Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **4**, December **2025**

Enhanced Fusion Approach for Multi-Focus Image with Object Semantic Detection and Attentive Transformer Network

Mohit A. Bhavsar[1,*]

[1]Research Scholar, Faculty of Engineering and Technology,
Ganpat University, Gujarat, India
mab01ganpatuniversity@gmail.com

Dr. Amrutbhai N. Patel[2]

[2]Associate Professor, Faculty of Engineering and Technology,
Ganpat University, Ganpat Vidyanagar, 384012, Gujarat, India
amrut.patelganpatuniversity@yahoo.com

Dr. Anand J. Patel[3]

[3]LDRP Institute of Technology and Research,
KSV University, Gandhinagar, 382025, Gujarat, India
patelanandpec@yahoo.com

[*]Corresponding author: Mohit A. Bhavsar

Received May 27, 2025, revised September 29, 2025, accepted October 3, 2025.

Abstract. *By combining images from different focus distances, multi-focus image fusion is essential for improving computer vision; yet, it faces difficulties such as blur and inconsistent results. Hence a novel "Enhanced Fusion Approach for Multi-Focus Image with Object Semantic Detection and Attentive Transformer Network" is proposed. In the semantic-based object detection, existing methods struggle to differentiate objects with identical colors but varying focus depths due to ignoring semantic interactions. So, a novel Versatile Object Class Cum Edge Detection Net leveraging a Single Shot Multi Adaptive Anchor Box Detector (SSAD) for precise object detection and Dual Out Branch U-Net for edge detection has been introduced, which enhances semantic interaction and improves edge detection, mitigating difficulties in distinguishing objects with similar colors. Then, the Aggregated Fuzzy Decision Map Generator combines focus measures with object semantics to create decision maps, ensuring overall focus quality. Moreover, boundary detection in multi-focus images is crucial for object segmentation and recognition, but edge detection and fusion algorithms struggle to accurately identify boundaries, causing halo artifacts and texture discontinuities. Thus, a novel Visio Quaternion Attentive Stitching, combining Vision Transformer, Quaternion Fourier Transform and FusionGAN has been introduced to mitigate depth and scale disparities and remove halo artifacts during fusion. The proposed approach achieves excellent results in improvement of object and boundary detection, with 0.94 of object detection accuracy, 0.92 of edge detection precision when compared to other state-of-the-art models.*

Keywords: Multi-focus image fusion; object detection; semantic interaction; edge detection; boundary delineation; halo artifacts; texture discontinuities.

1. Introduction.

Multi-focus image fusion is a vital and challenging image processing task aimed at enhancing the visual quality and information content of images captured at varying focal planes. In scenarios where it is difficult to obtain a single, fully-focused image, such as macro photography, microscopy, or surveillance, multi-focus image fusion plays a pivotal role in synthesizing a single image that retains the sharpest details and relevant information from multiple source images. By intelligently combining distinct focus regions, this fusion process enables improved object detection, scene interpretation, and visual analysis, making it indispensable in a wide range of applications, including medical imaging, remote sensing, robotics, and computer vision. The fusion techniques employed in multi-focus image fusion are continually evolving to address real-world challenges and harness the full potential of multi-focus image data. Due to the constraints of image-capturing systems, only objects inside the depth of field (DOF) are typically focused or clear, while things beyond the DOF are defocused or blurred. Using the MFIF approach, the full-focus picture may be produced from a set of partly-focused photos. This technique is effective for obtaining the fused image with an expanded depth of field. Additional uses of the combined pictures include particular computer vision tasks like object recognition and image segmentation [1-4].

In multi-focus image fusion, the decision map serves as a critical component that guides the fusion process, facilitating the selection of pixel values from multiple source images. This map acts as a spatial filter, determining which regions of the resulting fused image should be dominated by data from one source image or another, based on their respective focus qualities. The decision map is typically generated through various algorithms that assess the sharpness or clarity of pixels in each source image and assign weights accordingly. By incorporating contextual information about the varying focus levels across the scene, the decision map ensures that the fused image seamlessly combines the most focused and informative details from each input image, ultimately resulting in an image with improved clarity and a more comprehensive representation of the scene. The generation of an accurate and well-structured decision map is fundamental to the success of multi-focus image fusion techniques in various applications, including microscopy, photography, and computer vision [5-8].

Many image fusion technologies have been presented over the past few decades by academics from across the world. These technologies may be divided into two groups: deep learning-based fusion technologies and classical algorithms based on fusion technologies. Both the spatial domain fusion technologies and the transform domain fusion technologies may be used to categorise the conventional algorithm based on fusion technologies. The purpose of spatial domain fusion methods, which mostly rely on pixel-level data or picture blocks, is to generate a weight map, which is then used to weigh the source image to produce a fused image. In some scenarios, traditional picture fusion algorithms can handle the demands of future image processing tasks. However, there has been a bottleneck in the development of these algorithms. On the one hand, in order to improve fusion performance, the fusion algorithm uses more complex transformations or representations, which makes it challenging to fulfil real-time requirements in actual applications. Deep learning has made progress in picture fusion, denoising, and ship recognition in recent years thanks to the gradual deepening theory and the potent feature learning capability [9-12].

Fusion algorithms for multi-focus images employ a range of techniques to combine information from images captured at different focal depths. These techniques include pixel-based methods, such as weighted averaging and minimum/maximum selection, as well as more advanced approaches like gradient-based fusion, frequency-domain fusion, and deep learning-based fusion. Each method has its advantages and challenges. For instance, pixel-based methods are simple but may struggle to handle complex scenes. Gradient-based techniques capture edge information effectively but can introduce artifacts in smooth regions. Frequency-domain fusion methods leverage the power of Fourier or wavelet transforms but may be computationally intensive. Deep learning-based fusion, while promising, requires large datasets and can be challenging to train effectively. Overall, the challenge in multi-focus image fusion lies in achieving a balance between preserving fine details, handling complex scenes, and ensuring computational efficiency, depending on the chosen fusion technique and application context [13-15]. As the computer vision, medical imaging, and remote sensing domains increasingly rely on high-resolution imaging, the ability to seamlessly integrate multiple focal planes into a single, sharp image becomes essential. Current fusion techniques often suffer from artifacts, loss of information, and inadequate handling of complex scenes. Therefore, the development of more advanced fusion algorithms is necessary to enhance image quality, improve object recognition, and aid in critical decision-making processes.

This work introduced an enhanced fusion approach for multi-focus image with object semantic detection that lies in its integration of semantic object understanding with edge-aware and boundary-preserving fusion techniques, which together improve both object and boundary detection and ensure high-quality fused images. Unlike existing methods, this approach simultaneously addresses object differentiation, edge preservation, and seamless fusion using a combination of adaptive anchor detection, dual-branch U-Net semantic labeling, and quaternion-based attentive stitching. The main contribution of this content is as follows:

- In the case of eliminating the difficulties in distinguishing different objects with the same color, Versatile Object Class Cum Edge Detection Net is introduced, where Single Shot Multi Adaptive Anchor Box Detector adapts anchor aspect ratios and sizes to capture even small objects and a Dual Out Branch U-Net enables precise edge and boundary recognition, thus ensuring accurate object class identification, semantic object interaction, and edge detection.
- Aggregated Fuzzy Decision Map Generation generated a decision map that integrates both focus measures and semantic information, ensuring optimal focus quality for every pixel in the multi-focus image, thus overcomes the limitations of conventional decision maps that neglect semantic context.
- Boundary detection challenges in multi-focus images, stemming from blurry foregrounds overlapping clear backgrounds, are addressed by Visio Quaternion Attentive Stitching, in that ViT and QFT enables capturing both spatial dependencies and frequency domain information and FusionGAN effectively eliminates halo artifacts, reduces depth and scale disparities, and produces high-quality fused images with smooth transitions between focus levels.

The synergy of semantic detection, fuzzy decision-making, and advanced attention-based fusion demonstrates a significant advancement in both performance and image quality. The paper's content is organized as follows: Section 2 covers the literature review; Section 3 explains the methodology and operation of the proposed approach; and Section 4 covers the evaluation, performance analysis, and comparison components of the proposed framework. Section 5 serves as the paper's conclusion.

2. Literature Survey.

Pan Wu et al [16] put forth the multi-focus image fusion technique TSFA, which adds shallow feature attention modules to the spatial information to produce an accurate decision map and employed Swin Transformer blocks to compute the self-attention map. This technique fused multi-focus images end-to-end without requiring laborious post-processing through training. In order to optimise the network's ability to recognise clear pixels while keeping the original image's intricate texture and preventing information deletion, a joint loss function was implemented. Comparing the nine SOTA approaches quantitatively and qualitatively, TSFA properly recognised the focusing target, exhibited no chromatic aberration or decision map mistake, and displayed the most reliable and sophisticated performance. However, the recognition ability for small objects needs to be improved.

Shuaiqi Liu et al [17] created a novel adaptive feature concatenate attention network called AFCANet, to create aesthetically attractive, completely focused images. AFCANet adaptively learnt cross-layer features while retaining the texture characteristics and semantic content of images. The encoder-decoder network served as the backbone network of AFCANet. Also included a powerful channel attention module to fully understand the encoder output and hasten network convergence in the midst of the encoder-decoder network. The texture information of the image was considered and a more accurate decision map was generated by applying the pixel-based spatial frequency fusion rules to fuse the adaptive features learnt by the encoder. However, maintaining depth information in the fused image has been a challenging one in this approach.

Shuaiqi Liu et al [18] suggested a multifocus colour image fusion approach based on low-vision image reconstruction and focus feature extraction, which combined the structural gradient. To do the low vision image reconstruction using the super-resolution approach, the source images were first fed into the deep residual network (ResNet). Then, using a rolling guiding filter, an end-to-end restoration model was employed to enhance the image features and preserve the image's edges. Additionally, the source image and the reconstructed image were used to create a different image. The focus region detection approach based on structural gradient was then used to build the fusion decision map. In order to create a fusion image, weighted fusion was utilised to combine the source image with the fusion decision map. However, determining the correct weights was challenging, especially when the characteristics of the source image were not well-understood.

Xuejiao Wang et al [19] suggested to use a Transformer-based feedback mechanism for multi-focus image fusion. By using the powers of the transformer and convolutional neural network to their greatest potential, the characteristics of focussed and defocused areas were learned. The features recovered by the fusion feedback blocks were used in other blocks of the same sub-network as well as between two sub-networks, and self-feedback blocks in the network increase the processing efficiency of feedback information from the same sub-network and conduct feature reuse. Grayscale and Lytro datasets were used to compare the efficacy of the approach against seven sophisticated multi-focus fusions. The findings demonstrated that the technique outperforms other cutting-edge techniques in terms of both visual quality and impartial assessment. However, this approach struggles to capture long-range dependencies and contextual information that is crucial for understanding complex multi-focus scenes.

Rama Mohan et al [20] created an effective image fusion technique in the multiresolution (MR) domain, by combining the DTCWT (Dual Tree Complex Wavelet Transform) algorithm with the qshiftN and MPCA algorithms. Using the MR method, multifocus input images were divided into high and low-frequency components. The DTCWT with qshiftN technique was used to fuse the input images' decomposed frequency components. The multi-focus source images sift invariance and directional characteristics were both preserved by the suggested fusion procedure. Finally, the MPCA algorithm was employed to improve the characteristics of the fused image. Different multifocus images were used to evaluate the suggested approach, and the evaluated metrics were compared to previously published technologically sophisticated algorithms. However, this approach does not perform well in scenarios where objects or regions of interest appear in different positions or orientations across input images.

Ming Lv et al [21] proposed a unique distance-weighted regional energy and structural tensor-based multi-focus image fusion technique. Low-frequency sub-bands were fused using the structural tensor-based fusion rule, while high-frequency sub-bands were fused using the distance-weighted regional energy-based fusion rule. The suggested technique was tested using 20 pairs of photos from the Lytro dataset, and the fusion outcomes showed that it produces state-of-the-art fusion performance in terms of image information, definition, and brightness, enabling the seamless fusion of multi-focus images. However, this method was sensitive to focus measure errors which leads to artifacts and reduced fusion quality.

Zhai et al [22] proposed a unique approach to multi-focus image fusion that makes use of asymmetric soft sharing and interactive transformers. Initially, a locally enhanced interactive method was developed to make use of the transformer's benefits in global context modeling while improving its limitations in terms of diversity and efficiency. More precisely, it mitigated the inadequate local feature perception and redundant computational cost of the current technique while simultaneously overcoming some of the transformer's shortcomings in domain-specific tasks by employing cross-scale and cross-domain computation strategies. Second, a multi-task learning technique with asymmetric soft sharing was adopted by the proposed approach to address the issues of fusion image distortion and artifacts. However, it is nearly hard to take images with every item focused because of the shallow depth-of-field (DoF) of optical lenses and image sensors.

Kiran et al [23] proposed a unique multi-focus image fusion technique in the frequency partition (FP) domain that was based on the quarter shift dual-tree complex wavelet transform (qshiftN DTCWT) and laplacian pyramid (LP). The initial images were then broken down into row and column coefficients using an FP domain frequency factor. Second, row and column images were combined using qshiftN DTCWT, which turned out to be a particularly effective multiresolution transform for image fusion because of its directional and shift invariant properties. Finally, an all-in-focus image was created using the LP algorithm to improve the performance of the qshiftN DTCWT in the FP-based technique. However, conventional wavelet-based fusion techniques produce ringing distortions in the fused image because of their weak shift invariance and directionality.

Zheng et al [24] Using Cross, Split, and Shuffle stages to improve gradient propagation and combination, Stack-YOLO is a one-stage object identification technique that uses less memory and computing while increasing accuracy and convergence speed. With the introduction of ASPPF pooling, serial procedures took the place of parallel pooling to preserve detailed information and learn feature weights adaptively. Global matching costs were used by Fast-OTA to improve computational efficiency and a composite factor model was used to scale supply and demand for resources. When combined, these advancements maximize object detection tasks' resource needs, speed, memory consumption, and accuracy of detection. However, optimizing real-time object identification on edge devices, central processing units, GPU clusters, and cloud GPUs remained an extremely demanding task because of variables including processing speed, storage capacity, and image quality.

Liu et al [25] proposed a unique Transformer network for crowd localization called Cross-scale Vision Transformer (CsViT), which builds long-range context dependencies on the combined feature maps while concurrently fusing multi-scale information throughout both the encoder and decoder phases. To achieve this, this method designed a multi-scale encoder that fuses various scales' feature maps at corresponding positions to produce combined feature maps. In the meantime, a multi-scale decoder was designed to integrate the tokens at various scales when modeling long-range context dependencies. Moreover, Multi-scale SSIM (MsSSIM) loss was developed in this work to improve the similarity at many scales and calculate head regions adaptively. However, it has issues with perspective, occlusion, and illumination.

Overall, from this literature study, it is found that in the existing approaches, the ability of detection of small objects needs to be improved [16], maintaining depth information in the fused image is a challenging one [17], it is challenging determine the correct weight when the source image is not well understood [18], capturing long-range dependencies and contextual information of the muti-focus image is still a problem [19], does not perform well in scenarios where objects or regions of interest appear in different positions or orientations across input images [20], are sensitive to focus measure errors which can lead to artifacts and reduced fusion quality [21], It's difficult to take images when every object is sharply focused [22], results in ringing distortions in the fused image because of their directionality and poor shift-invariance [23], It was difficult to optimize real-time object recognition because of things like processing speed [24], and possess problems with lighting, occlusion, and perspective [25]. Hence, there is a need for an improvement in the fusion algorithm to obtain better-fused images.

