**OPTIC DISC DETECTION IN THE BINARY IMAGE OF AN EYE USING U-NET**

A minor project submitted to

In partial fulfillment of the requirements for the Degree Of

**Master of Computer Application (MCA)**

By

|  |  |
| --- | --- |
| **Name** | **Enrollment no.** |
| Saikat Santra | 12022010010033 |
| Abhrajit Kar | 12022010010047 |
| Sanjukta Mukherjee | 12022010010032 |

Under the guidance of

**Mr. Supratim Ghosh**



DEPARTMENT OF COMPUTER APPLICATION

Sem-3, 2023-2024

**DECLARATION CERTIFICATE**

This is to certify that the work presented in the document entitled “Optic disc detection in the binary image of an eye using U-net” in partial fulfilment of the requirement for the award of degree of Master of Computer Application of Institute of Engineering & Management is an authentic work carried out under my supervision and guidance.

To the best of my knowledge the content of this document does not form a basis for the award of any previous Degree to anyone else.

Date:

Dept. of Computer Application

Institute of Engineering and Management

**Dr. Soumi Dutta**

Dept. of Computer Application

Institute of Engineering and Management

**CERTIFICATE OF APPROVAL**

The forgoing thesis “**OPTIC DISC DETECTION IN THE BINARY IMAGE OF AN EYE USING U-NET**” is hereby approved as a creditable study of research topic and has been presented in satisfactory manner to warrant its acceptance as prerequisite to the degree for which it has been submitted.

It is understood that by this approval, the undersigned do not necessarily endorse any conclusion drawn or opinion expressed therein but approve the work for the purpose for which it is submitted.

Viva-Voice held on…………………

**(Internal Examiner) (External Examiner)**

**ACKNOWLEDGEMENT**

We would like to express our special thanks of gratitude to our Guide, **Mr. Supratim Ghosh** who helped us a lot in this project, his valuable suggestions helped us to solve toughchallenges and without his help this project could not have been completed in time. Aspecial thanks to our Head of Department **DR. Soumi Dutta** who gave us the goldenopportunity to do this wonderful project on the topic “**OPTIC DISC DETECTION IN THE BINARY IMAGE OF AN EYE USING U-NET**”, which helped us to gain a significant knowledge inthe aforesaid subjects. Secondly, we would like to thank our friends who helped us a lotin finalizing this project within the given time frame.

|  |  |
| --- | --- |
| **Name** | **Enrollment no.** |
| Saikat Santra | 12022010010033 |
| Abhrajit Kar | 12022010010047 |
| Sanjukta Mukherjee | 12022010010032 |

**ABSTRACT**

The retinal image analysis has been of great interest in the recent past for medical optical diagnosis. Different techniques have been developed for segmentation of eye structures such as optic discs which prove to be an efficient and reliable solution for optical diagnosis. While optic disc detection is primarily used in medical and ophthalmological applications, it can also find applications in non- medical context. Optic disc detection in the video of an eye can be used in fields of computer vision, image processing and surveillance.

In this project we are aiming towards building a model in which we detect the location of the optic disk in the image and video of an eye and use it in human-computer interaction field. Eye movement tracking is a type of indirect human-machine interaction which can be used to better comprehend a person’s focus, purpose in context sensitive scenarios. This present model is trained using the binary image of an eye. This is done in two stages. First, the ground truth is formulated and a mask is obtained.

Next a supervised learning algorithm is used to train a model which takes the binary image of an eye as input and returns the accurate location of the optic disc and its masked image. This model has proved to be very efficient and gives an overlap of more than 90% on an average in returning the detected and masked binary image. For future modifications and development IDRiD dataset for fundus images and RVD dataset for videos will be used to train the model for optic disc detection in video of an eye.

**TABLE OF CONTENTS**

**Contents Page Number**

**Chapter 1:**

Introduction 1-2

````````````````````````````````````````````````````````````````````````````````````````````````````

**Chapter 2:**

Objective 3

````````````````````````````````````````````````````````````````````````````````````````````````````

**Chapter 3:**

Dataset Preparation 4-5

````````````````````````````````````````````````````````````````````````````````````````````````````

**Chapter 4:**

Methodology 6-13

````````````````````````````````````````````````````````````````````````````````````````````````````

**Chapter 5:**

Results 14-15

````````````````````````````````````````````````````````````````````````````````````````````````````

**Chapter 6:**

Conclusion  16

````````````````````````````````````````````````````````````````````````````````````````````````````**Chapter 7:**

