JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 02
2024
Research Article
https://doi.org/10.56979/702/2024

CNN and Gaussian Pyramid-Based Approach For Enhance Multi-Focus Image Fusion

Kahkisha Ayub[1*], Muhammad Ahmad[2], Fawad Nasim[2], Shameen Noor[1], and Kinza Pervaiz[1]

[1]Department of Software Engineering, Superior University, Lahore, 54000, Pakistan.
[2]The Superior University Lahore, 54000, Pakistan.
[*]Corresponding Author: Kahkisha Ayub. Email: kahkiayub12@gmail.com

Received: April 21, 2024 **Accepted:** August 12, 2024 **Published:** September 01, 2024

Abstract: Achieving sharp focus in both foreground and background areas simultaneously in digital photography is challenging due to the limited depth of field (DOF) in DSLR cameras. This often results in blurred images with isolated details, hindering the capture of clear, high-quality photographs. Existing multi-focus image fusion models struggle with issues related to image quality and managing input images taken from varying angles. To address these challenges, we introduce a novel multi-focus image fusion method that integrates Convolutional Neural Networks (CNNs) with Gaussian Pyramid techniques. Our approach follows a four-step process: visual search, initial segmentation, compression analysis, and final synthesis. The Gaussian pyramid technique enhances edge detection by progressively reducing the image size and facilitating the identification of automatic objects. By leveraging these combined techniques, our method ensures the generation of uniformly focused images. We rigorously evaluate our model using metrics such as Perceptual Image Quality Evaluator (PIQE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and connectivity metrics. Our enhanced model demonstrates superior performance, achieving a 4.70% improvement in PIQE compared to previous CNN-based methods. This confirms the effectiveness of our approach in producing clear, high-quality fused images.

Keywords: CNN; Image Fusion; Gaussian Pyramid; Multi Fusion; Mat-Lab

1. Introduction

Capturing high-resolution images of all subjects is becoming increasingly difficult with devices such as single-lens digital cameras [1]. Generally, under the same contrast conditions, only deep-field (DOF) objects appear sharp, while others appear flat. The basic method for obtaining full-field images is multi-field image fusion, which means combining images of the same sequence captured on different focal lenses. Multi-view is an important area of image fusion, and many of the developed methods can be employed for various image fusion applications, for instance, optical infrared and ultra-local imaging [2]. This two-pronged approach highlights the importance of studying image synthesis in multiple ways, making it an active area of research in the image-processing community. In the last few years, many imaging techniques have been proposed [3], [4], which mainly consist of two categories: temporal domain (TD) and spatial domain (SD). Classical TD stands for Transform domain fusion methods and depends on multiple-transform theories (MST) [5], which have been used for more than three decades since the fusion model is based on the Laplacian pyramid (LP). Since then, many MST-based surveys have been published. Important examples are the matrix particle (MP) method, the discrete weight transform (DWT), the two-threshold convolutional transform (DTCWT), and the nonlinear convolutional transform (NSCT) [6], [7]. This MST process usually follows three steps: reduction, decomposition, and reconstruction. The basic idea is that the level of activity of the image can be evaluated by analysts with different options for different parameters [8]. In addition to the selection of the MST series, the design of the algorithms for the synthesis of thermocouples is of great importance, and many studies have focused on this.

Another class of domain methods is based on image segmentation [9], which contrasts with spectral methods, whose fusion quality depends on the spectrum. In the last few years, new pixel-based (SD) methods using contour data have been introduced, which have achieved advanced outcomes and results in multimodal image fusion. To heighten the integration performance, these techniques typically use challenging integration strategies to quantitatively measure their performance. As is well known, power scaling and integration laws are two important factors for both transformation domain and wide domain image fusion methods. The major issue while designing by hand is a challenging task and in some ways, it is almost impossible to create a suitable design that includes all the necessary elements. Similar work was done by Yu Li et al. [10] the multi-focus image fusion presented in them is rooted in a deep CNN approach. This process starts with preparing high-quality images and corresponding color versions for CNN training. This ensures that the reticle can accurately distinguish between in-focus and out-of-focus areas. The CNN architecture is designed to match source images to a focus map and extract quality information at the pixel level. The network consists of several complex layers capable of extracting and processing complex image details. The focus map is generated by a CNN that compares the intensity level of the source images to ensure reliable pixel accuracy. This deep learning approach overcomes the limitations of traditional manual approaches and more complex performance and integration rules. Turbulence is very reliable and the procedures are just a robustness check to ensure high-quality synthesis results. In addition, the execution speed of this method and concurrent processing are fast enough for practical applications. This study briefly investigates the ability of the trained CNN model to perform other image fusion job, along with infrared visible image rendering, medical images, and multi-image fusion. In this paper, we compare the hybrid methods identified in CNN and their best performance. We evaluate our parameter using several parameters, including PIQE stands for pattern-aware image, PSNR stands for peak signal-to-noise ratio, SSIM stands for structural similarity-index, and entropy [11]. Our experimental results show that this approach not only works effectively with different images, but also avoids images similar to ours with all methods.

