

Improved Model Based on GoogLeNet and Residual Neural Network ResNet

Xuehua Huang, Hunan City University, China*

ABSTRACT

To improve the accuracy of image classification, a kind of improved model is proposed. The shortcut is added to GoogLeNet inception v1 and several other ways of shortcut are given, and they are GRSN1_2, GRSN1_3, GRSN1_4. Among them, the information of the input layer is directly output to each subsequent layer in the form of shortcut. The new improved model has the advantages of multi-size and small convolution kernel in the same layer in the network and the advantages of shortcut to reduce information loss. Meanwhile, as the number of inception blocks increases, the number of channels is increased to deepen the extraction of information. The GRSN, GRSN1_2, GRSN1_3, GRSN1_4, GoogLeNet, and ResNet models were compared on cifar10, cifar100, and mnist datasets. The experimental results show that the proposed model has 3.07% improved to ResNet on data set cifar10, 2.08% on data set cifar100, 17.69% improved to GoogLeNet on data set cifar10, 28.47% on data set cifar100.

KEYWORDS

Deep Neural Network, GoogleNet Inception, Residual Neural Network, GRSN1, Recognition

INTRODUCTION

Deep learning is a research hot spot at present, and its research results have been widely used in image recognition, speech recognition, and other fields. This paper concerns the study of the application of deep neural network algorithm in image classification. An improved neural network model based on GoogLeNet and residual neural network ResNet is proposed. First, the mainstream algorithms of deep neural network in image recognition is introduced in this paper and the analysis of the advantages and disadvantages of GoogLeNet series algorithm and residual neural network algorithm. Based on the advantages of the two kinds of algorithms, an improved model GRSN is proposed by introducing the shortcut into the GoogLeNet inception block algorithm. Secondly, the detailed network structure and parameter setting of GRSN is shown. In addition, the experimental analysis is carried out. Then, the over-fitting situation is optimized. Finally, the conclusion and prospect are given.

The BP neural network proposed by Rumelhart and McClelland in 1986 is a fully connected network, but this network cannot provide great performance or high accuracy in image recognition. To address this problem, the CNN (Convolution Neural Network) was proposed, which imitates the higher sensitivity to main features of human eyes and adopts a shared weight structure to significantly reduce the weight parameters and improve the performance and accuracy of image recognition. In

DOI: 10.4018/IJICINI.313442

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

1998, LeCun proposed the LeNet, which is a typical CNN structure, and it consists of two convolution layers, two pooling layers, and two fully-connected layers. When this network was applied to the handwritten digit recognition of people in the United States, the error rate was lower than 1%. However, its recognition performances on Chinese characters and more complex images need to be enhanced. In 2012, the AlexNet developed by Hinton and his student Alex Krizhevsky won the International Conference on Computer Vision (ICCV) contest. Their algorithm adds more layers and adopts a parallel structure to run on GPU, so its image recognition performance is significantly improved. VGG uses small convolution kernels to increase the numbers of layers and channels of the network (Simonyan, 2014). The GoogLeNet algorithm won the championship in the same year. In this algorithm, the inception structure block is introduced, and the convolution kernels of different sizes are used in the same layer of the network to extract a feature of different sizes and improve the perception of the model. By introducing the 1×1 convolution kernels, dimension reduction can also be achieved. There are four versions of GoogLeNet: v1, v2, v3 and v4, and the main difference between these versions is in the inception module.

The recognition accuracy of the algorithm can be improved by increasing the number of network layers. LeNet has 5 layers, AlexNet has 8 layers, VGG has 16 layers and 19 layers structure, and inception v1 has 22 layers. However, He Kaiming, the designer of ResNet, conducted an experiment on the cifar10 data set and found that the image recognition error rate of a 56-layer convolution network was higher than that of a 20-layer convolution network (He et al., 2016). Therefore, it will lead to the degradation in the performance of the neural network model by simply increasing the network layers, causing the later features to lose the original information of previous features. To address this problem, he proposed adding a bypass between the two blocks. In addition to the ResNet algorithm, the residual neural network algorithms also include DenseNet and DarkNet. The main differences between these algorithms are in terms of shortcut and residual block. In the deep residual network ResNet, a shortcut is added between every two blocks, while in the deep residual network DenseNet, a shortcut is added to each subsequent block. Therefore, the accuracy of DenseNet is further improved.

The performance of a residual neural network (RNN) can be generally improved from three aspects: residual unit, residual connection (Zhou, 2022) and the overall network structure (Guo, 2020). For residual unit, investigation and improvement are mainly carried out in terms of the convolution layer, BN normalization layer, activation function, shortcut, and the overall structure of the residual unit; the residual connection is mainly improved upon from the perspectives of the add-type residual connection, Concat-type residual connection and topology. The overall structure of the residual neural network is improved upon from various aspects, including the parallel residual network, multi-scale residual network, random deep network, multi-path residual network, and wide residual network. At the same time, RNN has also been combined with the Inception structure and attention mechanism, as Chen et al.(2018), Fei et al.(2019), Zong et al. (2020), Jie et al.(2017), Woo et al.(2018), Guan et al. (2020), Wang et al.(2020), Cheng et al. (2022), and Qiao et al.(2021) discussed in their research and applied in various fields. Chen et al. (2018) proposed the DRINet for segmentation of medical images, and they used a residual Inception module to aggregate feature maps from different branches, which improved the segmentation performance of their model. Fei et al. (2019) etc. proposed a residual Inception encoder-decoder network for medical imaging synthesis, and in this model, the residual Inception block consists of a path with two 3×3 convolution layers and an Inception shortcut with 1×1 convolution layer, which has solved the problem that the input and output feature maps have different channels. Zong et al. (2020) proposed hard exudate segmentation of images. This network uses U-Net as the backbone, and each unit in the shrinking and expanding paths is composed of residual units with Inception modules, which can not only reduce the number of parameters, but also improve the representation ability of the network by introducing more linear mappings.

