Coarse-to-fine classification for diabetic retinopathy grading using convolutional neural network

Zhan Wu, Gonglei Shi, Yang Chen, Fei Shi, Xinjian Chen, Gouenou Coatrieux, Jian Yang, Limin M. Luo, Shuo Li

Coarse-to-fine Classification for Diabetic Retinopathy Grading Using Convolutional Neural Network

Zhan Wu[a], Gonglei Shi[b], Yang Chen[a,b,c,d] chenyang.list@seu.edu.cn, Fei Shi[e], Xinjian Chen[e] xjchen@suda.edu.cn, Gouenou Coatrieux[f], Jian Yang[g], Limin Luo[a,b,c,d], Shuo Li[h]

[a] School of Cyberspace Security, Southeast University, Nanjing, Jiangsu, China

[b] Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing, China

[c] Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing, China

[d] Centre de Recherche en Information Biomedicale Sino-Francais (LIA CRIBs), Rennes, France

[e] School of Electronics and Information Engineering, Soochow University, Suzhou, China

[f] Mines-Telecom, Telecom Bretagne, INSERM U1101 LaTIM, Brest, France

[g] School of Optoelectronics, Beijing Institute of Technology, Beijing, China

[h] Department of Medical Imaging, Western University, London, Canada

Abstract

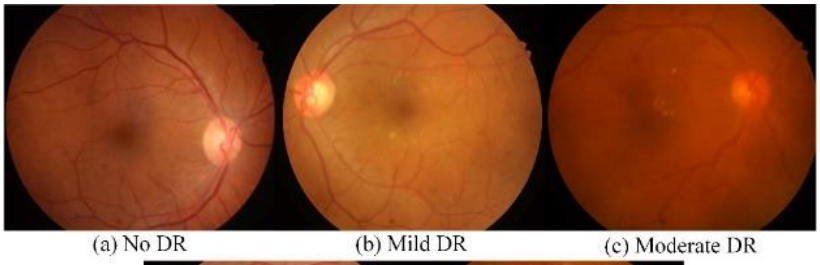
Diabetic retinopathy (DR) is the most common eye complication of diabetes and one of the leading causes of blindness and vision impairment. Automated and accurate DR grading is of great significance for the timely and effective treatment of fundus diseases. Current clinical methods remain subject to potential time-consumption and high-risk. In this paper, a hierarchically Coarse-to-fine network (CF-DRNet) is proposed as an automatic clinical tool to classify five stages of DR severity grades using convolutional neural networks (CNNs). The CF-DRNet conforms to the hierarchical characteristic of DR grading and effectively improves the classification performance of five-class DR grading, which consists of the following: (1) The Coarse Network performs two-class classification including No DR and DR, where the attention gate module highlights the salient lesion features and suppresses irrelevant background information. (2) The Fine Network is proposed to classify four stages of DR severity grades of the grade DR from the Coarse Network including mild, moderate, severe non-proliferative DR (NPDR) and proliferative DR (PDR). Experimental results show that proposed CF-DRNet outperforms some state-of-art methods in the publicly available IDRiD and Kaggle fundus image datasets. These results indicate our method enables an efficient and reliable DR grading diagnosis in clinic.

***Keywords*—Diabetic retinopathy grading; Coarse-to-fine classification; Convolutional neural networks; Fundus images.**

1. INTRODUCTION

Diabetic retinopathy (DR), a complication of diabetes, is one of the main causes of blindness in humans [1]. Individual with diabetes is more likely to develop into DR disease [2-3]. In recent years, clinical studies have indicated that the accurate classification for DR grading is important because it reveals DR severity levels to improve the selection of the appropriate therapeutic options (photocoagulation [4], vitrectomy [5], and injecting medicine into the eyes [6]) [7].

DR grading can be classified through the numbers, sizes, and types of lesions on the surface of the retinas from fundus images [8]. According to the clinical International Clinical Diabetic Retinopathy Disease Severity Scale [9], five stages of DR severity grades can be specified including no diabetic retinopathy (No DR), mild non-proliferative DR (NPDR), moderate NPDR, severe NPDR, and PDR. The samples of fundus images with increasing DR severity grades are shown in Fig. 1. In the subclinical phase, patients do not have apparent retinopathy [10]. In the clinic phase, the earliest DR stage is mild NPDR characterized by microaneurysms (MAs) that occurs due to leakage from tiny blood vessels of the retina. Then the DR severity grade is further evolved into moderate NPDR, where MAs begin to increase and additional lesions (hemorrhages (HMs), exudates (EXs)) appear. When developing to severe NPDR, amounts of MAs, HMs, and EXs diffuse on the surface of retina. From NPDR to PDR, an obvious signal is the growth of new blood vessels, which can cause a severe visual loss for patients [11]. These physiological changes in retinopathy leading to visual differences in fundus images can effectively divide five stages of DR severity grades [12].



(a) No DR (b) Mild DR (c) Moderate DR

(d) Severe DR (e) PDR

Fig. 1. The representative samples of diabetic retinopathy (DR) with increasing severity grades. Grade No DR represents no apparent DR lesions in (a). The DR lesions begin to expand and evolve from grade mild DR (b) to PDR (e) in the fundus images.

The current methods for DR grading have achieved significantly improved performance [2], [13]. However, accurate classification for DR grading remains challenges because: 1) the insufficiency of training samples limits the classification performance of automatic DR grading [14]. It is difficult to achieve excellent performance on the DR grading task compared with the classification tasks which have millions of data such as the ImageNet Challenge [15]; 2) the classification performance of DR grading suffers from inter-class similarities and intra-class variations [16]. The classification for DR grading is complicated because there are significant visual differences in sizes of lesions among the fundus images of the same class and visual similarities in shapes and colors between the fundus images of two different classes [17].

