



OPEN Multi scale multi attention network for blood vessel segmentation in fundus images

Giri Babu Kande¹, Madhusudana Rao Nalluri^{2,3}✉, R. Manikandan⁴, Jaehyuk Cho⁵✉ & Sathishkumar Veerappampalayam Easwaramoorthy⁶

Precise segmentation of retinal vasculature is crucial for the early detection, diagnosis, and treatment of vision-threatening ailments. However, this task is challenging due to limited contextual information, variations in vessel thicknesses, the complexity of vessel structures, and the potential for confusion with lesions. In this paper, we introduce a novel approach, the MSMA Net model, which overcomes these challenges by replacing traditional convolution blocks and skip connections with an improved multi-scale squeeze and excitation block (MSSE Block) and Bottleneck residual paths (B-Res paths) with spatial attention blocks (SAB). Our experimental findings on publicly available datasets of fundus images, specifically DRIVE, STARE, CHASE_DB1, HRF and DR HAGIS consistently demonstrate that our approach outperforms other segmentation techniques, achieving higher accuracy, sensitivity, Dice score, and area under the receiver operator characteristic (AUC) in the segmentation of blood vessels with different thicknesses, even in situations involving diverse contextual information, the presence of coexisting lesions, and intricate vessel morphologies.

In the development of a computer-aided fundus image analysis system, the automated delineation of retinal vasculature holds utmost importance. This is because the analysis of retinal blood vessels can unveil the presence of various pathologies, including arteriosclerosis, vein occlusions, diabetic retinopathy, age-related macular degeneration, glaucoma, and hypertensive retinopathy^{1–3}. Table 1 provides a comprehensive listing of symptoms and diseases diagnosable through the characteristics of retinal vasculature. Furthermore, the physical attributes of retinal blood vessels, encompassing their length, diameter, branching patterns, shape, and tortuosity, are leveraged in various applications such as biometric identification, classification of arteries and veins, and registration of fundus images^{4–7}. Hence, the development of reliable techniques for segmenting the retinal vasculature is of utmost importance.

The segmentation of retinal vasculature involves accurately identifying the locations of blood vessels pixel by pixel, irrespective of differences in vessel sizes, branching patterns, and angles. Figure 1 illustrates the substantial challenges that complicate the precise segmentation of retinal vasculature.

1. Fundus images featuring blood vessels of varying widths pose a challenge for segmentation with clinical precision.
2. Central light reflexes observed in retinal images due to fluctuations in light intensity can complicate segmentation. Typically, a blood vessel is characterized by a central line flanked by two edges. However, when a blood vessel exhibits a central light reflex, it may appear as two distinct vessels.
3. Blood vessels located in regions of poor contrast due to uneven background lighting present a significant segmentation challenge, particularly in low-contrast scenarios. This challenge becomes more pronounced with smaller vessels.
4. The incidence of various pathologies such as hemorrhages and microaneurysms within retinal images can lead to false segmentation as actual blood vessels.

¹Vasireddy Venkatadri Institute of Technology, Nambur 522508, India. ²School of Computing, Amrita Vishwa Vidyapeetham, Amaravati 522503, India. ³Department of Computer Science & Engineering, Faculty of Science and Technology (IcfaiTech), ICFai Foundation for Higher Education, Hyderabad, India. ⁴School of Computing, SASTRA Deemed University, Thanjavur 613401, India. ⁵Department of Software Engineering & Division of Electronics and Information Engineering, Jeonbuk National University, Jeonju-si 54896, Republic of Korea. ⁶Department of Data Science and Artificial Intelligence, Sunway University, 47500 Petaling Jaya, Selangor Darul Ehsan, Malaysia. ✉email: madhu031083@gmail.com; chojh@jbnu.ac.kr

| S.No | Symptoms | Disease |
|------|--|---|
| 1 | Arteriovenous crossings and artery width | Hypertensive retinopathy |
| 2 | Tortuous blood vessels | Hypertension, atherosclerosis, diabetic retinopathy (DR), cerebral vessel disorders |
| 3 | Growing of new blood vessels | Neovascularization (a severe stage of DR) |
| 4 | Microvasculature around the fovea leaks or becomes dilated | Retinal telangiectasia |
| 5 | Artery vein ratio and crossings | High blood pressure or cardiovascular risk |
| 6 | Retinal vein occlusion | Diabetes and high blood pressure |
| 7 | Arterial embolism | Tissue damage, a stroke, blindness or even death |

Table 1. Significance of retinal vasculature in disease identification.

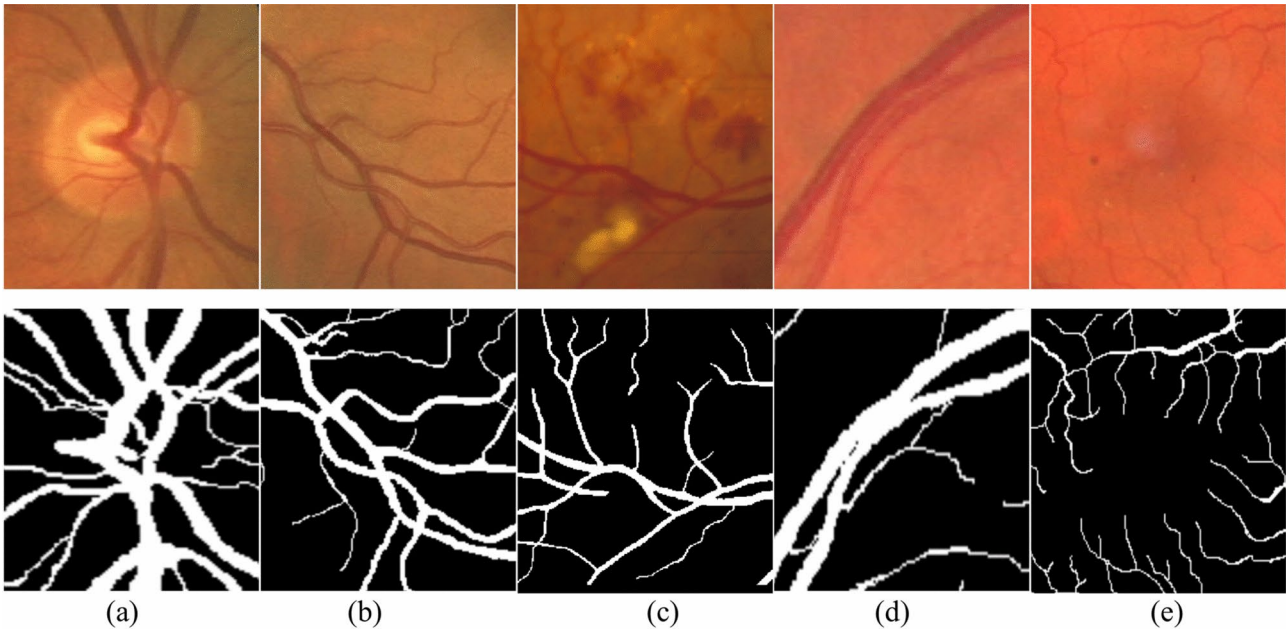


Fig. 1. Challenges present in the segmentation of retinal vasculature. The first row presents cropped regions from fundus images, and the second row represents corresponding cropped regions of the ground truth. (a) Blood vessels with variable thicknesses (b) centre reflex and cross-over vessels (c) Blood vessels along with pathologies (d) Vessels in close proximity (e) Tiny blood vessels with blurred boundaries.

5. Retinal vasculature exhibits highly intricate branching patterns, often intersecting with one another, making detection challenging at bifurcations and cross-overs. Additionally, vessels in close proximity may align parallel to each other, creating the illusion of a single thick vessel.

These challenges underscore the need for dependable and precise methods for retinal vascular segmentation. Currently, the majority of retinal vessel segmentation methods rely on deep learning techniques owing to their ability to automatically learn hierarchical features, leading to substantial improvements in accuracy and efficiency. SegNet⁸, UNet⁹, and their modified variants are notable deep-learning models frequently utilised for vessel segmentation tasks. Recent developments in this domain have brought forth novel architectures aimed at bolstering performance. As an illustration, a Recurrent Convolutional Neural Network (RU-Net)¹⁰ has been introduced, leveraging the efficacy of the U-Net architecture. DU-Net¹¹ employed deformable convolution to segment vessels across different scales. A dual encoder U-Net was proposed to capture spatial information and preserve semantic information¹². Enhancements in capturing spatial context data were achieved with Dense UNet¹³, whereas S-UNet¹⁴ used a saliency mechanism. OCE-Net¹⁵, was adept at extracting complex orientation and contextual information from blood vessels. Conversely, fusion and multitask segmentation networks were employed for vessel segmentation¹⁶. A shared decoder and pyramid-like loss were used to enhance vessel extraction significantly¹⁷. Automatic vascular network segmentation is achieved through a directed graph search-based multi-attentive neural network method¹⁸. Lightweight convolution blocks were employed to decrease computation intricacy and improve the segmentation outcomes¹⁹. The quest for precise vasculature segmentation faces a formidable obstacle in recognising diverse thicknesses. Multi-scale networks aim to capture a broader spectrum of data^{20–22}. Hu²⁰ introduced a multi-scale framework to gather more detailed multi-scale vascular characteristics. Multi-scale inputs with side-output layers were employed for detecting the boundaries of the cup and disk jointly²¹.

Zhang²² incorporated an attention-guided filter into the network, enhancing its capability to focus on relevant features. The ResDO-Unet²³ model utilises an attention-based fusion block to capture multiscale features effectively. A dice loss function based on multi-scale is employed to integrate characteristics from different layers²⁴. Filter operating at multiple scales are applied to address the variation in vessel thickness²⁵. Furthermore, a segment-level loss function was used to improve the detection accuracy of tiny vasculature²⁶. While multi-scale architectures have become more adept at segmenting a wider array of blood vessels, they frequently encounter a traditional compromise: focusing on thin vessels might be slightly inferior in their ability to detect thicker blood vessels.

The identified challenges pertaining to the delineation of vasculature from colour fundus images particularly necessitated innovation at every possible stage of the methodology of the solution. Inspired by the recent progress in the application of deep learning in computer vision, we developed an architecture that enabled the capturing of diverse, rich multi scale features of the vessels under various conditions in the encoder blocks, replaced skip connections with residual learning paths to make information from the output of the encoders more compatible to fuse with the output of the decoders at various depths, and the fusion is enabled by image dependent spatial attention maps derived by attention blocks. With these proposed modifications and the other adopted preprocessing steps along with the customised training strategies we are presenting below the contributions of our paper in the segmentation of blood vessels from color fundus image:

1. The multi-scale squeeze and excitation, dubbed as MSSE encoder/decoder, block is the foremost contribution to the field of research which can capture information under different scales and rank the channels of the feature map based on the tasks at hand, an important requirement in the tasks like delineation of vessels. The thickness of vessels varies in between 3 to 30 pixels and they could be surrounded by various structures, different kernels of MSSE blocks learn different amounts information of these structures and rank them based on amount of vessel information that they capture.
2. A customized input pipeline is proposed to counteract the challenges caused by fraudulent vessels and the unavailability of rich contextual information. The pipeline essentially consists of (i) employing morphological operations to minimise the effect of central light reflexes in the vessels and (ii) generating the different combinations of grayscale, CLAHE, and shade-corrected images extracted under the preprocessing step to enhance the feature representation power of the model.
3. A cohort of fixed-size patches cropped at random locations is generated as part of the input pipeline using which the model is trained. The cohort potentially is a large augmented dataset that helps the model to be less sensitive to fraudulent vessels caused by a variety of pathologies.
4. Deeper neural networks often face challenges with feature propagation to the later stages, leading to inferior feature propagation. Moreover, they are prone to the vanishing gradient issue. These problems complicate the preservation of intricate vessel morphology through different layers, especially within the U-net architecture, where transferring features between the encoder and decoder across various depths becomes intricate. By employing the B-Res path, we effectively mitigated these issues, resulting in improved outcomes.
5. Decoder outputs are utilised in two different ways. The Spatial Attention Block (SAB) learns the spatial attention by taking information from B-Res path and decoder as inputs. Then, the outputs from the SAB and the decoder at lower depth are used to reconstruct high-resolution feature maps. B-Res paths along with these SABs ensure that the features learned by encoders are more compatible with those of the decoders, leading to better generalisation of the vessels.

Methods

In this paper, the Multi-scale multi attention (MSMA) Net architecture is proposed as a solution to tackle the inherent complexities of vascular segmentation. Built upon the U-Net architecture⁹, proposed MSMA Net architecture undergoes substantial structural modifications intended at improving algorithmic performance. The abbreviation 'MSMA' succinctly encapsulates the core attributes of our model, emphasizing its distinctive features: encoders & decoders equipped with multi-scale feature extraction capability and incorporation of squeeze and excitation & spatial attention blocks in the bottleneck residual paths in place of skip connections for emphasizing more attentive features across channel and spatial dimensions. To evaluate the effectiveness of the MSMA Net, training and testing are carried out. Firstly, Preprocessing and augmentation techniques are employed on the data to augment image diversity. Model training is then conducted to ascertain the optimum parameter weights. Lastly, using test image set, the model is tested and analysed. Figure 2 depicts the proposed system.