The improvement of the GoogLeNet algorithm mainly includes improvement of the network structure and optimization of the network parameters, such as improvement of the structure of the inception block and improvement of the convolution size and the corresponding parameters. Zhang

and Yu (2021) employed the improved GoogLeNet network structure for recognition of table tennis motions. Moreover, the GoogLeNet algorithm can also be combined with other algorithms to improve certain application performances. For example, in the paper *Image Registration and Fusion of High Sensitive Ultraviolet Image Based on GoogLeNet Model and WT-Canny* written by Hou Sizu and Liu, Y. T. (2021), the method has not only incorporated the idea of GoogLeNet network model, but also introduced methods such as transfer learning and wavelet transform to improve the image recognition performance. Han et al. (2019) proposed an improved GoogLeNet network structure, and also introduced the idea of transfer learning to improve the recognition performance for remote sensing images. Furthermore, to address the redundant parameters and large computational overhead of GoogLeNet algorithm, Peng et al. (2019) introduced the pruning algorithm to simplify the network structure and reduce the number of parameters.

Due to their respective advantages, the GoogLeNet algorithm and ResNet algorithm can be combined. At present, there are not many improved algorithms combining these two algorithms. Among such improved algorithms, the most famous are the Inception-Resnet-v1 network and the Inception-Resnet-v2 network. The Inception-Resnet-v1 network model integrates the residual idea into Inception-v3, while the Inception-Resnet-v2 network model incorporates the residual idea on the basis of Inception-v4. Wang et al. (2022) proposed introducing the attention mechanism based on the Inception-ResNet-v2 algorithm to learn the importance of different channel features. Moreover, Wu et al. (2020) proposed combining Inception with the ResNet network structure and incorporating the Stem, SE-Block, and Reduction modules to improve the classification performance of lung CT images. Wang (2020) proposed incorporating the basic structure of CNN to the residual network, while Liu and Huang, L. C. (2020) proposed combining the two networks of GoogLeNet and ResNet to improve the accuracy. Habibzadeh et al. (2018) proposed using a pre-trained Inception-ResNet for the classification of four major white blood cells, and Inception-V1-ResNet achieved an accuracy of 99.84%. Szegedy et al. (2017) incorporated a residual module into the Inception structure and found that the residual module can increase the speed of training Inception network but cannot significantly improve the accuracy. Zhuang et al. (2017) integrated Google Inception v3 with different Long Short Term Memory (LSTM) networks to construct ConvDLSTM; the model was uniformly trained on a crowd image set, and the temporal and spatial information was unified in a neural network structure, which can effectively detect the crowd information. Zhang et al. (2019) proposed the fusion of an improved Inception module and ResNet.

The fusion of network structures is not simply combining different networks, but also involves improving the network structure and parameter settings. Based on the above studies, an improved model GRSN is proposed in this work.

THE IMPROVED MODEL GRSN

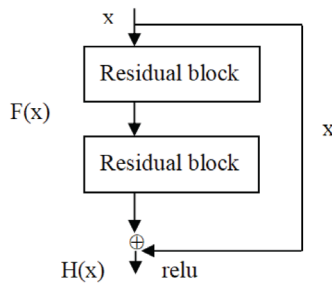
The Theory of Improved Model GRSN

Combining the advantages and disadvantages of the GoogLeNet and residual neural network algorithms, an improved network structure GRSN based on GoogLeNet and residual neural network is proposed.

The core block of ResNet is shown in Figure 1. The output result $H(x)$ is composed of the upper stacked convolution output $F(x)$, and the output of the direct upper layer X . It effectively alleviates the model degradation caused by the increase of neural network model layers and makes the neural network develop to a deeper level.

In view of the loss of information with the increase of the number of network layers, a shortcut to reduce the loss of information is proposed, so that the information of the upper layer can directly reach the next layer without matrix multiplication and nonlinear transformation. It is possible to increase the number of layers, thus improving the accuracy of image recognition. The problem of gradient extinction or gradient explosion is largely avoided as the number of network layers increases

Figure 1. The core block of ResNet



in Residual networks, which makes it possible to train extremely deep networks. Although the layers of residual neural network are deepened, the width is narrow, and multi-scale feature extraction is worse than inception block.

The disadvantage of GoogLeNet algorithm is that the width of network structure is wide, but the depth is insufficient, and the efficiency of parameter calculation is low. The inception block structure of GoogLeNet v1 algorithm is shown in Figure 2. Convolution kernels of different sizes are used in the same layer network to extract features of different sizes. However, with the increase of the number of layers, the information is lost in deep neural network.

Combined with the advantages of multi-size and small convolution kernels in GoogLeNet algorithm and information loss reduced in the residual network ResNet algorithm, an improved deep neural network structure GRSN is proposed. That is to say, the shortcut is added on the basis of inception block, which not only retains the advantages of GoogLeNet algorithm, but also has the advantages of residual neural network algorithm. Since convolution kernels of different sizes are used in the same layer, features of different sizes are extracted, which improves the perception of the model and the accuracy of image recognition. On the premise of the same performance, small convolution is adopted to reduce the weight parameters, the computation, and the network complexity. At the same time, the inter layer residual shortcut is increased to reduce the loss of information, making it possible to increase the number of layers, thus further improving the accuracy of image recognition.

Compared with GoogLeNet and ResNet in accuracy, the improved model GRSN has better performance and accuracy than GoogLeNet and ResNet.

Structure of Improved Model GRSN

In the improved model GRSN, the Residual block is replaced by the inception module—that is, in Figure 1 the Residual block is replaced by the inception block in Figure 2, as shown in Figure 3.

Figure 2. The core block of GoogLeNet Inception v1

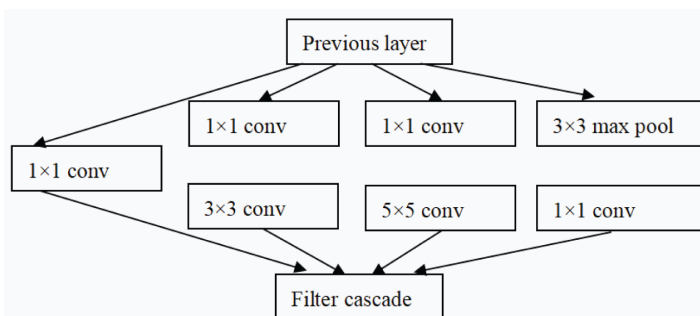
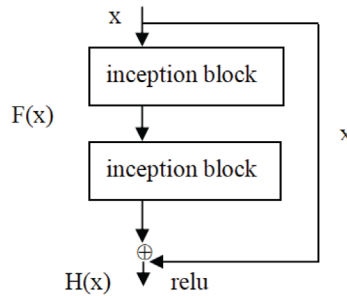


Figure 3. The core block of GRSN



A shortcut is added into the inception block, and the information from the previous layer is added to the inception block via the shortcut. It needs to be discussed how many inception blocks are added to a shortcut from the previous layer and to which layer the shortcut outputs to. In fact, there is a 1×1 convolution kernel in each inception block, as shown in Figure 2, which means that input information of the previous block is added to each inception block to reduce the loss.

