electronics

Article

Fundus Image Generation and Classification of Diabetic Retinopathy Based on Convolutional Neural Network

Peiming Zhang [1,†], Jie Zhao [2,†], Qiaohong Liu [3,*], Xiao Liu [1], Xinyu Li [1], Yimeng Gao [1] and Weiqi Li [1]

[1] School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; zhangpm@usst.edu.cn (P.Z.); liuxiao_bme@163.com (X.L.); xinyuli_01@163.com (X.L.); 15305278033@163.com (Y.G.); wqli@usst.edu.cn (W.L.)
[2] Department of Ophthalmology, Shidong Hospital of Shanghai Yangpu District, Shanghai 200438, China; jeslord@163.com
[3] School of Medical Instruments, Shanghai University of Medicine and Health Sciences, No. 279, Zhouzhu Road, Pudong New Area, Shanghai 200237, China
* Correspondence: liuqh@sumhs.edu.cn
† These authors contributed equally to this work.

Abstract: To detect fundus diseases, for instance, diabetic retinopathy (DR) at an early stage, thereby providing timely intervention and treatment, a new diabetic retinopathy grading method based on a convolutional neural network is proposed. First, data cleaning and enhancement are conducted to improve the image quality and reduce unnecessary interference. Second, a new conditional generative adversarial network with a self-attention mechanism named SACGAN is proposed to augment the number of diabetic retinopathy fundus images, thereby addressing the problems of insufficient and imbalanced data samples. Next, an improved convolutional neural network named DRMC Net, which combines ResNeXt-50 with the channel attention mechanism and multi-branch convolutional residual module, is proposed to classify diabetic retinopathy. Finally, gradient-weighted class activation mapping (Grad-CAM) is utilized to prove the proposed model's interpretability. The outcomes of the experiment illustrates that the proposed method has high accuracy, specificity, and sensitivity, with specific results of 92.3%, 92.5%, and 92.5%, respectively.

Keywords: fundus image generation; fundus image classification; diabetic retinopathy; generative adversarial network; class activation map

check for updates

Citation: Zhang, P.; Zhao, J.; Liu, Q.; Liu, X.; Li, X.; Gao, Y.; Li, W. Fundus Image Generation and Classification of Diabetic Retinopathy Based on Convolutional Neural Network. *Electronics* **2024**, *13*, 3603. https://doi.org/10.3390/electronics13183603

Academic Editors: Gemma Piella and Maria Filomena Santarelli

Received: 17 July 2024
Revised: 14 August 2024
Accepted: 3 September 2024
Published: 11 September 2024

Copyright: © 2024 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 is a chronic disease, and its characteristic is elevated blood sugar levels caused by the pancreas's inability to secrete enough insulin. Diabetes can be divided into two different categories: Type 1 diabetes mellitus (T1DM), where the body is unable to secrete insulin, and Type 2 diabetes mellitus (T2DM), where insulin can be secreted but cannot be effectively utilized [1]. Over the past few decades, the incidence of diabetes has risen significantly, and both patients of T1DM and T2DM are in potential risk of developing diabetic retinopathy [2]. By 2019, the number of diabetes patients increased to 463 million, and the number of adult patients is estimated to increase to 700 million by 2045. Such a trend reflects a sharp increase in the number of diabetes patients worldwide, highlighting the growing public health problem posed by the disease.

The fundus is the only part in which the blood circulation and microvascular structure can be directly observed in human bodies, making it significant in medical diagnosis. DR is a common and severe concurrent disorders of diabetes resulting from microvascular injury in the retinal blood vessels. DR can be sorted into five different stages: normal, mild, moderate, severe or non-proliferative, and proliferative [3]. Without proper screening and treatment, it can slowly progress through these stages. During DR, various lesions gradually appear in the eyes, such as microaneurysms with mild DR, bleeding and exudate

in moderate DR, neovascularization in non-proliferative DR, and fragile blood vessels and scar tissue in proliferative DR [3]. These pathological changes distort the retina step by step, and will do further damage to the macula.

Regular screening is essential for diabetes patients in order to ensure the early detection of DR and prevent complications. Traditional fundus examinations typically require patients to visit a hospital for fundus imaging, which is then diagnosed by ophthalmologists based on these images. Manual diabetic retinopathy detection necessitates highly skilled practitioners for accurate assessment, and even experienced ophthalmologists are at risk of misdiagnosis and missed diagnoses. However, with the increasing number of patients, the shortage of specialized ophthalmologists has become a growing issue, leading to a heavier workload for doctors. These drawbacks not only affect the timeliness and accuracy of diagnoses but also enhance the burden on the healthcare system.

Computer-aided diagnosis technology for the fundus is of great significance. With the improvement of deep learning models and advancements in computational power, the accuracy of DR grading has significantly increased, further demonstrating the potential of deep learning in the auxiliary diagnosis of diabetic retinopathy. In 2016, Gulshan et al. [4] applied deep learning to create an ensemble of 10 convolutional neural networks (CNNs) with Inception-v3 [5] architecture for the automated detection of diabetic retinopathy and diabetic macular edema in retinal fundus photographs. Their work demonstrated that an algorithm based on deep learning had high sensitivity and specificity for detecting referable diabetic retinopathy. CNNs have a powerful capability for automatic DR classification due to their ability to learn complex features and avoid manual feature extraction. Subsequently, some classical CNNs, such as FCN [6], VGG-16 [7], ResNet [8], et al., are used as the preferred solution for automatic classification tasks, including DR grading. Costa et al. [9] converted a pre-trained FCN into a weakly-supervised model that can not only detect DR in eye fundus images but can also pinpoint the regions of images that contain lesions. Wang et al. [10] proposed a CNN-based method for the joint learning of multi-level tasks, including DR grading, image super-resolution, and lesion segmentation, in which ResNet-18 is used to realize the DR grading subnet. Rocha et al. [11] focused on addressing the challenges in medical image analysis, including low contrast, poor lighting, and high noise levels. Their study used a VGG16 network to classify retinal fundus images into relevant categories. Additionally, various attention mechanisms, such as SE [12], CBAM [13], ECA [14], and CA [15], have been adopted to enhance the feature representation ability and focus on the important regions, resulting in the performance advancement of CNNs. He et al. [16] constructed the CABNet for DR grading, which addresses the imbalanced DR data distribution problem in an end-to-end manner. Bhati et al. [17] utilized an attention block followed by a SE block to construct a DKCNet for the multi-label classification of ODIR-5K fundus images, which explores discriminative region-wise features without adding an extra computational cost. Hai et al. [18] designed a two-stage DRGCNN to achieve marked enhancements in DR grading performance. The first component combines the EfficientNetV2-M model with the category attention module to encode the input fundus retinal images into feature vectors. The second component is a binocular features fusion network, which is used to fuse and extract features from the feature vectors of the left and right eyes to generate the final grading result. Zang et al. [19] designed a transformer-guided category relation attention network (CRA-Net) to alleviate the class-imbalance problem for diabetic retinopathy grading.