Future Scope 17

````````````````````````````````````````````````````````````````````````````````````````````````````

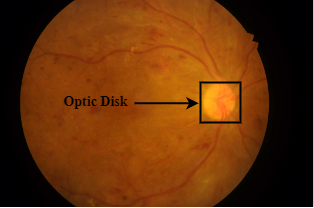
**References**  18

**CHAPTER 1:**

**INTRODUCTION**

The optic disc, sometimes called the optic nerve head, is a round section at the back of the eye. It is where the retina and optic nerve connect. The optic disc is also where the retina’s main artery and vein enter the eye. There is a layer of tissue at the back of each eye, opposite to the pupil, called the retina. This tissue is responsible for taking the light that enters the eye and turning it into the images you see.

There are two main areas in the retina: the macula and the peripheral retina. The macula, which is located in the center of the eye, is responsible for seeing images directly in front of you. The peripheral retina, which makes up the rest of the retina, is responsible for peripheral vision.



The optic disc Is a round, slightly raised area at the border of the macula and the peripheral retina, and it can be yellow-orange or pink. It is the only spot on the retina that has no rods or cones, making it a “blind spot.”

Fig. 1: Fundus image of eye (Source:IDRiD dataset)

When light enters the eye through the lens, it hits the retina at the back of the eye. The photoreceptors take the light and turn it into electrical signals. These signals then travel through the retina to the optic disc. The optic nerve takes these electrical signals and delivers them to the brain, which interprets them into the images we see.

There is a small indentation at the center of the optic disc, called the physiologic cup. This is where the optic nerve, made up of over a million nerve fibers, connects to the retina. The ground truth of the project is based on this criterion. To detect the optic disc, we locate the cluster of nerves which connect the optic disc to the retina. The binary image is preprocessed and converted into grayscale. The most intense region is considered as the target part.

After detecting the region of interest, a mask is created around the region. A set of 1317 binary images are processed using the ground truth and a dataset including binary images and their masks are created. This dataset is then used to train the U-net image segmentation model. A set of 565 binary images are used to test the model built. Accuracies of all images are calculated and saved in a file. On an average more than 90% overlap is achieved. Overlap of a single image is calculated based on the difference between the predicted mask and the true mask.

**CHAPTER 2:**

**OBJECTIVE**

With the massive increase in the invention of computers arises the need to communicate and interact with the computer. Human – Computer interaction is now not only confined to printers and keyboards. There are now inputs like voice, gestures and facial expression visuals.