In this paper, several simple models are studied and experiments carried out, as shown in Figures 4–7. In the model in Figure 4, shortcut from the input layer is added to the inception block in each layer (Huang et al., 2021). In Figure 5, shortcut from the input layer is added to the inception block for every two layers. In Figure 6, shortcut from the input layer is added to the inception block for every three layers. And in Figure 7, shortcut from the input layer is added to the inception block for every four layers. More inception blocks can be added to the models in Figures 4, 5, 6, and 7, but we skip more blocks here. In these figures, x represents the image input, which can also be an intermediate layer. In GoogLeNet, there are four versions of inception block (v1, v2, v3, and v4), and the v1 inception block is used in the improved model of GRSN. The model improved according

Figure 4. Structure 1 of improved model GRSN

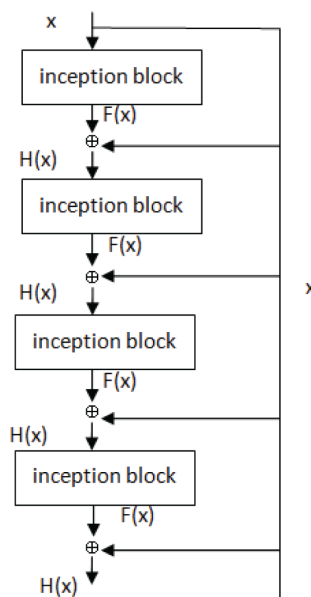


Figure 5. Structure 2 of improved model

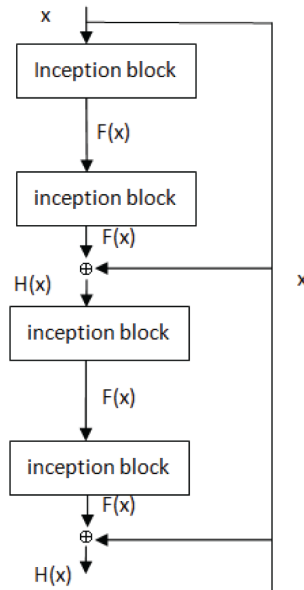
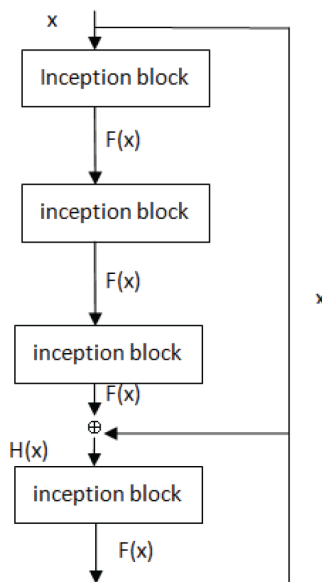


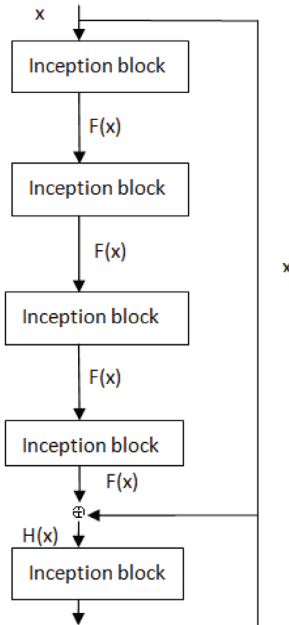
Figure 6. Structure 3 of improved model



to the structure in Figure 4 is GoogLeNet-v1-Residual Network, abbreviated as GRSN. The models improved according to the structures of Figures 5, 6, and 7 are GRSN1_2, GRSN1_3 and GRSN1_4, respectively. Our work can also be applied to v2, v3 and v4 of GoogLeNet Inception.

After being processed by the inception block, the size of the input image may change, so the sizes may not match by superimposing the input and output images. When they do not match, the image

Figure 7. Structure 4 of improved model GRSN



size can be adjusted via the 1×1 convolution kernel, so the structure in Figure 4 can be modified to the structure in Figure 8. Similarly, in Figures 5, 6, and 7, the sizes might not match after superimposing the input and output images, and in this case, the image sizes can also be modified using the 1×1 convolution kernel, as shown in Figures 9, 10, and 11, respectively.

After improvement of the GoogLeNet model by adding shortcut, the subsequent experiments prove that the performances of models in Figures 8–11 are all improved, but the improvement of the model in Figure 8 is the most significant.

Specific Network Structure of GRSN

The batch normalization (BN) can accelerate the convergence of data and alleviate over-fitting, which is generally performed after the convolution layer and before the activation layer. The dropout operation is mainly to alleviate over-fitting. During the training of the neural network, some neurons are temporarily discarded from the neural network according to a certain probability. When the neural network is used, the discarded neurons are reconnected. Pooling operation is used to reduce the amount of feature data in CNN.

On the basis of Figures 4 and 8, the specific structure of GRSN is presented in Figure 12 in further details (assuming the input size is $32 \times 32 \times 3$). The network structure of GRSN mainly includes the following parts:

1. The inception block structure of GoogLeNet is used in our model. The inception block obtains features of different sizes using multi-dimensional convolutions, which can enhance the performance. Sixteen convolution kernels of sizes 1×1 , 3×3 and 5×5 are used. By using small convolution kernels, the same size of the receptive field as the large convolution kernel can be obtained, but it requires less parameters, thus improving the operating efficiency. The inception block has a step size of 1, and adopts zero padding.

