A System for Disguised Face Recognition with Convolution Neural Networks

Kuo-Ming Hung
Department of Information Management
Kainan University
Kainan Road, Taoyuan City 33857
Taiwan
hkming@mail.knu.edu.tw

Jin-An Wu
Department of Electrical and Computer Engineering
Tamkang University
Tamsui Dist., New Taipei City 25137
Taiwan
wl00262188@gmail.com

Chia-Hung Wen, Li-Ming Chen
Department of Electrical and Computer Engineering
Tamkang University
Tamsui Dist., New Taipei City 25137
Taiwan
zjrnbckw1@gmail.com
ms0071799@hotmail.com

ABSTRACT
Face recognition technology has been quite advanced in recent years and has been applied to various daily necessities and applications. However, people may make a false positive feature of the masked camouflage face because of makeup or wearing different equipment. In this paper, a two-stage disguise face recognition method based on CNN is proposed for the disguised face wearing equipment. In the first stage, we train a network that identifies the type of equipment and extracts the remaining faces that are not disguised. In the second stage of identification, the extracted remaining faces use the identified network for identity identification. The experimental results show that the proposed method has reached an average of 97.6% accuracy in the first stage of equipment type recognition. In the second stage of disguise face identification, 72.4% identification rate was obtained. The proposed method in this paper has reached the identification rate of the disguise identification research in recent years. The results of the above two stages show that the proposed method can effectively identify the type of disguise worn when people wear disguise. Then, the facial information of the disguise is removed to achieve a certain identity recognition effect.

CCS Concepts
• **Computing methodologies → Artificial intelligence → Computer vision → Computer vision problems → Object identification**

Keywords
Convolutional neural network; Disguised face recognition; Disguised type detection; Deep learning

1. INTRODUCTION
Face recognition has been attracting lot of people to study for many years. Thanks to the continuous improvement of the calculation methods and techniques of image processing, the computing speed of computers has been continuously improved. Many applications and technologies related to face images have been developed. These technologies have produced many related applications in different fields [1]-[4], such as identity recognition, face tracking, expression recognition, 3D cartoon face making, video games and movie special effects production. As Apple Computer introduced the Face ID function into the next-generation mobile phone iPhone X, all mobile phone brand manufacturers have also adopted the function of face recognition into their new phone model. Research have also begun to study and challenge various face recognition, such as twins, different skin colors and disguise of head-mounted items. At present, the face recognition technology has a very high recognition rate. There are many ways to extract face features [3], [5], [6], such as 3D depth features, geometric features, thermal imaging training model features, etc. However, the disguised face masked or deviated due to the disguise causes the recognition rate to decrease. Face disguise generally falls into two categories. One is the use of items (such as wigs, sunglasses, hats, masks, etc.) to cover part of the face area. Such disguise can reduce the number of facial features. The other type is disguise using facial makeup (such as military disguise makeup, etc.). Disguised Face Identification (DFI) is a challenging problem because it involves intentional or unintentional facial feature changes and shadows. Machine vision is similar to human beings in judging the face of a disguised face. As disguise increases the difficulty of identification, it is easy to confuse or misjudge the identity of the pretender [7].

There are many ways to identify disguise faces. Tejas I. Dhamecha et al. [5], [6] proposed the use of both face images and infrared image. Make the face part into features and disguise two parts, and then find the features. The experimental results have a good recognition rate in various disguise types. Billy Y.L. Li et al. [3] used Microsoft's Kinect sensor to obtain face depth data. Their method was combined with facial shape and texture data. The sparse representation classifier is used to repair data errors and data loss, and has a good recognition rate for different expressions, illuminations, and sunglasses disguise. In this paper, we train and achieve the identification of pretenders using the convolutional neural network described above. This paper studies the above-mentioned Convolutional Neural Network to train and identify pretenders.

2. RELATED WORK
2.1 Deep Normalization and Convolutional Neural Network (DNCNN)
In 2017, Z. Yin et al. proposed a new Deep Normalization and Convolutional Neural Network (DNCNN) [8]. DNCNN is used to detect smoke images, which is different from the traditional way of detecting smoke. This method can simultaneously extract the

DMIP '18, November 12–14, 2018, Okinawa, Japan

ACM ISBN 978-1-4503-6578-9/18/11…$15.00
DOI: https://doi.org/10.1145/3299852.3299858

features of smoke images and identify smoke. In DNCNN, a deep neural network is specially constructed to extract features directly from the original pixels of smoke and non-smog images. From the experimental results, the DNCNN method was used on the dataset of the smoke image to achieve an error rate of less than 0.6% and an identification rate of up to 96.37%.

