



*Research article*

## Weakly supervised training for eye fundus lesion segmentation in patients with diabetic retinopathy

Yu Li<sup>1,\*</sup>, Meilong Zhu<sup>1</sup>, Guangmin Sun<sup>1,\*</sup>, Jiayang Chen<sup>1,2</sup>, Xiaorong Zhu<sup>3,4</sup> and Jinkui Yang<sup>3,4</sup>

<sup>1</sup> Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

<sup>2</sup> School of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China;

<sup>3</sup> Beijing Tongren Hospital, Beijing 100730, China

<sup>4</sup> Beijing Institute of Diabetes Research, Beijing 100730, China

\* **Correspondence:** Email: [yuli@bjut.edu.cn](mailto:yuli@bjut.edu.cn), [gmsun@bjut.edu.cn](mailto:gmsun@bjut.edu.cn).

**Abstract:** *Objective:* Diabetic retinopathy is the leading cause of vision loss in working-age adults. Early screening and diagnosis can help to facilitate subsequent treatment and prevent vision loss. Deep learning has been applied in various fields of medical identification. However, current deep learning-based lesion segmentation techniques rely on a large amount of pixel-level labeled ground truth data, which limits their performance and application. In this work, we present a weakly supervised deep learning framework for eye fundus lesion segmentation in patients with diabetic retinopathy. *Methods:* First, an efficient segmentation algorithm based on grayscale and morphological features is proposed for rapid coarse segmentation of lesions. Then, a deep learning model named Residual-Attention Unet (RAUNet) is proposed for eye fundus lesion segmentation. Finally, a data sample of fundus images with labeled lesions and unlabeled images with coarse segmentation results is jointly used to train RAUNet to broaden the diversity of lesion samples and increase the robustness of the segmentation model. *Results:* A dataset containing 582 fundus images with labels verified by doctors, including hemorrhage (HE), microaneurysm (MA), hard exudate (EX) and soft exudate (SE), and 903 images without labels was used to evaluate the model. In ablation test, the proposed RAUNet achieved the highest intersection over union (IOU) on the labeled dataset, and the proposed attention and residual modules both improved the IOU of the UNet benchmark. Using both the images labeled by doctors and the proposed coarse segmentation method, the weakly supervised framework based on RAUNet architecture significantly improved the mean segmentation accuracy by over 7% on the lesions. *Significance:* This study demonstrates that combining unlabeled medical images with coarse

segmentation results can effectively improve the robustness of the lesion segmentation model and proposes a practical framework for improving the performance of medical image segmentation given limited labeled data samples.

**Keywords:** diabetic retinopathy; deep learning; fundus image; lesion segmentation; weak supervision

---

## 1. Introduction

Diabetic retinopathy (DR) is a kind of specific change in fundus lesions and is a serious and syndromic manifestation of diabetes. Recent studies have also shown that DR affects not only the neural retina but also retinal pigment epithelium (RPE) cells in the outer layer of the retina, including changes in melanosomes and lipofuscin granules in RPE cytochrome granules [1,2]. According to a clinical meta-eye study [3], DR is a serious threat to human vision, leading to a large amount of blindness. Therefore, regular detection of retinopathy is of great significance for the treatment of patients with diabetic retinopathy. However, manual screening for diabetic retinopathy faces challenges such as an imbalance in the doctor-patient ratio and the varying experience of ophthalmologists.

The clinical manifestations of DR mainly include microaneurysm (MA), hemorrhage (HE), hard exudate (EX), soft exudate (SE) and other lesions [4]. Studies suggest that the accurate segmentation of key lesions plays a crucial role in the early detection and diagnosis of DR [5–7]. In recent years, with the continuous development of computer vision technology, more image processing algorithms have been widely used in analyzing medical images for detection, segmentation and classification tasks [8–13]. Shankar et al. applied histogram-based segmentation to fundus images and proposed a synergic deep learning model to grade the severity levels of DR [14]. Hire et al. proposed an exudate segmentation method based on ant colony optimization [15]. Imani et al. used a morphological component analysis algorithm to separate lesions from vessels and detect exudate regions [16]. However, the complexity of lesion appearance and the interference of vessels, noise and imaging artifacts challenge the robustness of previous lesion segmentation algorithms.

Recent deep learning technology development has further improved the performance of fundus segmentation and classification tasks, providing key information for the diagnosis of fundus diseases [17–19]. Tavakoli et al. studied the automatic detection of MA using a combination of a matching-based approach and a deep learning model and compared the effect of two different image preprocessing methods [20]. Yu et al. proposed an end-to-end deep semantic edge learning image segmentation architecture based on ResNet and a skip-layer architecture to address the problem of edge pixels belonging to more than one semantic class [21]. Mo et al. proposed cascaded deep residual networks that fuse multilevel hierarchical information to segment exudates accurately and efficiently to recognize DME [22]. Chen et al. proposed an encoder-decoder architecture with atrous separable convolution for semantic image segmentation, resulting in a faster and stronger segmentation result. Nevertheless, at this time, the number of fundus images of patients with pixel-level lesion markers is relatively insufficient, and the existing training samples cannot fully reflect the diversity of practical datasets, limiting the segmentation accuracy of a deep learning model to a certain extent. Therefore, methods are required to fully take advantage of the large amount of diverse unlabeled fundus images and improve the robustness of current deep learning-based fundus lesion segmentation

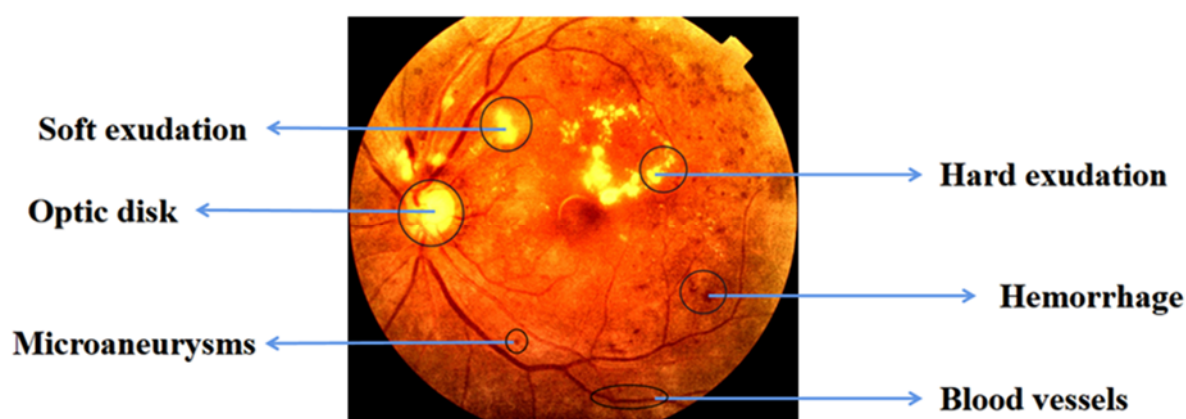
algorithms [12,13].

In this study, we first designed an efficient lesion segmentation algorithm based on the grayscale and morphological features of fundus lesions and used this algorithm to perform coarse segmentation on unlabeled fundus datasets to expand the number of fundus image samples for training a fine lesion segmentation model based on deep learning. Then, we developed a UNet-based deep learning model that applied the residual structure and attention mechanism for fundus lesion segmentation. The performance of the model was improved by weakly supervised learning, which takes advantage of both a labeled preprocessed fundus image and an unlabeled fundus image with coarse segmentation results.

