Weakly Supervised Segmentation of Geographic Atrophy on SD-OCT scans

Dmitrii Lachinov[1], Christoph Grechenig[1], Magdalena Baratsits[1,2], Stefan Sacu[1,2], Hrvoje Bogunović[1], Ursula Schmidt-Erfurth[1]

[1]Department of Ophthalmology, Medical University of Vienna, Austria
[2]Vienna Clinical Trial Center, Medical University of Vienna, Austria

Purpose

To segment geographic atrophy (GA) lesions on spectral-domain optical coherence tomography (SD-OCT) images with deep learning trained with weak labels in the form of atrophy lesion size. Also, we compare the proposed approach with existing fully-supervised methods.

Methods

We constructed a convolutional neural network for **GA segmentation from 3D OCT volumes**. The task of the network is to **regress** the area of the GA expressed as a percentage of the field of view occupied by the GA. The CNN follows the U-Net[1] architecture and consists of residual blocks[2]. The final output layer has a sigmoid activation followed by the mean operation over spatial dimensions (Fig. 1).
The neural network is trained with **weak labels** - incomplete, noisy, or labels that don't correspond to the given task. In our case, we want to get atrophy pixel-level segmentation, but we have information about image-level atrophic area size only. The example is shown in Fig. 2.
To solve this challenging task, the neural network must learn to highlight patterns on the image by finding a relationship between the provided image-level weak label and the desired segmentation mask.

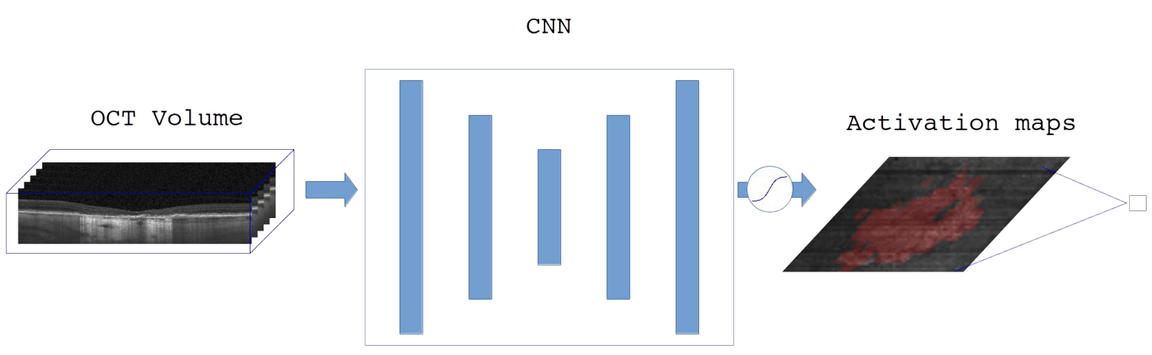

Figure 1. The schematic representation of the approach employed.

We compare the proposed method against a series of fully-supervised methods: its fully supervised version, Fully Convolutional Network (FCN)[3], UNet with 2D and 3D convolutions[1,4], and Selected Dimensions[5] baseline. Besides, we added a comparison with the semi-supervised Deep Voting method [6].

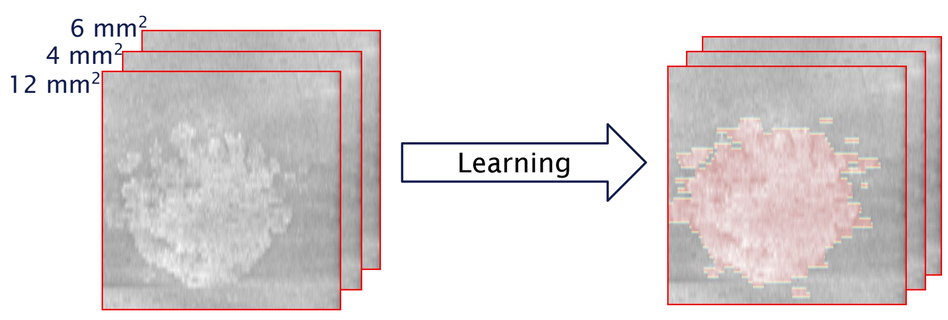

Figure 2. The example of weak labels and desired outcome.

Data

Scans taken at different time points from a follow-up with Geographic Atrophy at baseline:

- I92 SD-OCT scans,
- 37 patients,
- 70 eyes in total.

We split patients with stratification by study center and lesion size:

- Training: 25
- Validation: 5
- Testing: 7

Results

The **GA was successfully segmented** solely with the provided lesion size as a label. The proposed method achieved a Dice score (DSC) of 0.61 on a validation dataset. At the same time, a fully supervised version achieved a DSC of 0.78, and a semi-supervised version a DSC of 0.5. The results are summarised in Table 1.
For further analysis of the results, we provide the boxplot of the Dice Scores for each model (Fig. 3). The dashed line split the plot into three parts: models trained with full supervision - segmentation masks; a single model trained with weak image-level labels; and a model trained with partly-annotated data, where labels were missing for some samples.
As can be seen from the boxplot (Fig. 3), the test set has multiple challenging samples that influence the score the most. The overall performance of the models depends on how well it generalizes to handle the hard samples.

