Student Conference Proceedings
Vol 1, No 1, 2025, Article ID 1952, 4 Pages

Student Conference Lübeck

Proceedings Article

Anomaly-guided image segmentation on retinal OCT images

Bennet Kahrs[a,c,*] · Julia Andresen[b] · Timo Kepp[c] · Heinz Handels[b,c,*]

[a] Student of Medical Informatics, Universität zu Lübeck, Luebeck, Germany
[b] Institute of Medical Informatics, Universität zu Lübeck, Luebeck, Germany
[c] German Research Center for Artificial Intelligence, Luebeck, Germany
[*] Corresponding author, email: bennet.kahrs@student.uni-luebeck.de; heinz.handels@uni-luebeck.de

Received 04 February 2025; Accepted 23 September 2025; Published online 29 September 2025

Abstract

Medical image segmentation can profit from the integration of contextual information provided by anomaly detection methods. This work investigates different strategies for using anomaly detection results as guidance in supervised segmentation of pathological retinal scans. Three state-of-the-art anomaly detection algorithms, differing in architecture, training methods, and original use cases, are evaluated in five guidance strategies, including an Attention-MobileNet architecture. The best-performing approach achieves an improvement of 6.7 % in the detection rate score compared to the baseline. This framework offers an enhanced tool for automated analysis of retinal scans, supporting earlier diagnosis and improved treatment planning.

I. Introduction

Optical Coherence Tomography (OCT) has become a standard imaging technique for diagnosing and monitoring retinal diseases. It is a non-invasive imaging technology that captures high-resolution images of biological tissues, such as the retina or skin [1]. For patients suffering from progressive retinal diseases, fluid accumulation within the retina is a critical marker that requires accurate volumetric segmentations for reliable analysis [2]. However, manual fluid delineation is time-consuming and prone to inter-rater variability. Therefore, there is a need for automated segmentation frameworks.

In recent years, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have revolutionized the field of medical image analysis. CNNs, due to their ability to learn features from raw image data, have shown significant success in tasks such as image classification, object detection, and segmentation. Among the various CNN architectures, lightweight models like MobileNet [3] have gained attention due to their efficiency in balancing performance and computational cost. In this work, we employ MobileNet as the baseline model for segmenting retinal fluids. However, as shown in the work of Seeböck et al. [4], contextual information, such as anomaly detection scores, can improve the performance of the segmentation model.

Therefore, this work evaluates several novel strategies to incorporate anomaly detection results into supervised segmentation models. We show that anomaly guidance can enhance the ability of segmentation models to detect subtle and complex fluid patterns, leading to more accurate and consistent segmentation results.

II. Materials and Methods

To investigate the role of anomaly detection in supervised segmentation, this section outlines the datasets used and the methodologies developed for pathological retinal OCT segmentation and anomaly guidance.

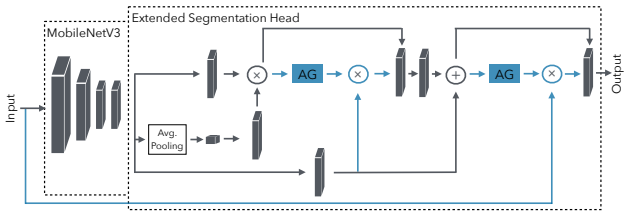

Figure 1: Structure of Attention-MobileNet based on MobileNet [3]. Gray coloured parts are from the original MobileNet, while blue highlighted parts refer to extended components from the Attention U-Net [5]. The AG blocks correspond to the Attention Gates of [5]. For better overview, some labels of the MobileNet are omitted.

II.I. Datasets

To encompass a wide range of biomarker representations, three datasets featuring different retinal diseases but similar fluid types are combined. These datasets contain multiple OCT volumes acquired with a Spectralis OCT scanner, as well as expert segmentations of the retina and pathological fluids. The resulting dataset includes data from a non-public age-related macular degeneration (AMD) study, a non-public longitudinal chronic serous chorioretinopathy (CSCR) study and the publicly available RETOUCH [2] challenge, comprising a total of 204 pathological OCT volumes.

In particular, there are 39 volumes with intraretinal fluid (IRF), 135 volumes with subretinal fluid (SRF), and 141 volumes with pigment epithelium detachment (PED). Since the number of B-scans is more crucial than the number of volumes, the methods presented in this paper operate on individual 2D slices (B-scans) rather than on entire 3D volumes. A total of 7023 B-scans are available, with 2098 pathological, 370 IRF, 1245 SRF, and 1073 PED B-scans, respectively. Due to the under-representation of the fluid classes, non-pathological B-scans are excluded from the model's statistics and evaluation.

