IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 6, 2020, accepted September 8, 2020, date of publication September 10, 2020, date of current version September 23, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3023273

U-net Based Method for Automatic Hard Exudates Segmentation in Fundus Images Using Inception Module and Residual Connection

YONGSHUO ZONG[1,4], JINLING CHEN[1], LVQING YANG[1,2,3], SIYI TAO[1], CIERYOUZHEN AOMA[1], JIANGSHENG ZHAO[1], AND SHUIHUA WANG[5], (Senior Member, IEEE)

[1]School of Information, Xiamen University, Xiamen 361005, China
[2]State Key Laboratory of Process Automation in Mining, Beijing 102600, China
[3]Beijing Key Laboratory of Process Automation in Mining & Metallurgy, Beijing 102600, China
[4]Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[5]School of Architecture Building and Civil Engineering, Loughborough University, Loughborough LE11 3TU, U.K.

Corresponding authors: Lvqing Yang (lqyang@xmu.edu.cn) and Shuihua Wang (shuihuawang@ieee.org)

This work was supported in part by the Open Fund of National (Beijing) Key Laboratory in 2019 for the Research on big data analysis method of concentration and metallurgy under Grant BGRIMM-KZSKL-2019-03, in part by the Fujian province industrial field Regional Development Project in 2019 for the intelligent mine construction and industrialization based on internet of things and virtual reality under Grant 2019H4021, and in part by Fundamental Research Funds For The Central University under Grant 22120190211.

ABSTRACT Diabetic retinopathy (DR) is an eye abnormality caused by chronic diabetes that affected patients worldwide. Hard exudate is an important and observable sign of DR and can be used for early diagnosis. In this paper, an automatic hard exudates segmentation method is proposed in order to aid ophthalmologists to diagnose DR in the early stage. We utilized the SLIC superpixel algorithm to generate sample patches, thus overcoming the difficulty of the limited and imbalanced dataset. Furthermore, a U-net based network architecture with inception modules and residual connections is proposed to conduct end-to-end hard exudate segmentation, and focal loss is utilized as the loss function. Extensive experiments have been conducted on the IDRiD dataset to evaluate the performance of the proposed method. The reported sensitivity, specificity, and accuracy achieve 96.38%, 97.14%, and 97.95% respectively, which demonstrates the effectiveness and superiority of our method. The achieved segmentation results prove the potential of the method for clinical diagnosis.

INDEX TERMS Deep learning, diabetic retinopathy, exudates segmentation, superpixel, U-net.

I. INTRODUCTION

Diabetic Retinopathy (DR) is a serious ocular abnormality associated with chronic diabetes. Patients suffering from it will possibly lose their sight gradually and even go blind [1]. Although DR is treatable with timely diagnosis and intervention, the symptoms of vision impairment can be easily overlooked in the early stage of the disease. Thus, it is necessary to arrange regular examinations for diabetics to delay or relieve the risk of blindness. However, the limited number of clinicians currently is far from enough for the diagnosis of the large number of patients, as there have been more than 400 million diabetics all over the world [2]. Therefore, an automatic diagnosis technology needs to be developed to aid medical specialists.

Hard exudate (HE) is regarded as one of the most prominent features caused by DR. It is formed by macromolecular substances leaking from blood vessels into the eyeball after the retinal vessels are ruptured. As shown in Figure 1, HE can be observed as bright spots or clumps with sharp edges in the fundus image. Nevertheless, various factors such as

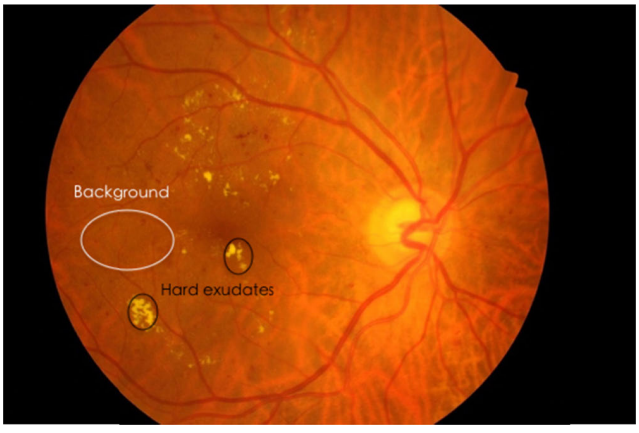

FIGURE 1. Hard exudates and background circled in black and white respectively.

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

uneven illumination and equipment noise constrain the accuracy and effectiveness of DR diagnosis through raw fundus images [3], [4]. Automatic detection of HE is therefore a useful and necessary auxiliary diagnostic approach for DR.

Traditional image processing approaches of HE detection includes four main categories: threshold-based [5]–[8], cluster-based [9]–[11], morphological-based [12]–[17], and region-growth [18]–[22]. Machine learning methods have also been introduced to detect HE [23]–[28], where professional knowledge is demanded to design hand-crafted features. These methods require exhausting work and complicated parameter settings but commonly not capable of possessing satisfying generalization. Without the need for expert knowledge, deep learning methods, especially convolutional neural networks (CNNs) [29]–[32] and U-Net [33]–[35], have been widely explored in recent years. Compared to traditional methods, detection and segmentation based on deep learning methods perform better in generalization ability and robustness. Despite this, approaches based on CNN fail to achieve satisfactory efficiency due to huge time and computational consumption. In addition, though faster and more efficient than CNN, current U-Net based methods have not achieved satisfying results as the state-of-the-art methods.

In order to solve the above-mentioned problems and develop a more efficient and accurate method for HE segmentation, we propose a novel U-Net based network architecture to implement end-to-end segmentation. As for the difficulty of limited and imbalanced training data, a SLIC superpixel algorithm is applied to generate sample patches to enlarge the dataset, with the relationship among adjacent pixels preserved.

The proposed method consists of the following three steps. After image preprocessing, the SLIC superpixel algorithm is applied to segment the images, and then sample patches are generated based on the superpixel. Then, the network architecture is proposed to predict the segmentation results of each patch. Finally, the neighboring prediction results of patches are spliced together to complete images for the final results. Extensive experiments are conducted on the publicly available dataset IDRiD [36] to verify the performance of the proposed method.

The main contribution of our work can be summarized as follows: (1) superpixels are utilized to cluster the images by considering image pixels possessing similar characteristics regionally as natural entities. Hence, patches extracted based on this possess more contextual information and are beneficial for further segmentation; (2) proportion of sampled patch is carefully chosen and focal loss function is applied, both of which help to overcome the difficulty of the imbalance of the dataset; (3) the proposed network architecture with inception modules and residual connections is capable of extracting multi-scale features and combing low-level and high-level features, thus achieving better performance.

The rest of this paper is organized as follows: Section II analyzes the advantages and disadvantages of related researches. Section III presents the proposed method in detail. Section IV validates the performance of our method with extensive experiments. Finally, the whole process and future work are concluded in Section V.

