Impact of Data Augmentation on Retinal OCT Image Segmentation for Diabetic Macular Edema Analysis

Daniel Bar-David[1]([✉]), Laura Bar-David[2], Shiri Soudry[2,3,4], and Anath Fischer[1]

[1] Faculty of Mechanical Engineering, Technion Israel Institute of Technology, Haifa, Israel
danielba@technion.ac.il
[2] Department of Ophthalmology, Rambam Health Care Campus, Haifa, Israel
[3] Clinical Research Institute at Rambam, Rambam Health Care Campus, Haifa, Israel
[4] Ruth and Bruce Faculty of Medicine, Technion Israel Institute of Technology, Haifa, Israel

Abstract. Deep learning models have become increasingly popular for analysis of optical coherence tomography (OCT), an ophthalmological imaging modality considered standard practice in the management of diabetic macular edema (DME). Despite the need for large image training datasets, only limited number of annotated OCT images are publicly available. Data augmentation is an essential element of the training process which provides an effective approach to expand and diversify existing datasets. Such methods are even more valuable for segmentation tasks since manually annotated medical images are time-consuming and costly. Surprisingly, current research interests are primarily focused on architectural innovation, often leaving aside details of the training methodology. Here, we investigated the impact of data augmentation on OCT image segmentation and assessed its value in detection of two prevalent features of DME: intraretinal fluid cysts and lipids. We explored the relative effectiveness of various types of transformations carefully designed to preserve the realism of the OCT image. We also evaluated the effect of data augmentation on the performance of similar architectures differing by depth. Our results highlight the effectiveness of data augmentation and underscore the merit of elastic deformation, for OCT image segmentation, reducing the dice score error by up to 23.66%. These results also show that data augmentation strategies are competitive to architecture modifications without any added complexity.

Keywords: Deep learning · Data augmentation · Elastic deformation · OCT · DME

1 Introduction

Diabetic retinopathy (DR) and one of its major sight-threatening complications, diabetic macular edema (DME), are the leading causes of vision loss in individuals with diabetes

mellitus. Early diagnosis and prompt treatment of DME are critical to prevent permanent vision loss [1]. The clinical features of DME include retinal thickening or fluid retention, sometimes with a cystic pattern, which may be accompanied by deposition of protein or lipid within the central retinal tissue.

Optical coherence tomography (OCT) is an ophthalmological imaging modality based on optical reflectivity, which can provide cross-sectional images and three-dimensional volumetric data of the retina [2]. Because of its simplicity, availability, and ability to provide abundance of information, OCT has become essential in the clinical practice, enabling better diagnosis and management of various retinal conditions, including DME.

Interpretation of OCT images requires trained retina experts and can be complex even for experienced clinicians. Moreover, human readings are notably time-consuming, with variable repeatability and interobserver agreement [3]. The application of computer-aided diagnosis (CAD) systems to medical imaging can significantly facilitate their interpretation, including detection of ophthalmic diseases such as DME. In the past few years, the use of deep-learning models in CAD has greatly improved the ability to detect clinical abnormalities in medical imaging, resulting in improved results [4]. To date, however, no retinal CAD system has become commercially available for routine clinical use, largely due to methodological challenges [5].

A major obstacle to the implementation of deep-learning algorithms for medical image analysis is the absence of large, annotated datasets required for training of neural networks. This partly stems from the level of expertise and extent of effort required for proper data interpretation and labeling, but also from ethical considerations required by data protection laws. Consequently, there are only a few publicly available OCT datasets collected from multiple imaging devices, most of which often comprise a relatively limited number of scans and represented pathologies [6].

The performance of deep-learning models in computer vision depends on the neural networks training, architecture and model scaling [7]. Moreover, architectural innovation is broadly regarded as the main focus of research interest, leaving aside critical details of the training methodology [8]. Specifically, only a few studies investigated the efficiency of data augmentation in convolutional neural network training for image classification and segmentation. The data augmentation type is often stated, but little is explained about the method, range, and frequency of the process [9].

