Incremental Learning For Fundus Image Segmentation

Javier Civit-Masot, Luis Muñoz-Saavedra, F. Luna-Perejón, Juan M. Montes-Sánchez, M. Domínguez-Morales
Robotics and Computer Technology Lab
Avda. Reina Mercedes s/n, E.T.S. Ingeniería Informática, Universidad de Sevilla, Sevilla, Spain
Email: {jcivit, luimunsaa, fralunper, jmontes, mdominguez}@atc.us.es

***Abstract*—Automated Fundus image segmentation is traditionally done in the image acquisition instrument and, thus, in this case it only needs to be able to segment data from this acquisition source. Cloud providers support multi GPU and TPU virtual machines making attractive the idea of cloud-based segmentation an interesting possibility. To implement this idea we need to make correct predictions for fundus coming from different sources. In this paper we study the possibility of building a web base segmentation service using incremental training, i.e, we initially train the system using data from a single data set and, afterwards, perform retraining with data from other acquisition sources. We are able to show that this type of training is efficient and can provide good results suitable for web-based segmentation.**

***Keywords*—Deep learning; Incremental Learning; U-Net; Image Segmentation; Eye Fundus; Optic disc; Glaucoma Detection**

I. Introduction

Segmentation is the process of detection of limits within an image. Deep Learning methods for segmentation are limited by the number and the variety of the training images. For cloud services it is necessary to train with new data set samples from different acquisition sources periodically. The objective of this paper is to study the effects of incremental training where we initially train with single dataset and later refine the training process by adding data from additional datasets.

A. Cloud based Image Segmentation

There are two scenarios for image segmentation in a clinical set up. In the first case the segmentation tool is marketed together with the acquisition instrument. Thus, we can train using images provided by the instrument linked to the segmentation software. In a second case segmentation is implemented as a web-based service. In this case we need to segment images coming from different sources.

Some works on combined dataset training in [1] are available but they are not related to image segmentation. In [2], [3] different data sets are use to train and test with each data set independently. [4] Uses several datasets but training is performed only with a combined dataset.

In this paper we will first train the system using a single data set and perform predictions over that dataset and also on images acquired with other instruments. Later we will perform a small (5 epoch) retraining using images from the other dataset and study the influence in the results. We apply these techniques to the detection of optical disc in eye fundus images.

B. U-Net Networks

For this study we will a generalized U-Net. U-Net [5] has been used for many medical segmentation problems including glaucoma [2] and diabetic macular edema [6].U-Net can be considered as a type of deconvolutional network [7]. In these networks a set of convolutional layers outputs a deep small representation of the original image. This highly encoded representation is decoded to the original image size using a set of upsampling layers. This network is shown in Figure 1. Its structure is describe in the original paper but has been widely modified by many researchers (e.g. [2], [3]

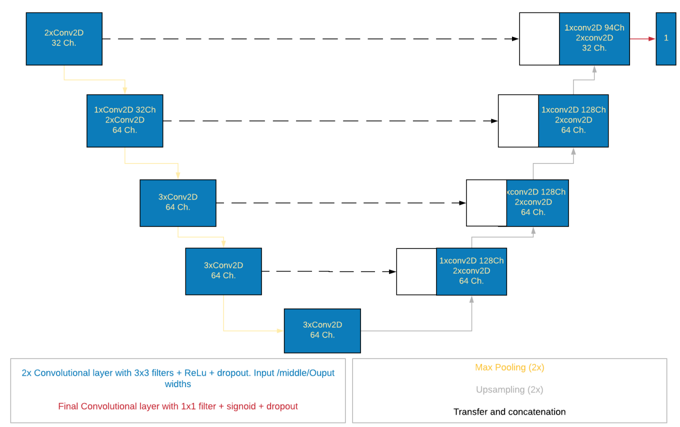

Fig. 1: U-Net Architecture

C. Fundus image glaucoma indicators

Glaucoma is a set of diseases that provoke damage to the optic nerve at the back of the eye causing loss of vision over time. [8]. Intraocular Hypertension (IH) is the most significant risk factor associated to glaucoma.

IH causes damage to the beginning of the the optic disc (OD) which is the beginning of the optic nerve. Optic disc can be visualized using many techniques including fundus color photography. The OD is made up by two subregions (Figure 2) a peripheral area (neuroretinal border) and a white central region (optic cup - OC).

