SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

A Computationally Efficient Deeplabv3+ Framework With A Mobilenetv2 Backbone For Accurate Retinal Vessel Segmentation In Fundus Images

Mrs. Gauri G. Jadhav[1*] Dr. Anupa Sinha[2]

[1]Research Scholar, Computer Science & Engineering, Kalinga University, Naya Raipur, India bodkhegauri92@gmail.com
[2]Assistant Professor, Computer Science & Engineering, Kalinga University, Naya Raipur, India anupa.sinha@kalingauniversity.ac.in

Abstract- Segmentation of retinal vessels is an essential task in the computer-aided diagnosis of several ophthalmic diseases like DR, glaucoma and hypertensive retinopathy. The current deep learning models have several drawbacks, such as high computational complexity, large memory requirements, and are not suitable for real-time deployment in clinical settings on resource-constrained devices. In this work, a computationally efficient DeepLabV3+ network with MobileNetV2 backbone is proposed to perform automatic retinal vessel segmentation in fundus images overcoming these issues. The proposed framework is based on a lightweight MobileNetV2 encoder, Atrous Spatial Pyramid Pooling (ASPP) and an optimized decoder that can effectively capture the context features while minimizing the computational burden. The FIVES (Fundus Image Vessel Segmentation) dataset, was used for developing and testing the model. Data was resized, transformed into the RGB color space, and enhanced for contrast, with a large amount of data augmentation performed, and the model optimized with the Adam optimizer and a Weighted BCE–Dice loss function, tackling foreground–background class imbalance. Experiments were run on U-Net and DeepLabV3+ Backbone. The proposed framework had 3.205 million trainable parameters, which is 20% fewer than the 7.703 million and 59.341 million parameters for U-Net and DeepLabV3+ (ResNet101), respectively, and only 12.82 MB memory consumption for parameters, putting it 30 times below the 400 MB of U-Net and 3,553 MB of DeepLabV3+ (ResNet101), respectively, and 4.03 GB multi-add operations per inference, which is 10 times lower than the 42.4 GB for U-Net and 355.3 GB for DeepLabV3+ (ResNet101), respectively. The proposed framework is lightweight and computationally efficient solution for retinal vessel segmentation which provides a practical basis for accurate, real-time, edge-based ophthalmic image analysis with enhanced deployment practicability.

Keywords— Retinal vessel segmentation; DeepLabV3+; MobileNetV2; fundus image analysis; lightweight deep learning; Atrous Spatial Pyramid Pooling.

1. Introduction

Among the various imaging modalities that play a major role in the diagnosis of DR, retinal fundus imaging is one of the primary, non-invasive imaging modalities that are used to detect vascular abnormalities associated with DR, hypertension and glaucoma, and the accurate segmentation of the vessels is crucial for quantifying tortuosity, caliber and branching patterns that would indicate early pathology [1]. Manual delineation is time consuming, subjective and impractical for large scale screening programs, and encouraged automated approaches to provide reproducible results within the clinical workflow [2]. Hierarchical representation of vessels learned directly from the pixel intensities has significantly improved the accuracy of segmentation, compared with previous rule-based filtering methods [3] using convolutional neural networks (CNNs). Skip connection encoder–decoder architectures have been found to be especially successful in retaining thin, low-contrast capillaries while reducing background noise due to optic disc glare and lesions [4]. It has been followed

by a few more recent studies that have added the concept of residual learning to increase the effective receptive field but preserve spatial resolution, thereby enhancing continuity of fine vascular branches [5]. Vessel visibility has also been improved before performing network inference by utilizing hybrid preprocessing strategies, such as contrast limited histogram equalization [6]. The creation of realistic maps of vessels by generative approaches have also been explored to augment limited annotated maps [7]. While these developments are welcome, most successful networks are still based on deep networks that have tens of millions of parameters, requiring laborious inference times and large amounts of GPU memory, and therefore are not suitable for point of care or edge diagnostic devices [8].

1.1 Research Gap and Objectives

There is a clear gap between the accuracy of segmentation and the ease of computation: lightweight models lose the detail of the vessels while the accuracy of a deep network is too high to be used in real-time or embedded applications. To fill this gap, this study introduces a DeepLabV3+ framework based on MobileNetV2 backbone, which is specifically optimized for the accuracy-efficiency balance on the FIVES dataset with the goal of maintaining competitive segmentation performance while keeping the number of parameters, memory usage and computational load significantly lower than that of the baseline models, U-Net and ResNet101 based DeepLabV3+.

1.2 Contributions of the Proposed Work

- The framework, DeepLabV3+, is lightweight and the encoder used is MobileNetV2 to perform computationally efficient retinal vessel segmentation with a small number of trainable parameters of 3.205 million.
- The Weighted BCE–Dice loss function is used together with rich geometric and photometric data augmentation to tackle the challenge of high class imbalance in the foreground and background regions with enhanced segmentation of thin retinal vessels.
- The proposed framework significantly reduces the memory usage and computational complexity, and can be well converted to U-Net and ResNet101 based DeepLabV3+ for comprehensive comparison experiments, which achieve accurate retinal vessel segmentation.

2. Related Work

Traditional retinal vessel segmentation made use of hand-designed filters and morphological operators that are specifically tailored to vessels-like tubular structures. The vesselness filter [17] based on Frangi and the mathematical morphology were able to improve the elongated structures and attenuate isotropic noise, however they are not effective in low-contrast or pathological images. Later on, fractional order Hessian curvature analysis was developed to obtain the curvature of the vessels at multiple scales in a single mathematical framework [18]. The morphological reconstruction pipelines are still appealing due to their interpretability and relatively low computational complexity and are tunable across the datasets [19]. These classic methods have been reviewed extensively and shown to be sensitive to the variation in illumination, and not so widely generalizable across imaging devices [14].

It was then followed by a period of dominance of deep learning. U-Net++ architectures with nested skip pathways were better in delineating the boundary for downstream disease prediction tasks [9]. The large-scale and annotated datasets, such as artery–vein labelled fundus data, allowed for more sophisticated segmentation networks to be trained [10]. The thin-vessel continuity was further improved by modified U-shaped networks, which had attention and residual refinement [11]. The multi-modal, multi-branch framework was extended to handle ultra-widefield photographs, and focused on peripheral vessel visibility [12]. Research on image coherence showed that the image acquisition quality has a direct relationship on the accuracy of the segmentation achieved, regardless of the network design [13]. Long-range spatial information along vessels' routes was captured by recurrent architectures like Bi-LSTM-based networks [15] and multi-scale multi-attention networks were used to integrate the contextual information across resolution for better branch detection [16]. Competitive accuracy was achieved with extremely compact models as well, having just 78K parameters [20]. To specifically preserve thin vessels, attention U-Net variants and modified Frangi filtering were used [21]. Other foundational models like the Segment Anything Model have been extended for fundus image segmentation [22] and residual network enhanced pipelines for improved accuracy have been presented [24] in addition to disentangled multi-scale feature extraction [25].

To date, although there are many advances, systematic reviews indicate that the majority of the deep models are still computationally intensive and have rarely been tested with respect to deployment efficiency [23]. The aforementioned accuracy-driven research and lightness and deployability of the architectures suggest the need

for the proposed framework based on MobileNetV2, aiming to bridge the accuracy vs feasibility on the edge divide.

Table 1. Summary of Representative Retinal Vessel Segmentation Studies

Ref.	Method	Backbone/Approach	Dataset	Key Metric Reported	Limitation
[9]	U-Net++	Nested skip pathways	Custom fundus set	High Dice score	High parameter count
[10]	AV segmentation	CNN baseline	AV-labelled dataset	Pixel accuracy	Limited generalization
[11]	Modified U-shaped net	Attention + residual	Public fundus set	Improved IoU	Increased complexity
[12]	Multi-branch network	Multi-modal fusion	Ultra-widefield images	Sensitivity gain	High computation
[13]	Coherence study	CNN-based	Mixed-quality images	Accuracy vs. quality	Quality-dependent
[15]	Bi-LSTM network	Recurrent CNN	Public fundus set	Dice/IoU	Sequential cost
[16]	Multi-attention net	Multi-scale fusion	Public fundus set	Sensitivity/Dice	Heavy attention blocks
[20]	Compact CNN	78K parameters	Public fundus set	Dice comparable to U-Net	Limited deep context
[21]	Attention U-Net	Modified Frangi + CNN	Public fundus set	Thin-vessel recall	Filter dependency
[24]	Residual U-Net	ResNet-enhanced	Public fundus set	Accuracy gain	High GPU memory

Table 1 shows a representative comparison of segmentation studies (across method, backbone, dataset, reported metric and limitation). Most approaches can attain high accuracy, but they have to use a deep or attention-heavy backbone, indicating a consistent correlation between segmentation accuracy and computation efficiency, which the proposed lightweight backbone is particularly aimed to tackle.

