# Deep learning-based contour detection for retinal layer segmentation and sickle cell retinopathy detection in OCT images

Ashuta Bhattarai
Department of Computer and Information Sciences
University of Delaware
ashutab@udel.edu

**Abstract**

Contour detection is a frequently sought-after technology in everyday life. In this research, we present a novel contour detection method specialized in obtaining continuous contour lines. Our work is focused on designing the architecture of a new U-Net variant called RLUNet. Unlike existing segmentation methods, our network can produce natural and uninterrupted contour lines without post-processing the results. The critical factor behind this concept is a novel activation function that converts discrete signals into smooth and continuous contours. We find the application of our method in medical image segmentation problems. Hence, we train our network on cross-sectional scans obtained through the optical coherence tomography (OCT) exam to extract contours of the retinal layers. Our work on some of the most popular OCT datasets shows promising results. We also used the contours extracted from OCT images to develop an algorithm for detecting Sickle-cell Retinopathy, which is an unexplored territory in the deep learning community. Using private hospital data and manual annotations from expert ophthalmologists, we successfully evaluated our SCR diagnostic algorithm with high accuracy.

**Keywords:** Contour detection, U-Net, optical coherence tomography

## 1 Introduction

Contour detection, a fundamental concept in image analysis and understanding, comes in different names - edge detection, boundary detection and in some cases, object segmentation. Over the past decade, contour detection techniques

have been employed in several applications such as medical image segmentation, object recognition and, scene understanding [1]. With the exponential growth and success of Artificial Intelligence (AI), an increasing number of Deep Learning (DL) based contour detection systems have been applied in medical image research [2, 3], particularly in diagnostic studies [4, 5]. Our focus in this work is to develop a robust contour detection model for retinal scans taken during an Optical Coherence Tomography (OCT) exam.

OCT [6] is a non-invasive optical imaging modality that performs high-resolution, cross-sectional tomographic imaging of the eye in real-time. The OCT technology is used as a diagnostic tool designed to assist ophthalmologists in identifying retinal diseases, such as Sickle Cell Retinopathy (SCR). The proposed research aims to develop a robust algorithm to localize the retinal layers affected by SCR using deep learning segmentation. Under microscopy, a cross-section of a human retina has 10 distinct layers. An OCT examination produces a sequence of retinal cross-section images or B-scans. The number of cross-sections produced is dependent on the OCT instrument and the pathology. For example, the SPECTRALIS OCT (www.know-the-eye.com/instruments/spectralis-oct/) generally produces 31 cross-sections (or A-scans) to examine SCR. An ophthalmologist can detect and locate SCR by observing the change in thickness of contours in the B-scans. A B-scan affected by SCR has one or more abnormally thin inner retinal layer regions than the outer reflective layer. An image containing 11 retinal layers scanned during an OCT exam is shown in figure 1. Among these layers, there are four that are studied to diagnose SCR. These are: ILM ($1^{st}$ layer), OPL ($6^{th}$ layer), PR1($8^{th}$ layer) and BM ($11^{th}$ layer). But, the retinal macula also contains a pit-like structure called fovea which also has a thin distance between the inner retinal layers - ILM and OPL. Unlike the injury due to SCR, the fovea is an important part of the retina that supports our vision. The difference is that fovea has a deep and symmetrical indentation between the two layers whereas, an SCR related injury has a shallow and asymmetrical indentation. The indentation in the fovea is the thinnest part of the retina, whereas the latter is only relatively thinner than nearby areas.

To develop an automatic SCR detection method, the following components are required:

1. Distinct segmentation of retinal layers with continuous contours
2. Thickness comparison between the retinal layers at multiple areas of the contours
3. Correct distinction between an unusual retinal thinning caused by SCR and normal retinal thinning present in the fovea

The SPECTRALIS OCT performs not only retinal scans but also generates automatic layer segmentation. However, this segmentation is typically inaccurate for several reasons, such as poor image quality or obstruction due to blood vessels and damaged retinal tissues. With incorrect layer segmentation, automatic SCR identification and localization becomes unreliable. Thus,

ophthalmologists are compelled to examine each scan manually. However, as far as we know, no research related to deep learning-based SCR detection has been published yet. Thus, the motivation of this work lies in eliminating the challenges above using deep learning approaches to support and promote SCR diagnostic research.

The major contributions of our work are listed below:

1. A novel UNet based architecture (RLUNet) capable of accurate retinal layer segmentation verified on multiple datasets.
2. A novel spline-based activation function (PClu) that can produce continuous segments or contours.
3. Probably the first deep learning-based diagnostic research in Sickle Cell Retinopathy (SCR).

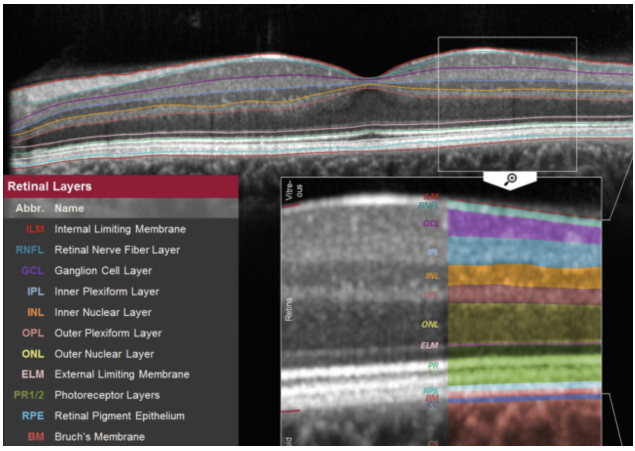



**Fig. 1**: Image illustrating 11 retinal layers in an OCT B-scan

## 2 Related Works

There are a large literature on contour detection and object segmentation methods. Beyond OCT scans and retinal layer segmentation, there have been several kinds of images and application areas where novel ideas for contour detection was studied. We discuss some of these works in section 2.1. In Sections 2.2 and 2.3 we discuss some of the most influential computer vision and deep learning literature related to RLS respectively.