## 2. Methodology

### 2.1. Characteristics of fundus lesions

The fundus images used in our study were taken from patients with retinopathy using color fundus photography with a relatively complete fundus structure. In these images, MA usually appears as dark red round dots with distinct boundaries. HE refers to the leakage of blood from abnormal vessels, also mostly in the form of dark red spots, except some in the form of flames. The shape and size are irregular, and the density within the area is not uniform. The HE severity varies with its location. EX is the leakage of lipoprotein and other substances from abnormal blood vessels. It is usually bright yellow or yellowish-white and is usually clumped and ring-shaped, with obvious boundaries. SE is caused by fiber layer ischemia of the retinal nerve, which is white or yellow-white cotton-flocculent with a fuzzy boundary, also known as cotton wool spots [23]. Typical examples of the four lesions on a fundus image are shown in Figure 1.



**Figure 1.** Diagram of fundus retinopathy.

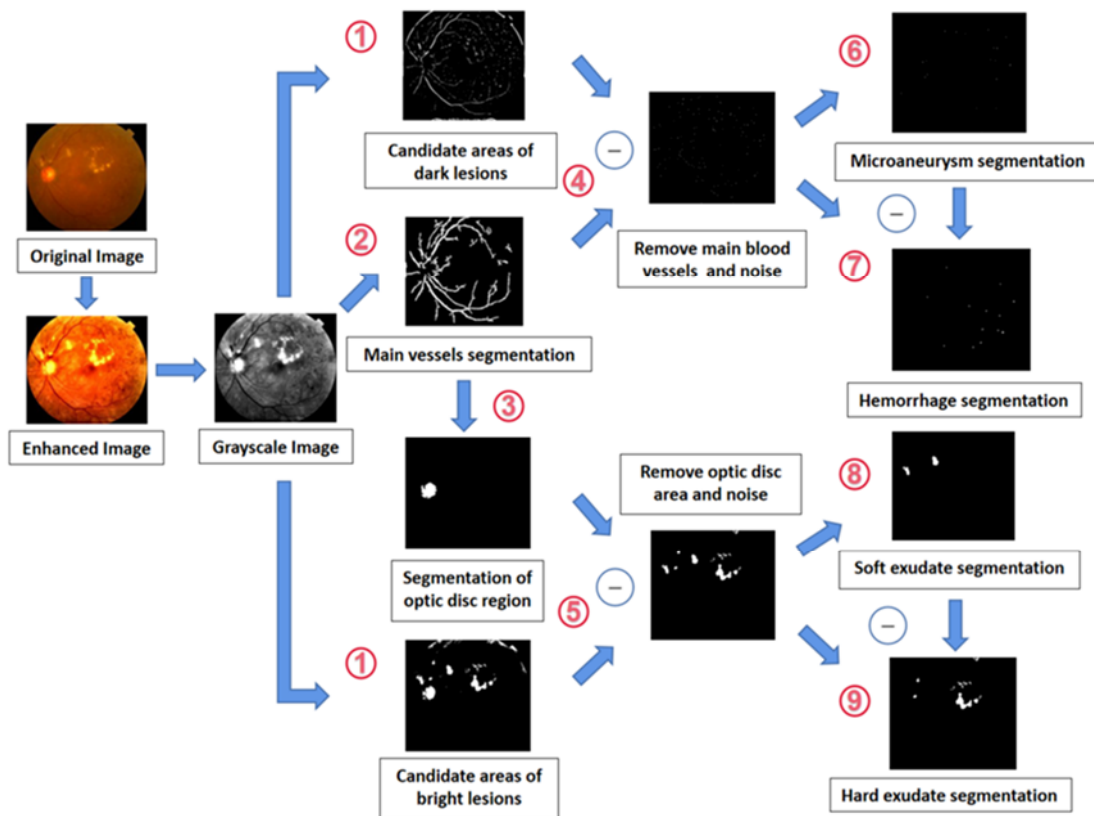
### 2.2. Image preprocessing

Each fundus image was first enhanced to improve the detectability of lesions. Based on the traditional adaptive histogram equalization (CLAHE) algorithm [24], which is widely used in the preprocessing of image segmentation tasks [25], algorithms including gamma correction, contrast

transformation enhancement, intensity range adjustment, etc. were applied to enhance the local details and suppress noise in each fundus image.

### 2.3. Lesion coarse segmentation algorithm based on grayscale and morphological features

Since the morphology and brightness features carry most lesion information, in this study, we use the grayscale image as the input for the proposed segmentation algorithm, which can not only retain the gradient information but also reduce the feature dimension. Fundus lesions can be classified into dark lesions and bright lesions according to their grayscale characteristics in enhanced fundus images [26]. The candidate areas of dark lesions mainly include HE and MA, as well as some fundus blood vessels and image noise/artifacts. The candidate areas of bright lesions mainly include EX and SE, as well as the optic disc region and noise/artifacts. The proposed lesion segmentation process is shown in Figure 2. The numbers in the diagram represent the key steps, which are described below.



**Figure 2.** Flow chart of the lesion segmentation algorithm.

1) Dual thresholds (TL and TH) were used to perform preliminary segmentation on the preprocessed grayscale images to obtain the candidate regions of dark and bright lesions, respectively:

$$f(x, y)_{rc} = \begin{cases} 255, & f(x, y) < TL \\ 0, & f(x, y) > TH \end{cases} \quad (1)$$



$$f(x, y)_{wc} = \begin{cases} 255, & f(x, y) > TH \\ 0, & f(x, y) < TH \end{cases} \quad (2)$$

where  $f(x, y)_{rc}$  is a candidate dark lesion region and  $f(x, y)_{wc}$  is a candidate bright lesion region. 2) A threshold (TB) was selected to binarize the grayscale image to extract all blood vessels and their approximate areas in the fundus image by:

$$f(x, y)_{bc} = \begin{cases} 255, & f(x, y) < TB \\ 0, & f(x, y) > TB \end{cases} \quad (3)$$

Then, multiple sizes of kernels (MK) were selected to carry out continuous opening and closing morphological operations on the binarized images to connect the small breakpoints in the blood vessels and remove the discrete noise pixels:

The erosion operator [27]:

$$[\varepsilon_B(X)](x) = \min \{X_B\} \quad (4)$$

and the dilation operator [27]:

$$[\delta_B(X)](x) = \max \{X_B\} \quad (5)$$

where  $[\varepsilon_B(X)](x)$  represents the structural element  $B$  that corrodes from a child element  $x$  of set  $X$ ,  $[\delta_B(X)](x)$  represents the expansion operation of structural element  $B$  on a child element  $x$  in set  $X$ , and  $X_B$  represents the  $x$  value in the structural element  $B$ .

