ACADEMIC
World Academic Union

ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 14, No. 3, 2019, *pp*.217-226

Face age and gender recognition based on improved VGGNet algorithm

Yulin Li

School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing, 210044, China

(Received February 20 2019, accepted June 24 2019)

Abstract. Recognition of age and gender based on face image is one of the hotspots of current artificial intelligence research. In this paper, an improved VGG+SENet algorithm is proposed to simplify the identification of age and gender algorithm by simplifying VGGNet model, improving the loss function and embedding the SENet module. Compared with other models, the improved network structure and loss function model proposed in this paper can quickly and accurately obtain output recognition results. Experimental results on multiple benchmark face datasets show that the proposed improved VGG+SENet algorithm has higher recognition accuracy than other related models based on deep learning.

Keywords: VGGNet, SENet, Age estimate, Gender identification

1. Introduction

Computer vision technology is widely used in various fields. Among them, face recognition [1] has made an important breakthrough in computer vision technology. At present, face recognition has been widely used in real life, such as identity authentication, network payment, public security monitoring, image tracking and so on. Our research interests focus on apparent age and gender estimates. Age and gender classification play a very important role in social life. According to the classification, we can know whether the person being contacted is young or old, whether it is “Ms.” or “Mr.”. Facial expressions have an important influence on the ability to estimate these individual classifications. The current facial expression analysis and recognition model capabilities are still far from meeting the needs of commercial applications [2].

VGGNet is a well-known convolutional neural network model applied to image classification. In 2018, He et al. proposed a facial expression recognition method based on VGGNet deep convolutional neural network for the low recognition rate of traditional convolutional neural networks in facial expression database. With a deeper network structure and 3*3 small convolution kernels and 2*2 small pool cores, the recognition rate is significantly improved, and the number of parameters is only slightly larger than the shallow layer. Shirkey et al.[4] combine feature-based gender identification methods with histogram-based age prediction methods to achieve desired goals. Through actual age estimation experiments, they demonstrate the effectiveness of the proposed method. In 2019, Chen et al.[5]proposes a face recognition algorithm based on SVM combined with VGG network model extracting facial features, which can not only accurately extract face features, but also reduce feature dimensions and avoid irrelevant features to participate in the calculation. Ghazi et al.[6] presented a comprehensive study to evaluate the performance of deep learning based face representation under several conditions including the varying head pose angles, upper and lower face occlusion, changing illumination of different strengths, and misalignment due to erroneous facial feature localization. Rothe et al.[7] used deep learning on the IMDB-WIKI dataset to solve the estimation of facial age in static face images.With the development of computer hardware and artificial intelligence technology, even under the conditions of large-scale training data, the computing power can satisfy the face recognition algorithm technology of practical application.

Aiming at the complexity of the existing deep learning neural network, the VGGNet network structure of the convolutional neural network is improved, the network structure is simplified, the loss function is improved, the SENet module is embedded, and a face age and gender recognition based on the improved VGG+SENet algorithm is proposed. Experiments show that under the large-scale data set, the improved VGG+SENet deep learning algorithm has a high recognition rate and can meet the requirements of actual face recognition applications.

2. VGGNet neural network introduction

VGGNet is a deep convolutional network jointly developed by Oxford Visual Geometry Group and DeepMind. In 2014, it won the second place in the classification project and the first place in the positioning project in the ILSVRC(ImageNet Large Scale Visual Recognition Challenge) competition, and the image feature extraction ability is very good, and its performance in multiple migration learning tasks is excellent for GoogLeNet. Therefore, extracting features from images, the VGG model is the preferred algorithm. The disadvantage of the VGG model is that there are many parameters, and the parameter quantity is more than 138 M. It requires a large storage space and the training time is relatively long. The model structure of VGGNet [8] is shown in Figure 1.

