Pusan National University

Computer Science and Engineering

Technical Report 2024-10

Development of an AI Model for Glaucoma Detection through Segmentation

저자1 : Aldo Sigfrido Calderoni Echeverri

저자2 : 배민준

저자3 : Mahboubeh Bagheri

지도교수 : 황원주

목 차

1. 서론

1.1. 연구 배경

Glaucoma is a chronic ophthalmology disease caused by damage to the optic nerve. This can lead to permanent blindness. As one of the primary causes of blindness in developed countries, it can result in irreversible vision impairment if not diagnosed and treated early. Patients may not notice symptoms until months or years after nerve damage has occurred, making timely detection crucial. According to published data, the number of glaucoma patients is estimated to reach 110 million by 2040 [1], underscoring the importance of prevention and early treatment.

To ensure timely treatment, early screening for glaucoma is essential. The three main techniques used for detection are intraocular pressure (IOP) assessment, visual field testing, and optic nerve head (ONH) evaluation [2]. Of these, ONH assessment, which involves analyzing the Cup-to-Disc Ratio (CDR) from fundus images, is particularly valuable. The CDR, calculated as the ratio of the vertical cup diameter to the vertical disc diameter, provides a precise indicator of glaucomatous damage, as glaucoma patients typically exhibit a significantly larger CDR compared to healthy individuals, as shown in Figure 1.

IOP measurement, while commonly used, has limitations as some patients with glaucoma may have normal IOP levels. Hence, CDR is a more accurate marker for detecting glaucoma progression.

Manual segmentation of the optic disc (OD) and optic cup (OC) for CDR calculation is time-consuming and requires subjective judgment, making it challenging for less experienced physicians to achieve accurate results. Automation of this process can help improve the accuracy, speed, and consistency of glaucoma diagnosis, enhancing patient care and enabling early intervention.

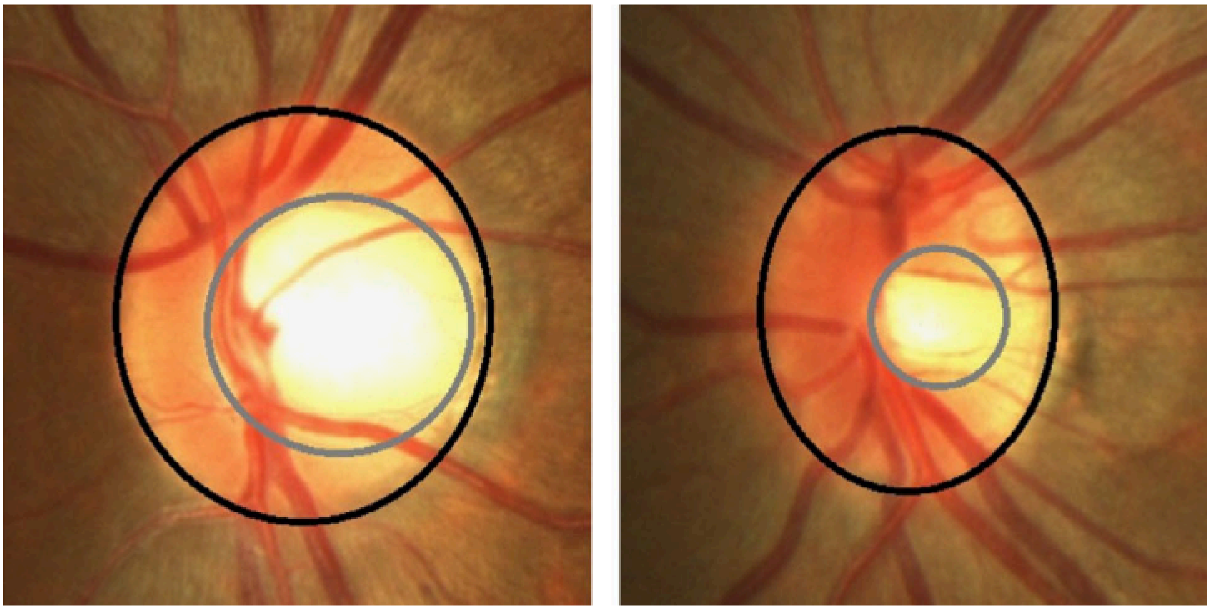


Figure 1. The black line indicates the optic disc area and the grey line indicates the optic cup area. [8]

1.2. 기존 문제점

Early diagnosis is crucial to prevent permanent optic nerve damage. Manual segmentation of the optic disc and cup is time-intensive, requiring skilled professionals and about eight minutes per eye. Automated segmentation can expedite mass screenings and offer vital support in regions with limited access to specialists. Moreover, the clinical significance of this project lies in its potential to transform glaucoma screening practices. Automated segmentation of the optic disc and cup enables rapid and accurate assessment of CDR, which is a critical parameter in glaucoma diagnosis. This approach can enhance early detection, allowing for timely interventions and potentially preventing vision loss in many patients [3].