Pre-processing

Notwithstanding the proficiency of deep learning frameworks in yielding favourable outcomes with input images, the efficiency of segmentation tasks can be significantly augmented by employing suitable pre-processing methodologies. To explore the effects of preprocessing, we carried out experiments both without and with pre-processing to evaluate its impact. Our framework precisely integrates three pre-processing methods:

1. Morphological opened (MO) Image, 2. CLAHE (Contrast Limited Adaptive Histogram Equalization), and 3. Shade Correction.
1. Central light reflexes appear in retinal images owing to variations in light intensity. Typically, a blood vessel is characterised by a singular centerline flanked by two edges. Nonetheless, when a blood vessel exhibits a central light reflex, it appears to be divided into two distinct vessels. To address this issue, a morphological opening procedure is employed on the green channel, utilising a disk having a radius of 2 as the structural element.

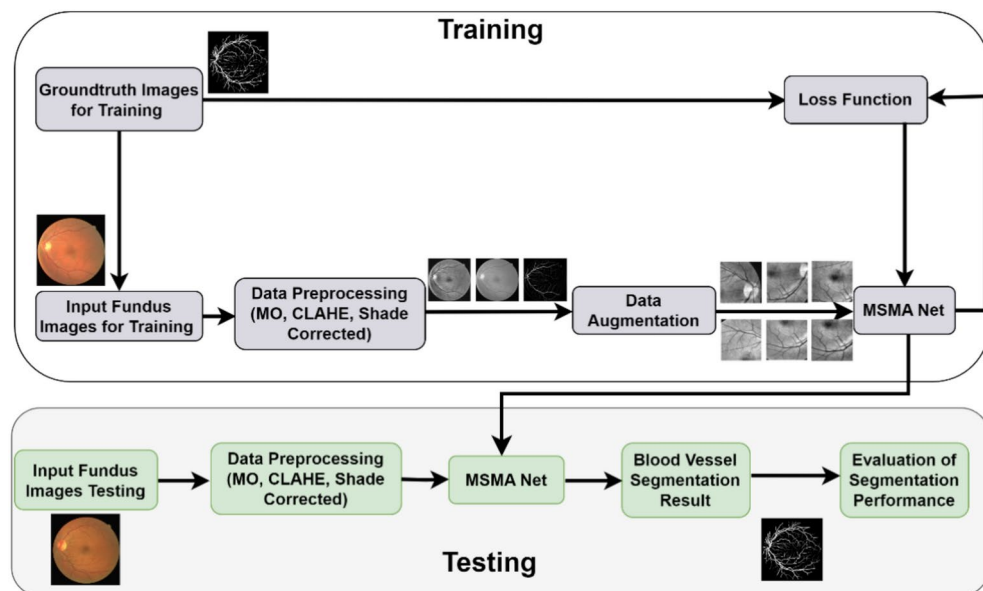


Fig. 2. Framework of the proposed system.

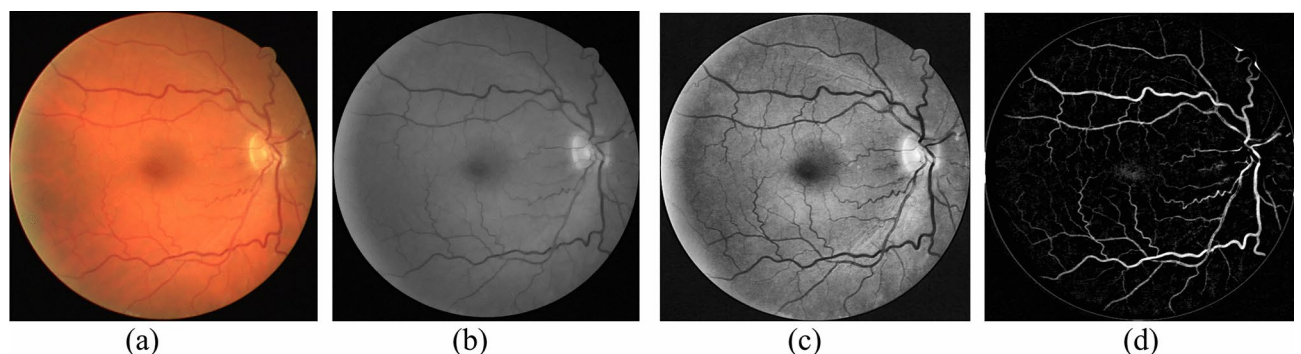


Fig. 3. (a) Retinal image. Outcomes of pre-processing methods: (b) Morphological opening (MO); (c) CLAHE; and (d) Shade-corrected.

- We utilize the CLAHE algorithm²⁷ to enhance the gray scale image, leveraging its capability to enhance the contrast between vasculature and background.
- The background image is estimated by applying a median filter to the green channel using a 25×25 kernel. Then this background image is subtracted from the green channel. The shade-corrected image is then obtained by resetting all pixels with positive values to zero. As a result, the final image may have negative values for pixels with intensities lower than the background.

Figure 3 illustrates the preprocessing strategies of the proposed method.

MSMA net model

Because of its popularity in semantic segmentation, many downstream tasks have employed the U-Net model as a baseline, requiring tailored modifications for each task. Figure 4a depicts the proposed MSMA Net model aimed at segmenting vasculature from digital fundus images. This model integrates multi-scale squeeze and excitation (MSSE) blocks in both the encoder and decoder. Attention gates and B-Res paths are integrated into the skip connections to enable feature flow between the encoders and decoders at all depths. These modifications are prompted by the challenges presented by the diverse complex patterns related to retinal vasculature.

MSSE block

Multi-scale feature extraction As Fig. 4b depicts, we derive inspiration from the objective of capturing blood vessels across various scales in the design of the proposed Multi-Scale Squeeze and Excitation (MSSE) block. Inspired by the inception module²⁸, the MSSE block improves the convolution block of the U-Net architecture by comprising a series of smaller and lightweight 3×3 convolutions. The outputs from the second and third 3×3 convolutional blocks effectively mimic the results of 5×5 and 7×7 convolutional operations, respectively²⁹.

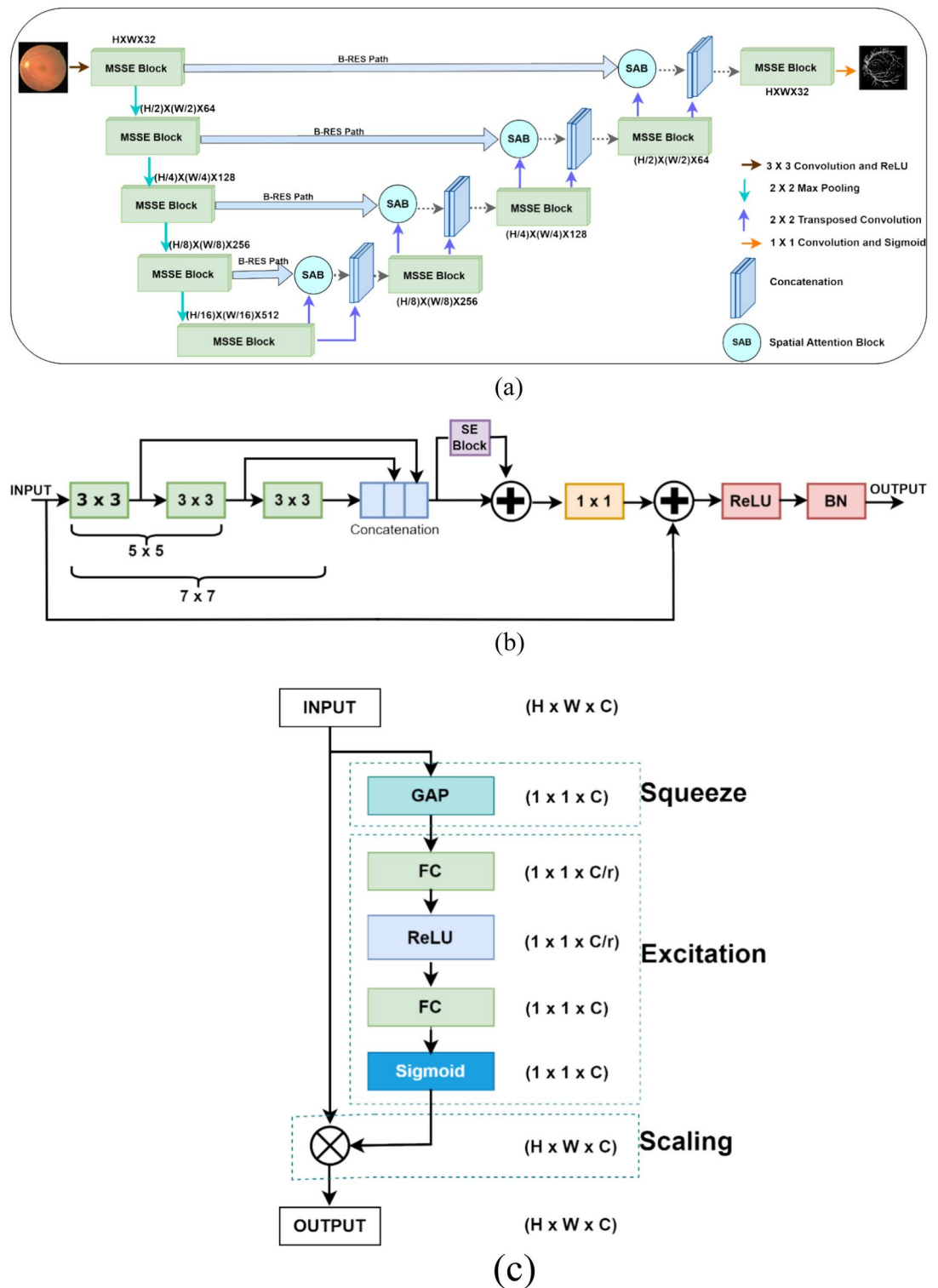
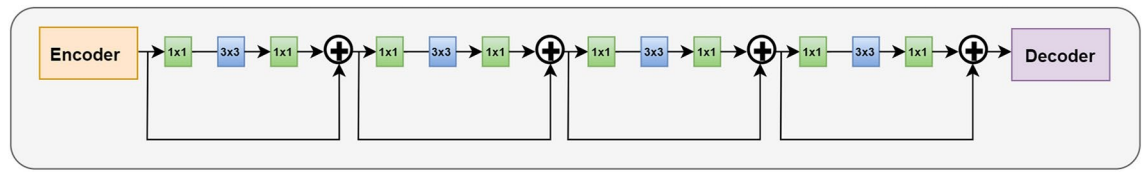
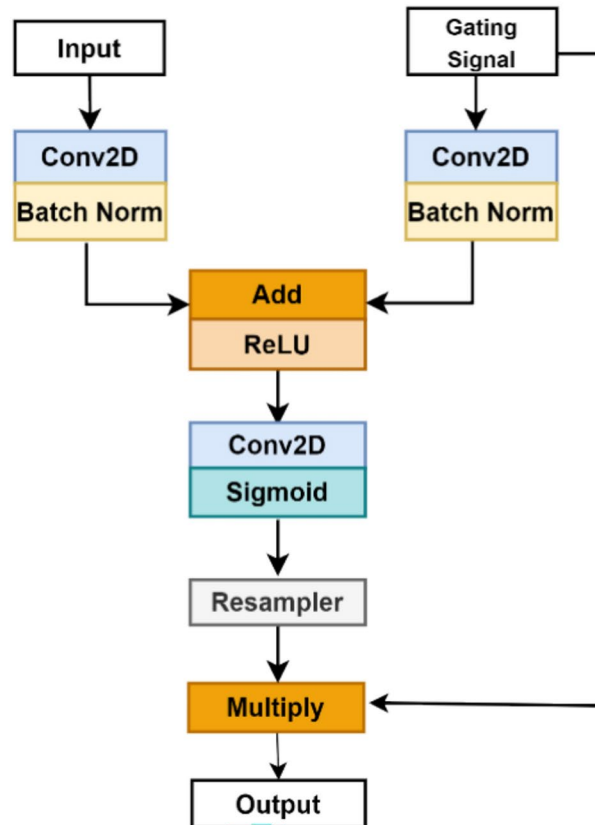


Fig. 4. (a) Proposed MSMA net model. The U-net's two-convolution layer sequence is replaced with multi-scale blocks with squeeze and excitation blocks. In addition, we utilized B-res paths and spatial attention blocks instead of skip connections. (b) Multi-Scale Squeeze and Excitation (MSSE) block. To obtain the features at various scales, a sequence of smaller and lightweight 3×3 convolutions is employed. A residual connection is used to enhance the learning process and preserve specific unchanged knowledge. To enhance the learning capacity of the model and prioritise more significant features of the object Squeeze and Excitation block is used. (c) squeeze-and-excitation block. (d) B-respath-instead of simply merging the feature maps from the encoder with those of the decoder, B-Res Path uses a more sophisticated approach. The encoder features go through a series of Bottleneck Residual layers, which help to minimize the semantic gap between the decoder features and their corresponding encoder features. Additionally, the residual connections make learning easier and are particularly useful in deep convolutional networks. (e) Spatial Attention Block.