DNCNN proposes a 14-layer neural network architecture as shown in Figure 1. The network architecture consists of eight normalization and convolution layers and three pooling layers and three fully connected layers.

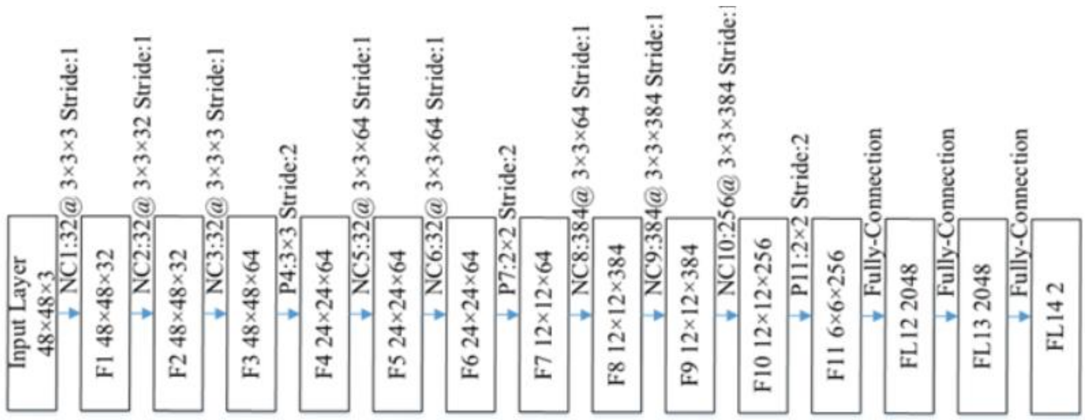


Figure 1. DNCNN Network

In order to reduce the problem of overfitting, DNCNN used image processing to increase the number of training samples, flip the original image horizontally, flip it vertically and rotate it by 90 degrees.

2.2 Viola-Jones Algorithm

In 2001, Paul Viola and Michael Jones published a framework called Robust Real-time Object Detection. Through this method, face recognition can achieve rapid and high detection rate.

This framework consists of three parts. The first part is Integral Image, which allows image features to be quickly calculated.

In the second part, the AdaBoost algorithm is used to find important features, including large majority of the available features and small set of critical features, and then produce extremely efficient classifiers.

The third part is the method of cascading multiple classifiers. This method allows the background area of the image to be quickly ignored. The experimental results confirmed that the face recognition system has better performance than the old methods. In this paper, we use this method to fulfill face position detection of the original image.

3. PROPOSED METHOD

3.1 Network Architecture

The architecture of the neural network has an important influence on the effectiveness of the identification system. In this paper, we construct a neural network that is identical to the reference mentioned in Section 2.3 [8]. The network uses a 14-layer Deep Normalized Convolutional Neural Network (DNCNN). The goal of the research and experiment in this paper is using DNCNN's network architecture to fulfill the disguise classification and the identity recognition application of disguised face.

The architecture of the neural network has an important influence on the effectiveness of the identification system. In this paper, we construct a neural network that is identical to the reference mentioned in Section 2.3 [8]. The network uses a 14-layer Deep Normalized Convolutional Neural Network (DNCNN). The goal of the research and experiment in this paper is using DNCNN's network architecture to fulfill the disguise classification and the identity recognition application of disguised face.