The RETOUCH dataset is limited to the Spectralis OCT scans and has been extended with retinal labels. For the longitudinal CSCR dataset, different numbers of volumes are given for each patient. Therefore, a five-fold cross-validation is used in an approximately 80:20 split between the training and test folds. To ensure a balanced distribution of fluid types in training and test sets, the relative pixel counts of each class are taken into consideration. To standardize image sizes, larger images were cropped to a common resolution of 496×512 pixels. Preprocessing steps included flattening the images along the Bruch's membrane, followed by data augmentation through random scaling and translation.

II.II. Architectures

For fast and efficient predictions, MobileNetV3 [3] is used as a baseline model. The model is trained for 40 epochs using Adam optimization with an initial learning rate of 1e−4. To enhance generalization, exponential learning rate decay is applied every tenth epoch with a gamma of 0.9, and Stochastic Weight Averaging is employed with an additional learning rate of 1e−2. The baseline loss function combines cross-entropy and Dice Similarity Coefficient (DSC) loss across all five classes, including the background class and retina label. This serves as the foundational loss function for the fluid segmentation task and is subsequently referred to as $\mathscr{L}_{\text{fluid}}$.

To obtain additional contextual information, three different anomaly detection methods are investigated. These three algorithms are (1.) f-AnoGAN [6], an unsupervised generative adversarial network that learns the healthy data distribution and a fast mapping of new data to its latent space; (2.) Natural Synthetic Anomalies (NSA) [7] trained to detect anomalies in a self-supervised manner using synthetic anomalies and Poisson image blending to create anomalies in healthy images; and (3.) FluidRegNet (FRN) [8], a registration framework specially designed for retinal fluids and adopted for anomaly detection. By registering healthy images to pathological ones, areas of large deformations effectively highlight anomalies.

II.III. Guidance methodologies

Anomaly detection guidance can provide additional contextual information to support segmentation performance, as demonstrated by these five methodologies.

Indirect anomaly guidance The first methodology evaluates the approach of Seeböck et al. [4]. With the aim of improving the semantic context, anomaly maps are added to the existing manual annotated labels. This results in an additional segmentation class for anomaly pixels. In [4], they extend existing ground truth labels in a distinct manner without overlaps between fluid segmentation and anomaly segmentation. In contrast, this work uses the overlapping anomaly map as an extra target map, while the model predicts an additional channel; which is learned with a separate binary cross-entropy loss. Resulting in the loss function $\mathscr{L}_{\text{indirect}} = \mathscr{L}_{\text{fluid}} + \mathscr{L}_{\text{ano}}$ where $\mathscr{L}_{\text{fluid}}$ indicates the baseline loss function for the original five classes and $\mathscr{L}_{\text{ano}}$ the additional binary cross-entropy loss for the spatial context. We will refer to it as *Indirect Anomaly Guidance (IDAG)*.

Direct anomaly guidance To increase the influence of anomaly detection on the segmentation task, the anomaly map is given as a direct input to the MobileNet. We call this *Direct Anomaly Guidance* (DAG). For this purpose, the network is modified by adding an extra input channel, and the image and the anomaly map are concatenated in advance. Otherwise, this approach is trained as described in section II.II.

10.18416/SCP.2025.1952

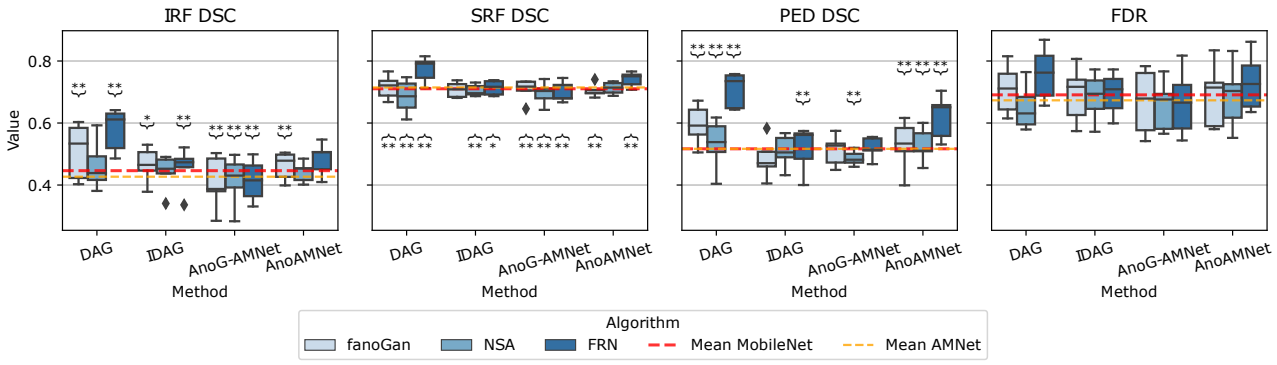

Figure 2: Influence of different anomaly detection methods on the segmentation results of MobileNet and AMNet. Significance to the MobileNet baseline is provided by stars ($* \, p < 0.05$, $** \, p < 0.01$) using Wilcoxon signed-rank test for DSC results only.