### 2.1 Contour Detection

Li et al. in [7] proposed Edgetrack, an active contour tracking system of human tongue ultrasound scans. The proposed model used edge gradient and intensity information in local regions around each snake element. Kang et al.

proposed a fast contour extraction model [8] based on Multiple Cue Inhibition (MCI) operator named speedMCI. Dollar et al. proposed a supervised boundary detection model called Boosted Edge Learning (BEL) [9] that learns by selecting and combining a large number of features across different scales. [10] proposed a fast algorithm for image contour detection. The authors, Catanzaro et al., combined local image analysis and generalized eigensolver to implement their algorithm on highly parallel commodity processors for speed, accuracy, and processing large images. [11] improved contour detection accuracy by computing sparse code gradients (SCG) that measures contrast using patch representations. Ren and Bo used K-SVD and orthogonal matching pursuit to calculate sparse codes and used SVM for classification. SCG extracted rich representations from pixels to learn how to measure local contrast to find contours.

Rupprecht et al. [12] proposed a deep learning network for interactive boundary extraction. The network was trained to predict a vector originating from a point in the evolving contour and directing towards the closest point on the object of interest boundary. UNET proposed by Ronneberger et al. [2] is a widely used encoder-decoder network architecture for medical image segmentation. Several variations of U-net, such as [3] used stacked dilated convolutions to outperform its predecessor. Our proposed model is also based on dilated convolutions on UNET. Gong et al. [1] presented a review of then-existing contour detection approaches, including deep CNN based algorithms. The authors divided the related papers into three categories: pixel-based, edge-based, and region-based contour detection. The authors used their study to analyze and predict the future of contour detection. Kokkinos [13] proposed a multi-resolution FCNN architecture that improved performance by combining deep learning with grouping and integrating normalized cut techniques within a deep network. Yang et al. [14] proposed a fully convolutional encoder-decoder network to detect higher-level object contours, which could generalize well to images with unseen sub-classes. Deng et al. [15] proposed a CNN network to produce sharp boundaries without any post-processing. To train their network, Deng et al. used a novel loss function obtained by combining weighted cross-entropy loss and dice coefficient.

## 2.2 Computer vision based RLS

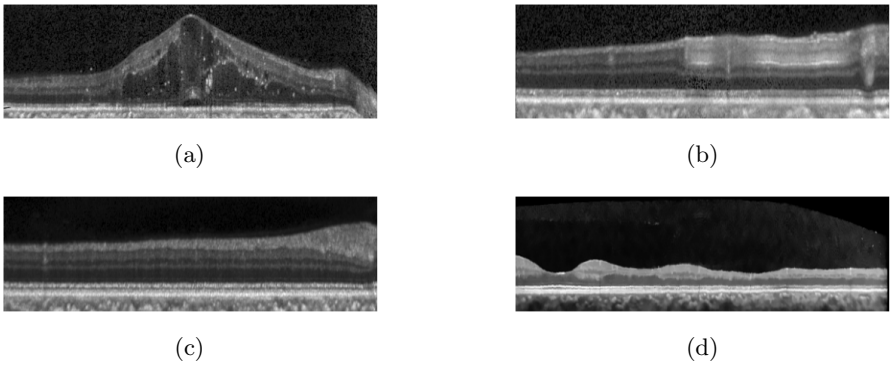
In [16] Giovinco et al.. extracted the contours of SD-OCT layers by approximating gradients through a series of computer vision operations like total variation denoising and edge enhancement. The authors performed the final segmentation using region fusion and complex diffusion algorithms. Sun et al.. in [17] perform 3D segmentation on OCT volume data. This method extracts retinal layer boundary surfaces sequentially with a decreasing search region of volume data. Authors Tian et al. [18] and Chiu et al. [19] used graph theory and graph-based shortest path search to segment eight retinal layers. For optimization, the former used inter-frame spatial dependency techniques, whereas the latter used dynamic programming. Lang et al. and Liu et al. [20, 21] used

a random forest classifier to learn boundary pixels between the retinal layers and to generate a probability map for each boundary point.

## 2.3 Deep learning-based RLS

Hsia et al. [22] used Mask R-CNN model [23] to segment choroid layers (ILM, CiB (Choroidal inner Boundary) and CoB (Choroidal outer Boundary) on SD-OCT images. SK et al. proposed a custom U-Net architecture called DRUNET [24] consisting of a standard convolutional block and a residual block in each down-sampling and up-sampling layer. DRUNET captured both local (tissue texture) and contextual (spatial arrangement of tissue) information to segment six optical nerve head (ONH) tissue layers. Li et al. [25] used an improved Xception-65 network along with an atrous spatial pyramid pooling module giving multiscale feature information to extract retinal boundaries in an encoder-decoder block. Kugelman et al. [26] proposed patch-based classification and semantic segmentation methods to segment choroidal tissues automatically. Gopinath et al. [27] proposed DL techniques to automate the pre-processing steps for OCT-based RLS methods using a combination of CNN and LSTM networks. The automated pre-processing steps included image denoising, ROI extraction, image flattening, edge detection and so on. Ngo et al. [28] proposed a deep feature-learning regression network to predict the corresponding retinal boundary pixel of an image segment using features such as intensity, gradient and, adaptive normalized intensity score (ANIS).

[29–32] studied the impact of various aspects of a CNN model in RLS and SD-OCT images. Authors Hamwood et al. in [29] studied the effects of CNN architecture modification (especially image patch size) in optimizing RLS. The research concluded that increasing patch size improved the performance of classification and reliable segmentation. Yanagihara et al. [30] studied and described methodological challenges in using the ML model for RLS, such as lack of a large SD-OCT dataset from multiple devices, lack of standard post-processing protocols between devices. The authors also presented a possible solution to these problems as a conclusion of their research. Kugelman et al. [31] studied and examined the image quality factors affecting the robustness of deep learning-based RLS methods. The authors pointed out that not using diverse images during training and evaluation challenges the prospect of implementing DL methods in clinical practice. The authors concluded that noise, low contrast, gamma reduction, etc., negatively impact the results and that image augmentation helps build more resilient models. Maloca et al. [32] studied the ambiguity in ground truth labels of OCT images and the impact it has on the RLS systems that use deep learning methods. The authors used Trainable Traceable Relevance Explainability (T-REX) technique on a DL based RLS network to study ground truth generation from multiple graders, calculation of hamming distance among the graders and ML algorithms and, smart data visualization. The authors concluded with visualizations that the ambiguity in ground truth has a significant impact on ML results.



