Notice to Readers

"Chaining a U-Net With a Residual U-Net for Retinal Blood Vessels Segmentation"
By Gendry Alfonso Francia, Carlos Pedraza, Marco Aceves, and Saúl Tovar-Arriaga
Published in IEEE *Access*, Volume 8
DOI: 10.1109/ACCESS.2020.2975745

This paper includes authors who, prior to final publication, were prohibited from publishing with IEEE. Due to the nature of this violation, reasonable effort should be made to remove all past references to this paper, and refrain from future references to this paper.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 2, 2020, accepted February 15, 2020, date of publication February 21, 2020, date of current version March 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975745

Chaining a U-Net With a Residual U-Net for Retinal Blood Vessels Segmentation

GENDRY ALFONSO FRANCIA, (Student Member, IEEE), CARLOS PEDRAZA, (Senior Member, IEEE), MARCO ACEVES, (Senior Member, IEEE), AND SAÚL TOVAR-ARRIAGA, (Senior Member, IEEE)
Faculty of Engineering, Autonomous University of Querétaro, Santiago de Querétaro 76010, Mexico
Corresponding author: Saúl Tovar-Arriaga (saul.tovar@uaq.mx)

This work was supported by the National Council of Reseach of Mexico (CONACyT).

ABSTRACT Retina images are the only non-invasive way of accessing the cardiovascular system, offering us a means of observing patterns such as microaneurysms, hemorrhages and the vasculature structure which can be used to diagnose a variety of diseases. The main goal of this paper is to automate retinal blood vessel segmentation with a good tradeoff between blood vessel classification and training time in the presence of high unbalanced classes. In this work, a novel methodology is proposed using two convolutional neural networks (CNN's), chained to each other. The second CNN has been designed with residual network blocks, which joined to the information flow from the first, give us metrics like recall and F1-Score, which are, in most cases, superior to state of the art in vessel segmentation task. We tested this work on two public datasets for blood vessel segmentation in retinal images showing that this work outperforms many of other contributions by other authors.

INDEX TERMS Retina vessel segmentation, convolutional neural network, U-Net, residual block, F1-Score.

I. INTRODUCTION

Cardiovascular diseases (CVD's) are a whole of interrelated pathologies, including coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic and congenital heart diseases, which are undoubtedly a world leading cause of death [1]. These diseases represent a set of disorders of the heart and blood vessels are most prevalent in developing countries. Habits like smoking, poor diet, sedentary lifestyle, and others are factors that negatively influence CVD's. Therefore, modifications of these factors can reduce their occurrence.

The predictive models which have obtained the most acceptable results in their attempts to predict these events have been the Pooled Cohort equations [2], Framingham [3] and the SCORE project [4]. All of these models try to predict the probability of a CVD, through a follow-up period of 10 years, which avoids predicting events in short periods of time.

The retina is the only place in the human body where veins and arteries can be seen directly. This gives us the opportunity to appreciate their structure, for example the shape of microvascular changes related to the development of cardiovascular diseases (CVD) [5].

There are other CVD related patterns that can be seen in retina images. One example is the hemorrhage, a disorder of the eye in which bleeding occurs in the light-sensitive tissue on the back wall of the eye [6]. It can be related to diabetic retinopathy, which cause the formation of small fragile blood vessels, which can be easily damaged by high blood sugar level, potentially causing the growth or creation of new blood vessels [7]. Another related disease is hypertensive retinopathy, which consists of damage to the retina from high blood pressure which may result in deformation of the retinal blood vessels. The recognition of this disease may therefore be important in cardiovascular risk stratification of hypertensive patients. Hypertension affect the eyes in several ways, some examples of them are: focal arteriolar narrowing, arteriovenous nicking, hemorrhages, microaneurysms and hard exudates. Related to hypertensive retinopathy, others affection are retinal vein occlusion, artery occlusion, arteriolar emboli, and macroaneurysms [8].

Segmenting blood vessels into retinal images can give us important clues in the diagnosis of ophthalmological diseases such as diabetes, microaneurysms, arteriosclerosis and hypertension. However, this is a task that retina specialists perform

The associate editor coordinating the review of this manuscript and approving it for publication was Soon Xin Ng.

manually, a time-consuming process which allows for more margin for human error, even when performed by the most experienced doctors. In this paper, we focus on the automatic segmentation of blood vessels in retina images.

Currently, the use of deep networks has shown significant advances and proven results in the detection of patterns in highly complex images with many classes of objects, specifically with convolutional neural networks (CNN's) [9]. These results have been facilitated by increased computational capacity through graphic processing units (GPUs).

In segmentation of medical images, the U-Net architecture proposed by [10] has been successful as well as the introduction of residual networks [11]. Many variations of the CNNs have been proposed, some of these will be reviewed below.