Although currently existing models for glaucoma segmentation of the optic cup (OC) and optic disc (OD) show promising results, they are not yet practical for clinical use due to an error rate of around 5–15%. We aim to enhance these models by integrating novel techniques to improve segmentation accuracy and reduce this error margin, making the models more reliable for real-world applications.

1.3. 연구 목표

The primary objective of this research is to develop an effective and reliable system for detecting glaucoma through image segmentation techniques. This involves designing and implementing a machine learning-based segmentation model to accurately identify and segment key anatomical structures in retinal fundus images, particularly the optic disc and optic cup, which are essential for diagnosing glaucoma. The research aims to improve the precision of the segmentation model, enabling it to effectively distinguish between healthy and glaucomatous optic nerve structures, with a particular focus on early-stage glaucoma detection for timely intervention.

The model's performance will be assessed using standard evaluation metrics such as the Dice coefficient, Intersection over Union (IoU), and accuracy, ensuring it meets clinically relevant standards.

2. 연구 배경

In our project, we emphasize the ingenuity of our approach through the integration of several advanced techniques. Specifically, we employed CBAM (Convolutional Block Attention Module) to enhance the model's focus on critical regions, transformer blocks to capture long-range dependencies in the image, and DC-UNet to improve feature propagation and segmentation accuracy. These components differentiate our model from conventional U-Net architectures, offering a novel combination that has not been previously applied in glaucoma segmentation. This innovation underscores the uniqueness of our work and its potential to advance medical image analysis.

2.1. U-Net

U-Net[4] is a convolutional neural network architecture designed specifically for medical image segmentation tasks. It has become one of the most widely used models for biomedical image analysis due to its ability to produce highly accurate segmentation results even with limited training data. The U-Net architecture follows a symmetric encoder-decoder structure. The encoder, often referred to as the contracting path, is responsible for capturing context by gradually reducing the spatial dimensions of the input image while increasing the depth of the feature maps. This is achieved through successive layers of convolution, max pooling, and non-linear activation functions, which allow the network to learn increasingly abstract representations of the input data.

On the other side, the decoder, known as the expansive path, reconstructs the segmentation map by up-sampling the feature maps and combining them with high-resolution features from the encoder through skip connections. These skip connections play a critical role in U-Net's success, as they ensure that spatial information lost during down-sampling in the encoder is preserved and passed to the decoder, leading to more precise segmentation boundaries. Figure 2 shows the visualization of U-Net structured model.

