

Article

A Multi-Task Learning and Multi-Branch Network for DR and DME Joint Grading

Xiaoxue Xing ^{1,*}, Shenbo Mao ¹ , Minghan Yan ¹, He Yu ¹ , Dongfang Yuan ¹, Cancan Zhu ¹, Cong Zhang ¹, Jian Zhou ¹ and Tingfa Xu ²

¹ College of Electronic Information Engineering, Changchun University, Changchun 130012, China; 230402214@mails.ccu.edu.cn (S.M.); 200401085@mails.ccu.edu.cn (M.Y.); yuh82@ccu.edu.cn (H.Y.); 210401131@mails.ccu.edu.cn (D.Y.); 210401102@mails.ccu.edu.cn (C.Z.); 210401120@mails.ccu.edu.cn (C.Z.); 200401084@mails.ccu.edu.cn (J.Z.)

² Image Engineering & Video Technology Lab, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China; xutingfa@bit.edu.cn

* Correspondence: xingxx@ccu.edu.cn

Abstract: Diabetic Retinopathy (DR) is one of the most common microvascular complications of diabetes. Diabetic Macular Edema (DME) is a concomitant symptom of DR. As the grade of lesion of DR and DME increase, the possibility of blindness can also increase significantly. In order to take the early interventions as soon as possible to reduce the likelihood of blindness, it is necessary to perform both DR and DME grading. We design a joint grading model based on multi-task learning and multi-branch networks (MaMNet) for DR and DME grading. The model mainly includes a multi-branch network (MbN), a feature fusion module, and a disease classification module. The MbN is formed by four branch structures, which can extract the low-level feature information of DME and DR in a targeted way; the feature fusion module is composed of a self-feature extraction module (SFEN), cross-feature extraction module (CFEN) and atrous spatial pyramid pooling module (ASPP). By combining various features collected from the aforementioned modules, the feature fusion module can provide more thorough discriminative features, which benefits the joint grading accuracy. The ISBI-2018-IDRiD challenge dataset is used to evaluate the performance of the proposed model. The experimental results show that based on the multi-task strategy the two grading tasks of DR and DME can provide each other with additional useful information. The joint accuracy of the model, the accuracy of DR and the accuracy of DME are 61.2%, 64.1% and 79.4% respectively.

Keywords: DR; DME; joint grading; multi-branch network; multi-task learning



Citation: Xing, X.; Mao, S.; Yan, M.; Yu, H.; Yuan, D.; Zhu, C.; Zhang, C.; Zhou, J.; Xu, T. A Multi-Task Learning and Multi-Branch Network for DR and DME Joint Grading. *Appl. Sci.* **2024**, *14*, 138. <https://doi.org/10.3390/app14010138>

Academic Editors: Jose-Maria Buades-Rubio and Antoni Jaume-i-Capó

Received: 8 November 2023

Revised: 15 December 2023

Accepted: 21 December 2023

Published: 22 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Diabetes poses a substantial health risk for a significant portion of the global population. At least one-third of people with diabetes have diabetes-related eye disease. The most common cause of blindness in diabetic patients is DR. As the number of diabetics increases globally, it is likely that DR will continue to play a significant role in vision loss and the resulting functional impairment for decades to come. As the level of DR lesion increases, patients will successively develop symptoms such as blurred vision, visual field defects, obscured and distorted vision, and dark shadows until blindness. Therefore, it is important to perform DR grading to take the appropriate therapeutic options, such as photocoagulation, vitrectomy, injecting medicine into the eyes, and so on. The number, size, and kind of lesions visible on the surface of the retinas in fundus pictures can be used to categorize DR grades. Figure 1 shows the possible lesions in the fundus image. The International Clinical Diabetic Retinopathy Severity Scale (ICDRS) is a uniform standard for DR grading, according to which DR can be classified as 0–5, i.e., no lesions, mild lesions, moderate lesions, severe lesions, and proliferative [1–3].

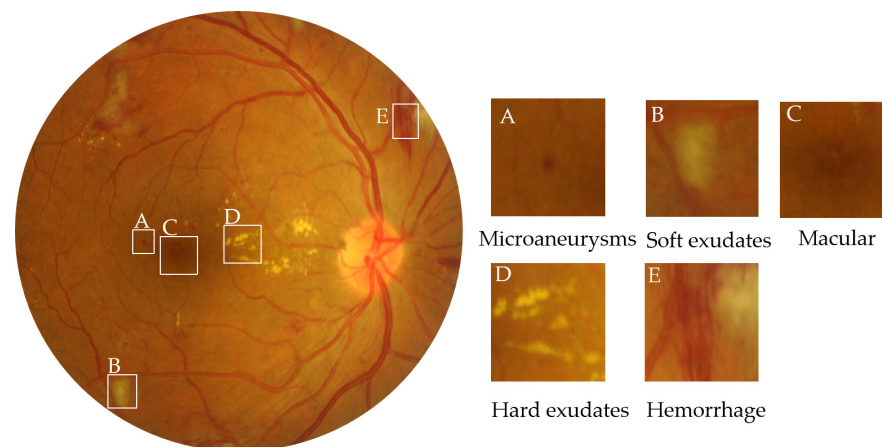


Figure 1. Fundus images from the IDRiD dataset with hard exudates, soft exudates, hemorrhage, microaneurysms, and other lesions. The severity of DR is related to microaneurysms, hemorrhage, soft exudates and hard exudates. The severity of DME is determined by the distance from hard exudates to the macular area.

DME is a concomitant symptom of DR and is the most common cause of visual impairment or loss. It refers to retinal thickening or hard exudative deposits caused by the accumulation of extracellular fluid within the central macular sulcus, which can be classified into 3 levels by the distance of exudate from the macular center [4]: 0 (normal, no obvious hard exudate), 1 (mild, hard exudate outside the circle with a radius of one optic disc diameter from the macular center), and 2 (severe, hard exudate within a circle with a radius of one optic disc diameter from the macular center). DR and DME have a more complex relationship, as shown in Figure 2. As shown in Figure 2a,b, there is a small amount of hard exudate in the macular area of the fundus image, and its DR and DME labels are 2. The DR grade of 2 means that the degree of DR is moderate. The DME grade of 2 means the lesion level of DME is severe, that likely causes the patient to go blind. So, in this circumstance, if only DR grading is done, the extent of the patient's lesion will be misjudged. As seen in Figure 2c,d, there is no exudate in the macular area, which has DR and DME grades of 4 and 0. It means normal with a grading label of 0 for DME, but a grading label of 4 for DR means proliferative, which can also lead to vision loss. In such a case, if only DME grading is done, the degree of the patient's lesion will also be misjudged. Therefore, it is important to automatically grade both DR and DME to assist physicians in taking the appropriate therapeutic options.

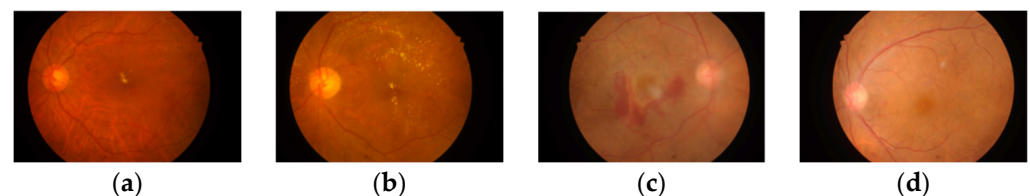


Figure 2. Example images of the degree of DR and DME pathology, (a,b) labeled DR2 DME2, (c,d) labeled DR4 DME0. (a) DR2 DME2; (b) DR2 DME2; (c) DR4 DME0; (d) DR4 DME0.

