An Improved Approach for Semantic Segmentation of Fundus Lesions using R2U-Net

Alejandro Pereira[1], Carlos Santos[2], Marilton Aguiar[1], Daniel Welfer[3], Marcelo Dias[1], Rafaela de Menezes[2], Paulo Roberto Ferreira[1], Fábio Rossi[2], Marcos d'Ornellas[3], Carlos Jesus Haygert[4], Juliano Kazienko[5], Fábio Comim[6], Aurélio Hoppe[7], Juliano Weber[8], Leandro Tavares[1], Leonardo Guths[9], and Rodrigo Guerra[10]

***Abstract*—Diabetic Retinopathy (DR) is a microvascular complication related to diabetes that affects approximately 33% of individuals with this condition and, if not detected and treated early, can lead to irreversible vision loss. Fundus lesions such as Hard and Soft Exudates, Hemorrhages, and Microaneurysms typically identify DR. The development of computational methods to segment these lesions plays a fundamental role in the early diagnosis of the disease. This paper proposes a new approach that uses an R2U-Net combined with data augmentation techniques for segmenting fundus lesions. We trained, adjusted, and evaluated the proposed work in the DDR dataset, achieving an accuracy of 99.87% and a mean Intersection over Union ($mIoU$) equal 59.69%. Furthermore, we assessed it in the IDRiD dataset, achieving an $mIoU$ of 49.92%. The results obtained in the experiments highlight the potential contribution of the model in the lesion annotations for creating new DR datasets, which is essential given the scarcity of annotations in publicly available datasets.**

***Index Terms*—diabetes mellitus, diabetic retinopathy, fundus image, lesions segmentation, deep learning**

I. Introduction

Diabetic Retinopathy (DR) represents the main ocular complication associated with Diabetes Mellitus, occurring in approximately 33% of diabetic patients. On a global scale, DR affects more than 100 million people, making it one of the leading causes of blindness and visual impairment if healthcare professionals do not identify and treat it early [1]. The severity of DR manifests through the presence of various types of retinal lesions, including Hard Exudates (EX), Hemorrhages (HE), Soft Exudates (SE), and Microaneurysms (MA) in the fundus of the eye [2]. Recently, researchers have emphasized the remarkable feature learning capability of deep learning in image segmentation. This way, encoder-decoder architectures based on U-Net can efficiently segment the previously mentioned lesions, demonstrating promising results in medical image segmentation applications [3]. However, obtaining DR data for model training is challenging since the different public datasets available in the literature present a small number of labeled lesions and problems associated with the quality of the images. These problems often negatively affect the results obtained by predictors based on deep neural networks.

In this context, the paper's main contribution is a new method for segmentating fundus lesions with greater precision using an R2U-Net [4] and data augmentation techniques. Therefore, the proposed work aims to assist in diagnosing DR and optimize the lesion annotation process to enable the creation of new DR datasets.

II. Related Work

Given the challenges presented in the previous section, the following will review works available in the literature whose purpose is the segmentation of fundus lesions. Among the pieces found is the one proposed by Porwal et al. [6], which presents results of deep learning models used for classification, detection, and segmentation of objects in fundus images, wherein the segmentation challenges the U-Net architecture [7] has been explored.

The study by Li et al. [8] reported the introduction of a new dataset for Diabetic Retinopathy named DDR (Dataset for Diabetic Retinopathy). This study also evaluated deep learning models for retinal lesions' classification, detection, and segmentation. The segmentation task employed the DeepLab-v3+ [9] and HED [10] models. The results presented demonstrate the challenge faced by the models in identifying retinal lesions, mainly in Microaneurysms, where they were obtained in the test set IoU values equal to 0.0325 and 0.0110, respectively, in the DeepLab-v3+ and HED models, highlighting the complexity of the task.

*This work is partly financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) Finance Code 001. This study was partially financed by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brazil.

