Research Article

Research on Face Recognition Classification Based on Improved GoogleNet

Zhigang Yu, Yunyun Dong, Jihong Cheng, Miaomiao Sun [iD], and Feng Su

Yantai Nanshan University, Yantai 265713, China

Correspondence should be addressed to Miaomiao Sun; yuzhigang@nanshan.edu.cn

Face recognition is a relatively mature technology, which has some applications in many aspects, and now there are many networks studying it, which has indeed brought a lot of convenience to mankind in all aspects. This paper proposes a new face recognition technology. First, a new GoogLeNet-M network is proposed, which improves network performance on the basis of streamlining the network. Secondly, regularization and migration learning methods are added to improve accuracy. The experimental results show that the GoogLeNet-M network with regularization using migration learning technology has the best performance, with a recall rate of 0.97 and an accuracy of 0.98. Finally, it is concluded that the performance of the GoogLeNet-M network is better than other networks on the dataset, and the migration learning method and regularization help to improve the network performance.

1. Introduction

In recent years, with the development of the Internet, people have been in the era of big data, which has brought about an explosive increase in the amount of information, and in some access control and other aspects, people often use biometrics for identity authentication for a reason because people's faces or fingerprints are unique. In this regard, face recognition is the main recognition method, which brings great convenience to people's life. It mainly uses optical imaging of human faces to perceive and recognize people. At present, this technology is mainly applied to criminal investigation, surveillance systems, and secure payment.

The traditional face recognition [1] technology is mainly to extract feature points for face recognition and now the main application of deep learning technology [2–4]. Due to the large amount of data and high computing power, the precision aspects of deep learning have been greatly improved. In [5–7], an improved additive cosine interval loss function was proposed to improve the additive cosine interval loss function. By subtracting a value from the cosine value of the angle between the feature and the target weight and adding a value to the cosine value of the angle between the feature and the nontarget weight, the value is a number between 0 and 1, and select the best value through experiments to achieve the purpose of reducing the distance between classes and increasing the distance between classes. In [8–12], a face recognition model combining singular value face and attention convolutional neural network is proposed. The algorithm first uses a normalized singular value matrix to represent facial features, then inputs the features into the deep convolutional neural network with the attention module added, and improves the robustness of the network through cross-channel and spatial information fusion. Finally, the classification and recognition of face images is completed through the iterative training of the network [13, 14]. Through experiments on two commonly used databases, it is confirmed that the algorithm proposed in this paper has better recognition performance and better lighting robustness [15].

2. Network Improvement

2.1. GoogLeNet network. The GoogLeNet network model is to increase the width of the network; its main part is the inception structure, which can improve the accuracy of the network. The structure is shown in Figure 1.
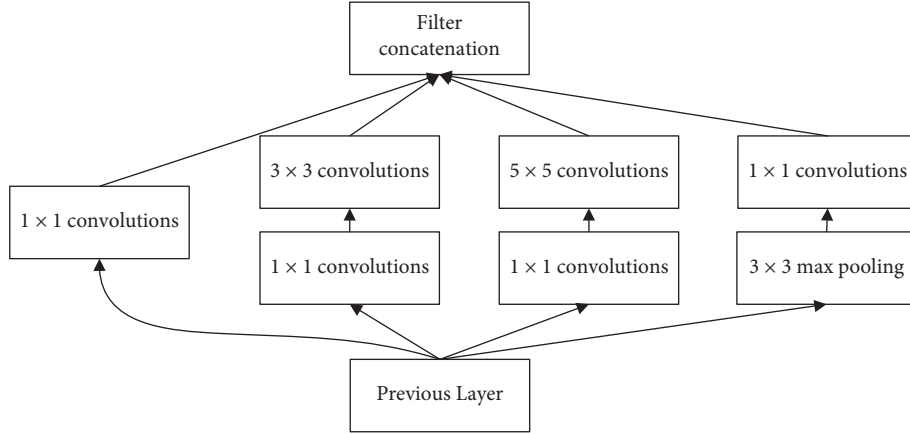
Figure 1: GoogLeNet Inception structure.

