A Study of Machine Learning based Face Recognition for User Authentication

Chung-Pyo Hong[*†]

[*†] Chief Technology Officer, LG Electronics

ABSTRACT

According to brilliant development of smart devices, many related services are being devised. And, almost every service is designed to provide user-centric services based on personal information. In this situation, to prevent unintentional leakage of personal information is essential. Conventionally, ID and Password system is used for the user authentication. This is a convenient method, but it has a vulnerability that can cause problems due to information leakage. To overcome these problem, many methods related to face recognition is being researched. Through this paper, we investigated the trend of user authentication through biometrics and a representative model for face recognition techniques. One is DeepFace of FaceBook and another is FaceNet of Google. Each model is based on the concept of Deep Learning and Distance Metric Learning, respectively. And also, they are based on Convolutional Neural Network (CNN) model. In the future, further research is needed on the equipment configuration requirements for practical applications and ways to provide actual personalized services.

Key Words : Machine Learning, Deep Learning, Neural Network, Authentication, Face Recognition

1. Introduction

According to brilliant development of smart devices, many related services are being devised. And, almost every service is designed to provide user-centric services based on personal information. In this situation, to prevent unintentional leakage of personal information is essential. Conventionally, ID and Password system is used for the user authentication. This is a convenient method, but it has a vulnerability that can cause problems due to information leakage. Therefore, in recent years, as a part of utilizing the characteristics of the face-to-face authentication method, biometric based Multi-Factor Authentication is gradually being used [1]. These days, as machine learning technologies are being applied to the biometrics field, FAR (False Acceptance Rate) and FRR (False Rejection Rate) are reduced. Based on this circumstances, many methods of authenticating users by using biometrics alone have been introduced. Biometric recognition refers to personal identification using biometric information such as fingerprint, iris, hand vein, retina, handwriting, face, and voice. Recently, a fingerprint or face recognition is used as a biometric technology for user identification in a smart phone.

Specially, face recognition technology can be utilized in many ways, for its characteristics of easy identification without contact [2]. However, there are difficulties in authenticating the user through the face due to various factors such as light and shadow. Fig. 1. shows the examples of face images that can be obtained in a real environment [3].

This paper introduces several studies of face recognition model, specially based on machine learning techniques. In Section 2, we review the basic concept machine learning. Section 3 shows representative face recognition schemes. Finally, conclusions are presented in Section 4.

[†]E-mail: cphong0399@gmail.com

Fig. 1. Examples of face images that can be obtained in a real environment.

2. Background

In this section, we describe basic concept of machine learning techniques, and we also describe the object recognition scheme based on vision information.

2.1 Machine Learning

Machine learning is a method of operating a machine through reasoning ability based on the results of learning. In other words, when a specific data set is provided to a machine, the machine learns the related rules by itself and outputs the result of applying the rule for additional data.

Among these machine learning techniques, deep learning, which mimics human neurons and configures multiple layers of learning layers between inputs and outputs to provide more advanced results, is in the spotlight [4]. Fig. 2. presents the diagram of Deep Learning Model [5].

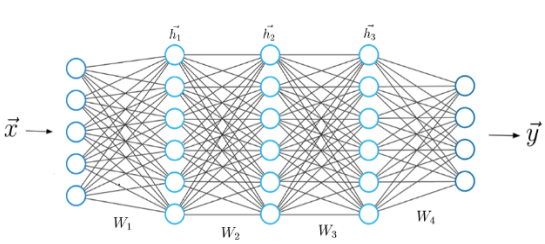

Fig. 2. A Diagram of Deep Learning Model.

Among the machine learning models based on deep learning, a convolutional neural network (CNN) is an important technique used for object recognition. The next section describes the basic concept of CNN and object recognition through it [6].

2.2 Convolutional Neural Network

The convolutional neural network (CNN) adopts a feed-forward network method that extracts a topological feature from an input image. CNN removes features from the input image and then classifies the features extracted through a classifier [6,7]. Therefore, it has strong characteristics in basic shape transformation such as scaling, transformation, squeezing, rotation, and deformation. The CNN structure consists of a convolutional layer, a pooling layer and a fully connected layer [8]. The convolutional layer, consisting of the weights of the local links and inter-functions, is a key part of CNN. In this layer, the input function expression is interpreted, and a number of functional maps are constructed. The pooling layer located in the middle of the two convolutional layers.