[1]Postgraduate Program in Computing, Federal University of Pelotas - UFPel, Pelotas, Brazil. (e-mail: {adspereira, marilton, marcelo.sdias, paulo.ferreira, lwtavares}@inf.ufpel.edu.br)

[2]Federal Institute of Education, Science and Technology Farroupilha - IFFar, Alegrete, Brazil. (e-mail: {carlos.santos, fabio.rossi}@iffarroupilha.edu.br; rafaela.2022302825@aluno.iffar.edu.br).

[3]Department of Applied Computing, Federal University of Santa Maria - UFSM, Santa Maria, Brazil. (e-mail: daniel.welfer@ufsm.br; ornellas@inf.ufsm.br).

[4]Department of Clinical Medicine, Federal University of Santa Maria - UFSM, Santa Maria, Brazil. (e-mail: carlos.jesus-haygert@ufsm.br).

[5]Industrial Technical College, Federal University of Santa Maria - UFSM, Santa Maria, Brazil. (e-mail: kazienko@ctism.ufsm.br).

[6]Department of Clinical Medicine, Federal University of Minas Gerais - UFMG, Belo Horizonte, Brazil. (e-mail: comimfv@gmail.com).

[7]Department of Information Systems and Computing, Regional University of Blumenau - FURB, Blumenau, Brazil. (e-mail: aureliof@furb.br).

[8]Federal Institute of Education, Science and Technology Farroupilha - IFFar, Santo Ângelo, Brazil. (e-mail: juliano.weber@iffarroupilha.edu.br).

[9]Postgraduate Program in Computing, Federal University of Rio Grande do Sul - UFRGS, Porto Alegre, Brazil. (e-mail: leonardo.guths@ufrgs.br)

[10]Center of Computational Sciences, Federal University of Rio Grande - FURG, Rio Grande, Brazil. (e-mail: rodrigo.guerra@furg.br).

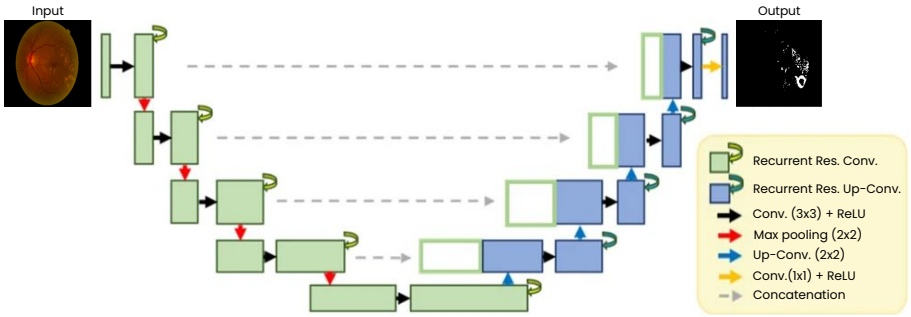

Fig. 1. R2U-Net architecture diagram. Source: Adapted from Shim et al. [5].

In Anand et al. [11], the base U-Net architecture received some modifications to improve the accuracy in the segmentation of blood vessels, Hard Exudates, and Microaneurysms in fundus images. The study utilized the following public datasets: IDRiD, DIARETDB1, STARE, ChaseDB1, DRIVE, and HRF. Although the work did not segment Hemorrhages and Soft Exudates, the results presented suggest great potential in the segmentation of retinal lesions with the U-Net architecture, highlighting the accuracies obtained in the IDRiD set of 99.86% in the segmentation of Exudates Hard and 99.9% in targeting Microaneurysms.

Nair et al. [12] proposed a new neural network called MesU-Net for retinal image segmentation. This model combines U-Net and MesNet. The authors segmented Hard Exudates, Hemorrhages, Microaneurysms, and blood vessels, obtaining accuracies of 97.6%, 98.1%, 99.2%, and 83.7%, respectively. Furthermore, they used public datasets to train the models, with the IDRiD set dedicated to segmenting fundus lesions and the DRIVE set focused on vessel segmentation.