It can be seen from Figure 1 that the dimensionality reduction of the 11 convolution kernel can reduce the amount of parameters and increase the depth of the network. By branching and merging, the network width can be increased, which is conducive to the improvement of the network accuracy. This is the Inception-v1 structure. Table 1 shows the specific network structure, where type represents type, depth represents depth, pooling represents pooling, fc represents fully connected, and softmax represents the output layer. The final result will be output in the form of probability.

Since then, the GoogLeNet network has been continuously improved, and the Inception-v2 structure, the Inception-v3 structure, and the Inception-v4 structure have been successively proposed. Among them, the Inception-v2 structure mainly adds the batch normalization layer, Inception-v3 mainly replaces the two-dimensional convolution kernel with a one-dimensional convolution kernel, and Inception-v4 mainly adds the idea of residual network. This article chooses the GoogLeNet network of Inception-v4 structure. The GoogLeNet network mentioned below refers to the GoogLeNet Inception-v4 network.

Table 1: GoogleLenet network architecture.

Type	Size/stride	Output	Depth	Pooling
Conv	$7 \times 7/2$	$112 \times 112 \times 64$	1	
Max pool	$3 \times 3/2$	$56 \times 56 \times 64$	0	
Conv	$3 \times 3/1$	$56 \times 56 \times 192$	2	
Max pool	$3 \times 3/2$	$28 \times 28 \times 192$	0	
Inception (3a)		$28 \times 28 \times 256$	2	32
Inception (3b)		$28 \times 28 \times 480$	2	64
Max pool	$3 \times 3/2$	$14 \times 14 \times 480$	0	
Inception (4a)		$14 \times 14 \times 512$	2	64
Inception (4b)		$14 \times 14 \times 512$	2	64
Inception (4c)		$14 \times 14 \times 512$	2	64
Inception (4d)		$14 \times 14 \times 528$	2	64
Inception (4e)		$14 \times 14 \times 832$	2	128
Max pool	$3 \times 3/2$	$7 \times 7 \times 832$	0	
Inception (5a)		$7 \times 7 \times 832$	2	128
Inception (5b)		$7 \times 7 \times 1024$	2	128
Avg pool	$7 \times 7/1$	$1 \times 1 \times 1024$	0	
Dropout 40%		$1 \times 1 \times 1024$	0	
Fc		$1 \times 1 \times 1000$	1	
Softmax		$1 \times 1 \times 1000$	0	

2.2. GoogLeNet Network Improvement. The experiment in this article is running on three GPUs, so it is necessary to apply grouped convolution technology. In grouped convolution, sometimes the information exchange between groups is not very convenient, and it will also increase the size of the model and increase the size of the model. This paper uses channel shuffle to improve the packet volume, so as to avoid the problem perfectly.

The channel shuffle convolution is completely different from the previous convolution. In the previous convolution, a convolution core has to copy many, respectively, corresponding to the number of input channels, and then, superimpose the results, so there are many parameter quantities. The channel shuffle convolution is a convolution core, which only corresponds to one channel, so the parameter quantities are less.

Figure 2 shows the channel shuffle. We can see that there are labels in the figure, which are on the far left, where gconv is packet convolution.

In Figure 2, there are three groups of pictures. Figure 2(a) represents grouping convolution, which is the most traditional way. Figure 2(b) is the information exchange process after adding the improved idea in this paper, and Figure 2(c) is the effect after adding the improved idea in this paper.

Therefore, the improvement idea of this paper is that all GoogLeNet networks use channel shuffle. This paper calls the improved network googLeNet-m network.