II. RELATED WORK

Automatic detection and segmentation of hard exudates have been studied before, developed methods include traditional methods and deep learning based methods.

A. TRADITIONAL METHODS

Traditional methods, consisting of image processing methods, such as threshold-based segmentation [5]–[8], cluster-based detection [9]–[11], morphological-based segmentation [12]–[17], region-growth detection [19]–[22], and machine learning methods that require hand-crafted features, have been extensively studied.

In the thresholding segmentation, local or global grayscale is regarded as the dominant characteristic. García *et al.* [8] utilized global and adaptive threshold simultaneously to segment candidate regions at first, and then employed a series of features and radial basis functions (RBF) to classify the true regions. However, it is difficult to select an appropriate and accurate threshold because the brightness and contrast of the image are not consistent.

Clustering-based methods recalculate the internal distance of various types after classifying the whole image set according to given rules, so as to update the clustering center until the convergence of each class center. Osareh *et al.* [10] adopted a fuzzy C-means (FCM) clustering algorithm to divide the exudates from background. These algorithms feature the disadvantages that they are usually sensitive to noise and computationally intensive. Also, the choice of the initial center is important but difficult as the location and characteristics of the center of the classes are unknown.

Morphological approaches identify exudates using extracted brightness and grayscale characteristics. Harangi and Hajdu [16] integrated mathematical morphology and active contours into a novel framework to segment exudates from retinal images. Zhang *et al.* [17] selected lesion candidates using morphological operators for subsequent detection in which multi-feature classification was applied. Despite the fast and effective effects of these methods in terms of computation, they do not take other characteristics into consideration, thus resulting in high noise sensitivity.

Segmentation based on region growing has been proved feasible especially when combined with the artificial neural network [19]. In these methods, the feature of spatial grayscale contiguity was applied for segmentation [20]. Additionally, edge detection can be employed at the same time to extract hard exudates [21] for optimization. Nevertheless, it is inclined to result in over-segmentation and the algorithm is relatively more time-consuming.

Machine learning methods have also been extensively studied to detect hard exudates, including support vector machine (SVM) [23], [24], linear discriminant classifiers [25], [26], Naive Bayes classifier [27] and random forest algorithm [28].

Giancardo *et al.* [24] extracted feature vectors based on color, wavelets and exudate probability, so that diabetic macular edema can be automatically diagnosed according to exudates. In [25], Saìnchez *et al.* improved Fisher's linear discriminant analysis using color features for classification. Besides, Harangi *et al.* [27] selected the most appropriate descriptors out of more than 50 other ones to train a boosted naive Bayes classifier.

These methods usually share a common idea that multi-dimensional feature vectors are constructed for each pixel or clustering of pixels. However, these feature vectors, which are often built on the basis of color, shape, size, and any other relevant information, require exhausting work and complicated parameter settings. In addition, poor generalization poses another challenge.

B. DEEP LEARNING METHODS

Deep learning methods, especially convolutional neural networks (CNNs) [29]–[32] and U-net [33]–[35], have been applied for exudates detection. They have been widely adopted in recent years because they do not require hand-crafted features.

Prentašicì and Lončaricì [29] built an 11-layer neural network model and used the output to generate an exudate probability map. Then, outputs of anatomical landmark detected were incorporated to optimize the exudate probability map. Gondal *et al.* [30] realized both image-level and lesion-level detection by improving well-performing o_O CNN architecture [31], by removing the dense layers and added a global average pooling (GAP) layer to the traditional CNN architecture. With o_O architecture employed as well, Quellec *et al.* [32] introduced a CNN visualization based solution where optimized heatmaps were produced for more accurate CNN predictions.

Although CNN based methods have achieved high accuracy, most of the CNN methods used the sliding-window algorithm to overcome the difficulty of limited and imbalanced training data, thus causing unsatisfying efficiency.

U-Net [33] is a commonly used backbone network in the field of medical image, consisting of contraction and expanding units, which can capture not only local but global context information of the images. Zabihollahy *et al.* [34] used modified U-net to conduct segmentation and detection of the hard exudates and removed optic disc for better results. To make full use of detailed information and context perceptions simultaneously, Yan *et al.* [35] combined the local and global U-Net decoders so that the two streams could be enhanced mutually. Although methods based on U-Net are fast and efficient, the segmentation results remain much to be improved compared to the traditional state-of-the-art methods.

In this paper, U-Net backbone is adopted to implement end-to-end segmentation, thus largely reducing the time consumption and designing for better performance. And superpixel is applied to extract data samples, which overcomes the difficulty of the small dataset and imbalanced data.

III. METHODOLOGY

A. PREPROCESSING

Since we are only interested in the retinal fundus, the redundant black background around fundus images needs to be cropped. Given that the image width is w, we select the largest pixel value from the w/64 leftmost and the w/16 rightmost parts of the image, add 10 to this value as a threshold. Regions with larger pixel values than the threshold are regarded as the foreground, and then the minimum bounding rectangle (MBR) of foreground area can be obtained. The area inside the MBR is cropped from the original image for further processing. After cropping, the size of the images is approximately 2848×3400. While the red channel is more saturated and the blue channel is darker, only the green channel [37] is adopted because it shows the highest contrast between the blood vessels and the background, which helps to reduce the interference of blood vessels in the process of classification [38]. To enhance the contrast between exudates and non-exudates, an image enhancement algorithm named Contrast Limited AHE (CLAHE) [39], which is a variant of adaptive histogram equalization (AHE), is adopted to reallocate lightness values with the clip limit of 8.0 and the grid size of 8×8. Subsequently, we apply gamma correction to compensate for the loss of brightness caused by uneven illumination and enhance the contrast. For a given input image I, the formulation is given by

$$f(I) = I^{\gamma} \tag{1}$$

where the gray-level coefficient γ is set to $1/2.2$ and $f(I)$ is the output image. Figure 2 shows four different stages during the image preprocessing. It can be seen that after the preprocessing, the image contrast between background and hard exudates is more obvious, which is beneficial for further segmentation.

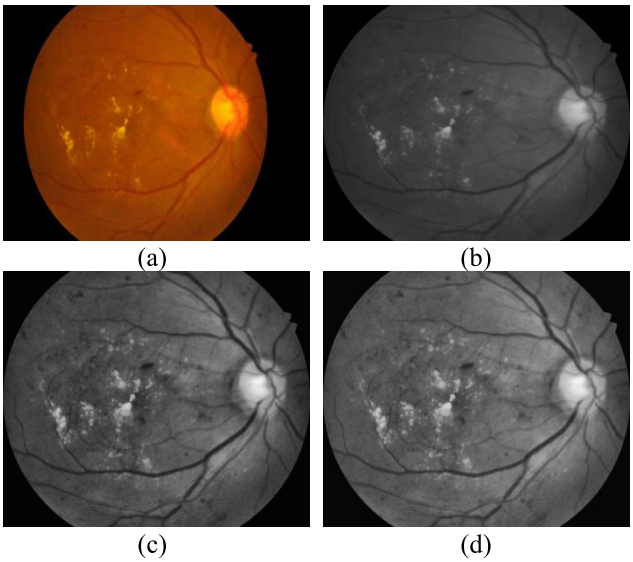

FIGURE 2. Four stages of preprocessing: (a) original RGB color retinal image, (b) cropped green-channel image, (c) enhanced green-channel image, (d) final image after gamma correction.