(d)



(e)

Figure 4. (continued)

Consequently, we combine the outputs from these three convolutional blocks through concatenation to capture spatial features across different scales.

Multi-scale feature recalibration The multi-scale convolution block extracts finer details from the input, which can potentially limit the learning ability of the model. As depicted in (Fig. 4c), squeeze and excitation (SE) block is integrated to enhance the learning capacity of the model and prioritise more significant features of the object. The SE block achieves this by recalibrating the feature responses channel wise and modelling interdependencies among the channels through squeeze and excitation operations. This block serves as a computational block, tasked with learning to precisely emphasize informative features on a channel-wise basis while suppressing less useful ones.

Within each SE block, an initial squeeze operation is applied to aggregate the global spatial data of input features into a channel feature. During the squeeze operation, input features undergo a transformation wherein global average pooling is applied to the input feature map having size $W \times H \times C$ that results in $1 \times 1 \times C$ attention vector. Let input feature map $M = [M_1, M_2, M_3, \dots, M_c]$ where $M_c \in \mathbb{R}^{H \times W}$ is c^{th} channel from input feature map of size $H \times W$. The Squeeze operation is performed by channel-wise average pooling resulting in a one-dimensional vector Z , where $Z: \mathbb{R}^{H \times W \times c} \rightarrow \mathbb{R}^c$. For c^{th} -channel, the element in the vector is given by

$$Z_c = F_{Sque}(M_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W M_c(i, j) \quad (1)$$

The Z represents the aggregated input feature map M of shape $H \times W \times C$ squeezed into a $1 \times 1 \times C$ descriptors that are statistically significant in learning the relative importance of the individual channels. These descriptors depend on the vessel instance and its morphology.

The excitation operation is designed to concentrate on modelling inter-channel dependencies, with the goal of capturing the interactions and influences among different channels. It typically involves learning a set of parameters (weights) that are applied to the squeezed features to emphasize informative channels and suppress less relevant ones. A simple gating mechanism employing two fully connected layers each equipped with ReLU and sigmoid activation functions is implemented to enhance network generalisation and isolate each channel effectively.

$$E = F_{Ex}(Z, W) = \sigma(g(Z, W)) = \sigma(W_2 \delta(W_1 Z)) \quad (2)$$

where, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$, r —the dimensionality reduction ratio serves as a hyperparameter that controls the computational cost-controlling capacity of the SE block. In our implementation, we have chosen a reduction ratio of $r=8$. This selection has been determined based on its optimal segmentation performance. We obtain the final output of the SE block by rescaling M with the activations E :

$$\widetilde{M}_c = F_{Scale}(M_c, E_c) = M_c E_c \quad (3)$$

where $\widetilde{M} = [\widetilde{M}_1, \widetilde{M}_2, \widetilde{M}_3, \dots, \widetilde{M}_c]$ and $F_{Scale}(M_c, E_c)$ denotes channel-wise multiplication between $E_c \in [0,1]$ and $M_c \in \mathbb{R}^{H \times W}$.

Furthermore, drawing inspiration from the results, we incorporated a residual connection. This addition is inspired by the remarkable performance it has demonstrated in medical image segmentation³⁰. The residual structure effectively mitigates the degradation problem and enhances the model's capacity to grasp complex spatial information. Lastly, the ReLU (Rectified Linear Unit) activation function is used to activate the output of the block and is subjected to batch normalization.

Bottleneck residual block

In the U-Net's expansion phase, transposed convolution can lead to imprecise spatial information recreation. To address this, skip connections are employed to merge spatial details from the downsampling phase, introducing the challenge of carrying redundant low-level features from the initial layers. To mitigate this issue, bottleneck residual block (B-Res Path) and attention gate are integrated into the skip connections, actively suppressing activations in irrelevant regions. The bottleneck block described in Fig. 4d first reduces the dimensionality of input features using a 1×1 convolution, then a 3×3 convolution is employed to extract in depth feature maps, and the original dimensionality is restored using another 1×1 convolution. The basic idea is to keep these blocks as lean as possible to reduce parameters and improve depth. Our decision to gradually decrease the number of bottleneck residual blocks along the B-Res paths is influenced by the potential reduction in the semantic gap between encoder and decoder features as we traverse the inner shortcut paths. We utilised 4, 3, 2, and 1 bottleneck residual blocks along the four B-Res paths, incorporating 32, 64, 128, and 256 filters in the respective paths to ensure alignment with the number of feature maps in the encoder-decoder of the proposed architecture.

Spatial attention block

B-Res Path followed by spatial attention block (SAB) illustrated in Fig. 4e reduces the inflow of redundant features, optimising information flow between downsampling and transposed convolution paths for more precise spatial information reconstruction.

Attention coefficients, $\alpha_i \in [0,1]$ identify and prioritise important image regions, preferentially preserving key feature activations while discarding irrelevant responses³¹. The result of SABs is obtained by performing a multiplication operation between the input feature maps and the attention coefficients on an element-wise basis: The gating vector $g_i \in \mathbb{R}^{F_g}$ consists of contextual data for pruning lower level feature responses and is used for each pixel i to identify focus regions. We use additive attention to get the gating coefficient and is as follows:

$$q_{att}^l = \varphi^T(\sigma_1(W_x^T x_i^l + W_g^T g_i + b_g)) + b_\varphi \quad (4)$$

$$\alpha_i^l = \sigma_2(q_{att}^l(x_i^l, g_i; \mathcal{O}_{att})) \quad (5)$$

where $\sigma_2(x_{i,c}) = \frac{1}{1 + \exp(-x_{i,c})}$ is sigmoid activation function.

SAB uses a set of parameters \mathcal{O}_{att} encompassing: linear transformations $W_x \in \mathbb{R}^{F_l \times F_{int}}$, $W_g \in \mathbb{R}^{F_g \times F_{int}}$, $\varphi \in \mathbb{R}^{F_{int} \times 1}$ and bias terms $b_\varphi \in \mathbb{R}$, $b_g \in \mathbb{R}^{F_{int}}$. For the input tensors, channel-wise $1 \times 1 \times 1$ convolutions are used to compute the linear transformations.

As training advances, the network hones its ability to focus on the desired region. The differentiable nature of the SAB allows it to be trained through backpropagation, refining attention coefficients to better highlight relevant regions. Consequently, the network becomes more proficient over time in assigning increased importance to specific regions of interest, enhancing its capability to focus on and extract pertinent information from input data.

Loss function

We have selected binary cross-entropy as the loss function for blood vessel segmentation task, as it is a binary classification problem and can be expressed as

$$L_{CE}(\hat{y}, y) = - \sum_k [y_k \log \hat{y}_k + (1 - y_k) \log(1 - \hat{y}_k)] \quad (6)$$

In Eq. (6), the penalty for misclassification is high, thereby exerting a greater influence on the overall loss. Here, \hat{y}_k represents the prediction scaled within the range of [0,1], y_k signifies the binary representation of the true label.

Results

This section delves into the experimental setup and analysis of our work. It begins by discussing the relevant data. Subsequently, the methods undertaken for preparing the training data, encompassing preprocessing and random cropping are outlined. A detailed explanation of the MSMA Net model implementation is presented in the third subsection. Finally, the evaluation parameters employed to assess the performance of MSMA Net model are presented.

Evaluation datasets

The performance evaluation of the proposed MSMA Net architecture is conducted on five standard benchmark datasets: DRIVE³², STARE³³, CHASE_DB1³⁴, HRF³⁵ and DR HAGIS³⁶. The datasets feature a diverse range of image content, presenting various challenges for analysis. The DRIVE³² dataset comprises forty color retinal images captured using a cannon camera at a 45° Field of View(FOV), each possessing a size of 565 × 584 pixels. It is partitioned into a training set and a test set, each containing twenty images. In the training set, there is one expert annotation per image, while in the test set, two annotations by different observers are available, with the first observer's annotation used for evaluation.

The STARE³³ dataset includes 20 images captured using a TopCon camera at a 35-degree FOV, having a resolution of 700 × 605 pixels. Half of these images are pathological, posing challenges for retinal blood vessel segmentation. Each image in this dataset has two manual segmentations, with the annotation given by first observer considered as the ground truth. The CHASE_DB1³⁴ dataset comprises 28 color retinal images taken using a Nidek NM-200-D camera at a 30-degree FOV, with dimensions of 999 × 960 pixels. It includes two sets of ground truths annotated by different observers, with the first annotation used for assessment. HRF³⁵ dataset includes 45 high-resolution fundus images with a resolution of 3304 × 2336 acquired from healthy patients, as well as those with diabetic retinopathy and glaucoma. The dataset provides only one set of ground truth annotations for blood vessels, optic disc, and fovea. DR HAGIS³⁶ dataset comprises a total of 40 high-resolution retinal fundus images with a resolution of 4288 × 2848 for blood vessel segmentation. This dataset also provides only one set of ground truth annotations for blood vessels.

Of all the five datasets, STARE³³ has a greater number of pathological images with severe lesions that have more interference with blood vessel segmentation. DRIVE³² contains numerous thin blood vessels, frequently leading to segmentation oversight by the algorithm. CHASE_DB1 mainly comprises retinal images of children, wherein thin blood vessel walls make centre vessel reflection to form, thereby impacting the segmentation results. These distinct challenges enable a comprehensive evaluation of the MSMA Net method against existing approaches.

Data preparation

Given the five datasets discussed earlier, the proposed MSMA Net model is expected to undergo training on smaller datasets than those required for other downstream tasks involving deep learning models, posing a limitation which may lead to overfitting of proposed MSMA Net model. To overcome this limitation, we extracted patches from the high-resolution images of the five datasets for training, opting for patch-based training instead of training on entire images. This approach allows our model to gain generalizability while maintaining consistency with the original image sizes.

The fundamental concept involves augmenting the training data by deriving patches from every image in the training set, maintaining constancy with original image dimensions by establishing the dimensions of the patches extracted as 64 × 64 pixels and generating every patch by randomly choosing its centre. From the DRIVE³², we randomly selected 200,000 patches for training. For the STARE³³, 250,000 image patches are chosen, and the same number of patches are selected for CHASE_DB1³⁴, HRF³⁵ and DR HAGIS³⁶ databases. The selected patches comprehensively cover various regions of retinal images, including vessel-containing and boundary areas, providing a diverse representation of image features for training. To complement these training patches, corresponding ground truth patches are extracted. This approach ensures that our MSMA Net model is trained on a wide spectrum of image features, enhancing its ability to generalize to unseen data. Visualization of sample patches can be seen in Fig. 5, indicating their varied origins from diverse regions within the complete images.

Implementation details

The proposed MSMA Net model is trained and evaluated on NVidia GeForce RTX 3070TI GPU, an Intel Core i9-12900HQ CPU @ 3.80 GHz processor, and 32 GB of RAM. The implementation is based on the Keras framework with TensorFlow as the backend. Training utilizes the Adam optimizer with Xavier weight initialization. The Adam optimizer is utilized for training due to its efficiency and adaptability in handling large datasets and high-dimensional parameter spaces. Xavier initialization is employed to set the initial weights, ensuring that the

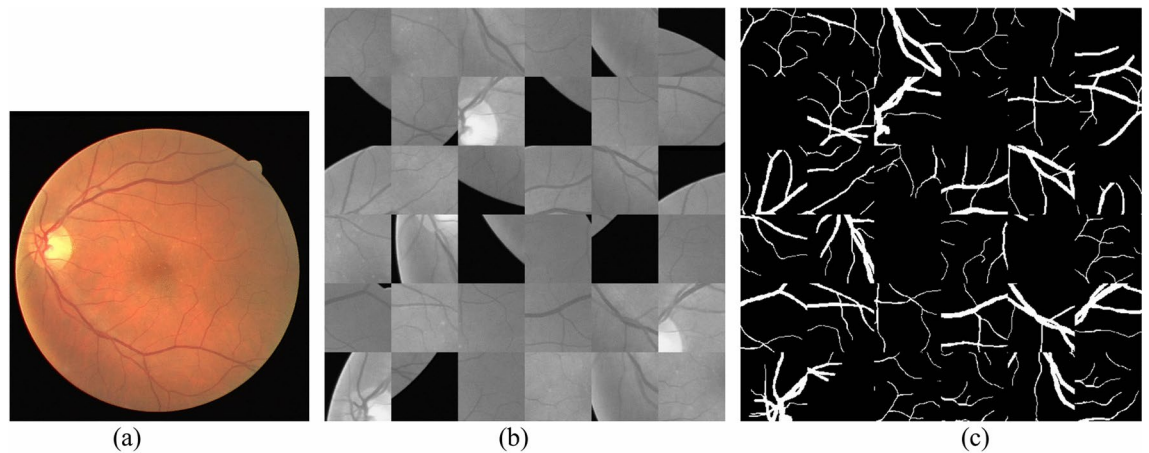


Fig. 5. Example retinal image along with some sample patches extracted from the corresponding pre-processed image. In the fundus image (a) Retinal Image in which patches are extracted with random centres. (b) Extracted Patches. (c) Corresponding groundtruths.

weights are neither too small nor too large, which helps in achieving faster convergence. Specific optimization parameters include a weight decay of 10^{-5} and a momentum of 0.9. A weight decay of 10^{-5} is applied to prevent overfitting by adding a small penalty to the loss function based on the magnitude of the weights. A momentum of 0.9 is used to accelerate convergence by considering the past gradients in the current update, thereby smoothing out the optimization process.