Figure 8. Structure 1 of improved model GRSN

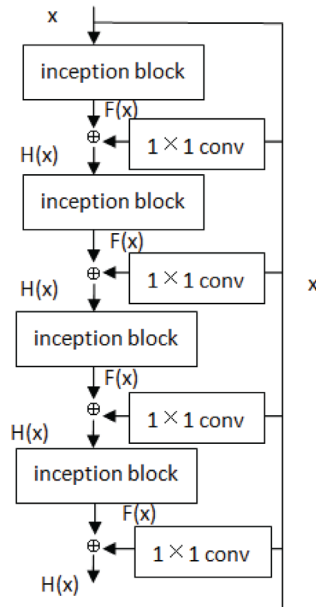
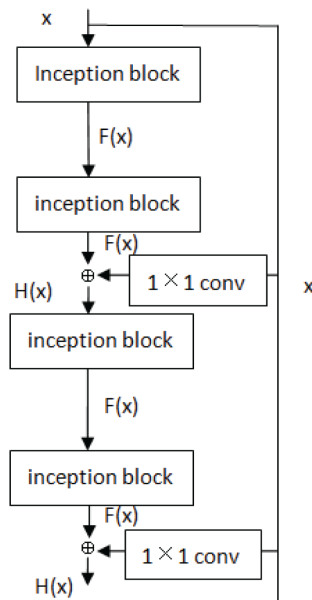


Figure 9. Structure 2 of improved model GRSN



2. The wider and the deeper the model is, the more network model parameters there will be. Limited data will be provided for training, which tends to cause over-fitting. Therefore, after each inception block, by adding the ReLU activation function, the BN (batch normalization) operation and the Dropout function, it can reduce over-fitting, and improve accuracy and efficiency.

Figure 10. Structure 3 of improved model GRSN

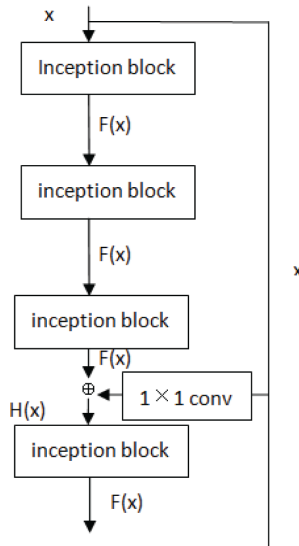
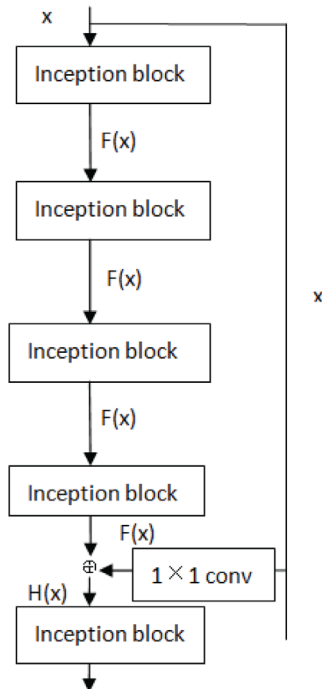


Figure 11. Structure 4 of improved model GRSN



3. By incorporating the residual idea of ResNet to our network structure, the information from the input layer can be directly added to each inception block. The original input information is directly transmitted to the subsequent layers. The purpose is to reduce the problems of information loss and gradient disappearance due to increase of layers. In fact, the residual unit in the GRSN

model consists of the activation function, the BN layer, the Dropout function, and the inception block. Shortcut is added to each residual unit layer. It is worth noting that the output size of the network may change after each residual unit. If the input image and the output image have the same size, they can be added directly; if they have different sizes, they need to be adjusted by 1×1 convolution.

4. With the increase of network layers, the number of channels is also increased; this results in more diverse features extracted from the model. Due to this reason, the accuracy is significantly increased. For example, in the first two residual units of network, the image size is adjusted via $64 \ 1 \times 1$ convolutions. After these two residual units, the channel number is doubled, and the size needs to be adjusted via $128 \ 1 \times 1$ convolutions. After 4 residual units, the size needs to be adjusted via $256 \ 1 \times 1$ convolutions, and so on. Due to hardware limitations and limited memory, in our network structure the channel number will not be doubled after 6 residual units. Therefore, after 6 residual units, all sizes are adjusted using $256 \ 1 \times 1$ convolutions.
5. The SOFTMAX function is employed in the last fully connected layer.

The other models have similar specific structures, only with a different shortcut, so we will not elaborate on these models here.

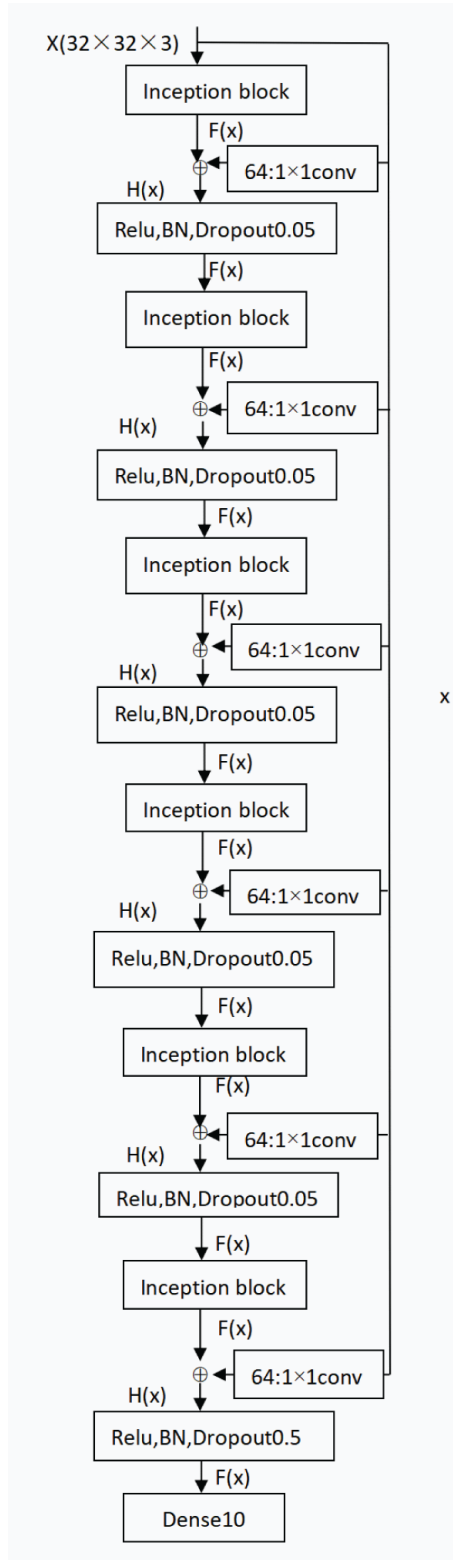
Parameter Settings of GRSN

For the input image of size $32 \times 32 \times 3$, the parameters of the first five layers of GRSN are listed in Table 1, and the parameters of other layers can be derived similarly. The parameter changes are shown in Table 1.

