2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Detection of Fundus Lesions through a Convolutional Neural Network in Patients with Diabetic Retinopathy

Carlos Santos[1,2], *Member, IEEE,* Marilton Sanchotene de Aguiar[2], Daniel Welfer[3], and Bruno Monteiro Belloni[4]

***Abstract*— Diabetic Retinopathy is a major cause of vision loss caused by retina lesions, including hard and soft exudates, microaneurysms, and hemorrhages. The development of a computational tool capable of detecting these lesions can assist in the early diagnosis of the most severe forms of the lesions and assist in the screening process and definition of the best treatment form. This paper proposes a computational model based on pre-trained convolutional neural networks capable of detecting fundus lesions to promote medical diagnosis support. The model was trained, adjusted, and evaluated using the DDR Diabetic Retinopathy dataset and implemented based on a YOLOv4 architecture and Darknet framework, reaching an mAP of 11.13% and a mIoU of 13.98%. The experimental results show that the proposed model presented results superior to those obtained in related works found in the literature.**

I. Introduction

Diabetes causes a disease that affects the eyes named Diabetic Retinopathy (DR), a significant cause of vision loss in working-age adults. Oftamologists identify DR by eye exams that aim to identify lesions of the retina, including hard exudates (EX), soft exudates (SE), microaneurysms (MA), and hemorrhages (HE). Solutions presented in the literature assist in diagnosing DR through deep neural networks, such as [1]–[5]. Although researchers use deep neural networks to detect lesions in the fundus images, they still have limitations in the results obtained, mainly due to the low representativeness of the attributes extracted from the images used for training the models. In this context, this work aims to present a computational model based on pre-trained convolutional neural networks capable of detecting lesions of the fundus of the eye to promote more efficient and more accurate medical diagnosis support than analogous works found in the literature. The main contribution of this work is to present a model of convolutional neural network based on a One-Shot detector, and applying the concept of transfer learning for resource extraction in the upper layers of the network, and obtaining characteristics of the fundus lesions in the posterior layers through training in the public Dataset for Diabetic Retinopathy (DDR) [5].

*This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

[1]Federal Institute of Education, Science and Technology Farroupilha, Alegrete, Brazil. (e-mail: carlos.santos@iffarroupilha.edu.br).

[2]Postgraduate Program in Computing (PPGC), Federal University of Pelotas, Pelotas, Brazil. (e-mail: marilton@inf.ufpel.edu.br).

[3]Departament of Applied Computing, Federal University of Santa Maria, Santa Maria, Brazil. (e-mail: welfer@gmail.com).

[4]Federal Institute of Education, Science and Technology Sul-Rio-Grandense, Passo Fundo, Brazil. (e-mail: bruno.belloni11@gmail.com).

II. Related Work

Benzamin et al. [1] presents a deep learning model for identifying hard exudates present in fundus images affected by DR. The authors developed an 8-layer convolutional neural network. The main limitation of the work was not to detect soft exudates, hemorrhages, and microaneurysms. Porwal et al. [2] used deep learning models for segmentation, classification, and detection of fundus lesions during the 'IDRiD: Diabetic Retinopathy - Segmentation and Grading Challenge' is presented. The detection challenge aimed to obtain the location of the optical disc (OD) and the fovea. The winning team presented a method based on ResNet-50 and VGG architecture. The work presented by Porwal et al. [2] was limited to presenting only the detection fovea and OD. Mateen et al. [3] proposed a pre-trained framework based on a convolutional neural network to detect exudates in fundus images through learning transfer. Inception-v3, ResNet-50 and VGG-19 architectures were used. The main limitation of the work was not to detect hemorrhages and microaneurysms. Li et al. [5] presented a new Diabetic Retinopathy dataset called DDR and evaluated state-of-the-art deep learning models for classification, segmentation, and DR lesions detection. The authors tested state-of-the-art object detection models to assess performance on the DDR dataset, including Faster R-CNN, SSD, and YOLO. They presented results with the mean Average Precision (mAP) and mean Intersection over Union (mIoU) metrics obtained in the validation and test sets. Researchers applied deep neural networks to identify DR, but the deep learning models used presented limitations in the results presented. Although deep learning can analyze medical images, it still has limitations, with a gap between research and clinical application. This problem is associated with how deep learning automatically extracts the most discriminating resources from the training examples. This work intends to present a model capable of identifying the fundus lesions employing digital image processing techniques and convolutional neural networks that detect the lesions with greater precision than the works found in the literature.

