Face Recognition using FaceNet (Survey, Performance Test, and Comparison)

Ivan William
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
ivan_willi@yahoo.com

De Rosal Ignatius Moses Setiadi
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
moses@dsn.dinus.ac.id

Eko Hari Rachmawanto
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
eko.hari@dsn.dinus.ac.id

Heru Agus Santoso
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
heru.agus.santoso@dsn.dinus.ac.id

Christy Atika Sari
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
atika.sari@dsn.dinus.ac.id

***Abstract*—Face recognition that is technology used for recognizing human faces based on certain patterns and re-detect faces in various conditions. Face recognition is currently becoming popular to be applied in various ways, especially in security systems. Various methods of face recognition have been proposed in researches and increased accuracy is the main goal in the development of face recognition methods. FaceNet is one of the new methods in face recognition technology. This method is based on a deep convolutional network and triplet loss training to carry out training data, but the training process requires complex computing and a long time. By integrating the Tensorflow learning machine and pre-trained model, the training time needed is much shorter. This research aims to conduct surveys, test performance, and compare the accuracy of the results of recognizing the face of the FaceNet method with various other methods that have been developed previously. Implementation of the FaceNet method in research using two types of pre-trained models, namely CASIA-WebFace and VGGFace2, and tested on various data sets of standard face images that have been widely used before. From the results of this research experiment, FaceNet showed excellent results and was superior to other methods. By using VGGFace2 pre-trained models, FaceNet is able to touch 100% accuracy on YALE, JAFFE, AT & T datasets, Essex faces95, Essex grimace, 99.375% for Essex faces94 dataset and the worst 77.67% for the faces96 dataset.**

Keywords— Face Recognition, Face Detection, FaceNet, Deep Convolutional Network, TensorFlow, Deep Learning

I. INTRODUCTION

Face recognition is one branch of computer science is an ability to recognize or identify the person's identity by analyzing the pattern-based facial contours of human faces [1]. The development of face recognition methods in the last two decades shows very rapid progress. Initially, the method was running very slowly with the results of poor accuracy so that it cannot be applied daily life until now it can be applied in real-time by producing excellent accuracy [3]. Currently, face recognition is used as a technology to provide multiple security in various practices likes verification of identity, access authority, observation, to replace passwords and identity cards that are no longer safe. The use of face recognition has the benefit of verifying personal data because, inhuman faces things like irises, retinas, faces are very unique to each other [2].

Face recognition has many methods in its application today. For example, Principal Component Analysis (PCA) [4], Linear Discriminant Analysis (LDA) [5], and Deep Learning [6] [7] [8]. PCA is a dimensional reduction method that is often used to reduce dimensions on a large enough dataset. You do this by changing variables in large datasets to smaller parts but still using the information on large datasets. The PCA method is very widely used because the concept is very simple and has an algorithm that is relatively efficient at the computational stage [9].

In signal processing, PCA is known as the transformation of the Karhunen-Lo`eve. The advantage of using PCA is the low complexity in grouping images, the small database representation because what is used is only image trainees that are stored in the form on a reduced basis, the lack of noise due to the chosen variation is only the maximum by ignoring the small variation base. The disadvantage of using PCA is a covariance matrix that is difficult to evaluate accurately, the difficulty of capturing invariance on PCA unless the data trainee provides information explicitly. One of the PCA methods is Eigenfaces. Eigenfaces is unsupervised and it ignores all the class labels. Eigenfaces use all data in its entirety by using SVD for dimensional reduction. However, Eigenfaces are not optimized for class separation. LDA or known as Fisher's Linear Discriminant or fisher faces and developed by Fisher which is a classic technique and is important in the discriminant analysis (DA) or as a classification [10]. LDA is a supervised data that is very strong against dimension reduction. By using supervised, LDA is able to work quite well on a dataset containing more number of face images. The purpose of the LDA method is to detect groups of test objects based on the closest average. Each group has an average vector of each vector characteristic for each object. The closest measurement is using a distance metric. The advantages of LDA are the simple, fast and portable method, and are good for the initial project. The disadvantage of LDA is an old algorithm. PCA is mainly used for feature extraction while the LDA is used for classification. There are many face recognition methods in Deep Learning, such as the use of Convolutional Neural Networks (CNN) and Artificial Neural Network (ANN).

