IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication X 00, 0000, date of current version Sept 21, 2024.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Advancing Retinal Vessel Segmentation with Diversified Deep Convolutional Neural Networks

TANZINA AKTER TANI[1], (Student Member, IEEE), and JELENA TEŠIĆ[1], (Senior Member, IEEE)
[1]Department of Computer Science, Texas State University, San Marcos, TX 78666, USA
Corresponding author: Tanzina Akter Tani (e-mail: tanzinatani@txstate.edu).

ABSTRACT Retinal vessel segmentation is crucial for the diagnosis and monitoring of ophthalmic illnesses. Deep learning algorithms have been extensively utilized in automated segmentation to improve effectiveness and efficiency. In this paper, we introduce the use of DeepLabV3+ architecture to segment retinal blood vessels and enhance its performance by applying six different deep neural network backbones: ResNet50, DenseNet121, MobileNetV2, Xception, Xception with lower features (XceptionLF), and Xception lower features with overlapping regions (XceptionLFOR) patches. We also demonstrate the robustness of placing the Swin Transformer into the DeepLabV3+ model. The integration of XceptionLF and XceptionLFOR into the pipeline enhances the semantic segmentation of retinal images by enabling the merging of global and patch-specific features along with features from both lower and higher resolutions. The enhancements enable our proposed best model, XceptionLFOR, to obtain an 89.23% dice score, which represents a significant advancement in applying advanced deep-learning techniques for medical imaging. The XceptionLFOR model achieves a higher performance and better $F1$ score (0.49%) over the state-of-the-art for the FIVES benchmark evaluation despite using lower image resolution (256 resolution patches from 512-resolution images). The use of lower resolution balances computational efficiency with enhanced performance, enabling faster processing and deployment in resource-constrained environments. The findings in this paper point in the right direction in improving semantic segmentation for retinal vessel images, and they highlight the potential to improve early diagnosis and treatment outcomes for ocular illnesses.

INDEX TERMS DeepLabV3+, DCNN, FIVES dataset, Retinal vessel segmentation, Swin Transformer.

I. INTRODUCTION

The segmentation of retinal blood vessels is essential for the early detection of ophthalmic illnesses, such as diabetic retinopathy, hypertension, muscular degeneration, and glaucoma. It is critical for reducing eyesight impairment and improving patient outcome [1]. Blood vessel segmentation allows the quantitative analysis of retinal blood vessels, such as vessel diameter, branch pattern, and changes over time. This information is helpful in following the progression of disease and determining the efficacy of treatment [1]. Retinal blood vessel analysis can also reveal information about one's overall cardiovascular health. Changes in vessel characteristics may signal the development of some cardiovascular illnesses, making it an essential diagnostic tool [2]. Manual blood vessel segmentation is a time-consuming and labor-intensive procedure, whereas automated semantic segmentation approaches lessen the burden on medical practitioners while offering robustness and objectivity. Image segmentation is a computer vision task that groups pixels in an image. Semantic segmentation assigns semantic labels to the pixel groups to further identify the shapes and objects in the image [3]. The profound convolutional neural network breakthrough helped advance the field in the past couple of years; as outlined in Section II, there is room for improvement as the Deep Convolutional Neural Networks (DCNNs) struggle to deal with multiscale information and explain missing vessels [4]; and the thin, low-contrast vessels provide additional challenges since typical segmentation approaches may lose spatial information [4].

Contributions: Medical imaging frequently uses a variety of semantic segmentation models, including U-Net, SegNet, FCNs, and DeepLabV3+. In these segmentation models, backbones are crucial because they serve as feature extractors by converting input images into high-dimensional representations using the spatial and contextual information needed for accurate segmentation. Here, we propose to modify and advance the retinal semantic segmentation with different backbones based on the DeepLabV3+ methodology. Selecting and

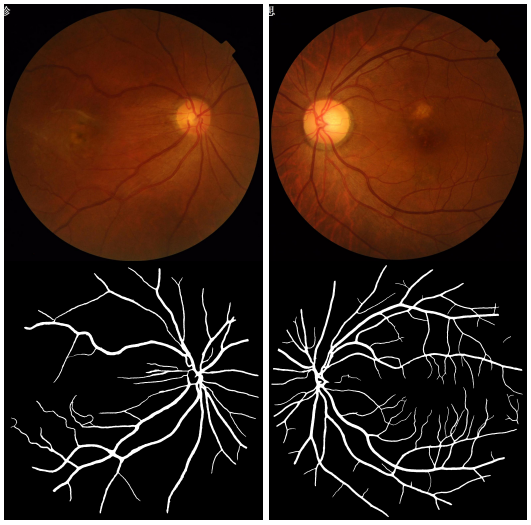

FIGURE 1. Two samples from the FIVES dataset (top) and the ground truth (bottom).

fine-tuning several backbones allowed us to leverage their varied strengths, which improved the overall segmentation performance and robustness. Paper contributions are:
Introduced the novel *XceptionLF* backbone that combines Xception's low features with a convolutional block to extract features at different levels of semantic resolution.
Introduced the novel *XceptionLFOR*, which extends the XceptionLF by incorporating patches from overlapping regions of images, thus including both global and grid local features in the segmentation process.
Integrated four DCNN backbones, DCNNs-ResNet50, MobileNetV2, DenseNet121 and Xception and demonstrated their robustness in the DeepLabV3+ integration for the task.
Demonstrated the effectiveness and robustness of the Swin Transformer within DeepLabV3+ on the latest benchmark for the retinal vessel segmentation task for the first time.
Our proposed XceptionLFOR model outperforms the state-of-the-art on the FIVES benchmark [5], achieving a 0.49% higher $F1$ score despite using lower-resolution images (256 resolution patches from 512 resolution images). This balance of computational efficiency and effectiveness enables faster processing and deployment. The supplementary code can be found in the GitHub repository as *RetinaVseg* [6].
Overview: Related work and state of the art are outlined in Section II, the proposed methodology is outlined in Section III, proof of concept setup is presented in Section IV, experimental results and discussion is presented in Section V and finally the conclusion and future work in Section VI.

II. RELATED WORK

In this section, we present a comprehensive review of the related work in vessel segmentation. The methods evaluated on multiple benchmarks are primarily reported using DRIVE [24] in related works, as it contains the most significant number of samples. Only one paper does not have the metric for DRIVE thus we report their findings on the STARE [25]. The focus is on how well the models perform in terms of true positives, false positives, and false negatives, not how well the model classifies the background (true negatives). Thus, here we report the F1 score, sensitivity, MCC, and AUC scores if the authors reported them. We group the related work by the principal methodology used in (1) U-Net-based Networks in Section II-A, (2) Transformer-based Networks in Section II-B, and (3) DeepLabV3+ -based Networks in Section II-C. Only two existing methods report evaluation cores on the FIVES [5] benchmark, and we outline the methods in Section II-D.

A. U-NET BASED NETWORKS

The integration of the U-Net architecture with Residual Networks and RCNN has shown improved performance without increased computational complexity [7] with 81.71% F1 score, 77.92% sensitivity, 97.84% AUC for the DRIVE [24] vessel segmentation benchmark.

The *Vessel-Net* incorporates the inception-residual convolutional blocks into a U-like encoder-decoder architecture that has four supervision paths to maintain rich features during optimization [8]. Vessel-Net achieved 80.38% sensitivity and 98.21% AUC for the DRIVE.