Presently, for most computer vision problems, basic transformations such as random flipping, rotating, scaling, shifting or adjusting contrast are valuable regularizers which can generalize the model and reduce overfitting by expanding and diversifying datasets without acquiring new images [10].

Elastic deformation is a more complex approach for data augmentation, introducing higher-order transformation. Utilization of elastic deformation for training of convolutional neural networks was first introduced on the MNIST handwritten digit dataset [11] where after deformation the image still appeared sufficiently plausible to represent a real digit. Along with basic transformations, elastic deformation is particularly suitable for non-rigid objects, yet at the same time it is complicated to construct since it alters the inner elements of the image. Medical images deal with objects which can inherently undergo natural transformations that can be described as elastic deformations. Indeed, different methods of elastic deformation have been applied for medical image registration [12]. Yet, due to the difficulty of achieving elastic deformation methods

for images with complex morphology, and perhaps also due to the indeterminate value of this approach thus far, elastic transformation for data augmentation has been less commonly used. A previously reported elastic deformation technique for data augmentation was applied on OCT scans of the optic nerve head to render the network invariant with atypical morphology [13]. Recently, an elastic deformation method was clinically validated for OCT images of patients suffering from DME [14]. However, to date, the impact of transformations on neural network performance for OCT image analysis, and particularly the benefit of elastic deformation, has been relatively understudied.

Here, to provide a systematic approach for OCT data segmentation for DME, we explored the benefits of data augmentation with a particular examination of the added value of elastic deformation. We first investigated the impact of diverse data augmentation methods on an established neural network for segmentation of OCT images from subjects with DME, and determined the relative effectiveness of the different approaches. We then evaluated the impact of data augmentation in relation with the depth of the neural networks by comparing the performance of two similar architectures differing by depth in OCT image segmentation.

2   Methods

2.1   Data Augmentation

Basic transformations are augmentation techniques commonly applied to most learning algorithms, as they are intuitive, easy to understand, and straightforward to implement. The inner composition of the image is essentially unaffected, but represents a variation in the image acquisition process such as the subject position or a physical property of the photographic system. Several basic transformation methods were evaluated on OCT scans and are described in Table 1.

Table 1. Description of the augmentations applied during neural network training

	Augmentation	Description
A	No transformation	Original OCT scan
B	Horizontal flip	Horizontally flip the OCT scan
C	Rotation	Randomly rotate the OCT scan in the range $\pm 15°$
D	Shift	Randomly translate horizontally and vertically by up to 10% of the image height and width
E	Scale	Randomly scale sampled from the interval [0.9,1.1]
F	Brightness, contrast, saturation	Modify the brightness, contrast and saturation by a random factor [0.75, 1.25]
G	Noise	Add gaussian noise with a variance of 0.07
H	Basic transformations	Combine and apply transformations A to G together
I	Elastic deformation	Apply elastic transformation with an intensity $\sigma = 9$
J	All transformations	Combine and apply transformations B to G and I together

Elastic deformation is a higher level of data augmentation that modifies the inner elements of the image, thus potentially affecting its intrinsic pattern and altering its realism. OCT retina images often represent anatomically complex features and thus following the introduction of additional distortion to the inner structure, practicing retina specialists highly-familiar with typical features of clinical OCT images can evaluate their authenticity to avoid potential bias. A recent study reported a process of clinical validation of the degree of elastic deformation that can be applied to OCT scans with DME while preserving their realistic value [14].