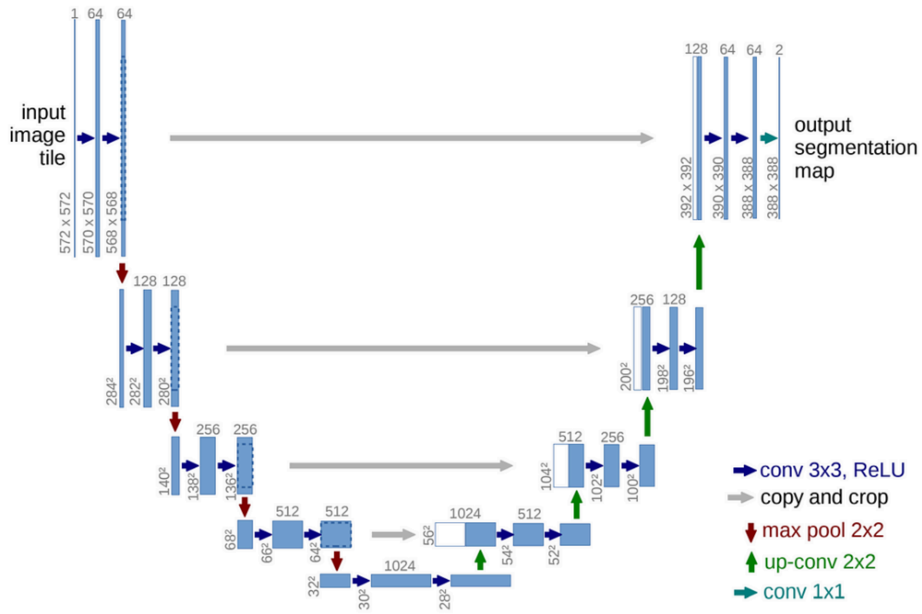


Figure 2. U-Net model illustration

2.2. CBAM

Convolutional Block Attention Module [5] is an efficient attention mechanism used in convolutional neural networks (CNNs) to enhance model performance by focusing on the most important features within the data. CBAM applies attention to both the channel and spatial dimensions of feature maps. It operates in two stages: first, the channel attention mechanism determines which feature map channels are most relevant to the task at hand, enhancing those that carry valuable information and suppressing less useful ones. Following this, the spatial attention mechanism identifies which specific regions within the feature maps are important, highlighting significant areas while downplaying irrelevant ones. By combining these two attention processes, CBAM allows the model to focus more effectively on informative features, both in terms of channels and spatial regions. This results in improved performance for tasks such as image classification and object detection. Additionally, CBAM is designed to be lightweight and can be easily integrated into existing CNN architectures, providing a boost in accuracy without adding significant computational complexity. Figure 3 illustrates main two attention modules in CBAM.

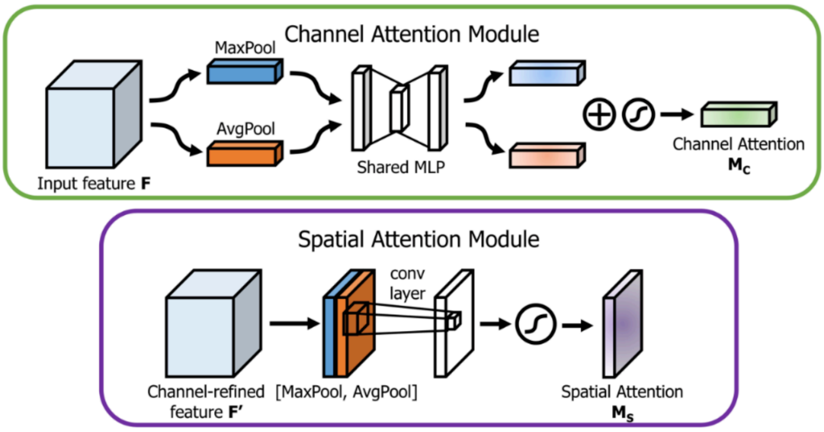


Figure 3. CBAM illustration of "CBAM: Convolutional Block Attention Module"

2.3. Transformer

Transformer is a deep learning architecture introduced in 2017 in the paper "Attention is All You Need," [6] which transformed natural language processing (NLP). Unlike traditional recurrent neural networks (RNNs), Transformers use a self-attention mechanism that processes all input tokens simultaneously, enabling greater parallelization. The architecture consists of an encoder and a decoder, each made up of multiple identical layers. The encoder generates continuous representations of input sequences using multi-head self-attention, capturing relationships between words regardless of their distance. The decoder produces output sequences while maintaining autoregressive properties through masked self-attention. Transformers excel at capturing long-range dependencies and have achieved state-of-the-art results in various NLP tasks, such as translation and summarization. They have also inspired numerous adaptations, leading to powerful models like BERT and GPT. This versatility has established the Transformer as a foundational component in modern AI research and applications. Figure 4 illustrates the model architecture of the transformer.