3. **Enhanced Fusion Approach for Multi-Focus Image with Object Semantic Detection and Attentive Transformer Network.** Multi-focus image fusion is vital in computer vision, combining images from various focus distances to enhance depth of field and overall quality. Despite its importance in tasks like object recognition, challenges include blur and inconsistencies. Efficient fusion is essential for clear, informative imagery across diverse applications. Hence a novel "Enhanced Fusion Approach for Multi-Focus Image with Object Semantic Detection and Attentive Transformer Network" is proposed in this approach to overcome the limitations of the existing multi-focus image fusion methods and improve object and boundary detection and the quality of fused images. In Semantic-based object detection, Current methods for multi-focus image fusion often fail to effectively distinguish between objects with the same color but at different depths of focus. This is because the decision map is typically created based on focus measures or sharpness metrics, neglecting the semantic interaction of objects within the scene. As a result, contextual aspects of objects are not considered, leading to poor object detection performance. So, a novel, Versatile Object Class Cum Edge Detection Net is presented. in which a new object class detection approach namely Single Shot Multi Adaptive Anchor Box Detector (SSAD) is introduced in which dynamically adjusts anchor box aspect ratios and scales to detect small and varied-sized objects in multi-focus images. This ensures that objects, regardless of size or focus depth, are properly identified and incorporated into the fusion process something not addressed in prior works. After the object classes are identified, the image is fed to a Dual Out Branch U-Net for semantic labeling and edge detection, which semantic segmentation and edge detection output branches utilize convolutional layers with softmax activation functions and ReLU activation functions for effective image classification and edge detection. This dual-branch setup enables the model to capture object meaning and structural boundaries in a unified framework capability not present in previous multi-focus image fusion methods. This combined approach uniquely captures semantic interactions while improving edge detection, overcoming challenges of differentiating objects with similar colors.

Then, the Aggregated Fuzzy Decision Map Generator generates a decision map using regular focus measures and object semantic interaction, ensuring satisfaction and overall focus quality for each pixel or region in a multi-focus image. Moreover, boundary detection in multi-focus images is crucial for object segmentation and recognition, enabling precise delineation of objects and regions, aiding in subsequent analysis and processing tasks. Edge detection and fusion algorithms struggle to accurately identify boundaries in multi-focus images, causing halo artifacts and texture discontinuities in the fused image. This is due to difficulties in determining which parts of the foreground and background should be included in the final fused image, resulting in texture discrepancies and a lack of seamless transitions. Therefore, a novel Visio Quaternion Attentive Stitching is introduced in which the Vision Transformer encoder captures spatial relationships between input images and the decision map, whereas traditional fusion methods rely on local pixel or region-based focus metrics, while this method captures global semantic and spatial relationships, reducing depth and scale inconsistencies Quaternion Fourier Transform separates amplitude and phase information for multi-focus image fusion. This separation enhances the preservation of texture and phase details, eliminating fusion artifacts like halo effects and texture discontinuities. FusionGAN captures spatial relationships, amplitude, and phase information, producing a fused image. This Visio Quaternion Attentive Stitching method ensures a smooth transition by reducing depth and scale discrepancies and eliminating halo artifacts.

An image processing method that fuses near- and far-focused images to produce a fused image with improved clarity and detail is shown in Figure 1. Two near-focused images are first processed using SSAD and U-Net for Versatile Object Class Cum Edge Detection. It utilizes an Aggregated Fuzzy Decision Map Generator to optimize focus quality. The Visio Quaternion Attentive Stitching fusion

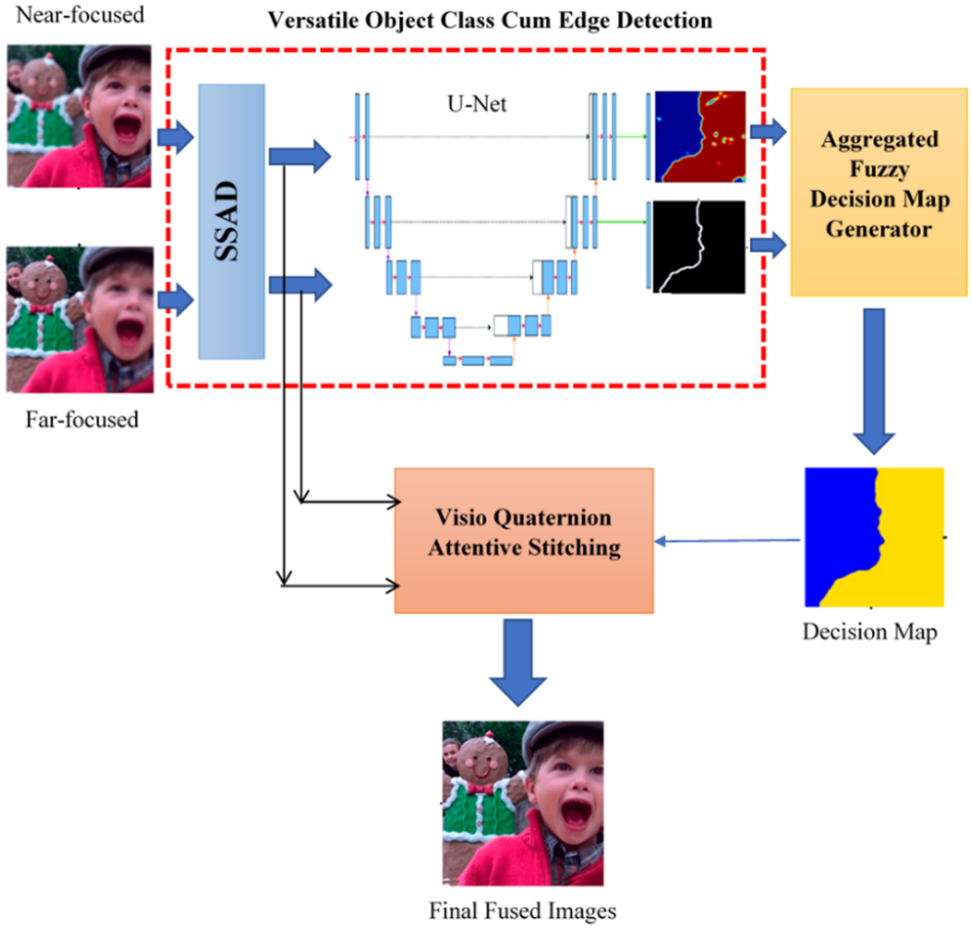
Near-focused
Versatile Object Class Cum Edge Detection
SSAD
U-Net
Aggregated Fuzzy Decision Map Generator
Far-focused
Visio Quaternion Attentive Stitching
Decision Map
Final Fused Images

FIGURE 1. The architecture of an image processing method that fuses near- and far-focused images

approach minimizes depth disparities and halo artifacts, resulting in higher-quality fused images. More details on these techniques will be explained in further sections.

3.1. **Versatile Object Class Cum Edge Detection Net.** The Versatile Object Class Cum Edge Detection Net is an innovative technique introduced in the proposed methodology to address the limitations of existing multi-focus image fusion methods. Unlike existing multi-focus image fusion methods that rely solely on focus measures, the proposed Versatile Object Class Cum Edge Detection Net combines semantic-based object detection and edge detection. Here the introduction of SSAD dynamically adapts anchor ratios and sizes to detect objects of varying shapes and scales, including small objects, enhancing object recognition in multi-focus images. And the Dual Out Branch U-Net allows simultaneous semantic segmentation and edge detection. This ensures accurate boundary delineation, resolving challenges of distinguishing objects with similar colors and reducing halo artifacts and texture discontinuities in the fused image.

In this technique, two main components are object class detection and edge detection. The purpose of object class detection, facilitated by the Single Shot Multi Adaptive Anchor Box Detector (SSAD) is a novel approach introduced to enhance object detection in multi-focus image datasets. It dynamically adjusts anchor aspect ratios and sizes to accurately capture the diverse shapes and sizes of objects present in the scene. By adapting the anchor boxes to match the characteristics of each object, including small ones, SSAD ensures comprehensive coverage and precise identification of object classes. This adaptive nature enables SSAD to effectively capture subtle details and distinctions in object morphology, contributing to more robust and accurate object detection in multi-focus images across a wide range of scales and sizes. The mathematical representation of SSAD is described below in equation (1):

$$L(x, c, l, g) = \frac{1}{N} \left(L_{conf}(x, c) + \alpha L_{loc}(x, l, g) \right) \tag{1}$$

The equation provided is the overall loss function $L(x, c, l, g)$ for the Single Shot Multi Adaptive Anchor Box Detector (SSAD). Where,

- $L_{conf}(x, c)$ represents the confidence loss, which measures the accuracy of the predicted objectness scores (confidence scores) for each anchor box.
- $L_{loc}(x, l, g)$ term denotes the localization loss, which quantifies the disparity between the predicted bounding box coordinates (center coordinates l and dimensions g) and the ground truth bounding box coordinates.
- N represents the total number of anchor boxes.
- α is a hyperparameter that controls the relative importance of the confidence loss and localization loss in the overall loss function.

The goal of training the SSAD model is to minimize this loss function $L(x, c, l, g)$, which encourages accurate predictions of objectness scores and precise localization of bounding boxes. By optimizing this loss function through techniques like stochastic gradient descent, the SSAD model learns to effectively detect object classes in multi-focus images.

The image is transmitted to a Dual Out Branch U-Net for edge detection and semantic labeling once the object classes have been identified. In this network architecture for multi-focus image fusion, two output branches are utilized for semantic segmentation and edge detection, respectively. The semantic segmentation branch concludes with a convolutional layer employing a softmax activation function. This activation function computes class probabilities for each pixel, facilitating the identification of different object classes within the image. Conversely, the edge detection branch concludes with a convolutional layer using the Rectified Linear Unit (ReLU) activation function. ReLU is chosen specifically to enhance edge responses, enabling the network to effectively detect and highlight the boundaries between objects, contributing to accurate edge detection in the fused image. Equation (2) describes the mathematical form of Dual Out Branch U-Net below:

$$D_a = D_a + (Y_a \cdot D_a + Y_b \cdot D_b) \tag{2}$$

- D_a represents the feature maps produced by the first branch of the network, which is responsible for semantic segmentation.
- D_b represents the feature maps generated by the second branch, which is focused on edge detection.
- Y_a and Y_b represent the predicted output tensors obtained from the semantic segmentation and edge detection branches, respectively.

This equation indicates that D_a is updated by adding the element-wise product of Y_a and D_a, representing the relevance of the semantic segmentation predictions to the feature maps of the first branch. Similarly, the product of Y_b and D_b is added to D_a, incorporating the relevance of the edge detection predictions from the second branch. This update mechanism allows the Dual Out Branch U-Net to refine its feature representations based on both semantic segmentation and edge detection outputs, ultimately improving its performance in generating high-quality fused images.

The Versatile Object Class Cum Edge Detection Net enhances object detection by incorporating semantic interaction, enabling better differentiation between objects of similar colors. Additionally, it improves edge detection, addressing challenges in accurately delineating object boundaries within multi-focus images. By deriving equation (1) and (2) we get Versatile Object Class Cum Edge Detection Net equation which described in equation (3):

$$O = \frac{1}{N}(L_{conf}(x, c) + \alpha L_{loc}(x, l, g) + D_a + (Y_a \cdot D_a + Y_b \cdot D_b)) \tag{3}$$

This equation represents the combined output of the network, incorporating both the confidence loss (L_{conf}) and the localization loss (L_{loc}), along with the adjustments made by the D_a and D_b terms.

The method of the Versatile Object Class Cum Edge Recognition Net, which enhances object recognition and boundary delineation in multi-focus images by combining object class detection, edge detection, and feature refinement, is illustrated in Figure 2. With the addition of edge detection and semantic information, this output improves item discrimination and boundary delineation in images.

The architecture of the Versatile Object Class Cum Edge Detection Net is shown in Figure 3. A Single Shot Multi Adaptive Anchor Box Detector (SSAD) is introduced by Versatile Object Class Cum Edge recognition Net for object class recognition in multi-focus image datasets. This technique captures object classes for even small objects by modifying anchor aspect ratios and sizes to fit object shapes and sizes. For semantic labeling and edge detection, the network employs a Dual Out Branch U-Net, which enhances edge recognition and reduces the challenge of differentiating objects with the same color. The

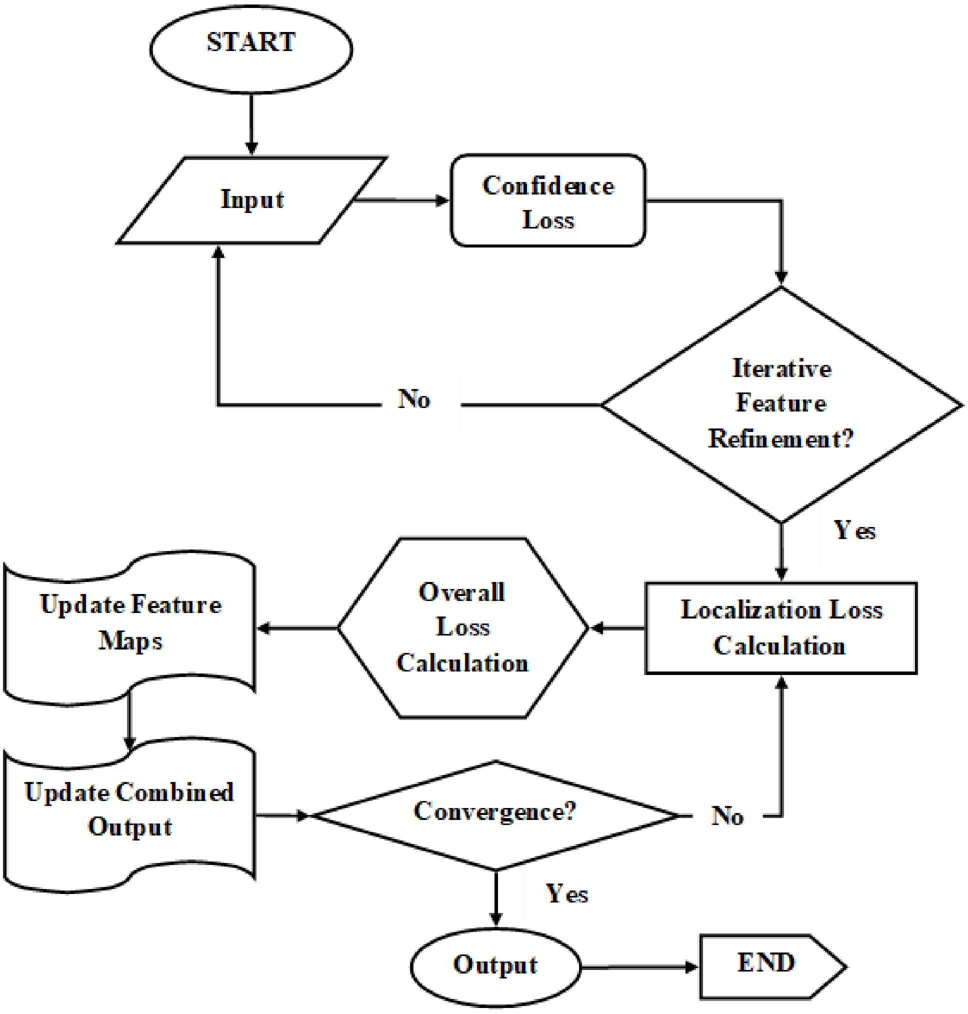

Figure 2. Process flow of Versatile Object Class Cum Edge Detection Net

Aggregated Fuzzy Decision Map Generator approach, which is covered in more detail in a later section, is then used to construct the decision map.

