



Retinal Vessel Segmentation Using CNN And U-Net Architecture

**Bhupathi Rayudu Inaganti¹, Prasanth Yenumula², Selvam K³, Jahnvi Bandaru⁴,
Varaha Varshini Naidu Polamarasetty⁵**

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram,
Guntur 522302, India¹⁻⁵

Abstract: The extent, width, curvature, and branching pattern of the retina's blood vessels, among other structural characteristics, plays a major role in the assessment of diseases related to diabetes and heart, hypertension. Through our research, we offer a procedure for segmenting the retinal vascular system from FCNs. From every retinal image, thousands of patches are gathered, and these patches are rotated before being sent via the network for Data Augmentation. For vessel segmentation, two FCNs are used: LadderNet Architecture and U-Net Architecture. Three well-known datasets: STARE, CHASE_DB1, and DRIVE are used to evaluate our methodology. When compared to the other previously mentioned methodologies, our strategy indicates better performance.

Keywords: Retina, Vessels, Convolutional, U-Net, Ladder-Net, Ophthalmology.

I. INTRODUCTION

The extent, width, curvature, and branching pattern of the retina's blood vessels, among other structural characteristics, plays a major role in the assessment of diseases related to diabetes and heart, hypertension [9]. Segmenting vessels is important ahead of collecting the structural characteristics of the retina's blood vessels from retinal pictures. Even for seasoned medical professionals, manually segmenting vessels in retinal scans is time-consuming and prone to mistakes. Therefore, it is essential to create accurate and automated techniques for separating blood vessels from retinal images. The analysis of biological pictures is one area of computer vision where image segmentation is very important, to assign labels to each pixel in an image. Direct feature learning from input data is possible with DCNN, allowing for a progressively complicated ladder of features.. Recent research has also used deep convolutional neural networks to segment retinal vessels. A 10-layer CNN was used by Melinscak et al. [8] to handle vascular segmentation. Liskowski and Krawiec [5] investigated a wide variety of architectural designs to draw attention to background information, with a seven-layer no- pooling Convolutional Neural Network appearing as the most successful. Fu et al.[2] 's method of coupling a standard seven-layer Convolutional Neural Network using a random conditional field that was reconstructed as an RNN allowed them to depict distant pixel relationships. Using operations on rotation to supplement input and SWT decomposition to provide additional channels to the network, the novel Oliveira et al.[11] architecture uses an encoder and decoder based on an FCN[7].

We offer a fully convolutional network-based method for segmenting retinal arteries in this research. Thousands of patches are extracted from each retinal image and then entered inside the network and rotated to provide data augmentation. For vessel segmentation, U-Net [14] and Ladder Net are two completely convolutional network designs. CHASE DB1[12], STARE[3], DRIVE [10] are three publicly available datasets used to assess the efficiency of our technique.

As key assessment criteria, accuracy and area under the ROC curve are used. When compared with all current approaches, our technology demonstrated improved performance. Following is the arrangement of the remaining parts. In Section 2, we detail our methods and the FCN [7] we used for vascular segmentation of the retinal. Section 3 of the experiment setup is discussed. Our findings are displayed in Section 4. Finally, Section 5 brings us to a conclusion.

II. METHOD

As retinal vessel segmentation is a binary classification problem, we have allocated labels for each vessel and background: the vessel as 1(one), and the background will be depicted as 0(zero). Our proposed approach for segmenting retinal arteries consists of three basic steps: 1) preprocessing 2) patch extraction 3) Data Augmentation 4) Classification.



A. DATASET

The technique we are using is rigorously trained and also tested using three popular datasets: the STARE [3] dataset is used for the training of our model and the CHASE DB1[12] and DRIVE [10] datasets are used to evaluate the generalization of our model. Sample photos from 3 datasets are shown in Figure 1, together with the corresponding field of view (FOV) masks and ground truth vessel segmentation.

Pixels beyond the field of view (FOV) are not scanned and are not essential for vessel segmentation, therefore we are only concerned with those inside the FOV. Initially, FOV masks will be present in the DRIVE[10] dataset but not in STARE[3], CHASE DB1[12] datasets. In all three datasets, there are two sets of hand segmentations available, and we utilize the first annotations as the underlying reference for all of them.