Various learning rates (e.g., 0.001, 0.0005, and 0.0001) were tested to observe their impact on convergence speed and stability. The chosen initial learning rate of 0.0001 was found to provide a good balance between convergence speed and stability. The dynamic adjustment strategy ensured that the model could fine-tune its parameters as training progressed. Training iterations were monitored using validation accuracy. Early stopping was employed to halt training when the validation accuracy plateaued, preventing overfitting and reducing unnecessary computational load. Various batch sizes (e.g., 64, 128 and 256) were experimented with to determine the impact on training stability and convergence speed. A mini-batch size of 256 was selected as it provided a good balance between computational efficiency and the benefits of batch normalization, leading to faster and more stable convergence.

As five different datasets are being considered in this work, the train-test splitting is carried out with little extra care to improve the performance of MSMA Net model. As per the guidelines provided by the data provider, 20 images from the DRIVE dataset are designated for training, while the remaining 20 are allocated for testing purposes. In the case of the STARE database, its twenty images are evenly divided between training and testing sets. As for the CHASE_DB1 database, twenty images are earmarked for training, leaving eight for testing purposes. For HRF database, thirty images are earmarked for training, leaving fifteen for testing purposes. In the case of the DR HAGIS database, its forty images are evenly divided between training and testing sets.

To perform a thorough evaluation, various metrics are utilized, encompassing accuracy (Acc), sensitivity (Se), specificity (Sp), Dice score (Dice), F1-score (F1), and the area under the ROC curve (AUC). These are defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Se = \frac{TP}{FN + TP} \quad (8)$$

$$Sp = \frac{TN}{FP + TN} \quad (9)$$

$$Precision = \frac{TP}{FP + TP} \quad (10)$$

$$Recall = \frac{TP}{FN + TP} \quad (11)$$

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (12)$$

The evaluation relies on classifying every pixel as true positive (TP), false positive (FP), true negative (TN), and false negative (FN). TP and TN represent the number of pixels correctly classified as vessels and background, respectively, while FP and FN denote the number of pixels misclassified as vessels and background, respectively. For binary segmentation, Dice score is equal to the F1-score^{37,38}. Importantly, during the calculation of these

parameters, only the pixels inside the FOV of the five datasets are considered. This emphasis on FOV guarantees that the assessment is performed solely on the pertinent regions of interest within the fundus images.

Experiments

The experiments are organized into eight sections, with each section thoroughly elaborated with accompanying details.

Experiment 1 - ablation study - analysis of different pre-processing combinations

The analysis of different preprocessing combinations and their influence on segmentation outcomes is examined, and Table 2 provides a summary of test results for every combination. Performance evaluation metrics are calculated based on the annotations provided by the first observer, serving as the groundtruth. A significant enhancement in experimental outcomes is observed when comparing No. 0 with No. 1, indicating the superiority of employing MO images over RGB images. MO helps in removing small noise elements from the images, which ensures that the training data is cleaner. This Improves overall vessel detection by enhancing vessel boundaries, leading to better AUC and dice scores. Additionally, the combination of CLAHE and shade correction have shown positive results. It's crucial to highlight that model performance diminishes when both MO and shade correction are used as evidenced by No. 6. CLAHE increases local contrast and enhances vessel visibility, leading to higher sensitivity and specificity. It improves detection accuracy by making vessels more distinguishable from the background. Shade correction addresses uneven illumination across the image, ensuring that the entire vessel structure is uniformly visible. This helps in creating a more consistent dataset where the model can learn to detect vessels under different lighting conditions. Hence, experiments No. 5 and No. 7 demonstrate the synergy between CLAHE and shade correction.

The comprehensive preprocessing pipeline (MO + CLAHE + Shade Correction) leads to the highest AUC, Acc, Dice, Se, and Sp values, showcasing its effectiveness in improving the segmentation of retinal vessels. Consequently, all the three preprocessing strategies are collectively utilized in subsequent experiments.

Experiment 2 - outcomes of the proposed MSMA net method

For a more intuitive assessment of the proposed MSMA Net model, Fig. 6 presents six examples of segmentation across the DRIVE³², STARE³³, CHAS_DB1³⁴, HRF³⁵ and DR HAGIS³⁶ datasets. The visualization results demonstrate that our segmentation.

methodology excels across various types of fundus images. Leveraging probability maps of blood vessels, the efficacy of the proposed segmentation method in extracting multiscale vessel structures is corroborated. Table 3 offers a comprehensive summary of average performance metrics across all images in the test set within the DRIVE³², STARE³³, CHAS_DB1³⁴, HRF³⁵ and DR HAGIS³⁶ datasets. As delineated in Table 3, the average metrics of our.

algorithm steadily outperforms those of the second human expert. This highlights our method's ability to maintain a minimal false positive (FP) rate and effectively distinguish between vessel and non-vessel pixels when compared directly with second human expert.

Experiment 3 - ablation study-each block in the network

An ablation study is conducted, wherein each network block is tested individually to evaluate the effect of incorporating each block in this study. In this context, a four-layer U-Net model serves as the baseline network. By integrating various network blocks, such as the multiscale residual block, SE block, and spatial attention block, different network configurations of the proposed MSMA Net are built for the purpose of ablation analysis.

Table 4 presents the specific experimental outcomes of the ablation study. When examining the ablation results, it becomes evident that the inclusion of these network blocks has enhanced the segmentation performance of the proposed MSMA Net on the five public datasets in comparison to the baseline network. For the DRIVE dataset, the proposed MSMA Net model achieved scores of 0.9883 (98.83%), 0.8821, 0.9890, and 0.8692 on Acc, Se, Sp, and dice, respectively. The most notable enhancement is observed in the Se metric, which improved from

| Database | Number (No.) | MO | CLAHE | Shade corrected | AUC | Acc | Dice | Se | Sp |
|-------------------------|--------------|-------------|-------|-----------------|--------|--------|--------|--------|--------|
| DRIVE ³² | 0 | Color Image | | | 0.9849 | 0.9690 | 0.8116 | 0.7673 | 0.9882 |
| | 1 | ✓ | | | 0.9871 | 0.9692 | 0.8269 | 0.8473 | 0.9807 |
| | 2 | | ✓ | | 0.9872 | 0.9709 | 0.8304 | 0.8199 | 0.9852 |
| | 3 | | | ✓ | 0.9883 | 0.9722 | 0.8333 | 0.7994 | 0.9860 |
| | 4 | ✓ | ✓ | | 0.9893 | 0.9827 | 0.8537 | 0.8792 | 0.9886 |
| | 5 | | ✓ | ✓ | 0.9865 | 0.9727 | 0.8455 | 0.8712 | 0.9871 |
| | 6 | ✓ | | ✓ | 0.9889 | 0.9739 | 0.8488 | 0.8681 | 0.9879 |
| | 7 | ✓ | ✓ | ✓ | 0.9898 | 0.9883 | 0.8692 | 0.8821 | 0.9890 |
| STARE ³³ | 8 | ✓ | ✓ | ✓ | 0.9961 | 0.9825 | 0.8837 | 0.8612 | 0.9883 |
| CHASE_DB1 ³⁴ | 9 | ✓ | ✓ | ✓ | 0.9879 | 0.9854 | 0.8699 | 0.8638 | 0.9889 |
| HRF ³⁵ | 10 | ✓ | ✓ | ✓ | 0.9872 | 0.9817 | 0.8538 | 0.8553 | 0.9895 |
| DR HAGIS ³⁶ | 11 | ✓ | ✓ | ✓ | 0.9870 | 0.9769 | 0.8378 | 0.8615 | 0.9886 |

Table 2. Test results with different image pre-processing methods.

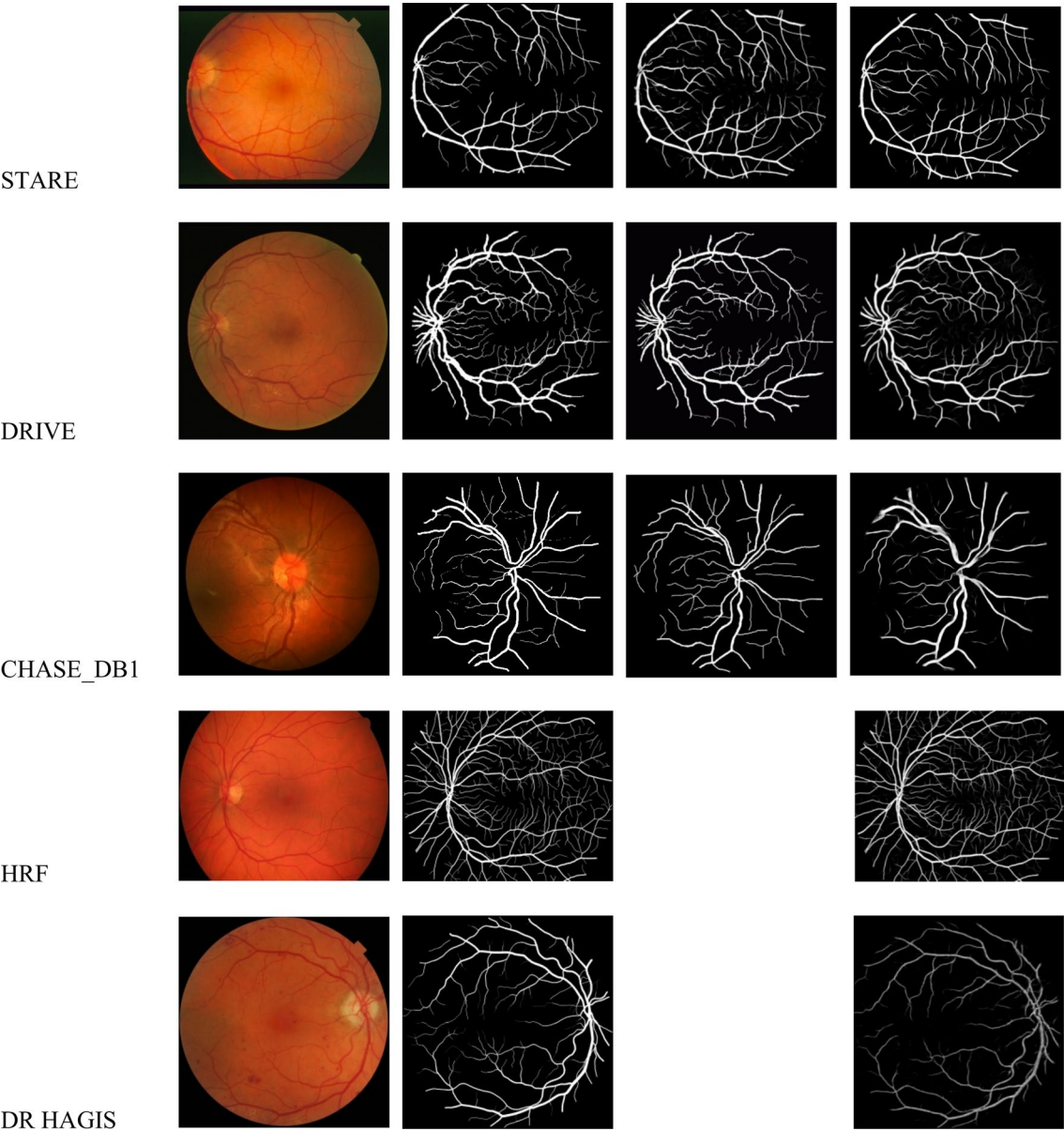


Fig. 6. Retinal images are selected randomly with corresponding manual ground truths and Segmentation results of MSMA Net method (Column-wise) 1. Original retinal images, 2. First manual ground-truths. 3. Second manual ground-truths (No 2nd manual groundtruths for HRF and DR HAGIS datasets). 4. Segmentation outcomes of the proposed MSMA Net method.

| Dataset | Observer/method | Acc | AUC | Dice | Se | Sp |
|-------------------------|--------------------------|--------|--------|--------|--------|--------|
| DRIVE ³² | Second human observer | 0.9527 | – | – | 0.7760 | 0.9816 |
| | Proposed MSMA Net method | 0.9883 | 0.9898 | 0.8692 | 0.8821 | 0.9890 |
| STARE ³³ | Second human observer | 0.9349 | – | – | 0.8952 | 0.9384 |
| | Proposed MSMA Net method | 0.9873 | 0.9961 | 0.8847 | 0.8691 | 0.9922 |
| CHASE_DB1 ³⁴ | Second human observer | 0.9545 | – | – | 0.8105 | 0.9711 |
| | Proposed MSMA Net method | 0.9854 | 0.9879 | 0.8799 | 0.8638 | 0.9901 |
| HRF ³⁵ | Proposed MSMA Net method | 0.9872 | 0.9817 | 0.8538 | 0.8553 | 0.9895 |
| DR HAGIS ³⁶ | Proposed MSMA Net method | 0.9870 | 0.9769 | 0.8378 | 0.8615 | 0.9886 |

