Retinal Blood Vessel Segmentation Using Pix2Pix GAN*

Dan Popescu, *Member, IEEE*, Mihaela Deaconu, Loretta Ichim, *Member, IEEE*, and Grigore Stamatescu, *Member, IEEE*

Abstract— **In the past few years, scientists have developed new models that are based on the U-Net structure and require a small number of ground truth images for training. Regarding retinal blood vessel segmentation, these models have proven higher accuracy in small vessels and microvascularization detection. After investigating GAN-type neural networks, the authors chose the Pix2Pix structure to implement the neural network for segmenting blood vessels in retinal images. For exemplification, we trained and tested the GAN model using retinal images from three public databases: CHASE, DRIVE and STARE. Training and testing of Pix2Pix GAN, as well as image preprocessing, were performed in Matlab R2020a. Globally, the proposed neural network had an accuracy of 92.36%, considering a total of 188 images for training and testing.**

I. Introduction

Diagnosis of retinal diseases is mainly based on examination methods such as ophthalmoscopy and fundus photography. Diabetic retinopathy, hypertensive retinopathy, central retinal vessel occlusion, and hemorrhages are some of the retinal vascular conditions which require specialized clinical care and efficient equipments for detection. Different methods are proposed for vessel segmentation in retinal images. In [1] two simple methods are purposed for vessel detection and segmentation in images taken from DRIVE and STARE databases: one based on Gabor wavelets, and other based on thresholding. Vessel segmentation and evaluation were used in the optic disc detection because the number of vessel pixels is greater in the optic disc region. For this purpose in [2] the blood vessels were segmented using a thresholding method. In [3] the classification accuracy for different usually used feature and for different classical classifiers were investigated. The best performance was obtained for multiscale rotation invariant LBP (Local Binary Pattern) as feature and rotation forest algorithm as classifier.

In the past decade medical imaging techniques have largely benefitted from artificial intelligence models, like neural networks. Scientists have proposed numerous classification and detection methods based on convolutional neural networks (CNNs), thus leading to fast advancements in the development of high-performance diagnostic imaging equipments. Thus, the authors in [4] proposed a CNN architecture for blood vessel segmentation in retinal images. The network structure is designed on 7 layers using MatConvNet. The retinal image was partitioned in patches of dimension 27 × 27 pixels using the sliding box algorithm. The whole network was trained using a number of 400,000 patches. This CNN was implemented using GPU Programming in MATLAB. A dense conditional random field was used in [5] to segment the blood vessels in retinal images from DRIVE and STARE databases.

Retinal blood vessels are intricate anatomical structures, having different shapes and diameters across the vascular tree. Automatic detection of such structures is a complex process, and basic CNNs are not able to perform fine segmentation of small blood vessels. Furthermore, training such networks is time-consuming and requires a large set of ground truth images, which are not available for all retinal images in public databases. In order to overcome these drawbacks, researchers have created neural networks that are specialized in such medical image segmentation. These models include a U-Net architecture and require fewer images for training; in addition, they produce more precise segmentation maps, especially in the case of retinal capillaries.

In this paper, we study the working principles of basic GANs and review some existing models which incorporate U-Net and GAN architectures. Then, we propose and evaluate the performance of a Pix2Pix GAN using retinal images from three publicly available databases: CHASE, DRIVE and STARE.

II. Related Works

A. U-Net Neural Network

U-Nets are a type of fully CNNs, generally used for semantic segmentation. The first U-Net model (Fig. 1) was created in 2015 by German researchers Olaf Ronneberger, Philipp Fischer, and Thomas Brox, for the purpose of automatic segmentation of biomedical images [6]. The network is named after its conformation, which resembles the "U" letter. The architecture comprises a descending path, called encoder, and an ascending path, called decoder. The encoder is a simple CNN, which performs convolutions and pooling operations on the input image, resulting in a feature vector. The decoder performs normal convolutions and transposed convolutions, also known as "up-convolutions". Thus, the size of the feature vector is gradually increased until it reaches the size of the input image [6], [7].