III. R2U-Net Architecture

R2U-Net was developed based on the Deep Residual Network [13], Recurrent Convolutional Neural Network (RCNN) [14] and U-Net [7] models in an attempt to optimize the advantages of these three deep learning models. Among the main factors that differ the architecture from R2U-Net to U-Net, the following stand out: (1) Use of RCLs (*Recurrent Convolutional Layers*) and RCLs with residual units; (2) Efficient method of accumulating characteristics in RCL units; and, (3) Elimination of the cutting and copying unit present in U-Net.

Adopting RCLs and RCLs with residual units replacing conventional convolutional layers gives the architecture a more efficient and deeper structure. Furthermore, the efficient feature accumulation method in RCL units incorporated by R2U-Net improves convergence during the training and testing phases. Such an implementation significantly contributes to enhanced feature extraction and representation. Finally, the cut and copy unit in U-Net is replaced exclusively by concatenation operations. These architectural modifications aim to optimize the effectiveness of R2U-Net, exploring state-of-the-art strategies to improve the model's performance, especially in medical image segmentation.

Figure 1 illustrates the architectural model of the neural network that makes up the proposed approach. This structure comprises two main parts: (1) Coding, represented by modules in green, and (2) Decoding, highlighted by modules in blue. Additionally, the architecture incorporates recurring and residual operations. Recurrent operations enable the reuse of information in different areas of the network, promoting a more comprehensive understanding of contexts. On the other hand, residual processes introduce shortcuts between layers to reduce challenges associated with performance degradation in deeper networks. Experimental results presented by the authors indicate that such operations positively impact the training and testing phases of the model without implying an increase in the number of network parameters.

IV. Materials and Methods

The proposed work used an R2U-Net as a basis to segment retinal lesions associated with DR. We conducted both experiments on hardware consisting of an AMD Ryzen 5 2600X Six-Core @ 12x 3.6GHz processor, an NVIDIA TITAN Xp GPU video card with 12 GB of VRAM, and 16 GB of RAM. To further enhance the accuracy of the proposed work, we incorporated a data augmentation step.

A. Datasets

Two datasets of DR were used in this work, both publicly available and consisting of color images of the fundus of the eye obtained from eye exams. The Dataset for Diabetic Retinopathy (DDR) [8] has 13,673 fundus images, but 757 have lesion annotations at the pixel level. They annotated the following lesions: Hard Exudates, Soft Exudates, Hemorrhages, and Microaneurysms. The Indian Diabetic Retinopathy Image Dataset (IDRiD) was created from clinical exams acquired in an Indian ophthalmology clinic. This set has 516 images of the fundus of the eye. Of these images, 81 have annotations of the lesions at the pixel level.

B. Data Augmentation

Among the main challenges faced in constructing computational models for segmenting fundus lesions associated with DR, the small number of lesions labeled in retinal images stands out. To overcome this problem, we performed a data augmentation step on the images in the training set, aiming to

improve the model's effectiveness in the segmentation of retinal lesions. At this stage, we applied several techniques to expand the available data in the library, named Albumentations: Fast and Flexible Image Augmentations [15]. We employed the following methods: Horizontal Flip, Vertical Flip, Elastic Transform, Grid Distortion, and Optical Distortion.

C. Architecture Training and Adjustment

For semantic segmentation of fundus lesions, an R2U-Net was used, as mentioned previously, thus training a model for each type of lesion. In the training processes, we divided images and masks with lesion annotations from the DDR set into Training, Validation, and Test sets in a proportion of 50:20:30, respectively. Furthermore, we configured the network input size to accommodate images with dimensions of 256×256×3 in the RGB (Red, Green, Blue) color space, and we resized both pictures accordingly. We trained each model for 50 epochs, using a learning rate 0.001 and a batch size of 4. Furthermore, the optimizer was Adam, combined with the ReLU activation function. We exclusively utilize the validation set to adjust the proposed models, aiming to optimize the hyperparameters and achieve improved results. Table I presents the hyperparameters used in this work's training.

