ACIS
Advances in Computational and Intelligent Systems
Vol. 1, No. 1, July 2025, pp. 18~26
ISSN: 3093-7388, DOI: https://doi.org/10.56313/acis.v1i1.3

Deep Learning-Based Semantic Focus Fusion for High-Quality Multifocus Image

Ismail[1], Kamarul Hawari Ghazali[2]
[1]Electrical Department, Politeknik Negeri Padang, Indonesia
[2]Center for Advance Industrial Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan Pahang

Article Info

Article history:

Received March 18, 2025
Revised month May 4, 2025
Accepted month May 5, 2025

Keywords:

Multifocus Image Fusion
Semantic Focus Fusion
Structural Similarity Index

ABSTRACT

Multifocus Image Fusion (MIF) is a technique in image processing that enhances image clarity by integrating multiple images taken at different focal distances into a single, sharp image. Several traditional fusion methods, including spatial and transform domain techniques, are available; however, these approaches often struggle with preserving fine details and preventing artifacts. This paper proposes a CNN-based Semantic Focus Fusion (SFF) method, leveraging deep learning architectures to improve focus region classification and eliminate blurring effects. The model is trained using a structured dataset and optimized through hyperparameter tuning to ensure efficient convergence. The effectiveness of the proposed method is evaluated using the Structural Similarity Index (SSIM), demonstrating superior fusion quality compared to conventional methods. Experimental results confirm that the proposed approach achieves high perceptual clarity, better structural consistency, and improved edge preservation with an efficiecny approximately 0.92. This study highlights the advantages of deep learning in MIF, offering a robust solution for applications in medical imaging, surveillance, and industrial inspection.

Corresponding Author:

Ismail
Email: indoismail@gmail.com

1. INTRODUCTION

In applications that require image analysis, high-quality images are essential for precise interpretation and decision-making in fields such as medical diagnostics, quality inspection in the automotive industry, fruit grading in agriculture, military surveillance, and traffic monitoring. The depth of field (DOF) limitations of traditional cameras contribute to blurring in objects outside a certain focus range. In image analysis, this issue significantly impacts the accuracy and reliability of interpretation [1].

Depth of field (DOF) refers to the range of distances in a scene where objects appear reasonably sharp [2]. It is determined by several factors, including sensor size, focal length, and camera lens aperture. A deep DOF ensures that both the foreground and background remain in focus, while a shallow DOF, commonly used in portrait photography, produces a sharp subject with a blurred background. Although DOF is beneficial for creative photography, it presents a significant challenge for computer vision systems that require precise object detection and classification.

The limitations of conventional cameras make it difficult to accurately capture scenes with multiple subjects at varying distances, leading to inaccuracies in focus and depth perception [3]. Therefore, understanding the factors that influence DOF is crucial for optimizing imaging techniques in scientific and industrial applications. The following Figure 1 illustrates an example of an image with good and poor DOF, demonstrating the impact of depth of field on image clarity and subject focus.

Figure 1. Example of a focused image used in the fusion process. The image illustrates a scene where specific regions are sharply focused, while others are intentionally blurred.

The depth of field (DOF) in an image is influenced by several key factors, including aperture size, focal length, and subject distance. A wider aperture (lower f-number) creates a shallower DOF, making the subject stand out against a beautifully blurred background—a technique often used in portrait photography. In contrast, a smaller aperture (higher f-number) increases DOF, ensuring that more elements in the scene remain sharp and in focus, which is useful for landscapes or scientific imaging.

Focal length also plays a crucial role, shorter focal lengths (such as those in wide-angle lenses) provide a greater DOF, capturing more of the scene in focus. On the other hand, longer focal lengths (telephoto lenses) create a shallower DOF, naturally drawing attention to the subject while softening the background. The distance between the camera and the subject further impacts DOF—when the camera is closer to the subject, the background becomes more blurred, whereas increasing the distance results in a deeper DOF, keeping more of the scene in focus [4]. Understanding these elements allows photographers, videographers, and engineers to precisely control focus for both artistic and practical applications. Whether it's creating a dramatic portrait with soft backgrounds, ensuring sharp focus in product photography, or capturing detailed industrial or medical images, mastering DOF is essential for achieving the desired visual effect.