3.2 Identification System Process

This paper is a disguised face recognition system based on Deep Normalized Convolutional Neural Network (DNCNN). This system needs to train two sets of DNCNN identification networks. The function of the first group of disguise classification network is to identify the type of disguise that inputs the full face image. The network divides the face disguise input image into three categories. There are no disguise, disguise on the upper half and disguise on the lower half. After the recognition of the disguise classification is completed, the system removes the disguise upper half or lower half of the image. Leave no disguise half face image that was inputted to the second group identification network. Training the second set of identification network with images of half faces. Before starting the training and identification of the above two sets of DNCNN identification networks, the original image samples of the portraits need to be pre-processed. The image pre-processing uses the Viola-Jones face detection algorithm [9] to first find the block of the face position. Then take the face block image and the half face image. In order to reduce the shortage of training samples. We rotate the face center to generate face images at different angles and increase the training data of the network. The system captures a face training image every 1 degree of rotation. The rotation range is between -15 degrees and +15 degrees, and 31 full-face images have been generated. Refer to Figure 2 and Figure 3. The face image is divided into an upper half face and a lower half face. According to different identities, we classifiy the face images. The identification system flow diagram of this paper is shown in Figure 4 and Figure 5.

Figure 2. face training image after rotation

Figure 3. half face training images after rotation in different degrees

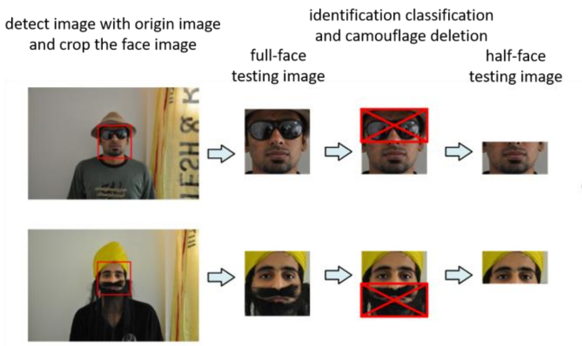
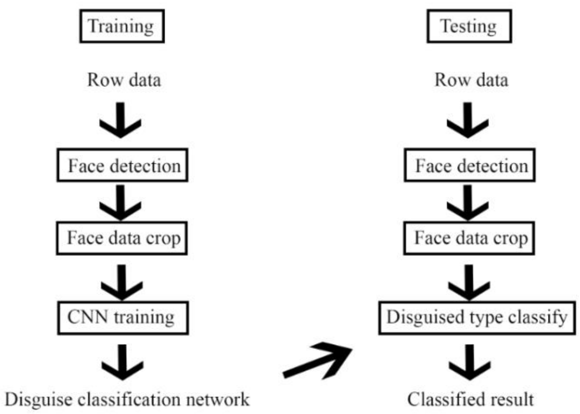
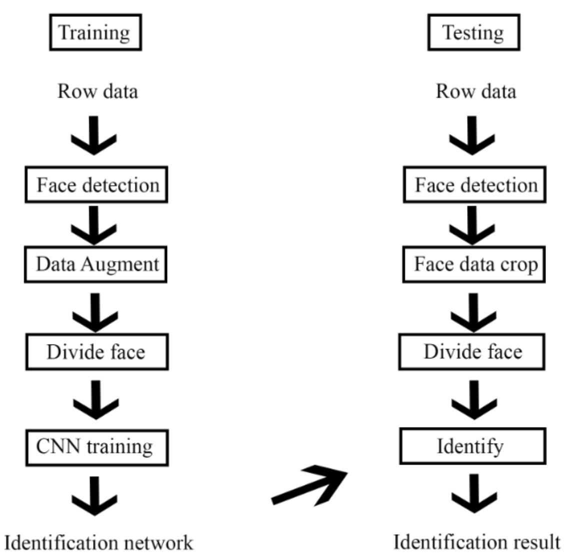


Figure 4. Full-face and half-face generation process



(a) Equipment Type Recognition Net Training and Testing



(b) Identification Net Training and Testing

Figure 5. The Block Diagram of Our Proposed System

The first step of the experiment is to collect data, which is divided into non-disguised and disguised face images of people who need to be identified. According to different people, they were classified by different folders.

The original image has a size 3264×1836, and each original image will be judged by the Viola–Jones face detection algorithm. The original image has a resolution of 3264×1836, and each original image detect the face position using the Viola–Jones method. The red line in Figure 2 is the coordinate of the output face area detected by the Viola–Jones method. We adjust the image resolution to match the network parameters (192×192, 128×128, 64×64). All face image data is pre-processed for full face and half face. It includes face disguise training data (full face), no disguise training data (half face) and untrained face disguise test data (full face), which are used as raw data for training and testing.

