Artificial Intelligence Research and Development
A. Cortés et al. (Eds.)

doi:10.3233/FAIA220352

EDBNet: Efficient Dual-Decoder Boosted Network for Eye Retinal Exudates Segmentation

Mohammed Yousef Salem ALI [a,1], Mohamed ABDEL-NASSER [a,b], Aida VALLS [a], Marc BAGET [c] and Mohammed JABREEL [a]

[a] *ITAKA, Departament d'Enginyeria Informàtica i Matemàtiques Universitat Rovira i Virgili, Tarragona, Catalonia, Spain*
[b] *Department of Electrical Engineering, Aswan University, 81528 Aswan, Egypt*
[c] *IISPV, Hospital Universitari Sant Joan de Reus, Spain*

Abstract. Diabetic retinopathy (DR) is one of the most common causes of vision loss or blindness globally. Early detection of retinal eye lesions like hard exudates, soft exudates, microaneurysms, and hemorrhages is crucial to detect DR in a human eye. Therefore, accurate segmentation of lesions from eye fundus images is essential to develop efficient automated DR detection systems. This paper presented a novel hard and soft exudates lesions segmentation method called Efficient Dual-Decoder Boosted Network (EDBNet). EDBNet is composed of the following main components: 1) pre-trained ImageNet ResNet50 encoder with Atrous Spatial Pyramid Pooling (ASPP), 2) UNet decoder block with Gated Skip Connections mechanism to enhance capture more details of fundus images, 3) dual-decoder boosted to improve the performance segmentation of retinal lesion in the eye fundus images, and fusion outputs of the dual-decoder boosted to generate enhanced exudates segmentation. The effectiveness of the proposed framework is assessed on the IDRiD publicly dataset in terms of accuracy, Area Under Precision-Recall (AUPR), IOU, and Dice metrics. EDBNet obtains 99.8, 74.4, 78.0, and 87.6% of soft exudates, respectively. For hard exudates, EDBNet achieves 99.5, 85.3, 80.3, and 89.1%, respectively. The experimental results also demonstrate that EDBNet outperforms many state-of-the-art methods.

Keywords. Fundus Images, Exudates, Lesion Segmentation, Deep Learning, Diabetic Retinopathy

1. Introduction

Diabetic retinopathy (DR) is a disease that highly affects human eye vision. The early detection and treatment of DR are essential to prevent total vision loss [1]. DR produces different retinal lesions called microaneurysm (MA), haemorrhage (HE), hard exudate (EX), and soft exudate (SE). Ophthalmologists inspect eye fundus images to detect the signs of such lesions. However, the complex structure of lesions, various sizes, differences in brightness and the inter-class similarity with other fundus tissues add more dif-

[1] Corresponding Author: Mohammed Yousef Salem Ali. E-mail: horbio10@gmail.com

ficulties to the manual analysis of many fundus images. In addition, manual detection of the tiny lesions is extensive and consumes both time and effort.

Modern computer-aided diagnosis (CAD) systems can analyze medical images and provide a diagnosis as accurate as ophthalmologists with many years of experience [2]. Deep learning (DL) technologies have become the cornerstone of several modern CAD systems. Several DL-based automated systems have recently been proposed for segmenting eye retinal lesions. Most of them use Convolutional Neural Networks (CNNs) like UNet [3] to automatically learn representative and high-level features from the input fundus images to achieve accurate segmentation. For instance, The authors of [4] proposed CARNet for multi-lesion segmentation. CARNet feeds the whole image and patch image into ResNet50 and ResNet101 networks, respectively, using a single attention refinement decoder. They used IDRiD, E-ophtha and DDR datasets for evaluation. EAD-Net [5] presented a CNN-based system divided into an encoder module, dual attention module, and decoder module. They evaluated their work on two datasets: the E_ophtha_EX dataset for exudates and the IDRiD dataset for four kinds of lesions. In [6] authors developed a weakly-supervised framework for fundus lesion segmentation using grayscale and morphological features of lesions and a deep neural network with an attention mechanism and residual module. They evaluated their system by 1485 images extracted from the Messidor dataset and labelled by them. Paper [7] introduced scale-aware attention with different backbones to re-weight multi-scale features of decoders dynamically, and they evaluated it on IDRiD, E-ophtha and DDR datasets.

