Research Postprint on Deep Learning-Based Face Recognition Algorithms

Authors: Hu Yazhou, Zhou Yali, Zhang Qizhi

Date: 2019-04-01T00:00:00+00:00

Abstract

Deep learning-based face recognition technology is currently one of the hot research topics in the field of artificial intelligence. Given the complexity of face images in real-world environments with respect to pose, illumination, and resolution, the Inception-ResNet-V1 network architecture was improved, with related work including dataset creation and hyperparameter tuning completed, and experimental research conducted on a domestic service robot platform. Experimental results indicate that the improved network architecture achieved an accuracy of 99.22% on the LFW test set, surpassing the 99.05% of the original network architecture; an accuracy of 99.20% on the Asian face dataset, surpassing the 97.10% of the original network architecture; and a false recognition rate of 3.43% on the self-constructed non-matching face dataset, lower than the 12.28% of the original network architecture. It can be observed that, compared with the original network architecture, the improved network architecture enhances face recognition accuracy and reduces the false recognition rate.

Full Text

Face Recognition Algorithm Based on Deep Learning

Hu Yazhou, Zhou Yali, Zhang Qizhi
(School of Automation, Beijing Information Science & Technology University, Beijing 100192, China)

Abstract: Face recognition technology based on deep learning is currently one of the hot topics in artificial intelligence research. Considering the complexity of angle, lighting, and resolution in real-world face images, this paper improves the Inception-ResNet-V1 network structure while completing dataset production, hyperparameter tuning, and related work, with experimental research conducted on a home service robot platform. Experimental results demonstrate that the improved network achieves 99.22% accuracy on the LFW test set, surpassing the original network' s 99.05%; reaches 99.20% accuracy on the Asian face dataset, exceeding the original network' s 97.10%; and achieves a false recognition rate of 3.43% on a self-built mismatched face dataset, lower than the original network' s 12.28%. These results show that compared with the original network structure, the improved architecture enhances face recognition accuracy while reducing false recognition rates.

Keywords: home service robot; face recognition; deep learning; Inception-ResNet-V1

0 Introduction

Home service robots [1] are special-purpose robots equipped with visual sensors, compasses, LiDAR, and other sensors designed to serve humans by performing household tasks. Face recognition is a biometric identification technology based on human facial feature information [2]. It involves using cameras or capture devices to collect images or video streams containing faces, then automatically detecting faces within these images or streams through relevant algorithms to obtain facial position information, followed by feature extraction, feature comparison, and output of facial information. This technology, also known as portrait recognition or facial identification, has become a hot research topic in artificial intelligence, with numerous scholars and researchers continuously optimizing existing algorithms or proposing new ones to improve the speed and accuracy of face detection and recognition while reducing false detection rates.

With the development of artificial intelligence, both face recognition technology and home service robot research have advanced considerably. Applying face recognition technology to home service robots has also become a research focus in computer vision and robotics. In countries with advanced home service robot research, such as Canada, Japan, and the United States, most research on various home service and industrial robots is jointly undertaken by enterprises and research institutes, effectively combining research, industrial development, and market demand to create a virtuous cycle where research serves enterprises and enterprises drive research forward. With favorable development environments in both academia and industry, these countries have achieved significant research results in home service robots.

In 2008, Canada introduced the perfect robot wife Aiko, a female robot with beautiful hair and delicate features capable of cleaning and household chores, skilled in mathematical calculations, able to recognize multiple family members, read newspapers, and provide directions [3]. In May 2010, Japan' s Fujitsu company launched a home medical rehabilitation robot teddy bear that could quickly recognize human facial expressions and respond accordingly. Sensors on the teddy bear could collect user emotional information and perceive the surrounding environment to find appropriate ways to interact with users [3]. In May 2017, iRobot, a leading American home robot company, announced the launch of a new sweeping robot in the Chinese market. iRobot' s Roomba sweeping robot could effectively remove dust, and the company developed the VSLAM image displacement positioning system that uses cameras for automatic positioning and map construction. However, since all data acquisition is completed through the camera on top of the device, any operational error in the camera directly affects positioning accuracy and consequently impacts the robot' s performance.