2.3. Training Strategy

2.3.1. Activation Function. There are many kinds of activation functions. Their function is to increase the nonlinearity of the network. Only in this way can the network depth be meaningful. However, the previously proposed activation functions have various shortcomings, which are continuously improved. When it comes to the ReLu function, there are also defects when the input is less than 0, that is, the output is all 0.
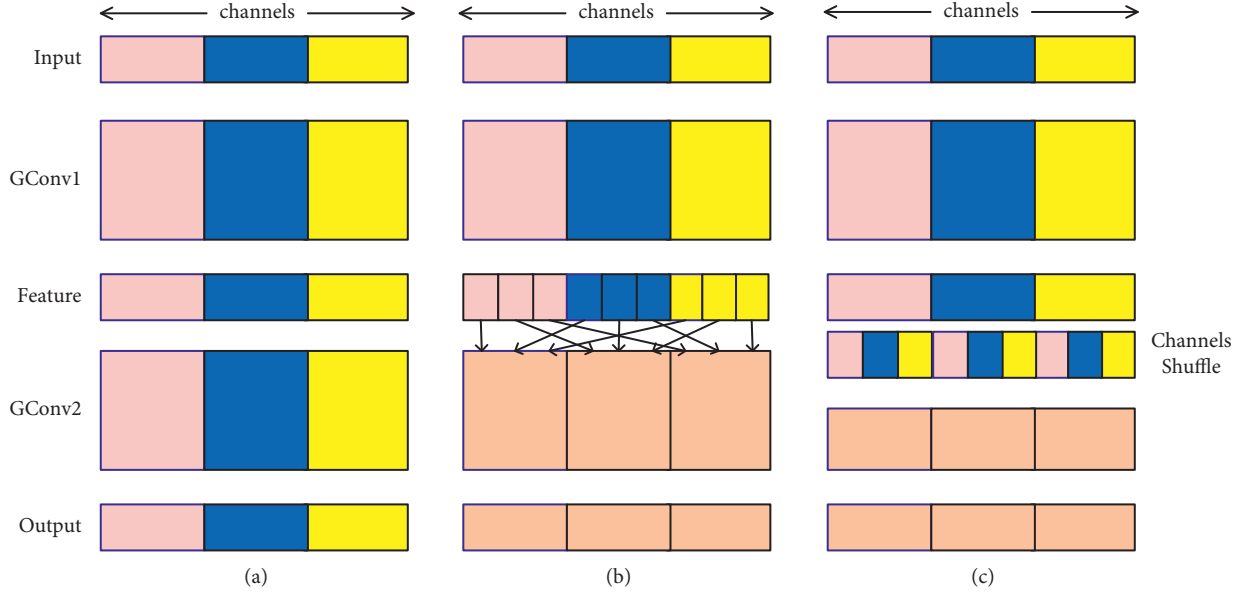
FIGURE 2: Channel shuffle schematic diagram. (a) Grouped convolution. (b) Apply shuffle. (c) Shuffle effect.

In order to solve the shortcomings of the ReLu function and then continue to propose improvements, in this paper, the random corrected linear unit (ReLU) transfer function is selected. In ReLU, the slope of the negative value is random during training, and it becomes fixed in the subsequent test.

2.3.2. Learning Rate. Learning rate is a very important super parameter. In deep learning training, a good initial learning rate is very important. If the initial learning rate is too large, it will lead to training shock. If the initial learning rate is too small, it will lead to training difficulty convergence. Therefore, it is necessary to select an appropriate initial learning rate, and the training method of the learning rate will also be very important. Therefore, this paper uses cosine function as periodic function to change the learning rate up and down and speed up the convergence.

2.3.3. Loss Function and Regularization. The cross-entropy loss function used in this paper is as follows:

$$C = -\frac{1}{n} \sum_{x} [y \ln a + (1 - y)\ln(1 - a)], \tag{1}$$

where C is the output, representing the loss, y is the expectation, and a represents the reality.

In the training process, we often encounter the problem of high training accuracy and low test accuracy, that is, overfitting. In this case, we can add our regular term to the loss function:

$$L = C + \frac{\lambda}{2n} \sum_{w} w^2. \tag{2}$$

2.3.4. Optimizer. The optimizer is used to adjust parameters in deep learning training. After the loss is obtained through the loss function, the parameters will be adjusted through the optimizer to optimize the network performance and finally achieve the convergence effect. Therefore, the selection of the optimizer is very important. If a poor optimizer is selected, the training will be difficult to converge or the convergence effect will be poor.

This paper selects the Adam optimization method. This optimizer belongs to the second-order moment category, which is better than the previous first-order optimizer. It also has momentum term and has its own unique advantages, which will make the parameters more stable.

3. Experiment Analysis

3.1. Transfer Learning. This article mainly uses IMDB WIKI face dataset for training and testing. This article uses transfer learning to train the GoogLeNet-M network and uses the ImageNet dataset to perform transfer learning training on the GoogLeNet-M network. The original data below all represent the IMDB WIKI face dataset.

