Scientific Research Publishing

Journal of Computer and Communications, 2023, 11, 16-28
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Diabetic Retinopathy Classification Based on Bilinear Cross Attention Network

Zhiyuan Ren, Chen Xing

School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
Email: chuju1994@qq.com

How to cite this paper: Ren, Z.Y. and Xing, C. (2023) Diabetic Retinopathy Classification Based on Bilinear Cross Attention Network. *Journal of Computer and Communications*, **11**, 16-28.
https://doi.org/10.4236/jcc.2023.115002

Received: April 24, 2023
Accepted: May 21, 2023
Published: May 24, 2023

Abstract

Computer-aided diagnostic systems can assist doctors in diagnosing and treating DR cases more effectively, thereby improving work efficiency, reducing the burden on doctors during examinations, and alleviating problems related to uneven distribution of medical resources and shortage of doctors. In this article, we propose a classification method for diabetic retinopathy based on a bilinear multi-attention network. This method uses two backbone networks to extract features, and cross-shares the features using two attention modules to further deepen feature extraction. The non-local attention module is added to address the limitations of traditional convolutional neural networks in capturing global information. By paying attention to highly correlated pathological areas globally, performance improvement can be achieved. We achieved an accuracy of 91.7% on the Messidor dataset.

Keywords

Diabetic Retinopathy, Deep Neural Network, Styling, Deep Neural Network

1. Introduction

The prevalence of diabetes is continuously increasing worldwide. According to the 2021 Global Diabetes Map published by the International Diabetes Federation (IDF), approximately 537 million adults aged 20 to 79 years old have diabetes globally, and it is expected that the number of people with diabetes will increase to 783 million by the year 2045 [1]. As described by the International Diabetes Federation (IDF) and other studies, approximately 50% of diabetes patients are unaware of their condition [2]. Diabetic retinopathy (DR) refers to a series of ocular complications caused by damage to the retina in the eyes due to long-term diabetes. More than 30% of diabetic patients suffer from diabetic retinopathy [3]. The main manifestation of diabetic retinopathy is damage to the

blood vessels in the retina, leading to swelling, leakage, and abnormal growth of new blood vessels. If DR is not treated with targeted therapy in its early stages, it may lead to complete loss of vision or blindness. However, the early symptoms of DR are often not obvious, which makes it easy for patients to miss the optimal diagnosis timing. Therefore, early detection and participation in treatment are key to delaying or even preventing DR-induced blindness. However, traditional medical diagnostic mechanisms require high-level professional equipment, medical resources, and specialized ophthalmological services, resulting in imbalances in the distribution of medical resources worldwide and a contradiction between limited numbers of ophthalmologists and a growing number of patients [4]. Therefore, using computers to perform rapid, simple, and accurate classification of diabetic retinopathy has extremely important research value.

The research community has begun to introduce artificial intelligence to assist in the diagnosis of diabetic retinopathy. In recent years, deep learning technology has rapidly developed and has been widely applied in fields such as image processing and natural language processing, providing more reliable and efficient solutions for computer-aided diagnosis systems. By using deep learning algorithms, these systems can automatically analyze large amounts of medical imaging data, quickly and accurately detecting signs of diabetic retinopathy, assisting doctors in diagnosis and treatment, saving time and energy for doctors, shortening patient waiting times, and providing more convenient medical services for patients in some remote areas while reducing waste of medical resources.

2. Related Work

The automatic classification of diabetic retinopathy primarily relies on the characterization of retinal lesions in fundus images. These features include vascular changes, microaneurysms, hemorrhages, exudates, neovascularization, and vitreous opacities. These features can be extracted through image processing and feature extraction techniques, and then input into a classifier for classification.