China' s home service robots started relatively late, and there remains a certain gap compared to foreign service robot development levels. Nevertheless, the broad application prospects attract an increasing number of researchers to this field. A representative example is the intelligent service robot "KeJia" independently developed by the University of Science and Technology of China [4]. This robot possesses core capabilities including natural language human-computer interaction, automatic reasoning and knowledge acquisition, environmental perception and modeling, and robot control. Sun@Home [5] is a home service robot research team from Beijing Information Science & Technology University. Since 2010, the team' s home service robots have participated in over ten domestic and international robot competitions and public demonstrations, achieving remarkable results and receiving praise from peers both at home and abroad. The team has conducted in-depth research on key technologies including target segmentation, indoor positioning and navigation, mobile object retrieval, sound source localization and speech recognition, object classification and recognition, and face recognition, with some technologies reaching domestic leading levels.

Face recognition is a typical problem of image pattern analysis, understanding, and classification computation involving multiple interdisciplinary fields. In 1965, Chan et al. [6] published a technical report at Panoramic Research Inc., marking the beginning of face recognition research. Over the past two to three decades, an increasing number of research institutes and internet companies have invested in face recognition research and development, achieving remarkable results in algorithm research and application innovation. For example, in 1991, a research group led by Pentland at MIT proposed the milestone Eigenface method [7]; in 1996, a group led by Kriegman proposed the equally significant FisherFace method [8]; in 1997, Peng Hui et al. from Tsinghua University [9] further improved the Eigenface method by proposing the use of inter-class scatter matrix as the generation matrix, which reduced the dimensionality of the generation matrix and improved feature computation speed without affecting face recognition accuracy; in 2005, a team led by Jain at Michigan State University conducted extensive research on 3-D face recognition methods [10], achieving prominent results. Traditional algorithms provided a solid foundation for face recognition research.

In recent years, due to the widespread application of deep learning, significant breakthroughs have been made in face recognition. In 2014, at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Facebook proposed the DeepFace [11] algorithm based on deep learning, achieving 97.35% recognition accuracy on Labeled Faces in the Wild (LFW), approaching human eye recognition accuracy of 97.52% on LFW. This algorithm uses convolutional networks to predict output vectors, using the highest hidden layer as facial features. This layer, trained to distinguish a large number of face categories, contains rich inter-class variation information and possesses strong generalization capabilities.

In June 2014, Tang Xiaoou led the Computer Vision Research Group at the Chinese University of Hong Kong to develop a deep learning face recognition algorithm called DeepID [12,13], achieving 99.15% recognition accuracy on the LFW database. Prior to this, Tang Xiaoou's research group had developed GaussianFace, a face recognition technology based on Gaussian processes, achieving 98.52% recognition accuracy. This marked the first time that a computer automatic recognition algorithm surpassed human eye recognition accuracy.

In 2015, Google proposed the FaceNet [14] deep learning face recognition algorithm. Unlike other deep learning methods applied to faces, FaceNet did not use traditional Softmax for classification learning and then extract a certain layer as features. Instead, it directly performed end-to-end learning of an encoding method from image to Euclidean space, outputting face feature vectors for subsequent face recognition, verification, and clustering.

In 2017, Liu et al. [15] proposed the SphereFace deep learning face recognition algorithm, which introduced normalized weights and angular margin to improve the traditional Softmax, thereby achieving the recognition criterion where the maximum intra-class distance is smaller than the minimum inter-class distance. This algorithm demonstrates high recognition accuracy and has maintained top verification accuracy on small datasets (with less than 500,000 training samples).

In 2018, Tencent proposed the CosFace [16] deep learning face recognition algorithm based on large margin cosine loss. This algorithm optimized the loss function for face recognition by using cosine distance as the loss function. Cosine distance focuses more on directional differences, demonstrating strong generalization capability in face recognition. The algorithm performed excellently, achieving high recognition accuracy not only on LFW but also improving the MegaFace: million-scale face recognition dataset accuracy to 98%, surpassing the 91% record held by Russia' s Vocord company.