Table 1. Parameters of the first five layers of GRSN

Structure	Convolution Kernel Size	Strides	The Number of Fillers	Channel Numbers	Output Size
input				3	$32 \times 32 \times 3$
The first Inception block	1×1 3×3 5×5	1	all zero padding	16	$64 \times 32 \times 32$
The 2 nd Inception block	1×1 3×3 5×5	1	all zero padding	16	$64 \times 32 \times 32$
Double the number of channels				32	
The 3 rd Inception block	1×1 3×3 5×5	1	all zero padding	32	$128 \times 32 \times 32$
The fourth Inception block	1×1 3×3 5×5	1	all zero padding	32	$128 \times 32 \times 32$
Double the number of channels				64	
The fifth Inception block	1×1 3×3 5×5	1	all zero padding	64	$256 \times 32 \times 32$

Figure 12. Detailed structure of GRSN



EXPERIMENTAL ANALYSIS

Experimental Environment

The experimental environment is shown in Table 2.

Experimental Data

In order to better compare the performance of the algorithm, we use three typical general data sets, namely MNIST, cifar10, cifar100.

The MNIST data set is composed of handwritten digits of 250 different people. There are 70000 28×28 grayscale pictures of handwritten digits 0 to 9 in MNIST datasets, including 60,000 training sets and 10,000 test sets.

The cifar10 data set has a total of 60,000 images and tags, of which 50,000 images and tags are used for training, and 10,000 images and tags are used for testing. The data set includes 10 categories, including dog, cat, bird, airplane, frog, horse, boat, deer, truck and automobile. with 6000 pictures in each category.

The cifar100 data set contains 100 categories, each with 600 32×32 pixel color pictures. These 100 categories are further divided into 20 super categories, such as fish, flowers, household furniture, and fruits and vegetables.

Experiment Process

Three comparative experiments were performed. The settings of hyper-parameters should be determined based on comparisons of many experiments. In the GRSN network model, batch_size is set to 128, the parameter of the last Dropout layer is set to 0.5, and the parameters of other Dropout layers are set to 0.05. The Adam optimizer is used to optimize the model, and the cross-entropy loss function is employed. During the experiment, the accuracy and performance of various models on the test set were compared, and the images were converged.

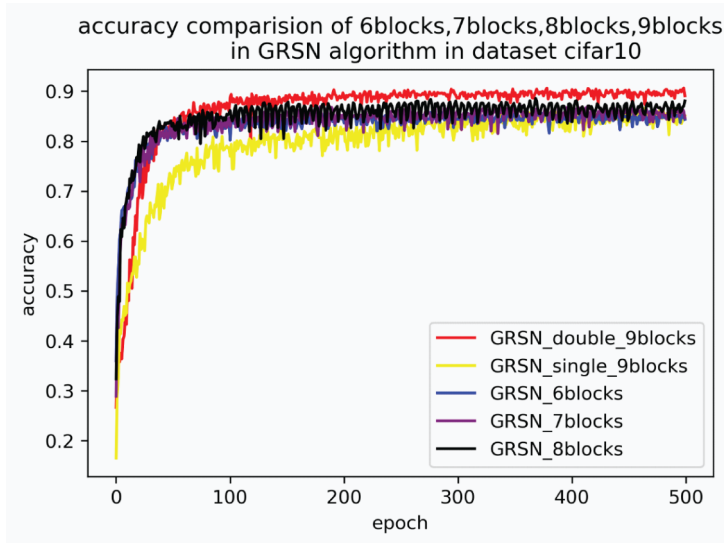
Comparison of the Classification Accuracy of Various Network Models on Data Set Cifar10

1. The accuracy of GRSN models with 6, 7, 8, and 9 residual blocks on the cifar10 data set were tested, and these models are GRSN_6blocks, GRSN_7blocks, GRSN_8blocks, and GRSN_9blocks. In this experiment, the amount size of data sent in each round was 128. With the increase of layers, the channel number also increased, shortcut was added to each block, data in the data set was not enhanced, and the other parameters were the same. As shown in the figure 13, the accuracy presents an increasing trend with the increase of layers. In the meantime, GRSN_double_9blocks and GRSN_single_9blocks also performed data enhancement for the

Table 2. Experimental environment

Category	Name	Type	Performance
hardware	CPU	Intel core i7-10700	2.90GHZ 8 cores and 16 threads
	Video card	GeForce RTX 3070 8G	
	RAM	32.00G	
software	system	Win10 64位	
	Deep learning framework	Tensorflow Keras	

Figure 13. Comparison of the classification accuracy of the GRSN models containing 6, 7, 8, and 9 residual blocks on cifar10



GRSN model containing 9 residual blocks. Due to hardware limitations, the amount size of data sent in each round was changed from 128 to 64, and the channel number increased with the increase of layers in GRSN_double_9blocks, while the channel number didn't increase with layers in GRSN_single_9blocks. The results show that data enhancement did not have significant impact on the classification accuracy, and the classification performance can be improved by increasing the channel number.

2. Comparison of the classification accuracy of the GRSN1_2, GRSN1_3, GRSN1_4 and GRSN models containing the same number of residual blocks on cifar10.

In this experiment, in the GRSN1_2, GRSN1_3, and GRSN1_4 models, the channel number did not increase with the increase of layers, data in the cifar10 data set was not enhanced either, and the data size sent in each round was 128. In the GRSN model, data in the cifar10 data set was enhanced, the channel number increased appropriately with the increase of layers, and the data size sent in each round was 64. According to the figure 14, the GRSN model significantly outperforms the GRSN1_2, GRSN1_3, and GRSN1_4 models. Therefore, compared to the model by adding one shortcut for several blocks, the performance is better by adding one shortcut for each block. At the same time, GRSN_double outperforms GRSN_single because the number of channels increases with layers in GRSN_double.