3.2. **Aggregated Fuzzy Decision Map Generator.** The Aggregated Fuzzy Decision Map Generator is an inventive technique introduced in this proposed methodology to enhance multi-focus image fusion by effectively integrating information from various focus levels and incorporating object semantic interaction into the decision-making process. This technique fuses several key components to ensure robust and accurate determination of focused regions. Firstly, it incorporates regular focus measures, traditional metrics assessing pixel focus quality based on sharpness or contrast. Secondly, it considers object semantic interaction, analyzing contextual relationships between objects to distinguish between similar-colored objects and improve object detection. Finally, fuzzy membership functions are utilized to handle decision-making uncertainty, assigning degrees of membership to pixels or regions based on focus quality and semantic relevance. This comprehensive approach enhances decision-making accuracy and facilitates precise delineation of focused regions, ultimately improving the overall quality of the fused image. The Aggregated Fuzzy Decision Map Generator combines regular focus measures $L_{conf}(x,c)$ with object semantic interaction $L_{obj}(x)$ and fuzzy aggregation $L_{fuzzy}(x)$ to create a decision map $D(x)$ for multi-focus image fusion which shown in equation (4):

$$D(x) = Aggregation\left(L_{conf}(x,c), L_{obj}(x), L_{fuzzy}(x)\right) \tag{4}$$

This equation represents the aggregation of different aspects of focus quality and semantic interaction to produce a comprehensive decision map for determining focused regions in the multi-focus image.

To identify focused regions in the multi-focus image, a decision map is created by combining various characteristics of focus quality and semantic interaction, as shown in Figure 4. To improve focus levels

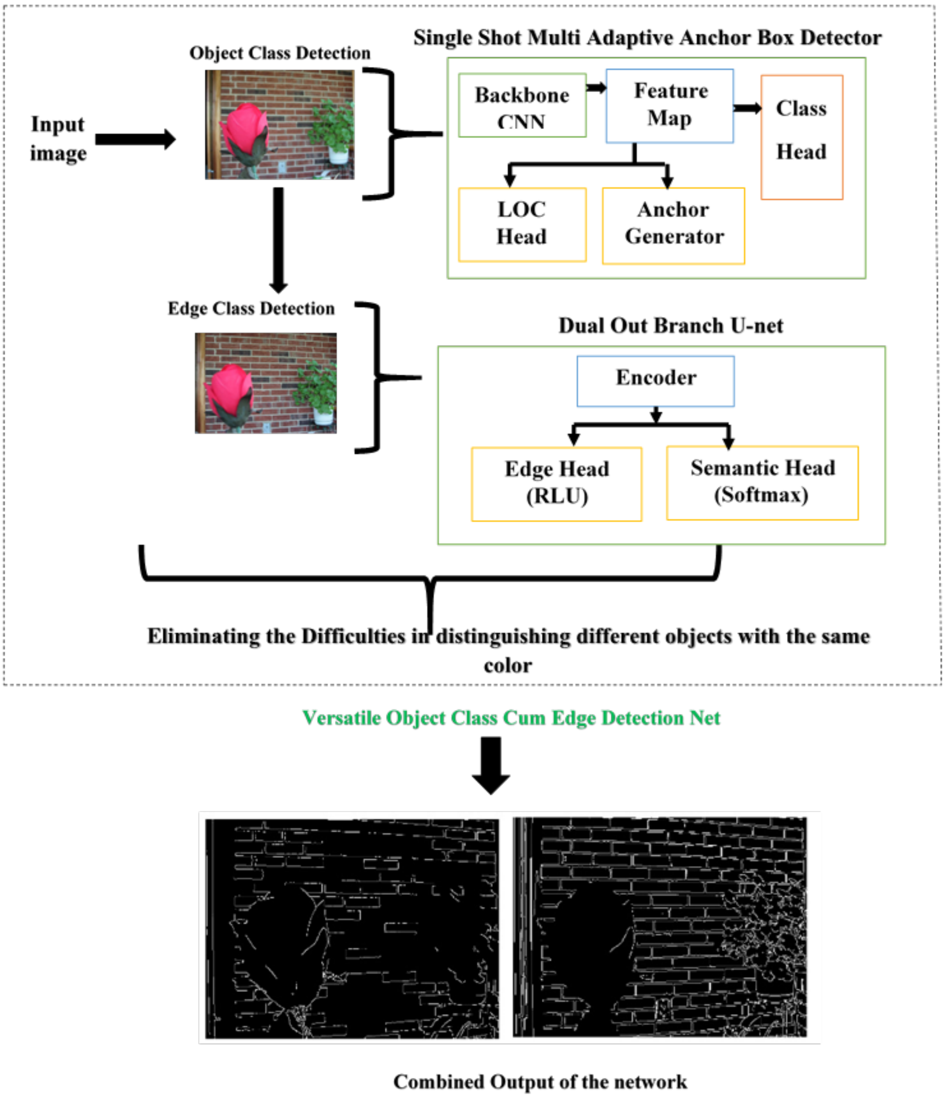
Input image
Object Class Detection
Single Shot Multi Adaptive Anchor Box Detector
Backbone CNN
Feature Map
Class Head
LOC Head
Anchor Generator
Edge Class Detection
Dual Out Branch U-net
Encoder
Edge Head (RLU)
Semantic Head (Softmax)
Eliminating the Difficulties in distinguishing different objects with the same color
Versatile Object Class Cum Edge Detection Net
Combined Output of the network

FIGURE 3. Architecture of Versatile Object Class Cum Edge Detection Net

and enable precise focus area determination, this output creates a decision map that allocates focus quality and satisfaction to each pixel in a multi-focus image. Ultimately, this improves the fused image quality. Finally, a technique Visio Quaternion Attentive Stitching is introduced in this approach to reduce disparities and to remove halo artifacts which will be explained in detail in the next section.

3.3. **Visio Quaternion Attentive Stitching.** The Visio Quaternion Attentive Stitching technique is a creative technique introduced in this proposed methodology to address the challenges associated with multi-focus image fusion, particularly in reducing halo artifacts and texture discontinuities while ensuring a seamless transition between different focus levels. Its purpose is to effectively combine information from the original input images and the decision map generated by the Aggregated Fuzzy Decision Map Generator, utilizing advanced techniques to mitigate depth and scale disparities inherent in multi-focus imagery.

Figure 5 depicts Visio Quaternion Attentive Stitching's architectural design. In this approach, while a technique Vision Transformer employs an encoder architecture equipped with a self-attention mechanism, enabling it to capture intricate spatial dependencies among patches within the input data. By analyzing these relationships, Vision Transformer generates feature representations that encapsulate the contextual information of each patch. This allows Vision Transformer to effectively compare the decision map, which highlights focused regions, with the near-focused and far-focused original images. Through this comparison, Vision Transformer mitigates depth and scale disparities inherent in multi-focus imagery,

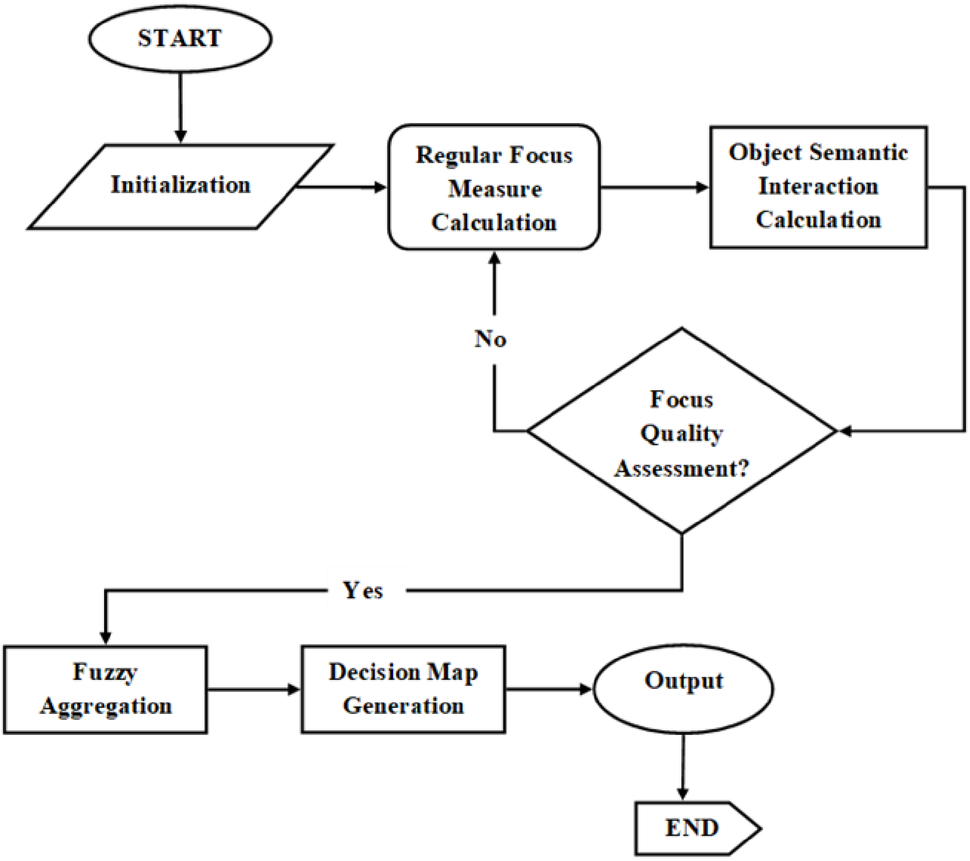
START
Initialization
Regular Focus Measure Calculation
Object Semantic Interaction Calculation
Focus Quality Assessment?
No
Yes
Fuzzy Aggregation
Decision Map Generation
Output
END

Figure 4. Process flow of Aggregated Fuzzy Decision Map Generator

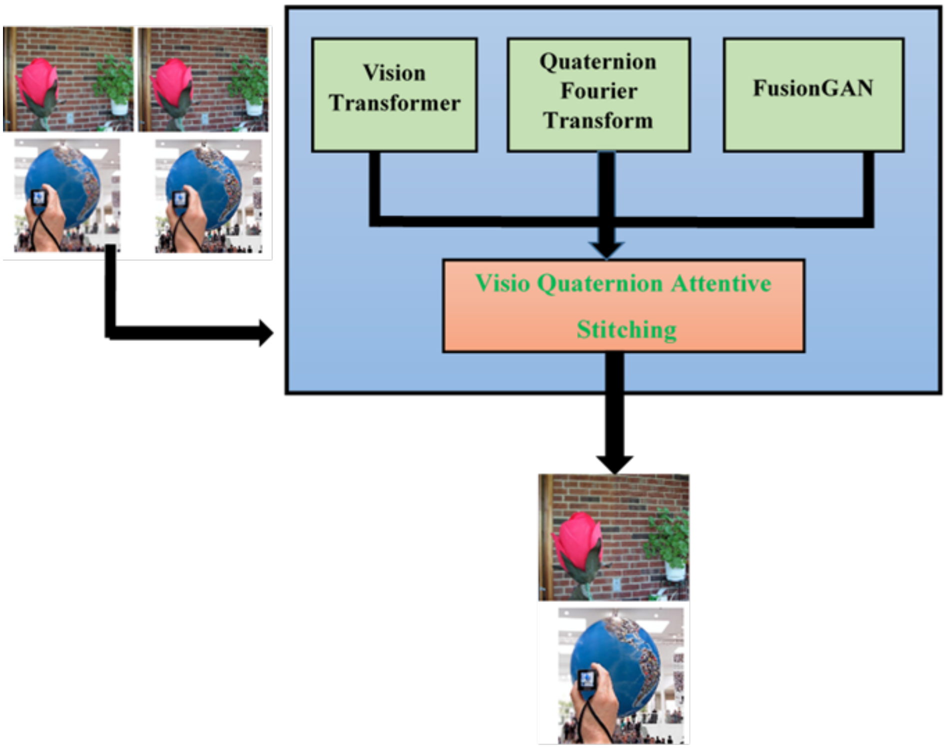
Vision Transformer
Quaternion Fourier Transform
FusionGAN
Visio Quaternion Attentive Stitching

Figure 5. Architecture of Visio Quaternion Attentive Stitching

facilitating the seamless integration of information across different focus levels and reducing texture discontinuities in the final fused image. The general form of vision transformer is expressed in the equation (5):

$$ViT = \sum_{j=1}^{N} \frac{Sim(Q_i, K_j)}{\sum_{j=1}^{N} Sim(Q_i, K_j)} V_j \tag{5}$$

- ViT represents the output feature vector obtained by the transformer.
- N signifies the total number of key-value pairs, where each pair (Q_i, K_j) corresponds to a different patch in the input image.
- Q_i represents the query vector associated with the $i-th$ patch in the input image.
- K_j denotes the key vector associated with the $j-th$ patch in the input image.
- V_j represents the value vector associated with the $j-th$ patch in the input image.
- $Sim(Q_i, K_j)$ signifies the similarity score between the query vector Q_i and the key vector K_j.

The equation calculates a weighted sum of the value vectors V_j based on the similarity scores, where the weights are determined by the softmax of the similarity scores. This weighted sum is used to generate the output feature vector ViT. In essence, the equation captures the essence of the self-attention mechanism in the Vision Transformer, where each query vector attends to all key-value pairs, and the output is a weighted sum of the value vectors based on their similarity to the query vector.

Parallelly, the Quaternion Fourier Transform is employed on the original input image to decompose it into separate components of amplitude and phase information. This transformation is particularly advantageous for multi-focus image fusion tasks as it effectively separates the spatial frequency components of the image. By extracting both amplitude and phase information, Quaternion Fourier Transform enables a more comprehensive representation of the content of the image, facilitating the elimination of halo artifacts commonly encountered in fused images. This decomposition process helps enhance the clarity and quality of the final fused image by mitigating unwanted artifacts and preserving essential details across different focus levels. The Quaternion Fourier Transform is formulated as in equation (6):

$$|a\rangle = \sum_{j=0}^{N-1} a_j |j\rangle \xrightarrow{QFT} |b\rangle \geq \sum_{k=0}^{N-1} b_k |k\rangle \tag{6}$$

- $|a\rangle$ represents the original input image or signal, which is expressed as a linear combination of basis states $|j\rangle$. These basis states represent different spatial frequencies or components of the input signal.
- The coefficients a_j indicate the contribution of each basis state to the overall signal.
- The term $\xrightarrow{QFT}$ denotes the Quaternion Fourier Transform operator, which acts on the input signal to transform it into the frequency domain.
- After applying the Quaternion Fourier Transform operator to the input signal $|a\rangle$, the resulting transformed signal is represented as $|b\rangle$. Similar to the original signal, $|b\rangle$ is expressed as a linear combination of basis states $|k\rangle$, where b_k are the coefficients indicating the contribution of each basis state to the transformed signal.

Overall, this equation illustrates how the Quaternion Fourier Transform operates on the input signal to decompose it into different frequency components, preserving both amplitude and phase information in the transformed domain.

In the final stage, the FusionGAN integrates spatial relationships, amplitude, and phase information extracted from the input images and the decision map. This process enables the FusionGAN to synthesize a comprehensive representation of the scene, incorporating both the structural details captured by the spatial relationships and the frequency characteristics conveyed by the amplitude and phase information. By leveraging generative adversarial networks (GANs), the FusionGAN refines this representation to generate the fused image. GANs utilize a dual-network architecture consisting of a generator and a discriminator, which work collaboratively to produce high-quality and visually appealing outputs while maintaining consistency with the input data and decision map. Equation (7) represents the mathematical description of FusionGAN:

$$V_{FusionGAN}(G) = \frac{1}{N} \sum_{n=1}^{N} (D_{\theta_D}(I_f^n) - c)^2 \tag{7}$$

- G represents the generator network of the FusionGAN, responsible for generating the fused image.

- N is the total number of input images or input patches used in the fusion process.
- D_{θ_D} refers to the discriminator network of the FusionGAN, which evaluates the realism of the generated images.
- I_f^n represents the $n-th$ input image or input patch used in the fusion process.
- c denotes the target value, often set to 1 for real images and 0 for generated (fake) images.
- $\sum$ denotes summation, indicating that the loss function is computed as the sum of individual loss terms for each input image or patch.

The loss function is calculated as the mean squared difference between the discriminator's output when evaluating the generated image $G(I_f^n)$ and the target value c. This loss function guides the training process of the FusionGAN, encouraging the generator to produce fused images that are realistic and indistinguishable from real images, as assessed by the discriminator.