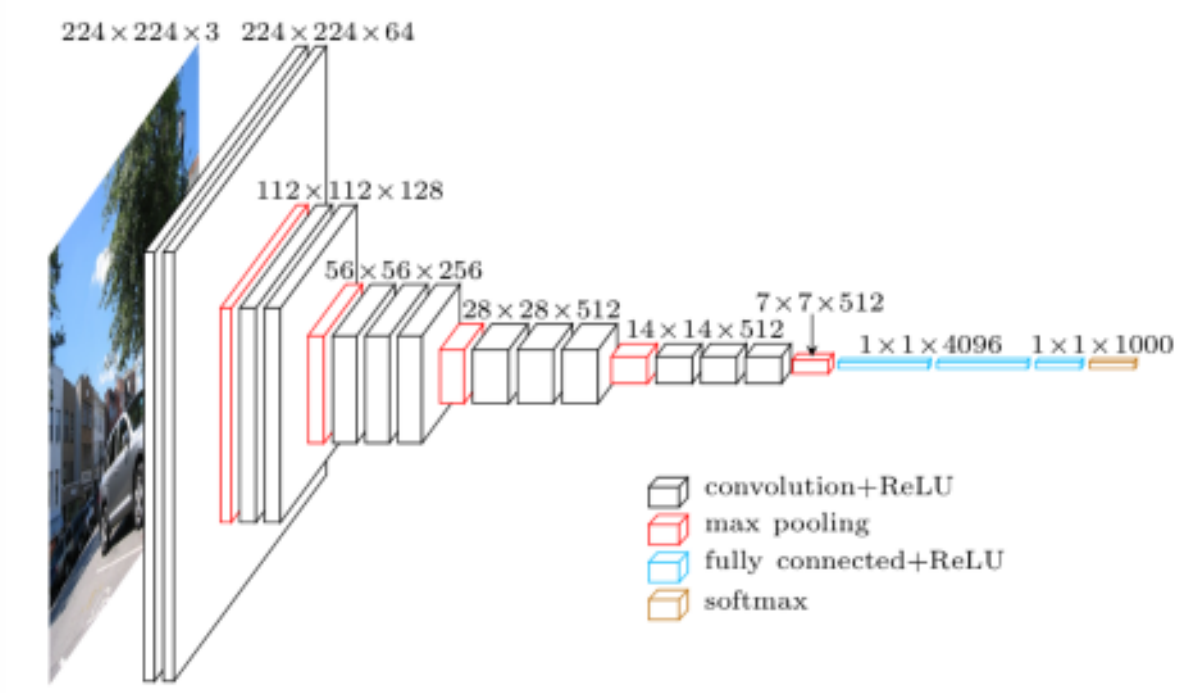

Fig. 1.VGGNet model structure

The structure of the VGGNet model is described as follows:

1) The input to the network is 224 × 224 RGB pictures, all of which are averaged.

2) The network has a total of 5 maximum pooling layers and 13 convolution layers. Three fully connected layers and one SoftMax classifier layer.

3) The convolution kernel has a size of 3×3 in the convolutional layer, a step size of 1 (stride=1), and a complement of 0 (pad=1).

4) The pooling layer uses MaxPooling, but not all convolution layers have a pooling layer. The pooling window is 2×2 and the step size is 2, that is, non-overlapping pooling is adopted.

5) All hidden layers are equipped with a ReLU layer.

6) After the first and second fully connected layers, dropout technology is also used to prevent network overfitting.

3. Face recognition based on improved VGG+SENet algorithm

3.1 Insufficient VGGNet network

Because VGGNet has good portability and promotion characteristics, VGGNet is chosen as the neural network model for face recognition. VGGNet has excellent image extraction effects, but has the following shortcomings:

1) There are many network layers, and the amount of calculation is large during training, and the convergence is slow.

2) There are many network parameters, and the memory required for the storage network is relatively large, and the training time is relatively long.

3) The performance of the face recognition model trained by SoftMax is poor. VGGNet uses the SoftMax classifier to classify images.SoftMax is an extension of logistic regression to solve multi-classification problems [9]. For ordinary image classification, SoftMax is simple and effective, however, when it is applied to face image recognition, since the characteristics of the face image are large within the class and the difference between the classes is small, the feature obtained by SoftMax training has a large inter-class distance, but the intra-class distance is not compact enough. The intraclass distance may be greater than the distance between classes, which may lead to face recognition errors.