Throughout the up-sampling process, the feature information extracted in each convolution layer of the encoder is concatenated with the up-sampled feature map in the corresponding up-convolution layer in the decoder. Therefore, U-Net produces a more accurate segmentation map.

*Research has been funded by University POLITEHNICA of Bucharest

D. Popescu is with the Faculty of Automatic Control and Computer Science, University Politehnica of Bucharest, 060042 Bucharest, Romania (phone: +40766218363; dan.popescu@upb.ro).

M. Deaconu, L. Ichim, and G. Stamatescu are with the Faculty of Automatic Control and Computer Science, University Politehnica of Bucharest, 060042 Bucharest, Romania (g.deaconu@gmail.com, loretta.ichim@upb.ro, and grigore.stamatescu@upb.ro).

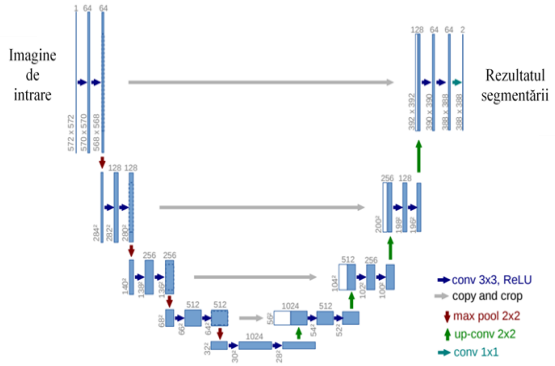


Figure 1. U-Net architecture [6].

B. GAN Neural Network

The GAN model was proposed in 2014 by Ian Goodfellow and his colleagues [8]. This model consists of two neural networks which are trained simultaneously in order to generate realistic artificial data based on the input.

The first network is a generator that models a transfer function. It takes noise as input and generates data, following the target data distribution as closely as possible. The second network in the GAN model is a discriminator which must distinguish between real and generated data.

Ideally, the total loss of GAN is 0.5 and it is based on the following equation (1):

$$\min_G \max_D L(G,D) = P_{x\sim r}[\log D(x)] + P_{n\sim g}\log([1 - D(G(x,n))]) \quad (1)$$

where x is the input data, n is the noise, $P_{x\sim r}$ is the probability that data are real, and $P_{n\sim g}$ is the probability that data are generated. The objective of the generator is to minimize the error between real and generated data, while the discriminator must detect the differences, thus maximizing the error [8], [9].

C. Conditional GANs

Conditional GANs (cGANs) were developed in 2014 by Mehdi Mirza and Simon Osindero [10]. These networks are characterized by an additional input layer in the generator and discriminator nets. In the context of image generation and classification, the generator takes noise as input and produces a close copy of the target image. In this case, the information in the additional input can be a feature vector or a label which specifies the attributes on which the generator should focus. Given the target image and the conditioning information, the discriminator can better distinguish between the real and generated images, because it also checks whether the image contains the specified features.

Conditional GANs have proven high performance in image generation based on object features. In the process of image segmentation, cGANs learn to generate accurate segmentation maps of objects and features of interest [11].

D. U-GAN Neural Networks

Compared to a simple U-Net, a neural network which comprises both GAN and U-Net architectures perform much better in image segmentation applications. Many researchers have obtained notable results in retinal blood vessel segmentation, using GAN architectures with a U-Net type generator.

Jaemin Son et al. developed a cGAN with a U-Net generator, for blood vessels and optic disc segmentation in retinal images [12]. In their model called RetinaGAN, the generator takes a retinal image as input and must generate a segmentation map of blood vessels or optic disc. The discriminator's input is a pair made of a retinal image and its segmented image (the real mask). Using ground truth images as conditional information, RetinaGAN can perform fine segmentation of blood vessels. After the training process, good performances were obtained and the network managed to detect the microvascularization, which is usually difficult to observe.