The study of DR and DME auto-grading based on deep learning has also evolved with the development of convolutional neural networks (CNN). A CNN-based model was first proposed by Pratt et al. [5] to classify the five-level of DR, where they used a class-weighting strategy to update the parameters of each batch during backpropagation to compensate for the class imbalance in the dataset and reduce overfitting. The model designed by Gargeya and Leng [6] had five residual blocks, which first extracted fundus lesion features and then fed the extracted features into a decision tree for secondary classification with and without

lesions. Gulshan et al. [7] used the pre-trained InceptionV3 model on the ImageNet [8] dataset to perform DR classification. Zhang et al. [9] built a high-quality dataset and performed two and four classifications for DR by using an integrated model. Li et al. [10] proposed an integrated algorithm based on multiple improved Inception-v4 for DR lesion detection on retinal fundus images. Wang Z et al. [11] proposed a CNN-based method to simultaneously diagnose DR lesions and highlight suspicious areas. Lin et al. [12] designed an attention fusion-based network with better noise immunity for DR grading, which fused the lesion features by CNN and color fundus images for 5-level DR grading. Zhou Y et al. [13] used a semi-supervised learning method to improve the performance of DR grading and lesion segmentation through collaborative learning.

For DME grading, Perdomo et al. [14] used a two-stage approach to classify DME. In the first stage, an eight-layer CNN model was trained on the e-ophtha database to detect the hard exfiltration, which used 48×48 pixel RGB slices as input; in the second stage, the pre-trained CNN model was used as predictor to generate a new dataset of 1200 grayscale mask images; in the final stage, the DME detection model based on the AlexNet architecture was trained, which consists of a 17-layer CNN using a 512×512 pixel RGB fundus image plus the previously generated grayscale mask as the 4th input channel for training. Mo J et al. [15] constructed a deep residual network through cascading to identify DME. In the model, a fully convolved residual network was used to perform hard exudate mask segmentation, and then according to the segmentation results the region with the highest pixel-centered probability was cropped and fed it into another deep residual network for grading. He X et al. [16] proposed a multiscale collaborative grading model. A multiscale feature extraction model was designed to extract different features, including hard exudation masks, macular masks, and macular images. XGBoost classifier was used to perform the classification training according to the features and the original images.

There has been a lot of research on DR and DME grading and some progress has been made. However, most works are concern at the separate grading of DR or DME. In fact, physicians perform the diagnosis of DR and DME at the same time. Therefore, it is necessary to design a computer vision-based method to automatically grade DR and DME simultaneously. In this paper, a multi-task learning and multi-branch network (MaMNet) is proposed to achieve simultaneous grading of DR and DME.

The contribution of this paper are as follows:

- (i) Build a multi-task learning and multi-branch network (MaMNet) for simultaneous grading of DME and DR. By tapping the relationship between the two grading tasks of DME and DR, multi-task learning can make models more robust and improve the grading accuracy.
- (ii) Design a four-architecture multi-branch network to increase the expression of the underlying features of DME and DR.
- (iii) Design a feature fusion module to fuse the self-features, the cross-feature, and the global features involved in DR or DME to enhance the joint grading accuracy.

The rest sections are organized as follows. Section 2 focuses on the application of multi-task learning, multi-branch networks, and attention mechanisms in medical image processing. Section 3 provides a detailed description of MaMNet. Section 4 illustrates the relevant experimental results. Section 5 concludes this paper.

2. Related Work

2.1. Multi-Task Learning

Multi-task learning is an inductive migration mechanism first proposed by Caruana R. [17] in 1997 with the main goal of improving generalization performance. In multitask learning, if there are certain relationships between the tasks, additional useful information can be dug to train more robust models for better performance.

Multitask learning has been demonstrated to be useful in the field of medical imaging [18–25]. For example, Tabarestani et al. [22] proposed a multitask learning approach with multimodality to predict clinical scores for the development of Alzheimer's disease.

Compared to several other established methods, the presented model would yield minimal prediction error by capturing correlations between different modalities. He et al. [23] designed a multi-task deep transfer learning model for the early prediction of neurodevelopment in very preterm infants. The experimental results showed that the model could effectively identify risk with 81.5% accuracy. Estienne T. et al. [24] proposed a deep learning-based dual-task structure to deal with both alignment and tumor segmentation problems. Evaluated on BraTS 2018 and OASIS 3 datasets, the results demonstrated that the proposed method significantly improved the alignment performance of tumor location and had better segmentation effects. Jin C et al. [25] proposed a multi-task model to accomplish tumor segmentation and response prediction, which was validated in two independent cohorts of 160 and 141 patients and showed that the Area Under the Curve (AUC) could reach 0.95 and 0.92, respectively.

2.2. Multi-Branch Network

In order to extract richer feature information, more and more multi-branching network models have been designed and achieved better results. Hao P et al. [26] proposed a multi branch fusion network for screening MI in 12 lead electrocardiogram images, and the results showed that the proposed method had a good effect. J Zhuang [27] proposed a network called LadderNet for retinal vessel segmentation, which was based on a multi-branch structure. LadderNet had multiple branches consisting of encoders and decoders. Compared with other advanced methods, its segmentation effect was better than other methods. Yang Z et al. [28] proposed a multi-scale convolutional neural network integration (EMS-Net) to classify breast histopathology microscopy images into four categories. The model first converted each image into multiple scales, and then fine-tuned the pre-trained DenseNet-161, ResNet-152, and ResNet-101 at each scale, respectively, and finally used them to form an integrated model. This algorithm was tested on BACH Challenge dataset, and achieved accuracy levels of 90.00% and 91.75%, respectively.

2.3. Attention Mechanism

Bahdanau D et al. [29] first used attention mechanism (AM) for machine translation, and it is now an important part of most deep learning network designs. AM is mostly applied in models used for medical image studies to extract useful features and ignore distracting information. Sinha A et al. [30] designed a medical image segmentation model based on self-directed attention. This method could combine local features with respective global relations to highlight interdependent channel maps. The results show that the model effectively improved the accuracy of segmentation. Cai [31] proposed an improved version of Unet based on multi-scale and attention mechanism for medical image segmentation (MA-Unet). MA-Unet used attention gates (AG) to fuse local features with corresponding global relations, which could attenuate the semantic ambiguity caused by skip-join operations. Compared with other advanced segmentation networks, this model had better segmentation performance and less number of parameters. Valanarasu J et al. [32] proposed a gated attention-based model for medical image segmentation. The model extended the existing architecture by introducing additional control mechanisms in the self-attention module. The segmentation performance was tested on three different datasets and the evaluation results prove that the proposed model was superior to other segmentation models.

3. Proposed Method

Propose a multi-task learning and multi-branch network model (MaMNet) for the joint grading of DME and DR. The MaMNet model is shown in Figure 3, which consists of three main parts: multi-branch network, feature fusion module, and disease grading module. The MbN has a four-branch architecture, by fusing different branches which can extract the low-level features information of DR and DME in a targeted way. The feature fusion module includes SFEN, CFEN, and ASPP. By combining various features that were collected from the aforementioned modules, the feature fusion module can provide more

thorough discriminative features to improve joint grading accuracy. The disease grading module is composed of a combination of GA (Global Average Pooling Layer) and FC (Fully Connected Layer).

The proposed method can be divided into four steps:

Firstly, color fundus images are filtered by the averaging filter, and then the original and the filtered images are weighted and superimposed to obtain the input images.

