A new deep learning approach for the retinal hard exudates detection based on superpixel multi-feature extraction and patch-based CNN

Chenxi Huang
School of informatics, Xiamen University, China
supermonkeyxi@xmu.edu.cn

Yongshuo Zong
Department of Computer Science, Tongji University, China
zongys@outlook.com

Yimin Ding
School of Electronic and Information Engineering, Tongji University, China
1940161867@qq.com

Xin Luo
School of Electronic and Information Engineering, Tongji University, China
907383882@qq.com

Kathy Clawson
School of Computer Science, University of Sunderland, St Peter Campus， United Kingdom
Email: Kathy.Clawson@sunderland.ac.uk

Yonghong Peng *
Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, United Kingdom
Email: Y.Peng@mmu.ac.uk

Abstract Diabetic Retinopathy (DR) is a severe complication of chronic diabetes which causes significant visual deterioration and, when coupled with delayed treatment, may lead to blindness. Exudative diabetic maculopathy, a form of macular edema where hard exudates (HE) develop, is a frequent cause of visual deterioration in DR. The detection of HE comprises a significant role in the DR diagnosis. In this paper, an automatic exudates detection method based on superpixel multi-feature extraction and patch-based deep convolutional neural network is proposed. Firstly, candidate superpixels are generated on each resized image using the superpixel segmentation algorithm called Simple Linear Iterative Clustering (SLIC). Then, 25 features extracted from resized images and patches are generated on each feature. Patches are subsequently used to train a deep convolutional neural network, which distinguishes hard exudates from the background. Experiments conducted on three publicly available datasets (DiaretDB1, e-ophtha EX and IDRiD) demonstrate that our proposed methodology achieved superior HE detection when compared with current state-of-art algorithms.

Keywords: retinal hard exudates, superpixel, feature extraction, deep learning, automatic diagnosis.

1. Introduction

Diabetes is regarded as one of the most ubiquitous chronic diseases in the world [1], [2] and the number of diabetes patients has exceeded 400 million world-wide [3]. Diabetic Retinopathy (DR) is a serious complication of diabetes characterized by specific fundus changes, including blot hemorrhages, microaneurysms and hard exudates. DR can cause blindness in severe cases [4], and therefore the early diagnosis and treatment of DR is of great significance.

Hard exudates (HE) are one of the most important features of DR. HE appear as massive or dot-shaped highlights, and are caused by leakage of macromolecular substances (lipid and proteinaceous material) from retinal vessels into the eyeball after the blood vessel wall is damaged. Detection of HE, achieved via fundus image examination, is considered an effective method for DR diagnosis, with early detection contributing to improved patient outcomes and the reduction of medical costs [5]. Figure 1 shows an example of a retinal fundus image with hard exudates.

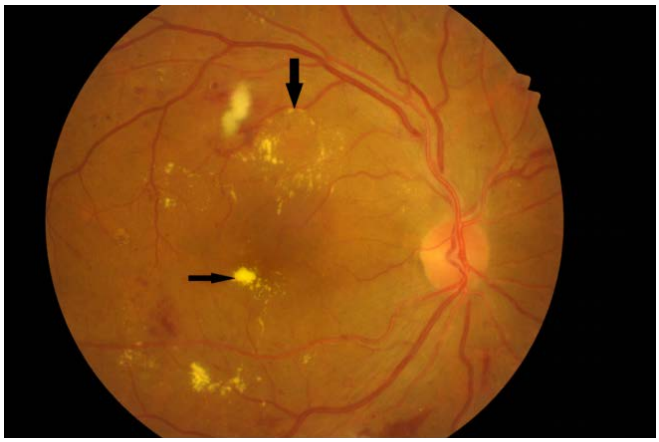


Fig. 1. An example of retinal fundus image. Hard exudates are the yellow-white spots marked by black arrows.

In clinical diagnosis, the quality of collected fundus imagery is often limited by environmental constraints and physiological structures, such as lighting and equipment conditions, making it difficult for doctors to accurately and efficiently analyze the raw images directly [6][7]. Therefore, it is necessary to propose an automatic HE detection methodology to aid clinical diagnosis.

Early medical image processing technologies mainly focused on reduction of equipment noise, detection of lesion area and contrast enhancement, thus improving the

image quality or enhancing certain information within the image [8][9]. However, traditional methods often cannot fully and accurately express the features of the target object due to limitations of manually-extracted features [10].

Current HE detection methods may be categorized under four main approaches: morphology-based [11]-[14], threshold-based [15]-[18], cluster-based [19]-[22], and region-based [23]-[25]. The application of graph theory and machine learning to HE detection, although widely explored, is constrained due to insufficient numbers of retinal samples. Small sample sizes reduce the effect with such methods can achieve satisfying results. Deep learning method, as a widely studied technology in recent years, is able to learn and extract potential features of target objects effectively [26] from original images. Compared to traditional methods, deep learning has stronger generalization ability and better robustness through network feature extraction. However, current studies evaluating deep learning for hard exudate detection are limited by the lack of image data available [27][28].

In order to solve the above problems, this paper presents an automatic HE detection method using patch-based deep Convolutional Neural Networks (CNN) for multi-feature classification, which has effectively expanded the training samples and comprehensively characterized the images.

The proposed method consists of four steps: (1) after image resizing, the SLIC superpixel segmentation algorithm is utilized for retinal image segmentation, and superpixels are generated as candidates; (2) a total of 25 features (both pixel level and superpixel level) are extracted to characterize the candidates; (3) patches are generated on each extracted feature; and (4) a CNN is established and trained to distinguish exudates from normal background. Experiments are conducted on three public datasets to evaluate the effectiveness and robustness of the proposed method.

Our main contribution may be summarized as follows: (1) superpixel is applied as another interpretation of the image, which treats the pixels in areas with similar characteristics as an entity and therefore reduces noise disturbance and improves stabilization; (2) a considerable number of patches are generated on each extracted feature and used as inputs for CNN learning, thus expanding the training samples largely and making deep learning methods feasible with small sample retinal databases.

The remainder of this paper is organized as follows. Section 2 discusses the related studies. Our proposed method is elaborated in Section 3. Section 4 presents experimental results and analysis. Conclusions and avenues for future work are offered in Section 5.

2. Related Work

In broad terms, we categorize the prior work in this field into Traditional Approaches, Superpixel-Based Approaches, and Deep Learning Approaches.

2.1 Traditional Approaches

Traditional HE detection methods can be divided into four main categories, specifically morphological-based segmentation, threshold-based segmentation, cluster-based detection, and region-growth detection. In addition to four main categories of approach, a number of novel methods have been proposed and are discussed below.

(1) Morphological-based segmentation. Walter et al. [11] used operations such as local window variance calculation combined with morphological operations to accurately segment the contours of the target. After image pre-processing and noise reduction, Imani et al. [12] used morphological component analysis to match and divide normal physiological structures of a fundus image from the lesion area, and subsequently segmented HE using Kirsch edge detection. Sreng et al. [13] applied maximum entropy thresholding combined with morphological reconstruction, to segment foreground and background regions (on the green channel) and to obtain the hard exudate segmentation results. Welfer [14] adopted a morphological-based method utilizing a stepwise refinement strategy. While morphological methods have the advantage of high computational efficiency and fast execution, they only consider brightness / grayscale information and ignore other exudate characteristics. Such approaches are therefore very sensitive to noise. Despite achieving good performance, when using morphological methods the key issue of choosing structural elements of appropriate size still exists.

(2) Threshold-based segmentation. Threshold-based segmentation is a simple but widely used method. In order to segment the hard exudate from the background, Sánchez et al. [15] adopted a dynamic threshold technique. To avoid the influence of contour pixels, the final threshold is determined according to the Gaussian components of two higher mixing weights. Phillips et al. [16] utilize image sharpening and shadow correction

technology, combined with local thresholding and global thresholding technology to achieve automatic HE detection. García et al. [17] used a combination of global threshold and adaptive threshold to segment candidate HE regions, and then determined the true exudation region through a series of features and radial basis functions (RBF). Sharib Ali et al. [18] proposed a method based on the statistical atlas to segment the exudate, obtained the distance map through the average atlas image, and then introduced the post-processing scheme to segment the exudate. However, it is not easy to choose a suitable and robust threshold due to the inconsistency of image brightness and contrast caused by the imbalance of illumination in the fundus image.

