Research Article

LUPU-Net: a new improvement proposal for encoder-decoder architecture

Saadet Aytaç Arpacı [a],* iD and Songül Varlı [a] iD

ªComputer Engineering Department, Yıldız Technical University, Istanbul, 34220, Turkey

ABSTRACT

Many network designs in recent years have offered deeper layered solutions. However, models that achieve high-performance results with fewer layers are preferred due to causing less processing load for the system. The U-Net authors succeeded in efficiently creating a model with fewer layers. However, the U-Net architecture also requires improvement to become more efficient. For this purpose, we offer a novel encoder-decoder architecture based on the U-Net and the LU-Net. Furthermore, we propose using a reduced number of up-sampling operations, which were utilized together with the down-sampling operations intensively in the encoder section in our previous research, in the encoder part. The proposed architecture was evaluated on the IOSTAR dataset for the segmentation of retinal vessels. The preprocessing and data augmentation processes were applied to the images before training. The U-Net, LU-Net, and the proposed model were evaluated by using the accuracy, sensitivity, specificity, Dice, and Jaccard metrics. The proposed model achieved performance metric values such as an accuracy of 97.29%, a sensitivity of 81.10%, a specificity of 98.94%, a Dice coefficient of 84.66%, and a Jaccard coefficient of 73.41%. The proposed model obtained improved results compared with the other models, especially for test samples.

1. Introduction

Artificial neural network studies started in 1943 with [1] research, and these networks were introduced by [2] as the term "deep learning" in the year 2000. In the 1980s, the "Neocognitron" model, which could learn with a hierarchical multi-layered network, was proposed [3]. Later, the convolutional neural network (CNN) was developed based on the "Neocognitron" model.

CNN is a breakthrough in visual pattern recognition. Convolutional neural networks basically consist of three types of layers: 1) The convolution layer produces the feature maps (the activation maps); 2) the activation functions cause the values obtained from the layers to be kept within specific value ranges; 3) the down-pooling layer reduces spatial resolution.

Many architectures have been developed based on CNN. One of these architectures is the VGG16 classification model [4]. The VGG16 architecture uses 13 convolution layers and 3 x 3 convolution filters for feature extraction, a rectified linear unit (ReLU) layer follows each convolution layer, and the architecture has maximum

pooling layers for down-sampling. Fully connected layers are located in the last three layers [4].

With the semantic segmentation process, each pixel of an image is associated with a class tag. Thus, the desired region in the image is determined with its boundaries. One of the first samples of semantic segmentation was created by the Fully Convolutional Networks (FCN) authors [5]. The FCN authors [5] replaced the fully connected layers of existing CNN architectures such as VGG16 with the convolutional layers and used the transposed convolution for up-sampling. They did not add more decoder layers to the FCN architecture; therefore, the high-resolution features were disregarded by the model. This problem was solved with an improved encoder-decoder network. The U-Net [6] model, which was developed from the FCN architecture, was proposed in 2015. The U-Net contains some differences compared with the FCN: 1) The encoder-decoder sections are symmetrical; 2) it uses the "concatenation" operation instead of the "add" method to connect the encoder and decoder layers. Another encoder-decoder network architecture is SegNet [7], which was

proposed in 2017. The encoder part of this architecture is the same as the 13 convolution layers in the VGG16 network. The SegNet has 13 convolution layers in the decoder part corresponding to the encoder path. In the U-Net architecture, all activation maps are transferred to the corresponding decoder layer and concatenated with the up-sampled feature maps through the transposed convolution operation. The difference between SegNet with U-Net is that SegNet uses the pooling indices in the encoder section (calculated in the maximum pooling layer) for up-sampling in the corresponding layers in the decoder section. The up-sampled maps are then subjected to the convolution operation with the trainable filters to create the dense feature maps, as in U-Net. Some researches have been done in recent years with SegNet. For example, SegNet was used for retinal vessel segmentation in [8, 9] researches. The jump connections in the U-Net architecture can be seen as an advantage for the model to obtain more feature information. Even if the number of blocks of the model is reduced, the U-Net can achieve high performance at the pixel level [10-12]. The simple and flexible composition of the architecture allows efficient, new designs to be developed based on this model. For example, separate side paths were added to the encoder and the decoder sections in the M-Net [13] architecture. In this architecture, gradually reduced inputs with the maximum pooling operation in the first side path were given to the encoder section, and the outputs from the decoder section were given to the second side path. In the IterNet [14] architecture, the U-Net model was iteratively implemented to be four times deeper than the U-Net. Besides the raw input image, segmented vessel images were also used as input. Their model has achieved a 79.70% sensitivity rate on the CHASE-DB1 dataset [14]. The authors in [15] reference proposed an improvement to use the convolution operation instead of the maximum pooling process in the original U-Net model to retain more feature information. The authors in [16] reference used the bilinear interpolation function for the up-sampling operation. The theoretically expected situation in deep architectures is that if the number of layers of the network increases, the training error decreases. Contrary to this situation, the experiments were indicating the problem of the vanishing gradient. The Res-Net and the Dense-Net models tried to alleviate this problem. The input features were added as a shortcut link to the output of the next two convolutional layers in the Res-Net [17] model. The authors in [18, 19] references, who interpreted the Res-Net model differently, integrated their methods into the U-Net architecture for retinal vessels' segmentation. The MResU-Net [18] model has achieved an 81.01% sensitivity rate on the STARE dataset. The weighted Res-UNet [19] model has achieved a 77.15% sensitivity rate on the DRIVE dataset. In the Dense-Net [20] model, each

layer in a block is connected with all previous layers. The Dense-Net was integrated into the U-Net model for retinal vessels' segmentation by [21, 22] researchers. The proposed model in [21] has achieved a 76.72% and an 89.67% sensitivity rate on the DRIVE and the CHASE-DB1 dataset, respectively. The proposed model in [22] reference has achieved a 79.86% sensitivity rate on the DRIVE dataset. The U-Net ++ authors [23] used the dense convolution blocks in the U-Net model's jump connections. The authors of the [24] reference proposed creating the dense feature maps in the U-Net model's encoder path and transferring them to the decoder path through the jump connections.