II. BACKGROUND

Numerous works for retinal blood vessels segmentation have been proposed, some using traditional machine learning methods. Orlando *et al.* [12], used a fully connected model and a conditional random field with learned parameters of the method using a support vector machine classifier. Oliveira *et al.* [13], proposes an unsupervised method combining a matched filter, a Frangi's filter and a Gabor Wavelet filter to enhance the images.

Currently the mainstream techniques used for segmentation make use of CNNs, for example the work of Cai *et al.* [14], whose architecture is based on the VGG network [15]. Dasgupta and Singh [16] proposed a multi-label inference task, combining a CNN with a structured prediction. In [17], Alom *et al.* introduce residual blocks and complement it with recurrent residual convolutional layers. However, Zhuang in [18] went beyond this and stacked two U-Nets with residual blocks, increasing the paths for information flow. Another important work was proposed by Khanal and Estrada in [19], which used of stochastic weights to get a good balance between background pixels and vessels, by using a second reduced network, for ambiguous classification pixels. Table 1 summarizes these works and others which are based on their methods and reported performance metrics.

The present work proposes an architecture based on a common U-Net and U-Net with residual blocks, where both networks are linked. The first part consists of feature extraction, while the second focuses on the detection of new characteristics or those that were ambiguous (or could not be specified), by importing the information flow from the previous network.

III. METHODOLOGY

For the development of this work we used the DRIVE [22] and CHASE [23] databases which are free and have been used in competitions in the search and detection of different pathologies that can be identified in retinal images. These images can be used in the training and testing of a deep learning network.

The general intention of this study is to follow a workflow in which the retinal images are obtained, then subjected to pre-processing. At this point, various techniques can be applied as feature extraction, highlighting certain patterns and/or omitting others. The most important characteristics are then selected, before the image is introduced to the CNN architecture. Next comes the training and evaluation of the images passed in batch, the calculation of and the adjustment of the weights. All of this process is a cycle subject to changes in the architecture until the best results are achieved. Finally, the results can be used and interpreted by specialists.

TABLE 1. Overview of papers for retinal image segmentation. Metric reported were accuracy, recall and F1-Score among others for DRIVE dataset.

Author	Method	Metric
Oliveira *et al.* [13], 2016	Combination of filters (matched, Frangi's & Gabor Wavelet). Deformable models and Fuzzy C-means for segmentation.	Acc = 0.9402
Orlando *et al.* [12], 2016	Fully connected conditional random field with SVM.	F1 = 0.7857
Maninis *et al.* [14], 2016	CNN based on VGG network.	F1 = 0.8210
Soomro *et al.* [20], 2017	CNN.	Recall = 0.746
Dasgupta *et al.* [16], 2017	CNN with structured prediction.	F1 = 0.8074
Alom *et al.* [17], 2018	CNN with residual block and recurrent residual convolutional layers.	F1 = 0.8171
Jin *et al.* [21], 2018	Deformable CNN.	F1 = 0.7883
Khanal *et al.* [19], 2019	DNN.	F1 = 0.8259
Zhuang [18], 2019	Chain of multiple U-Nets.	F1 = 0.8202

A. PRE-PROCESSING

The successful behavior of a neural network is dependent on the preprocessing of the images before their use. Such preprocessing can improve the quality of the images and therefore allow the CNN to detect certain characteristics.

The first step is to convert the RGB input image to grayscale, thus achieving a higher contrast of blood vessels with respect to the background of the original image. The formula used was the following:

$$Img = R * 0.299 + G * 0.587 + B * 0.114 \quad (1)$$

where R, G, and B are the corresponding channels of the input image. In this equation, the green channel (G) is highlighted, because it provides the highest contrast [24].

The second step is data normalization, which is useful in classification algorithms, particularly in backpropagation neural network types. The normalization of the input values

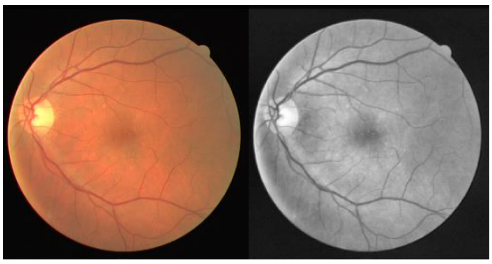

FIGURE 1. Result of preprocessing. Original image on the left, preprocessed image on the right.

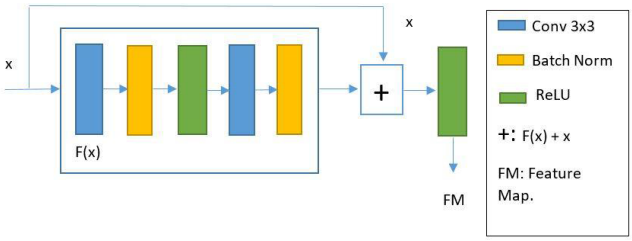

FIGURE 2. Residual block used in the present work. A feature map x passes a sequence of layers (blue: convolutional, yellow: batch normalization, green: ReLU activation). After that original characteristic x is added with results obtained through the layers. Finally, a ReLU activation is applied.

for each measured attribute in the training tuples will contribute to accelerate the training phase [25].