(3) Cluster-based detection. Within the literature, clustering methods have been applied to classify and quantify image collections, for example to calculate the area of target regions of interest, and to identify regions with specific characteristics.. Kumari et al. [19] used a text clustering algorithm to divide the fundus image into background class and bright target class for HE segmentation. Osareh et al. [20] used fuzzy C-means clustering to segment the HE after color normalization and local contrast enhancement on the color fundus image. Zhang et al. [21] utilized a combination of local contrast enhancement and fuzzy C-means to cluster images in *LUV* color space. Xie et al. [22] proposed a genetic algorithm and clustering combined retinal vascular network segmentation method. Despite their common application, clustering algorithms are usually sensitive to noise, and the location and characteristics of the center of the class are unknown, so a priori assumptions are required. If the initial center selection is not appropriate, both algorithm convergence and segmentation performance will be poor.

(4) Region-growth detection. This method first sets the appropriate seed point, and then expands the seed point according to the brightness and edge characteristics of the visual cup to obtain the visual cup outline. Sinthanayothin et al. [23] used a recursive region growing and segmentation algorithm, combined with a "moat operator", to automatically detect the characteristics of DR. Li et al. [24] compared detection across four color spaces, namely RGB, *Lab*, *Luv*, and *HVC*. *Luv* color space was used for final HE detection. Lowell et al. [25] used template matching and a deformable contour model to locate and segment the optic nerve head boundary. The disadvantage of the area growth methods is that it often causes over-segmentation, and the common noise and grayscale unevenness in the fundus image will exacerbate this tendency. Moreover, the space and time costs of

the algorithm are relatively large, and the shadow part in the image often interferes greatly with the detection accuracy.

(5) Additional methods. Giancardo et al. [29] proposed a 3-level detection technique, which applied: image histogram equalization for candidate region extraction; multiscale local binary pattern feature extraction; and classifier based training using diseased regions for DR diagnosis. Vimala et al. [30] extracted the four features of energy, difference, homogeneity, and standard deviation based on the gray level co-occurrence matrix, and used support vector machines to classify HE. Ege et al. [31] used Bayesian methods, Mahalanobis distance, and KNN classifiers to detect exudates. Niemeijer et al. [32] applied lesion probability maps to distinguish HE from other bright structures.

2.2 Retinal fundus detection using superpixels

The above-mentioned HE detection methods analyze image pixels discretely. This may be regarded as counterproductive. Combinations of pixels within images may be viewed as natural entities, at least regionally. Moreover, shape variegation and sharp edges are specific features of hard exudates. It is therefore desirable to consider the boundaries of exudates when developing robust and accurate detection methodologies.

The concept of a superpixel, proposed by Ren et al. [33], provides an alternative representation of images whereby individual pixels are aggregated. Superpixels may be viewed as local image subregions that are consistent and capable of maintaining certain local structural features. Superpixel representation can effectively reduce pixel level noise disturbances and provide a basis for the rapid and accurate diagnoses of hard exudates.

A novel superpixel segmentation algorithm called Simple Linear Iterative Clustering (SLIC) is proposed by Achanta et al. [34]. SLIC algorithm can efficiently generate compact and uniform superpixels with customized region size and regularity, achieving good boundary compliance and recall.

Borsos et al. [35] used an improved SLIC algorithm to generate uniform superpixels to segment retinal white lesions. Sheng et al. [36] used superpixels as the basic unit of the vessel segmentation scheme to detect low-contrast narrow blood vessels. Yan et al. [37] used superpixels to preprocess the fundus image and designed a method for detecting different anomalies at the pixel level from different retinal image modes without adjusting

parameters. Zhou W et al. [38], segmented the image into superpixels and extracted multi-channel features for further classification.

2.3 *Detection method based on deep learning*

Deep learning for HE recognition has been applied in both supervised and unsupervised contexts. In many practical implementations, a mixed network structure is often utilized. Advantages of such methods include: the effective use of data for distributed training; easier learning of latent feature expressions than traditional methods; and enhanced segmentation accuracy when differentiating between blood vessels and a variety of lesions and potential regions of interest. .

Prentašić et al. [39] partition the fundus image into 65x65 blocks (the center pixel of the block is the pixel to be detected), to construct the input sample of the depth network, and employ a 10 layer convolutional network to for pixel level HE detection. Maji et al. [40] use convolutional neural networks for block-level blood vessel extraction. Niemeijer et al. [41] propose an automated detection system based on deep learning, which can distinguish between fundus exudate and cotton plaque and has a performance level close to that of a retina expert. Yu [42] identify candidate regions using morphological methods, followed convolutional neural network detection. The identification of candidate regions prior to learning reduces the required number of network calculations, but introduces the potential for error propagation.

Although many algorithms for the detection of hard exudates have been proposed, one problem remains to be solved. The majority of existing deep methods focus on image enhancement or extract features manually, thus lacking of deep feature extraction of pertinent information content within fundus photos. Due to the lack of availability of large image datasets, deep learning methods have not been fully utilized for HE detection. Therefore, the deep level features and other complex content are scarcely extracted by the previous works.

3. Method

Our proposed HE detection methodology encompasses the following steps: (1) superpixel generation, (2) multi-feature extraction, (3) patch generation, and (4) CNN classification. After resizing all images to 512×512 resolution, SLIC is utilized for retinal

image segmentation and superpixels are generated on each resized image. Subsequently, a total of 25 features (both pixel level and superpixel level) are extracted to characterize the candidates and sample training patches are generated from every feature. Finally, a convolutional neural network is established and trained to predict whether the central pixel of the patch belongs to either the background or hard exudates. The process of our method is illustrated in Figure 2.

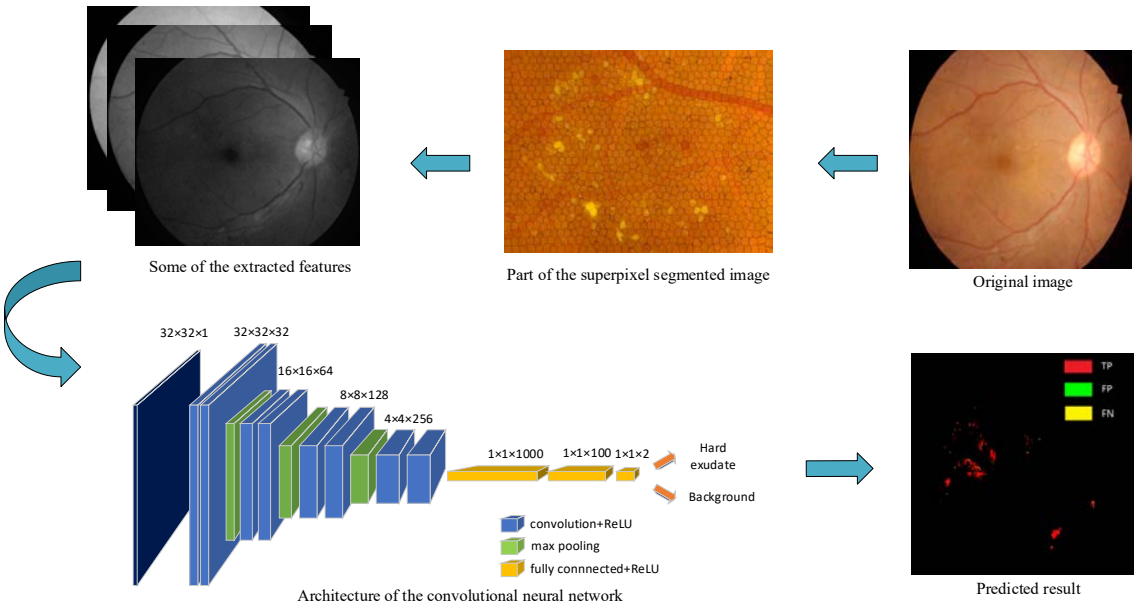


Fig. 2. Proposed methodology

3.1 Superpixel generation

Given its previously reported efficiency and high boundary recall, we adopt the SLIC [34] algorithm for superpixel generation. To acquire superpixels with regular size and compact structure, the following distance measure is adopted. For a given image I, represented in CIE Lab color space with number of pixels N, we initialize M cluster centers and define the grid interval G, where $G = \sqrt{M/N}$. Given pixel m and n-th cluster center, the color distance d_{LAB} and the coordinate distance d_{XY} are calculated as