By integrating these techniques, the Visio Quaternion Attentive Stitching method aims to minimize texture discontinuities, minimize halo artifacts, and minimize disparities in depth and scale in the final fused image. In the end, this method improves the overall quality of the fused image by maintaining small details across the scene and guaranteeing a smooth transition between various focus levels. By combining equation (5), (6) & (7) the derived equation for Visio Quaternion Attentive Stitching method is expressed in the equation (8):

$$VQAS = V_{FusionGAN}\left(ViT(I_f, Q_i, K_j) \odot QFT(I_f, a_j, b_k)\right) \tag{8}$$

- I_f represents the input image or patch.
- Q_i and K_j denote the queries and keys extracted by ViT, respectively.
- a_j and b_k represent the amplitude and phase information obtained through QFT, respectively.
- $V_{FusionGAN}$ is the FusionGAN module applied to the features extracted by ViT and QFT.
- $\odot$ denotes element-wise multiplication, combining the features obtained from ViT and QFT.

The derived equation (8) for Visio Quaternion Attentive Stitching integrates features from ViT, spatial relationships from QFT, and FusionGAN for image fusion, addressing depth and scale disparities while reducing halo artifacts. Algorithm 3 describes the Visio Quaternion Attentive Stitching.

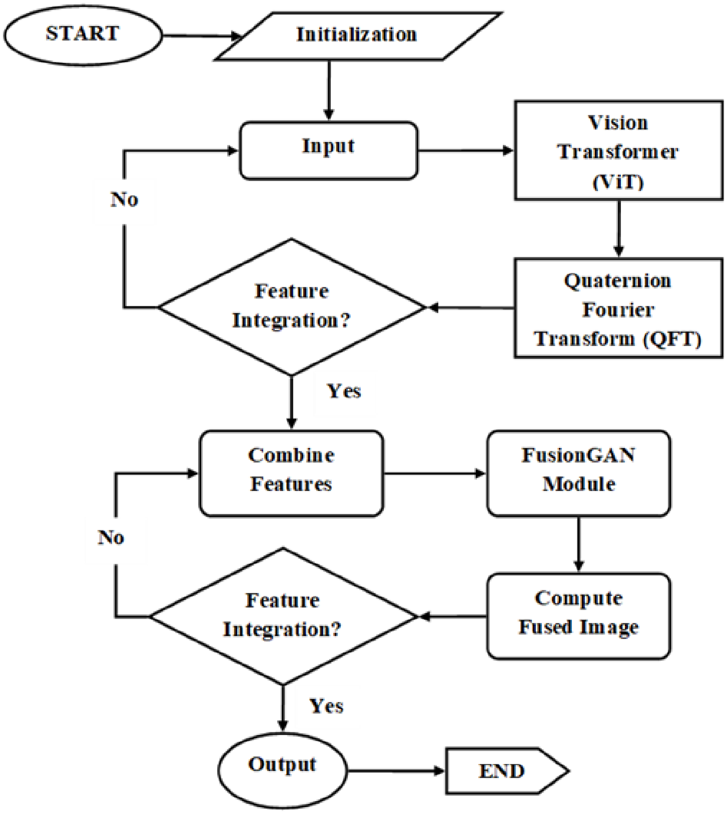

Figure 6. Process flow of Visio Quaternion Attentive Stitching

The Visio Quaternion Attentive Stitching algorithm's stages are shown in Figure 6. It begins with input multi-focus images, processes them using ViT and QFT, combines the extracted features, uses the FusionGAN module to create the fused image, and outputs the fused image. The output of this method is a high-quality fused image with no halo artifacts, a smooth transition between focus levels, and decreased depth and scale disparities. Pseudocode 1 for the overall proposed model Attentive Transformer Network is provided below.

Pseudocode 1: Object Semantic Detection and Attentive Transformer Network

```
Input: Multi-focus image dataset
Output: High-quality fused image
Initialize empty decision map $D(x)$, query $Q_i$, keys $K_j$,
// Versatile Object Class Cum Edge Detection Net (VOCED Net)
for every single image $I_f$ in the multi-focus dataset do:
    Compute confidence loss: $L_{conf}(x, c)$
    Compute localization loss: $L_{loc}(x, l, g)$
    Compute overall loss:
        $L(x, c, l, g) = \frac{1}{N}\left(L_{conf}(x, c) + \alpha L_{loc}(x, l, g)\right)$
    Extract feature maps using Dual Out Branch U-Net:
        $D_a = D_a + (Y_a \cdot D_a + Y_b \cdot D_b)$ // $D_a$: semantic features, $D_b$: edge features
    Combine outputs of network
        $O = \frac{1}{N}(L_{conf}(x, c) + \alpha L_{loc}(x, l, g) + D_a + (Y_a \cdot D_a + Y_b \cdot D_b))$
// Aggregated Fuzzy Decision Map Generator (AFDMG)
Initialize empty decision map $D(x)$
For each pixel do
    Compute regular focus measure $L_{conf}(x, c)$
    Compute object semantic interaction $L_{obj}(x)$
    Apply fuzzy aggregation $L_{fuzzy}(x)$
End for
Aggregate measures to generate decision map
$D(x) = Aggregration\left(L_{conf}(x, c), L_{obj}(x), L_{fuzzy}(x)\right)$
// Visio Quaternion Attentive Stitching (VQAS)
Apply ViT on $I_f$
    $(Q_i, K_j) \leftarrow ViT$ // Extract queries $Q_i$ and keys $K_j$
Apply QFT on $I_f$
    $(a_j, b_k) \leftarrow QFT$ // Extract amplitude $a_j$ and phase $b_k$
Compute combined feature representation
    F_combined $= ViT(I_f, Q_i, K_j) \odot FT(I_f, a_j, b_k)$
Feed F_combined and decision map into FusionGAN
    $VQAS = V_{FusionGAN}\left(ViT(I_f, Q_i, K_j) \odot QFT(I_f, a_j, b_k)\right)$
End for
Return fused image
```

Overall, the proposed methodology introduces a novel Versatile Object Class Cum Edge Detection Net, addressing limitations in existing multi-focus image fusion methods. By leveraging a Single Shot Multi Adaptive Anchor Box Detector (SSAD) for object detection and Dual Out Branch U-Net for edge detection. The Aggregated Fuzzy Decision Map Generator creates decision maps, ensuring overall focus quality. Additionally, a Visio Quaternion Attentive Stitching guarantees a smooth transition, halo artifacts are eliminated and depth and scale discrepancies are reduced. These innovations improve object and boundary detection and the quality of fused images, in multi-focus image fusion. The performance and comparison evaluation of this method is explained in further sections.

4. **Results and Discussion.**

This section focused on how multi-focus image fusion techniques enhance the quality of fused images as well as object and boundary detection. It examines the Versatile Object Class Cum Edge Detection Net, the Aggregated Fuzzy Decision Map Generator, and the Visio Quaternion Attentive Stitching for enhanced objects. This work attempts to overcome problems with object recognition and boundary detection in multi-focus image fusion approaches by leveraging the analytical power of MATLAB.

4.1. **Experimental Setup.**

- Software: MATLAB
- OS: Windows 10 (64-bit)
- Processor: Intel i5
- RAM: 8GB RAM

4.2. **Dataset Description.** The experiments utilize the Lytro Color Multi-Focus Dataset [29], a benchmark collection derived from the Lytro light-field camera gallery, specifically designed for multi-focus image fusion (MFIF) research in color images. The dataset consists of 20 pairs of high-resolution RGB images (each 520×520 pixels) capturing complex real-world scenes with diverse focus depths ranging from natural landscapes to indoor objects and challenging elements such as semi-transparent fences with fine edges and varying blur characteristics. For training 70% data used, Validation 15% and Test 15% respectively.

Distinctive features of this dataset include:

- Colour and focus diversity, enabling evaluation in RGB space rather than traditional grayscale.
- High visual complexity with intricate structures that test edge preservation and spatial detail retention.
- Real-world scenes that enhance the robustness evaluation of fusion algorithms.

Focus annotations are provided via manual or Laplacian-based focus maps to aid in training and evaluation.

4.2.1. *Experimental procedures.* To ensure each image had the same input size, all the images were initially resized to a specific resolution (e.g., 256×256). The pixel values were then normalized to the range of [0,1] and provided as input to the network. To enhance diversity in the dataset and limit overfitting the data was augmented. This consisted of using random cropping, flipping (horizontal), and rotation to augment the data. Ground-truth segmentation masks for the object boundaries in the images were created using a pre-trained edge detector that were further delineated by the author for sample validation. The Versatile Object Class Cum Edge detectors were then trained using the pre-trained models using stochastic gradient descent with momentum (SGD-M), as well as using the Adam optimizer, as a baseline of comparison. The SSAD module was used to adaptively generate anchor boxes, with a starting learning rate of 1e-4, which was reduced by 0.1 every 20 epochs, to a final of 100 epochs. The Dual-Out branch U-Net for semantic segmentation and edge detection was trained simultaneously and was modified by using a multi-tasks loss (Eq. (1) and Eq. (2)) that summarizes confidence, localization, and boundary refinement losses.

In the case of the Visio Quaternion Attentive Stitching, the Vision Transformer (ViT) was pre-trained on ImageNet patches before being fine-tuned on the multi-focus pairs. The Quaternion Fourier Transform (QFT) was conducted in the frequency domain via FFT-based quaternion decomposition. The Fusion-GAN was then trained in an adversarial way following a generator–discriminator framework where the generator synthesized the fused image and the discriminator assessed realism against the ground truth reference images. Training was stabilized using birtrophic Wasserstein loss with gradient penalty and spectral normalization.

4.3. **Simulation Results.** These findings highlight the benefits of using the Versatile Object Class Cum Edge Detection Net, the Aggregated Fuzzy Decision Map Generator, and the Visio Quaternion Attentive Stitching in multi-focus image fusion methods. These advances provide significant improvements in performance metrics, including enhanced fused images and better object and boundary detection.

Figure 7 illustrates two distinct images capturing different depths of the object to enable clear fusion. Initially, both images prioritize focusing on the foreground image, rendering the background image blurred. Subsequently, attention shifts to the background image for precise fusion, resulting in the foreground image appearing blurred. Ultimately, leveraging the proposed techniques, the objective is to generate a seamlessly fused image lacking of blurring, ensuring clear boundary detection.

Figure 8 illustrates the simulation outcomes of the edge detection technique. Each image showcases stark contrasts with bright outlines delineating objects against dark backgrounds. The sequence demonstrates edge detection applied to both foreground and background images. The clear detection of edges signifies the effectiveness of the proposed techniques in generating a fused image lacking blurring, ensuring clarity and precision.

Figure 9 presents the simulation results of the final fused image, showcasing the essence of multi-focus image fusion. This process involves amalgamating multiple images captured at different focal distances

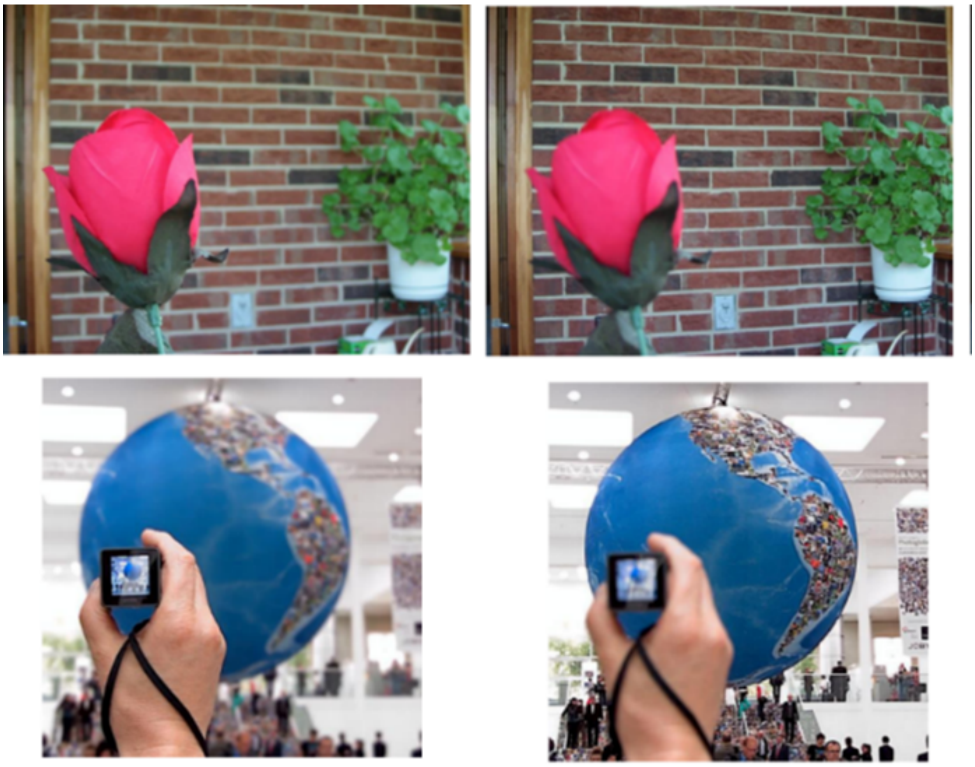

FIGURE 7. Simulation results of two different objects

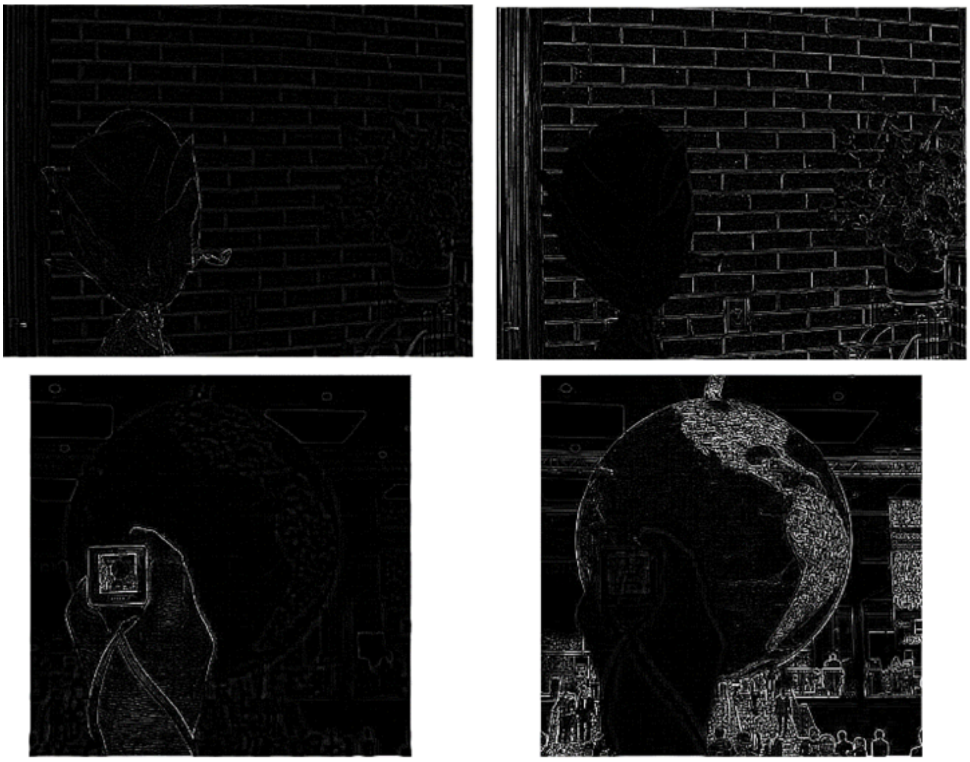

FIGURE 8. Edge detection technique

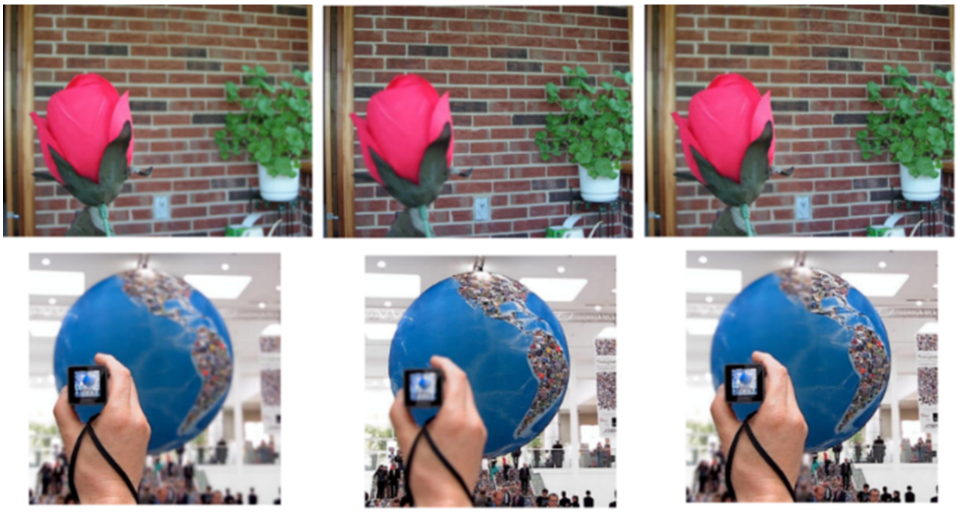

Figure 9. Simulation results of the final fused image

to produce a unified image with an extended depth of field. In the Top Row (Rose), the focal point transitions from nearby elements such as petals to distant features like walls and plants, while in the Bottom Row (Globe), the focus shifts from close-up objects like the camera screen to distant ones. By these proposed techniques, these image sets are fused to create a singular image where both nearby and distant objects are sharply in focus.