In the field of automatic classification of diabetic retinopathy, deep learning has become a hot topic and frontier, providing new ideas and methods for research and application in medical imaging. For example, Neelu K had proposed a deep learning model based on PCA-Firefly for early detection of diabetic retinopathy. The model had used a preprocessing stage to convert the original images to grayscale and had reduced the dimensionality of input images using the PCA algorithm. Then, the Firefly algorithm had been used for parameter adjustment to optimize the learning process of the deep neural network, classifying 64,000 fundus images from the UC Irvine Machine Learning Repository and Messidor into proliferative (PDR) and non-proliferative (NPDR) categories with an accuracy of 96% [5]. Li Y H *et al*. had proposed a system that used a DCNN to detect DR and classify it as NPDR and PDR, using algorithms such as maximum pooling and support vector machine (SVM), trained and run on 34,000

images from Kaggle datasets including pre-processing techniques such as resizing and color normalization, achieving a classification accuracy of 91% [6]. Transfer learning had also been widely applied in the classification of diabetic retinopathy. Since the dataset for diabetic retinopathy is often small, transfer learning techniques can transfer knowledge from pre-trained deep learning models and apply them to new tasks, thus improving classification performance. For example, Nikhil M N had used a CNN to classify DR categories on 500 images from Kaggle datasets. The HE filtering algorithm had been used for image preprocessing, and the stochastic gradient descent with momentum optimization algorithm had been used for training. By combining the VGG16, AlexNet, and Inception v3 architectures, an accuracy of 81.1% had been achieved [7]. Somasundaram K *et al*. had applied transfer learning and deep convolutional neural networks to classify diabetic retinopathy. They had used a publicly available dataset containing 1200 images, which had been divided into 5 different categories, and had fine-tuned a pre-trained CNN model for classification. The authors had compared three different pre-trained models (VGG-16, VGG-19, and ResNet-50), and had performed data augmentation to improve model performance. The classifier of the ResNet-50 model had achieved the best accuracy of 96.41% on the test set [8].

In the field of diabetic retinopathy classification, attention mechanisms can help models better identify diabetes-related lesions and thus achieve more accurate classification. Specifically, through attention mechanisms, the model can focus attention on areas related to diabetic retinopathy during the learning process, thereby improving the accuracy of classification. For example, Mohammad T had proposed a method that used multi-scale convolutional neural networks and attention mechanisms to detect diabetic retinopathy. In this method, lesion images had been inputted into multi-scale CNNs, allowing image features to be extracted at different scales. After feature extraction, an attention mechanism had been applied to weight the feature map to enhance attention to key regions. Finally, a fully connected layer had been used to classify the features for automatic detection of diabetic retinopathy, exhibiting high performance and robustness [9]. Wang Z *et al*. had proposed a new deep learning model—Zoom-in-Net—to address the detection problem of diabetic retinopathy. The model was based on convolutional neural networks (CNN), which had first extracted different scale feature information from low to high by performing "pyramid"-like multi-scale convolution operations on the input image. Then, the model had adopted an attention mechanism to locate lesion regions and further extract and classify these regions' feature information. Finally, Zoom-in-Net had merged all results to obtain the final diagnosis [10].

Although the aforementioned models achieved good results in classifying diabetic retinopathy, there is still room for further optimization, such as insufficient basic feature extraction of the primary network, the inability of regular convolution and attention mechanisms to effectively capture the highly corre-

lated regions of pathological areas distributed globally in medical images, and the problem of single-feature extraction. To address these issues, we propose a Bilateral Cross Multi-Attention Fusion Network (BCM-AFNet) model that combines multiple attention modules and an improved loss function. Through multiple ablation experiments and comparative experiments, the improved model's performance on diabetic retinopathy classification tasks was verified. The proposed BCM-AFNet achieved advanced scores on the Messidor dataset and exhibited good stability.

3. Methods

3.1. Backbones

We used two transfer learning pre-trained backbones, Xception [11] and ResNet50 [12], to extract fundamental features of diabetic retinopathy images. Xception was employed as the first primary network backbone, while ResNet50 was used as the second primary network backbone. These backbones are popular deep learning models in image processing that have been pre-trained on large datasets such as ImageNet and can effectively capture essential characteristics in images. The extracted features can be utilized for various subsequent tasks such as classification, detection, segmentation and so on.

3.2. Channel Attention Module

The channel attention mechanism assigns more attention and resources to channels that are considered to make a greater contribution to the final result, based on observed facts within a given feature map. EAC-Net [13] is a cross-channel attention mechanism that emphasizes high-dimensional information interaction and can reduce a significant number of parameters without decreasing accuracy compared to SE-Net [14]. SE-Net uses global pooling and a set of fully connected layers to calculate the weight of each channel, thus achieving cross-channel attention mechanism. This method requires global pooling operations for all feature maps, which leads to a decrease in feature map resolution and a loss of information. In contrast, ECA's cross-channel attention mechanism achieves local cross-channel interaction by considering the interaction between each channel and its neighbors without reducing the feature map resolution. This cross-channel interaction mechanism enables ECA-Net to better utilize channel correlations and improve feature representation without reducing the input feature map dimension. Additionally, the attention module of SE-Net requires a large number of parameters to learn the relationship between channels, resulting in a larger network size. ECA-Net reduces the number of parameters without reducing the feature map resolution through cross-channel interaction mechanisms and parameter compression. The structure of the ECA attention module is shown in **Figure 1**.