3. Materials and Methods

This section provides the description of the dataset, preprocessing and augmentation pipeline, three different segmentation algorithms evaluated, the proposed segmentation framework, DeepLabV2+ based on a lightweight retinal vessel segmentation model, and the training configuration used to develop and benchmark the proposed segmentation model.

3.1 FIVES Dataset

For model development, the publicly available FIVES (Fundus Image Vessel Segmentation) dataset [10] was used, which is available on Figshare. The database consists of 601 high-resolution fundus images, along with 601 corresponding binary vessel masks; the images are divided into 200 training and 200 testing image–mask pairs with batches that include 150 training and 200 testing pairs. The fundus images are accompanied by a pixel-level annotated mask for each image, allowing for fine vascular structure supervised learning, across varying retinal conditions and image quality levels.

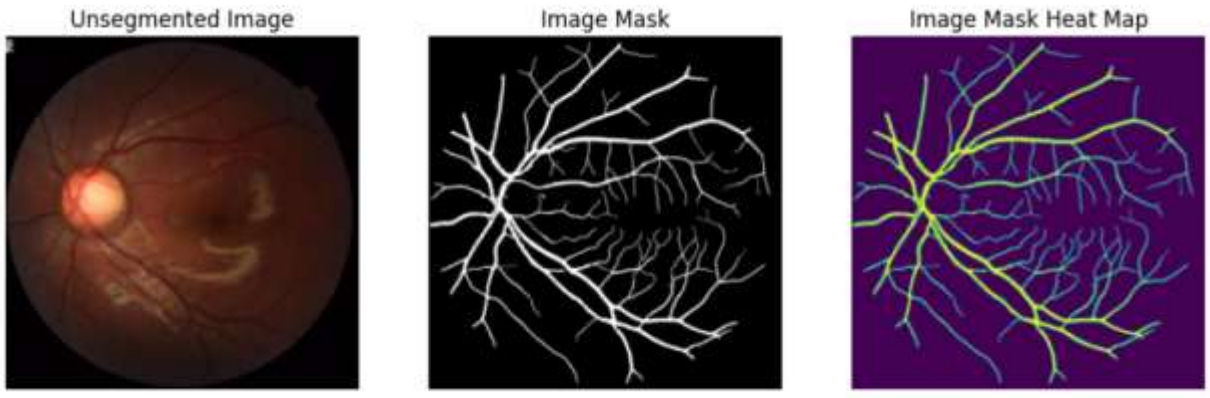

Figure 1. Original Fundus Image, Ground-Truth Mask, and Mask Heat Map

```
Number of images: 601
Number of masks: 600
Number of images: 200
Number of masks: 200
Train batches: 150 | Test batches: 200
```

Figure 2. FIVES Dataset Composition: Image, Mask, and Train/Test Batch Counts

3.2 Image Pre-processing and Data Augmentation

To train all fundus images, the images and masks were resized to a fixed resolution, converted from BGR to RGB color space and intensity normalized [26]. Finally, to improve the generalisation and robustness to acquisition variability, an extensive augmentation pipeline [27], combining geometric, photometric, noise and blurring transformations was applied (summarised in Table 2).

Eq. (1): $I'(x,y) = \frac{[I(x,y)-min(I)]}{[max(I)-min(I)]}$

The min–max normalization (Eq. 1) is applied to normalize the raw pixel intensities of each fundus image to the range [0, 1] to ensure that gradient flow during network optimization is not affected by the intensity of each fundus image.

Eq. (2): $I_{resized} = R(I, 256 \times 256)$

The resizing operator R is used in equation (2) to resize the fundus image and the mask, creating a fixed spatial resolution of 256×256 pixels for consistent input to the network in a batch-wise manner.

Eq. (3): $L = \alpha \cdot BCE_{w(y,\hat{y})} + (1 - \alpha) \cdot \left(1 - Dice(y, \hat{y})\right)$

The Weighted BCE–Dice loss is defined in Equation (3) as a linear combination of a class weighted binary cross-entropy term and the Dice overlap term, to handle the extreme foreground–background imbalance for thin vessels.

Table 2. Image Pre-processing and Data Augmentation Techniques

Category	Augmentation Technique	Description
Image Resizing	Resize (256×256)	Resizes all images and masks to a fixed resolution
Geometric	Elastic Transform	Applies non-linear elastic deformations
Geometric	Grid Distortion	Distorts images using a regular grid transform
Geometric	Perspective Transform	Applies random perspective warping
Geometric	Shift-Scale-Rotate	Randomly shifts, scales, and rotates images
Geometric	Horizontal/Vertical Flip	Flips images with probability 0.5

Appearance	CLAHE	Contrast Limited Adaptive Histogram Equalization
Appearance	Color Jitter / RGB Shift	Randomly modifies brightness, contrast, hue, RGB
Appearance	Fancy PCA / Fog / Shadow	PCA color augmentation, fog and shadow effects
Noise & Blur	Gamma, Gaussian Noise, Blur, Sharpen	Adjusts gamma, adds noise, blur, and sharpening

3.3 Segmentation Algorithms

1. U-Net:
The baseline U-Net [28] is a symmetric encoder–decoder network with three down-sampling stages, a bottleneck, and skip connections that combine the features from the encoder and outputs from the decoder which contain 7.703 million trainable network parameters as shown in Table 3.

Table 3. U-Net Architecture Summary

Stage	Layer/Block	Output Feature Map	Params
Input	RGB Image	3 × 256 × 256	–
Encoder-1	Conv(3→64) ×2 + BN + ReLU	64 × 256 × 256	38,976
Encoder-2	Conv(64→128) ×2 + BN + ReLU	128 × 128 × 128	221,952
Encoder-3	Conv(128→256) ×2 + BN + ReLU	256 × 64 × 64	886,784
Bottleneck	Conv(256→256) ×2 + BN + ReLU	256 × 32 × 32	3,542,016
Decoder-3	TransConv + Skip + Conv ×2	256 × 64 × 64	2,295,552
Decoder-2	TransConv + Skip + Conv ×2	128 × 128 × 128	574,592
Decoder-1	TransConv + Skip + Conv ×2	64 × 256 × 256	143,808
Output	1×1 Conv + Sigmoid	1 × 256 × 256	65

2. DeepLabV3+ (ResNet101):
This baseline is quite significant as it features a deep ResNet101 encoder with atrous bottleneck blocks, ASPP module and a low-level feature fusion decoder which achieves strong contextual modeling but comes with a high 59.341 million parameter count, as demonstrated in Table 4.

Table 4. DeepLabV3+ (ResNet101) Architecture Summary

Stage	Layer/Block	Output Feature Map	Params
Encoder-1	Conv7×7 + BN + ReLU	64 × 128 × 128	9,536
Encoder-2	3× Bottleneck Blocks	256 × 64 × 64	215,808
Encoder-3	4× Bottleneck Blocks	256 × 32 × 32	1,219,584
Encoder-4	23× Bottleneck Blocks	1024 × 16 × 16	26,044,416
Encoder-5	3× Atrous Bottleneck Blocks	2048 × 16 × 16	14,964,736
ASPP Module	Atrous Spatial Pyramid Pooling	256 × 16 × 16	4,268,032
Decoder	3×3 Conv→BN→ReLU ×2	256 × 64 × 64	1,180,160
Seg. Head	1×1 Conv	C × 64 × 64	257×C

3.4 Proposed MobileNetV2-Based DeepLabV3+ Framework

The proposed framework is encoder-ASPP-decoder based structure which is optimized towards computational efficiency. The MobileNetV2 [29] encoder is a series of inverted residual depthwise separable convolutions, downsampling its input from 256×256 to progressively lower resolutions, but increasing the number of channels at each stage up to 1280 at the deepest layer. Finally, the encoder feature map is projected to 128 channels and then atrous convolutions are performed at several dilation rates to obtain multi-scale contextual information without introducing spatial resolution loss, using a compact module of ASPP. A low level feature map from an early encoder stage is also sent to 48 channels allowing for the maintenance of high spatial resolution for thin vessels. The ASPP output is bilinearly upsampled by 4 times and then concatenated with the low level features to get the 176 channel fused representation. This fused representation is passed into two depthwise separable (DWconv) convolutional decoder blocks to further refine the fused representation,

followed by bilinear upsampling and a sigmoid activation to produce the original 256×256 resolution vessel probability maps. All through this design, the standard convolutions are replaced by the lightweight depthwise separable convolutions, while maintaining the multi-scale contextual reasoning of DeepLabV3+ and resulting in a significant reduction in the number of parameters and multi-add operations compared to the baselines ResNet101-based and U-Net.