(a) Grid displacement   (b) Displacement field   (c) Overlay on the scan and mask   (d) Resulting scan and mask
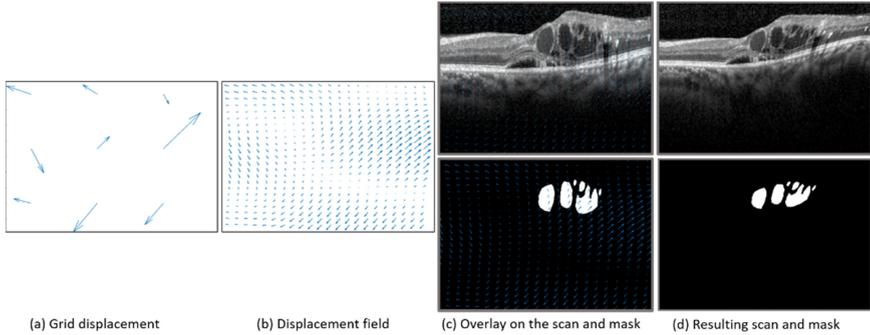
Fig. 1. Elastic deformation process (a) Generate a displacement grid; (b) then a smooth displacement field is interpolated using spline interpolation on the grid displacement; (c) Superimposing the displacement field on the mask and OCT scan by using bilinear interpolation; (d) Result: deformed OCT scan and mask

The outline of the elastic deformation method is as follow. First, a uniform 2D grid of 3×3 control points is generated from a normal distribution of mean $\mu = 0$ and standard deviation σ (Fig. 1(a)). Then a displacement field is created by using spline interpolation between values of the 3×3 grid (Fig. 1(b)). Finally, the displacement grid is applied on the original OCT scan and on the mask by using bilinear interpolation (Fig. 1(c)), resulting in a deformed OCT image and mask (Fig. 1(d)). To keep the elastic deformation realistic for OCT images with DME, the maximum deformation intensity σ is equal to 9 [14].

2.2 Segmentation Network

To assess the impact of various types of transformations on segmentation tasks, we use a convolutional neural network (CNN) based on U-net architecture [15]. We choose the U-net model since it has gaineds tremendous recognition and popularity in recent years in medical image segmentation. Indeed, most of segmentation neural networks of OCT scans, rely on its encoder decoder design removing or adding layers [16], adding skip connection [17], modifying convolution [13] or pooling [18].

The segmentation network receives as input an OCT scan, and outputs a segmentation map that predicts for each pixel if it belongs to intraretinal fluid cysts (IRF), intraretinal

lipid (IRL) or background. The symmetric structure of the network is shown in Fig. 2 including the contracting path for analyzing context information, the expanding path for synthesizing the output of the contracting and merging path that transfers local and accurate information. The proposed architecture differs from the original U-net in the following manner: batch normalization is added after each block of convolution, conv-transposed is used instead of up-convolution and filters are resized.
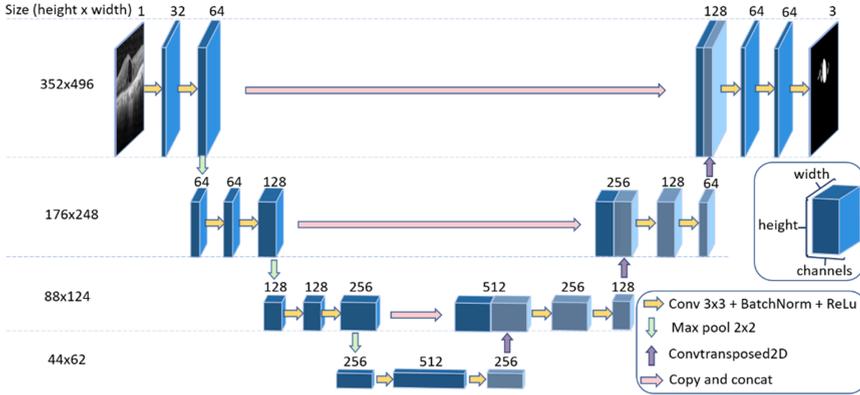
Fig. 2. Illustration of the auto encoder for segmentation based on u-net. Each blue block corresponds to a multi feature map. The model gets an OCT scan and the output is a mask. (Color online figure)