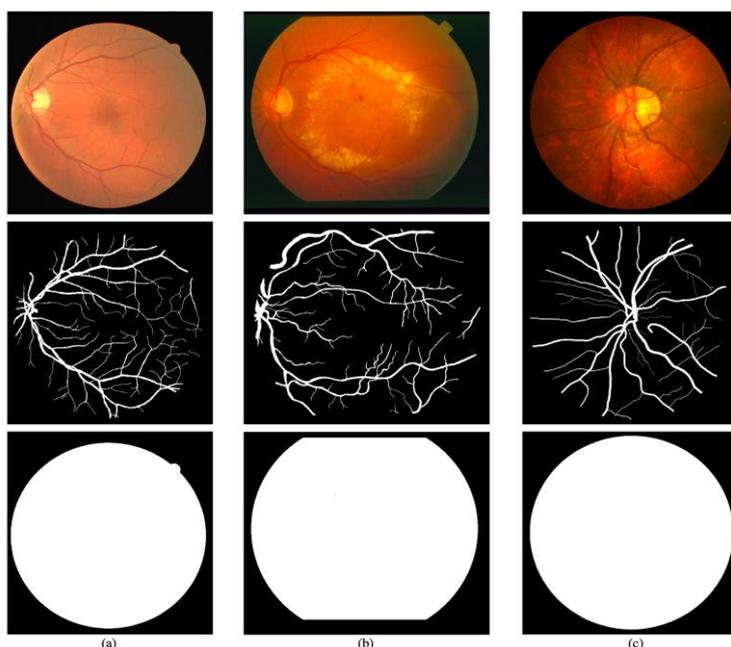


Fig.1 depicts pictures from three datasets 1)DRIVE[10] 2)STARE[3] 3) CHASE DB1 [12] and the masks of FOV.

The DRIVE[10] dataset has 40 photos with a resolution of 565 x 584. The STARE dataset contains twenty photographs for vascular segmentation tasks, with ten images classed as normal and ten as unwell. Each image has a resolution of 700 x 605 pixels. Twenty-eight photos from 14 children's two eyes were included in the CHASE DB1 collection. Every picture has a 999 × 960 resolution.

The test set is specifically used to evaluate the performance of our method since every one of the 40 photos in the training and test sets of the DRIVE dataset consists of 20 pictures. STARE[3] and CHASE DB1[12], on the other hand, do not discriminate between training and test sets. K-Fold Cross Validation is the method we utilized to validate our strategy. This method will split our dataset into K folds of the same size, with each fold acting as a test set and the remaining K folds acting as the algorithm's training set. The evaluation model will be trained and tested k-times for each fold and this outcome will be used for future processes. The model's final evaluation result was derived by averaging the outcomes. k is set to 5 in the STARE[3] dataset, and each fold has four images, two of which are sick and two of which are healthy.

B. PRE-PROCESSING

The CHASE DB1 [12] dataset has seven photos per fold, three of which are of the left and right eyes and four of which are of the other. In this dataset, K is set to four. The four stages in raw picture processing are grayscale conversion, standardization, CLAHE, and Gamma Correction. Figure 2. Shows the final product of pictures for each phase Grayscale Transformation As seen in Fig. 2(a), retinal images are typically RGB images with three channels. By converting these images to grayscale images, it is easier to proceed with subsequent preprocessing phases and minimizes computer expenditures during training (Fig. 2(b)). The Z-score approach is used to standardize each pixel of the grayscale photographs. Consolidating pixel value dispersion by Z-score standardization speeds up training and improves generalization.