The main steps of face recognition are face detection, face cropping, feature extraction, feature comparison to obtain cosine distance, and threshold-based determination of whether they are the same person. The flow chart of face recognition algorithm implementation is shown in Figure 1. The face recognition application in this paper is deployed on home service robots. Since home environments are extremely complex and susceptible to lighting, angle, and face size variations, traditional face recognition algorithms such as LBP face recognition and FisherFace exhibit poor robustness, resulting in poor recognition performance and severe false recognition in home environments, making them unsuitable for home service robot applications.

Deep learning convolutional neural network face recognition [17] demonstrates excellent robustness and adapts to multi-scenario recognition, effectively reducing recognition accuracy degradation caused by lighting and angle variations. However, the primary metric for measuring face recognition accuracy is LFW accuracy. While deep learning face recognition achieves high accuracy on LFW, testing face recognition models on service robots shows significantly decreased accuracy that cannot fully meet existing requirements. The main reason is that face images in real-world environments exhibit far greater complexity in angle, lighting, and resolution than the LFW dataset.

Addressing these issues and considering the specific application scenarios of home service robots, this paper proposes an improved Inception-ResNet-V1 network structure for face feature vector extraction, while completing Asian face dataset training, hyperparameter tuning, and related work. Compared with FaceNet, our experimental results outperform the original network structure on both the LFW dataset and the actual home environment face dataset collected by our robot.

1 Algorithm Implementation

Inception-ResNet-V1 Network Structure

The Inception-ResNet-V1 network structure is shown in Figure 2 [Figure 2: see original paper]. This network combines Inception and ResNet architectures. The pooling layer performs downsampling on the input, reducing output data dimensions without losing significant features extracted by the network. The Dropout layer is a deep learning regularization method that reduces overfitting. The Softmax classifier performs multi-class classification by mapping the output of multiple neurons to the interval (0,1).

Using identical convolutional kernels in the network structure leads to single-scale features, loss of partial useful features, and weakened feature extraction capability. Employing different convolutional kernels can obtain receptive fields of various sizes for multi-scale feature fusion, improving model generalization ability.

The Inception network stacks 1×1, 3×3, and 5×5 convolutional kernels with 3×3 pooling layers, increasing both network width and scale adaptability. The Inception network structure is shown in Figure 3 [Figure 3: see original paper].

GoogLeNet [19] is the representative work of Inception, demonstrating that sufficient network depth is a prerequisite for good model performance. However, beyond a certain depth, simple network stacking degrades performance. Moreover, as network depth increases, computational parameters increase, leading to decreased network performance. At a certain point, deeper networks result in higher training errors, primarily because deeper networks exhibit more pronounced gradient vanishing.

The ResNet [18] network, proposed in 2015 and winner of the ImageNet classification task, was designed to effectively solve gradient vanishing when increasing network depth. The core structure of ResNet is the residual network, shown in Figure 4 [Figure 4: see original paper].

The combination of Inception and ResNet networks [18] includes a 6-layer convolutional stem network, 5-layer Inception-ResNet-A, Reduction-A networks, 10-layer Inception-ResNet-B, Reduction-B networks, 5-layer Inception-ResNet-C networks, Average Pooling layer, Dropout layer, and Softmax classifier. This network structure combines the advantages of Inception and ResNet, increasing network width for stronger adaptability and generalization while deepening the network to enhance feature extraction capability and classification accuracy.

To enhance the network' s nonlinear mapping capability while limiting network size, a fully connected layer is added after feature extraction. Each neuron in this layer connects to all neurons in the previous layer, with no connections between neurons in the same layer. The formula is:

$$o_j^l = f\left(\sum_{i=1}^{n} w_{ji}^l x_i^{l-1} + b_j^l\right)$$

where n is the number of neurons in the previous layer, l represents the current layer, w_{ji}^l is the connection strength between neuron j in the current layer and neuron i in the previous layer, b_j^l is the bias term, and $f(\cdot)$ represents the activation function.