The opening operation [27]:

$$g(x, y)_o = \text{open}[f(x, y), B] = \text{dilate}\{\text{erode}[f(x, y), B], B\} \quad (6)$$

and the closing operation [27]:

$$g(x, y)_c = \text{close}[f(x, y), B] = \text{erode}\{\text{dilate}[f(x, y), B], B\} \quad (7)$$

Finally, to obtain more accurate blood vessel segmentation results, the center point of each remaining region was selected for polynomial fitting verification:

Polynomial fitting verification [28]:

$$\Delta = |y_i - y_{x_i}| = |y_i - (a_0 + a_1x_i + a_2x_i^2)| \leq L \quad (8)$$

where  $\Delta$  represents the offset,  $x_i$  and  $y_i$  represent the central coordinates of each region,  $y_{x_i}$  represents the output of the polynomial fitting verification with  $x_i$  as its input,  $a_0$ ,  $a_1$ ,  $a_2$  represent three constants in the polynomial fitting verification, and  $L$  represents the maximum allowable offset. The objective function of polynomial fitting is minimizing the mean square error (MSE) [29]:

$$Q(a_0, a_1, a_2) = \sum_{i=1}^m (y(x_i) - y_i)^2 = \sum_{i=1}^m (a_0 + a_1x_i + a_2x_i^2 - y_i)^2 \quad (9)$$

3) The approximate convergence point of the blood vessels in the image was obtained by calculating the horizontal and vertical histograms:

$$(x, y)_{optic} = (\max_i \{\sum_{j=1}^a p_j\}, \max_j \{\sum_{i=1}^b p_i\}) \quad (10)$$

where  $i$  and  $j$  represent the row and column of the image, respectively,  $a$  and  $b$  represent the number of pixels in each row and column, respectively, and  $p$  represents the pixel value.

Then, to guarantee that in fundus images with different angles and directions, the proportion of space occupied by the optic disc is always fixed, we took the convergence point as the center and an empirical percent value ( $P\%$ ) of the image resolution as the radius to draw a circular region as the candidate location region of the optic disc.

$$h(x, y)_{circle} = \begin{cases} 255, & \sqrt{(y_i - y_c)^2 + (x_i - x_c)^2} \leq p\% * resolution \\ 0, & \sqrt{(y_i - y_c)^2 + (x_i - x_c)^2} > p\% * resolution \end{cases} \quad (11)$$

where  $h(x, y)_{circle}$  is a circle candidate region of the optic disc,  $x_i, y_i$  represent the coordinates of each pixel in the image, and  $x_c, y_c$  represent the coordinates of the convergence point.

Finally, an AND operation was performed to obtain the final segmentation results of the optic disc region:

$$g(x, y)_{disc} = f(x, y)_{wc} \wedge h(x, y)_{circle} \quad (12)$$

where  $h(x, y)_{circle}$  is the circle candidate region of the optic disc.

4) The main blood vessels in the dark lesion candidate region were extracted and removed using a vessel segmentation algorithm. The noise was suppressed and removed through a morphological operation, including expansion and corrosion, connected domain calculations, including area calculations, contour calculations and edge processing.

5) The optic disc in the bright lesion candidate region was removed using an optic disc segmentation algorithm, and the noise was also reduced through a morphological operation, connected domain calculation and edge processing methods.

6) After acquiring each lesion's connected domains in an image, the area of the connected domains and the coordinates of the four vertices of the minimum enclosing rectangle were calculated. The screening and identification of MA were realized based on three limiting parameters. The first two limiting parameters filtered the shape of the candidate connected domain based on the characteristic that MA is mostly round or oval, and the third limiting parameter filtered the candidate area of the connected domain based on the assumption that MA is usually small.

The 1<sup>st</sup> limiting parameter:

$$contours(i) = \begin{cases} remain, & ARA = \frac{h_i}{w_i} < \lambda \\ remove, & ARA = \frac{h_i}{w_i} > \lambda \end{cases} \quad (13)$$

The 2<sup>nd</sup> limiting parameter:

$$contours(i) = \begin{cases} remain, & RA = \frac{area\ of\ A_i}{area\ of\ B_i} < \mu \\ remove, & RA = \frac{area\ of\ A_i}{area\ of\ B_i} > \mu \end{cases} \quad (14)$$

The 3<sup>rd</sup> limiting parameter:

$$ontours(i) = \begin{cases} remain, & \alpha < I = area\ of\ contours(i) < \beta \\ remove, & else \end{cases} \quad (15)$$

where ARA is the aspect ratio of the area, RA is the ratio of the area,  $I$  is a closed interval,  $\lambda, \mu, \alpha, \beta$  are parameters,  $h$  and  $w$  represent the length and width of each connected domain, respectively,  $A_i$  represents the area of the circular region drawn with the center of each connected domain as the midpoint and the mean distance between the center point and each contour as the radius, and  $B_i$  represents the area of each connected domain.

Finally, the remaining connected domains were regarded as the MA segmentation results:

$$g(x, y)_{MA} = contours(remain) \quad (16)$$

7) HE identification and segmentation was realized using an XOR operation:

$$g(x, y)_{HE} = f(x, y)_{rc} \oplus g(x, y)_{MA} \quad (17)$$

8) Two groups of threshold values ( $TG1(TG1A, TG1A^)$  and  $TG2(TG2B, TG2B^)$ ) were selected to obtain the difference in the segmentation edges:

$$f(x, y)_{gt1} = \begin{cases} 255, & f(x, y) > TG1A \\ 0, & f(x, y) < TG1A \end{cases} \quad (18)$$

$$f(x, y)_{gt1^} = \begin{cases} 255, & f(x, y) > TG1A^ \\ 0, & f(x, y) < TG1A^ \end{cases} \quad (19)$$

$$f(x, y)_{gt2} = \begin{cases} 255, & f(x, y) > TG2B \\ 0, & f(x, y) < TG2B \end{cases} \quad (20)$$

$$f(x, y)_{gt2^} = \begin{cases} 255, & f(x, y) > TG2B^ \\ 0, & f(x, y) < TG2B^ \end{cases} \quad (21)$$

$$f(x, y)_{dif} = (f(x, y)_{gt1} \oplus f(x, y)_{gt1^}) \oplus (f(x, y)_{gt2} \oplus f(x, y)_{gt2^}) \quad (22)$$

After that, an XOR operation was performed to obtain the SE contours:

$$f(x, y)_{cse} = f(x, y)_{wc} \oplus f(x, y)_{dif} \quad (23)$$

Finally, the SE lesion segmentation result was obtained by filling the inner regions of the contours:

$$g(x, y)_{SE} = \begin{cases} 255, & f(x, y) \in f(x, y)_{cse} \\ 0, & f(x, y) \notin f(x, y)_{cse} \end{cases} \quad (24)$$

9) The EX lesion segmentation result was obtained by conducting an XOR operation:

$$g(x, y)_{EX} = f(x, y)_{wc} \oplus g(x, y)_{SE} \quad (25)$$

## 2.4. Lesion segmentation based on deep learning

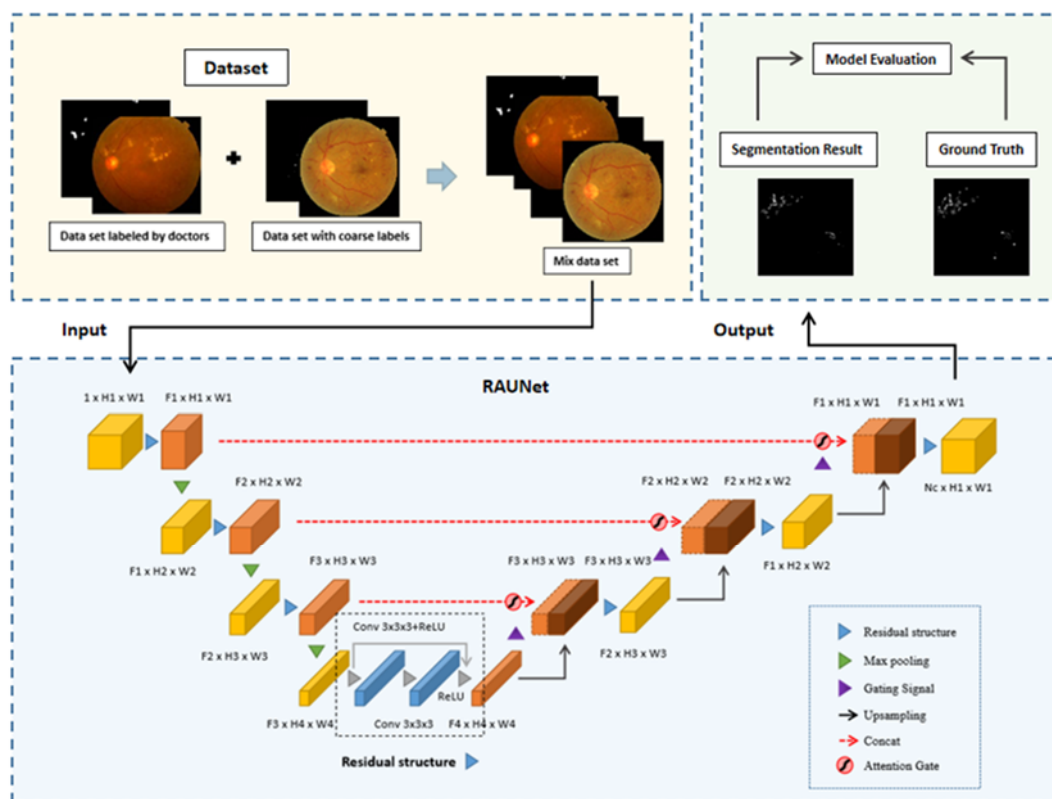
### 2.4.1 Framework of RAUNet

UNet is a conventional fully convolutional network for semantic segmentation, which was originally proposed for medical image segmentation [30]. UNet has a large number of feature channels in the upsampling part, which allows the network to propagate context information to higher resolution layers. By supplying the expansive path with finer spatial information, skip connections are applied to

boost UNet's segmentation accuracy of target borders and improve gradient flow. Therefore, UNet can be trained on fewer data samples [31]. The encoder-decoder architecture and skip connection in UNet also help to capture multiscale information in images [32]. To further boost the training process of the deep neural network, we introduced a residual module to better solve the problem of network degradation, avoid gradient dispersion and improve the network fitting ability. At the same time, to better allocate the limited information processing resources to the important parts of the model and improve the detection ability of small target lesions, we introduced an attention mechanism. The proposed deep learning-based lesion segmentation network model was named Residual-Attention UNet (RAUNet).

#### 2.4.2 Finetuning the segmentation model based on weakly supervised training

To improve the robustness of the segmentation model, in this study, we proposed a framework to fine-tune the deep learning model by jointly using the labeled dataset and data with coarse labels obtained using the proposed grayscale and morphological feature-based segmentation method. Generally, weak supervision includes incomplete supervision, inexact supervision and inaccurate supervision [33]. The proposed training framework takes a mixed dataset as the input to RAUNet and fine-tunes the weights of the whole network through batch iteration using the Adam optimizer. Therefore, it can be considered a weak supervision structure. The hyperparameters of the deep learning model, including training duration, learning rate, step size, etc., were manually optimized to achieve better model robustness and generalization. The specific training process is shown in Figure 3.



**Figure 3.** Flow chart of the weakly supervised training-based segmentation model.

### 2.4.3 Evaluation indices

The accuracy and intersection over union (IOU) were selected as the two main evaluation indices for lesion segmentation:

$$\text{Accuracy} = (\text{Area of (prediction)} \cap \text{Area of (label)}) / \text{Area of (label)} \quad (26)$$

$$\text{IOU} = (\text{Area of (prediction)} \cap \text{Area of (label)}) / (\text{Area of (prediction)} \cup \text{Area of (label)}) \quad (27)$$

The receiver operating characteristic (ROC) curve was also drawn, and the area under the curve (AUC) was calculated to better evaluate the actual performance of the model.

## 3. Results and discussion

### 3.1. Data samples

Three groups of data samples were used as input to the deep learning model:

- 1) A total of 528 labeled fundus images from Beijing Tongren Hospital manually labeled by professional ophthalmologists.
- 2) A total of 532 fundus images from the OIADDR dataset [34] extracted by professionals and annotated by ophthalmologists.
- 3) A total of 425 fundus images from the Messidor dataset [35] with coarse segmentation labels of lesions derived using the proposed segmentation method based on grayscale and morphological features.

We reduced the resolution of each image to  $320 \times 320$  to accommodate images with poor resolution. All 1485 fundus images were randomly divided into a training set and verification set at a ratio of 4:1. All the images in the verification set were labeled manually by ophthalmologists.

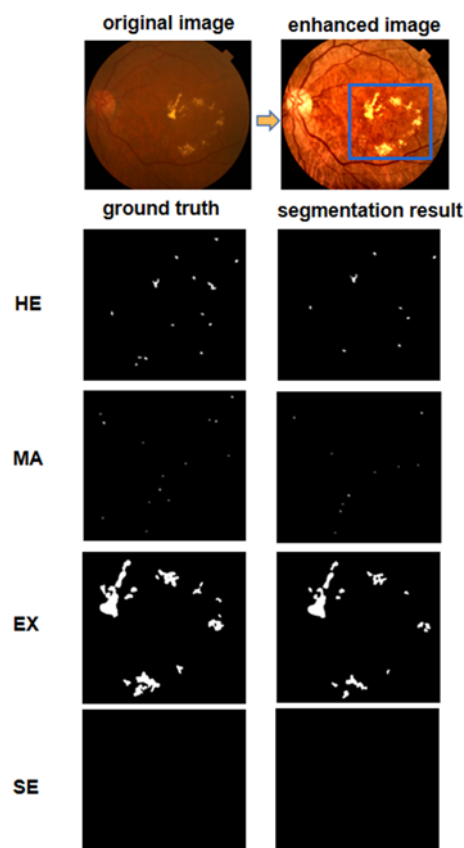
### 3.2. Coarse segmentation of fundus lesions based on grayscale and morphological characteristics

A total of 528 fundus images with a resolution of  $1400 \times 1200$  labeled by doctors were used to evaluate the proposed grayscale and morphological feature-based segmentation algorithm. The processing time of a single image was approximately 7 s on a PC platform with an RTX 2080 graphics card. The parameters of the proposed coarse segmentation method were set empirically and are shown in Table 1. The overall segmentation accuracies for HE, MA, EX and SE were 27.80, 37.67, 36.30 and 45.32%, respectively, with IOUs of 18.87, 7.58, 29.90 and 41.71%, respectively. The effect of image preprocessing and the actual segmentation effect are shown in Figure 4.

**Table 1.** Parameters selected for the coarse segmentation.

Coarse segmentation	Parameters
Preliminary segmentation thresholds (TL and TH)	50 and 200
Binarization threshold (TB)	15
Open and close operations kernels (MK)	(5,11,23)
Linear fitting parameters (L)	10
The radius of the circle (P)	1/15 of the image resolution
The aspect ratio of the area (ARA)	$\lambda = 1.75$
The ratio of the area (RA)	$\mu = 0.5$
The area of the connected domain ( $I$ )	$\alpha = 8, \beta = 53$
groups of thresholds for SE (TG1and TG2)	159, 160 and 199, 200

After image preprocessing, the image details were greatly enhanced, and the influence of noise in each image was suppressed; therefore, the visual interpretation of an image was improved. Comparing the segmentation results and the ground truths, we can see that the proposed segmentation algorithm based on the grayscale and morphological features can segment most of the key lesions in an image effectively, and the coarse segmentation results derived were generally in accordance with the ground truth.