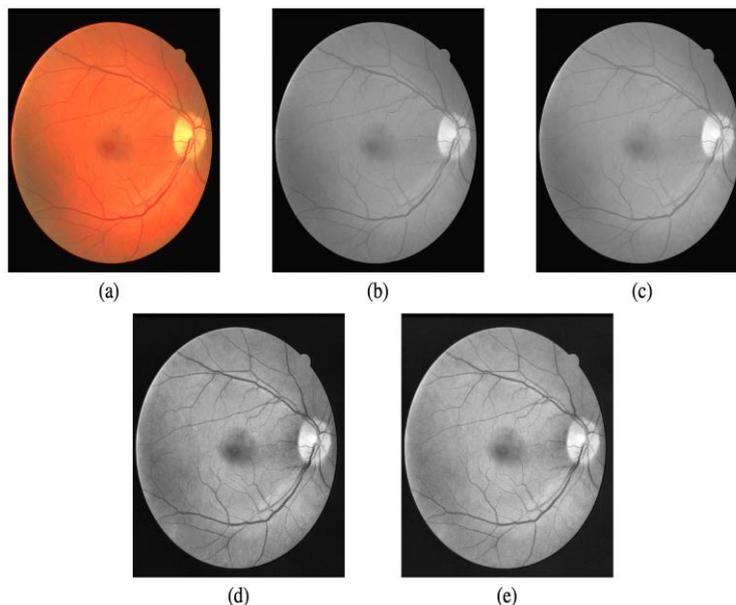


Fig.2 shows the outcomes of preprocessing at each stage: (A) displays the original retinal picture. (b) The conversion to grayscale. The symbol for standardising is (c). CLAHE is in (d). (e) Gamma correction.

Min-Max normalization is used to transform all the pixel values to the range within $[0.0,1.0]$. Fig.2 (c) displays the picture of the standardization outcome. Adaptive histogram equalization, also known as CLAHE may be applied to both real-world and medical images. It creates several histograms for various areas of the image and then spreads the brightness values using those histograms. Contrarily, AHE tends to exaggerate noise and contrast in almost constant regions of the image. By limiting contrast amplification, CLAHE lessens noise amplification[13]. Figure 2(d) displays the application of the output picture from CLAHE.

Gamma adjustment is a contrast-enhancing method that exponentially increases pixel intensity. Fig. 2 (e) displays the gamma correction's outcome.

C. DATA AUGUMENTATION

We built a collection of patches from pre-processed photos and put these patches, rather than whole pictures, into the net for training because each retinal vascular segmentation dataset only contains a small number of images (twenty in STARE, twenty-eight in CHASE DB1, and forty in DRIVE). Additionally divided into pictures, patches are tested for inference. In the training process, twenty thousand 48×48 patches are chosen at random from each image. Even though some patches may overlap, this strategy dramatically enhanced the training dataset, making it feasible and efficient to train the fully convolutional network.

Images are divided into 48 by 48 -pixel patches during testing and supplied into the network for inference. Cutting photos into patches that don't have any spaces or overlaps between them is the simplest method. When segmentation maps for the patches are combined with the segmentation mask for the entire image, bias is produced and certain pixels that were once inner pixels of the complete picture become boundaries of patches. For instance, the resulting segmentation mask occasionally appears to be made up of patches rather than being a whole picture segmentation. Additional information on this topic is provided in Section 4.

We used data augmentation in addition to patch extraction by rotating the patches since it is crucial for teaching the network the required invariance and resilience qualities. All of the patches, including the initially extracted and rotated patches are introduced to the network after being randomly shuffled, along with the ground truth segmentation mask patches.

D. U-NET AND LADDERNET ARCHITECTURE

The vessel segmentation is performed using U-Net[14] and Ladder Net[15], two entirely convolutional network architectures. Below, a detailed description of these two networks' architectures may be found. Our U-Net architecture integrates elements from both the completely twisted network presented by Oliveira et al. [11] and U-Net[14], combining the



benefits of both systems. In Fig. 3, the network architecture is displayed. The network’s decoder and encoder are shown as a contracting route on the left. This encoder-decoder design’s main goal is to mix information at various abstraction levels. A feature map is produced by each convolutional layer and is then enlarged by several pooling operators in the encoder. These maps are useful for identifying objects in images because of their vast and compressed feature sets, but localization- specific information is lost.