In addition to using the attention mechanism to solve the problem of data imbalance, generative adversarial networks (GANs) [20], which are utilized for medical image compositing, can settle problems such as the lack of major and various annotation datasets, as well as issues with various data distribution imbalances. Various data synthesis methods have been proposed in different medical imaging fields, for instance, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and chest X-rays [21–24]. As mentioned above, the existing large-scale fundus image datasets primarily contain images of healthy fundi, leading to an imbalance in the data distribution

among categories and an overall insufficiency in data volume. These issues severely impact the accuracy of classification models. Additionally, the current neural networks often lack interpretability and cannot correspond to the standards used in clinical diagnoses. Therefore, to address the above issues, a conditional generative adversarial network with a self-attention mechanism and the name SACGAN is proposed to augment the number of diabetic retinopathy fundus images, thereby addressing the problems of insufficient and imbalanced data samples. Additionally, a diabetic retinopathy grading model named DRMCNet is developed to classify lesion severity using ResNeXt-50 as the baseline and integrating a channel attention mechanism within the multi-branch residual blocks to enhance the recognition of subtle lesion features and reduce the distractions of information that is not relevant to the theme in the original images. Consequently, both the generalization and robustness of the model are improved.

2. Models and Methods

Figure 1 presents the workflow designed in this study, which contains a pre-processing stage, fundus image generation, fundus image classification, and class activation map model interpretability. In the pre-processing stage, the original dataset is cleaned and enhanced to improve the image quality. To settle the issue of imbalanced training data, a conditional generative adversarial network that integrates a self-attention mechanism is designed to generate new samples and increase the samples' number, which is low. Then, for the fundus image classification, ResNeXt-50 combines a channel attention mechanism with multi-branch residual blocks to grade the severity of DR. Finally, Grad-CAM is used to prove the interpretability of the proposed model.

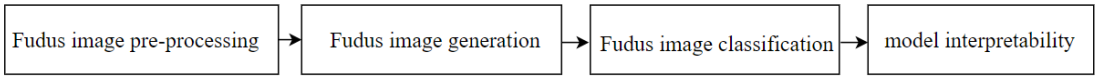


Figure 1. The workflow utilized in this study.

2.1. Images Dataset

In this study, the Eye-PACS dataset [25] was employed as a training set and validation set. The Eye-PACS dataset was provided by multiple hospitals, whose patients are Caucasian and other ethnicities. The Eye-PACS dataset consists of three-channel high-resolution color fundus images that were taken under diverse imaging conditions. The total image number is 35,126, divided into 5 different varieties according to the international grading standard for DR, including no obvious retinopathy, mild non-proliferative diabetic retinopathy (NPDR), moderate NPDR, severe NPDR, and proliferative DR (PDR), as shown in Table 1.

Table 1. Diabetic retinopathy disease category numbers in the Eye-PACS dataset.

Disease Category	Findings Observable upon Dilated Ophthalmoscopy	Numbers
No apparent retinopathy	No abnormalities	25,810
Mild NPDR	Microaneurysms only	2443
Moderate NPDR	Severer than just microaneurysms but milder than severe NPDR	5292
Severe NPDR	Any of the following and no signs of proliferative retinopathy: • More than 20 intraretinal hemorrhages in each of 4 quadrants • Definite venous beading in 2 or more quadrants • Conspicuous IRMA in 1 or more quadrants	873
PDR	One of the following two: • Neovascularization • Vitreous/preretinal hemorrhage	708

For further verification of the model's generalization performance, another dataset named OIA-DDR [26] was used as the test set. It was established by professional personnel extraction, ophthalmologist annotation, and computer expert modeling and evaluation using 1.6 million fundus images from 400 clinical hospitals in 26 provinces of China. The OIA-DDR dataset contains 13,673 fundus images, which is currently the largest publicly available fundus image dataset in China, and is divided into 5 categories, as shown in Figure 2.