Although there are many advantages offered by deep learning techniques, especially those based on the UNet models, there is still a problem of dealing with tiny lesions like retinal exudates, so we need to develop a boosted mechanism for dealing with them. This paper proposes an accurate eye retinal exudates segmentation model called EDBNet, that takes as an input a fundus image and produces its corresponding lesion mask. The novelties of the proposed DL model are the following:

1. EDBNet uses dual decoders to boost the performance and produce an accurate mask of the input image. The main unit of each decoder is the Gated Skip Connection (GSC) network [8]. The reason behind using the GSC is the ability to focus on the most valuable feature from the decoder based on the features from the previous level.
2. The first decoder receives the skip connections from the encoder, where the second decoder takes the output of the first one as skip connections in a cascading manner. Hence, although we only use two decoders in this work, EDBNet can generalize to any level of cascaded boosted decoders. Such cascading and composing of multiple layers gives the network the ability to learn representations of data with multiple levels of abstraction [9].
3. The final output is obtained by fusing the output of the two decoders, which can be seen as a kind of online ensembling technique. We mean by online that the fusion operation is done in both the training and the inference phases.

We conducted extensive experiments on the well-known publicly available IDRiD dataset. We achieved competitive performance in one kind of lesion compared with the state-of-the-art systems in the IDRiD challenge and outperformed them in another.

The remainder of this paper is organized as follows. Section 2 explains the proposed Eye Retinal Exudates Segmentation system. Then, section 3 presents the experiments and results. Finally, the conclusion and future work are provided in section 4.

2. EDBNet

This section describes the proposed model, EDBNet. As shown in Figure 1, it is composed by the following parts: **Backbone** (colored in blue), also known as encoder layer, which aims to encode the input image and produce feature maps at multiple levels of scales. **Neck** (colored in purple), which is an Atrous Spatial Pyramid Pooling (ASPP) layer that helps to extract high-resolution features. **Head**, also known as decoder layer, a boosted dual decoder based on Gated-Skip connections network [8] followed by an output layer. We explain in detail each part in the following subsections.

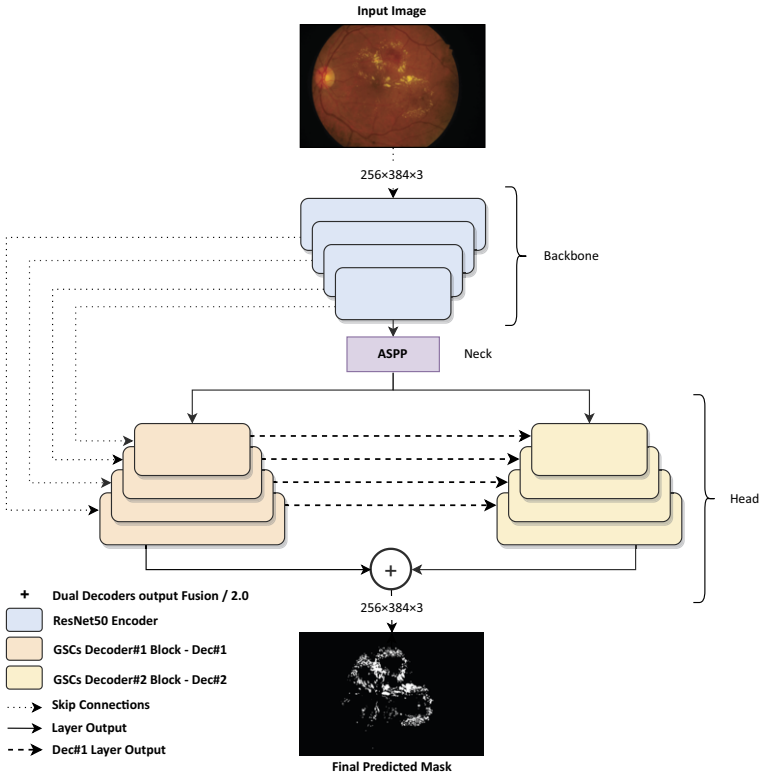

Figure 1. Structure of the EDBNet framework for Eye Retinal Exudates Segmentation.