Secondly, the input images are fed into the MbN to extract the underlying features F_{DME} and F_{DR} . The results of branch 1 and branch 3 are fused to obtain the F_{DME} , and the results of branch 1 and branch 4 are fused to obtain the F_{DR} .

Thirdly, we use SFEN and CFEN to get the self-feature S_{DME} and cross-feature C_{DME} of F_{DME} . And then, F_{DME} , S_{DME} , and C_{DME} are concatenated to get the comprehensive discriminative features D_{DME} . Similarly, the S_{DR} and C_{DR} are obtained from the SEFN and CFEN. The ASPP is used to extract the global lesion feature G_{DR} . S_{DR} , C_{DR} , and G_{DR} are concatenated to get the D_{DR} .

Finally, the D_{DME} and D_{DR} are sent to the grading module of DR and DME respectively for disease grading.

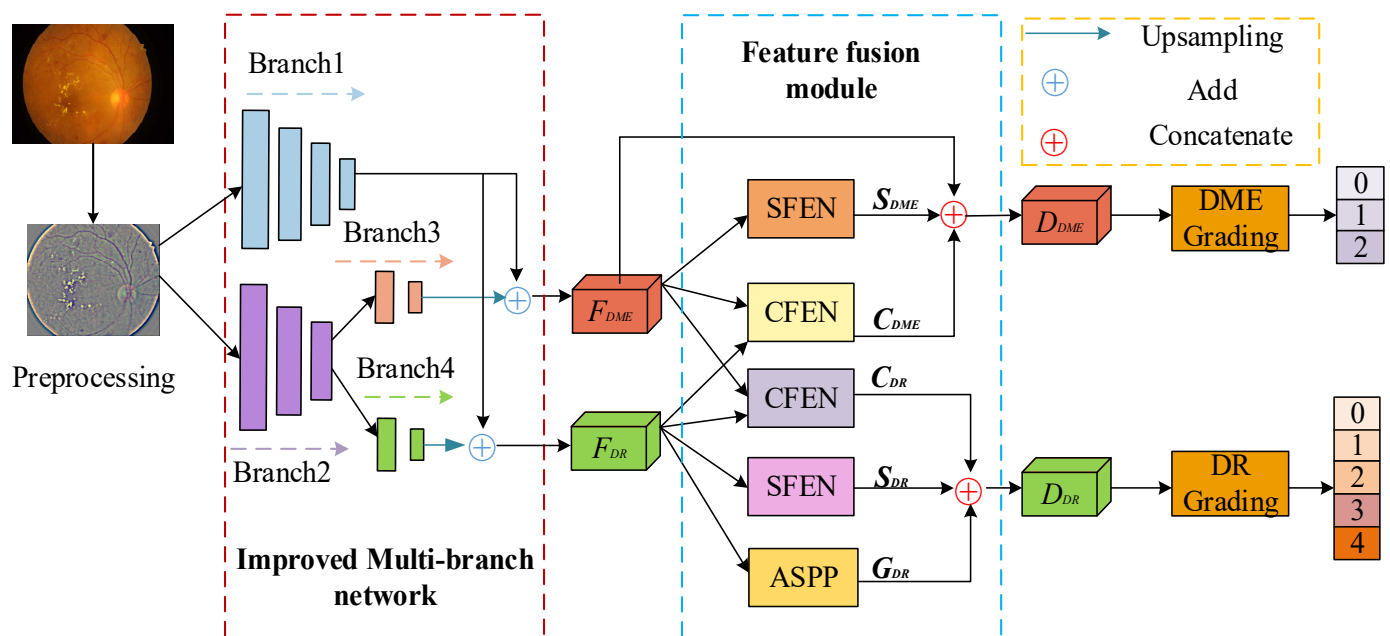


Figure 3. Structure of MaMNet model. It consists of three main parts: multi-branch network, feature fusion module, and disease grading module.

3.1. Proposed Multi-Branch Network

To extract richer underlying lesion features in fundus images, inspired by [33,34], a multi-branch network is designed based on the convolutional module in this paper. As shown in Figure 3, the multi-branch network has a four-branch architecture, and the specific network settings are shown in Table 1. Here, branch 1 has 10 convolutional layers, branch 2 has 7 convolutional layers, branch 3 and branch 4 have 6 convolutional layers, and the convolution kernel size is 3×3 . MbN takes the pre-processed fundus images with the size $224 \times 224 \times 3$ as the input images, and the output is a feature map with the size of $28 \times 28 \times 512$. Branch 1 is shared by DR and DME classification, and branch 3 and branch 4 also share the output of branch 2. This share operation can reduce the model training parameters. The output of branch 3 is up-sampled and fused with the output of branch 1 to get the DME-graded bottom feature F_{DME} . Branch 4 and branch 1 also perform the same operation to get the DR-graded bottom feature F_{DR} .

Table 1. Network structure settings.

Branch 1	Branch 2	Branch 3	Branch 4
Input ($224 \times 224 \times 3$)		Input (branch 2–output)	
conv3 – 64×2	conv3 – 64×2	conv3 – 512×3	conv3 – 512×3
Maxpool (pool_size = (2, 2), strides = 2)			
conv3 – 128×2	conv3 – 128×2	conv3 – 512×3	conv3 – 512×3
Maxpool (pool_size = (2, 2), strides = 2)		Maxpool (pool_size = (3, 3), strides = 1)	
conv3 – 256×3	conv3 – 256×3	UpSampling	
Maxpool (pool_size = (2, 2), strides = 2)			
conv3 – 512×3			
Output	Output	Output	Output

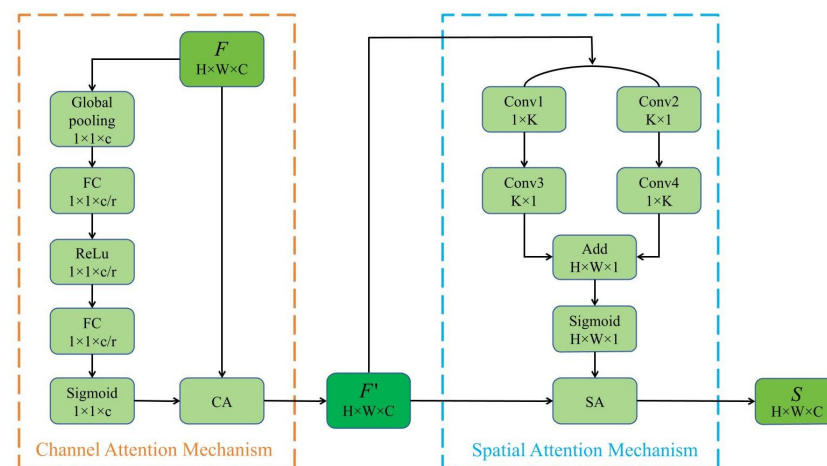
3.2. Feature Fusion Module

In order to improve the grading performance, a multi-feature fusion operation [35] is designed in this paper. According to the different classification criteria of DR and DME, we adopt different fusion operations. The DME grading task fuses the S_{DME} , C_{DME} and F_{DME} to obtain the discriminative feature D_{DME} , and the DR grading task fuses S_{DR} , C_{DR} , and G_{DR} to acquire the D_{DR} .

3.2.1. Self-Feature Extraction Network

DR and DME have their own specific characteristics. DR grading is based on the presence of exudate, hemorrhage, microaneurysms, and other lesions. DME is graded according to the location relationship between the central macular and the hard exudate. However, the features F_{DR} and F_{DME} extracted by the MbN are only the underlying representation of the fundus images, and the detailed self-features of each disease are not easy to capture. To better learn the representative characteristics of each disease, in this paper, the SFEN module is designed to obtain specific discriminatory characteristics of DME and DR.