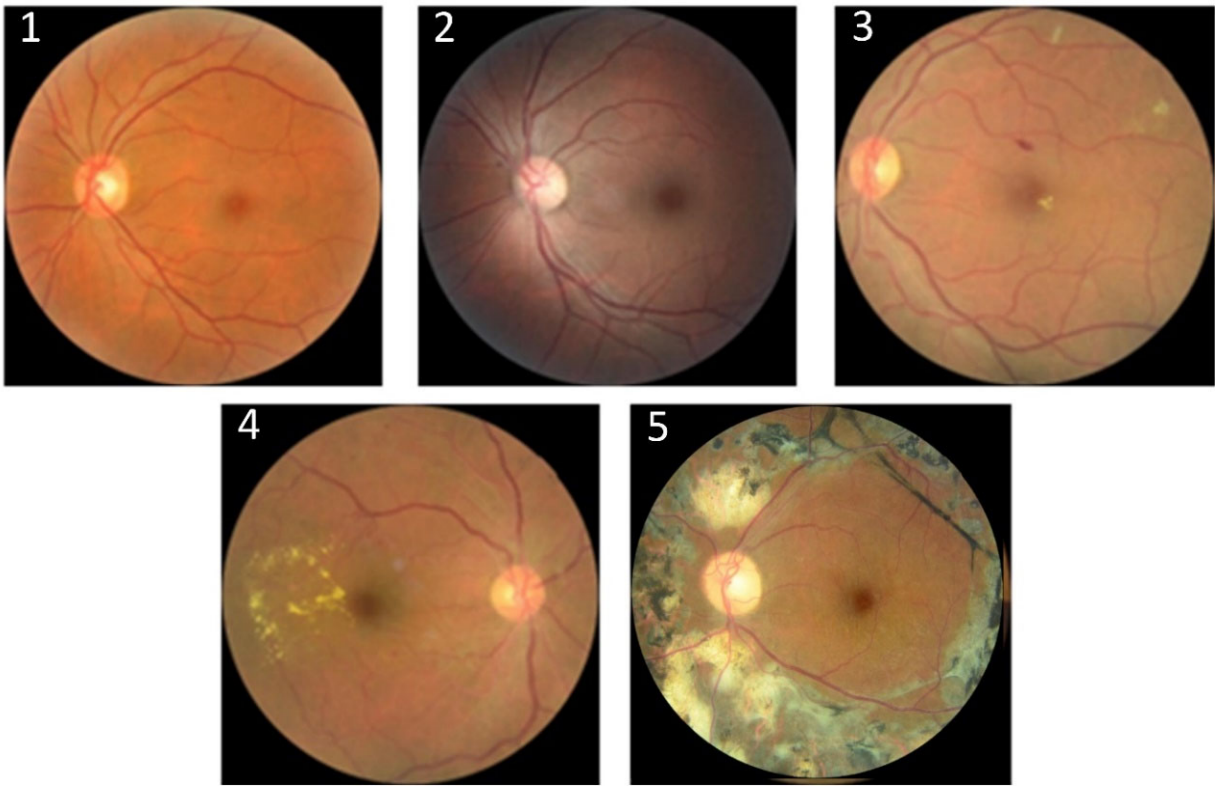


Figure 2. Images of DR: (**1**) no apparent retinopathy; (**2**) NPDR; (**3**) moderate NPDR; (**4**) severe NPDR; (**5**) PDR.

2.2. Pre-Processing

In the process of fundus diagnosis and treatment, doctors distinguish the type of diseases and judge how severe the disease is by observing clear fundus images. However, because of the influences of the capturing environment, imaging equipment, and the patient's physiological condition, some obtained images are degraded, such as with noise, blur, and distortion. Therefore, it was necessary to perform data cleaning and enhancement on the dataset to improve the image quality.

Firstly, some severely invalid or incomplete images caused by environmental factors were cleaned. Secondly, due to the various sizes of the original images in the dataset, the black borders around the fundus images were cropped, and all the image sizes were unified to 512 × 512. Next, the method of contrast finite adaptive histogram equalization was introduced to augment the image contrast and highlight their feature details. Lastly, uneven brightness and detail loss caused by different intensities of lighting in the same image were reduced by gamma correction. Figure 3 presents some typical cases after the preprocessing operation which clearly identify the fundus optic disc, macula, and major blood vessels.

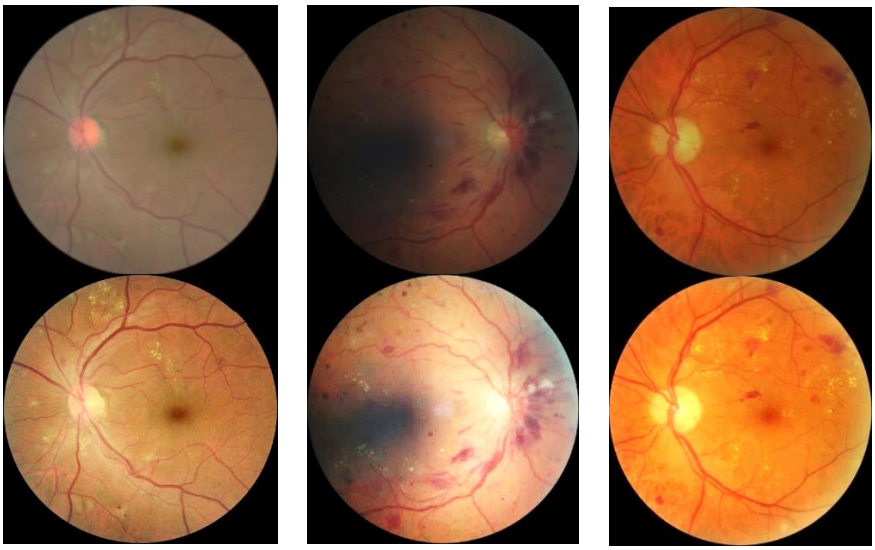


Figure 3. Typical cases after the preprocessing operation.

2.3. Fundus Image Generation

As shown in Table 1, there is serious imbalance in the Eye-PACS dataset, which would decrease the classifier's recognition accuracy, especially for the small sample categories. Therefore, in this study, a new conditional generative adversarial network (GAN) with a self-attention mechanism was designed for image generation to increase the number of images with few sample categories and to correct the problem of imbalanced dataset distribution.

2.3.1. GAN-Based Model

A generative adversarial network was proposed to generate data that do not exist in the real world. The main structure of a GAN includes a generator model (generator, G) and a discriminator model (discriminator, D), where the generator produces images via inputting noise signals stochastically, and the discriminator is responsible for determining whether the input image is a generated image or an authentic image. Both the generator and discriminator continuously iterate and optimize until the discriminator cannot distinguish the source of the input image. Therefore, the core of the network is the game between the generator and the discriminator.

The training of a GAN includes generator training and discriminator training. The training objective has two different portions, including to acquire the generator weight parameters that can deceive the discriminative network to the greatest extent possible, and the parameters that maximize the classification accuracy of the discriminative network. The objective function formula is as follows:

$$\min_{G}\max_{D} V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)] + E_{z\sim P_z(x)}[\log(1 - D(G(z)))] \quad (1)$$

where z is the random noise, x is the classifier's acceptance of the image, $P_z(z)$ represents the distribution of random noise generated by the network input, and P_{data} represents the mathematical distribution of the real image.

However, a traditional GAN cannot constrain the categories of generated images, resulting in unclear generation goals and poor controllability. A CGAN that can control the categories of network-generated images is proposed. The objective function formula is as follows:

$$\min_{G}\max_{D} V(D,G) = E_{x\sim P_{data}(x)}[\log D(x|y)] + E_{z\sim P_z(x)}[\log(1 - D(G(z|y)))] \quad (2)$$

where z is the random noise, x is the classifier's acceptance of the image, $P_z(z)$ represents the distribution of random noise generated by the network input, and P_{data} represents the mathematical distribution of the real image.

Compared to a traditional GAN, a CGAN embeds the category information at the input of the generator and discriminator, which can not only distinguish the authenticity of the generated images but can also judge the correctness of the generated image categories. The proposed conditional generative adversarial network with a self-attention mechanism, named SACGAN, is shown in Figure 4, which contains a generator and a discriminator. Both category labels and random noise are input into the generator, and the obtained virtual DR image and real DR image are used as inputs for the discriminator to distinguish the authenticity of images.