The SA-UNet is a lightweight network for vessel segmentation that incorporates a spatial attention module and structured dropout convolutional blocks to improve feature refinement and prevent overfitting [10]. The SA-UNet struggles with noise and fine details with an F1 score of 82.63%, MCC score of 80.97%, and sensitivity of 82.12% for the DRIVE.

The *FR-UNet* is a deep learning approach for segmenting thin, low-contrast vessels using full image resolution and a multi-resolution convolution interactive mechanism [11], with 83.16% F1 score, 83.56% sensitivity and 98.89% AUC for the DRIVE.

The *DA-Res2UNet* model utilizes Res2blocks for multiscale information extraction and dual attention for better focus and uses a GAN-based image generator to explain the segmentation process and identify errors [3]. The DA-Res2UNet model has issues with low-quality data, and it gave 82.77% F1 score, 81.28% MCC, 81.50% sensitivity, and 98.77% AUC for the DRIVE.

The *MDUNet* model is the transformer-based model that combines cross-dimensional transformation and self-attention mechanisms [13]. MDUNet utilizes an encoder-decoder structure with Dense Blocks, HR Blocks, and ASPP modules for rich feature extraction and fusion. The model's F1 score is 81.52%, sensitivity is 80.22%, and AUC is 98.39% for the DRIVE. The MDUNet does not perform well on lesion fundus images with significant morphological differences.

The dual-encoder dynamic-channel GCN retains edge information with a dynamic channel graph convolutional network and improves feature synthesis across channels [14]. The proposed approach enhances the fine detail of vessel segmentation with added computational complexity, and its

DRIVE F1 score is 82.88%, MCC is 80.32%, sensitivity is 83.50%, and AUC is 98.66%.

The *FES-Net* architecture balances performance and computational cost by reducing trainable parameters. The FES-Net processes input images using four prompt convolutional blocks (PCBs) and a shallow up-sampling approach, bypassing conventional image enhancement [16]. The FES-Net is computationally intensive as it achieved 83.10% F1 score, 82.89% sensitivity, and 98.32% AUC on the DRIVE.

The *MFI-Net* is a multiscale Feature Interaction Network, a U-shaped network with Pyramid Squeeze-and-Excitation and Coarse-to-Fine modules [32]. The MFI-Net is computationally intensive, and its F1 score is 83.15%, sensitivity is 81.70%, and AUC is 98.36% for DRIVE.

The *LEA U-Net* is a U-Net that incorporates a Local Feature Enhancement module and attention mechanisms for better retinal vessel segmentation with 82.3% F1 score and 79.83% sensitivity on the DRIVE dataset [34].

The *DTI* method is the dual-threshold iterative method that improves vessel connectivity by identifying weak vessel pixels. The reliance on the manual selection of thresholds may introduce subjectivity and variability for the DTI [2]. Researchers improved the segmentation of retinal vessels when they integrated the Bi-FPN network into U-net and applied multiple preprocessing techniques to avoid overfitting, achieving 80.64% sensitivity and 97.87 AUC for the DRIVE.

The *AACA-MLA-D-UNet* uses multi-level attention and adaptive channel attention to improve segmentation [15] to achieve 83.03% F1 score, 80.70% MCC, 80.46% sensitivity, and 98.27% AUC for DRIVE. The AACA-MLA-D-UNet model has difficulty in accurately detecting and maintaining the connectivity of vessel structures, leading to potential fragmentation and incomplete segmentation.

The *PLVS-Net* uses prompt blocks with asymmetric and depth-wise convolutions to segment retinal vessels [33] to score the sensitivity of 82.50% and AUC of 98.15% on the DRIVE benchmark. The PLVS-Net is efficient and lightweight. Still, it may struggle with more complex tasks.

The *MRC-Net* combines multi-resolution features, recurrent fusion, and adversarial learning for high performance at low cost [35]. MRC-Net sensitivity is 82.50%, and AUC is 98.25% for DRIVE.

The *AFFD-Net* is a dual-decoder network with multiscale Feature Extraction, Feature Fusion, and Attention-enhancing modules to segment retinal vessels [36]. The AFFD-Net sensitivity is 84.19%, and the AUC is 98.41% on the DRIVE.

The *CRAUNet* used DropBlock, multiscale Fusion Channel Attention, and a cascaded U-Net for refined retinal vessel segmentation [37]. CRAUNet achieved 83.02% F1 score, 79.54% sensitivity, and 98.30% AUC for DRIVE but struggled with thin vessels in noisy areas.

The *LadderNet* is the best and most resource-intensive of the six improved U-Net models in [38] with 79.84% F1 score, 78.27% sensitivity, and 98.02% AUC for the STARE benchmark (no results reported for DRIVE).

The *ColonSegNet* lightweight-based model is efficient for low-end hardware but may struggle with complex vessel structures [39], and its MCC is 79.60%, sensitivity is 84.91%, and AUC is 98.50% for DRIVE.

The *BFMD SN U-net with GCI-CBAM* incorporates BFMD SN U-net with the switchable normalization for faster convergence, block feature map distortion to prevent overfitting, and GCI-CBAM for better feature refinement [17]. The model is susceptible to block size and distortion probability in BFMD, and its F1 score is 82.60%, sensitivity is 83%, and AUC is 98.71% for the DRIVE.

B. TRANSFORMER BASED NETWORKS

The *G2ViT* model is a Vision Transformer-based model that integrates convolutional and graph neural networks, U-net encoder, Vision Transformer, and MEFA and MLF2 modules to edge information and feature fusion [18]. The method relies on pre-extracted graph structures, limiting its applicability when such data is unavailable, and its F1 score is 81.42%, sensitivity is 84.36%, and AUC is 98.97% on the DRIVE.

The *HT-Net* combines convolutional neural network and transformers, proposes an effective self-attention mechanism as well as novel Feature Fusion and Refinement Blocks [19]. The transformers lack inductive bias and cannot be effectively trained on small datasets without extensive annotated data. HT-Net F1 score is 82.79%, sensitivity is 82.56%, and AUC is 98.72% for DRIVE.

The *PVT-FCN* model combines the Pyramid Vision Transformer (PVT) and FCN-Transformer models [20]. The PVT-FCN model is computationally expensive as it employs an ensemble model. It effectively captures discriminative features, resulting in an F1 score of 82.56% and a sensitivity of 81.66% for DRIVE.

The *TiM-Net* incorporated transformers into the M-Net for capturing long-range relationships and employs the dual-attention mechanism for noise reduction [21], resulting in a sensitivity of 78.05% and AUC of 96.81% for DRIVE benchmark. The TiM-Net model lost some vessel details due to continuous upsampling.

The *PCAT-UNet* is a U-Net-based model that combines convolution branches with patches-based transformers to improve retinal vessel segmentation [22]. This hybrid technique benefited from both local and global feature extractions at the expense of increased computations with the integration of the transformer and CNN components. The PCAT-UNet F1 score is 81.60%, sensitivity is 85.76%, and AUC is 98.72% for DRIVE.

C. DEEPLABV3+ BASED NETWORKS

The *SVSN* model is a lightweight CNN using an encoder-decoder structure with spatial pyramid pooling inspired by DeepLabV3+ architecture [9]. SVSN model captures multi-scale contextual information without pre- or post-processing, effectively segmenting both large and tiny retinal vessels [9]. The SVSN model sensitivity is 82.95%, AUC is 97.10% for DRIVE. The performance of the model heavily depends on

extensive data augmentation, which may affect its generalizability to unseen data.