In view of the problems existing in VGGNet, an improved VGG+SENet face age and gender recognition algorithm is proposed.

3.2 Network structure improvement

In order to reduce the amount of calculation, it is necessary to reduce the network parameters. From the experience of neural network extraction of image features, the closer to the underlying network, the more effective the extracted features. VGGNet extracts network features with significant effects. In order to maintain this feature, the underlying network is not changed, and the upper layer network is modified as much as possible. The network model contains convolutional layers, pooling layers, and fully connected layers. Compared with the convolutional layer, the full connection layer will generate more parameters due to more connections. Therefore, reducing the full connection layer is the most effective way to reduce network parameters.In order to reduce the parameters, refer to the model DeepID, GoogLenet, to reduce the three fully connected layers to a fully connected layer, and to modify the last largest pooled layer in front of the fully connected layer to be the max pooling layer. The kernel size of the max sampling layer is 7×7. By using the mean pooling layer instead of the maximum pooling layer, the image extraction feature can be maintained to the greatest extent, so that the network parameters can be effectively reduced while maintaining the network feature extraction capability as much as possible.The network parameters of the improved VGG+SENet are shown in Table 1.

Table 1.Improved VGG+SENet network parameters

Layer	Layer type	Input	Kernel	Step	Padding	Output
0	Input	$224\times224\times3$				$224\times224\times3$
1	Conv3-64	$224\times224\times3$	$3\times3\times64$	1	1	$224\times224\times64$
2	Conv3-64	$224\times224\times64$	$3\times3\times64$	1	1	$224\times224\times64$
3	MaxPool2	$224\times224\times64$	2×2	2	0	$224\times224\times64$
4	Conv3-128	$112\times112\times64$	$3\times3\times128$	1	1	$112\times112\times128$
5	Conv3-128	$112\times112\times128$	$3\times3\times128$	1	1	$112\times112\times128$
6	MaxPool2	$112\times112\times128$	2×2	2	0	$56\times56\times128$
7	Conv3-256	$56\times56\times128$	$3\times3\times256$	1	1	$56\times56\times256$
8	Conv3-256	$56\times56\times256$	$3\times3\times256$	1	1	$56\times56\times256$
9	Conv3-256	$56\times56\times256$	$3\times3\times256$	1	1	$56\times56\times256$
10	MaxPool2	$56\times56\times256$	2×2	2	0	$28\times28\times256$
11	Conv3-512	$28\times28\times256$	$3\times3\times512$	1	1	$28\times28\times512$
12	Conv3-512	$28\times28\times512$	$3\times3\times512$	1	1	$28\times28\times512$
13	Conv3-512	$28\times28\times512$	$3\times3\times512$	1	1	$28\times28\times512$
14	MaxPool2	$28\times28\times512$	2×2	2	0	$14\times14\times512$
15	Conv3-512	$14\times14\times512$	$3\times3\times512$	1	1	$14\times14\times512$
16	Conv3-512	$14\times14\times512$	$3\times3\times512$	1	1	$14\times14\times512$
17	Conv3-512	$14\times14\times512$	$3\times3\times512$	1	1	$14\times14\times512$
18	AvePool2	$14\times14\times512$	7×7	2	0	$2\times2\times512$
19	FC	$2\times2\times512$				2/8

By reducing the fully connected layer and replacing the maximum pooling layer with the max pooling layer, the parameter values are greatly reduced, and the number of parameters is reduced from the original 138 M to 14 M, which will effectively reduce the time required for model calculations and the storage space required to save model parameters. Comparison of structural parameters before and after improvement in Table 2.

Table 2 comparison of structural parameters before and after improvement.

Layer	VGGNet	VGG+SENet
0~17	Unchange	Unchange
18	Max Pooling	Ave Pooling
19	Full Connect	Full Connect
20	Full Connect	—
21	Full Connect	—
22	SoftMax	SoftMax
Total parameter /M	138	14

3.3 Loss function improvement

VGGNet uses the SoftMax classifier to classify images. The features obtained by SoftMax training have a large inter-class distance, but the intra-class distance is not compact enough, and the intra-class distance may be greater than the inter-class distance. Therefore, for ordinary image classification, SoftMax is simple and effective.However, due to the similarity of face images, it is easy to cause recognition errors. Therefore, the classification function can be enhanced by improving the loss function.