A common used metric for medical image segmentation is the Dice coefficient that compares the pixel wise agreement between a segmentation model prediction and their corresponding ground truth. The formula for Boolean data is defined as follow:

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{1}$$

Where (TP) is True positive, (FP) False positive and (FN) False negative. The best segmentation is reached when dice $= 1$ while dice $= 0$ refers to a wrong segmentation. The dice loss function is defined as follow:

$$Loss_{dice} = 1 - \frac{1}{l} \sum_{j=1}^{l} \frac{\sum_{i=1}^{N} 2y_i^j \hat{y}_i^j + \varepsilon}{\sum_{i=1}^{N} y_i^j + \sum_{i=1}^{N} \hat{y}_i^j + \varepsilon} \tag{2}$$

Where y is the ground truth, $\hat{y}$ is the predicted segmentation, l is the total number of labels and N is the number of pixels. ε avoids the division by zero.

2.3   Comparison of Shallow Network Versus Deep Network

Following Alexnet architecture [19], researchers have mostly created deeper and more complex networks to increase performance [20–22]. These architectural expansions are

beneficial for improving performance but make the network less efficient by increasing complexity and decreasing speed. Recently, studies have emphasized model efficiency by optimizing training methods and scaling strategies [8, 9, 23]. Here, a comparison of two segmentation networks differing by their depth is presented, where the objective is to determine if architectural differences can be overshadowed by data augmentation. The "deep network" refers to the segmentation network presented in Sect. 2.2 (Fig. 2) and consists in 22,807,617 parameters. To provide an objective comparison, the "shallow network" was constructed by removing the last level of the deep network (supplementary Table 1). As a result, the number of parameters for the shallow network drop by 35% to 14,940,737 parameters.

3 Evaluation

3.1 Dataset and Training Process

OCT scans were obtained from patients treated at the Retina service of the department of Ophthalmology, Rambam Health Care Campus, Haifa, Israel from 2016 to 2019. B-scans were extracted from the Heidelberg Spectralis device using a 49-line raster macula scan. The size of each OCT image used in this study consists of 352×496 pixels and no subsampling is applied.

OCT volume-scans of 120 subjects affected by DME were randomly extracted where only a single cross-section image of the macula was selected per each scan. Two of the most prevalent clinical features associated with DME, namely intraretinal fluid cysts (IRF) and intraretinal lipid (IRL) deposits, were manually segmented by a trained ophthalmologist and reviewed by a retinal expert. The data was randomly split into three sets: 60% for the training, 20% for the validation and 20% for the test set.

The network was trained with a batch size of 8 using Adam optimizer. Data augmentation was performed online at each epoch during the training session to remove memory constraints. The probability that an image undergoes a transformation is 0.5. We used pytorch library on a single NVIDIA Titan V GPU.

3.2 Evaluation of Data Augmentation Impact on Segmentation

To evaluate the impact of data augmentation on OCT segmentation, the dice score, sensitivity (Se) and specificity (Sp) metrics were calculated.

Table 2 summarizes results obtained for each transformation on the test set for the shallow and deep network. Specificity is close to one because there are many more background pixels than object pixels. Sensitivity is the true positive rate and measures the proportion of object pixels that are correctly identified. When each basic transformation (B-G) is applied separately, the dice score is only slightly improved compared to the baseline (A). However, when they are combined together (H) there is a significant increase over the baseline (A). Paradoxically, even applied alone, elastic deformation (I) performs as well as all basic transformations (H). Best performances are obtained with a combination of all transformations (J).

