An improved LinkNet-based approach for Retinal Image Segmentation

Srijarko Roy
Department of Computational Intelligence
SRM Institute of Science and Technology
sr8962@srmist.edu.in

Ankit Mathur
Department of Computational Intelligence
SRM Institute of Science and Technology
am9964@srmist.edu.in

S Velliangiri
Department of Computational Intelligence
SRM Institute of Science and Technology
vellians@srmist.edu.in

***Abstract*- The characteristics of Retinal Blood Vessels helps identify various eye ailments. The proper localization, extraction and segmentation of blood vessels are essential for the treatment of the eye. Manual segmentation of blood vessels may be error-prone and inaccurate, leading to difficulty in further treatment, and causing problems for both operators and ophthalmologists. We present a novel method of semantic segmentation of Retinal Blood Vessels using Linked Networks to account for spatial information that is lost during feature extraction. The implementation of the segmentation technique involves using Residual Networks as a feature extractor and Transpose Convolution and Upsample Blocks for image-to-image translation thereby giving a segmentation mask as an output. The use of Upsample Blocks arises from its ability to give noise-free output while 18 layered Residual Networks using skip connections are used in the feature extraction without the vanishing gradient issues. The main feature of this architecture is the links between the Feature Extractor and the Decoder networks that improve the performance of the network by helping in the recovery of lost spatial information. Training and Validation using the Pytorch framework have been performed on the Digital Retinal Images for Vessel Extraction (DRIVE) Dataset to establish quality results.**

Index Terms – Semantic segmentation, feature extraction, linked networks, spatial information and noise.

I. INTRODUCTION

Only the retinal vascular system can produce masked blood vessels using non-invasive imaging techniques, and it offers abundant information on the health of the eye. For the detection of various disorders, retinal vascular segmentation is of utmost importance. Retinal scans are therefore often utilized to identify early indicators of systemic vascular disease. Accurate segmentation of the arteries are necessary to make systemic vascular disorders easier to diagnose. As a result, in the realm of medical imaging, automated segmentation from fundus images has grown in popularity. Manual segmentation of arteries may be incorrect and error-prone, making subsequent treatment challenging. Segmentation in the retinal vascular system is essential for identifying numerous disorders.

Thus, early indicators of systemic vascular disease are routinely found using retinal scans. The precise segmentation of the arteries is necessary for the identification of systemic vascular disorders. Hence, in the realm of medical imaging, it has become increasingly common to automatically segment blood vessels from retinal images.

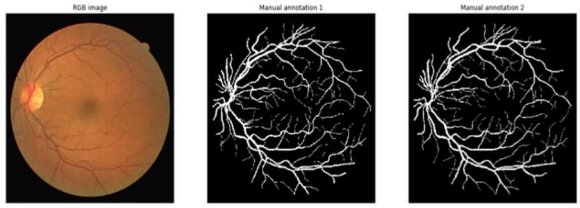

Fig.1. Retinal Image with Manual Annotations

Existing systems for Retinal Blood Vessel Segmentation such as UNETs [1], CNNs [2], and SVMs [3] have a common problem of the loss of important spatial information during the feature extraction stage due to multiple downsampling. Hence, the existing models are not able to achieve the accuracy that they could have if the spatial information has been retained. Retention of this spatial information can lead to increased accuracy of segmentation thereby improving the diagnosis of various eye ailments. Certain networks such as UNETs use Transpose Convolutions [4] for the Upsampling task.

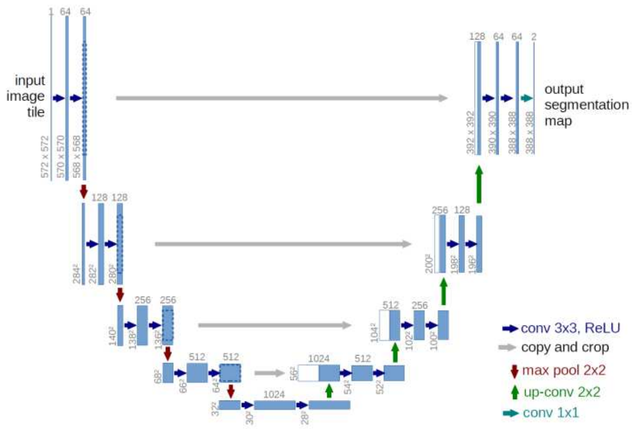

Fig.2. U-Net Architecture

The usage of Transpose Convolution often leads to noisy outputs. In our paper, we have tried an implementation of a Semantic Segmentation Network to segment blood vessels present in the retina for proper treatment and reduction of operator fatigue. As a counter to the problem of lost spatial information, we propose using an improved LinkNet [5] based architecture for semantic segmentation of the Retinal Blood Vessels. The network makes use of skip connections [6] to link each and every Encoder Block to its corresponding