In order to get better classification results, different categories should be separated to minimize intra-class distance. SoftMax Loss can be used to separate different categories. Center Loss can increase the distance between classes and reduce the distance within the class. Center Loss is based on the SoftMax Loss classification and maintains a class center for each category. During training, gradually reduce the distance of class members from the class center and increase the distance between other class members and the center.

The SoftMax Loss function [10] can be written as follows:

$$L_s = -\sum_{i=1}^{m} \log \frac{e^{w_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{w_j^T x_i + b_{y_i}}} \tag{1}$$

The Center Loss function [10] can be written as follows:

$$L_c = \frac{1}{2} \sum_{i=1}^{m} \left\| x_i - c_{y_i} \right\|_2^2 \tag{2}$$

The mixing loss function can be written as follows:

$$\begin{aligned} L &= L_s + \lambda L_c \\ &= -\sum_{i=1}^{m} \log \frac{e^{w_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{w_j^T x_i + b_{y_i}}} + \frac{\lambda}{2} \sum_{i=1}^{m} \left\| x_i - c_{y_i} \right\|_2^2 \end{aligned} \tag{3}$$

Where: λ is the weight of Center Loss.

Referring to the model CaffeFace [10], the final SoftMax classification layer of VGGNet was adjusted and modified into SoftMax Loss and Center Loss joint monitoring training to achieve better training results. The improved VGG+SENet network structure is shown in Figure 2.

DOI https://doi.org/2024-JICS-22416 | Generated on 2025-04-05 22:12:15

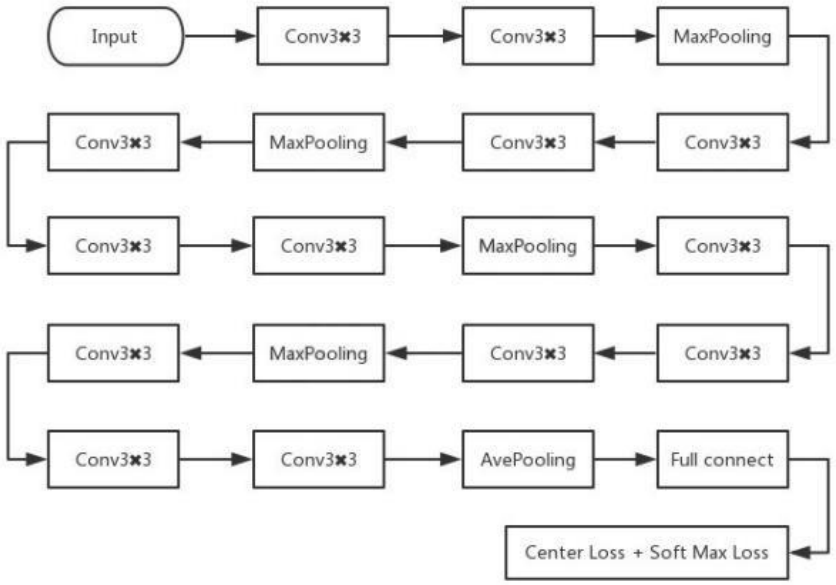

Fig.2.Improved VGGNet network structure

4. Embed SENet module

SENet starts with the relationship between feature channels and hopes to explicitly model the interdependencies between feature channels. In addition, instead of introducing a new spatial dimension to fuse between feature channels, a new "feature recalibration" strategy was adopted. Specifically, it is to learn the importance of each feature channel automatically by learning, and then to enhance the useful features according to this importance level and suppress the features that are not useful for the current task. In general, it is to enable the network to use the global information to selectively enhance the feature channel and suppress the useless feature channel, thus enabling feature channel adaptive calibration.The basic structure of the SENet module is shown in Figure 3.