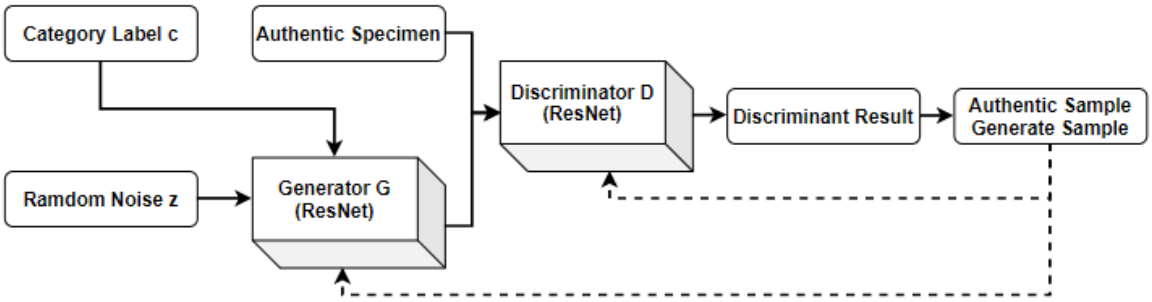


Figure 4. The architecture of the proposed SACGAN.

The generator of SACGAN consists of two convolutional layers, five upsampling residual modules, and a self-attention module, as displayed in Figure 5. First, the label information c and random noise z are concatenated as inputs of the generator. The concatenated result is subsequently processed by a convolutional layer, three upsampling residual modules, a self-attention module, two upsampling residual modules, and another convolutional layer. The upsampling residual module contains batch normalization, the LeakyReLU activation function, 2D convolution, identity mapping, and upsampling. The skip structure of residual modules can increase the network depth and alleviate the gradient vanishing while solving the instability problem of traditional CGAN training and accelerating the network convergence. The self-attention module is utilized to obtain global information and improve the quality of the generated feature maps. During the upsampling process, multiple upsampling and convolution operations are combined to replace the traditional deconvolution step, thereby reducing edge defects in the generated image and enhancing the generated image's quality.

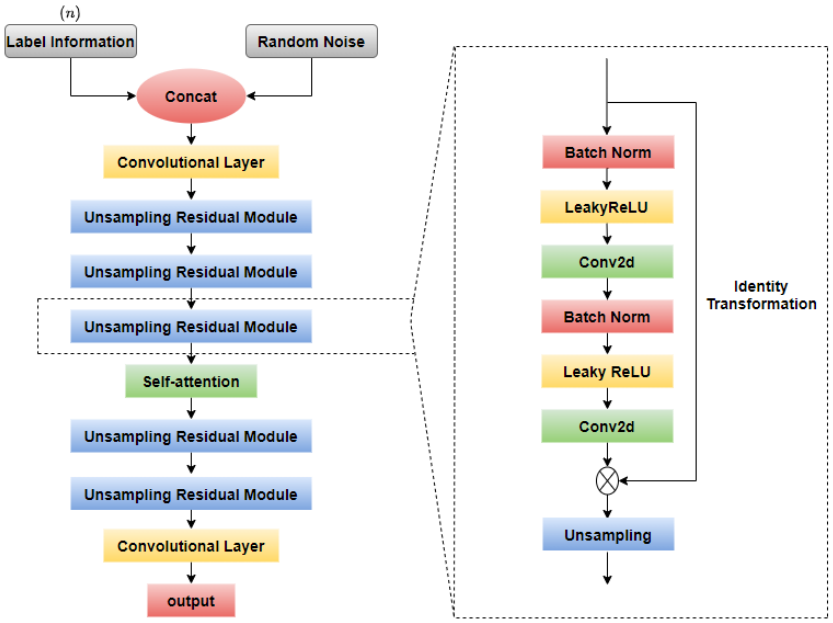


Figure 5. The generator architecture of the proposed SACGAN.

The discriminator of the SACGAN is made up of a convolutional layer, four residual modules, a self-attention module, adaptive average pooling, and a fully connected layer, as displayed in Figure 6. The input DR images are processed by a convolutional layer with the LeakyReLU activation function, and the obtained feature maps are fed into a residual module composed of 2D convolution, batch normalization, the LeakyReLU activation function, and identity mapping. The self-attention module is utilized between the four residual modules to obtain the global information of the feature maps. An adaptive average pooling layer is beneficial for cutting down the parameters of the ultimate fully connected layer. Lastly, a fully connected layer is utilized to determine the authenticity of the fundus image.

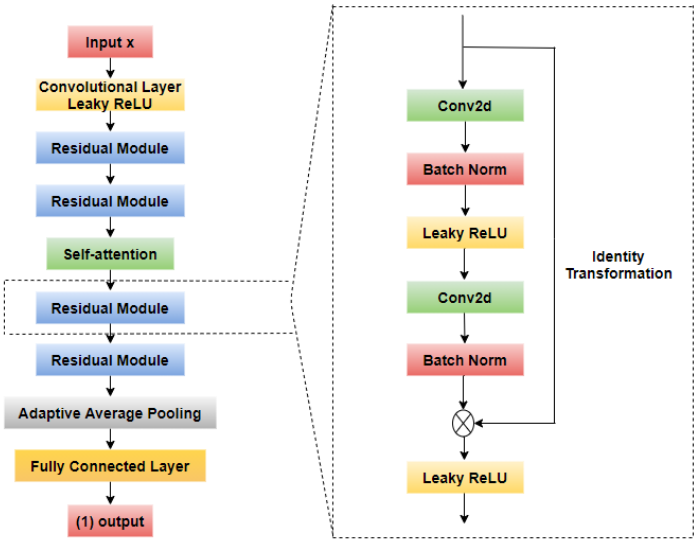


Figure 6. The discriminator architecture of the proposed SACGAN.

2.3.2. Self-Attention Module

In the generator and discriminator of the proposed SACGAN, a self-attention module is devised to capture the overall information by learning the relationship between a pixel and all other pixel positions. The self-attention module can compensate for the shortcoming of the residual module with the intrinsic locality of convolution operations and enhance the quality of the generated feature maps. The structure of the proposed self-attention module is presented in Figure 7.