In medical imaging and automated inspection, a limited depth of field (DOF) can affect the ability to capture fine details necessary for accurate analysis. For example, in medical diagnostics, failing to achieve a clear focus across an entire microscopic slide (Figure 2) can lead to misinterpretations in image analysis, potentially affecting diagnostic accuracy [5]. Similarly, in quality inspection within the automotive industry, parts with varying depths did not captured in a single frame, affecting defect detection processes. In precision agriculture, automated grading systems rely on clear, fully focused images to accurately classify fruit quality, detect diseases, or assess ripeness levels. The same applies to military surveillance and traffic monitoring, where identifying objects across different distances requires overcoming DOF constraints to enhance situational awareness.

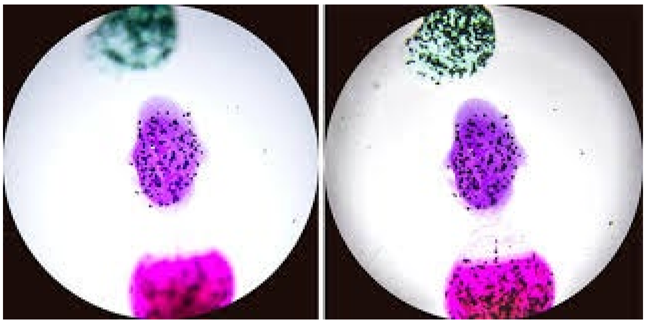

Figure 1. Microscopic slides

Multifocus Image Fusion (MIF) provides a solution to the depth of field (DOF) problem [6]. MIF generates a single, high-quality image where all regions remain in focus by combining multiple images of the same scene captured at different focal distances. This technique preserves details across various depth levels, enhancing the clarity and interpretability of the scene.

Several MIF techniques have been introduced, including deep learning-based methods, transform domain approaches, and spatial domain techniques. However, achieving seamless focus fusion without introducing artifacts or inconsistencies remains a challenge. These limitations emphasize the need for more advanced computational techniques, such as deep learning-based Semantic Focus Fusion (SFF), which leverages deep learning architectures to enhance focus classification and fusion accuracy.

For the improvement of image focus, image fusion has been introduced to enhance image clarity. Normally the approach is consisting of transform domain methods, spatial domain approaches, and hybrid methods that combine several different approaches [7]. Normally, the approach consists of transform domain methods, spatial domain approaches, and hybrid methods that combine several different techniques.

Simple pixel-level fusion is one of the oldest methods and it selects the sharpest pixel from a number of input images by using local contrast or intensity gradients. To find focus regions, techniques like Laplacian-based fusion use sharpness metrics that are obtained from Laplacian filters. Similarly, to identify the areas of an image that are most focussed, gradient-based techniques examine edge information.

Another method that can be used is frequency-based transformations, which are utilized in transform domain approaches to analyze and combine targeted areas. The Discrete Wavelet Transform (DWT), which breaks down images into multi-scale representations and then selects high-frequency coefficients from the most focused areas to accomplish fusion [8][9], [10]. The Curvelet Transform is another well-liked technique that works better on images with noticeable geometric patterns because it captures directional characteristics more well than wavelets. To get the best results, these techniques necessitate precise parameter adjustment and frequently create artifacts.

Utilizing the advantages of both spatial and transform domain approaches is the goal of hybrid approaches. To enhance fusion quality, the Guided Filtering-based method combines focus measurements with edge-preserving filtering. Furthermore, region-based techniques divide images into several regions using watershed algorithms or superpixel clustering, then choose fuse the areas that are of interest. Although these techniques increase boundary smoothness, they can be computationally costly.

Conventional multifocus image fusion approaches have advanced, but they still have drawbacks such computational inefficiency, loss of texture information, and imprecise border recognition. In order to increase the accuracy of focus identification and fusion, researchers are now investigating deep learning-based methods that make use of convolutional neural networks (CNNs) and semantic segmentation models [11], [12], [13].

In this study, the Structural Similarity Index (SSIM) is used as a metric to evaluate the performance of the fusion method. SSIM, introduced by Wang et al. [14], is a perceptual image quality metric that measures the similarity between two images. Unlike traditional metrics such as Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR), which quantify absolute differences, SSIM assesses image quality based on luminance, contrast, and structural information, making it more consistent with human visual perception. The SSIM score ranges from -1 to 1, with a value of 1 indicating perfect similarity between the reference and test images. Due to its ability to capture perceptual differences effectively, SSIM is widely used in various image processing tasks, including image fusion.