1.1. Related Works

Accurate and automatic classification for DR grading aims to achieve automated DR severity grades for improving the diagnosis efficiency and precision [18]. Traditional approaches for DR grading need to design manual features and classify the DR grades using common classifier or their variants such as support vector machines (SVMs), random forest (RF). Acharya *et al*. [19] employed a higher-order spectra method to extract the features from fund images of 300 subjects and used the SVMs for five-class classification of DR grading. Adarsh *et al*. [20] recognized the retinal blood vessels and pathologies (exudates and MAs) from fundus images as DR features and classified the DR severity grades by the SVMs. De la *et al*. [13] employed the local binary patterns (LBP) to extract local features and trained the random forest classifier for DR detection, which achieves the excellent performance using 71 fundus images. All of these show great potential with the development of methods for DR grading, however, they excessively depend on prior knowledge.

Recently, Convolutional Neural Networks (CNNs) have been proved effective methods for many medical imaging tasks, including feature recognition [21], image analysis [22], and lesion detection [23]. Chandrakumar T *et al*. [24] using CNN models deployed with dropout layer techniques obtained the excellent performance for classifying five stages of DR grades. Pratt *et al*. [2] proposed a CNN model with data augmentation and achieved a sensitivity of 95% and an accuracy of 75% on the publicly available Kaggle dataset. Zhou *et al*. [25] designed the Multi-Cell architecture to gradually increase the depth of CNNs and the resolution of input images, which effectively reduces computation complexity and relieves the influence of gradient vanishing problem.

The above studies have made significant contributions for DR grading. In this paper, the CF-DRNet is proposed to further improve the classification performance as to the five stages of DR severity grades. According to the hierarchical attributes of DR grading, a coarse-to-fine network is designed to hierarchically classify five stages of DR severity grades and reduce the influence of data imbalance problem. Moreover, rich physiological structures (blood vessels, macula lutea, and optic papillae) can be observed in fundus

images, which are redundant and seriously influence the DR grading classification. To suppress the interference of this irrelevant information, the Coarse Network with attention gate module is designed and this can effectively highlights the lesion areas and extracts discriminative lesion features for DR grading.

1.2. Contributions

In this paper, a hierarchically Coarse-to-fine DR network (CF-DRNet) based on CNNs is proposed to automatically classify five stages of DR severity grades from fundus images with two complementary sub-networks including the **Coarse Network** and the **Fine Network**. The pre-trained Coarse Network with attention modules performs the two-class classification including grade No DR and DR to highlight the discriminative lesion regions and efficiently exploit localized lesion information in fundus images. The grade DR can be further divided into four stages of severity grades including mild NPDR, moderate NPDR, severe NPDR, and PDR. These four-class classification is performed by the pre-trained Fine network. The main contributions can be generalized as follows:

1. For the first time, we propose a CF-DRNet to automatically and hierarchically classify five-stage DR grades. It enables a reliable and effective DR grading on the fundus images. It can help to develop accurate and automatic DR diagnosis and evaluation schemes for clinical physicians.
2. The Coarse Network and the Fine Network are designed using the pertained CNNs for two-class and four-class classification respectively. These two classification tasks hierarchically perform the five-stage DR grading, which effectively alleviates data imbalance problem and improves the five-stage DR classification performance.
3. Self-gated soft-attention mechanism modules are introduced in the pre-trained Coarse Network for two-class classification (No DR, DR) to effectively highlight the lesion features and suppress irrelevant information, which efficiently improves the two-class classification performance.

Experiment results show that, for five-fold cross-validation, proposed CF-DRNet for DR grading achieves classification with accuracy of 56.19%, sensitivity of 64.21%, and specificity of 87.39% in IDRiD database and accuracy of 83.10%, sensitivity of 53.99%, and specificity of 91.22% in Kaggle database. For the IDRiD database, the same training set and testing set are employed in the Grading Challenge of ISBI-2018 and the proposed CF-DRNet obtains the performance with the accuracy of 60.20%, which can achieve the second place [39]. These demonstrate that the CF-DRNet has the potential to become highly competitive method for five-stage DR grading.

The remainder of this paper is organized as follows: In Section 2, proposed framework for DR grading is discussed. The implementation details are reported in Section 3. In Section 4, the evaluation of proposed CF-DRNet is given to validate the performance. Finally, we draw conclusions about proposed CF-DRNet and discuss related future work in Section 5.

II. Methods

The proposed framework includes the following interdependent parts: (1) The **Preprocessing module** performs the operations of normalization, image enhancement, and data augmentation for fundus images. (2) The **CF-DRNet module** is utilized for DR classification, which consists of two sub-networks including Coarse Network and Fine Network. The pre-trained Coarse Network with attention modules performs the two-class classification. The pre-trained Fine Network with larger-size images is designed for these four stages of DR severity grades. (3) The **Aggregation module** performs label fusion from the Coarse Network and the Fine Network for final five-class DR grading classification. The main workflow of proposed CF-DRNet is depicted in Fig. 2.