The authors of U-Net [6] demonstrated that the results of designs with many levels in applications could be accomplished with fewer layers as well. This situation provides an advantage in that it brings less processing load to achieve the result. However, the U-Net architecture also needs improvement to achieve more efficiency. For example, for cardiac ventricular segmentation, the LU-Net authors [25] proposed some improvements. In another improvement example, many feature maps obtained from the down-sampling and up-sampling operations were transferred to the decoder section of the U-Net in the reference [24] research, but here, as the number of layers increased, so did the processing load. Therefore, within this research, we tried the up-sampling operation applied in the reference [24] research in order to achieve results with less processing load. In continuation, the outcomes of the experiments were appraised by ours.

In this manuscript, we proposed a new model based on the U-Net, LU-Net architectures, and the reference [24] research for retinal vessel segmentation. This article has two essential contributions: Utilizing the up-sampling operation in the encoder part to get precise location information, and introducing a hybrid architecture improved with the up-sampling operation. The rest of this paper describes the proposed architecture, the used dataset, presents the experimental results, and the discussion section.

2. Materials and Methods

2.1 U-Net

The U-Net architecture comprises an encoder path on the left side and a decoder path on the right side. The encoder section is like a characteristic convolutional network architecture. Four down-sampling processes occur in the encoder path with a 2 x 2 max-pooling operation. Each block in the encoder path consists of two 3 x 3 convolution layers and a ReLU as the activation function. According to the number of feature maps in the first layer, the number of feature channels doubles after each down-sampling operation. The decoder path is symmetrical to the encoder path, and four up-sampling

operations occur in this section. The feature map size is increased in the decoder path by the 2 x 2 up-sampling process, and these feature maps are concatenated with the feature maps in the corresponding block in the encoder path. A more precise output is created by integrating the contextual information obtained from the encoder path with the localization information obtained from the decoder path, thanks to these jump connections, which are the reason for the segmentation success of the U-Net. In continuation of the up-sampling process, two 3 x 3 convolution operations occur, in which ReLU is applied as an activation function. At the output, a 1 x 1 convolution layer is used. The encoder path and decoder path are bridged by two 3 x 3 convolution layers [6].

2.2 LU-Net

For cardiac ventricular segmentation, the LU-Net authors [25] offered three improvement methods on the U-Net model. According to their application [25], first, the SE-Net [26] module, which is a feature weighting process, was applied before each pooling operation in the encoder path. When the feature maps pass to the SE-Net module, the global average pooling process performs to obtain the mean pixel value per channel, and the feature vector occurs. The feature vector enters two consecutive fully connected layers. In the first, linearity is achieved by the ReLU function, and in the second, a series of weight values between 0 and 1 is obtained by the sigmoid function. Finally, the weight values are multiplied by the activation maps to generate the new feature maps comprising the weight information. The LU-Net authors included the SE-Net module because it suppresses the ineffective information and gives more weight to the beneficial information on the channel, so the sensitivity of the network increases. In this way, the effectiveness of the extracted features is increased, and the amount of calculation is reduced. Secondly, in the LU-Net research [25], the feature maps at various scales were obtained from the input (apart from the encoder path). Then these feature maps were concatenated with the output of the corresponding pooling layer in the encoder path. In this stage, a down-sampling process takes place with the convolution layer with a kernel size of 2 x 2 and a step size of 2 to obtain the feature maps of different scales. Thirdly, the deconvolution process in the decoder section of the U-Net was replaced with 3 x 3 and 1 x 1 convolution operations. Additionally, the up-sampling process was applied to the network model with the inverse unpooling operation. During U-Net's encoding and decoding processes, some important details are lost with the convolution and pooling operations, resulting in imperfect segmentation boundaries. To solve this problem, the authors of the LU-Net proposed the second and third improvements to the U-Net model [25].

2.3 Proposed Model

In this research, the U-Net [6], LU-Net [25] architectures, and the reference [24] research were used to create the proposed network model. Figure 1 shows the proposed LUPU-Net network.