2. METHODOLOGY

In this paper, we propose a method for multifocus image fusion based on a Convolutional Neural Network (CNN) architecture designed to classify focus and blurred regions in multifocus images. The methodology involves several critical steps, including dataset collection, preprocessing, CNN-based feature extraction, focus region classification, and image fusion, as shown in Figure 1.

Figure 3. The research methodology

2.1. Dataset Collection and Preparation

The data used in this study is taken from the Nejati repository [15]—which comprises 2,304 pairs of multifocus photographs and the label images that go with them are used in this investigation. The sample of data is shown in figure 4. With a 512 x 512-pixel resolution, each multifocus image pair consists of two grayscale images: one with a crisp foreground and blurred background, and the other with a sharp background and blurred foreground [16].

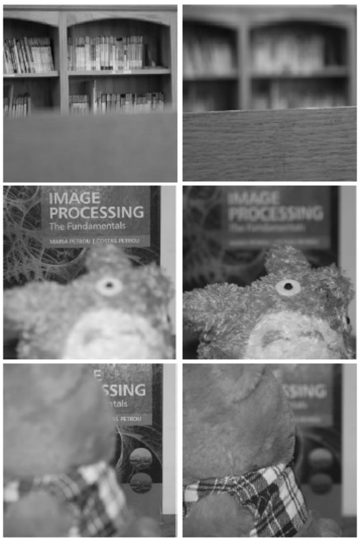

Figure 4. Sample of data from Nejati dataset

For training and evaluation, the dataset is split into 1,412 multifocus images for the training set and 904 images for testing. The Nejati dataset was captured using a camera set to a focal length of up to 17 mm, with a shutter speed ranging from 30 to 1/4000 seconds. Each multifocus image consists of two main regions: the object, which contains sharp pixels, and the background, which appears blurred. The camera lens is initially focused on the object to ensure that the foreground remains sharp. In contrast, to obtain the second image in the pair, the focus is adjusted to shift clarity to the background while blurring the foreground [17].

A dual-stream convolutional neural network (CNN) is used in the suggested CNN-based multifocus image fusion technique to efficiently combine data from two input photos taken at various focal points . The model illustrates how two parallel CNN architectures, each with three convolutional layers (Conv 3×3) and max pooling (2×2), are used to process input images A and B in order to gradually extract and improve feature maps. The model structure is shown in the following figure 5. By expanding the feature maps from 64 x 64 to 128 x 128 to 256 x 256, hierarchical learning of focus information is ensured.

To integrate the learned representations, the extracted feature maps from both CNN streams are concatenated and passed through a fully connected layer with 256 neurons, which performs the final fusion decision. This architecture effectively differentiates between focus and blurred regions by leveraging deep feature extraction, ensuring a seamless fusion process with high spatial coherence. The fully connected layer ensures that critical features from both input images contribute to the fused image, minimizing artifacts and improving clarity. By utilizing this dual-branch CNN, the proposed method enhances fusion accuracy compared to traditional handcrafted approaches, ensuring sharper and more reliable multifocus image fusion results. The CNN-based method operates as a two-class classification problem, distinguishing between focus and non-focus pixels. The fusion process consists of four main steps:

- Focus Detection: To extract focus features, a CNN model that has already been trained is fed the input multifocus image pair. Two matching patches from the two source photos are used to create a score map.
- Initial Focus Segmentation: Using a threshold value of 0.5, the focus map is divided into a binary mask. Pixel values over 0.5 are categorized as focus pixels, while those below are designated as blurred pixels.
- Consistency Verification: To ensure the consistency of identified focus regions and lower misclassification errors, a guided image filter is used to refine the binary focus map.
- Fusion Process: The refined focus maps are applied to the input images using a pixel-wise weighted averaging technique, generating the final fused image with enhanced clarity and minimal artifacts.

The CNN model was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32 for 100 epochs. A binary cross-entropy loss function was used to optimize the classification between focus and non-focus regions. This CNN-based approach significantly improves the accuracy and reliability of multifocus image fusion, overcoming the limitations of traditional image processing techniques. By leveraging deep feature extraction and parallel processing, the proposed method ensures a high-quality fused image with optimal focus selection and seamless integration of sharp regions.