To sum up, this article makes three key contributions:

1. **Addressing Image Size Discrepancies:** The proposed method effectively handles the challenge of fusing images that differ in size, ensuring seamless integration regardless of varying dimensions. This capability is crucial for maintaining consistency and accuracy in the fusion process.
2. **Enhancing Image Quality and Noise Reduction:** By combining CNNs with Gaussian Pyramid techniques, our approach significantly improves the quality of the fused images. It not only enhances the clarity and detail but also effectively reduces noise, resulting in cleaner and more visually appealing images.
3. **Superior Performance over Existing Methods:** Our method demonstrates substantial improvements over previous approaches, as evidenced by better performance metrics. The enhanced model yields higher image quality and more accurate fusion results, setting a new standard in multi-focus image fusion.

1.1. Problem Statement

It is particularly difficult to capture a single image at all focal lengths in scenes with objects at different distances from the camera. Many image fusion methods often lead to artifacts and loss of detail, especially at the boundary between in-focus and out-of-focus regions. In my experience, traditional methods do not preserve background information well and often do not have enough variability in merged images.

The rest of the article is organized as follows. In Section 2, we introduce convolutional neural networks (CNNs). In Section 3 we will discuss the literature review. Section 4 presents the proposed method. Section 5 presents the results. We conclude the paper in the last section.

2. Literature Review

Feng Zhou et al. [12] presented a novel approach using a fully convolutional pyramidal neural network (PFCN) for multifocal image fusion. Their approach involves a new network architecture that combines a coding subnet and a pyramid fusion subnet. The encoder extracts hierarchical features from input images and captures important spatial details at multiple levels. These features are followed by a pyramid fusion network that uses multidimensional spatial information to create a high-resolution, full-resolution image. The model uses a gradient jitter reduction feature to improve edge protection and reduce friction. Extensive research has been conducted to evaluate the effectiveness of traditional methods in

terms of quantitative and qualitative visual metrics. The results showed that as spatial details become less precise, PFCN performs better on multifocal image fusion tasks. Sneha Singh and R.S.

Anand [6] presented a two-stage clustering framework and a case study was presented. The method emphasizes the preservation of layer information such as texture, detail, brightness, and color, which improves the synthesis accuracy. In the proposed method, a novel clustering method based on n-norms is used to classify the keywords with textual information. Visual features are extracted from the lowest and most basic levels and compared with the conservation rules based on the Visual Structure (VS) model. Experimental results demonstrate the effectiveness of the noise removal method and the preservation of ambiguity without sacrificing computational resources. Comparative analysis shows that the proposed method outperforms state-of-the-art methods by preserving the key information, increasing the discrimination, and improving the accuracy of multimodal image fusion.

Hongmei et al. [13] proposed fixed the auto-encoding system severely impacting image fusion models that violate manual compositing rules. Traditional autoencoder-based convolutional networks use dual-stream KNNs with the same encoder structure, which makes feature extraction for convolutional operations difficult. Moreover, they often fail to capture the characteristics of still and real images. To address this problem, a new automatic encoding method based on image segmentation has been developed. This model includes an encoder module that combines CNN and Transformer channels to represent local and global features from input images. In addition, it includes distinct features and improvements created for images and animations that preserve the unique characteristics of each scene. This feedback is fed into the combiner and fed into the decoder to output the modified image. Experimental results in three cases show that this method maintains good retrieval accuracy and intelligibility and outperforms many state-of-the-art methods in both pilot and experimental studies.