First, the ECA module performs a global average pooling (GAP) operation on the input feature map channel-wise to obtain a $1 \times 1 \times C$ feature vector, where C

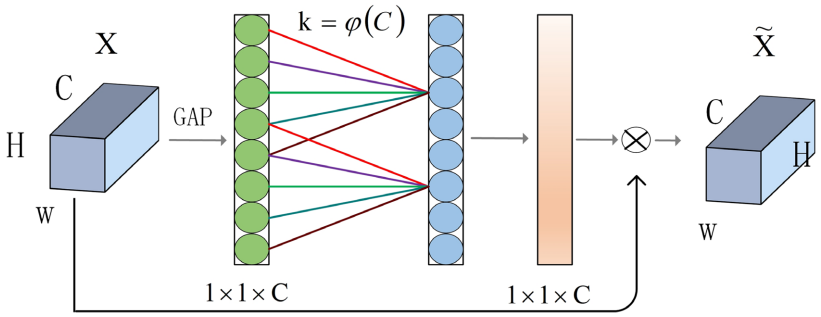

Figure 1. ECA structure diagram.

is the number of channels in the feature map. Then, the resulting feature vector from the pooling operation is subjected to a one-dimensional convolutional layer for cross-channel information interaction, yielding a second $1 \times 1 \times C$ feature vector. The size k of the one-dimensional convolutional kernel in the ECA module is determined by an adaptive function that dynamically adjusts the kernel size based on the statistical information of the input feature map, allowing for adaptive modulation of the interaction rate. The adaptive kernel size is computed using Equation (1).

$$k = \varphi(C) = \frac{1}{2}\left|1 + \log_2(C)\right| \tag{1}$$

Next, the second feature vector is normalized using the sigmoid activation function and multiplied element-wise with the original features. This approach enhances important features, suppresses irrelevant ones, and ultimately improves the quality and accuracy of feature representation.

3.3. Non-Local Attention Module

In convolutional neural networks, local connections are often used between adjacent layers to capture local features of an image. By gradually increasing the number of feature extraction layers, the network's receptive field expands and more global information is obtained. However, traditional convolutional neural networks have limitations in capturing global information due to loss of information in the sampling and layer-by-layer transmission processes. In the field of medical image analysis, pathological regions are typically highly correlated and often distributed globally. Traditional convolutional operations may not fully take these factors into account, thereby failing to effectively capture the global pathological information and correlations between different regions. To address these issues, this study introduces non-local attention [15], which considers all feature points/regions for weighted calculation, allowing the network to globally attend to the image and effectively capture global pathological information and correlations between different regions. This operation can be represented mathematically by Equation (2).

$$y_i = \frac{1}{C(x)} \sum_j f\left(x_i, x_j\right) g\left(x_j\right) \tag{2}$$

In the formula, x represents the input feature map. x_i denotes a vector with the same dimensionality as x in terms of channels. It is a length-C vector, where C is the number of channels in x, corresponding to the features at position i in the feature map. y_i is the output vector, also a length-C vector. i and j represent spatial positions in the feature map. In the formula, i and j sum over all possible positions in the feature map. $f\left(x_m, x_n\right)$ is a function used to compute the similarity relationship between two positions x_m and x_n in the feature map. The output of the function is a normalized value representing the similarity weight between these two positions. We use an embedded Gaussian function to calculate the similarity. The embedded Gaussian function takes feature vectors x_i and x_j as inputs and generates a scalar value that serves as the similarity weight. The definition of the embedded Gaussian function is given by Equation (3).

$$f\left(x_i, x_j\right) = \mathrm{e}^{\theta(x_i)^{\mathrm{T}} \cdot \phi(x_i)^{\mathrm{T}}} \tag{3}$$

Here, θ and ϕ are two independent neural networks that map the input feature vectors x_i and x_j to new feature vectors, which are then used to compute similarity weights. $g(x_j)$ computes the feature at position j. It is a neural network that maps the input feature vector x_j to a new feature vector. $C(x)$ is a normalization term that sums over the function f for all positions j in the feature map, as given by Equation (4).

$$C(x) = \sum_j f\left(x_i, x_j\right) \tag{4}$$

The non-local operation can be used to capture long-range dependencies in the input feature map by computing a weighted sum of features at all positions in the feature map, where the weights are determined by the similarity between position pairs. The structure of the non-local attention module is shown in **Figure 2**.