Table 3. Average evaluation metrics for DRIVE³², STARE³³, CHASE_DB1³⁴, HRF³⁵ and DR HAGIS³⁶.

| Database | Basic U-net | MS network | SE block | SAB | Dice | Se | Sp | Acc | AUC |
|-----------|-------------|------------|----------|-----|--------|--------|--------|--------|--------|
| DRIVE | ✓ | | | | 0.8175 | 0.7653 | 0.9811 | 0.9536 | 0.9736 |
| | | ✓ | | | 0.8354 | 0.8738 | 0.9759 | 0.9814 | 0.9873 |
| | | ✓ | ✓ | | 0.8532 | 0.8788 | 0.9837 | 0.9860 | 0.9885 |
| | | ✓ | | ✓ | 0.8577 | 0.8794 | 0.9876 | 0.9869 | 0.9890 |
| | | ✓ | ✓ | ✓ | 0.8692 | 0.8821 | 0.9890 | 0.9883 | 0.9898 |
| STARE | ✓ | | | | 0.8133 | 0.7639 | 0.9796 | 0.9588 | 0.9759 |
| | | ✓ | | | 0.8742 | 0.8580 | 0.9801 | 0.9807 | 0.9943 |
| | | ✓ | ✓ | | 0.8768 | 0.8588 | 0.9879 | 0.9808 | 0.9950 |
| | | ✓ | | ✓ | 0.8790 | 0.8603 | 0.9887 | 0.9816 | 0.9953 |
| | | ✓ | ✓ | ✓ | 0.8837 | 0.8612 | 0.9883 | 0.9825 | 0.9961 |
| CHASE_DB1 | ✓ | | | | 0.7792 | 0.7870 | 0.9777 | 0.9604 | 0.9779 |
| | | ✓ | | | 0.8583 | 0.8199 | 0.9882 | 0.9771 | 0.9849 |
| | | ✓ | ✓ | | 0.8610 | 0.8473 | 0.9855 | 0.9788 | 0.9871 |
| | | ✓ | | ✓ | 0.8638 | 0.8551 | 0.9894 | 0.9825 | 0.9870 |
| | | ✓ | ✓ | ✓ | 0.8699 | 0.8638 | 0.9889 | 0.9854 | 0.9879 |
| HRF | ✓ | | | | 0.7825 | 0.7499 | 0.9790 | 0.9673 | 0.9524 |
| | | ✓ | | | 0.8010 | 0.7884 | 0.9805 | 0.9699 | 0.9716 |
| | | ✓ | ✓ | | 0.8138 | 0.8056 | 0.9831 | 0.9704 | 0.9765 |
| | | ✓ | | ✓ | 0.8291 | 0.8253 | 0.9879 | 0.9718 | 0.9849 |
| | | ✓ | ✓ | ✓ | 0.8538 | 0.8553 | 0.9895 | 0.9817 | 0.9872 |
| DR HAGIS | ✓ | | | | 0.7849 | 0.7630 | 0.9741 | 0.9465 | 0.9581 |
| | | ✓ | | | 0.8047 | 0.8500 | 0.9799 | 0.9599 | 0.9737 |
| | | ✓ | ✓ | | 0.8197 | 0.8511 | 0.9796 | 0.9614 | 0.9809 |
| | | ✓ | | ✓ | 0.8206 | 0.8552 | 0.9879 | 0.9622 | 0.9827 |
| | | ✓ | ✓ | ✓ | 0.8378 | 0.8615 | 0.9886 | 0.9769 | 0.9870 |

Table 4. Experimental outcomes of the ablation study.

0.7653 to 0.8788, primarily attributable to the integration of the SE block. With the addition of both SE and SA blocks, Se experienced an increase from 0.7653 to 0.8821. Concurrently, Acc and dice also saw improvements, reaching 0.9854 and 0.8799 respectively.

The blocks introduced have demonstrated substantial enhancements in performance on the STARE dataset. Although Sp does not show improvement, key metrics such as Acc, Se, and the dice score exhibit significant enhancements. In particular, sensitivity showed a substantial increase, reaching 0.8691. Similarly, for the CHASE_DB1 dataset, the proposed network exhibits a comparable trend. Apart from Specificity, the introduction of the proposed network blocks leads to improvements in other evaluation metrics. Se has improved from 0.7876 to 0.8638, indicating a heightened ability for detailed segmentation. Taken together, these ablation study results convincingly demonstrate the efficacy of the proposed blocks in enhancing the outcomes of the MSMA Net model. Moreover, the proposed approach yields AUC values exceeding 0.97, specifically 0.9898 for the DRIVE³² dataset, 0.9961 for the STARE³³ dataset, and 0.9879 for the CHASE_DB1³⁴ dataset. These high AUC values demonstrate the MSMA Net method's strong generalisation capacity and demonstrate how well it segments vessels in all the datasets.

The ablation study clearly shows the incremental improvements each module provides. Even without the SE and Spatial Attention blocks, MSMA Net improves upon the Basic U-Net by incorporating multi-scale kernels in the convolutional layers. This allows the model to capture features at different scales, enhancing its ability to detect vessels of varying thicknesses and structures. Improved detection of small and thin vessels increases the true positive rate, reducing false positives thus enhancing dice score, se and sp. Removing only the Spatial Attention block shows that SE blocks contribute more to the model's performance in terms of sensitivity and accuracy, suggesting that SE blocks are crucial for enhancing channel-wise feature calibration and robustness. Enhanced feature recalibration allows the model to focus on crucial vessel features, improving detection rates. The Spatial Attention Block further refines the feature maps by focusing on spatially significant regions in the image, which helps in distinguishing vessels from similar-looking structures. This results in better segmentation accuracy and an improved AUC by enhancing the model's ability to correctly identify true vessel regions. The full MSMA Net model achieves the best performance across all metrics, demonstrating the synergistic effect of combining SE and Spatial Attention blocks. This combination ensures robust feature calibration and spatial relevance, enhancing overall segmentation performance.

Experiment 4 - evaluation against existing methods

To provide an unbiased evaluation of the proposed MSMA Net model's effectiveness, our approach is compared to a wide range of current methodologies using the DRIVE³², STARE³³, CHAS_DB1³⁴, HRF³⁵ and DR HAGIS³⁶ datasets. The assessment parameters are averaged to determine the quantitative results for each approach, with

the highest value for each parameter indicated in bold. Based on the quantitative findings, it is evident that deep learning-based approaches outperform conventional methods.

For the DRIVE³², STARE³³, CHAS_DB1³⁴, HRF³⁵ and DR HAGIS[36]dataset, the AUCs of the MSMA Net method are 0.9898, 0.9961, 0.9879, 0.9872 and 0.9870 respectively. These findings indicate that, the MSMA Net model outperforms all other approaches. However, when it comes to the CHASE_DB1³⁴ database, our MSMA net approach shows a slightly lower AUC compared to MRA Net. This discrepancy arises from the potent detection ability of the proposed network for minute vascular structures, resulting in the retention of more suspected vascular elements.

Our approach outperforms other methods with higher accuracy, sensitivity, and dice scores on five datasets. Accuracy measures the similarity between predicted and actual labels. Our MSMA Net model differentiates vascular from background regions, with detailed contour information for vascular endings. Our model's high segmentation accuracy and strong generalization capabilities are demonstrated across different datasets.

The sensitivity indicates the model's proficiency in classifying vessel pixels. Since there are significantly fewer vessel pixels compared to background pixels in retinal images, achieving high sensitivity is crucial. The MSMA Net approach has the highest sensitivity, demonstrating its ability to precisely classify a greater number of vessel pixels. As vascular pixels constitute only around 10% of all pixels in the image, it can be challenging to segment vessels using a classifier. Therefore, a high sensitivity is essential, particularly in a CAD system, where accurate detection of blood vessels is crucial without introducing false positives.

Out of all the approaches, the MSMA Net model had the utmost dice score, demonstrating its superior capability to accurately segment background and vascular structures. Our approach establishes an exceptional balance amid precision and recall, as evidenced by the dice score, which effectively balances these two aspects. Tables 5, 6, and 7 validate that our model has a lower Sp than some of the approaches. This implies that our methodology minimises the possibility of misidentifying vascular pixels as background while having a greater detection rate for vascular pixels.

Experiment 5 - impact of cross-testing

It is not practical to retrain the model for every new patient's retinal image analysis in clinical practice. Moreover, medical apparatus utilised in different healthcare facilities is typically produced by diverse manufacturers. Therefore, a reliable technique should include the capability to efficiently analyse images acquired from various equipment sources. It is essential to assess the suitability of the approach for use with any retinal image in real-world situations. Therefore, the model's practical applicability relies depends on critical evaluation criteria such as robustness and generalisation.

Analysis of Table 8 unequivocally proves that the MSMA Net model surpasses other approaches in all cross-testing outcomes. When the MSMA Net is trained using the DRIVE³² dataset and then tested on the STARE³³ dataset, it achieves the maximum AUC and accuracy. However, its specificity is slightly inferior compared to³⁹. The results obtained from cross-testing highlight the robustness and exceptional ability of the proposed MSMA Net framework to generalize and remain effective when presented with unfamiliar data.

| S.No | Methods | Year | Dice | Se | Sp | Acc | AUC |
|------|---------------------------|------|--------|--------|--------|--------|--------|
| 1 | HED ⁴⁶ | 2016 | 0.8089 | 0.7627 | 0.9801 | 0.9524 | 0.9758 |
| 2 | DeepVessel ³⁹ | 2016 | – | 0.7612 | 0.9768 | 0.9523 | 0.9752 |
| 3 | JL-UNet ⁴³ | 2018 | 0.8102 | 0.7653 | 0.9818 | 0.9542 | 0.9752 |
| 4 | CC-Net ⁴¹ | 2018 | N.A | 0.7625 | 0.9809 | 0.9528 | 0.9678 |
| 5 | AttUNet ³¹ | 2018 | 0.8232 | 0.7946 | 0.9789 | 0.9564 | 0.9799 |
| 6 | DenseUNet ⁴² | 2018 | 0.8279 | 0.7985 | 0.9805 | 0.9573 | 0.9810 |
| 7 | Yan et al. ⁴³ | 2019 | N.A | 0.7631 | 0.9820 | 0.9533 | 0.9750 |
| 8 | BTS-DSN ⁴⁴ | 2019 | 0.8208 | 0.7800 | 0.9806 | 0.9551 | 0.9796 |
| 9 | DUNet ¹¹ | 2019 | 0.8249 | 0.7984 | 0.9803 | 0.9575 | 0.9811 |
| 10 | CTF-Net ⁴⁷ | 2020 | 0.8241 | 0.7849 | 0.9813 | 0.9567 | 0.9788 |
| 11 | CSU-Net ⁴⁸ | 2021 | 0.8251 | 0.8071 | 0.9782 | 0.9565 | 0.9801 |
| 12 | MC-DMD ⁴⁵ | 2022 | NA | 0.721 | 0.983 | 0.954 | 0.952 |
| 13 | WA Net ⁴⁹ | 2022 | 0.8269 | 0.7962 | 0.9875 | 0.9708 | 0.9849 |
| 14 | Bridge Net ⁵⁰ | 2022 | 0.8283 | 0.7853 | 0.9818 | 0.9565 | 0.9834 |
| 15 | OCE-Net ¹⁵ | 2023 | 0.8302 | 0.8018 | 0.9826 | 0.9581 | 0.9821 |
| 16 | MRA Net ⁵¹ | 2023 | NA | 0.8488 | 0.9907 | 0.9698 | 0.9899 |
| 17 | DA-Res2UNet ⁵² | 2023 | 0.8277 | 81.50 | 98.56 | 97.04 | 98.77 |
| 18 | ResDO-UNet ²³ | 2023 | 0.8229 | 0.7985 | 0.9791 | 0.9561 | NA |
| 19 | RCAR-UNet ⁵³ | 2024 | 0.8047 | 0.7487 | 0.9836 | 0.9537 | 0.9759 |
| 20 | MSR U-Net ⁵⁴ | 2024 | 0.8344 | 0.8717 | 0.9886 | 0.9722 | 0.9883 |
| 21 | Proposed MSMA Net | 2024 | 0.8692 | 0.8821 | 0.9890 | 0.9883 | 0.9898 |

Table 5. Performance evaluation of the MSMA Net method on the DRIVE³² dataset.