Figure 4 shows the specific architecture of the SFEN module, which is composed of the CA (channel attention) and SA (spatial attention) mechanisms. SFEN takes the feature map F as the input and uses CA and SA to enhance the inter-channel and interspatial relationship of the features related to each disease.

**Figure 4.** SFEN Module. It consists of two parts, the channel attention and spatial attention mechanisms.

The CA operation is mainly divided into 3 steps:

- (1) F is processed by the global averaging pooling to obtain the feature map with global sensory field Z . The calculation is described in Equation (1).

$$Z = f_{GA}(F) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{i,j} \quad (1)$$

where, Z denotes the global feature, f_{GA} denotes the global averaging pooling operation, F denotes the input feature map, H is the height of F , W is the width of F , and $F_{i,j}$ is the feature vector of the pixel in the i -th row and j -th column.

- (2) Z is processed by the use of two FC operation to get the dependencies between individual channels. The process can be given by Equation (2).

$$CA = \sigma_1(W_2\sigma_2(W_1Z)) \quad (2)$$

here, $W_1 \in R^{1 \times 1 \times \frac{c}{r}}$ is used for dimensionality reduction and $W_2 \in R^{1 \times 1 \times c}$ is used for dimensionality increase. r denotes the scaling factor (set to 4 here), σ_1 is the sigmoid function, and σ_2 is the ReLU activation function.

- (3) The corresponding weight CA is multiplied with the original input feature F to obtain the channeled disease grading discriminant feature F' . The calculation is described in Equation (3).

$$F' = CA \otimes F \quad (3)$$

The SA [36,37] is used to focus on the more important location of the disease feature, which is complementary to the CA. The structure of SA is shown in the blue area of Figure 4. The SA operation is mainly divided into 3 steps:

- (4) The input features F' are convolved by two convolution branches to obtain two spatial feature maps C_1 and C_2 . Every convolution branch has two convolutional layers, one with a kernel of $1 \times k$ and the other with $k \times 1$. This convolutional layer design can enlarge the acceptance domain without increasing the training parameters. The corresponding calculation can be given by Equations (4) and (5).

$$C_1 = conv_3\left(\left(conv_1\left(F', \omega_1^1\right), \omega_1^2\right)\right) \quad (4)$$

$$C_2 = conv_4\left(\left(conv_2\left(F', \omega_2^1\right), \omega_2^2\right)\right) \quad (5)$$

where, $conv_1$ and $conv_4$ are 1×9 convolution, $conv_2$ and $conv_3$ are 9×1 convolution.

- (5) The feature map obtained by fusing C_1 and C_2 is converted to get the spatial weight SA .

$$SA = \sigma_3(C_1 + C_2) \quad (6)$$

where, σ_3 is the sigmoid activation function, the spatial feature weight $SA \in R^{H \times W \times 1}$ belongs to $[0, 1]$.

- (6) F' is multiplied with SA to obtain the disease-specific self-feature maps S .

$$S = SA \otimes F' \quad (7)$$

3.2.2. Cross Feature Extraction Network

DME is one of the common complications of DR, and there is an internal relationship between the two diseases. The two diseases are associated with hard exudate. As the area of hard exudate increases, the risk of DME may become greater, which means an increased likelihood of DR severity. Meanwhile, as the distance between the hard exudate and the central macular recess decreases, there is a greater likelihood of signs of pathological DR, which means the deterioration in DME may lead to the deterioration in DR. Therefore, the CFEN module is designed in this paper to capture the cross features between the two

diseases, i.e., to find out the corresponding feature of one disease from the characteristics of another disease.

As shown in Figure 5, F_{DME} and F_{DR} are taken as the input of CFEN, and the cross features of DR and DME are generated by the learning of the module. Figure 5 gives the detailed structure of the CFEN module for DME classification. The structure for DR grading is similar to DME grading. We take the DME as an example to illustrate the process of CFEN. Firstly, the CA operation is performed on F_{DME} to get channel attention weights CA_{DME} , and then the CA_{DME} of DME is multiplied with F_{DR} to achieve the cross feature C_{DME} . The calculation of C_{DME} and C_{DR} can be described in Equations (8) and (9).

$$C_{DME} = CA_{DME} \otimes F_{DR} \quad (8)$$

$$C_{DME} = CA_{DME} \otimes F_{DR} \quad (9)$$

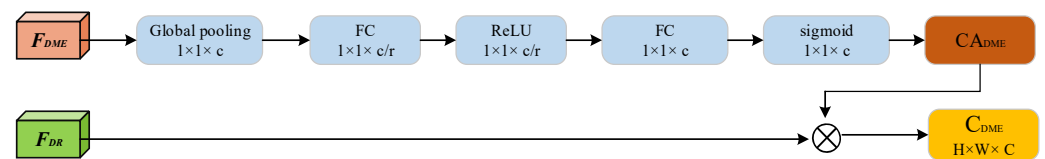


Figure 5. CFEN Module. It is the detailed structure of the CFEN module for DME classification, which consists of two branches.

3.2.3. Atrous Spatial Pyramid Pooling Module

ASPP first appeared in the semantic segmentation network Deeplab [38], which can extract global features of the input images through dilated convolution with different expansion rates. On the one hand, the discriminative features of DR grading are hemorrhage, hard exudate, soft exudate, etc., which are distributed in various locations of retinal images. It is necessary to consider the global nature when extracting lesion features. On the other hand, larger dilation rates would lead to the loss of smaller lesion features, which is unfavorable to DR grading. So, to get better grading performance, the modified ASPP is designed in this paper to extract the global context information. The schematic diagram of the modified ASPP module is shown in Figure 6.

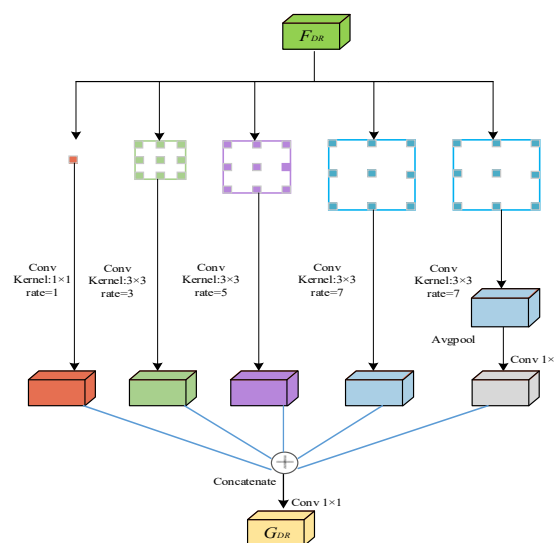


Figure 6. Schematic diagram of the modified ASPP structure. It consists of five branches, and the expansion rates of these five branches are 1, 3, 5, 7, and 7 respectively.

As seen in Figure 6, the modified ASPP has a five-branch architecture. The underlying features F_{DR} are firstly processed by the dilated convolution with different expansion rates

to obtain the image feature maps with different receptive fields. The expansion rates of these five branches are 1, 3, 5, 7, and 7 respectively. The size of the convolution kernels of these five branches is 1×1 , 3×3 , 3×3 , 3×3 and 3×3 . Then, for the fifth branch, we apply the average pooling and 1×1 operations to generate the feature map. Finally, these five feature maps are concatenated together and a 1×1 convolutional operation is used to obtain the output feature G_{DR} .