Cong Wu et al. proposed a similar model, called U-GAN, which is based on the same architecture as RetinaGAN, and also good performances were reported in retinal blood vessel segmentation [13]. The generator is a U-Net with a depth of 3, and the discriminator is a CNN comprising dense block layers, where each layer receives at input the outputs from all preceding layers.

E. Pix2Pix GAN

In 2016, Phillip Isola et al. developed a conditional GAN, called Pix2Pix (pixel to pixel), which is able to perform image to image translation tasks [11]. This technique refers to generating new images based on reference and target images and had numerous applications such as: image enhancement, image super-resolution, black-and-white image coloring, feature transfer, fine image segmentation, etc.

Pix2Pix network contains a U-Net type generator, but unlike ordinary GANs, it does not take noise as input. The input for Pix2Pix GAN (Fig. 2) consists of a reference and a target image, and the noise is indirectly introduced by the dropout layers during network training and testing [11], [14]. The discriminator is a PatchGAN. Basically, it is a deep CNN that performs classification, but instead of processing the entire image at once, it classifies patches of the input image as real or generated. Usually, the patch size is set to 70×70 pixels, and the final prediction is calculated as the mean of predictions obtained for each patch. For a cGAN, the loss function is (2):

$$L_{cGAN}(G,D) = P_{x\sim r}[\log D(x,r)] + P_{n\sim g}\log([1 - D(G(x,n))]) \quad (2)$$

In the case of Pix2Pix GAN, the discriminator is trained like in the simple GAN case, having the purpose to maximize the error between real and generated images.

The loss function of the generator is a combination of adversarial loss and L1 function. The L1 loss function (3), also known as the least absolute error, minimizes the sum of errors and it is chosen here because it produces less blur in the generated images [11]:

$$L_{L1}(G) = E_{x,r,n}[\|r - G(x,n)\|_1] \quad (3)$$

where, r is the real (target) image and $G(x,n)$ is the generated image. The final loss function of Pix2Pix GAN is (4):

$$L_{Pix2Pix} = arg\ \min_G \max_D L_{cGAN}\ (G, D) + \lambda L_{L1}(G) \quad (4)$$

where, λ is a hyperparameter that specifies the weight of L1 error [11], [14].

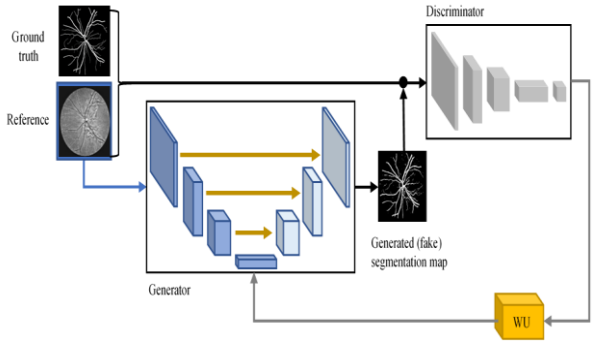


Figure 2. Pix2Pix GAN architecture (training phase).

Compared to U-Net, the Pix2Pix GAN generator has two ways of updating its weights in the convolutional filters during training: one is the internal circuit, similar to U-Net, which uses an improved backpropagation through the skip connections, and the other is an external path (WU – weight update in Fig. 2) which provides it with the comparison results between ground truth and fake images from the discriminator. Thus, the generator learns to produce segmentation maps that imitate more closely the target images.

III. Materials and Method

Considering different databases, in this section we describe the method for segmentation of blood vessels in retinal images using Pix2Pix GAN. The process consists of three main steps: image preprocessing, network training and network testing. The training and testing results are discussed in the next section.

A. Materials

The images used for training and testing are taken from CHASE, DRIVE and STARE databases. Some examples are given in Fig. 3.

CHASE database (CHASE-DB1) [15] contains 28 retinal images with a resolution of 960×999 pixels. Each image has two manual segmentations of blood vessels, done by two different experts.

