Blood Vessel Segmentation Using U-Net for Glaucoma Diagnosis with Limited Data

Lukas SCHIESSER[a,1], Jens Julian STORP[b], Kemal YILDIRIM[c], Julian VARGHESE[c] and Nicole ETER[b]

[a] *University of Osnabrück, Germany*
[b] *Department of Ophthalmology, University of Münster Medical Center, Germany*
[c] *Institute of Medical Informatics, University of Münster, Münster, Germany*

Abstract. Glaucoma is one of the leading causes of blindness worldwide. Therefore, early detection and diagnosis are key to preserve full vision in patients. As part of the SALUS study, we create a blood vessel segmentation model based on U-Net. We trained U-Net on three different loss functions and used hyperparameter tuning to find their optimal hyperparameters for each loss function. The best models for each of the loss functions achieved an accuracy of over 93%, Dice scores around 83% and Intersection over Union scores over 70%. They each identify large blood vessels reliably and even recognize smaller blood vessels in the retinal fundus images and thus pave the way for improved glaucoma management.

Keywords. Glaucoma, U-Net, Deep Learning, Ophthalmology, Blood vessel segmentation, segmentation

1. Introduction

Glaucoma is one of the most common chronic eye diseases worldwide, causing irreversible visual field defects and blindness. One of the major risk factors for glaucoma progression is an elevated intraocular pressure (IOP) [1]. In medical practice, monitoring of chronic diseases usually requires multiple, repetitive measurements over long time periods in order to adequately evaluate the activity level of a disease and to adjust treatment schemes accordingly. SALUS ("*Selbsttonometrie und Datentransfer bei Glaukompatienten zur Verbesserung der Versorgungssituation*") is a two-arm, multicenter, randomized clinical trial evaluating the state of medical care of glaucoma patients in Germany [2]. In short, glaucoma patients were randomly assigned to one of two groups: 1. inpatient cohort, 2. outpatient cohort. The inpatient group received standard Goldmann applanation tonometry over the course of three consecutive days to assess IOP. Individuals in the outpatient cohort were given iCare Home devices [3], which they used to assess IOP by themselves. The SALUS trial will follow-up on participants until May 2023 and aims to publish results of the comparison of both study arms by the end of 2023. As part of this study, a multi-step deep learning framework is created to assist ophthalmologists with the diagnosis of glaucoma. Although glaucoma is often associated with increased IOP, it is also important to look at physiological changes in the retina. As one-third to half of all patients with early-stage glaucoma

---

[1] Corresponding Author: Lukas Schießer, E-mail: lschiesser@uos.de.

experience progressing visual field loss despite having adequate IOP compared to control groups [4]. Therefore, non-IOP-dependent methods like evaluating the retinal vasculature are needed since vascular changes can be observed in early stages in the glaucoma's pathogenesis [5]. The segmentation of blood vessels in the papillary region of the retina can aid ophthalmologists in this task. This research aims to contribute to non-IOP-dependent methods by developing models for blood vessel segmentation as means of observing vascular changes more efficiently.

2. Methods and Material

Data. We used the DRIVE dataset [6] to train and test our models. It contains 40 images taken from the retinas of human subjects with corresponding expert generated ground truth segmentation maps indicating the location of blood vessels in each image. It is split into 20 training and 20 test images. The dataset is commonly used in medical image analysis and has been widely used to research automated blood vessel segmentation. We additionally used 5 random images from the STARE dataset [7] as a validation dataset for hyperparameter tuning.

    Preprocessing and Data Augmentation. All images were first preprocessed using contrast limited adaptive histogram equalization (CLAHE). Since the dataset is very small, we enlarged the dataset by using different combinations of image augmentation techniques, namely, rotating the images by at most 10° as well as changing the brightness, contrast, and saturation of the fundus images using the albumentations package [8]. Thus, we created 60 additional images for training, totaling 80 training images. No image augmentation techniques were applied to the test and validation images.

    U-Net is a powerful convolutional neural network commonly used for biomedical image segmentation [9]. Its architecture is based on an encoder-decoder network. The encoder which is also sometimes called the *contractive path* downsamples the input image, while the decoder, also sometimes called the *expansive path,* upsamples the encoded representation producing a segmentation mask. U-Net's encoder and decoder networks are connected by a series of skip connections which help to preserve spatial information and improve accuracy of the segmentation. These connections pass information from the encoder to the decoder at multiple scales which allows the decoder to make more accurate predictions by taking into account both global and local information from the input image [10]. U-Net can already achieve results comparable with sliding-window based convolutional networks when trained on an extremely small dataset of a few hundred images. When adding data augmentation, preprocessing and enhancement techniques, it outperforms existing state-of-the-art methods on several biomedical image segmentation challenges [9]. A pretrained architecture can be employed as U-Net's encoder path to improve the model's general performance as measured by accuracy and *Intersection over Union* (IoU) score [11]. We chose a ResNet-34 architecture pretrained on the ImageNet dataset as our encoder path.