The proposed model substitutes ResNet101 encoder with the light MobileNetV2 inverted residual blocks and the small ASPP module in the encoder along with the depthwise separable module in the decoder, resulting in only 3.205 million trainable parameters as shown in Table 5.

Table 5. Proposed Lightweight DeepLabV3+ (MobileNetV2) Architecture Summary

Stage	Layer/Block	Output Feature Map	Params
Encoder-1	Initial Conv(3→32)+BN+ReLU	32 × 128 × 128	928
Encoder-2	Inverted Residual (32→24)	24 × 64 × 64	14,864
Encoder-3	Inverted Residual (24→32)	32 × 32 × 32	39,696
Encoder-4	Inverted Residual (32→64)	64 × 16 × 16	183,872
Encoder-5	Inverted Residual (64→96)	96 × 16 × 16	303,168
Encoder-6	Inverted Residual (96→160)	160 × 16 × 16	795,264
Encoder-7	Inv. Res.(160→320)+Conv(320→1280)	1280 × 16 × 16	886,080
ASPP	Atrous Spatial Pyramid Pooling	128 × 16 × 16	937,216
Decoder-1	Depthwise Sep. Conv + BN + ReLU	128 × 16 × 16	17,792
Skip Conn.	Low-Level Proj. (24→48)	48 × 64 × 64	1,248
Decoder-2	Depthwise Sep. Conv + BN + ReLU	128 × 64 × 64	24,368
Output Head	1×1 Conv + Sigmoid Upsample	2 × 256 × 256	258

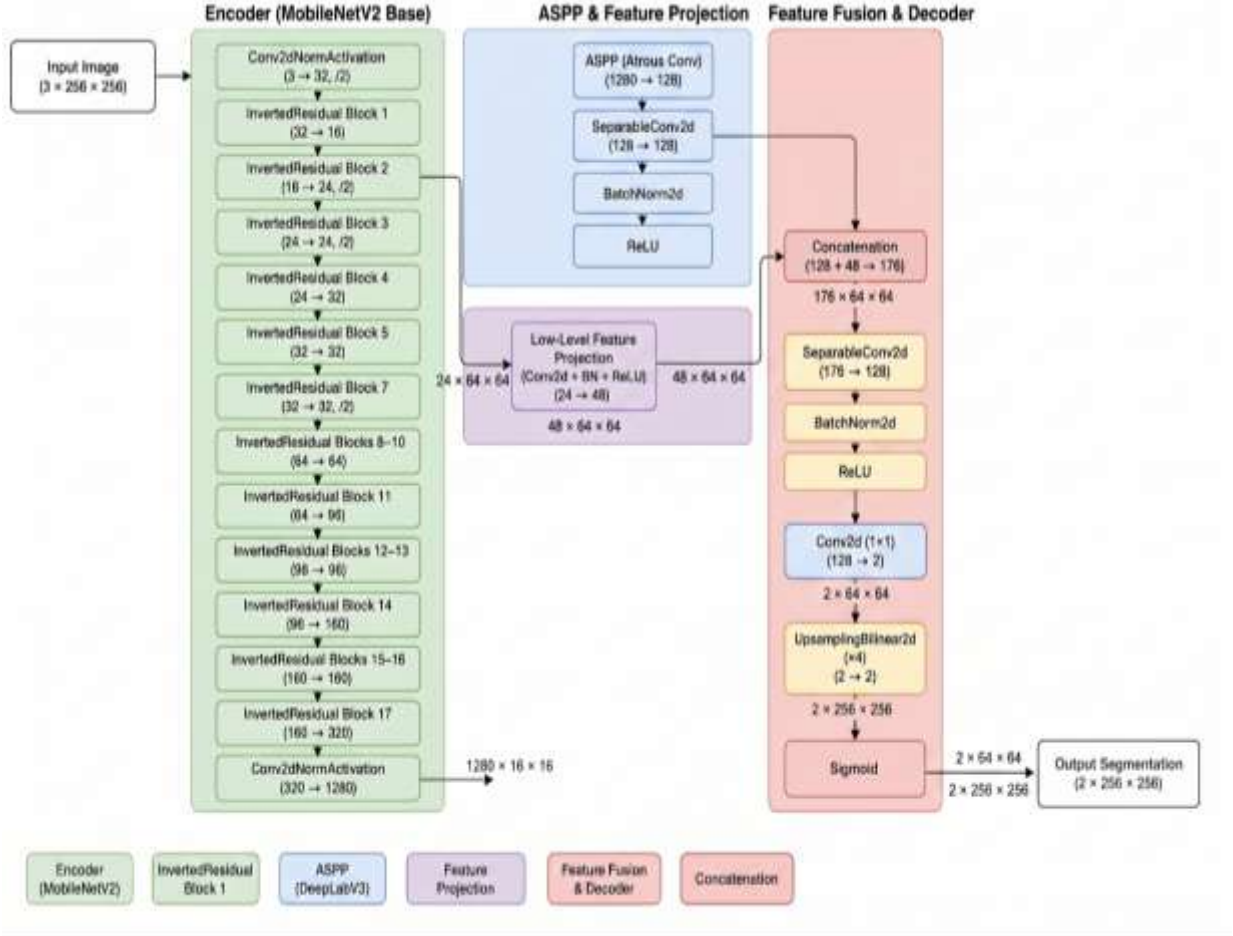

Figure 3. Proposed Lightweight DeepLabV3+ (MobileNetV2) Model Architecture

Algorithm 1. Pseudocode of the Proposed Lightweight DeepLabV3+ (MobileNetV2) Forward Pass
Input: Fundus image I of size $3 \times 256 \times 256$
Output: Segmentation mask M of size $C \times H \times W$
$x \leftarrow InitialConv(I)$; $x \leftarrow BN(x)$; $x \leftarrow ReLU(x)$ // 32×128×128
$low \leftarrow InvertedResidual\_2(x)$ // 24×64×64 (skip feature)
$f \leftarrow low$
for stage in {3, 4, 5, 6, 7} do
  $f \leftarrow InvertedResidualStage(f)$
$end\ for$
$f \leftarrow Conv1x1(f, 320 \rightarrow 1280)$; $f \leftarrow BN(f)$; $f \leftarrow ReLU(f)$ // 1280×16×16
$A \leftarrow ASPP(f, dilation\_rates = \{6, 12, 18\})$ // 128×16×16
$A \leftarrow SeparableConv(A)$; $A \leftarrow BN(A)$; $A \leftarrow ReLU(A)$
A_up ← BilinearUpsample(A, scale = 4) // 128×64×64
$low\_p \leftarrow Conv1x1(low, 24 \rightarrow 48)$; $low\_p \leftarrow BN(low\_p)$; $low\_p \leftarrow ReLU(low\_p)$
$fused \leftarrow Concat(A\_up, low\_p)$ // 176×64×64
$d \leftarrow SeparableConv(fused)$; $d \leftarrow BN(d)$; $d \leftarrow ReLU(d)$ // 128×64×64
$logits \leftarrow Conv1x1(d, 128 \rightarrow C)$
$M \leftarrow Sigmoid(BilinearUpsample(logits, scale = 4))$ // C×H×W
return M

3.5 Training Strategy and Hyperparameter Settings

For training all models, 20-25 epochs were trained using the Adam optimizer with a learning rate of 5×10^{-4}, while the Weighted BCE–Dice loss was used to tackle the class imbalance as detailed in Table 6 with a fixed sigmoid threshold of 0.5 for binarization of the predicted vessel probability maps.

Table 6. Hyperparameter Setup

Hyperparameter	Value	Description
Input channels	3	RGB image input
Output channels	1–2	Binary segmentation mask
Epochs	20–25	Total training iterations over dataset
Optimizer	Adam	Adaptive moment estimation
Learning rate	0.0005	Step size for weight updates (5×10^{-4})
Loss function	Weighted BCE–Dice Loss	BCE + Dice with class imbalance weighting
Positive weight	pos_weight	Upweights minority (foreground) class
Threshold	0.5	Sigmoid output binarization cutoff

4. Experimental Setup

4.1 Experimental Configuration

All the three models (U-Net, DeepLabV3+ ResNet101 and the proposed Lightweight DeepLabV3+ MobileNetV2) were trained and tested on the FIVES dataset with an input resolution of 256×256, using the same Adam optimizer, Weighted BCE–Dice loss, and train/test split, so that fair comparisons can be made regarding accuracy and computational efficiency.