4.4. **Performance evaluation of the proposed model.** The performance examination of multi-focus image fusion approaches includes the Versatile Object Class Cum Edge Detection Net, Aggregated Fuzzy Decision Map Generator, and the Visio Quaternion Attentive Stitching. Furthermore, enhanced boundary and object identification are included for sophisticated multi-focus image fusion.

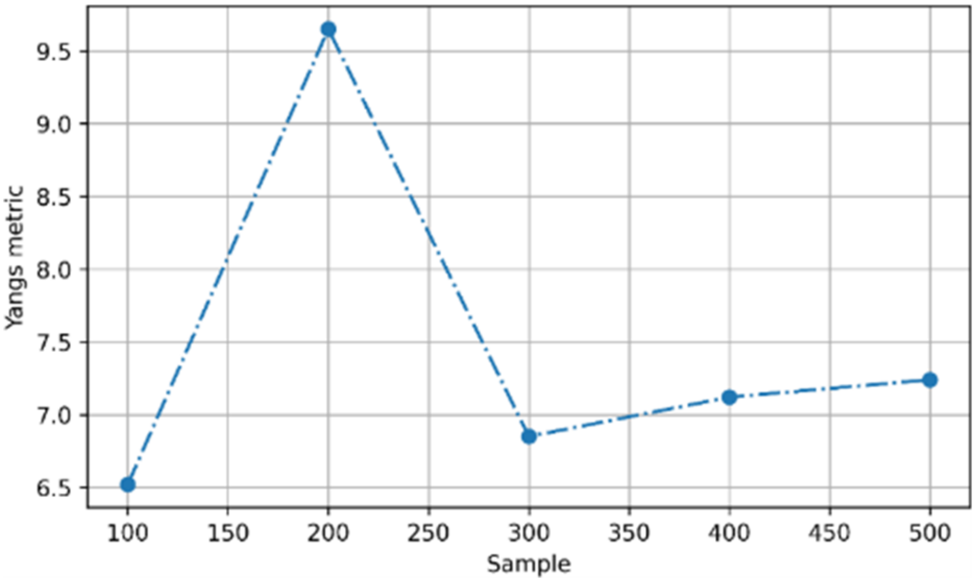

Figure 10. Yang's metric in the proposed model

In Figure 10, the yang's metric in the proposed model is described. With a sample of 100, the yang's metric of the proposed model is 6.5, at a sample of 200, the yang's metric of the proposed model raised to 9.7. Again, it dropped to 6.9, when the sample is 300, and finally the yang's metric reached to 7.25 when the sample is 500. The observed variations in Yang's metric across different sample sizes reflect the complexity and variability inherent in image fusion tasks. This variation indicates that the model has reached its capacity to learn from the data.

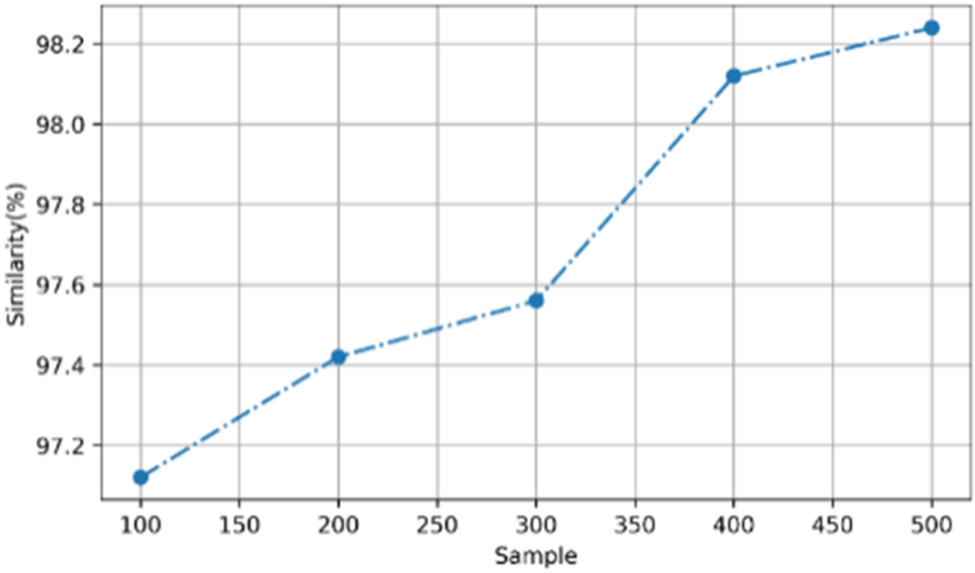

Figure 11. Similarity in the proposed model

The proposed model's Edge-based similarity metric is explained in Figure 11. With a sample of 100, the similarity of the proposed model is 97.1%, at a sample of 500, the similarity of the proposed model raised to 98.3%. This analysis highlights how the Versatile Object Class Cum Edge Detection Net, uses a convolutional layer with a ReLU activation function to promote edge responses and effectively capture edge information in the multi-focus image, which increases the similarity under varying samples.

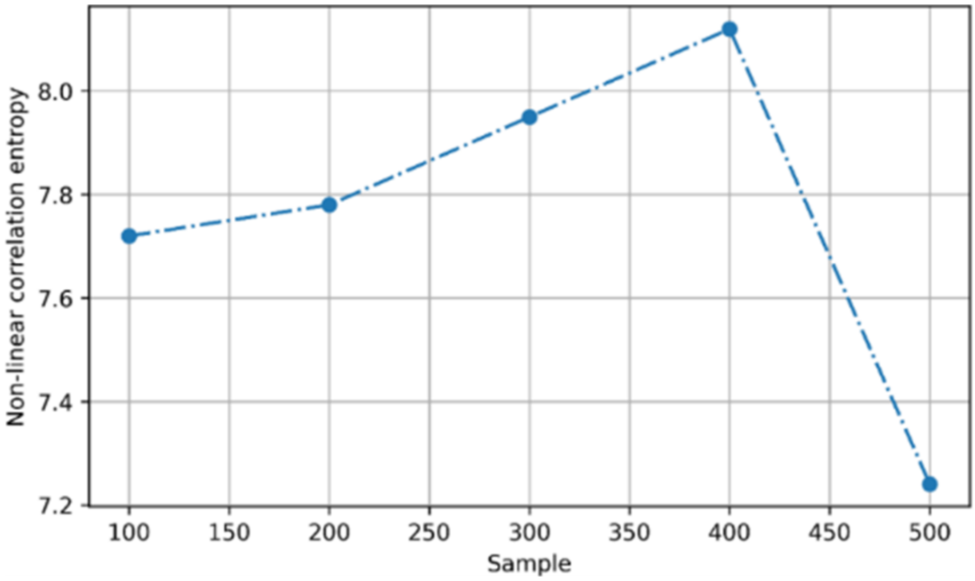

Figure 12. Non-linear correlation information entropy in the proposed model

Figure 12 explains the Non-linear correlation information entropy of the proposed model. With a sample of 100, the Non-linear correlation information entropy of the proposed model is 7.7, at a sample of 400, the Non-linear correlation information entropy of the proposed model raised to 8.15. Again, it dropped to 7.25, when the sample is 500. This investigation demonstrates how the Non-linear Correlation Information Entropy under varied samples is reduced using the Aggregated Fuzzy Decision Map Generator by efficiently merging information from multiple attention levels and object semantics.

The proposed model's Normalized Mutual Information is explained in Figure 13. With a sample of 100, the Normalized mutual information of the proposed model is 5.39, at a sample of 500, the Normalized mutual information of the proposed model raised to 6.21. This study shows that by integrating fuzzy

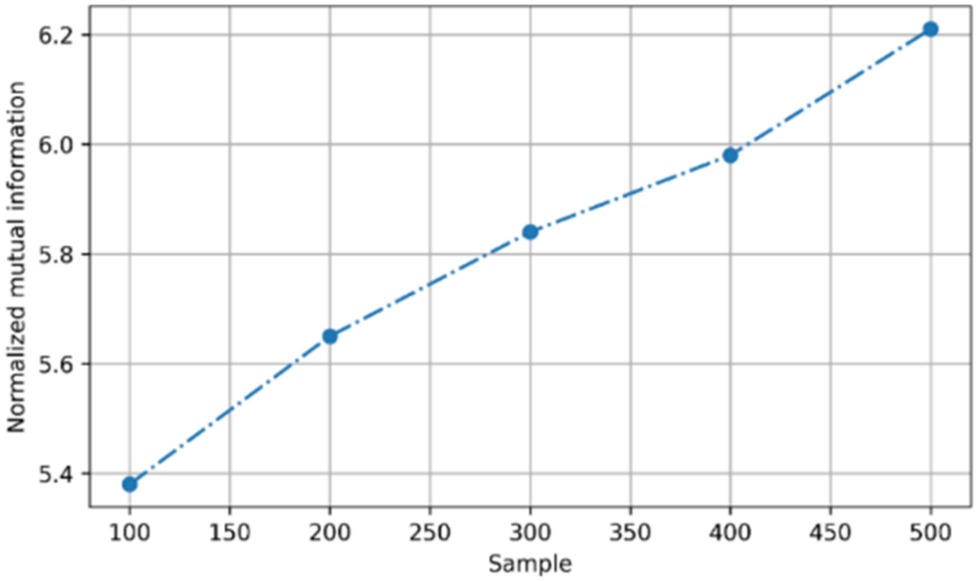

Figure 13. Normalized mutual information in the proposed model

membership functions and semantic interaction into the decision-making process, the Aggregated Fuzzy Decision Map Generator improves the Normalized mutual information under a variety of samples.

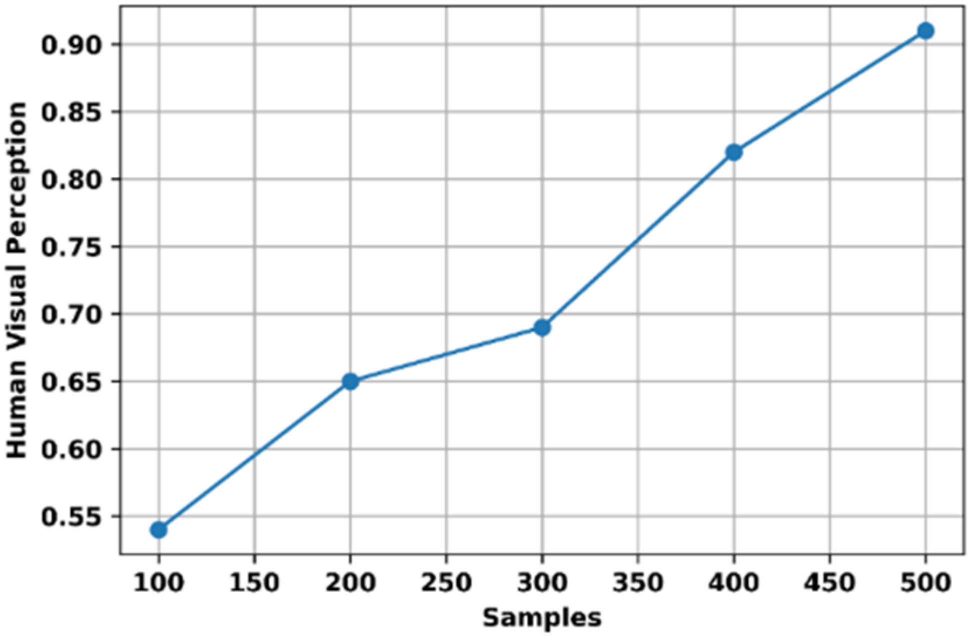

Figure 14. Human visual perception in the proposed model

Figure 14 explains how the human visual perception works in the proposed model. The human visual perception of the proposed model is 0.54 with a sample size of 100 and increases to 0.91 with a sample size of 500. This study demonstrates how Visio Quaternion Attentive Stitching enhances human visual perception across a range of samples by emphasizing texture continuity and spatial coherence in the final fused image.

The operation of Mutual information in the proposed model is explained in Figure 15. With a sample size of 100, the proposed model's mutual information is 0.61; with a sample size of 500, it rises to 0.94. This study shows how regular focus measurements and object semantic interaction, combined with an aggregated fuzzy decision map generator, to improve Mutual information across a variety of samples. The result is a decision map that precisely identifies focused regions inside the image.