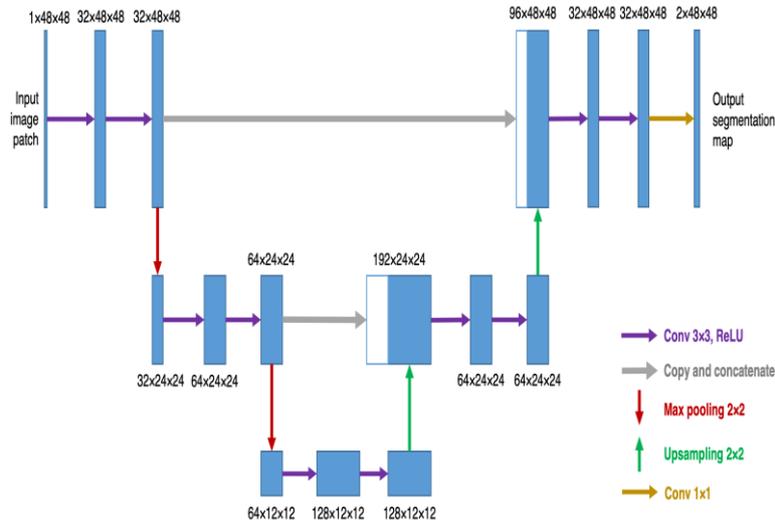


Fig.3 shows the architecture of U-Net.

Replicated feature maps are shown in white boxes. Different sorts of processes are indicated by different arrow colours. As a result, the decoder uses skip connections to merge the encoder’s lightweight, higher clarity low-level layers with rich, lower-clarity high-level feature maps. Improved localization performance may be achieved by obtaining spatial data and low-level information from the lightweight layers with increased clarity. The decoder structure’s main motivating factor is this. As described in the sections above, the input layer of our network is a one-channel patch created from retina images that have already undergone processing. The input layer is then carried by two expanding routes, an encoder and a decoder. Following each pair of sequential 3x3 convolutions, the contracting technique performs a downsampling step of stride 2 and a max pooling operation of 2x2. Following each maximum pooling, the contracting route’s feature channel count doubles, going from 32 to 64 to 128. Steps in the expanding path are as follows: two successive 3x3 convolutions (without padding), a 2x2 upsampling of the feature map, and a concatenation utilizing skip connections with the matching feature map from the contracting route. Half as many feature channels are included in the longer route between each upsampling. Because the expanding path is symmetric to the route, it then produces a U-shaped pattern that resembles the U-Net.. The activation function was the activation ReLu, which was used between two successive convolutional layers, one after the other, with a dropout rate of 0.2 to reduce overfitting. A group of U-Nets can be likened to LadderNet [15] , a Multi- Branch Fully Convolutional Network.

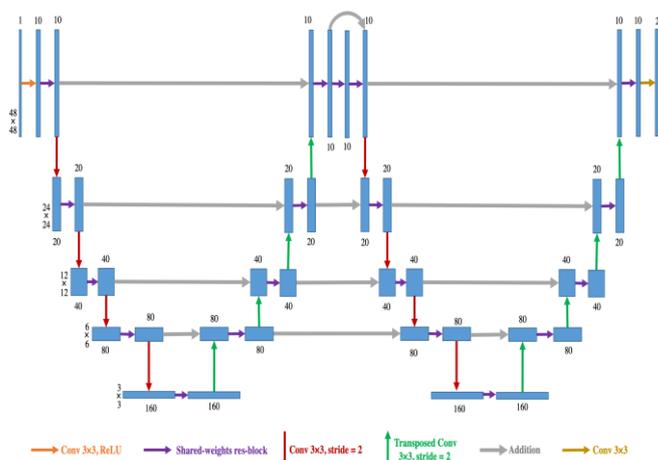


Fig. 4 Architecture of the LadderNet.



The blue box depicts the multi-channel feature maps. A number of channels will be represented at the top of the box. Each layer's width and height are displayed on the left. Unlike U-Net [14], which combines the feature maps from both encoder and decoder branches, feature maps are calculated by adding the values from two branches. LadderNet [15] may also be seen as an assortment of many FCN's [7]. Since LadderNet [15] has access to far more data than U-Net does, it can gather more sophisticated features. There will be more parameters and a more challenging training procedure with more encoder-decoder branches and U-Nets. To address this problem, we used shared-weight residual blocks in LadderNet (Fig. 5). The shared-weight residual block uses far fewer parameters than a traditional residual block to give accurate results by combining all the benefits.