(a) (b)

(c) (d)

**Fig. 2**: B-scan of: (*a*) DME patient (*b*) MS patient (*c*) Healthy patient (*d*) SCR patient

## 3 Dataset

We have trained and tested the proposed segmentation method on multiple benchmarked datasets [33, 34] containing OCT scans of retinal layers. The dataset by Chiu et al. [33] from Vision and Image Processing Lab at Duke University includes 110 OCT B-scans from 10 different individuals with severe diabetic macular edema (DME). This dataset provides manually segmented labels of 7 retinal layers. The original images in this dataset have a resolution of (768x224) pixels.

Similarly, the dataset by He et al. [34] from Image Analysis and Communications Lab (IACL) at Johns Hopkins contains OCT images from 14 healthy controls and 21 multiple sclerosis (MS) patients. It provides a total of 1715 B-scans and the segmentation labels of 9 retinal layers. The original images in this dataset have a resolution of (1024x128) pixels.

Figure 2 shows scans from a healthy, DME and ME patient taken from the datasets mentioned earlier. Since DME and MS are not the pathologies of interest for this research, we treat the scans of all healthy and diseased individuals the same. Furthermore, the retinal of the patients with these pathologies may have been affected in varying degrees as shown in figure 2a, 2b, providing us with a large variety of retinal structures in the dataset.

Additionally, we also worked on a private dataset from Nemours Children's Hospital, which contains the OCT scans from 22 patients belonging to an age group of 1-30 years. This dataset includes labelled scans from both eyes of a patient. The dataset is labelled manually by experienced ophthalmologists actively working with children suffering from SCR. The original images in this dataset have a resolution of (2032 x 596) pixels. The dataset contains B-scans of patients diagnosed with various stages of SCR. Hence, some scans show severely damaged retina (figure 2d) while others may even appear healthy. Each subject in this dataset may have had multiple OCT exams. The exam takes 31 B-scans from each eye of the patient. Hence we have 31 cross-sectional

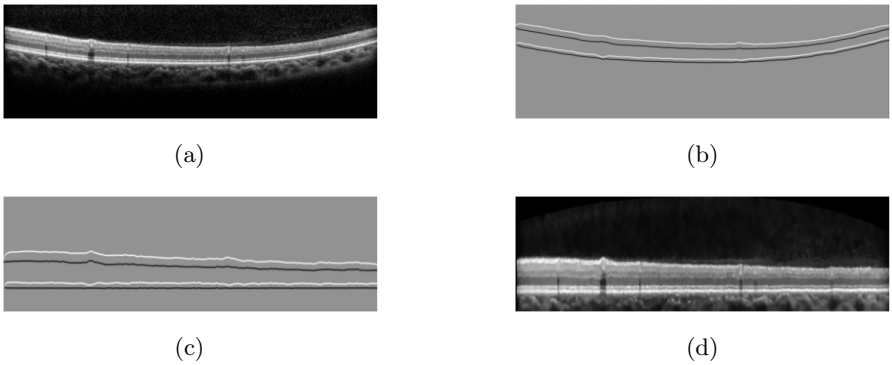


**Fig. 3**: Flattening steps: ($a$) Original image ($b$) After applying filters, threshold and sobel operator ($c$) After shifting columns ($d$) Flattened image

images for each eye of the 22 patients in our dataset. Most images in this dataset have segmentation labels of 10 or 11 retinal layers. But, the images from some severely affected patients may only have 4 segmentation labels viz, ILM, OPL, PR1 and BM, which are required in diagnosing SCR.

## 3.1 Pre-processing

The pre-processing step is required to feed clean data to the deep learning network to increase the training and inference speed. We perform image flattening, the most commonly followed pre-processing operation for OCT images, along with general image enhancement techniques.

A human retinal has strong natural curvatures, as shown in figure 3a. Strong curvatures and irregularities in the OCT images make the segmentation process more complicated than it has to be. Furthermore, SCR is detected using the rate of change in distance between the ILM and OPL layers. The distance measurement at multiple points of the contours can be complex and error-prone if the contours are strongly curved. Therefore, it is common practice to flatten the retinal scans before operating on the OCT images. Among all the layers in the retinal, BM is one of the most hyper-reflective layers. It is also the most stable layer with minimum irregularities that transition from lighter to darker regions towards the bottom boundary. Additionally, the first layer, ILM, is the easiest to detect using computer vision techniques since it transitions from darker to lighter regions. Thus, we exploit these properties to estimate the retinal boundaries and flatten the entire image to the BM boundary. First, we convert the images into grayscale and denoise using the Non-local Means Denoising algorithm [35]. This algorithm efficiently clears out white Gaussian noise. Then, we apply the OTSU thresholding algorithm to binarize the image and use the Sobel operator to calculate gradients along the y-axis to obtain the top and bottom boundaries. Then we remove outliers using morphological tools and median filtering. Once we have a sparse

estimate of the BM layer, we fit a polynomial curve along the BM boundary. Then, we make BM parallel to X-axis by shifting individual columns of the denoised image down to the plane tangential to the lowest point of the BM curve. The intermediate steps of this procedure are illustrated in figure 3. We use this denoised, flattened image for training and inference. The ground

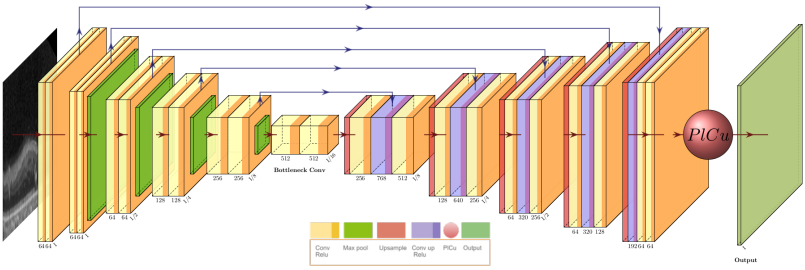



**Fig. 4**: RLUNet architecture

truth labels for each contour is converted into a segmentation mask of shape ($batch\_size$, $num\_class$, $height$, $width$). We dilate the masks to enhance the contour points before training. We resize the images to $(256, 512)$ pixels.

