EPA (Elegant Persian Algorithms)

Development of an AI Model for Glaucoma Detection through Segmentation

Computer science and Engineering Department

2024-2025

Wonju Hwang Professor

202155546 Aldo Sigfrido Calderoni Echeverri
201924487 Minjun Bae
202155549 Mahboubeh Bagheri

Index

1. Introduction

1.1. Necessity

Glaucoma is a critical ophthalmic condition that, if not diagnosed and treated early, can lead to irreversible blindness. It is one of the primary causes of blindness in developed countries. Accurate segmentation of structures related to glaucoma, such as the optic disc and optic cup, is essential for early detection and monitoring. [1] Traditional methods of detecting glaucoma involve measuring intraocular pressure (IOP); however, the Cup-to-Disc Ratio (CDR) has proven to be a more precise indicator. Automating the segmentation of these structures can significantly enhance the accuracy and efficiency of glaucoma diagnosis. [2]

1.2. Motivation

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][4]

2. Background

2.1. Preprocessing

Data preprocessing in computer vision refers to the transformation and preparation of raw image data into a suitable format for deep learning models. This involves a series of steps designed to enhance the quality and consistency of the data, ensuring that machine learning models can interpret the images effectively. Common preprocessing steps include resizing and scaling, data augmentation, noise reduction and filtering, standardization, etc. Data preprocessing is relevant to computer vision models because real-world image data often contains variability, noise, and inconsistencies that can degrade the performance of deep learning models. However, preprocessing methods ensure improved accuracy, generalization, and efficiency. [5]

2.2. Segmentation

Segmentation in the field of deep learning refers to the process of classifying each pixel in an image into a category, making it a crucial task in computer vision. In medical imaging, for example, segmentation helps in isolating structures such as the optic disc and the optic cup.

Deep learning-based segmentation often relies on convolutional neural networks (CNNs), which excel at recognizing patterns and structures within images. Architectures like U-Net and its variants are particularly popular for segmentation tasks because of their ability to capture both local and global context through their symmetric, encoder-decoder structure. These networks have dramatically improved the accuracy and efficiency of segmentation tasks compared to traditional methods. [6]

2.3. Visualization Techniques

Visualization involves displaying the retinal images alongside their corresponding segmentation masks. This step helps in verifying the quality and consistency of the annotations. Tools such as matplotlib in Python can be used for this purpose. By examining the images, we can identify common patterns and variations in the appearance of the optic disc and cup, which are critical for effective model training. [7]

3. Related Research

3.1. Segmentation and Classification of Glaucoma Using U-Net

- U-Net for Optic Cup Segmentation: U-Net effectively segments the optic cup from retinal fundus images.
- DenseNet-201 for Feature Extraction: A pre-trained DenseNet- 201 model is used to extract relevant features from retinal images.

- DCNN for Classification: A deep convolutional neural network (DCNN) classifies glaucoma based on segmented OD and OC features.
- Importance of Segmentation: Accurate OD and OC segmentation is crucial for calculating the cup-to-disc ratio (CDR), an important glaucoma indicator.
- Performance: The model, combining U-Net and DCNN, outperforms other deep learning models like VGG-19 and ResNet-152v2 in accuracy and precision. [8]

3.2. Optic Disc and Optic Cup Segmentation for Glaucoma Detection from Blurred Retinal Images Using Improved Mask R-CNN

- Improved Mask R-CNN: Utilizes an enhanced Mask R-CNN framework with a DenseNet-77 backbone for segmenting OD and OC from blurred retinal images.
- Data Preprocessing and Augmentation: Artificial blurriness is added to training samples to improve robustness against real- world blur.
- DenseNet-77 Backbone: Effective for feature extraction, aiding in the accurate localization and segmentation of OD and OC regions.
- Evaluation and Results: Tested on the ORIGA dataset, achieving high precision (0.965), recall (0.963), F-measure (0.97), and IoU (0.972), even under challenging conditions.
- Conclusion: The method offers robust performance for early glaucoma detection, even with blurred images, making it suitable for clinical use. [9]

3.3. ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data

- Architecture: ResUNet-a combines ResNet and UNet, integrating atrous convolutions and attention mechanisms to improve segmentation accuracy in complex image data.
- Performance: The model is particularly effective in remote sensing tasks, outperforming state-of-the-art models in semantic segmentation.
- Innovations: The paper introduces novel loss functions to accelerate training and enhance model performance.

