PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Deep-learning for thyroid microstructure segmentation in 2D OCT images

Tampu, Iulian Emil, Eklund, Anders, Haj-Hosseini, Neda

Iulian Emil Tampu, Anders Eklund, Neda Haj-Hosseini, "Deep-learning for thyroid microstructure segmentation in 2D OCT images," Proc. SPIE 11630, Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XXV, 116301Z (5 March 2021); doi: 10.1117/12.2576854

SPIE.

Event: SPIE BiOS, 2021, Online Only

Deep-learning for thyroid microstructure segmentation in 2D OCT images

Iulian Emil Tampu^a, Anders Eklund^{a,b,c}, and Neda Haj-Hosseini^a

^aDept. of Biomedical Engineering, Linköping University, Sweden 581 85

^bDept. of Computer and Information Science, Linköping University, Sweden 581 83

^cCenter for Medical Image Science and Visualization, Linköping University, Sweden 581 85

ABSTRACT

Optical coherence tomography (OCT) can provide exquisite details of tissue microstructure without traditional tissue sectioning, with potential diagnostic and intraoperative applications in a variety of clinical areas. In thyroid surgery, OCT could provide information to reduce the risk of damaging normal tissue. Thyroid tissue's follicular structure alters in case of various pathologies including the non-malignant ones which can be imaged using OCT. The success of deep learning for medical image analysis encourages its application on OCT thyroid images for quantitative analysis of tissue microstructure. To investigate the potential of a deep learning approach to segment the follicular structure in OCT images, a 2D U-Net was trained on *b-scan* OCT images acquired from *ex vivo* adult human thyroid samples affected by a range of pathologies. Results on a pool of 104 annotated images showed a mean Dice score of 0.74 ± 0.19 and 0.92 ± 0.09 when segmenting the follicular structure and the surrounding tissue on the test dataset ($n=10$ images). This study shows that a deep learning approach for tissue microstructure segmentation in OCT images is possible. The achieved performance without requiring manual intervention encourages the application of a deep-learning method for real-time analysis of OCT data.

Keywords: Deep-learning, U-Net, segmentation, optical coherence tomography, tissue microstructure, thyroid, intraoperative.

1. INTRODUCTION

The major clinical application of OCT has so far been in ophthalmology, where it is commonly used nowadays for the diagnosis, among others, of glaucoma and retinal diseases.¹ Technological advancements have facilitated the application of OCT also in imaging of the coronary artery wall and visualization of the gastrointestinal tract.^{1,2} The implementation of quantitative analysis tools along with the hardware improvements, would widen the range of implementation of OCT in clinics. The fast development of deep learning methods for medical image analysis³ could speed up the development of analysis tools, complementing the qualitative evaluation of the OCT images with quantitative information. Deep learning has been implemented in ophthalmology⁴⁻⁶ and cardiology,⁷ where convolutional neural networks (CNNs) were used for segmentation and identification of pathological conditions in OCT images. Other medical fields could also benefit from similar implementations by using OCT as a tool for imaging of tissue microstructure with micrometer resolution and without traditional tissue sectioning.

A promising application of OCT is in the field of surgery, where it could provide diagnostic information both intraoperatively and during pathology examination. An example is thyroid surgery, where tissue identification is challenging with the risk of harming parathyroid glands and vocal cords as well as performing unnecessary thyroidectomies, that can result in high patient morbidity and need for life-long medication. Several studies have addressed the changes of the thyroid tissue microstructure due to various diseases, showing that using OCT it is possible to differentiate between normal and diseased thyroid tissue.^{8,9} In a recent work, it was shown that the microstructural features visible in the OCT images highly agreed with histopathology for a variety of pathologies.¹⁰ Moreover, quantitative 3D analysis of the thyroid follicular structure was performed *ex vivo* using traditional image processing techniques, resulting in the quantification of tissue microstructure features which showed distinction between normal and pathological tissue. However, this method has several limitations

Corresponding author: Iulian Emil Tampu, iulian.emil.tampu@liu.se

Figure 1. Training and testing pipeline. *b-scan* OCT images along with the annotations are augmented and used to train 5 models using a 5-fold cross validation scheme. The segmentation performance was evaluated on the test dataset, where test-time augmentation was used to improve performance. The final segmentation for each test sample is obtained as the ensemble of the predictions of the 5 models. M1-M5 refers to the 5 models trained through cross-validation.

including: (1) manual intervention is needed for the selection of the region of interest to analyze, (2) the slow processing of each OCT volume, which includes noise reduction, segmentation and its refinement, hinders the use of the method in real-time applications, and (3) follicular segmentation is poor in deep regions of the sample.