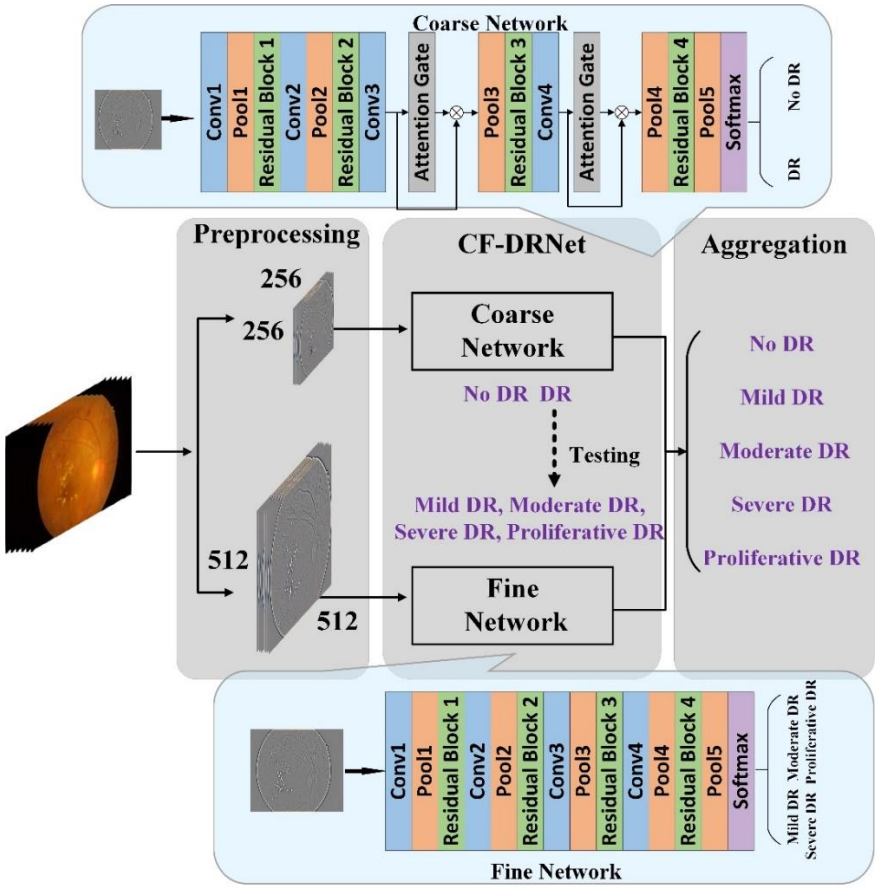


Fig. 2. The CF-DRNet performs automatic five stages of DR grading, which has three interdependent parts: Preprocessing module, CF-DRNet module, and Aggregation module. The CF-DRNet module includes the Coarse Network and the Fine Network. Coarse Network performs the two-class classification to judge whether there is a DR disease or not. The Fine Network performs four-class classification to classify four stages of DR severity grades (mild NPDR, moderate NPDR, severe NPDR, and PDR). Final outputted five-class classification results will be obtained via label fusion operation.

2.1. Preprocessing Module

The preprocessing module follows three steps including image enhancement, image normalization, and data augmentation. The image enhancement performs the noise

reduction from varying illumination. The normalization is utilized to reduce computation complexity. Data augmentation is employed to tackle the overfitting problem.

2.1.1. Image enhancement

To remove the irrelevant information from varying illumination, a contrast-enhanced image $I'(x,y;\sigma)$ is obtained as follows [26]:

$$I'(x,y;\sigma)=\alpha I(x,y)+\beta G(x,y;\varepsilon)*I(x,y)+\gamma, \quad (1)$$

where $*$ represents the convolution operator and $G(x, y;\varepsilon)$ is a Gaussian filter. α is employed to adjust to local average color from one fundus image, β is utilized to highlight pixel values of lesion areas, γ is the pixel bias, and ε is the scale of the Gaussian filter. The values of the parameters are chosen as: $\alpha=4$, $\beta=-4$, $\varepsilon=512/20$ and $\gamma=128$ according to [27]. The representative original image and the enhanced fundus image can be seen in Fig. 3.

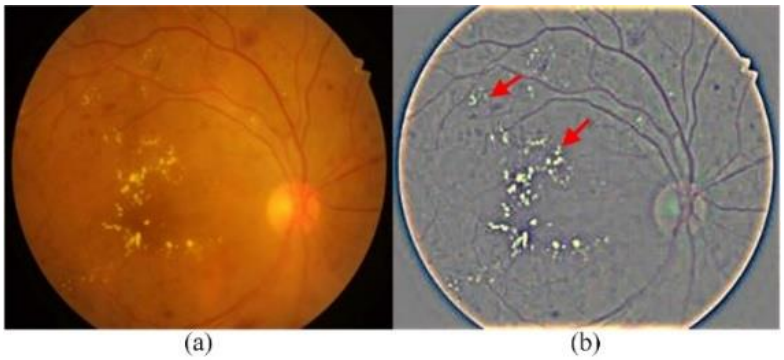


Fig. 3. Comparison between the Original image (a) and the image after enhancement (b). The red arrows indicate the underlying lesions can be singularized after preprocessing.

2.1.2. Image normalization

To lower computational complexity, the fundus images are normalized into 256×256 pixels for the Coarse Network and 512×512 pixels for the Fine Network via bilinear interpolation [28], respectively.

As to the fundus images with three channels, each pixel value of each channels is normalized into the range of (0, 1). The normalization formula is defined by Eqs (2):

$$y=\frac{x\text{-}MinValue}{MaxValue\text{-}MinValue} \quad (2)$$

where x is the input pixel value of one fundus image, *MinValue* is the minimum pixel value of this fundus image, *MaxValue* is the maximal pixel value of this fundus image, y is the pixel value output after normalization.

2.1.3. Data Augmentation

Data augmentation is applied upon training datasets to tackle the over-fitting and data imbalance problems in the case of limited training dataset [29]. In our experiment, the transformations including translation, stretching, rotation and flipping are employed to the

labelled dataset. A summary of the transformations with the parameters is given in TABLE 1.

TABLE 1
Data Augmentation Parameters

Transformation Type	Description
Rotation	Randomly rotate an angle of 0°-360°
Flipping	0 (without flipping) or 1(with flipping)
Rescaling	Randomly with scale factor between 1/1.6 and 1.6
Translation	Randomly with shift between −10 and 10 pixels

2.2. CF-DRNet Module

CF-DRNet consists of two sub-networks including the Coarse Network and the Fine Network. The Coarse Network performs two-class classification to determine the presence of DR lesions. The Fine Network performs four-class classification including mild NPDR, moderate NPDR, severe NPDR, and PDR.