It has a function extraction effect and plays a role of reducing the size of the function map and increasing the robustness of the function extraction. The size of the functional map of the pooling layer is determined according to the moving stage of the kernel. The convolutional neural network classifier has one or more fully connected layers, and every single neuron in one layer is connected to every single neuron in the next layer [9]. Fig. 3. shows a Typical Convolutional Neural Network (CNN) model [10].

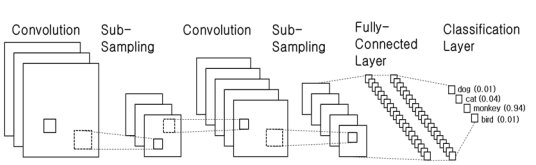

Fig. 3. A Typical Convolutional Neural Network.

3. Face Recognition Scheme

Among machine learning-based face recognition techniques, Google's FaceNet and Facebook's DeepFace are typical examples. FaceNet is based on distance metric learning, however DeepFace is based on deep learning model. We will introduce each one through the following sections. And also, the face data sets which can be used for learning phase of each face recognition model.

3.1 Deep Learning Approach

DeepFace uses CNN technology. Before providing learning data or inference data to a CNN-based network, it performs preliminary work to build accurate data. This includes 3D facial geometry-based landmark extraction and alignment through rotation (Affine). Then, learning is performed through a convolution network composed of 9 layers. DeepFace achieved a 97.35% accuracy on the Labeled Faces in the Wild (LFW) data set, leading the way in a highly successful era of face recognition [11]. However, it is a disadvantage of learning about 120 million neural network parameters. Fig. 4. shows the 3D landmark extraction and alignment steps [11].

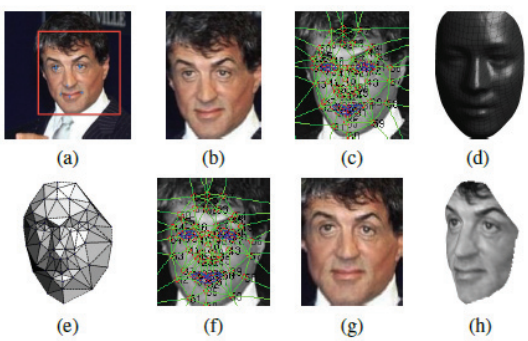

Fig. 4. 3D landmark extraction and alignment.

3.2 Distance Metric Learning Approach

FaceNet is a face recognition model based on Distance Metric Learning, as mentioned earlier. This is based on the redefinition of the loss function. The loss function utilized by FaceNet is a triplet-loss technique. This is based on the fact that the embedding value output through CNN has a smaller Mahalanobis Distance than if it is the same person [12]. The basic structure is composed of 22 layers of Deep Network. Fig. 5. is a structure of FaceNet Network Model and Fig. 6. is a conceptual diagram of Triplet-Loss mechanism [13].

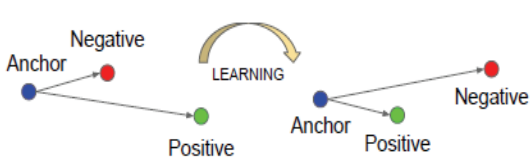

Fig. 5. Machine Learning Concept based on Triplelet-Loss.

layer	size-in	size-out	kernel	param	FLPS
conv1	220×220×3	110×110×64	7×7×3, 2	9K	115M
pool1	110×110×64	55×55×64	3×3×64, 2	0	
rnorm1	55×55×64	55×55×64		0	
conv2a	55×55×64	55×55×64	1×1×64, 1	4K	13M
conv2	55×55×64	55×55×192	3×3×64, 1	111K	335M
rnorm2	55×55×192	55×55×192		0	
pool2	55×55×192	28×28×192	3×3×192, 2	0	
conv3a	28×28×192	28×28×192	1×1×192, 1	37K	29M
conv3	28×28×192	28×28×384	3×3×192, 1	664K	521M
pool3	28×28×384	14×14×384	3×3×384, 2	0	
conv4a	14×14×384	14×14×384	1×1×384, 1	148K	29M
conv4	14×14×384	14×14×256	3×3×384, 1	885K	173M
conv5a	14×14×256	14×14×256	1×1×256, 1	66K	13M
conv5	14×14×256	14×14×256	3×3×256, 1	590K	116M
conv6a	14×14×256	14×14×256	1×1×256, 1	66K	13M
conv6	14×14×256	14×14×256	3×3×256, 1	590K	116M
pool4	14×14×256	7×7×256	3×3×256, 2	0	
concat	7×7×256	7×7×256		0	
fc1	7×7×256	1×32×128	maxout p=2	103M	103M
fc2	1×32×128	1×32×128	maxout p=2	34M	34M
fc7128	1×32×128	1×1×128		524K	0.5M
L2	1×1×128	1×1×128		0	
total				140M	1.6B

Fig. 6. Structure of FaceNet Network Model.