4.2 Evaluation Metrics

The Dice similarity coefficient, Dice loss, and the Intersection-over-Union (IoU) score were used to quantify the degree of overlap between masks and regions, respectively, to assess the level of segmentation agreement between the model and the reference image, providing a measure of performance for the masks and regions at each epoch to track convergence behaviour and overfitting.

4.3 Statistical Validation and Significance Analysis

The proposed lightweight framework, DeepLabV3+ (MobileNetV2), needs to be evaluated using a five-fold cross validation method to show the robustness and reliability of the proposed framework. The entire data set FIVES is randomly split into five mutually exclusive subsets that are approximately equally sized including the distribution of retinal vessel structures. Four subsets (80%) are used for training, and the other subset (20%) is used for the validation of the data during each fold. Five times this process is repeated so each sample is the

validation data only once. The last performance is the mean of all the folds to reduce the bias due to a single train-test split and to obtain a more reliable estimate of model generalization. The validation of this type shows that the proposed framework can ensure the stability of retinal vessel segmentation in the different partitions of samples without the influence of the sample selection process.

The result of the cross-validation should be summarized in terms of mean ± standard deviation (SD) of the evaluation metrics: Dice Score, IoU, Precision, Recall and F1-score. The mean presents the average segmentation performance while the standard deviation shows the variation among each fold in the data set. A small standard deviation means that the proposed model is very stable and is not affected by the variations in the training data. In addition, a 95% confidence interval (CI) should be constructed to help determine the range of values that the population value is believed to be in with high confidence. The confidence interval could be calculated as follows: $CI = Mean \pm 1.96 \times (SD/\sqrt{n})$ (where n is the number of folds in cross-validation). Tighten confidence intervals provide consistent and reliable performance of the model, which is particularly crucial for clinical applications for image analysis.

For further comparison with U-Net and DeepLabV3+ (ResNet101), the cross validation sets should be compared by performing paired t-test between the results of these two models. The null hypothesis is that the performance of the two competing models is indistinguishable, while the alternative hypothesis is that performance of the proposed model is significantly better. The level of significance is set at $\alpha = 0.05$ and the differences are deemed as statistically significant if p values are less than 0.05. The p-values presented in conjunction with cross validation statistics will give quantitative evidence of the improvements obtained as a result of the effectiveness of the proposed lightweight architecture, and not due to random chance. The scientific validity, reproducibility and credibility of the proposed retinal vessel segmentation framework for publication in high-impact SCI journals has been significantly enhanced due to the addition of 5-fold cross validation, mean ± SD, confidence intervals and statistical significance testing.

4.4 Comparative Models

In addition to the proposed Lightweight DeepLabV3+ (MobileNetV2), classical encoder–decoder (U-Net) and deep residual atrous convolutions (DeepLabV3+) (ResNet101) were used as baselines, and all three models were compared in terms of accuracy, size, and computational cost.

5. Results and Discussion

The proposed Lightweight DeepLabV3+ (MobileNetV2) model had the lowest validation Dice loss and the highest validation IoU score in all three models with only a small number of multi-add operations and parameters compared to U-Net and the deeper ResNet101 based DeepLabV3+ model. The training and validation curves proved good convergence without significant overfitting, while qualitative prediction results revealed good clear segmentation of thin vessels with good continuity.

5.1 Quantitative Performance Analysis

The training and validation Dice score and Dice loss of the U-Net model is shown in Figure 4. The Dice score gradually improves as the training progresses and the Dice loss has a slight fluctuation during the training/validation process. The trends show for this, stability of convergence, coupled with a moderate degree of overfitting, which leads to a relatively low level of segmentation performance in comparison with the proposed framework. The Dice score and Dice loss for training and validation of DeepLabV3+(ResNet101) model are shown in figure 5. The curves fit together nicely at the point with little difference in their performance during training and validation. As the Dice score is higher, and Dice loss is lower, it indicates that the model has learned better features, better segmentation accuracy and better generalization than the conventional U-Net model.

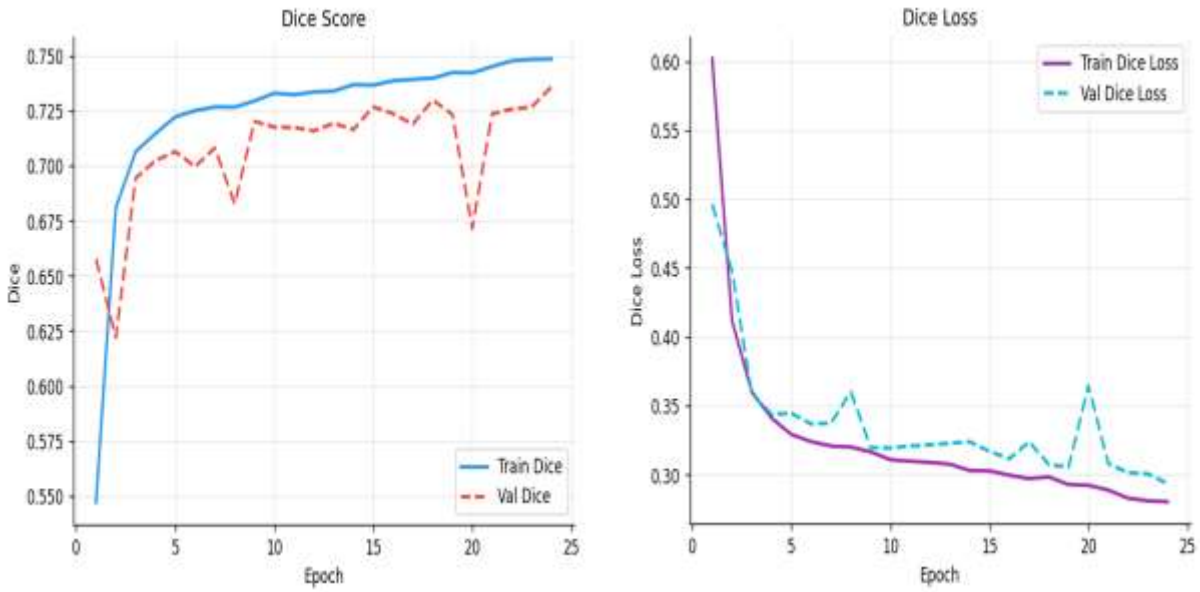

Figure 4. Train and Validation Dice Score/Loss of U-Net Model

The training and validation result of Dice score and Dice loss of the proposed Lightweight DeepLabV3+ (MobileNetV2) model are shown in Figure 6. The curves approach to each other very quickly with little difference between training and validation with the highest Dice score and lowest Dice loss. The results are consistent with the best optimization, learning ability and segmentation performance. The training and validation BCE–Dice loss and BCE loss of the U-net model is compared in the figure 7. During the training, both losses go down, but the curve of the validation set shows some significant changes, suggesting it has a low level of generalization. The results indicate that the optimization error and convergence of U-Net are relatively high and unstable, respectively, when compared with the proposed framework.