    Training. To segment the blood vessels, we employed a U-Net network with a ResNet-34 encoder pretrained on the ImageNet dataset using the Segmentation Models package [12]. We decided to use a mini-batch approach. We further used the Adam optimizer with a learning rate of 0.001. We compared three different loss functions for training the model, namely, Tversky Loss, Dice Loss, and Focal Loss. To find the best-performing versions of our model for all three loss functions, we employed

hyperparameter tuning using Ray Tune [13] during which we also investigated the effect of the number of epochs on the maximum IoU score of our models.

Hyperparameter tuning. As already mentioned, we decided to use the number of epochs and the alpha values of the Tversky and Focal Loss for hyperparameter tuning. The Dice Loss has no tunable hyperparameters. The Tversky Loss's alpha value is used to weigh the relative importance of false positives and false negatives. Since blood vessel segmentation is a very complex task due to the widely ranging sizes of blood vessels in the retina, we want to penalize false negatives more than false positives [14]. A similar reasoning can be applied to the Focal Loss's alpha value. They weigh the different classes of the segmentation task. Since we were more interested in the correct identification of our foreground class (the blood vessels) than the background class, we weighed them more strongly. For finetuning the number of training epochs, we decided to keep the number of epochs trained relatively narrow to avoid overfitting and keep training times short. An overview of which hyperparameter values were tested for which loss can be found in Table 1. The hyperparameter tuning was evaluated for all three loss functions separately. We searched the hyperparameter space with respect to maximizing IoU scores using grid search evaluating the performance of each model on the validation dataset.

Table 1. All possible hyperparameters used during hyperparameter tuning.

Loss function	Hyperparameter	Range
Tversky loss	Epochs	10, 15, 20, 25, 30, 35, 40
	Alpha	0.5, 0.6, 0.7, 0.8, 0.9
Dice loss	Epochs	10, 15, 20, 25, 30, 35, 40
Focal loss	Epochs	10, 15, 20, 25, 30, 35, 40
	Alpha	0.5, 0.6, 0.7, 0.8, 0.9

3. Results

After hyperparameter tuning, we used the best hyperparameters to retrain the models from scratch. The best Tversky Loss combination was trained for 25 epochs with an alpha value of 0.6. It achieved an IoU score of 0.8674 on the validation data. The best Dice Loss combination was trained for 35 epochs and achieved an IoU score of 0.8639 on the validation data. Finally, the Focal Loss performed best during the tuning process with an IoU score of 0.8715. Its alpha value was 0.5 which implicitly results in an 0.5 for the second class in the weighting vector. The model was trained for 30 epochs. The models were then evaluated on a held-out test dataset. The model trained with the Focal Loss function performed best with an IoU score of 0.718. Table 2 includes all test metrics computed for the three different loss functions with their best hyperparameter combination.

Table 2. Best Hyperparameter combination and corresponding metrics during testing

Model	Hyperparameter Configuration	IoU score	Dice Score	Accuracy
Tversky Loss	Alpha: 0.6, Epochs: 25	0.705	0.826	0.929
Dice Loss	Epochs: 35	0.712	0.830	0.931
Focal Loss	Alpha: 0.5, Epochs: 30	0.718	0.836	0.930

4. Discussion

By using hyperparameter tuning, we created three models with solid performance on the task of blood vessel segmentation in fundus images. All three models achieve an accuracy of around 93%. However, an IoU score of just over 70% and a dice score of just over 82% is not necessarily considered satisfactory. Nevertheless, these results still attest a good performance of our models, especially given our small dataset. Looking at the predicted blood vessel segmentation masks and comparing them to the ground truth masks (see Figure 1), we can visually confirm a good performance.

Nevertheless, the reliability of our validation metrics on which we base our decision during hyperparameter tuning on might be subject to improvements. Since we only used five images during validation, we cannot count on the same statistical properties that we normally count on with larger validation datasets. Therefore, we must regard our validation metrics more carefully than with larger datasets. This could be avoided by creating datasets with a larger number of data points, e.g., by merging multiple datasets together, acquiring more images with ground truth masks.