Optic disc detection widely serves the purpose of medical optical diagnosis. But detecting the optic disc and tracking the eye movement can also serve in non-medical applications. The optic disc detection in a video can primarily help in areas of computer vision, image processing and surveillance. Tracking the eye movement, including the optic disc can be used for eye-based control interfaces in applications like gaming, virtual reality and human computer interaction research.

This project aims at:

1. Detecting the optic disc in the image of an eye.
2. Building a model to find masked image output locating the optic disc position.
3. Building a model which can be used for human-computer interaction. A model which detects the movement of the optic disc in every frame throughout the video.

In this project the first two goals of the project which are detecting the optic disc and building a model to find masked images locating the optic disc are completed and the documentation for the same has been provided.

**CHAPTER 3:**

**DATASET PREPARATION**

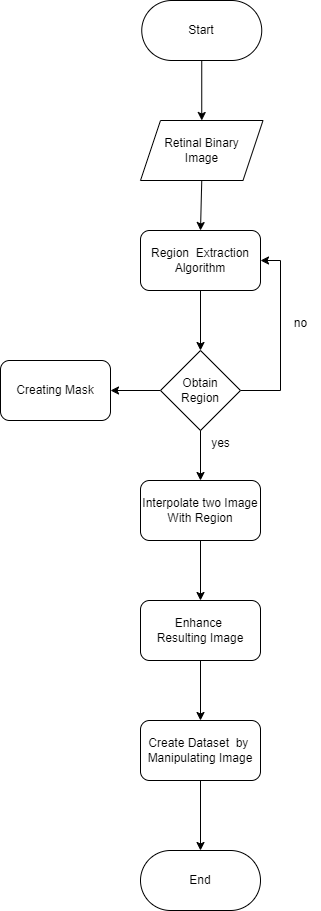


Fig. 2: Flow chart representing steps of dataset preparation

For working on this project, a dataset of retinal binary images is used. To increase the size of the dataset, data augmentation is done on the initial binary image dataset.

In this case to execute data augmentation, after the desired region is obtained and mask their mask is obtained, two images are interpolated based on the positions of their regions of interest. These images were then mirrored to artificially produce images for the training dataset.

A white lines on a black background

Description automatically generatedA close-up of a brain

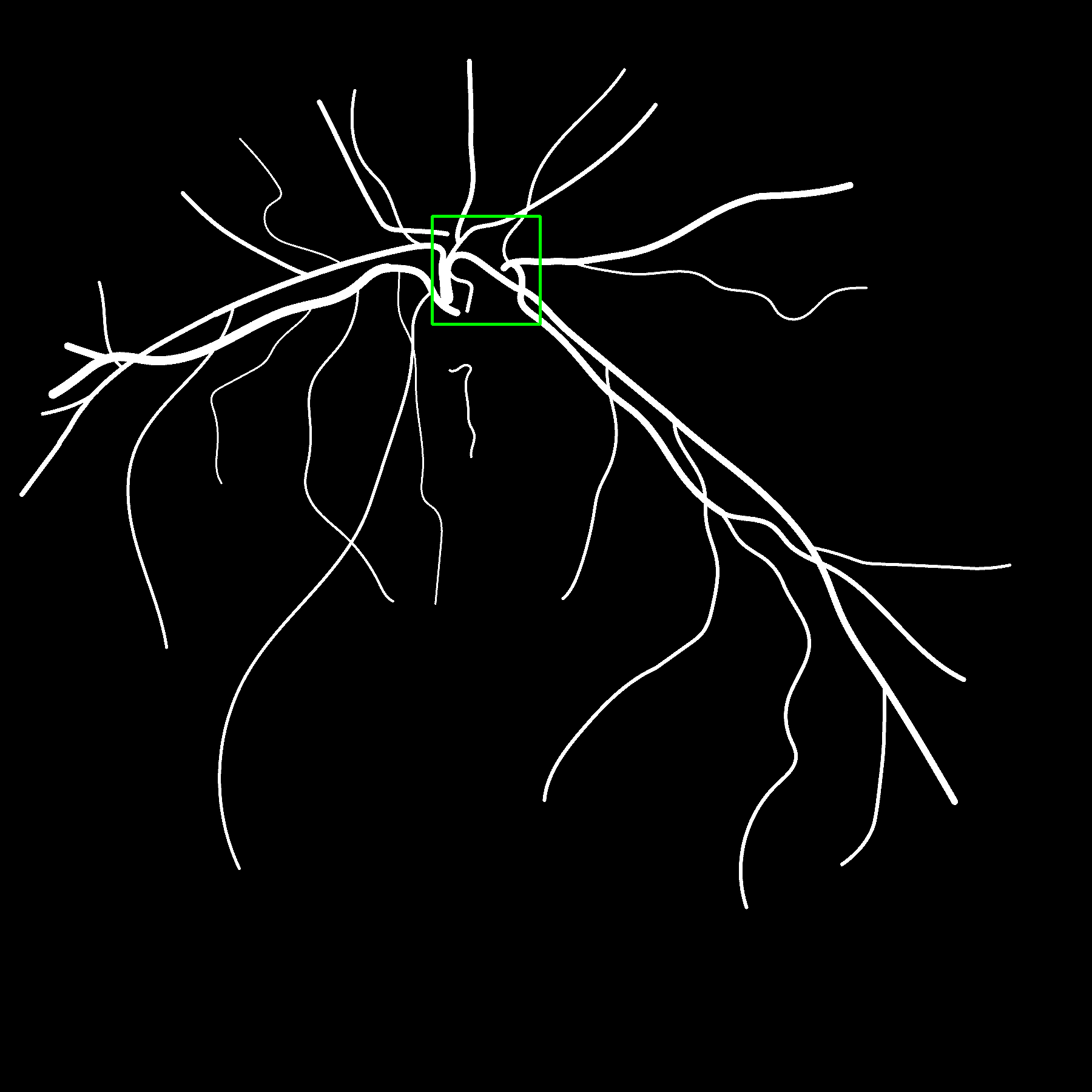
Description automatically generated

Image 1



Combined Image Resulting Image

Image 2

Fig. 3: Data augmentation by overlapping images

A total of 1994 such images were produced among which only 1882 were usable to train the model. From these 1882 images, 70% images (1317) were considered for training the model and another 30% (565) was used to test the model. The original binary image dataset can be downloaded from GitHub with the address

https://github.com/SaikatSantra9/binary-image-of-optic-disk/tree/main/Binary%20images

**CHAPTER 4:**

**METHODOLOGY**

The methodology part is essentially divided into two sections: 1) Locating the optic disc by considering the part with maximum intensity. 2) Training and testing the model. After this we calculate the overlap of the model.

**4.1 Locating the optic disc**

To detect and locate the optic disc in the binary image the following steps were followed:

**4.1.1 Reading and Preprocessing the Image:**

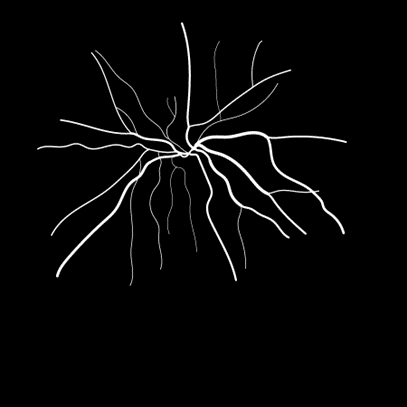
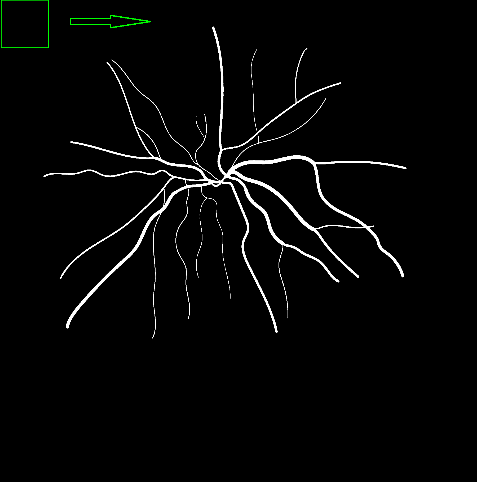