3.3 Large Size Public Face Data Set

In practice, a large data set is required for face learning. Companies that are easy to collect face data, such as Google and Facebook, use their own data. However, these data are not available from an academic standpoint. To solve this problem to some extent, public data sets exist. The most widely used among them are LFW [14] and YouTube Face [15].

The most widely used LFW was released in 2009. It consists of 13,233 photos of 5,749 celebrities. Since it is data obtained in a natural environment, it provides various features such as lighting, shadows, facial expressions, and poses that can be encountered in real situations. Unfortunately, learning and verification data are not distinguished, so they are mainly used for verification of recognition performance. The verification accuracy using LFW in face recognition through recent research reaches 99.73%. Table 1. shows the Data Sets for Face Recognition [3]

4. Conclusion

Through this paper, we investigated the trend of user authentication through biometrics and a representative model for face recognition techniques. Various studies

Table 1. Data Sets for Face Recognition

Dataset	Image#	Face#	Source
MegaFace	1,027,060	690,572	Flickr
MegaFace2	4,753,320	672,057	Flickr
CASIA WebFac	494,414	10,575	Celebrity Search
LFW	13,233	5,749	Yahoo News
FaceScrub	106,863	530	Celebrity Search
YouTube Faces	3,425(Video)	1,595	YouTube
CelebFaces	202,599	10,177	Celebrity Search
DeepFace	4,400,000	4,000	Internal
FaceNet	500,000,000	10M	Internal
IJB-A	5,712 2,085(Video)	500	Internal
IJB-B	67,000 7,000(Video)	1,845	Internal
IJB-C	138,000 11,000(Video)	3,531	Internal
VGGFace	2,600,000	2,622	Google & YouTube
VGGFace2	3,310,000	9,131	Google & YouTube

are in progress and as a result, it was found to be very useful. In the future, further research is needed on the equipment configuration requirements for practical applications and ways to provide actual personalized services.

References

1. Hyung-Jin Mun, "A Study on the User Identification and Authentication in the Smart Mirror in Private", Journal of Convergence for Information Technology, vol.9, no.7, pp.100-105, 2019.
2. H. J. Moon, S. H. Kim, "Face Recognition : A Survey", Korea Information Processing Society Review, vol.20, no.3, pp.14-23, 2013.
3. H. I. Kim, J. Y. Moon, J. Y. Park, "Research Trends for Deep Learning-Based High-Performance Face Recognition Technology", Electronics and Telecommunications Trends, vol.33, no.4, pp.43-53, 2018.
4. H. S. Choi, Y. H. Cho, "Analysis of Security Problems of Deep Learning Technology", Journal of the Korea Convergence Society, vol.10, no.5, pp.9-16, 2019.
5. Artem Oppermann, "What is Deep Learning and How does it work?", (https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac), Nov. 2019.
6. Y. H. Lee, Y. S. Kim, "Comparison of CNN and YOLO for Object Detection", Journal of the Semiconductor & Display Technology, vol.19, no.1, pp.85-92, 2020.
7. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Communications of the ACM, vol. 60, no.6, 2017.
8. Rahul Haridas, Jyothi RL, "Convolutional Neural Networks: A Comprehensive Survey", International Journal of Applied Engineering Research, vol.14, no.3, pp.780-789, 2019.
9. Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, Stefan Carlsson, "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition", IEEE Conference on Computer Vision and Pattern Recognition, pp.512-519, 2014.
10. Hyochang Ahn, Yong-Hwan Lee, "A Research of CNN-based Object Detection for Multiple Object Tracking in Image", Journal of the Semiconductor & Display Technology, pp.110-114, 2019.
11. Yaniv Taigman, Ming Yang, Marc'Aurelio, Lior Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", CVPR 2014, 2014.
12. Jae Sung Choi, "Predictive Maintenance of the Robot Trouble Using the Machine Learning Method", Journal of the Semiconductor & Display Technology, 2020.
13. Florian Schroff, Dmitry Kalenichenko, James Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering", CVPR 2015, 2015.
14. Labeled Faces in the Wild, UMass(http://vis-www.cs.umass.edu/lfw/)
15. YouTube Faces DB, TAU(https://www.cs.tau.ac.il/~wolf/ytfaces/)

접수일: 2020년 6월 25일, 심사일: 2020년 6월 26일,
게재확정일: 2020년 6월 26일