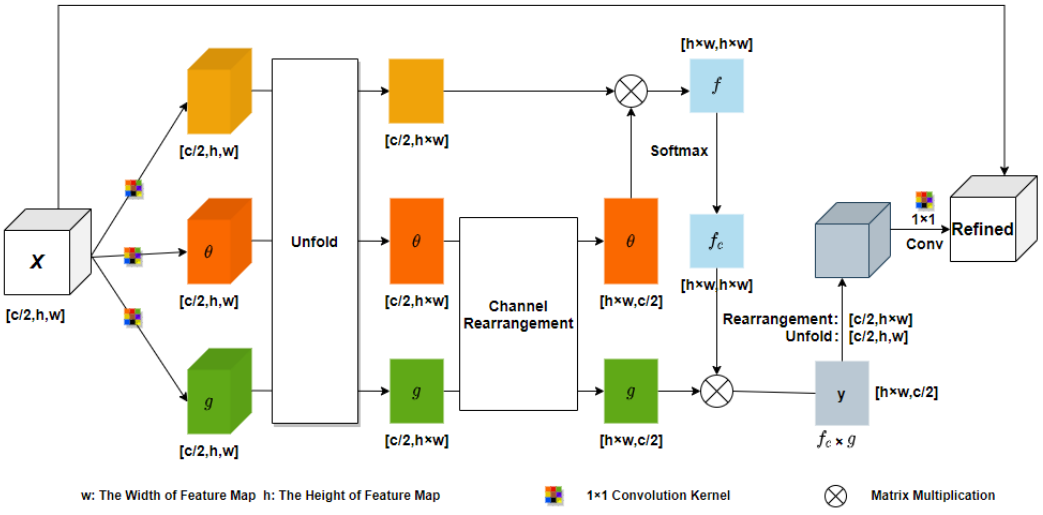


Figure 7. The structure of the proposed self-attention module.

2.4. Fundus Image Classification

2.4.1. CNN-Based Model

For the classification of DR, a diabetic retinopathy grading model with the name DRMCNet was developed, which uses ResNeXt-50 as the baseline and integrates a channel attention mechanism within the multi-branch residual blocks to enhance the recognition of subtle lesion features and eliminate the disturbance of the information not related to the research in the original images. The ResNeXt-50 structure utilizes the stacking network strategy of VGG and the split–transform–merge design concept of the Inception model, which improves the expressive ability of the network through parallel residual convolution blocks without significantly increasing the parameter complexity. This multi-branch residual learning strategy effectively copes with the common issues of gradient disappearance, gradient explosion, and network degradation in traditional deep neural networks and optimizes computing efficiency and parameter usage. The channel attention mechanism aims to improve the capacity of the model to study crucial features in fundus images. By embedding a channel attention module after the convolution module of ResNeXt, the weights of the feature maps extracted by the model are adjusted to prioritize the more informative features. Compared to traditional models, which require large-scale data and high computing resources, the proposed DRMCNet can reduce the dependence on big data by simplifying the learning path and optimizing convolution operations and can improve the usability and efficiency of actual medical image processing. The proposed DRMCNet structure is shown in Figure 8.

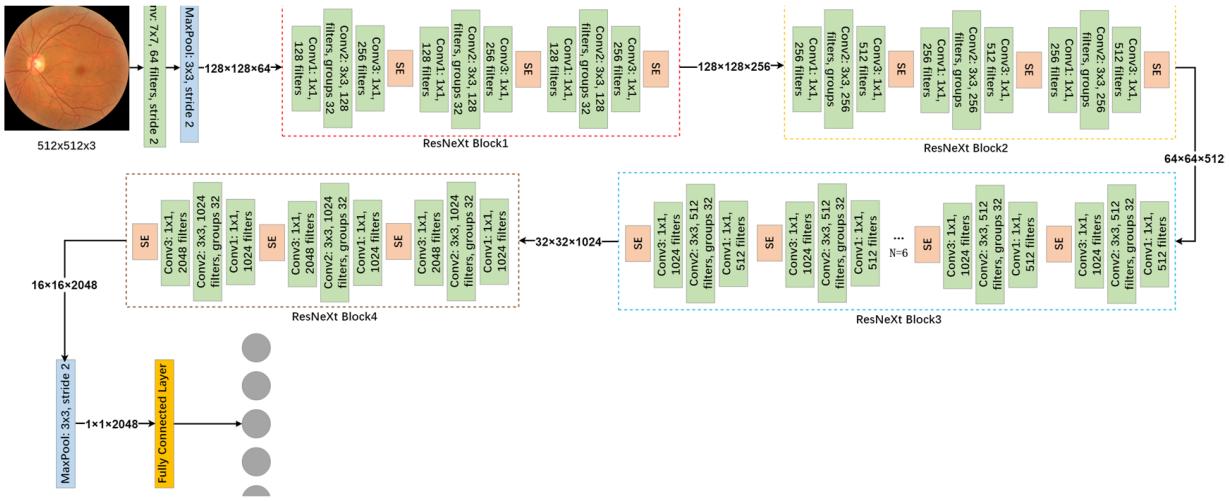


Figure 8. The proposed DRMCNet structure.

2.4.2. Channel Attention Mechanism

In the proposed DRMCNet, ResNeXt-50 is used as the main backbone and combines the channel attention mechanism to improve the recognition efficiency and accuracy of DR. As indicated in Figure 9, the channel attention module first compresses and reduces the spatial dimension of the feature map through overall average pooling and overall maximum pooling to gain two independent one-dimensional channel features. These two features are fed into a shared neural network comprised of two convolutional layers and, afterward, a ReLU activation function in between. The number of neurons in the first convolutional layer is set to C/r to decrease the number of parameters and alleviate overfitting, while the second convolutional layer is restored to the original number of channels C to ensure the integrity of the features. Through this structure, the network can generate weights for each channel and then output them through the Sigmoid function to obtain the ultimate channel attention weights. These weights are multiplied with the original feature map to highlight important features and suppress secondary information.

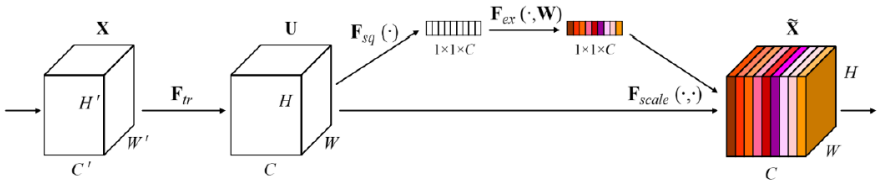


Figure 9. The channel attention mechanism.

2.4.3. Class Activation Map

To make the proposed classification model more transparent and interpretable, gradient-weighted class activation mapping (Grad-CAM) is introduced to generate 'visual explanations' for the decisions of the proposed DRMCNet. Figure 10 shows a flow chart of Grad-CAM, which calculates the predicted class score and the gradient of the last convolutional layer's feature map to determine the significance of each feature map to a specific class. The heat activation diagram produced by Grad-CAM can highlight the image regions that the DRMCNet focuses on during prediction and presents the visual explanations.