B. SAMPLING

In order to overcome the difficulty of the small dataset, sampling is performed on the original images to obtain more patches as samples. At the same time, we utilized the SLIC [40] superpixel algorithm to guide the patch generation process, instead of directly cropping rectangle patches on the original images.

The term superpixel is proposed by Ren and Malik [41] to aggregate individual pixels possessing similar features in interested low-level space, such as color, brightness, and texture [40]. It regards pixels as natural entities regionally so that the connection relationships within images can be preserved for further patch extraction.

We adopt the Simple Linear Iterative Clustering (SLIC) algorithm proposed by Achanta *et al.* [44], which has better performance than previous approaches [41]–[43] in adhering to image boundaries, speed, and memory efficiency. Given a customized region size and regularity, the SLIC algorithm can generate uniform and compact superpixels with similar features and clear boundaries. Specifically, a larger region size can result in lager superpixel segmentation, and the compactness coefficient relates to the regularity of superpixels. In our experiment, the number of segments and the compactness coefficient is set to 150 and 0.5 separately, which suits the patch size. As shown in Figure 3, the yellow polygons are superpixel subregions obtained after segmentation through the SLIC algorithm.

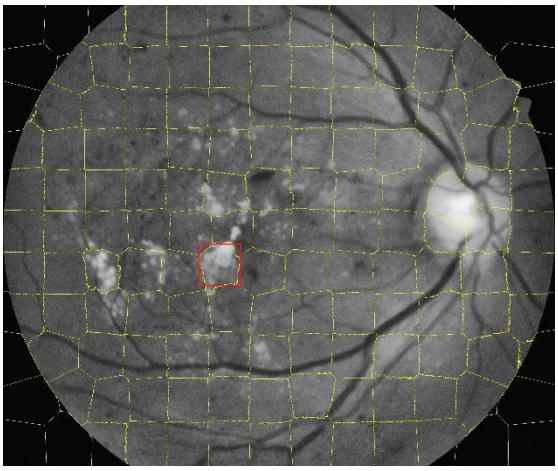

FIGURE 3. SLIC superpixel segmentation result with region number of 150 and compactness of 0.5. The red rectangle is an exudate included patch generated on the basis of the superpixel.

Next, we divide the superpixels into two categories according to the corresponding ground truth. For a fundus image I, there are N segmented superpixel subregions $X = (X_1, X_2, \ldots, X_i, \ldots, X_N)$, $i\epsilon(1, N)$. The i-th superpixel consists of M pixels $X_i = \{p_i(1), \ldots, p_i(k), \ldots, p_i(M)\}$ with $p_i(k)\epsilon I$. If there exists at least one pixel $p_i(k)$ belonging to the hard exudate according to the ground truth inside the superpixel X_i, $i\epsilon(1, N)$, we regard the superpixel X_i as a hard exudate included one. Otherwise, X_i will be treated as non-exudate background.

In this way, original images are represented in the superpixel level, which takes the information of adjacent regions into consideration. Therefore, the region of each extracted patch possesses similar features, which provides a conducive foundation for further hard exudates segmentation due to the clustering effect [45].

Although the superpixel can effectively interpret and represent image information, its irregular shape makes it unable to be directly fed to the network. To overcome this problem, we generate sample patches with the superpixel centroids as the center of the generated patches. Thus, a set of N patches $Y = (Y_1, Y_2, \ldots, Y_i, \ldots, Y_N)$, $i\epsilon(1, N)$ is obtained based on the superpixels $X = (X_1, X_2, \ldots, X_i, \ldots, X_N)$, $i\epsilon(1, N)$, where each image patch $Y_i\epsilon R^{h\times w}$ is a rectangle R with a size of height h and width w. Both of the h and w are to 256, and therefore the size of each patch is 256×256.

As illustrated in Figure 3, the red rectangle is one of the patches obtained based on superpixels in the yellow line. Similarly, the patches are classified into ones with and without hard exudates, which is shown in Figure 4. If a superpixel X_i contains hard exudate pixels, the relevant patch Y_i will be considered as a hard exudate patch, and vice versa.

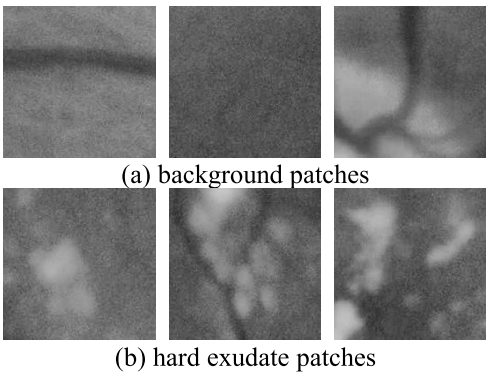

FIGURE 4. Two kinds of patches: (a) background patches, (b) hard exudate patches.

Finally, a total of 900 patches are generated for training, and in order to avoid the imbalance between positive and negative samples, the proportion of hard exudate patches and background ones is set to 2 : 1.

C. NETWORK ARCHITECTURE

In this subsection, we proposed a network architecture with U-Net [33] as the backbone for the segmentation of the hard exudates, which is shown in Figure 5. It takes the generated patches as input and outputs binary images as segmentation results. The size of both input and output images are of 256×256.

The proposed 9-unit network architecture is composed of a contracting path and an expansive path. The former path extracts features of the input patches, and then the latter path carries out the process of up-convolution. Between the contracting path and an expansive path, skip connections are serving as bridges for information propagation. These skip