**Figure 4.** Image enhancement and segmentation result.

### 3.3. Comparison with state-of-the-art models

The proposed RAUNet segmentation model was compared with existing state-of-the-art segmentation models. The IDRiD [36] open segmentation dataset was employed for the experiment (54 for training and 27 for testing). FCRN, CASENet, DeepLabv3+, LSeg and the proposed RAUNet model were used for training and testing. As seen in Table 2, RAUNet proposed in this paper was superior to the other four models in the segmentation AUC of all four types of lesions, especially for the two types of small target lesions, HE and MA.

**Table 2.** Test result comparison.

Model	EX	HE	SE	MA	Mean
	AUC	AUC	AUC	AUC	mAUC
CASENet [21]	0.7483	0.4486	0.3269	0.4013	0.4813
FCRN [22]	0.5469	0.4189	0.5163	0.3386	0.4552
DeepLabv3+ [37]	0.7125	0.4762	0.5932	0.1602	0.4855
LSeg [11]	0.7945	0.6374	0.7113	0.4627	0.6515
RAUNet	0.9321	0.8018	0.8479	0.6176	0.7976

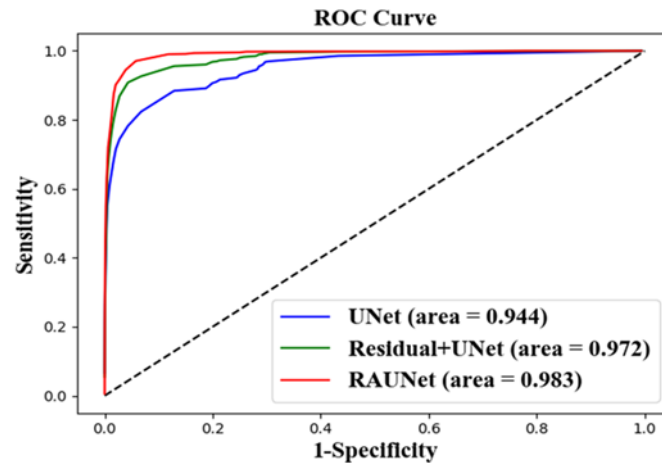
### 3.4. RAUNet ablation experiments

To verify the contribution of each improvement of the proposed segmentation model, ablation experiments were conducted. The original UNet, UNet with a residual module and the proposed RAUNet (UNet with a residual module and attention mechanism) were trained and evaluated with labeled dataset A (381 for training and 147 for testing).

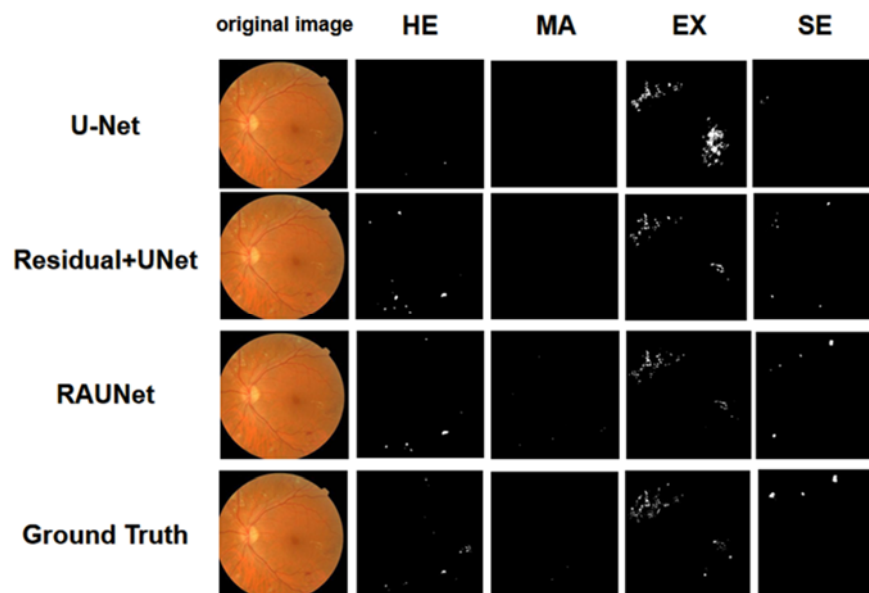
**Table 3.** Ablation test results.

Model	EX		HE		SE		MA		Mean		
	Acc.	IOU	Acc.	IOU	Acc.	IOU	Acc.	IOU	mAcc.	mIOU	mTime/pic (s)
UNet	0.8762	0.6140	0.4847	0.3342	0.6133	0.5728	0.5408	0.4558	0.6288	0.4942	2.9861
Res-UNet	0.7845	0.6161	0.4728	0.3582	0.6749	0.5793	0.5510	0.4694	0.6208	0.5058	1.0386
RAUNet (Res.+Att.)	0.8531	0.6448	0.5402	0.3892	0.6743	0.5961	0.5461	0.4809	0.6534	0.5278	6.4896





**Figure 5.** ROC curves of the four networks.



**Figure 6.** Comparison of the segmentation results derived using different networks.

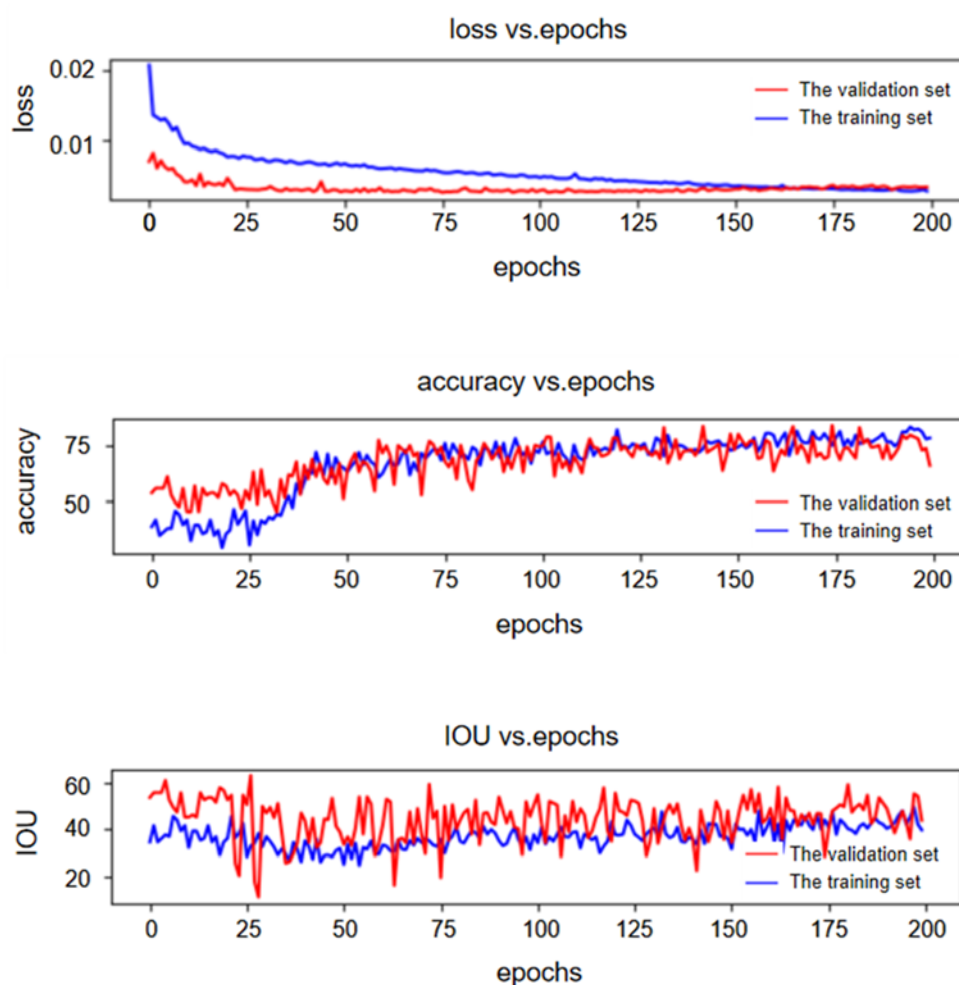
It can be observed from Table 3 that compared with UNet as the benchmark, for all lesions, the IOU between the segmentation result and ground truth was gradually improved by introducing the residual module and attention mechanism. The IOU obtained by RAUNet was significantly higher than that of the other networks. For the mean segmentation accuracy of all lesions, the result derived by the proposed RAUNet is the highest. We also calculated the mean time required for each model to process a single image. As shown in the last column of Table 3, the processing time of the proposed segmentation algorithm is larger than that of UNet and Res-UNet due to the added attention mechanism. However, we believe this is acceptable since diagnosing DR usually does not require strict real-time processing, and the processing time of the proposed algorithm can be largely reduced by improving the hardware and computation efficiency.