- Applications: Tailored data augmentation techniques are discussed to boost the model's robustness, especially for large-scale image segmentation tasks. [10]

3.4. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

- Architecture: SegNet is an encoder-decoder network designed for semantic pixel-wise segmentation. The architecture consists of an encoder network, a decoder network, and a pixel-wise classification layer.
- Encoder Network:
  - ➢ Structure: Comprises a series of convolutional layers followed by max-pooling layers.
  - ➢ Purpose: Extracts features from the input image and reduces its spatial resolution while increasing the depth of the feature maps.
- Decoder Network:
  - ➢ Structure: Mirrors the encoder network with a series of upsampling layers followed by convolutional layers.
  - ➢ Purpose: Reconstructs the spatial dimensions of the image and refines the feature maps to produce the segmented output.
- Pooling Indices:
  - ➢ Feature Preservation: During max-pooling in the encoder, the indices of the maximum values are stored. These indices are used in the decoder to ensure precise upsampling, improving the spatial accuracy of the segmentation.
- Pixel-wise Classification:
  - ➢ Output: After decoding, the final layer is a softmax layer that assigns a class to each pixel based on the refined feature maps.
- Applications: SegNet is widely used in autonomous driving, medical image analysis, and any field requiring accurate image segmentation. [11]

3.5. Building the Initial U-Net Model

The U-Net architecture is selected for its proficiency in biomedical image segmentation, even with limited data. The initial model comprises a contracting path that captures the context and a symmetric expanding path that allows precise localization. This structure is particularly effective for segmenting the optic disc and cup in retinal images.

Initial Model Configuration:

- Input Size: (128, 128, 3)
- Convolutional Filters: 64 filters in the initial layer, doubling after each pooling operation
- Pooling Size: 2x2
- Activation Function: ReLU
- Loss Function: Dice loss, which directly optimizes the overlap between predicted and actual segmentation masks. [12]

4. Proposed Methodology

For this project, we selected the U-Net architecture as our primary algorithm for several reasons:

- Proven Effectiveness: The U-Net architecture has consistently demonstrated high proficiency in biomedical image segmentation tasks, making it an ideal choice for segmenting the optic disc and cup in retinal images. [13]
- Extensive Validation: Numerous studies and articles, including those cited in this report, have validated the U-Net model's performance in similar applications, highlighting its robustness and accuracy. [12][13]
- Symmetric Design: The U-Net's unique structure, comprising a contracting path that captures context and an expanding path that ensures precise localization, is particularly effective for the detailed segmentation required in glaucoma detection.

The U-Net's balance between capturing broad contextual information and maintaining high spatial accuracy aligns well with our project's requirements. Therefore, it was chosen over other architectures like CNN and AFNN, which, although potentially effective, do not have the same level of validation in the specific context of retinal image segmentation. [13][14]

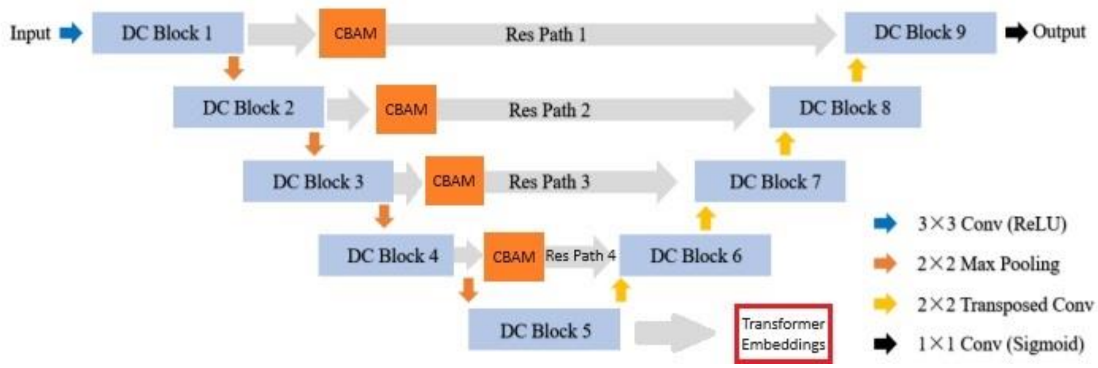
Enhancement Strategy: Hybrid Model Approach