Decoder Block, concatenating the lost spatial information during Upsampling while using Upsample Blocks as our approach, instead of Transpose Conv counters the problem of noisy output.

II. RELATED WORK

Significant advancements have been achieved over the years in the field of Computer Vision with respect to biomedical image segmentation. Implementations of various architectures for Retinal Image Segmentation have been made with increasing accuracy with each implementation. [7] has been used by many architectures for their development over the years. A major breakthrough started when the ridge-based segmentation method was used by Joes Staal et. al. in [8], achieving an accuracy of 0.9441. Centerline detection and morphological reconstruction were used a few years later by Mendonca et. al. [9], to improve the accuracy of segmentation to 0.9452. An attempt to further improve morphological reconstruction was made in [10] by using mathematical morphological theory and intensity transformation algorithm. The method used True Positive Fraction as the evaluation metric achieving a value of 0.8214 for the same. Further research was done on the DRIVE Dataset during the time with Morlet Transforms [11] by Ghaderi et. al. which achieved an AUC ROC of 0.90 to 0.96. Research shifted from DRIVE Dataset to other custom datasets when OctaveUNets [12] were used for Retinal Image Segmentation. Trained and validated on a custom dataset, OctaveUNets achieved an accuracy of 0.9524, thus leading to further improvements in this field. With OctaveUnets achieving high accuracy, neural networks scheme for pixel classification and moment invariant-based features for pixel representation-based methods were experimented with by Marin et. al. on the DRIVE Dataset. Training and Validation with this architecture led to an accuracy of 0.9452.

New datasets like STARE [13] started to be used alongside DRIVE for model improvement. Xinge et. al. used both DRIVE and STARE datasets for their research with semi-supervised learning and radial projection [14], and achieved a better accuracy on STARE Dataset, 0.9434 on DRIVE and 0.9497 on STARE. With better results on STARE, new datasets such as CHASE-DB [15] started to be used for training. [16] achieved an accuracy of 0.9480, 0.9534 and 0.9469 on DRIVE, STARE and CHASE -DB datasets respectively using ensemble methods and Gabor Filters. Multiscale line-detection [17] was used by Nguyen et. al. on STARE Dataset achieving an accuracy of 0.9326. CLAHE followed by a 2D Gabor wavelet was used in [18] on DRIVE and STARE datasets to get an accuracy of 0.9477 and 0.9509 respectively. Further research we're conducted using a Drive dataset with a B-COSFIRE filter [19], unsupervised iterative segmentation with tophat reconstruction and global thresholding [20] and Holistically Nested Edge Detection [21]. They achieved an accuracy of 0.9442, 0.9494 and 0.9435 respectively. Multilevel CNN with Conditional Random Fields [22] was used on DRIVE, STARE and CHASE-DB datasets with an average accuracy of 0.9523. SVMs were used on DRIVE Dataset, using accuracy and specificity as evaluation metrics, achieving values of 0.9538 and 0.9773 respectively. Moving to a more data-centric approach, UNETs with data-specific augmentations [23] were used on DRIVE, achieving an AUC ROC of 0.9790. UNETs were further improved in [24], achieving an accuracy of 0.9650 on DRIVE Dataset. Modified SUSAN edge detector and shape analysis was used in [25] on DRIVE and STARE datasets achieving an accuracy of 0.9633 and 0.9610 on them respectively. This was improved upon by Deformable ConvNets [26] achieving 0.9628 and 0.9690 accuracy on DRIVE and STARE datasets.

In our paper, we have tried to employ skip connections to concatenate lost spatial information during feature extraction with their corresponding decoder blocks to convert the lost information into useful information for better segmentation.