The binary image is loaded using OpenCV and is converted into grayscale. The image dimensions (height, width) are obtained and other variables needed for further processing are initialized. A window of size 178 is defined for analysis.

Fig. 4: Binary image of the eye

**4.1.2 Locating the Optic Disc:**

The intensity of each window is measured. The This window is iterated throughout the entire image horizontally one row at a time. window with the maximum intensity is considered as the region of optic disc.

Fig. 5: Raster scan performed on the image

**4.1.3 Highlighting the Optic Disc**

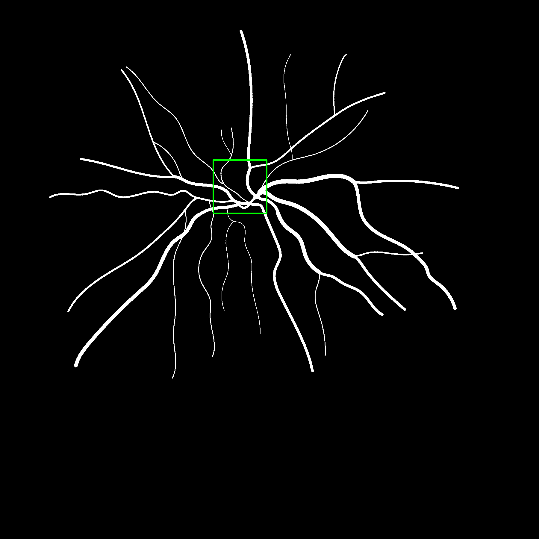
After we get the region of maximum intensity, we enclose it within a rectangle to highlight it from the rest of the image. The colour and thickness of the rectangle is adjusted and the most intense region of the rectangle is extracted.

Fig. 6: Binary image of eye highlighting the optic disc

**4.1.4 Masking:**

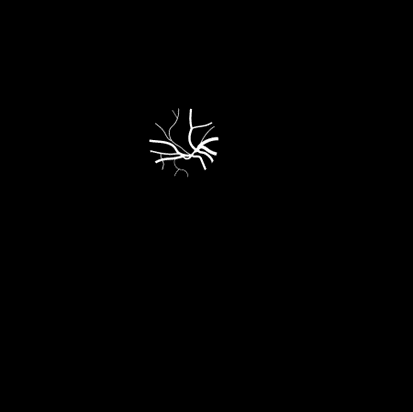
Now a circular mask is created. Its radius is defined based on the detected region’s central point. This circular mask is then applied to the image.

Fig. 7: Circular mask showing location of the optic disc

**4.2 Training and Testing the model**

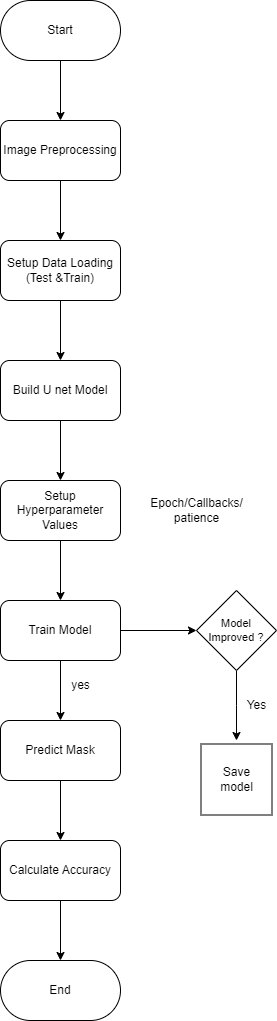


Fig. 8: Flow chart representing steps of building the model

For training and testing the model U-Net image segmentation model is used. U-net is a semantic segmentation technique which was originally proposed for medical image segmentation. U-Net was first introduced in the paper, U-Net: Convolutional Networks for Biomedical Image Segmentation in the year 2015. It is a simple model with an encoder (for downsampling) and a decoder (for upsampling) with skip connections. The above flowchart represents the set of steps to be followed for training and testing the model. The brief description about each step is given below:

**4.2.1 Image pre-processing:**

In this step the data is prepared for the machine learning model. Firstly, the desired width, height and number of channels for the images are defined. In this case, the images are expected to be of size 128x128 pixels with only 1 channel (binary). Even though the current model is trained using binary images, coloured images could be used by increasing the colour channel size to 3. The directories to the training and test dataset are then set. NumPy arrays are initialized to store image data (X) and corresponding masks (Y). X is initialized with zeroes and has the shape (number of training samples, 128,128,1) indicating binary images. Y is initialized with zeroes and has the shape (number of training samples, 128,128,1) indicating binary masks.