The *DeepLabV3+* network is used for single-channel images and two-class pixel classification to enhance the blood vessel segmentation [12]. Also, images are preprocessed with CLAHE and data augmentation for better training and refine the output using morphological closing operations. The proposed approach achieved 80% sensitivity and 65.5% dice on the DRIVE. Note that additional post-processing is still needed to refine results and remove noise [12].

D. MODELING FOR FIVES BENCHMARK

The FIVES Benchmark [5] is a relatively new benchmark, and it has been used only in two studies to date.

The *SCOPE* is a graph-based neural network that preserves continuity and connectivity in vessel segmentation. It is the first known study to evaluate FIVES dataset segmentation with 85% dice and 85% sensitivity scores [1].

The *SGAT-Net* integrates CNN with the transformers and includes the Stimulus-Guided Adaptive Module (SGA-Module) for extracting detailed features, the SGAP-Former for enriching contextual embeddings, and the SGAFF for effective feature fusion [23]. The study used 2048×2048 image resolution to divide it into 512×512 patches for FIVES dataset analysis, which increases the computational complexity as well as the correlation of test and train datasets. The SGAT-Net achieved 90.51% F1 score and 91.62% sensitivity for FIVES and 83.32% F1 score, and 86.32% sensitivity for the DRIVE.

III. METHODOLOGY

Here, we enhance the backbone of the semantic segmentation architecture by using modified Deep Convolutional Neural Networks (DCNN) and incorporating Swin Transformers into the encoder. These enhancements represent novel approaches for retinal vessel segmentation.

A. STATE OF THE ART COMPONENTS

DeepLabV3+ [26], introduced in 2018, is an advanced semantic segmentation model that improves performance by integrating numerous components from prior generations. It starts with an encoder using a backbone network, which extracts higher features from input images utilizing pre-trained weights from massive datasets like ImageNet. The model effectively captures multiscale contextual information by using atrous (dilated) convolutions that expand the receptive field without increasing parameters. The Atrous Spatial Pyramid Pooling (ASPP) module is a critical component of the architecture, as it performs parallel atrous convolutions at varying rates and adds global context via image-level features. To improve segmentation results, particularly object boundaries, DeepLabV3+ includes a decoder module that combines low-level features from the backbone with high-level ASPP features, followed by a series of 3x3 convolutions and upsampling to the original image scale. Owing to its capacity to extract precise feature representations from datasets, pretrained **DCNN models** like ResNet, Xception, DenseNet, and MobileNet are frequently applied as backbone models in segmentation tasks. The robust feature extraction capabilities of these backbones significantly enhance the performance of segmentation models. For example, ResNet's [27] deep architecture and residual connections capture detailed and hierarchical features, making it a familiar candidate for segmentation methods like U-Net and DeepLab. By utilizing depthwise separable convolutions, Xception [28] offers rich feature maps and efficient computation that improves segmentation in models like DeepLabV3+. The DeepLabV3+ uses MobileNet or DenseNet as a backbone. DenseNet [29] enhances gradient flow and feature reuse, yielding parameter-efficient and effective segmentation models. MobileNet [30], designed for lightweight and efficient performance, is ideal for real-time applications and resource-constrained deployment. The **Swin Transformer** as a backbone of the network has been shown to improve the segmentation results [31].

With this transformer method, local and global contexts can be captured effectively. The self-attention mechanism employed in the Swin Transformer is defined in Eq. 1.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

where Q is query, K is key, V is values, and d_k is the critical dimension. Thus, the Swin Transformer model computes self-attention over progressively larger areas due to the hierarchical structure formed by combining windows.

B. PROPOSED MODELING PIPELINES

The DeepLabV3+ model maintains efficacy while reducing computing complexity by using depth-wise separable convolutions. The model produces cutting-edge results on benchmark datasets such as PASCAL VOC 2012 and Cityscapes while balancing effectiveness and efficiency [26]. The pretrained backbones outlined in Section III-A can be used to ensure high-quality and reliable segmentation results across a range of applications by reducing training time and processing resources while also improving performance, particularly in a scenario with limited data.

The deep learning segmentation model DeepLabV3+ with different Deep Convolutional Neural Network (DCNN) backbones was employed as a baseline. Each of the following DCNNs is instantiated without the top layer, utilizing pretrained ImageNet weights. It was customized to receive input through a specific input tensor using a particular size tensor of 256×256×3.

Baseline backbones: DeepLabV3+ model with ResNet50 backbone in the encoder extracts high-level feature from the 'conv4_block6_2_relu' layer outputs. The Atrous Spatial Pyramid Pooling (ASPP) block then processes the features to capture extensive semantic information. Conversely, the decoder processed lower-level features from the 'conv2_block3_2_relu' layer of ResNet50 and then passed through a convolutional block, which used 48 filters with

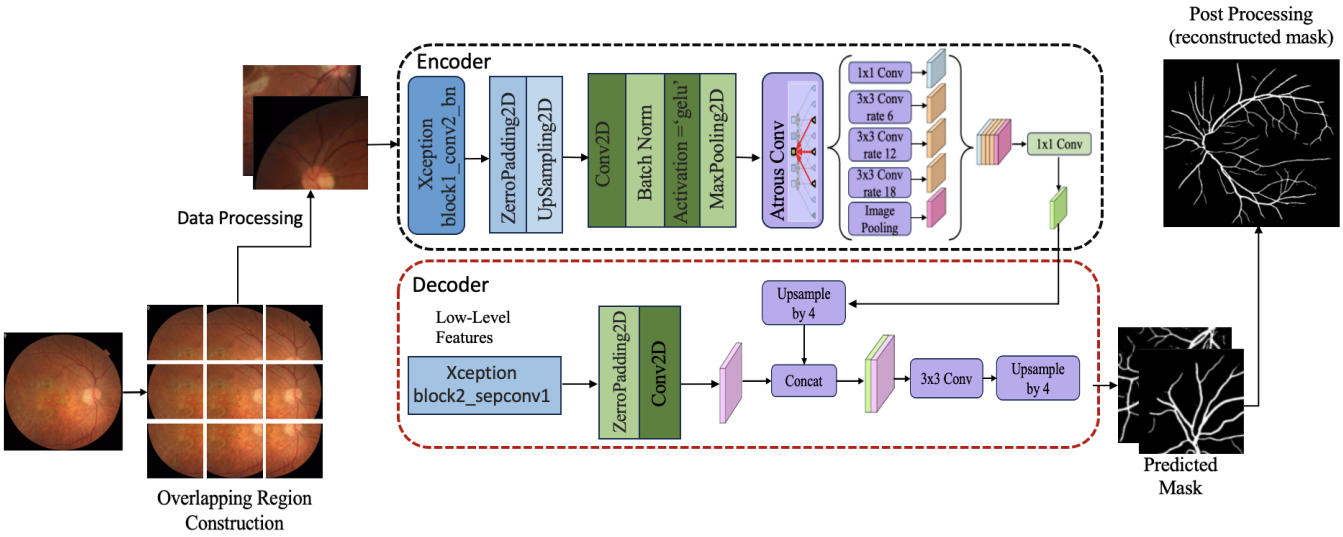

FIGURE 2. Modified DeepLabV3+ Architecture: Integration of Overlapping 256-Resolution Patches from 512-Resolution Images with Xception Low-Level Features and Additional Convolution Block, Followed by Post-Processing for Enhanced Retinal Segmentation

a 1x1 kernel size to refine and integrate textural and edge details. In the DeepLabV3+ model with MobileNetV2 backbone, the encoder collected high-level features from the 'block_4_project_BN' layer, which were subsequently processed using Atrous Spatial Pyramid Pooling (ASPP). Concurrently, the decoder extracted lower-level features from the 'block_2_project_BN' layer and passed through a convolution block equipped with 50 filters of 1x1 kernel. In the DenseNet121-based DeepLabV3+ model, the encoder leveraged higher-level features from the 'conv5_block6_1_conv' layer, while the decoder refined lower-level features from the 'conv2_block4_1_conv' layer. Next, the convolution block with 48 filters of size 1x1 enhances the detail and the texture captured by the features. The encoder extracted higher-level features from the 'block11_sepconv1' layer of the Xception backbone. In contrast, the decoder processed lower-level features from the 'block3_sepconv1' layer, enhanced by ZeroPadding2D, and refined through a convolutional block with 50 filters of 1x1 size.