Two techniques were carried out, min-max and z-score. The first one develops a linear transformation of the input data, preserving the relation of the original values:

$$v' = \frac{v - min_A}{max_A - min_A} \tag{2}$$

where v' is the normalized value from the original v value and its relation with the minimum and maximum values of an attribute A. In the case of this study, attribute varies in a scale between 0 and 255.

Z-score normalization is based on the mean and standard deviation of an A attribute, and is useful when there may be outliers governing the minimum and maximum extremes of an attribute:

$$v' = \frac{v - \acute{A}}{\sigma_A} \tag{3}$$

where $\acute{A}$ and σ_A are the mean and standard deviation respectively.

The third step is to apply a contrast-limited adaptive histogram equalization (CLAHE), which is an effective method for improving the details of a grayscale retina image in a uniform manner. The final stage is the adjustment of the brightness levels through a gamma correction. This becomes necessary because many images present an insufficient or uneven brightness, which can make the details of diagnoses undetectable [26].

The result of the pre-processing stage can be seen in figure 1:

Patches are then extracted, allowing for the acquiring of larger scale data sets and thus facilitating the training of arbitrarily configured neural networks. This procedure can be considered as an increase in data and is complemented with the random horizontal and vertical flipping of these patches in each iteration [27].

B. ARCHITECTURE

In this work a fully double convolutional neural network was used, with the peculiarity that the second uses a residual network. The original structure, also known as U-Net, was proposed by [10] and is composed of an encoding part, where both semantic and contextual information is captured as the number of filters is doubled; as well as a second decoding part, where the spatial information is restored and the characteristics obtained at the lowest levels are merged.

- **U-Net 1**: It is composed of four levels where a double convolution is repeatedly applied with a 3×3 filter, with a padding equal to zero, a stride equal to one and a max pooling operation with a 2×2 filter and stride of two. Input image had one channel and 572×572 of size. In the first level the numbers of channels used were set in 24 and the image size was 568×568 after double convolutions. This process was repeated by three levels and the results for each level were 48 channels and 280×280, 96 channels and 136×136, 192 channels and 64×64, in the bottleneck 384 channels and 28×28. The decoding part follows the previous structure, but in the opposite direction. The numbers of channels were the same of contracting part but image size get reduced, the values for each level were 52×52, 100×100, 196×196 and 388×388. It also adds the characteristics found in each of the different encoding levels. After each convolution layer, a batch normalization layer and a ReLU activation function are applied, significantly improving the training time as well as the stability and generalization capacity of the model [28].
- **Residual block**: the idea arises from a degradation problem that comes when models begin to converge, as demonstrated in [11]. To reduce this problem, a residual block was proposed where the expected output is composed of:

$$FM\ (x) = F\ (x) + x \tag{4}$$

  where $FM\ (x)$ is the expected feature map, from applying two convolutional layers to the input features, represented as $F\ (x)$ and to this transformation, the original input x was added. The addition of the original feature map attenuates the problem of degradation that appears in the models. Figure 2 details the process used in this work.
- **U-Net 2 with Residual blocks**: The output of the first network, as well as the maps of characteristics of each level constitute the entrance to this second one. The number of channels and image size at each level kept