2.2.1. Coarse Network

The overall dataset can be divided into two classes including grade No DR and grade DR. The Coarse Network performs this two-class classification to determine the presence of DR lesions.

The Coarse Network is designed based on the ResNet-18 proposed by kaiming He et.al [30], which contains four convolution layers, five pooling layers, four residual block, and two attention gate modules. The ResNet network can well tackle the degradation problems via shortcut connection, and Fig. 4 depicts the structure of the Residual block [31,32]. For clarity, $H_L(x)$ denotes the transformation function of the L^{th} building block, and x is the input of the L^{th} building block. The desired output for the Residual block is set as $F_L(x)$. The residual block explicitly forces the output to fit the residual mapping, i.e., the stacked nonlinear layers are used to learn the following transformation:

$$F_L(x)=H_L(x)-x. \tag{3}$$

Therefore, the transformation for the L^{th} building block is:

$$H_L(x)=F_L(x)+x. \tag{4}$$

The residual block consists of convolution layers with the kernel size of 1×1 and 3×3. The convolution layers with the kernel size of 1×1 are used to reduce channel numbers into n ($n<m$) and the convolution layers with the kernel size of 3×3 are employed for extracting spatial features and returning to the input channel number m. Limited to the small size of training data, all the residual blocks pre-trained on the ImageNet dataset [15] are fine-tuned for two-class classification.

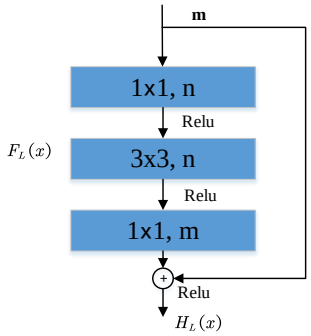


Fig. 4. A building block of residual network.

The pooling layer is used for down-sampling to reduce computation complexity [33]. From the detail design of the Coarse Network in TABLE 2, we can see that the data size is down-sampled from 256×256 pixels to 4×4 pixels throughout the pooling layers.

The attention gate module proposed by Oktay O *et al*. [34] is applied in the Coarse Network. The ResNet with the attention gate module can learn to enhance the lesion features and suppress irrelevant information for fundus images. The activation maps $A^{l_1}=\{a_i\}\in R^{C_1\times H_1\times W_1}$ of a chosen layer $l_1\in\{1,\ldots,L_1\}$ extract the local features, where C_1, H_1, W_1 are the numbers of channel, height, width, and a_i is the pixel-wise feature vectors of the feature map A^{l_1}. The global feature maps $B^{l_2}=\{b_i\}\in R^{C_2\times H_2\times W_2}$ of a chosen layer $l_2\in\{1,\ldots,L_2\}$ related to the lesion region of interests are extracted, where C_2, H_2, W_2 and b_i are the numbers of channel, height, width and the pixel-wise feature vectors of feature maps B^{l_2}. The attention gate combines the global features A^{l_1} and the local features B^{l_2} via the pixel-wise additive operation to compute the compatibility score $T=\{t_i^l\}_{i=1}^n\in R^{1\times H_1\times W_1}$. The t_i^l can be obtained by Eqs (5):

$$t_i^l=\mu\sigma_1(e)+b_2 \\ =\mu\sigma_1(W_aA^{l_1}+W_bB^{l_2}+b_1)+b_2 \quad , \tag{5}$$

where σ_1 is the Rectified Linear Unit (ReLU) nonlinear activation function [35] can be described as follows:

$$\sigma_1(x)=max(0,e)\,, \tag{6}$$

where e represents the input of the ReLU nonlinear activation function. $\mu\in\mathbb{R}^{1\times C_{int}\times H_1\times W_1}$, $W_a\in\mathbb{R}^{C_{int}\times C_1\times H_1\times W_1}$, and $W_b\in\mathbb{R}^{C_{int}\times C_2\times H_2\times W_2}$ are learnable weight parameters to match the dimension of between A^{l_1} and B^{l_2}, b_1 and b_2 are the learnable bias parameters.

Two attention gate modules are employed in the Coarse Network. The outputs of the attention gate modules are fused with its input via product fusion operation and the fusion results serve as the input of the next layer. For clarity, one product fusion function $y=f(x^a,x^b)$, two feature maps x^a and x^b, and a fusion feature map y are defined, where $x^a\in R^{H\times W\times D}$, $x^b\in R^{H\times W\times D}$, $y\in R^{H\times W\times D}$ (W, H, and D are the width, height, and channel number of feature maps). The function $y=f(x^a,x^b)$ stacks the two features at the same

location (i, j) across the feature channel d :

$$y_{i,j,d}=x^{a}_{i,j,d}\times x^{b}_{i,j,d}\,. \tag{7}$$

The Softmax layer is used to normalize feature maps into the range of (0, 1) so that the output vector y_m represents the probability of the m^{th} class [36]. The operation for the Softmax layer can be written as:

$$y_m=\frac{e^{x}}{\sum_{m=1}^{2}e^{x}}, \tag{8}$$

where y_m is the output probability of the m^{th} class, x represents the input neurons of the upper layer.

The cross-entropy loss function is selected as the objective function of the Coarse Network to accelerate training. The cross-entropy loss function of Coarse Network $loss_C$ is given by Eqs (9):

$$loss_C=-\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{2}I(l_i=k)log\ p(k|x_i)\,, \tag{9}$$

where m is the number of samples in per min-batch, l_i stands for the class label (0-1) of the image x_i, $I(\bullet)$ is an indicator function which equals one if l_i is equal to k. The detailed configuration of the Coarse Network is listed in TABLE 2.

TABLE 2
Configurations of the CF-DRNet.