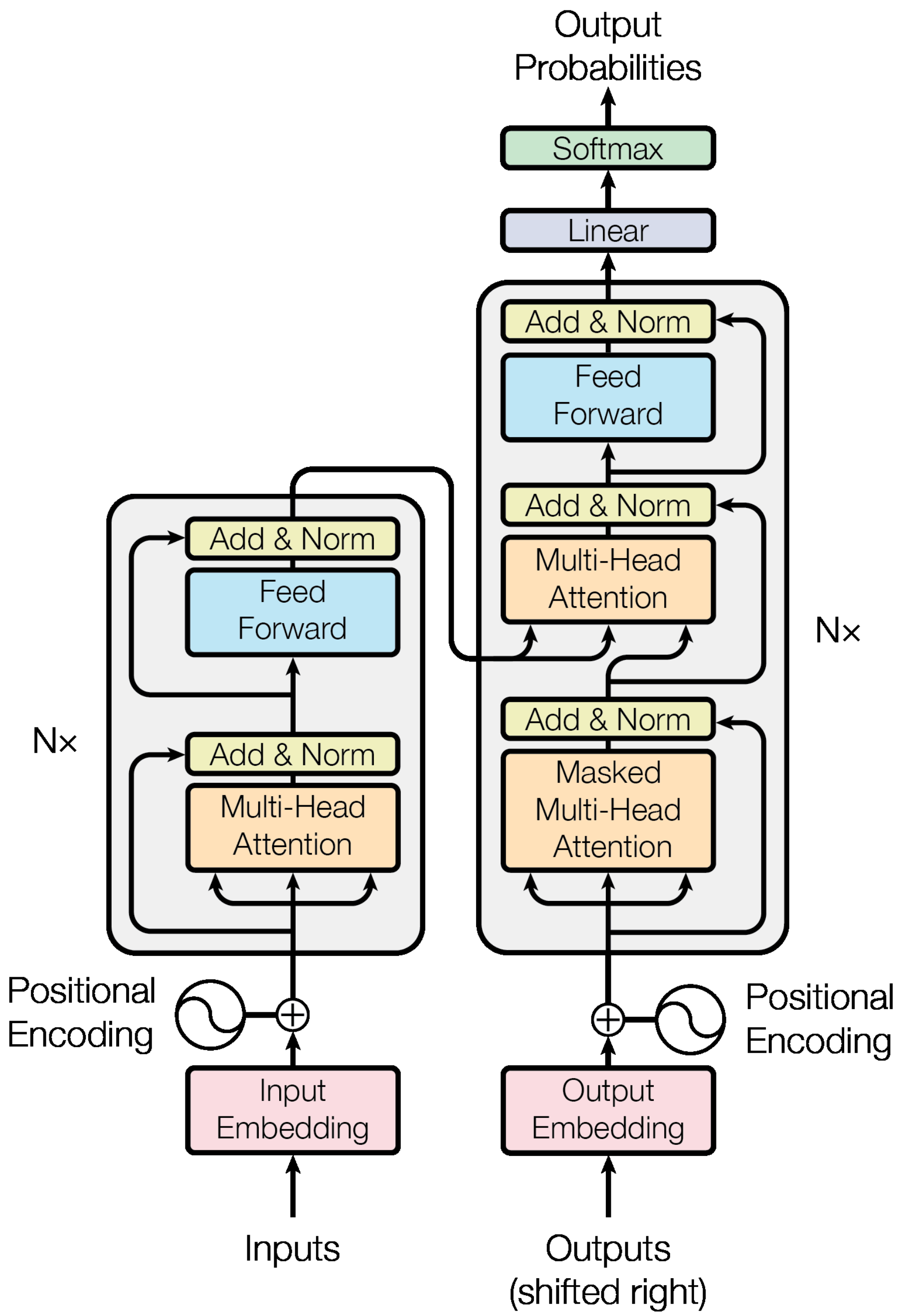


Figure 4. An illustration of Transformer from "Attention Is All You Need" [6]

2.4. DC U-Net

Dense Convolutional U-Net, is an advanced variant of the traditional U-Net architecture, specifically designed for medical image segmentation tasks. Dense Convolutional indicates the integration of dense connections into the U-Net framework. Dense connections, derived from DenseNet architecture [7], ensure that each layer is connected to every subsequent layer in a feed-forward manner, promoting feature reuse and efficient learning of complex representations. This allows the network to capture both low-level and high-level features more effectively, which is crucial for detailed segmentation in medical images.

Like the standard U-Net, DC U-Net follows an encoder-decoder structure. The encoder compresses the input image into smaller, feature-rich representations, while the decoder reconstructs the segmentation map. To ensure spatial information is preserved during the reconstruction, skip connections are included between corresponding layers in the encoder and decoder. The inclusion of dense connections in this architecture facilitates better feature propagation, leading to improved accuracy, especially in cases where subtle details are critical.

DC U-Net is particularly effective in medical imaging tasks such as organ or lesion segmentation, where accurate identification of regions is vital. Its dense connectivity and efficient feature reuse make it well-suited for handling limited datasets, a common challenge in the medical field. By combining the strengths of dense connections and U-Net's encoder-decoder design, DC U-Net provides enhanced performance for detailed segmentation tasks like glaucoma detection and tumor segmentation.