3.3. Disease Classification

The comprehensive features D (D_{DME} or D_{DR}) obtained through the feature fusion module are fed into the disease grading module to achieve the grading results. According to the grading task, DME outputs three types of labels, and DR outputs five types of labels. Figure 7 shows the detailed structure of the disease grading module. DR and DME have the same structure.

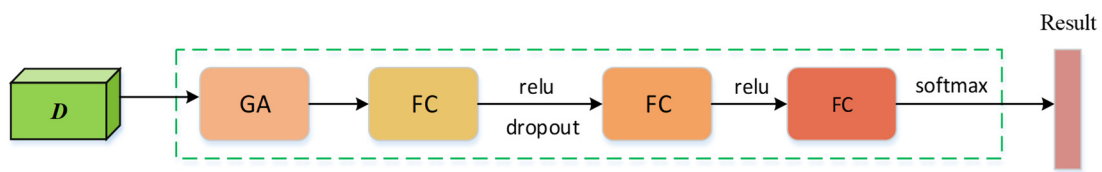


Figure 7. Disease classification module. It consists of GA, FC, relu and softmax activation functions, and a drop-out layer.

The grading module shown in Figure 7 includes a global average pooling layer (GA), a fully connected layer (FC), relu and softmax activation functions, and a drop-out layer with a drop rate of 0.5.

For DR and DME grading, we utilize weighted L_{DR} and L_{DME} as the total loss of the model, which can be given by Equation (10).

$$L = \alpha L_{DR} + \beta L_{DME} \quad (10)$$

where, L is the total loss of the model, L_{DR} is the loss of DR classification, L_{DME} is the loss of DME classification, and L_{DR} and L_{DME} are both cross-entropy loss functions.

The cross-entropy loss function is calculated by Equation (11).

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^M y_n^c \log \hat{y}_n^c \quad (11)$$

Here, y_n^c represents the image true label, $\hat{y}_n^c$ represents the probability that the image predicted disease classification belongs to label c , N is the number of images included in the training set, M is the number of disease classification label categories, for DR classification, M is 5; for DME classification, M is 3.

4. Experiment and Results

4.1. Experimental Settings

4.1.1. Datasets

The grading performance is tested on the IDRiD [39] dataset, which has 516 fundus images and is divided into a training set and a test set. The training set includes 413 images, and the test set contains 103 images. Each image in the training set and test set has two types of labels: DR-graded labels and DME-graded labels. According to the severity, DR has five levels of labeling (0,1,2,3,4) and DME has three levels of labeling (0,1,2).

4.1.2. Evaluation Metrics

In this paper, the Accuracy Ac and joint accuracy rate $Joint\ Ac$ are used to assess the grading performance of the proposed model.

The *Ac* is defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (12)$$

where, *TP* (true-positive) denotes the number of positive samples successfully predicted to positive samples. *FP* (false-positive) is the number of negative samples incorrectly predicted as positive sample. *TN* (true-negative) is the number of negative samples predicted correctly as negative samples. *FN* (false-negative) represents the number positive samples wrongly predicted as negative samples.

The *Joint Ac* is defined as follows:

$$Joint\ Ac = \frac{total(R \cap E)}{T} \quad (13)$$

where, *R* and *E* represent the predicted correct images in DR and DME grading test sets, respectively. *T* is the total number of images in the test set. When the DR prediction is correct, *R* is 1, and when the prediction is wrong, *R* is 0. Similarly, when the DME prediction is correct, *E* is 1, and when the prediction is wrong, *E* is 0. When both *R* and *E* take 1, $R \cap E$ is 1, which means both diseases are correctly predicted. $total(\cdot)$ indicates the total number of simultaneous correct predictions for both diseases.

4.1.3. Experimental Environment and Training Methods

In this paper, experiments are conducted on Google Colaboratory using GPU, and the overall framework of the model is based on Keras. The training set and test set are preprocessed and dimensionally transformed to get the input images, which have a size of 224×224 . The Adam optimizer is used to set the dynamic learning rate, which is 0.0001 at the beginning. If the loss of the test set does not decrease when the model training goes through 3 epochs, the learning rate will change to 0.5 times the original one, and the learning rate will decrease to 0.000001 at the minimum. To avoid the occurrence of model overfitting, EarlyStopping is used to end the training. The training will be stopped if the loss of the test set no longer decreases after 5 epochs. The total epoch is 50 and the batch size is 8. All the hyperparameter values are shown in Table 2.

Table 2. Hyperparameter values.

Parameter	Value
WARMUP_EPOCHS	10
LEARNING_RATE	1×10^{-4}
WARMUP_LEARNING_RATE	1×10^{-3}
ES_PATIENCE	5
RLROP_PATIENCE	3
DECAY_DROP	0.5
EPOCHS	50
BATCH_SIZE	8

4.2. Experimental Results and Analysis

In order to verify the effectiveness of the MaMNet model, four groups of experiments are set up in this paper.

The first group of experiments: For DR and DME grading, we utilize weighted L_{DR} and L_{DME} as the total loss of the MaMNet model. To select the optimal combination of weights, we set 16 different sets of values of loss weights for the experiments.

The second group of experiments: The performance of the SA module differs when the SA module convolution kernel parameter *K* is set to different values in the SFEN module. So, we conducted the experiment to decide the values of *K*.

The third group of experiments: To verify the validity of each module in the MaMNet model, we perform the ablation experiment.

The fourth group of experiments: To demonstrate the validity of the MaMNet for the joint classification of DR and DME, we compare MaMNet with 11 other models.

4.2.1. Loss Weight Setting Experiment

We use weighted L_{DR} and L_{DME} as the MaMNet model's overall loss for DME and DR classification. To select the optimal combination of weights, we set 16 different sets of values of loss weights for the experiments. The loss weights α and β for both DME-graded and DR-graded are given as starting values of 0.25 and are increased in steps of 0.25.

From the data in Table 3, it can be seen that the setting of α and β values has a great influence on the values of *Joint Ac*, *DR Ac*, and *DME Ac*. The values of *Joint Ac*, *DR Ac*, and *DME Ac* are different when α and β values are different. The aim of this paper is to grade DR and DME simultaneously, *Joint Ac* is a more important evaluation index. When $(\alpha, \beta) = (0.5, 1.0)$, *Joint Ac* achieves a maximum value of 0.612. So, we choose $(\alpha, \beta) = (0.5, 1.0)$ as the optimal combination weight of the MaMNet model.

Table 3. Training results for DR and DME with different weighted losses.

α	β	<i>Joint Ac</i>	<i>DR Ac</i>	<i>DME Ac</i>
0.25	0.25	0.5049	0.5437	0.7573
0.25	0.5	0.4466	0.5049	0.8155
0.25	0.75	0.5146	0.5534	0.7961
0.25	1.0	0.5146	0.5728	0.7670
0.5	0.25	0.4660	0.5243	0.7864
0.5	0.5	0.4951	0.5534	0.8058
0.5	0.75	0.5922	0.6117	0.8058
0.5	1.0	0.6117	0.6407	0.7961
0.75	0.25	0.4757	0.5436	0.7961
0.75	0.5	0.4466	0.5340	0.7961
0.75	0.75	0.5243	0.5340	0.7767
0.75	1.0	0.5728	0.5825	0.8155
1.0	0.25	0.4466	0.4854	0.7864
1.0	0.5	0.5340	0.5922	0.7961
1.0	0.75	0.5243	0.5437	0.7670
1.0	1.0	0.5437	0.5534	0.8058

The bold in the table represent the highest *Joint Ac*.