As glaucoma develops, the OC increases occupying a large part of the OD. The ratio of the OC radium to that of the OD is known as CDR (Cup to Disc ratio) and is a glaucoma indicator [9]. The measurement of the OC and OD radii is needed to calculate the CDR. OD and OC human segmentation is

 ISBN: 978-1-61208-763-4

difficult and leads to many errors. Machine based segmentation is, thus, attractive.Many machine learning (ML) alternatives are available for fundus image OD segmentation. [10], [2], [3], [11], [12].

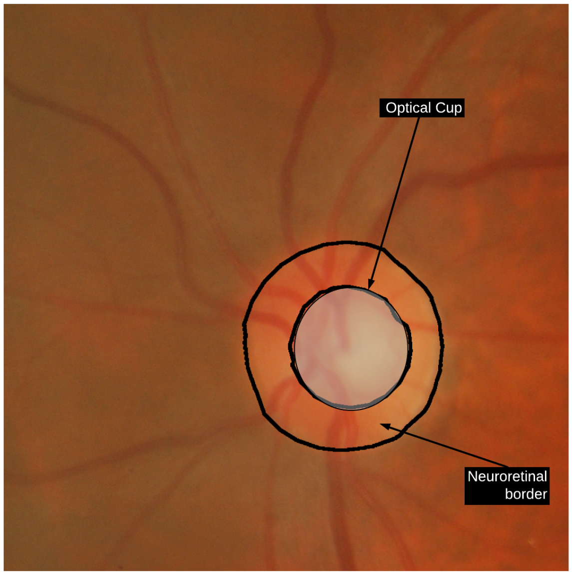

Fig. 2: Optic Disc and Cup

We do not introduce a new technique but studies the influence of retraining on the results. Thus we will only study the segmentation of the OD, although our technique is equally applicable to the OC case. The used approach based on [3] but with many major changes introduced in [4] to make it more adequate for a cloud implementation.

II. Materials and methods

In this paper we use a generalized 6 layer U-Net. It has only 40 channels in the first layer and the layer increment ratio, i.e. the ratio of the number of channels in a layer to that of the following one [13] is 1.1 instead of 2. Considerig that we resize the dataset images to 128x128 this reduces the number of trainable parameters to less than 1 million.

In this paper we test the feasibility of incremental training using the web resources and networks that we would deploy in a web service. This is ndifferent from other papers (e.g. [2], [3], [11], [12] where the data from a single source is used for training and testing.

We use Google Collaboratory python notebook environment for our implementation and apply a recursive flexible U-net model. We perform aggressive static and dynamic data augmentation modifying the approach proposed in [14].

For training and testing we use 96 image batches as this is suitable for GPU implementation. We use is 25 epochs, 256 training steps and 6 validation steps for each epoch. The employed optimizer is Adam with a 0.0007 learning rate. This values have proven adequate for training our U-Net and give excellent results with reasonable training times. We perform random sub-sampling based cross validation. Our loss function is the negative log of the Sørensen-Dice coefficient [15] (Dice).

We use the RIM-ONE v3 and DRISHTI data sets. RIM ONE [16] comes from the Spanish University of La Laguna and includes 159 images. DRISTI [17] comes from Aravind Eye Hospital, Madurai, India and includes 101 images. Both data sets have been tagged by expert ophtalmologists.

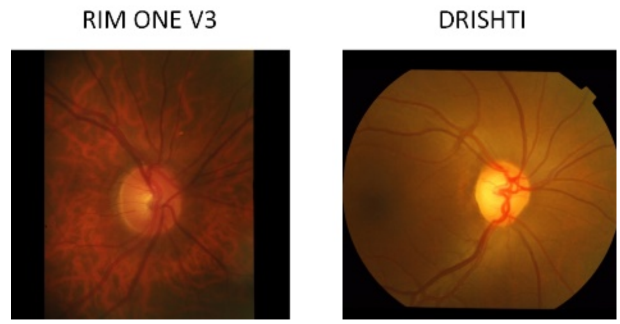

Fig. 3: Images from RIM ONE and DRISHTI datasets

Figure 3 shows that images in each dataset, which clearly have been captured using different devices and have, thus, different characteristics.Our processing approach is very similar to [4] but, in our case we don't crate a mixed dataset but train with a single dataset (DRISHTI) and then perform a small number of retrain steps using the RIM ONE dataset.