While the U-Net architecture is robust, incorporating advancements from other models can further improve its performance. We propose a three-fold enhancement strategy:

- Integration with Attention Mechanisms: Recent research has demonstrated that attention mechanisms can significantly improve segmentation accuracy by allowing the model to focus on the most relevant parts of an image. Integrating attention modules into the U-Net architecture—such as the Convolutional Block Attention Module (CBAM)—can refine the focus on critical features of the optic disc and cup, enhancing the segmentation quality. CBAM, for example, applies attention mechanisms in both the spatial and channel dimensions of feature maps. The channel attention focuses on ‘what’ is important in the feature maps by emphasizing more informative channels, whereas the spatial attention focuses on ‘where’ is important by highlighting the most relevant spatial locations in the feature map (e.g., regions containing the optic disc and cup). [15]
- Hybrid Architecture with Other Deep Learning Models: Combining the U-Net with another model can significantly enhance its performance in medical image segmentation by leveraging the strengths of multiple architectures. U-Net excels in precise localization and boundary detection, but it may miss out on global context or diverse feature representations that other models, like ResNet, can provide. By integrating these models, a hybrid system can capture a broader range of patterns, enhance critical feature identification, and reduce the impact of noise, leading to more accurate and robust segmentation. [16]
- Hybrid Architecture with Transformer-Based Models: Transformer models, known for their capacity to capture global dependencies, can complement the U-Net's local feature extraction capabilities. By combining the U-Net with a Transformer encoder, we can leverage the global context provided by Transformers to improve segmentation accuracy. This hybrid approach, known as U-Net with Transformer, can be beneficial for complex cases

where global context enhances feature extraction and improves segmentation performance. Transformers can be used in conjunction with U-Net's decoder to integrate global contextual information into the segmentation process. [16]

Here is a diagram that illustrates an example of the application of CBAM and a hybrid of the DC-UNet with a Transformer-based model:



[17]

5. Experimental Results

5.1. Data Collection

- ORIGA (Online Retinal Image Glaucoma Analysis) Dataset: This is a comprehensive collection of retinal images specifically designed for glaucoma analysis. It consists of 650 color fundus images, of which 168 are labeled as glaucomatous and 482 as non-glaucomatous. The dataset is meticulously annotated, providing pixel-level ground truth for the optic disc (OD) and optic cup (OC) segmentation, which is crucial for calculating the cup-to-disc ratio (CDR), a key indicator of glaucoma. ORIGA is widely used in the development and benchmarking of automated glaucoma detection algorithms, particularly those focusing on segmentation tasks, making it a valuable resource for both research and clinical applications. [18]
- REFUGE (Retinal Fundus Glaucoma Challenge) Dataset: It was created as part of the Retinal Fundus Glaucoma Challenge, a global competition aimed at advancing glaucoma detection and optic nerve head segmentation techniques. It comprises 1,200 retinal fundus images, divided into training,

validation, and testing sets, with 400 images each. The dataset includes detailed annotations for the optic disc and optic cup, enabling the calculation of CDR. Additionally, the dataset is enriched with clinical information, such as the presence of glaucoma, allowing for the development of both segmentation and classification models. REFUGE is widely recognized in the research community for fostering innovation in glaucoma screening methodologies. [19]

- G1020 Dataset: The G1020 dataset is a large-scale retinal image dataset curated to support research in glaucoma detection through optic disc and cup segmentation. It contains 1,020 high-resolution color fundus images, with detailed annotations for both OD and OC boundaries. The dataset is unique in its diversity, including images from multiple sources and representing various levels of glaucoma severity, making it particularly useful for developing robust segmentation models that generalize well across different populations. The G1020 dataset is instrumental in training and evaluating deep learning models aimed at improving the accuracy of glaucoma detection, particularly in calculating the cup-to-disc ratio. [20]