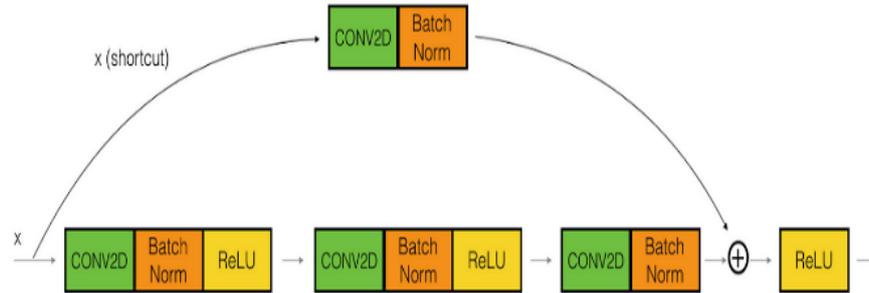


Fig. 5 shows a shared-weights residual block's intricate construction.

III. EXPERIMENTS

The suggested network was trained and tested using Keras2 with a TensorFlow3 backend. We obtained a pre-trained model for LadderNet from Zhuang's supporting GitHub repository.

A. TRAINING

As previously stated, 2000 patches are picked at random from each picture in the collection. Each patch has three rotations: Ninety(90), One-Hundred and Eighty(180), and Two Hundred and Seventy(270) degrees for data augmentation. The 20 photos in the training set are used to generate 160000 patches for the DRIVE dataset. This resulted in 128000 training patches from the original picture collection of the STARE dataset, which was split into a 4(four)-image test set and a 16(sixteen)-image training set. After splitting up the original picture collection into a testing set of seven photos and a training set of twenty-one images, a total of 168000 patches were made for the CHASE DB1 dataset. 10% of these extracted patches are used for validation across all trials, and 90% are used for the training purpose. The U-Net model which is trained will be trained using the three datasets for epochs of ten(10) and batch size of thirty-two patches. Training of the pre-trained LadderNet model will be done for four epochs with a thousand and twenty-four patch batch size.

B. METRICS OF EVALUATION

With the help of evaluation metrics, we have evaluated our model. The metrics we have used are shown in Fig.6.

Metric	Formula
True positive rate, recall	$\frac{TP}{TP+FN}$
False positive rate	$\frac{FP}{FP+TN}$
Precision	$\frac{TP}{TP+FP}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
F-measure	$\frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

Fig.6 The Evaluation Metrics used for reviewing the model performance.



IV. RESULTS

We only completely discuss our LadderNet results and briefly detail our U-Net results due to space limitations. Every image in Figs. 7,8 and 9 have two sets of rows with four columns each, and each of these four columns represents the segmentation level. The retinal picture is mostly shown in the first column, the ground and output probability maps are shown in the second and third columns, and the binary representation of the image is shown in the fourth column. Our outcomes are very accurate when compared with other methodologies.

A. Figures and Tables

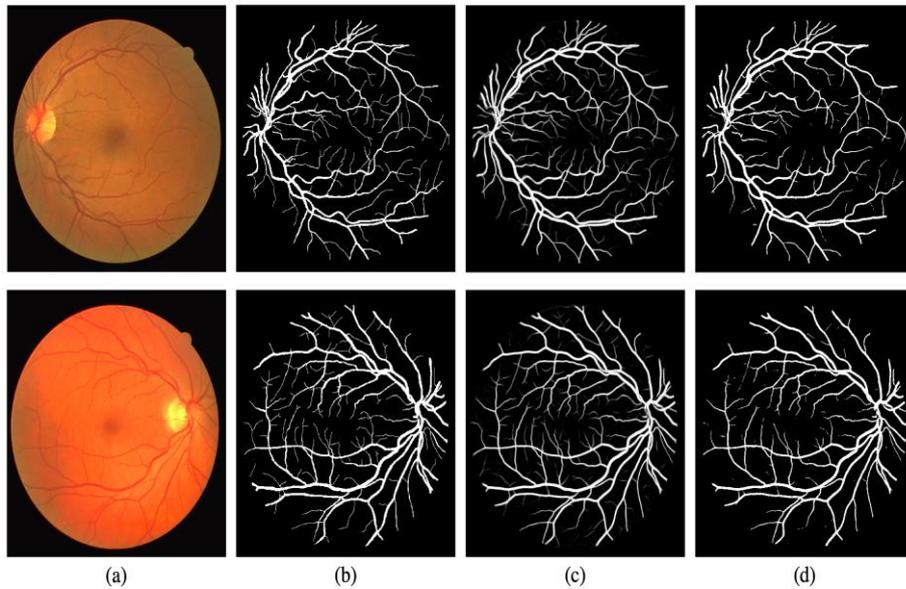


Fig.7 Results of Segmentation for the DRIVE Dataset using U-Net Architecture.