2.5. Incorporating Transformers, DC Block, and CBAM into the U-Net model

In order to take advantage of the previous tools and technologies, we decided to implement a new model that used transformers, a DC block, and the CBAM attention mechanism in the U-Net model. This decision was made after consulting in several papers what the best methods for medical segmentation were, a research that led us to the new hypothesis that their combination could bring a better solution for this task. Here we attach an image that displays a basic architecture of our model:

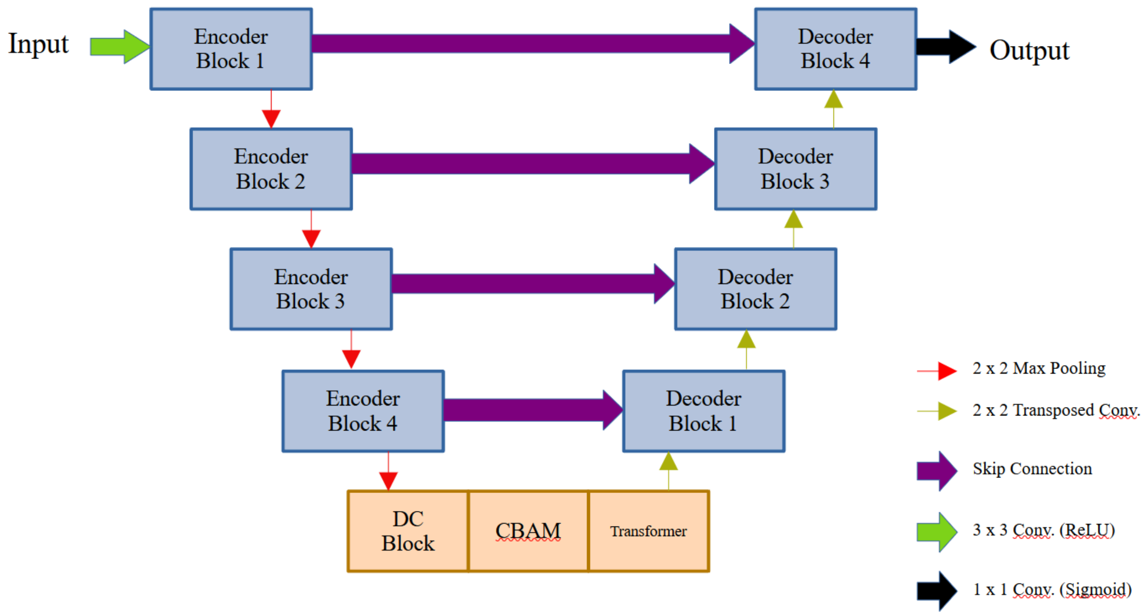


Figure 5. Diagram that illustrates an example of the application of CBAM and a hybrid of the DC-UNet with a Transformer-based mode

Here, in Figure 5 we attach the illustration of our model. When reading about the research, we hope that this illustration will aid your understanding of our work.

3. 연구 내용

3.1 Data Acquisition

To develop and validate our model for glaucoma segmentation, we gathered data from three publicly accessible ophthalmic datasets: REFUGE, ORIGA, and G1020. These datasets provide retinal fundus images with corresponding segmentation masks for optic disc (OD) and optic cup (OC) regions, crucial for assessing glaucoma.

- REFUGE: This dataset is designed specifically for glaucoma assessment, offering a diverse set of retinal images and accurate segmentation masks for OD and OC.
- ORIGA: It features a wide array of images with various stages of glaucomatous damage, offering a challenging dataset for model training and evaluation.

- G1020: This dataset contains high-resolution images that contribute to the precision of segmentation, aiding in the distinction between glaucomatous and non-glaucomatous conditions.

Images from these datasets were combined into a single dataset to ensure sufficient data diversity and volume, given the limited data available for each individual dataset. This approach also helped mitigate the class imbalance present in each dataset.

3.2 Image Preprocessing

To prepare the images for model training, we conducted several preprocessing steps that could ensure consistency and enhance our model performance:

1. Color Conversion: All images were converted from BGR to RGB format, which is essential for standardization as TensorFlow models typically operate on RGB data.
2. Resizing: Images were resized to 64 x 64 pixels, balancing computational efficiency with retention of critical anatomical details for accurate segmentation.
3. Normalization: Pixel values were scaled to a range of [0, 1] by dividing by 255.0 in order to improve model convergence and to enhance gradient stability during training.

3.3 Data Aggregation and Preparation

Following preprocessing, we aggregated the images and masks into a unified dataset and performed further data preparation steps to make the data compatible with our model:

1. Label Encoding and Reshaping: The segmentation masks were label-encoded and reshaped, which enabled compatibility with TensorFlow's categorical loss functions.
2. One-Hot Encoding: The masks were converted to one-hot encoded format with three channels representing the background, OD, and OC, respectively, due to the fact that this format is essential for multi-class segmentation.

3. Data Splitting: Finally, the dataset was divided into training and testing sets in an 80-20 ratio. This allowed for the evaluation of the model's generalization capabilities.