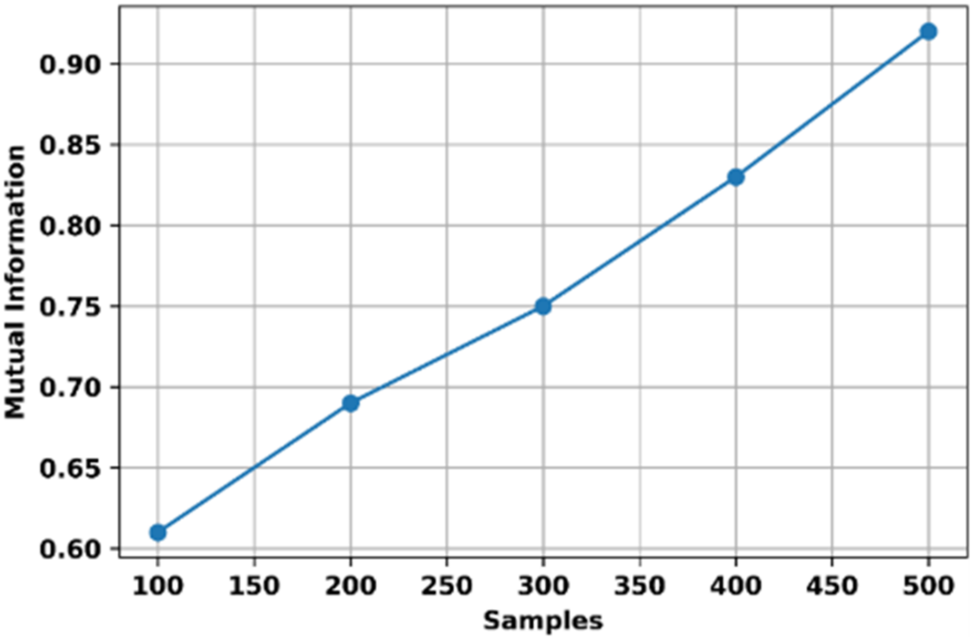

Figure 15. Mutual information in the proposed model

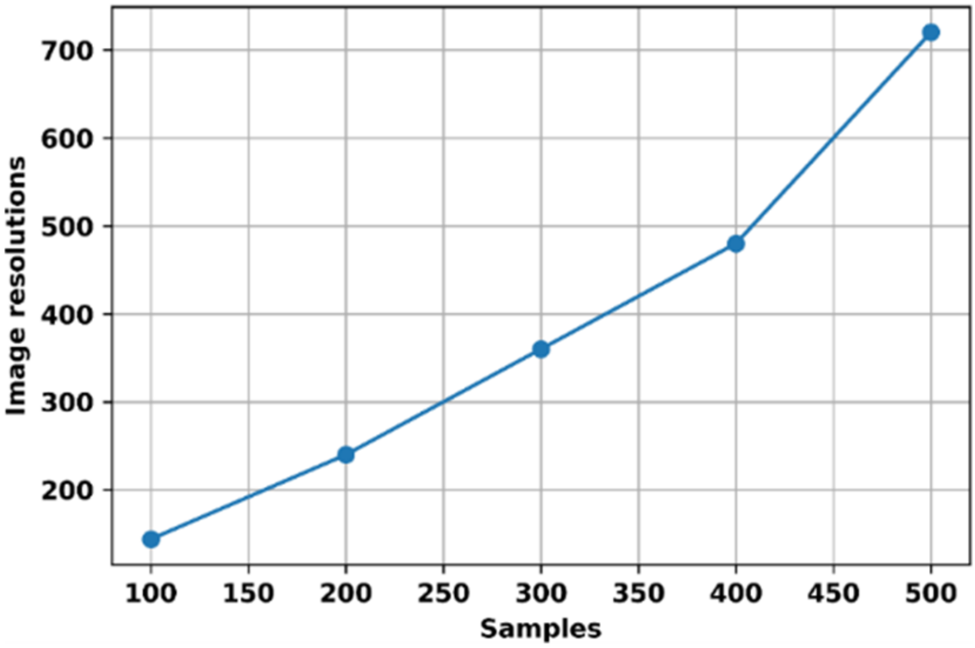

Figure 16. Image resolution in the proposed model

The operation of Image resolution in the proposed model is explained in Figure 16. With a sample size of 100, the proposed model's Image resolution is 144p; with a sample size of 500, it rises to 720p. This study shows how Visio Quaternion Attentive Stitching, improves Image resolution across a variety of samples, by improving depth perception and reducing texture discontinuities, ultimately leading to higher-quality fused images.

Figure 17 explains how Average fusion time (s) works in the proposed model. The average fusion time (s) of the proposed model is 12 s for a sample size of 100 and reduces to 3 s for a sample size of 500. This work demonstrates how spatial connections, amplitude, and phase information from the input images and decision map are directly integrated using Visio Quaternion Attentive Stitching to minimize Average fusion time (s) across a range of samples.

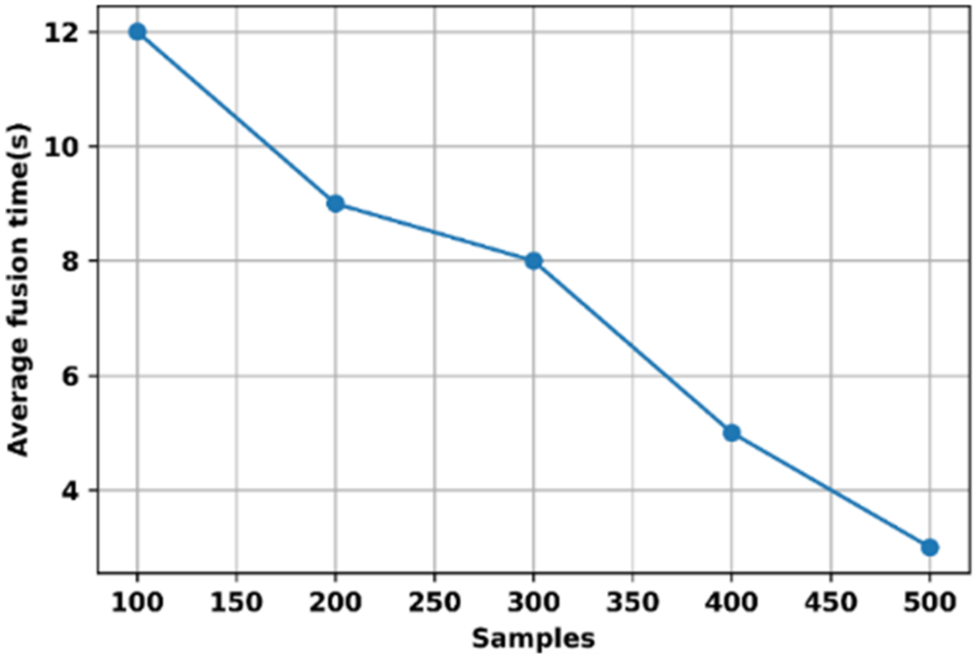

Figure 17. Average fusion time (s) in the proposed model

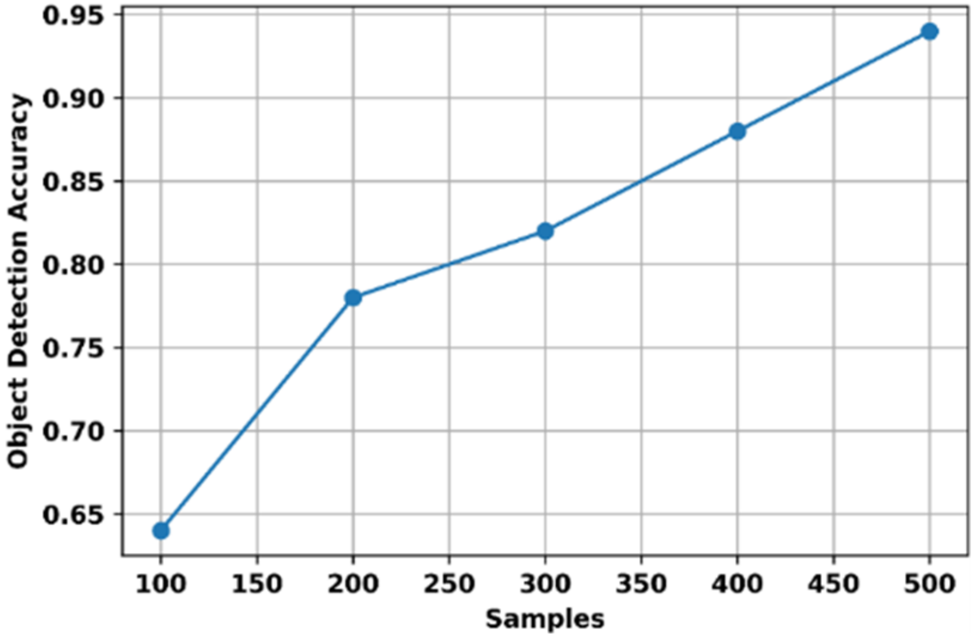

Figure 18. Object Detection Accuracy in the proposed model

The proposed model's Object Detection Accuracy is explained in Figure 18. The Object Detection Accuracy of the proposed model is 0.64 for a sample size of 100 and rise to 0.94 for a sample size of 500. This work shows how the Versatile Object Class Cum Edge Detection Net captures object classes, even for small objects, by varying anchor aspect ratios and sizes and using the Single Shot Multi Adaptive Anchor Box Detector (SSAD). This improves object detection accuracy across a variety of samples.

Figure 19 explains the Edge Detection Precision of the proposed model. For a sample size of 100, the proposed model's Edge Detection Precision is 0.60; for a sample size of 500, it increases to 0.92. The Versatile Object Class Cum Edge Detection Net improves edge detection precision in multi-focus image fusion tasks by combining object class detection and edge detection, utilizing semantic labeling and end-to-end learning, and improving the overall quality of the fused images across a range of samples.

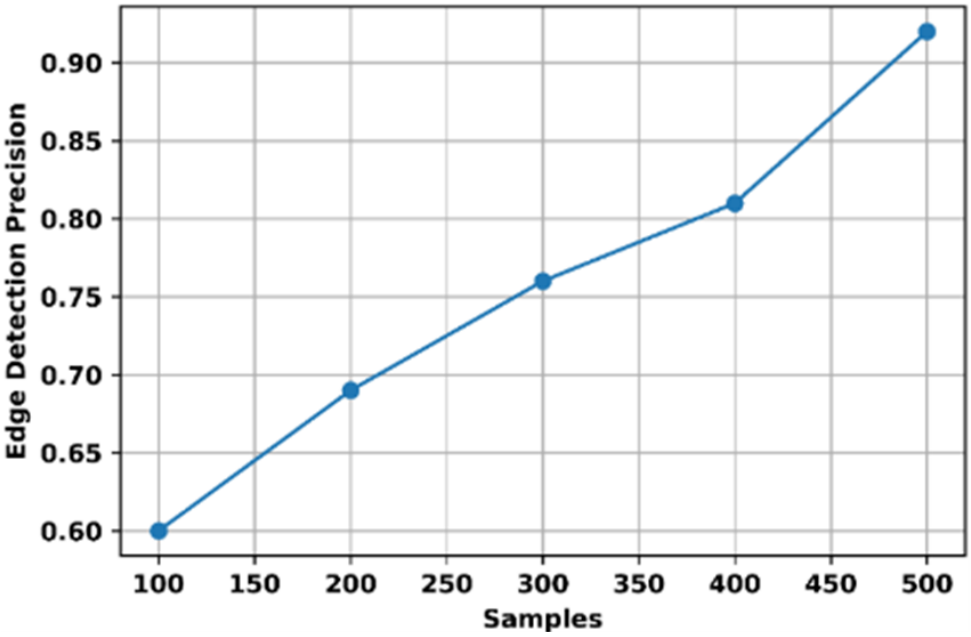

Figure 19. Edge Detection Precision in the proposed model

4.5. **Comparison analysis of the proposed model.** This section analyses the proposed method's comparison of Qualitative mutual information, Quality Assessment Based Fusion, Quantum Circuit Breaker, Object detection accuracy, Edge detection precision to existing CNN (Convolutional Neural Network), IMF (Improved Multiscale Fusion), FusionDN (Fusion Decomposition Network), U2fusion (U-2-Net based fusion) [26] and Quality Guided, Quality Metric, Quality Score, Sharpness Factor and Peak Signal-to-Noise Ratio to existing ReC (Resolution Enhancement Compression), SeA (Statistical Energy Analysis), SDD (Syntax Directed Definition) and U2F (Universal 2nd Factor) [27]. An in-depth investigation of these parameters shows how the proposed approach competes with or exceeds current methods in improving object and boundary detection in multi-focus image fusion.

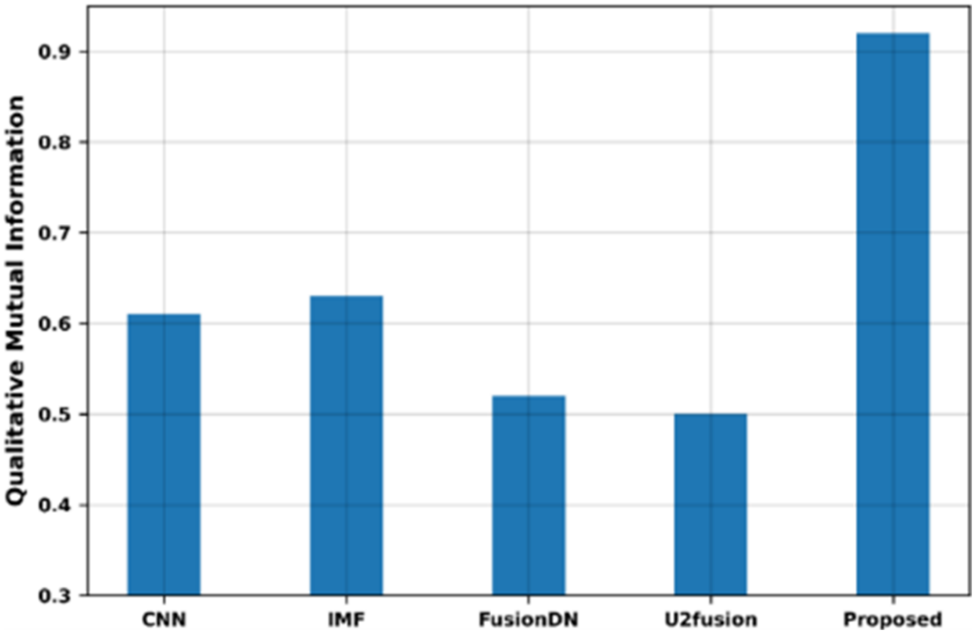

Figure 20. Comparison of Qualitative mutual information

The comparison of the proposed model's Qualitative mutual information with that of other current methods is shown in Figure 20. The proposed method's Qualitative mutual information is compared with that of other existing methods, including CNN, IMF, FusionDN, U2fusion. The proposed model's Qualitative mutual information comes up at 0.92, whereas the Qualitative mutual information of CNN, IMF, FusionDN, U2fusion is, 0.61, 0.63, 0.52, 0.50. Qualitative mutual information has been improved due to the Aggregated Fuzzy Decision Map Generator. The higher QMI indicates better preservation of meaningful content across focus regions. Existing methods rely primarily on low-level features, which fail to capture contextual object relationships, leading to lower mutual information.

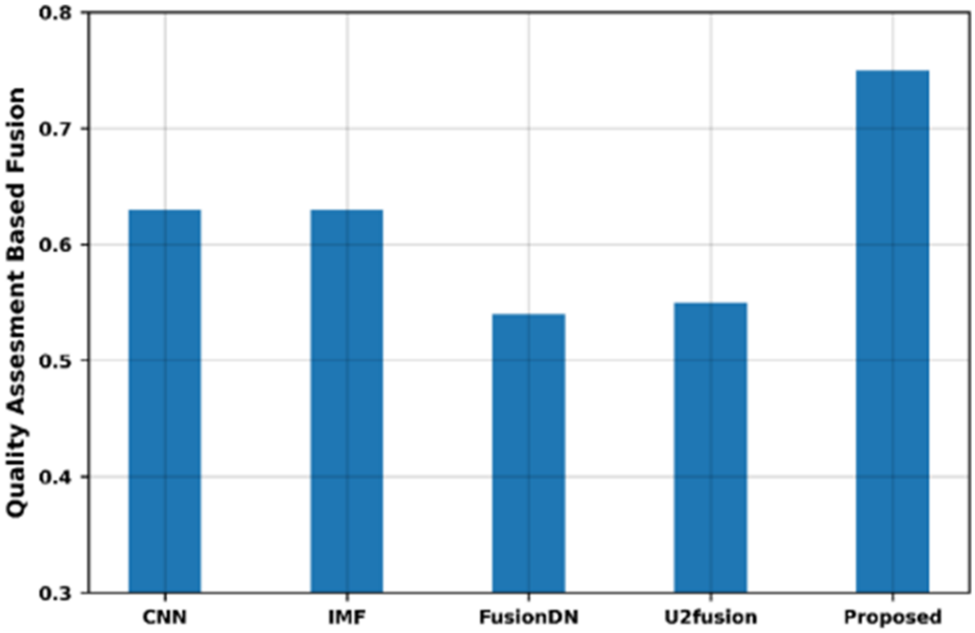

Figure 21. Comparison of Quality assessment-based fusion

Figure 21 compares the Quality assessment-based fusion of the proposed model with that of other existing approaches. The quality assessment-based fusion of the proposed method is compared with that of various current approaches, such as CNN, IMF, FusionDN, and U2fusion. The Quality evaluation-based fusion of CNN, IMF, FusionDN, and U2fusion is 0.63, 0.63, 0.54, and 0.55, whereas the quality assessment-based fusion of the proposed model yields a result of 0.75. By integrating semantic guidance from the Versatile Object Class Cum Edge Detection Net, the proposed framework ensures more context-aware focus selection, resulting in clearer and more visually coherent fused images, that existing models, this increases the Quality assessment-based fusion.

