Segmentation-Based Attention Mask for Enhancing Fundus Image Diagnosing

Elena Himbitskaya
Belarusian State University
Belarusian State Medical University
Minsk, Belarus
fpm.gimbicka@bsu.by

Kseniya Svistunova
Belarusian State University
Minsk, Belarus
fpm.svistuno@bsu.by

Siarhei Kezik
Belarusian State University
Belarusian State Medical University
Minsk, Belarus
sergeyk8008@gmai.com

Abstract. **Approach to improve the classification of ocular diseases from fundus images by leveraging semantic segmentation as an attention mechanism. Key anatomical structures – optic disc, optic cup, and retinal vessels – are segmented using deep learning models and combined into weighted attention masks. These masks guide a classifier based on EfficientNetB6 to focus on clinically relevant regions, resulting in significant improvements in diagnostic accuracy. The method enhances detection sensitivity for subtle disease features and increases model interpretability.**

Keywords: **deep learning, medical image analysis, computer vision, semantic segmentation, attention mechanism, medical image classification, retina pathology detection, computer-aided diagnosis**

I. INTRODUCTION

Fundus imaging is essential for non-invasive, high-resolution visualization of the retina, aiding diagnosis of many ocular diseases. Interpretation requires expert knowledge and can be time-consuming, especially where ophthalmologists are scarce [1]. Deep learning, particularly convolutional neural networks (CNNs), have made advances in medical image classification, including retinal disease detection. However, typical end-to-end models use global image features, potentially diluting important localized clinical information. This limits early disease detection when signs are subtle [2].

To address this, attention mechanisms focusing on clinically relevant areas are increasingly used [3]. This study applies semantic segmentation as attention masks highlighting key anatomical zones (optic disc, cup, vessels) to enhance classifier focus and accuracy.

A. Symptoms of Ocular Diseases in Fundus Imaging

Fundus photography visualizes retinal structures critical in diagnosing:

- Myopia: Axial elongation causes retinal thinning, peripapillary atrophy, tilted discs.
- Hypertension: Microvascular changes like arteriolar narrowing, AV nicking, hemorrhages, cotton wool spots.
- Diabetic Retinopathy: Microaneurysms, hemorrhages, exudates, neovascularization, macular edema.
- Glaucoma: Increased cup-to-disc ratio, neuroretinal rim thinning, disc hemorrhages.
- Cataract: Lens opacity reduces image contrast, causing blur.
- Age-Related Macular Degeneration: Drusen, pigment changes, subretinal hemorrhages, fluid accumulation.

Accurate segmentation of these features enables precise diagnostic algorithms [4].

II. MULTICLASS SEMANTIC SEGMENTATION METHODS

A. Optic Disc and Optic Cup Segmentation

A modified U-Net in PyTorch segments optic disc and cup from RGB images (256×256). The network outputs dual-channel masks indicating disc and cup regions. The architecture uses symmetric encoder-decoder blocks with skip connections, two convolutional layers per block, and ReLU activations.

Training used Binary Cross-Entropy loss with early stopping over 30 epochs. Performance metrics included:

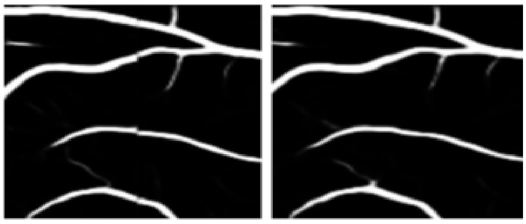
- Dice coefficient ~0.95 (disc), ~0.88 (cup).
- Binary IoU 0.92 (disc), 0.83 (cup).

The optic cup showed lower accuracy due to smaller size and class imbalance. Loss fluctuations indicate sensitivity to anatomical variability.

B. Blood Vessel Segmentation

The vessel segmentation model used a U-Net with ResNet18 backbone pretrained on ImageNet, trained in two stages: pretraining on public datasets (DRIVE [6], CHASE DB1 [7], HRF [8]), then fine-tuning on target images.

Images resized to 996×996, split into overlapping tiles (352×352) to reduce border artifacts, then merged with alpha blending for smooth transitions. Loss combined binary cross-entropy and Jaccard index [9], achieving high-quality vessel segmentation. This process is shown in Figure.



Left – result of segmentation model on neighboring tiles, right – tiles merged with alpha blending

III. EfficientNetB6 Classifier (Baseline)

We trained a classifier on a curated ODIR-5K dataset with seven classes:

- Pathological Myopia
- Hypertensive Retinopathy
- Diabetic Retinopathy
- Glaucoma
- Cataract
- Age-related Macular Degeneration (AMD)
- Normal (Healthy).

Images were resized to 224×224 and augmented with rotation (±30°), shifts (10 %), zoom (20 %), and flips.

Model used EfficientNetB6 backbone (no top), with global average pooling, dense layer (224 units, ReLU), dropout (0.3), and softmax output (7 classes) [10].