Moreover, the complex images generated by these clusters achieve a higher target detection rate, which means better subsequent use. Zicheng Nian and Cheolkon Jung [14] the authors proposed a CNN-based multi objective image fusion framework using illumination features, aiming at feature quantification for network training. The light field data are used to generate depth images and reflectance maps, which are then used to train a network that generates pixel-wise maps using fully convolutional networks, similar to binary segmentation methods which are located in the patch directory. For this purpose, a multifocal image dataset is constructed consisting of multifocal images and their ground truth. This technique shows excellent performance in terms of communication index (QMI), quantity of space (QSF), edge preservation (QAB/F), and quantum signals (QSSIM), which provides consistent and smooth edges and as well as more consistent evidence than the existing ones materials and methods. The proposed method significantly improves image quality, which makes it useful for underlying tasks such as target detection.

Shiveta Bhat and Deepika Koundal [15] provides an overview of the multiple image fusion (MFIF) process, which combines various source images at different levels to create a single target image for enhancement and information. The optical sensor's small depth of field (DOF) makes it difficult to extract all relevant information from a single image. It requires the use of multiple high-resolution images. The authors propose a new classification scheme that divides existing MFIF methods into four general categories: opportunistic classification, adaptive classification, deep learning, and their combinations, each with its shortcomings and challenges. The parametric parameters are considered “valid” and “not supported”. Thirty publicly available image pairs were used for comparative analysis of image fusion levels. The paper identifies several outstanding issues and provides suggestions for future work, highlighting the need for robust algorithms that can adapt to different datasets and provide better fusion results, thus increasing image depth of field. This work is a valuable resource for researchers seeking to develop new MFIF methods that overcome the limitations of existing methods. Shifeng et al.

[16] Proposed a new image classification algorithm aimed at improving object detection accuracy in diffraction images using non-subsampling contour transform (NSCT) and convolutional neural network (CNN). Initially, a fast adaptive filter and a pulsed convolutional neural network (PCNN) are used to reduce the noise in the refractive angle image. The linear polarization and polarization angle images are combined to generate a pattern of diffraction behavior, which is then compared to the intensity pattern using NSCT [19] [20]. CNN was used to extract features from high-resolution NSCT images, and zero-phase component analysis (ZCA) correlated these images. The coupling coefficient for the high-frequency sub-range was calculated by calculating the image formation energy map, the coupling coefficient for the low-frequency sub-range was calculated using the energy method, and the resulting image l' was reconstructed using NSCT. Experimental results show that the embedded image gradient is 51.3% higher

and the spatial frequency is 35.1% higher, indicating better image processing and segmentation performance. The proposed method shows sensitive and effective results to achieve the desired object recognition performance despite the increased computation time due to the use of VGG-19 channels in the above frequency sub band. Future work could focus on improving real-time detection and extraction methods.

Hao Zhang et al. [17] they investigated the impact of deep learning on image classification, which can improve image quality and data classification in many tasks, such as image description, object recognition, disease detection, and deep understanding. It provides an overview of recent developments in deep learning image classification and organizes them by their applications and technologies. Studies have shown the effectiveness of machine learning techniques such as generative adversarial networks and Autoencoders to improve fusion performance. This includes qualitative and quantitative assessments of representative approaches to various integration practices, highlighting their strengths and identifying current challenges. The main challenges discussed include the need for unregistered fusion algorithms to handle unregistered speech images, strategies for efficiently combining images of different resolutions, and developing integration techniques that meet application requirements certain. This paper highlights the importance of real-time capabilities in image fusion algorithms and the need for reliable quality assessment metrics for future research and application development in this area. Muhammad Ahmad et al. [18] Proposed a method based on fuzzy based focus measure (FBHFM) to fuse multiple images. They used particle swarm optimization (PSO) to determine the optimal block size for feature optimization. This method combines different Laplacian, gray level transform, and optical focus measurements with fuzzy methods to produce a well-focused image. Their systems have been tested using state-of-the-art technology and proven to work efficiently at low cost.

3. Methodology

The proposed CNN model is designed to effectively learn and extract focused regions from multi-focus input images using Gaussian pyramid decomposition and various convolutional and pooling operations. Below is a detailed description of the CNN model architecture according to the provided diagram and your research objectives.