III. PROPOSED METHOD

The proposed architecture is an improvement upon LinkNet for semantic segmentation. Here "conv" describes a convolution operation. Batch normalization is employed in between each 'conv' and ReLU activation follows in order to induce non-linearity. Our LinkNet comprises an Encoder, which is essentially a ResNet18 network, and a Decoder which is an improvement upon the decoder present in existing LinkNet.

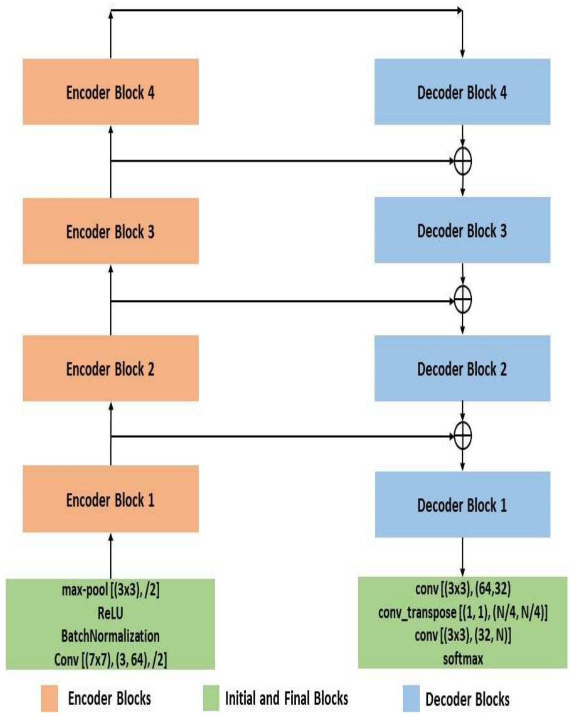

Fig.3.LinkNet Architecture

The Encoder Architecture, as shown in Fig. 4 consists of an initial block along with 4 encoder blocks for the purpose of feature extraction.

The Initial Block performs 'conv' on the input, with a (7x7) filter, and stride of 2. Batch Normalization and ReLU activation follows to introduce non-linearity, and a (3x3) kernel spatial max-pooling.

TABLE I. LINKMODULES

Modules	Block	Actions	Out Channel
Encoder	Input-Block Encoder Blocks	Convolution operation and max-pool Feature Extraction with skip connections	64 64, 128, 256, 512
Decoder	Upsample Final Segmentation Block	Upscale channels concatenated to the required size Uses softmax to produce final output mask	256, 128, 64, 64 1

The residual blocks follow the Initial Block, which are used for feature extraction. Each blocks perform 3 convolutional operations with a (3x3) filter and skip-connections which is preceded by a strided convolution.

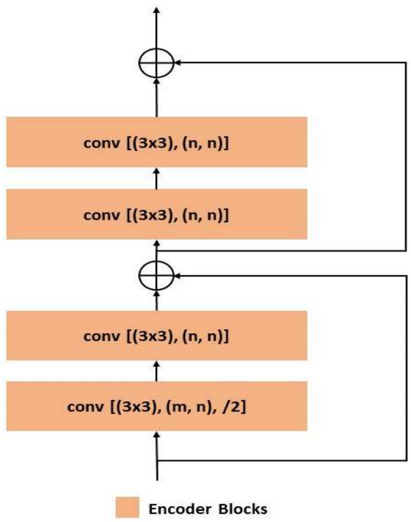

Fig.4. Encoder Block

Linking the encoder blocks to decoder blocks has been implemented as in existing LinkNet, for the recovery of the spatial information that was lost due to multiple downsampling. We have used strided convolutions for enabling the linking.

The Decoder Architecture takes the encoder output as its input. It consists of 4 decoder blocks followed by a final segmentation block for semantic segmentation.

Each of the decoder blocks has 2 convolutions with (1x1) filter, followed by 'Upsample' with a scale factor of 2 in trilinear mode. The Upsample operation is an improvement upon the existing LinkNet, introduced in [27], to counter noisy outputs as a result of transpose convolution. The final segmentation block follows the decoder blocks.