4.2.2. K-Parameter Setting Experiment

The performance of the SA module differs when the SA module convolutional kernel parameter K is set to different values in the SFEN module. Therefore, we conducted the experiments to decide the values of K . The values of K are set to 5, 7, 9, and 11.

From Table 4, it can be seen that the *Joint Ac*, *DR Ac*, and *DME Ac* values have been significantly affected by the K value setting. Different K can get different *Joint Ac*, *DR Ac*, and *DME Ac*. The larger K can't get the better the results for the DR and DME grading task. When K is 9, the values of *Joint Ac*, *DR Ac*, and *DME Ac* are greater than the values at $K = 5, 7$, and 11. So, we choose $K = 9$ as the optimal result.

Table 4. Convolution kernel parameter K setting.

K	<i>Joint Ac</i>	<i>DR Ac</i>	<i>DME Ac</i>
5	0.534	0.563	0.777
7	0.534	0.583	0.796
9	0.612	0.641	0.796
11	0.563	0.583	0.786

The bold in the table represent the highest *Joint Ac* and *DR Ac*.

4.2.3. Module Ablation Experiments

In order to verify the effectiveness of CFEN and ASPP of MaMNet, the ablation experiments are conducted in this section, the loss weights are set to the optimal loss weights i.e., $(\alpha, \beta) = (0.5, 1.0)$.

The following conclusions can be drawn from Table 5:

- (1) When the CFEN module is added to the branch for DME grading, DME *Ac* improved by 0.029 and DR *Ac* improved by 0.010; when the CFEN module is added to the subnet for DR grading, *Joint Ac* improved by 0.009 and DR *Ac* improved by 0.019. The above results demonstrated that the CFEN module explored more correlation features for DME grading and DR grading.
- (2) Due to the addition of the ASPP module, *Joint Ac*, DME *Ac*, and DR *Ac* are improved by 0.020, 0.019, and 0.047 respectively in the joint grading course. From the experimental data, it can be seen that ASPP can better extract global lesion features and improve the grading performance.
- (3) Due to the CFEN and ASPP module, the highest values of *Joint Ac*, DR *Ac*, and DME *Ac* are obtained. These results demonstrate that MaMNet can effectively extract the self-features, correlated features between DME and DR, and the global features of DR to enhance the grading performance.

Table 5. Ablation experiments.

	Methods	<i>Joint Ac</i>	DR <i>Ac</i>	DME <i>Ac</i>
1	DR: MbN + SFEN DME: MbN + SFEN	0.583	0.602	0.767
2	DR: MbN + SFEN DME: MbN + SFEN + CFEN	0.583	0.612	0.796
3	DR: MbN + SFEN + CFEN DME: MbN + SFEN	0.592	0.621	0.777
4	DR: MbN + SFEN + CFEN DME: MbN + SFEN + CFEN	0.592	0.621	0.777
5	DR: MbN + SFEN + CFEN + ASPP DME: MbN + SFEN + CFEN	0.6117	0.6407	0.7961

4.2.4. Comparison Experiments

To verify the effectiveness of MaMNet for joint grading of DME and DR, the proposed method MaMNet is compared with other eleven methods, namely, LzyUNCC [40], VRT [40], Mammoth [40], HarangiM1 [40], AVSASVA [40], VGG16 [41], ResNet50 [42], InceptionV3 [43], Xception [44], and DenseNet121 [45]. The specific experimental results are shown in Table 6.

Table 6. Comparison of results.

Method	<i>Joint Ac</i>	DR <i>Ac</i>	DME <i>Ac</i>
LzyUNCC [40]	0.631	0.748	0.806
VRT [40]	0.553	0.592	0.816
Mammoth [40]	0.515	0.544	0.835
HarangiM1 [40]	0.476	0.553	0.748
AVSASVA [40]	0.476	0.553	0.806
HarangiM2 [40]	0.408	0.476	0.728
VGG16 [41]	0.524	0.583	0.767
ResNet50 [42]	0.524	0.592	0.757
InceptionV3 [43]	0.437	0.563	0.767
DenseNet121 [44]	0.456	0.485	0.699
Xception [45]	0.467	0.515	0.738
Proposed Approach	0.612	0.641	0.796

The bold in the table represent the best and second-best results.

The first to sixth row shows the results of the methods used by the top five ranked teams in challenge-2 of the ISBI 2018 IDRiD competition (which is ranked based on joint accuracy). The seventh to eleventh row lists the results of the five classical neural networks in deep learning, which are trained with migration learning.

It can be seen from Table 6 that compared with the 11 advanced methods MaMNet can achieve the second-best results. Although MaMNet doesn't achieve the best results, it still has the following advantages: (1) MaMNet is an end-to-end network that can achieve simultaneously DR and DME grading. The LzyUNCC team trains DR and DME grading networks separately; (2) The best training model of MaMNet takes only 15 min to finish the training and achieves better results. The model [46] proposed by the LzyUNCC team has a large amount of computation and requires more time to obtain the classification results; (3) MaMNet only uses the IDRiD dataset provided by the competition for training. The LzyUNCC team first used 35,000 retinal images from the Kaggle dataset to train the initial model and then fine-tuned it with the IDRiD dataset. In summary, it can be demonstrated that MaMNet performs well on DR and DME joint classification.

4.2.5. Visualization of the Best Model Results

The confusion matrices of DME and DR are shown in Figure 8. In general, both the matrices have a diagonal tendency, i.e., the predictions are close to the ground truth. DME1 and DR1 are the most misclassified grade.

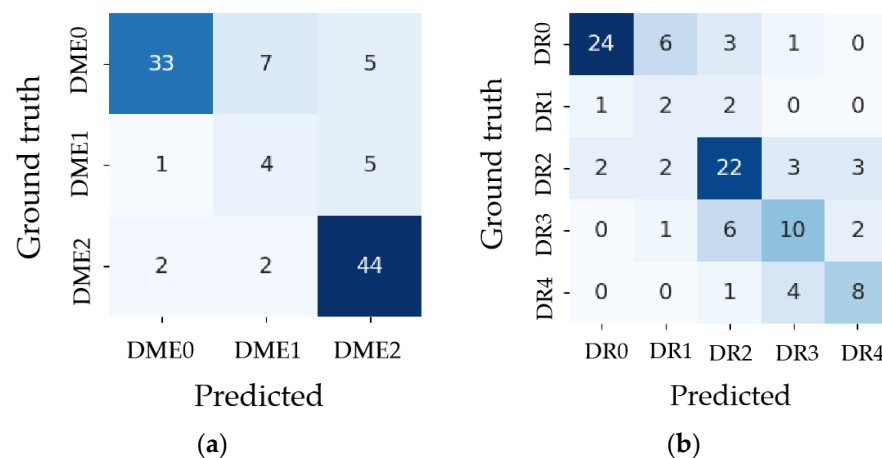


Figure 8. Confusion matrixes of MaMNet's predictions. (a) DME's Confusion Matrix; (b) DR's Confusion Matrix.

The receiver operating characteristic (ROC) curve and the area under the precision-recall curve (AUPR) of our method on the IDRiD dataset are displayed in Figure 9. These curves exhibit the stable training and classification performance of the proposed model.

Visualization of some graded prediction scores is listed in Table 7. The table is divided into four columns: the input fundus image, the true label for DR and DME, the predicted label, and the graded score for the disease level. The grading score is defined as the probability that the model predicts for each disease. The bold data represent the grading scores of the correctly classified category.