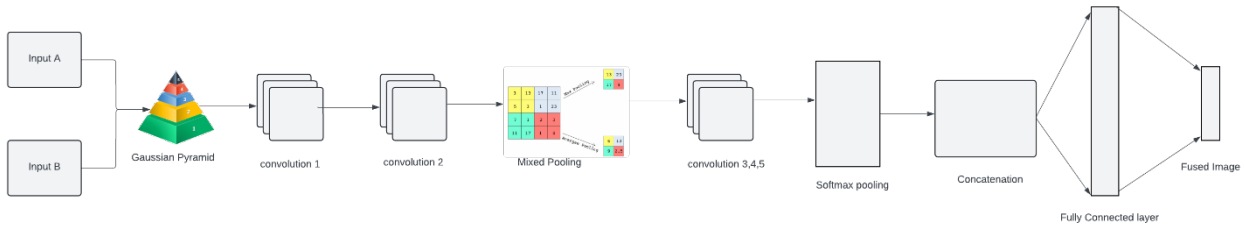

Figure 1. Enhance CNN Model

In the multi-focus process using CNN and Gaussian pyramid techniques, the input layer depend on two images, Input A and Input B, each with different focal regions. Gaussian pyramid decomposition is employed to create a multi-scale representation of these images, capturing details across various resolutions through Gaussian filtering and subsampling. Convolutional layers follow, beginning with Convolution 1 (3x3 filters, 32 filters, ReLU activation) extracting low-level features like edges. Subsequent layers (Convolution 2 to 5) deepen in complexity with increasing filter sizes and numbers (64, 128, 256, 512 filters), culminating in deep feature maps encoding high-level semantic information. Mixed pooling, combining max and average pooling, balances feature detail and generalization, while Softmax pooling highlights focused regions via a Softmax function. A concatenation layer integrates features from different scales and operations, feeding into fully connected layers (1024, 512, 256 neurons, ReLU activation) for final decision-making based on focus measures. The output layer produces a fused image by selecting and combining the best-focused regions from both input images.

3.1. Network Design

3.1.1. Focus Detection

Suppose "A" and "B" illustrate two images as a source. According to proposed fusion algorithm, if the origin image is color, it will be converted to grayscale first. Now A^ and B^ represent the grayscale versions of A and B. Grayscale images are fed into CNN models A^ and B^ to obtain the resulting mapping.

Each value in S ranges from 0 to 1 and indicates the sharpness quality of the spot corresponding to the 16 × 16 dimension in the source image. Values of 0 or close to 0 indicate that the source image patch A^ or B^ is more focused. For doublet adjacent coordinates in "S", the parallel coordinates in original based image are scaled to the two-pixel level.

To obtain a focus map (denoted FM) of the equivalent size as the origin image, the significance of per coefficient in "S" is divided by average of the overlapping pixels of all pixels in its corresponding patch in FM (Figure 2). Accurately detects the brightness of the focus, with brighter areas closer to 1 which represent white color or 0 represent black color, simultaneously plain region value 0.5 is gray.

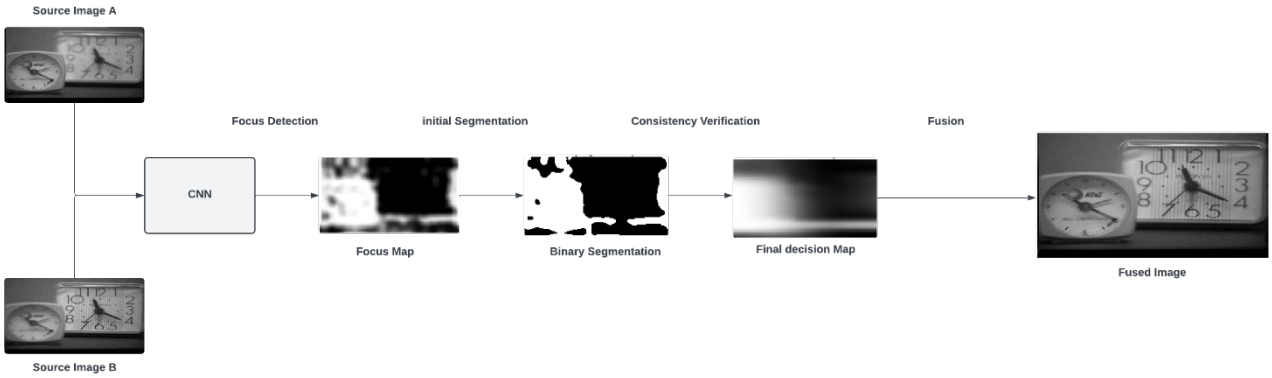

Figure 2. Diagram of Proposed Enhanced Algorithm of CNN Model

3.1.2. Initial Segmentation

In order to sustain as much beneficial information as much as possible, the FM should be existed further procession. In this approach, we use a "mixed integration" strategy for FM processing. This means that the output of the integration function is multiplied by the retention probability p. The process is described by the following formulas [10]:

$$d_k^l = p.Max(x_k) \ominus p.Avg(x_k) \quad (1)$$

$$d_k^l = p.y_k^l \quad (2)$$

Specifically, a defined threshold value is 0.5 is used to segment FM into a binary map B according to the taxonomy rule of CNN experienced model.