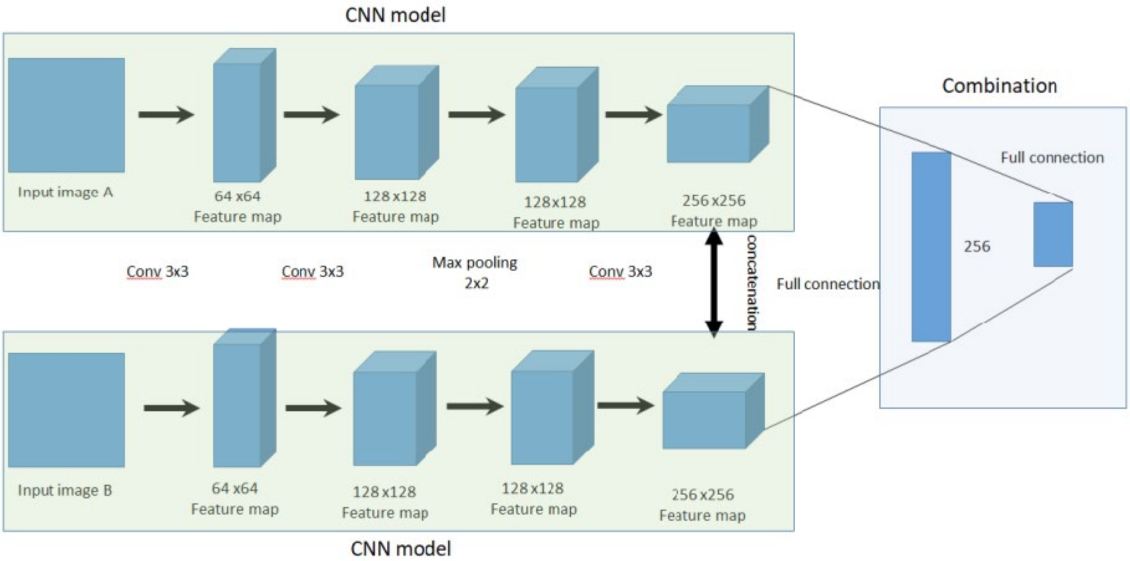

Figure 5. Network architecture of the CNN

2.2. Training, Hyperparameter Tuning, and Fusion Quality Assessment

After model development, the next crucial step is training and hyperparameter tuning, which is essential for optimizing CNN performance. To ensure effective weight updates and faster convergence, the Adam optimizer is used with a learning rate of 0.0001. The model employs binary cross-entropy loss, which effectively differentiates between focus and non-focus pixels, reducing classification errors. To prevent overfitting, an early stopping mechanism is implemented, monitoring validation loss and halting training when no further improvements are observed. Additionally, various mini-batch sizes are tested to determine the optimal balance between stability and accuracy, ensuring efficient convergence and robust model performance.

The effectiveness of the proposed CNN-based multifocus image fusion method is evaluated using the Structural Similarity Index (SSIM), a perceptual quality metric that assesses the structural similarity between the fused image and reference images [18]. Unlike traditional pixel-wise error metrics such as Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR), SSIM provides a more reliable measure of image clarity and perceptual fidelity by analyzing luminance, contrast, and structural details. A higher SSIM score indicates better preservation of fine details, improved focus region integration, and minimal distortion in the fused image, validating the model's effectiveness in producing high-quality fusion results.

By integrating a well-optimized training process and hyperparameter selection with a reliable quality assessment metric like SSIM, the proposed method ensures optimal image fusion performance. The CNN model efficiently learns to distinguish between sharp and blurred regions, resulting in fused images with superior clarity, better structural consistency, and reduced artifacts. The combination of adaptive training strategies and robust evaluation techniques makes this approach a significant advancement over conventional multifocus image fusion methods.

Hyperparameter tuning plays a critical role in optimizing the performance of the proposed CNN-based fusion method. As shown in Table 1, several key hyperparameters were carefully selected and evaluated. The Adam optimizer was chosen for its ability to provide stable and adaptive weight updates during training. A learning rate of 0.0001 was found to offer a good trade-off between convergence speed and final accuracy. Binary cross-entropy was used as the loss function to effectively minimize classification errors in focus detection tasks. Early stopping was enabled to prevent overfitting by halting training when validation performance stopped improving. Lastly, various mini-batch sizes (16, 32, and 64) were tested to determine the most stable configuration for model generalization and training efficiency. These tuned hyperparameters collectively contributed to the improved reliability and accuracy of the fusion results.