Even though we have selected these datasets for our project and they have been enough until now, we remain open to exploring various datasets that can enhance our model's accuracy and robustness. Publicly available datasets such as Drishti-GS, RIM-ONE, IDRiD, RIGA, and others could be integrated to increase the volume and variability of training data. [21]

5.2. Preprocessing

During visualization, preprocessing needs such as resizing, normalization, and augmentation are identified. Preprocessing ensures that the images are uniform in size and intensity, which is important for the neural network to learn effectively. Common preprocessing steps include:

- Resizing: Adjusting the image dimensions to a fixed size (e.g., 128x128 pixels).

- Normalization: Process of adjusting the pixel values in an image to a common scale. This is a crucial step in the preprocessing pipeline to ensure that the model can learn effectively and make accurate predictions. The Min-Max Normalization technique rescales the pixel values to a range between 0 and 1 or -1 and 1.
- Data Augmentation: Applying transformations such as rotation, flipping, and cropping to artificially increase the dataset size and variability. [22]

5.3. Evaluation Metrics and Loss Function

- Pixel Accuracy: is a fundamental evaluation metric used in image segmentation tasks, including medical image segmentation, to measure the proportion of correctly classified pixels in a segmented image relative to the total number of pixels. It is calculated by comparing the predicted segmentation mask with the ground truth mask, and it is expressed as:

  Pixel Accuracy (PA) = (NCCP) / (TNP),

  where NCCP stands for “Number of Correctly Classified Pixels,” and TNP means “Total Number of Pixels.”

- Dice Coefficient: Measures the overlap between the predicted segmentation mask and the ground truth. It is calculated as

  Dice=2×|A∩B||A|+|B|Dice=|A|+|B|2×|A∩B|,

  where AA and BB are the predicted and ground truth masks, respectively.

- Intersection over Union (IoU): Evaluates the ratio of the intersection and union of the predicted and actual segmentation masks. It is calculated as

  IoU=|A∩B||A∪B|IoU=|A∪B||A∩B|

- Loss Function: The dice loss function will be employed, as it directly optimizes the overlap between the predicted and actual segmentation masks, thereby improving the model's accuracy in identifying the optic disc and cup. [13]

5.4. Model Comparison and Optimization

Until now, we have implemented several models to check the results of the pixel accuracy in different models like U-Net, ResUNet, DC-UNet, and SegNet, both with the CBAM and without it. The range of values we have got is from 88% to 90.52%. The values for each model were the following:

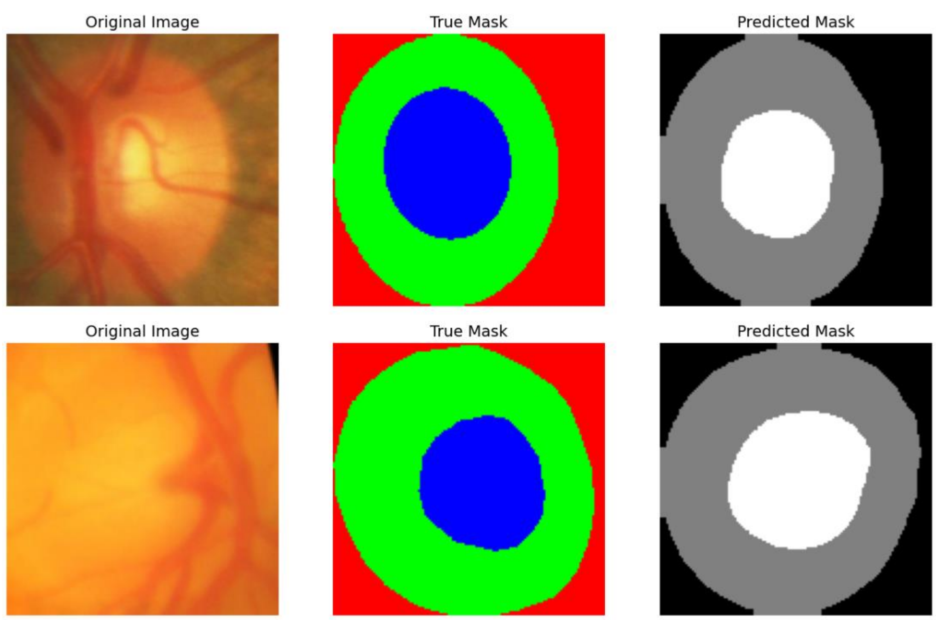
- SegNet with CBAM = 90.52%
- U-Net with CBAM = 90.30%
- DC-UNet with CBAM = 90.02%
- U-Net = 89.85%
- ResUNet = 88.00%