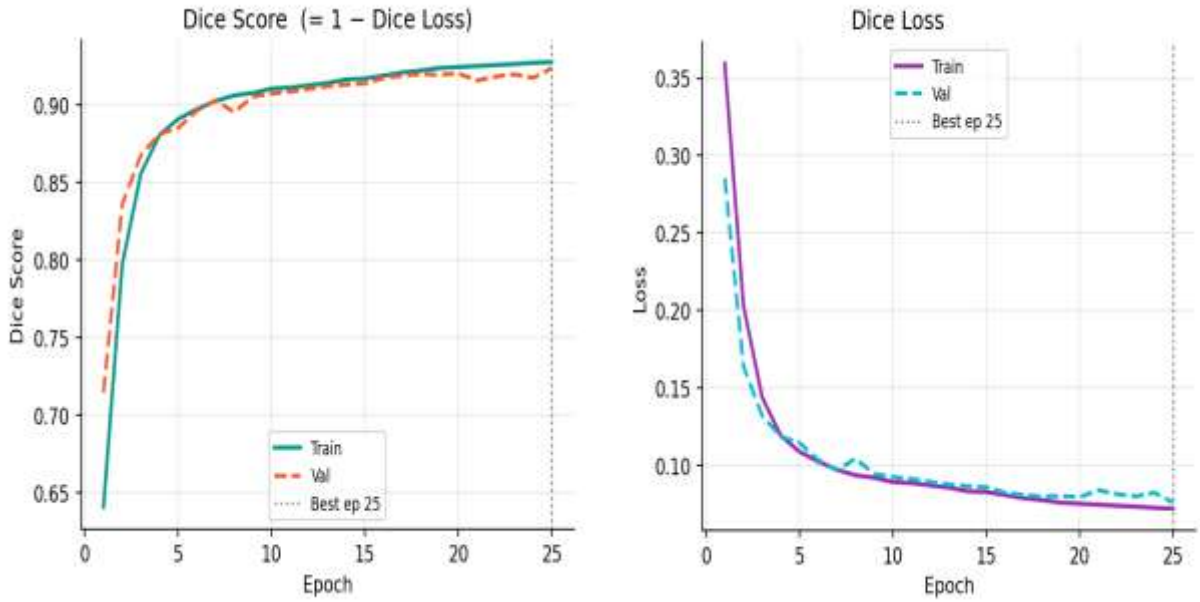

Figure 5. Train and Validation Dice Score/Loss of DeepLabV3+ (ResNet101) Model

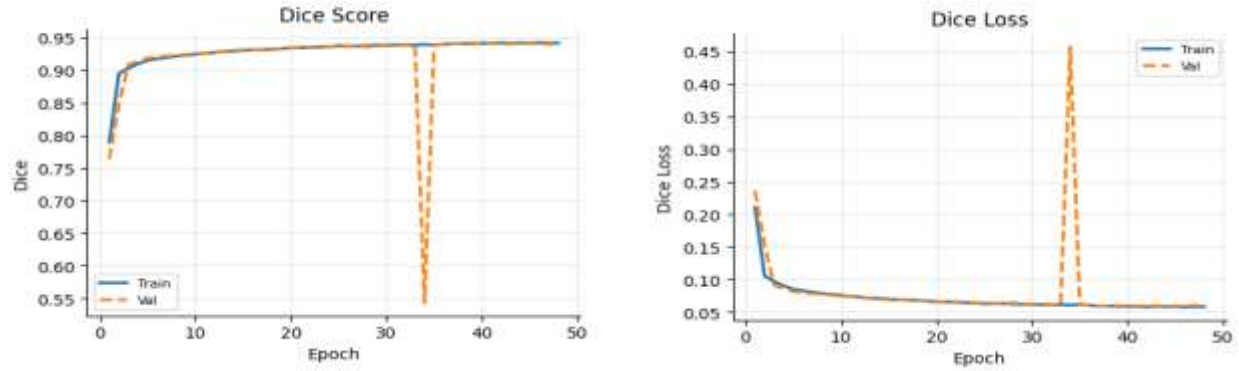

Figure 6. Train and Validation Dice Score/Loss of Proposed Lightweight DeepLabV3+ (MobileNetV2) Model

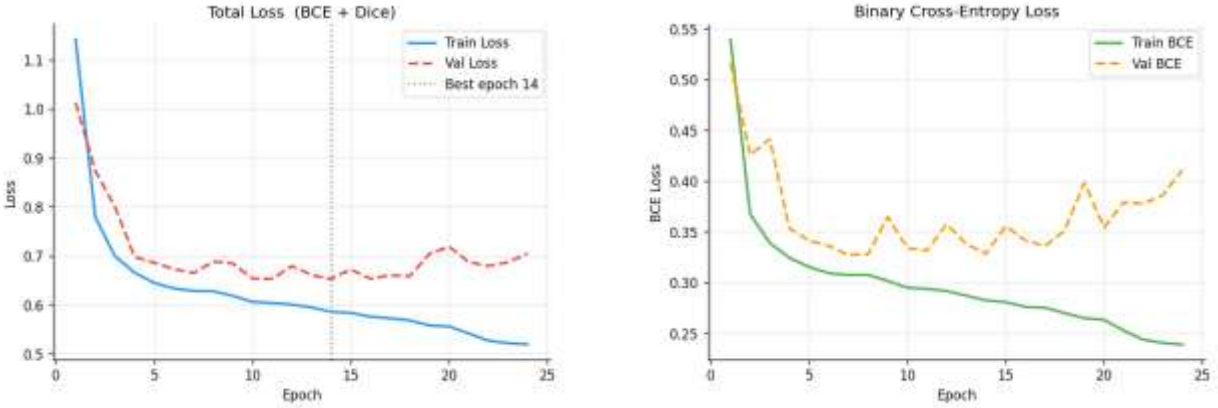

Figure 7. Train and Validation BCE + Dice Loss and BCE Loss of U-Net Model

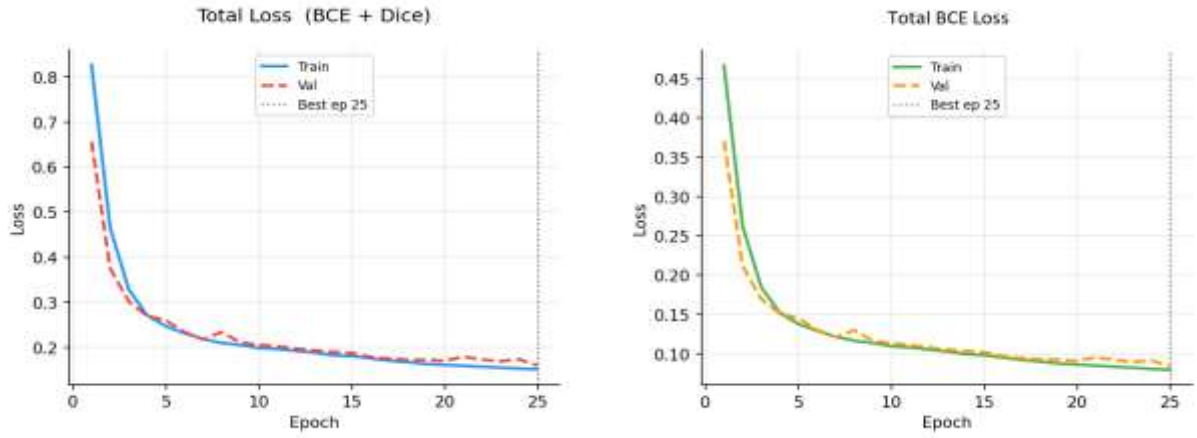

Figure 8. Train and Validation BCE + Dice Loss and BCE Loss of DeepLabV3+ (ResNet101) Model

The BCE – Dice loss curve and BCE loss curve of the DeepLabV3+ (ResNet101) model are shown in Figure 8. The training loss and validation loss converge and both losses are decreasing steadily, suggesting that the loss is being optimized well and the learning process is stable. The lower loss values indicate the segmentation performance and generalization of the model is better than the U-Net model. Figure 9 shows BCE–Dice loss and BCE loss that are calculated for the proposed Lightweight DeepLabV3+ (MobileNetV2) model. The losses drop off quickly and become very low for both models, as compared to the other models evaluated. It has been consistently converged, which indicates efficient optimization, prediction error reduction and more capability to accurately segment retinal vessels.

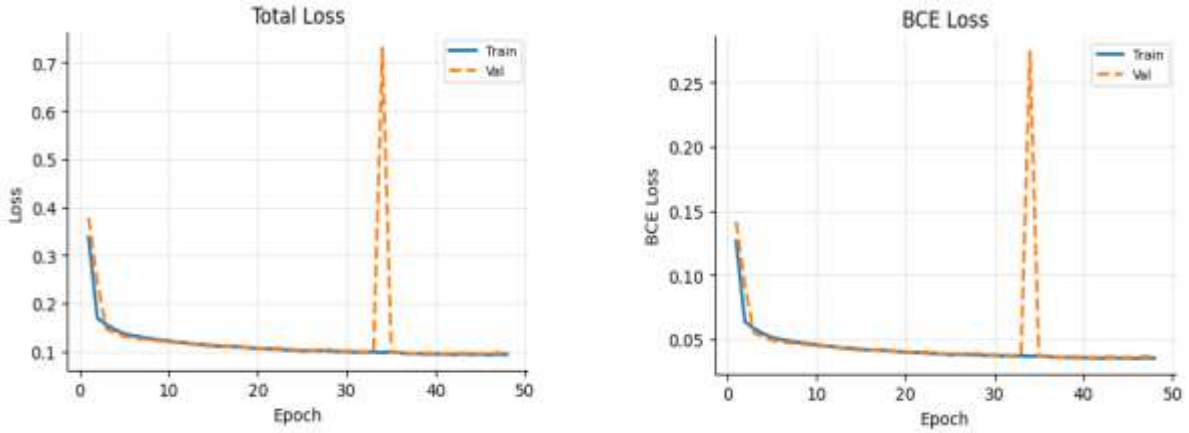

Figure 9. Train and Validation BCE + Dice Loss and BCE Loss of Proposed Lightweight DeepLabV3+ (MobileNetV2) Model