TABLE I
HYPERPARAMETERS USED IN THE PROPOSED WORK.

Hyperparameters	**Value**
Activation Function	ReLU
Batch Normalization	True
Batch Size	4
Filters for each down and upsampling levels	[64, 128, 256, 512, 1024]
Input Size	(256, 256, 3)
Learning Rate	0.001
Number of Epochs	50
Optimizer	Adam
Output Activation	Sigmoid

D. Performance Metrics

We evaluated the proposed work using the metrics: Accuracy (Acc), Sensitivity (Sen), Precision (Pre), Dice Coefficient (DC), and Intersection over Union (IoU). For a deeper understanding of these metrics, it is crucial to understand the concepts of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

V. RESULTS

We compared the proposed work with various state-of-the-art models that perform semantic segmentation of objects. In the experiments, we utilized the following models: (1) HED [8]; (2) DeepLab-v3+ [8]; (3) U-Net [7]; (4) U-Net++ [16]; (5) Attention U-net [17]; and, (6) R2U-Net [4]. Furthermore, the U-Net models, U-Net++ and Attention U-net, were also evaluated using pre-trained weights on the ImageNet [18] dataset. The Table II presents the results obtained in the segmentation of Hard Exudates, Hemorrhages, Soft Exudates, and Microaneurysms using the metrics Acc, Sen, Pre, DC, and IoU in dataset DDR validation set. The values highlighted in bold represent the best results obtained in each metric in the different types of lesions. When analyzing Table II, it is evident that the proposed work achieved better results than other models in most metrics. Accuracies of 0.9997, 0.9998, 0.9992, and 0.9997 were achieved for EX, HE, SE, and MA, respectively, with a mean Intersection over Union ($mIoU$) for the classes equal to 0.5969. Figure 2 compares the predictions made by the proposed work during the segmentation of fundus lesions and the ground truth of these images from the test set of the DDR dataset. When comparing columns (b) and (c), despite not having detected some lesions, it is notable that the proposed work managed to segment most of the instances in the selected images accurately. Finally, experiments were conducted on the IDRiD set to assess the accuracy of the proposed work on diverse datasets. Thus, our model trained on the DDR set was used to segment the lesions contained in the *dataset* IDRiD validation set. Table III presents the results obtained with the metrics IoU and $mIoU$. The proposed work achieved IoU values of 0.4983, 0.4992, 0.4993, and 0.4998 for the classes EX, HE, SE, and MA, with a $mIoU$ of 0.4992.

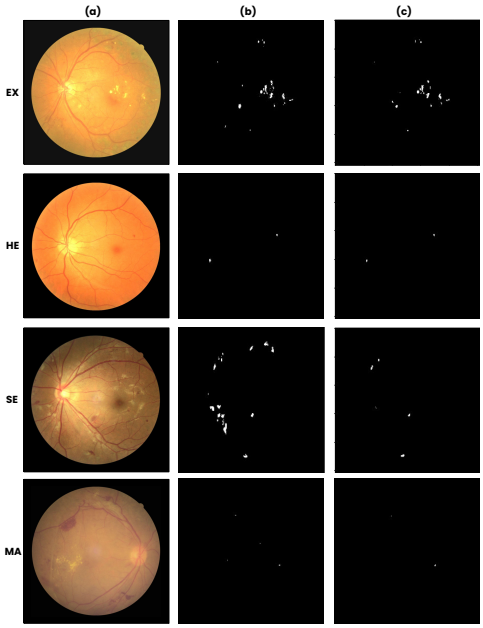

Fig. 2. Visual comparison between the fundus lesion instance segmentations performed by the proposed work with ground truth in images from the test set of the DDR dataset. (a) Original images; (b) Ground Truth; and (c) Proposed work. Source: Own authorship.