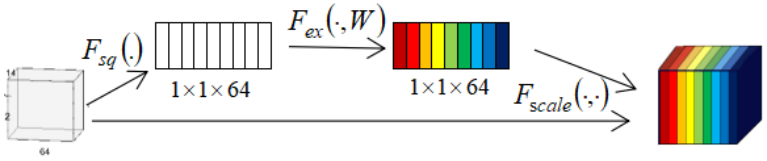

Fig.3.SENet module basic structure diagram

The Squeeze operates as follows:

$$u_c = v_c * X = \sum_{s=1}^{C'} v_c^{\ s} * x^s \tag{4}$$

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i, j) \tag{5}$$

The Excitation operation is as follows:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \tag{6}$$

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \bullet u_c \tag{7}$$

Where, $X = [x_1, x_2, \ldots\ldots, x_c]$ and $F_{scale}(u_c, s_c)$ refers to the corresponding channel product between the feature map $u_c \in R^{W \times H}$ and the scalar s_c.

5. Experiment and result

5.1 Model training

This article experiments using python3.6 in Windows 10 environment, the processor is Intel Core i5-7300HQ quad-core processor, RAM is 32G, the experimental data set is IMDB-WIKI face , select tensorflow as a deep learning platform. The IMDB-WIKI face database consists of an IMDB database and a Wikipedia database. The IMDB face database contains 460,723 face images, while the Wikipedia face database contains 62,328 face databases for a total of 523,051 face databases, each image in the IMDB-WIKI face database is labeled with the age and gender of the person. It is by far the largest public data set for age prediction, and it is of great significance for the study of age and gender classification.

Gender classification is divided into two categories, namely male and female. The gender tags are ‘Male’ and ‘Female’. Age classification Because the face image data set is large enough, this article divides the age into 8 categories, namely 8 age groups. Age tags are '0~10', '11~20', '21~30', '31~40', '41~50', '51~60', '61~70', '71~100' . The IMDB-WIKI face data set is shown in Figure 4. The face image data of the age group of 20 to 50 years old is mostly, and there are more than 10,000 face images in each age group.

Before training, first align the pictures, and image alignment is a more important step in face recognition [11]. Here, the face recognition alignment tool is used to align the IMDB-WIKI face database. The alignment effect is shown in Figure 5.