2.1. Backbone: The Encoder Layer

We use a ResNet50 [10] encoder that was pre-trained on the ImageNet dataset as a backbone for our model. We selected this model because ResNet is the state-of-the-art backbone for many computer vision tasks [4,11].

The main goal of the backbone in our proposed model is to encode the input eye fundus image and extract abstracted and meaningful features at different levels of scales. The key advantage of ResNet50 is the residual connection which is used to add the output from an earlier layer to a later layer which helps to avoid the gradient vanishing problem.

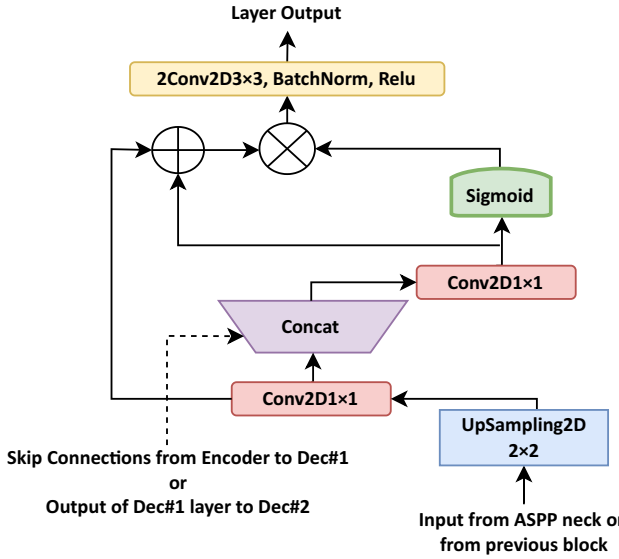

Figure 2. Gated Skip Connections network.

2.2. Neck: ASPP

To help extract high-resolution feature maps and enhance capturing the contextual information of the tiny lesions lost after multiple scales. We add the Atrous Spatial Pyramid Pooling (ASPP) [12] as a bridge between the ResNet encoder and the dual-decoder resolutions in the EDBNet framework.

2.3. Head: The Decoder Layer

The main new component of our methodology is the design of the head part. EDBNet employs a shared ResNet50 encoder to improve the performance without computation, and memory overload and dual-decoder boosted for lesions segmentation. The dual-decoder boosted has the same internal Gated Skip Connections (GSCs) architecture in each decoder block. Figure 2 presents the GSCs mechanism of each decoder block. GSCs modified the standard UNet decoder with a boosted feature maps production to enhance the discrimination between the lesion and background pixels for exudates segmentation.

EDBNet has four encoder blocks and four GSCs mechanism blocks for each decoder. In dec#1, the GSCs mechanism receives feature maps from the corresponding ResNet encoder layers. It concatenates them with the feature maps produced by the previous block (either the ASPP neck block or a previous dec#1 block). For dec#2, it receives feature maps from the corresponding dec#1 layers. Then concatenate them with the previous block's feature maps (either the ASPP neck block or a previous dec#2 block). We can express these feature maps as $S_1 \in \mathrm{R}^{h \times w \times f}$, and $S_2 \in \mathrm{R}^{h/2 \times w/2 \times 2f}$. Then, S_2 feeds to UpSampled2D transposed convolution layer with a kernel size of 2×2 to produce feature maps $\hat{S_2}$. $\hat{S_2}$ and S_1 should have the same width and height to perform the concatenation process as follows:

$$C = \varphi_{1\times1}([S_1||\hat{S_2}]) \tag{1}$$

In this expression, $\varphi_{1\times1}$ stands for the convolution operation with a kernel size of 1×1 and $||$ refers to the concatenation operation. C feature maps are fed to a *sigmoid* activation function to generate the weights ϑ, which helps to improve the discrimination between the lesion pixels and background pixels for EX or SE segmentation tasks. These weights are multiplied by the summation of S_1, $\hat{S_2}$ as follows:

$$D = \vartheta * (S_1 + \hat{S_2}) \tag{2}$$

After that, the improved feature maps of D are fed into two convolution layers, batch normalization and rectified linear unit activation function. The final output blocks of both dec#1 and dec#2 are followed by a sigmoid activation function and fed into a fusion process by the average weighted aggregation to take the benefits of multiple information sources and generate an optimal joint lesion segmentation [13]. In addition, the fusion aims to produce one final output with a fewer output channels as follows:

$$M = (M_1 + M_2)/2.0 \tag{3}$$

Where M stands for the final output of the proposed framework, whereas M_1 and M_2 indicate the output masks of dec#1 and dec#2, respectively. The final output mask is a binary image that includes the EX or SE lesions and has a size identical to the size of the input image size (384×256).