3.4 Model Architecture

Our model integrates several advanced components–It combines the TransUNet structure with DC blocks and CBAM modules so as to leverage both local and global feature representations.

1. Base Architecture (TransUNet V2):
    - Encoder-Decoder Structure: TransUNet's encoder-decoder structure provides an efficient means of downsampling and upsampling that allows for detailed segmentation maps.
2. Encoder Enhancements:
    - DC Blocks: These Dilated Convolutional Blocks expand the receptive field, which enable the network to capture fine-grained details while maintaining computational efficiency. By aggregating multi-scale features, these blocks improve the segmentation precision of glaucoma-related regions.
    - CBAM (Convolutional Block Attention Module): This module applies attention mechanisms to both channel and spatial dimensions; this is a tool that helps our model focus the network on critical image regions.
3. Transformer Integration:
    - Patch Embedding: The encoder outputs are transformed into smaller patches, which are embedded into a lower-dimensional space. This process prepares the data for transformer processing.
    - Transformer Blocks: Multi-head attention within the transformer captures dependencies across different image regions, which enables a comprehensive understanding of spatial relationships that are crucial for segmenting OD and OC areas.
4. Decoder Enhancements:
    - Skip Connections and Up-sampling: Transposed convolutions progressively reconstruct the segmentation map, and skip connections from the encoder

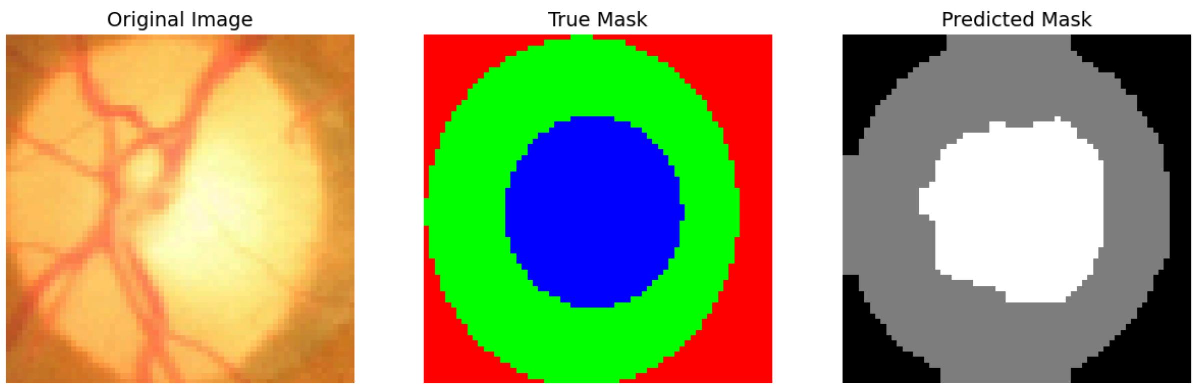
ensure that high-resolution features are preserved.