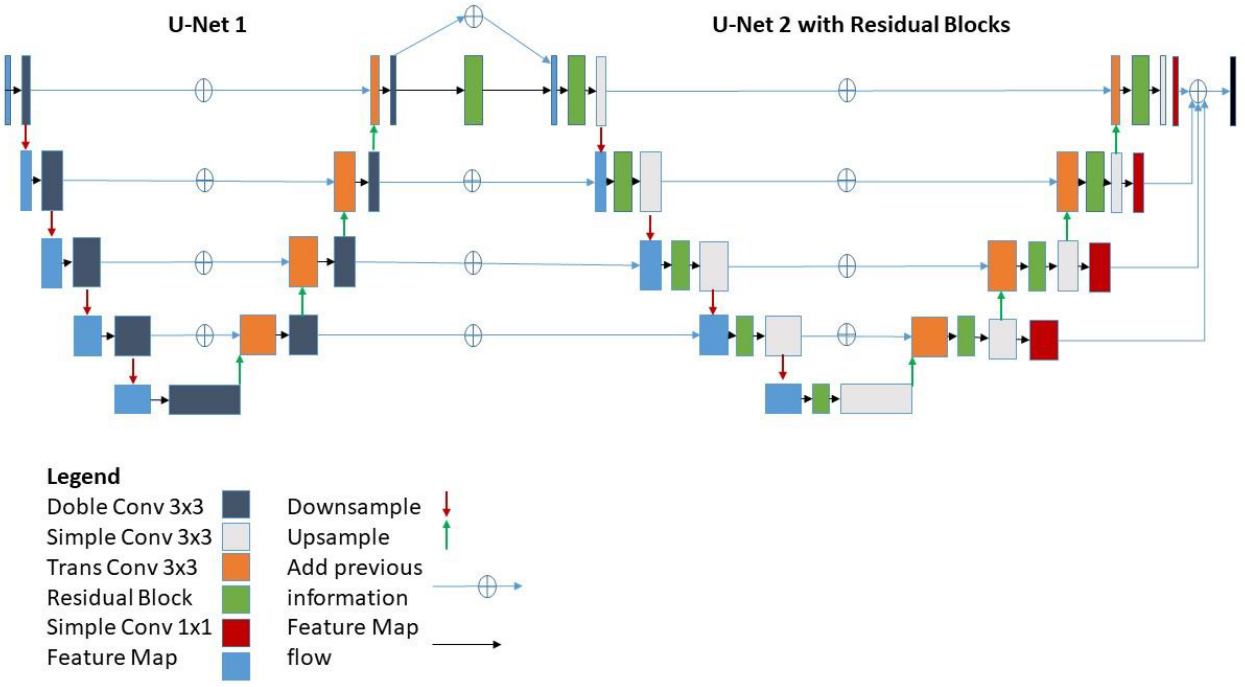

FIGURE 3. Proposed architecture. It consists of a U-net following with a U-net with residual blocks. In the first U-net, feature extraction is carried out, while in the second U-net with residual blocks the detection of new characteristics or those that were ambiguous is performed.

equal to decoding part of first half of the architecture proposed. The main difference is that a residual block is added at each new level, both contraction and expansion. This time a padding equal to one is used and in the last layer a 1×1 filter convolution is applied in order to produce an output of two channels, since a binary classification task is carried out (see Figure 3).

This classification problem has the peculiarity that it presents unbalanced classes, because the majority of pixels represent the background of the images, while the minority represent the vasculature of a retina image; for this reason, a negative log-likelihood loss function will be used. This type of loss function is defined by the following equation:

$$loss\,(x, y) = -\,(logy) \tag{5}$$

This function maximizes the overall probability of the data, by giving a high loss value when classification is wrong or unclear and a low loss value when prediction matches the expected by the model. The logarithm performs the penalizing part, the lower the probability, the greater the logarithm. Since these probabilities have values between zero and one, and the logarithms in that range are negative, the negative sign is used to convert them into positive values.

To handle the problem of class unbalance, the weight attribute is provided, and each class is assigned both the prediction and the reference. The approach proposed in [19] is follow

$$w_{rand\,(1,\alpha,s)} \tag{6}$$

where the weight w, varies randomly between one and the value of α with a step equal to s. This dynamic variation of the weights prevents the net from falling into local minima. In order to obtain the logarithmic probabilities, the *LogSoftmax* function is applied to the last layer of the neural network.

IV. MATERIALS AND EXPERIMENTATION

For the present investigation we worked with two public databases for the segmentation of retina images, DRIVE and CHASEDB that are described below:

- DRIVE: Consists of 40 color images of retina, with dimensions of 565×584. This set is already divided into 20 images for training, which were separated into 15 to train the proposed neural network and 5 to validate them, as well as 20 other images for tests.
- CHASEDB: Consists of 28 images of retina of 14 children, centered in the optic nerve, each with a dimension of 999×960 pixels.

The equipment used in all tests is a PC with an Intel(R) Core (TM) i5-8400 processor CPU@2.80 GHz, with 16 GB RAM and an NVIDIA GeForce GTX 1070 graphics card, with 8 GB VRAM.

The neural network was initialized according to the parameters proposed by [19], and ADAM optimization was used, proposed in [29], with a learning rate of 0.001. The setting of a learning rate value is the most important task for the configuration the model, yet unfortunately this value cannot be analytically calculated and the closest value must instead be found by means of trial and error. Normally the learning

rate is less than 1.0 and greater than 10^{-6}, learning rate decay was applied and was adjusted every 30 epochs to a ratio of 0.7. In total, 250 epochs were performed with an early stop configured to 35 iterations with no change in the loss function. Batch size is also an important hyperparameter, as is the definition of the number of samples that will be propagated through the network. In this case selected value was four since less memory is required and typically networks train faster with mini-batches because the weight is updated after each propagation.

For training as well as for validation and testing, the default configuration was respected in DRIVE. However, CHASEDB used the cross validation technique using a stratified 5-fold configuration, which was generated randomly, such that in each fold 5 images were left for testing and the remaining 23 images were split into training and validation. The validation process was repeated 5 times with the same architecture and trained from scratch for each fold. Finally, the results were then averaged to produce a single estimation. The value of 5 for k in k-fold was set empirically and according to the proposals of most of the state of the art, a higher value for k would have led to more training time. All the implementation was done in PyTorch (version 1.1.0), which is an open-source automatic learning library, used for applications such as computer vision.