| S.No | Methods | Year | Dice | Se | Sp | Acc | AUC |
|------|-----------------------------|------|--------|--------|--------|--------|--------|
| 1 | HED ⁴⁶ | 2016 | 0.8268 | 0.8076 | 0.9822 | 0.9641 | 0.9824 |
| 2 | Orlando etal. ⁵⁵ | 2017 | N.A | 0.7680 | 0.9738 | 0.9519 | |
| 3 | JL-UNet ²⁶ | 2018 | N.A | 0.7581 | 0.9846 | 0.9612 | 0.9801 |
| 4 | AttUNet ³¹ | 2018 | 0.8136 | 0.8067 | 0.9816 | 0.9632 | 0.9833 |
| 5 | CC-Net ⁴¹ | 2018 | N.A | 0.7709 | 0.9848 | 0.9633 | 0.9700 |
| 6 | DenseUNet ⁴² | 2018 | 0.8232 | 0.7859 | 0.9842 | 0.9644 | 0.9847 |
| 7 | Yan etal. ⁴³ | 2019 | N.A | 0.7735 | 0.9857 | 0.9638 | 0.9833 |
| 8 | BTS-DSN ⁴⁴ | 2019 | 0.8362 | 0.8201 | 0.9828 | 0.9660 | 0.9872 |
| 9 | DUNet ¹¹ | 2019 | 0.8230 | 0.7892 | 0.9816 | 0.9634 | 0.9843 |
| 10 | MC-DMD ⁴⁵ | 2022 | NA | 0.734 | 0.976 | 0.961 | 0.964 |
| 11 | WA Net ⁴⁹ | 2022 | 0.8211 | 0.7834 | 0.9908 | 0.9752 | 0.9906 |
| 12 | Bridge Net ⁴⁹ | 2022 | 0.8289 | 0.8002 | 0.9864 | 0.9668 | 0.9901 |
| 13 | OCE-Net ¹⁵ | 2023 | 0.8341 | 0.8012 | 0.9865 | 0.9672 | 0.9876 |
| 14 | DA-Res2UNet ⁵² | 2023 | 83.96 | 82.69 | 98.85 | 97.65 | 98.83 |
| 15 | ResDO-UNet ²³ | 2023 | 0.8172 | 0.7963 | 0.9792 | 0.9567 | NA |
| 16 | RCAR-UNet ⁵³ | 2024 | 0.7850 | 0.6979 | 0.9905 | 0.9594 | 0.9801 |
| 17 | MSR U-Net ⁵⁴ | 2024 | 0.8748 | 0.8585 | 0.9914 | 0.9813 | 0.9952 |
| 18 | Proposed MSMA Net | 2024 | 0.8837 | 0.8612 | 0.9883 | 0.9825 | 0.9961 |

Table 6. Performance evaluation of the MSMA Net method on the STARE³³ dataset.

| S.No | Methods | Year | Dice | Se | Sp | Acc | AUC |
|------|---------------------------|------|--------|--------|--------|--------|--------|
| 1 | HED ⁴⁶ | 2016 | 0.7815 | 0.7516 | 0.9805 | 0.9597 | 0.9796 |
| 2 | JL-UNet ²⁶ | 2018 | N.A | 0.7633 | 0.9809 | 0.9610 | 0.9781 |
| 3 | Att UNet ³¹ | 2018 | 0.8012 | 0.8010 | 0.9804 | 0.9642 | 0.9840 |
| 4 | Dense UNet ⁴² | 2018 | 0.7901 | 0.7893 | 0.9792 | 0.9611 | 0.9835 |
| 5 | Yan et al. ⁴³ | 2019 | N.A | 0.7641 | 0.9806 | 0.9607 | 0.9776 |
| 6 | BTS-DSN ⁴⁴ | 2019 | 0.7983 | 0.7888 | 0.9801 | 0.9627 | 0.9840 |
| 7 | DUNet ¹¹ | 2019 | 0.7932 | 0.7735 | 0.9801 | 0.9618 | 0.9839 |
| 8 | WA Net ⁴⁹ | 2022 | 0.8069 | 0.8041 | 0.9873 | 0.9758 | 0.9890 |
| 9 | Bridge Net ⁵⁰ | 2022 | 0.8263 | 0.8132 | 0.9840 | 0.9667 | 0.9893 |
| 10 | OCE-Net ¹⁵ | 2023 | 0.8196 | 0.8138 | 0.9824 | 0.9678 | 0.9872 |
| 11 | MRA Net ⁵¹ | 2023 | NA | 0.8533 | 0.9856 | 0.9755 | 0.9893 |
| 12 | DA-Res2UNet ⁵² | 2023 | 0.8188 | 0.8318 | 0.9867 | 0.9770 | 0.9912 |
| 13 | ResDO-UNet ²³ | 2023 | 0.8236 | 0.8020 | 0.9794 | 0.9672 | NA |
| 14 | RCAR-UNet ⁵³ | 2024 | 0.7470 | 0.7475 | 0.9798 | 0.9566 | 0.9695 |
| 15 | MSR U-Net ⁵⁴ | 2024 | 0.8683 | 0.8551 | 0.9891 | 0.9781 | 0.9835 |
| 16 | Proposed MSMA Net | 2024 | 0.8699 | 0.8638 | 0.9889 | 0.9854 | 0.9879 |

Table 7. Performance evaluation of the MSMA Net method on the CHASE_DB1³⁴ dataset.

Experiment 6 - performance on pathological images

In Fig. 7.a, images depicting diabetic retinopathy are displayed, featuring both bright and dark lesions. Comparative analysis reveals that, in contrast to true vessels, the probability values generated by the MSMA Net model for both dark and bright lesions are notably lower. Consequently, the segmentation outcomes are anticipated to be less affected by the presence of such lesions. We compare results of our method with results of⁷ and⁴⁰. Our MSMA Net model demonstrates superior vascular segmentation compared to methods of⁷ and⁴⁰. The performance assessment highlights a substantial enhancement in vascular segmentation outcomes with MSMA Net model. These findings validate the resilience of our approach in effectively managing pathological fundus images.

Our proposed MSMA Net method, though successfully delineated the vessels that are surrounded by complex contextual information caused by several pathologies, certainly had limitations in handling the task when the vessels are being overlapped with hemorrhages. The results of the proposed method on two such images are shown in the figure. The proposed method had difficulty in detecting the continuity of blood vessels in the locations of vessel and hemorrhage overlap resulting in broken segmented vessels as shown in (Fig. 7b). Moreover, on rare occasions model also suffered predicting the false positives (both are enclosed in yellow boxes).

| S.No | Train Set | Test Set | Methods | Se | Sp | AUC | Acc | Dice |
|------|-------------------------|-------------------------|-------------------------|--------|--------|--------|--------|--------|
| 1 | STARE ³³ | DRIVE ³² | Li ⁵¹ | 0.7273 | 0.9810 | 0.9677 | 0.9486 | – |
| 2 | | | Fraz ³⁴ | 0.7242 | 0.9792 | 0.9697 | 0.9456 | – |
| 3 | | | MSR U-Net ⁵⁴ | 0.7904 | 0.9822 | 0.9827 | 0.9684 | 0.7944 |
| 4 | | | Mo ³⁹ | 0.7412 | 0.9799 | 0.9653 | 0.9492 | – |
| 5 | | | MSMA Net | 0.8009 | 0.9868 | 0.9830 | 0.9685 | 0.8196 |
| 1 | | CHASE_DB1 ³⁴ | Li ⁵¹ | 0.7103 | 0.9665 | 0.9565 | 0.9415 | – |
| 2 | | | Mo ³⁹ | 0.7032 | 0.9794 | 0.9690 | 0.9515 | – |
| 3 | | | MSR U-Net ⁵⁴ | 0.7668 | 0.9805 | 0.9758 | 0.9657 | 0.7597 |
| 4 | | | MSMA Net | 0.7972 | 0.9927 | 0.9804 | 0.9657 | 0.8081 |
| 1 | | HRF ³⁵ | MSMA Net | 0.7360 | 0.9810 | 0.9660 | 96.10 | 0.7629 |
| 1 | | DR HAGIS ³⁶ | MSMA Net | 0.7484 | 0.9780 | 97.11 | 95.42 | 0.7748 |
| 1 | DRIV ³² | STARE ³³ | Li ⁵¹ | 0.7027 | 0.9820 | 0.9671 | 0.9545 | – |
| 2 | | | Fraz ³⁴ | 0.7010 | 0.9770 | 0.9660 | 0.9493 | – |
| 3 | | | MSR U-Net ⁵⁷ | 0.7883 | 0.9822 | 0.9807 | 0.9686 | – |
| 4 | | | Mo ³⁹ | 0.7009 | 0.9843 | 0.9751 | 0.9570 | – |
| 5 | | | MSMA Net | 0.8048 | 0.9889 | 0.9835 | 0.9689 | 0.8256 |
| 1 | | CHASE_DB1 ³⁴ | Mo ³⁹ | 0.7003 | 0.9750 | 0.9671 | 0.9478 | – |
| 2 | | | MSR U-Net ⁵⁴ | 0.7492 | 0.9786 | 0.9709 | 0.9633 | – |
| 3 | | | Li ⁵¹ | 0.7118 | 0.9791 | 0.9628 | 0.9429 | – |
| 4 | | | MSMA Net | 0.7764 | 0.9829 | 0.9785 | 0.9650 | 0.7906 |
| 1 | | HRF ³⁵ | MSMA Net | 0.7540 | 0.9750 | 0.9530 | 0.9580 | 0.7758 |
| 1 | | DR HAGIS ³⁶ | MSMA Net | 0.7237 | 0.9703 | 0.9592 | 0.9459 | 0.7554 |
| 1 | CHASE_DB1 ³⁴ | DRIVE ³² | MSR U-Net ⁵⁴ | 0.7316 | 0.9812 | 0.9701 | 0.9598 | 0.7308 |
| 2 | | | Mo ³⁹ | 0.7003 | 0.9750 | 0.9671 | 0.9478 | – |
| 3 | | | Li ⁵¹ | 0.7118 | 0.9791 | 0.9628 | 0.9429 | – |
| 4 | | | MSMA Net | 0.7564 | 0.9829 | 0.9785 | 0.9620 | 0.7966 |
| 1 | | STARE ³³ | Mo ³⁹ | 0.7032 | 0.9794 | 0.9690 | 0.9515 | – |
| 2 | | | MSR U-Net ⁵⁴ | 0.7583 | 0.9857 | 0.9799 | 0.9654 | 0.7635 |
| 3 | | | Fraz ³⁴ | 0.7103 | 0.9665 | 0.9565 | 0.9415 | – |
| 4 | | | Li ⁵¹ | 0.7240 | 0.9768 | 0.9553 | 0.9417 | – |
| 5 | | | MSMA Net | 0.7664 | 0.9899 | 0.9815 | 0.9668 | 0.7995 |
| 1 | | HRF ³⁵ | MSMA Net | 0.7735 | 0.9857 | 0.9833 | 0.9638 | 0.7842 |
| 1 | | DR HAGIS ³⁶ | MSMA Net | 0.7890 | 0.9720 | 0.9610 | 0.9560 | 0.8011 |
| 1 | HRF ³⁵ | DRIVE ³² | MSMA Net | 0.7735 | 0.9857 | 0.9833 | 0.9638 | |
| 2 | | STARE ³³ | MSMA Net | 0.8068 | 0.9838 | 0.9812 | 0.9656 | 0.8295 |
| 3 | | CHASE_DB1 ³⁴ | MSMA Net | 0.7476 | 0.9946 | 0.9807 | 0.9611 | 0.7616 |
| 4 | | DR HAGIS ³⁶ | MSMA Net | 0.7854 | 0.9813 | 0.9794 | 0.9563 | 0.8177 |
| 1 | DR HAGI ³⁶ | DRIVE ³² | MSMA Net | 0.7500 | 0.9722 | 0.9638 | 0.9523 | 0.7605 |
| 2 | | STARE ³³ | MSMA Net | 0.7544 | 0.9818 | 0.9744 | 0.9607 | 0.7812 |
| 3 | | CHASE_DB1 ³⁴ | MSMA Net | 0.7484 | 0.9780 | 0.9711 | 0.9542 | 0.7639 |
| 4 | | HRF ³⁵ | MSMA Net | 0.7237 | 0.9703 | 0.9592 | 0.9459 | 0.7295 |

Table 8. Comparison of cross-testing results: One dataset is used for training, and the testing is done on the remaining four datasets.

Experiment 7 - segmentation of small blood vessels

Although several segmentation methodologies have been proposed, two visual constraints have consistently been noted in all methods: 1. Incapability to detect small blood vessels, and 2. diminished sensitivity. The only way to overcome these disadvantages is by efficiently segmenting small vessels, which are crucial for achieving high sensitivity. Experimental findings validate the proposed MSMA Net model’s effective identification of vessels, regardless of their size, leading to superior performance. The proposed MSMA Net model result can be directly compared to the gold standard image. It provides ophthalmologists with an excellent opportunity to observe the progression of retinal diseases. In Fig. 8a, the comparison between the ground truth and small blood vessels generated by MSMA Net method reveals that the results of our method clearly display small blood vessels. The accurate detection of these small vessels significantly enhances the effectiveness of our approach.