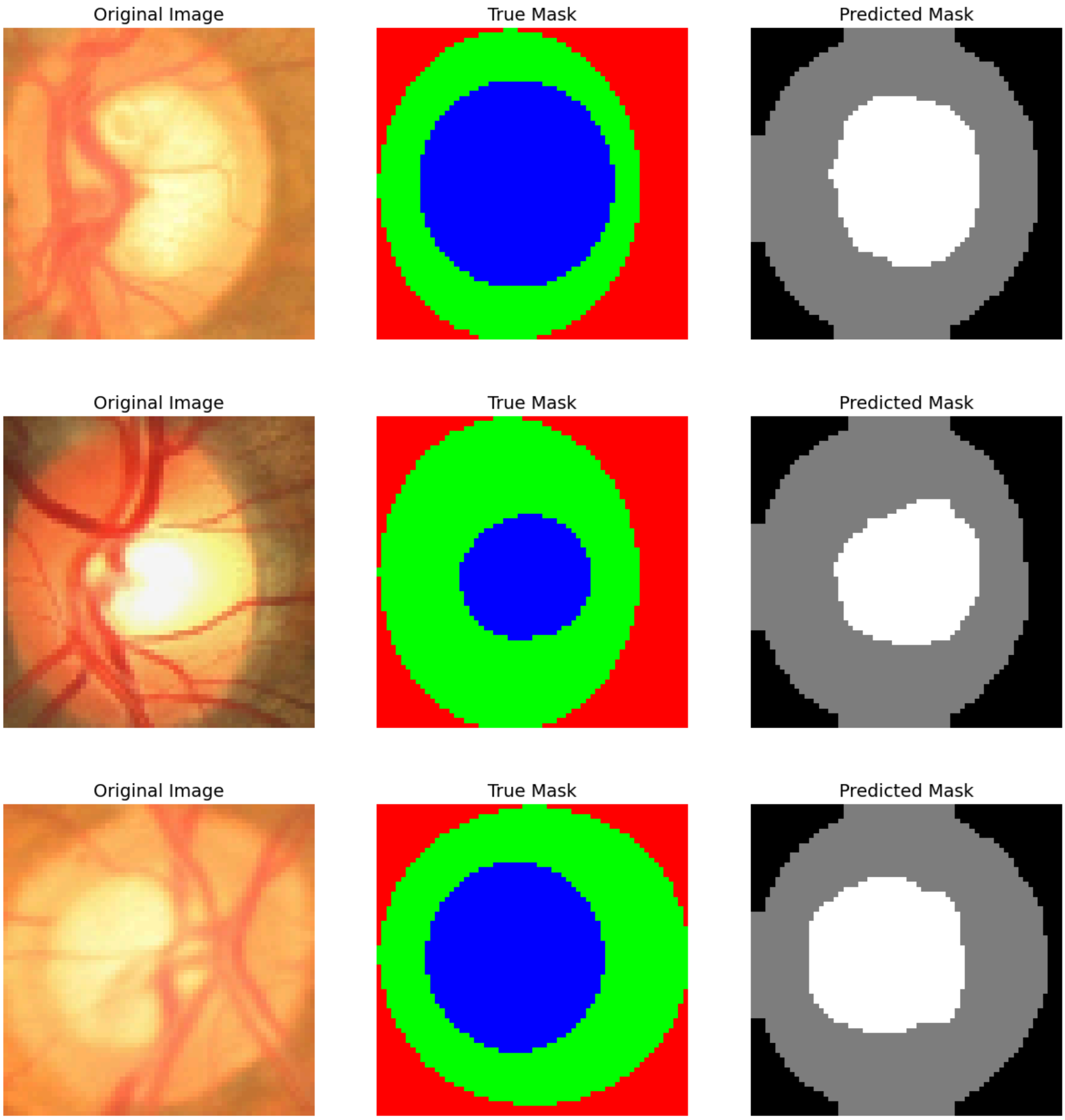
3.5 Loss Function

This model employs the categorical cross-entropy loss function to evaluate and optimize segmentation accuracy across three classes: background, OD, and OC. This loss function calculates the divergence between predicted and actual class probabilities for each pixel, which improves the model's ability to accurately differentiate between the three regions. Categorical cross-entropy is particularly effective for multi-class segmentation tasks, and this quality ensures precise gradient updates based on the difference between predictions and ground truth. This choice of loss function complements the model's architecture by promoting accurate pixel-wise classification.

3.6 Visualization

To evaluate our model's performance, we generated visualizations of segmentation predictions compared to ground truth masks. These visualizations were created using matplotlib and were instrumental in refining the model by providing qualitative assessments as well as quantitative metrics. Here we attach the following images as an example:





4. 연구 결과 분석 및 평가

In this section, we provide an analysis of the results obtained from various experiments using our proposed model and compare its performance against existing models like U-Net. To evaluate the performance of the models, we used three key metrics: accuracy, Intersection over Union (IoU), and the Dice coefficient.

Accuracy measures the overall correctness of the model, reflecting the percentage of correctly classified pixels in the image. However, since it may not fully capture the model's performance in the context of image segmentation, we also consider IoU and the Dice coefficient, which provide more specialized insights into segmentation quality.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

IoU quantifies the overlap between the predicted segmentation and the ground truth, measuring the ratio of the intersection of the two sets to their union. A higher IoU indicates a more accurate segmentation by the model, with values ranging from 0 (no overlap) to 1 (perfect overlap).

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

The Dice coefficient, on the other hand, is similar to IoU but gives slightly more weight to correctly classified pixels, making it sensitive to small segmentation errors. It is defined as twice the area of overlap divided by the total number of pixels in both the predicted and ground truth masks. Like IoU, Dice values range from 0 to 1, with higher values indicating better segmentation performance.

$$Dice = \frac{2|A \cap B|}{|A| + B\vee}$$

By using these metrics, we can comprehensively assess the segmentation capability of our proposed model compared to U-Net, highlighting improvements or shortcomings in accuracy and overlap with ground truth segmentations.