FaceNet is one of the uses of face recognition based on Deep Learning. That is a one-shot learning method using Euclidean space to calculate the similarity distance for each face. FaceNet is a fairly new method, introduced by Google

researches in 2015, using Deep Convolutional Network method. On the study [8], two different architectures were used, namely The Zeiler & Fergus network method and latest of Inception network method. CNN in train uses Stochastic Gradient Descent (SGD) applying backpropagation and AdaGrad standards. FaceNet is able to provide accuracy of up to 99.63% from dataset Labeled Faces in the Wild (LFW) and 95.12% on the YouTube Faces DB. The advantage of using the FaceNet method is that this model only requires minimal alignment in terms of cutting the face area quite tightly. The disadvantage of FaceNet is that the training performance is quite heavy because it uses a CPU. This study aims to propose the FaceNet method and test it on a public image dataset. Then compare with previous face recognition methods to find out how much face recognition accuracy increases on FaceNet.

II. THEORY

A. FaceNet

FaceNet is a method that uses deep convolutional networks to optimize its embedding, compared to using intermediate bottleneck layers as a test of previous deep learning approaches. This method is called one-shot learning. In more detail, this method can use a small sample of face images to produce the initial model, and when there are new models, the initial model can be used without retraining. FaceNet directly trains the face using the Euclidean space where the distance consists of similarities between facial models. When the results of similarities between face models are obtained, it will be easy to carry out face recognition and classification using FaceNet attached become feature vectors.

In the training process, FaceNet applies triplets by matching face to face with the online novel triplet mining method. Of course, this triplet consists of a collection of anchor images, where each image consists of positive and negative images. Fig. 1 shows the structural model used in FaceNet. FaceNet consists of batch layers as input and deep architecture which is deep CNN followed by L2 normalization, that become the result of face embedding [8]. FaceNet also pursued by the triplet loss when the training process, see Fig. 2.

Triplet loss training methods have three main elements namely anchor, positive and negative. This triplet loss works by minimizing the distance between anchors positively and maximizing the distance between anchors negatively. Where this positive has the same identity as the anchor and negative has a different identity from the anchor.

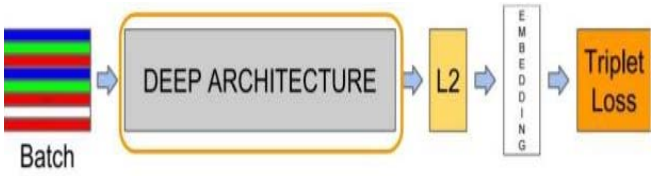

Fig. 1. The model structure of the FaceNet

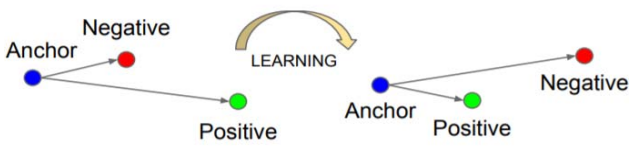

Fig. 2. The triplet loss training

FaceNet trains its output directly into concise 128-dimensional embedding by apply triplet based loss method depend on LMNN [11]. It formed by two thumbnails of compared faces and thumbnails that do not match and the loss aim to distinguish between positive and negative pairs using a range of limit. Thumbnails were cut tightly on the face field, it didn't need 2D or 3D adjustment, apart from the ratio and translation implemented.

