Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 162 (2019) 932–939

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Information Technology and Quantitative Management (ITQM 2019)

Identity authentication on mobile devices using face verification and ID image recognition

Xing Wu[a,b,*], Jianxing Xu[a], Jianjia Wang[a], Yufeng Li[a], Weimin Li[a], Yike Guo[a]

[a]*School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China*
[b]*Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai, 200444, China*

Abstract

Identity authentication is of great importance in the digital age where ID information is commonly used in finance, insurance, transportation and other fields. Challenges of identity authentication lie in the verification of the ID card provided by the user and the information extraction from the user's ID card. To meet the challenge, an identity authentication framework is proposed, which can extract and verify personal information through face verification and ID image recognition. The identity authentication is realized by the proposed face verification model which is called Inception-ResNet Face Embedding (IRFE). IRFE uses an Inception-ResNet structure to ensure a good feature extraction aiming at accurate face verification. Moreover, a robust ID card extraction method named Morphology Transformed Feature Mapping (MTFM) is proposed to extract ID information. Experimental results demonstrate that the proposed IRFE and MTFM outperform state-of-the-art methods both in face verification and in ID extraction.

Keywords: ID Card; Object Detection; Face Verification; Text Extration; Biometrics

1. Introduction

Mobile devices are widely used to perform remote operations benefited from their portability [1]. Mobile applications, particularly mobile payments, are indispensable for our daily lives. Meanwhile, the phenomenon of identity theft (unauthorized use of other people's identification information or confirmed fraud) has grown very rapidly in many countries. In many interactive scenarios, especially financial scenarios, the user's identity information needs to be authenticated in case of identity theft [2].

* Corresponding author. Tel.: +86-21-66135538; fax: +86-21-66135273.
E-mail address: xingwu@shu.edu.cn.

10.1016/j.procs.2019.12.070

The question that naturally follows is that can we automatically verify the user's identity information to avoid identity theft. To achieve this goal, we need to extract the identity information from the photo taken by the user, and to verify that the ID information belongs to the user. Every Chinese citizen holds an ID card that contains the holder's identity information and his face portrait which is shown in Fig. 1(a).

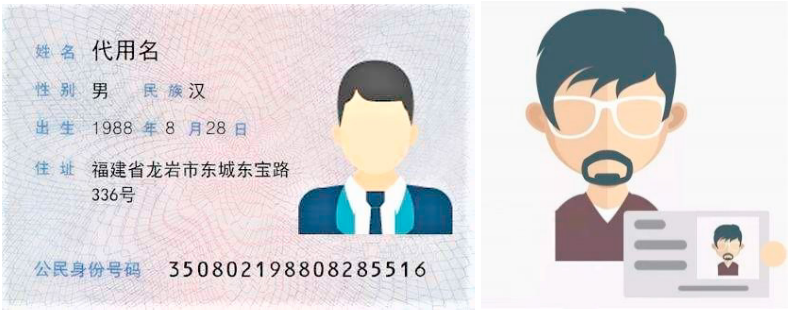

Fig. 1. (a) The sample of citizens' ID card; (b) The sample photo that should be taken by the user on mobile devices for verification

According to the design of Chinese ID cards, a unified framework is developed that can perform personal identity authentication automatically with face verification and ID card recognition. In particular, the framework only needs one photo taken by the user as input and then it will return the recognized information and the face verification result as output. The user is supposed to use a mobile device to take a photo with his hand-held ID card. The sample photo is shown in Fig.1(b). Once this photo is uploaded to the proposed system, the identity authentication result will be achieved in seconds automatically.

The proposed IRFE model adopts the novel face feature embedding structure with a strong feature representation for face verification. The proposed MTFM method is able to extract the ID information with high accuracy. Thus IRFE and MTFM are combined to achieve promising results in identity verification.

The remaining of this paper is organized as follows. In Section 2, the related works are discussed. In Section 3, a detailed description of the proposed authentication framework is presented. In Section 4, the implementation details and experimental results are demonstrated. Section 5 is the conclusion of our work.

2. Related works

Identity information extraction refers to reading the text in ID card images. It belongs to the research field of text detection and recognition, which have been studied over the last few decades [3, 4, 5]. Current researches focus on the standard horizontal direction of the ID card [6, 7, 8, 9], and the ID card area occupies the entire image. Thus most detection methods are based on horizontal projection and vertical projection [4, 10]. There are no previous researches for skewed ID image recognition within natural backgrounds. Considering the lack of ID card datasets, an image morphology-based method is proposed to detect texts on ID cards. The task of Chinese character recognition is achieved by the deep learning method based on Convolutional Neural Networks (CNNs). The ResNet [11] is a novel architecture in deep CNN models, which has made many breakthroughs in the domain of computer vision. The ResNet-50 is integrated into the character recognition procedure.