3. The classification accuracy of the GoogLeNet, ResNet, and GRSN models containing 10 blocks on data set cifar10 were compared. In this experiment, the performances of GoogLeNet doubling the channel number (GoogLeNet_double) and GoogLeNet not doubling the channel number (GoogLeNet_single) were evaluated and compared with the performances of ResNet and GRSN. As shown in Figure 15, GRSN_double has the best performance. In GRSN_double, data in the cifar10 data set was enhanced, the channel number increased appropriately with the increase of layers, and the size of data sent in each round was 64. The difference between GRSN_single and GRSN_double is that the channel number did not increase with the increase of layers in GRSN_single, and the improvement of the classification performance of GRSN_single was not significant.

Figure 14. Comparison of the classification accuracy of the GRSN1_2, GRSN1_3, GRSN1_4, and GRSN models on cifar10

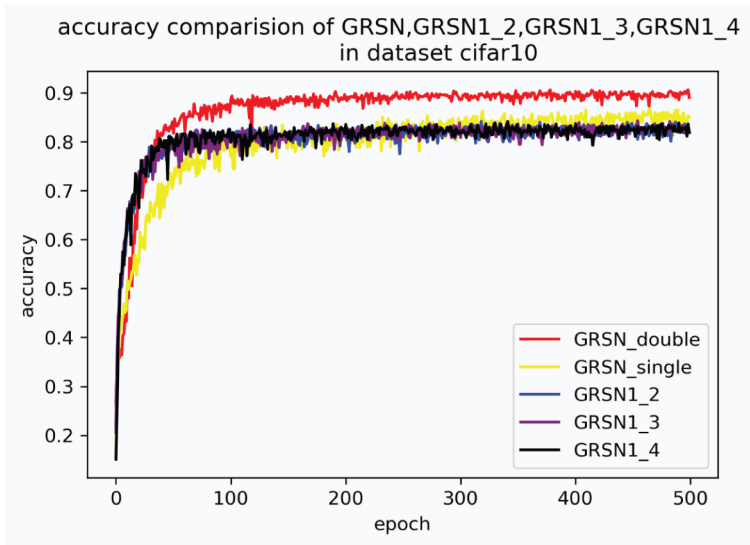
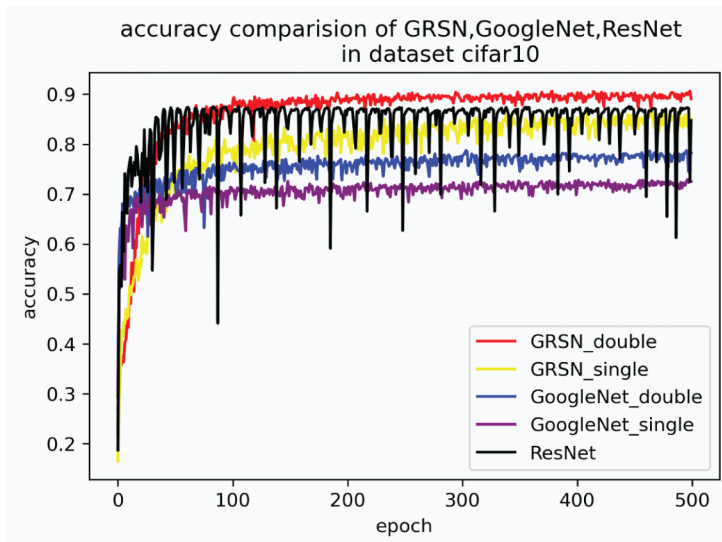


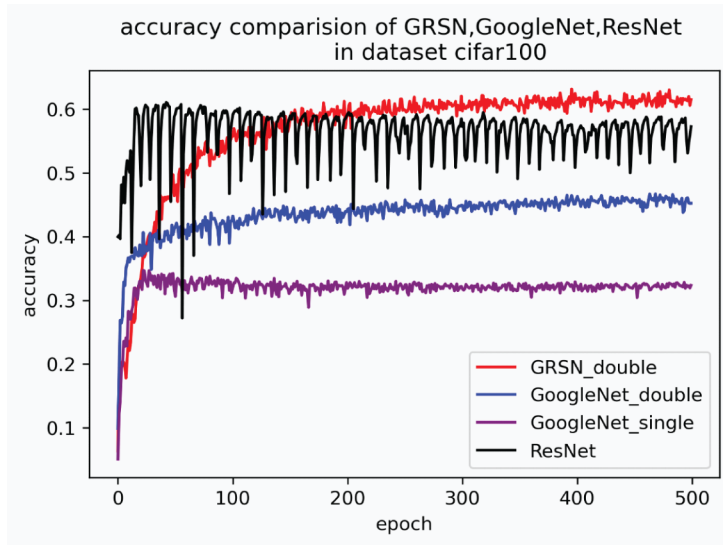
Figure 15. Comparison of the classification accuracy of GoogleNet, ResNet and GRSN on cifar10



Comparison of the Classification Accuracy of Various Network Models on Data Set Cifar100

1. The classification accuracy of GoogLeNet, ResNet, and GRSN on data set cifar100 were compared. Both GoogLeNet and ResNet had 10 blocks, and GRSN contained 9 blocks. In GoogLeNet, the channel number increased approximately with the increase of layers like in GoogLeNet_double. The results are shown in the figure 16, and the GRSN model GRSN_double still has the best performance.

Figure 16. Comparison of the classification accuracy of GoogleNet, ResNet, and GRSN on cifar100

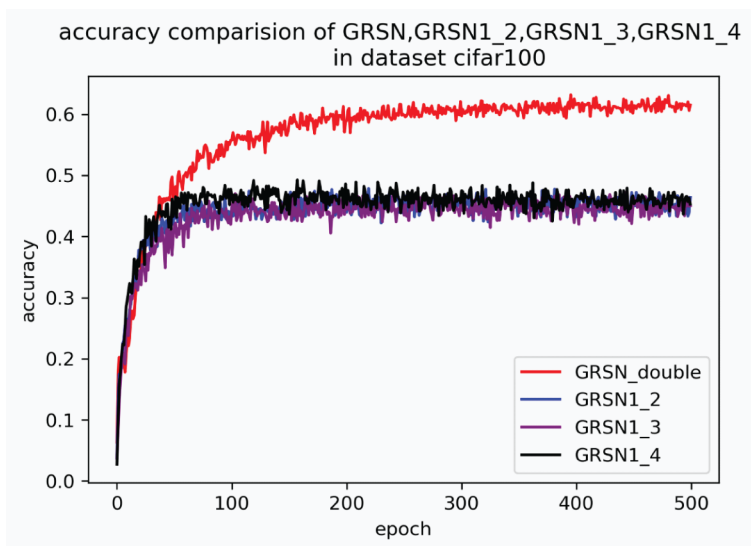


2. The classification accuracy of GRSN, GRSN1_2, GRSN1_3 and GRSN1_4 on data set cifar100 were compared, and the results are shown in the figure 17.