C. IMPROVING THE BACKBONE DCNN

The main contribution to the DeepLabV3+ backbone is the introduction of the Xception backbone with Lower Features, *XceptionLF*, and the Xception backbone with Lower Features & Overlapping Regions, *XceptionLFOR*. In the XceptionLF backbone, only the lower-level features from Xception were used in both the encoder and decoder. In the encoder, lower-level features were initially extracted from the 'block1_conv2_bn' layer as illustrated in Figure. 2. These features were then first padded using ZeroPadding2D and then upscaled with UpSampling2D to enhance and restore detail. They were further processed through a Conv2D layer with 50 filters of size 3x3, followed by batch normalization and GELU activation, with a subsequent MaxPooling2D step to refine the feature representation. After processing, the Atrous Spatial Pyramid Pooling (ASPP) module processes the features. In parallel, in the decoder, 'block2_sepconv1' outputs lower-level features as an input to a convolution block using 48 filters of size 1x1. This systematic approach to optimizing lower-level features ensures detailed and effective segmentation across the model. The model training with *XceptionLFOR* backbone experiment used the previously described Xception backbone with Lower Features. XceptionLFOR does not use the entire image as an input, and 256X256 overlapping regions are extracted from the image. Section IV outlines the patch processing steps. The model's predictions on the test set were also performed using these patches. During post-processing, the patches were reassembled into complete images. Figure. 2 illustrates the processing pipeline.

D. SWIN TRANSFORMER AS ENCODER

We integrated the Swin Transformer into the encoder backbone of the DeepLabV3+ pipeline, as illustrated in Figure. 3. The encoder processed the input image by dividing it into 16x16 patches and embedding them. Then, we added two blocks of Swin Transformer before merging the patches. Swin Transformer, implemented from scratch, could capture local and global features through hierarchical self-attention. A block consisting of Conv2D-BatchNorm-GELU activation-Maxpooling layers further refined the features before the decoder extracted low-level features from a pre-trained Xception block. Next, the pipeline merged the features with the encoder output and upscaled through convolutional and upsampling layers. The final output is a prediction mask for segmenting retinal vessels. Combining both approaches leveraged their strengths: the Swin Transformer captures long-range dependencies and multiscale features, while the Xception decoder efficiently extracts features and ensures precise segmentation.

In summary, we have introduced seven different back-

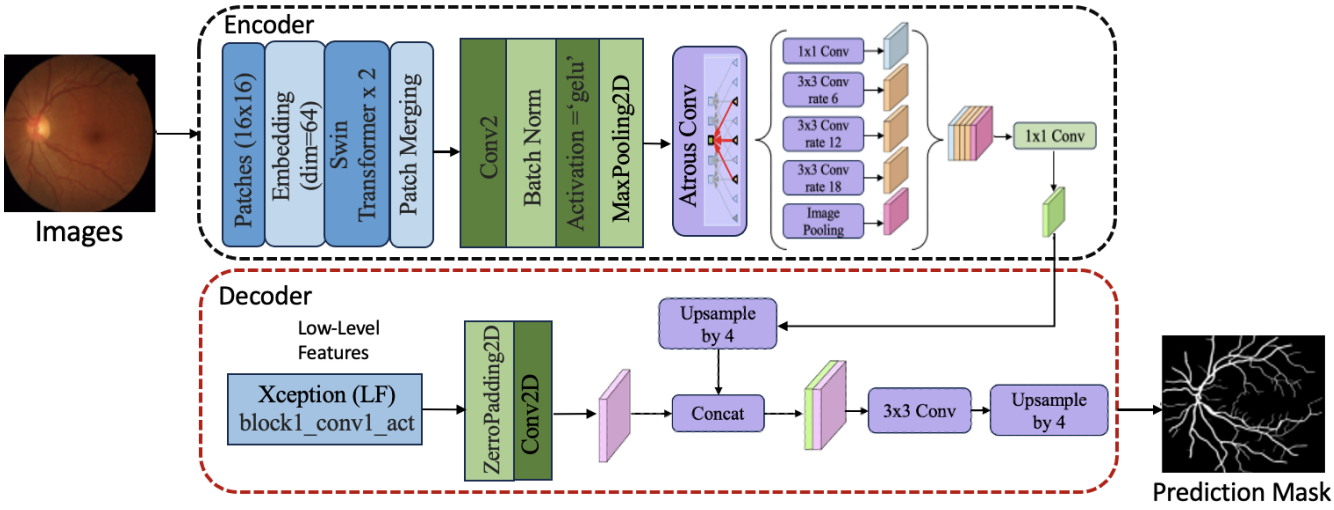

FIGURE 3. **Proposed Model Architecture: Integration of Swin Transformer Blocks with Additional Convolution Block in DeepLabV3+ Encoder for Enhanced Retinal Segmentation**

bones, four of which followed the DeepLabV3+ architecture and utilized the higher features in the encoder and lower features in the decoder. Our novel contributions include the XceptionLF model, which integrated lower features of Xception with an additional convolution block, and the XceptionLFOR model, an extension of XceptionLF that incorporated patches from overlapping regions to enhance performance. Additionally, we experimented with Swin Transformer with an additional convolution block as a new variant encoder for the DeepLabV3+ model in context to retinal vessel segmentation.

IV. PROOF OF CONCEPT

A. BENCHMARKS AND DATA PROCESSING

A Fundus Image Dataset for AI-based Vessel Segmentation (FIVES) [5] is one of the most extensive benchmarks for retinal blood vessel image segmentation. The dataset consists of 800 RGB fundus images with pixel-wise annotation. The dataset provider divided the original photos and their corresponding ground truth images into training and testing sets using a 75:25 split. Each image in the dataset has a resolution of 2048×2048 and is labeled either healthy or Age-related Macular Degeneration (AMD), Diabetic Retinopathy (DR), or glaucoma. In our experiments, we focused on binary segmentation instead of segmenting based on individual classes. Figure. 1 illustrates the samples of FIVES datasets.

The training dataset of FIVES [5] is split into train and validation sets 80:20. As a result, the train set has 480 images, the validation set has 120 images, and the test set has 200 images. We normalized the images and masks to [0,1] and applied a few augmentation techniques, such as random horizontal and vertical flips and random rotation within the -30 to 30 degrees range. We resize the image input to 512×512 pixels for XceptionLFOR and to 256×256 resolution for all other approaches. For the XceptionLFOR approach, we extract the smaller patches of 256×256 pixels using a stride of 128 pixels to ensure that each patch overlaps half of its predecessor, creating overlapping regions in both horizontal and vertical directions. This results in $\frac{512-256}{128} + 1 = 3$ patches per dimension and a total of nine patches per image. The approach captures sufficient context around the edges of each patch through these overlapping regions, which is crucial for tasks like image segmentation, where preserving edge details is paramount. Both images and their corresponding masks underwent the same procedure. This strategy ensures that the training and validation data are well-prepared, allowing the model to recognize patterns across different scales and conditions efficiently, as illustrated in the initial step of Figure. 2. We also evaluate our approaches without re-training on the 20 RGB fundus training set of the DRIVE dataset [24] to see how the models perform for the domain adaptations scenario. in an entirely new unseen dataset.

