both datasets.

Method	MFI - WHU	Lytro
Proposed (Additive fusion)	0.07s	0.007s
Proposed (Multiplicative fusion)	0.026s	0.027s

The results show that the algorithm achieves very low execution time, owing to its reliance exclusively on local convolution operations and simple algebraic computations. In particular, the Additive Fusion variant requires approximately 0.007 s to generate a fused image, whereas the Multiplicative Fusion variant requires 0.026 s. This difference can be attributed to the higher computational cost associated with the multiplicative and exponential operations involved in the Multiplicative FR, compared with the weighted summation used in Additive Fusion.

4. Discussion

The obtained results allow analyzing the proposed-methodology performance from three complementary perspectives: i) fusion quality assessed through objective metrics, ii) qualitative analysis focused on edge preservation, contrast, and information transfer, and iii) computational cost. Taken together, the results demonstrate that the proposed method provides superior performance compared with state-of-the-art methodologies when considering the balance between fusion quality and computational efficiency. In particular, the obtained PSNR and SSIM values indicate a high level of fidelity between the fused images and the fully focused reference images. Furthermore, it is worth noting that these metrics are not reported by the compared state-of-the-art methods, making it difficult to directly assess their accuracy with respect to a ground-truth image.

When analyzing the no-reference metrics, the proposed method achieves a well-balanced performance across NMI, Q_{ABF}, and Q_{CB}, improving information preservation, edge retention, and local contrast enhancement. Although the obtained NMI values are not the highest among the evaluated methods, they remain very close to those reported by approaches with significantly higher computational complexity. This suggests that the proposed method efficiently integrates the most relevant information from the source images while avoiding unnecessary redundancy in the fused result.

It is important to highlight the slightly lower values obtained for the SF metric. Although high SF values are often associated with a high level of detail, this metric does not differentiate between meaningful information and noise components. In this context, the results suggest that the proposed method avoids artificially enhancing high-frequency components, thereby promoting a more controlled fusion process. As a result, noise amplification is reduced while preserving the structures and details that are truly relevant to the scene.

The qualitative analysis reinforces the findings obtained from the quantitative metrics, revealing clear differences between the two proposed fusion strategies. Additive Fusion produces smoother and more homogeneous images, with gradual transitions between regions, thereby enhancing visual consistency and reducing the presence of artifacts. In contrast, Multiplicative Fusion, due to its nonlinear nature, emphasizes fine details and local contrast, resulting in improved edge and texture definition. These differences confirm that both approaches exhibit complementary characteristics. While Additive Fusion prioritizes smoothness and visual consistency, Multiplicative Fusion favors

 doi:10.20944/preprints202606.0829.v1

detail preservation and structural discrimination. Therefore, the most suitable strategy can be selected according to the specific requirements of the target application.

From a computational perspective, the proposed method demonstrates high efficiency, achieving very low execution times that make it suitable for real-time applications. The difference observed between the Additive and Multiplicative Fusion schemes is directly related to the complexity of the operations involved in each approach. Nevertheless, both variants maintain low processing times, confirming the feasibility of the proposed method for deployment in computationally constrained environments.

5. Conclusions

In this work, a MFIF method based on local focus measurement and a soft decision map is proposed, incorporating two fusion strategies: Additive Fusion and Multiplicative Fusion. Both variants are evaluated on the MFI-WHU and Lytro datasets using reference-based and non-reference metrics, qualitative assessment, and computational cost analysis. The results show that the proposed method achieves competitive and well-balanced performance compared with state-of-the-art approaches, particularly in edge preservation (Q_{ABF}) and contrast enhancement (Q_{CB}), while maintaining adequate shared information (NMI) and high structural fidelity (SSIM). Although the SF values are slightly lower, this suggests a controlled fusion process that preserves relevant structures without unnecessarily enhancing high-frequency components. The qualitative results show that the two fusion strategies have complementary behavior. Additive Fusion produces smoother and more visually homogeneous images, making it suitable when visual consistency is prioritized. In contrast, Multiplicative Fusion better preserves fine details and enhances local contrast, which is useful when structural definition is critical. From a computational perspective, the method achieves low execution times because it relies on simple local operations, making it suitable for real-time applications without specialized hardware. Overall, the proposed framework offers a favorable balance between fusion quality, structural preservation, perceptual quality, and computational efficiency. Future work will explore hybrid architectures that integrate deep learning techniques to further improve representation capability and fusion performance.

Author Contributions: B.L.M. participated in all steps of the research method's conceptualization, the materials and methods, the experimentation, the validation, the writing—original draft; L.M.L.C., S.S.C. and E.C.Y., participated in the conceptualization, resources, supervision, and investigation; B.L.M., L.M.L.C., S.S.C, M.L.-R., C.R.D. and E.C.-Y., participated in the writing—review and editing, visualization, and formal analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Secretariat of Science, Humanities, Technology and Innovation (SECIHTI), Mexico, under Scholarship 1079866.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the financial support of the Secretariat of Science, Humanities, Technology and Innovation (SECIHTI), Mexico, under Scholarship 1079866.