Improved Inception-ResNet-V1 Network Structure

This paper utilizes the CASIA-WebFace dataset, which contains 10,575 IDs (all face images of the same person stored in the same folder, referred to as an ID) and 494,414 images. This dataset suffers from insufficient face images, uneven distribution, and noise. Using FaceNet' s official network structure achieves 99.05% accuracy on LFW. Through research and testing on this dataset, we found that network structure significantly impacts final recognition accuracy. Therefore, this paper adds an improved Inception network module under the Stem network to increase network depth and width for obtaining more salient features, as shown in Figure 5 [Figure 5: see original paper]. Additionally, we modify the layer numbers and Inception sub-network weight parameters in the three sub-networks Inception-ResNet-A, Inception-ResNet-B, and Inception-ResNet-C, increasing Inception-ResNet sub-network layers to 20 to extract more effective feature data. A 1×1 convolutional dimensionality reduction layer is added to the Inception-ResNet-C output layer, which performs cross-channel aggregation to effectively reduce parameters, improve computational speed, and reduce model size while maintaining accuracy.

Our network structure is used for face feature extraction, which is the critical step determining face recognition accuracy. Inaccurate feature extraction leads to biased results when calculating Euclidean or cosine distances. This paper employs cosine distance as the feature comparison algorithm. Euclidean distance calculates the absolute distance between two vectors, where closer to 0 indicates greater similarity. However, two different images of the same person may have large Euclidean distances due to lighting and angle factors, causing the system to incorrectly determine they are different people. The Euclidean distance formula for points in n-dimensional space is:

$$d = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$$

Cosine distance uses the cosine of the angle between two vectors in vector space to measure differences between individuals, focusing more on directional differences than distance or length. In experiments, using cosine distance as the feature comparison algorithm yields better accuracy than Euclidean distance under identical test datasets, parameters, and network structures. The cosine distance formula for points in n-dimensional space is:

$$\cos\theta = \frac{\sum_{k=1}^{n} x_k y_k}{\sqrt{\sum_{k=1}^{n} x_k^2}\sqrt{\sum_{k=1}^{n} y_k^2}}$$

2 Experimental Setup and Results

2.1 Experimental Platform

Figure 6 [Figure 6: see original paper] shows the Sun@Home robot from Beijing Information Science & Technology University' s home service robot team. The experimental platform consists primarily of a Kinect-v2 camera [20], adjustable lifting mechanism, 3-degree-of-freedom robotic arm, omnidirectional wheel chassis, and a LiDAR with a 270° detection range. This experiment uses the Kinect-v2 camera sensor. Compared to the first-generation Kinect-v1 released in 2010, the second-generation Kinect sensor can obtain maximum camera resolution of 1920×1080, higher than Kinect-v1' s 640×480; supports up to 30 fps frame rate; has a detection range of 0.5-4.5 m; and detection angles of 70° horizontal and 60° vertical. With superior performance metrics, Kinect-v2 was selected as the image information acquisition device for the home service robot. By obtaining RGB image data from Kinect-v2 and processing it through image preprocessing, face detection, face cropping, feature extraction, feature comparison, and threshold determination, the system ultimately determines facial position information and corresponding person IDs in images.

2.2 Asian Face Dataset Creation

Currently, most publicly available face datasets are foreign, such as VGGFace2 and CASIA-WebFace. Although network models trained on these datasets achieve good recognition rates on the LFW test set, they perform very poorly on Asian face datasets or in actual home service robot environments. Therefore, establishing an Asian face dataset is crucial.

In May 2018, Glint opened 93,000 IDs with 2.8 million Asian celebrity images, currently the largest open Asian face dataset, providing significant help to academia and industry. However, statistical analysis revealed this dataset consists entirely of online celebrity data lacking real-world environment face images, with extremely unbalanced sample distribution—the most frequent ID contains over 3,000 face images while the least frequent has only 2. Such imbalance affects final accuracy to some extent. Based on the Asian celebrity dataset and combined with face-containing images collected using home service robots, this paper constructs a multi-scenario, multi-angle face dataset. Sample balancing processing was applied to the self-built dataset by deleting IDs with insufficient face images and augmenting IDs with fewer images through data enhancement algorithms including noise addition. The final dataset contains 100,000 IDs with 3.5 million images, referred to as OurDatas. Face image acquisition is shown in Figure 7 [Figure 7: see original paper].

