Face Gender Recognition Postprint Based on Multi-layer Feature Fusion and Adjustable Supervision Function Convolutional Neural Network

Authors: Shi Xuechao, Zhou Yatong, Chi Yue

Date: 2018-05-02T00:00:00+00:00

Abstract

To further improve the accuracy of gender recognition, we propose a Convolutional Neural Network model that combines multi-layer feature fusion with an adjustable supervision function mechanism, termed L-MFCNN, and apply it to facial gender recognition. Unlike traditional Convolutional Neural Networks (CNN), L-MFCNN combines the feature outputs from multiple shallow intermediate convolutional layers with that of the final convolutional layer, thereby fusing features across multiple convolutional layers. This approach not only leverages the holistic semantic information from deep convolutions but also incorporates detailed local texture information from shallow convolutions, resulting in more accurate gender recognition. Furthermore, L-MFCNN employs Large-Margin Softmax Loss with an adjustable target supervision function mechanism as its output layer. By exploiting its capacity to modulate different margins, this mechanism effectively guides the learning process of the deep convolutional network, reducing intra-class distances within the same gender while increasing inter-class distances between different genders, thus achieving superior gender recognition performance. Experimental results on multiple face datasets demonstrate that L-MFCNN achieves higher recognition accuracy compared to other traditional convolutional network models. The L-MFCNN model also offers novel insights and directions for future research in facial gender recognition.

Full Text

Preamble

Face Gender Recognition Based on Multi-Layer Feature Fusion Convolutional Neural Network with Adjustable Supervisory Function

Shi Xuechao, Zhou Yatong†, Chi Yue
(Tianjin Key Laboratory of Electronic Materials and Devices, School of Electronics & Information Engineering, Hebei University of Technology, Tianjin 300401, China)

Abstract: To further improve the accuracy of gender recognition, this paper proposes a convolutional neural network model based on multi-layer feature fusion with an adjustable supervisory function (L-MFCNN) for face gender recognition. Unlike traditional convolutional neural networks (CNNs), L-MFCNN combines the output features of multiple shallow intermediate convolutional layers with the final convolutional layer output, fusing multi-layer convolutional features. This approach not only leverages the holistic semantic information from deep convolutions but also incorporates detailed local texture information from shallow convolutions, enabling more accurate gender recognition. Additionally, L-MFCNN introduces the Large-Margin Softmax Loss with an adjustable target supervisory function mechanism as its output layer. By leveraging its ability to adjust different margin values, this loss effectively guides deep convolutional network learning to produce smaller intra-class distances within the same gender and larger inter-class distances between different genders, thereby achieving superior gender recognition performance. Experimental results on multiple face datasets demonstrate that L-MFCNN achieves higher recognition accuracy than other traditional convolutional network models. The L-MFCNN model also provides new ideas and directions for future face gender recognition research.

Key Words: face gender detection; multi-layer feature fusion; convolutional neural network (CNN); deep learning

0 Introduction

Face gender recognition is a critical step in face attribute analysis and represents a binary classification problem that automatically discovers and analyzes facial attributes from face image information. It has found applications in video surveillance, intelligent user interfaces, demographic statistics, and other domains. Face gender recognition encompasses various methods ranging from pattern recognition to deep learning, including artificial neural networks (ANN), principal component analysis (PCA), Bayesian decision theory, support vector machines, AdaBoost algorithms, and convolutional neural networks. Among these, Golomb et al. trained a two-layer artificial neural network, marking the first application of neural networks to gender recognition and achieving promising results on a small face dataset. Bruunelli et al. employed a three-layer backpropagation network for gender recognition on face images of different resolutions, attaining 93% accuracy on 30 low-resolution test images sized 8×6. Tamura et al. proposed using extreme learning for face gender recognition, achieving favorable accuracy rates. However, these methods using fully connected neural networks for face gender recognition ignored the two-dimensional correlations among pixels in images, resulting in insufficient classification accu-

racy.