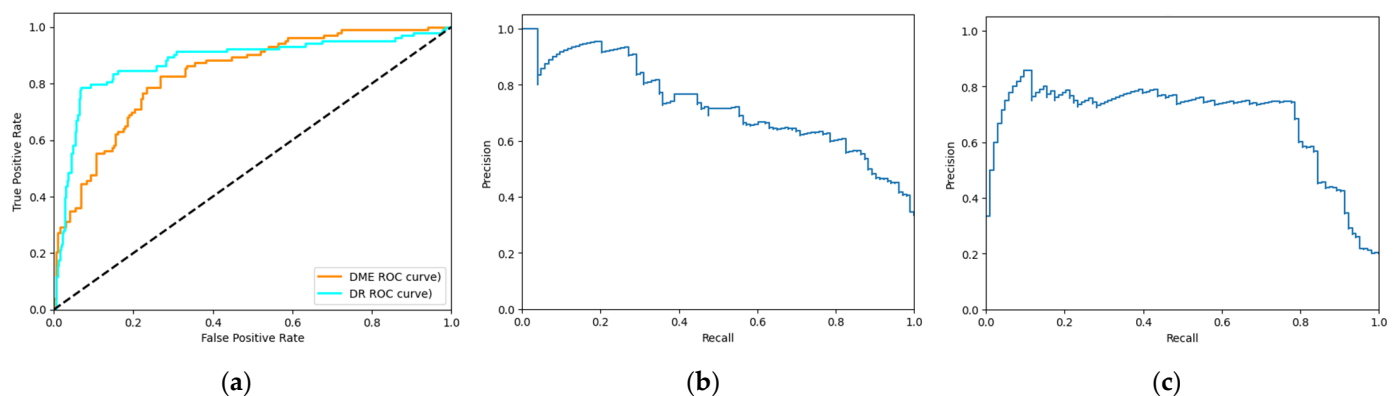
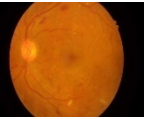
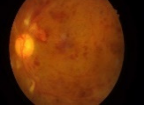


Figure 9. ROC curve and AUCPR curves of DR and DME. (a) ROC Curve; (b) DR's AUPR Curve; (c) DME's AUPR Curve.

It can be seen from Table 7 that MaMNet can correctly distinguish the severity of DR and DME and obtain consistent results with the true labels of fundus images. When the grading level is correctly predicted, this level can achieve a higher grading score, while the other levels result in essentially 0.

Table 7. Visualization of the best model prediction results.

Input Image	True Label	Predicted Label	Various Grading Scores
	DR3 DME2	DR3 DME2	DR 0.000 0.000 0.093 0.697 0.209 DME 0.000 0.052 0.947
	DR2 DME2	DR2 DME2	DR 0.010 0.008 0.889 0.072 0.021 DME 0.002 0.026 0.972
	DR0 DME0	DR0 DME0	DR 0.912 0.022 0.043 0.016 0.007 DME 0.902 0.065 0.033
	DR2 DME2	DR2 DME2	DR 0.006 0.005 0.924 0.047 0.017 DME 0.000 0.008 0.991
	DR0 DME0	DR0 DME0	DR 0.991 0.003 0.004 0.001 0.000 DME 0.970 0.018 0.012
	DR4 DME2	DR4 DME2	DR 0.000 0.000 0.017 0.297 0.686 DME 0.000 0.000 1.000

The bold in the table represent the grading scores of the correctly classified category.

5. Conclusions

In this paper, a joint grading model of DR and DME based on multi-tasking and multi-branching MaMNet is proposed. Thanks to the multi-task learning strategy, MaMNet increases the grading accuracy by tapping more correlated information of DR and DME. A multi-branch network is designed to extract the underlying features of DR and DME in a targeted manner. The feature fusion module is constructed to obtain more comprehensive

and disease-specific discriminative features. To validate the performance of MaMNet, experiments are conducted on the challenging IDRiD dataset. The experimental results show that compared with the other 11 advanced methods, the proposed method has better grading performance, and its joint accuracy, DR accuracy, and DME accuracy can reach 61.2%, 64.1%, and 79.6%, respectively.

The dataset IDRiD used for model training is a public dataset with both DR and DME classification labels, only 413 images are used for training. To further improve the grading accuracy, the fundus images of diabetic patients with labels of both DR and DME disease categories will be collected subsequently to build a large dataset for joint grading and to train the model. In addition, we also want to explore the relationship between multiple diseases that occur in one fundus image in the future. At the same time, considering the existing issue of low accuracy, in future research directions, it may be beneficial to utilize classifiers such as Support Vector Machines (SVM), Naive Bayes (NB), and k-Nearest Neighbors (KNN) for the classification of DR and DME using features extracted through MaMNet [47].

In the last, for a new researcher, applying existing network architectures to new cases can be challenging. Here, a procedural guide is presented. The initial step entails inputting training images and labels for model training. Upon achieving a certain accuracy threshold, the trained model is stored for subsequent testing. In the subsequent step, the qualified model from the initial phase is employed on the test set to acquire corresponding labels for the test set images. This facilitates an assessment of the accuracy of the test results. Once a specific precision level is reached, the third step can be undertaken, involving the application of the final qualified model to a grading task with unknown labels. The final grading prediction results can assist physicians in conducting preliminary diagnosis and treatment for patients.

Author Contributions: Conceptualization, X.X. and M.Y.; methodology, X.X. and M.Y.; software, S.M., M.Y. and H.Y.; validation, S.M., D.Y., J.Z. and C.Z. (Cancan Zhu); formal analysis, S.M. and M.Y.; investigation, C.Z. (Cong Zhang); writing—original draft preparation, M.Y.; writing—review and editing, X.X., M.Y. and S.M.; supervision, T.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Project of Jilin Provincial Department of Science and Technology (No.2022LY402L06).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

DR	Diabetic Retinopathy
DME	Diabetic Macular Edema
MaMNet	Multi-task learning and Multi-branch Networks
MbN	Multi-branch Network
SFEN	Self-Feature Extraction module
CFEN	Cross-Feature Extraction module
ASPP	Atrous Spatial Pyramid Pooling module
GA	Global Average Pooling Layer
FC	Fully Connected Layer