The receiver operating characteristic (ROC) curves shown in Figure 5 are also in accordance with the above results. The proposed RAUNet obtained the largest AUC, followed by UNet combined with the residual and attention modules.

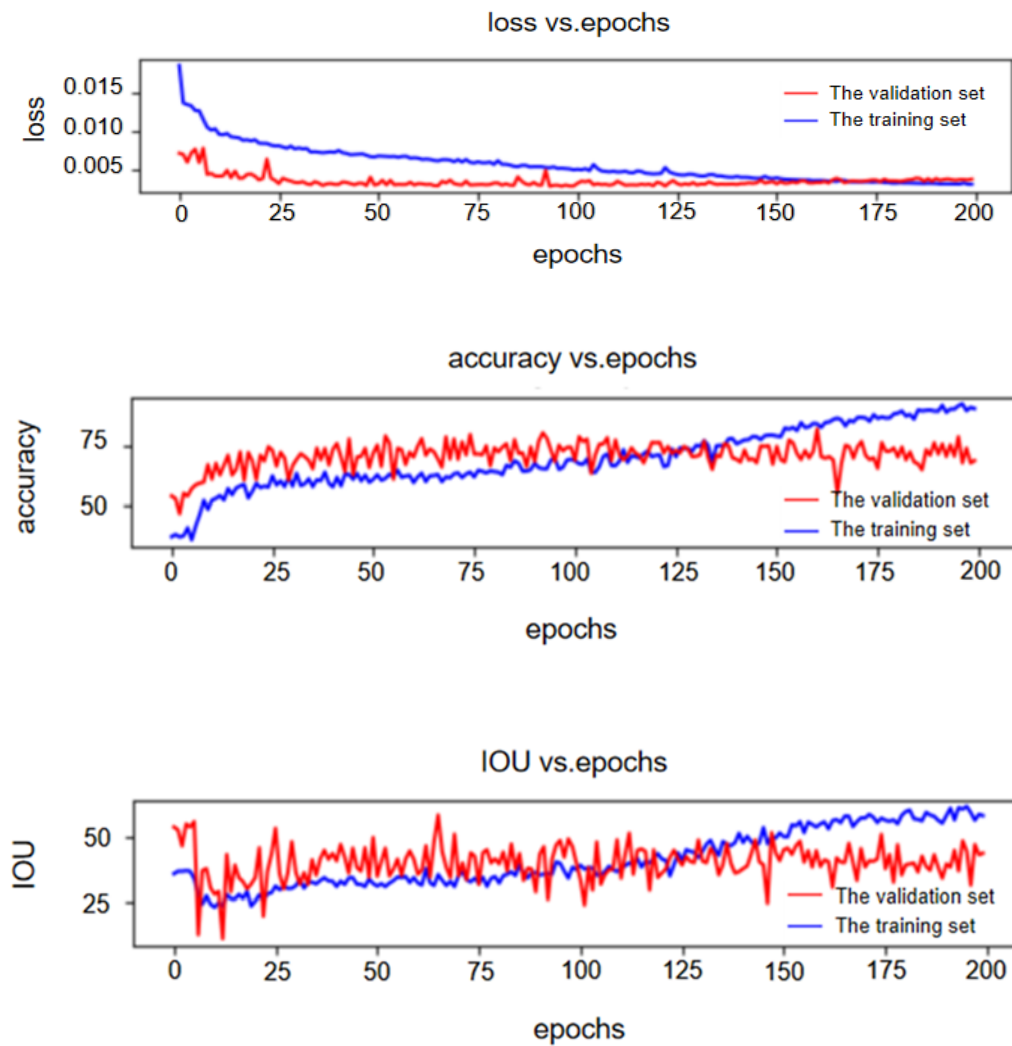
Some typical examples of the segmentation results of the four networks are shown in Figure 6. It can be observed that introducing residual and attention modules improved the UNet visual segmentation result. The residual module increased the model's sensitivity to HE and SE while reducing the false alarm rate of EX. The attention mechanism largely improved the segmentation for small lesions such as HE and MA. Therefore, the proposed RAUNet further obtained results closer to the ground truth labels than the compared networks for all kinds of lesions.

### 3.5. Weakly supervised segmentation of fundus lesions based on deep learning

The weakly supervised training model was evaluated on the basis of controlling the validation set unchanged. The loss, accuracy and IOU varying with the number of iteration epochs of the weakly and fully supervised models (taking EX as an example) are shown in Figures 7 and 8, respectively, where red represents the result derived by the verification set and blue represents that of the training set.



**Figure 7.** Changing indicator curves for the weakly supervised training model with increasing number of training epochs.



**Figure 8.** Changing indicator curves for the fully supervised training model with increasing number of training epochs.