The best model in terms of pixel accuracy was SegNet with CBAM and the worst model was ResUNet. However, U-Net (both with CBAM and without it) could be train in a significantly smaller time. Now, the biggest increment in the pixel accuracy was observed after the implementation of the CBAM, which improved the quality of the U-Net model by 0.45%; on the other hand, the worst change to U-Net was its combination with ResNet, which worsened the accuracy by 1.85%.

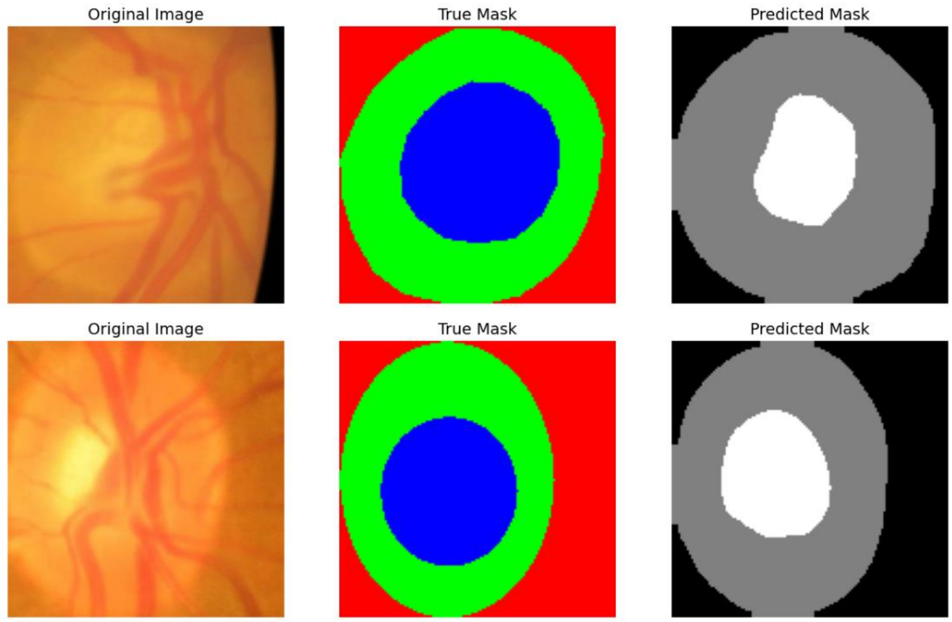
5.5. Predicted Masks

Here are the results of U-Net, DC-UNet, and SegNet, all of them using CBAM (attention mechanism):