The proposed model's quantum circuit breaker is compared with those of other current methods in Figure 22. The proposed method's quantum circuit breaker is compared to that of many existing techniques, including CNN, IMF, FusionDN, and U2fusion. The proposed model's quantum circuit breaker produces a result of 0.76, while the quantum circuit breaker of CNN, IMF, FusionDN, and U2fusion is 0.56, 0.56, 0.51, and 0.49. This improvement of the proposed is the use of Quantum circuit breaker, which is provided by the Visio Quaternion Attentive Stitching.

Figure 23 compares the Object detection accuracy of the proposed model with various existing techniques. The object detection accuracy of the proposed approach is compared with several other approaches, such as CNN, IMF, FusionDN, and U2fusion. The object detection accuracy of the proposed model yields a value of 0.94, whereas CNN, IMF, FusionDN, and U2fusion give results of 0.42, 0.62, 0.86, and 0.73, respectively. This detection accuracy of the proposed fusion model is attributed to the dual-branch U-Net structure that jointly optimizes semantic labeling and edge detection. By considering semantic interactions between objects and their edges, the model effectively distinguishes similarly colored objects at different depths, a limitation in many traditional fusion methods.

The proposed model's Edge detection precision is compared with that of many other methods in Figure 24. The proposed method's Edge detection precision is compared with several alternative methods, including CNN, IMF, FusionDN, and U2fusion. The proposed model's Edge detection precision is 0.92, while CNN, IMF, FusionDN, and U2fusion provide values of 0.73, 0.73, 0.74, and 0.82, in that order. This high detection precision of the proposed model is due to the use of dual-branch U-Net, that separates semantic labeling from edge detection, improving boundary delineation without compromising object

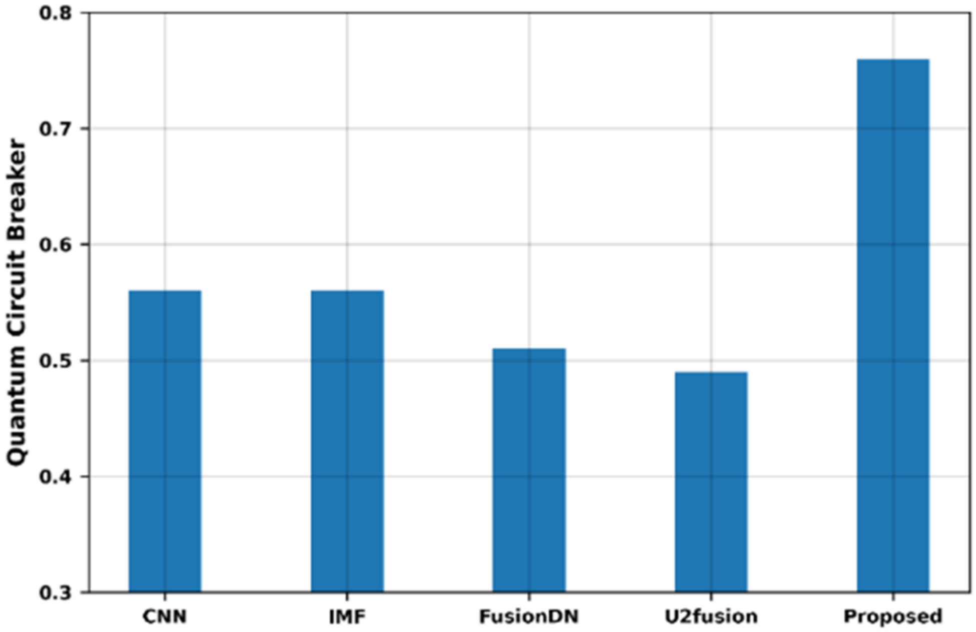

Figure 22. Comparison of Quantum circuit breaker

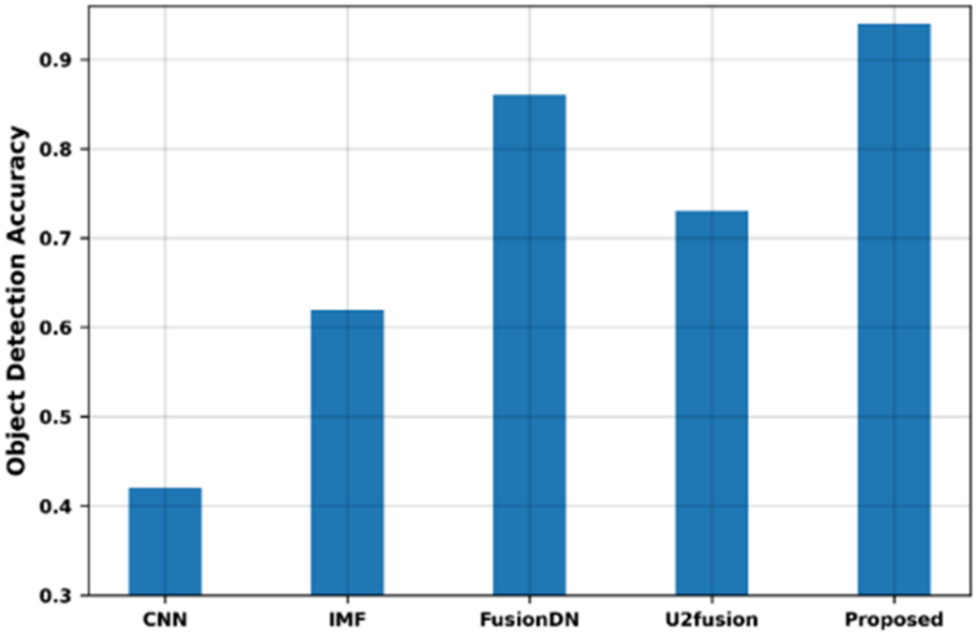

Figure 23. Comparison of Object detection accuracy

classification than existing models. This enhanced edge precision is vital for minimizing halo artifacts and improving the visual integrity of the fused image.

Figure 25 compares the Quality guided of the proposed model with several alternative approaches. The Quality guided of the proposed method is compared with several different approaches, including ReC, SeA, SDD, and U2F. The Quality guided of the proposed model is 0.71, while the values provided by ReC, SeA, SDD, and U2F are 0.35, 0.31, 0.31, and 0.42, respectively. This improvement in the proposed model is the result of the Aggregated Fuzzy Decision Map Generator, which integrates semantic context into focus-based decision making. The proposed method uses object semantics in directing, or fusion, which results in more contextually relevant, more focused regions to decision making than traditional methods that only rely on sharpness.

The Quality metric of the proposed approach is compared with several other methods in Figure 26. The proposed method's Quality metric is compared with that of various other techniques, such as ReC,

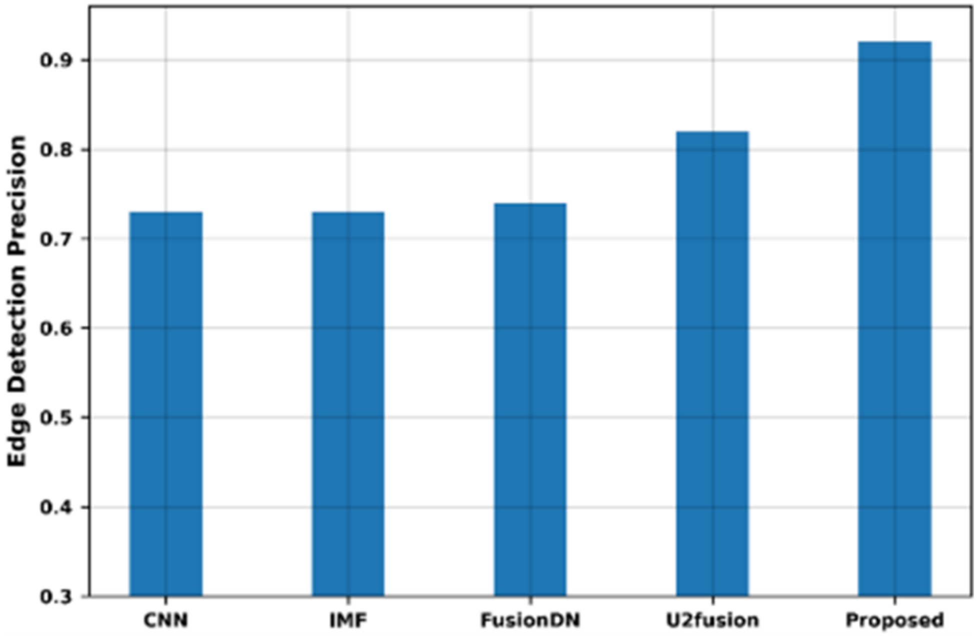

Figure 24. Comparison of Edge detection precision

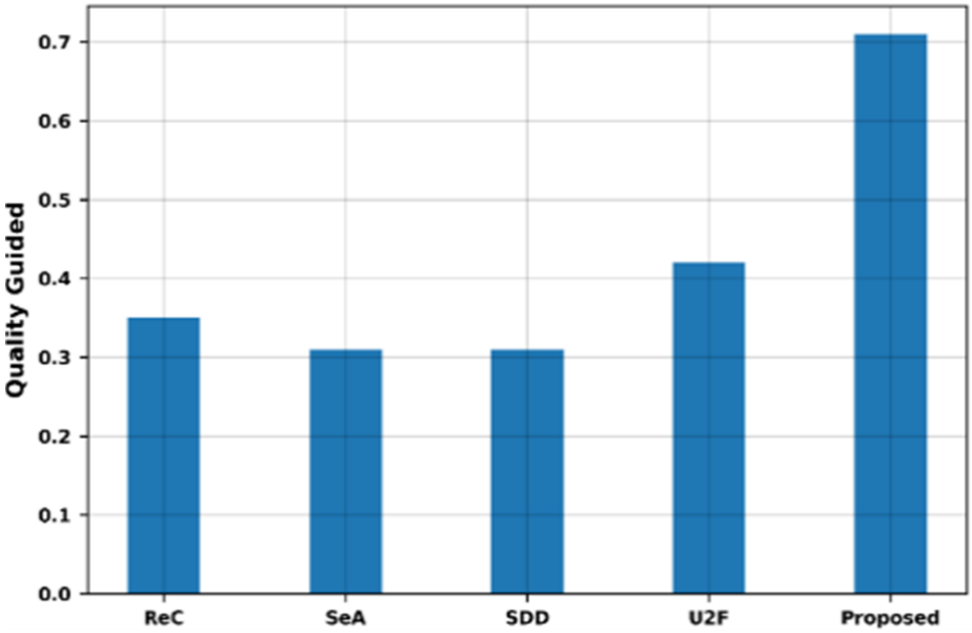

Figure 25. Comparison of Quality guided

SeA, SDD, and U2F. The Quality metric of the proposed model is 0.74, while ReC, SeA, SDD, and U2F yield values of 0.48, 0.37, 0.46, and 0.54 respectively. The improved score comes from the Versatile Object Class Cum Edge Detection Net, which makes it easier to tell the difference between objects and edges, making sure that the meaning is clear and reducing areas where unclear. This makes the quality metric better than traditional models.

In Figure 27, the Quality score of the proposed method is compared with many different approaches. The Quality score of the proposed method is compared with many other approaches, including ReC, SeA, SDD, and U2F. The proposed approach has a Quality score of 0.92 and yields values of 0.78, 0.69, 0.70, and 0.83 for ReC, SeA, SDD, and U2F, respectively. Aggregated Fuzzy Decision Map Generator offers a higher quality score.

The proposed method's Sharpness factor is compared with different alternatives in Figure 28. The proposed method's Sharpness factor is compared with several other strategies, such as ReC, SeA, SDD,

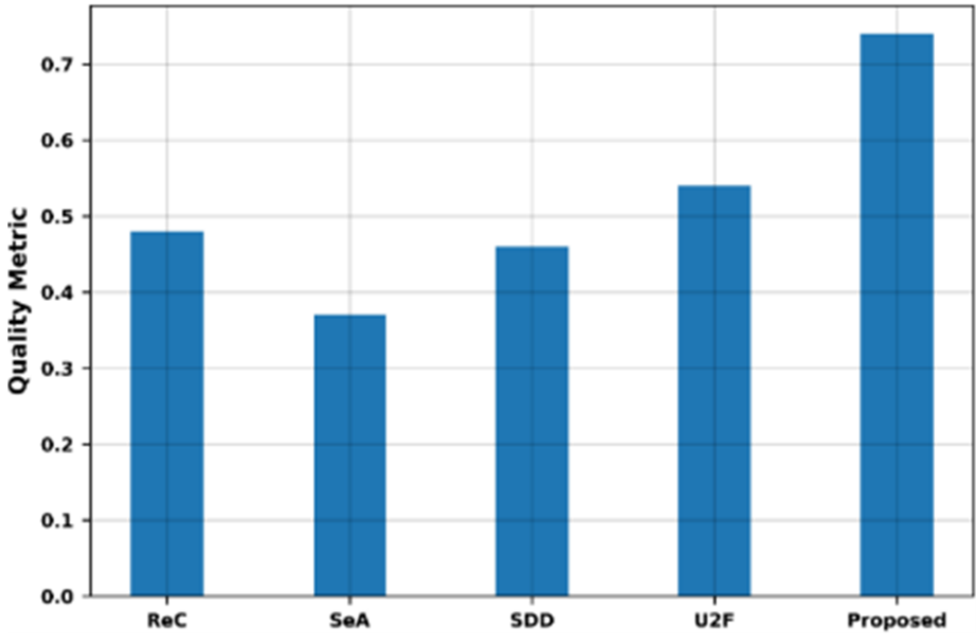

Figure 26. Comparison of Quality metric

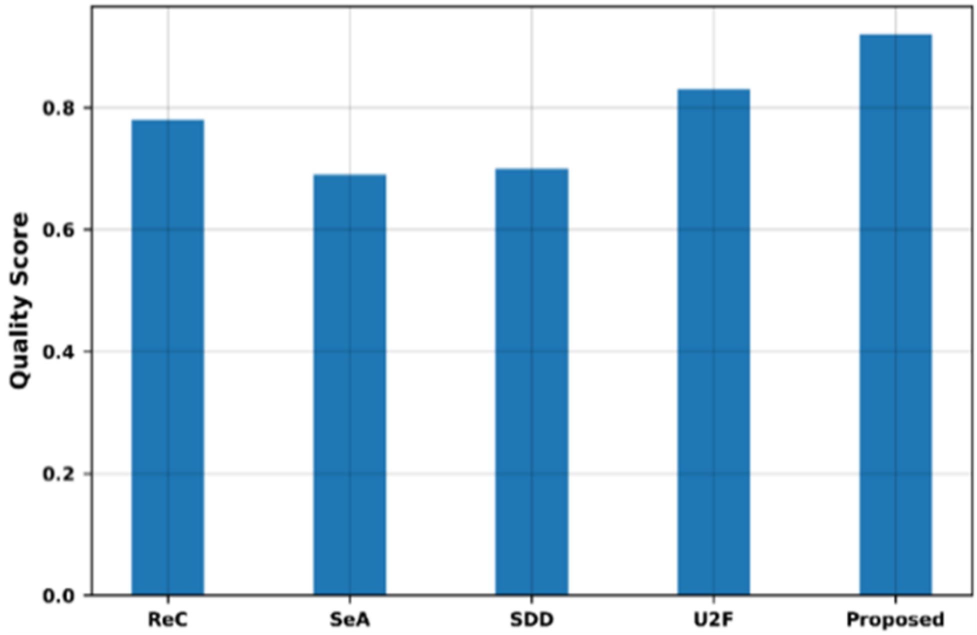

Figure 27. Comparison of Quality score

and U2F. In addition to producing values of 6.22, 13.88, 9.99, and 8.01 pixels per unit distance for ReC, SeA, SDD, and U2F, respectively, the proposed method has a Sharpness factor of 18.5 pixels per unit distance. A greater Sharpness factor is provided by the Aggregated Fuzzy Decision Map Generator.