Currently, deep learning has achieved breakthrough results in computer vision and object recognition. An increasing number of researchers have introduced convolutional neural network algorithms to face-related domains, including face detection, keypoint localization, face recognition, and facial age estimation. Nevertheless, research on face gender recognition based on convolutional neural networks remains in its early stages. Verma et al. utilized a 6-layer deep convolutional network for face gender recognition, achieving higher accuracy than previous traditional methods. Wang Jimin et al. employed a simple 2-layer convolutional network for face gender recognition, which only leveraged the sparse connectivity and weight sharing characteristics of traditional convolutional networks, resulting in relatively low gender recognition accuracy. Dong Lanfang et al. adopted a combination of deep convolutional networks and random forests for face gender recognition. Although this approach achieved high recognition precision, its implementation was overly cumbersome and complex. Zhang Ting et al. proposed a 9-layer cross-connected convolutional neural network (CCNN) model that connected the output of an intermediate pooling layer across two convolutional layers and a fully connected layer. However, due to feature loss during the pooling process, the classification results were not significantly better than those of traditional convolutional networks.

This paper proposes a face gender recognition algorithm based on multi-layer feature fusion and an adjustable supervisory function mechanism in convolutional neural networks. The proposed algorithm can be considered an improvement over traditional convolutional neural networks. It comprehensively utilizes multi-layer feature information from both shallow and deep convolutions by combining the outputs of the second convolutional layer (Conv2), fourth convolutional layer (Conv4), and fifth convolutional layer (Conv5). This multi-layer feature fusion approach enhances the semantic feature information of images ultimately input to the fully connected layer, thereby improving classification accuracy. Furthermore, the algorithm introduces the Large-Margin Softmax Loss with an adjustable mechanism as the target supervisory function. This effectively guides network learning to achieve smaller intra-class distances within the same gender and larger inter-class distances between different genders. Simultaneously, this objective function can adjust different margin mechanisms to prevent overfitting during network training and further enhance the robustness of the proposed algorithm.

The significance of this research is twofold: (a) it constructs a multi-layer convolutional fusion architecture that leverages multi-layer feature fusion to enhance the facial feature extraction process and improve face gender recognition accuracy; and (b) it introduces Large-Margin Softmax Loss to replace traditional Softmax loss as the final objective function to effectively supervise the model training process, thereby obtaining a more discriminative gender recognition model.

1.1 Convolutional Neural Networks

Convolutional neural networks consist of forward propagation and backward propagation, with convolutional layers and pooling layers alternating. A pooling layer follows each convolutional layer to reduce computation time and establish spatial and structural invariance.

Forward propagation involves a single calculation from input parameters to output results. The output of the upper layer serves as the input to the current layer, which then computes its output through an activation function. Considering the squared error loss function, for a classification problem with c classes and N training samples, the error function is expressed as:

$$E = \frac{1}{2N} \sum_{n=1}^{N} \sum_{k=1}^{c} (y_k^n - \hat{y}_k^n)^2$$

where y_k^n represents the k-th dimension of the label corresponding to the n-th sample, and $\hat{y}_k^n$ represents the k-th output of the network corresponding to the n-th sample.

Backpropagation updates the convolutional layers. The feature maps from the previous layer are convolved with a trainable kernel, and the result after passing through an activation function forms the output feature maps of the current layer. Each output map may be related to the convolution of several feature maps from the previous layer. The general form of a convolutional layer is:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right)$$

where k represents the convolution kernel, M_j denotes a selection of input features, and b is a bias term.

The downsampling operation does not alter the feature map data but merely reduces its size. If the sampling operator size is $n \times n$, then after one downsampling operation, the feature map size becomes $1/n$ of the original feature, expressed as:

$$x_j^l = f\left(\text{down}(x_j^{l-1}) * W_j^l + b_j^l\right)$$

where down$(\cdot)$ represents a downsampling function.