### 3.2 Augmentation

Image augmentation is performed to increase the number of training images and generate more dataset variations. We normalize the images based on the ImageNet standard, perform random horizontal flips and, apply random noise and gaussian blur to the images.

## 4 Methodology

### 4.1 RLUNet

#### 4.1.1 Network Architecture

The proposed segmentation model, Retinal Layer U-net (RLUNet), is based on a U-net type architecture. A vanilla U-net [2] architecture contains two modules: a downsampling encoder and an upsampling decoder. The encoder consists of multiple repetitions of two 3x3 convolution layers, each followed by a rectified linear unit (ReLu) and a 2x2 max-pooling layer. The decoder consists of an upsampling layer concatenated with feature maps from the corresponding encoder layer and a couple $3x3$ convolution layers followed by ReLu. The final layer is a $1x1$ convolution layer mapping the final feature vector to the required number of classes. The RLUNet architecture (illustrated in figure 4) uses components of the resnet-18 network pre-trained on the ImageNet dataset. The network freezes all resnet-18 layers and removes

the last layer. The structure of RLUNet encoder is 5 levels deep. The encoder adopts the frozen layers: $0-3, 3-5, 5, 6$, and 7 in its five respective levels. Each resnet block in the encoder is followed by a convolution block consisting of a $1x1$ convolution layer and ReLu activation. The RLUNet decoder also has five levels. The first four levels consist of a bicubic upsample layer, a non-padded $1x1$ convolution layer, an encoder concatenation layer of the corresponding level and, a $3x3$ dilated convolution layer with padding. Each convolution layer is followed by ReLu activation. The final level consists of an upsample layer, concatenation layer and a 3x3 convolution layer that outputs a 64-dimensional feature map. The feature map is passed onto a $1x1$ convolution layer whose output is the same size as the original input but with $n$ channels where $n$ is the number of segmentation classes. The layer is followed by a novel activation function layer called Polynomial Continuity Unit (PlCu). PlCu is defined in equation 1 and figure 5.

$$PlCu(X,Y) = \begin{cases} Y & if\ X_{i+1}-X_i=1 \quad \forall i \in len(row)-1 \\ \\ I(Y,1) & otherwise \end{cases} \tag{1}$$

where, $(X,Y) = [(x,y)_1, (x,y)_2, ...(x,y)_n]$ are the coordinates of the image, $I(t,k)$ is a spline based interpolation function where $t$ is the discrete input and $k$ is the interval.

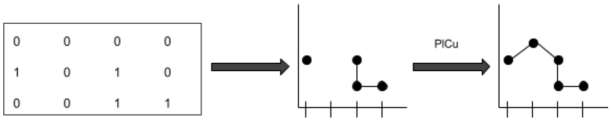
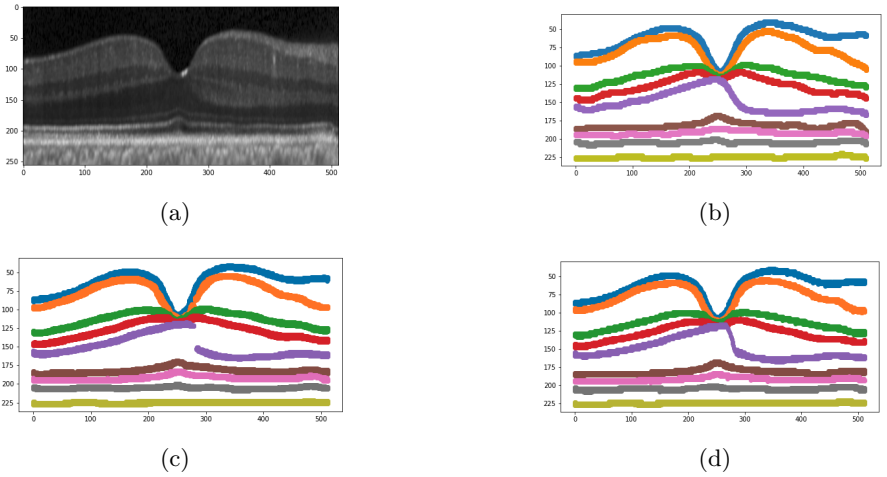



**Fig. 5**: Discrete to continuous contour conversion using PlCu. It interpolates image pixels across all the columns.

PlCu converts a discrete, discontinuous signal into a smooth continuous signal. Since retinal layer segmentation is an end-to-end contour detection problem, PlCu function helps in achieving smooth and continuous contours. For retinal layer segmentation, RLUNet without PlCu would predict discontinuous contours containing small gaps within the lines. An illustration of retinal segmentation with and without using PlCu function is shown in figure 6.

### 4.1.2 Training

For training RLUNet, we assign a separate channel for all retinal layers of interest. Each pixel in a channel can have binary values where 0 indicates background and 1 indicates a point in a contour. The training dataset goes through several preprocessing and augmentation stages described in Section 3. The network uses the Adam algorithm [36] for gradient-based optimization with adaptive learning rate (initially 1e-4). The learning rate is reduced by a certain factor if the training does not show improvement over time or, in other words, if the gradient forms a plateau.



(a) (b) (c) (d)

**Fig. 6**: (*a*) OCT scan (*a*) Ground truth contours (*c*) Without PlCu (*d*) With PlCu

The training is performed in batches of 16, over 200 epochs on average for different datasets. We use 80 samples belonging to 8 random subjects and 1029 samples belonging to 21 random subjects for training He et al.. [34] and Chiu et al.. [33] dataset, respectively. The training uses three loss functions: Binary Cross-Entropy (BCE), Dice Loss (DL) and, Intersection over Union Loss (IoU).