V. EVALUATION AND RESULTS

After building a model, it is of upmost importance to evaluate its performance. The problem presents the peculiarity of a high class unbalance, so the selection of the correct metrics may be critical. According to [30], recall or sensitivity, precision and the harmonic mean between these two metrics called F1-Score are suitable, useful metrics when working with unbalanced classes. In addition, we include accuracy. Their calculation is based on four possible interpretations of the data:

- ➢ True Positive (TP): The sample label is positive and is classified as such.
- ➢ True negative (TN): the sample label is negative and is classified as such.
- ➢ False positive (FP): the sample label is negative, but is classified as positive.
- ➢ False negative (FN): the sample label is positive, but is classified as negative.

Based on these values the metrics mentioned above can be calculated:

- ➢ Recall: tells us how many relevant samples are selected.

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

- ➢ Precision: tells us how many predicted samples are relevant.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

TABLE 2. Comparison with other state of the art works on DRIVE.

Method	Precision	Recall	F1-Score	Accuracy
Orlando et al [12]	0.7854	0.7897	0.7857	-
Maninis et al [14]		**0.8261**	0.8210	0.9541
Dasgupta et al [16]	-	0.7691	0.8074	0.9533
Yang et al [31]	-	0.7631	-	0.9538
U-Net [17]	-	0.7537	0.8142	0.9531
Residual U-Net [17]	-	0.7726	0.8149	0.9553
Recurrent U-Net [17]	-	0.7751	0.8155	0.9556
R2U-Net [17]	-	0.7792	0.8171	0.9556
Zhuang [18]	-	0.7856	0.8202	0.9561
Jin et al [21]	**0.8529**	0.7963	0.8237	0.9566
Khanal et al [19]	0.8284	0.8235	**0.8259**	0.9693
Proposed	0.8341	0.8160	0.8250	**0.9696**

TABLE 3. Comparison respect to training time.

Architecture	DRIVE	CHASEDB
Khanal et al [19]	1:34 hours	19:46 hours
Proposed	0:37 hours	9:08 hours

- ➢ F1-Score: is the harmonic mean between recall and precision.

$$F1_{Score} = \frac{2 * TP}{2 * TP + FP + FN} \tag{9}$$

- ➢ Accuracy: measures how many observations, both positive and negative, were correctly classified.

$$ACC = \frac{(TP + TN)}{TP + FP + FN + TN} \tag{10}$$

In Table 2 we can see a comparison of several similar works based on the DRIVE dataset, according to the metrics described above.

One can conclude, from the balanced values of precision and recall obtained by using the F1-Score metric and a high value, that we were properly classifying veins or background pixels of the input image. As it can be seen, in the comparative table 2, our proposal obtained the second-highest margin of F1-Score and the highest accuracy among all related works, it can be concluded that this classification proposal was in most cases successful, and not at the expense of increasing the number of false positives and false negatives. It should also be noted that under the technical conditions in which the experiments of this study were carried out, the execution time was reduced by 57 minutes on the set of DRIVE images, and by 10 hours 38 min on CHASEDB, in comparison to [19]. This study therefore achieved similar results in a shorter processing period, as shown in Table 3.

Table 4 shows the same comparisons in the case of the CHASEDB dataset:

Figures 4 and 5, show qualitative outputs of proposed architecture. In Figure 4 one can see an example of a result

TABLE 4. Comparison with other state-of-the-art works on CHASEDB.

Method	Precision	Recall	F1-Score	Accuracy
Orlando et al [12]	0.7438	0.7277	0.7332	-
Yang et al [31]	-	0.7641	-	0.9607
U-Net [17]	-	**0.8288**	0.7783	0.9578
Residual U-Net [17]	-	0.7726	0.7800	0.9553
Recurrent U-Net [17]	-	0.7459	0.7810	0.9622
R2U-Net [17]	-	0.7756	0.7928	0.9634
Zhuang [18]	-	0.7978	0.8031	0.9656
Jin et al [21]	0.7630	0.8155	0.7883	0.9610
Khanal et al [19]	**0.8550**	0.8143	0.8245	0.9759
Proposed	0.8366	0.8258	**0.8312**	**0.9766**

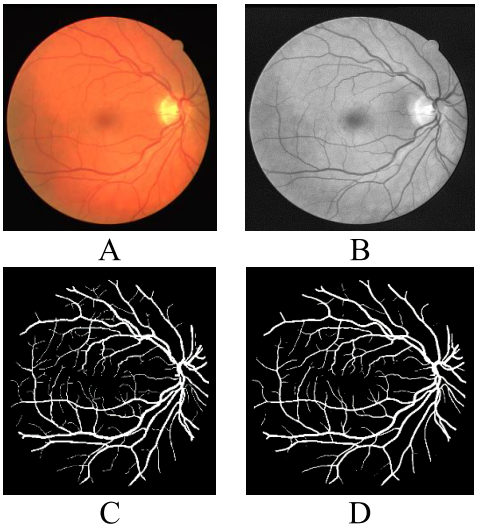

FIGURE 4. Test results on DRIVE dataset. A: original image, B: preprocessed image, C: ground truth and D: segmented output.

obtained from the DRIVE dataset. In Figure 5, the result for CHASEDB is shown.