1.2 Weight Updates

Convolutional neural networks typically use the backpropagation algorithm and gradient descent to update weights and biases. The gradient descent algorithm

requires computing the partial derivatives of the loss function with respect to the weights and biases at each node, expressed as:

$$\frac{\partial J}{\partial W_{ij}^l} = \frac{\partial J}{\partial b_i^l} \cdot \frac{\partial b_i^l}{\partial W_{ij}^l}$$

$$\frac{\partial J}{\partial b_i^l} = \frac{\partial J}{\partial y_i^l} \cdot \frac{\partial y_i^l}{\partial b_i^l}$$

where W_{ij}^l is the weight of node j in layer l, b_i^l is the bias of node i in layer l, and γ is a regularization parameter. The backpropagated error is then:

$$\delta_i^l = f'(z_i^l) \sum_{j=1}^{s_{l+1}} \delta_j^{l+1} W_{ji}^{l+1}$$

where l represents the layer number, W denotes weights, b is a bias, and f is the activation function.

1.3 Theoretical Analysis of Multi-Layer Feature Fusion

For traditional CNNs, the process involves mapping and filtering images layer by layer, with the final mapped features in the last layer serving as the extraction result. Throughout this mapping process, different CNN layers learn different features. Early shallow networks contain more hierarchical information, such as edge and texture details, while the final layer outputs more abstract semantic information. [Figure 1: see original paper] illustrates the visualization of filter features from different layers, clearly revealing the distinct information provided by different network layers. Each convolutional layer exhibits different representation features for the same input image. Moreover, convolutional layers preserve more spatial information than fully connected layers.

Additionally, as shown in [Figure 1: see original paper], as convolutional layers deepen, the represented features change accordingly, evolving from obvious texture information in shallow layers to concentrated semantic information in deeper layers. The CNN convolution process is essentially a continuous filtering process, but the shallow features filtered out during this process are not necessarily useless for final recognition or classification. For instance, edge and texture information from the first and second shallow layers still contain valuable information and possess certain representational capabilities for images. Therefore, this paper considers fusing the feature information from shallow convolutional layers with deep features through multi-layer feature fusion to enhance the overall model' s representational capacity for images.

1.4 Multi-Layer Feature Fusion Convolutional Neural Network

Traditional convolutional neural networks only utilize deep convolutional feature information to build classifiers, indirectly abandoning the detailed texture information from shallow convolutional layers. To address this limitation, this paper proposes a face gender recognition algorithm based on a multi-layer feature fusion convolutional neural network. The deep convolutional model of this algorithm is shown in [Figure 2: see original paper]. The model comprises one input layer (data), five convolutional layers (Conv1, Conv2, Conv3, Conv4, Conv5), three pooling layers (Max1, Max2, Max3), two upsampling layers (Upsampling1, Upsampling2), one fusion layer (Concat), one fully connected layer (FC), and one output layer (Large-Margin Softmax Loss).

The input layer receives image information, which is then processed through five convolutional layers and two pooling layers for feature extraction. The outputs of Conv4 and Conv5 undergo 2× upsampling operations to obtain feature maps of the same size as the Conv2 output. These three feature components are then processed through the fusion layer (Concat) for multi-layer feature fusion. Finally, the fused features are classified by the fully connected layer (FC) and fed to the output layer, where the two nodes represent the categories of the input image.

The convolution kernel sizes and sliding stride parameters for each layer of the multi-layer feature fusion convolutional neural network are designed as shown in .