Network architectures were modified such that: For the encoder part, we used the blocks in the encoder path of the LU-Net model up to the bridging section. Thus, the use of useful and effective information is provided. For the decoder path, we used the bridging section of the U-Net model and the blocks in the decoder path. Thus, the processing load that the convolution process would bring to the system is lightened. Each block in the network has two 3 x 3 convolution layers, and we used the ReLU as the activation function.
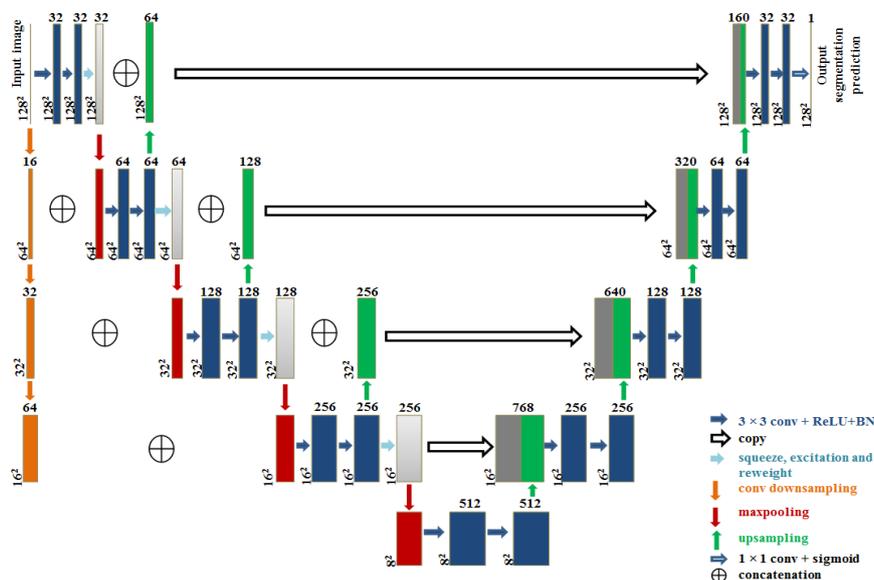

Figure 1. LUPU-Net model

For down-sampling in the encoder path, we used four 2 x 2 max-pooling operations, three convolutional down-sampling operations with a kernel size of 2 x 2 and a step size of 2.

We propose that the feature maps obtained by the up-sampling process from the blocks in each level of the encoder path pass to the decoder path through the jump connections. In this process, the up-sampling operation is applied to these feature maps before passing the feature maps to the SE-Net module. In continuation, these up-sampled feature maps are concatenated with feature maps obtained from the SE-Net module at the upper level. The precise localization information of the features in the encoder path concatenates with the contextual information obtained from the encoder path blocks. In this way, the segmentation of the images is made better. Later, all the feature maps are transferred from the encoder section to the decoder part.

The number of feature maps in the first layer of the LUPU-Net network is 32. After each down-sampling operation, the number of feature maps increases twice. During the training of the network, the pre-training process was not applied. In this research, the "UpSampling2D" class was used for up-sampling. As the optimization algorithm of the model, the Adam method [27] was used by starting it at a learning rate of 0.001. The batch normalization (BN) process was applied after the convolution layers to increase the convergence speed of the network. At the output, we used a convolution layer of 1 x 1 with the sigmoid activation function, which reduced all feature map information to a two-element vector. The performance of the model trained with 100 iterations was evaluated. Our segmentation research was carried out using the Keras library [28], which uses the TensorFlow [29] framework on the backend.

2.4 Dataset

In this research, the IOSTAR dataset, which is publicly available on the website [30], was used. These high contrast images in the dataset based on the Scanning Laser Ophthalmoscopy (SLO) technique have a resolution of 1024 x 1024 pixels with a 45° FOV. The dataset contains 30 SLO images. The ground truth images were created by experts, and these images were included in the dataset. More information about the dataset is available in [31, 32] references.

2.5 Preprocessing

Since SLO images show retinal vascular structures with a stronger contrast, these images do not require preprocessing, but preprocessing was applied to the images in this experiment to get more vascular detail. Figure 2 shows the preprocessing stages. In this process [33], in the first stage, the images were converted to the HSV color space and the luminosity was adjusted by applying gamma correction to the value channel. Then the images were converted from the HSV color space to the RGB color space. In the second stage, the images were converted from the RGB color space to the LAB color space, and the CLAHE process was applied to the L channel to increase the vessel contrast. The image pixel values were normalized to the range [0-1]. 128 x 128-size single-channel gray levels of the images were used in the evaluation of the models. The OpenCV library [34] was used during the preprocessing of the images.
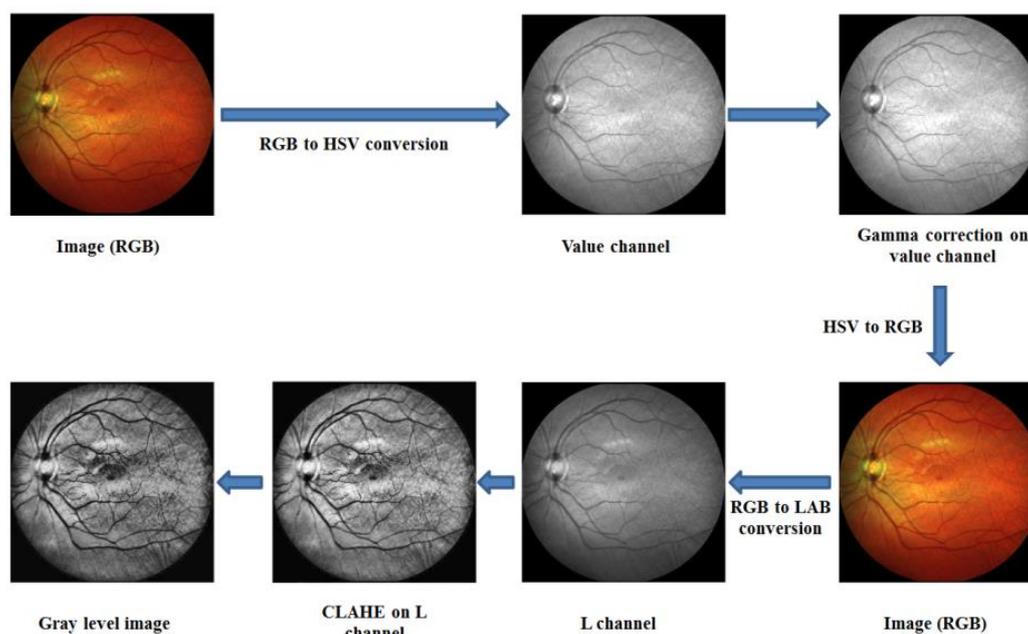