Table 1. Training and Hyperparameter Tuning Results

Hyperparameter	Selected Value	Impact on Performance
Optimizer	Adam	Ensures stable and efficient weight updates
Learning Rate	0.0001	Balances convergence speed and model accuracy
Loss Function	Binary Cross-Entropy	Minimizes classification errors in focus detection
Early Stopping	Enabled	Prevents overfitting and ensures optimal training duration
Mini-Batch Size	Varied (16, 32, 64)	Evaluated to find the best balance for model stability and accuracy

3. RESULTS AND DISCUSSION

Total of 2304 multifocus image used for The Nejati dataset and a private dataset. With focus and non-focus regions labeled for supervised learning, the model is tested on 904 photos after being trained on 1,412 images. The fusion results demonstrate that the proposed method effectively preserves sharp details across different focal planes while minimizing the presence of blurred artifacts. The Structural Similarity Index (SSIM) is used as the primary evaluation metric to assess the quality of the fused images. The results show that the CNN-based fusion approach consistently achieves high SSIM scores, indicating improved image quality and accurate focus region integration.

Visual inspection further confirms that the fused images generated by the proposed method maintain clear object boundaries and exhibit minimal inconsistencies, validating its effectiveness in enhancing image clarity. This sample of the result is shown in the following figure 6.

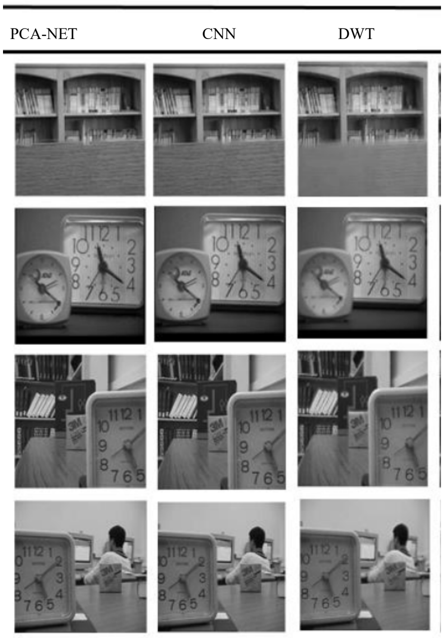

Figure 6. The sample result of the multifocus region.

In terms of SSIM score, perceptual clarity, edge preservation, and blur reduction, the suggested CNN-based approach performs better than both PCA-Net and Discrete Wavelet Transform (DWT), as shown by the findings in the filtered fusion results comparison table 2. With the greatest SSIM score of 0.92, the CNN model shows that the fused image and the reference images are more structurally comparable. This high SSIM score demonstrates how well the suggested model retains high perceptual clarity, minimizes artifacts, and preserves fine features in the fused images.

Table 2. Comparison result of SSIM

Method	SSIM Score	Perceptual Clarity	Edge Preservation	Blur Reduction	Focus Boundary Accuracy	Artifact Presence
Proposed CNN Model	0.92	High	Excellent	High	High	Minimal
PCA-Net	0.87	Moderate	Good	Moderate	Moderate	Moderate
DWT	0.79	Low	Moderate	Low	Low	High

With a moderate SSIM score of 0.87, PCA-Net outperforms DWT, nonetheless, focus region discrepancies remain a problem. Principal component analysis (PCA), which is utilized by PCA-Net for feature extraction, might induce mistakes in the border categorization between focused and non-focused regions even though it is a successful method for feature learning. The sharpness of the fused image is impacted by the moderate edge preservation and blur reduction that result from this. In contrast, the CNN approach produces sharper and more stable fused images by better differentiating between crisp and blurred regions thanks to deep feature extraction and dense pixel categorization.