Training parameters:

- Loss: sparse categorical crossentropy
- Optimizer: Adam, lr=1e-4
- Epochs: 30
- Batch size: 8
- Early stopping with patience 5.

Results: Overall test accuracy 78.3 %; best accuracy for Myopia (94 %) and AMD (90 %); poorest for Diabetic Retinopathy (29 %) and Healthy (69 %).

IV. EfficientNetB6 Classifier Using Semantic Attention Mask

To enhance performance, we generated semantic attention masks per image using pretrained segmentation models for optic disc, cup, and vessels. Each mask was weighted based on clinical relevance:

- Disc: 0.9
- Cup: 1.0
- Vessels: 0.8
- Background: 0.2

A composite attention mask M is computed as a weighted sum of the individual segmentation maps:

$$M = w_{\text{disc}} \cdot \text{mask}_{\text{disc}} + w_{\text{cup}} \cdot \text{mask}_{\text{cup}} + w_{\text{vessels}} \times$$

$$\times \text{mask}_{\text{vessels}} + w_{\text{background}} \cdot (1 - \text{combined\_mask}),$$

where combined_mask denotes the union of all binary object masks. The attention-enhanced image I_{attn} is then obtained via element-wise multiplication:

$$I_{attn}(x, y) = I(x, y) \cdot M(x, y)$$

This operation suppresses less informative background regions while amplifying features in diagnostically critical areas.

The modified images I_{attn} are fed into the same EfficientNetB6 classifier described earlier. The training configuration – loss function, optimizer, learning rate, and number of epochs – remains unchanged to ensure a consistent comparison with the baseline model.

This strategy allows the network to concentrate on regions most likely to contain pathological changes, and leads to measurable improvements in classification accuracy, especially in complex or borderline cases.

Results: Overall accuracy improved to 92.2 %, with highest accuracies for Cataract (96.6 %) and Myopia (96.0 %), and the lowest still above 84 % for Diabetic Retinopathy and Healthy classes.

V. Conclusion

This study demonstrates that using attention masks generated through semantic segmentation significantly enhances the classification of retinal diseases from fundus images. By highlighting key anatomical structures – the optic disc, optic cup, and retinal vessels – these masks effectively guide the classifier to focus on clinically relevant regions, which leads to substantial improvements in diagnostic accuracy.

In addition to improving accuracy, the use of attention masks increases the interpretability of the model by revealing which image regions contribute most to its decisions. Such transparency is essential in medical settings, providing clinicians with justifiable diagnostic results. Moreover, the system's ability to adapt by adjusting mask weights based on classification errors enhances its robustness and effectiveness in real-world applications.

Acknowledgment

The authors would like to thank the group of Laboratory of Information and Computer Technologies of Belarusian State Medical University for their valuable comments and help, in particular, Grigory Karapetyan and Ivan Kosik.

REFERENCES

[1] I. B. Gurevich, V. V. Yashina, S. V. Ablameyko, A. M. Nedzved, A. M. Ospanov, A. T. Tleubaev, A. A. Fedorov, N. A. Fedoruk, "Development and experimental investigation of mathematical methods for automating the diagnostics and analysis of ophthalmological images", Pattern Recognit. Image Anal., no. 28(4), 2018, pp. 612–636.

[2] R. Sarki, et al., "Automatic detection of diabetic eye disease through deep learning using fundus images : a survey", IEEE Access, 2020, vol. 8, pp. 151133–151149. DOI: 10.1109/AC-CESS.2020.3015258.

[3] D. Ting, et al., "Artificial intelligence and deep learning in ophthalmology", Br. J. Ophthalmol, vol. 103, 2019, pp. 67–75. DOI: 10.1136/bjophthalmol-2018-313173.

[4] Ambika Selvakumar. Understanding Optic Disc Pallor - Shades of White [Online]. Available: https://www.eophtha.com/posts/understanding-optic-disc-pallor-shades-of-white.

[5] V. V. Starovoitov, Y. I. Golub, M. M. Lukashevich, "Digital fundus image quality assessment", System analysis and applied information science, 2021, vol. 4, pp. 25–38.

[6] Paperswithcode Available at: https://paperswithcode.com/datas et/ drive

[7] Paperswithcode. Available: https://paperswithcode.com/dataset/chase-db1.

[8] Paperswithcode. Available: https://paperswithcode.com/dataset/hrf.

[9] Adrian Rosebrock, Deep Learning for Computer Vision, PyImageSearch, 2017, 330 p.

[10] M. Tan, Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks", 36th International Conference on Machine Learning, ICML, June 2019.

UNITED INSTITUTE OF INFORMATICS PROBLEMS
OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

PRIP'2025

PATTERN RECOGNITION AND INFORMATION PROCESSING

EMBEDDING THE FUTURE

PROCEEDINGS
OF THE 17th INTERNATIONAL CONFERENCE

16–18 September 2025, Minsk, Belarus

Minsk
UIIP NASB
2025