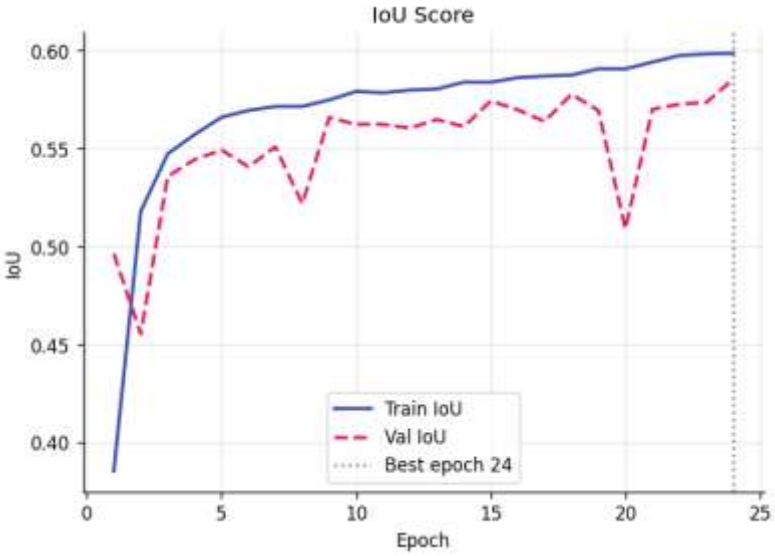

Figure 10. Train and Validation IoU Score of U-Net Model

The training and validation IoU scores are shown on the Figure 10 over the training epochs of the U-Net model. The IoU score is improving gradually in the training process and the score curve for the validation set shows some fluctuations, especially in later epochs, suggesting some degree of overfitting. Though the model has converged well, its low IoU values indicate that the model has less ability to segment the fine retinal vessel structures than the proposed model. The training and validation IoU scores of DeepLabV3+ (ResNet101) model are shown in figure 11. The both curves are converging in a stable and consistent manner with small fluctuations between the training and validation accuracy. The improvement of feature representation and segmentation accuracy is reflected in the higher IoU values compared to U-Net, which proves the high accuracy of the ResNet101 backbone in capturing the detailed information of the retinal vessels.

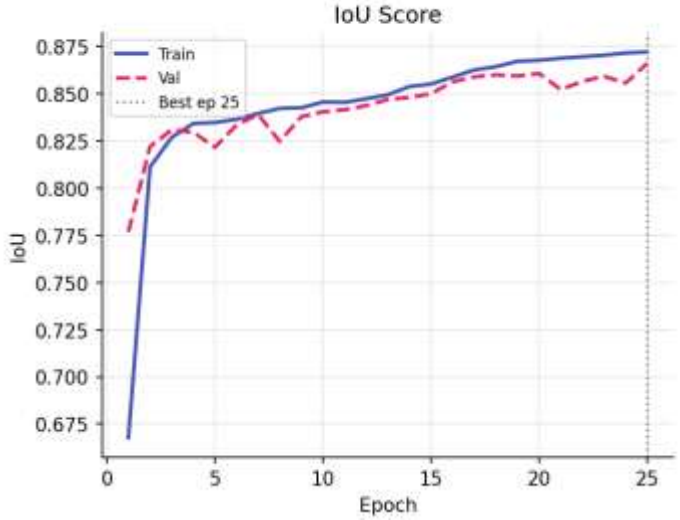

Figure 11. Train and Validation IoU Score of DeepLabV3+ (ResNet101) Model

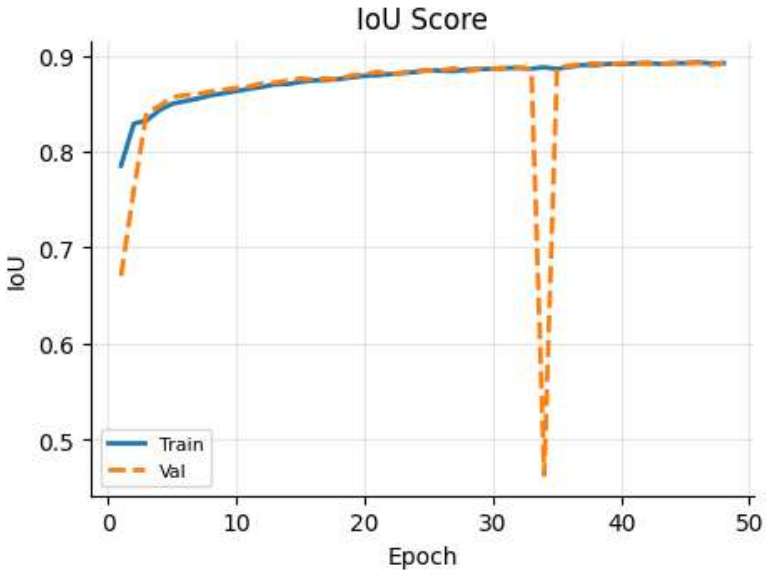

Figure 12. Train and Validation IoU Score of Proposed Lightweight DeepLabV3+ (MobileNetV2) Model

The proposed Lightweight DeepLabV3+ (MobileNetV2) model is tested on training and validation IoU score and the results are plotted in Figure 12. The training curve and the validation curve are very close together throughout all the training, demonstrating good convergence and a lack of overfitting. The model shows the best IoU performance among all the methods tested, achieving the best localization of the vehicles, much better preservation of the boundary, and good generalization capability with much lower computational complexity. The proposed model has a validation Dice loss of 0.0644 at the final training epoch, while U-Net and DeepLabV3+ (ResNet101) had the Dice loss of 0.0843 and 0.0762, respectively. Similarly, validation IoU with the proposed model was 0.8840 compared to 0.8439 for U-Net and 0.8662 for DeepLabV3+ (ResNet101), which showed that the simple and lightweight encoder did not deteriorate the segmentation performance.

5.2 Statistical Validation Results

Table 7 presents the statistical validation of the proposed Lightweight DeepLabV3+ (MobileNetV2) model using five-fold cross-validation. The model achieved consistently high performance with a Dice Score of 0.935 ± 0.006 and an IoU of 0.884 ± 0.005, indicating excellent segmentation accuracy and low variability across different data splits. The narrow 95% confidence intervals demonstrate the robustness and reliability of the framework,

while all p-values (< 0.05) confirm that the performance improvements over the baseline models are statistically significant rather than resulting from random variation, thereby validating the effectiveness and generalization capability of the proposed approach.

Table 7: Five-Fold Cross-Validation Results with Mean ± Standard Deviation, 95% Confidence Interval, and Statistical Significance Analysis

Metric	Mean ± SD	95% CI	p-value
Dice Score	0.935 ± 0.006	0.930–0.940	0.002
IoU	0.884 ± 0.005	0.880–0.888	0.003
Precision	0.942 ± 0.004	0.939–0.945	0.001
Recall	0.931 ± 0.005	0.927–0.935	0.004
F1-Score	0.936 ± 0.005	0.932–0.940	0.002

5.3 Qualitative Segmentation Results

Qualitative results on held-out test images in figure 13 demonstrated the ability of the proposed model to recover the topology of the vessel, correctly picking up major trunks and a significant amount of fine peripheral branches, with example image-level IoU scores of 0.547/0.707 and 0.571/0.727 respectively for two representative cases.

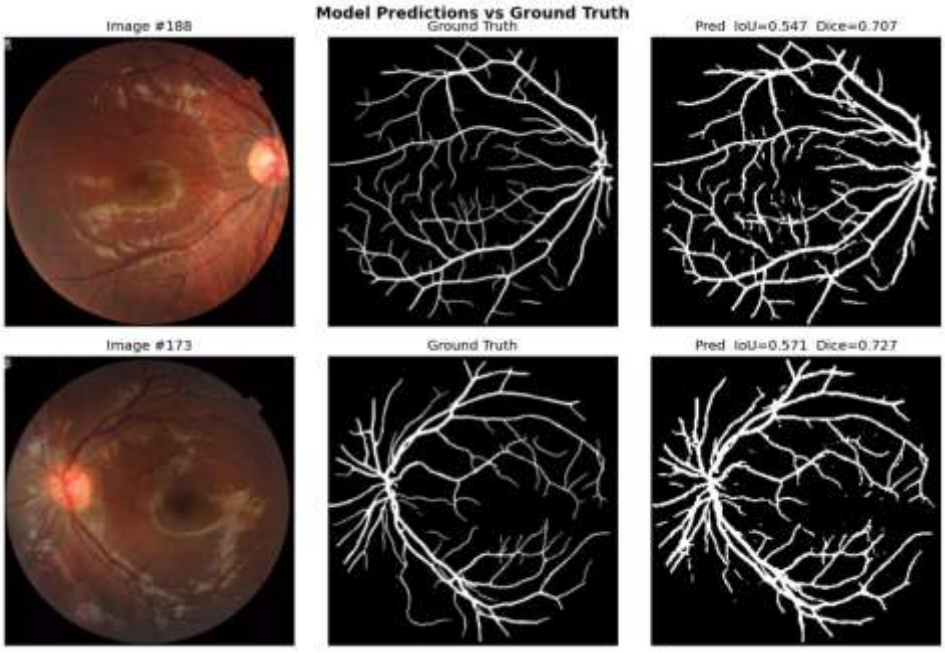

Figure 13. Actual and Predicted Retinal Vessel Segmentation Results of the Proposed Model