Attention-MobileNet This method introduces a novel way to focus attention on the MobileNet, called *Attention-MobileNet* (AMNet). As shown in Fig. 1 the MobileNet is modified by Attention Gates (AGs) from the Attention U-Net [5]. The AGs learn to predict soft attention maps that are multiplied with the feature maps provided by the skip connections to suppress unrelated information and focus on the more relevant information. In Fig. 1, the blocks and connections that differ from the original MobileNet are shown in blue. Two AGs are inserted, which work on different resolutions due to their position in that architecture – similar to the Attention U-Net. In addition to the AGs, a further skip connection is introduced that inserts the initial image at the end of the network just in front of the last layer block. This architecture is trained like the MobileNet and as described in section II.II using the foundation loss $\mathcal{L}_{\text{fluid}}$.

Anomaly-guided Attention-MobileNet *Anomaly-guided AMNet* (AnoG-AMNet) combines the IDAG approach with the application of AMNet. To support the attention mechanism and achieve a stronger influence of the anomaly detection guidance, anomaly maps are used to train the AGs. In this way, AMNet is used to learn the capability of anomaly detection. This results in the following loss function $\mathcal{L}_{\text{guided}} = \mathcal{L}_{\text{fluid}} + \mathcal{L}_{\text{AG1}} + \mathcal{L}_{\text{AG2}}$, where $\mathcal{L}_{\text{AG}_i}$ is the MSE loss of the AG and the scaled anomaly detection map.

Anomaly-Attention-MobileNet The last proposed methodology, Anomaly-Attention-MobileNet (AnoAMNet), extends the idea of contextual information in AGs a step further. It modifies the AMNet by replacing the AGs with direct input of anomaly maps. This means that the anomaly map is used in the forward path of the AMNet and scaled before being used as input on the corresponding AG position. Like the AMNet, DAG, and baseline approach, this network is trained using just the $\mathcal{L}_{\text{fluid}}$ loss function.

III. Results and discussion

The segmentation performances are evaluated by the DSC. For detailed comparison, the DSC is computed for each class considered separately, depending on the existence of the class in the B-scan. This results in three scores for the IRF DSC, SRF DSC, and PED DSC. As a further metric, the fluid detection rate (FDR) is used, which evaluates the detection capabilities at instance level. The detection rate is defined as the ratio of the number of fluids that are segmented/detected to the overall number of fluids. One fluid is thereby counted as such when all pixels are connected in an eight neighbourhood. Segmentation objects that are not connected are counted as two separate fluids. A fluid is counted as detected when at least one pixel of the label overlaps with the prediction.

Fig. 2 summarises the results. The baseline model and the AMNet serve as a reference and are highlighted in red and orange. The baseline achieves average scores of 44.6 (±5.1), 70.1 (±2.4), 51.7 (±2.4), and 69.1 (±5.9), for IRF DSC, SRF DSC, PED DSC, and FDR, respectively. AMNet obtained similar mean scores of 42.7 (±6), 71.5 (±1.4), and 51.7 (±1.4), for IRF DSC, SRF DSC, PED DSC, as well as a FDR of 67.3 (±5.9). However, AMNet was observed to reduce the number of false positives. The highest performance, with an FDR of 75.8 % (±7.5 %), was achieved by the DAG using the FRN algorithm. This approach improved segmentation across all fluid classes in DSC and FDR, outperforming all other methods. The improvement is qualitatively shown in Fig. 3, where the baseline is compared to the DAG methodology using FRN. The second-best performance was achieved by AnoAMNet, which improved the IRF DSC by 3.8 % and PED DSC by 10.3 % over the baseline. Both DAG and AnoAMNet demonstrated consistent improvements in all metrics. In general, the guidance approaches (IDAG, AnoG-AMNet), which are limited to the loss function, performed worse, failing to surpass the baseline.