III. Materials and Methods

The proposed model was developed based on the YOLOv4 architecture, according to the block diagram of the proposed model illustrated in Fig. 1, and the Darknet [6] framework. We trained the model for 8,000 epochs and 64 lots per epoch, with a learning rate of 0.0001 and a momentum rate of 0.949.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

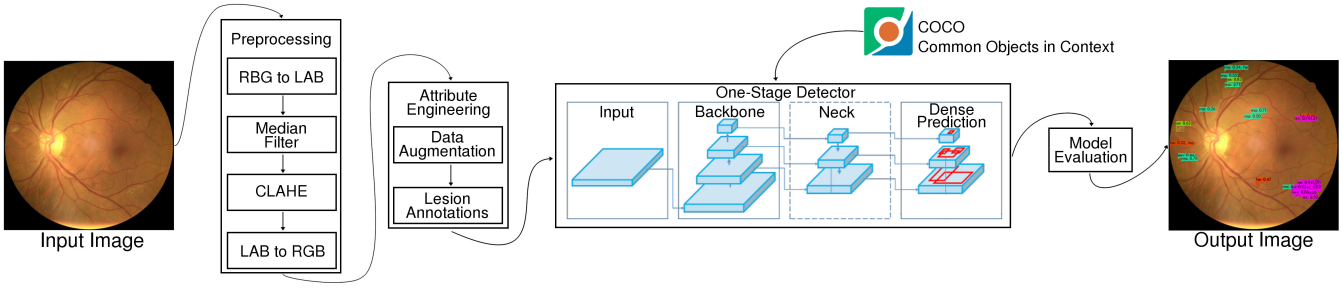

Fig. 1. Block diagram of the proposed model.

The size of the anchors of the bounding boxes varies between 1 and 200. After detection, we reached a confidence percentage for each identified lesion. We use a core i7 microcomputer with 16GB of RAM and an NVIDIA GTX 1080 TI GPU. The YOLOv4 architecture is a One-Shot detection model capable of detecting objects without a preliminary step instead of a Two-Stage detector that uses a preliminary stage where regions of importance are detected and then classified to verify whether we detect an object in these areas. The advantage of a One-Shot detector is the speed with which it can make inferences in real-time. Besides, another feature of the model is the possibility to work on edge devices and with low-cost hardware whose training with just one GPU [6].

A. Dataset and Image Preparation

We obtained the fundus lesions images used in this work from the DDR fundus image dataset. This dataset has 757 images with annotations of the fundus in JPEG format with variable sizes. The training of the proposed model has challenges, such as the small number of examples of lesions, and the fact that they do not have a well-defined format, varying according to the stage of the disease. As a result, we augmented the dataset from the derivation of the original images and the annotation of DR lesions applying translational and rotation transformations to the original images, resulting in eight new images for each annotated image in the DDR dataset. After, it was necessary to balance the samples of lesions since there was a significantly higher EX number than SE. Finally, we trained our model to detect lesions in the fundus of the eye using bounding boxes drawn around the region of interest of the objects.

B. Pre-Processing

As illustrated by Fig. 1, we used the median filter with a 3×3 kernel to smooth the image. Also, we applied contrast limited adaptive histogram equalization (CLAHE) to the L component of the LAB color space of DDR images to enhance the contrast. Besides, to measure quantitatively the best contrast obtained in the images after the enhancement, we used the metric Measure of Enhacement [7]. Furthermore, finally, we converted the images of the fundus to the RGB color space.

C. The Proposed Approach

After the pre-processing and attribute engineering step previously described, we obtain the input images in the architecture. Then, we rescale the images to the size of 608×608 pixels at the network entrance to reduce dimensionality and the computational cost during training; however, without reducing the accuracy of the model, avoiding high rates of error of the classifier. Besides, we partitioned the dataset in 50:20:30 proportion for training, testing, and validation, respectively. Finally, we use the network Backbone as an extractor of pre-trained resources in a set of image classification data, useful for detecting objects in the last layers of the network. We obtained the set of weights used for pre-training from Common Objects in Context (COCO) [8], and the Backbone used in the experiments was a CSPDarknet53 [9] with activation function Mish [10].