2.3 Experimental Results

To verify the proposed algorithm, we trained on the CASIA-WebFace dataset and OurDatas dataset using Ubuntu 16.04, GTX1080 GPU, and the TensorFlow deep learning framework. Training parameters are shown in Table 1 .

Table 1 Training Parameters

```
--image_size 160
--model Inception_ResNet_V1
--optimizer ADAM
--batch_size 90
--keep_probability 0.4
--random_flip
```

Where: `--image_size` is the normalized face image size—after face localization, cropped face images are normalized to 160×160 pixels; `--model Inception_ResNet_V1` is the network feature extraction structure; `--optimizer` is the optimizer, with ADAM selected for its ability to iteratively update neural network weights based on training data; `--batch_size` processes 90 face images per batch; `--keep_probability` is the dropout parameter to reduce overfitting, set to 0.4 in this paper, meaning 40% of neuron parameters are frozen during each iteration; `--random_flip` is an image preprocessing operation.

2.3.1 Comparison Results on General Test Sets Our test sets include the LFW dataset (containing 5,749 IDs and 13,233 face-containing images) and data collected by home service robots and related competition scenes (containing 2,400 IDs and 9,137 face-containing images, hereinafter referred to as OurFace).

Training sets use the public CASIA-WebFace dataset and the self-built OurDatas dataset. Recognition rate verification employs ten-fold cross-validation, which divides the validation dataset into 10 parts, using 9 parts as training data and 1 part as test data in rotation. Experimental results are shown in Tables 2 and 3.

Table 2 Comparison of CASIA-WebFace Dataset Results

Training Dataset	Test Dataset	Original Network	Improved Network
CASIA-WebFace	LFW	99.05%	99.22%

Table 3 Comparison of OurDatas Dataset Results

Training Dataset	Test Dataset	Original Network	Improved Network
OurDatas	LFW	97.78%	99.10%
OurDatas	OurFace	97.10%	99.20%

Tables 2 and 3 demonstrate that using CASIA-WebFace and OurDatas as training sets and LFW and OurFace as test sets, our experimental results consistently outperform the original network, with particularly significant improvements on the Asian face test set.

2.3.2 RoboCup Robot World Cup On-Site Test Results Figure 8 [Figure 8: see original paper] shows the on-site test at the 2018 RoboCup Robot World Cup China Competition. The competition includes a home service robot multi-person identification project designed to test home service robots' ability to autonomously identify target persons in unfamiliar environments, marking target persons' positions and IDs as well as strangers' position information. The robot memorizes a designated target person through voice interaction and Kinect v2, after which the target person enters a small crowd of 5-10 people with various postures (sitting or standing). The robot must locate the crowd and accurately identify both target persons and strangers.

As shown in Figure 8, our algorithm can accurately identify target persons (marked as Operator) and others (marked as strangers). At the RoboCup Robot World Cup competition site, due to lighting, angle, face resolution, and other factors, FaceNet' s official model and traditional algorithms exhibited very high false recognition rates, mistakenly identifying strangers as target persons and causing recognition errors. Our network structure model was trained with a multi-angle, multi-lighting dataset and incorporated face alignment processing, effectively reducing lighting and angle impacts and improving accuracy. The outstanding performance at RoboCup fully validates the reliability of our algorithm. This paper collected on-site face recognition test sets from the RoboCup Robot World Cup over three years, with experimental comparison results shown in Table 4 .

Table 4 Statistical Results of Recognition Rate

Algorithm	Recognition Rate
Traditional Algorithm	60.0%
Original Network	73.3%
Improved Network	96.7%

2.3.3 Different Face False Recognition Test Although the official original network structure achieves good accuracy on the LFW test set, it performs poorly in actual testing on Chinese faces. Feature vectors of different faces are very similar, failing to achieve small intra-class distances and large inter-class distances, frequently misidentifying different people as the same person. The improved network structure and dataset proposed in this paper can effectively distinguish between the same and different persons. The same person yields high cosine distance scores (normalized to 0-1, with values closer to 1 indicating greater similarity), while different persons yield low scores. Through matching test datasets, a reasonable distance threshold was determined. This paper sets the cosine distance threshold at 0.8, with values above 0.8 judged as the same person and below 0.8 as different persons. Figures 10 [Figure 10: see original paper] and 11 contain test images of two different persons. Our experimental results accurately identify them as different persons, while the official model incorrectly judges them as the same person, causing comparison errors. Experimental results are shown in Figures 9 [Figure 9: see original paper] and 10.