B. Multi-task Cascaded Convolutional Neural Networks (MTCNN)

Convolutional Neural Network is a neural network design that aims to rate visual images and datasets. MTCNN or Multi-task Cascaded Neural Networks is a developed version of CNN [12]. Shows that the proposed MTCNN model show the innate bond between disclosure and the adjustment of the usage frames to increase the performance. Purpose of the proposed MTCNN is to form an avalanched structure and use it as material for multi-task knowledge to forecast the location of the face and it is marked in a coarse-to-fine method. Also, MTCNN aims the bond of 2 tasks. And in its application, MTCNN is able to detect real-time with fairly high accuracy [13]. The MTCNN model made from three networks first is the Proposal Network (P-Net) which functions to get the face area and give some bounding boxes to the face. Second is the Refine Network (R-Net), which functions to remove some bounding boxes on the face by calibrating them and leaving only an accurate bounding box. And the last is the Output Network (O-Net). The workings of the O-Net are different from the previous layers, the O-Net takes the result of the R-Net in the form of a boundary box and divides it into three different layers: the first layer for face probabilities in the box, the second layer to give the boundary coordinates in the box, the last layer for the coordinates of the five landmarks of the face.

C. TensorFlow

TensorFlow is one of a public library for large-scale analytics and machine learning computing using dataflow graphs. The dataflow graph is to represent the computational, shared state, and operation calculations that affect certain states. When data flow graphs are mapped in multiple device clusters, computational techniques can use one or more CPUs or GPUs and Tensor Processing Units (TPUs) on one desktop [14]. TensorFlow combines a set of models and machine learning algorithms and deep learning (or neural networks). TensorFlow was found and developed by Google Brain team researchers and architects from Google's Machine Intelligence Research organization with the aim of carrying out research in machine learning and deep neural networks.

TensorFlow may develope and do deep neural networks for partition the handwritten digit, image identification, word attaching, recursive neural networks, sequence-to-sequence methods for machine interpretation, natural language processing, and Partial Differential Equation (PDE) depend on simulation [15].

III. IMPLEMENTATION AND RESULTS

In testing using the FaceNet method in this research, many public datasets such as YALE [16], JAFFE [17], AT&T [18], Georgia Tech [19], dan Essex [20] were used to obtain the accuracy of this face recognition method.

TABLE I. FACE IMAGE DATABASE USED

Dataset	The number of persons	Total number of images in the databases	Sample images
Yale Database	15	164	
JAFFE	10	213	
AT&T	40	400	
Georgia Tech	50	750	
Essex_faces94	153	3078	
Essex_faces95	72	1440	
Essex_faces96	152	3016	
Essex_grimace	18	360	

The face image dataset is used because it has been widely used in various face recognition studies and is widely used as a standard face image dataset. So by using this data set, it will be easy to compare with other methods that have been proposed in various previous studies. The face image sample used with the description of each dataset is shown in Table 1. To do face recognition with FaceNet, the following steps are carried out:

A. Preprocessing

In the pre-processing stage, face image detection is carried out on each image of the MTCNN. When face detection is successful, the original image on the dataset with the size of x pixels × y pixels is done by cutting according to the detected face area with a size of 182 pixels × 182 pixels. Fig. 3 is a description of this preprocessing process using the MTCNN method.

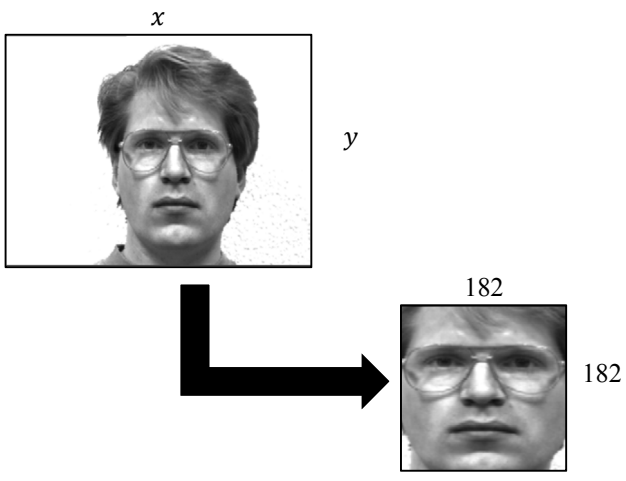

Fig. 3. Example of cropping the image on the preprocessing stage using MTCNN method