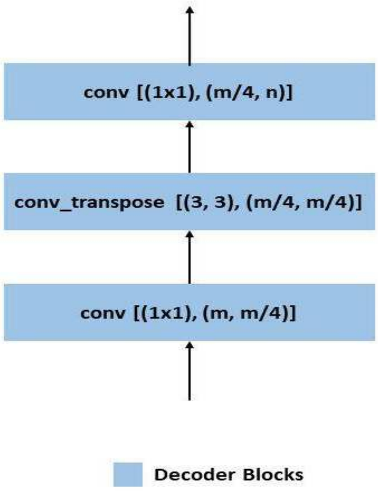

Fig.5. Decoder Block

The main semantic segmentation is performed on the output from decoder blocks, by the final segmentation block. The decoder output undergoes a convolution operation with a (3x3) filter, followed by Upsampling for noise reduction. The upsampled result undergoes yet another convolution with a (3x3) filter before a Softmax activation is applied.

IV. EXPERIMENTS

A. Dataset

For the purpose of segmenting retinal vessels, we employed the DRIVE dataset. It contains 40 JPEG colour fundus photos in total, including 7 occurrences of the aberrant disease. Each Image is downsized to 512x512 pixels for our model and has three colour channels. The data has been augmented using the argumentation module. Data Augmentation has been accomplished by doing HorizontalFlip, VerticalFlip and Rotate, thereby generating 160 images for training. The use of augmentations and pre-processing procedures like normalization and standardization has been made. As part of the pre-processing, all images were altered to have zero means and one unit of standard deviation. The model received basic normal Images to learn on.

Fig.6. Original, Horizontal Flip, and Vertical Flip Images

B. Metrics

Dice Loss: We now use Dice Loss as a statistic to assess how well our model is working. The dice coefficient is used to determine if two sets overlap or are similar. For issues including class disparity, this measure is particularly popular. For semantic segmentation, the predicted label and the actual label may be seen as two sets. The Dice coefficient has a value between 0 and 1, with 0 denoting complete overlap and 1 denoting entire overlap. The following formula is used to get the dice coefficient.

Dice Coefficient = $2 \mid T \cap P \mid / |T| + |P|$

The dice coefficient is used to calculate the generalized loss function, which is done by subtracting it from 1. The dice coefficient is increased by minimizing this dice loss since a greater value indicates a better overlap.

IoU Loss: We have used Intersection over Union as another metric for evaluating the performance of our model. IoU, as the name suggests, is calculated as a ratio of the overlap of the predicted label with the ground truth to their union, i.e. the total area they cover. Similar to the dice loss, has a value between 0 and 1, with 0 denoting complete overlap and 1 denoting entire overlap. The Intersection over Union value is calculated as:

Intersection over Union = $\mid T \cap P \mid / |T \cup P|$

The IoU loss function is calculated by subtracting the IoU Score from 1, and similar to the Dice Loss, a lower IoU loss gives a better overlap, hence is minimized.

C. Training details

The PyTorch framework was used to code the architecture, and the NVIDIA Tesla T4 GPU and CUDA were integrated combinedly to train it. With a batch size of 4, 80 images were used to train the model. Before being loaded into a data loader object for training, each image underwent pre-processing, was enhanced, normalised, and standardised.

To reduce the dice and IoU loss, the Adam optimizer was used to train the model for 100 iterations at a learning rate of 0.0001.

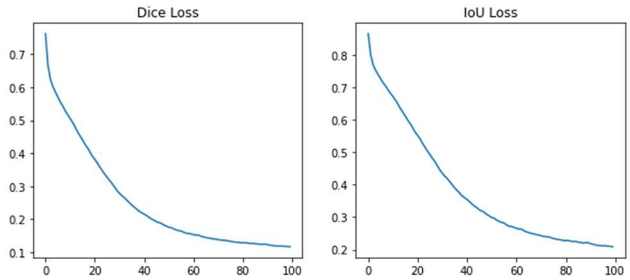

Fig.7. Training Progression (Dice Loss and IoU Loss)

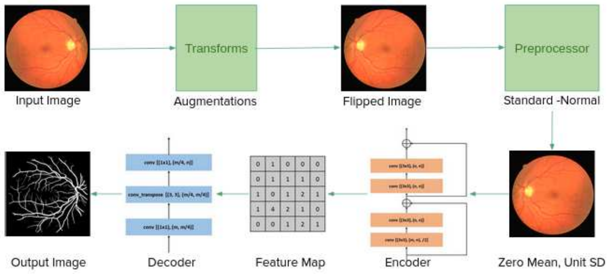

Fig.8. Training Workflow

D. Experimental Results

Validation was performed on DRIVE and STARE datasets using NVIDIA Tesla T4 GPU. The weights were hosted and downloaded via the gdown module for further tests. Figure 9 shows the test results of the input retinal images from DRIVE and validation results from STARE have been mentioned in Table II.