Fig. 2 shows the performance of the methodologies according to the anomaly detection algorithm used. It is

10.18416/SCP.2025.1952

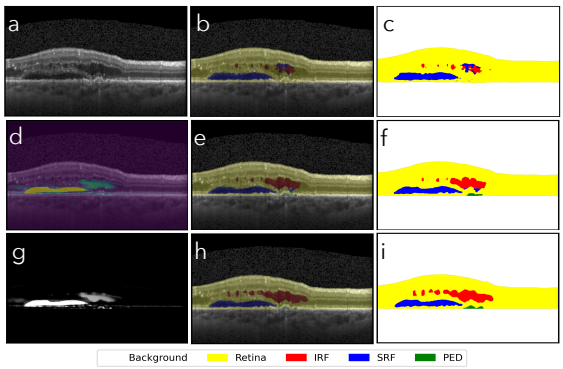

Figure 3: Comparison of the baseline prediction and the DAG prediction using FRN results for guidance. The first column (a, d, g) displays the input image alongside the FRN anomaly map (for guidance methodologies). The second and third columns present the labels and their overlays on the input image, respectively. The first row (b, c) shows the baseline predictions, the second row (e, f) shows the DAG predictions using FRN, and the third row (h, i) presents the ground truth.

quite noticeable that the performance of the loss-guided methodologies hardly deviates from the baseline. The results for NSA are close to each other and to the baseline, indicating a lower information gain. FRN achieves widely scattered and higher results, indicating higher information gain. The same is true for more direct methodologies, such as DAG and AnoAMNet, where the anomaly map is used as input for the segmentation architectures.

However, direct anomaly guidance approaches have the disadvantage of high overhead due to the need for anomaly detection maps for prediction. In contrast, indirect guidance approaches require anomaly detection maps only during training. This enables these methodologies to be used even without existing anomaly maps.

Typical issues evident in both qualitative and quantitative aspects include the misclassification of IRF, as shown in 3. All approaches struggle to accurately classify IRF regions. Despite the presence of fluid pathologies that are visually apparent to humans, these models are unable to detect them reliably. This limitation may be attributed to the small amount of training data available, particularly for the IRF class (39 volumes). However, this hypothesis requires further investigation in future studies to confirm and address the underlying causes.

IV. Conclusion

Anomaly detection plays a crucial role in medical image computing, particularly for tasks involving pathological image data. This work demonstrates that integrating anomaly detection guidance can significantly enhance supervised segmentation of medical images. By evaluating state-of-the-art anomaly detection algorithms and developing novel guidance methodologies, we show that using strong anomaly feature maps improves segmentation performance. Our findings suggest future research on contextual information for fluid relations and positions, highlighting the potential of anomaly detection not only as a diagnostic aid but also as a complementary tool to improve segmentation frameworks, paving the way for more accurate and efficient automated medical image analysis.

Acknowledgements

The work has been carried out at German Research Center for Artificial Intelligence, Luebeck, Germany and supervised by the Institute of Medical Informatics, Universität zu Lübeck. Research funding: The author states no funding involved.

Author's statement

Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee. AI tools were used for the linguistic fine-tuning.

References

[1] F. Hajizadeh, Ed., Atlas of Ocular Optical Coherence Tomography. Springer International Publishing, 2022, doi:10.1007/978-3-031-07410-3.

[2] H. Bogunović, F. Venhuizen, S. Klimscha, *et al.* RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge. *IEEE Transactions on Medical Imaging*, 38(8):1858–1874, 2019, doi:10.1109/TMI.2019.2901398.

[3] A. Howard, M. Sandler, G. Chu, *et al.*, Searching for MobileNetV3, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1314–1324, 2019. doi:10.1109/ICCV.2019.00140.

[4] P. Seeböck, J. I. Orlando, M. Michl, *et al.* Anomaly guided segmentation: Introducing semantic context for lesion segmentation in retinal OCT using weak context supervision from anomaly detection. *Medical Image Analysis*, 93, 2024, doi:10.1016/j.media.2024.103104.

[5] O. Oktay, J. Schlemper, L. L. Folgoc, *et al.*, Attention U-Net: Learning where to look for the pancreas, in *Medical Imaging with Deep Learning*, 2018.

[6] T. Schlegl, P. Seeböck, S. M. Waldstein, *et al.* f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54:30–44, 2019, doi:10.1016/j.media.2019.01.010.

[7] H. M. Schlüter, J. Tan, B. Hou, *et al.*, Natural synthetic anomalies for self-supervised anomaly detection and localization, in *European Conference on Computer Vision*, Springer, 474–489, 2022. doi:10.1007/978-3-031-19821-2_27.

[8] J. Andresen, J. Ehrhardt, C. von der Burchard, *et al.*, FluidRegNet: Longitudinal registration of retinal OCT images with new pathological fluids, in *Medical Imaging with Deep Learning*, 2024.