In order to determine the correctness of the network performance, migration learning, and regularization proposed in this article, this article sets up several sets of comparative experiments as follows. First use the ImageNet dataset for migration learning and then use the original data training to join the regularized GoogLeNet-M network. GoogLeNet network directly trains the original dataset. Directly train the DenseNet network of the original dataset. Train the ResNet network directly on the original dataset. There are six sets of comparative experiments in total.

The experiment process uses three GPUs, that is, the grouped volume points are divided into three groups. The batch sizes are set to 192, each group of convolution is responsible for 64, and the initial learning rate is set to 0.01.

After the training converges, the model parameters are retained, and then, the original dataset is trained with an initial learning rate of 0.0001 to achieve the effect of

migration learning. For the same training 600 epochs, batch sizes are set to 192, each group convolution is responsible for 64, and the initial learning rate is set to 0.01.

3.2. Analysis of Results.
The final accuracy and loss curve after passing the experiment are shown in Figures 3 and 4:

In Figure 3, Pre-T-GoogLeNet-M represents the regularized GoogLeNet-M network that uses the ImageNet dataset for migration learning, and Pre-GoogLeNet-M represents the GoogLeNet-M network that uses the ImageNet dataset for migration learning.

The network performance from high to low is Pre-T-GoogLeNet-M, Pre-GoogLeNet-M, GoogLeNet-M, DenseNet, GoogLeNet, and ResNet. Among them, the comparison of the two network performances of Pre-T-GoogLeNet-M and Pre-GoogLeNet-M shows that the regularization proposed in this article is successful. The comparison of Pre-T-GoogLeNet-M, Pre-GoogLeNet-M, and GoogLeNet-M shows that the migration learning method proposed in this article helps to improve network performance. The comparison of GoogLeNet-M and GoogLeNet network shows that the improvement of the model in this article has effectively improved the network performance. The comparison of GoogLeNet-M, DenseNet, and ResNet shows that the improvement of the model in this paper surpasses other networks in terms of face recognition and classification.

Common evaluation criteria include recall, precision, and $F1$ value. Table 2 is the confusion matrix of the classification results. Through the confusion matrix of the classification results, the formulas of these three evaluation criteria can be obtained.

The TP in Table 2 refers to the actual positive case and the prediction also positive case. FN refers to the fact that the actual is a positive example and the prediction is a negative example. FP refers to the fact that the actual is a negative example and the prediction is a positive example. TN refers to the fact that the actual is a negative example, and the prediction is a negative example.

The recall rate refers to the proportion of images predicted to be lesions that are correctly predicted. The specific formula is as follows:

$$recall = \frac{TP}{TP + FN}. \tag{3}$$

The accuracy rate is as follows:

$$precision = \frac{TP}{TP + FP}. \tag{4}$$

Recall rate and accuracy rate are a pair of contradictory standards. It is difficult to achieve double high. When the recall rate needs to be improved, the accuracy rate needs to be sacrificed, and when the accuracy rate needs to be improved, the recall rate needs to be sacrificed.