3. Experimental Results and Discussions

This section describes the conducted experiments to evaluate the effectiveness of the proposed model, including the description of the dataset, the experimental setup, the evaluation metrics and the analysis of the obtained results.

3.1. Dataset, pre-processing, and experimental setup

We used the popular Indian Diabetic Retinopathy Image Dataset (IDRiD) in our experiments [14]. It composed of 81 high-resolution retinal fundus images of 4288×2848. Each image contains at least one mask labelled as one out of four types of DR lesions EX, SE, MA, and HE. The dataset was split into 54 images as the training set and the rest of 27 as the testing set.

We used the following training pipeline (including some data augmentation techniques to enrich the data and improve the regularity of the model) to process the images in the training set. First, each image is divided into four non-overlapped sub-images, and the corresponding sub masks are constructed. We ignored the negative sub-images, i.e., the sub-images only with the background mask. Hence, each example in the training process is a sub-image with the size of 2144×1424 pixels with its corresponding sub mask. Next, we resized the sub-images and the sub masks to 384×256. The interpolation mechanism used is cubic for the images and the nearest neighbour for the masks.

After that, we applied horizontal flipping, rotation, Gaussian noise and grayscale augmentation techniques for 12 times. The total of training data calculated as: $((4 \times 54) - 20\%$ (for validation)$) \times 12$.

We trained each model for 50 epochs using Adam optimizer and a batch size of 4. The learning rate is set to 0.001. We sampled a subset (20%) from the training set and

used it as a validation set to save the best checkpoint of the trained models. We used the binary cross-entropy as a loss function to train the models.

During the inference phase, we only resize the input image to 768×512 and perform a full image segmentation process (i.e., no image splitting or image augmentation is used during the inference).

3.2. Evaluation Metrics

In this study, we used the most common evaluation metrics in the lesions segmentation task, specifically:

- Area Under Precision-Recall curve (AUPR): it is known to be a realistic measure for lesion segmentation performance like exudates [15].
- Pixel Accuracy (ACC): It can be defined as the percent of pixels in the image which were correctly classified. Formally it is defined as the following:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{4}$$

- Intersection-over-Union (IoU): also known as the Jaccard index, is basically a method to compute the percent overlap between the ground truth mask and the predicted mask.

$$IoU = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{5}$$

- Dice Coefficient: also referred to as F-score, it is the harmonic mean of precision and recall. It can be expressed as follows:

$$Dice = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{6}$$

- TPR/Sensitivity: denotes the proportion of real lesion pixels classified as lesion pixels. Formally it is defined as the following:

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

- FNR: denotes the model incorrectly predicts the negative class (background pixels). Formally it is defined as the following:

$$FNR = 1 - \mathrm{TPR} \tag{8}$$

In these expressions, TP refers to the true positive (the pixels were labelled as foreground, i.e., retinal lesion pixels, and correctly classified). The term FP means false positive (the pixels were labelled as background and misclassified as foreground). Also, TN is a true negative, which refers to the healthy pixels correctly classified by the network. whereas FN is a false negative representing the lesion pixels misclassified as healthy pixels. Finally, the TPR refers to the true positive rate of TP lesion pixels prediction, and the FNR indicates the false negative rate.

3.3. Different Decoder Architectures Evaluation of Retinal Exudates Segmentation

In this section, we evaluate the performance of the used decoder architectures on the test set images of the IDRiD dataset. Table 1 presents the performance of EX and SE retinal lesion segmentation models with the IDRiD dataset. We conducted three different experiments of decoder-side architecture for each EX and SE retinal lesions segmentation—Unet, UNet + GSCs, and EDBNet. We achieved the result of 74.4% of

Table 1. Performance comparison on IDRiD dataset. Values in bold highlight idicates to the highest case of accuracy