We use the Neck to extract different resource maps from different backbone stages, constituting extra layers between the Backbone and the Head. The Neck used in the experiments was the Path Aggregation Network (PANet) [11], which allows a better propagation of information from the bottom to top or top to bottom. Head is the network in charge of making a dense prediction (final prediction) and is composed of a vector containing the predicted bounding box (center coordinates, height, width), the forecast confidence score, and the label. The Head used in the experiments was the YOLOv3 based on anchors [12].

Besides, we used Cross-Iteration Batch Normalization [13], and regularization employing the Drop Block technique, in which we hide sections of the image from the first layer, i.e., we discard resources in a contiguous region of a resource map [14]. Finally, to remove the bounding boxes representing the same object, keeping the most accurate one, we used the Non-Maximum Suppression [15] technique. The loss function adopted for the bounding boxes was the Complete Intersection Over Union Loss [16], which aims to provide faster convergence and superior performance in detecting lesions.

D. Pre-training

We performed transfer learning to train the proposed model initializing the weights of the architecture with weights from the COCO challenge dataset. COCO provides a large set of annotated image data for object detection tasks. We modified the output of the proposed model to suit the detection task and preserved the knowledge of the upper layers.

IV. Results and Discussions

We ran the experiments using the DDR public dataset. First, the proposed model was implemented using the Darknet framework and trained with a deep neural network architecture based on the YOLOv4 model. After that, we performed transfer learning based on the pre-trained weights in the COCO dataset. We evaluated our model with the AP (Average Precision), mAP, and mIoU metrics to compare the results.

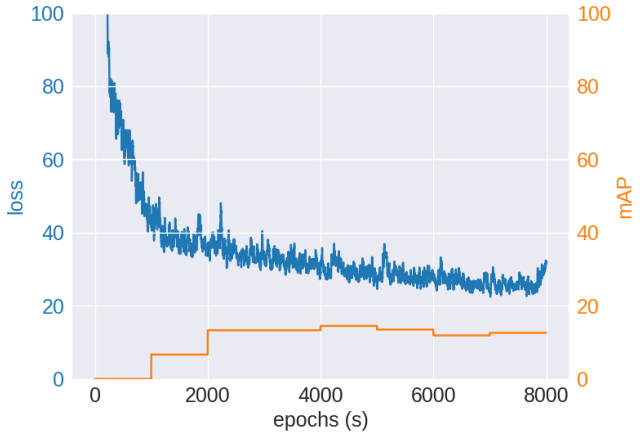

Fig. 2. Loss function×epochs×mAP in the test step.

Fig. 2 shows the graph of the loss function (blue line), as a function of 8,000 epochs performed in the training stage, and values of the mAP (orange line) for detecting lesions in the fundus of the eye. Also, while it occurs the interactions of architecture training, the loss value decreases. In general, the mAP has a more significant increase at the beginning of the training. It then tends to reduce until there is stabilization due to the adjustment of the model during training. It is possible to verify that the highest value obtained for mAP was approximately 11%, while the lowest value obtained for the loss function was 31.26%. After the end of the training, it was possible to get the following AP values for each lesion: 5.72% (EX), 21.62% (SE), 1.52% (MA), and 15.66% (HE). We obtained 11.13% of the mAP with a limit value of IoU@0.5, and 13.98% of the mIoU, as shown in Table I. We used a test dataset to evaluate the performance of the proposed model. We adopted the threshold value IoU equal to 0.5 (IoU@0.5) and mIoU to assess the detection quality. Besides, we used the mAP metric, often used to measure the accuracy of object detectors [17], and [12], which has the purpose of calculating the average precision obtained in all evaluated classes.

We used the Intersection over Union (IoU) metric to measure the accuracy of an object detector in a specific dataset. We also measure the proposed model with various metrics to evaluate its performance in detecting fundus lesions. Table I presents the results obtained by the proposed model with the metrics AP, mAP, and mIoU, as well as the results of the approaches found in the literature that used One-Shot detection architecture. As noted, the values obtained by the proposed model were higher than those obtained by related works. For example, in [5] the authors report an mAP of 0.35% for the model trained in the DDR dataset. On the other hand, we carried out our experiments on the same dataset, and we obtained an mAP of 11.13%.

TABLE I

AP, mAP, and mIoU metrics obtained in the detection of lesions in the fundus of the eye in the validation set.