SSIM metrics were performed in order to make understandable the segmentation process, and due to the difficulty of differentiating between a first phase with only U-Net and a second one with the addition of the residual blocks. The structural similarity index (SSIM) analyzes the viewing distance, edge information between the reference and the test image. Is a perceptual metric that quantifies image quality degradation caused by processing such as data compression [32]. The SSIM ranges from 0 (completely different) to 1 (identical patches). The higher the value, the better. Figure 6 shows a comparison between ground truth and a result from the first U-Net, whilst Figure 7 compares the ground truth with a result from all proposed architecture over DRIVE database.

In the following image a detailed piece of segmentation on images from DRIVE can be observed.

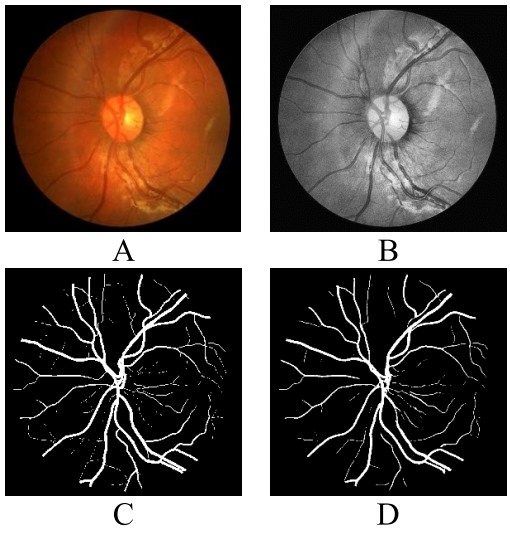

FIGURE 5. Test results on CHASEDB dataset. A: original image, B: preprocessed image, C: ground truth and D: segmented output.

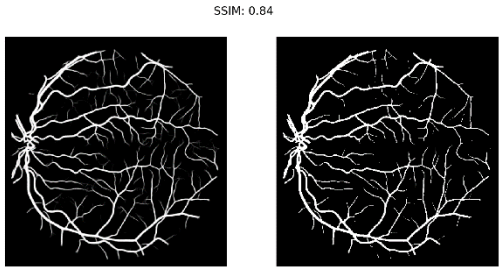

FIGURE 6. SSIM value for ground truth and result from U-Net 1.

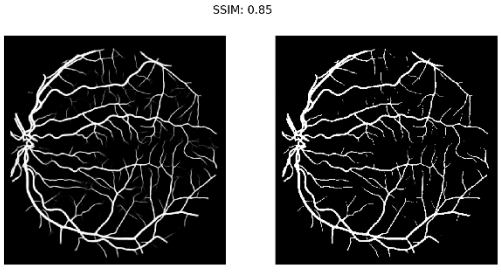

FIGURE 7. SSIM value for ground truth and result from entire architecture.

The effect of lesion near vessels was also evaluated in the segmentation process. The images used for experimentation were taken from the DRIVE and CHASE databases. In the 40 photographs found in DRIVE, 33 present no sign of illness and the remaining 7 present signs of mild early diabetic retinopathy and pigment epithelium changes, pigmented scar in fovea, or choroidiopathy. None of the images in CHASE_DB show any presence of lesion. In the images which presented signs of lesion, the segmentation was performed without any problem. The results can be observed in the following images.

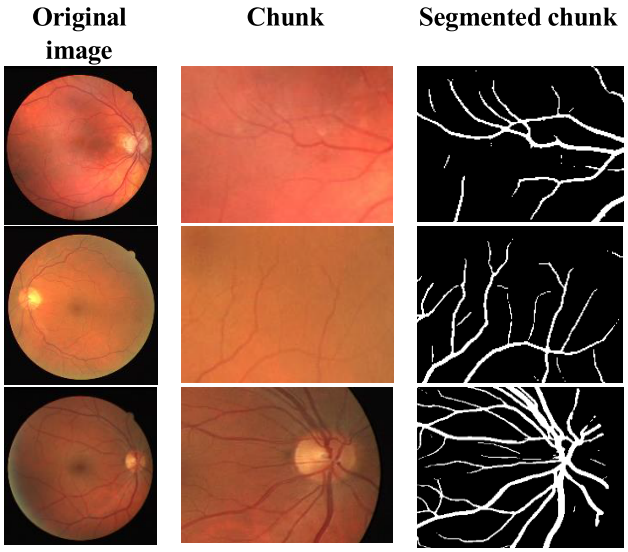

FIGURE 8. Piece of segmentation where thin, medium and large vessel structures can be seen.