Methods	SE				EX			
	ACC	AUPR	IOU	Dice	ACC	AUPR	IOU	Dice
UNet	99.8±0.001	70.7±0.289	74.3±0.064	85.3±0.049	99.5±0.002	81.4±0.149	76.5±0.024	86.7±0.015
UNet + GSCs	99.8±0.001	73.4±0.271	76.5±0.071	86.7±0.053	99.5±0.002	82.0±0.149	78.1±0.023	87.7±0.014
EDBNet	**99.8**±0.001	**74.4**±0.273	**78.0**±0.072	**87.6**±0.053	**99.5**±0.002	**85.3**±0.157	**80.3**±0.009	**89.1**±0.006

SE lesion segmentation and 85.3% of EX lesion segmentation with the AUPR metric that common uses of the IDRiD dataset challenge. For the Accuracy metric of SE and EX, we reached 99.8 and 99.5%, respectively. Regarding the Dice metric, we obtained 87.6% of SE and 89.1% of EX. We noted our GSCs mechanism enhanced the performance of SE and EX lesions segmentation when it has added to the UNet decoder, as shown in the second row of Table 1. At the same time, when we used the EDBNet, the performance of SE and EX lesions segmentation improved compared to the method that just used GSCs in one decoder, as shown in the third row. As a result, the SE lesion segmentation performance has improved by 1%, 1.5%, and 0.9% of AUPR, IOU, and Dice metrics, respectively.

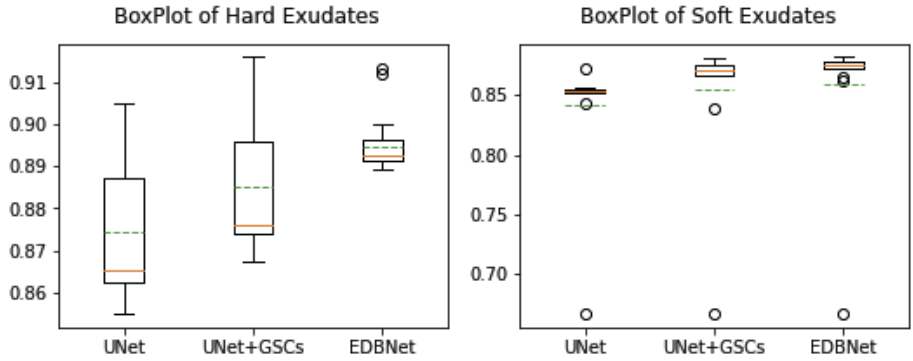

Figure 3. BoxPlot of Dice for Hard and Soft Exudates Segmentation results (green dashed lines indicate the mean, and the oranges indicate the median). All values outside the whiskers are considered as outliers, which are marked with the (○) symbol.

On the other hand, the EX lesion segmentation performance has significantly enhanced the AUPR, IOU, and Dice metrics of 3.3%, 1.8%, and 1.4%, respectively.

The EDBNet reduced the standard deviation from ±0.015 to ±0.006 with Dice and ±0.024 to ±0.009 with IOU metrics of EX segmentation. These effects reveal that EDBNet can present more precise and robust segmentation than the UNet and UNet + GSCs models.

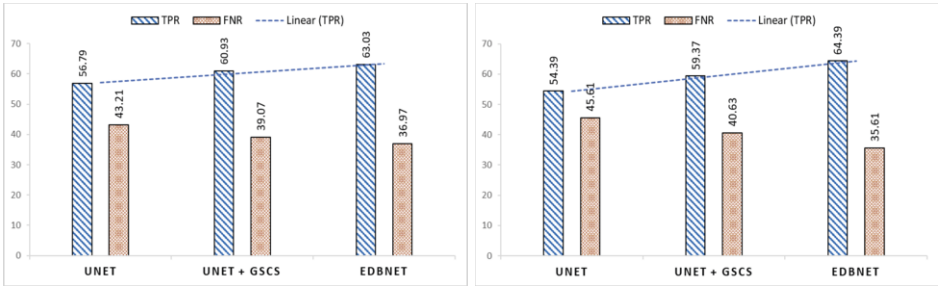

Figure 4. Average of TPR and FNR with different decoder architectures (EX left, SE right).