CF-DRNet					
		Coarse Network		Fine Network	
Layer		**Kernel Size, Channel Number**	**Output Size**	**Kernel Size, Channel Number**	**Output Size**
Data		-	256×256	-	512×512
Conv 1		3×3, 128	128×128	3×3, 128	256×256
Pool 1		2×2, 128	64×64	2×2, 128	128×128
Residual Block-1	Conv 1-1	1×1, 64	64×64	128×128	128×128
	Conv 1-2	3×3, 64	64×64	128×128	128×128
	Conv 1-3	1×1, 256	64×64	128×128	128×128
Conv2		3×3,256	64×64	3×3,256	128×128
Pool 2		2×2, 256	32×32	2×2, 256	64×64
Residual Block-2	Conv 2-1	1×1,128	32×32	1×1, 128	64×64
	Conv 2-2	3×3, 128	32×32	3×3, 128	64×64
	Conv 2-3	1×1,512	32×32	1×1,512	64×64
Conv3		3×3,512	32×32	3×3,512	64×64
Attention Gate 1		-,512	32×32	-	-
Pool 3		2×2, 512	16×16	2×2, 512	32×32
Residual Block-3	Conv 3-1	1×1, 256	16×16	1×1, 256	32×32
	Conv 3-2	3×3,256	16×16	3×3,256	32×32
	Conv 3-3	1×1,1024	16×16	1×1,1024	32×32
Conv 4		3×3, 1024	16×16	3×3, 1024	32×32
Attention Gate 2		-,1024	16×16	-	-
Pool 4		2×2, 1024	8×8	2×2, 1024	16×16
Residual	Conv 4-1	1×1, 512	8×8	1×1, 512	16×16

Block-4	Conv 4-2	3×3, 512	8×8	3×3, 512	16×16
	Conv 4-3	1×1, 2048	8×8	1×1, 2048	16×16
Pool 5		2×2, 2048	4×4	2×2, 2048	8×8
Softmax		2 Neurons		4 Neurons	

2.2.2. Fine Network

The Fine Network is designed to further divide the grade DR outputted from the Coarse Network into four severity grades including grade mild NPDR, moderate NPDR, severe NPDR, and PDR. In order to dig into more elaborate inter-class differences for these four grades, the input size of the Fine Network is resized into 512×512 pixels for four-class classification.

The pre-trained Fine Network based on the ResNet-18 is fine-tuned for four-class classification. The data size is down-sampled from 512×512 pixels to 8×8 pixels throughout the pooling layers.

The cross entropy loss of Fine Network is employed as the objective function of the Fine Network to accelerate training. The cross-entropy loss function is given by Eqs (10):

$$loss_F = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{4} I(l_i=k)\log p(k|x_i), \quad (10)$$

where m is the number of samples in per min-batch, l_i stands for the class label (0-3) of the image x_i, $I(\bullet)$ is an indicator function which equals one if l_i is equal to k. The configurations of the Fine Network are depicted in TABLE 2.

2.3. Aggregation Module

Overall testing set is divided into two classes including grade No DR and grade DR. The results of grade No DR is obtained via the two-class classification performance from the Coarse Network. Grade DR is further divided into four sub-classes including grade mild NPDR, moderate NPDR, severe NPDR, and PDR and these four grade results are obtained by the four-class classification performance from Fine Network. Final five-class classification results are summarized to realize five stages of DR grading.

III. Implementation Details

In this section, the implementation of proposed CF-DRNet and the training/testing process is described. The computer platform is configured as follows: CPU was Inter(R) Core(TM) i7-5930K 3.5GHz; GPU was NVIDIA 2080TI with 11G memory. All codes were written under Python 3.6, and we used Tensorflow r1.4 as the deep learning library. The CUDA edition used here was 10.0.

During the training phase, the weight parameters are learned using mini-batch stochastic gradient descent with momentum (set to 0.9). The base learning rate is set to 10^{-3} and iteratively decreases until the loss stops decreasing. All the models were trained using 10000 iterations.

For the Coarse Network, five grades of fundus images and corresponding labels (No DR, mild NPDR, moderate NPDR, and severe NPDR) in the dataset are transformed into

two grades of fundus images and corresponding labels (No DR, DR). The batch size is set to 8 during the training phase and all the preprocessed test images of 256×256 pixels are input into the Coarse Network in the testing phase. Finally, the fundus images from the testing set are classified as two classes (No DR, DR).

For the Fine Network, the batch size is set to 4 in the training phase. In the testing phase, the images and labels of the grade DR from the results of the Coarse Network are further divided into four grades (Mild NPDR, Moderate NPDR, and Severe NPDR). All the testing images of 512×512 pixels are input into the Fine Network. Finally, the fundus images from the testing set are classified as four stages of DR severity grades.

IV. Experiment and Results

4.1. Database Description

The Kaggle fundus database contains 88400 fundus images taken under a variety of imaging conditions [12]. These fundus images were provided via EyePACS [37], which is a free platform for retinopathy screening. Every subject provides the left and right fields of human eyes. The IDRiD fundus database provided by the retinal specialists at an Eye Clinic located in India [38] contains 516 fundus images from thousands of examinations. There are five grades (No DR, Mild NPDR, Moderate NPDR, Severe NPDR, PDR) in these two databases. The detailed data distribution of collected fundus images can be seen in TABLE 3 and TABLE 4.

Five-fold cross-validation is used. We use 80 percent of the fundus images for training and 20 percent of the images for testing. It is noted that no data is overlapping between the training dataset and testing dataset.