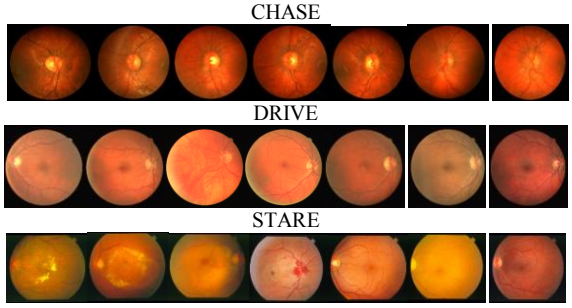


Figure 3. Examples of learning images from different databases.

DRIVE (Digital Retinal Images for Vessel Extraction) [16] is a set of 40 images, where 33 are normal and 7 show signs of diabetic retinopathy. The set is divided into 20 images for training and 20 images for testing and only manual segmentation is available for each training image. The image resolution is 565×584 pixels.

STARE database (STructured Analysis of the Retina) [17] contains 400 images of retina with different diagnoses. From this dataset, 20 images are selected for blood vessel segmentation. They have a resolution of 605×700 pixels and are provided with two manual segmentations from experts.

B. Retinal Image Preprocessing

Usually, the goal of image preprocessing is contrast enhancement and reduction of network training time. Further, we present our preprocessing method, but it is worth mentioning that this stage can be skipped, especially in the case of more advanced neural networks, like GANs. Fig. 4 shows the enhancement steps of the image representation in the preprocessing stage consisting in: green channel extraction, adaptive histogram equalization with a limited contrast amplification to reduce the noise amplification (CLAHE), and gamma correction.

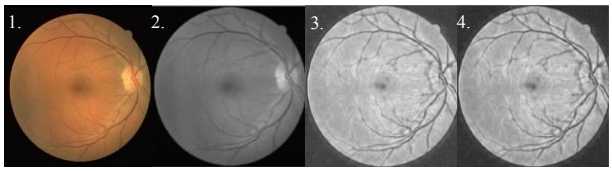


Figure 4. Retinal image enhancement: 1. Original image; 2. Green channel extraction; 3. CLAHE; 4. Gamma correction.

The first step in retinal image preprocessing is the normalization. The neural network will work with values ranging from 0 to 1 and therefore the computation complexity will be decreased [18]. Next, we extract and process the green channel, because it provides the best contrast between blood vessels and background. Some marginal rows and columns are removed in order to reduce the black area in the field of view and to obtain square images. Afterwards, the images are resized to 512×512 pixels using bicubic interpolation [19].

For further enhancement of contrast in the green channel, CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm is applied. The image is divided into blocks of 32×32 pixels, and for each block the adaptive histogram equalization is computed. The limit of the contrast was set to 0.2 and the alpha parameter of the Rayleigh distribution was set to 1 [20]. Finally, a gamma correction is performed, with a gamma value of 1.2 [18].

C. Network Training

For the training set we selected 20 images from CHASE database, 30 images from DRIVE database, 100 images from STARE database, and their corresponding ground truth annotations from one expert. The total numbers of training images are of 150 and the numbers of testing images are of 38 for all databases.

As we mentioned, the network was implemented according to the method described by Phillip Isola et al. [11].

Fig. 2 shows the Pix2Pix GAN architecture in the context of retinal blood vessel segmentation [21]. The generator is a U-Net type network with a depth of 8 and it takes a pair of images as input the preprocessed retinal images. The network uses Leaky ReLU activation function in the encoder and ReLU activation function in the decoder and aims at minimizing the L1 error between the ground truth and generated segmentations.

The discriminator is a convolutional PatchGAN that takes the ground truth and generated images as input. It tries to maximize the error, thus forcing the generator to produce more accurate images.

Training and testing of Pix2Pix GAN, as well as image preprocessing, were performed in Matlab R2020a. The training options are set as follows: optimization algorithm is Adam, initial learn rate is 0.0002, λ parameter for L1 loss is 100, maximum number of epochs for training is 200, each epoch having 60 iterations (one iteration correspond to one learned pair of images). The experiments were performed on a laptop with the following hardware specifications: Intel Core i5-7200U CPU @ 2.50GHz, 4GB RAM, Windows 10 Pro 64-bit operating system. Given these conditions, the training time for Pix2Pix GAN was approximately 48 hours.