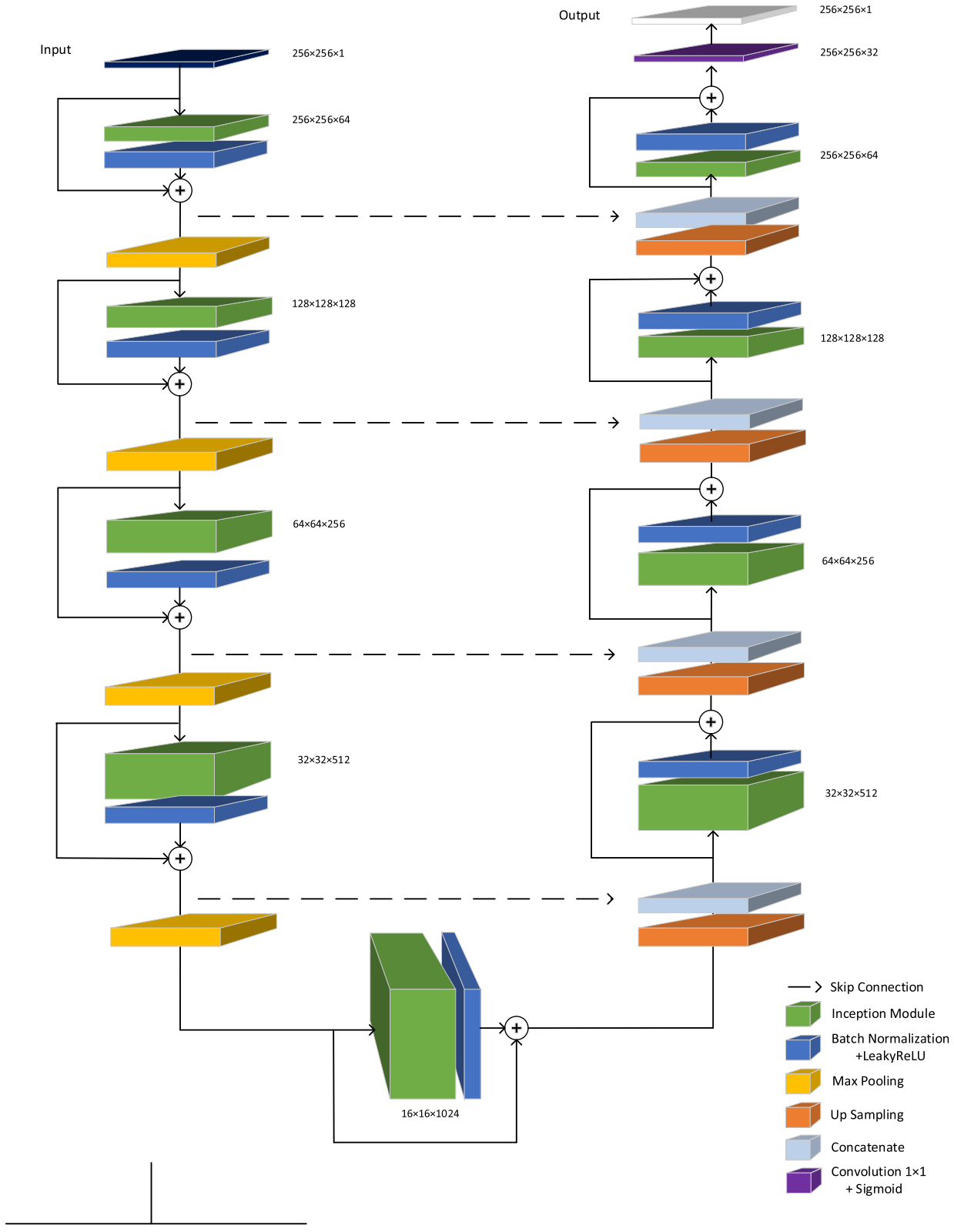

FIGURE 5. The architecture of the network based on U-Net backbone.

connections are capable of combining low-level details with high-level semantic information. The contracting path consists of five basic units, where a max-pooling layer is inserted between every two units along the path to down-sample the feature map. Correspondingly, there are four basic units in the expansive path with an up-sampling layer and a concatenation

layer before each unit, thus making use of feature maps from both the lower level and the relevant contracting path. In the end, we use a 1×1 convolution layer with a sigmoid activation to generate the output prediction results.

Each of the abovementioned units in both of the two paths is constructed by the residual unit with inception [46] module, which is shown in Figure 6(b). It consists of an inception module instead of a common convolution module, followed by batch normalization (BN) [47] and LeakyReLU activation [48]. By using the LeakyReLU as an activation function, the network converges at an earlier stage. Besides, borrowed from ResNet [49], we add an identity mapping to the unit.

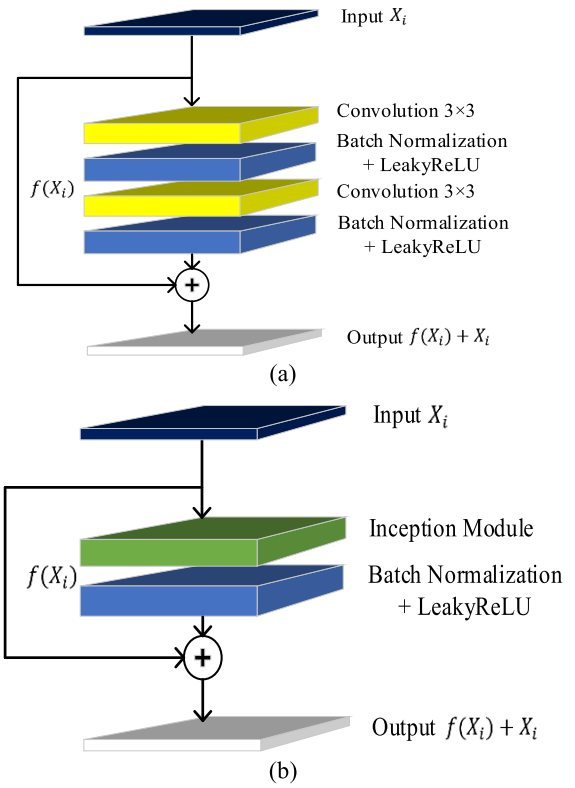

FIGURE 6. (a) Residual unit of U-Net with plain convolution, (b) Residual unit with inception module.

Figure 6(a) shows a residual unit with plain convolution, while it is effective in most cases, however, due to the severely uneven and imbalanced distribution of hard exudates in the patches, it is necessary to extract features from different scales. Therefore, the convolution operation in U-Net is substituted by an inception module, which is shown in Figure 6(b) and its internal structure is shown in Figure 7. A 1×3, a 3×1, and two 3×3 filter kernels are designed to extract multi-scale features. These features are then combined together and a 1×1 convolution kernel is used for dimension reduction. The inception module is not only able to efficiently reduce the number of parameters [50], but also capable of increasing the expression ability of the network by introducing more linear mappings. Moreover, the direct connection

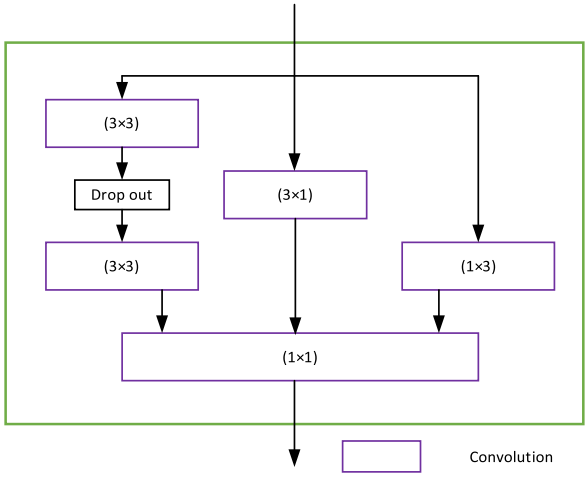

FIGURE 7. Internal structure of the inception module.

of the original U-Net is replaced by the residual connection in the proposed architecture. As illustrated in Figure 6, we denote the input of the residual as X_i, the residual function as $f(\cdot)$. Therefore, the output of the residual is $f(X_i)+X_i$ after the addition.