**4.2.2 Setup data loading (test and train):**

The X and Y input and their corresponding masks are resized and three sets of data is returned. The first set of data includes the set of binary images for training the model (X\_train). The second set of data includes the set of corresponding masks of the binary images for training the model (Y\_train). The third set of data includes the set of binary images for testing the model (X\_test). The final output of the masks for the binary images used for testing are then saved in (Y\_test).

**4.2.3 Build U-Net Model:**

The model is built using U-net. First the class value is set to 1 which indicates that the model is designed for binary segmentation, where each pixel in the output is assigned to one of two classes (foreground or background). Hence, since the goal is to return a binary mask the value of class is set to 1.

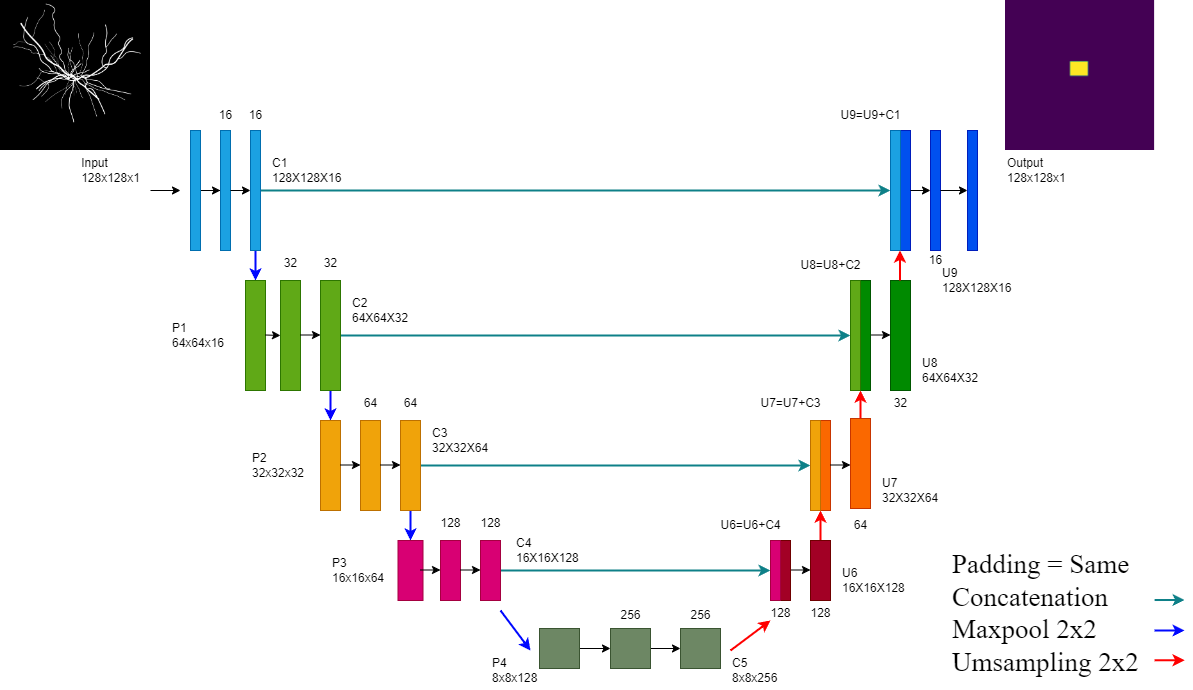


Fig. 9: Developed U-Net architecture

Contraction path (Encoder) of the U-net architecture is then defined. The encoder is responsible for capturing and extracting features from the input image. Here each layer is designed to progressively reduce the spatial dimensions of the input and increasing the number of feature channels. The encoder captures features at different levels of abstraction through a series of convolutional and pooling operations and using ReLU activation function. The model is prepared to later recover spatial details through the expansive path.

Expansive path (Decoder) of the U-Net architecture is defined next. The expansive path is responsible for upsampling and recovering spatial information from the features captured in the contraction path. The expansive path applies transposed convolution (also known zs fractionally strided convolution or deconvolution) to upsample the feature maps. This operation doubles the spatial dimensions. Next it concatenates the upsampled feature maps with the corresponding feature maps from the contracting path. The skip connections formed by concatenation helps in preserving information during the upsampling process.

Lastly the output layer of the U-Net model is defined. It is a convolutional layer with a sigmoid activation function. It takes the features from the expansive path and produces the final segmentation output with the specified number of classes which in this case is 1. Class value is 1 as final output is a binary mask and hence sigmoid activation function is applied which squashes the output values in the range [0,1] and is used for binary classification problems.

**4.2.3.1 Comparison between two developed models**

Two U-Net models were built to find the masks to the binary images with slight variations in terms of layer configuration.