VI. CONCLUSIONS

This work introduced a convolutional neural network model for the segmentation of instances of fundus lesions associated with Diabetic Retinopathy using an R2U-Net combined with data augmentation techniques, which improved the model's accuracy, including the use of small batch sizes. The best results obtained by the proposed work were in the dataset DDR validation set, with accuracies equal to 99.97%, 99.98%, 99.92%, and 99.97% for Exudates Hard, Hemorrhages, Soft Exudates,

TABLE II

RESULTS OBTAINED IN THE SEGMENTATION OF HARD EXUDATES, HEMORRHAGES, SOFT EXUDATES, AND MICROANEURYSMS COMPARED TO OTHER MODELS USING THE METRICS Acc, Sen, Pre, DC, AND IoU IN THE VALIDATION SET OF THE DDR DATASET.

Models	EX					HE					SE					MA				
	Acc	*Sen*	*Pre*	*DC*	*IoU*	*Acc*	*Sen*	*Pre*	*DC*	*IoU*	*Acc*	*Sen*	*Pre*	*DC*	*IoU*	*Acc*	*Sen*	*Pre*	*DC*	*IoU*
HED [8]	-	-	-	-	0.0948	-	-	-	-	0.2183	-	-	-	-	0.0379	-	-	-	-	0.0204
DeepLab-v3+ [8]	-	-	-	-	0.2910	-	-	-	-	0.2819	-	-	-	-	0.2756	-	-	-	-	0.0429
U-Net [7]	0.9994	0.1666	0.3064	0.1218	0.5429	0.9984	0.0000	0.0000	0.0000	0.4990	0.9990	0.0000	0.0000	0.0000	0.4992	0.9997	0.0000	0.0000	0.0000	0.4999
U-Net++ [16]	0.9995	0.0000	0.0000	0.0000	0.4995	0.9984	0.0000	0.0000	0.0000	0.4990	0.9990	0.0000	0.0000	0.0000	0.4992	0.9997	0.0000	0.0000	0.0000	0.4999
Attention U-net [17]	0.9995	0.1434	0.3737	0.1279	0.5448	0.9984	0.0000	0.0000	0.0000	0.4990	0.9990	0.0149	0.5928	0.0238	0.5066	0.9997	0.0000	0.0000	0.0000	0.4999
R2U-Net [4]	0.9992	0.3631	0.2948	0.2073	0.5720	0.9962	0.0740	0.1270	0.0509	0.5137	0.9991	0.1313	0.7759	0.1682	0.5601	0.9997	0.0408	**0.5528**	0.0629	0.5200
U-Net with imageNet weights	0.9994	**0.4639**	0.2295	0.2119	0.5757	0.9205	0.0690	0.4838	0.0020	0.0035	**0.9992**	**0.2656**	0.7786	**0.3054**	**0.6184**	0.9997	0.0319	0.2575	0.0465	0.5158
U-Net++ with imageNet weights	0.9995	0.0000	0.0000	0.0000	0.4995	0.9984	0.0000	0.0000	0.0000	0.4990	0.9990	0.0000	0.0000	0.0000	0.4992	0.9997	0.0000	0.0000	0.0000	0.4999
Attention U-net with imageNet weights	0.9993	0.4197	0.1966	0.1842	0.5644	0.9982	0.0247	0.6314	0.0412	0.5118	**0.9992**	0.2357	0.7889	0.2912	0.6091	0.9997	0.0115	0.1909	0.0177	0.5052
Proposed work	**0.9997**	0.3234	**0.4809**	**0.2705**	**0.6053**	**0.9988**	**0.2284**	**0.7585**	**0.2991**	**0.6094**	**0.9992**	0.2230	**0.8604**	0.2796	0.6072	**0.9997**	**0.2298**	0.3418	**0.2185**	**0.5656**

TABLE III

RESULTS OBTAINED IN THE SEGMENTATION OF FUNDUS LESIONS COMPARED TO THE OTHER MODELS USING THE METRICS IoU E $mIoU$ IN THE IDRID DATASET VALIDATION SET.