B. Training

After the preprocessing process is complete on all data sets. The next step is to train the model. In this research, pre-trained data of existing models are used to simplify the process. The data set in the pre-trained model also influences the accuracy of face recognition. So in this research two types of pre-trained models were taken from CASIA-WebFace [21] and VGGFace2 [22], as a comparison of whether they will actually influence the quality of the results. Table 2 displays more detailed descriptions of the pre-trained models used.

TABLE II. PRE-TRAINED MODELS

Model Name	LFW Accuracy	Training dataset	Architecture
20180408-102900	0.9905	CASIA-WebFace	Inception ResNet v1
20180402-114759	0.9965	VGGFace2	Inception ResNet v1

Then training the dataset of images that have been preprocessed and then collected in a folder called pre_img with labels according to their names and then trained using TensorFlow machine learning assistance. The results of this process will produce files in the .pkl format, as an illustration, see Fig. 4.

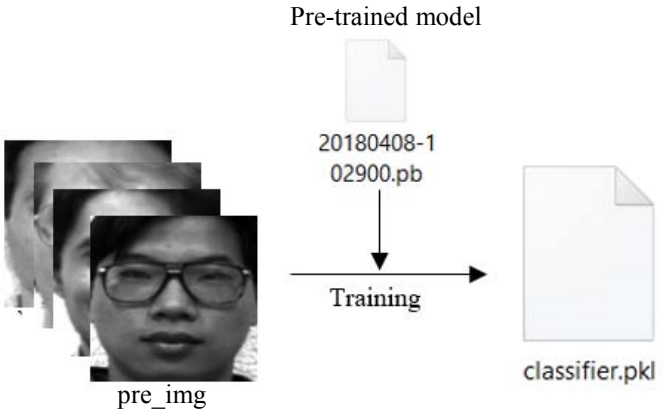

Fig. 4. Training process

C. Results

At this stage, an accuracy testing of face faces in each face database is carried out. There are eight public face databases that have been widely used in previous research. This dataset is selected because it makes it easier during the comparison process with other methods. The results of the accurate measurement of face recognition on each face image dataset are presented in table 3. The measurement technique used is by using Euclidean space. Every face image that has been processed using Euclidean space will be compared with the index label on each face. Using the numpy.mean function (numpy.equal (best class indices, label)) where best_class_indices is the result of Euclidean space measurements, and labels are the original labels of each face. Based on the results of FaceNet testing in Table. 3, this shows that FaceNet accuracy is very accurate to use in the face recognition method. But in the Essex_faces96 dataset, the accuracy is not maximal, this is possible because in this dataset there are several face labels containing different faces from each other. Therefore differences in facial conditions that are very extreme will greatly affect the accuracy of using FaceNet.

TABLE III. FACE RECOGNITION RESULTS USING FACENET IN EACH FACIAL IMAGE DATABASE

Dataset	Total images in the database	Total images successfully aligned	FaceNet rate (in %)	
			Casia-WebFace	VGGFace2
Yale	164	164	98.9	100
JAFFE	213	213	100	100
AT&T	400	400	97.5	100
Georgia Tech	750	750	100	100
Essex faces94	3078	3059	99.37	99.37
Essex faces95	1440	1439	99.65	100
Essex faces96	3016	3016	76.86	77.67
Essex grimace	360	360	100	100

IV. COMPARISON WITH PREVIOUS METHODS

After obtaining the results of accuracy on the FaceNet test, the next step that will be carried out in this study is to compare the FaceNet method with other face recognition previously proposed. The other methods that are compared are:

1) Local Texture Description Framework-based Modified Local Directional Number (LTDF_MLDN) + Nearest Neighborhood Classifier (NNC) using Euclidean, Manhattan, Minkowski, G-statistics, and chi-square. (Only the best value distance metrics is taken) [23].
2) Eigenfaces use PCA and KPCA (Kernel Principal Component Analysis) methods [24].
3) The string of Grammar Fuzzy K-Nearest Strength (sgFKNN) [25].
4) PCA + SVM (Support Vector Machine) [4].