In the initial model Batch Normalization and ReLU activation function was used only after the first convolution in each block of the contraction path. Similarly in this model the dropout was used only after the first convolution in each block of the contraction path. In case of the expansive path Dropout, ReLU activation functions were used after each concatenation in the upsampling path. This decoder included Dropout layers after each convolution layer in the upsampling path. This model however could only result in an overlap of 73% on an average so slight alterations were made to make the model give more accurate results.

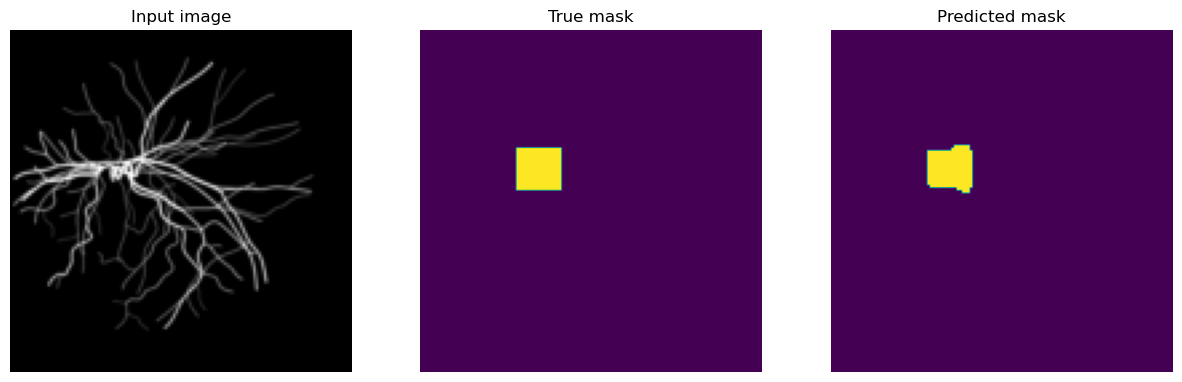


Fig. 10: Output from the initial model displaying input image its true mask and predicted mask

In the final model Batch Normalization and ReLU activation function was used after each convolutional layer in the contraction path. The dropout was used after each convolutional layer with varying dropout rates in the encoder. This provided regularization benefits and prevented model from getting overfitted. This model had more consistent use of Batch Normalization and ReLU activation throughout the architecture. In the decoder additional convolutional layers were removed and only upsampling layers, Batch Normalization and ReLU was used. This model served the purpose and resulted in an overlap of 92% on an average and this model was chosen.

**4.3.4 Setup Hyperparameter values & train the model:**

The earlyStopping callback monitors the validation loss and stops the training process early if there is no improvement for a certain number of epochs which is defined by patience which in this case is 2. model.fit method trains the model on the provided x\_train and y\_train dataset. The validation data x\_test and y\_test is used to monitor the performance of the model on unseen data during training. Epochs are set to 30 which is the number of times the entire training dataset is passed forward and backward through the neural network. This training setup helps to prevent overfitting by stopping training early if the validation loss stops decreasing.

**4.3.5 Predict mask:**

After training the model creates a binary mask where each element is True if the corresponding element in prediction > 0.5, and false otherwise. So the predicted\_mask is a binary mask where each pixel is either 0 or 1, based on the corresponding pixel in the model’s prediction was above or below the threshold of 0.5. The model chooses at random input from the test dataset and after prediction the mask is displayed. Three images namely input image, true mask and predicted mask is displayed on the screen.

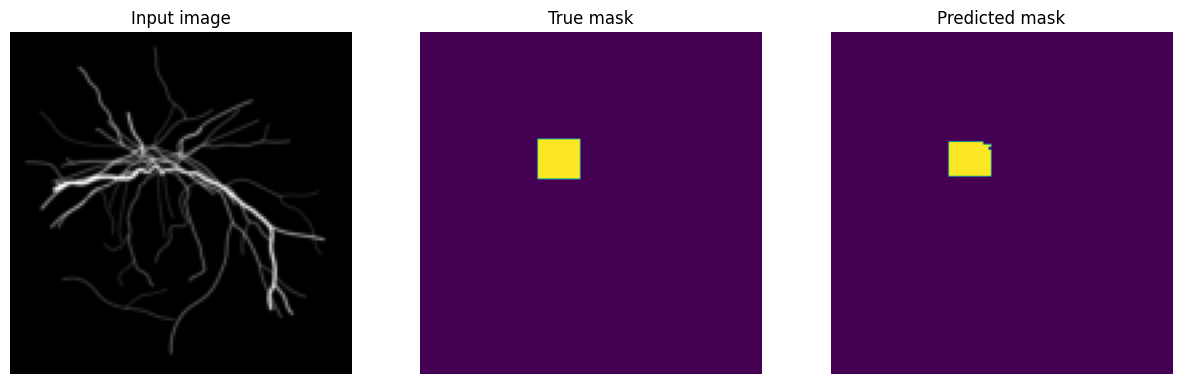


Fig. 11: Output from the final model displaying input image its true mask and predicted mask

**4.3.6 Calculate accuracy:**

The model calculates the Intersection Over Union (IoU) between the groundtruth mask and the predicted mask. The IoU is a common evaluation metric for segmentation tasks. It calculates the ratio of the intersection to the union of the true and predicted mask. The IoU value is a measure of how well the predicted mask aligns with the true mask.