We experimented with several models, including U-Net, U-Net with CBAM, ResUNet, DC-UNet with CBAM, SegNet with CBAM, and TransDC-UNet with CBAM. The following table summarizes the performance of these models across different evaluation metrics, including Pixel Accuracy, Intersection over Union (IoU), and Dice Coefficient:

4.1. Comparison of experimental results

Model	Pixel-Accuracy	IoU	Dice Coefficient
U-Net	0.9001	-	-
U-Net with CBAM	0.9064	0.8289	0.8812
ResUNet (Used ResNet34)	0.9001	0.7928	0.8722
DC-UNet with CBAM	0.8756	0.7788	0.8467
SegNet with CBAM	0.9043	-	-
TransDC-UNet with CBAM	0.8498	0.7403	0.7765

From the table, we observe that U-Net with CBAM achieved the best results in terms of both IoU (0.8289) and Dice Coefficient (0.8812), indicating improved segmentation quality when attention mechanisms are employed. However, while some models demonstrated promising results, the performance of our proposed TransDC-UNet with CBAM did not surpass the baseline U-Net. In fact, the accuracy metrics show a slight decrease compared to U-Net and U-Net with CBAM.

The Pixel Accuracy values across all models remain relatively consistent, with minor variations. However, the Intersection over Union (IoU) and Dice Coefficient provide a more nuanced view, reflecting the differences in the model's ability to segment key anatomical structures, such as the optic disc and optic cup, with precision.

Although the performance of our model was lower than expected in terms of accuracy, we emphasize that the goal of this project was not merely to surpass U-Net in accuracy but to introduce ingenuity through the integration of state-of-the-art techniques such as CBAM, transformer blocks, and DC-UNet. The novelty of our approach lies in combining these elements, which, despite not yielding higher accuracy in this iteration, opens up new avenues for future refinement and optimization.

Future work will focus on improving the generalization capabilities of our model, exploring different combinations of hyperparameters, and further enhancing the model's structure to better capture the complexities of medical images.

5. 결론 및 향후 연구 방향

In this research, we aimed to enhance the segmentation model for glaucoma detection by integrating innovative techniques, including CBAM, transformer blocks, and DC-UNet. Although our model did not demonstrate an improvement in accuracy compared to the baseline U-Net model, the core achievement lies in the ingenuity of our approach. We introduced enhancements to the traditional U-Net model that incorporated modern techniques, which hold potential for further exploration.

As a summary of our model, the integration of CBAM allowed the model to focus on the most informative regions of the image, enhancing the attention mechanism both spatially and channel-wise. Similarly, the inclusion of transformer blocks facilitated the capture of long-range dependencies within the data, which is crucial in medical image analysis. Additionally, the DC-UNet structure provided the benefits of dense connections, enabling feature reuse and more effective learning, especially with limited datasets.

However, we provide the following suggestions for future work:

1. To obtain more data. 2,000 images were not enough, but this can be improved with a larger dataset that future medical research could bring about.
2. To consider more ways for preprocessing. Considering that nowadays there is not enough data to train in the medical field, data augmentation is not optional. However, alternative preprocessing methods could be utilized in the future as research in this field grows and improves.

Despite the lack of accuracy improvement, this study has demonstrated the potential of combining these advanced techniques in a novel way. The contribution of this work lies in the innovative architecture designed to push the boundaries of traditional models for medical image segmentation. Future work can build on this foundation to optimize the model further and explore its applications in different medical domains.

6. 참고 문헌

[1] Pachade, S., Porwal, P., Kokare, M., Giancardo, L., Mériaudeau, F.: NENet: nested efficientnet and adversarial learning for joint optic disc and cup segmentation. Med. Image Anal. 74, 102253 (2021)

[2] Jiang, Y., Duan, L., Cheng, J., Gu, Z., Xia, H., Fu, H., Li, C., Liu, J.: JointRCNN: a region-based convolutional neural network for optic disc and cup segmentation. IEEE Trans. Biomed. Eng. 67(2), 335–343 (2020)

[3] Jun Cheng, Jiang Liu, Yanwu Xu, Fengshou Yin, Damon Wing Kee Wong, Ngan-Meng Tan, Dacheng Tao, Ching-Yu Cheng, Tin Aung, Tien Yin Wong, "Superpixel Classification Based Optic Disc and Optic Cup Segmentation for Glaucoma Screening", 18 Feb 2013

[4] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", 18 May 2015

[5] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon, "CBAM: Convolutional Block Attention Module", 17 Jul 2018

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, "Attention Is All You Need", 12 Jun 2017

[7] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, "Densely Connected Convolutional Networks", 25 Aug 2016

[8] Xiaoyue Ma, Guiqun Cao, Yuanyuan Chen, "A review of optic disc and optic cup segmentation based on fundus images", 1 Nov 2023