$$d_{LAB} = \sqrt{(l_m - l_n)^2 + (a_m - a_n)^2 + (b_m - b_n)^2} \quad (1)$$

$$d_{XY} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2} \quad (2)$$

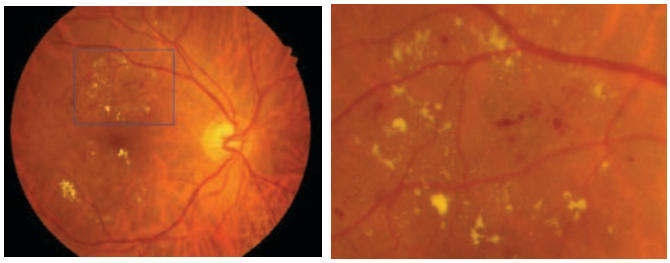
where (l_m, a_m, b_m) is the color value vector of pixel m, and (x_m, y_m), (x_n, y_n) are respectively the corresponding position of the pixel m and the cluster center in the original image. Thus, at the xy coordinate system, the distance between the n-th cluster center and the nearest pixel m which is denoted as D_{mn}, could be calculated within a vicinal area of $2G \times 2G$:

$$D_{mn} = \sqrt{d_{LAB}^2 + (\frac{\omega}{G})^2 d_{XY}^2} \tag{3}$$

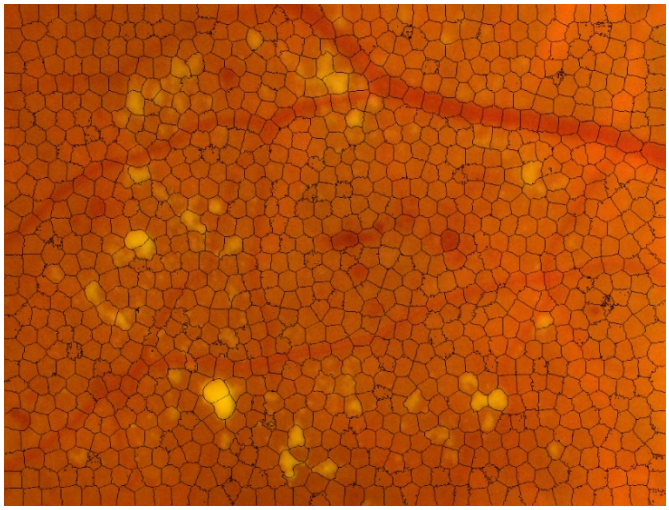
where ω is weight factor introduced to control the compactness of a superpixel.

Next, the new generation of the cluster centers is determined by the distinguished centroid of the pixels, which is iteratively repeated until the distance between the neighboring cluster centers becomes constant.

In addition, there are two parameters to be set in the generation process: the regularizer and the region size. The former represents the spatial proximity and the latter is related to the size of superpixel segmentation. A larger regularizer results in more compact superpixels, while a smaller regularizer means less regular shape and size, which can influence both computational complexity and classification performance. Considering both the time consumption and accuracy [38], we select the region size of 30 and the regularizer of 0.001. Figure 3 shows the segmentation results under this condition.



(a) original image (b) cropped image



(c) segmentation result with region size of 30
and regularizer of 0.001

Fig. 3. SLIC process

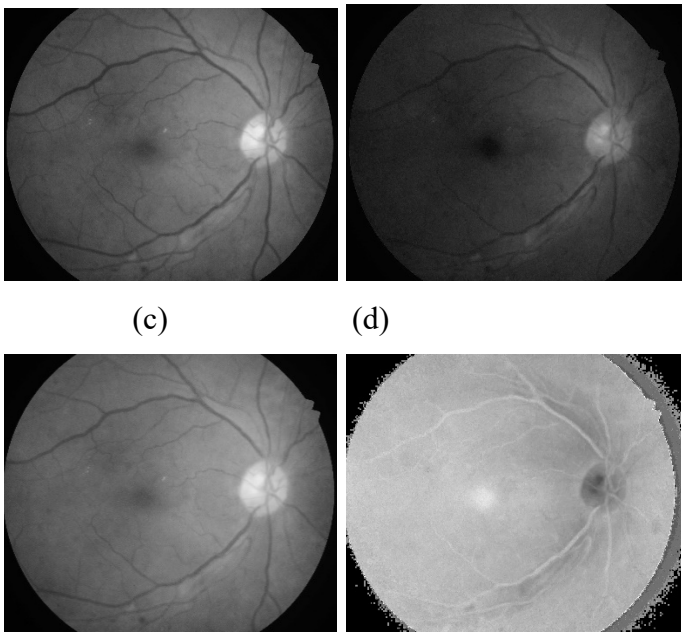
3.2 Multi-feature extraction

Due to uneven illumination and contrast of the fundus images collected, it is difficult to obtain satisfying results directly using the original image through deep learning methods. Therefore, in addition to extracting different color channels from the original image, which is denoted as pixel level features, we also use superpixel as another representation of images. The features extracted at the levels of both pixel and superpixel can fully describe the characteristics of sharp edges and high contrast of hard exudates.

3.2.1 pixel level features

As Figure 4 shows, the background contrast of exudates varies with different channels. Hence, it is beneficial to extract a set of features for each pixel in different color channels to better characterize the image.

(a) (b)



(c) (d)

(e) (f)

Fig. 4. Different color channels of the original image.
(a) original image, (b-d) red, green, blue channel image
(e) grayscale image, (f) saturation channels image.

Optical artifacts are inevitably introduced by the fundus cameras and may largely interfere with the classification process. Such artefact may not appear in RGB color space, but can be detected in the saturation channel [43] of *HSV* color space. Futhermore, while saturation values usually follow the change of imaging sources, other structures in the saturation channel are brighter than the reflections for the same type of image, which suggests the normalization of the saturation channels becomes feasible. Hence, by separating the saturation channel with its global mean value, the normalized saturation feature is obtained and coded as sat.

Grayscale image and red, green, blue channel images are selected as additional pixel level features to compensate the lost information in saturation channel.

3.2.2 Superpixel level features

It is critical to take exudate boundaries into consideration in the process of detection, because boundary shape and acute edge are specific HE features which cannot be detected through discrete pixel analysis alone. Therefore, superpixel level features are extracted in addition to pixel level features, to reflect the local structural features in small regions.

While the blue channel is darker and the red channel is more saturated, green channel [44] is usually adopted in retinal image. Green channel is helpful to extract feature[45] and reduce the interference of blood vessels in the classification process [46], due to the highest contrast it has between the background and the blood vessels. Simultaneously, the intensity of the green channel image may vary largely between different images due to the Photographing environment. Hence, characterizing the retinal image from different perspectives in green channels is necessary. For each green channel image, the standard deviation and normalized mean value in each candidate (coded as STD and μ) are obtained for extracted features.

In order to differentiate the exudate candidates and the non-exudate candidates and better characterize the candidates, the contextual feature is proposed. The contextual feature [38], calculated in green channel after superpixel generation, is adopted. Assume that the quantity of the spatial compact candidate regions with the similar size is Q and are coded as $Reg_i(i = 1,2,3, \dots, Q)$. d_j is defined as the intensity distance, which is calculated by:

$$d_j = avg\_Reg_i - avg\_Reg_j \tag{4}$$

where avg_Reg_i denote the average gray value of the i-th superpixel region.