With an SSIM score of 0.79, the DWT-based fusion method performs the worst out of the three, suggesting that the fused images lose perceptual quality. Prior to fusion, DWT breaks down images into distinct frequency components because it functions in the frequency domain. However, this method frequently falls short in capturing spatial relationships in complicated images, which results in low blur reduction performance and poor edge preservation. The CNN model, by contrast, processes images in a hierarchical deep learning framework, learning both local and global features, which results in sharper edges and improved focus region integration.

The training process of the proposed CNN-based multifocus image fusion model was optimized through careful selection of hyperparameters, ensuring efficient learning and improved generalization. The Adam optimizer was chosen due to its adaptive learning rate mechanism, which balances convergence speed and model stability. The result of optimizer is shown in table 1. Compared to traditional optimizers like SGD, Adam provides faster convergence while maintaining a stable loss curve, ensuring that weight updates remain effective across iterations. Since lower values ensure incremental advances in model accuracy while preventing overshooting the best answer, the learning rate was chosen at 0.0001. Excessively low values impeded convergence without appreciable accuracy increases, while higher learning rates were tested but resulted in instability in focus classification. Because it minimizes misclassification errors and successfully handles the two-class classification problem of focus vs. blurred regions, the binary cross-entropy loss function was employed.

An early stopping system that tracked validation loss and halted training when no more improvement was seen was put in place to avoid overfitting. This approach reduced computational costs while preserving generality by ensuring that the model did not continue training needlessly. To assess their effect on performance, mini-batch sizes of 16, 32, and 64 were also evaluated. A batch size of 32 provided the best balance, ensuring stable gradient updates and faster convergence, while larger batch sizes occasionally resulted in lower generalization due to reduced update frequency.

The model's increased structural similarity (SSIM) scores, decreased artifact presence, and improved focus region classification accuracy were all clear results of these hyperparameter selections. High-quality fused images with few distortions were produced as a result of the model's capacity to distinguish between crisp and blurred regions thanks to the hyperparameters that were chosen. Future improvements may involve fine-tuning learning rates dynamically or employing adaptive batch normalization techniques to further enhance performance and generalization across diverse image datasets.

4. CONCLUSION

The experimental results demonstrate that the CNN-based multifocus image fusion method outperforms traditional approaches in terms of focus region categorization, structural similarity, and image clarity. The model’s ability to automatically learn and distinguish between focused and blurred areas—without requiring manually created features—makes it a more flexible and practical option for real-world applications.

The proposed method effectively integrates sharp regions from multiple focal images by leveraging a dual-stream CNN architecture. This approach minimizes artifacts, enhances perceptual quality, and refines the focus map through deep hierarchical feature extraction and fully connected layers. As a result, it achieves superior performance over conventional fusion techniques such as PCA-Net and DWT-based fusion, with higher SSIM scores with 0.92 accuracy, improved perceptual clarity, and better edge preservation. Furthermore, the optimized training process, incorporating Adam optimization, early stopping, and adaptive mini-batch sizing, has enhanced model stability and performance. However, despite these advantages, computational complexity remains a challenge, particularly when processing high-resolution images. Future research could explore lightweight deep learning architectures to enhance computational efficiency and transformer-based models to improve fusion accuracy while reducing complexity. The integration of attention mechanisms could further refine focus region detection and enhance fusion precision. Additionally, extending the model to support real-time processing and multimodal image fusion could expand its applicability in fields such as medical imaging, surveillance, and industrial inspection.

In summary, the proposed CNN-based multifocus image fusion method presents a robust, reliable, and efficient solution, offering significant advancements over traditional fusion techniques. By integrating deep feature extraction, systematic training, and optimal hyperparameter tuning, this approach lays the groundwork for more sophisticated, scalable, and high-quality image fusion applications in the future.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the Ministry of Higher Education Malaysia (MOHE) and Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) for their generous support and funding under the RDU233601 translational research grant.