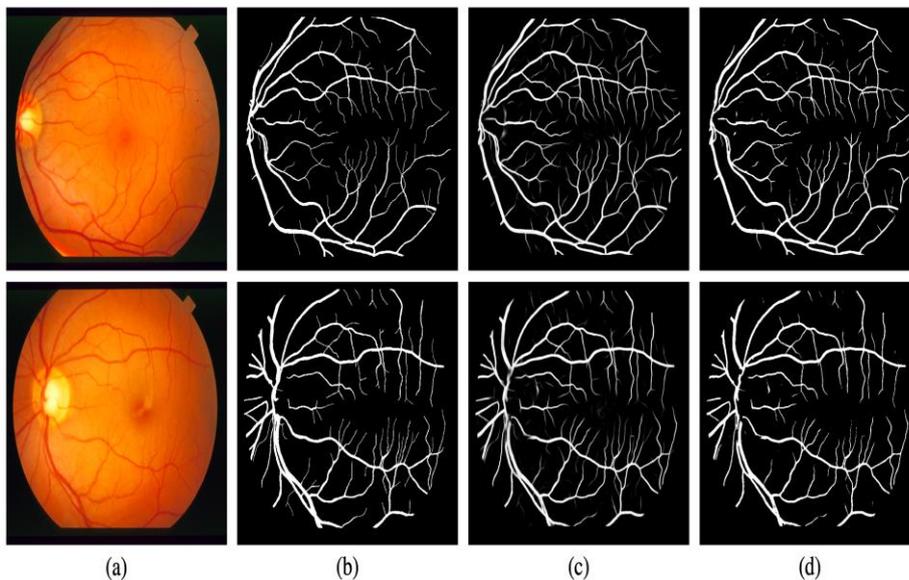


Fig.8 Results of Segmentation for the DRIVE Dataset using U-Net Architecture.



Table 1. Results of segmentation using the DRIVE dataset.

Method	F1 score	Sensitivity	Specificity	Accuracy	Area under curve
Li et al.[6]	-	0.7568	0.9815	0.9524	0.9732
Fu et al.[2]	-	0.7601	-	0.9520	-
M2U-Net [4]	0.8095	-	-	0.9634	0.9712
Oliveira et al.[11]	-	0.8030	0.9806	0.9573	0.9826
R2U-Net [1]	0.8173	0.7791	0.9815	0.9558	0.9782
My findings, U-Net	0.8455	0.7954	0.9936	0.9691	0.9891
My findings, LadderNet	0.8320	0.7954	0.9921	0.9643	0.9902

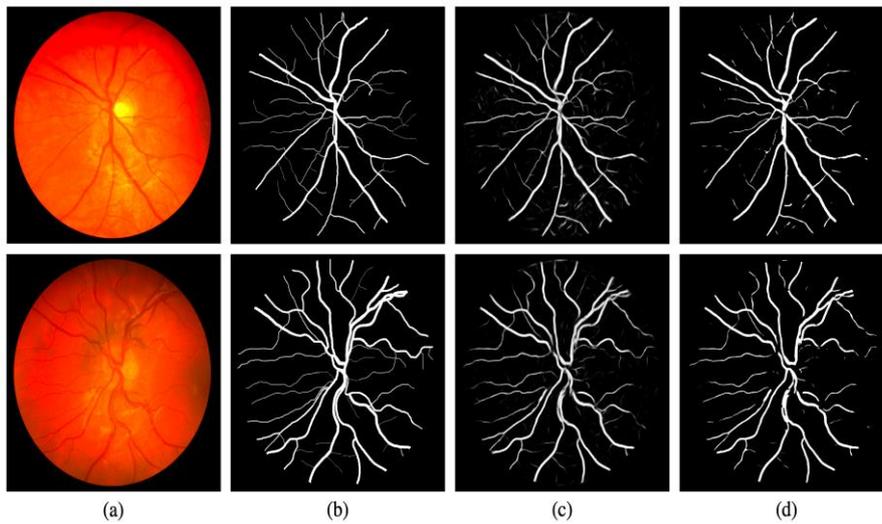


Fig.9 Results of Segmentation for the CHASE DB1 Dataset using U-Net Architecture.