In addition, dropout [51] is utilized between the two 3×3 filter kernels to avoid overfitting, and the dropout rate is set to 0.5.

The imbalance of the data consists of two aspects. First, most of the sampled patches only contain background, and patches containing hard exudates are of a small majority. As described before, this can be solved by choosing the appropriate proportions of samples. The other comes from the imbalanced distribution of hard exudates in each patch containing hard exudates. For most of these patches, hard exudates only consist of a small proportion, most of which are less than 10%, thus making the accurate segmentation even more difficult. In order to solve this problem, Focal Loss (FL) [52] is introduced in our proposed network as the loss function, which is shown below:

$$FL(p_s) = -w_s(1-p_s)^{\alpha}\log(p_s) \tag{2}$$

where

$$p_s = \begin{cases} p, & s = 1 \\ 1-p, & s = 0, \end{cases}$$

and p is defined as the estimated probability for the class with label $s = 1$. w_s is derived by

$$w_s = \begin{cases} w, & s = 1 \\ 1-w, & s = 0, \end{cases}$$

with w dealing with the weight of positive and negative samples. The variable s in the above two formulas denotes the type of ground-truth class, for which $s = 1$ means that the class is an easily classified sample, and otherwise it is difficult to divide. The index number α ($\alpha \geq 0$) works as a tunable

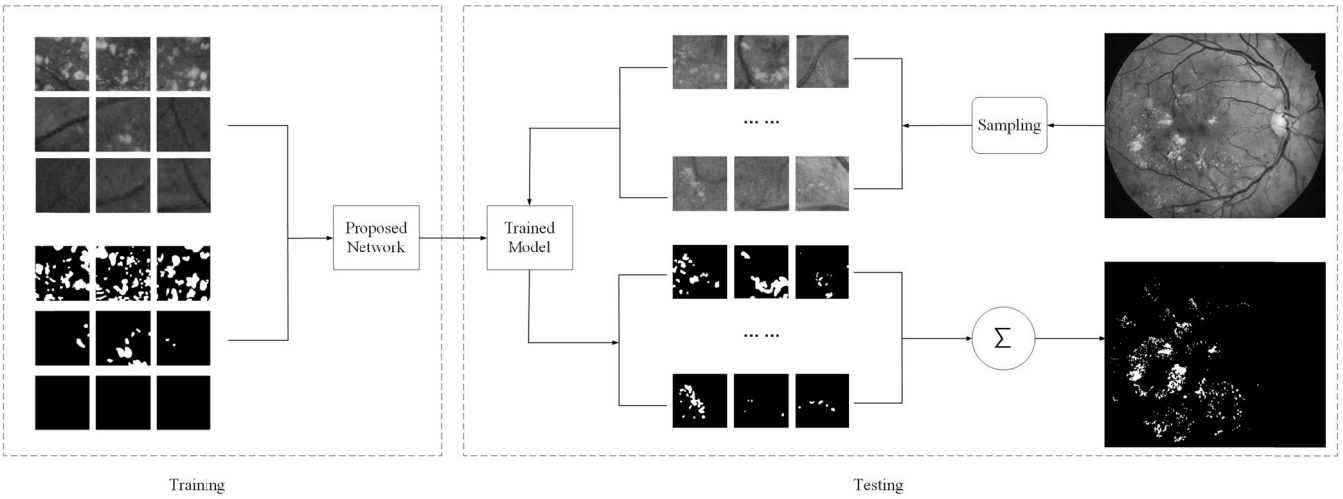

FIGURE 8. The training and testing process.

focusing parameter to relieve the loss dominated by a large number of simple samples. Compared with the cross-entropy (CE), the focal loss tackles the imbalance between not only positive and negative samples but also hard and simple examples, thus more suitable for our scenario.

Finally, we apply the Adam (adaptive moment estimation) Optimizer [53] to train the network, with a learning rate of $1e^{-4}$. The network takes 800 epochs to converge.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATASET

IDRiD (Indian Diabetic Retinopathy Image Dataset) [36] is a public dataset available online. This dataset provides pixel-level typical diabetic retinopathy lesions and normal retinal structures. A total of 81 images are given labeled with pixel-level ground truths, among which 54 images are training set and 27 images are testing set. Images in this dataset feature a resolution of 4288 × 2848 pixels and a 50° field of view.

B. EVALUATION METRICS

Metrics including *sensitivity*, *specificity*, and *accuracy* are adopted to evaluate the performance of the proposed method. The related formulas are defined as follows:

$$sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$specificity = \frac{TN}{TN + FP} \tag{4}$$

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{5}$$

where TP (*True Positive*), TN(*True Negative*) represent the correctly detected hard exudates and non-exudate background respectively, while FP (*False Positive*), FN(*False Negative*) denote the number of wrongly detected ones as exudates and background.

C. RESULTS AND ANALYSIS

The training and testing process of our experiment is described in Figure 8. After preprocessing and sampling, we have generated **900** sample patches of size **256** × **256** for training through the SLIC superpixel algorithm. Then we classify them into patches including hard exudates and non-exudate backgrounds, which is used to train the network. For testing, in order to obtain the final segmentation results, we splice the neighboring prediction results of patches together to complete images, during which we take the average value of overlapped areas and set the threshold value to **0.5**.

The first experiment compares the results of whether the patches are randomly cropped out or generated using the SLIC superpixel algorithm. Figure 9(b) is the enlarged image of the blue rectangle in Figure 9(a), the red rectangle is the patch cropped randomly and the pink patch is generated through the above-mentioned method. Random crops fail to consider the contextual information of pixels, and may introduce irrelevant parts in the same patch, thus undermining the segmentation performance. On the contrary, patches extracted based on superpixels possess necessary information

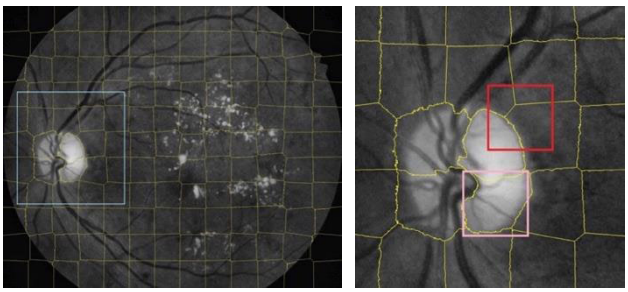

FIGURE 9. Two kinds of patch generation methods: random cropping and generating based on superpixels.