1. Binary Cross Entropy: BCE can be defined as the average of negative log of the corrected probabilities. The mathematical formula for multiclass BCE is shown in equation 2.

$$BCE = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{M} y_{ij} log(p_{ij}) \tag{2}$$

where, $N$ = number of rows, $M$ = number of classes

2. Dice Loss: Dice score measures the relative similarity between the predicted pixels and the ground truth, where a score of 1 means the images are perfectly alike. Dice coefficient can efficiently handle imbalanced data since it measures the per cent overlap and not the pixel overlap. Dice loss can be defined as shown in equation 3.

$$DL = 1 - \frac{2yy' + \epsilon}{y + y' + \epsilon} \tag{3}$$

where, $\epsilon$ is added to avoid division by zero cases.

3. Intersection over Union loss: IoU is a measurement metric that compares the similarity of an image to its dissimilarity. However, because IoU is not differentiable, it cannot be used as a loss function. Yu et al. in [37] have defined gradients for IoU so that it can be used to train object detection networks. IoU loss for segmentation problem is simply defined as in equation 4.

$$IoULoss = -ln\frac{Intersection(y, y')}{Union(y, y')} \tag{4}$$

The total loss is then calculated as a linear sum of the three individual losses, as shown in equation 5.

$$Loss = w_1 * BCE + w_2 * DL + w_3 * IoULoss \tag{5}$$

Here $w_1$,$w_2$ and $w_3$ are the weights or coefficients of BCE, Dice Loss and Iou loss respectively. The weights are adjusted in such as way that the value of loss stays between 0 and 1. For retinal layer segmentation, we achieved the best results when $w_1 = 0.3,\ w_2 = 0.3$ and $w_3 = 0.4$.

**Algorithm 1** Multilevel Gradients for SCR Identification (MGrSI)

**Data**: $C_{ontours} = \{ILM, ...OPL, ...BM\}$ $where$ $C.shape = (AXB)$
**Data**: $ILM = \{ilm_1, ...ilm_n\}$, $OPL = \{opl_1, ...opl_n\}$
**def Step**$_1(th_1)$
  $d \Leftarrow dist(ILM - OPL)$
  $d' \Leftarrow \frac{d}{dx}\, d'$
  $p_1 \Leftarrow maxima(d')$
  $return$ $p_1$
**def Step**$_2(th_2)$
  $m_i \Leftarrow Slope(ILM)$
  $I_p \Leftarrow \{i_{p1}, i_{p2}...\}$ $\forall m_i = 0 \pm \alpha$
  $m_o \Leftarrow Slope(OPL)$
  $O_p \Leftarrow \{o_{p1}, o_{p2}...\}$ $\forall m_o = 0 \pm \alpha$
  $p_2 \Leftarrow I_p \cap O_p$
  $return$ $p_2$
**def Step**$_3(th_3)$
  $C \Leftarrow C.resize((\frac{A}{2}, \frac{B}{2}))$
  $p_3 \Leftarrow Step_1(th_3)$
  $return$ $p_3$
**def Step**$_4(th_4)$
  $p_4 \Leftarrow Step_2(th_4)$
  $return$ $p_4$
**main()**
  $p_{final} \Leftarrow (p_1 \cap p_2 \cap p_3 \cap p_4)$

### 4.2 Sickle cell retinopathy detection

Manually, SCR is detected by carefully studying the change in thickness between the ILM and OPL layers of a B-scan. The distance between ILM and OPL layers is thinner in an affected area than in the adjacent non-affected areas. To detect this pathology using computer vision, we develop an algorithm - Multilevel Gradients for SCR Identification (MGrSI). This method analyzes the contour maps of images obtained after running inference with a trained RLUNet. Our algorithm can be divided into multiple steps, as shown in our pseudo code 1. First, we calculate the distance ($d$) between the ILM and OPL layers across the image width (line 3). Then, we calculate the rate of change in distance ($d'$) and the peaks or maxima ($p_1$) using a generous threshold ($th_1$). Points ($p_1$) provides the affected areas for step 1. Then, in the second step (line 8), we analyze the ILM and OPL layers individually. If the retina is injured due to SCR, the ILM layer should concave upwards, and the OPL layer should concave downwards at the same point/s along the X-axis. Hence we determine the point/s in ILM ($I_p = \{i_{p1}, i_{p2}...\}$) and OPL ($O_p = \{o_{p1}, o_{p2}...\}$) where the slope are equal to zero (i.e., $m = 0 \pm \alpha$, where $\alpha$ is a small number used for thresholding leniency). The intersection of these points ($(p_2) = I_p \cap O_p$) give us the affected areas for step 2. For the third (line 15) and the fourth (line 19) steps, we decrease the image resolution by 50% and repeat the first two steps but, this time with a smaller threshold. This gives us the affected areas ($p_3$) and ($p_4$). We then consider the intersection of all four sets of points (with a certain threshold) to get our final results ($P$). A pseudo-code for our algorithm is shown in reference 1.

## 5 Experiments and Results

We comprehensively evaluated our segmentation method on the two benchmarked datasets described in Section 3. We used multiple metrics such as Intersection over Union (IoU), Confusion matrix and, Dice coefficient to analyze the performance of our method. We have already discussed IoU and Dice coefficient in section 4.1.2. So, we will discuss the confusion matrix here.