B. MODEL EVALUATION METRICS

First, we introduce the quantitative performance metrics for model evaluations. The following formulas use the number of true positives TPP), true negatives TNN), false positives FP), and false negatives FN) in the evaluation set:

Dice Score is an essential metric for comparing the similarity of two segmentations as it measures the overlap between the predicted and ground truth masks. The formula is:

$$\text{Dice} = \frac{2 \cdot |\text{Ground Truth} \cap \text{Predicted Mask}|}{|\text{Ground Truth}| + |\text{Predicted Mask}|}$$

Precision measures the ratio of accurately predicted positive observations to the total predicted positives. **Sensitivity** is the proportion of accurately predicted positive observations to total observations in the actual class. The precision and sensitivity are two critical metrics for medical image analysis. Higher precision reduces the false positives, which is essential to avoid any false alarms and give more reliable clinical

diagnoses. On the other hand, higher sensitivity reduces the false negative, which is crucial for not missing diseases.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Sensitivity} = \frac{TP}{TP + FN}$$

F1 score is the harmonic mean of precision and sensitivity, which balances the trade-off between them:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$$

Accuracy measures the ratio of correctly predicted observation over total observation, and it is used as a measure of performance during the model training:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

MCC: The Matthew Correlation Coefficient (MCC) balances binary classification by considering all the confusion matrix categories. The MCC provides a more comprehensive evaluation for vessel segmentation as the pixels of vessels are a lot fewer than background pixels.

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

C. HYPER PARAMETER TUNING

We tune parameters by searching for the best batch size, learning rate, and optimizer on the FIVES [5] dataset. The hyper-parameter space ranges in batch sizes (8,16), learning rates (0.0001, 0.001), and optimizers (Adam, AdamW, RMSprop) using the ResNet50 model with 50 epochs. We have obtained the best dice score for batch size 8, Adam optimizer, and 0.001 learning rate. Next, we conduct all experiments with 100 epochs using these hyper-parameters with the dice loss and binary cross-entropy loss combined as the loss function. We have used the Google Colab Pro service, which has 53G.B.B of system RAM and 22.5 GB of dedicated GPU RAM (L4 GPU configuration), for the hyper-parameter tuning.

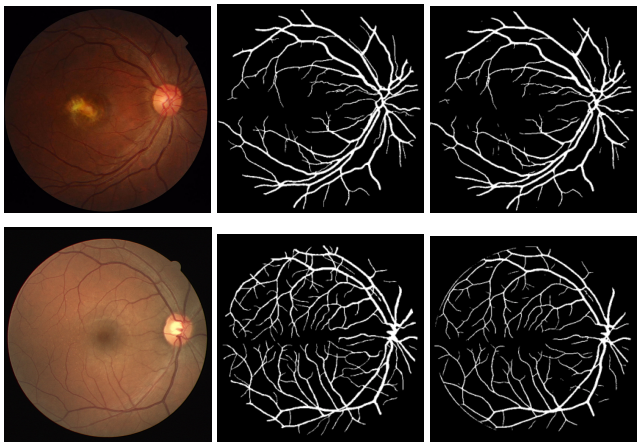

FIGURE 4. Examples of the prediction mask when compared to ground truth for the FIVES (top) image and the DRIVE (bottom) image.

TABLE 1. Training Time of Different Approaches

Model	Training Time (minutes)
ResNet50	24.25
MobileNetV2	23.63
DenseNet121	23.17
Xception	23.86
SwinTransformer	91.27
XceptionLF	48.95
XceptionLFOR	234.62

D. COMPUTING

The evaluation of training times across different backbones in the DeepLabV3+ segmentation model given in Table 1 indicated substantial differences influenced by their architectural complexities. Base models such as ResNet50, MobileNetV2, DenseNet121, and Xception demonstrated moderate performance with short and comparable training durations—24.25, 23.63, 23.17, and 23.86 minutes, respectively—highlighting their efficiency with the low computational load. The *XceptionLFOR* pipeline used patches from overlapping regions and thus requires a considerably longer training time of almost 4 hours. Still, the XceptionLFOR approach showed the best results on the FIVES benchmark, which confirms that more complex computation does lead to improvement in modeling. Compared to SwinTransformer, which took around 91 minutes to train, the *XceptionLF* modeling displayed a more balanced approach by keeping the training under 50 minutes without compromising its performance. This efficiency made the XceptionLF particularly appealing in circumstances where resource constraints are a concern.

These findings illustrate a crucial trade-off in deep learning architectures: while advanced features can enhance model capabilities, they also demand more computational resources. Therefore, selecting an appropriate backbone architecture is essential for effectively balancing performance with resource constraints.

V. EXPERIMENTAL RESULTS AND DISCUSSION

All the experimental methods were evaluated on the FIVES [5] and DRIVE [24] datasets for testing with different metrics. Each experiment in the FIVES test dataset employed a batch size of 8, except for XceptionLFOR, where a batch size of 1 is applied due to the need to reconstruct the prediction patch mask as a whole. Table 2 summarizes results on the 200 images from the FIVES test dataset.

TABLE 2. Performances of different approaches(FIVES dataset)

Backbone	Precision	Sensitivity	MCC	Dice
ResNet50	84.43	78.90	80.26	81.57
MobileNetV2	83.26	78.10	79.21	80.54
DenseNet121	86.72	77.53	80.70	81.86
Xception	88.23	78.33	81.93	83.01
SwinTransformer	90.21	81.61	84.78	85.62
XceptionLF	89.60	85.10	86.38	87.21
XceptionLFOR	**90.48**	**91.53**	**90.34**	**89.23**