Table 2. Performance of the shallow and deep segmentation models on IRF and IRL

	Augmentation		Shallow network			Deep network		
			Dsc (%)	Se (%)	Sp (%)	Dsc (%)	Se (%)	Sp (%)
A	No transformation	IRL	50.14	56.73	99.86	53.17	56.25	99.88
		IRF	70.2	73.46	99.58	75.2	79.13	99.62
B	Horizontal flip	IRL	57.32	**61.81**	99.88	58.1	57.45	**99.91**
		IRF	70.97	76.39	99.57	75.64	80.44	99.49
C	Rotation	IRL	56.18	57.86	99.88	57.36	59.43	99.89
		IRF	72.06	80.91	99.54	76.68	80.06	99.61
D	Shift	IRL	52.62	54.78	99.89	54.61	57.7	99.88
		IRF	72.51	71.07	**99.74**	75.22	79.51	99.54
E	Scale	IRL	56.57	58.76	98.87	57.57	59.86	99.88
		IRF	71.82	70.5	99.68	76.63	78.26	**99.69**
F	Brightness, contrast, saturation	IRL	50.28	59.74	99.82	54.6	56.23	99.9
		IRF	72.66	74.08	99.6	76.45	80.78	99.59
G	Noise	IRL	53.33	50.39	**99.92**	55.69	55.44	99.89
		IRF	70.43	78.73	99.54	75.7	79.66	99.62
H	Basic transformations	IRL	58.47	59.15	99.9	60.25	60.69	99.9
		IRF	76.02	**82.08**	99.51	78.05	81.8	99.63
I	Elastic deformation	IRL	58.69	57.74	**99.92**	60.2	59.17	**99.91**
		IRF	76.1	81.34	99.57	78.64	**82.75**	99.63
J	All transformations	IRL	**60.23**	61.06	99.9	**62.03**	**67.07**	99.87
		IRF	**77.25**	81.94	99.55	**79.36**	82.48	99.61

The chart presented in Fig. 3 compares our shallow and deep network for IRF and IRL. For both features and with the same transformations applied during the training, the deep network (in red) is always greater than or equal to the shallow network (in blue). But with all transformations (J) the shallow network succeeds to perform better than the deep network with each single basic transformation (A–G) and is comparable to elastic deformation and all basic transformations performance (H, I).

To get a quantitative sense of the effect of basic transformations (H), elastic deformation (I) and all transformations (J), the improvement in the test set accuracy was calculated in Table 3. Improvement over baseline is greater for shallow than deep network for IRF and IRL. Data augmentation has resulted by a reduction of 23.66% in the dice score error (IRF, shallow network). All transformations (J) reduced the dice score error by up to 3.9% compared to basic transformations and elastic deformation.

Figure 4 illustrates an example of segmentation outputs of the IRF and IRL features for the deep network with no augmentation (A) and with all transformations (J).
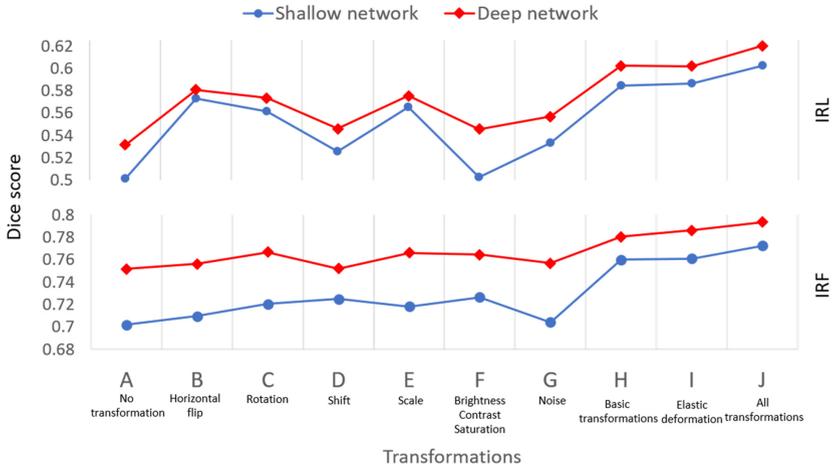
Fig. 3. Dice score values for distinct transformations on IRL and IRF. Each color represents a segmentation network: in blue the shallow and in red the deep. (Color figure online)

Table 3. Improvement when transformations are added. Difference of the dice score error (%).