Figure 2. The preprocessing stages applied to the SLO images

The last 10 images in the IOSTAR dataset were used for testing. The data augmentation processes such as horizontal flipping, cropping a random area of the image, and rotating the image left and right by a maximum of 10º were applied to the 20 images. The 1000 images were obtained through the data augmentation process. The Augmentor library [35] was used for the data augmentation process. 20% of the dataset images were used for validation.

2.6 Evaluation Metrics

As used in the reference manuscript [36], accuracy as in Equation (1), sensitivity as in Equation (2), and specificity as in Equation (3) were used to measure the performance of the models in this research, as well. Furthermore, we measured the models' performance for the retinal vessels' segmentation with the Dice metric as in Equation (4) [37], and the Jaccard coefficient (IoU) as in Equation (5) [38]. The accuracy is the ratio of the sum of the number of pixels in the correctly defined vessel and background areas to the sum of the total number of pixels [36]. The sensitivity gives the rate of determination of the vessel area. The specificity gives the rate of determination of the background area. The Dice and Jaccard metrics measure the overlap between the segmented prediction image and the ground truth images. These metrics are calculated as follows:

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN} \qquad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (2)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (3)$$

$$Dice = \frac{2 * TP}{(2 * TP) + FP + FN} \qquad (4)$$

$$Jaccard = \frac{TP}{TP + FP + FN} \qquad (5)$$

TP, TN, FP, FN are the quantities of true positive, true negative, false positive, and false negative at the pixel level, respectively.

3. Results

The performance of the proposed network on vessel segmentation was evaluated using the IOSTAR dataset. Table 1 shows the evaluation metric values obtained on the IOSTAR data set with the U-Net, LU-Net, and LUPU-Net models. The accuracy, sensitivity, specificity, Dice, and

Jaccard metric values obtained with three models show minor differences. We used the same basic parameters, values, and classes for the three models in the experiments. For example, the activation function, the initial quantity of the feature maps, the down-sampling and up-sampling classes, the optimization function, the loss function, etc. The three models in the experiments differ only architecturally. Thus, we can explain the minor differences between the values.

Figure 3 shows the segmentation results obtained by three models for a test sample. Models can predict most retinal vessels.

In the result segmentation prediction images, the pixel values are obtained in the range of [0-1]; therefore, a binary thresholding process was applied to clarify the background and vessel separation with the values of 0 and 1. In this post-processing process, 0.10 was used as the threshold value. Table 2 shows the average accuracy, sensitivity, specificity, Dice, and Jaccard metric values obtained after the post-processing stage from the images in Figure 3.

Figure 4 shows graphically the average Dice similarity coefficients obtained from the models for ten test images. Figure 5 shows graphically the average Jaccard similarity coefficients obtained from the models for ten test images. The differences that did not appear in Table 1 metric value results became more apparent in the test samples. Figure 4 and Figure 5 show many test samples are segmented more efficiently by the LUPU-Net model. Therefore, the LUPU-Net increases the overlap between the ground truth and prediction images. These results show that the proposed architecture contributes to the improvement of retinal vessel segmentation.

4. Discussion

In this research, we created a hybrid network architecture based on the U-Net and the LU-Net models, and we applied a new improvement proposal based on the reference [24] research to this hybrid model. We used the IOSTAR dataset to measure the efficiency of this new model (LUPU-Net). Table 3 shows the outcomes of other studies [39-41] based on the U-Net architecture on the IOSTAR dataset and the results obtained from the LUPU-Net model. The findings show that the proposed model performed well in terms of pixel discrimination.

The researchers in reference [42] evaluated their proposed model for melanocytic lesion and retinal vessel segmentation without using image preprocessing. Their model has achieved a 94% Dice coefficient, an 88.2% Jaccard coefficient for melanocytic lesion segmentation, and an 82.6% Dice coefficient, a 70.8% Jaccard coefficient for retinal vessel segmentation. These comparative results show that the segmentation of a narrow and thin region is more difficult than a large region.