References

1. Wilkinson, C.P.; Ferris, F.L.; Klein, R.E.; Lee, P.P.; Agardh, C.D.; Davis, M.; Dills, D.; Kampik, A.; Pararajasegaram, R.; Verdaguer, J.T. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* **2003**, *110*, 1677–1682. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Ciulla, T.A.; Amador, A.G.; Zinman, B. Diabetic retinopathy and diabetic macular edema: Pathophysiology, screening, and novel therapies. *Diabetes Care* **2003**, *26*, 2653–2664. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Li, T.; Bo, W.; Hu, C.; Kang, H.; Liu, H.; Wang, K.; Fu, H. Applications of deep learning in fundus images: A review. *Med. Image Anal.* **2021**, *69*, 101971. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Wu, L.; Fernandez-Loaiza, P.; Sauma, J.; Hernandez-Bogantes, E.; Masis, M. Classification of diabetic retinopathy and diabetic macular edema. *World J. Diabetes* **2013**, *4*, 290. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Pratt, H.; Coenen, F.; Broadbent, D.M.; Harding, S.P.; Zheng, Y. Convolutional neural networks for diabetic retinopathy. *Procedia Comput. Sci.* **2016**, *90*, 200–205. [\[CrossRef\]](#)
6. Gargeya, R.; Leng, T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* **2017**, *124*, 962–969. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Gulshan, V.; Peng, L.; Coram, M.; Stump, M.C.; Du, W. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **2016**, *316*, 2402–2410. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Fei-Fei, L. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [\[CrossRef\]](#)
9. Zhang, W.; Zhong, J.; Yang, S.; Gao, Z.; Hu, J.; Chen, Y.; Yi, Z. Automated identification and grading system of diabetic retinopathy using deep neural networks. *Knowl.-Based Syst.* **2019**, *175*, 12–25. [\[CrossRef\]](#)
10. Li, F.; Wang, Y.; Xu, T.; Dong, L.; Yan, L.; Jiang, M.; Zou, H. Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs. *Eye* **2022**, *36*, 1433–1441. [\[CrossRef\]](#)
11. Wang, Z.; Yin, Y.; Shi, J.; Fang, W.; Li, H.; Wang, X. Zoom-in-net: Deep mining lesions for diabetic retinopathy detection. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, 11–13 September 2017.
12. Lin, Z.; Guo, R.; Wang, Y.; Wu, B.; Chen, T.; Wang, W.; Wu, J. A framework for identifying diabetic retinopathy based on anti-noise detection and attention-based fusion. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, 16–20 September 2018.
13. Zhou, Y.; He, X.; Huang, L.; Liu, L.; Zhu, F.; Cui, S.; Shao, L. Collaborative learning of semi-supervised segmentation and classification for medical images. In Proceedings of the 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
14. Perdomo, O.; Ojalora, S.; Rodríguez, F.; Arevalo, J.; González, F.A. A novel machine learning model based on exudate localization to detect diabetic macular edema. *Ophthalmic Med. Image Anal. Int. Workshop* **2016**, *3*, 137–144.
15. Mo, J.; Zhang, L.; Feng, Y. Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks. *Neurocomputing* **2018**, *290*, 161–171. [\[CrossRef\]](#)
16. He, X.; Zhou, Y.; Wang, B.; Cui, S.; Shao, L. Dme-net: Diabetic macular edema grading by auxiliary task learning. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, 13–17 October 2019.
17. Caruana, R. Multitask learning. *Mach. Learn.* **1997**, *28*, 41–75. [\[CrossRef\]](#)
18. Tan, C.; Zhao, L.; Yan, Z.; Li, K.; Metaxas, D.; Zhan, Y. Deep multi-task and task-specific feature learning network for robust shape preserved organ segmentation. In Proceedings of the 15th IEEE International Symposium on Biomedical Imaging, Washington, DC, USA, 4–7 April 2018.
19. Liu, L.; Dou, Q.; Chen, H.; Olatunji, I.E.; Qin, J.; Heng, P.A. Mtmr-net: Multi-task deep learning with margin ranking loss for lung nodule analysis. In Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis (DLMIA)/8th International Workshop on Multimodal Learning for Clinical Decision Support (ML-CDS), Granada, Spain, 20 September 2018.
20. Chen, Q.; Peng, Y.; Keenan, T.; Dharssi, S.; Agro, E.; Wong, W.T.; Lu, Z. A multitask deep learning model for the classification of age-related macular degeneration. *AMIA Summits Transl. Sci. Proc.* **2019**, *2019*, 505–514.
21. Xu, X.; Zhou, F.; Liu, B.; Bai, X. Multiple organ localization in CT image using triple-branch fully convolutional networks. *IEEE Access* **2019**, *7*, 98083–98093. [\[CrossRef\]](#)
22. Tabarestani, S.; Aghili, M.; Eslami, M.; Cabrerizo, M.; Barreto, A.; Rishe, N.; Adjouadi, M. A distributed multitask multimodal approach for the prediction of Alzheimer’s disease in a longitudinal study. *NeuroImage* **2020**, *206*, 116317. [\[CrossRef\]](#)
23. He, L.; Li, H.; Wang, J.; Chen, M.; Gozdas, E.; Dillman, J.R.; Parikh, N.A. A multitask, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. *Sci. Rep.* **2020**, *10*, 15072. [\[CrossRef\]](#)
24. Estienne, T.; Lerousseau, M.; Vakalopoulou, M.; Alvarez Andres, E.; Battistella, E.; Carré, A.; Deutsch, E. Deep learning-based concurrent brain registration and tumor segmentation. *Front. Comput. Neurosci.* **2020**, *14*, 17. [\[CrossRef\]](#)
25. Jin, C.; Yu, H.; Ke, J.; Ding, P.; Yi, Y.; Jiang, X.; Li, R. Predicting treatment response from longitudinal images using multi-task deep learning. *Nat. Commun.* **2021**, *12*, 1851. [\[CrossRef\]](#)
26. Hao, P.; Gao, X.; Li, Z.; Zhang, J.; Wu, F.; Bai, C. Multi-branch fusion network for Myocardial infarction screening from 12-lead ECG images. *Comput. Methods Programs Biomed.* **2020**, *184*, 105286. [\[CrossRef\]](#)

27. Zhuang, J. LadderNet: Multi-path networks based on U-Net for medical image segmentation. *arXiv* **2018**, arXiv:1810.07810.
28. Yang, Z.; Ran, L.; Zhang, S.; Xia, Y.; Zhang, Y. EMS-Net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images. *Neurocomputing* **2019**, *366*, 46–53. [\[CrossRef\]](#)
29. Chaudhari, S.; Mithal, V.; Polatkan, G.; Ramanath, R. An attentive survey of attention models. *ACM Trans. Intell. Syst. Technol. (TIST)* **2021**, *12*, 1–32. [\[CrossRef\]](#)
30. Sinha, A.; Dolz, J. Multi-scale self-guided attention for medical image segmentation. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 121–130. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Cai, Y.; Wang, Y. Ma-unet: An improved version of unet based on multi-scale and attention mechanism for medical image segmentation. In Proceedings of the Third International Conference on Electronics and Communication; Network and Computer Technology (ECNCT 2021), Xiamen, China, 5 April 2021.
32. Valanarasu, J.M.J.; Oza, P.; Hacıhaliloglu, I.; Patel, V.M. Medical transformer: Gated axial-attention for medical image segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021.
33. Wu, Z.; Su, L.; Huang, Q. Cascaded partial decoder for fast and accurate salient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
34. Xie, H.; Zeng, X.; Lei, H.; Du, J.; Wang, J.; Zhang, G.; Lei, B. Cross-attention multi-branch network for fundus diseases classification using SLO images. *Med. Image Anal.* **2021**, *71*, 102031. [\[CrossRef\]](#)
35. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
36. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large kernel matters—improve semantic segmentation by global convolutional network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
37. Zhao, T.; Wu, X. Pyramid feature attention network for saliency detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
38. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [\[CrossRef\]](#)
39. Porwal, P.; Pachade, S.; Kamble, R.; Kokare, M.; Deshmukh, G.; Sahasrabudhe, V.; Meriaudeau, F. Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research. *Data* **2018**, *3*, 25. [\[CrossRef\]](#)
40. Diabetic Retinopathy: Segmentation and Grading Challenge. Available online: <https://idrid.grand-challenge.org/Leaderboard/> (accessed on 27 March 2019).
41. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
43. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
44. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
45. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
46. Yu, F.; Wang, D.; Shelhamer, E.; Darrell, T. Deep layer aggregation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
47. Afreen, S.; Bhurjee, A.K.; Aziz, R.M. Gene selection with Game Shapley Harris hawks optimizer for cancer classification. *Chemom. Intell. Lab. Syst.* **2023**, *242*, 104989. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.