1.5 Selection of Output Layer Loss Function

The most commonly used output layer in previous convolutional networks is the cross-entropy Softmax Loss. By defining the i-th output feature f_i and its label y_i, the output layer softmax loss expression is:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right)$$

where f_j represents the j-th element of the final fully connected layer' s category output vector f, and N is the number of training samples. Since f is the output of the fully connected layer' s activation function, f_j can be expressed as $W_j^T x$, and the final loss function can be represented as an inter-class angular expression:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{e^{\|W_{y_i}\|\|x_i\|\cos(\theta_{y_i})}}{\sum_j e^{\|W_j\|\|x_i\|\cos(\theta_j)}}\right)$$

Although softmax is widely used in deep convolutional networks, this formulation cannot effectively learn features that are compact within classes and separable between classes. The goal of softmax is to ensure $W_{y_i}^T x > W_j^T x$, i.e., to obtain the correct classification result for x through inequality. The advantage of the Large-Margin Softmax Loss supervisory function is that it adds a positive integer variable m to create a decision margin, thereby imposing stricter constraints on the above inequality, i.e.:

$$\|W_1\|\|x\|\cos(m\theta_1) > \|W_2\|\|x\|\cos(\theta_2)$$

where θ_i represents the angle between W_i and x. If $\|W_1\|\|x\|\cos(\theta_1) > \|W_2\|\|x\|\cos(\theta_2)$ can be satisfied, then $\|W_1\|\|x\|\cos(m\theta_1) > \|W_2\|\|x\|\cos(\theta_2)$ must also be satisfied. Such constraints impose higher requirements on the learning process of W_1 and W_2, thereby creating a wider classification decision boundary between class 1 and class 2. The Large-Margin Softmax Loss expression is:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{e^{\|W_{y_i}\|\|x_i\|\cos(\psi(\theta_{y_i}))}}{e^{\|W_{y_i}\|\|x_i\|\cos(\psi(\theta_{y_i}))} + \sum_{j\neq y_i} e^{\|W_j\|\|x_i\|\cos(\theta_j)}}\right)$$

where $\psi(\theta)$ can be expressed as:

$$\psi(\theta) = \begin{cases} \cos(m\theta), & 0 \le \theta \le \frac{\pi}{m} \\ \cos(\theta), & \frac{\pi}{m} < \theta \le \pi \end{cases}$$

Thus, different margin m values can be adjusted to modify the classification boundary and control the learning difficulty to prevent network training overfitting.

2 Experimental Results and Analysis

To verify the recognition performance of the proposed L-MFCNN model, this paper compares it with traditional CNNs and the cross-connected convolutional neural network (CCNN) model proposed by Zhang Ting et al. Experiments were conducted on six face datasets: AR (Aleix Martinez and Robert Benavente) dataset, ORL (Olivetti Research Laboratory) dataset, UMIST (University of Manchester Institute of Science and Technology) dataset, FERET (Face Recognition Technology) dataset, LFW (Labeled Faces in the Wild) dataset, and CelebFace dataset. Training was performed on a WinFast gs4800 server with a 1.2 GHz GPU and 12 GB of video memory, while testing was conducted on a personal computer with a 3.6 GHz CPU, 8 GB of RAM, pre-installed Windows 10 Ultimate 64-bit operating system. The deep learning framework was Caffe, with a software programming environment of Python 2.7.5 and MATLAB 2014a.

Both the proposed L-MFCNN network and the traditional CNN network adopt the five-layer convolutional AlexNet architecture, with consistent parameters for each layer. The activation function uses the ReLU activation function: $f(z) = \max(0, z)$, which can be approximated by $f(z) = \log(1 + e^z)$. The penalty coefficient is set to $\lambda = 0.00001$, the momentum coefficient is $\beta = 0.9$, and the initial learning rate is $lr = 0.01$. The maximum number of iterations for the UMIST and FERET datasets is set to 10,000, with a model saved every 100 iterations; for the LFW and CelebFace datasets, the maximum number of iterations is 100,000, with a model saved every 1,000 iterations.

2.1 Dataset Preparation

The AR, ORL, UMIST, and FERET datasets are relatively small datasets, all converted to grayscale images. The LFW and CelebFace datasets are large datasets, and for these, RGB color images are used for training. All dataset images are converted to size 128×128. The training and testing sample allocations for the six datasets are shown in , and sample images from each dataset are displayed in [Figure 3: see original paper].