For performance evaluation we calculated the accuracy, precision, sensitivity, and specificity, based on the following formulas (5)-(8):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

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

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

where TP (true positive) – correctly generated blood vessel pixel in the segmentation map, TN (true negative) – correctly generated background pixel, FP (false positive) – background pixel incorrectly translated to blood vessel pixel, FN (false negative) – blood vessel pixel incorrectly translated to background pixel.

IV. Experimental Results

Fig. 5 and Fig. 6 show the evolution of loss functions during training. Analyzing the plots, we can observe that L1 loss decreased during training, but the total GAN loss was higher than 0.6, because the discriminator learned slightly faster than the generator.

The network was tested on 8 retinal images from CHASE dataset, 10 retinal images from DRIVE dataset, and 20 retinal images from STARE. For exemplification a number of 12 images (I_1- I_8 from CHASE and I_9 - I_12 from DRIVE) are presented in Fig. 7.

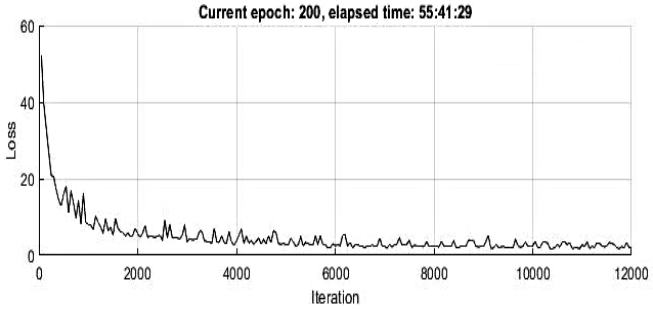


Figure 5. Pix2Pix GAN training plots. Generator total.

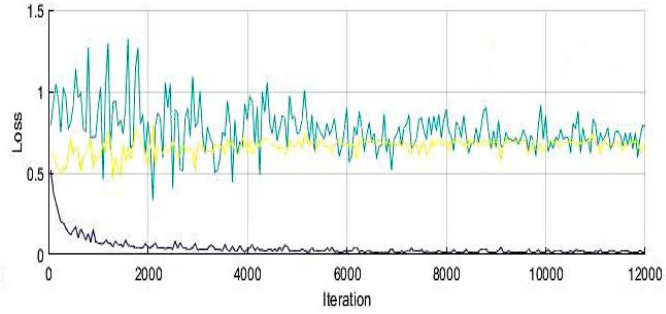


Figure 6. Pix2Pix GAN training plots. L1 loss (black), GAN loss (green), and discriminator loss (yellow).

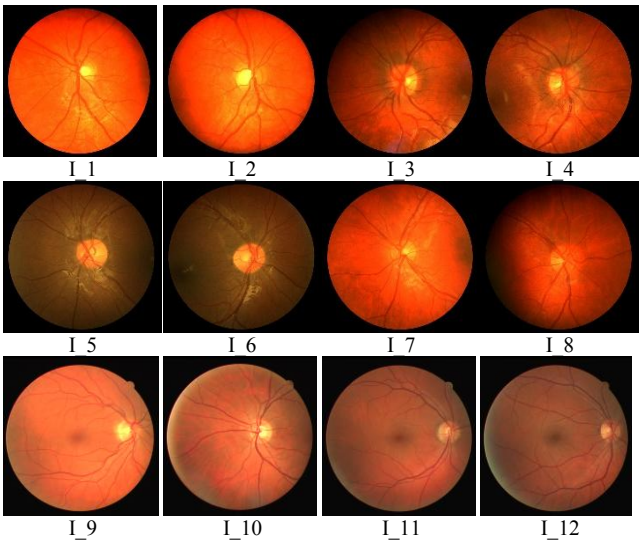


Figure 7. Testing images.