Index	Augmentation	IRF Δ (%)		IRL Δ (%)	
		Shallow	Deep	Shallow	Deep
A	No transformation	0	0	0	0
H	Basic transformations	9.53	11.49	16.7	15.1
I	Elastic deformations	19.8	13.87	17.1	15
J	All transformations	23.66	16.77	20.2	18.9
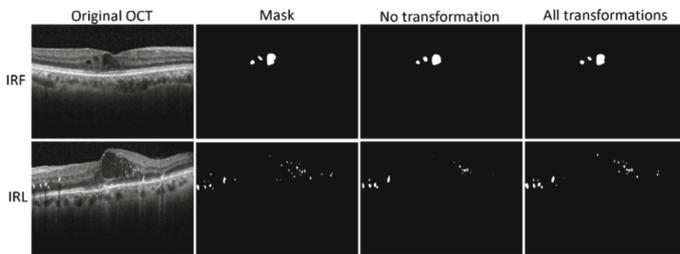
Fig. 4. Examples of IRF and IRL segmentation results with no transformation (A) and with all transformations (J) during the training for the deep network.

4 Discussion

This work investigated the contribution of data augmentation methods on OCT images from subjects with DME for improving model performance, and showed their competitiveness to CNN architecture modification. Recent works, using deep-learning for automated segmentation of OCT scans have shown promising results, but thus far most studies prioritized architecture modification over training methods. Moreover, even when data augmentation methods have been applied, studies generally briefly mention or describe them putting more emphasis on architecture modifications [24–26]. Whereas architectural improvements are indeed essential, the paucity of annotated datasets makes data augmentation crucial in the field of deep-learning of medical images. The results of our study highlighted the positive impact of data augmentation on OCT image segmentation when the transformations applied are carefully designed to preserve the realism of the images and to avoid bias.

Best results for IRF and IRL segmentation were achieved when all transformations were applied. Yet, sometimes transformations may add less value depending on the segmentation task. Moreover, as IRL features are smaller than IRF, manual and computational segmentation are less accurate, resulting in a lower dice score.

Each basic transformation applied separately improved much less performance than elastic deformation (I). Also, all basic transformations (H) applied together yielded similar results as elastic deformation (I). Its complexity compared to basic transformations has probably limited its application to OCT scans, but these results showed that they should be more commonly apply.

The comparison of two similar architectures differing by depth showed that architecture improvements can be overshadowed by data augmentation. Indeed, the shallow network using all transformations (J) for training outperformed the deep network without transformation (A) and with each basic transformation (B-G). Moreover, the performance was very tie for the shallow network using all transformations (J) and the deep network using all basic transformations (I).

Further studies will determine the proper balance between training methodology and architecture modification to reduce complexity and increase efficiency.

References

1. Cheloni, R., Gandolfi, S.A., Signorelli, C., Odone, A.: Global prevalence of diabetic retinopathy: protocol for a systematic review and meta-analysis. BMJ Open **9**(3), e022188 (2019). https://doi.org/10.1136/bmjopen-2018-022188
2. Huang, D., et al.: Optical coherence tomography. Science **254**(5035), 1178–1181 (1991). https://doi.org/10.1126/science.1957169
3. Patel, P.J., Browning, A.C., Chen, F.K., da Cruz, L., Tufail, A.: Interobserver agreement for the detection of optical coherence tomography features of neovascular age-related macular degeneration. Investig. Ophthalmol. Vis. Sci. **50**(11), 5405–5410 (2009). https://doi.org/10.1167/iovs.09-3505

4. Lynch, C.J., Liston, C.: New machine-learning technologies for computer-aided diagnosis. Nat. Med. **24**(9), 1304–1305 (2018). https://doi.org/10.1038/s41591-018-0178-4

5. Stolte, S., Fang, R.: A survey on medical image analysis in diabetic retinopathy. Med. Image Anal. **64**, 101742 (2020). https://doi.org/10.1016/j.media.2020.101742