TABLE 1. Comparision of whether to use superpixel.

sample method	sensitivity	specificity	accuracy
random crop	94.96%	95.33%	95.47%
superpixel based	**96.38%**	**97.14%**	**97.95%**

TABLE 2. Comparison of varying patch sizes.

size	number of segments	sensitivity	specificity	accuracy
128 × 128	500	89.67%	91.34%	89.05%
480 × 480	50	94.52%	94.70%	94.53%
360 × 360	80	95.19%	96.51%	96.42%
256 × 256	**150**	**96.38%**	**97.14%**	**97.95%**

TABLE 3. Comparision of different sampling proportions.

HE: BG	sensitivity	specificity	accuracy
1: 2	94.82%	97.82%	96.80%
1: 1	95.63%	97.35%	96.94%
2: 1	**96.38%**	**97.14%**	**97.95%**

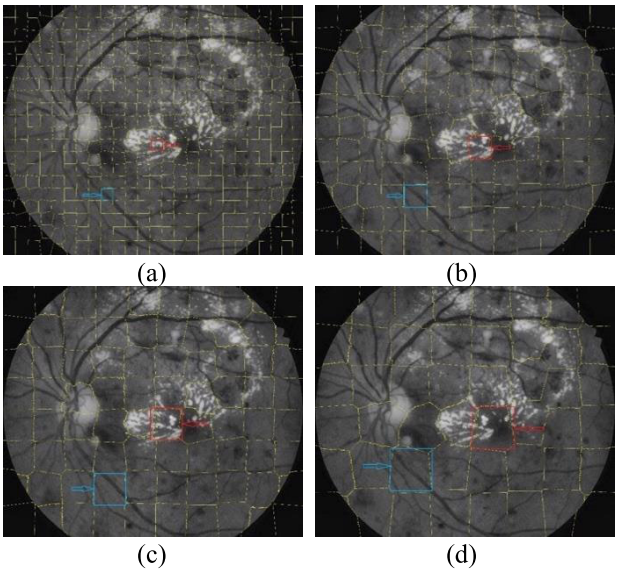

FIGURE 10. Different sample results with varying patch size: (a) 128 × 128, (b) 256 × 256, (c) 360 × 360, (d) 480 × 480.

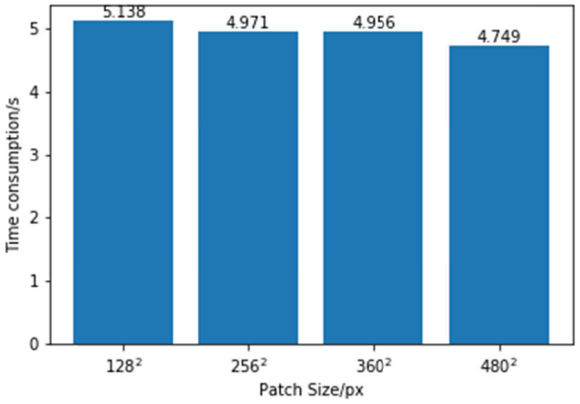

FIGURE 11. Time consumption of SLIC segmentation for different patch sizes per image.

for better segmentation, because the SLIC algorithm has clustered the regions with similar features into superpixels. Table 1 demonstrates the experimental results based on different patch generation methods.

In terms of the size of the patches, we choose the possible range of the patch size firstly, then carry out experiments to determine the best patch size. As depicted in the four subfigures of Figure 10, we mark an exudate included patch and a background one with red and blue rectangles respectively. The corresponding performance is revealed in Table 2, where we can see that the segmentation results become better when the size increase from 128 × 128 to 256 × 256, but begin to fall gradually when the size decrease to 360 × 360 and 480 × 480. The patch size of 256 × 256 performs best due to its appropriate size that is capable of including the essential information for segmentation and is not too large to introduce unnecessary interference.

In addition, the time consumption of the SLIC segmentation algorithm for different patch sizes per image is shown in Figure 11, with Python3.6 and skimage module on a desktop of 2.60Ghz CPU with 16 G RAM. It can be seen that the time consumption varies slightly, which means different patch sizes do not influence the efficiency of the sampling process.

Experiments regarding the proportions of hard exudate patches (HE) to background patches (BG) have been carried out to overcome the imbalance of data between positive and negative samples. It can be seen from Table 3, when the proportion of HE to BG is 2:1, the sensitivity and accuracy reach the highest, thus achieving the best segmentation result. When the proportion is 1:2, the patches that only contain background are relatively excessive, and therefore lead to

TABLE 4. Comparison of different network architectures' evaluation.

network architecture	sensitivity	specificity	accuracy
U-Net	92.15%	93.63%	93.87%
U-Net + residual connection	93.40%	94.25%.	94.11%
U-Net + inception module	96.26%	96.87%	97.03%
Proposed network	**96.38%**	**97.14%**	**97.95%**

TABLE 5. Comparision of different methods' performance on the IDRiD dataset.

Authors	year	method	sensitivity	specificity	accuracy
Zabihollahy et al. [34]	2019	U-net based	96.15%	80.77%	88.46%
Song et al. [54]	2020	HED[56] based	95.79%	-	-
Singh et al. [55]	2020	CNN based	96.32%	95.84%	96.12%
Ours	**2020**	**U-net based**	**96.38%**	**97.14%**	**97.95%**