Another error source could be the vast amount of image augmentation that was used to increase the size of the training dataset. By applying multiple augmentation techniques, we increased our training data size from 20 to 80. This can lead to overfitting and a worse performance [15]. Further experiments with differing amounts of image augmentation will be conducted in follow-up studies.
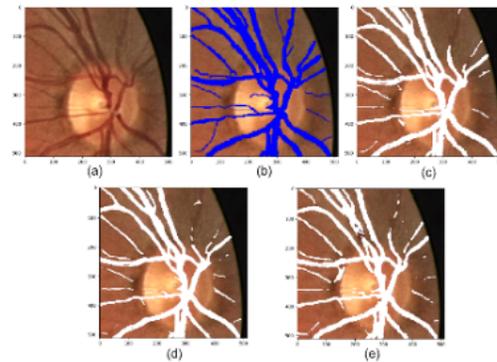

Figure 1. Comparison of predicted segmentation masks (white) with the ground truth mask (blue). (a) Input image to the model (preprocessed with CLAHE) (b) Ground truth mask (c) Segmentation mask predicted by model trained with Dice Loss (d) Segmentation mask predicted by model trained with Tversky Loss (e) Segmentation mask predicted by model trained with Focal Loss

5. Conclusion and Outlook

Our experiments show that U-Net can accurately segment blood vessels, even with small datasets. Our models perform comparably to other approaches and correctly identify major blood vessels in the papillary region while also segmenting smaller vessels reliably. While segmenting the entire retinal fundus image would yield better results, we are constrained by our larger multi-step framework for diagnosing glaucoma as part of the SALUS study.

In the future, we could extend our U-Net model to include data about arterial and venous blood vessels, which might be advantageous during the diagnosis process as

Chan et al. suggest [5]. The RITE dataset provides ground truth masks for the DRIVE dataset's images with these classes.

References

[1]   Weinreb RN, Aung T, Medeiros FA. The Pathophysiology and Treatment of Glaucoma: A Review. JAMA [Internet]. 2014 [cited 2022 Nov 30];311:1901. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2014.3192.

[2]   Oldiges K, Steinmann M, Duevel JA, et al. SALUS—a non-inferiority trial to compare self-tonometry in glaucoma patients with regular inpatient intraocular pressure controls: study design and set-up. Graefes Arch Clin Exp Ophthalmol [Internet]. 2022 [cited 2022 Nov 30];260:3945–3955. Available from: https://link.springer.com/10.1007/s00417-022-05759-7.

[3]   Valero B, Fénolland J-R, Rosenberg R, et al. Fiabilité et reproductibilité des mesures de la pression intraoculaire par le tonomètre Icare® Home (modèle TA022) et comparaison avec les mesures au tonomètre à aplanation de Goldmann chez des patients glaucomateux. Journal Français d'Ophtalmologie [Internet]. 2017 [cited 2022 Nov 30];40:865–875. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0181551217303984.

[4]   Gupta N, Weinreb RN. New definitions of glaucoma. Curr Opin Ophthalmol. 1997;8:38–41.

[5]   Chan KKW, Tang F, Tham CCY, et al. Retinal vasculature in glaucoma: a review. BMJ Open Ophthalmol [Internet]. 2017 [cited 2022 Dec 5];1:e000032. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5721639/.

[6]   Staal JJ, Abramoff MD, Niemeijer M, et al. Ridge based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging. 2004;23:501–509.

[7]   Hoover AD, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Transactions on Medical Imaging. 2000;19:203–210.

[8]   Buslaev A, Iglovikov VI, Khvedchenya E, et al. Albumentations: Fast and Flexible Image Augmentations. Information [Internet]. 2020;11. Available from: https://www.mdpi.com/2078-2489/11/2/125.

[9]   Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [Internet]. arXiv; 2015 [cited 2022 Jun 29]. Available from: http://arxiv.org/abs/1505.04597.

[10] Chen L-C, Zhu Y, Papandreou G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation [Internet]. arXiv; 2018 [cited 2022 Dec 6]. Available from: http://arxiv.org/abs/1802.02611.

[11] Iglovikov V, Shvets A. TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation [Internet]. arXiv; 2018 [cited 2022 Dec 7]. Available from: http://arxiv.org/abs/1801.05746.

[12] Iakubovskii P. Segmentation Models Pytorch [Internet]. GitHub repository. GitHub; 2019. Available from: https://github.com/qubvel/segmentation_models.pytorch.

[13] Liaw R, Liang E, Nishihara R, et al. Tune: A Research Platform for Distributed Model Selection and Training. arXiv preprint arXiv:180705118. 2018;

[14] Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks [Internet]. arXiv; 2017 [cited 2022 Dec 10]. Available from: http://arxiv.org/abs/1706.05721.

[15] Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data [Internet]. 2019 [cited 2022 Dec 11];6:60. Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0.