Face verification is a sub domain of face recognition. It focuses on verifying whether the two face images belong to the same person which is also known as one-to-one face verification [12]. Face recognition focus on choosing who this person is in the existing dataset. It's also called one-to-many face identification [13]. In the proposed authentication framework, face verification is indispensable. In order to ensure the claimed identity information belongs to the user, it’s necessary to verify the faces on the captured ID photo to avoid fraudulent attackers.

Modern face recognition methods [14, 15] often regard CNNs as robust feature extractors. A CNN is trained with a large face image dataset in [16] and it works as a feature extractor for the face verification. FaceNet is a unified embedding for face recognition and clustering [12], and it directly maps a face image into a compact Euclidean space where distances directly correspond to a measure of face similarity. This idea is borrowed by the proposed IRFE, in which the feature extractor is updated by replacing GoogLeNet from FaceNet with Inception-ResNet-v1 [17], making it more efficient for feature extracting. A Light CNN framework is presented in [18] to learn a compact embedding on the large-scale face data with massive noisy labels. A deep cascaded multitask framework, MTCNN, is proposed in [19] to exploit the inherent correlation in face tasks. There are pros and cons of previous researches, their strengths are adopted and weaknesses are ignored to develop the unified identity verification framework.

3. The identity authentication framework

3.1. Framework overview

In order to verify the user’s identity information before identity theft, the key task is to ensure that the identity information provided by the user belongs to him. The IRFE is implemented to realize face verification and ensure the security. The face verification part is accomplished by judging the similarity between the face of the user and the face on his ID card, which is shown in Fig.1(b). Afterwards the texts on the ID card are extracted. To deal with a skewed ID card in the photo, a skew correction method based on MTCNN is proposed. Furthermore, a text detection method based on image morphology is implemented to detect texts on ID images. Then we use the projection algorithm to segment each character from the ID card. Finally, a deep CNN model is constructed to recognize more than 3,500 Chinese characters covering most frequently used Chinese characters. With the proposed IRFE and MTFM, personal identity is verified to avoid identity theft.

3.2. Face verification phase

The face verification phase is first performed in the proposed framework to guarantee the reliability of identity information. The face detection is the prerequisite of face verification because the captured photo contains two faces as shown in Fig.1(b). Once two faced are detected, they will be compared with each other.

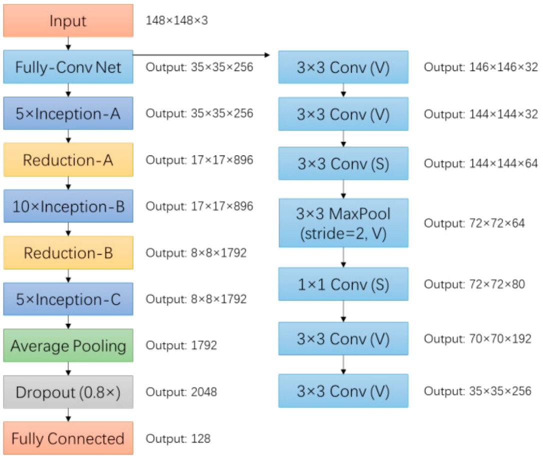

Fig. 2. The overall structure of face embedding model

Face verification can be a classification problem in judging the similarity between two faces. Thus we improve the architecture of deep convolutional neural network described in FaceNet to achieve the face feature expression. The GoogLeNet Inception structure is replaced by Inception-ResNet-v1 to build the IRFE model and generate better face features (i.e. face embeddings). Once the face embedding is obtained, the question is transformed into measuring the distance between two face embeddings. The verification result is obtained by thresholding the distance.

The details of Fully-Conv Net are shown in Fig. 2. The input image size is 148×148 with 3 channels. The inception and reduction structure are shown in Fig. 3(a) and Fig. 3(b). "5×Inception-A" means there are 5 same Inception structures in the unit. V means the "Valid" padding and S is the "Same" padding. The output shape of each layer is summarized on the right of each layer.