Conflicts of Interest: The authors declare no conflicts of interest

Abbreviations

The following abbreviations are used in this manuscript:

MFIF	Multi-Focus Image Fusion
FMO	Focus Measure Operator
LFM	Local Focus Mesasure
FR	Fusion Rule
NSCT	Non-Subsampled Contourlet Transform
MFI-WHU	Multi-Focus Image Dataset from Wuhan University

 doi:10.20944/preprints202606.0829.v1

GT	Ground Truth
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure
NMI	Normalized Mutual Information
QABF	Quality Assessment Based on Edge Information
QCB	Chen–Blum Metric
SF	Spatial Frequency
GPU	Graphics Processing Unit
CPU	Central Processing Unit

References

1. Ojeda-Castañeda, J. *Wavefront Shaping and Pupil Engineering*; Springer Series in Optical Sciences; Springer Berlin Heidelberg: Berlin, Heidelberg, 2021; Vol. 235; ISBN 978-3-662-63800-2.
2. Ojeda-Castañeda, J.; Yepez-Vidal, E.; Gómez-Sarabia, C.M. Multiple-Frame Photography for Extended Depth of Field. *Appl. Opt.* **2013**, *52*, D84, doi:10.1364/AO.52.000D84.
3. Ledesma-Carrillo, L.; Guzmán-Cabrera, R.; Gómez-Sarabia, C.M.; Torres-Cisneros, M.; Ojeda-Castañeda, J. Tunable Field Depth: Hyperbolic Optical Masks. *Appl. Opt.* **2017**, *56*, A104, doi:10.1364/AO.56.00A104.
4. Zhou, Y.; Yu, L.; Zhi, C.; Huang, C.; Wang, S.; Zhu, M.; Ke, Z.; Gao, Z.; Zhang, Y.; Fu, S. A Survey of Multi-Focus Image Fusion Methods. *Appl. Sci.* **2022**, *12*, 6281–6281, doi:10.3390/app12126281.
5. Xia, H.H.; Gao, H.; Shao, H.; Gao, K.; Liu, W. Multi-Focus Microscopy Image Fusion Based on Swin Transformer Architecture. *Appl. Sci.* **2023**, *13*, 12798, doi:10.3390/app132312798.
6. Wu, P.; Jiang, L.; Hua, Z.; Li, J. Multi-Focus Image Fusion: Transformer and Shallow Feature Attention Matters. *Displays* **2023**, *76*, 102353, doi:10.1016/j.displa.2022.102353.
7. Xie, X.; Guo, B.; He, S.; Gu, Y.; Li, Y.; Li, P. One-Shot Multi-Focus Image Stack Fusion via Focal Depth Regression. *Eng. Appl. Artif. Intell.* **2025**, *162*, 112667, doi:10.1016/j.engappai.2025.112667.
8. Li, H.; Shen, T.; Zhang, Z.; Zhu, X.; Song, X. EDMF: A New Benchmark for Multi-Focus Images with the Challenge of Exposure Difference. *Sensors* **2024**, *24*, doi:10.3390/s24227287.
9. Li, X.; Li, X.; Ye, T.; Cheng, X.; Liu, W.; Tan, H. Bridging the Gap between Multi-Focus and Multi-Modal: A Focused Integration Framework for Multi-Modal Image Fusion. **2024**, doi:10.48550/arXiv.2311.01886.
10. Wang, C.; Yuan, R.; Sun, Y.; Jiang, Y.; Chen, C.; Lin, X. A New Method of Multi-Focus Image Fusion Using Laplacian Operator and Region Optimization. *J. Comput. Commun.* **2018**, *06*, 106–118, doi:10.4236/jcc.2018.65009.
11. Bavirisetti, D.P.; Dhuli, R. Multi-Focus Image Fusion Using Multi-Scale Image Decomposition and Saliency Detection. *Ain Shams Eng. J.* **2018**, *9*, 1103–1117, doi:10.1016/j.asej.2016.06.011.
12. Piao, W.; Han, Y.; Hu, L.; Wang, C. Quantitative Evaluation of Focus Measure Operators in Optical Microscopy. *Sensors* **2025**, *25*, doi:10.3390/s25103144.
13. Zang, Y.; Zhou, D.; Wang, C.; Nie, R.; Guo, Y. UFA-FUSE: A Novel Deep Supervised and Hybrid Model for Multi-Focus Image Fusion. **2021**, doi:10.1109/TIM.2021.3072124.
14. Lv, M.; Li, L.; Jin, Q.; Jia, Z.; Chen, L.; Ma, H. Multi-Focus Image Fusion via Distance-Weighted Regional Energy and Structure Tensor in NSCT Domain. *Sensors* **2023**, *23*, doi:10.3390/s23136135.
15. Wan, H.; Tang, X.; Zhu, Z.; Li, W. Multi-Focus Image Fusion Method Based on Multi-Scale Decomposition of Information Complementary. *Entropy* **2021**, *23*, 1362, doi:10.3390/e23101362.
16. Chen, H.; Du, X.; Huang, H.; Zhao, T. A Real-Time High-Resolution Multi-Focus Image Fusion Algorithm Based on Multi-Scale Feature Aggregation. *Appl. Sci.* **2025**, *15*, 6967, doi:10.3390/app15136967.
17. Tian, B.; Yang, L.; Dang, J. Fine-Grained Multi-Focus Image Fusion Based on Edge Features. *Sci. Rep.* **2023**, *13*, 2478, doi:10.1038/s41598-023-29584-y.
18. Liu, T.; Chen, M.; Duan, Z.; Cui, A. Multi-Focused Image Fusion Algorithm Based on Multi-Scale Hybrid Attention Residual Network. *PLOS ONE* **2024**, *19*, e0302545, doi:10.1371/journal.pone.0302545.
19. Zhang, Y.; Liu, Y.; Sun, P.; Yan, H.; Zhao, X.; Zhang, L. IFCNN: A General Image Fusion Framework Based on Convolutional Neural Network. *Inf. Fusion* **2020**, *54*, 99–118, doi:10.1016/j.inffus.2019.07.011.
20. Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2Fusion: A Unified Unsupervised Image Fusion Network. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 502–518, doi:10.1109/TPAMI.2020.3012548.