The GRSN_double model has the best performance. In this model, shortcut was added for each block, the number of channel increased with the increase of block number, the data was enhanced, and the size of data sent in each round was 64.

There was one less experiment on data set cifar100 than on data set cifar10. The first experiment on cifar10 was not conducted on cifar100, and the 9-block GRSN model was directly compared with other models.

Figure 17. Comparison of the classification accuracy of GRSN, GRSN1_2, GRSN1_3, and GRSN1_4 on cifar100



Comparison of the Classification Accuracy of Various Network Models on Data Set MNIST

This experiment was performed on data set MNIST, comparing the performances of GoogLeNet and GRSN, and the results are shown in the Figure 18.

The accuracy of both models was very high, at more than 99%. The highest and average accuracy of GRSN was 0.03 and 0.1 higher than the highest and average accuracy of GoogLeNet.

Result Analysis

The results of the three experiments described above are summarized in Table 3.

According to the results of the above three experiments, we can reach the conclusion that the best performances can be achieved in the GRSN model by using the following settings: add shortcut to each block; use data enhancement; set the size of data in each round as 64; the channel number increases appropriately with the number of network layers; use ReLU as the activation function; set the Dropout parameter of each previous layer as 0.05; and set the Dropout parameter of the last layer as 0.5.

CONCLUSION

By studying the CNN, LeNet, AlexNet, VGG, GoogLeNet, ResNet, and DenseNet algorithms and models, this paper proposes the GRSN, GRSN1_2, GRSN1_3, and GRSN1_4 models, with GRSN having the best effect. By comparing GRSN, GRSN1_2, GRSN1_3, and GRSN1_4 models, the best way to introduce the shortcut is obtained. GRSN is compared with GoogLeNet Inception v1 and ResNet on datasets cifar10, cifar100 and MNIST, the classification accuracy was significantly improved. Compared with ResNet, on data set cifar10, the highest accuracy of GRSN was 3.07 higher, and its average accuracy rate was 2.39 higher. Compared with ResNet, on data set cifar100, the highest accuracy of GRSN was 2.08 higher, and its average accuracy rate was 0.24 higher. Compared with GoogLeNet, on data set cifar10, the highest accuracy of GRSN was 17.69 higher, and its average accuracy rate was 15.99 higher. Also compared with GoogLeNet, on data set cifar100, the highest accuracy of GRSN was 28.47 higher, and its average accuracy rate was 24.83 higher.

Figure 18. Comparison of the classification accuracy of the GoogleNet and GRSN models on MNIST

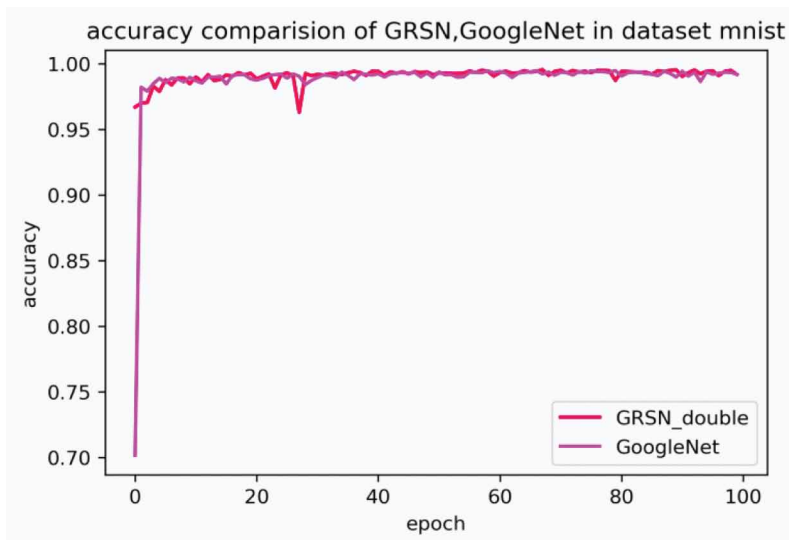


Table 3. Experimental summary

Model Name	Data set	Highest classification accuracy%	Average classification accuracy%	Instructions
GRSN_double_9blocks	Cifar10	90.64	86.47	As the number of layers increases the number of channels, the size of batch is 64, data enhancement, shortcut added to each block
GRSN_single_9blocks		86.65	79.60	Data enhancement, shortcut added to each block
GRSN_6blocks		86.15	83.23	GRSN model with 6 blocks, as the number of layers increases the number of channels, the size of batch is 128
GRSN_7blocks		87.09	83.73	GRSN model with 7 blocks, as the number of layers increases the number of channels, the size of batch is 128
GRSN_8blocks		88.59	84.86	GRSN model with 8 blocks,as the number of layers increases the number of channels, the size of batch is 128
GoogLeNet_double		78.73	75.49	As the number of layers increases the number of channels, the size of batch is 128
GoogLeNet_single		72.95	70.48	The size of batch is 128
GRSN1_2		84.19	80.46	Add shortcut every two blocks, the size of batch is 128
GRSN1_3		84.32	80.49	Add shortcut every three blocks, the size of batch is 128
GRSN1_4		84.18	80.61	Add shortcut every four blocks, the size of batch is 128
ResNet		87.57	84.08	Batch size=128
GRSN_double_9blocks	Cifar100	63.18	56.75	
GoogLeNet_double		46.72	43.50	
GoogLeNet_single		34.71	31.92	
ResNet		61.10	56.51	
GRSN1_2		47.85	43.97	
GRSN1_3		47.15	43.19	
GRSN1_4		49.26	45.11	
GoogLeNet_single	MNIST	99.62	99.30	
GRSN_double_9blocks		99.65	99.40	

Furthermore, these experiments and results will be extended to GoogLeNet Inception v2, v3 and v4—that is, adding shortcut from the input to each inception block on the basis of GoogLeNet v2, v3, and v4 and increasing the number of channels. The accuracy of the improved model is higher than that of GoogLeNet and ResNet.