We also calculate s_j , the spatial distance between Reg_i and its neighbor region Reg_j, corresponding to d_j:

$$s_j = sqrt(||b_i - b_j||_2^2)\ j \epsilon P(i) \tag{5}$$

where $||\cdot||_2^2$ denotes a quadratic term of the l_2-norm and b_i denotes the barycenter position of the i -th superpixel region.

Based on the equations above, the contextual feature C_i can be calculated by:

$$C_i = \frac{1}{P_i} \times \frac{avg\_Reg_i}{avg} \times \sum_{j \epsilon P_{(i)}} \frac{d_j}{s_j} \tag{6}$$

where P_i denotes the total number of pixels in Reg_i and avg denotes the global mean value of green channel image. The neighbor of Reg_i is defined as the superpixels whose barycenters locate in a round area, which is 7 times of the region size. Thus, the set of neighbors of Reg_i is coded as $P(i)$. Figure 5 shows a concept picture of the contextual feature, where the current candidate region Reg_i is enclosed by the blue dotted line, and the yellow triangle demonstrates the position of the barycenter of Reg_i . The

corresponding neighbor region is enclosed by the blue circle, where red dots indicates the barycenter position of the neighbor region.

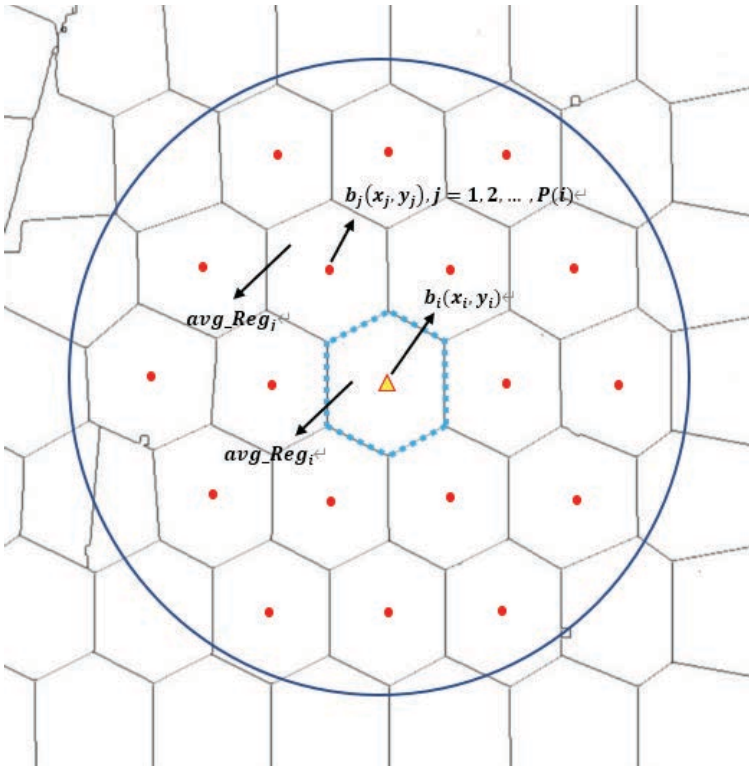


Fig. 5. Illustration of the contextual feature.

To remove the effect of varying retinal pigmentations and different image acquisition processes, the global mean parameter is adopted. To our knowledge, the avg_Reg_i/avg and the gray distance will be larger when Reg_i belongs to bright structures and vice versa. Thus, the total weighted gray distance will be larger if the bright structure turns out to be an exudate. **Equal** (5) demonstrates that the larger the contextual feature C_i is, the candidate region is more probably to be an exudate.

Figure 6 illustrates the original image segmentated by SLIC algorithm, and maximum, minimum, median and mean value of the candidate in grayscale image.

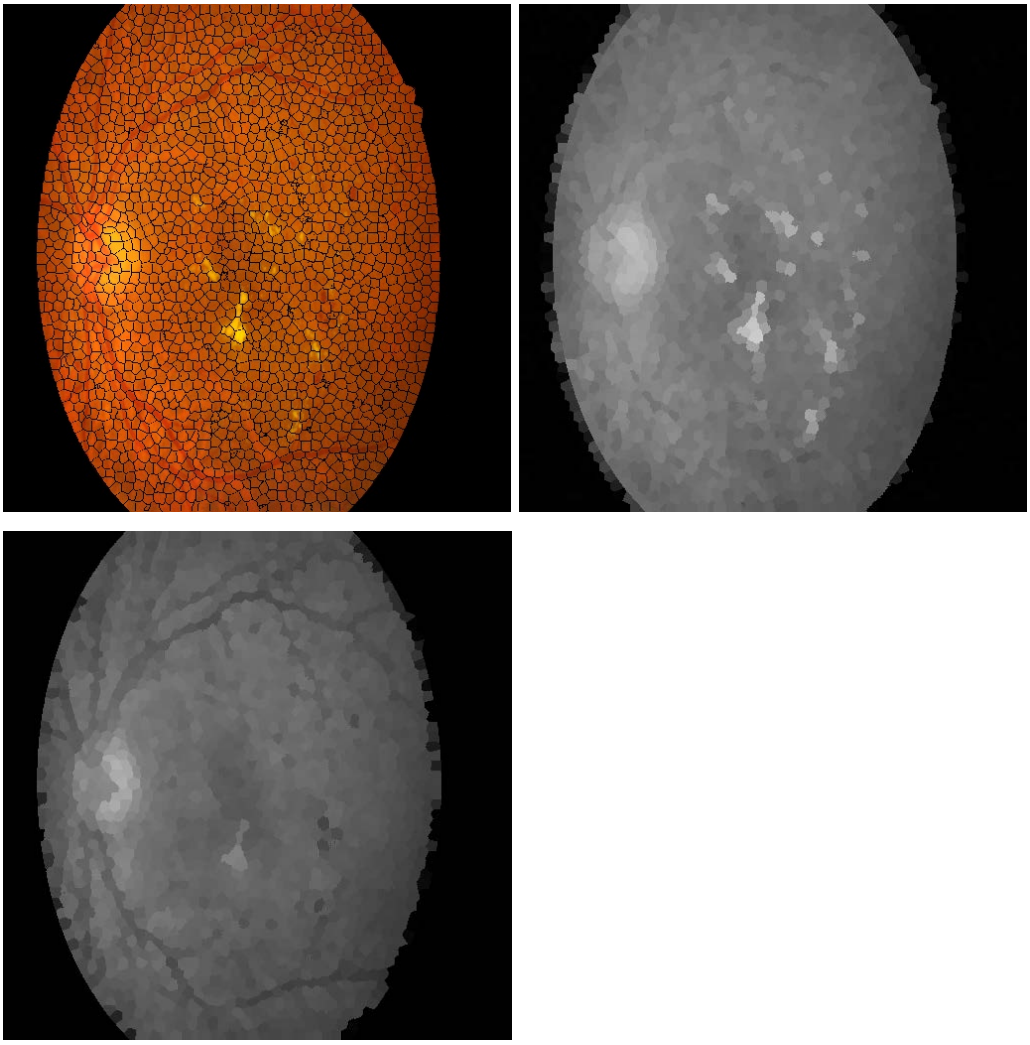


Fig. 6. Segmentation results of SLIC

Given that the background contrast of exudates vary across different channels, we extract features across multiple color spaces and channels. Specifically, a total of 25 features (grayscale, RGB and saturation) are extracted from the original images for each candidate superpixel. A full list of calculated features is presented in Table 1. Each feature calculated has the same resolution as the original image.