Models	**AP**					**mIoU**
	EX	**SE**	**MA**	**HE**	**mAP**	
YOLO [5]	0.39	0	0	1.01	0.35	0.05
SSD [5]	0	2.27	0	0.07	0.59	0.15
Proposed Approach	5.72	21.62	1.52	15.66	**11.13**	**13.98**

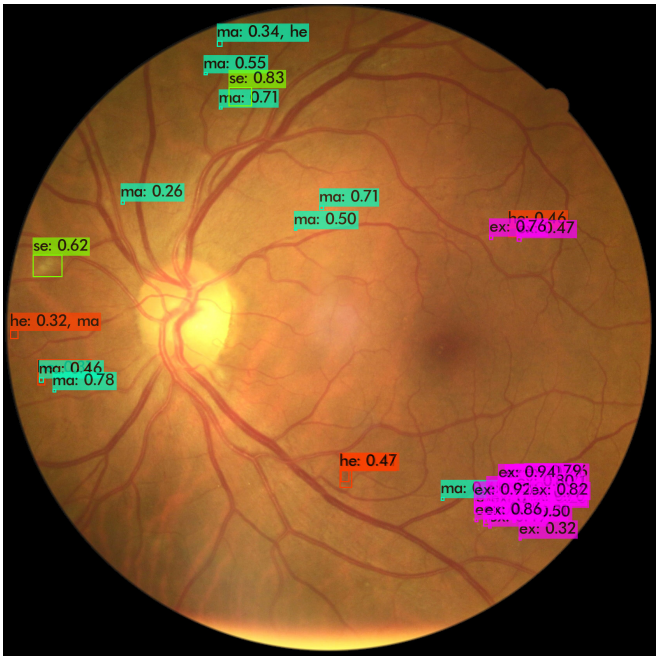

Fig. 3. Detected lesions and the degree of confidence added in each object located in the fundus image.

It is possible to verify that the proposed model also obtained better results in the metrics of AP and mIoU. After filtering (by IoU and confidence limits) and obtaining the total number of bounding boxes in the test step, we consider the values of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) to quantitatively evaluate the results. Then, we calculate the Accuracy, which considers among all the Positive class classifications that the model made, how many are correct; the Recall, which assumes among all situations of class Positive as expected value, how many are correct; and the F1-Score, which calculates the harmonic average between Precision and Recall. Table II shows the results obtained with the metrics Precision, Recall, and F1-Score during the experiments carried out with the model proposed in the validation and test sets.

Note that in the test set, the value of the F1-Score measure followed the average value of Precision and Revocation, so the accuracy obtained by the model was reliable. Fig. 3

TABLE II

THE PROPOSED MODEL RESULTS IN THE VALIDATION AND TEST SETS, ACCORDING TO PRECISION, RECALL, AND F1-SCORE METRICS.

Set	Precision	Recall	F1-Score
Validation	0.23	0.11	0.15
Test	0.20	0.11	0.15

shows lesions detection examples and the degree of confidence obtained in each object located in an image of the fundus. Confidence levels ranged from 25% to 94%. Fig. 3 shows the results of the detections made by the proposed model, where it is possible to observe the identification of several lesions. The model took 263.55 milliseconds to perform the inference. With the minimum confidence limit set to 25%, we identified the following lesions with their respective confidence percentages. The experimental results showed that the proposed model has greater precision in detecting SEs, HEs, and EXs lesions but a relatively low precision in detecting MAs, suggesting that the model is inefficient in learning MAs. Microaneurysms in mild DR are challenging to detect as they are tiny objects. Besides, there are lesions of the same type with very different shapes and sizes, which generates a high number of FP and FN and decreases the model's ability to predict. Even with data augmentation, which improves deep neural network architecture to extract attributes and recognize patterns, there is still the problem of not having a significant number of different examples for each type of lesion. This issue makes it difficult to train the neural network and directly impacts the performance of the model in making new inferences and adequately identifying the different types of fundus lesions. All these characteristics added to the great variety of size, intensity, shape, and contrast of the lesions justify the low results obtained during the experiments.