2.2 Experimental Testing and Analysis

To verify the impact of multi-layer feature fusion and the introduction of Large-Margin Softmax Loss as the output layer on gender recognition performance, experiments were divided into four groups. The first group involves visualization analysis of feature maps from each convolutional layer of the L-MFCNN model. The second group compares models using only multi-layer feature fusion for training and testing. The third group builds upon the second group by additionally replacing traditional softmax loss with Large-Margin Softmax Loss for comparative validation. The fourth group analyzes the male and female recognition performance of the L-MFCNN model on each dataset.

2.2.1 Visualization Analysis of L-MFCNN Convolutional Layers In the first group of experiments, we performed visualization analysis on the feature maps output by the trained multi-layer feature fusion L-MFCNN model. The feature maps corresponding to the outputs of convolutional layers Conv2, Conv4, and Conv5 were visualized, and the final output feature map after fusing these three layers was also visualized. The visualization results are shown in [Figure 4: see original paper].

From the visualization results, we can clearly analyze that: the feature maps output by the shallow Conv2 layer contain strong fine-grained texture and edge information from the original image; the deeper Conv4 layer outputs feature maps showing increased semantic information; the deep Conv5 layer outputs feature maps where the original image' s edge and texture information is barely discernible, representing the holistic semantic information of deep convolutions. However, when the features from Conv2, Conv4, and Conv5 layers are fused, the resulting image, as shown in Figure 4: see original paper, contains both the

fine-grained texture information from shallow layers and the holistic semantic information from deep convolutional layers. This enhances the model' s discriminative capability for images and enables effective differentiation of facial gender images.

2.2.2 Comparative Analysis of Multi-Layer Feature Fusion The second group of experiments compares the multi-layer feature fusion model (MFCNN) with traditional CNN and cross-connected convolutional neural network (CCNN) models. All three models are based on the five-layer convolutional AlexNet architecture. The traditional CNN maintains its original network structure; the cross-connected CNN (CCNN) connects the second hidden layer (first pooling layer) to the final fully connected layer; and the proposed multi-layer feature fusion model concatenates features extracted from three convolutional layers (Conv2, Conv4, Conv5) through a fusion layer (Concat). The gender recognition results of the three models on the six face datasets are shown in [Figure 5: see original paper].

As shown in [Figure 5: see original paper], the multi-layer feature fusion MFCNN model achieves higher recognition accuracy than CNN and CCNN on all six datasets: AR, ORL, UMIST, FERET, LFW, and CelebFace. The MFCNN model achieves recognition rates of 99.13% and 99.28% on the small datasets AR and UMIST, respectively, and recognition accuracies of 98.21%, 90.16%, and 90.08% on the relatively larger datasets FERET, LFW, and CelebFace, representing improvements of 1.63%, 2.2%, and 1.33% over the cross-connected CCNN model, respectively. The cross-connected CCNN model only utilizes features extracted from two pooling layers for recognition and classification, without adding much new feature information from the original image, resulting in limited accuracy improvement. In contrast, MFCNN fuses features extracted from three convolutional layers (Conv2, Conv4, Conv5), leveraging both the holistic semantic information from deep convolutional layers (Conv4, Conv5) and the detailed local texture information from the shallow convolutional layer (Conv2), thereby achieving better classification performance.

2.2.3 Comparative Analysis of Multi-Layer Feature Fusion with Large-Margin Softmax Loss Supervisory Function The third group of experiments builds upon the first group by introducing the Large-Margin Softmax Loss output layer into the multi-layer feature fusion model to form the L-MFCNN model, which is then compared with traditional CNN and cross-connected CNN (CCNN) models. All three models are based on the five-layer convolutional AlexNet architecture. The traditional CNN maintains its original network structure; the cross-connected CNN (CCNN) connects the second hidden layer (first pooling layer) to the final fully connected layer; and the L-MFCNN model fuses features from three convolutional layers (Conv2, Conv4, Conv5) through a fusion layer (Concat) and uses Large-Margin Softmax Loss as the output layer for inter-gender class supervision. By adjusting different margin m values to modify the classification boundary, higher gender

recognition accuracy is obtained. The gender recognition results of the five models on the six face datasets are shown in [Figure 6: see original paper].