Feature	Description	Feature Level
gray	gray scale image	Superpixel level
R	red channel of the original image	

G	green channel of the original image	
B	blue channel of the original image	
Sat	normalized saturation channel of the original image	
max_gray max_R max_G max_B	maximum value of the candidate in gray, R, G, B images, respectively.	Pixel level
min_gray min_R min_G min_B	minimum value of the candidate in gray, R, G, B images, respectively.	Pixel level
μ_gray μ_R μ_G μ_B	mean value of the candidate in gray, R, G, B images, respectively.	
M_gray M_R M_G M_B	median value of the candidate in gray, R, G, B images, respectively.	
μ_S	mean value of the candidate in saturation image	

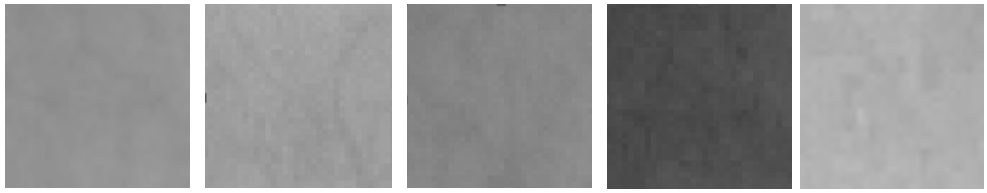
μ	normalized mean value of the candidate in G
STD	standard deviation of the candidate in G
C	contextual feature

Table 1. Summary of extracted features

3.3 Patch generation

Due to the limited number of the retinal images, it is ineffective to apply deep learning method directly on the extracted features. Therefore, we proposed a patch-based approach for network training.

After multi-feature extraction, training sample patches of resolution 32×32 are generated on each extracted feature. A resolution of 32×32 is chosen empirically because it includes the necessary adjacent information around the hard exudates whilst excluding interference. By comparing with ground truth, the patches are labeled and divided into two groups: background and hard exudates. If the central pixel of a patch (coordinated as $(17,17)$) belongs to the exudates, the image is labeled as 1; otherwise, the image will be labeled as 0. On each feature, 2500 patches of background and 2500 patches of hard exudates are generated. Figure 7 shows some of the generated patches in gray, red, green, blue and saturation channel, which belong to either background or hard exudate.



(a) background patches

(b) hard exudates patches

Fig. 7. Generated patches

To circumvent issues associated with patch-region definition around fundus image edges, zero padding is applied as illustrated in Figure 8. Within Figure 8, the blue border represents the pixel of zero padding, and the white box and black box respectively represent the patches of background and hard exudate generated from grayscale image. In this way, every pixel of the images can be predicted.

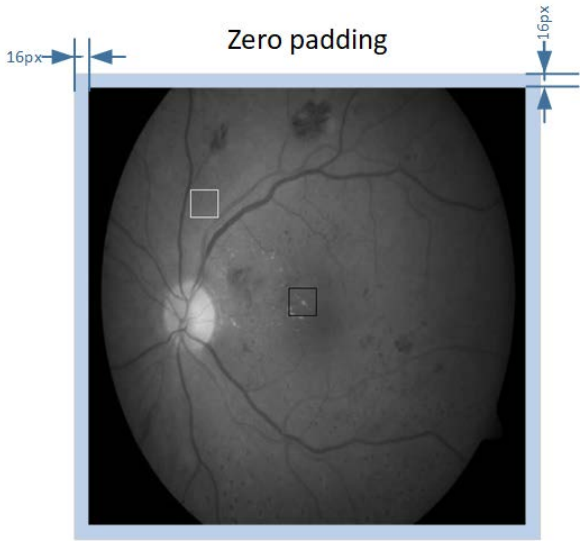


Fig. 8. Zero padding and patch generation

3.4 CNN classification

After patch generation, an 8-layer CNN network, as presented in Figure 9, is developed and trained on the patches. The first layer is the input patch. Then, there are two convolutional layers with ReLU, followed by a maxpooling layer. By applying maxpooling methods, the size of feature map is reduced to half after every two layers but kept constant on the last layer. After that, fully connected layers with ReLU are applied. As Figure 9 illustrates, max pooling layers are introduced after every second convolutional layer (with the exception of the last), to reduce the feature map size. The CNN is trained so that prediction results indicate whether the central pixel of the input patch is hard exudate or background. In this way, we convert a whole image detection task to a local binary classification task, which is more operable and practical.

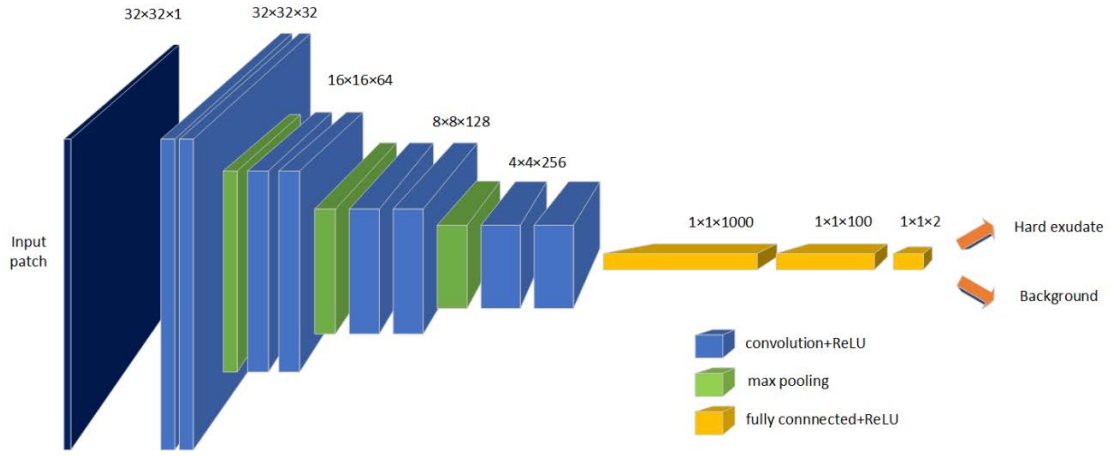


Fig. 9. architecture of the convolutional neural network

Drop out , with a rate of 0.5, is adopted to reduce the complexity of neural networks and to avoid overfitting [47], and our network is trained using the Adam optimizer (learning rate $= 0.0001$) and with Cross Entropy loss function. To improve training efficiency, mini-batch normalization is used, where batch size = 50. The training images are grouped into 5 sets, each of which is trained for 200 epochs, and training accuracy of 99.74% is achieved by the network.

The network predicts whether the central pixel of each patch belongs to the background or hard exudates. Then these predictions are reconstructed to an image of 512×512. In this way, prediction results can be obtained and compared to the ground truths, which are also resized to the resolution of 512×512.

4. Experimental results and discussions

We evaluate our HE detection pipeline using three diabetic retinopathy image databases, as described in Section 4.1, and present a discussion and comparison of results (Section 4.2) to demonstrate the superiority of our approach.

4.1 DATABASES

The DiaretDB1[48] database is a publicly available database, which gathers high-quality retinal fundus images of diabetic retinopathy. In this dataset, the fixed resolution of the database is 1500×1152 with the field of view of 50°. A total of 89 fundus images are divided into 47 images of the training set and 42 of the testing set.

The e-ophtha EX [49] database is an open-source database proposed by Zhang et al, which contains color eye fundus with four different resolution varying from 2544×1696 to 1440×960 in a 45° field of view. This database involves a total of 82 color fundus images with 47 images with exudates and 35 images without exudates, and experts' annotations are given on the 47 images. Randomly, 40 images are selected to comprise the training set and 42 images are selected for testing set.

IDRiD (Indian Diabetic Retinopathy Image Dataset) [50] is a public database that consists of typical diabetic retinopathy lesion images and retinal fundus structures at a pixel level. A total of 81 images from IDRiD are divided into 54 images of the training set and 27 of the testing set. The images have a resolution of 4288×2848 pixels, with 50° field of view.

Among these datasets, DiaretDB1 is annotated in lesion-level, while IDRiD and e-ophtha are annotated in pixel-level.