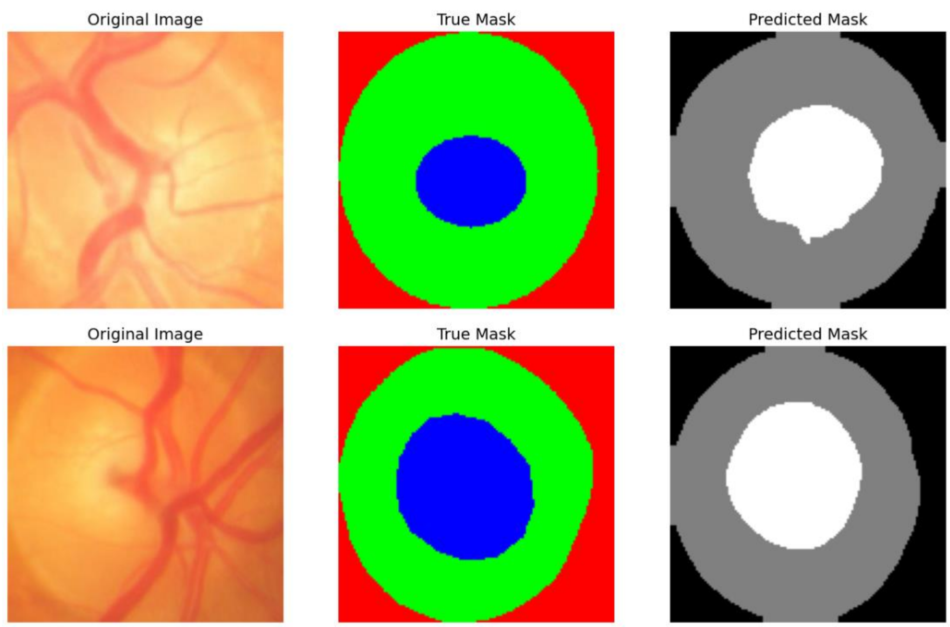
U-Net:



DC-UNet:



SegNet:



6. Conclusion

6.1. Some Realistic Constraints

Because medical image datasets often have limited training samples compared to traditional image datasets (for example, the Drishiti- GS dataset for glaucoma has only 0.1K samples, whereas natural image datasets like MS-Coco contain 330K training images), this disparity can lead to inferior performances and overfitting in domain generalization tasks. [23]

6.2. Future Work

Future work will focus on enhancing the model's robustness and generalizability through:

- Cross-Dataset Evaluation: Testing the model on different datasets to ensure its applicability across various demographic groups and imaging conditions.
- Ensemble Methods: Other combinations with the U-Net (including Transformer-based models) will be implemented to see if they can work better than ResUNet and DC-Unet, and CBAM will be used in any of them to check their result with this attention mechanism that has helped increment each model's accuracy. Also, the U-Net has been selected instead of SegNet because it has been much lighter and faster to train than SegNet. This latter obtained a little better pixel accuracy, but the time for training was significantly longer (which also occurred when training the DC-Unet), and this small improvement in accuracy did not make up for the time it takes. Hence, the U-Net model seems better fit to the glaucoma detection project necessities despite such small difference in their accuracy. Moreover, other metrics will be added as mentioned in the previous section, which will help measure the efficiency of each model with more precision.

Since this project aims to develop an AI-based segmentation model capable of accurately detecting and segmenting the optic disc and optic cup in retinal images for glaucoma screening, by leveraging deep learning techniques, particularly the U-Net architecture, the model aims to facilitate early diagnosis and monitoring of glaucoma, thereby potentially preventing irreversible vision loss in affected individuals.

References

1. Chen, Y., Bai, Y., Zhang, Y. (2024). *Optic disc and cup segmentation for glaucoma detection using Attention U-Net incorporating residual mechanism*. Retrieved on August 21, 2024, from https://peerj.com/articles/cs-1941/
2. Gao, Y. et al. (2019). *Accurate Optic Disc and Cup Segmentation from Retinal Images Using a Multi-Feature Based Approach for Glaucoma Assessment*. Retrieved on August 21, 2024, from https://www.mdpi.com/2073-8994/11/10/1267#metrics
3. Sreng, S. et al. (2020). "Deep Learning for Optic Disc Segmentation and Glaucoma Diagnosis on Retinal Images." *Applied Sciences 10*, no. 14: 4916. Retrieved on August 21, 2024, from https://doi.org/10.3390/app10144916
4. Miag-ull. (2020). *RIM-ONE DL*. Retrieved on August 21, 2024, from https://github.com/miag-ull/rim-one-dl
5. Rudas, I., Eigner, G., and Kovacs, L. (2016). SMC 2016 and SMC Junior 2016 [Conference Reports]. *IEEE Systems, Man, and Cybernetics Magazine*, vol. 2, no. 1, pp. 41-44, Jan. 2016. Retrieved on August 21, 2024, from doi: 10.1109/MSMC.2016.2538501
6. Hafeez, S., and Kathirisetty, N. (2022). "Effects and Comparison of different Data pre-processing techniques and ML and deep learning models for sentiment analysis: SVM, KNN, PCA with SVM and CNN." *2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)*, Hyderabad, India, 2022, pp. 1-6. Retrieved on August 21, 2024, from doi: 10.1109/ICAITPR51569.2022.9844192
7. Litjens, G. et al. (2017). *A survey on Deep Learning in Medical Image Analysis*. Retrieved on August 21, 2024, from https://arxiv.org/abs/1702.05747
8. Sudhan, M. B. et al. (2022). *Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model*. Retrieved on August 21, 2024, from https://pubmed.ncbi.nlm.nih.gov/35222876/
9. Nazir, T. et al. (2021). *Optic Disc and Optic Cup Segmentation for Glaucoma Detection from Blur Retinal Images Using Improved Mask R-CNN*. Retrieved on August 21, 2024, from https://www.researchgate.net/publication/353452193_Optic_Disc_and_Optic_Cu