In this work, we address these limitations by using a deep-learning method for the segmentation of the follicular structure using 2D *b-scans* as input for the model. Deep-learning methods can have long training times, ranging from hours to days. However, once the model is trained, the time needed for segmenting one image is in the order of hundreds of milliseconds on a common computer hardware, opening up possibilities for real-time applications. In addition, the proposed method does not require any manual intervention, with the OCT data entering the segmentation model only after isotropic resampling. The aim of this work was to implement and train a 2D U-Net model¹¹ for thyroid follicular structure segmentation in OCT *b-scans*, in order to assess the possibility of using a similar model for real-time applications.

2. MATERIAL AND METHODS

Images of thyroid tissue samples from 22 random adult patients (20 female and 2 male), age 53 ± 18 years (mean $\pm$ standard deviation), were used in this study. The tissue samples were collected under local ethical approval (Dnr 2012/237-31, 2014/452-32 and 2015/463-32). Samples were affected by a variety of pathologies including goiter, adenoma, papillary cancer and hashimoto thyroiditis. 3D OCT images were acquired as described in¹⁰ using a spectral domain OCT imaging system (Telesto II, Thorlabs, Inc., NJ, USA). Each OCT volume was then resampled to a 0.01mm isotropic resolution with resulting volumes of $250 \times 250 \times 200$ voxels in the x, y and z dimensions, respectively. Among the resampled volumes, a total of 104 *b-scans* (x, z) with size 250×200 pixels were randomly selected for training and testing. Each 2D image was labeled into three classes identifying follicles, surrounding non-follicular tissue and the non-tissue region. This latter includes the portion of the *b-scans* that images the space between the OCT probe and the tissue sample, as well as the zero-padding introduced by the resampling procedure. The reason for assigning separate labels (annotations) to the surrounding non-follicular tissue and the non-tissue region is motivated by the future perspective of using this dataset for a separate analysis of the non-follicular tissue regions. Annotations were performed using a two step approach: initially, an unsupervised segmentation method as described in¹² was used to produce a rough segmentation of the three classes described above. In the second step the segmentations were manually refined to obtain the final annotations.

A 2D encoder-decoder U-Net¹¹ architecture was selected in this work given its success in various medical image segmentation tasks. Each layer in the encoder path consisted of two 3×3 convolutions each with subsequent batch normalization and rectified linear unit (ReLU) activation function, and a max pooling operation with kernel 2×2 and stride two in each direction. In the decoder path, each layer contained a 2×2 up-convolution with stride two in each direction followed by batch normalization and concatenation with the corresponding decoder feature

map, and two 3×3 convolutions with batch normalization each followed by a ReLU. The bottleneck, where the encoder and the decoder paths meet, consisted of two 3×3 convolutions with batch normalization and dropout set to 0.2. The overall model included 3 layers, with 64 initial filters that doubled after each max pooling operation resulting in 256 feature maps in the network bottleneck.

A 5-fold cross validation scheme was employed where 10% of the data was kept for testing. Given the small number of annotated images available, data augmentation was used at training time by means of random mirroring, rotations ($[-15, 15]$ degrees) and elastic deformation (strength of displacement $[0, 5]$).¹³ During training, the network prediction and the ground truth were compared using a weighted generalized Dice loss function,¹⁴ where weights for the different classes were computed on the training set to alleviate the class imbalance problem. Adam optimizer¹⁵ was used during training with a constant learning rate set to 10^{-5} . Training was stopped early if the validation accuracy did not improve over 30 epochs. During testing, the final segmentation was obtained as an ensemble of the individual segmentation produced by the models trained during cross-validation. In addition, test-time augmentation was also used to improve the quality of the segmentation.¹⁶ Model training and testing were performed using a NVIDIA RTX 2080 Ti graphics card with 11 GB memory. A summary of the training and testing pipeline can be seen in Figure 1.