The ratio between OD and OC diameters (CDR), is one of the most popular glaucoma indicators. We define a clinically significant parameter (RRP -Radii Ratio parameter) based on the ratio between the radius of the machine segmented and that of the ground truth segmented discs. The RPP is defined as the percent of test images for which the radius error is below 10%.

We compare our work with the results from papers that use Deep Learning for optic disc segmentation and uses DRISHTI dataset. Zilly et. al. [12] use a light three layer CNN with sophisticatd pre and post-processing and apply it independently to both datsets. Sevastopolsky [3] uses a very light U-Net architecture and provides results for the RIM ONE data set. Al-Bander [2] uses a widely modified dense U-Net and provides results for both datasets. Shankaranarayana [11] uses a modified residual U-Net and provides results for the RIM ONE dataset.

III. Results

We want to find out how our system behaves when it is trained with a dataset and then lightly retrained with the other and compare our results with those obtained when a single dataset (i.e. RIM ONE or DRISHTI) is applied to train the system. We will compare our results to those by other authors who use a single data set for training and validation. The Dice coefficient is used to estimate the similarity between the correct and predicted disc. This figure of merit, also known as

 ISBN: 978-1-61208-763-4

F1 score, is widely used and allows us to compare our results with those from other work. Dice coefficient is defined as:

$$DC = \frac{2TP}{2TP + FP + FN} \quad (1)$$

In this equation TP indicates true positives, FP false positives, and FN false negatives.

In Table I results for Disc segmentation for our two study cases are shown. In the first one train using just the DRISHTI data set and validate using the part of that dataset not used for training and the other dataset. In the second scenario we so a short (5 epoch) retrain using the RIM ONE and data set. Thus our scenarios are the following:

- 75% of the DRISHTI dataset is used for training and after validation is carried out first with the rest of DRISHTI data set and then with the full RIM ONE data set.
- 75% of the RIM ONE data set is used to retrain the network and then we validate with the rest of the test part of both data sets.

TABLE I. OD segmentation Dice (Mean/Best/Worst) and RRP

	Dice-DRI	Dice-RIM
DRI-Trained	0.98	0.64
RIM-retrained	0.89	0.80

We can see in table I that when we train with the DRISHTI dataset results when testing with images from that dataset get good Dice coefficient values. Specifically we get a mean dice value of 0.98 (DRISHTI) but only 0.64 (RIM1) for OD segmentation. The situation is worst than it looks as in the worst case for RIM the segmentation thus not produce any pixel. When we retrain the network with the other data set results are 0.89 (DRISHTI) and 0.80 (RIM). For the worst segmentation case we get a Dice value of 0.69. Thus, we can see that with a very light retraining the network can quickly learn the specific characteristics of the second dataset.

TABLE II. OD segmentation RRP

	RRP-DRI	RRP-RIM
DRI-Trained	100%	23%
RIM-retrained	89%	80%

In Table II we show the the percentage of predictions that estimate the OD radius with an error smaller than 10% as a percentage. This data is clinically relevant as the CDR (ratio between the cup and disc radii) is a widely used glaucoma indicator. We can see that without retraining only 23% of the predicted disk segmentations have radium error bellow 10%. After the quick retrainning this value increases to 80%.

In table III we include results from other papers that have performed OD segmentation using Deep Learning methods and have trained with one of the datasets used in our study. These papers have, in every case, trained and tested with each data set independently.

Although we use networks with a small number of trainable parameters, when training with a single dataset we get results for that dataset that are similar to those obtained by other research papers. When training with the DRISHTI dataset we obtained a Dice value of 0.98 for OD segmentation. This value is slightly above 0.97 [12].

TABLE III. OD segmentation Dice comparison.

Author	**DRI**	**RIM ONE**
Zilly et al. [12]	0.97	-
Al-Bander [2]	0.95	0.90
Sevastopolsky [3]	-	0.94
Shankaranarayana et al. [11]	-	0.98
Drishti Trained	0.98	0.64
RIM Retrained	0.89	0.80

The most significant results in table III come after the retrainning that is not performed in the other studies. The results obtained when we do a quick retrain show that, in this casem we get good prediction results for all the test images. This demonstrates that it will not be feasible to create a service using training data captured with a single acquisition device.

IV. CONCLUSIONS AND FUTURE WORK

We have shown that by performing a fast retrain when adding data from a new dataset, and by preprocessing images and performing static and dynamic data augmentation, we can implement disc segmentation with an equivalent performance to that reported by researchers who use a single dataset both for evaluation and testing.