**Table 4.** Experimental results of weakly supervised and fully supervised models.

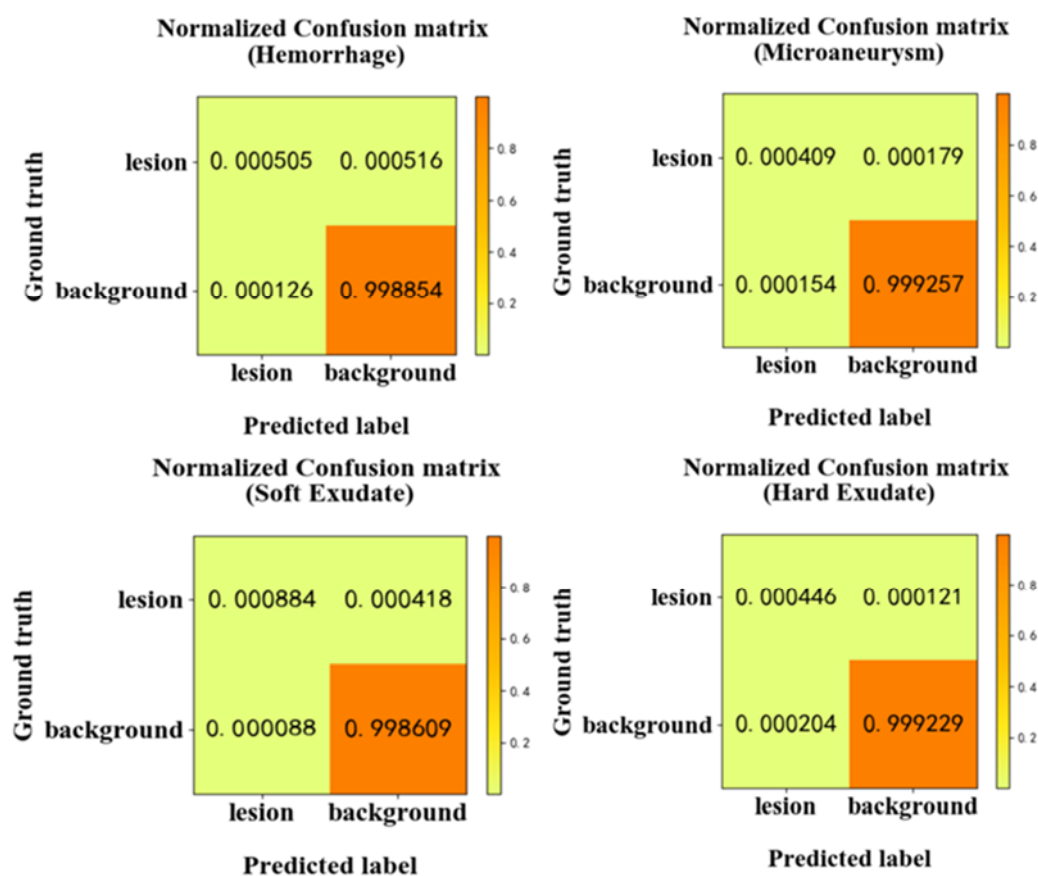
Model	EX		HE		SE		MA		Mean	
	Acc.	IOU	Acc.	IOU	Acc.	IOU	Acc.	IOU	mAcc.	mIOU
RAUNet (Fully supervised)	0.7061	0.5821	0.4129	0.3918	0.6946	0.6331	0.5955	0.5737	0.6023	0.5452
RAUNet (weakly supervised)	0.7714	0.5849	0.6184	0.4466	0.743	0.6403	0.5789	0.5684	0.6779	0.5601

As shown in Figure 7, for the proposed weakly supervised model, the evaluation indices on the validation and training datasets showed a generally synchronous upward trend with the increase in the number of training epochs. However, as shown in Figure 8, for the fully supervised learning model,

when the number of training epochs reached a certain value, the evaluation indices of the validation set no longer increased with that of the training set, indicating a certain degree of overfitting.

A comparison of the segmentation results derived from the fully and weakly supervised training is shown in Table 4 (weakly supervised: 1200 for training and 285 for testing; fully supervised: 381 for training and 285 for testing). For HE, EX and SE, the segmentation accuracy and IOU of the segmentation results using weakly supervised training were higher than those using fully supervised training. However, the segmentation result of MA was not improved by weakly supervised learning. This may be attributed to the small size of MA, which may be caused by the effect of noise on grayscale and morphological feature-based segmentation. Overall, the segmentation performance was significantly improved by the proposed weakly supervised model. The coarse segmentation results derived from the proposed coarse segmentation model sufficiently take advantage of the large amount of unlabeled data samples and increase the diversity of the lesions by adding small turbulence on the segmentation label, which can avoid overfitting and improve the robustness of the proposed weakly supervised model.

The confusion matrices of the four lesions obtained using weakly supervised training are shown in Figure 9. It can be seen that most of the correctly segmented lesion areas are significantly higher than the misclassified and missed areas. The misclassification rate for HE is higher, resulting in the lowest accuracy and IOU, which is partly due to the complex appearance of this kind of lesion.



**Figure 9.** The confusion matrices of the four lesions obtained using weakly supervised training.

#### 4. Conclusions

Lesion segmentation in fundus images is beneficial to the diagnosis and treatment of patients. However, due to the low contrast, small size and variant appearance characteristics of the lesions, conventional image segmentation algorithms usually cannot obtain satisfactory segmentation results. In this study, a weakly supervised framework was proposed for fundus lesion segmentation using grayscale and morphological features of lesions and a deep neural network.

Experimental results proved that: 1) the CLAHE image enhancement algorithm effectively improved the contrast between lesions and background; 2) an attention mechanism can improve the segmentation accuracy by increasing the sensitivity of the model to small lesions, and a residual module can improve the multiscale feature extraction ability of the network and avoid the problem of gradient dispersion of a deep neural network; 3) the introduction of a coarse labeled training dataset derived from the proposed segmentation method based on the grayscale and morphological features effectively increased the diversity of lesion samples and further improved the generalization and robustness of the segmentation model using weakly supervised training.

#### Acknowledgments

This work was supported by grants from National Key R & D Program of China (2017YFC0909600), National Natural Science Foundation of China (8151101058, 11527801) and the Strategic Priority Research Program of Chinese Academy of Sciences (XDB41020104).

#### Conflict of interest

The authors declared that they have no conflicts of interest or competing interests to this work.

#### References

1. R. K. Meleppat, K. E. Ronning, S. J. Karlen, K. K. Kothandath, M. E. Burns, E. N. Pugh Jr, et al., In situ morphologic and spectral characterization of retinal pigment epithelium organelles in mice using multicolor confocal fluorescence imaging, *Invest. Ophthalmol. Visual Sci.*, **16** (2020). <https://doi.org/10.1167/iovs.61.13.1>
2. R. K. Meleppat, K. E. Ronning, S. J. Karlen, M. E. Burns, E. N. Pugh Jr, R. J. Zawadzki, In vivo multimodal retinal imaging of disease-related pigmentary changes in retinal pigment epithelium, *Sci. Rep.*, **11** (2021), 16252. <https://doi.org/10.1038/s41598-021-95320-z>
3. S. Fu, Analysis of 56 cases of type 2 diabetes mellitus with ocular lesions as the first manifestation, *Clin. Focus*, **22** (2007), 256–257. <https://doi.org/10.3969/j.issn.1004-583X.2007.04.013>
4. V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *JAMA*, **316** (2016), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
5. S. Sengupta, A. Singh, H. A. Leopold, T. Gulati, V. Lakshminarayanan, Ophthalmic diagnosis using deep learning with fundus images – A critical review, *Artif. Intell. Med.*, **102** (2020), 101758. <https://doi.org/10.1016/j.artmed.2019.101758>