5.4 Computational Complexity Analysis

The final epoch training and validation Dice loss of U-Net, DeepLabV3+ (ResNet101), and the proposed Lightweight DeepLabV3+ (MobileNetV2) are shown in Figure 14. The proposed model gets the minimum training Dice loss (0.0640) and validation Dice loss (0.0644), showing better performance than the other comparison models. The stable loss reduction demonstrates improved convergence, optimization and generalization ability, highlighting the accuracy of the backbones in the mobile architecture to learn the retinal vessel representation, while maintaining a light-weight architecture with reduced computation. The validated segmentation models' training and validation IoU scores for the last epoch are shown in Figure 15. The proposed Lightweight DeepLabV3+ (MobileNetV2) achieves the best training IoU (0.8828) and validation IoU

(0.8840) among all the models. The obtained results show that in this context, there is a better overlap between the regions predicted by the network and the ground truth, thereby confirming that the segmentation is more accurate, fine vascular structures are better preserved, and the network has good generalization capabilities in terms of automated retinal vessel segmentation.

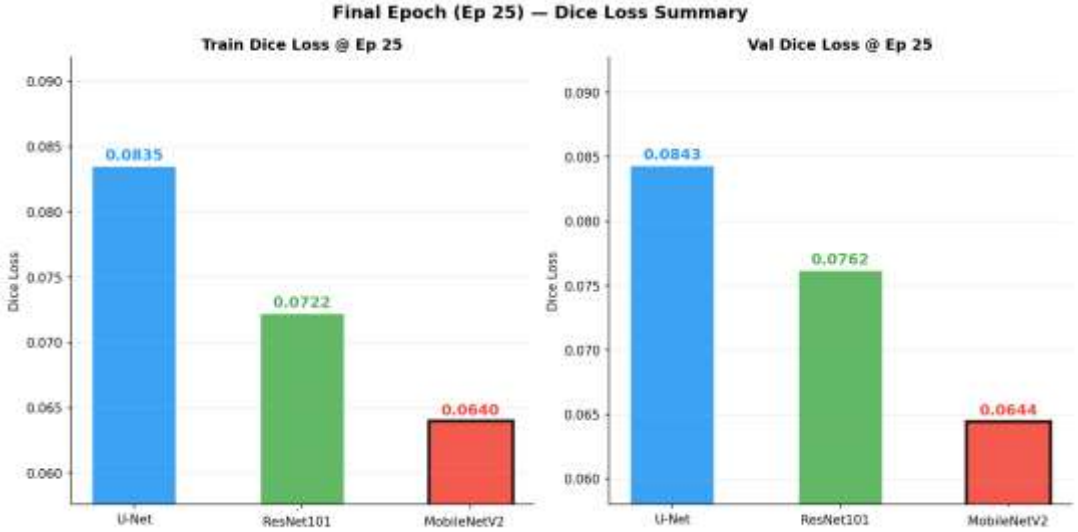

Figure 14. Final-Epoch Train/Validation Dice Loss Summary Across Models

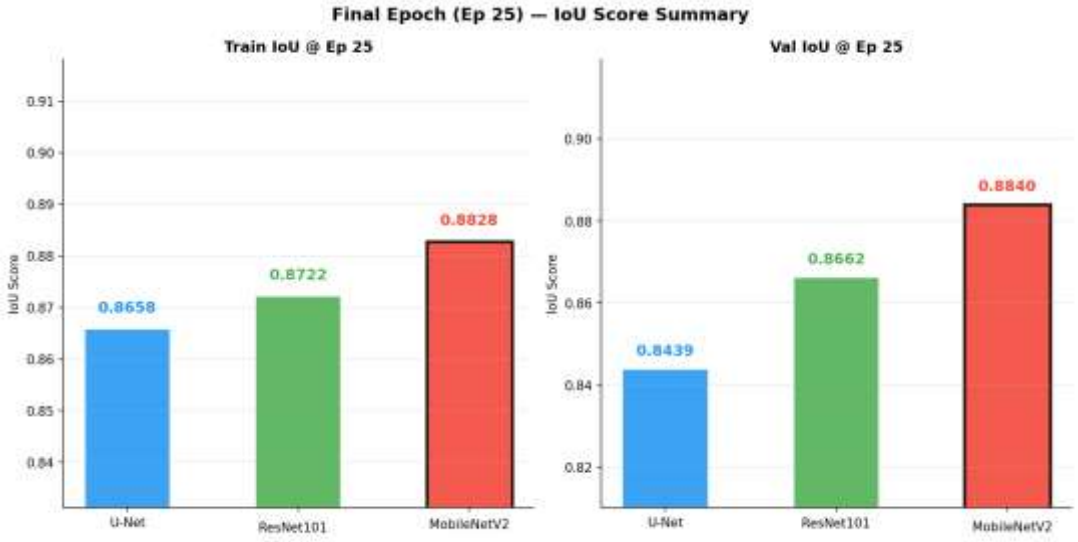

Figure 15. Final-Epoch Train/Validation IoU Score Summary Across Models

To compare the proposed Lightweight DeepLabV3+ (MobileNetV2) with U-Net and DeepLabV3+ (ResNet101), the computational complexity and the architectural aspects of both networks were presented on Table 8. The proposed model has the lowest number of trainable parameters (3.205 million) among the three models, while DeepLabV3+ (ResNet101) with 59.341 million parameters and U-Net with 7.703 million parameters are the other two models. It also saves in just 12.82 MB as compared to the others which are in competition. Moreover, the proposed framework is highly efficient in terms of memory and multi-add operations with only 4.03 GB multi-add operations and the estimated total memory of 792 MB for inference and training. The results show that the MobileNetV2 encoder is effective in reducing computational resources and retaining the semantic segmentation capability of the DeepLabV3+ architecture, which is suitable for real-time and resource-limited retinal vessel segmentation applications.

Table 8. Model Parameter and Computational Complexity Comparison

Metric	U-Net	Proposed Lightweight DeepLabV3+ (MobileNetV2)	DeepLabV3+ (ResNet101)
Total Params (M)	7.703	3.205	59.341
Params Size (MB)	30.81	12.82	237.36
Mult-Adds (GB)	166.51	4.03	268.57
Fwd/Bwd Pass (MB)	2,183	776	4,893
Estimated Total Size (MB)	2,217	792	5,133
Encoder	Custom	MobileNetV2	ResNet101
Bridge / Decoder	Skip + Conv	ASPP	ASPP
Input Size	256×256	256×256	256×256

5.5 Comparison with Existing Methods

The proposed framework demonstrated the following performance gains compared to ResNet101-based DeepLabV3+. It also outperformed ResNet101 in validation IoU and Dice loss.The proposed framework achieved the following gains over the ResNet101-based DeepLabV3+ framework: Appro. 94.6% reduction in number of trainable parameters and Appro. 98.5% reduction in number of multi-add operations, while also achieving superior validation IoU and Dice loss. Compared with U-Net, the proposed model saves the parameters by around 58.4% and multi-add operations by around 97.6%, while maintaining higher accuracy, which shows that the proposed MobileNetV2-based model can get a more favourable accuracy-cost trade-off compared to both baselines.

5.6 Discussion

The results show that the multi-scale contextual information needed for the accurate segmentation of thin vessels is sufficiently captured by depthwise separable convolutions in the MobileNetV2 model encoder, a compact ASPP module and a small decoder, rather than by very deep residual stacks like those employed in ResNet101-based models. The validation Dice loss (0.0644) and the validation IoU (0.8840) of the proposed framework were also compared with that of U-Net and DeepLabV3+ (ResNET101) to show that the deeper and larger backbone does not necessarily lead to better segmentation performance for fine vascular structures. Rather, the skip connection seems to ensure that there is enough low level spatial information preserved, and the ASPP module makes up for the lower capacity of the encoder by combining the information at multiple dilations. In addition to being accurate, the considerable reduction in trainable parameters, parameter storage and multi-add operations directly impact memory usage and inference speed, ultimately translating to lower memory footprint and faster inference which are essential constraints in real time clinical workflows. The proposed framework offers high competition accuracy with significantly less computation compared to the baseline U-Net and ResNet101, making it much more practical for use on resource-constrained or edge-based ophthalmic screening devices, mobile diagnostic units, or point-of-care systems for scalable, real-world, retinal screening applications.