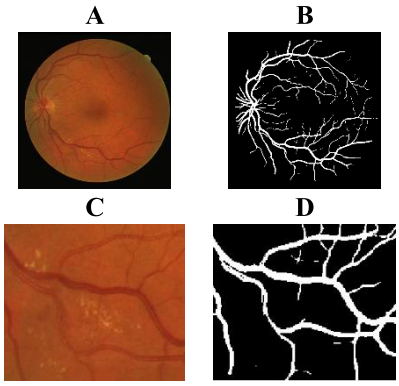

FIGURE 9. Test results on CHASEDB dataset. A: Original image with background diabetic retinopathy, B: Segmented image, C: Area of the lesion and D: Area of the lesion in segmented image.

As one can see in these images, there is no evidence of any lesion in the segmented area. One can therefore conclude that this level of lesion does not affect the segmentation of vessels.

VI. CONCLUSION

The task of achieving precise segmentation is an arduous one, which has a high processing cost in algorithm training. The main contribution of this work is the addition of a new U-Net network, connected to the first one, with the peculiarity that residual blocks were added to it, therefore attenuating the degradation problem. Moreover, connections are established at all levels, so that the information obtained in each of the previous levels is added to the new characteristics identified. Likewise, for the final output, a coupling of each level of this last U-Net network with residual blocks is made. This constant flow of information allows us to avoid or minimize the natural information loss that occurs in the contraction of images.

The results of this study are very similar to those of the highest-performance methods, however they were obtained with a considerably shorter training time. We were able to reduce this training time thanks to the proposed architecture, in which two fully-connected convolutional neural networks were linked, with an encoder-decoder design, with the peculiarity of adding residual network blocks in the second one.

The pre-processing of input images is very helpful in the task of segmentation, which consists of working with grayscale images, normalizing them, applying CLAHE and gamma adjustment. Due to the few images available, the increase in data was also significant, this time working with patches of the original data and flipping them randomly.

The use of dynamic weights rounded the final result, achieving a high F1-Score and precision values, from which one can conclude that the segmentation task was carried out with high reliability values.

The experiments were conducted on two public datasets, DRIVE and CHASEDB. The research team hopes to continue experimenting, for example increasing the number of filters in order to determine if new patterns can be detected.

ACKNOWLEDGMENT

The authors would like to thank the Mexican Advanced Imaging Laboratory for Ocular Research (MAILOR) and the Laboratorio Nacional de Visualización Científica Avanzada (LAVIS) for their advice.

REFERENCES

[1] J. Stewart, G. Manmathan, and P. Wilkinson, "Primary prevention of cardiovascular disease: A review of contemporary guidance and literature," *JRSM Cardiovascular Disease*, vol. 6, Jan. 2017, Art. no. 204800401668721.

[2] D. Preiss and S. L. Kristensen, "The new pooled cohort equations risk calculator," *Can. J. Cardiol.*, vol. 31, no. 5, pp. 613–619, May 2015.

[3] R. B. D'Agostino, R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro, and W. B. Kannel, "General cardiovascular risk profile for use in primary care: The framingham heart study," *Circulation*, vol. 117, no. 6, pp. 743–753, Feb. 2008.

[4] R. Conroy, "Estimation of ten-year risk of fatal cardiovascular disease in europe: The SCORE project," *Eur. Heart J.*, vol. 24, no. 11, pp. 987–1003, Jun. 2003.

[5] D. S. W. Ting and T. Y. Wong, "Eyeing cardiovascular risk factors," *Nature Biomed. Eng.*, vol. 2, no. 3, pp. 140–141, Mar. 2018.

[6] M. B. Wankhade, "Analysis of disease using retinal blood vessels detection," *Int. J. Eng. Comput. Sci.*, vol. 5, no. 12, pp. 19644–19647, Dec. 2016.

[7] T. Laibacher, T. Weyde, and S. Jalali, "M2U-Net: Effective and efficient retinal vessel segmentation for resource-constrained environments," 2018, *arXiv:1811.07738*. [Online]. Available: https://arxiv.org/abs/1811.07738