4.2 RESULTS AND ANALYSIS

A. Metrics

Within existing literature, evaluation metrics relating to HE detection in DR images may examine image-based criterion or pixel-based criterion. The former regards HE detection an image classification problem, with accuracy calculated at image level. The latter focuses on image segmentation, whose accuracy should be considered at pixel-level. In the image-based situation, an image without hard exudate is regarded as healthy, and regarded as diseased if there is any hard exudate detected. Three common criteria are adopted to evaluate the performance of the method:

$$Accuracy = \frac{TP+TN}{P+N} \tag{7}$$

$$Specificity = \frac{TN}{TN+FP} \tag{8}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{9}$$

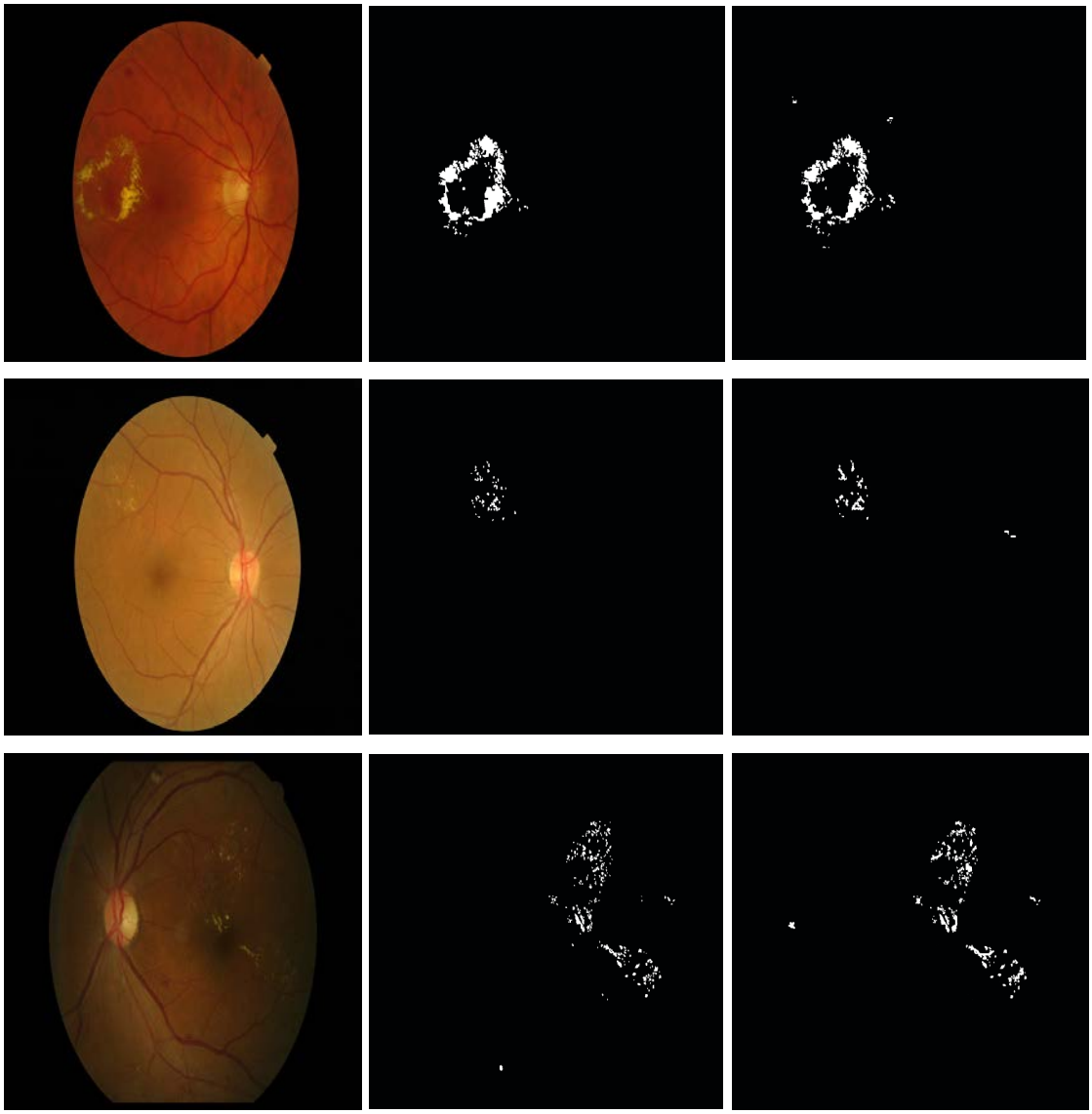
where TP ($True\ Positive$) denotes the diseased images that are correctly found; FP ($False\ Positive$) denotes the diseased images that are wrongly found; TN

($True\ Negative$) denotes the healthy images that are correctly found, and FN ($False\ Negative$) denotes the images that are wrongly found as healthy ones. In pixel-based scenarios, TP denotes the correctly detected hard exudates, FP denotes the background that are correctly detected, TN denotes the non-exudates that are wrongly detected as exudates and FN denotes the exudates that are wrongly detected as non-exudates.

Furthermore, the Receiver Operating Characteristics (ROC) curve is introduced to validate the effectiveness of the proposed method. In the ROC graph, the horizontal axis consists of 1-specificity and the vertical axis consists of sensitivity. The AUC is the value of the area under the ROC curve.

B. Pixel-based experiments

For each database, the training set is divided into 5 groups, each of which is trained for 200 epochs with early stop. 25 features of each testing image are predicted by the network and therefore 25 prediction results are acquired for each pixel. The final prediction result of each pixel is determined through considering all of the 25 prediction results and choosing the value with a larger proportion. The pixel-based experiments are carried out on the e-ophtha EX and IDRiD databases. On the e-ophtha EX database, an example of the prediction results is shown in Figure 10, where the first column (a) are the resized original images, the second column (b) illustrate corresponding ground truths and the last column (c) represents the predicted images.



(a) resized original images (b) corresponding groundtruths (c) predicted images

Fig. 10. Example of the prediction result

System accuracy, specificity, sensitivity and AUC are displayed in Table 2. As the results indicate, while Zhou's method obtains slightly higher AUC when $region\ size = 10$ and $regularizer = 0.001$, ours performs better when $region\ size = 30$ and $regularizer = 0.001$. As discussed above, a larger region size means less computation time consumption. According to Zhou et al., the computation time when $region\ size = 10$ is longer than the case when $region\ size = 30$ by approximately 90 times. Thus, our method achieves higher AUC more efficiently.

Approach (Regisonsize, regularizer)	Sensitivity	Specificity	Accuracy	AUC
Zhou et al [30]. (10, 0.001)	-	-	-	0.9684
Zhou et al. [30]. (30, 0.001)	-	-	-	0.9670
Ours (10, 0.001)	98.33%	91.17%	97.65%	0.9703
Ours (30, 0.001)	97.96%	90.84%	97.58%	0.9682

Table 2. Results comparisons of pixel-based criterion on e-ophtha EX

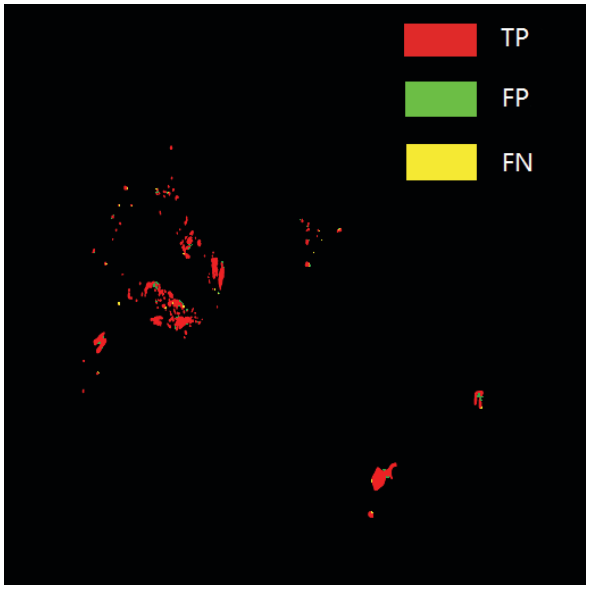


Fig. 11. result of pixel-based validation.