6. J. Son, J. Y. Shin, H. D. Kim, K. Jung, K. Park, S. J. Park, Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images, *Ophthalmology*, **127** (2020), 85–94. <https://doi.org/10.1016/j.ophtha.2019.05.029>
7. L. M. Devi, K. Wahengbam, A. D. Singh, Dehazing buried tissues in retinal fundus images using a multiple radiance pre-processing with deep learning based multiple feature-fusion, *Opt. Laser Technol.*, **138** (2021), 106908. <https://doi.org/10.1016/j.optlastec.2020.106908>
8. A. V. Varadarajan, P. Bavishi, P. Ruamviboonsuk, P. Chotcomwongse, S. Venugopalan, A. Narayanaswamy, et al., Predicting optical coherence tomography-derived diabetic macular edema grades from fundus photographs using deep learning, *Nat. Commun.*, **11** (2020), 130. <https://doi.org/10.1038/s41467-019-13922-8>
9. H. N. Veena, A. Muruganandham, T. S. Kumaran, A review on the optic disc and optic cup segmentation and classification approaches over retinal fundus images for detection of glaucoma, *SN Appl. Sci.*, **2** (2020), 1476. <https://doi.org/10.1007/s42452-020-03221-z>
10. Q. Wu, A. Cheddad, Segmentation-based deep learning fundus image analysis, in *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, (2019), 1–5. <https://doi.org/10.1109/IPTA.2019.8936078>
11. S. Guo, T. Li, H. Kang, N. Li, Y. Zhang, K. Wang, L-Seg: An end-to-end unified framework for multi-lesion segmentation of fundus images, *Neurocomputing*, **349** (2019), 52–63. <https://doi.org/10.1016/j.neucom.2019.04.019>
12. C. Ployout, R. Duval, F. Cheriet, A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images, *IEEE Trans. Med. Imaging*, **38** (2019), 2434–2444. <https://doi.org/10.1109/tmi.2019.2906319>
13. R. Wang, B. Chen, D. Meng, L. Wang, Weakly supervised lesion detection from fundus images, *IEEE Trans. Med. Imaging*, **38** (2019), 1501–1512. <https://doi.org/10.1109/TMI.2018.2885376>
14. K. Shankar, A. R. W. Saitet, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, H. M. Pandey, Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model, *Pattern Recognit. Lett.*, **133** (2020), 210–216. <https://doi.org/10.1016/j.patrec.2020.02.026>
15. M. Hire, S. Shinde, Ant colony optimization based exudates segmentation in retinal fundus images and classification, in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, (2018), 1–6. <https://doi.org/10.1109/ICCUBEA.2018.8697727>
16. E. Imani, H. Pourreza, A novel method for retinal exudate segmentation using signal separation algorithm, *Comput. Methods Programs Biomed.*, **133** (2016), 195–205. <https://doi.org/10.1016/j.cmpb.2016.05.016>
17. V. Sathananthavathi, G. Indumathi, R. Rajalakshmi, Abnormalities detection in retinal fundus images, in *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, (2017), 89–93. <https://doi.org/10.1109/ICICCT.2017.7975165>
18. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, et al., Attention U-Net: Learning where to look for the pancreas, preprint, arXiv:1804.03999v3.
19. N. Ilyasova, A. Shirokanev, N. Demin, R. Paringer, Graph-based segmentation for diabetic macular edema selection in OCT images, in *2019 Fifth International Conference on Frontiers of Signal Processing (ICFSP)*, (2019), 77–81. <https://doi.org/10.1109/ICFSP48124.2019.8938047>

20. M. Tavakoli, S. Jazani, M. Nazar, Automated detection of microaneurysms in color fundus images using deep learning with different preprocessing approaches, in *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*, **11318** (2020), 113180E. <https://doi.org/10.1117/12.2548526>
21. Z. Yu, C. Feng, M. Y. Liu, S. Ramalingam, CASENet: deep category-aware semantic edge detection, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 1761–1770. <https://doi.org/10.1109/CVPR.2017.191>
22. J. Mo, L. Zhang, Y. Feng, Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks, *Neurocomputing*, **290** (2018), 161–171. <https://doi.org/10.1016/j.neucom.2018.02.035>
23. N. Kasabov, N. M. Scott, E. Tu, S. Marks, N. Sengupta, E. Capecchi, et al., Evolving spatio-temporal data machines based on the NeuCube neuromorphic framework: Design methodology and selected applications, *Neural Networks*, **78** (2016), 1–14. <https://doi.org/10.1016/j.neunet.2015.09.011>
24. L. K. Abood, Contrast enhancement of infrared images using Adaptive Histogram Equalization (AHE) with Contrast Limited Adaptive Histogram Equalization (CLAHE), *Iraqi J. Phys.*, **16** (2018). <https://doi.org/10.30723/ijp.v16i37.84>
25. O. Ramos-Soto, E. Rodríguez-Esparza, S. E. Balderas-Mata, D. Oliva, A. E. Hassanien, R. K. Meleppat, et al., An efficient retinal blood vessel segmentation in eye fundus images by using optimized top-hat and homomorphic filtering, *Comput. Methods Programs Biomed.*, **201** (2021), 105949. <https://doi.org/10.1016/j.cmpb.2021.105949>
26. X. Fan, J. Gong, Y. Yan, Red lesion detection in fundus images based on convolution neural network, in *2019 Chinese Control And Decision Conference (CCDC)*, (2019), 5661–5666. <https://doi.org/10.1109/CCDC.2019.8833280>
27. P. Maragos, 3.3 - Morphological filtering for image enhancement and feature detection, in *Handbook of Image and Video Processing (Second Edition)*, (2005), 135–156. <https://doi.org/10.1016/B978-012119792-6/50072-3>
28. L. Cheng, J. Xiong, L. He, Non-gaussian statistical timing analysis using second-order polynomial fitting, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, **28** (2009), 130–140. <https://doi.org/10.1109/TCAD.2008.2009143>
29. A. Valizadeh, Z. J. Wang, Minimum mean square error detector for multimessage spread spectrum embedding, in *2009 Sixteenth IEEE International Conference on Image Processing (ICIP)*, (2009), 121–124. <https://doi.org/10.1109/ICIP.2009.5414115>
30. O. Ronneberger, U-Net convolutional networks for biomedical image segmentation, in *Bildverarbeitung für die Medizin 2017*, (2017), 3. [https://doi.org/10.1007/978-3-662-54345-0\\_3](https://doi.org/10.1007/978-3-662-54345-0_3)
31. L. Han, Y. Chen, J. Li, B. Zhong, Y. Lei, M. Sun, Liver segmentation with 2.5D perpendicular UNets, *Comput. Electr. Eng.*, **91** (2021), 107118. <https://doi.org/10.1016/j.compeleceng.2021.107118>
32. Y. Zhang, H. Lai, W. Yang, Cascade UNet and CH-UNet for thyroid nodule segmentation and benign and malignant classification, in *MICCAI 2020: Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data*, (2021), 129–134. [https://doi.org/10.1007/978-3-030-71827-5\\_17](https://doi.org/10.1007/978-3-030-71827-5_17)
33. Z. H. Zhou, A brief introduction to weakly supervised learning, *Natl. Sci. Rev.*, **5** (2018), 44–53. <https://doi.org/10.1093/nsr/nwx106>



34. T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, H. Kang, Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening, *Inf. Sci.*, **501** (2019), 511–522. <https://doi.org/10.1016/j.ins.2019.06.011>
35. E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, et al., Feedback on a publicly distributed image database: the messidor database, *Image Anal. Stereol.*, **33** (2014), 231–234. <https://doi.org/10.5566/ias.1155>
36. P. Porwal, S. Pachade, M. Kokare, G. Deshmukh, F. Mériaudeau, IDRiD: Diabetic retinopathy – segmentation and grading challenge, *Med. Image Anal.*, **59** (2020), 101561.
37. L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in *Computer Vision – ECCV 2018*, 833–851. [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49)



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)