Table 1. The performance metric values obtained from the models

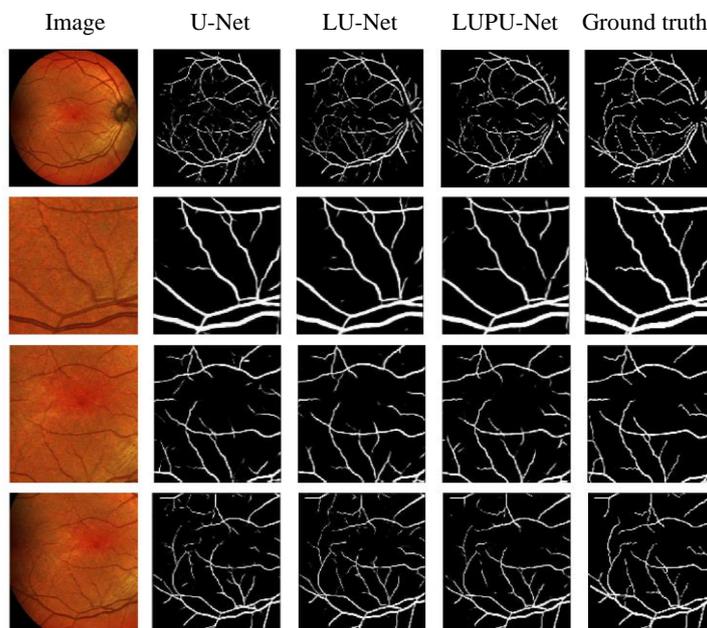| Models	Evaluation Metrics (%)				
	Accuracy	Sensitivity	Specificity	Dice	Jaccard (IoU)
U-Net	97.30	80.52	99.01	84.63	73.37
LU-Net	97.29	80.52	98.99	84.52	73.21
LUPU-Net	97.29	81.10	98.94	84.66	73.41

Figure 3. The segmentation results were obtained by three models for a test sample. The first column demonstrates the fundus images, the second column demonstrates the segmentation result images obtained by the U-Net model, the third column demonstrates the segmentation result images obtained by the LU-Net model, the fourth column demonstrates the segmentation result images obtained by the LUPU-Net model, and the fifth column demonstrates the ground truth images

Fundus photographs obtained with standard fundus cameras show complexities such as low contrast difference between the vessel and the non-vessel pixels, uneven background illumination, vessel width variation, brightness, and shape [43]. For this reason, many studies based on the U-Net [10, 11, 16, 19, 21, 22] have applied different preprocessing techniques to achieve more vascular detail. Compared with the standard fundus photographs, the images obtained by the SLO method are higher-quality images with high contrast and resolution. However, changes in vessel thicknesses, transitions and bifurcations in vessels, the low contrast difference between the vessel and the background for thin vessels are valid problems for the SLO images as well. For this reason, we applied image preprocessing to the IOSTAR dataset images to obtain more vessel details. The researchers in [39- 41] did not use image preprocessing.

The authors of reference [39] segmented these images in the U-Net model by dividing the images into overlapping patches of 128 x 128 pixels. The authors of reference [41] created the datasets by only producing a small number of image patches without using any other data augmentation method. They evaluated the dataset segmentation by the U-Net model after 1000 epochs. The authors in reference [40] utilized the IOSTAR dataset without using any data augmentation process and changing the image size. They created a dense layered and densely connected architecture based on the Res-Net, Dense-Net, and U-Net models and evaluated the IOSTAR dataset segmentation after 300 epochs.

Table 2. The average metric values obtained after the post-processing stage of the images in Figure 3

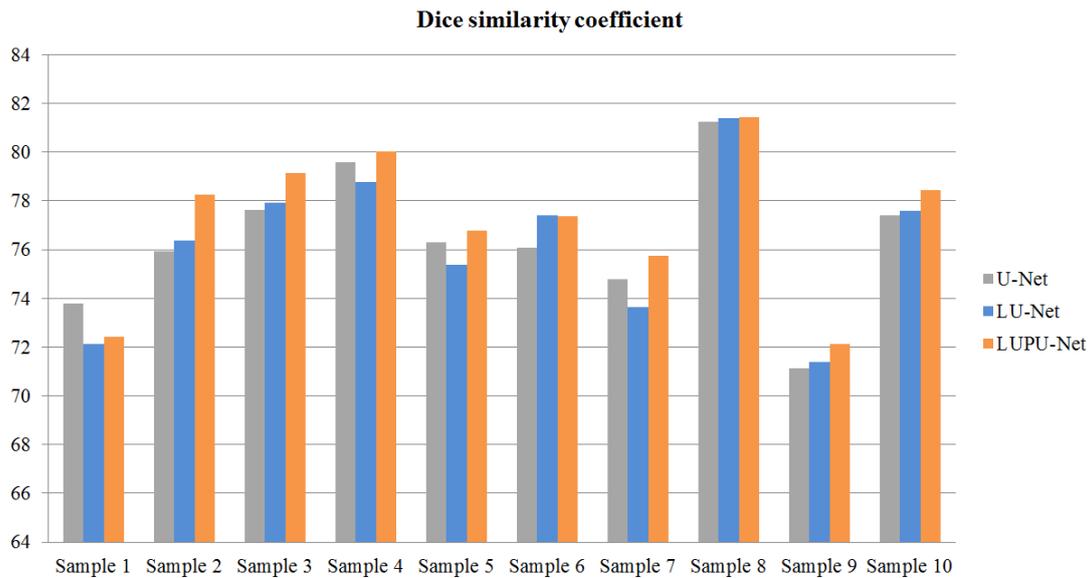| Models	Evaluation Metrics (%)				
	Accuracy	Sensitivity	Specificity	Dice	Jaccard (IoU)
U-Net	95.45	84.50	96.74	79.57	66.31
LU-Net	95.19	84.25	96.56	78.78	65.14
LUPU-Net	95.60	83.22	97.12	80.03	66.84