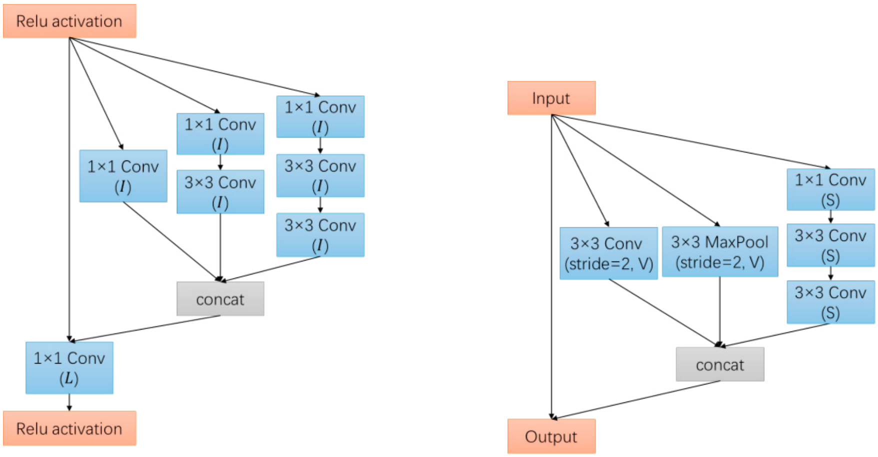

Fig. 3. (a) The Inception structure in face embedding model; (b) The Reduction structure in face embedding model

In Fig 3(a), "I" represents the different Inception section and "L" represents the output length. I = 32, L = 256 when Inception-A, I = 128, L = 896 when Inception-B and I = 192 L = 1792 when Inception-C.

Fig 3(b) shows the reduction structure where "V" denotes the "Valid" padding and "S" denotes the "Same" padding. It is a sampling unit in the model. The units are connected to construct the face embedding model.

After constructing the deep model, it is then trained on face dataset to extract face features. The 1×1 conv-layers are added between the standard convolutional layers. The dimension of output embedding is set to 1×1×128. The output of IRFE is a feature representation of the input data. Benefitting from the deep model, the face images are densely embedded into 128-dimension space, resulting in an efficient feature representation.

3.3. ID image recognition in MTFM

After the verification of the user's face, the information on his ID card should be extracted for authentication. In most cases, the photos captured by users are not horizontal, causing a low text recognition rate. This drives us to horizontally correct the ID card first before recognition. The proposed image skew correction method based on MTCNN is a very effective method to horizontally correct ID card images. The aligned face is used to calculate the card direction and locate the card area, and the horizontally corrected image can be obtained by rotating the corresponding angle to the original image.

Image morphology operation is widely used in image processing, which helps to extract image components that are meaningful for expressing the shape of region-of-interest in an image. In pace with the image morphology operation, the subsequent image recognition can obtain the most distinguishable shape features of the target object.

With the face detection of the ID card, a complete ID image without background can be obtained according to the standard configuration of Chinese ID card. Then the top hat operator (the difference value between the original image and the close operation image) is applied on the image to obtain the black foreground region. The close operation is then used to decrease the space between the words. Then the Otsu method is applied to threshold the image. By applying the close and erode operations, the text fields of ID card are then separated. Thereby text lines are detected. After the detection of text lines, in order to obtain a single character in each line of text, the projection method is used to segment characters; thereby a single character image is obtained after text segmentations.

In the character recognition phase, due to the large number of Chinese characters, some Convolutional Neural Network models are tested to achieve the best recognition task of more than 3,500 characters. We implement the ResNet-50 model to deal with the problem and it achieves the best recognition rate. The ResNet is different from traditional CNN structure owing to the shortcut connection module which is shown in Fig. 4.

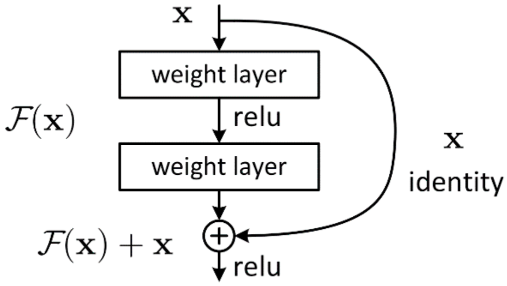
x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Fig. 4. The short connection block of ResNet

Deep CNN models are complex since they often have millions of parameters. The complexity also makes deep CNN models difficult to fit in training data. As the depth of the model increases, the parameters increase rapidly. The short connection block is designed to reduce parameters, meanwhile, increase the depth of the network. Also, the well-designed model is trainable. The results show that this idea is concise and effective.