V. CONCLUSIONS

This paper presented a convolutional neural network model based on a One-Shot detector to detect the fundus lesions caused by Diabetic Retinopathy. The training and validation process of the model used the DDR public dataset, in which we partitioned it into a 50:20:30 ratio, and the identification of the lesions reached an mAP of 11.13%. A YOLOv4 deep neural network architecture and the Darknet framework implemented the proposed model, achieving an Average Precision of 5.72% for Hard Exudates, 21.62% for Soft Exudates, 1.52% for Microaneurysms, and 15.66% for Hemorrhages. The Intersection over Union average was 13.98%. The experiments carried out achieved promising results, surpassing the related works found in the literature and demonstrating that detecting DR lesions in the fundus of the eye can be performed with good precision using deep neural networks that detect objects in One-Shot. As future work, we intend to use new structures for the Backbone, Neck, and Head of the architecture used in the proposed model. We also intend to experiment with other public datasets and models that perform object detection.

REFERENCES

[1] A. Benzamin and C. Chakraborty, "Detection of Hard Exudates in Retinal Fundus Images Using Deep Learning," *2018 IEEE International Conference on System, Computation, Automation and Networking, ICSCA 2018*, 2018.

[2] P. Porwal, S. Pachade, M. Kokare, G. Deshmukh, J. Son, W. Bae, L. Liu, J. Wang, X. Liu, L. Gao, T. B. Wu, J. Xiao, F. Wang, B. Yin, Y. Wang, G. Danala, L. He, Y. H. Choi, Y. C. Lee, S. H. Jung, Z. Li, X. Sui, J. Wu, X. Li, T. Zhou, J. Toth, A. Baran, A. Kori, S. S. Chennamsetty, M. Safwan, V. Alex, X. Lyu, L. Cheng, Q. Chu, P. Li, X. Ji, S. Zhang, Y. Shen, L. Dai, O. Saha, R. Sathish, T. Melo, T. Araújo, B. Harangi, B. Sheng, R. Fang, D. Sheet, A. Hajdu, Y. Zheng, A. M. Mendonça, S. Zhang, A. Campilho, B. Zheng, D. Shen, L. Giancardo, G. Quellec, and F. Mériaudeau, "IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge," *Medical Image Analysis*, vol. 59, 2020.

[3] M. Mateen, J. Wen, N. Nasrullah, S. Sun, and S. Hayat, "Exudate Detection for Diabetic Retinopathy Using Pretrained Convolutional Neural Networks," *Complexity*, vol. 2020, 2020.

[4] H. Wang, G. Yuan, X. Zhao, L. Peng, Z. Wang, Y. He, C. Qu, and Z. Peng, "Hard exudate detection based on deep model learned information and multi-feature joint representation for diabetic retinopathy screening," *Computer Methods and Programs in Biomedicine*, vol. 191, p. 105398, 2020. [Online]. Available: https://doi.org/10.1016/j.cmpb.2020.105398

[5] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511–522, 2019. [Online]. Available: https://doi.org/10.1016/j.ins.2019.06.011

[6] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *ArXiv e-prints*, 2020, arXiv:2004.10934. [Online]. Available: http://arxiv.org/abs/2004.10934

[7] K. Lentz and A. Grigoryan, "A new measure of image enhancement," *IASTED International Conference on Signal Processing & Communication*, pp. 19–22, 01 2000. [Online]. Available: https://bit.ly/3tlcgNn

[8] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8693 LNCS, no. PART 5, pp. 740–755, 2014.

[9] C. Y. Wang, H. Y. Mark Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2020-June, pp. 1571–1580, 2020.

[10] D. Misra, "Mish: A Self Regularized Non-Monotonic Activation Function," *ArXiv e-prints*, 2019, arXiv:1908.08681. [Online]. Available: http://arxiv.org/abs/1908.08681

[11] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 8759–8768, 2018.

[12] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *ArXiv e-prints*, 2018, arXiv:1804.02767. [Online]. Available: http://arxiv.org/abs/1804.02767

[13] Z. Yao, Y. Cao, S. Zheng, G. Huang, and S. Lin, "Cross-Iteration Batch Normalization," *ArXiv e-prints*, 2020, arXiv:2002.05712. [Online]. Available: http://arxiv.org/abs/2002.05712

[14] G. Ghiasi, T. Y. Lin, and Q. V. Le, "Dropblock: A regularization method for convolutional networks," *Advances in Neural Information Processing Systems*, vol. 2018-December, pp. 10 727–10 737, 2018.

[15] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS - Improving Object Detection with One Line of Code," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-October, pp. 5562–5570, 2017.

[16] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," *In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

[17] Y. Konishi, Y. Hanzawa, M. Kawade, and M. Hashimoto, "SSD: Single Shot MultiBox Detector," *Eccv*, vol. 1, pp. 398–413, 2016.