3.1.3. Consistency verification

Segmented binary maps may contain spurious pixels that can be conveniently remove by using the signature elimination technique. By default, any region smaller than the boundary of a given region belongs to a binary map. Sometimes there are small holes in the source image. In this case, the user can set the default value manually or, if necessary, set the default value to zero, which means that the default option is not used. In the next segment, we show the resulting classifiers can accomplish higher accuracy. In area, layer is generally H × 0.01 × W, in which H and W are the height and width of every single layer.

You can see the first score card that appears after using this strategy. Using the source resolution map and the weighted averaging method, the fused image showed some unwanted artifacts at the border of sharp and un-sharp areas. To solve this problem, we use a targeted kernel to enhance the quality of the previous "S" (Score map). A vector kernel is an edge-preserving kernel that transfers structural information from a vector image to a filtered output of the input image. The preliminary collected images serve a guide for initial review of the decision map. The manual filtering algorithm consists of two arbitrary parameters: local window width is "r" and the parameter ε is regularization. In this analysis, our team set r is equivalent "8" and ε is equivalent "0.1". The first results show a significant improvement.

3.1.4. Fusion

Decision map (D), the corresponding image represent as F is computed apply the pixel-scale averaging rule:

$$F(x,y) = D(x,y).A(x,y)+(1-D(x,y)).B(x,y) \quad (3)$$

Figure 1. (1) Input Image 1, (2) Input image 2, (3) output Fused Image(F)

4. Results and Discussion

Table 1. Evaluation Matrix

Model	PIQE	PSNR	SSIM	Entropy
Yu Lie et al.	38.5	29.61	0.67	7.5
Enhance Model	**36.69**	**30.29**	**0.68**	7.5

According the **Table 1.** Considering the image processing model using the "Children" test dataset, two models were compared: Yu Liu et al. [10] and the Enhance Model. According to the findings of Yu Liu et al. [10] this model obtained a PIQE score of 38.5, indicating its perceived image quality as judged by human observers. It recorded a PSNR of 29.61 dB, indicating good fidelity to the original image, with an SSIM of 0.67, indicating moderate structural similarity, and an entropy of 7.5, indicating high information content of the image. In contrast, the Enhance model outperforms the previous studies. An average PIQE score of 36.69 was obtained, indicating a slight improvement in image quality. The Enhance model also recorded a PSNR of 30.29 dB, indicating improved visual quality compared to the original image, and a SSIM of 0.68, indicating artifact similarity. The entropy remains at 7.5, indicating that the generated images contain similar or unexpected information.

Overall, these results show that the **Enhanced model** is better by predicting image quality and texture while keeping the content level constant. These results provide important insight into the performance of these models and are suitable for image processing, especially where image quality is critical to human perception.

5. Conclusions

This study presents an improvement over the traditional image fusion method. Our Enhance model improved the quality of fused image. This study demonstrates the potential of the Gaussian pyramid method to improve the quality of fused images. By combining features and using deep fusion and fusion operations, the proposed model captures and integrates class information from different sources.

Future work aims to find options to improve the accuracy of the fusion process by considering network planning and better imaging.