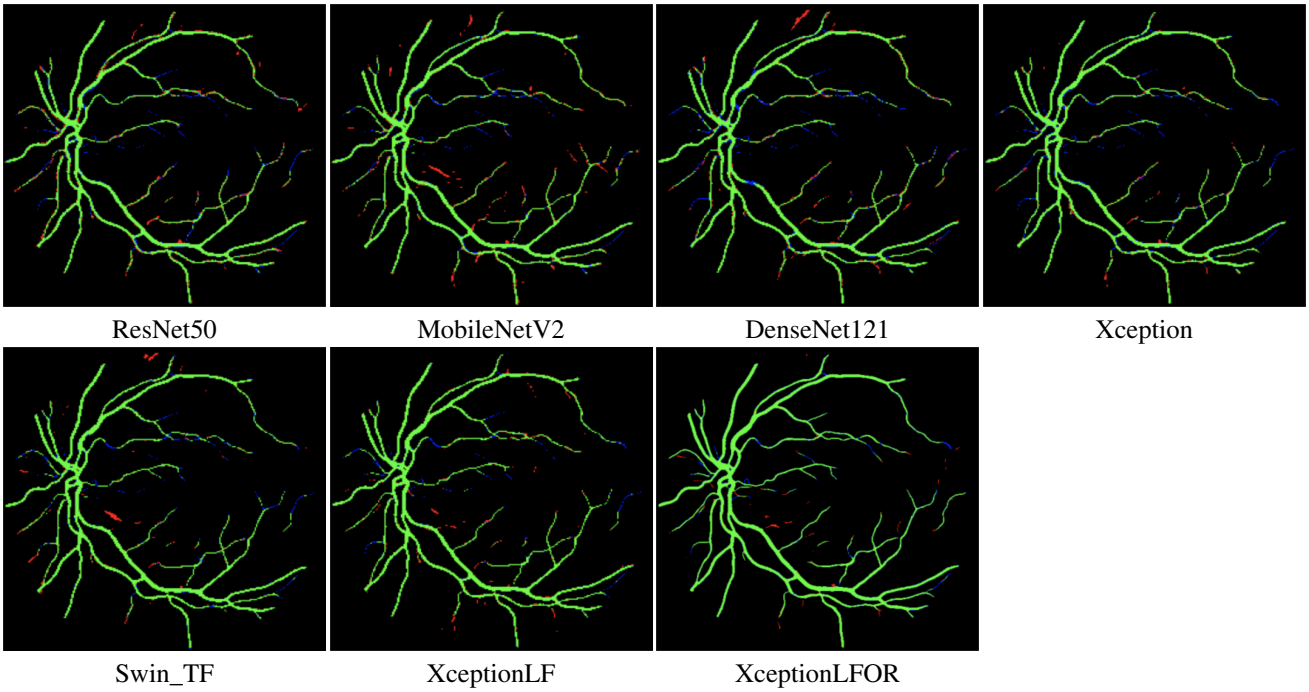

FIGURE 5. TP(green)-vessel correctly classified, FN(blue)-vessel classified as background, FP(red)-non-vessel classified as a vessel with different backbones.

MobileNetV2, ResNet50, and DenseNet121 gave the lowest performances among all the experiments of different backbones, scoring approximately 81% dice score. Compared to all DCNN backbones with higher features in the encoder and low features in the decoder, Xception performed the best with 83.01% dice score. The Swin Transformer in the encoder as the backbone is comparable to the XceptionLF and XceptionLFOR with the 85.62% dice score, almost the same as XceptionLF. The reason for the improved results is that the Swin Transformer can capture both local and global contexts, and the self-attention mechanism can help focus on the essential features of vessels. The performance improved a lot after using Xception with lower features and an additional convolution block (XceptionLF) instead of higher features in the encoder. Dice score and MCC gave an improvement of 4.2% and 4.45%, respectively.

There are several reasons for providing better results for retinal vessel segmentation. Lower features capture fine-grained information such as edges, textures, and small structures that are more critical for segmenting the thin and intricate structures of retinal vessels accurately. On the other hand, high-level features are adequate for understanding the overall image context, but they cannot capture the fine-grained information essential for the vessel segmentation task. Furthermore, low features can preserve more spatial information. The additional convolution block in the encoder also helped to further process and refine those lower features by enhancing their representation. In contrast, higher features usually involve pooling and other operations that can degrade spatial resolution, resulting in the loss of critical details required for precise segmentation. Nevertheless, utilizing low features helped to develop a lightweight model that reduces the computation complexity of training.

The image patching input to the XceptionLF model resulted in a 2.02% improvement in dice score for the test set. Both precision and sensitivity improved a lot for XceptionLFOR by providing 90.48% and 91.53% scores, respectively. Also, the model showed a 90.34% MCC score by outperforming all the other approaches, which indicates that this model is handling imbalance problems more effectively. The reason for the excellent performance is that patch-based training enabled more extensive and varied augmentation, which can help the model reduce over-fitting. The model performed better in segmentation since it is more adaptable to changes in the appearance of the vessel due to its robustness. The 256-resolution patching module captures the local details better to segment delicate vessel structures.

TABLE 3. Performances of different approaches (DRIVE dataset)

Backbone	Precision	Sensitivity	MCC	Dice
ResNet50	79.99	66.01	70.37	72.33
MobileNetV2	77.11	66.15	68.95	71.21
DenseNet121	79.71	65.73	70.01	72.05
Xception	81.39	65.07	70.54	72.32
SwinTransformer	78.18	**72.67**	73.14	75.33
XceptionLF	81.83	69.51	73.32	75.17
XceptionLFOR	**84.06**	71.69	**75.72**	**77.34**

Table 3 outlines the results of the applied methods on the DRIVE dataset for testing. Base models like ResNet50, MobileNetV2, and DenseNet121 gave competitive precision

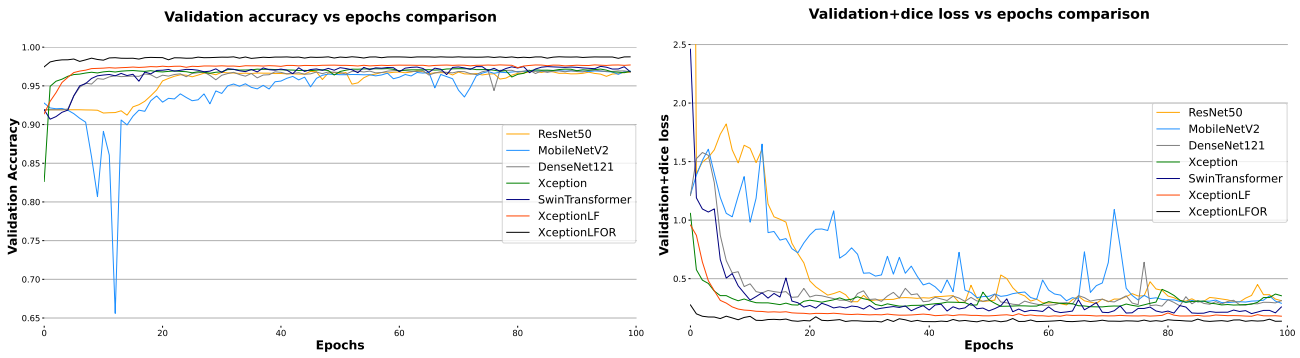

FIGURE 6. **Validation accuracy (y-axis) vs. epochs (x-axis) on the left, and validation+dice loss (y-axis) vs. epochs (x-axis) on the right, for proposed approaches (legend).**

but lag in sensitivity and dice score. On the other hand, the XceptionLFOR model outperforms all other models across most metrics, achieving the highest precision 84.06%, MCC 75.72%, and dice score 77.34%. It also demonstrates a sensitivity of 71.69%, second only to the SwinTransformer, which has the highest sensitivity of 72.67% but lower performance in other metrics. The model showed over 5.01% dice score improvement against the ResNet50 backbone baseline. The XceptionLF model has MCC 73.32% as it stands out for its superior performance in most evaluation metrics.

The evaluation of the DRIVE dataset showed that the model is capable of performing well in different unseen retinal blood vessel datasets. Figure. 4 illustrates the output visualization result of FIVES and DRIVE using the best experiment model, XceptionLFOR. Indicate that the predicted mask is efficient in segmenting vessels from images.

A. ABLATION STUDY

The proposed methodologies were implemented sequentially, effectively incorporating most of the ablation study within the process. Initially, the baseline model was developed, followed by architectural modifications. Finally, region-based image patches were integrated into the model, allowing for a comprehensive evaluation of each modification's impact on overall performance.