Figure 29 compares the Peak-signal-to-noise ratio of the proposed approach with several alternatives. The Peak-signal-to-noise ratio of the proposed method is compared with several alternative approaches, including ReC, SeA, SDD, and U2F. The proposed method has a Peak-signal-to-noise ratio of 25.6 dB and yields values of 18.5 dB, 16.18 dB, 14.23 dB, and 19.21 dB for ReC, SeA, SDD, and U2F, respectively. The Visio Quaternion Attentive Stitching module decreases the halo artifacts and enhances frequency-domain information integration, that leads to higher PSNR and visually pleasing results compared to the existing models.

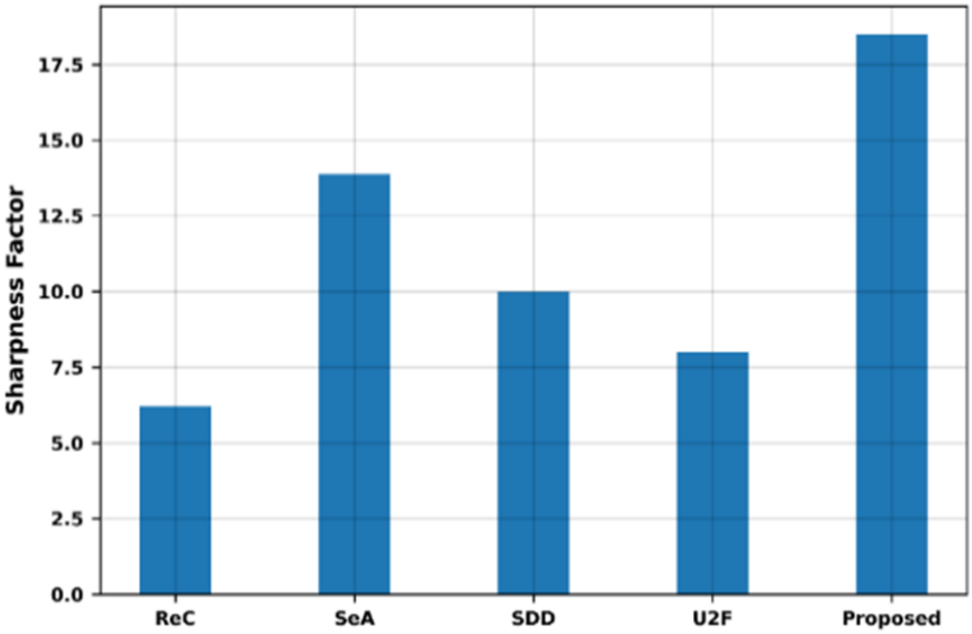

Figure 28. Comparison of Sharpness factor

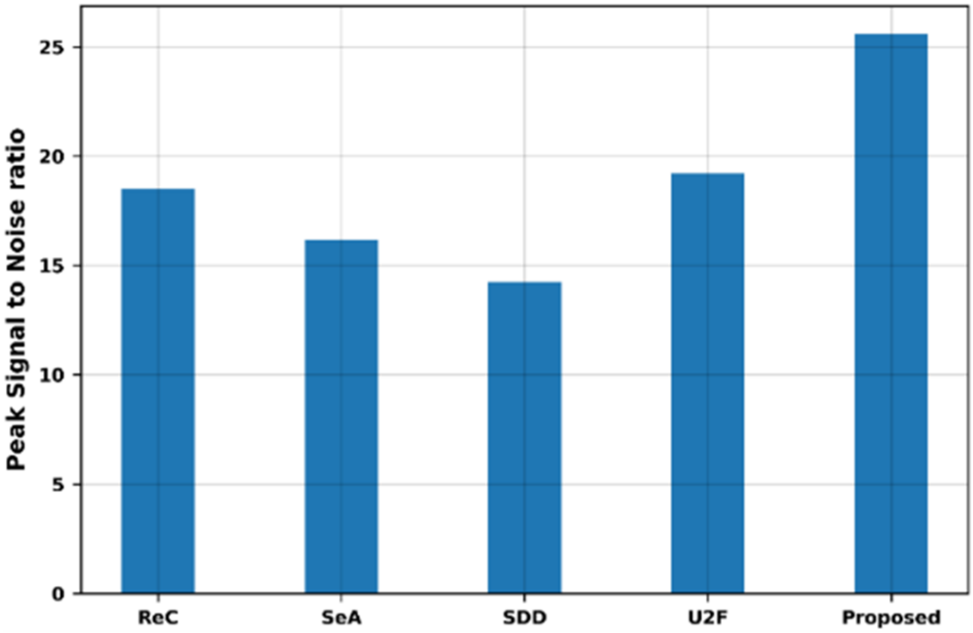

Figure 29. Comparison of Peak-signal-to-noise ratio

4.6. **Discussion.** Analysis of the comparative evaluation of proposed Enhanced Fusion Strategy for Multi-Focus Image with Object Semantic Detection and Attentive Transformer Network demonstrates that the model displayed significant improvements in all image quality, object detection, and edge retaining results compared to various contemporary baseline methods, including CNN, IMF, FusionDN, U2fusion, ReC, SeA, SDD, and U2F.

As captured in Table 1, the approach presented attained a Qualitative Mutual Information score of 0.92, a fusion score based on quality assessment of 0.75, and a quantum circuit breaker score of 0.76, all of which outperformed conventional fusion methods. The detection accuracy of 0.94 for objects and precision of 0.92 for edge detection further highlight the superiority of proposed Versatile Object Class Cum Edge Detection Net and the Aggregated Fuzzy Decision Map Generator and VQAS. These measures showcase that the approach successfully maintains object boundaries, suppresses halo artifacts, and ensures semantic consistency in multi-focus images.

Table 1. Comparative Performance of Proposed Model vs Baseline Models (for fusion and detection metrics)

Metric	**Proposed**	**CNN**	**IMF**	**FusionDN**	**U2Fusion**
Qualitative Mutual Info	0.92	0.61	0.63	0.52	0.50
Quality Assessment Fusion	0.75	0.63	0.63	0.54	0.55
Quantum Circuit Breaker	0.76	0.56	0.56	0.51	0.49
Object Detection Accuracy	0.94	0.42	0.62	0.86	0.73
Edge Detection Precision	0.92	0.73	0.73	0.74	0.82

Table 2. Comparison of proposed model with ReC, SeA, SDD, U2F (for quality metrics)

Metric	**Proposed**	**ReC**	**SeA**	**SDD**	**U2F**
Quality Guided	0.71	0.35	0.31	0.31	0.42
Quality Metric	0.74	0.48	0.37	0.46	0.54
Quality Score	0.92	0.78	0.69	0.70	0.83
Sharpness Factor (px/unit)	18.5	6.22	13.88	9.99	8.01
PSNR (dB)	25.6	18.5	16.18	14.23	19.21

Examining the quality metrics (Table 2), the new method decidedly beats traditional methods in yielding the highest quality guided (0.71), quality metric (0.74), quality score (0.92), sharpness factor (18.5 pixels/unit), and PSNR (25.6 dB). The measures reflect a significant enhancement in image clarity preservation, texture details, and overall perceptual quality. This enhancement of the suggested model is the use of Aggregated Fuzzy Decision Map Generator and VQAS, which altogether produce improved overall image quality through the combination of low- and high-level features. The large enhancement in the sharpness factor shows that the fused image retains fine textures, which is very important in applications such as object recognition. Larger PSNR values imply that noise and artifacts are well reduced, creating visually pleasing fused images. It is the synergy of semantic-aware feature extraction, fuzzy decision mapping, and quaternion attentive fusion that is behind these improvements. It is the best to use in order to optimize multi-focus image fusion object and boundary detection. The method is not only technically robust but also practically applicable in real-world scenarios requiring high-fidelity image fusion.

4.6.1. *Limitations.* Despite its improvements over existing multi-focus image fusion techniques, the proposed Enhanced Fusion Approach has certain limitations. While this framework assumes similar quality in the source images and similar exposure levels, extreme noise, changes in lighting, and motion blur disrupts the accuracy of the fusion and produce artifacts. Hyperparameter tuning for the aggregation of fuzzy decision maps and quaternion attention methods requires careful experimentation, limiting the method's generalizability across contrasting imaging domains.

5. **Conclusion.** In relation to the improvement of the quality of the object and boundary detection in the multi-focus image fusion, a unique novel "Enhanced Fusion Approach for Multi-Focus Image with Object Semantic Detection and Attentive Transformer Network" was proposed. The proposed method Versatile Object Class Cum Edge Detection Net was implemented for the improvement of edge detection which removed the difficulties in distinguishing different objects with the same color, while SSAD was constructed for object class detection captured the object classes for even small object and Dual Out Branch U-Net was introduced for edge detection which shaped an effective edge detection, with 0.94 of object detection accuracy, 0.92 of edge detection precision and 0.74 of the quality metric. Then, using regular focus measurements and object semantic interaction, the Aggregated Fuzzy Decision Map Generator created a decision map that guarantees overall focus quality and satisfaction for every pixel or region in a multi-focus image, with 0.92 of qualitative mutual information, 0.75 of quality assessment-based fusion, 0.71 of quality guided, 0.92 of quality score and 18.5 pixels per unit distance of sharpness factor. Furthermore, a Visio Quaternion Attentive Stitching guaranteed a smooth transition, decreased the disparities, and eliminated halo artifacts, with 0.76 quantum circuit breakers and 25.6 dB of Peak-signal-to-noise ratio. These innovative methods solve the problems and enhance the object's quality and boundary detection. It is a significant development in the area of multi-focus image fusion techniques.

REFERENCES

[1] H. Jung, Y. Kim, H. Jang, N. Ha, K. Sohn, “Unsupervised deep image fusion with structure tensor representations,” *IEEE Trans. Image Process.* vol. 29, 2020, pp. 3845–3858.

[2] 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, 2020, pp. 71–91.

[3] M.A. Azam, K.B. Khan, S. Salahuddin, E. Rehman, S.A. Khan, M.A. Khan, S. Kadry, and A.H. Gandomi, “A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics,” *Computers in biology and medicine,* vol. 144, 2022, pp. 105253.

[4] L. Tang, J. Yuan, H. Zhang, X. Jiang, and J. Ma, “PIAFusion: A progressive infrared and visible image fusion network based on illumination aware,” *Information Fusion,* vol. 83, pp. 79–92, 2022.

[5] W. Lv, Y. Wang, X. Chen, X. Fu, J. Lu, P. Li, “Enhancing vascular visualization in laser speckle contrast imaging of blood flow using multi-focus image fusion,” *J. Biophotonics,* vol. 12, no. 1, 2019, pp. e201800100.

[6] L. Tang, Y. Deng, Y. Ma, J. Huang, and J. Ma, “SuperFusion: A versatile image registration and fusion network with semantic awareness,” *IEEE/CAA Journal of Automatica Sinica,* vol. 9, no. 12, pp. 2121–2137, 2022.

[7] V. Vs, J.M.J. Valanarasu, P. Oza, and V.M. Patel, “Image fusion transformer,” *In 2022 IEEE International Conference on Image Processing (ICIP),* pp. 3566–3570, 2022, October. IEEE.

[8] J. Li, G. Yuan, H. Fan, “Multifocus image fusion using wavelet-domain-based deep cnn,” *Comput. Intell. Neurosci.,* 2019, pp. 23.

[9] J. Li, X. Guo, G. Lu, B. Zhang, Y. Xu, F. Wu, D. Zhang, “Drpl: deep regression pair learning for multi-focus image fusion,” *IEEE Trans. Image Process.,* Vol. 29, 2020, pp. 4816–4831.

[10] K. Bhalla, D. Koundal, B. Sharma, Y.C. Hu, and A. Zaguia, “A fuzzy convolutional neural network for enhancing multi-focus image fusion,” *Journal of Visual Communication and Image Representation,* vol. 84, 2022, pp. 103485.

[11] D. Rao, T. Xu, and X.J. Wu, “Tgfuse: An infrared and visible image fusion approach based on transformer and generative adversarial network,” *IEEE Transactions on Image Processing.*

[12] J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma, “SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer,” *IEEE/CAA Journal of Automatica Sinica,* vol. 9, no. 7, pp. 1200–1217, 2022.

[13] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, “U2Fusion: A unified unsupervised image fusion network,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 44, no. 1, pp. 502–518, 2020.

[14] Y. Gao, Y. Su, Q. Li, J. Li, “Single fog image restoration with multi-focus image fusion,” *J. Vis. Commun. Image Represent.* Vol. 55, 2018, pp. 586–595.

[15] N. Alseelawi, H.T. Hazim, and Salim H.T. ALRikabi, “A Novel Method of Multimodal Medical Image Fusion Based on Hybrid Approach of NSCT and DTCWT,” *International Journal of Online & Biomedical Engineering,* vol. 18, no. 3, 2022.

[16] P. Wu, L. Jiang, Z. Hua, and J. Li, “Multi-focus image fusion: Transformer and shallow feature attention matters,” *Displays,* vol. 76, pp. 102353, 2023.

[17] S. Liu, W. Peng, Y. Liu, J. Zhao, Y. Su, and Y.D. Zhang, “AFCANet: An adaptive feature concatenate attention network for multi-focus image fusion,” *Journal of King Saud University-Computer and Information Sciences,* pp. 101751, 2023.

[18] S. Liu, J. Ma, Y. Yang, T. Qiu, H. Li, S. Hu, and Y.D. Zhang, “A multi-focus color image fusion algorithm based on low vision image reconstruction and focused feature extraction,” *Signal Processing: Image Communication,* vol. 100, pp. 116533, 2022.

[19] X. Wang, Z. Hua, and J. Li, “Multi-focus image fusion framework based on transformer and feedback mechanism,” *Ain Shams Engineering Journal,* vol. 14, no. 5, pp. 101978, 2023.

[20] C.R. Mohan, S. Kiran, and A.A. Kumar, “An Enhancement Process for Multi-Focus Images Resulted from Image Fusion using qshiftN DTCWT and MPCA in Multiresolution Domain,” *Procedia Computer Science,* vol. 218, pp. 2713–2722, 2023.

[21] M. Lv, L. Li, Q. Jin, Z. Jia, L. Chen, and H. Ma, “Multi-focus image fusion via distance-weighted regional energy and structure tensor in NSCT domain,” *Sensors,* vol. 23, no. 13, pp. 6135, 2023.

[22] H. Zhai, W. Zheng, Y. Ouyang, X. Pan, and W. Zhang, “Multi-focus image fusion via interactive transformer and asymmetric soft sharing,” *Engineering Applications of Artificial Intelligence,* vol. 133, pp. 107967, 2024.

[23] S. Kiran, and A.A. Kumar, "All-in-Focus Imaging using qshiftN DTCWT and LP in the Frequency Partition Domain," *In 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom),* pp. 754–759, 2022, March. IEEE.

[24] C. Zheng, "Stack-YOLO: A friendly-hardware real-time object detection algorithm," *IEEE Access,* 2023.

[25] S. Liu, Y. Lian, Z. Zhang, B. Xiao, and T.S. Durrani, "Cross-scale vision transformer for crowd localization," *Journal of King Saud University-Computer and Information Sciences,* pp. 101972, 2024.

[26] X. Jin, X. Xi, D. Zhou, X. Ren, J. Yang, and Q. Jiang, "An unsupervised multi-focus image fusion method based on Transformer and U-Net," *IET Image Processing,* vol. 17, no. 3, pp. 733–746, 2023.

[27] X. Li, X. Li, T. Ye, X. Cheng, W. Liu, and H. Tan, "Bridging the gap between multi-focus and multi-modal: a focused integration framework for multi-modal image fusion," *In Proceedings of the IEEE/CVF winter conference on applications of computer vision* pp. 1628–1637, 2024.