TABLE 3
The Data Distribution of Fundus Images from the Kaggle Database

	Images	No DR	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Raw images	88400	65130	6185	13105	2075	1905
Train	70720	52104	4948	10484	1660	1524
Training Augmentation	101915	52104	12370	13629	11620	12192
Test	17680	13026	1237	2621	415	381

TABLE 4
The Data Distribution of Fundus Images from the IDRiD Databases

	Images	No DR	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Raw images	516	168	25	168	93	62
Train	415	135	20	135	75	50

Training Augmentation	3237	1620	400	405	412	400
Test	101	33	5	33	18	12

4.2. Results and Evaluation

With the Kaggle and IDRiD fundus database, the classification performance is measured. According to the property measurement, we conduct the following experiments: (1) For analyzing the effectiveness of the Coarse Network for DR grading, we compare the performance of between the basic ResNet and the Coarse Network (the ResNet attached with the attention gate). (2) The proposed Coarse-to-fine network and the sub-networks (the Coarse Network and the Fine Network) are analyzed for five-class classification of DR grading. (3) Finally, proposed CF-DRNet is compared with the other methods.

To further clarify the evaluation metrics, the classification performance is measured using sensitivity (SENS), specificity (SPEC), accuracy (ACC), which can be defined by Eqs (11) - (13):

$$SEN = \frac{TP}{TP + FN}, \quad (11)$$

$$SPEC = \frac{TN}{FP + TN}, \quad (12)$$

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)}, \quad (13)$$

Where:

- TP, True Positives: the number of positive samples is correctly classified.
- FP, False Positives: the number of negative samples is wrongly classified as positive.
- TN, True Negatives: the number of negative samples is correctly classified.
- FN, False Negatives: the number of positive samples is wrongly classified as negative.

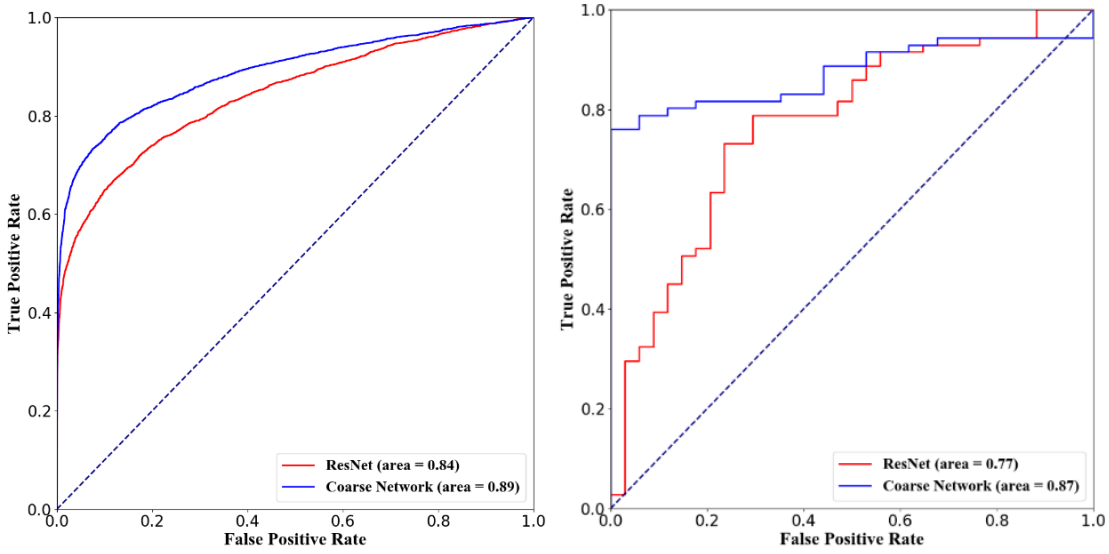
4.2.1. The effectiveness of the Coarse Network

TABLE 5
The Performance of the Coarse Network and the Normal ResNet 50 for Two-class Classification in the IDRiD and Kaggle Database.

Methods	Kaggle		
	ACC	SEN	SPEC
ResNet	84.86%	91.90%	72.56%
Coarse Network	88.61%	64.43%	96.92%
	IDRiD		
	ACC	SEN	SPEC
ResNet	73.33%	83.09%	68.02%
Coarse Network	80.00%	85.91%	76.78%

For two-class classification for DR grading, the attention gate module is applied in the Coarse Network based on the ResNet. The performance of the proposed Coarse Network and the basic ResNet is shown in TABLE 5. It can be seen that the Coarse Network in virtue of attention gate module has better performance than the basic ResNet for DR grading compared with ResNet in terms of accuracy (3.75% improvement) and specificity (24.36% improvement) in Kaggle Database, accuracy (6.67% improvement), sensitivity (2.82% improvement) and specificity (8.76% improvement) in IDRiD Database.

The statistical analysis is done under the help of Receiver Operating Characteristics (ROC) curves and Area Under Curve (AUC) for the proposed Coarse Network and the Basic ResNet in Fig. 5. As illustrated in this figure, the Coarse Network achieves superior performance (AUC: 0.87 in the IDRiD database, 0.89 in the Kaggle database) over the ResNet (AUC: 0.77 in the IDRiD database, 0.84 in the Kaggle database). The ROC curves of the Kaggle database are smoother than the ROC curves of the IDRiD database because of the larger data size of the testing set for the Kaggle database.



(a) (b)

Fig. 5. The ROC Curves and AUC Values of the Coarse Network (Blue), ResNet (Red) in the Kaggle Database (a) and in the IDRiD Database (b).

4.2.2. The effectiveness of the CF-DRNet and its subnetworks

The classification performance of proposed CF-DRNet in terms of SENS, SPEC and ACC for DR grading and its subnetworks (Coarse Network and Fine Network) in the IDRiD and Kaggle database are given in TABLE 6. As can be seen in this table, the Coarse Network and the Fine Network achieve excellent performance to fulfill the two-class classification and four-class classification in the Kaggle and IDRiD database. The overall CF-DRNet obtains the classification performance with the sensitivity of 53.99%, specificity of 91.22%, and accuracy of 83.10% in the Kaggle and the performance with the sensitivity of 64.21%, specificity of 87.39%, and accuracy of 56.19% in the IDRiD database, respectively.