Furthermore, there can be observed circular artifacts at the edges of the field of view in the generated segmentations (Fig. 8). Removal of these artifacts required a postprocessing step in which we applied circular masks on the field of view in the images produced by the network.

Starting from the images in Fig. 7 we present the results containing the test image pairs (processed retinal images - references, ground truth – manually segmented) and the GAN segmented images (Fig. 9). We can see that the network performs very well the segmentation in case of large blood vessel.

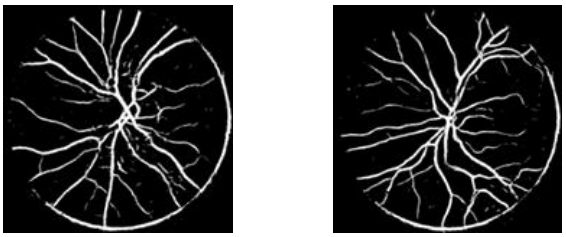
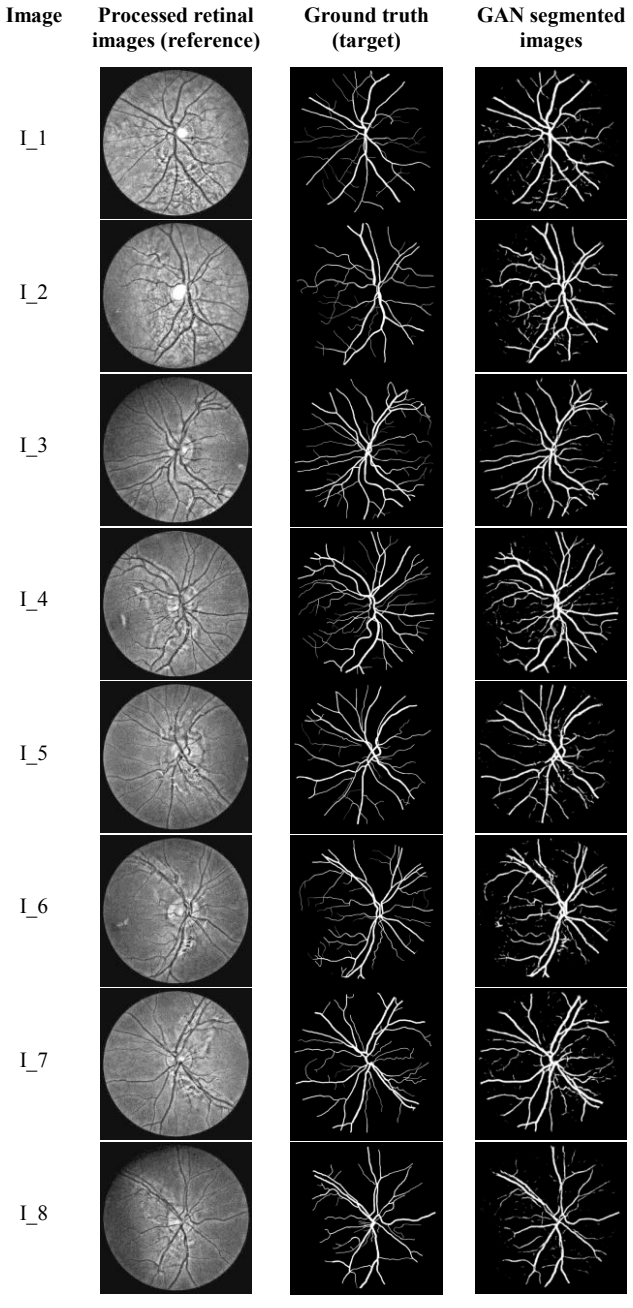


Figure 8. Raw segmentations generated by Pix2Pix GAN showing circular artifacts at the edge of the field of view.

Detection and segmentation of small vessels is less accurate, some of the background pixels being mistaken as vessel pixels (false positive) and vice versa (false negative). This can be due to high contrast and heterogeneity in the retinal image background.