[8] T. Y. Wong and P. Mitchell, "The eye in hypertension," *Lancet*, vol. 369, no. 9559, pp. 425–435, Feb. 2007.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1–9.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[12] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 16–27, Jan. 2017.
[13] W. S. Oliveira, J. V. Teixeira, T. I. Ren, G. D. C. Cavalcanti, and J. Sijbers, "Unsupervised retinal vessel segmentation using combined filters," *PLoS ONE*, vol. 11, no. 2, Feb. 2016, Art. no. e0149943.
[14] J. Cai, L. Lu, Z. Zhang, F. Xing, L. Yang, and Q. Yin, "Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, vol. 1. Cham, Switzerland: Springer, 2016, pp. 140–148.
[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556
[16] A. Dasgupta and S. Singh, "A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 248–251.
[17] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*. [Online]. Available: https://arxiv.org/abs/1802.06955
[18] J. Zhuang, "LadderNet: Multi-path networks based on U-net for medical image segmentation," 2018, *arXiv:1810.07810*. [Online]. Available: https://arxiv.org/abs/1810.07810
[19] A. Khanal and R. Estrada, "Dynamic deep networks for retinal vessel segmentation," 2019, *arXiv:1903.07803*. [Online]. Available: https://arxiv.org/abs/1903.07803
[20] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, M. A. U. Khan, M. Paul, and L. Zheng, "Boosting sensitivity of a retinal vessel segmentation algorithm with convolutional neural network," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2017, pp. 1–8.
[21] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, Aug. 2019.
[22] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
[23] C. G. Owen, A. R. Rudnicka, R. Mullen, S. A. Barman, D. Monekosso, P. H. Whincup, J. Ng, and C. Paterson, "Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (CAIAR) program," *Investigative Opthalmol. Vis. Sci.*, vol. 50, no. 5, p. 2004, May 2009.
[24] N. El Abbadi and E. Al Saadi, "Automatic early diagnosis of diabetic retinopathy using retina fundus images," *Eur. Acad. Res.*, vol. 2, no. 9, pp. 11397–11418, 2014.
[25] J. H. M. Kamber, *Data Mining: Concepts and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2006.
[26] M. Zhou, K. Jin, S. Wang, J. Ye, and D. Qian, "Color retinal image enhancement based on luminosity and contrast adjustment," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 521–527, Mar. 2018.
[27] T. B. Sekou, M. Hidane, J. Olivier, and H. Cardot, "From patch to image segmentation using fully convolutional networks—Application to retinal images," 2019, *arXiv:1904.03892*. [Online]. Available: https://arxiv.org/abs/1904.03892
[28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 1, 2015, pp. 448–456.
[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[30] A. Tharwat, "Classification assessment methods," in *Applied Computing and Informatics*. Riyadh, Saudi Arabia: King Saud Univ., 2018.
[31] Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1427–1436, Jul. 2019.
[32] G. P. Renieblas, A. T. Nogués, A. M. González, N. Gómez-Leon, and E. G. del Castillo, "Structural similarity index family for image quality assessment in radiological images," *J. Med. Imag.*, vol. 4, no. 3, Jul. 2017, Art. no. 035501.

GENDRY ALFONSO FRANCIA (Student Member, IEEE) received the B.Sc. degree in informatics from the University of Informatics Science, Cuba. He is currently pursuing the Master of Science degree in artificial intelligence wtih the Autonomous University of Querétaro. His research interests include automatic illness diagnosis, machine learning, computer vision, and deep learning.

CARLOS PEDRAZA (Senior Member, IEEE) received the B.Sc. degree in electronics engineering from the Celaya Institute of Technology, Mexico, the M.Sc. degree in electrical engineering from the FIMEE Universidad de Guanajuato, and the Ph.D. degree in mechanical engineering from the University of Tsukuba, Japan. He is a full-time Professor with the Engineering Faculty, Autonomous University of Querétaro, Mexico. He has been a member of the National Systems of Researchers (SNI), since 2004, and the Mexican Science Academy (AMC), since 2007. His research interests include machine vision, 3D reconstruction, image processing, and deep and machine learning.

MARCO ACEVES (Senior Member, IEEE) received the B.Sc. degree in telematics engineering from Colima University, Mexico, and the M.Sc. and Ph.D. degrees in the field of intelligent systems from the University of Liverpool, U.K. He is a full-time Professor with the Engineering Faculty, Autonomous University of Querétaro, Mexico. His research interest includes intelligent and embedded systems. He has been a member of the National Systems of Researchers (SNI), since 2009. He is also the Honorary President of the Mexican Association of Embedded Systems (AMESE) and a Board Member for many Institutions and Associations.

SAÚL TOVAR-ARRIAGA (Senior Member, IEEE) received the B.Sc. degree in electronics engineering from the Queretaro Institute of Technology, Mexico, the M.Sc. degree in mechatronics from the University of Siegen, Germany, and the Ph.D. degree from the University of Erlangen–Nuremberg, Germany. He is currently a full-time Professor with the Engineering Faculty, Autonomous University of Querétaro and the Manager of the master's degree in artificial intelligence. His research interests include automatic illness diagnosis, surgical robotics and machine learning applications. He has been a member of the National Systems of Researchers (SNI), since 2013. He is also the Chair of the IEEE Computational Intelligence Society Queretaro Chapter and the past President of the IEEE Queretaro Section.