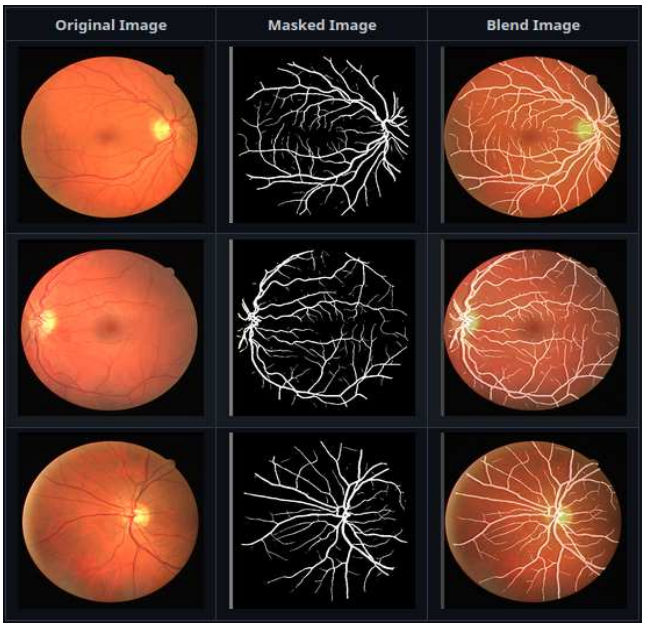

Fig.9 Test Results

V. CONCLUSION AND FUTURE WORKS

Effective segmentation of retinal image can be extensively used in medical imaging, particularly in ophthalmology, for diagnosis, monitoring, and treatment of retinal diseases. In our research, we have tried to improve the segmentation accuracy by improving upon LinkNet, which will help in better diagnosis of various eye ailments. This work can be further improved by collection of a binary classification dataset with classes healthy and infected, and a binary classification model that would classify the output segmented mask into one of the aforementioned classes.

TABLE II. VALIDATION RESULTS

Metrics	DRIVE	STARE
Dice Loss	0.1164	0.1277
IoU Loss	0.2086	0.1962

TABLE III. COMPARISON WITH SOTA ARCHITECTURES

Architecture	Metric
UNETs	0.9790 (AUC ROC)
Lattice NN with Dendrite	0.81 (F1 Score)
Multi-level CNN with	0.9523 (Accuracy)
Multi-scale Line Detection	0.9326 (Accuracy)
CLAHE	0.9477 (Accuracy)
Modified SUSAN edge	0.9633 (Accuracy)
Improved LinkNet	0.8836 (Dice Score)

REFERENCES

[1] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham.

[2] Yamashita, R., Nishio, M., Do, R.K.G. *et al.* Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9**, 611–629(2018).

[3] Zhang, Y. (2012). Support Vector Machine Classification Algorithm and Its Application. In: Liu, C., Wang, L., Yang, A. (eds) Information Computing and Applications. ICICA 2012. Communications in Computer and Information Science, vol 308. Springer, Berlin, Heidelberg.

[4] Gao, Hongyang & Yuan, Hao & Wang, Zhengyang & Ji, Shuiwang. (2019). Pixel Transposed Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. 1-1. 10.1109/TPAMI.2019.2893965.

[5] A. Chaurasia and E. Culurciello, “Linknet: Exploiting encoder representations for efficient semantic segmentation,” pp. 1–4, 2017.

[6] A. E. Orhan and X. Pitkow, “Skip connections eliminate singularities,” 2017.

[7] https://drive.grand-challenge.org/DRIVE/

[8] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," in IEEE Transactions on Medical Imaging, vol. 23, no. 4, pp. 501-509, April 2004, doi: 10.1109/TMI.2004.825627.

[9] A. M. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," in IEEE Transactions on Medical Imaging, vol. 25, no. 9, pp. 1200-1213, Sept. 2006, doi: 10.1109/TMI.2006.879955.

[10] I. Abdurrazaq, S. Hati and C. Eswaran, "Morphology approach for features extraction in retinal images for diabetic retionopathy diagnosis," 2008 International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, 2008, pp. 1373-1377, doi: 10.1109/ICCCE.2008.4580830.