The L-MFCNN model relies on the Large-Margin Softmax Loss output layer as a supervisory function, adjusting different margin m values to generate different decision margins and obtain wider classification decision boundaries. In experiments, the best results were achieved when m was set to 2 or 3. As shown in [Figure 6: see original paper], when $m = 3$, the L-MFCNN model achieves higher recognition accuracy than all other models on the AR, ORL, UMIST, FERET, LFW, and CelebFace datasets. The L-MFCNN model reaches recognition rates of 99.36%, 99.42%, and 99.38% on the small datasets AR, ORL, and UMIST, respectively, representing improvements of 0.17%, 0.31%, and 0.03% over the standalone feature fusion MFCNN model. On the relatively larger datasets FERET, LFW, and CelebFace, the recognition accuracies are 99.12%, 92.22%, and 90.88%, respectively, representing improvements of 0.27%, 0.54%, and 0.42% over the MFCNN model. This demonstrates that L-MFCNN achieves better gender recognition performance than multi-layer fusion MFCNN.

2.2.4 Analysis of L-MFCNN Performance on Male and Female Recognition Across Datasets The fourth group of experiments compares and illustrates the recognition performance of L-MFCNN, standalone multi-layer fusion MFCNN, and cross-connected CCNN models on male and female images in the six datasets. Male and female images from each dataset' s test set were separately selected and tested using the trained convolutional models from the second group of experiments. The test results are shown in [Figure 7: see original paper].

As shown in [Figure 7: see original paper], when the margin m is set to 3, the L-MFCNN model achieves classification accuracy for both male and female test sets that is no lower than that of CNN, CCNN, and MFCNN models across all six datasets. On the relatively larger datasets FERET, LFW, and CelebFace, the male classification accuracies are 98.88%, 90.16%, and 90.08%, respectively, which are 1.36%, 2.02%, and 3.3% higher than those of traditional CNN. Additionally, the recognition performance on female test sets is relatively lower than that on male test sets across all three models, because the number of female training samples on these three datasets (FERET, LFW, CelebFace) is significantly smaller than that of male training samples, resulting in relatively better learning of the male class during convolutional network training.

3 Conclusion

To further improve the accuracy of gender recognition, this paper proposes the L-MFCNN model, which combines multi-layer feature fusion with an adjustable target supervisory function mechanism based on Large-Margin Softmax Loss. The model integrates the output features of multiple shallow intermediate convolutional layers with the final convolutional layer output, fusing multi-layer convolutional features. By utilizing both the holistic semantic information from

deep convolutions and the detailed local texture information from shallow convolutions, more accurate gender recognition results are achieved. Additionally, the introduction of Large-Margin Softmax Loss with an adjustable mechanism as the output layer leverages its ability to adjust different margin values to effectively guide deep convolutional network learning. This produces smaller intra-class distances within the same gender and larger inter-class distances between different genders, resulting in better recognition and classification performance. Experimental results demonstrate that the L-MFCNN model achieves higher recognition accuracy than other models across six face datasets.