The $F1$ value is an evaluation method that comprehensively considers the recall rate and accuracy rate:
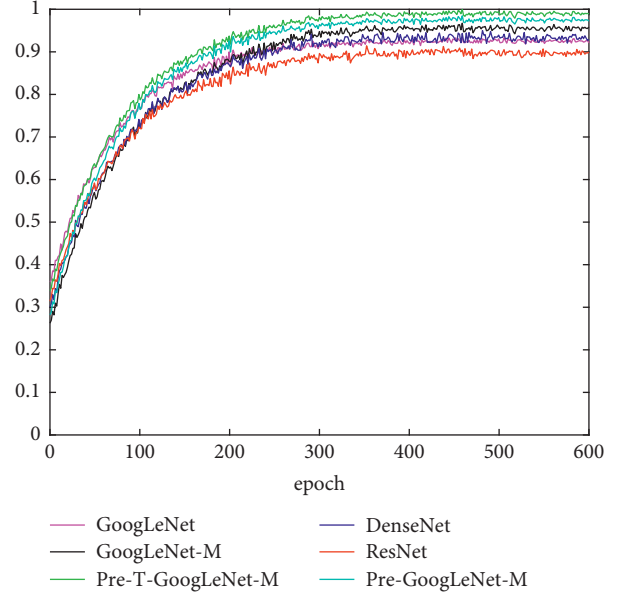
$$F1 = \frac{2 * R * P}{R + P}, \tag{5}$$
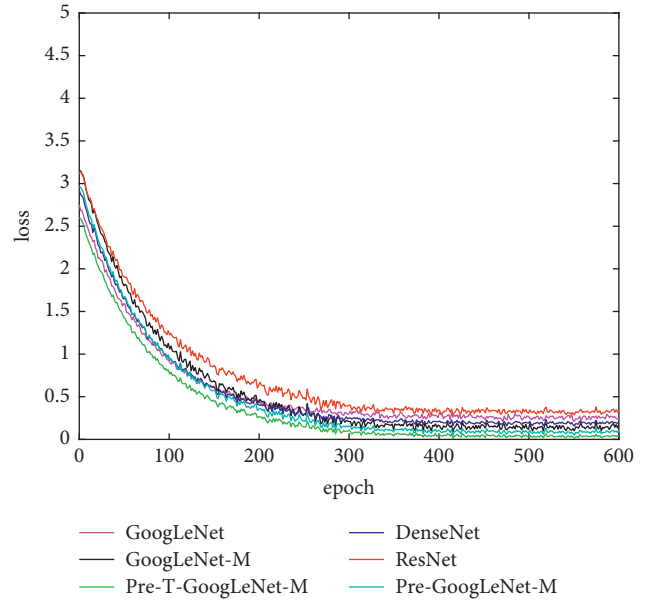

FIGURE 3: Accuracy curve.

FIGURE 4: Error curve.

where R represents the recall rate and P represents the precision rate.

The recall rate, precision rate, and $F1$ value obtained from the final experimental results are shown in Table 3:

From Table 3, it can be concluded that the order of $F1$ values of each model from high to low is the same as the order obtained above, followed by Pre-T-GoogLeNet-M, Pre-GoogLeNet-M, GoogLeNet-M, DenseNet, GoogLeNet, and ResNet.

The model evaluation criteria are not only those mentioned above but also an ROC curve commonly used in deep learning. The area of the curve can be used as the evaluation criteria for the performance of the model. The x-axis and Y-

TABLE 2: Confusion matrix of classification results.

The true situation	Forecast result	
	Positive example (p)	Negative example (N)
Positive example (P)	TP (true positives)	FN (false negative example)
Negative example (N)	FP (false positives)	TN (true negative example)

TABLE 3: F1 values for each model.

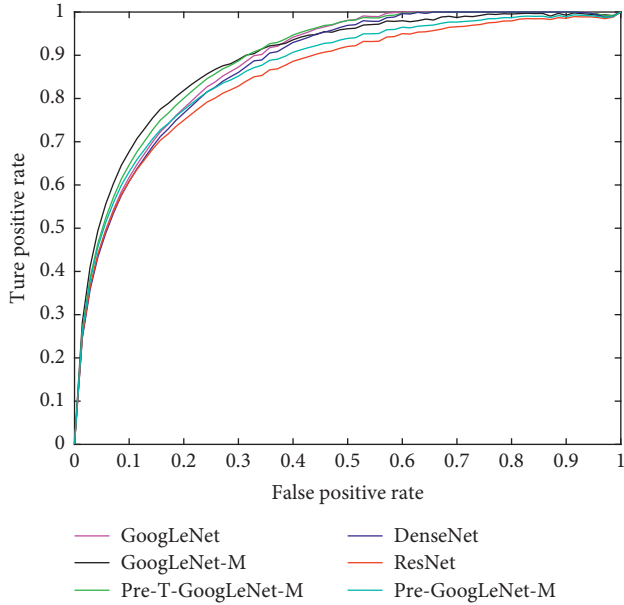| Model	Recall rate	Accuracy	F1 value
Pre-T-GoogLeNet-M	0.974	0.987	0.980
Pre-GoogLeNet-M	0.971	0.976	0.973
GoogLeNet-M	0.929	0.932	0.930
DenseNet	0.914	0.926	0.920
GoogLeNet	0.912	0.927	0.919
ResNet	0.907	0.913	0.910

FIGURE 5: ROC curve.

axis of the ROC curve are FPR (false positive rate) and TPR (true positive rate), respectively.