A Confusion matrix is a table that summarizes the performance of a supervised algorithm based on True Positives (TP), True Negatives (TN), False Positive (FP) and, False Negatives (FN). This matrix is used to visualize important performance metrics like recall, specificity, accuracy, and precision, each of which are mathematically defined as:

$$Recall(R) = \frac{TP}{TP + FN} \quad Specificity(S) = \frac{TN}{TN + FP}$$
$$Accuracy(A) = \frac{TP + TN}{TP + TN + FP + FN} \quad Precision(P) = \frac{TP}{TP + FP}$$

We also plot receiver operating characteristic (ROC) and detection error tradeoff (DET) curves for our test results. ROC curves [38] project true positive rate on the Y-axis and false positive rate on the X-axis. They are generally used in binary classification to study the output of a classifier. In our case, we get binarized output for each class. Hence, ROC can be directly used. DET curves project false negative rate on the Y-axis and false positive rate on the X-axis. DET curves [39]

helps to deduct the rate at which false-negative error will improve when willing to accept an increase in false-positive error rate (or vice-versa).

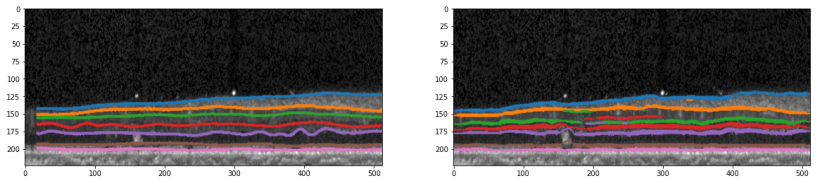
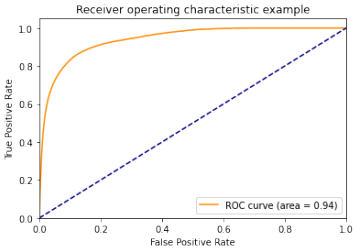
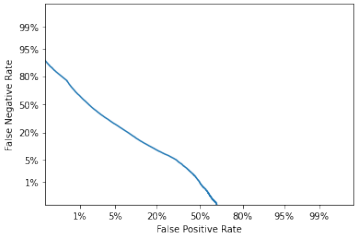


**Fig. 7**: Retinal layer segmentation in Chiu et al. dataset. Left: Ground truth Right: Segmentation from RLUNet

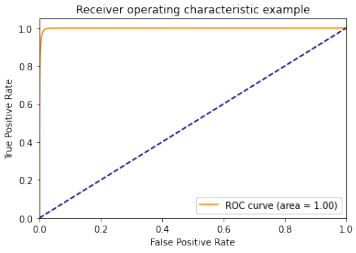
To evaluate RLUNet on He et al.. dataset, we used 20 B-scans taken from 2 randomly selected subjects. Upon running the tests, we got an average IoU of 0.58 and a dice coefficient of 0.66. Since this dataset has only 110 samples, we used 80 images for training, so our network suffered from overfitting, resulting in moderate performance. The quantitative and qualitative results of our experiment on this dataset are presented in table 1 and figure 7 respectively. Similarly, the ROC and DET curves are shown in figure 8a and figure 8b.



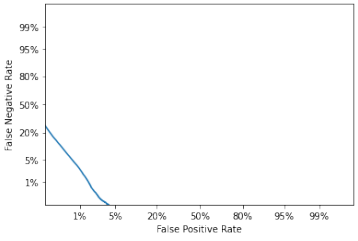

(a) Area under the curve (AUC) = 0.94




(b) DET curve




(c) Area under the curve (AUC) = 1.0




(d) DET curve

**Fig. 8**: ROC and DET curves for He et al. (top) and Chiu et al. (bottom) datasets.

**Table 1**: Evaluation table

| | Confusion Matrix Rates (%) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Dataset | TPR | TNR | FPR | FNR | R | S | A | P | IoU | Dice |
| DU | 70 | 99 | 0.4 | 30 | 0.51 | 0.99 | 0.98 | 0.5 | 0.58 | 0.66 |
| JHU | 80 | 99 | 0.2 | 21 | 0.85 | 0.99 | 0.99 | 0.80 | 0.72 | 0.83 |
| SCR[1] | 75 | 98 | 0.4 | 24 | 0.75 | 0.98 | 0.98 | 0.68 | - | - |

Note: The table presents the test results of retinal layer segmentation for the three datasets: [33], [34] and, Nemours for SCR. Here R, S, A and P refer to Recall, Specificity, Accuracy and Precision, respectively.

[1]Since TN cannot be objectively determined in this case, we considered TN to be the percentage of data that is neither TP, FP or FN.

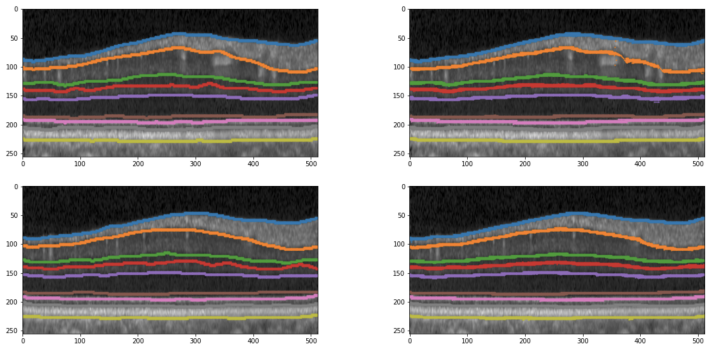
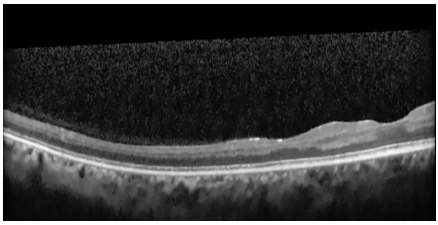


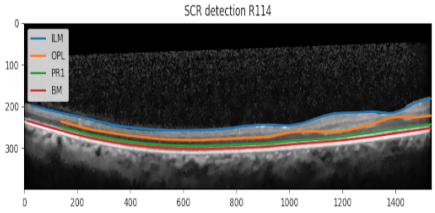
**Fig. 9**: Retinal layer segmentation in He et al. dataset. Left: Ground truth Right: Segmentation from RLUNet

To evaluate Chiu et al.. dataset, we use 686 B-scans taken from 15 randomly selected subjects. Some examples of retinal layer segmentation using our trained model are illustrated in figure 9. Applying RLUNet, we achieved an IoU score of 0.83 and a Dice coefficient of 0.72. An overall evaluation of our segmentation model is depicted in table 1. The related curves are shown in figure 8c, 8d.