To see the comparative results of all tests on each data set of facial images can be seen in Table. 4. Based on the results of the comparison method presented in Table 4, it is seen that the accuracy produced using the FaceNet method is very good for each dataset that has been tested. FaceNet is able to achieve almost 100% accuracy for each test. This is because this method compares for each face after face with the Tensorflow pre-trained model and machine assistance. The similarity of the results of accuracy on the JAFFE dataset with other methods, because this dataset does not show significant differences in each face, and what distinguishes only facial expressions. Even though certain datasets have different facial conditions (such as the use of accessories) that are different, this has proven to not affect accuracy in this method. It's just that the results of this accuracy will be affected if one face has a very different image as in the Essex faces96 dataset. Pre-trained models also greatly affect the accuracy of FaceNet. Using VGGFace2 pre-trained models produces better accuracy compared to the Casia-WebFace pre-trained model. VGGFace2 pre-trained models actually look superior compared to the Casia-WebFace pre-trained model. Where it has been shown in Table 2 that the results of LFW, VGGFace2 pre-trained accuracy measurements have a slight difference of 0.006. Although the accuracy of the LFW results is very slight, it has a greater impact on the testing of facial image recognition.

TABLE IV. COMPARISON OF THE FACENET METHOD WITH THE METHOD IN PREVIOUS RESEARCH ON EACH DATA SET OF FACIAL IMAGES

Dataset	LTDF_MLDN + NNC [23]	Eigen faces [24]		sgFKNN [25]	PCA + SVM [4]	FaceNet	
		PCA Algorithm	*KPCA Algorithm*			*Casia-WebFace*	*VGGFace2*
Yale Database	-	88.26	97.25	-	93	98.9	100
JAFFE	100 (Manhattan & chi-square)	71.2	80	100	-	100	100
AT&T	97.5 (chi-square)	-	-	99.25	98.75	97.5	100
Georgia Tech	83.73 (chi-square)	-	-	79.57	-	100	100
Essex faces94	-	70.0	77.0	-	-	99.37	99.37
Essex faces95	99.09 (g-statistics)			-	-	99.65	100
Essex faces96	-			-	-	76.86	77.67
Essex grimace	100 (Manhattan & chi-square)			-	-	100	100

V. CONCLUSION

This study aims to conduct a survey and test the performance of the relatively very new face recognition method called FaceNet. FaceNet is introduced by Google researches by integrating machine learning in processing face recognition. Tests were carried out on several public datasets, such as YALE, JAFFE, AT & T, Georgia Tech, and Essex. Two pre-trained models are also used in testing, namely CASIA-WebFace and VGGFace2. From the test values, it shows that the exactness using FaceNet algorithm is very good where at each dataset the face image produces recognitions that can reach up to 100%. The accuracy of the FaceNet method is also strongly influenced by the pre-trained data model where VGGFace2 produces better average recognition accuracy. But in one data set only found an accuracy of about 77%, this is possible because each face label has many differences, while the training method on FaceNet uses triplet loss that will minimize the gap of anchor and positive, also maximize the gap of anchor and negative image, where is the difference a very significant face can be considered a negative image. However, based on comparison with other methods, the FaceNet method is proven to have the best performance.

REFERENCES

[1] J. B. Wilmer, "Individual Differences in Face Recognition: A Decade of Discovery," *Current Directions in Psychological Science,* vol. 26, no. 3, 2017.

[2] "Eigenfaces and Beyond," in *Face Processing*, Academic Press, 2006, pp. 55-86.