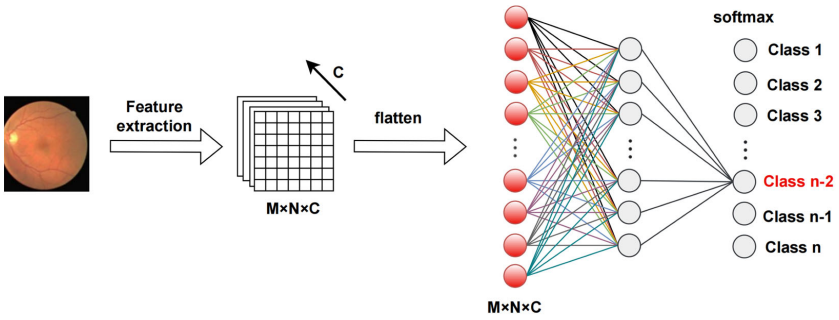


Figure 10. Flow chart of Grad−CAM.

3. Results

3.1. Hardware and Resources

The proposed approach was developed using a computing device equipped with an Intel(R) Core i7-10700X CPU processor, a single 8G graphics memory NVIDIA GeForce RTX 3070, and Python 3.7.0 and Pytorch 2.0 software in Windows 10 OS.

Adaptive moment estimation (Adam) was utilized to establish the optimization of the gradient descend technique in the SACGAN training period. The learning rate of both the generator and discriminator was 1×10^{-3}, and the size of batch was set to 32. The frequency of the training iteration epochs was 1000. During the training process of the DRMCNet, the stochastic gradient descent (SGD) optimizer was used to accelerate the model convergence. The initial value of the learning rate was 1×10^{-3}, and the batch size was set to 32. The frequency of iterations was set to 200.

3.2. Evaluation Indicators

The proposed approach was verified by three quantitative evaluations, listed in Formulas (3) to (5), denoting accuracy, sensitivity, and specificity, respectively. Accuracy is the percentage of true predications out of all predications, and sensitivity is the accurate prediction of actually positive samples. Specificity means the proportion of actually negative samples and samples judged to be negative.

$$Accuracy = \frac{TP + PN}{TP + TN + FP + FN} \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

where TP, FP, TN, and FN represent true positive, false positive, true negative, and false negative, respectively. Meanwhile, the confusion matrix is presented to ensure visualization of the classification of performance.

3.3. Fundus Image Generation Result

To solve the severe imbalance issue of the five categories of fundus image numbers, the SACGAN is proposed to generate new images to augment the number of small sample categories. Table 2 shows the original data distribution, the sample numbers after data pre-processing, and the image generation numbers. The final sample numbers of the five categories are approximately the same, which is beneficial for enhancing the accuracy of fundus image classification model training.

Table 2. Eye-PACS dataset distribution after fundus image generation.

Dataset	Health	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Eye-PACS	25,810	2443	5292	873	708
Data pre-processing	3821	2034	3733	658	643
Image generation	3821	3813	3733	3678	3528

Figure 11 shows the fundus image generation results of the original CGAN and the proposed SACGAN. Figure 11a–c illustrate the images generated in the 10th, 100th, and 1000th iterations by the original CGAN, respectively. And the image generated in 1000th iterations by the proposed SACGAN is shown in Figure 11d. Figure 11a is composed of 16 dark purple blocks that do not provide any fundus information, but according to Figure 11b, there is a round fundus outline and optic disc on the left side of the image. Although some structures, i.e., the optic disc, blood vessel, and macula lutea, are observed in Figure 11c, the borders of the different regions are relatively vague, and the key feature regions do not fully and clearly stand out. In contrast, the generated image with distinct borders and fundus structure in Figure 11d is clearer than that in Figure 11c, which meets the demand of the classification model. The generated image by the proposed SACGAN not only contains the basic structures, like the optic disc and blood vessels, but also highlights lesions, such as exudates and bleeding points, which can effectively supplement the original dataset to settle the imbalanced distribution issue.

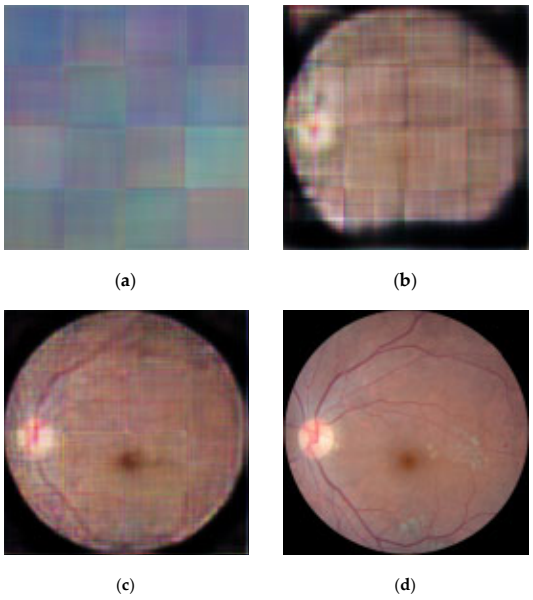


Figure 11. Image generation processes using different methods: (**a**) 10 iterations by CGAN; (**b**) 100 iterations by CGAN; (**c**) 1000 iterations by CGAN; (**d**) 1000 iterations by SACGAN.

3.4. Fundus Image Classification Result

To validate the effectiveness and generalization of the DRMCNet, 350 fundus images from the OIA-DDR dataset were randomly selected as the test data. Several relevant classification networks, i.e., DenseNet-121, Inception-V3, RestNet-50, and RepVGG were compared with the proposed DRMCNet. By investigating the comparison results given in Table 3 below, it is evident that the accuracy, sensitivity, and specificity of the proposed approach on the OIA-DDR dataset is the highest and obviously improves the classification performance. The proposed DRMCNet utilizes ResNeXt-50 as the baseline, which integrates a channel attention mechanism and multi-branch residual blocks to improve the classification performance. Compared to the ResNet, the residual learning part of ResNeXt uses multi-branch polymerization residual blocks to replace the traditional single-path convolution, which enables the network to learn more features and increase its feature representation ability. The ResNeXt mainly utilizes not only the stacking network strategy of VGG but also the strategy of split–transform–merge of Inception, integrating the advantages of each convolutional neural network model and thereby improving the performance of the model. Therefore, the proposed classification network DRMCNet is a great improvement, with an accuracy of more than 92%.

Table 3. Comparison of various approaches.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
DenseNet-121	83.4	82.3	82.3
Inception-V3	87.3	89.4	87.2
ResNet-50	89.5	90.7	85.6
RepVGG	90.1	91.2	91.2
Proposed method	92.3	92.5	92.5

To validate the effectiveness of the channel attention module in the proposed network, the ablation experiment is displayed in Table 4. After the channel attention module was added to ResNeXt-50, the accuracy, sensitivity and precision were all optimized, indicating that the channel attention module was able to enhance the classification by utilizing the channel information efficiently.