3. RESULTS

Results show a Dice score of 0.74 ± 0.19 and 0.92 ± 0.09 for the follicular structure and the surrounding tissue, respectively. Figure 2 shows four randomly selected test images highlighting the segmentation accuracy for the follicular structure, along with the segmented surrounding tissue (transparent yellow) and non-tissue (transparent blue) region. The segmented follicular surfaces show the correctly segmented pixels (green), false positives (dark blue) and false negatives (red). The dashed region in Figure 2.c identifies the zero-padding introduced by the resampling. The network correctly identifies this region as a non-tissue region.

Figure 2. Representative OCT test images (top row) with corresponding segmentation results (bottom row). The dashed region in (c) identifies the zero-padding introduced by the resampling. Surrounding non-follicular tissue and non-tissue regions are shown yellow and blue in transparency, respectively. Segmentation of the follicular structure is illustrated in detail, with correctly classified pixels in green, false positives in blue and false negatives in red. Dice scores are 0.75, 0.80, 0.81 and 0.76 for (a), (b), (c) and (d), respectively. Scale bars are 500 micrometers in all images.

4. DISCUSSION

OCT *b-scans* were used in this study given the prospect of using the implemented segmentation method during real-time applications. 2D *b-scans* are the common form of output data from intraoperative OCT systems^{17,18} that allow faster acquisition and less complex systems. The time needed by the segmentation method to process an image of 200×200 pixels is 200 ms on a current laptop without graphics card, which is reasonable for real-time applications and even for the rendering of the follicular structure alongside the OCT 2D *en face* and volumetric data.

As discussed in¹⁰, OCT imaging depth is influenced, among others, by the tissue structure including follicular and tissue density. These ultimately impact how distinguishable follicles are from the surrounding tissue, thus affecting their segmentation. OCT *b-scans* were selected as input to the network considering the importance of the depth information contained in such images. The deep-learning model could use this information to generate a spatial relationship between parts in the image, thus improving the segmentation. The proposed method was able to segment follicles as deep as 1.5 mm, as can be seen in Figure 2.

With respect to the model training, the weighted loss function does not penalize the network when close regions are segmented as one. In the case of follicular structure segmentation, being able to segment each follicle independently is important and impacts metrics derived from segmentation results, such as follicular count and volume variability. Thus, it is reasonable to consider a modified loss function that emphasizes the error in segmenting close regions as one. Possible loss function implementations could be based on J statistics, as suggested in literature.¹⁹

To improve the segmentation of follicles deep in the sample, a 3D U-Net architecture could also be used with the advantage of exploiting all the spatial information available in the volumetric data. The major challenge to be addressed for such implementation is the creation of labels for training data. Manual annotation of 3D data is labor intensive. Semi-automatic annotation of 2D images, as the one proposed here, can be used to obtain an initial rough 3D segmentation that can be refined in the next step.

Another possible approach to improve segmentation performance, especially for the follicular structure, is to use a network that combines the information from different kernel sizes. As it was noted during network hyperparameter optimization, changes in kernel size affect the ability of the network when segmenting large follicles or small ones close to each other. With large kernels, such as 5×5 and 7×7 , big follicles were better segmented, whereas small adjacent follicles were correctly identified with a smaller kernel. Literature shows that models incorporating multi-scale features perform better at segmentation.²⁰

In conclusion, a 2D U-Net was implemented and trained on OCT *b-scans* for the segmentation of the follicular structure and surrounding non-follicular tissue. The method allowed a fast and reliable segmentation of the tissue microstructure, without the need for user intervention. The comparable performance and high efficiency of the procedure encourage the implementation of deep-learning methods for real-time analysis of OCT images.

ACKNOWLEDGMENTS

The study was supported by Forskningsrådet i Sydöstra Sverige (FORSS - 931466), AIDA (2019-1908), Åke Wiberg Stiftelse (M19-0455, MT20-0034).

REFERENCES

- [1] Zysk, A. M., Nguyen, F. T., Oldenburg, A. L., Marks, D. L., and Boppart, S. A., "Optical coherence tomography: a review of clinical development from bench to bedside," *Journal of biomedical optics* **12**(5), 051403 (2007).
- [2] Gora, M. J., Suter, M. J., Tearney, G. J., and Li, X., "Endoscopic optical coherence tomography: technologies and clinical applications," *Biomedical optics express* **8**(5), 2405–2444 (2017).
- [3] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I., "A survey on deep learning in medical image analysis," *Medical image analysis* **42**, 60–88 (2017).