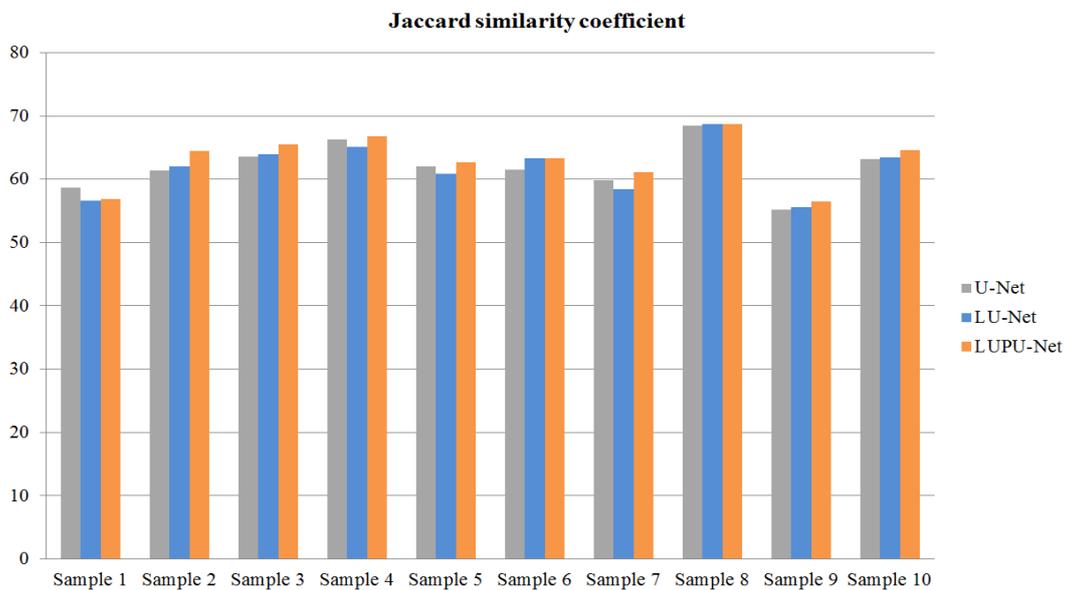Figure 4. Models' average Dice similarity coefficients for ten test samples

Figure 5. Models' average Jaccard similarity coefficients for ten test samples

Table 3. The performance metric values of our model and previous researches on the IOSTAR dataset

Researches	Evaluation Metrics (%)				
	Accuracy	Sensitivity	Specificity	Dice	Jaccard (IoU)
Meyer et al. (2017) [39]	96.95	80.38	98.01	-	-
Guo et al. (2020) [40]	97.13	80.82	98.54	-	-
Brea et al. (2020) [41]	95.00	86.00	95.00	71.00	-
This research (2021)	97.29	81.10	98.94	84.66	73.41

A sufficient quantity of augmented data with the correct method can create high-efficiency segmentation. The data augmentation methods generally applied to the images are flipping, rotation, translation, cropping, and elastic deformation [12, 15, 19, 24, 42, 44, 45]. However, investigations into different data augmentation methods like the translation-reflection technique [16], etc. have

been continued as well. Through the data augmentation methods we used in this research, a successful result was obtained. We used the batch normalization and data augmentation operations in the model to prevent the overfitting problem.

The feature maps obtained through the down-sampling and up-sampling operations were used by the authors of

reference [24]. They concentrated a lot of feature maps on the jump connections to use more feature information. Their model has achieved a 97.87% accuracy, an 84.11% sensitivity, a 99.39% specificity, an 88.70% Dice, a 79.69% Jaccard value. Within these results, we can say that the use of more feature information produces more successful results. However, this situation causes more processing load for the system. In this research, we tried a model using fewer operations. In this manner, the processing load to reach the result was reduced. Based on the research of [24] reference, we proposed using the up-sampling operation in the encoder section, and we evaluated our proposal together with a hybrid architecture based on the U-Net and the LU-Net. The LUPU-Net architecture improved results, particularly for test samples. These results are seen in Figure 4 and Figure 5. The data augmentation methods, the amount of data, the quality of the image, and using the various libraries, classes, optimization algorithms, loss functions, activation functions, etc., all have an impact on the model segmentation performance. Therefore, the models were evaluated under the same conditions.

The retinal vascular system provides important markers for the diagnosis of some diseases (diabetic retinopathy, hypertension, cardiovascular diseases) for ophthalmologists. Fast and accurate detection of changes in the vessels is a factor that increases the success of the diagnosis. The process of segmenting the retinal vasculature from the image for better examination is performed by trained experts traditionally, and this segmentation action is a time-consuming procedure. Even among professional experts, the segmentation of the same image differs significantly [43]. Therefore, the development of automated techniques in retinal vessel segmentation continues in order to support the clinician's decisions. We can say that the method we propose will contribute to the field of ophthalmology. However, we also recommend using the new method for other medical image segmentation applications in the future.

5. Conclusion

This paper introduces a novel method for segmentation. We created an encoder-decoder architecture and used the up-sampling operation in the encoder section. The encoder-decoder architecture proposed in this research, which was developed based on the U-Net, LU-Net models and our previous research, is a hybrid design. The proposed LUPU-Net architecture was evaluated on the IOSTAR dataset for retinal vessel segmentation. The preprocessing and data augmentation processes were applied to the images in the IOSTAR dataset before training. Under the same conditions, we compared the U-Net, LU-Net, and LUPU-Net models. The LUPU-Net architecture obtained improved results, especially for test samples.