3.3 Training Network

After the above training images were pre-processed and classified, we can input all the full-face and half-face images that need to be trained into two sets of 14-layer DNCNN networks to start training. In this paper, we need to train two networks. The first is the disguise classification network, and the other is the identification network. Two DNCNN networks are the network architectures mentioned in Section 2.3 of this paper. Their network configuration and parameter settings are shown in Figure 1. In the Figure 1, the eight layers F1, F2, F3, F5, F6, F8, F9, and F10 are Normalization and Convolutiono layers (NC). The three layers F4, F7, and F11 are the Max-Pooling layer. The three layers FL12, FL13 and FL14 are Fully-Connected layers (FC).

Take NC8:384@3×3×64 in the Figure 1 as an example. NC8 represents the 8th layer of Normalization and Convolution layer. 384@3×3×64 represents 384 filters with 3×3×64 size in this NC layer. P7: 2×2 means that the 7th layer is the Max-Pooling layer with 2×2 mask size.FL12:2048 represents 2048 neurons in the 12th layer of the fully connected layer. In output phase, we modify the final output of full connection layer with 3 and 20 classification output results in equipment type recognition net and identification net, individual. When all the training images are entered into the DNCNN network, the training is completed. The closer the accuracy is to 100%, the better the LOSS curve is closer to 0. Then, we can use this trained network to identify the identity of the disguise.

3.4 Test Identification Network

When the network training is finished, we can start testing to the system. The process of testing the network is the same as when training the network. The system uses the same resolution and untrained disguise face images to be input into the first set of disguise classification networks. Test the data with the principles of Open Test. The system will recognize whether the half face is disguised, leaving only the half face without disguise. Then input the result of first stage into identification network of the second stage and output the identification result.

4. EXPERIMENTAL RESULT

4.1 Experimental Environment

The experimental development environment and program used in this paper are described as below. The central processing unit (CPU) used for the computer hardware is Intel's CORE(TM) i7-4790 CPU @ 3.60GHz dual-core CPU and NVIDIA GeForce GTX 970 with memory 8G. In the part of the software, the operating system of the computer used is Microsoft Windows 8.1

Enterprise Edition (64-bit). Development of experimental programs using Matlab version R2018a (64-bit) pre-processing image database and training convolutional neural network (CNN).

4.2 Database

The database used in this paper is IIIT-Delhi Disguise Version 1 face database (ID V1) [10]. This database contains 75 different people, both male and female. Each subject has 1 or 2 non-disguised color photos, and 6 to 9 disguised color photos. The participants being photographed can randomly choose their own props on disguise at ple, so the portrait of disguise types of each object are not exactly the same. The database has a total of 682 portrait images as Figure 6.

Figure 6. portrait image in database

4.3 Experimental Result

This identification system requires training two sets of DNCNN networks. The function of the first set of networks is to identify the type of the input face image's disguise. The network divides the face disguise input images into three categories: no disguise, upper-face disguise (wearing sunglasses or glasses), and lower-face disguise (wearing a mask or fake beard). The recognition rate of the disguise classification network can reach 97.59%, as shown in Table 1.

Table 1. Identification Rate of Disguise Classification

Categories	Number of testing image	Number of correct Rrcognition	Recognition rate (%)
upper-face	22	20	90.9 %
lower-face	24	24	100.0 %
no disguise	37	37	100.0 %
Sum / average rate	83	81	97.6 %

The second group of networks can identify the identity of the person. We remove the half face of the classification result with the disguise and input the remaining half face into the second group of networks to identify the person. We used three different disguise facial image sizes (96×96, 128×128, 192×192) and two different numbers of fully connected layer neuron settings (1024, 2048) to do multiple sets of experimental comparisons. The experimental results show that the recognition rate when the image resolution is 96×96 and the number of fully-connected layer (FC) neurons is set to 1024 is 72.4%.

According to the disguise identification result of [11], the accuracy is 70%. In this paper, result we got was as Table 2 achieving 72.4% recognition rate.

Table 2. Identification rate of disguised identity

Dataset / Method	Cap + Mask / Cap + Glasses / Cap + Scarf identification rate (%)
Proposed 96 x 96 FC1024	72.4 %
Proposed 96 x 96 FC2048	62.1 %
Proposed 128 x 128 FC1024	68.9 %
Proposed 128 x 128 FC2048	62.1 %
Proposed 192 x 192 FC1024	48.3 %
Proposed 192 x 192 FC2048	44.8 %

We also used the Cumulative Match Characteristic Curve, which is often used when evaluating the identification system, to compare the results of this experiment. The main difference between them is the image resolution and the number of neurons in the fully connected layer, as shown in Figure 7.