Table 4. Ablation experiment using the proposed network.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
ResNeXt-50	89.9	90.8.	88.9
Proposed method	92.3	92.5	92.5

The confusion matrix on OIA-DDR is given in Figure 12 to make the classification performance of the proposed approach visible. From the values on the left diagonal, it is evident that the most classification results of the proposed model are correct. Only a few over-diagnoses are distributed in the lower left of the diagonal, and a few misdiagnoses are distributed in the upper right corner of the diagonal. The results of the confusion matrix prove the superiority of the proposed model as well.

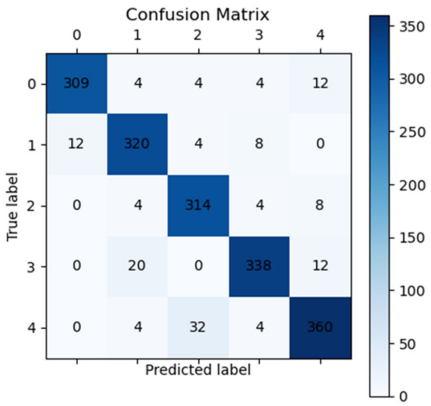


Figure 12. Confusion matrix of the classification result.

3.5. Additional Generalization Experiment with the Clinical Data

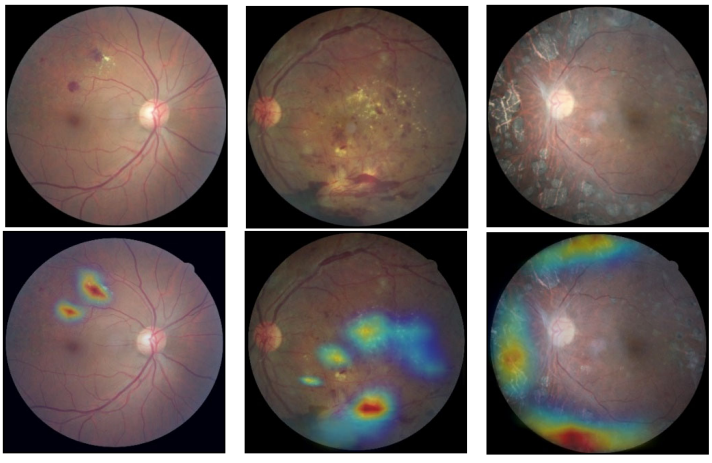
To further demonstrate the practicality and generalization of the proposed method, a private fundus dataset collected from Shidong Hospital of Shanghai was used for validation. This study was approved by the Ethics Committee of the Center (approval number: IRB-AF63-V1.0). The dataset comprised 688 fundus images, which were manually annotated by two doctors and verified by experienced ophthalmologists. In this experiment, the private dataset was only used for testing. The proposed method was compared with DenseNet-121, Inception-V3, RestNet-50, and RepVGG. Table 5 provides the quantitative results of the accuracy, sensitivity, and precision using different methods. The proposed method achieved the best numerical measurements compared to the other methods. This experiment with the clinical data demonstrates that the proposed method has effectiveness and a generalization ability, which can help doctors enhance their work efficiency. This study not only verifies the possibility of using advanced deep learning techniques to effectively diagnose diabetic retinopathy but also emphasizes the importance of continuous technological innovation in the field of medical image analysis.

Table 5. Quantitative results of the different methods on a private clinical dataset.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
DenseNet-121	81.7	80.6	82.4
Inception-V3	87.0	90.2	84.3
ResNet-50	87.6	87.4	84.6
RepVGG	89.9	86.1	90.8
Proposed method	90.6	91.5	93.2

3.6. Classification Based on Visualization

To conduct an interpretability analysis of the proposed model, the Grad-CAM was calculated to produce an interpretable heat activation diagram of the classification network. Figure 13 indicates the mapping results of the proposed model for proliferative diabetic retinopathy. The highlighted red area indicates that the features in this part of the feature map are an important basis for network classification, which is the attention region of the classification network. The color that gradually changes from yellow to blue indicates a gradual decrease in the importance of the local region in the classification results. It can also be noticed in the activation heat map in Figure 13 that the proposed model accurately highlighted and marked the focus of diabetic retina, such as bleeding spots, hard exudates, proliferative lesions, etc., which reflects the model's ability to identify the focus and further explains the rationality of the model classification. Such an approach has the capacity to meet the clinical diagnosis and treatment requirements and assist doctors in achieving rapid diagnosis.



(**a**) (**b**) (**c**)

Figure 13. Class activation map of the proliferative diabetic retinopathy. (**a**) Hard exudates and bleeding points; (**b**) microaneurysm; (**c**) proliferative lesions.

4. Discussions

With the development of deep learning technology in the medical field, assisted clinical diagnosis by using deep learning technology has gradually been accepted by clinical staff. To address the problems of data imbalance, small lesion features, and poor model interpretability in the auxiliary diagnosis of diabetic retinopathy, this study proposes a new corresponding solution that includes a pre-processing stage, fundus image generation, fundus image classification, and class activation map model interpretability. Although the proposed mothed achieves excellent performance in handling DR classification problems, there is still much work to be carried out in this field. First, by using the proposed SACGAN to generate categories that lack samples, which can alleviate the imbalanced DR data distribution problem to a certain extent, the possible bias caused by the generated data is still a focus of future research. In the process of expanding data, it is still necessary to reasonably control the ratio of generated images to original images. Second, improving the robustness of the model on diverse data is also an important direction of future research. By collaborating with more clinical practices, more diverse fundus image data could be collected to verify and improve the practicality and generalization ability of the proposed model. Third, the diagnosis and classification of DR in combination with specific lesions of diabetic retinopathy, such as microaneurysms, cotton wool spots, exudates, and hemorrhages, will be considered. A classification model that can automatically classify DR with reference to lesions could be designed to further improve the accuracy of DR diagnosis and treatment, and increase the innovation and practicality of research work.

5. Conclusions

In this paper, a new approach founded on a convolutional neural network is proposed to grade diabetic retinopathy. First, data cleaning and enhancement were conducted to enhance the quality of image and decrease the possibility of unnecessary interference. Second, a brand new conditional generative adversarial network with a self-attention mechanism was designed to increase the number of categories with fewer sample sizes, which addresses the sample imbalance problem. Next, an improved ResNeXt-50 combined with a channel attention mechanism and a multi-branch convolutional residual module was proposed to classify the diabetic retinopathy. Finally, Grad-CAM was utilized to prove the interpretability. The proposed method was further validated on the public OIA-DDR dataset, and the experimental results verify the advantages of the proposed method, indicating that the proposed approach is superior to other relevant approaches in diabetic retinopathy grading. Though image generation by the proposed SACGAN can balance the sample number of different category, the generated data can bring some false information, which effects the classification accuracy. Thus, in future work, associated researchers should collect more clinical data to expand the datasets and balance the numbers of data to train and verify the classification model, which would result in more accurate grading results.