References

[1] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521 (7533): 436-444.
[2] Golomb B A, Lawrence D T, Sejnowski T J. SexNet: a neural network identifies sex from human faces [C]// Advances in Neural Information Processing Systems. 1991: 572-579.
[3] Brunelli R, Poggio T. HyberBF networks for gender classification [C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. 1992: 553-556.
[4] Tamura S, Kawai H, Mitsumoto H. Male/female identification from 8×6 very low resolution face images by neural network [J]. Pattern recognition, 1996, 29 (2): 331-335.
[5] Osuna E, Freund R, Girosi F. Training support vector machines: an application to face detection [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2002: 130-136.
[6] Farfade S S, Saberian M J, Li L J. Multi-view face detection using deep convolutional neural networks [C]// Proc of the 5th ACM on International Conference on Multimedia Retrieval. New York: ACM Press, 2015: 643-650.
[7] Liao S, Jain A K, Li S Z. A fast and accurate unconstrained face detector [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2016, 38 (2): 211-223.
[8] Zhang K, Zhang Z, Li Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23 (10): 1499-1503.
[9] Viola P, Jones M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57 (2): 137-154.
[10] Sun Yi, Liang Ding, Wang Xiaogang, et al. DeepID3: face recognition with very deep neural networks [J]. Computer Science, 2015.
[11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2015: 3431-3440.
[12] Levi G, Hassner T. Age and gender classification using convolutional neural networks [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015: 34-42.
[13] Rothe R, Timofte R, Van Gool L. Deep expectation of real and apparent

age from a single image without facial landmarks [J]. International Journal of Computer Vision, 2016: 1-14.
[14] Verma A, Vig L. Using convolutional neural networks to discover cognitively validated features for gender classification [C]// Proc of IEEE International Conference on Soft Computing and Machine Intelligence. 2014: 33-37.
[15] 汪济民, 陆建峰. 基于卷积神经网络的人脸性别识别 [J]. 现代电子技术, 2015, 38 (7): 81-84.
[16] 董兰芳, 张军挺. 基于深度学习和随机森林的人脸年龄和性别分类研究 [J//OL]. 计算机工程, : 1-6 (2017-05-23).
[17] 张婷, 李玉鑑, 胡海鹤, 等. 基于跨连卷积神经网络的性别分类模型 [J]. 自动化学报, 2016, 42 (6): 858-865.
[18] Liu W, Wen Y, Yu Z, et al. Large-margin softmax loss for convolutional neural networks [C]// Proc of International Conference on International Conference on Machine Learning. 2016: 507-516.
[19] Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.
[20] Shen W, Wang X, Wang Y, et al. Deepcontour: a deep convolutional feature learned by positive-sharing loss for contour detection [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3982-3991.
[21] Mansanet J, Albiol A, Paredes R. Local deep neural networks for gender recognition [J]. Pattern Recognition Letters, 2016, 70: 80-86.
[22] Hassner T, Harel S, Paz E, et al. Effective face frontalization in unconstrained images [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4295-4304.
[23] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks [C]// Proc of the 14th International Conference on Artificial Intelligence and Statistics. 2011: 315-323.
[24] Sanguansat P. Face hallucination using bilateral-projection-based two-dimensional principal component analysis [C]// Proc of IEEE International Conference on Computer and Electrical Engineering. 2008: 876-880.
[25] Hightower J, Borriello G. Location systems for ubiquitous computing [J]. Computer, 2001, 34 (8): 57-66.
[26] Shen X, Lin Z, Brandt J, et al. Detecting and aligning faces by image retrieval [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3460-3467.
[27] Phillips P J, Wechsler H, Huang J, et al. The FERET database and evaluation procedure for face-recognition algorithms [J]. Image and vision computing, 1998, 16 (5): 295-306.
[28] Huang G B, Ramesh M, Berg T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments, Technical Report 07-49 [R]. Amherst: University of Massachusetts, 2007.
[29] Guo Y, Zhang L, Hu Y, et al. Ms-celeb-1m: a dataset and benchmark for large-scale face recognition [C]// Proc of European Conference on Computer Vision. Springer International Publishing. 2016: 87-102.
[30] Zeiler M D, Fergus R. Visualizing and understanding convolutional

networks [J]. Computer, 2013, 8689: 818-833.

Note: Figure translations are in progress. See original paper for figures.