4. Implementation and experiments of the proposed framework

4.1. Face and character datasets

Labeled Faces in the Wild (LFW) is the de-facto academic test set for face verification [20]. It is mainly used to study face recognition in unconstrained situations. The LFW database mainly collects images from the Internet, not laboratories. It contains more than 13,000 face images. Each image is identified with the name of the corresponding person. Among them, 1,680 people correspond to more than one image, that is, about 1,680 people contain more than two faces. VGGFace2 [21] is another large-scale face recognition dataset which contains more than 9,000 identities and 3.3 million faces, spanning a wide range of different ethnicities, accents, professions and ages. CASIA-WebFace [22] is also a public large dataset with about 10,000 people and more than 500,000 images. Our face embedding model is trained on this largest dataset.

For the purpose of good character recognition, a character dataset is needed. However, the ID card dataset is hard to obtain because ID cards are personal belongings. To achieve the goal, we use the method of text image generation to generate a text dataset used in the experiment. The produced dataset contains a total of 3,562 categories of character, including 3,500 Chinese characters, 52 English uppercase and lowercase characters, and 10 Arabic numerals. The dataset contains a total of 363,210 text images of single-channel data, each of which is 32×32 pixels in size. At a 5:1 ratio, the total dataset is randomly divided into a training set and a test set. At the same time, we also use data enhancement operations when making training data. Random image rotation, random cutting, perspective transformation, deformation and others are applied to each image, which greatly increases the training data. The data enhancement also makes the training set contain more image features, which makes the trained model more robust.

4.2. Face verification

In order to make the face embedding model trainable, a Softmax layer is followed after the fully connected layer. The loss function is defined as cross-entropy loss, and then the back-propagation is used to update parameters of the model. After training on the CASIA-WebFace and VGGFace2 dataset, we can get a well-trained model.

Table 1. Test accuracy results in public datasets

Network	Training Dataset	Validation rate on LFW
FaceNet	CASIA-WebFace	89.4%
IRFE	CASIA-WebFace	**96.1%**
FaceNet	VGGFace2	90.3%
IRFE	VGGFace2	**97.5%**

For more comparable tests, face verification results are tested on the open large-scale dataset LFW. The original FaceNet is implemented. The verification performance in Table 1 shows that the proposed IRFE increases the face feature representation and outperform state-of-the-art face verification methods with 97.5% validation rate.

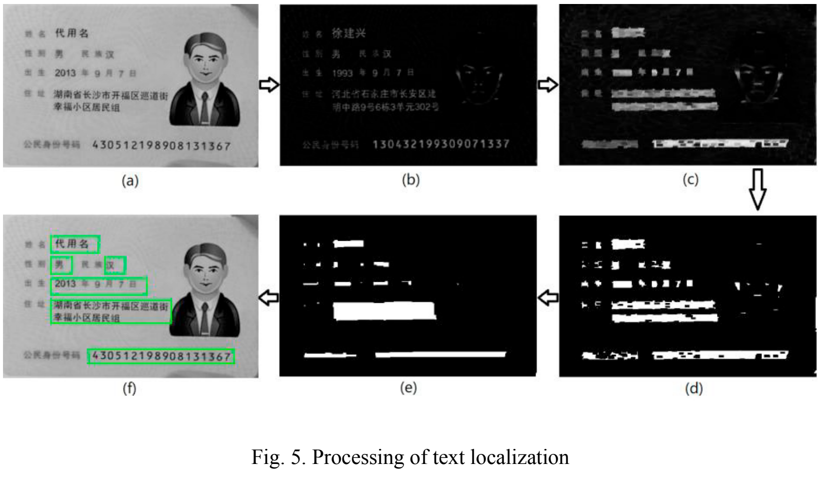

Fig. 5. Processing of text localization

4.3. Text recognition

For images that have been located in the ID area, a series of morphological transformations are implemented to detect text areas on the ID card.

Firstly, we apply the morphological black hat operation on the image to find the dark areas on a light background (Fig. 5(b)). Secondly the rectangular core is used to apply the close operation to eliminate the gap between the letters (Fig. 5(c)). Thirdly, the Otsu threshold is applied to obtain the binary image (Fig. 5(d)). Finally another close and erode operation is applied to get the final text area.

Two convolutional networks are implemented and experimental results are compared with previous methods. As shown in Table 2, a comparison of ID image recognition accuracy is listed. The recognition performance is tested in our self-collected dataset which consists of 47 ID card images.

The LeNet-5 [23] is implemented by applying ReLU activation function instead of the sigmoid function. Other implementation details of LeNet-5 are the same as described in its original paper.