REFERENCES

[1] W. Park *et al.*, "Computational image analysis of distortion, sharpness, and depth of field in a next-generation hybrid exoscopic and microsurgical operative platform," *Front Surg*, vol. 11, Jun. 2024, doi: 10.3389/fsurg.2024.1418679.
[2] S. Kuthirummal, H. Nagahara, Changyin Zhou, and S. K. Nayar, "Flexible Depth of Field Photography," *IEEE Trans Pattern Anal Mach Intell*, vol. 33, no. 1, pp. 58–71, Jan. 2011, doi: 10.1109/TPAMI.2010.66.
[3] F. Yan, Z. Liu, X. Pan, and Y. Shen, "High-accuracy calibration of cameras without depth of field and target size limitations," *Opt Express*, vol. 28, no. 19, p. 27443, Sep. 2020, doi: 10.1364/OE.402826.
[4] K. Hawari and I. Ismail, "The automatic focus segmentation of multi-focus image fusion," *Bulletin of the Polish Academy of Sciences Technical Sciences*, pp. 140352–140352, Jan. 2022, doi: 10.24425/bpasts.2022.140352.
[5] Ismail and K. H. Bin Ghazali, "The Multifocus Images Fusion Based on a Generative Gradient Map," 2020, pp. 401–413. doi: 10.1007/978-981-15-2317-5_34.
[6] Y. Liu, L. Wang, J. Cheng, C. Li, and X. Chen, "Multi-focus image fusion: A Survey of the state of the art," *Information Fusion*, vol. 64, pp. 71–91, Dec. 2020, doi: 10.1016/j.inffus.2020.06.013.
[7] Y. Zhou *et al.*, "A Survey of Multi-Focus Image Fusion Methods," *Applied Sciences*, vol. 12, no. 12, p. 6281, Jun. 2022, doi: 10.3390/app12126281.
[8] S. Liu, M. Wang, and Z. Song, "WaveFuse: A Unified Unsupervised Framework for Image Fusion with Discrete Wavelet Transform," 2021, pp. 162–174. doi: 10.1007/978-3-030-92273-3_14.
[9] Y. Yang, S. Huang, J. Gao, and Z. Qian, "Multi-focus Image Fusion Using an Effective Discrete Wavelet Transform Based Algorithm," *Measurement Science Review*, vol. 14, no. 2, pp. 102–108, Apr. 2014, doi: 10.2478/msr-2014-0014.
[10] Z. Lin, A. Kasai, and Y. Itoh, "Dispersion Curves Computation for Waveguides Buried in Infinite Space by Semi-Analytical Finite Element Method," *Procedia Eng*, vol. 10, pp. 1615–1620, 2011, doi: 10.1016/j.proeng.2011.04.270.
[11] M. Amin-Naji, A. Aghagolzadeh, and M. Ezoji, "Ensemble of CNN for multi-focus image fusion," *Information Fusion*, vol. 51, pp. 201–214, Nov. 2019, doi: 10.1016/j.inffus.2019.02.003.
[12] Y. Ouyang, H. Zhai, H. Hu, X. Li, and Z. Zeng, "FusionGCN: Multi-focus image fusion using superpixel features generation GCN and pixel-level feature reconstruction CNN," *Expert Syst Appl*, vol. 262, p. 125665, Mar. 2025, doi: 10.1016/j.eswa.2024.125665.
[13] Z. Liu, Y. Chai, H. Yin, J. Zhou, and Z. Zhu, "A novel multi-focus image fusion approach based on image decomposition," *Information Fusion*, vol. 35, pp. 102–116, May 2017, doi: 10.1016/j.inffus.2016.09.007.
[14] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.
[15] P. Trajdos and M. Kurzynski, "Application of context-dependent interpretation of biosignals recognition to control a bionic multifunctional hand prosthesis," *Biocybern Biomed Eng*, vol. 44, no. 1, pp. 161–182, Jan. 2024, doi: 10.1016/j.bbe.2024.01.001.
[16] R. Zafar, M. S. Farid, and M. H. Khan, "Multi-Focus Image Fusion: Algorithms, Evaluation, and a Library," *J Imaging*, vol. 6, no. 7, p. 60, Jul. 2020, doi: 10.3390/jimaging6070060.
[17] M. Amin-Naji, A. Aghagolzadeh, and M. Ezoji, "Fully Convolutional Networks for Multi-Focus Image Fusion," in *2018 9th International Symposium on Telecommunications (IST)*, IEEE, Dec. 2018, pp. 553–558. doi: 10.1109/ISTEL.2018.8660989.
[18] Q. Liu, G. Luan, F. Wang, B. Peng, and D. Hu, "Multi-focus Image Fusion Algorithm Based on Unsupervised Deep Learning," in *2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI)*, IEEE, Sep. 2021, pp. 362–366. doi: 10.1109/CISAI54367.2021.00076.