3.4. Bilinear Cross Attention Fusion Network

We propose a BCM-AFNet model for diabetic retinopathy classification by cross-fusing two transfer learning backbone networks with attention mechanisms. We use ECA channel attention modules and non-local attention modules.

Firstly, the output of backbone network 1 is randomly dropped out by 20%

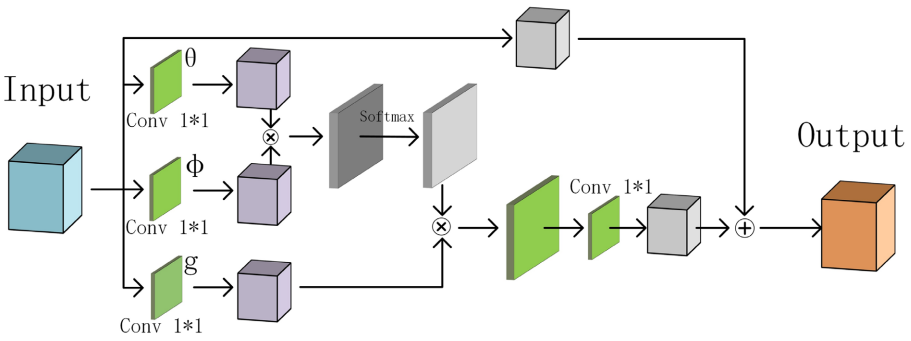

Figure 2. Structure diagram of the non-local attention module.

using a dropout layer and used as the input C_{Ein} to the ECA module. The channel-wise global average pooling is applied to obtain C'_{Ein} which is then subjected to one-dimensional convolution with dynamically computed convolutional kernels to enable cross-channel interactivity for obtaining the second feature vector C''_{Ein}. Next, the Sigmoid activation function is used to normalize C''_{Ein} which is element-wise multiplied by the original feature and outputted as the channel feature information C_{out}, as shown in Equation (5). Here, $\otimes$ denotes element-wise multiplication and σ represents the Sigmoid activation function.

$$C_{\text{out}} = (C_{\text{Ein}}) \otimes \sigma(C''_{\text{Ein}}) \tag{5}$$

Simultaneously, the output of backbone network 2 is introduced as the input to the non-local attention module after being randomly dropped out by 20% using a dropout layer. A 1 × 1 convolution layer is used to embed the input feature map to obtain the embedded feature map M_{nin}. Then, we use two additional 1 × 1 convolution layers θ and φ to transform the embedded feature map to obtain $\theta(M_{\text{nin}})$ and $\phi(M_{\text{nin}})$, respectively. The convolution kernels of θ and φ both have a size of 1 × 1, so the width and height remain unchanged while the number of channels is halved after transformation. Next, we perform matrix multiplication on $\theta(M_{\text{nin}})$ and $\phi(M_{\text{nin}})$ to obtain a weighted correlation matrix. By applying the Softmax operation to this matrix, we can obtain an attention map M'_{nin} for each element in the matrix, which represents the similarity weight between each position in the input feature map and other positions, as shown in Equation (6). We then use another 1 × 1 convolution layer g to transform this attention map to obtain the final output of the non-local attention module, denoted as M''_{nin}, as shown in Equation (7). The final output feature map is transformed using another 1 × 1 convolution layer and added to the input feature map to obtain the ultimate output M_{out}, as shown in Equation (8). Here, $\otimes$ denotes element-wise multiplication, and $\oplus$ denotes element-wise addition. In this process, the role of the attention map is to adjust the relative weights between different positions in the input feature map, thereby enhancing the network's ability to extract important information from the input data.