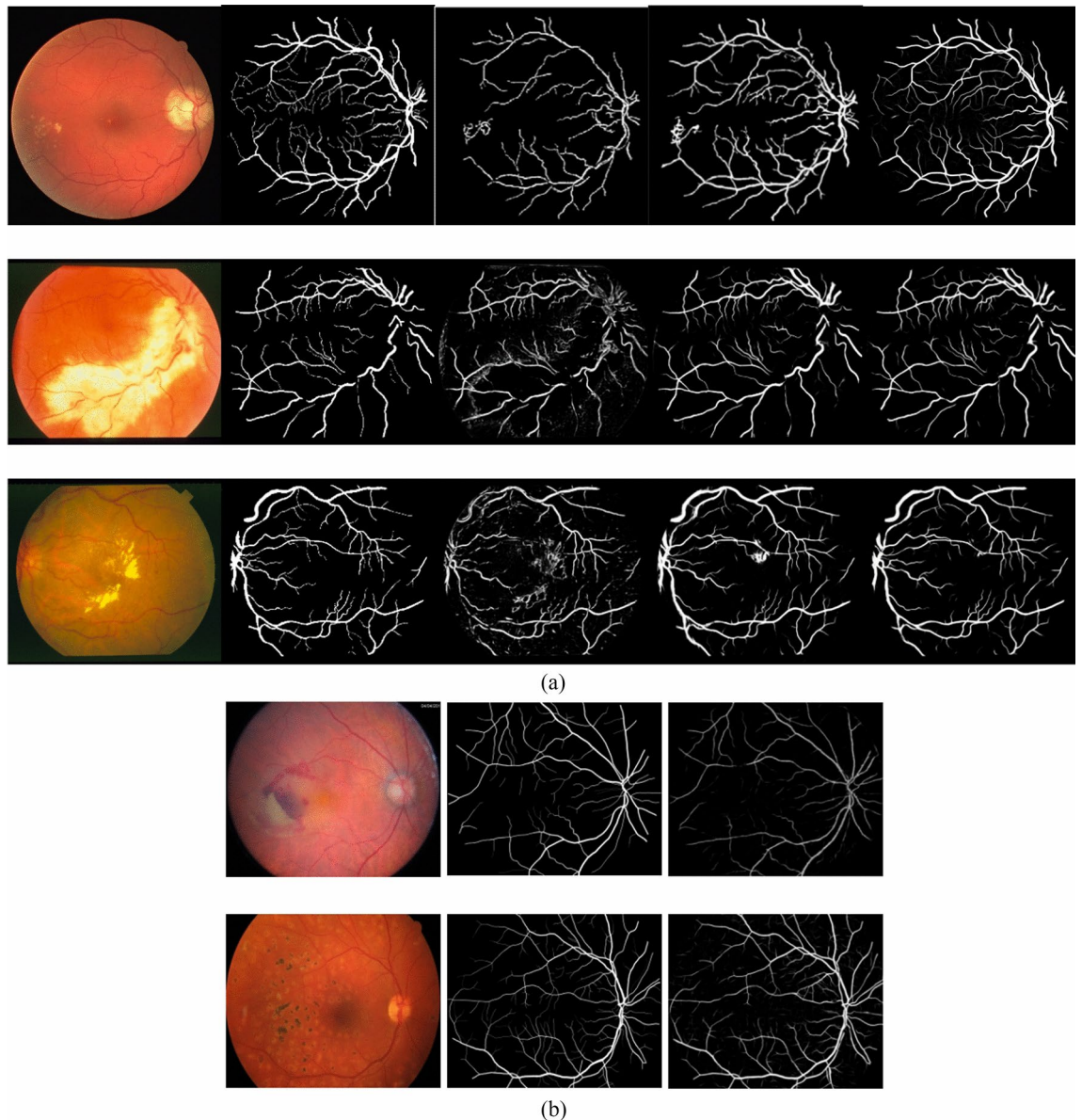


Fig. 7. (a) Vessel segmentation in Images with Pathologies containing dark and bright lesions. (Column-wise) 1. Retinal images, 2. Manual ground-truths, 3. Segmented vessels from Nuygen's method⁷, 4. Segmented vessels from Soomro's method⁴⁰, and 5. Segmented vessels from our MSMA Net method. (b). Vessel segmentation in Images with hemorrhages. (Column-wise) 1. Retinal images, 2. Manual groundtruth, 3. Segmented vessels from our MSMA Net method. Both images have hemorrhages overlapping with the vessels. The regions enclosed by yellow boxes in the third column images are showing the broken vessels indicating True Negatives and isolated patches impersonating vessels indicating False Positives.

Experiment 8 - performance on various challenges

This experiment analyses how well the proposed method performs under various challenges. The model demonstrates the capability to enhance vascular structures with varying morphological details despite encountering several limitations. These limitations include vessels exhibiting central reflex, poor contrast micro-vessels, cross-over vessels, the branching of vessels, and very closed parallel vessels that tend to be detected as one large vessel. Figure 8b presents the results of the proposed method. The proposed MSMA Net method efficiently tackles challenges while preserving the quality of signals in the vascular regions. MSMA Net employs pre-processing techniques specifically designed to mitigate the effects of noise and uneven illumination, such as shade correction, CLAHE, and MO on green channel images. These methods help to standardize the illumination across the image and reduce the impact of noise, thereby improving the overall quality of the input data for the model.

This is demonstrated through significant performance improvements with advanced pre-processing. From Table 1, in the DRIVE database, the combination of shade correction, CLAHE, and MO increased AUC by

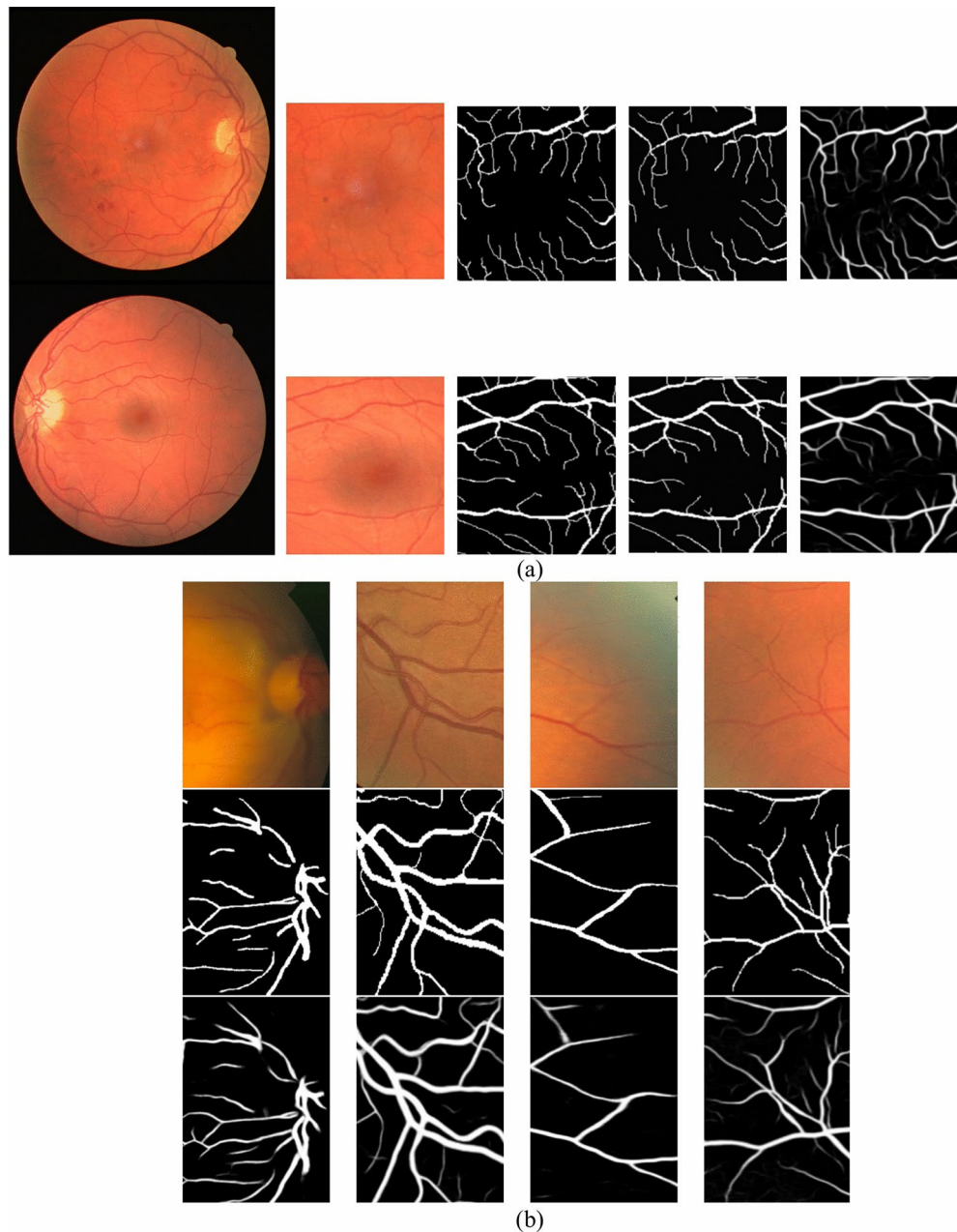


Fig. 8. (a) Observation-Segmentation of small vessels: The retinal images and cropped regions featuring obscured blood vessel appearance are depicted in the first and second columns respectively. Following them, the corresponding ground truth from the first and second experts are displayed in the third and fourth columns respectively. The fifth column presents the segmented output of our MSMA Net method. (b) Blood vessel segmentation in challenging images. (Row-wise) 1. Retinal image patches with central light reflex and low-varying contrast, 2. Corresponding patches from ground-truths, and 3. Segmented vasculature using the proposed MSMA Net method.

0.50%, Accuracy by 2.00%, dice score by 7.10%, Sensitivity by 15.00%, and Specificity by 0.08%. Before applying these techniques (original color fundus image), the metrics were 0.9849, 0.9690, 0.8116, 0.7673, and 0.9882, respectively. Similar enhancements in metrics were observed in other databases as well.

The multi-scale convolution block enables the model to process information across various receptive fields, which helps in recognizing and enhancing vascular structures despite the presence of noise and uneven illumination, ensuring that both small and large vessel structures are accurately segmented. SE blocks recalibrate the feature maps by emphasizing the most informative channels, enhancing the network's ability to distinguish between vascular and non-vascular regions. This recalibration helps in maintaining the integrity of the signal in vascular regions, even when the images are affected by noise and uneven illumination. The SAB (Spatial Attention Block) enhances the model's ability to focus on relevant spatial regions of the image, effectively filtering

out irrelevant noise. By emphasizing the spatial dimensions of the features, the SAB ensures that the network concentrates on the vascular structures, thereby improving segmentation performance even in challenging conditions.

Combination of both pre-processing methods and specialized network blocks enhances the robustness of MSMA Net against noise and uneven illumination. From Table 4, in the DRIVE database, the full model configuration achieved an AUC of 0.9898, Accuracy of 0.9883, dice Score of 0.8692 and Sensitivity of 0.8821. Compared to the basic U-Net, which had an AUC of 0.9736, Accuracy of 0.9536, dice score of 0.8175 and Sensitivity of 0.7653, the full model showed increases of 1.66% in AUC, 3.63% in Accuracy, 6.33% in dice score and 15.26% in Sensitivity. Similar enhancements in metrics were observed in other databases. These enhancements highlight the effectiveness of MSMA Net's advanced techniques in ensuring high-quality segmentation under challenging conditions.

Discussion and conclusion

Accurate segmentation of blood vessels in retinal images is essential for promptly identifying, diagnosing, and treating retinal disorders that pose a threat to vision. Numerous deep learning methodologies have been proposed with the objective of achieving accurate segmentation of retinal vasculature. However, achieving this objective faces challenges due to difficulties in deriving multiscale features effectively, information loss from pooling operations, and insufficient handling of local contextual information through skip connections. To overcome the challenge posed by the limited number of images in widely recognized benchmark datasets such as DRIVE³², STARE³³, CHAS_DB1³⁴, HRF³⁵ and DR HAGIS³⁶, we developed an innovative framework for automatically segmenting retinal vessels. The proposed MSMA Net model demonstrates significant advancements in the segmentation of retinal blood vessels, addressing critical challenges faced in this domain. The comprehensive approach, incorporating innovative preprocessing, data augmentation, and network architecture improvements, has led to marked enhancements in model performance.

Key advantages of the MSMA Net model are summarized as follows:

Preprocessing and data augmentation

Enhanced image quality

The use of diverse preprocessing techniques such as shade correction, CLAHE, and MO significantly improves the quality of input data and generates a wide range of image variations, from noise-reduced and high-contrast to uniformly illuminated images. This diversity in the training data helps the model learn from a broad spectrum of scenarios, leading to better vessel delineation.

Augmented dataset

To address the limited availability of datasets for blood vessel segmentation, we implemented a customized input pipeline strategy. This strategy involved extracting fixed-size patches cropped at random locations. As a result, we were able to create a dataset thousands of times larger than the original one. The proposed model successfully segmented vessels, particularly showed its ability in segmenting thin vessels hidden within low-contrast regions, with improved accuracy. Additionally, it effectively suppressed the generation of noise in cases where vessels were surrounded by similar vessel-like pathologies.