We also define a clinically significant parameter (Radii Ratio parameter- RRP) that can be useful to estimate the quality of the CDR estimations and thus, to give some confidence on the quality of the system for glaucoma prediction.

This work shows the importance of retraining when adding new image sources to the segmentation system. In a real clinical segmentation service scenario, we would have to start training the network with the initially available data and retrain it when new images from different instruments become available. The possibility of improving the network architecture by the inclusion of residual blocks [18] or the combination of a these blocks and a conventional U-Net [19] has been shown effective in several medical segmentation applications and could potentially improve the performance of our segmentation process. The robustness of these networks when analyzing images from many different instruments is still an open issue for the future.

ACKNOWLEDGMENT

Development in Cloud environment was supported by GoogleCloud platform research credit program.

REFERENCES

[1] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8789–8797.

[2] B. Al-Bander, B. Williams, W. Al-Nuaimy, M. Al-Taee, H. Pratt, and Y. Zheng, “Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis,” *Symmetry*, vol. 10, no. 4, p. 87, 2018.

[3] A. Sevastopolsky, “Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network,” *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 618–624, 2017.

[4] J. Civit-Masot, F. Luna-Perejón, S. Vicente-Díaz, J. M. Rodríguez Corral, and A. Civit, “Tpu cloud-based generalized u-net for eye fundus image segmentation,” *IEEE Access*, vol. 7, pp. 142 379–142 387, 2019.

[5] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[6] C. S. Lee, A. J. Tyring, N. P. Deruyter, Y. Wu, A. Rokem, and A. Y. Lee, “Deep-learning based, automated segmentation of macular edema in optical coherence tomography,” *Biomedical optics express*, vol. 8, no. 7, pp. 3440–3448, 2017.

[7] H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1520–1528.

[8] H. A. Quigley and A. T. Broman, “The number of people with glaucoma worldwide in 2010 and 2020,” *British journal of ophthalmology*, vol. 90, no. 3, pp. 262–267, 2006.

[9] H. A. Quigley, “Better methods in glaucoma diagnosis,” *Archives of Ophthalmology*, vol. 103, no. 2, pp. 186–189, 1985.

[10] M. C. V. S. Mary, E. B. Rajsingh, J. K. K. Jacob, D. Anandhi, U. Amato, and S. E. Selvan, “An empirical study on optic disc segmentation using an active contour model,” *Biomedical Signal Processing and Control*, vol. 18, pp. 19–29, 2015.

[11] S. M. Shankaranarayana, K. Ram, K. Mitra, and M. Sivaprakasam, “Joint optic disc and cup segmentation using fully convolutional and adversarial networks,” in *OMIA 2017*, ser. Fetal, Infant and Ophthalmic Medical Image Analysis. Springer International Publishing, 2017, Conference Proceedings, pp. 168–176.

[12] J. Zilly, J. M. Buhmann, and D. Mahapatra, “Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation,” *Computerized Medical Imaging and Graphics*, vol. 55, pp. 28–41, 2017.

[13] A. G. Howard and et al., “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.

[14] B. Zoph, E. D. Cubuk, G. Ghiasi, T.-Y. Lin, J. Shlens, and Q. V. Le, “Learning data augmentation strategies for object detection,” *arXiv preprint arXiv:1906.11172*, 2019.

[15] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[16] F. Fumero, S. Alayón, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, “Rim-one: An open retinal image database for optic nerve evaluation,” in *2011 24th international symposium on computer-based medical systems (CBMS)*. IEEE, 2011, Conference Proceedings, pp. 1–6.

[17] J. Sivaswamy, S. Krishnadas, G. D. Joshi, M. Jain, and A. U. S. Tabish, “Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation,” in *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*. IEEE, 2014, Conference Proceedings, pp. 53–56.

[18] P. Xiuqin, Q. Zhang, H. Zhang, and S. Li, “A fundus retinal vessels segmentation scheme based on the improved deep learning u-net model,” *IEEE Access*, vol. 7, pp. 122 634–122 643, 2019.

[19] S. Kim, W. Bae, K. Masuda, C. Chung, and D. Hwang, “Fine-grain segmentation of the intervertebral discs from mr spine images using deep convolutional neural networks: Bsu-net,” *Applied Sciences*, vol. 8, no. 9, p. 1656, 2018.

 ISBN: 978-1-61208-763-4