Additionally, this paper tested 350 groups of different faces to evaluate the false recognition rate of misidentifying different faces as the same person, with experimental results shown in Table 5 .

Table 5 Statistical Results of False Recognition Rate

Algorithm	Test Set Size	False Recognition Count	False Recognition Rate
Original Network	350	43	12.28%
Improved Network	350	12	3.43%

3 Conclusion

This paper proposes an improved Inception-ResNet-V1 face feature extraction network structure and creates Asian face training and test sets. Experimental results demonstrate that the improved network structure, combined with Asian face dataset training and hyperparameter tuning, has been fully validated on the LFW test set, OurFace test set, and actual home service robot environments, with results consistently outperforming the original FaceNet network. The improved network structure has been applied to Beijing Information Science & Technology University' s Sun@Home home service robot, achieving excellent results in multiple domestic service robot competitions and fully validating the effectiveness of our face recognition algorithm.

References

[1] Xu Yuanyuan, Kong Lingfu, Gao Shengnan. Study on Chinese instruction parser issues for home service robot [J/OL]. Application Research of Computers, 2019, 36(1): [2018-01-10].

[2] Xiao Bing, Wang Yinghui. Survey of human face recognition [J]. Application Research of Computers, 2005, 22(8): 1-5.

[3] Hu Xiaorong. Research and implementation of family service robot target recognition and face recognition in complex scene [D]. Hefei: Hefei Polytechnic University, 2011.

[4] Fang Chen. Intelligent robot "Jiajia" [J]. Scientific World, 2013, 16(11): 50-55.

[5] Zhang Qizhi, Zhou Yali. Sun@Home home service robot, Beijing University of Information Technology [J]. Robot Technology and Application, 2014, 36(3): 49-50.

[6] Chan H, Bledose W. A man-machine facial recognition system: some preliminary result [R]. California: Palo Alto, Panoramic Research Inc, 1965.

[7] Turk M, Pentland A. Eigenfaces for recognition [J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71-86.

[8] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2002, 19(7): 711-720.

[9] Peng Hui, Zhang Chang. Automatic face recognition based on K-L transform [J]. Journal of Tsinghua University: Natural Science Edition, 1997, 37(3): 67-70.

[10] Lu Xiaoguang, Jain A K, Colbry D. Matching 2.5D face scans to 3D Models [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2005, 28(1): 31-43.

[11] Taigman Y, Yang Ming, Ranzato M, et al. DeepFace: closing the gap to human-level performance in face verification [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE Press, 2014: 1701-1708.

[12] Sun Yi, Wang Xiaogang, Tang Xiaoou. Deep learning representation from predicting 10,000 classes [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE Press, 2014: 1891-1898.

[13] Sun Yi, Chen Yuheng, Wang Xiaogang, et al. Deep learning face representation by joint identification-verification [C]//Proc of International Conference on Neural Information Processing Systems. [S.l.]: MIT Press, 2014: 1988-1996.

[14] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE Press, 2015: 815-823.

[15] Liu Weiyang, Wen Yandong, Yu Zhiding, et al. SphereFace: deep hypersphere embedding for face recognition [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE Press, 2017: 6738-6746.

[16] Wang Hao, Wang Yitong, Zhou Zheng, et al. CosFace: large margin cosine loss for deep face recognition [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2018: 5265-5274.

[17] Chen Yaodan, Wang Lianming. Face recognition based on convolutional neural network [J]. Journal of Northeast Normal University: Natural Science Edition, 2016, 48(2): 70-76.

[18] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the impact of residual connections on learning [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE Press, 2016: 4278-4284.

[19] Szegedy C, Liu Wei, Jia Yangqing, et al. Going deeper with convolutions [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE Press, 2014: 01-09.

[20] Yang Wenlu, Guo ming. A real-time face recognition system based on Kinect [J]. Computer Applications and Software, 2014, 35(5): 64-67.

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