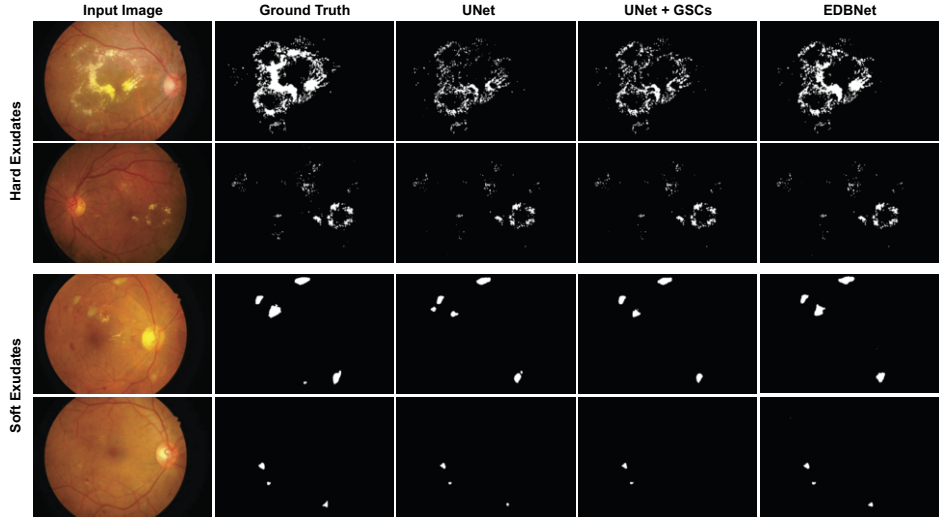

Figure 5. Hard and Soft Exudates Segmentation results.

Figure 4 shows that the EDBNet model of the dual-decoder yields the highest True Positive Rate (TPR/Sensitivity) and lowest True Negative Rate (TNR) for all testing set images of the IDRiD dataset. The dashed line indicates the performance enhancement when we used the GSCs decoder and the EDBNet. In addition, we present some sample masks obtained for the three SE and EX lesion segmentation models to demonstrate the efficacy as shown in Figure 5. Figure 5 shows that the final output masks of EDBNet are better segmented than those final output masks of UNet+GSCs and UNet baseline models.

Finally, we show the boxplots of the Dice of the proposed model, UNet, and UNet+GSCs. As shown in Figure 3 among the tested models, the proposed model has the highest mean, median, and smallest standard deviation of the Dice for the EX and has the highest mean and median for the SE.

It is engaging to determine the statistical significance of the differences in performance between the proposed EDBNet and UNet + GSCs (the second best model) in terms of Dice. To do so, we used Student's t-test (significance level < 0.05) to specify

the difference between Dice values. The p-values obtained are less than 0.05 for EX and higher than 0.05 for SE, indicating a statistical significance for Ex but not for SE.

3.4. Comparison with existing methods

To confirm the effectiveness of the proposed method, we conduct a comparison between EDBNet and state-of-the-art on the AUPR metric (same as the one used in the IDRiD competition) as shown in Table 2. The comparison includes the top 5 teams on the IDRiD competition [16] (the first five rows in the table), as well as CARNet [4], EAD-Net [5], L-Seg [17], and SAA [7]. EDBNet surpasses all the state-of-the-art results of segmenting the SE retinal lesion segmentation by 1.55% of the AUPR metric. Regarding the EX, EDBNet achieves acceptable performance. The results of EDBNet

Table 2. Comparison with state-of-the-art results on IDRiD

Method	AUPR on SE	AUPR on EX
VRT (1st) [16]	0.6995	0.7127
PATech (2nd) [16]	-	**0.8850**
IFLYTEK-MIG (3rd) [16]	0.6588	0.8741
SOONER (4th) [16]	0.5395	0.7390
SAIHST (5th) [16]	-	0.8582
CARNet [4]	0.7125	0.8675
EAD-Net [5]	0.6083	0.7818
L-Seg [17]	0.7113	0.7945
SAA [7]	0.7281	0.8792
Proposed	**0.7436**	0.8534

are comparable with IFLYTEK-MIG, CARNet, and SAA, while they had good results of EX lesion segmentation but not with SE. On the other hand, PATech and SAIHST have the best results for EX retinal lesion segmentation but did not introduce SE retinal lesion segmentation results to compare. Therefore, as shown in Table 2 there is no method with the best results for the two retinal lesions segmentation at the same time.