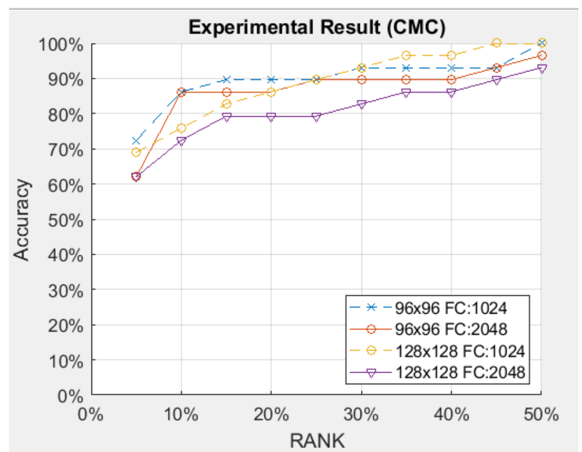


Figure 7. CMC curve of experiment result

5. CONCLUSION

This paper applies the Deep Normalized and Convolutional Neural Network (DNCNN) to disguise face recognition. We reach a level similar to [11]. The results of this experiment, in the work of classification of disguise, can obtain a recognition rate at 97.6%. Identification of the identity wearing a disguise shield can achieve 72.4% recognition rate.

In the future, we can use a database with more adequate image data to increase the robustness of the identification system. Research related topics such as side face recognition and face recognition from different angles also can be studied.

6. REFERENCES

[1] B. B. (Intel), Archived - code sample: facial recognition using Intel® RealSense™ SDK, Intel, 16 March 2015. [Online]. Available: https://software.intel.com/en-us/articles/code-sample-facial-recognition-using-intel-realsense-sdk.

[2] Zhang, X. Q. and Ren, D. A survey of approaches to feature extraction in face recognition. *International Conference on Computer Science and Network Technology*, 2012.

[3] Li, Billy Y. L., Mian, A. S. Using kinect for face recognition under varying poses, expressions, illumination and disguise. *Applications of Computer Vision* (WACV), vol. 7, no. 5, pp. 1498-1510, 2013.

[4] Sheng, G. and Kai, W. (Intel), Archived - Intel® RealSense™ Archived - Intel® RealSense SDK-based real-time face tracking and animation, Intel, 14 11 2016. [Online]. Available: https://software.intel.com/en-us/articles/intel-realsense-sdk-based-real-time-face-tracking-and-animation.

[5] Dhamecha, T. I., Nigam, A., Singh, R., and Vatsa, M. Disguise detection and face recognition in visible and thermal spectrums, P*roceedings of International Conference on Biometrics*, 2013.

[6] Dhamecha, T. I., Singh, R., Vatsa, M., and Kumar, A. Recognizing disguised faces: human and machine evaluation, PLoS ONE, 2014.

[7] Mahmood, S., Sruti, B., Lujo, B., and Michael, K. R. Adversarial generative nets: neural network attacks on state-of-the-art face recognition, arXiv: 1801.00349v1, 31 Dec. 2017.

[8] Yin, Z. J., Wan, B. Y., Yuan, F. N., Xia, X., and Shi, J. Y. A deep normalization and convolutional neural network for image smoke detection, *Digital Object Identifier 10.1109/ IEEE ACCESS*. 2017.2747399, 2017.

[9] Viola, P. and Jones, M. Robust real time object detection, in Second International Workshop On Statistical and Computational Theories Of Vision, Vancouver, Canada, 2001.

[10] Goswami, G., Mayank, V., and Richa, S., RGB-D face recognition with texture and attribute features, *IEEE Trans. Information Forensics and Security*, pp. 1629-1640, 9 10 2014.

[11] Singh, A., Patil, D., Reddy, G. M., and Omkar, S. N. Disguised face identification (dfi) with facial keypoints using, *IEEE International Conference on Computer Vision Workshops*, pp. 1648-1655, 2017.