$$M'_{\text{nin}} = \text{soft}\max\left(\theta(M_{\text{nin}}) \otimes \phi(M_{\text{nin}})\right) \tag{6}$$

$$M''_{\text{nin}} = g(M_{\text{nin}}) \otimes M'_{\text{nin}} \tag{7}$$

$$M_{\text{out}} = \text{conv}(M''_{\text{nin}}) \oplus M_{\text{nin}} \tag{8}$$

Subsequently, we take the features C_{out} obtained from Equation (5) and process them through the Non-local attention module to extract Non-local features MC_{out} with channel characteristics, as shown in Equation (9). This method is capable of capturing long-range dependency relationships between features while preserving channel information during feature extraction. Additionally, we apply the feature M_{out} obtained from fEquation (8) to the channel attention module ECA module to extract channel feature vectors CM_{out} with Non-local features, as shown in Equation (10).

$$MC_{\text{out}} = \text{conv}\left(g\left(C_{\text{out}}\right) \otimes \text{softmax}\left(\theta\left(C_{\text{out}}\right) \otimes \phi\left(C_{\text{out}}\right)\right)\right) \oplus C_{\text{out}} \tag{9}$$

$$M_{\text{out}} = \text{conv}\left(M''_{\text{nin}}\right) \oplus M_{\text{nin}} \tag{10}$$

Next, we will combine the base features BS_1 obtained from main network 1 with the base features BS_2 obtained from main network 2 through a 50% random dropout using Dropout. This results in BS_1' and BS_2'. We then concatenate them with the channel attention-derived C_{out}, the Non-local attention-derived M_{out}, the feature CM_{out} obtained through Non-local channel cross-attention, and the feature MC_{out} obtained through channel Non-local cross-attention. The fusion of these features is achieved through concatenation, resulting in the fused feature RC_{out}, as shown in Equation (11).

$$RC_{\text{out}} = \text{concat}\left(\left(BS_1'\right),\left(BS_2'\right),\left(C_{\text{out}}\right),\left(M_{\text{out}}\right),\left(CM_{\text{out}}\right),\left(MC_{\text{out}}\right)\right) \tag{11}$$

After obtaining the fused feature, we pass it through a global average pooling layer and then into a fully connected layer. We use the Softmax function to perform classification. Thus, our BCM-AFNet ultimately outputs as shown in Equation (12).

$$BCMAF_{\text{out}} = \text{softmax}\left(GAP\left(RC_{\text{out}}\right)\right) \tag{12}$$

Our proposed BCM-AFNet model structure, as shown in **Figure 3**, is based on a bilinear backbone network and utilizes two interlaced attention modules, namely ECA channel attention and Non-local attention, to further extract features. It achieves the classification task by fusing all the extracted features. This comprehensive attention mechanism not only adequately captures the features but also obtains feature information at different scales and spatial levels, thereby improving the model's performance.

4. Experiments

4.1. Dataset and Preprocessing

Messidor DataSet [16] is a publicly available dataset for diabetic retinopathy (DR) classification. This dataset was jointly released by researchers from three ophthalmic departments in France and contains 1200 digital retinal images. These images are from the eyes of patients with varying degrees of DR, classified into four levels: normal, mild DR, moderate DR, and severe DR. Each image has an associated annotation file that includes patient ID, age, gender, image quality, and DR level. It also includes diabetic macular edema (DME) labels, which are divided into three levels to assess the risk of macular edema.

For retinal fundus images, due to their complex structure and the relatively fine positions of hemorrhages and lesions, we adopt a method of cropping the original images to reduce noise and interference from useless information on the retinal image features. This can eliminate redundant black background areas, make the image clearer, and reduce the impact of noise and useless information. Additionally, since the original image has a high resolution that is not suitable