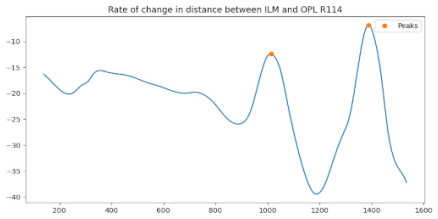
We tested the Nemours dataset for SCR detection and localization. The detection process outputs whether a particular B-scan is convincingly affected by SCR (as determined by an expert in this field) and what parts or sections within the B-scan show signs of the pathology. We localize the pathology visually by superimposing a couple of vertical lines enclosing the affected area for our evaluations. Some images of our results are shown in figure 10. Our expert ophthalmologist prepared the ground truth values of SCR localization. Our expert also manually evaluated the results from our method to see which ones were correct and which ones were wrong since it was impossible to determine the affected area by ourselves objectively. Based on this, we calculated the confusion matrix and other evaluation statistics for SCR localization as illustrated in table 1.



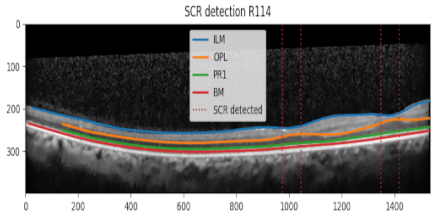
(a) Original image




(b) ILM, OPL, PR1 and BM contours




(c) Rate of change in distance and peak detection (Step 3)




(d) vertical lines enclosing area of injury

**Fig. 10**: Results and evaluations on Nemours dataset

# 6 Future work

The quantitative and qualitative evaluations of our algorithms show some promising performance. However, with more research and evaluations, we can obtain competitive scores on par with the current state of the art. In the case of retinal layer segmentation, exploring more datasets designed for multiple retinal pathologies could be an excellent first step. Our RLUNet architecture could also benefit from parameter tuning and k-fold training.

The future of this research is directed towards developing deep learning solutions on SCR detection and surveillance. Our work can be extended to monitor the progression of retinal pathology by processing repeated OCT scans from patients during follow up examinations. Additionally, we are looking forward to advancing our research by combining two retinal imaging modalities: retinal microanatomy observed on OCT and retinal circulation on angiography. The goal of this novel research is to employ deep learning techniques to reliably detect retinal anomalies and improve our understanding on SCR pathophysiology.

# References

[1] Gong, X.-Y., Su, H., Xu, D., Zhang, Z.-T., Shen, F., Yang, H.-B.: An overview of contour detection approaches. Int. J. Autom. Comput. **15**(6), 656–672 (2018). https://doi.org/10.1007/s11633-018-1117-z

[2] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation (2015)

[3] Wang, S., Hu, S.-Y., Cheah, E., Wang, X., Wang, J., Chen, L., Baikpour, M., Ozturk, A., Li, Q., Chou, S.-H., Lehman, C.D., Kumar, V., Samir, A.: U-Net Using Stacked Dilated Convolutions for Medical Image Segmentation (2020)

[4] Singh, R.P., Gupta, S., Acharya, U.R.: Segmentation of prostate contours for automated diagnosis using ultrasound images: A survey. J. Comput. Sci. **21**, 223–231 (2017)

[5] Zotin, A., Simonov, K., Kurako, M., Hamad, Y., Kirillova, S.: Edge detection in mri brain tumor images based on fuzzy c-means clustering. Procedia Computer Science **126**, 1261–1270 (2018). https://doi.org/10.1016/j.procs.2018.08.069. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 22nd International Conference, KES-2018, Belgrade, Serbia

[6] Fujimoto, J.G., Pitris, C., Boppart, S.A., Brezinski, M.E.: Optical coherence tomography: An emerging technology for biomedical imaging and optical biopsy. Neoplasia **2**(1), 9–25 (2000). https://doi.org/10.1038/sj.neo.7900071

[7] Li, M., Kambhamettu, C., Stone, M.: Automatic contour tracking in ultrasound images. Clinical Linguistics & Phonetics **19**(6-7), 545–554 (2005) https://doi.org/10.1080/02699200500113616. https://doi.org/10.1080/02699200500113616. PMID: 16206482

[8] Kang, X., Kong, Q., Zeng, Y., Xu, B.: A fast contour detection model inspired by biological mechanisms in primary vision system. Frontiers in Computational Neuroscience **12**, 28 (2018). https://doi.org/10.3389/fncom.2018.00028

[9] Dollar, P., Tu, Z., Belongie, S.: Supervised learning of edges and object boundaries. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 2, pp. 1964–1971 (2006). https://doi.org/10.1109/CVPR.2006.298

[10] Catanzaro, B., Su, B.-Y., Sundaram, N., Lee, Y., Murphy, M., Keutzer, K.: Efficient, high-quality image contour detection. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 2381–2388 (2009). https://doi.org/10.1109/ICCV.2009.5459410

[11] Ren, X., Bo, L.: Discriminatively trained sparse code gradients for contour detection. In: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. NIPS'12, pp. 584–592 (2012)

[12] Rupprecht, C., Huaroc, E., Baust, M., Navab, N.: Deep Active Contours

(2016)

[13] Kokkinos, I.: Pushing the Boundaries of Boundary Detection using Deep Learning (2016)

[14] Yang, J., Price, B., Cohen, S., Lee, H., Yang, M.-H.: Object Contour Detection with a Fully Convolutional Encoder-Decoder Network (2016)

[15] Deng, R., Shen, C., Liu, S., Wang, H., Liu, X.: Learning to predict crisp boundaries (2018)

[16] Giovinco, G., Savastano, M.C., Ventre, S., Tamburrino, A.: Automated detection of the retinal from OCT spectral domain images of healthy eyes. Journal of Modern Optics **62**(10), 837–850 (2015)