The LUPU-Net model can also be evaluated on a variety of

retinal vessel datasets in the future. Furthermore, the LUPU-Net model can also be evaluated for different medical image segmentation tasks other than ophthalmology. Our proposal to use the up-sampling operation in the encoder section can be implemented into the other encoder-decoder architectures, and the results obtained from these models can be compared. The LUPU-Net model can be compared with the other traditional methods in the future.

Declaration

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors also declared that this article is original, was prepared in accordance with international publication and research ethics, and ethical committee permission or any special permission is not required.

Author Contributions

S. A. Arpacı performed the investigation, developed the software, wrote the original draft and improved the research. S. Varlı supervised and improved the research.

Nomenclature

BN	: Batch Normalization
CNN	: Convolutional Neural Network
CLAHE	: Contrast Limited Adaptive Histogram Equalization
FCN	: Fully Convolutional Networks
FP	: False Positive
FN	: False Negative
HSV	: Hue, Saturation, Value
IoU	: Intersection-Over-Union
ReLU	: Rectified Linear Unit
Res-Net	: Residual Network
RGB	: Red, Green, Blue
SE-Net	: Squeeze-and-Excitation Network
SLO	: Scanning Laser Ophthalmoscopy
TP	: True Positive
TN	: True Negative
2D	: Two-dimensional

References

1. McCulloch, W.S. and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*. The Bulletin of Mathematical Biophysics, 1943. **5**: p. 115-133.

2. Aizenberg, I.N., N.N. Aizenberg, and J. Vandewalle, *Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications*. 2000, USA: Kluwer Academic Publishers.

3. Fukushima, K., *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*. Biological Cybernetics, 1980. **36** (4): p. 193–202.

4. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, in *3rd*

International Conference on Learning Representations, 2015, San Diego, CA: USA. p. 1-14.

5.   Long, J., E. Shelhamer, and T. Darrell, *Fully convolutional networks for semantic segmentation*, in *The IEEE Conference on Computer Vision and Pattern Recognition*, 2015, Boston: USA. p. 3431-3440.

6.   Ronneberger, O., P. Fischer, and T. Brox, *U-Net: Convolutional networks for biomedical image segmentation*, in *Medical Image Computing and Computer-Assisted Intervention*, 2015, Munich: Germany. p. 234-241.

7.   Badrinarayanan, V., A. Kendall, and R. Cipolla, *SegNet: A deep convolutional encoder-decoder architecture for image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017. **39** (12): p. 2481-2495.

8.   Khan, T.M., S.S. Naqvi, M. Arsalan, M.A. Khan, H.A. Khan, et al., *Exploiting residual edge information in deep fully convolutional neural networks for retinal vessel segmentation*, in *International Joint Conference on Neural Networks*, 2020, Glasgow: United Kingdom. p. 1-8.

9.   Ozgunalp, U., R. Fan, and A. Serener, *Semantic segmentation of retinal vessels using SegNet*, in *28th Signal Processing and Communications Applications Conference*, 2020, Gaziantep: Turkey. p. 1-4.

10. Xian-cheng, W., L. Wei, M. Bingyi, J. He, Z. Jiang, et al., *Retina blood vessel segmentation using a U-Net based convolutional neural network*, in *International Conference on Data Science*, 2018, Beijing: China. p. 1-11.

11. Gao, X., Y. Cai, C. Qiu, and Y. Cui, *Retinal blood vessel segmentation based on the gaussian matched filter and U-Net*, in *10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, 2017, Shanghai: China. p. 1-5.

12. Fu, W., K. Breininger, Z. Pan, and A. Maier, *Degenerating U-Net on retinal vessel segmentation*. [cited 2021 14 May]; Available from: https://doi.org/10.1007/978-3-658-29267-6_7.

13. Mehta, R. and J. Sivaswamy, *M-Net: A convolutional neural network for deep brain structure segmentation*, in *IEEE 14th International Symposium on Biomedical Imaging*, 2017, Melbourne: Australia. p. 437-440.

14. Li, L., M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, *IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks*, in *IEEE Winter Conference on Applications of Computer Vision*, 2020, Colorado: USA. p. 3656-3665.

15. Li, Q., S. Fan, and C. Chen, *An intelligent segmentation and diagnosis method for diabetic retinopathy based on improved U-Net network*. Journal of Medical Systems, 2019. **43**: p. 304.

16. Cai, Y., Y. Li, X. Gao, and Y. Guo, *Retinal vessel segmentation method based on improved deep U-Net*, in *Chinese Conference on Biometric Recognition*, 2019, Zhuzhou: China. p. 321-328.

17. He, K., X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, Las Vegas: USA. p. 770-778.

18. Li, D., D.A. Dharmawan, B.P. Ng, and S. Rahardja, *Residual U-Net for retinal vessel segmentation*, in *IEEE International Conference on Image Processing*, 2019, Taipei: Taiwan. p. 1425-1429.

19. Xiao, X., S. Lian, Z. Luo, and S. Li, *Weighted Res-UNet for high-quality retina vessel segmentation*, in *9th International*

Conference on Information Technology in Medicine and Education, 2018, Hangzhou: China. p. 327-331.

20. Huang, G., Z. Liu, L. Van Der Maaten, and K.Q. Weinberger, *Densely connected convolutional networks*, in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, Honolulu: USA. p. 2261-2269.