 doi:10.20944/preprints202606.0829.v1

21. Wang, X.; Hua, Z.; Li, J. Multi-Focus Image Fusion Framework Based on Transformer and Feedback Mechanism. *Ain Shams Eng. J.* **2023**, *14*, 101978, doi:10.1016/j.asej.2022.101978.
22. Nejati, M.; Samavi, S.; Shirani, S. Multi-Focus Image Fusion Using Dictionary-Based Sparse Representation. *Inf. Fusion* **2015**, *25*, 72–84, doi:10.1016/j.inffus.2014.10.004.
23. Zhang, H.; Le, Z.; Shao, Z.; Xu, H.; Ma, J. MFF-GAN: An Unsupervised Generative Adversarial Network with Adaptive and Gradient Joint Constraints for Multi-Focus Image Fusion. *Inf. Fusion* **2021**, *66*, 40–53, doi:10.1016/j.inffus.2020.08.022.
24. Shutao Li; Xudong Kang; Jianwen Hu Image Fusion With Guided Filtering. *IEEE Trans. Image Process.* **2013**, *22*, 2864–2875, doi:10.1109/TIP.2013.2244222.
25. Peña, F.A.G.; Fernández, P.D.M.; Ren, T.I.; Vasconcelos, G.C.; Cunha, A. A Multiple Source Hourglass Deep Network for Multi-Focus Image Fusion. **2019**, doi:https://doi.org/10.48550/arXiv.1908.10945.
26. Greer, M. The Economics (and Econometrics) of Cost Modeling. In *Electricity Cost Modeling Calculations*; Elsevier, 2011; pp. 115–148 ISBN 978-1-85617-726-9.
27. Li, X.; Zhou, F.; Tan, H.; Chen, Y.; Zuo, W. Multi-Focus Image Fusion Based on Nonsubsampled Contourlet Transform and Residual Removal. *Signal Process.* **2021**, *184*, 108062, doi:10.1016/j.sigpro.2021.108062.
28. Li, X.; Zhou, F.; Tan, H.; Zhang, W.; Zhao, C. Multimodal Medical Image Fusion Based on Joint Bilateral Filter and Local Gradient Energy. *Inf. Sci.* **2021**, *569*, 302–325, doi:10.1016/j.ins.2021.04.052.
29. Wang, Y.; Xu, S.; Liu, J.; Zhao, Z.; Zhang, C.; Zhang, J. MFIF-GAN: A New Generative Adversarial Network for Multi-Focus Image Fusion. *Signal Process. Image Commun.* **2021**, *96*, 116295, doi:10.1016/j.image.2021.116295.
30. Cheng, C.; Xu, T.; Wu, X.-J. MUFusion: A General Unsupervised Image Fusion Network Based on Memory Unit. *Inf. Fusion* **2023**, *92*, 80–92, doi:10.1016/j.inffus.2022.11.010.
31. Li, X.; Li, X.; Tan, H.; Li, J. SAMF: Small-Area-Aware Multi-Focus Image Fusion for Object Detection. **2024**, doi:10.48550/arXiv.2401.08357.
32. Yu, W.; Nie, Y.; Xiong, W. An Adaptive Region Division Multi-Focus Image Fusion Algorithm with Defocus Diffusion Mitigation Mechanism. In Proceedings of the Proceedings of the 2024 2nd Asia Conference on Computer Vision, Image Processing and Pattern Recognition; ACM: Xiamen China, April 26 2024; pp. 1–6.
33. Li, X.; Li, X.; Ye, T.; Cheng, X.; Liu, W.; Tan, H. Bridging the Gap between Multi-Focus and Multi-Modal: A Focused Integration Framework for Multi-Modal Image Fusion. **2024**, doi:10.48550/arXiv.2311.01886.
34. Li, H.; Li, X. Gradient-Based Multi-Focus Image Fusion with Focus-Aware Saliency Enhancement. **2025**, doi:10.48550/ARXIV.2509.22392.