4. Conclusions and future work

This paper presented a new deep learning architecture called EDBNet for image segmentation problems. It has been used for segmenting retinal exudates in fundus images of the human eye. EDBNet is composed of three main elements: ResNet50 backbone, ASPP neck, and dual-decoder using several GSCs in cascade. EDBNet framework led to high segmentation performance of both SE and EX retinal lesions for the IDRiD dataset, obtaining an accuracy, AUPR, IOU, and Dice metrics of 99.8, 74.4, 78.0, and 87.6% of soft exudates, respectively; while for hard exudates, we achieve 99.5, 85.3, 80.3, and 89.1%, respectively. EDBNet showed superiority with respect to other state-of-the-art models, with performance higher than 1.55% of the AUPR metric for the SE retinal lesion segmentation with the IDRiD dataset.

The EDBNet framework is not designed exclusively for retinal lesion segmentation on fundus images; it may also work well in cases of medical images with tiny objects.

Future work will include using the proposed exudates segmentation model to develop segmentation methods for four kinds of eye lesions: microaneurysm (MA), haemorrhage (HE), hard exudate (EX), and soft exudate (SE) to diagnose diabetic retinopathy.

Acknowledgements

This work has been funded by the research projects PI21/00064 and PI18/00169 from Instituto de Salud Carlos III & FEDER funds. The University Rovira i Virgili also supports this work with projects 2020PFR-B2-61 and 2019PFR-B2-61.

References

[1] Mary VS, Rajsingh EB, Naik GR. Retinal fundus image analysis for diagnosis of glaucoma: a comprehensive survey. IEEE Access. 2016.

[2] Jani K, Srivastava R, Srivastava S, Anand A. Computer aided medical image analysis for capsule endoscopy using conventional machine learning and deep learning. In: 2019 7th International Conference on Smart Computing & Communications (ICSCC). IEEE; 2019. p. 1-5.

[3] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer; 2015. p. 234-41.

[4] Guo Y, Peng Y. CARNet: Cascade attentive RefineNet for multi-lesion segmentation of diabetic retinopathy images. Complex & Intelligent Systems. 2022:1-21.

[5] Wan C, Chen Y, Li H, Zheng B, Chen N, Yang W, et al. EAD-net: a novel lesion segmentation method in diabetic retinopathy using neural networks. Disease Markers. 2021;2021.

[6] Li Y, Zhu M, Sun G, Chen J, Zhu X, Yang J. Weakly supervised training for eye fundus lesion segmentation in patients with diabetic retinopathy. Mathematical Biosciences and Engineering. 2022;19(5):5293-311.

[7] Bo W, Li T, Liu X, Wang K. SAA: Scale-Aware Attention Block For Multi-Lesion Segmentation Of Fundus Images. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). IEEE; 2022. p. 1-5.

[8] Jabreel M, Abdel-Nasser M. Promising crack segmentation method based on gated skip connection. Electronics Letters. 2020;56(10):493-5.

[9] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.

[10] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770-8.

[11] Pan X, Jin K, Cao J, Liu Z, Wu J, You K, et al. Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus fluorescein angiography based on deep learning. Graefe's Archive for Clinical and Experimental Ophthalmology. 2020;258(4):779-85.

[12] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence. 2017;40(4):834-48.

[13] Zhang J. Multi-source remote sensing data fusion: status and trends. International Journal of Image and Data Fusion. 2010;1(1):5-24.

[14] Porwal P, Pachade S, Kamble R, Kokare M, Deshmukh G, Sahasrabuddhe V, et al. Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. Data. 2018;3(3):25.

[15] Boyd K, Eng KH, Page CD. Area under the precision-recall curve: point estimates and confidence intervals. In: Joint European conference on machine learning and knowledge discovery in databases. Springer; 2013. p. 451-66.

[16] Porwal P, Pachade S, Kokare M, Deshmukh G, Son J, Bae W, et al. Idrid: Diabetic retinopathy–segmentation and grading challenge. Medical image analysis. 2020;59:101561.

[17] Guo S, Li T, Kang H, Li N, Zhang Y, Wang K. L-Seg: An end-to-end unified framework for multi-lesion segmentation of fundus images. Neurocomputing. 2019;349:52-63.