In the baseline model, we experimented with Xception using higher-level features in the encoder, as proposed in DeepLabV3+, achieving a dice score of 83.01%. In the modified version, XceptionLF, the encoder was replaced with lower-level features and a new convolutional block, which improved the dice score to 87.21%, shown in Table 2. Removing the convolutional block from XceptionLF resulted in a slight decrease in the dice score to 86.77%, along with a reduction in sensitivity to 83% and an increase in precision to 91.07%. In medical analysis, where sensitivity is crucial for correctly identifying diseases (reducing false negatives), this convolutional block plays a vital role in balancing precision and sensitivity. The MCC was 86%, close to those achieved by XceptionLF with the convolutional block. In the final experiment, XceptionLFOR, we incorporated image patches into the XceptionLF model, leading to improvements across all evaluation metrics (Table 2). Sensitivity increased by 6.43% and MCC by 3.96% over the XceptionLF model, demonstrating that adding region-based image patches significantly enhances retinal vessel segmentation.

B. MODEL PERFORMANCE

Figure. 6 shows the accuracy and loss of validation data for the proposed methods for 100 epochs. The DenseNet121 and MobileNetV2 backbone models showed some fluctuations in the plots, which means for some batches, it led to a noisier training update. Therefore, these two models need more fine-tuning. For other experimental models, the validation accuracy and loss have shown more stability, which indicates the model is training well and effectively generalizing the validation set, and the model does not overfit.

To examine the generalization capabilities of our proposed models and prevent overfitting, we trained all models on the train and validation split sets without utilizing cross-validation. Based on initial performance measures, we chose the two best-performing models for additional validation. To properly validate these models, we performed four-fold cross-validation on 600 images from the original given FIVES trainset. This method allowed us to confirm that the high performance reported during initial training was not the result of overfitting but instead of the models' capacity to generalize well across diverse subsets of the data. The consistent dice scores found in cross-validation, as shown in Table 4, imply that these models are robust and perform well across different data splits.

TABLE 4. **Dice Scores from Four-Fold Cross-Validation for the Best Two Models for the FIVES benchmark.**

Model	Fold-1	Fold-2	Fold-3	Fold-4
XceptionLF	86.87	87.12	87.17	87.61
XceptionLFOR	92.33	92.21	92.27	92.23

We used the color map depicted in Figure. 5 to provide a visual comparison of the predicted FIVES test results from our experimental models: ResNet50, MobileNetV2, DenseNet121, Xception, Swin_TF, XceptionLF, and Xcep-

tionLFOR. For each model, we use three color codes: green for vessels correctly classified, blue for vessels mistakenly classified as background (false negatives, FN), and red for non-vessels incorrectly classified as vessels (false positives, FP). We observe a noticeable reduction in FP and FN errors as we move from the first image (ResNet50) to the last image (XceptionLFOR) of Figure. 5, indicating an improvement in model performance. The ResNet50 and MobileNetV2 model shows a considerable amount of red and blue, suggesting a higher rate of misclassification. DenseNet121 shows a reduction in these errors but still exhibits a significant number of misclassified pixels. The Xception and Swin_TF models further reduce these errors, with fewer red and blue pixels visible. XceptionLF shows even more improvement, but it is the XceptionLFOR model that provides the most refined results—the model significantly minimizes FP and FN errors, with a higher proportion of green pixels indicating correctly classified vessels. Figure. 5 also shows that there is a need to highlight areas where the model needs to do better. Overall, the visualization clearly demonstrates the superior performance of the XceptionLFOR model in detecting retinal vessels accurately but also points to the ongoing challenge of perfecting vessel segmentation.

C. COMPARISON WITH STATE OF THE ART

From the related work reviews, we found only two papers that experimented with vessel segmentation using the FIVES dataset. Table 5 shows the evaluation results comparing our best three proposed experimental models with these two papers.

TABLE 5. Comparative analysis with related works

Reference	Approach	ACC	SN	PRE	F1	Dice
Yeganeh et al. [1]	SCOPE	–	85	90	–	85
J. Lin et al. [23]	SGAT-NeT	**98.86**	**91.62**	–	90.51	–
Proposed Model1	SwinTransformer	98.06	81.61	90.21	85.70	85.62
Proposed Model2	XceptionLF	98.24	85.10	89.60	87.29	87.21
Proposed Model3	XceptionLFOR	98.76	**91.53**	**90.48**	**91.00**	**89.23**

The first paper [1] used 512×512 resolution to experiment with their SCOPE models. Their sensitivity score is higher than our Swin Transformer and XceptionLF model; however, our proposed models outperformed other metrics, such as precision and dice scores. Also, our XceptionLFOR outperformed their SCOPE model by a 4.23% dice score. The second paper [23] used 512×512 size of patches from whole images with 2048x2048 resolution to train their SGAT-Net model, which is higher image resolution than our experiment model training. However, our model still could give a higher F1 score (91%) than their model.

Next, we evaluate the effectiveness of all models on the unseen 20 images of the DRIVE dataset. The results are summarized in Table 3. We report over 6% improvement in the dice score (72.33%) over the reported DeepLabV3+ model with ResNet18 backbone (66%) on the same dataset [24]. Thus, we conclude that applying CLAHE for preprocessing and morphological operations for post-processing proved less effective compared to our fine-tuned complex model utilizing RGB images [12]. Another paper [9] introduced a lightweight model SVSN, adapted the idea from DeepLabV3+, and achieved 82.95% sensitivity on the DRIVE. Our XpeptionLFOR model achieved a decent score despite not training with the dataset. The overall comparison with related works showed our proposed experimental models could efficiently segment blood vessel images in previously unseen images.

VI. CONCLUSION AND FUTURE WORK

Retinal blood vessel segmentation is a valuable application in diagnosing ophthalmic diseases. However, due to limited sample availability and image complexity, it remains challenging to achieve efficient automation results using deep learning approaches. To address these challenges, researchers are continuously experimenting with different techniques. In our study, we have experimented with one of the most extensive vessel datasets, FIVES [5], and proposed seven different backbones utilizing the deep semantic segmentation model DeepLabV3+. The six models utilized the DCNN backbones for effective feature extraction. Also, from the related work reviews, it is found none of the studies applied a Swin Transformer with a DeepLabV3+ model for vessel segmentation. The Swin Transformer integrated into the encoder without pre-training delivered comparatively good results. Our best model, XceptionLFOR, with lower features for both encoder and decoder, achieved 90.34% MCC, and 89.23% dice score. The model also performed well with the unseen dataset DRIVE. Therefore, our proposed methodologies can efficiently segment retinal blood vessels, which will aid in the diagnosis of early eye diseases for patients.

While the models were trained on the FIVES dataset and tested on both FIVES and DRIVE datasets, their generalizability to other datasets or diverse real-world clinical settings needs further investigation. The specific characteristics of these two datasets may not fully represent the variability found in other retinal images or broader clinical settings, hence additional datasets need to be explored to confirm the models' robustness across different data sources. In the future, we plan to experiment with varying transformers with pre-training and attention mechanisms to evaluate their efficiency in retinal vascular segmentation. We have also applied minimal data preprocessing techniques. The proposed models can test different augmentation approaches and preprocessing procedures.

REFERENCES

[1] Y. Yeganeh, et al., "Scope: Structural continuity preservation for medical image segmentation," arXiv preprint *arXiv:2304.14572*, 2023.
[2] K. Ren, et al., "An improved U-net based retinal vessel image segmentation method," *Heliyon,Heliyon*, vol. 8, no. 10, e11187, 21 Oct, 2022.
[3] R. Liu, et al., "DA-Res2UNet: Explainable blood vessel segmentation from fundus images," *Alexandria Engineering Journal* vol. 68, pp. 539-549, 2023.
[4] R. Wang, et al., "Medical image segmentation using deep learning: A survey." *IET image processing*, vol. 16, no. 5, pp. 1243-1267, 2022.