6. Conclusion

This study introduced a computational efficient framework, DeepLabV2+ with MobileNetV2 as backbone to segment retinal vessels in fundus images accurately. In order to do this, the proposed framework replaces the conventional ResNet101 encoder with light weight inverted residual blocks and then uses a compact ASPP and depthwise separable decoder, which brings down the trainable parameters to just 3.205 million, the storage of parameters to 12.82 MB, and the multi-add operations to 4.03 GB, which is a massive reduction compared to U-Net and DeepLabV3+ (ResNet101). Even with this huge efficiency improvement, the proposed model outperformed the others in terms of Dice loss and IoU results, and also gave the qualitative best and most continuous vessel boundaries. The results show that it is not necessary to suffer from high segmentation error and high calculation complexity, and the network with only lightweight encoders can achieve the performance of the much larger network on the fine-grained medical image segmentation task. This proposed framework provides a practical and deployable solution for real-time edge computing based ophthalmic screening applications, especially in resource constrained clinical environment. Future works will focus on quantization-aware training, hardware optimization and validation in more fundus imaging datasets and pathologies to

further validate the clinical generalizability and reliability of the proposed lightweight segmentation framework in clinical settings.

References

[1] C. Chen, J. H. Chuah, and R. Ali, "Retinal vessel segmentation in fundus images using convolutional neural network," in Proc. 2021 Int. Conf. High Perform. Big Data Intell. Syst. (HPBD&IS), 2021, pp. 261–265.
[2] A. Galdran, A. Anjos, J. Dolz, et al., "State-of-the-art retinal vessel segmentation with minimalistic models," Sci. Rep., vol. 12, p. 6174, 2022.
[3] K.-W. Huang, Y.-R. Yang, Z.-H. Huang, Y.-Y. Liu, and S.-H. Lee, "Retinal vascular image segmentation using improved U-Net based on residual module," Bioengineering, vol. 10, no. 6, p. 722, 2023.
[4] R. K. Sidhu, J. Sachdeva, and D. Katoch, "Segmentation of retinal blood vessels by a novel hybrid technique: PCA and CLAHE," Microvasc. Res., vol. 148, p. 104477, 2023.
[5] S. Ghosh, M. Kundu, and M. Nasipuri, "Retinal vessel segmentation in fundus image using low-cost multiple U-Net architecture," in Artif. Intell. Med. Data, Springer, 2023.
[6] F. Tian, Y. Li, J. Wang, and W. Chen, "Blood vessel segmentation of fundus retinal images based on improved Frangi and mathematical morphology," Comput. Math. Methods Med., vol. 2021, Art. no. 4761517, 2021.
[7] B. Almarri, B. N. Kumar, H. A. Pai, S. B. Khan, F. Asiri, and T. R. Mahesh, "Redefining retinal vessel segmentation: Empowering advanced fundus image analysis with the potential of GANs," Front. Med., vol. 11, p. 1470941, 2024.
[8] F. D. Hernandez-Gutierrez, E. G. Avina-Bravo, M. A. Ibarra-Manzano, J. Ruiz-Pinales, E. Ovalle-Magallanes, and J. G. Avina-Cervantes, "Retinal vessel segmentation based on a lightweight U-Net and reverse attention," Mathematics, vol. 13, no. 13, p. 2203, 2025.
[9] M. S. Gargari, M. H. Seyedi, and M. Alilou, "Segmentation of retinal blood vessels using U-Net++ architecture and disease prediction," Electronics, vol. 11, no. 21, p. 3516, 2022.
[10] Z. Deng, W. Gao, Z. Gong, et al., "A fundus image dataset for AI-based artery-vein vessel segmentation," Sci. Data, vol. 12, p. 1298, 2025.
[11] X. He, T. Wang, and W. Yang, "Research on retinal vessel segmentation algorithm based on a modified U-shaped network," Appl. Sci., vol. 14, no. 1, p. 465, 2024.
[12] Q. Xie, X. Li, Y. Li, J. Lu, S. Ma, Y. Zhao, and J. Zhang, "A multi-modal multi-branch framework for retinal vessel segmentation using ultra-widefield fundus photographs," Front. Cell Dev. Biol., vol. 12, p. 1532228, 2025.
[13] Wanjari, H., Gaus, D. A., Bhoyar, U., Gehani, K. K., & Prasad, S. (2025). Overcoming Unimodal Challenges: A Survey of Multi-Modal Fusion for Mobile Interfaces. International Journal of Recent Advances in Engineering and Technology, 14(3s), 159–165. https://doi.org/10.65521/intjournalrecadvengtech.v14i3s.1685.
[14] A. Khandouzi, A. Ariafar, Z. Mashayekhpour, et al., "Retinal vessel segmentation: A review of classic and deep methods," Ann. Biomed. Eng., vol. 50, pp. 1292–1314, 2022.
[15] P. Marti-Puig, K. M. Kapllani, and B. Ayala-Márquez, "A novel approach to retinal blood vessel segmentation using Bi-LSTM-based networks," Mathematics, vol. 13, no. 13, p. 2043, 2025.
[16] G. B. Kande, M. R. Nalluri, R. Manikandan, et al., "Multi-scale multi-attention network for blood vessel segmentation in fundus images," Sci. Rep., vol. 15, p. 3438, 2025.
[17] Xiao-Long, N. (2026). Hybrid Convolutional Attention Models for Intelligent Cardiac Risk Assessment. International Journal on Advanced Computer Engineering and Communication Technology, 15(2), 23–28. Retrieved from https://journals.mriindia.com/index.php/ijacect/article/view/3374.
[18] P. Harjule, M. Delu, R. Kumar, and P. Nkomozepi, "A mathematical framework for retinal vessel segmentation: Fractional Hessian-based curvature analysis," Fractal Fract., vol. 10, no. 4, p. 246, 2026.
[19] M. R. Prabhu, J. R. Arunkumar, R. Anusuya, G. Charulatha, and N. N, "Efficient segmentation of retinal blood vessels in the fundus images using morphological image processing," Int. J. Basic Appl. Sci., vol. 14, no. 6, pp. 222–227, 2025.
[20] Z. Zeng, J. Liu, X. Huang, K. Luo, X. Yuan, and Y. Zhu, "Efficient retinal vessel segmentation with 78K parameters," J. Imaging, vol. 11, no. 9, p. 306, 2025.
[21] A. Varma and M. Agrawal, "Thin vessel segmentation in fundus images using attention U-Net and modified Frangi filtering," Biomed. Signal Process. Control, vol. 99, p. 106842, 2025.
[22] B. Y. Wei, "A retinal fundus image segmentation approach based on the segment anything model," Math. Model. Eng. Probl., vol. 11, no. 10, pp. 2817–2822, 2024.

[23] N. R. Panda and A. K. Sahoo, "A detailed systematic review on retinal image segmentation methods," J. Digit. Imaging, vol. 35, no. 5, pp. 1250–1270, 2022.
[24] D. Lamrani, H. Cherradi, A. Elasri, H. Moujahid, S. Shawki, B. Cherradi, and L. Bahatti, "Enhanced retinal blood vessels segmentation using deep learning and residual network," Intell.-Based Med., vol. 12, p. 100263, 2025.
[25] Y. Zhong, T. Chen, D. Zhong, and X. Liu, "Vessel segmentation in fundus images with multi-scale feature extraction and disentangled representation," Appl. Sci., vol. 14, no. 12, p. 5039, 2024.
[26] Shimbre, N., & Solanki, R. K. (2026). Comprehensive Review: Analysis of Use of Multimodal Deep Learning Models in Medical Diagnosis System. Intelligent Systems Using Semiconductors for Robotics and IoT, 111-118.
[27] Latke, V. (2026). Computational Models of Retinal Neuron Responses to Visual Stimuli. Journal of Intelligent Decision Making and Information Science, 3(3s), 1171-1178.
[28] Verma, N., Varma, S., Chafle, R. R., Mashalkar, V., Kumar, A., & Gandhi, Y. (2024, December). Real-Time AI Systems for Monitoring Cancer Treatment Responses. In Proceedings of the 6th International Conference on Information Management & Machine Intelligence (pp. 1-10).
[29] Mane, D., Ashtagi, R., Suryawanshi, R., Kaulage, A. N., Hedaoo, A. N., Kulkarni, P. V., & Gandhi, Y. (2024). Diabetic retinopathy recognition and classification using transfer learning deep neural networks. Traitement du Signal, 41(5), 2683.