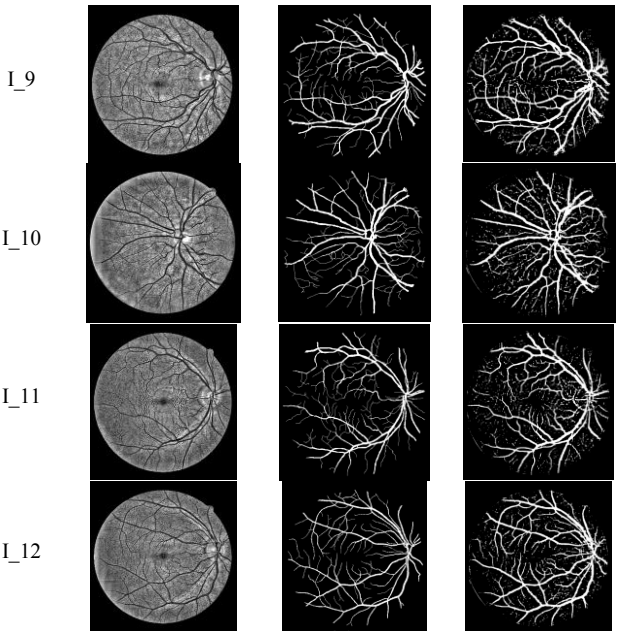


Figure 9. Segmentation results after testing Pix2Pix GAN.

Performance evaluation metrics are shown in Table I. Testing of Pix2Pix GAN gave a mean accuracy of 93.41%. The high precision and specificity values indicate that there are very few false positives, while the low sensitivity confirms that microvascularization is harder to detect. Similar results were reported in [4]. The accuracy was calculated at the pixel level classification.

TABLE I. PIX2PIX GAN PERFORMANCE EVALUATION METRICS (CHASE DATABASE)

Test image	Accuracy [%]	Precision [%]	Sensitivity [%]	Specificity [%]
I_1	92.71	98.96	90.24	97.97
I_2	92.98	99.17	90.42	98.40
I_3	94.19	97.67	93.46	95.63
I_4	92.79	97.97	90.96	96.35
I_5	93.66	98.22	92.08	96.74
I_6	93.03	98.12	91.10	96.69
I_7	93.04	98.64	90.73	97.56
I_8	94.85	97.93	94.40	95.80
Mean	**93.41**	**98.34**	**91.67**	**96.89**

In Table II we present the mean performances of Pix2Pix GAN for the three databases and the global performances (as the average of these – marked with gray). It can be observed that the performances are better in the case of CHASE database because the images have high resolution, uniformity, and are less affected by diseases.

Table III shows a comparison with U-Net and other works concerning the accuracy. It can be seen that the proposed method has good accuracy.

The first approach to optimize the network for the current application can be the discriminator parameters adjustment, so that it trains concurrently with the generator. The

architecture of the network can also be modified, as well as the λ scaling factor for L1 loss. Another improvement can be made in the preprocessing stage, like morphological operations, to reduce the false positive and false negative values. Training of Pix2Pix GAN can also be done using original color retinal images, without preprocessing.

TABLE II. MEAN PERFORMANCES EVALUATION TABLE STYLES

Mean performances	Accuracy [%]	Precision [%]	Sensitivity [%]	Specificity [%]
CHASE DB	93.41	98.34	91.67	96.89
DRIVE DB	92.14	94.98	83.42	96.06
STARE DB	91.52	93.25	81.78	94.21
GLOBAL	92.36	95.52	85.62	95.72

TABLE III. COMPARING ACCURACY WITH OTHER WORKS

Paper	U-Net	[3]	[4]	Our
Accuracy [%]/ Database	83/ DRIVE	90.7/ STARE	90.13/ DRIVE	92.36/ GLOBAL

V. CONCLUSION

In this paper we presented a method for automatic segmentation of blood vessels in retinal images using Pix2Pix GAN. The architectures and functioning principles of reference neural network models were described in order to understand how Pix2Pix GAN is implemented and trained. The network was neither specifically designed, nor optimized for retinal blood vessel segmentation. Nevertheless, it demonstrated promising results which make it suitable for further research and development in this field.