p_Segmentation_for_Glaucoma_Detection_from_Blur_Retinal_Images_Using_Improved_Mask-RCNN

10. Diakogiannis, F. I. et al. (2020). *ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data*. Retrieved on August 21, 2024, from https://arxiv.org/pdf/1904.00592
11. Badrinarayanan, V. (2016). *SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation*. Retrieved on August 21, 2024, from https://arxiv.org/pdf/1511.00561
12. Sudre, C. H. et al. (2017). *Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations*. Retrieved on August 21, 2024, from https://arxiv.org/abs/1707.03237
13. Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Retrieved on August 21, 2024, from https://arxiv.org/abs/1505.04597
14. Guan, H., & Liu, M. (2021). *Domain Adaptation for Medical Image Analysis: a Survey*. Retrieved on August 21, 2024, from https://arxiv.org/abs/2102.09508
15. Woo, S. et al. (2018). *CBAM: Convolutional Block Attention Module*. Retrieved on August 21, 2024, from https://arxiv.org/pdf/1807.06521
16. Ma, X., Cao, G., & Chen, Y. (2024). “A Review of Optic Disc and Optic Cup Segmentation Based on Fundus Images.” *The Institution of Engineering and Technology*. Retrieved on August 21, 2024, from https://www.researchgate.net/publication/380400097_A_review_of_optic_disc_and_optic_cup_segmentation_based_on_fundus_images
17. Lou, A., Guan, S., & Loew, M. (2020). *DC-UNet: Rethinking the U-Net Architecture with Dual Channel Efficient CNN for Medical Images Segmentation*. Retrieved on August 21, 2024, from https://arxiv.org/pdf/2006.00414
18. Zhang, Z. et al. (2010). ORIGA(-light): an online retinal fundus image database for glaucoma analysis and research. Retrieved on August 21, 2024, from https://ieeexplore.ieee.org/document/5626137

19. Fu, H. et al. (2019). *REFUGE: Retinal Fundus Glaucoma Challenge*. Retrieved on August 21, 2024, from https://ieee-dataport.org/documents/refuge-retinal-fundus-glaucoma-challenge
20. Bajwa, M. N. et al. (2020). *G1020: A Benchmark Retinal Fundus Image Dataset for Computer-Aided Glaucoma Detection*. Retrieved on August 21, 2024, from https://arxiv.org/abs/2006.09158
21. Sevastopolsky, A. (2017). *Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network*. Retrieved on August 21, 2024, from https://link.springer.com/article/10.1134/S1054661817030269
22. Lin, T. Y. et al. (2014). *Microsoft COCO: Common objects in context*. Retrieved on August 21, 2024, from https://arxiv.org/abs/1405.0312
23. Biloborodova, T. et al. (2023). *Generalization Ability in Medical Image Analysis with Small-Scale Imbalanced Datasets: Insights from Neural Network Learning*. Retrieved on August 21, 2024, from https://link.springer.com/chapter/10.1007/978-3-031-49011-8_19