From TABLE 6, it can be seen that the overall performance of the Kaggle database is higher than the case for the IDRiD database. The reason might be the difference in image quantity and clinician's experience. For the IDRiD database, the same training set and testing set are employed in the Grading Challenge of ISBI-2018 and our proposed CF-DRNet obtains the performance with the accuracy of 60.20%, sensitivity of 69.61%, and specificity of 88.78%, which can achieve the second place [39].

TABLE 6
The Comparative Performances of Proposed CF-DRNet and Its Subnetworks (Coarse Network and Fine Network) in the Kaggle and IDRiD Database.

Methods	Kaggle		
	ACC	SEN	SPEC
Coarse Network	88.61%	64.43%	96.92%
Fine Network	67.70%	53.64%	84.01%
CF-DRNet	83.10%	53.99%	91.22%
	IDRiD		
	ACC	SEN	SPEC
Coarse Network	80.58%	85.91%	77.30%
Fine Network	58.33%	58.33%	69.73%
CF-DRNet	56.19%	64.21%	87.39%

4.2.3. Comparison of the classification performance with state-of-the-art methods

For the five-class classification of DR grading, the proposed CF-DRNet is compared with other well-known classification methods: LBP+SVM [6], and VGG CNN [19]. LBP+SVM is a method which employs binarization algorithm of local pixels and uses the multiple binary SVMs with linear kernel for classification. The VGG CNN is a modified VGG network which consists of multiple convolutional and pooling layers, and fully connected layers.

All the methods are compared in the IDRiD and Kaggle database in TABLE 7. The parameters of the comparative methods were set according to their original works. It is observed in this table that deep learning based methods (e.g., VGG 16, CF-DRNet) achieve better performance than traditional method (e.g., LBP+SVMs) using manual features. Especially, it can be found that the CF-DRNet outperforms the other comparative methods.

TABLE 7
Comparative Performances of Proposed CF-DRNet and Other State-of-the-art Methods in the Kaggle and IDRiD Database.

Methods	Kaggle		
	ACC	SEN	SPEC
LBP+SVMs [6]	59.09%	59.09%	80.86%
VGG CNN [19]	74.12%	30.00%	90.12%
CF-DRNet	83.10%	53.99%	91.22%

	IDRiD		
	ACC	SEN	SPEC
LBP+SVMs [6]	51.43%	57.66%	86.25%
VGG CNN [19]	52.28%	62.26%	86.27%
CF-DRNet	56.19%	64.21%	87.39%

V. Conclusion and Discussion

In this paper, for the first time, we proposed a hierarchically coarse-to-fine DR network (CF-DRNet) for five-stage DR grading employing convolutional neural networks (CNNs). CF-DRNet consists of two subnetworks: one Coarse Network and one Fine Network. The Coarse Network with attention gate module is designed for two-class classification (No DR, DR) to enhance the lesion features and suppress irrelevant information for two-class classification of DR grading. The Fine Network with larger size input based on the pre-trained ResNet-18 performs four-class classification (mild NPDR, moderate NPDR, severe NPDR, and PDR) for DR grading.

Five-fold cross-validation is used in method validation. Experimental results show that proposed CF-DRNet outperforms the comparative methods [6, 19] in the IDRiD database and Kaggle database. The obtained results demonstrated a promising performance for DR grading in fundus images.

To further obtain improved classification performance, a more elaborate Fine Network needs to be designed, which can reduce the confusion among four stages of DR severity grades. In addition, DR grading will be explored with the help of DR lesion detection to realize automatic DR diagnosis in the future.

Acknowledgment

This research was supported in part by the State's Key Project of Research and Development Plan under Grant 2017YFA0104302, Grant 2017YFC0109202 and 2017YFC0107900, in part by the National Natural Science Foundation under Grant 61801003, 61871117and 81471752, in part by the China Scholarship Council under NO. 201906090145.

REFERENCE

[1] L. Seoud, J. Chelbi, and F. Cheriet, "Automatic grading of diabetic retinopathy on a public database," 2015.

[2] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional Neural Networks for Diabetic Retinopathy," in *Procedia Computer Science*, 2016, doi: 10.1016/j.procs.2016.07.014.

[3] M. de Bruijne, "Machine learning approaches in medical image analysis: From detection to diagnosis," *Medical Image Analysis*. 2016, doi: 10.1016/j.media.2016.06.032.

[4] X. D. Liu, Q. Lu, H. Di Xu, Q. Wang, and J. Zhao, “Mild form disseminated photocoagulation treatment for moderate non-proliferative diabetic retinopathy,” *Int. Eye Sci.*, vol. 18, no. 7, pp. 1313–1316, Jul. 2018, doi: 10.3980/j.issn.1672-5123.2018.7.36.

[5] J. M. Smith and D. H. W. Steel, “Anti-vascular endothelial growth factor for prevention of postoperative vitreous cavity haemorrhage after vitrectomy for proliferative diabetic retinopathy,” *Cochrane Database of Systematic Reviews*. 2015, doi: 10.1002/14651858.CD008214.pub3.

[6] S. Sivaprasad and S. Oyetunde, “Impact of injection therapy on retinal patients with diabetic macular edema or retinal vein occlusion,” *Clin. Ophthalmol.*, 2016, doi: 10.2147/OPTH.S100168.

[7] B. Antal and A. Hajdu, “An ensemble-based system for microaneurysm detection and diabetic retinopathy grading,” *IEEE Trans. Biomed. Eng.*, 2012, doi: 10.1109/TBME.2012.2193126.

[8] M. Usman Akram, S. Khalid, A. Tariq, S. A. Khan, and F. Azam, “Detection and classification of retinal lesions for grading of diabetic retinopathy,” *Comput. Biol. Med.*, vol. 45, no. 1, pp. 161–171, Feb. 2014, doi: 10.1016/j.compbiomed.2013.11.014.