Author Contributions: Conceptualization, P.Z.; Methodology, P.Z.; Software, X.L. (Xiao Liu); Formal analysis, X.L. (Xinyu Li); Investigation, X.L. (Xinyu Li); Resources, J.Z.; Writing–original draft, X.L. (Xiao Liu); Writing–review & editing, Q.L. and W.L.; Visualization, Y.G.; Project administration, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (grant No. 61801288, 12302417) and the Shanghai Pujiang Program (23PJ1409200).

Data Availability Statement: Data is unavailable due to privacy or ethical restrictions.

Acknowledgments: The authors would like to thank the anonymous reviewers and related editors for their constructive comments and suggestions that enhanced not only the technical content but also the presentation quality of this paper.

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

References

1. Roglic, G. Who global report on diabetes: A summary. *Int. J. Noncommun. Dis.* **2016**, *1*, 3–8. [CrossRef]
2. Teo, Z.L.; Tham, Y.C.; Yu, M.; Chee, M.L.; Rim, T.H.; Cheung, N.; Bikbov, M.M.; Wang, Y.X.; Tang, Y.; Lu, Y.; et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis. *Ophthalmology* **2021**, *128*, 1580–1591. [CrossRef] [PubMed]
3. Islam, S.M.S.; Hasan, M.M.; Abdullah, S. Deep learning based early detectionand grading of diabetic retinopathy using retinal fundus images. *arXiv* **2018**, arXiv:1812.10595.
4. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **2016**, *316*, 2402–2410. [CrossRef] [PubMed]
5. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
6. Alam, M.; Zhao, E.J.; Lam, C.K.; Rubin, D.L. Segmentation-assisted fully convolutional neural network enhances deep learning performance to identify proliferative diabetic retinopathy. *J. Clin. Med.* **2023**, *12*, 385. [CrossRef] [PubMed]
7. Theckedath, D.; Sedamkar, R.R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput. Sci.* **2020**, *1*, 79. [CrossRef]
8. Huang, Y.; Lin, L.; Cheng, P.; Lyu, J.; Tam, R.; Tang, X. Identifying the key components in resnet-50 for diabetic retinopathy grading from fundus images: A systematic investigation. *Diagnostics* **2023**, *13*, 1664. [CrossRef] [PubMed]
9. Costa, P.; Araujo, T.; Aresta, G.; Galdran, A.; Mendonça, A.M.; Smailagic, A.; Campilho, A. EyeWeS: Weakly Supervised Pre-Trained Convolutional Neural Networks for Diabetic Retinopathy Detection. In Proceedings of the 2019 16th International Conference on Machine Vision Applications (MVA), Tokyo, Japan, 27–31 May 2019.
10. Wang, X.; Xu, M.; Zhang, J.; Jiang, L.; Li, L.; He, M.; Wang, N.; Liu, H.; Wang, Z. Joint learning of multi-level tasks for diabetic retinopathy grading on low-resolution fundus images. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 2216–2227. [CrossRef] [PubMed]
11. Rocha, D.A.D.; Ferreira, F.M.F.; Peixoto, Z. Diabetic retinopathy classification using VGG16 neural network. *Res. Biomed. Eng.* **2022**, *38*, 761–772. [CrossRef]
12. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
13. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
14. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
15. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 13713–13722.
16. He, A.; Li, T.; Li, N.; Wang, K.; Fu, H. CABNet: Category Attention Block for Imbalanced Diabetic Retinopathy Grading. *IEEE Trans. Med. Imaging* **2020**, *40*, 143–153. [CrossRef] [PubMed]
17. Bhati, A.; Gour, N.; Khanna, P.; Ojha, A. Discriminative kernel convolution network for multi-label ophthalmic disease detection on imbalanced fundus image dataset. *Comput. Biol. Med.* **2023**, *153*, 106519. [CrossRef] [PubMed]
18. Hai, Z.; Zou, B.; Xiao, X.; Peng, Q.; Yan, J.; Zhang, W.; Yue, K. A novel approach for intelligent diagnosis and grading of diabetic retinopathy. *Comput. Biol. Med.* **2024**, *172*, 108246. [CrossRef] [PubMed]
19. Zang, F.; Ma, H. CRA-Net: Transformer guided category-relation attention network for diabetic retinopathy grading. *Comput. Biol. Med.* **2024**, *170*, 107993. [CrossRef] [PubMed]
20. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1–9.
21. Kazeminia, S.; Baur, C.; Kuijper, A.; van Ginneken, B.; Navab, N.; Albarqouni, S.; Mukhopadhyay, A. GANs for medical image analysis. *Artif. Intell. Med.* **2020**, *109*, 101938. [CrossRef] [PubMed]
22. Pesaranghader, A.; Wang, Y.; Havaei, M. CT-SGAN: Computed tomography synthesis GAN. In *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections, Proceedings of the First Workshop, DGM4MICCAI 2021, and First Workshop, DALI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, 1 October 2021*; Proceedings 1; Springer International Publishing: Cham, Switzerland, 2021; pp. 67–79.
23. Zhan, B.; Zhou, L.; Li, Z.; Wu, X.; Pu, Y.; Zhou, J.; Wang, Y.; Shen, D. D2FE-GAN: Decoupled dual feature extraction based GAN for MRI image synthesis. *Knowl.-Based Syst.* **2022**, *252*, 109362. [CrossRef]
24. Xue, Y.; Bi, L.; Peng, Y.; Fulham, M.; Feng, D.D.; Kim, J. PET Synthesis via Self-supervised Adaptive Residual Estimation Generative Adversarial Network. *IEEE Trans. Radiat. Plasma Med. Sci.* **2023**, *8*, 426–438. [CrossRef]

25. Cuadros, J.; Sim, I. EyePACS: An open source clinical communication system for eye care. In *MEDINFO 2004*; IOS Press: Amsterdam, The Netherlands, 2004; pp. 207–211.
26. Xia, X.; Zhan, K.; Fang, Y.; Jiang, W.; Shen, F. Lesion-aware network for diabetic retinopathy diagnosis. *Int. J. Imaging Syst. Technol.* **2023**, *33*, 1914–1928. [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.