[3] V. Bhandiwad and B. Tekwani, "Face Recognition and Detection using Neural Networks," pp. 879-882, 2017.

[4] X. Chen, L. Song and C. Qiu, "Face Recognition by Feature Extraction and Classification," *2018 12th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID),* pp. 43-46, 2018.

[5] M. E. Rane and A. J. Pande, "Multi-Modal Biometric Recognition of Face and Palm-Print Using Matching Score Level Fusion," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),* pp. 1-6, 2018.

[6] M. Korkmaz and N. Yilmaz, "Face Recognition by Using Back Propagation Artificial Neural Network and Windowing Method," *2015 2nd International Conference on Artificial Intelligence (ICOAI 2015),* vol. 4, no. 1, pp. 15-19, 2015.

[7] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," *Proceedings of the British Machine Vision Conference 2015,* no. Section 3, pp. 41.1-41.12, 2015.

[8] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 815-823, 2015.

[9] I. T. Jolli□e and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 374, p. 20150202, 2016.

[10] D. Chu, L.-Z. Liao, M. K.-P. Ng and X. Wang, "Incremental Linear Discriminant Analysis: A Fast Algorithm and Comparisons," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 26, pp. 2716-2735, 2015.

[11] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *In NIPS. MIT Press,* pp. 2, 3, 2016.

[12] F. Rahman, I. J. Ritun, N. Farhin, and JiaUddin, "AnAssistive Model for Visually Impaired People using YOLO and MTCNN," *ICCSP '19 Proceedings of the 3rd International Conference on Cryptography, Security and Privacy,* pp. 225-230, 2019.

[13] M. Ma and J. Wang, "Multi-view Face Detection and Landmark Multi-view Face Detection and Landmark," pp. 4200-4205, 2018.

[14] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker and V, "TensorFlow: A system for large-scale machine learning," *12th USENIX Symposium on Operating Systems Design and Implementation,* pp. 265-283, 2016.

[15] L. Yuan, Z. Qu, Y. Zhao, H. Zhang and Q. Nian, "A Convolutional Neural Network based on TensorFlow for Face Recognition," *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC),* pp. 525-529, 2017.

[16] "Yale face database," [Online]. Available: http://vision.ucsd.edu/content/yale-face-database.

[17] M. J. Lyons, S. Akemastu, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," *3rd IEEE International Conference on Automatic Face and Gesture Recognition,* pp. 200-205, 1998.

[18] F. Samaria and A. Harter, "Parameterisation of a Stochastic Model for Human Face Identification," *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision,* pp. 138-142, 1994.

[19] A. V. Nefian, "Georgia Tech Face Database," [Online]. Available: http://www.anefian.com/research/face_reco.htm.

[20] L. Spacek, "Computer Vision Science Research Projects," 2008. [Online]. Available: https://dces.essex.ac.uk/mv/allfaces/.

[21] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning Face Representation from Scratch," 2014.

[22] Q. Cao, L. Shen, W. Xie, O. M. Parkhi and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," *International Conference on Automatic Face and Gesture Recognition,* 2018.

[23] R. R. Rose, K. Meena, and A. Suruliandi, "An Empirical Evaluation of the Local Texture Description Framework-Based Modified Local Directional Number Pattern with Various Classifiers for Face Recognition," *Brazilian Archives of Biology and Technology,* vol. 59, no. 2, pp. 1-17, 2016.

[24] H. S. Dadi and K. M. P.G, "Performance Metrics for Eigen and Fisher Feature Based Face Recognition Algorithms," vol. 16, no. 6, pp. 157-167, 2016.

[25] P. Kasemsumran, S. Auephanwiriyakul, and N. Theera-Umpon, "Face Recognition Using String Grammar Fuzzy K-Nearest Neighbor," *2016 8th International Conference on Knowledge and Smart Technology (KST),* no. 2, pp. 55-59, 2016.