[17] Sun, Y., Zhang, T., Zhao, Y., He, Y.: 3d automatic segmentation method for retinal optical coherence tomography volume data using boundary surface enhancement. Journal of Innovative Optical Health Sciences **09**(02), 1650008 (2016) https://doi.org/10.1142/S1793545816500085. https://doi.org/10.1142/S1793545816500085

[18] Tian, J., Varga, B., Somfai, G., Lee, W., Smiddy, W., DeBuc, D.: Real-time automatic segmentation of optical coherence tomography volume data of the macular region. PLoS One **10**(8) (2015). https://doi.org/10.1371/journal.pone.0133908

[19] Chiu, S.J., Li, X.T., Nicholas, P., Toth, C.A., Izatt, J.A., Farsiu, S.: Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation. Opt. Express **18**(18), 19413–19428 (2010). https://doi.org/10.1364/OE.18.019413

[20] Lang, A., Carass, A., Hauser, M., Sotirchos, E.S., Calabresi, P.A., Ying, H.S., Prince, J.L.: Retinal layer segmentation of macular oct images using boundary classification. Biomed. Opt. Express **4**(7), 1133–1152 (2013). https://doi.org/10.1364/BOE.4.001133

[21] Liu, Y., Carass, A., He, Y., Antony, B., Filippatou, A., Saidha, S., Solomon, S., Calabresi, P., Prince, J.: Layer boundary evolution method for macular oct layer segmentation. Biomedical Optics Express **10**(3), 1064–1080 (2019). https://doi.org/10.1364/BOE.10.001064

[22] Hsia, W.P., Tse, S.L., Chang, C.J., Huang, Y.L.: Automatic segmentation of choroid layer using deep learning on spectral domain optical coherence tomography. Applied Sciences **11**(12) (2021). https://doi.org/10.3390/app11125488

[23] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN (2018)

[24] Devalla, S.K., Renukanand, P.K., Sreedhar, B.K., Subramanian, G., Zhang, L., Perera, S., Mari, J.-M., Chin, K.S., Tun, T.A., Strouthidis, N.G., Aung, T., Thiéry, A.H., Girard, M.J.A.: Drunet: a dilated-residual u-net deep learning network to segment optic nerve head tissues in optical coherence tomography images. Biomedical optics express **9**(7), 3244–3265 (2018). https://doi.org/10.1364/boe.9.003244

[25] Li, Q., Li, S., He, Z., Guan, H., Chen, R., Xu, Y., Wang, T., Qi, S., Mei, J., Wang, W.: DeepRetina: Layer Segmentation of Retina in OCT Images Using Deep Learning. Translational Vision Science & Technology **9**(2), 61–61 (2020)

[26] Kugelman, J., Alonso-Caneiro, D., Read, S.A., Hamwood, J., Vincent, S.J., Chen, F.K., Collins, M.J.: Automatic choroidal segmentation in oct images using supervised deep learning methods. Scientific Reports **9** (2019)

[27] Gopinath, K., Rangrej, S.B., Sivaswamy, J.: A deep learning framework for segmentation of retinal layers from oct images. In: 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), pp. 888–893 (2017). https://doi.org/10.1109/ACPR.2017.121

[28] Ngo, L., Cha, J., Han, J.-H.: Deep neural network regression for automated retinal layer segmentation in optical coherence tomography images. IEEE Transactions on Image Processing **29**, 303–312 (2020). https://doi.org/10.1109/TIP.2019.2931461

[29] Hamwood, J., Alonso-Caneiro, D., Read, S.A., Vincent, S.J., Collins, M.J.: Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of oct retinal layers. Biomed. Opt. Express **9**(7), 3049–3066 (2018). https://doi.org/10.1364/BOE.9.003049

[30] Yanagihara, R.T., Lee, C.S., Ting, D.S.W., Lee, A.Y.: Methodological Challenges of Deep Learning in Optical Coherence Tomography for Retinal Diseases: A Review. Translational Vision Science & Technology **9**(2), 11–11 (2020) https://arvojournals.org/arvo/content_public/journal/tvst/938366/i2164-2591-225-2-1873.pdf. https://doi.org/10.1167/tvst.9.2.11

[31] Kugelman, J., Alonso-Caneiro, D., Read, S.A., Vincent, S.J., Chen, F.K., Collins, M.J.: Effect of altered oct image quality on deep learning boundary segmentation. IEEE Access **8**, 43537–43553 (2020). https://doi.org/10.1109/ACCESS.2020.2977355

[32] Maloca, P.M., Müller, P.L., Lee, A.Y., Tufail, A., Balaskas, K., Niklaus, S., Kaiser, P., Suter, S.K., Zarranz-Ventura, J., Egan, C.A., Scholl, H.P.N., Schnitzer, T.K., Singer, T., Hasler, P.W., Denk, N.: Unraveling the deep learning gearbox in optical coherence tomography image segmentation towards explainable artificial intelligence. Communications Biology **4** (2021)

[33] Chiu, S.J., Allingham, M.J., Mettu, P.S., Cousins, S.W., Izatt, J.A., Farsiu, S.: Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. Biomed. Opt. Express **6**(4), 1172–1194 (2015). https://doi.org/10.1364/BOE.6.001172

[34] He, Y., Carass, A., Solomon, S.D., Saidha, S., Calabresi, P.A., Prince, J.L.: Retinal layer parcellation of optical coherence tomography images: Data resource for multiple sclerosis and healthy controls. Data in brief **22**, 601–604 (2019)

[35] Buades, A., Coll, B., Morel, J.-M.: Non-Local Means Denoising. Image Processing On Line **1**, 208–212 (2011). https://doi.org/10.5201/ipol.2011.bcm_nlm

[36] Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization (2017)

[37] Yu, J., Jiang, Y., Wang, Z., Cao, Z., Huang, T.: Unitbox: An advanced object detection network. Proceedings of the 24th ACM international conference on Multimedia (2016). https://doi.org/10.1145/2964284.2967274

[38] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Receiver Operating Characteristic (ROC). scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html

[39] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Detection error tradeoff (DET) curve. scikit-learn.org/stable/auto_examples/model_selection/plot_det.html