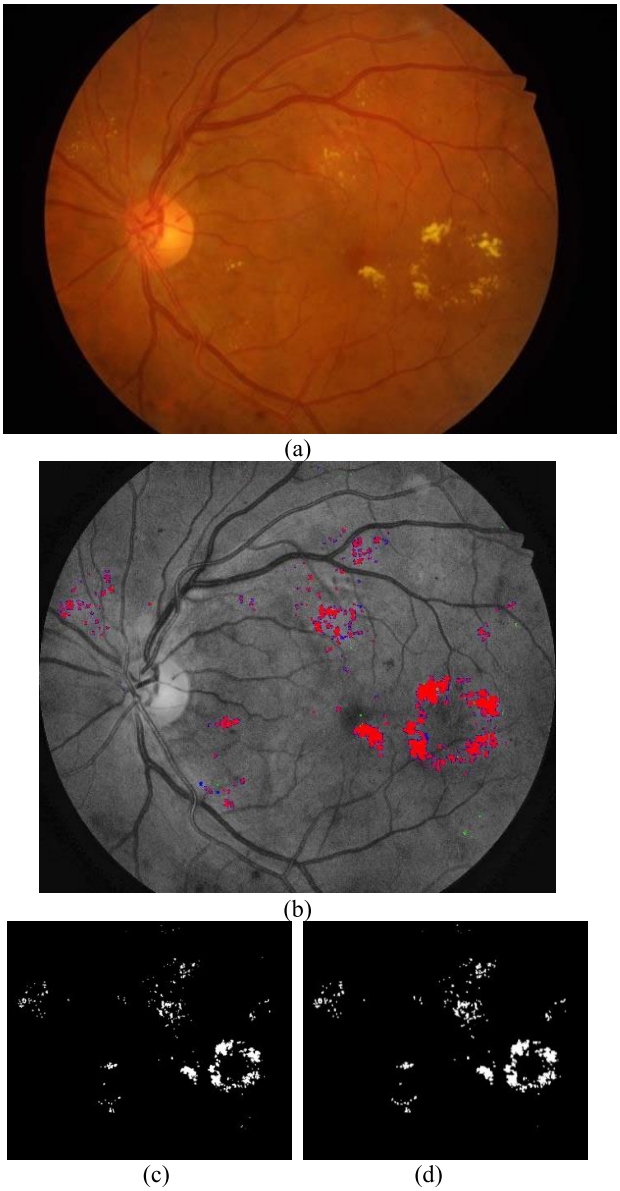

FIGURE 12. Prediction result: (a) Image IDRiD_70 from the IDRiD dataset, (b) corresponding pixel-based segmentation result, (c) ground truth label, (d) relevant segmentation result.

lower accuracy and sensitivity. In addition, when the proportion is 1:1, the same number of hard exudates and background patches are sampled. However, this makes training less targeted, and therefore weaken the network's ability to distinguish the hard exudates from the interference, such as optic disc and vessels. Hence, the ratio of 2 : 1 is chosen due to the highest *sensitivity* and *accuracy*.

Meanwhile, we have conducted experiments to compare the performance of different network architectures and demonstrate the effectiveness of the proposed network. As can be seen in Table 4, whether U-Net is combined with a residual connection or an inception module, the result will score higher in all the three dimensions *sensitivity*, *specificity*, and *accuracy*, especially when inception module is applied. The improvement of the performance by inception module can be attributed to its better ability to extract features from multi scales, and the increase in the width of the network. The building unit in the architecture of U-Net + residual network is shown in Figure 6(a), which is a plain 3×3 convolution unit of U-Net with an identity mapping. This structure, though optimized with batch normalization, fails to achieve a satisfying result mainly due to its same size filter kernels, which cannot extract features effectively in this scenario. In our proposed network, the U-Net backbone is optimized with both residual connections and inception modules, achieving the best performance with a 96.38% *sensitivity*, 97.14% *specificity*, and *accuracy* in 97.95%.

In the end, we compare the pixel-level segmentation results with other researches to prove the effectiveness of the proposed method. The latest researches conducted on the same dataset IDRiD with different methods, such as U-net based, CNN based and HED [56] based, are selected and compared. As can be seen in Table 5, the result of our method outperforms others' in terms of *sensitivity*, *specificity*, and *accuracy*. Furthermore, we have achieved a significant improvement compared with the previous U-net based method [34]. Figure 12 shows an original RGB color retinal image and its corresponding pixel-level segmentation result, where *TP* is marked in red, *FP* is marked in blue, and *FN* is marked in green.

V. CONCLUSION

In this paper, we propose a novel method of hard exudate segmentation based on an optimized U-Net architecture. Firstly, we generate sample patches based on the SLIC superpixel algorithm and distinguish them into hard exudate patches and non-exudate backgrounds. Then, we fed them to our network, where the U-Net backbone is improved

by utilizing both residual connections and inception modules. At last, we splice the neighboring prediction results of patches together to complete images for the final segmentation results. The proposed method is evaluated on the public dataset IDRiD with a series of comparative experiments. The experimental result in *sensitivity*, *specificity*, and *accuracy* achieves 96.38%, 97.14%, and 97.95% respectively, which demonstrates superior performance among current methods. Future work could be extended by applying attention gates into the network.