A black text on a white background

Description automatically generated

IoU is basically area of overlap divided by area of union. The IoU can be understood clearly with the help of the diagram below

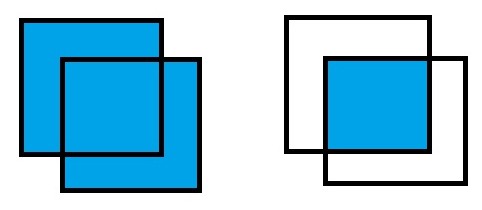


Fig. 12: Pictorial representation of Intersection over Union

In the numerator the area of overlap is calculated between the predicted mask and the ground truth mask. The area of union is the area bounded by both the predicted mask and the ground truth mask. Diving the area of overlap by area of union we get out required IoU.

Additionally, the percentage of the predicted mask in terms of the true mask in terms of the true mask is calculated and saved in a file to calculate average overlap.

**CHAPTER 5:**

**RESULTS**

The main objective of this project was focused on two things. First was to find the ground truth of the model which can correctly predict the location of the optic disc. Second was to build and train a model which efficiently returns the masked image for a binary image input.

These two problems were solved and implemented using the Python Programming language. The ground truth of the project gave an accuracy of around 70% - 73% which means the developed ground truth system could accurately locate the optic disc in around 70% of the images in the initial binary image dataset.

These images were then carried forward to train the image segmentation model. U-Net image segmentation model was implemented to finally predict the location of the optic disc and return a masked image for a binary image input. The model was trained using 1317 binary images along with their masks and was tested with 565 images. The overlap of each image was calculated using IoU between ground truth mask and predicted mask and gives an overlap of 92% on an average.

The below graph shows the training and validation loss over the epochs during the training of the neural network.

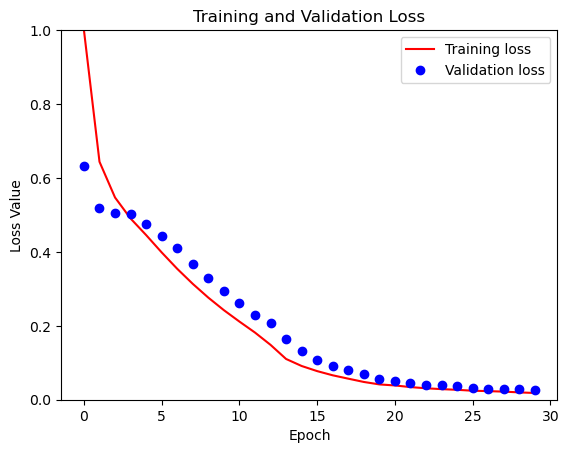


Fig. 13: Graph representing training and validation loss

The x-axis represents the number of epochs and the y-axis represents the loss value between the range 0-1. The red line depicts the training loss values while the blue dots depict the validation loss. The model is said to be working correctly if there is decreasing trend in both the training and validation loss as in this case.

**CHAPTER 6:**

**CONCLUSION**

The developed system can be seen fulfilling both the objectives of the project. The system effectively detects the location of the optic disc and returns the masked image for a binary input with an average overlap of 92%.