REFERENCES

- Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., & Rueckert, D. (2018). DRINet for medical image segmentation. *IEEE Transactions on Medical Imaging*, 37(11), 2453–2462. doi:10.1109/TMI.2018.2835303 PMID:29993738
- Cheng, J., Tian, S., Yu, L., Gao, C., Kang, X., Ma, X., Wu, W., Liu, S., & Lu, H. (2022). ResGANet: Residual group attention network for medical image classification and segmentation. *Medical Image Analysis*, 76, 102313. doi:10.1016/j.media.2021.102313 PMID:34911012
- Gao, F., Wu, T., Chu, X., Yoon, H., Xu, Y., & Patel, B. (2019). Deep residual inception encoder-decoder network for medical imaging synthesis. *IEEE Journal of Biomedical and Health Informatics*, 24(1), 39–49. doi:10.1109/JBHI.2019.2912659 PMID:31021777
- Guan, Q., & Huang, Y. (2020). Multi-label chest X-ray image classification via category-wise residual attention learning. *Pattern Recognition Letters*, 130, 259–266. doi:10.1016/j.patrec.2018.10.027
- Guo, Y. X., Yang, W., Liu, Q., & Wang, Y. (2020). Summary of residual network research. *Computer Application Research*, 37(5), 1292–1297.
- Habibzadeh, M., Jannesari, M., Rezaei, Z., Baharvand, H., & Totonchi, M. (2018). Automatic white blood cell classification using pre-trained deep learning models: Resnet and inception. In *Tenth International Conference on Machine Vision (ICMV 2017)* (Vol. 10696, pp. 274–281). SPIE. doi:10.1117/12.2311282
- Han, Y. C., Wang J., Shi, T., & Cai, Q. H. (2019). Remote sensing image classification method based on improved GoogLeNet. *Journal of Missiles and Guidance*, 39(5), 139–142, 153.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE Computer Society. doi:10.1109/CVPR.2016.90
- Hou, S. Z., & Liu, Y. T. (2021). Image registration and fusion in the UV imager based on GoogLeNet-WT-Canny. *Semiconductor Optoelectronics*, 42(5), 733–740.
- Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020). Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), 2011–2023. doi:10.1109/TPAMI.2019.2913372 PMID:31034408
- Hu, L. H., Wang, C. L., Sun, Z. Q., & Yang, A. J. (2020). Image recognition method based on wgan. *Control Engineering*, 27(12), 2168–2175.
- Huang, X., Chen, W., & Yang, W. (2021). Improved Algorithm Based on The Deep Integration of GoogLeNet and Residual Neural Network. *Journal of Physics: Conference Series*, 1757(1), 12069. doi:10.1088/1742-6596/1757/1/012069
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. doi:10.1145/3065386
- Liu, J. W., & Huang, L. C. (2020). Pathological image recognition of breast cancer based on Inception-ResNet-V2. *Software Guide*, 19(5), 225–229.
- Peng, D. L., & Wang, T. X. (2019). Pruning algorithm based on GoogLeNet model. *Control and Decision*, 34(6), 1259–1264.
- Qiao, S. B., Pang, S. C., Wang, M., Zhai, X., Yu, S. X., & Ding, T. (2021). Convolutional neural network model for brain CT image classification based on residual mixed attention mechanism. *Tien Tzu Hsueh Pao*, 49(5), 984–991.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition [Paper presentation]. The 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, United States.
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (pp. 4278–4284). ACM. doi:10.1609/aaai.v31i1.11231

- Wang, J., Bao, Y., Wen, Y., Lu, H., Luo, H., Xiang, Y., Li, X., Liu, C., & Qian, D. (2020). Prior-attention residual learning for more discriminative COVID-19 screening in CT images. *IEEE Transactions on Medical Imaging*, 39(8), 2572–2583. doi:10.1109/TMI.2020.2994908 PMID:32730210
- Wang, J., Huang, D. Q., Guo, X., & Yang, L. M. (2022). Urban traffic pavement state recognition algorithm based on improved Inception-ResNet-v2 Science. *Technology and Engineering*, 22(6), 2524–2530.
- Wang, T. H. (2020). *An improved CNN ResNet deep learning neural network and its application* [Unpublished master's thesis]. Jilin University, Changchun, China.
- Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *European Conference on Computer Vision* (pp. 3-19). Springer.
- Wu, Y. F., & Wu, C. M. (2021). Classification of lung CT images based on Inception-ResNet. *Electronic Technology and Software Engineering*, (6), 196–197.
- Xiao, Z. J., Yang, X. D., Wei, X., & Tang, X. L. (2021). Application of improved lightweight network in image recognition. *Computer Science and Exploration*, 15(4), 743–753.
- Zhang, A., & Yu, H. X. (2021). Spatiotemporal double flow table tennis motion recognition based on improved GoogLeNet network. *Computer Knowledge and Technology*, 17(33), 78–80.
- Zhang, X., & Hu, X. J. (2019). Pulse wave recognition using deep hybrid neural networks based on GoogLeNet and ResNet. *Computer Systems & Applications*, 28(10), 15–26.
- Zhen, Y. P., Li, G. Y., & Li, Y. (2019). Summary of Research on Application of Deep Learning in Image Recognition. *Computer Engineering and Applications*, 55(12), 20–36.
- Zhou, T., Liu, Y. C., Lu, H. L., Ye, X. Y., & Chang, X. Y. (2022). ResNet and its application in medical image processing: Research progress and challenges. *Journal of Electronics and Information*, 44(1), 149–167.
- Zhuang, N., Ye, J., & Hua, K. A. (2017). Convolutional DLSTM for crowd scene understanding. In *2017 IEEE International Symposium on Multimedia (ISM)* (pp. 61-68). IEEE. doi:10.1109/ISM.2017.19
- Zong, Y., Chen, J., Yang, L., Tao, S., Aoma, C., Zhao, J., & Wang, S. (2020). U-net based method for automatic hard exudates segmentation in fundus images using inception module and residual connection. *IEEE Access: Practical Innovations, Open Solutions*, 8, 167225–167235. doi:10.1109/ACCESS.2020.3023273