- [4] Venhuizen, F. G., van Ginneken, B., Liefers, B., van Grinsven, M. J., Fauser, S., Hoyng, C., Theelen, T., and Sánchez, C. I., "Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks," *Biomedical optics express* **8**(7), 3292–3316 (2017).
- [5] Roy, A. G., Conjeti, S., Karri, S. P. K., Sheet, D., Katouzian, A., Wachinger, C., and Navab, N., "Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomedical optics express* **8**(8), 3627–3642 (2017).
- [6] Garcia-Martin, E., Pablo, L. E., Herrero, R., Ara, J. R., Martin, J., Larrosa, J. M., Polo, V., Garcia-Feijoo, J., and Fernandez, J., "Neural networks to identify multiple sclerosis with optical coherence tomography," *Acta ophthalmologica* **91**(8), e628–e634 (2013).
- [7] Abdolmanafi, A., Duong, L., Dahdah, N., and Cheriet, F., "Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography," *Biomedical optics express* **8**(2), 1203–1220 (2017).
- [8] Zhou, C., Wang, Y., Aguirre, A. D., Tsai, T.-H., Cohen, D. W., Connolly, J. L., and Fujimoto, J. G., "Ex vivo imaging of human thyroid pathology using integrated optical coherence tomography and optical coherence microscopy," *Journal of biomedical optics* **15**(1), 016001 (2010).
- [9] Erickson-Bhatt, S. J., Mesa, K. J., Marjanovic, M., Chaney, E. J., Ahmad, A., Huang, P.-C., Liu, Z. G., Cunningham, K., and Boppart, S. A., "Intraoperative optical coherence tomography of the human thyroid: Feasibility for surgical assessment," *Translational Research* **195**, 13–24 (2018).
- [10] Tampu, I. E., Maintz, M., Koller, D., Johansson, K., Gimm, O., Capitanio, A., Eklund, A., and Haj-Hosseini, N., "Optical coherence tomography for thyroid pathology: 3d analysis of tissue microstructure," *Biomedical Optics Express* **11**(8), 4130–4149 (2020).
- [11] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in *[International Conference on Medical image computing and computer-assisted intervention]*, 234–241, Springer (2015).
- [12] Kim, W., Kanezaki, A., and Tanaka, M., "Unsupervised learning of image segmentation based on differentiable feature clustering," *IEEE Transactions on Image Processing* **29**, 8055–8068 (2020).
- [13] Jung, A. B., Wada, K., Crall, J., Tanaka, S., Graving, J., Reinders, C., Yadav, S., Banerjee, J., Vecsei, G., Kraft, A., Rui, Z., Borovec, J., Vallentin, C., Zhydenko, S., Pfeiffer, K., Cook, B., Fernández, I., De Rainville, F.-M., Weng, C.-H., Ayala-Acevedo, A., Meudec, R., Laporte, M., et al., "imgaug." <https://github.com/aleju/imgaug> (2020). Online; accessed 5-Feb-2020.
- [14] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Cardoso, M. J., "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *[Deep learning in medical image analysis and multimodal learning for clinical decision support]*, 240–248, Springer (2017).
- [15] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).
- [16] Perez, F., Vasconcelos, C., Avila, S., and Valle, E., "Data augmentation for skin lesion analysis," in *[OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis]*, 303–311, Springer (2018).
- [17] Rubinstein, M., Hu, A. C., Chung, P.-S., Kim, J. H., Osann, K. E., Schalch, P., Armstrong, W. B., and Wong, B. J., "Intraoperative use of optical coherence tomography to differentiate normal and diseased thyroid and parathyroid tissues from lymph node and fat," *Lasers in medical science*, 1–10 (2020).
- [18] Sommerey, S., Al Arabi, N., Ladurner, R., Chiapponi, C., Stepp, H., Hallfeldt, K. K., and Gallwas, J. K., "Intraoperative optical coherence tomography imaging to identify parathyroid glands," *Surgical endoscopy* **29**(9), 2698–2704 (2015).
- [19] Pena, F. A. G., Fernandez, P. D. M., Tarr, P. T., Ren, T. I., Meyerowitz, E. M., and Cunha, A., "J regularization improves imbalanced multiclass segmentation," in *[2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)]*, 1–5, IEEE (2020).
- [20] Yang, D., Liu, G., Ren, M., Xu, B., and Wang, J., "A multi-scale feature fusion method based on u-net for retinal vessel segmentation," *Entropy* **22**(8), 811 (2020).