21. Cheng, Y., M. Ma, L. Zhang, C. Jin, L. Ma, and Y. Zhou, *Retinal blood vessel segmentation based on densely connected U-Net*. Mathematical Biosciences and Engineering, 2020. **17** (4): p. 3088-3108.

22. Wang, C., Z. Zhao, Q. Ren, Y. Xu, and Y. Yu, *Dense U-Net based on patch-based learning for retinal vessel segmentation*. Entropy, 2019. **21** (2): p. 168.

23. Zhou, Z., M.M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, *UNet++: A nested U-Net architecture for medical image segmentation*, in *4th International Workshop on Deep Learning in Medical Image Analysis*, 2018, Granada: Spain. p. 3-11.

24. Arpacı, S.A. and S. Varlı, *Retinal vessel segmentation with differentiated U-Net network*, in *28th Signal Processing and Communications Applications Conference*, 2020, Gaziantep: Turkey. p. 1-4.

25. Zhang, J., J. Du, H. Liu, X. Hou, Y. Zhao, et al., *LU-NET: An improved U-Net for ventricular segmentation*. IEEE Access, 2019. **7**: p. 92539-92546.

26. Hu, J., L. Shen, and G. Sun, *Squeeze-and-excitation networks*, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, Salt Lake City: USA. p. 7132-7141.

27. Kingma, D.P. and J.L. Ba, *ADAM: A method for stochastic optimization*. [cited 2021 14 May]; Available from: https://arxiv.org/pdf/1412.6980.pdf .

28. Keras library. [cited 2021 14 May]; Available from: https://keras.io/ .

29. TensorFlow library. [cited 2021 14 May]; Available from: https://www.tensorflow.org/ .

30. IOSTAR retinal vessel segmentation dataset. [cited 2019 5 April]; Available from: http://www.retinacheck.org/download-iostar-retinal-vessel-segmentation-dataset .

31. Zhang, J., B. Dashtbozorg, E. Bekkers, J.P.W. Pluim, R. Duits, et al., *Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores*. IEEE Transactions on Medical Imaging, 2016. **35** (12): p. 2631-2644.

32. Abbasi-Sureshjani, S., I. Smit-Ockeloen, J. Zhang, and B. Ter Haar Romeny, *Biologically-inspired supervised vasculature segmentation in SLO retinal fundus images*, in *12th International Conference Image Analysis and Recognition*, 2015, Niagara Falls: Canada. p. 325-334.

33. Zhou, M., K. Jin, S. Wang, J. Ye, and D. Qian, *Color retinal image enhancement based on luminosity and contrast adjustment*. IEEE Transactions on Biomedical Engineering, 2018. **65** (3): p. 521-527.

34. OpenCV library. [cited 2021 14 May]; Available from: https://opencv.org/ .

35. Bloice, M.D., C. Stocker, and A. Holzinger, *Augmentor: An Image Augmentation Library for Machine Learning*. [cited 2021 14 May]; Available from: https://arxiv.org/abs/1708.04680 .

36. Soomro, T.A., A. J. Afifi, J. Gao, O. Hellwich, M. Paul, and L. Zheng, *Strided U-Net Model: Retinal Vessels Segmentation using Dice Loss*, in *Digital Image Computing:*

Techniques and Applications, 2018, Canberra: Australia. p. 1-8.

37. Sorensen, T., *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons*. Biologiske Skrifter, 1948.**5**: p.1–34.

38. Jaccard, P., *Lois de distribution florale dans la zone alpine*. Bull. Société Vaudoise Sci. Nat., 1902. **38**: p. 69–130.

39. Meyer, M.I., P. Costa, A. Galdran, A.M. Mendonça, and A. Campilho, *A deep neural network for vessel segmentation of scanning laser ophthalmoscopy images*, in *International Conference on Image Analysis and Recognition*, 2017, Montreal: Canada. p. 507-515.

40. Guo, C., M. Szemenyei, Y. Yi, Y. Xue, W. Zhou, et al., *Dense residual network for retinal vessel segmentation*, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020, Barcelona: Spain. p. 1374-1378.

41. Brea, L.S., D.A. De Jesus, S. Klein, and Tv. Walsum, *Deep learning-based retinal vessel segmentation with cross-modal evaluation*, in *Proceedings of the Third Conference on Medical Imaging with Deep Learning*, 2020, Montreal: Canada. p. 709-720.

42. Kim, J.U., H.G. Kim, and Y.M. Ro, *Iterative deep convolutional encoder-decoder network for medical image segmentation*, in *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2017, Jeju: Korea (South). p. 685-688.

43. Khan, K.B., A.A. Khaliq, A. Jalil, M.A. Iftikhar, N. Ullah, et al., *A review of retinal blood vessels extraction techniques: challenges, taxonomy, and future trends*. Pattern Analysis and Applications, 2019. **22** (3): p. 767-802.

44. Oliveira, A., S. Pereira, and C.A. Silva, *Augmenting data when training a CNN for retinal vessel segmentation: How to warp?*, in *IEEE 5th Portuguese Meeting on Bioengineering*, 2017, Coimbra. p. 1-4.

45. Arpacı, S.A. and S. Varlı, *Diabetic retinopathy classification with deep learning*, in *4th International Scientific Research Congress*, 2019, Yalova: Turkey. p. 311-321.