In the fully-supervised group of models, we can see that the Selected Dimensions (SD) model underperforms compared to other models. It's due to overfitting for easy samples, since the SD model has the largest number of learnable parameters among all methods. At the same time, the proposed method trained on pixel-level annotations **achieves the best performance**.
Analyzing the performance of the weakly- and semi-supervised models, we can conclude that the quality of the annotation heavily influences the final result. However, the weakly-supervised version of the proposed model trained on the image-level labels achieved substantial performance in the pixel-level segmentation task.

Data	Method	Dice Score
Fully supervised	Proposed	0.78
	FCN	0.73
	UNet	0.73
	UNet3D	0.67
	Selected Dimentions	0.67
Weakly-supervised	Proposed	0.61
Semi-supervised	Deep Voting	0.5

Table 1. The results on the testing dataset

Dice Score (DSC)

$$DSC = \frac{2TP}{2TP + FP + FN}$$

Where:
TP - True Positives
FP - False Positives
FN - False Negatives

Conclusion

We proposed a **weakly-supervised geographic atrophy segmentation algorithm** operating on SD-OCT that utilizes image-level lesion size information during the training and compared this approach to the existing ones. The algorithm can learn the concept of Geographic Atrophy from a regression task alone, without the help of human-annotated pixel masks. Weakly-supervised algorithms naturally don't stand the comparison with fully supervised ones. However, we can notice substantial improvement of the proposed method over the other method trained on incomplete or missing data.
The employed method is a baseline for further research on the capabilities of deep learning methods. The results give us an estimate of the data quality needed for the pixel-level segmentation. Also, we once more time demonstrated that the deep learning methods **can learn a task from its abstract description** or produce pixel-level annotations while trained on image-level labels. We believe that it can offer more interpretable deep learning approaches in medicine.

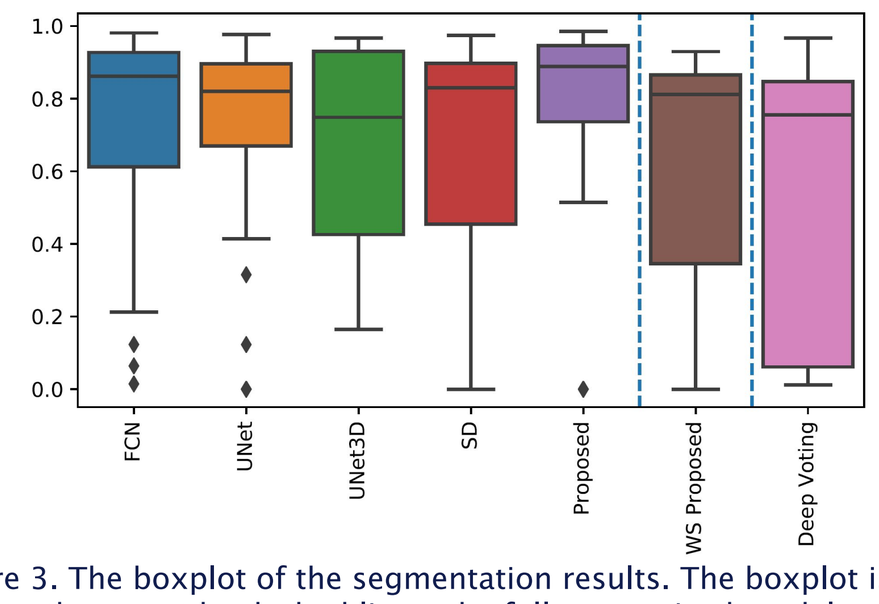

Figure 3. The boxplot of the segmentation results. The boxplot is split into three part by dashed lines: the fully supervised models, the weakly supervised model and the semi-supervised model.

References

1. Ronneberger, O., P. Fischer and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." MICCAI (2015).
2. He, Kaiming, X. Zhang, Shaoqing Ren and Jian Sun. "Deep Residual Learning for Image Recognition." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 770-778.
3. Shelhamer, Evan, J. Long and Trevor Darrell. "Fully Convolutional Networks for Semantic Segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2017): 640-651.
4. Çiçek, Özgün, A. Abdulkadir, S. Lienkamp, T. Brox and O. Ronneberger. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation." ArXiv abs/1606.06650 (2016): n. pag.
5. Liefers, Bart, Cristina González-Gonzalo, C. Klaver, B. Ginneken and C. Sánchez. "Dense Segmentation in Selected Dimensions: Application to Retinal Optical Coherence Tomography." MIDL (2018).
6. Ji, Z., Qiang Chen, Sijie Niu, T. Leng and D. Rubin. "Beyond Retinal Layers: A Deep Voting Model for Automated Geographic Atrophy Segmentation in SD-OCT Images." Translational Vision Science & Technology 7 (2018): n. pag.