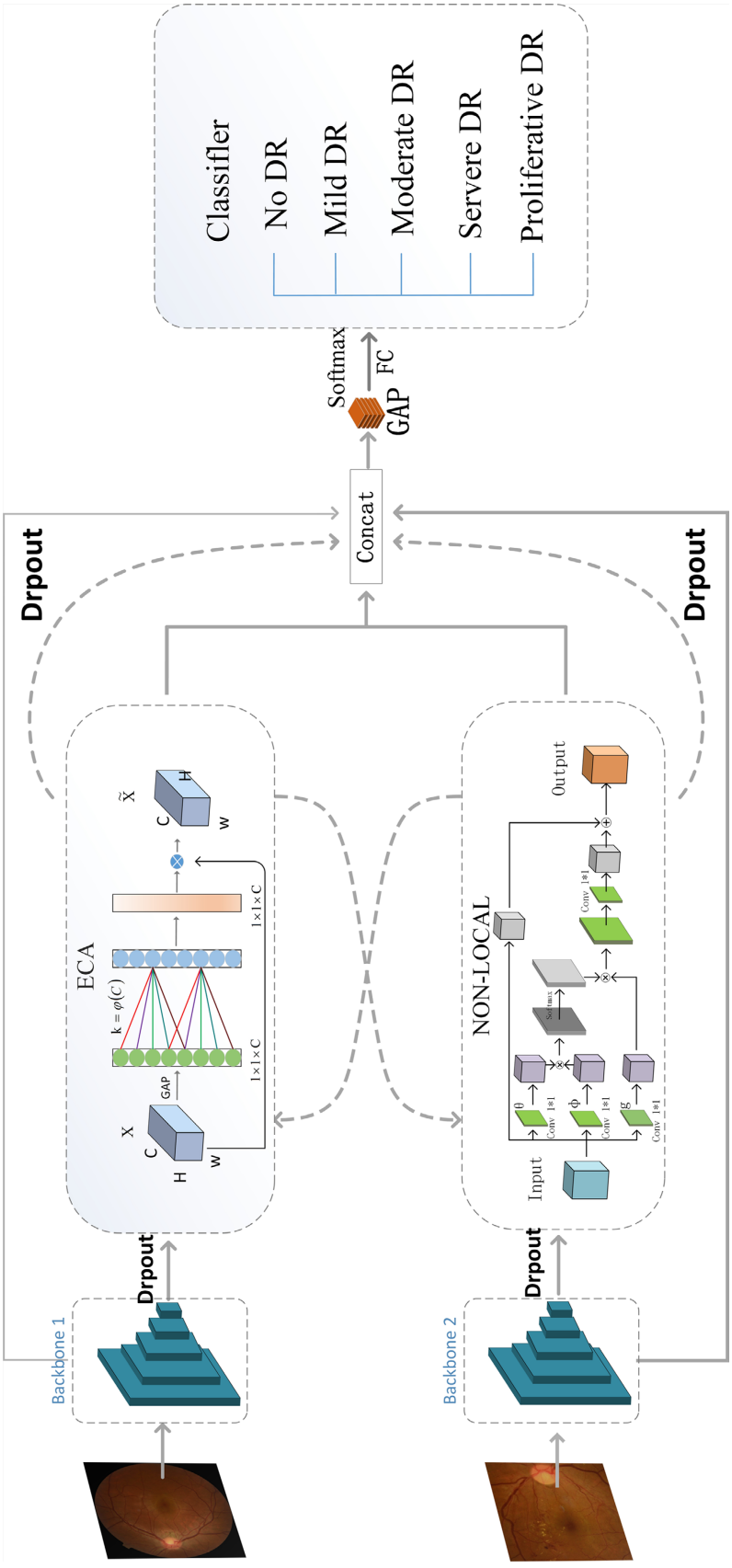

Figure 3. The model structure diagram of BCM-AFNet.

for model training, we use the "resize()" method in the OpenCV library to adjust the resolution to 512 × 512.

4.2. Experimental Analysis

We compared our method with other advanced methods on the Messidor dataset. Since this dataset does not classify diabetic retinopathy according to the standards established by the International Council of Ophthalmology, we conducted binary classification of DR based on whether or not referrals were required, as done in other papers. Specifically, classes 0 and 1 were classified as no referral needed, while classes 3 and 4 were classified as requiring referral. The metrics used in this study included accuracy, precision, recall, F1-score, sensitivity, specificity, and AUC. Table 1 shows the binary DR classification results (%) of different methods on the Messidor dataset. It can be seen that our method achieved the best performance in binary classification on the Messidor dataset. "-" indicates that their results were not reported in their paper.