[9] S. Haneda and H. Yamashita, “[International clinical diabetic retinopathy disease severity scale].,” *Nippon rinsho. Japanese J. Clin. Med.*, 2010.

[10] T. Kawaguchi, S. Horie, N. Bouchenaki, K. Ohno-Matsui, M. Mochizuki, and C. P. Herbort, “Suboptimal therapy controls clinically apparent disease but not subclinical progression of Vogt-Koyanagi-Harada disease,” *Int. Ophthalmol.*, vol. 30, no. 1, pp. 41–50, Feb. 2010, doi: 10.1007/s10792-008-9288-1.

[11] N. Salamat, M. M. S. Missen, and A. Rashid, “Diabetic retinopathy techniques in retinal images: A review,” *Artificial Intelligence in Medicine*, vol. 97. Elsevier B.V., pp. 168–188, 01-Jun-2019, doi: 10.1016/j.artmed.2018.10.009.

[12] V. Gulshan *et al.*, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA - J. Am. Med. Assoc.*, 2016, doi: 10.1001/jama.2016.17216.

[13] J. De La Calleja, L. Tecuapetla, M. Auxilio Medina, E. Bárcenas, and A. B. Urbina Nájera, “LBP and machine learning for diabetic retinopathy detection,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, doi: 10.1007/978-3-319-10840-7_14.

[14] R. A. Welikala *et al.*, “Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy,” *Comput. Med. Imaging Graph.*, vol. 43, pp. 64–77, Jul. 2015, doi: 10.1016/j.compmedimag.2015.03.003.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[16] R. F. Mansour, “Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy,” *Biomed. Eng. Lett.*, 2018, doi: 10.1007/s13534-017-0047-y.

[17] O. Faust, R. Acharya U., E. Y. K. Ng, K. H. Ng, and J. S. Suri, “Algorithms for the automated detection of diabetic retinopathy using digital fundus images: A review,” *J. Med. Syst.*, vol. 36, no. 1, pp. 145–157, Feb. 2012, doi: 10.1007/s10916-010-9454-7.

[18] G. M. Lin *et al.*, “Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy,” *J. Ophthalmol.*, 2018, doi: 10.1155/2018/2159702.

[19] R. Acharya U, C. K. Chua, E. Y. K. Ng, W. Yu, and C. Chee, “Application of higher order spectra for the identification of diabetes retinopathy stages,” *J. Med. Syst.*, 2008, doi: 10.1007/s10916-008-9154-8.

[20] P. Adarsh and D. Jeyakumari, “Multiclass SVM-based automated diagnosis of diabetic retinopathy,” in *2013 International Conference on Communication and Signal Processing*, 2013, pp. 206–210.

[21] S. Ji, W. Xu, M. Yang, and K. Yu, “3D Convolutional neural networks for human action recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, doi: 10.1109/TPAMI.2012.59.

[22] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 Fourth International Conference on 3D Vision (3DV)*, 2016, pp. 565–571.

[23] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in neural information processing systems*, 2015, pp. 91–99.

[24] T. Sajana, K. Sai Krishna, G. Dinakar, and H. Rajdeep, “Classifying diabetic retinopathy using deep learning architecture,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 6 Special Issue 4, pp. 1273–1277, 2019, doi: 10.35940/ijitee.F1261.0486S419.

[25] K. Zhou *et al.*, “Multi-Cell Multi-Task Convolutional Neural Networks for Diabetic Retinopathy Grading,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018, doi: 10.1109/EMBC.2018.8512828.

[26] M. J. J. P. Van Grinsven, B. Van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, “Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images,” *IEEE Trans. Med. Imaging*, 2016, doi: 10.1109/TMI.2016.2526689.

[27] S. H. Rasta, M. E. Partovi, H. Seyedarabi, and A. Javadzadeh, “A comparative study on preprocessing techniques in diabetic retinopathy retinal images: Illumination correction and contrast enhancement,” *J. Med. Signals Sens.*, 2015, doi: 10.4103/2228-7477.150414.

[28] P. Porwal *et al.*, “Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research,” *Data*, 2018, doi: 10.3390/data3030025.

[29] J. Salamon and J. P. Bello, “Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification,” *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 279–283, Mar. 2017, doi: 10.1109/LSP.2017.2657381.

[30] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[31] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[32] Z. Wu, C. Shen, and A. van den Hengel, “Wider or Deeper: Revisiting the ResNet Model for Visual Recognition,” *Pattern Recognit.*, 2019, doi: 10.1016/j.patcog.2019.01.006.

[33] T. Y. Lin, A. Roychowdhury, and S. Maji, “Bilinear CNN models for fine-grained visual recognition,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, vol. 2015 Inter, pp. 1449–1457, doi: 10.1109/ICCV.2015.170.

[34] O. Oktay *et al.*, “Attention U-Net: Learning Where to Look for the Pancreas,” Apr. 2018.

[35] J. Schmidt-Hieber, “Nonparametric regression using deep neural networks with ReLU activation function,” Aug. 2017.

[36] X. Wang, L. Gao, J. Song, and H. Shen, “Beyond Frame-level CNN: Saliency-Aware 3-D CNN with LSTM for Video Action Recognition,” *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 510–514, Apr. 2017, doi: 10.1109/LSP.2016.2611485.

[37] Y. Sun, Y. Chen, X. Wang, and X. Tang, “Deep learning face representation by joint identification-verification,” in *Advances in neural information processing systems*, 2014, pp. 1988–1996.

[38] J. Cuadros and G. Bresnick, “EyePACS: An adaptable telemedicine system for diabetic retinopathy screening,” *J. Diabetes Sci. Technol.*, 2009, doi: 10.1177/193229680900300315.

[39] “IDRiD - Leaderboard.” [Online]. Available: https://idrid.grand-challenge.org/Leaderboard/. [Accessed: 21-Nov-2019].