On the IDRiD database, the exudate detection result of IDRiD_76_EX from IDRiD database the using pixel-based criteria is shown in Figure 11, where TP is marked as red, FP as green and FN as yellow. As Figure 9 indicates, most of the exudate lesions are correctly detected as TP (in red), and a small part of pixels are predicted as FP (in green) and FN (in yellow). The comparisons of the results of pixel-based experiments on the IDRiD are displayed in Table 3. As it indicates, our approach achieves high diagnostic accuracy than 35], [47], [51] and [52], but with a smaller proportion of false

positive predictions than [47]. Therefore, the comprehensive performance of our method in sensitivity and specificity is superior. For realistic applications, higher sensitivity means a larger percentage of successfully detected lesions, and higher specificity means a greater proportion of healthy parts successfully judged as healthy parts. The satisfying results in these two parameters demonstrate the strong practicality in clinical diagnosis.

Approach	Sensitivity	Specificity	Accuracy	AUC
Borsos [35]	62.42%	98.99%	-	-
Benzamin [47]	98.29%	41.35%.	98.6%	-
Derksen [51]	-	-	89.13%	-
Farooq [52]	-	-	96.35%	-
Ours	98.40%	90.67%	98.19%	0.9674

Table 3. The comparison of pixel-based experiments on IDRiD database.

C. Image-based experiments

The image-based experiments are conducted on DiaretDB1 and IDRiD database. The comparisons between Zhou's method and ours on DiaretDB1 are listed in Table 4. As demonstrated in Table 4, the % sensitivity achieved by our approach is 98.01%, which constitutes an increase of 10.21% in comparison to Zhou et al., [38] and an increase of almost 28% in comparison to Welfer et al. [14]. In the experiments on the IDRiD database, we have also achieved approving results in AUC of 0.9650 and accuracy of 98.93%.

Authors	Sensitivity	Specificity	Accuracy
Welfer et al.[14]	70.48%	98.84%	
Zhou et al.[38]	88%	95%,	-
Ours	98.21%	91.38%	98.01%

Table 4. Comparison of the results of image-based criterion.

D. Patch size

We determined the approximate range of patch size empirically at first, then compared patch sizes of 24×24, 32×32, and 48×48, respectively. Our comparative analysis is shown in table 5. For pixel-based experiments, we selected the highest performing patch size of 32×32.

Patch size	Sensitivity	Specificity	Accuracy
24×24	97.13%	88.79%	96.64%
32×32	98.40%	90.67%	98.19%
48×48	96.42%	89.35%	97.08%

Table 5. Comparison of the results of image-based criterion.

As shown in table 5, the patch size of 32×32 is superior, because size of 24×24 is too small to include all necessary area to predict the central pixel, and size of 48×48 is relatively too large that it contains irrelevant areas that may interfere the prediction.

E. Zero-padding

In order to verify the necessity of zero-padding, we have also conducted the pixel-based and image-based experiments on IDRiD database without zero-padding. The results are shown in table 6 and 7.

Method	Sensitivity	Specificity	Accuracy
without zero-padding	97.36%	90.48%	97.62%
with zero-padding	98.40%	90.67%	98.19%

Table 6. Comparison of the results of pixel-based criterion.

Method	Sensitivity	Specificity	Accuracy
with zero-padding	98.21%	91.38%	98.01%

without zero-padding	98.21%	91.38%	98.01%

Table 7. Comparison of the results of image-based criterion.

For pixel-based experiments, the experimental result of sensitivity after zero-padding is 1.04% higher than that without zero-padding. That is because the process of zero-padding enables the prediction of hard exudates near the border, which increases the TP and decrease the FP. However, the results remain the same in image-based experiments, since the hard exudates never only appear in the area near the border in the database.

F. Analysis

The advantages of the proposed method can be attributed mainly to the following aspects. First, the number and characteristics of our extracted features is broad and covers multiple aspects of the image features. This in turn contributes to more accurate detection results. Furthermore, the hard exudate boundary recall method is applied in superpixel segmentation, where the proportion of the boundaries of ground truth images that fall in a superpixel boundary is calculated. Therefore, the performance of detection on the edges is significantly improved, thus leading to better accuracy. Finally, our Multi-feature extraction and patch generation methodologies sufficiently expand the training set, thus overcoming the difficulties associated with a lack of training images. When coupled with drop out, we are able to further circumvent problems such as overfitting.

5. Conclusion

This paper presents a novel hard exudate detection method of DR diagnosis based on multi-feature extraction and patch-based CNN. SLIC segmentation is adopted to acquire compact and regular-shaped superpixels, and 25 well-designed features are extracted to comprehensively describe the unique characteristics of retinal fundus images. After feature extraction, training patches are generated and used as inputs for CNN learning. The experiments, conducted on three public databases (e-ophtha EX, DiaretDB1 and IDRiD) demonstrate the stability and superiority of our proposed method, both for image-based and pixel-based classification. Future studies can be extended to two main aspects. First, different patch sizes can be tested to evaluate its impact on training.

Second, the features extracted from the same image can be assembled into an n-dimensional image as the input, and apply feature selection for the training.

Acknowledgments

This work was supported in part by the National Science and Technology Support program under grant 2015BAF10B01.

References

[1] Group T D . The Effect Of Intensive Treatment Of Diabetes On The Development And Progression Of Long-Term Complications In Insulin-Dependent Diabetes Mellitus[J]. New England Journal of Medicine, 1993, 329(14):977-986.

[2]Adrian Galdran,Hadi Chakor,Abdulaziz A. Alrushood,Ryad Kobbi,Argyrios Christodoulidis, Jihed Chelbi, Marc-André Racine, Ismail Benayed. Automatic classification and triage of diabetic retinopathy from retinal images based on a convolutional neural networks (CNN) method[J]. Acta Ophthalmologica,2019,97.

[3] Huang Chuang-Xin,Lai Kun-Bei,Zhou Li-Jun,Tian Zhen,Zhong Xiao-Jing,Xu Fa-Bao,Gong Ya-Jun,Lu Lin,Jin Chen-Jin. Long-term effects of pattern scan laser pan-retinal photocoagulation on diabetic retinopathy in Chinese patients: a retrospective study.[J]. International journal of ophthalmology,2020,13(2).

[4] Hui Wang,Guohui Yuan,Xuegong Zhao,Lingbing Peng,Zhuoran Wang,Yanmin He,Chao Qu,Zhenming Peng. Hard exudate detection based on deep model learned information and multi-feature joint representation for diabetic retinopathy screening[J]. Computer Methods and Programs in Biomedicine,2020,191.

[5] Sinthanayothin C , Boyce J F , Williamson T H , et al. Automated detection of diabetic retinopathy on digital fundus images[J]. Diabetic Medicine, 2002, 19(2):105-112.

[6] Kang Yafen, Fang Ying, Lai Xiaobo. Automatic Detection of Diabetic Retinopathy with Statistical Method and Bayesian Classifier. 2020, 10(6):1225-1233.

[7]Suriyasekeran Karkuzhali, Santhanamahalingam Senthilkumar, Duraisamy Manimegalai. Algorithms for Diagnosis of Diabetic Retinopathy and Diabetic Macula Edema- A Review.[J]. Advances in experimental medicine and biology,2020.