REFERENCES

[1] S. Gross, M. Klein, and D. Schneider, "Segmentation of blood vessel structures in retinal fundus images with Logarithmic Gabor filters," *Current Medical Imaging Reviews*, vol. 9, pp. 138-144, 2013.

[2] L. Ichim, D. Popescu, and S. Cirneanu, "Combining blood vessel segmentation and texture analysis to improve optic disc detection," in *5th IEEE International Conference on E-Health and Bioengineering* (*EHB*), Iasi, Romania, pp. 1-4, 2015.

[3] N. Hatami and M. H. Goldbaum, "Automatic identification of retinal arteries and veins in fundus images using local binary patterns," 2015, https://arxiv.org/ftp/arxiv/papers/1605/1605.00763.pdf.

[4] M. Savu, D. Popescu, and L. Ichim, "Blood vessel segmentation in eye fundus images," in *Proc. International Conference on Smart Systems and Technologies* (*SST*), Osijek, Croatia, pp. 245-251, 2017.

[5] L. Zhou, Q. Yu, X. Xu, Y. Gu, and J. Yang "Improving dense conditional random field for retinal vessel segmentation by discriminative feature learning and thin-vessel enhancement," *Comput. Methods Programs Biomed.*, vol. 148, pp. 13-25, September 2017.

[6] O. Ronneberger, P. Fischer, and T. Brox, "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), *Lecture Notes in Computer Science*, vol 9351, pp 234-241, November, 2015.

[7] H. Lamba, "Understanding semantic segmentation with UNET," 17 February 2019. [Online]. Available: https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47.

[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, and S. Ozair, "Generative adversarial nets," in *Proc. of the 27th International Conference on Neural Information Processing Systems*, 2014.

[9] J. Rocca, "Understanding generative adversarial networks (GANs)," 8 January 2019. [Online]. Available: https://towardsdatascience.com/understanding-generative-adversarial-networks-gans-cd6e4651a29.

[10] M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv:1411.1784, 2014.

[11] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, Hawaii, pp. 1125-1134, 2017.

[12] J. Son, S. J. Park, and K.-H. Jung, "Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks," *Journal of Digital Imaging*, vol. 32, pp. 499-512, 2019.

[13] C. Wu, Y. Zou, and Z. Yang, "U-GAN: generative adversarial networks with U-Net for retinal vessel segmentation," in *Proc. 14th International Conference on Computer Science & Education* (*ICCSE* 2019), Toronto, Canada, pp. 642-646, 2019.

[14] J. Brownlee, "A gentle introduction to Pix2Pix generative adversarial network," 29 July 2019. [Online]. Available: https://machinelearningmastery.com/a-gentle-introduction-to-pix2pix-generative-adversarial-network/.

[15] "Retinal image analysis," Kingston University London, [Online]. Available: https://blogs.kingston.ac.uk/retinal/chasedb1/.

[16] "DRIVE: Digital retinal images for vessel extraction," [Online]. Available: https://drive.grand-challenge.org/.

[17] "Structured analysis of the retina," [Online]. Available: https://cecas.clemson.edu/~ahoover/stare/.

[18] Y. Jiang, H. Zhang, N. Tan, and L. Chen, "Automatic retinal blood vessel segmentation based on fully convolutional neural networks," *Symmetry*, vol. 11, 1112, 2019.

[19] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," *Medical Image Analysis*, vol. 49, pp. 14-26, October 2018.

[20] M. A. Escobar, J. R. Guzmán Sepúlveda, J. R. Parra Michel, and R. Guzmán Cabrera, "A proposal to measure the similarity between retinal vessel segmentations images," *Nova Scientia*, vol. 11, no. 22, pp. 224-245, 2019.

[21] "Pix2Pix - Image to image translation using generative adversarial networks," [Online]. Available: https://github.com/matlab-deep-learning/pix2pix.