6. Yanagihara, R.T., Lee, C.S., Ting, D.S.W., Lee, A.Y.: Methodological challenges of deep learning in optical coherence tomography for retinal diseases: a review. Transl. Vis. Sci. Technol. **9**(2), 11 (2020). https://doi.org/10.1167/tvst.9.2.11

7. Tan, M., Le, Q.V.: EfficientNet: rethinking model scaling for convolutional neural networks. In: ICML 2019, pp. 10691–10700 (2019)

8. Bello, I., et al.: Revisiting ResNets: improved training and scaling strategies (2021). http://arxiv.org/abs/2103.07579. Accessed 02 June 2021

9. Zoph, B., Cubuk, E.D., Ghiasi, G., Lin, T.-Y., Shlens, J., Le, Q.V.: Learning data augmentation strategies for object detection. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12372, pp. 566–583. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58583-9_34

10. Dvornik, N., Mairal, J., Schmid, C.: On the importance of visual context for data augmentation in scene understanding. IEEE Trans. Pattern Anal. Mach. Intell. **43**(6), 2014–2028 (2021). https://doi.org/10.1109/TPAMI.2019.2961896

11. Simard, P.Y., Steinkraus, D., Platt, J.C.: Best practices for convolutional neural networks applied to visual document analysis. In: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR 2003, pp. 958–963 (2003). https://doi.org/10.1109/ICDAR.2003.1227801

12. Wang, M., Li, P.: A review of deformation models in medical image registration. J. Med. Biol. Eng. **39**(1), 1–17 (2018). https://doi.org/10.1007/s40846-018-0390-1

13. Devalla, S.K., et al.: DRUNET: a dilated-residual u-net deep learning network to digitally stain optic nerve head tissues in optical coherence tomography images. Biomed. Opt. Express **9**(7), 3244–3265 (2018). https://doi.org/10.1364/boe.9.003244

14. Bar-David, D., et al.: Elastic deformation of optical coherence tomography images of diabetic macular edema for deep-learning models training: how far to go? (2021). https://arxiv.org/abs/2107.03651v1

15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

16. Venhuizen, F.G., et al.: Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks. Biomed. Opt. Express **8**(7), 3292 (2017). https://doi.org/10.1364/BOE.8.003292

17. Tennakoon, R., Gostar, A.K., Hoseinnezhad, R., Bab-Hadiashar, A.: Retinal fluid segmentation in OCT images using adversarial loss based convolutional neural networks. In: Proceedings – International Symposium on Biomedical Imaging, pp. 1436–1440. May 2018. https://doi.org/10.1109/ISBI.2018.8363842

18. Asgari, R., Waldstein, S., Schlanitz, F., Baratsits, M., Schmidt-Erfurth, U., Bogunović, H.: U-Net with spatial pyramid pooling for Drusen segmentation in optical coherence tomography. In: Fu, H., Garvin, M.K., MacGillivray, T., Xu, Y., Zheng, Y. (eds.) OMIA 2019. LNCS, vol. 11855, pp. 77–85. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32956-3_10

19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst., 1–9 (2012). https://doi.org/10.1016/j.protcy.2014.09.007

20. Szegedy, C., et al.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9. June 2015. https://doi.org/10.1109/CVPR.2015.7298594

21. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
22. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: CVPR, pp. 4700–4708 (2017)
23. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: practical automated data augmentation with a reduced search space. In: CVPR Workshops, pp. 702–703 (2020)
24. He, Y., et al.: Fully convolutional boundary regression for retina OCT segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 120–128. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_14
25. Apostolopoulos, S., De Zanet, S., Ciller, C., Wolf, S., Sznitman, R.: Pathological OCT retinal layer segmentation using branch residual U-shape networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 294–301. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_34
26. Lu, D., et al.: Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network. Med. Image Anal. **54**, 100–110 (2019). https://doi.org/10.1016/j.media.2019.02.011