References

1. Vibashan VS, Jeya Maria Jose Valanarasu, Poojan Oza, and Vishal M. Patel," IMAGE FUSION TRANSFORMER", in 2022.
2. Liangzhi Li, Ling Han, Mingtao Ding, and Hongye Cao,"Multimodal image fusion framework for end-to-end remote sensing image registration",in 2023, doi:10.1109/TGRS.2023.3247642
3. Y. Pan, Y. Zhang, Y. Wei, and Q. Liu, "Saliency detection based on the fusion of spatial and frequency domain analysis," in *2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)*, Xiamen, China: IEEE, Oct. 2019, pp. 577–581. doi: 10.1109/EITCE47263.2019.9095064.
4. Xin Zhang , Xia Wang , Changda Yan , and Qiyang Sun, "EV-Fusion: A Novel Infrared and Low-Light Color Visible Image Fusion Network Integrating Unsupervised Visible Image Enhancement", in 2024 , IEEE SENSORS JOURNAL, VOL. 24, NO. 4
5. C. Ma, X. Mu, and D. Sha, "Multi-Layers Feature Fusion of Convolutional Neural Network for Scene Classification of Remote Sensing," *IEEE Access*, vol. 7, pp. 121685–121694, 2019, doi: 10.1109/ACCESS.2019.2936215.
6. Sneha Singh, Student Member, IEEE, R.S. Anand, "Multimodal Medical Image Fusion using Hybrid Layer Decomposition with CNN-based FeatureMapping and Structural Clustering" in 2019, OI 10.1109/TIM.2019.2933341, IEEETransactions on Instrumentation and Measurement.
7. WENCHENG WANG 1, (Member, IEEE), XIAOJIN WU1,XIAOHUI YUAN 2, (Senior Member, IEEE), AND ZAIRUI GAO," An Experiment-Based Review of Low-LightImage Enhancement Methods", in 2022, doi: *10.1109/ACCESS.2020.2992749.*
8. SimrandeepSingha,b, HarbinderSinghc,∗, GloriaBuenod, OscarDenizd, SartajvirSinghe et al., "Areview of image fusion: Methods, applications and performance metrics", in 2023.
9. Meilong Xu 1 , Linfeng Tang 1 , Hao Zhang, Jiayi Ma," Infrared and visible image fusion via parallel scene and texture learning ", in 2022.
10. Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017, doi: 10.1016/j.inffus.2016.12.001.
11. V. Vani and K. V. M. Prashanth, "Image Enhancement of Wireless Capsule Endoscopy Frames Using Image Fusion Technique," *IETE J. Res.*, vol. 67, no. 4, pp. 463–475, Jul. 2021, doi: 10.1080/03772063.2018.1554459.
12. F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Pyramid Fully Convolutional Network for Hyperspectral and Multispectral Image Fusion," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 5, pp. 1549–1558, May 2019, doi: 10.1109/JSTARS.2019.2910990.
13. H. Wang, L. Li, C. Li, and X. Lu, "Infrared and Visible Image Fusion Based on Autoencoder Composed of CNN-Transformer," *IEEE Access*, vol. 11, pp. 78956–78969, 2023, doi: 10.1109/ACCESS.2023.3298437.
14. Z. Nian and C. Jung, "CNN-Based Multi-Focus Image Fusion with Light Field Data," in *2019 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan: IEEE, Sep. 2019, pp. 1044–1048. doi: 10.1109/ICIP.2019.8803065.
15. S. Bhat and D. Koundal, "Multi-focus image fusion techniques: a survey," *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 5735–5787, Dec. 2021, doi: 10.1007/s10462-021-09961-7.
16. S. Wang, J. Meng, Y. Zhou, Q. Hu, Z. Wang, and J. Lyu, "Polarization Image Fusion Algorithm Using NSCT and CNN," *J. Russ. Laser Res.*, vol. 42, no. 4, pp. 443–452, Jul. 2021, doi: 10.1007/s10946-021-09981-2.
17. H. Zhang, H. Xu, X. Tian, J. Jiang, and J. Ma, "Image fusion meets deep learning: A survey and perspective," *Inf. Fusion*, vol. 76, pp. 323–336, Dec. 2021, doi: 10.1016/j.inffus.2021.06.008.
18. Muhammad Ahmad *et al.*, "Fuzzy Based Hybrid Focus Value Estimation for Multi Focus Image Fusion," *Comput. Mater. Contin.*, vol. 71, no. 1, pp. 735–752, 2022, doi: 10.32604/cmc.2022.019691.
19. Khan, I. U., Khan, Z. A., Ahmad, M., Khan, A. H., Muahmmad, F., Imran, A., ... & Hamid, M. K. (2023, May). Machine Learning Techniques for Permission-based Malware Detection in Android Applications. In 2023 9th International Conference on Information Technology Trends (ITT) (pp. 7-13). IEEE.
20. Shah, A. M., Aljubayri, M., Khan, M. F., Alqahtani, J., Sulaiman, A., & Shaikh, A. (2023). ILSM: Incorporated Lightweight Security Model for Improving QOS in WSN. Computer Systems Science & Engineering, 46(2).