REFERENCES

[1] H. Wang, G. Yuan, X. Zhao, L. Peng, Z. Wang, Y. He, C. Qu, and Z. Peng, "Hard exudate detection based on deep model learned information and multi-feature joint representation for diabetic retinopathy screening," *Comput. Methods Programs Biomed.*, vol. 191, Jul. 2020, Art. no. 105398.
[2] C.-X. Huang, "Long-term effects of pattern scan laser pan-retinal photocoagulation on diabetic retinopathy in chinese patients: A retrospective study," *Int. J. Ophthalmol.*, vol. 13, no. 2, pp. 239–245, Feb. 2020.
[3] Y. Kang, Y. Fang, and X. Lai, "Automatic detection of diabetic retinopathy with statistical method and Bayesian classifier," *J. Med. Imag. Health Informat.*, vol. 10, no. 5, pp. 1225–1233, May 2020.
[4] S. Karkuzhali, S. Senthilkumar, and D. Manimegalai, "Algorithms for diagnosis of diabetic retinopathy and diabetic macula edema—A review," in *Advances in Experimental Medicine and Biology*. New York, NY, USA: Springer, 2020.
[5] C. I. Sánchez, M. García, A. Mayo, M. I. López, and R. Hornero, "Retinal image analysis based on mixture models to detect hard exudates," *Med. Image Anal.*, vol. 13, no. 4, pp. 650–658, Aug. 2009.
[6] R. Phillips, J. Forrester, and P. Sharp, "Automated detection and quantification of retinal exudates," *Graefe's Arch. Clin. Exp. Ophthalmol.*, vol. 231, no. 2, pp. 90–94, 1993.
[7] S. Ali, D. Sidibé, K. M. Adal, L. Giancardo, E. Chaum, T. P. Karnowski, and F. Mériaudeau, "Statistical atlas based exudate segmentation," *Computerized Med. Imag. Graph.*, vol. 37, nos. 5–6, pp. 358–368, Jul. 2013.
[8] M. García, C. I. Sánchez, J. Poza, M. I. López, and R. Hornero, "Detection of hard exudates in retinal images using a radial basis function classifier," *Ann. Biomed. Eng.*, vol. 37, no. 7, pp. 1448–1463, Jul. 2009.
[9] C. JayaKumari and R. Maruthi, "Detection of hard exudates in color fundus images of the human retina," *Procedia Eng.*, vol. 30, pp. 297–302, Jan. 2012.
[10] A. Osareh, M. Mirmehdi, B. Thomas, and R. Markham, "Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural networks," in *Proc. Med. Image Understand. Anal. Conf.*, 2001, pp. 49–52.
[11] X. Zhang and O. Chutatape, "Top-down and bottom-up strategies in lesion detection of background diabetic retinopathy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2005, pp. 422–428, doi: 10.1109/CVPR.2005.346.
[12] T. Walter, J. Klein, P. Massin, and A. Erginay, "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina," *IEEE Trans. Med. Imag.*, vol. 21, no. 10, pp. 1236–1243, Oct. 2002.
[13] E. Imani and H.-R. Pourreza, "A novel method for retinal exudate segmentation using signal separation algorithm," *Comput. Methods Programs Biomed.*, vol. 133, pp. 195–205, Sep. 2016.
[14] S. Sreng, J.-I. Takada, N. Maneerat, D. Isarakorn, B. Pasaya, R. Panjaphongse, and R. Varakulsiripunth, "Automatic exudate extraction for early detection of diabetic retinopathy," in *Proc. Int. Conf. Inf. Technol. Elect. Eng.*, Oct. 2013, pp. 31–35.
[15] D. Welfer, J. Scharcanski, and D. R. Marinho, "A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images," *Computerized Med. Imag. Graph.*, vol. 34, no. 3, pp. 228–235, Apr. 2010.
[16] B. Harangi and A. Hajdu, "Automatic exudate detection by fusing multiple active contours and regionwise classification," *Comput. Biol. Med.*, vol. 54, pp. 156–171, Nov. 2014.
[17] X. Zhang, G. Thibault, E. Decencière, B. Marcotegui, B. Laÿ, R. Danno, G. Cazuguel, G. Quellec, M. Lamard, P. Massin, A. Chabouis, Z. Victor, and A. Erginay, "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Med. Image Anal.*, vol. 18, no. 7, pp. 1026–1043, Oct. 2014.
[18] B. M. Ege, O. K. Hejlesen, O. V. Larsen, K. Møller, B. Jennings, D. Kerr, and D. A. Cavan, "Screening for diabetic retinopathy using computer based image analysis and statistical classification," *Comput. Methods Programs Biomed.*, vol. 62, no. 3, pp. 165–175, Jul. 2000.
[19] D. Usher, M. Dumskyj, M. Himaga, T. H. Williamson, S. Nussey, and J. Boyce, "Automated detection of diabetic retinopathy in digital retinal images: A tool for diabetic retinopathy screening," *Diabetic Med.*, vol. 21, no. 1, pp. 84–90, Jan. 2004.
[20] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images: Original article," *Diabetic Med.*, vol. 19, no. 2, pp. 105–112, Mar. 2002.
[21] H. Li and O. Chutatape, "Automated feature extraction in color retinal images by a model based approach," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 246–254, Feb. 2004.
[22] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, and L. Kennedy, "Optic nerve head segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 2, pp. 256–264, Feb. 2005.
[23] A. D. Fleming, S. Philip, K. A. Goatman, G. J. Williams, J. A. Olson, and P. F. Sharp, "Automated detection of exudates for diabetic retinopathy screening," *Phys. Med. Biol.*, vol. 52, no. 24, p. 7385, 2007.
[24] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin, and E. Chaum, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Med. Image Anal.*, vol. 16, no. 1, pp. 216–226, Jan. 2012.
[25] C. I. Sánchez, R. Hornero, M. I. López, M. Aboy, J. Poza, and D. Abásolo, "A novel automatic image processing algorithm for detection of hard exudates based on retinal image analysis," *Med. Eng. Phys.*, vol. 30, no. 3, pp. 350–357, Apr. 2008.
[26] M. Niemeijer, B. van Ginneken, S. R. Russell, M. S. A. Suttorp-Schulten, and M. D. Abràmoff, "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis," *Invest. Ophthalmol. Vis. Sci.*, vol. 48, no. 5, pp. 2260–2267, 2007.
[27] B. Harangi, B. Antal, and A. Hajdu, "Automatic exudate detection with improved Naïve-Bayes classifier," in *Proc. 25th IEEE Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2012, pp. 1–4.
[28] X. Zhang, G. Thibault, E. Decencière, B. Marcotegui, B. Laÿ, R. Danno, G. Cazuguel, G. Quellec, M. Lamard, P. Massin, A. Chabouis, Z. Victor, and A. Erginay, "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Med. Image Anal.*, vol. 18, no. 7, pp. 1026–1043, Oct. 2014.
[29] P. Prentašić and S. Lončarić, "Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion," *Comput. Methods Programs Biomed.*, vol. 137, pp. 281–292, Dec. 2016.
[30] W. M. Gondal, J. M. Kohler, R. Grzeszick, G. A. Fink, and M. Hirsch, "Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2069–2073.
[31] *o_O CNN Solution*. Accessed: Jan. 16, 17. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathydetection/ discussion/15617
[32] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, "Deep image mining for diabetic retinopathy screening," *Med. Image Anal.*, vol. 39, pp. 178–193, Jul. 2017.
[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
[34] F. Zabihollahy, A. Lochbihler, and E. Ukwatta, "Deep learning based approach for fully automated detection and segmentation of hard exudate from retinal images," *Proc. SPIE*, vol. 10953, Mar. 2019, Art. no. 1095308, doi: 10.1117/12.2513034.
[35] Z. Yan, X. Han, C. Wang, Y. Qiu, Z. Xiong, and S. Cui, "Learning mutually local-global U-Nets for high-resolution retinal lesion segmentation in fundus images," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Venice, Italy, Apr. 2019, pp. 597–600.
[36] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, Jul. 2018.
[37] X. Jiang and D. Mojon, "Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 1, pp. 131–137, Jan. 2003.

[38] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
[39] K. Zuiderveld, "Contrast limited adaptive histogram equalization," Graphics Gems IV, Academic Press Professional, New York, NY, USA, Tech. Rep., 1994, pp. 474–485, doi: 10.1016/B978-0-12-336156-1.50061-6.
[40] Z. Tian, L. Liu, Z. Zhang, and B. Fei, "Superpixel-based segmentation for 3D prostate MR images," *IEEE Trans. Med. Imag.*, vol. 35, no. 3, pp. 791–801, Mar. 2016.
[41] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 1, Oct. 2003, pp. 10–17.
[42] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and supervoxels in an energy optimization framework," in *Proc. Eur. Conf. Comput. Vis.*, vol. 6315, 2010, pp. 211–224.
[43] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *Proc. Eur. Conf. Comput. Vis.*, vol. 5305, 2008, pp. 705–718.
[44] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[45] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and supervoxels in an energy optimization framework," in *Proc. Eur. Conf. Comput. Vis.*, vol. 6315, 2010, pp. 211–224.
[46] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
[47] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 2483–2493.
[48] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*. [Online]. Available: https://arxiv.org/abs/1505.00853
[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[50] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[51] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[52] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.
[53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015, pp. 1–15.
[54] S. Guo, K. Wang, H. Kang, T. Liu, Y. Gao, and T. Li, "Bin loss for hard exudates segmentation in fundus images," *Neurocomputing*, vol. 392, pp. 314–324, Jun. 2020.
[55] R. K. Singh and R. Gorantla, "DMENet: Diabetic macular edema diagnosis using hierarchical ensemble of CNNs," *PLoS ONE*, vol. 15, no. 2, Feb. 2020, Art. no. e0220677.
[56] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1395–1403, doi: 10.1109/ICCV.2015.164.