[11] R. Ghaderi, H. Hassanpour and M. Shahiri, "Retinal vessel segmentation using the 2-D Morlet wavelet and neural network," 2007 International Conference on Intelligent and Advanced Systems, Kuala Lumpur, Malaysia, 2007, pp. 1251-1255, doi: 10.1109/ICIAS.2007.4658584.

[12] Osareh, Alireza & Shadgar, Bita. (2009). Automatic blood vessel segmentation in color images of retina. Iranian Journal of Science and Technology. Transaction B: Engineering. 33.

[13] https://cecas.clemson.edu/~ahoover/stare/

[14] You, Xinge & Peng, Qinmu & Yuan, Yuan & Cheung, Yiu-ming & Lei, Jiajia. (2011). Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. Pattern Recognition. 44. 2314-2324. 10.1016/j.patcog.2011.01.007.

[15] Jiaqi Guo, Ziliang Si, Yu Wang, Qian Liu, Ming Fan, Jian-Guang Lou, Zijiang Yang, and Ting Liu. 2021. Chase: A Large-Scale and Pragmatic Chinese Dataset for Cross-Database Context-Dependent Text-to-SQL. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 2316–2331, Online. Association for Computational Linguistics.

[16] M. M. Fraz et al., "An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation," in IEEE Transactions on Biomedical Engineering, vol. 59, no. 9, pp. 2538-2548, Sept. 2012, doi: 10.1109/TBME.2012.2205687.

[17] Nguyen, Uyen & Bhuiyan, Alauddin & Park, Laurence & Ramamohanarao, Kotagiri. (2012). An effective retinal blood vessel segmentation method using multi-scale line detection. Pattern Recognition. 46. 703-. 10.1016/j.patcog.2012.08.009.

[18] Yu Qian Zhao, Xiao Hong Wang, Xiao Fang Wang, Frank Y. Shih, Retinal vessels segmentation based on level set and region growing, Pattern Recognition, Volume 47, Issue 7, 2014, Pages 2437-2446, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2014.01.006.

[19] Azzopardi, G., Strisciuglio, N., Vento, M., & Petkov, N. (2015). Trainable COSFIRE filters for vessel delineation with application to retinal images. Medical Image Analysis, 19(1), 46-57.

[20] S. Roychowdhury, D. D. Koozekanani and K. K. Parhi, "Iterative Vessel Segmentation of Fundus Images," in IEEE Transactions on Biomedical Engineering, vol. 62, no. 7, pp. 1738-1749, July 2015, doi: 10.1109/TBME.2015.2403295.

[21] Xie, S., & Tu, Z. (2015). Holistically-Nested Edge Detection.ArXiv.https://doi.org/10.48550/arXiv.1504.06375.

[22] H. Fu, Y. Xu, S. Lin, D. W. K. Wong and J. Liu, "DeepVessel: Retinal vessel segmentation via deep learning and conditional random field", Proc. MICCAI, pp. 132-139, 2016.

[23] Xiancheng, Wang, et al. "Retina blood vessel segmentation using a U-net based Convolutional neural network." Procedia Computer Science: International Conference on Data Science (ICDS 2018). 2018.

[24] P. Xiuqin, Q. Zhang, H. Zhang and S. Li, "A Fundus Retinal Vessels Segmentation Scheme Based on the Improved Deep Learning U-Net Model," in IEEE Access, vol. 7, pp. 122634-122643, 2019, doi: 10.1109/ACCESS.2019.2935138.

[25] Braovic, Maja & Stipanicev, Darko & Šerić, Ljiljana. (2018). Retinal blood vessel segmentation based on heuristic image analysis. Computer Science and Information Systems. 16. 14-14. 10.2298/CSIS180220014B.

[26] Jin, Q., Chen, Q., Meng, Z. et al. Construction of Retinal Vessel Segmentation Models Based on Convolutional Neural Network. Neural Process Lett 52,1005–1022(2020).

[27] P. Ravi, S. Roy, I. Dutta and K. Kottursamy, "Attention Mechanism, Linked Networks, and Pyramid Pooling Enabled 3D Biomedical Image Segmentation," 2022 IEEE/ACIS 23rd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Taichung, Taiwan, 2022, pp. 91-96, doi: 10.1109/SNPD54884.2022.10051771.