Table 2. Comparison of recognition results

Name	Method	Accuracy
Fang et al. [9]	SVM	71.1%
LeNet-5	CNN	73.9%
Cheng et al. [15]	Local Similarity Voting	92.0%
MTFM	Deep CNN	**96.83%**

Previous works such as [15] can only deal with the PIN recognition because that problem is much simpler: personal identification numbers have only ten categories of characters. When facing Chinese character recognition, which has more than 3,500 categories of character, it is a hard mission. Benefiting from the development of deep learning, it can be easily solved by well-designed MTFM. The ResNet is a deep model in computer vision and it is end-to-end trainable, improving many vision problems. The ResNet-50 is applied to our scheme and achieves the best recognition result among all similar works. From Table 2, we could safely draw a conclusion that the proposed MTFM outperform state-of-the-art character recognition methods with the highest 96.83% accuracy.

5. Discussion and conclusion

To avoid identity theft, traditional ID authentication scheme has a very tedious process. To meet the challenge, an automated personal identity authentication framework is proposed, which can extract and verify identity through one photo. Usually the user should take an ID image first to provide his identity, and then the liveness detection is verified by blinking or nodding. Finally, the face image captured by the user and the ID image are compared to ensure that the identity is true. The proposed framework has reduced the interaction process considerably. The identity verification is achieved by IRFE where face embedding is enhanced by adopting Inception-ResNet. The proposed ID recognition procedure MTFM provides an accurate information extraction result to improve identity verification accuracy. The procedure can deal with complicated situations such as skewed ID image, illumination and so on. Experimental results show that the proposed framework outperforms state-of-the-art methods with high accuracy to guarantee the reliability of identity verification.

Acknowledgements

The research presented in this paper is supported by the National Key R&D Program of China (No. 2017YFE0117500) and by the Research and Development Program in Key Areas of Guangdong Province (2018B010113001).

References

[1] Mohammedi M, Omar M, Bouabdallah A. Secure and lightweight remote patient authentication scheme with biometric inputs for mobile healthcare environments[J]. Journal of Ambient Intelligence and Humanized Computing, 2018, 9(5): 1527-1539.

[2] Zaeem R N, Manoharan M, Yang Y, et al. Modeling and analysis of identity threat behaviors through text mining of identity theft stories[J]. Computers & Security, 2017, 65: 50-63.

[3] Chen D, Luettin J. A survey of text detection and recognition in images and videos[R]. IDIAP, 2000.

[4] Jung K, Kim K I, Jain A K. Text information extraction in images and video: a survey[J]. Pattern recognition, 2004, 37(5): 977-997.

[5] Uchida S. Text localization and recognition in images and video[J]. Handbook of Document Image Processing and Recognition, 2014: 843-883.

[6] LI K, CHEN L, CAO J. ID card number identification based on gray scale multi-level[J]. Computer Engineering and Applications, 2015, 2015(13): 40.

[7] Ning M . Id Card Number Identification Based on Artificial Neural Network[C]// International Conference on Robots & Intelligent System. IEEE, 2016.

[8] Ryan M, Hanafiah N. An examination of character recognition on ID card using template matching approach[J]. Procedia Computer Science, 2015, 59: 520-529.

[9] Cheng Y, Qu Y, Shi H, et al. ID numbers recognition by local similarity voting[C]//2010 IEEE International Conference on Systems, Man and Cybernetics. IEEE, 2010: 3881-3888.

[10] Fang X , Fu X , Xu X . ID card identification system based on image recognition[C]// Industrial Electronics & Applications. IEEE, 2018.

[11] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[12] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 815-823.

[13] Hu W, Huang Y, Zhang F, et al. SeqFace: Make full use of sequence information for face recognition[J]. arXiv preprint arXiv:1803.06524, 2018.

[14] Cao K, Rong Y, Li C, et al. Pose-robust face recognition via deep residual equivariant mapping[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5187-5196.

[15] Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 2892-2900.

[16] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1701-1708.

[17] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Thirty-First AAAI Conference on Artificial Intelligence. 2017.

[18] Wu X, He R, Sun Z, et al. A light cnn for deep face representation with noisy labels[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2884-2896.

[19] 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.

[20] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments[C]. 2008.

[21] Cao Q, Shen L, Xie W, et al. Vggface2: A dataset for recognising faces across pose and age[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE, 2018: 67-74.

[22] Yi D, Lei Z, Liao S, et al. Learning face representation from scratch[J]. arXiv preprint arXiv:1411.7923, 2014.

[23] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.