[5] K. Jin, et al., "Fives: A fundus image dataset for artificial Intelligence based vessel segmentation," *Scientific data*, vol. 9, no. 1, pp. 475, 2022.
[6] T. Akter Tani, "Retinal vessel segmentation," Available: https://github.com/DataLab12/RetinaVseg, Accessed on: June 30, 2024.
[7] M. Z. Alom, et al., "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," arXiv preprint *arXiv:1802.06955*, 2018.
[8] Y. Wu, et al., "Vessel-Net: Retinal vessel segmentation under multi-path supervision," *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China*, October 13–17, 2019, Proceedings, Part I 22, pp. 264-272. Springer International Publishing, 2019.
[9] T. M. Khan, F. Abdullah, S. S. Naqvi, M. Arsalan and M. A. Khan, "Shallow Vessel Segmentation Network for Automatic Retinal Vessel Segmentation," *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, 2020, pp. 1-7.
[10] C. Guo, et al., "Sa-unet: Spatial attention u-net for retinal vessel segmentation," *2020 25th international conference on pattern recognition (ICPR)*, pp. 1236-1242. IEEE, 2021.
[11] W. Liu, et al., "Full-Resolution Network and Dual-Threshold Iteration for Retinal Vessel and Coronary Angiography Segmentation," in *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4623-4634, Sept. 2022.
[12] M. C. S. Tang, S. S. Teoh and H. Ibrahim, "Retinal Vessel Segmentation from Fundus Images Using DeepLabV3+," 2022 *IEEE 18th International Colloquium on Signal Processing & Applications (CSPA)*, Selangor, Malaysia, pp. 377-381, 2022.
[13] A. Jayachandran, S. Ratheesh Kumar, and T. Sudarson Rama Perumal, "Multi-dimensional cascades neural network models for the segmentation of retinal vessels in color fundus images," *Multimedia Tools and Applications*, vol. 82, no. 27, pp. 42927-42943, 2023.
[14] Y. Li, Y. Zhang, W. Cui, B. Lei, X. Kuang and T. Zhang, "Dual Encoder-Based Dynamic-Channel Graph Convolutional Network With Edge Enhancement for Retinal Vessel Segmentation," in *IEEE Transactions on Medical Imaging*, vol. 41, no. 8, pp. 1975-1989, Aug. 2022.
[15] Y. Yuan, L. Zhang, L. Wang, and H. Huang. "Multi-Level Attention Network for Retinal Vessel Segmentation," in *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 312-323, Jan. 2022.
[16] T. M. Khan, et al., "Feature Enhancer Segmentation Network (FES-Net) for Vessel Segmentation," *2023 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 160-167, IEEE, Nov 28. 2023.
[17] S. Deari, I. Oksuz, and S. Ulukaya, "Block Attention and Switchable Normalization Based Deep Learning Framework for Segmentation of Retinal Vessels," in *IEEE Access*, vol. 11, pp. 38263-38274, 2023.
[18] H. Xu, and W. Yun, "G2ViT: Graph Neural Network-Guided Vision Transformer Enhanced Network for retinal vessel and coronary angiograph segmentation," *Neural networks: the official journal of the International Neural Network Society*, vol. 176, p. 106356, 2024.
[19] X. Hu, L. Wang, and Y. Li, "HT-Net: A hybrid transformer network for fundus vessel segmentation," *Sensors*, vol. 22, no. 18, p. 6782, 2022.
[20] L. Du, et al., "Deep ensemble learning for accurate retinal vessel segmentation," *Computers in Biology and Medicine* vol. 158, p. 106829, 2023.
[21] H. Zhang, et al., "TiM-Net: Transformer in M-Net for Retinal Vessel Segmentation," *Journal of Healthcare Engineering 2022*, no. 1, p. 9016401.
[22] D. Chen, et al., "PCAT-UNet: UNet-like network fused convolution and transformer for retinal vessel segmentation," *PloS one*, vol. 17, no. 1, p. e0262689, 2022.
[23] J. Lin, et al., "Stimulus-guided adaptive transformer network for retinal blood vessel segmentation in fundus images," *Medical Image Analysis*, vol. 89, p. 102929, 2023.
[24] M. Niemeijer, et al., "Comparative study of retinal vessel segmentation methods on a new publicly available database," *Medical imaging 2004: image processing*, vol. 5370, pp. 648-656. SPIE, 2004.
[25] A. D. Hoover, V. Kouznetsova and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," in *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203-210, March 2000.
[26] LC. Chen et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," *Proceedings of the European conference on computer vision (ECCV)*, pp. 801-818, 2018.
[27] K. He, et al., "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.
[28] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251-1258, 2017.
[29] G. Huang, Z. Liu, L. Van Der Maaten, "Densely connected convolutional networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700-4708, 2017.
[30] K. Dong, et al., "MobileNetV2 model for image classification," *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, pp. 476-480. IEEE, 2020.
[31] Z. Liu, et al., "Swin transformer: Hierarchical vision transformer using shifted windows," *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012-10022, 2021.
[32] Y. Ye, et al., "MFI-Net: Multiscale feature interaction network for retinal vessel segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4551-4562, 2022.
[33] M. Arsalan, et al., "Prompt deep lightweight vessel segmentation network (PLVS-Net)," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 2, pp. 1363-1371, 2022.
[34] J. Ouyang, et al., "LEA U-Net: a U-Net-based deep learning framework with local feature enhancement and attention for retinal vessel segmentation," *Complex & Intelligent Systems*, vol. 9, no. 6, pp. 6753-6766, 2023.
[35] TM. Khan et al., "Retinal vessel segmentation via a Multi-resolution Contextual Network and adversarial learning," *Neural Networks*, vol 165, pp. 310-320, 2023.
[36] X. Zijian, et al., "AFFD-Net: A Dual-Decoder Network Based on Attention-Enhancing and Feature Fusion for Retinal Vessel Segmentation," *IEEE Access*, vol. 11, pp. 45871-45887, 2023.
[37] F. Dong, et al., "CRAUNet: A cascaded residual attention U-Net for retinal vessel segmentation," *Computers in Biology and Medicine*, vol. 147, p. 105651, 2022.
[38] LK. Singh et al., "Deep-learning based system for effective and automatic blood vessel segmentation from Retinal fundus images," *Multimedia Tools and Applications*, vol. 83, no. 2, pp. 6005-6049, 2024.
[39] K. Aurangzeb, et al., "An efficient and lightweight deep learning model for accurate retinal vessels segmentation," *IEEE Access*, vol. 11, pp. 23107-23118, 2022.
[40] J. Odstrcilik, et al. "Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database," *IET Image Processing*, vol. 7, no. 4, pp. 373-383, 2013.

TANZINA AKTER TANI received her B.Sc. degree in Computer Science & Engineering from East West University, Bangladesh, 2021. She is currently pursuing her Ph.D. in Computer Science at Texas State University, USA. Her research interests are computer vision, deep learning, and image processing.

JELENA TEŠIĆ, PH.D. is an Associate Professor at the Department of Computer Science, Texas State University. She received her Ph.D. from the University of California Santa Barbara, CA, USA. Dr. Tešić has authored over 80 peer-reviewed scientific papers and holds six US patents. Her research focuses on unstructured data representation and analysis at scale, machine learning, and network science. She is a senior IEEE member.