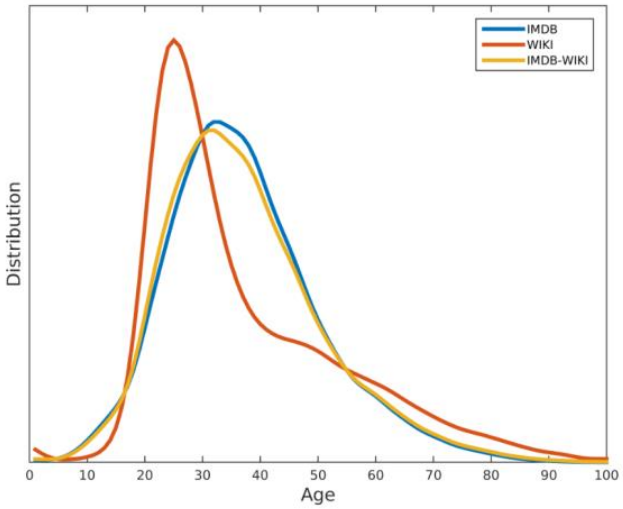

Fig.4.IMDB-WIKI face dataset age distribution map

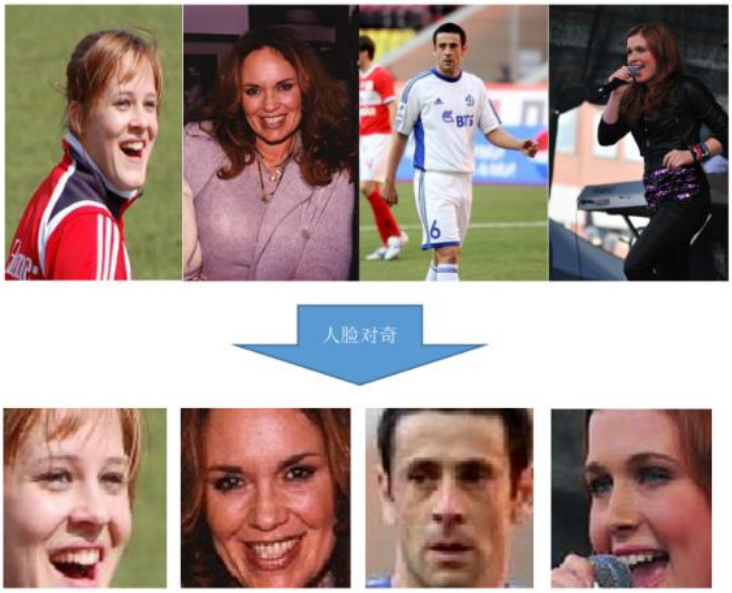

Fig.5.Training data face alignment

DOI https://doi.org/2024-JICS-22416 | Generated on 2025-04-05 22:12:15

5.2 Comparison of gender test results

In the gender classification task, first, the test is performed on the IMDB-WIKI data set. The pre-processing of the input image is to extract the face and scale it to 224×224 pixels, and then directly input the network for testing. Last,tested on the test data set, the resulting model confusion matrix is shown in Table 3.

Table 3. Confusion matrix of gender classification %

Tag value	female	male
female	97.4	2.6
male	1.5	98.5

As can be seen from Table 4, the final calculated accuracy on the IMDB-WIKI data set is 97.82. Compared with other small data sets, the huge data set can have better accuracy and utilize the model proposed in this paper. Better experimental results can be obtained. The model used in this paper is much simpler, and the time and amount of calculation required for training are greatly reduced.

Table 4. Comparison of gender classifications on different data sets

Model	Data set	Accuracy/%
References[12]	Adience	92.00
References[13]	IoG	89.10
This paper	IMDB-WIKI	97.82

5.3 Comparison of age test results

In the age classification task, the facial image face is first extracted and then the face is aligned, and the face image is all scaled to 224×224 pixels. The tested data set is the public IMDB-WIKI data set. In order to get the best results, the data in the dataset is filtered, the image data of the accurate age value is divided into the corresponding age groups, and then the image of the same data amount in the 8 types of labels is retained, and finally the data is shown in Table 5.

Table 5. Division of age classification

Classification	Age group	Picture quantity
0	0~10	1000
1	11~20	1000
2	21~30	1000
3	31~40	1000
4	41~50	1000
5	51~60	1000
6	61~70	1000
7	71~100	1000
Total	—	8000

The training data used in the paper contains 8 categories, a total of 8000 face images of different ages. By simplifying , improving the loss function and embedding SE module in the VGG+SENet model, the resulting confusion matrix of the optimal model is shown in Table 6.What is different from gender classification is that because the age recognition model uses well-preconditioned model parameters and effectively classifies age groups, it can still achieve better performance than traditional models.

Table 6. Confusion matrix of age classification %

Label	0	1	2	3	4	5	6	7
0	91.2	2.5	0.7	3.7	0.6	0	1.3	0
1	3.2	92.8	1.2	0.6	0	1.3	0	0.9
2	0	1.9	96.7	0.6	0	0.8	0	0
3	0	0.4	0.2	90.5	0.9	3.3	1.0	3.7
4	2.0	0	1.1	0	95.6	1.1	0	0.2
5	0	0.3	0	1.3	0	98.4	0	0
6	1.2	0	0.8	6.3	1.2	2.5	87.9	0.1
7	0	2.6	0	0.8	1.5	0.4	0.5	94.2

As can be seen from Table 7, the final calculation accuracy of the improved VGGNet model in the IMDB-WIKI data set is 90.8%. Compared with other neural network methods, the model can have higher accuracy and can utilize the model proposed in this paper. Better experimental results can be obtained. The model used in this article is much simpler and the amount of time and computation required for training is greatly reduced.

Table 7.Comparison of age classification of different neural networks

Model	Method	Accuracy/%
References[7]	VGG16-DEX	64.0
References[14]	LBP+SVM	56.3
This paper	VGG+SENet	90.8

Since this accuracy is still not very high, and the actual application is more concerned with the age value of each photo, rather than a rough age segment. In view of this consideration, this paper simplifies the final output layer into a neuron and directly outputs the predicted value of the age. The prediction results are shown in Table 8. It can be seen from Table 10, if the error is 10 years old, the model accuracy rate can be as high as 95.2%, which is obviously better than the existing other model test results.