[8]Rezatofighi S H,Roodaki A,Ahmadi Noubari H. An enhanced segmentation of blood vessels in retinal images using contourlet.[J]. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference,2008.
[9]S Karkuzhali,D Manimegalai. Distinguising Proof of Diabetic Retinopathy Detection by Hybrid Approaches in Two Dimensional Retinal Fundus Images.[J]. Journal of medical systems,2019,43(6).
[10]Sun,Sun,Chen,Pan,Song. An artificial target detection method combining a polarimetric feature extractor with deep convolutional neural networks[J]. International Journal of Remote Sensing,2020,41(13).
[11] WalterT, Klein J, Massin P, et al. A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina[J]. Medical Imaging, IEEE Transactions on, 2002, 21(10):1236–1243.
[12] Imani Elaheh,Pourreza Hamid-Reza. A novel method for retinal exudate segmentation using signal separation algorithm.[J]. Computer methods and programs in biomedicine,2016,133.
[13] Sreng S,Maneerat N,Isarakorn D,et al. Automatic exudate extraction for early detection of Diabetic Retinopathy2013:31-35.
[14]Daniel Welfer,Jacob Scharcanski,Diane Ruschel Marinho. A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images[J]. Computerized Medical Imaging and Graphics,2009,34(3).
[15]Clara I. Sánchez,María García,Agustín Mayo,María I. López,Roberto Hornero. Retinal image analysis based on mixture models to detect hard exudates[J]. Medical Image Analysis,2009,13(4).
[16]Phillips R,Forrester J,Sharp P. Automated detection and quantification of retinal exudates.[J]. Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie,1993,231(2).
[17] M.García et al."Detection of hard exudates in retinal images using a radial basis function classifier." Ann. Biomed. Eng., vol. 37, no. 7, pp. 1448–1463, 2009.
[18]S.Ali et al." Statistical atlas based exudate segmentation." Comput.Med. Imag. Graph., vol. 37, nos. 5–6, pp. 358–368, 2013.[12]

[19]C. JayaKumari,R. Maruthi. Detection of Hard Exudates in Color Fundus Images of the Human Retina[J]. Procedia Engineering,2012,30.
[20]Osareh A, Mirmehdi M, Thomas B, et al.Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural networks2001:49-52.
[21]Zhang X, Chutatape O. Presented at the IEEE Conference on Computer Vision and Pattern Recognition. 2005.
[22] Xie S H, Nie H. Retinal vascular image segmentation using genetic algorithm plus FCM clustering[C]//Proceedings of 2013 Third International Conference on Intelligent System Design and Engineering Applications. Los Alamitos : IEEE Computer Society Press, 2013: 1225-1228.
[23]Sinthanayothin C,Boyce J F,Williamson T H,Cook H L,Mensah E,Lal S,Usher D. Automated detection of diabetic retinopathy on digital fundus images.[J]. Diabetic medicine : a journal of the British Diabetic Association,2002,19(2).
[24]Li H, Chutatape O. Automated feature extraction in color retinal images by amodel based approach[J]. Biomedical Engineering, IEEE Transactions on, 2004,51(2):246–254.
[25]J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, L. Kennedy,Optic nerve head segmentation, IEEE Trans. Med. Imag. 23, 2 (2005) 256-264.
[26]Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. (2016). Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society.
[27]Khojasteh Parham,Passos Júnior Leandro Aparecido,Carvalho Tiago,Rezende Edmar,Aliahmad Behzad,Papa João Paulo,Kumar Dinesh Kant. Exudate detection in fundus images using deeply-learnable features.[J]. Computers in biology and medicine,2019,104.
[28]Khojasteh Parham,Aliahmad Behzad,Kumar Dinesh K. Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms.[J]. BMC ophthalmology,2018,18(1).
[29] L. Giancardo, E. Chaum, T. P. Karnowski, F. Meriaudeau, K. W. Tobin, and Y. Li, "Bright retinal lesions detection using color fundus images containing reflective features," in World Congresson Medical Physics and Biomedical Engineering. Berlin, Germany: Springer, 2009, pp. 292–295.

[30]Vimala G S A G,Mohideen S K. Automatic detection of Optic Disk and Exudate from retinal images using Clustering algorithm2013:280-284.
[31]Bernhard M. Ege,Ole K. Hejlesen,Ole V. Larsen,Karina Møller,Barry Jennings,David Kerr,David A. Cavan. Screening for diabetic retinopathy using computer based image analysis and statistical classification[J]. Computer Methods and Programs in Biomedicine,2000,62(3).
[32]M. Niemeijer et al., "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.
[33] Ren X, Malik J.Learning a classification model for segmentation[C]. 9th IEEE International Conference on Computer Vision. Washington DC; IEEE Computer Society,2003.
[34] Achanta,Radhakrishna, et al. "SLIC superpixels compared to state-of-the-artsuperpixel methods." Pattern Analysis and Machine Intelligence, IEEETransactions on 34.11 (2012): 2274-2282.
[35]Bálint Borsos,László Nagy,David Iclănzan,László Szilágyi. Automatic detection of hard and soft exudates from retinal fundus images[J]. Acta Universitatis Sapientiae, Informatica,2019,11(1).
[36]Sheng Bin,Li Ping,Mo Shuangjia,Li Huating,Hou Xuhong,Wu Qiang,Qin Jing,Fang Ruogu,Feng David Dagan. Retinal Vessel Segmentation Using Minimum Spanning Superpixel Tree Detector.[J]. IEEE transactions on cybernetics,2018.
[37]Qifeng Yan,Yitian Zhao,Yalin Zheng,Yonghuai Liu,Kang Zhou,Alejandro Frangi,Jiang Liu. Automated retinal lesion detection via image saliency analysis[J]. Medical Physics,2019,46(10).
[38] Wei Zhou, Chengdong Wu, Yugen Yi, Wenyou Du. Automatic Detection of Exudates in Digital Color Fundus Images Using Superpixel Multi-Feature Classification, Digital Object Identifier 10.1109/IEEE ACCESS,2017,pp.17077-17088.
[39]Pavle Prentaš ić,Sven Lončarić. Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion[J]. Computer Methods and Programs in Biomedicine,2016,137.

[40]Maji D, Santara A, Mitra P, et al. Ensemble of Deep Convolutional Neural Networks for Learning to Detect Retinal Vessels in Fundus Images[DB/OL]. arXiv.org. Mar 15, 2016.

[41]Niemeijer M,van Ginneken B,Russell S R, et al. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis[J]. Investigative ophthalmology and visual science, 2007, 48(5): 2260-2267.

[42]Yu S, Xiao D, Kanagasingam Y. Exudate detection for diabetic retinopathy with convolutional neural networks[C]. Engineering in Medicine & Biology Society. Conf Proc IEEE Eng Med Biol Soc,2017:1744.

[43]X. Zhang et al., ‘‘Exudate detection in color retinal images for mass screening of diabetic retinopathy,’’ Med. Image Anal., vol. 18, no. 7, pp. 1026–1043, 2014

[44]X. Jiang, D. Mojon. “Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images.” IEEE Trans. Pattern Anal. Mach. Intell., 25 (1) (2003), pp. 131-137

[45]M. Niemeijer, J. Staal, B. van Ginneken, M. Long, M.D. Abramoff. “Comparative study of retinal vessel segmentation methods on a new publicly available database.” Proc. SPIE Med. Imag., 5370 (2004), pp. 648-656

[46]J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, B. van Ginneken. “ Ridge-based vessel segmentation in color images for the retina.” IEEE Trans. Med. Imag., 23 (4) (2004), pp. 501-509

[47]Avula Benzamin, Chandan Chakraborty. Detection of Hard Exudates in Retinal Fundus Images Using Deep Learning, Image and Video Processing (eess.IV),2018

[48] T. Kauppi et al., ‘‘DIARETDB1 diabetic retinopathy database and evaluation protocol,’’ in Proc. Brit. Mach. Vis. Conf., 2007, pp. 1–18.

[49] Decenciere, E. , et al. "TeleOphta: Machine learning and image processing methods for teleophthalmology." Innovation and research in biomedical en 34.2(2013):196-203.

[50] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, F. Meriaudeau, Indian Diabetic Retinopathy Image Dataset (IDRiD): A database for diabetic retinopathy screening research, Data 3, 3 (2018) 25.

[51] Derksen, D., Inglada, J., Michel, J., 2020. Geometry Aware Evaluation of Handcrafted Superpixel-Based Features and Convolutional Neural Networks for Land Cover Mapping Using Satellite Imagery. Remote Sensing.. doi:10.3390/rs12030513
[52] Farooq, A., Jia, X., Hu, J., Zhou, J., 2019. Multi-Resolution Weed Classification via Convolutional Neural Network and Superpixel Based Local Binary Pattern Using Remote Sensing Images. Remote Sensing.. doi:10.3390/rs11141692