Innovative network architecture

Multi-scale feature extraction

The incorporation of the MSSE block allows the model to extract features using multi-scale kernels, enhancing its ability to process information across various receptive fields. This is crucial for detecting vessels of varying thicknesses and discriminating them from non-vessel pathologies.

Attention mechanisms

Taking into account the intricate nature of retinal vascular structures, delicate blood vessels, lighting conditions, limited to complex contextual information, and the potential for confusion with lesions, we explored the relationship between characteristic channels by SE blocks. The channel-wise informative feature emphasizing the power of SE block also has contributed to the increased model generalisation capability towards the detection of blood vessels under different conditions. The SE block in both encoders and decoders basically helped in balancing the pooling and interpolation operations to be more effective.

The SAB (Spatial Attention Block) further refines the feature maps by focusing on spatially significant regions in the image, which helps in distinguishing vessels from similar-looking structures. Having this attention gate at the juncture of encoder and decoder hierarchical feature fusion has helped the model to be more robust in detecting vessels under various conditions.

Performance and generalizability

Cross-dataset validation

The model's performance remains robust when trained on one benchmark dataset and tested on others (DRIVE, STARE, CHASE_DB1), demonstrating its superior generalizability across different benchmark datasets.

Robust vessel segmentation

The MSMA Net model excels in segmenting challenging small and edge vasculature, as well as other regions within retinal images. This is achieved with only a small increase in model complexity compared to the basic U-Net architecture.

The proposed MSMA Net model contains 4.092 million trainable parameters, with a model size of 15.61 MB and an inference speed of 1.730 s. While the MSMA Net method has shown remarkable performance in vessel segmentation for both normal and pathological fundus images, and those with challenging conditions such as varying lighting conditions, uneven illumination, poor contrast micro-vessels, cross-over vessels, and branching vessels, there are few areas requiring further improvement and future exploration.

Future work could focus on four important aspects which may improve the model's performance: (1) enhancing the model's ability by incorporating dual attention mechanisms that separately focus on vessels and lesion features, allowing the model to better distinguish between these structures, (2) developing mechanism to deal hemorrhages overlapped with the vessels which pose a great challenge in reducing the true negatives and false positives of the model predictions, (3) augmenting training data with synthetic lesions or using generative adversarial networks (GANs) to create varied lesion examples can improve model robustness, and (4) utilizing more robust feature extraction layers that are less sensitive to noise and illumination changes, such as more sophisticated convolutional layers or hybrid architectures combining CNNs with transformers, could further enhance performance.

Data availability

The datasets analysed during the current research work are available in the [DRIVE; STARE; CHASE_DB1; HRF; DrHAGIS] repositories, [<https://drive.grand-challenge.org/>; <https://cecas.clemson.edu/~ahoover/stare/>; <https://researchdata.kingston.ac.uk/96/>; <https://www5.cs.fau.de/research/data/fundus-images/>; <https://persona.lpages.manchester.ac.uk/staff/niall.p.mcloughlin/>].

Received: 16 April 2024; Accepted: 20 December 2024

Published online: 27 January 2025

References

- Lam, B. & Liew, Y. G. A. General retinal vessel segmentation using regularization-based multiconcavity modeling. *IEEE Trans. Med. Imaging* **29**, 1369–1381 (2010).
- Zhang, J. et al. Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. *IEEE Trans. Med. Imaging* **35**, 2631–2644 (2016).
- Bibiloni, P., González-Hidalgo, M. & Massanet, S. A survey on curvilinear object segmentation in multiple applications. *Pattern Recognit.* **60**, 949–970 (2016).
- Debayle, J. & Presles, B. Rigid image registration by general adaptive neighborhood matching. *Pattern Recognit.* **55**, 45–57 (2016).
- Trucco, E., Azegrouz, H. & Dhillon, B. Modeling the tortuosity of retinal vessels: does caliber play a role. *IEEE Trans. Biomed. Eng.* **57**, 2239–2247 (2010).
- Lajevardi, S. M., Arakala, A., Davis, S. A. & Horadam, K. J. Retina verification system based on biometric graph matching. *IEEE Trans. Image Process.* **22**, 3625–3635 (2013).
- Nguyen, K., Fookes, C., Jillela, R., Sridharan, S. & Ross, A. Long range iris recognition: a survey. *Pattern Recognit.* **72**, 123–143 (2017).
- Badrinarayanan, V., Kendall, A. & Cipolla, R. 'SegNet: A deep convolutional encoder–decoder architecture for image segmentation'. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 2481–2495 (2017).
- Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* (Springer, 2015).
- Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M. & Asari, V. K. Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. [arXiv:1802.06955](https://arxiv.org/abs/1802.06955) (2018).
- Jin, Q. et al. 'DUNet: A deformable network for retinal vessel segmentation'. *Knowl. Based Syst.* **178**, 149–162 (2019).
- Wang, B., Qiu, S. & He, H. Dual encoding U-Net for retinal vessel segmentation. In: *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* 84–92, (Springer, 2019).
- Yue, K., Zou, B., Chen, Z. & Liu, Q. Retinal vessel segmentation using denseU-net with multi scale inputs. *J. Med. Imag.* **6**, 1 (2019).
- Hu, J. et al. 'S-UNet: A bridge-style U-Net framework with a saliency mechanism for retinal vessel segmentation'. *IEEE Access* **7**, 174167–174177 (2019).
- Wei, X., Yang, K., Bzdok, D. & Li, Y. Orientation and context entangled network for retinal vessel segmentation. *Expert Syst. Appl.* **217**, 119443 (2023).
- Yang, L., Wang, H., Zeng, Q., Liu, Y. & Bian, G. A hybrid deep segmentation network for fundus vessels via deep-learning framework. *Neurocomputing* **448**, 168–178 (2021).
- Xu, G. X. & Ren, C. X. SPNet: A novel deep neural network for retinal vessel segmentation based on shared decoder and pyramid-like loss. *Neurocomputing* **523**, 199–212 (2023).
- Wang, G. et al. Automatic vessel crossing and bifurcation detection based on multi-attention network vessel segmentation and directed graph search. *Comput. Biol. Med.* **155**, 106647 (2023).
- Boudegga, H. et al. Fast and efficient retinal blood vessel segmentation method based on deep learning network. *Comput. Med. Imag. Graph.* **90**, 101902 (2021).
- Hu, K. et al. Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function. *Neurocomputing* **309**, 179–191 (2018).
- Fu, H. et al. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans. Med. Imag.* **37**, 1597–1605 (2018).
- S. Zhang, H. Fu, Y. Yan, Y. Zhang, Q. Wu, M. Yang, M. Tan, and Y. Xu. Attention guided network for retinal image segmentation. In: *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 797–805 (Springer, 2019).
- Liu, Y., Shen, J., Yang, L., Bian, G. & Yu, H. 'ResDO-UNet: A deep residual network for accurate retinal vessel segmentation from fundus images'. *Biomed. Signal Process. Control* **79**, 104087 (2023).
- Mou, L. et al. 'Dense dilated network with probability regularized walk for vessel detection'. *IEEE Trans. Med. Imag.* **39**, 1392–1403 (2020).
- Cherukuri, V., Kumar, B. G. V., Bala, R. & Monga, V. Deep retinal image segmentation with regularization under geometric priors. *IEEE Trans. Image Process.* **29**, 2552–2567 (2020).
- Yan, Z., Yang, X. & Cheng, K.-T. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Trans. Biomed. Eng.* **65**, 1912–1923 (2018).
- Pizer, S. M. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process* **39**, 355–368 (1987).

28. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. [arXiv:1602.07261](https://arxiv.org/abs/1602.07261), (2016).
29. Ibtehaz, N. & Rahman, M. S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw.* **121**, 74–87 (2020).
30. Drozdal, M., Vorontsov, E., Chartrand, G., Kadoury, S. & Pal, C. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications* (eds Drozdal, M. et al.) (Springer, 2016).
31. Oktay, O. et al. Attention UNet: Learning where to look for the pancreas. [arXiv:1804.03999](https://arxiv.org/abs/1804.03999) (2018).
32. Staal, J., Abramoff, M. D., Niemeijer, M., Viergever, M. A. & van Ginneken, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imag.* **23**, 501–509 (2004).
33. Hoover, A. D., Kouznetsova, V. & Goldbaum, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imag.* **19**, 203–210 (2000).
34. Fraz, M. M. et al. 'An ensemble classification-based approach applied to retinal blood vessel segmentation'. *IEEE Trans. Biomed. Eng.* **59**, 2538–2548 (2012).
35. Odstrcilik, J. et al. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. *IET Image Process.* **7**, 373–383 (2013).
36. Holm, S., Russell, G., Nourrit, V. & McLoughlin, N. DR HAGIS-a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients. *J. Med. Imag.* **4**, 014503 (2017).
37. Taha, A. A. & Hanbury, A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med. Imag.* **15**, 1–28 (2015).
38. Guns, R., Lioma, C. & Larsen, B. The tipping point: F-score as a function of the number of retrieved items. *Inform. Process. Manag.* **48**, 1171–1180 (2012).
39. Mo, J. & Zhang, L. Multi-level deep supervised networks for retinal vessel segmentation. *Int. J. Comput. Assist. Radiol. Surg.* **12**, 2181–2193 (2017).
40. Soomro, T. A. et al. Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation. *Expert Syst. Appl.* **134**, 36–52 (2019).
41. Feng, S., Zhuo, Z., Pan, D. & Tian, Q. Ccnet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features. *Neurocomputing* **392**, 268–276 (2020).
42. Li, X. et al. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE Trans Med Imaging* **37**, 2663–2674 (2018).
43. Yan, Z., Yang, X. & Cheng, K.-T. A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE J. Biomed. Health Inform.* **23**, 1427–1436 (2018).
44. Guo, S. et al. Bts-dsn: Deeply supervised neural network with short connections for retinal vessel segmentation. *Int. J. Med. Inform.* **126**, 105–113 (2019).
45. Madathil, S. & Padannayil, S. K. MC-DMD: A data-driven method for blood vessel enhancement in retinal images using morphological closing and dynamic mode decomposition. *J. King. Saud. Univ. Comp. Info. Sci.* **34**, 5223–5239 (2022).
46. Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, 1395–1403 (2015).
47. Wang K., Zhang, X., Huang, S., Wang Q. & Chen, F. Ctf-net: Retinal vessel segmentation via deep coarse-to- ne supervision network. *IEEE 17th International Symposium on Biomedical Imaging (ISBI)* 1237–1241 (2020).
48. Wang, Bo. et al. CSU-Net: A context spatial u-net for accurate blood vessel segmentation in fundus images. *IEEE J. Biomed. Health Inform.* **25**, 1128–1138 (2020).
49. Alvarado-Carrillo, D. E. & Dalmau-Cedeño, O. S. Width attention based convolutional neural network for retinal vessel segmentation. *Expert Syst. Appl.* **209**, 118313 (2022).
50. Zhang, Y. et al. Bridge-net: Context-involved U-net with patch-based loss weight mapping for retinal blood vessel segmentation. *Expert Syst. Appl.* **195**, 116526 (2022).
51. Feng, Q. L. B., Xie, L. P., Liang, P., Zhang, H. & Wang, T. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Trans. Med. Imag.* **35**, 109–118 (2016).
52. Liu, R., Wang, T., Zhang, X. & Zhou, X. DA-Res2UNet: Explainable blood vessel segmentation from fundus images. *Alex. Eng. J.* **68**, 539–549 (2023).
53. Ding, W. et al. RCAR-UNet: Retinal vessel segmentation network algorithm via novel rough attention mechanism. *Inform. Sci.* **657**, 120007 (2024).
54. Kande, G. B. et al. MSR U-Net: An improved U-Net model for retinal blood vessel segmentation. *IEEE Access* **12**, 534–551 (2024).
55. Orlando, J. I., Prokofyeva, E. & Blaschko, M. B. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Trans. Biomed. Eng.* **64**, 16–27 (2016).

Acknowledgements

This work was supported the Korea Environmental Industry & Technology Institute(KEITI), with a grant funded by the Korea government, Ministry of Environment(The development of IoT-based technology for collecting and managing big data on environmental hazards and health effects), under Grant RE202101551.

Author contributions

Giri Babu Kande: Problem finding, Conceptualization, Methodology, Implementation, Manuscript drafting. Madhusudana Rao Nalluri: Conceptualization, Methodology, Reviewing, Results validation, Supervision. R.Manikandan: Results validation, Supervision, Methodology, Reviewing. Jaehyuk Cho: Problem finding, Conceptualization, Methodology, Implementation, Manuscript drafting, Investigation, Conceptualization, Methodology, Reviewing, Results validation, Supervision. Sathishkumar Veerappampalayam Easwaramoorthy: review & editing, Problem finding, Conceptualization, Methodology, Implementation, Manuscript drafting.

Funding

This work was supported the Korea Environmental Industry & Technology Institute (KEITI), with a grant funded by the Korea government, Ministry of Environment (The development of IoT-based technology for collecting and managing big data on environmental hazards and health effects), under Grant RE202101551.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.R.N. or J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025