Table 2. Displays the segmentation results on STARE obtained by applying multiple DL approaches.

Method	F1 score	Sensitivity	Specificity	Accuracy	Area under curve
Fu et al. [2]	-	0.7415	-	0.9587	-
Li et al.[6]	-	0.7724	0.9846	0.9620	0.9875
My findings, U-Net	0.8456	0.7983	0.9912	0.9751	0.9910
My findings, LadderNet	0.7932	0.7813	0.9813	0.9734	0.9743



V. CONCLUSION

This work displays the segmentation of retinal arteries using FCNs based on patches. All patches from retinal vision are extracted and arranged in networks to exploit data augmentation. To segment vessels, we used the U-Net [14] and LadderNet[15] network architectures. By comparing our methodology in segmenting the retinal vessels with all the methodologies that were stated before, we have achieved higher accuracy and performance of the model.

ACKNOWLEDGMENT

I want to thank Professor **Selvam K and Dr. Pavan Kumar** Head of Computer Science Engineering for their valuable insights and assistance.

REFERENCES

- [1] Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation.arXiv preprint arXiv:1802.06955 (2018)
- [2] Fu, Huazhu Xu, Yanwu Lin, Stephen Wong, Damon Liu, Jiang. (2016). DeepVessel: Retinal Vessel Segmentation via Deep Learning and Conditional Random Field. pp. 132-139. Springer (2016)
- [3] Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched lter response.IEEE Transactions on Medical Imaging 19(3), pp.203-210 (2000)
- [4] Laibacher, T., Weyde, T., Jalali, S.: M2u-net:Effective and efficient retinal vessel segmentation for resource-constrained environments. ArXiv preprintarXiv:1811.07738 (2018)
- [5] Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. IEEE transactions on medical imaging 35(11), pp.2369-2380 (2016)
- [6] Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., Wang, T.: A crossmodality learning approach for vessel segmentation in retinal images.IEEE transactions on medical imaging 35(1),pp.109-118 (2015)
- [7] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431-3440 (2015)
- [8] Melinscak, Martina Prentasic, Pavle Loncaric, Sven. Retinal Vessel Segmentation using Deep Neural Networks. VISAPP 2015 - 10th International Conference on Computer Vision Theory and Applications (VISAPP 2015) (2015)
- [9] M.M. Fraz, P. Remagnino, A. Hoppe, B.Uyyanonvara, A.R. Rudnicka,C.G. Owen, S.A. Barman,Blood vessel segmentation methodologies in retinal images – A survey,Computer Methods and Programs in Biomedicine, 108(1), pp.407-433 (2012)
- [10] Niemeijer, M., Staal, J., van Ginneken, B., Loog, M., Abramo, M.D.:Comparative study of retinal vessel segmentation methods on a new publicly available database. In: Medical imaging 2004: image processing. vol. 5370,pp. 648-656. International Society for Optics and Photonics (2004)
- [11] Oliveira, A., Pereira, S., Silva, C.A.: Retinal vessel segmentation based on fully convolutional neural networks. Expert Systems with Applications112,pp.229-242 (2018)
- [12] Owen, C.G., Rudnicka, A.R., Mullen, R., Barman, S.A., Monekosso,D.,Whincup, P.H., Ng, J., Paterson, C.: Measuring retinal vessel tortuosity in10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program. Investigative ophthalmology visual science 50(5),pp. 2004-2010 (2009)
- [13] Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A.,Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer vision, graphics, and image processing 39(3),pp. 355-368 (1987)
- [14] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medicalimage computing and computer-assisted intervention. pp. 234- 241. Springer (2015)
- [15] Zhuang,J.: Laddernet: Multi-path networks based on u-net for medical image segmentation. arXiv preprint arXiv:1810.07810 (2018)