Table 8. Age regression analysis results

Result bias	Proportion/%	Correct rate/%
0	10.6	11.3
±1	12.4	24.8
±2	9.5	38.7
±3	10.7	56.4
±4	9.1	62.5
±5	9.6	73.6
±6	8.8	80.5
±7	9.5	85.7
±8	4.9	89.1
±9	7.2	92.8
±10	5.5	95.2

6. Conclusion

Aiming at the image classification deep learning algorithm VGGNet, this paper proposes an improved VGG+SENet algorithm to intelligently identify the mixed model of age and gender. The improved algorithm uses the IMDB-WIKI data set as training and test data. The results show that compared with the original model, the model reduces the parameters, improves the loss function and embeds the SENet module, which not only reduces the intra-class differences, but also increases the inter-class differences, the face recognition performance is better, the gender classification model training is faster, the proposed age regression model can achieve a correct rate of 95.2% at 10 years of error.

Future research will focus on two aspects: one is to continue to optimize the network structure, so that the accuracy is improved, and the second is to study the network compression algorithm and get a lighter network, so that the depth model proposed in this paper can run in front-end devices with limited computing resources.

References

[1] E. Eidinger, R. Enbar, and T. Hassner, Age and gender estimation of unfiltered faces,IEEE Trans. Inf. Forensics Secur. 9 (2014), pp.2170–2179.

[2] G. Levi, and T. Hassncer, Age and gender classification using convolutional neural networks, in: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 34–42.

[3] Jun. H, Shuai. L, Jinming. S, Yue. L, Jingwei. W, and Peng. J, Facial Expression Recognition Based on VGGNet Convolutional Neural Network. In 2018 Chinese Automation Congress (CAC) pp. 4146-4151. IEEE.

[4] Shirkey, Dhanashree. M, and S. R. Gupta. An image mining system for gender classification & age prediction based on facial features. International Journal of Science and Modern Enginnering 1.6 (2013): pp. 8-12.

[5] Chen H, and Haoyu C. Face Recognition Algorithm Based on VGG Network Model and SVM Journal of Physics: Conference Series. IOP Publishing, 2019, 1229(1): pp. 012015.

[6] Mehdipour Ghazi M, and Kemal Ekenel H. A comprehensive analysis of deep learning based representation for face

recognition, Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2016: pp.34-41.

[7] Rothe R, Timofte R, and Van Gool L. Dex: Deep expectation of apparent age from a single image Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015:pp. 10-15.

[8] Shirkey, Dhanashree. M, and S. R. Gupta. An image mining system for gender classification & age prediction based on facial features. International Journal of Science and Modern Enginnering 1.6 (2013): pp. 8-12.

[9] Xiangbin S, Xuejian F, and Deyuan Z, Image classification based on deep learning mixed model migration learning, System Simulation Study, 2016, 28(1): pp.167-173, 182.

[10] Wen Y, Zhang K, and Li Z, A discriminative feature learning approach for deep face recognition, Computer Vision ECCV,2016: pp.499-515.

[11] Zhongyu X, Yan Z, and Wei W, Face detection based on skin color model and threshold segmentation , Changchun Institute of Technology, Natural Science, 2013, 34(4):pp. 382-386.

[12] Ozbulak G,Aytar Y,Ekenel H K,How transferable are CNN-based features for age and gender classification, 2016 International Conference of the Biometrics Special Interest Group, Damrstadt,2016:pp.1-6.

[13] Bekhouche SE, Ouafi A, and Benlamoudi A, Facial age estimation and gender classification using multi-level local phase quantization, Proceedings of the Third International Conference on Control, Engineering & Information Technology (CEIT), Sousse, 2015 :pp.1-4.

[14] Shan C,Learning local features for age estimation on real-life faces,Proceedings of the Workshop on Mul-timodal Pervasive Video Analysis,Firenze,2010: pp. 23-28.

DOI https://doi.org/2024-JICS-22416 | Generated on 2025-04-05 22:12:15