Formula (6) is TPR, which is as follows:

$$TPR = \frac{TP}{TP + FN}. \qquad (6)$$

Formula (7) is FPR, which is as follows:

$$FPR = \frac{FP}{FP + TN}, \qquad (7)$$

where TPR can also be regarded as the recall rate, and the two have the same meaning.

The area of ROC curve is AUC, and the larger the area, the better. It can be seen from Figure 5 that the algorithm proposed in this paper is optimal.

The order of the area AUC is Pre-T-GoogLeNet-M, Pre-GoogLeNet-M, GoogLeNet-M, DenseNet, GoogLeNet, and ResNet, which shows that the model performance is also sorted in this way, which is the same as the results obtained by the previous evaluation methods.

4. Conclusion

This paper studies the application of deep learning models in face recognition classification. It mainly improves the GoogLeNet, obtains the GoogLeNet-M network to improve the grouping convolution method under multi-GPU applications, and uses regularization and migration learning techniques to improve model performance. The final experimental results prove that the algorithm used in this paper is feasible. The next step in this article should use a larger dataset to test the generalization ability of the network model.

Data Availability

The simulation experiment data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

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

References

[1] L. Qing, S. Shan, and X. Chen, "Face recognition under arbitrary illumination based on spherical harmonic fundamental image," *Chinese Journal of Computers*, vol. 29, no. 5, pp. 760–768, 2006.

[2] S. Pang and Z. Hu, "Face recognition: a novel deep learning approach," *Journal of Optical Technology*, vol. 82, no. 4, pp. 237–245, 2015.

[3] H. Zhou and R. Zheng, "Coal mine underground track detection algorithm based on improved bisenet," *Journal Of Hubei University For Nationalities (Natural Science Edition)*, vol. 39, no. 4, pp. 398–403, 2021.

[4] Z. Pu, J. Hou, and Y. Liu, "Application of deep learning algorithm in Tibetan emotion analysis," *Computer Science and Exploration*, vol. 11, no. 7, p. 9, 2017.

[5] Y. Li and D. Xu, "Deep learning algorithm for style transformation of woodcut prints," *Journal of computer aided design and graphics*, vol. 32, no. 11, p. 9, 2020.

[6] L. Teng and B. Zhang, "Face recognition based on local binary features and deep learning," *Journal of Inner Mongolia University of Science and Technology*, vol. 37, no. 2, pp. 163–169, 2018.

[7] q. Zhu, H. Wang, and G. Zhang, "Research on face recognition algorithm based on fast PCA and k-cv optimized SVM," *Journal of Hubei University for Nationalities (NATURAL SCIENCE EDITION)*, vol. 39, no. 2, pp. 193–198, 2021.

[8] H. Wang, y. Hu, and C. Deng, "Research and implementation of face recognition algorithm based on LBP and elm," *Journal of Henan University of Technology (Natural Science Edition)*, vol. 40, no. 5, pp. 139–145, 2021.

[9] R. Peng, Y. Peng, and A. Lu, "Face recognition system based on improved PCA + SVM," *Electronic Science and technology*, vol. 34, no. 12, pp. 56–61, 2021.

[10] L. Li and M. Liu, "Research on robustness of face recognition based on machine learning algorithm," *Journal of Chinese Academy of Electronic Sciences*, vol. 12, no. 2, p. 6, 2017.

[11] J. Chen, J. Zhou, and S. Shen, "Research on anti attack algorithm of deep learning face recognition system," *Small Microcomputer System*, vol. 40, no. 8, p. 6, 2019.

[12] S. Liang, Y. Liu, and L. Li, "Face recognition algorithm under unrestricted conditions based on LBP and deep learning," *Journal of Communications*, vol. 35, no. 6, p. 7, 2014.

[13] J. Su and Z. Liu, "Face recognition based on improved residual network and graph clustering," *Microprocessors*, vol. 42, no. 5, pp. 37–44, 2021.

[14] D. Zhang, S. Chen, and J. Li, "Deep learning face recognition based on improved additive cosine interval loss function," *Journal of Sensor Technology*, vol. 32, no. 12, pp. 1830–1835, 2019.

[15] Y. Zhu, X. Ni, and Y. Yao, "Face recognition combining singular value face and attention deep learning," *Small Microcomputer System*, vol. 41, no. 8, pp. 197–201, 2020.