The training and validation loss graph stands as an evidence that the developed model is an efficient solution to the problem. It also shows how the model is away from issues like overfitting and underfitting.

The methodology gives an efficient solution to the given problems and can be modified further for RGB image inputs and can be developed to extract more features other than just the optic disc which can help in the goal of human-computer interaction even more.

**CHAPTER 7:**

**FUTURE SCOPE**

The immediate future scope is to work with RGB image dataset input and then Video dataset input and continuously track the position of the optic disc in each frame. This will help in the ultimate goal of the project to build a model which tracks eye movement for human – computer interaction.

The existing model can be modified more by using different algorithms for finding the groundtruth to increase the ground truth accuracy also the neural network needs to be modified and worked on to decrease the computation time and achieve even better results. This will improve the overall performance of the developed system.

**REFERENCES**

* O. Ronneberger, P. Fischer, and T. Brox, “U-NET: Convolutional Networks for Biomedical Image Segmentation,” in *Lecture Notes in Computer Science*, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4\_28.
* G. Y. Kim, S. H. Lee, and S. M. Kim, “Automated segmentation and quantitative analysis of optic disc and fovea in fundus images,” *Multimedia Tools and Applications*, Apr. 2021, doi: 10.1007/s11042-021-10815-1.
* B. J. Bhatkalkar, S. V. Nayak, S. V. Shenoy and R. V. Arjunan, "FundusPosNet: A Deep Learning Driven Heatmap Regression Model for the Joint Localization of Optic Disc and Fovea Centers in Color Fundus Images," in IEEE Access, vol. 9, pp. 159071-159080, 2021, doi: 10.1109/ACCESS.2021.3127280.
* J. Dietter *et al.*, “Optic disc detection in the presence of strong technical artifacts,” *Biomedical Signal Processing and Control*, vol. 53, p. 101535, Aug. 2019, doi: 10.1016/j.bspc.2019.04.012.
* M. W. Khan, “RVD: a Handheld Device-Based FunDUs video dataset for retinal vessel segmentation,” *arXiv.org*, Jul. 13, 2023. <https://arxiv.org/abs/2307.06577>
* B. Jin, P. Li, P. Wang, L. Shi, and J. Zhao, “Optic disc segmentation using Attention-Based U-Net and the improved Cross-Entropy convolutional neural network,” *Entropy*, vol. 22, no. 8, p. 844, Jul. 2020, doi: 10.3390/e22080844.
* N. Modi and J. Singh, “Role of eye tracking in human computer interaction,” *ECS Transactions*, vol. 107, no. 1, pp. 8211–8218, Apr. 2022, doi: 10.1149/10701.8211ecst.
* M. Maynard-Reid, “U-Net image segmentation in Keras - PyImageSearch,” *PyImageSearch*, Mar. 23, 2023. <https://pyimagesearch.com/2022/02/21/u-net-image-segmentation-in-keras/>
* A. A. Awan, “A complete guide to data augmentation,” Nov. 23, 2022. https://www.datacamp.com/tutorial/complete-guide-data-augmentation