Models	*IoU*				*mIoU*
	EX	**HE**	**SE**	**MA**	
U-Net	0.4987	0.4999	0.4999	0.0000	0.3746
U-Net++	0.0000	0.0000	0.0000	0.0000	0.0000
Attention U-net	0.4989	0.4999	0.4995	0.0000	0.3745
R2U-Net	0.4963	0.4956	0.4994	0.4999	0.4978
U-Net with imageNet weights	0.4975	0.4595	0.4994	0.4999	0.4890
U-Net++ with imageNet weights	0.0000	0.0000	0.0000	0.0000	0.0000
Attention U-net with imageNet weights	0.4978	0.4997	0.4996	0.4999	**0.4992**
Proposed work	0.4983	0.4992	0.4993	0.4998	**0.4992**

and Microaneurysms, respectively, totaling an average class accuracy of 99.87% and a $mIoU$ of 59.69%. Furthermore, the proposed work was superior to the other models in the IDRiD set, reaching $mIoU$ equal to 49.92% and 49.90%, respectively, in the validation and test sets. These results demonstrate significant potential in the use of encoder-decoder architectural models, based on U-Net, for the segmentation of retinal lesions related to DR. Despite the promising results achieved by the proposed work, it is still necessary to further improve the accuracy of the model to improve the early diagnosis of Diabetic Retinopathy and optimize the lesion annotation process, to facilitate the creation of new DR datasets. In future work to achieve these objectives, it will be necessary to explore new architectural models, incorporate pre-processing steps to improve the extraction of image features within the neural network, and explore ensemble techniques combining different encoder-decoder deep neural network models to create an even more precise approach.

REFERENCES

[1] T.-E. Tan and T. Y. Wong, "Diabetic retinopathy: Looking forward to 2030," *Frontiers in Endocrinology*, vol. 13, p. 1077669, 2023.

[2] J. Kanimozhi, P. Vasuki, and S. M. M. Roomi, "Fundus image lesion detection algorithm for diabetic retinopathy screening," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 7407–7416, 2021.

[3] H. Wang, P. Cao, J. Yang, and O. Zaiane, "Mca-unet: multi-scale cross co-attentional u-net for automatic medical image segmentation," *Health Information Science and Systems*, vol. 11, no. 1, p. 10, 2023.

[4] 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," *arXiv preprint arXiv:1802.06955*, 2018.

[5] J.-H. Shim, W. S. Kim, K. G. Kim, G. T. Yee, Y. J. Kim, and T. S. Jeong, "Evaluation of u-net models in automated cervical spine and cranial bone segmentation using x-ray images for traumatic atlanto-occipital dislocation diagnosis," *Scientific Reports*, vol. 12, no. 1, p. 21438, 2022.

[6] 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.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[8] 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

[9] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11211 LNCS, pp. 833–851, 2018.

[10] S. Xie and Z. Tu, "Holistically-nested edge detection," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015 Inter, pp. 1395–1403, 2015.

[11] M. Anand *et al.*, "Channel and spatial attention aware unet architecture for segmentation of blood vessels, exudates and microaneurysms in diabetic retinopathy," 2023.

[12] A. T. Nair, A. ML, and A. K. MN, "Segmentation of retinal images using improved segmentation network, mesu-net." *International Journal of Online & Biomedical Engineering*, vol. 19, no. 15, 2023.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, Las Vegas, NV, USA, 27–30 June 2016, 2016, pp. 770–778.

[14] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3367–3375.

[15] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information (Switzerland)*, vol. 11, no. 2, 2020.

[16] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer, 2018, pp. 3–11.

[17] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[18] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, pp. 211–252, 2015.