The CKML model combines multiple convolutional kernels of different sizes in parallel to comprehensively capture global and local information in retinal images for diabetic retinopathy (DR) grading. Additionally, this model adopts multiple loss functions to optimize the network's performance for different evaluation metrics. Similarly, the VNXK model uses a deep convolutional neural network architecture based on the VGGNet model and adds extra convolutional kernels to capture more detailed features from retinal images for DR grading. Both models perform well in DR grading, surpassing traditional machine learning algorithms and other deep learning models. In this study, we propose a model that utilizes various attention modules and transfer learning techniques with a dual backbone network. Compared to these methods, our proposed model demonstrates significant improvements in AUC, accuracy, specificity, and sensitivity. The confusion matrix of our model, shown in Figure 4 validates the performance and stability of our proposed model.

5. Discussion

We propose a diabetic retinopathy (DR) classification method based on a bilinear

Table 1. Table of experimental results on the Messidor dataset.

Method	Table Column Head						
	AUC	ACC	Pre	Recall	F1	Spe	Sen
Lesion Based [17]	76.0	-	-	-	-	-	-
Fisher Vector [17]	86.3	-	-	-	-	-	-
VNXK [18]	88.7	89.3	-	-	-	89.2	89.3
CKML [18]	89.1	89.7	-	-	-	90.0	89.7
Dynamic Feature [19]	91.6	-	-	-	-	50.0	96.2
OUR Method	93.6	91.7	89.7	91.4	90.5	92.0	91.4

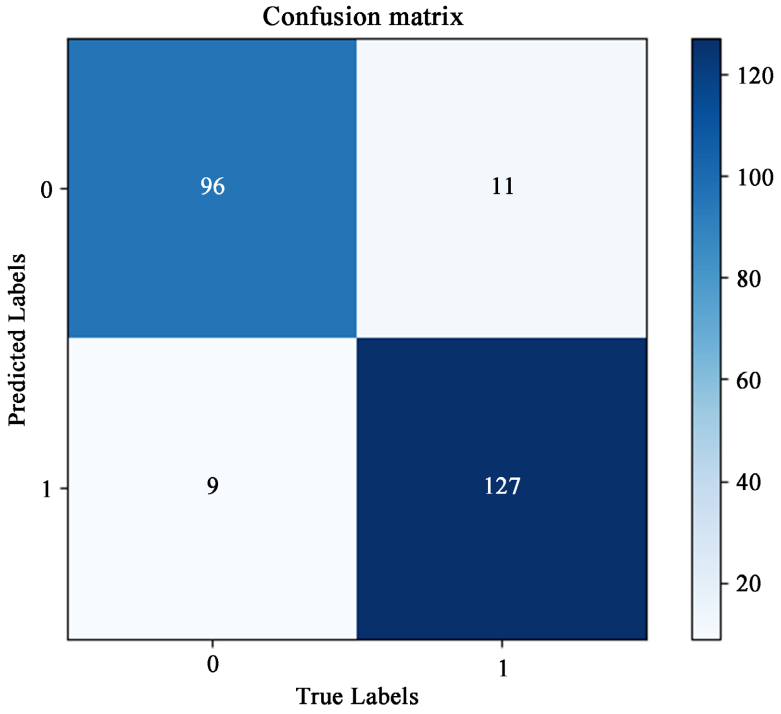

Figure 4. Confusion Matrix for the Messidor dataset.

multi-attention network that addresses the issue of insufficient feature extraction using a single backbone network, as well as the problem of ineffective capture of highly correlated pathological areas with a single convolutional and attention mechanism in medical images. First, we use two backbone networks to extract features and apply cross-attention strategies with two attention modules to deepen the feature extraction. In particular, we added a Non-local attention module to address the limitations of traditional convolutional neural networks in capturing global information, by focusing on highly correlated pathological regions for global attention to achieve improved performance. The addition of numerous Dropout layers reduces the risk of overfitting while reducing the number of parameters. Finally, through comparative experiments with other models, we achieved advanced results on the Messidor dataset.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Ogurtsova, K., Guariguata, L., Barengo, N.C., Ruiz, P.L.D., Sacre, J.W., Karuranga, S., *et al.* (2022) IDF Diabetes Atlas: Global Estimates of Undiagnosed Diabetes in Adults for 2021. *Diabetes Research and Clinical Practice*, **183**, 109-118. https://doi.org/10.1016/j.diabres.2021.109118

[2] Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., *et al.* (2019) Global and Regional Diabetes Prevalence Estimates for 2019 and Projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes At-

las, 9th Edition. *Diabetes Research and Clinical Practice,* **157**, Article ID: 107843. https://doi.org/10.1016/j.diabres.2019.107843

[3] Hill, L. and Makaroff, L.E. (2016) Early Detection and Timely Treatment Can Prevent or Delay Diabetic Retinopathy. *Diabetes Research and Clinical Practice*, **120**, 241-243. https://doi.org/10.1016/j.diabres.2016.09.004

[4] Resnikoff, S., Felch, W., Gauthier, T.M. and Spivey, B. (2012) The Number of Ophthalmologists in Practice and Training Worldwide: A Growing Gap Despite more than 200 000 Practitioners. *British Journal of Ophthalmology*, **96**, 783-787. https://doi.org/10.1136/bjophthalmol-2011-301378

[5] Gadekallu, T.R., Neelu, K., Bhattacharya, S., *et al*. (2020) Early Detection of Diabetic Retinopathy Using PCA-Firefly Based Deep Learning Model. *Electronics*, **9**, Article 274.

[6] Li, Y.H., Yeh, N.N., Chen, S.J. and Chung, Y.C. (2019) Computer Assisted Diagnosis for DR Based on Fundus Images Using Deep Convolutional Neural Network. *Mobile Information Systems*, **2019**, Article ID: 6142839. https://doi.org/10.1155/2019/6142839

[7] Nikhil, M.N. and Rose, A. (1906) Diabetic Retinopathy Stage Classification Using Convolutional Neural Network. *International Research Journal of Engineering and Technology*, 6, 5969-5974.

[8] Somasundaram, K., Sivakumar, P. and Suresh, D. (2021) Classification of Diabetic Retinopathy Disease with Transfer Learning Using Deep Convolutional Neural Networks. *Advances in Electrical and Computer Engineering*, **21**, 49-56. https://doi.org/10.4316/AECE.2021.03006

[9] Al Antary, M.T. and Arafa, Y. (2021) Multi-Scale Attention Network for Diabetic Retinopathy Classification. *IEEE Access*, **9**, 54190-54200. https://doi.org/10.1109/ACCESS.2021.3070685

[10] Wang, Z., Yin, Y.X., Shi, J.P., Fang, W., Li, H.S., Wang, X.G. (2017) Zoom-in-Net: Deep Mining Lesions for Diabetic Retinopathy Detection. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Quebec, 11-13 September 2017, 267-275. https://doi.org/10.1007/978-3-319-66179-7_31

[11] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 1800-1807. https://doi.org/10.1109/CVPR.2017.195

[12] He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[13] Wang, Q.L., Wu, B.G., Zhu, P.F., Li, P.H., Zuo, W.M. and Hu, Q.H. (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 11531-11539. https://doi.org/10.1109/CVPR42600.2020.01155

[14] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/CVPR.2018.00745

[15] Wang, X.L., Girshick, R., Gupta, A. and He, K. (2018) Non-Local Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7794-7803.

https://doi.org/10.1109/CVPR.2018.00813

[16] Decencière, E., Zhang, X., Cazuguel, G., *et al*. (2014) Feedback on a Publicly Distributed Image Database: The Messidor Database. *Image Analysis & Stereology*, **33**, 231-234. https://doi.org/10.5566/ias.1155

[17] Pires, R., Avila, S., Jelinek, H.F., Wainer, J., Valle, E. and Rocha, A. (2015) Beyond Lesion-Based Diabetic Retinopathy: A Direct Approach for Referral. *IEEE Journal of Biomedical and Health Informatics*, **21**, 193-200. https://doi.org/10.1109/JBHI.2015.2498104

[18] Vo, H.H. and Verma, A. (2016) New Deep Neural Nets for Fine-Grained Diabetic Retinopathy Recognition on Hybrid Color Space. 2016 *IEEE International Symposium on Multimedia* (*ISM*), San Jose, 11-13 December 2016, 209-215. https://doi.org/10.1109/ISM.2016.0049

[19] Seoud, L., Hurtut, T., Chelbi, J., Cheriet, F. and Langlois, J.P. (2015) Red Lesion Detection Using Dynamic Shape Features for Diabetic Retinopathy Screening. *IEEE Transactions on Medical Imaging*, **35**, 1116-1126. https://doi.org/10.1109/TMI.2015.2509785