Face Recognition Using ArcFace and FaceNet in Google Cloud Platform For Attendance System Mobile Application

Rosa Andrie Asmara[1] , Brian Sayudha[2], Mustika Mentari[3], Rizky Putra Pradana Budiman[4], Anik Nur Handayani[5], Muhammad Ridwan[6], and Putra Prima Arhandi[7]

[1,2,3,4,6,7] State Polytechnic of Malang, Malang 65141, Indonesia
[5] Malang University, Malang 65145, Indonesia
[1]rosa.andrie@polinema.ac.id
[2]briansayudha@gmail.com
[3]must.mentari@polinema.ac.id
[4]rizkyputrapb@gmail.com
[5]aniknur.ft@um.ac.id
[6]ridwan.mtte20@gmail.com
[7]putraprima@polinema.ac.id

Abstract. The attendance system process in Indonesia generally are still using a traditional method. Paper-based is used as a medium to perform attendance at every event. With this traditional method, there are still many shortcomings in terms of security and management. In terms of security, the traditional attendance system is still quite lacking due to the number of participants cheating by asking their relatives, such as examples of signatures that can still be imitated, or attendance checks can still be tricked because we can change them easily. Therefore, it is necessary to have an attendance system that can be carried out efficiently, safely, and easy to manage, with attendance being done online or using a smartphone. It can be implemented easier for event owners to manage the attendance track of participants, reduce the use of paper, which is quite significant, and secure the attendance system. CNN is an artificial neural network that is more often used in visual image analysis. CNN can distinguish visual images from one another with various aspects given. The models that we used for this application are ArcFace and FaceNet. Three different BackEnd Encoder and BackEnd recognized are used, RetinaFace, MTCNN, and OpenCV. From the experiment, we suggest the usage of ArcFace in RetinaFace for High Accuracy of recognition but with high-cost drawbacks, the longer computation time for encoding and recognition. As an alternative, ArcFace with MTCNN can be used with faster computation time but less accurately than RetinaFace.

Keywords: Attendance system, Smartphone, biometrics, facial recognition, CNN

1 Introduction

In this modern era, people are getting more familiar with the usage of technology, especially mobile devices. The development of mobile applications is in their best

R. Andrie Asmara et al. (Eds.): ATASEC 2022, AER 219, pp. 134–144, 2022.
https://doi.org/10.2991/978-94-6463-106-7_13

competition times ever. The number of industries and people involved in this area is getting more prominent over time, relevant with the more advancement in technology. More work can be done easier by implementing these mobile applications since almost all activities have many alternative apps and every people at least has a smartphone device. Indonesia is a country with a population of approximately 272 million people, with a device ownership rate of 76% for smartphones and a percentage of internet usage rate of 98.2% for a smartphone. However, with many existing devices, Indonesia is still outdated for this technological development, especially in a security system with biometrics. China, on the contrary, has developed more advanced apps with artificial intelligence (AI) that can identify people through facial recognition in public areas. To reduce errors and increase effectiveness at work, we need to take advantage of technology and implement it fundamentally in the field of an attendance system. This system can record student presence in school or any public events using facial recognition technology, so the event or school attendance can be carried out efficiently and safely without additional fraud, such as asking their relatives to represent them attending the event.

In Indonesia, the attendance system is still relatively underdeveloped because most attendance systems in Indonesia still use a piece of paper and give the signature there. However, this method causes many problems, such as management difficulty, presence cheating, and even someone's late presence still cannot be detected precisely. Therefore, the company surveyed to see which method was more profitable to reduce all the shortcomings in the manual attendance system. The results prove that 48% of companies choose online attendance because it can be managed quickly and accurately and reduce the costs incurred in attendance (such as reducing paper for attendance).

Face detection is a crucial procedure for other face-related technologies, including face alignment, facial recognition, facial animation, facial attribute analysis, and human-computer interaction. The accuracy of the face detection system directly impacts this technology, which is why the success of face detection is significant. In a general sense, face detection aims to determine whether faces are in the image and recognize those faces detected in images[1].

The face recognition technology as an identification tool still has considerable weaknesses in terms of face authentication. For example, traditional face recognition cannot effectively recognize a person's face in certain positions. To be able to maximize this, it is necessary to refine the existing facial recognition approach by relying on all the feature points on a person's face so that verification can be done very accurately[2].

On the journal entitled “Penerapan Facial Landmark Point untuk Klasifikasi Jenis Kelamin berdasarkan Citra Wajah”, The face of everyone contains several different information. Examples are expression, gender, age, and race. Therefore, in biometrics technology or biological data recognition technology, faces can be used as an identification [3]. Based on research entitled "Comparison of Geometric Features and Color Features for Face Recognition," performance analysis with Geometric Features and Color Features combined will make the results more accurate, and the mean accuracy between three models resulted: GNB with an average accuracy of 74.67 %, KNN K=5 w with an average accuracy of 72.1%, and SVM one to one with an average accuracy of 74.83%. Moreover, the excellent condition of the dataset must be set to gain optimum accuracy[4].

Based on research entitled "Haar Cascade and Convolutional Neural Network Face Detection in Client-Side for Cloud Computing Face Recognition", six types of test data (Normal Face, Expression Face, Face with Mask, Face with some Obstacles, Face with Glasses, and More than one face) has an average accuracy of 81.12%. Compared to using CNN, an accuracy of 86.53% was obtained. Haar Cascade can detect multi-user faces (more than one) and is superior by detecting simultaneously without interference which face is more dominant [5].

The previous research entitled "RetinaFace: Single-stage Dense Face Localization in the Wild" concluded that the RetinaFace method had outperformed the most advanced state-of-art methods for face detection today. With stable face detection in various poses, when RetinaFace is combined with existing state-of-art methods for face detection, it will undoubtedly increase face detection accuracy [2].

From this explanation, the method used in this research is CNN to recognize faces optimally. The system will be developed in mobile applications because it has been shown that many Indonesians access the internet. The attendance system using smartphones and mobile devices is designed for highly efficient face localization and recognition that can give high-speed real-time recognition[2]. This system will mainly develop for offline events (required for a participant to come to the event place) such as graduation, music festival, weddings, offline seminars, and support online events such as video conference seminars.

The attendance system will follow the plan that organizers choose, and it is hoped that this research and development can produce good facial recognition accuracy and a good attendance system.

2 Literature Study

2.1 CNN (Convolutional Neural Network)

CNN / Convnet is a Deep Learning algorithm, where algorithm can take an input image or computer vision and assign it to various aspects in the image and we can distinguish one image to another.

The preprocessing required in CNN is much lower that other method. The method is often used in image recognition systems. There are various architectures of CNN that available which become a key in building algorithm in Image recognition.

CNN works with preprocessing the inputted image with separated their primary color (Red, Green, and Blue). The role of CNN will convert the image into a form that is easier to process without losing its features, and we can get a perfect prediction later on. The CNN used the Kernel / Filter K to carry the involved element for convolution operation. The purpose of Convolutional Operation is to extract the high-level features from input images. The following process is the Pooling layer, which reduces the spatial size of the Convolved features, and the last process is the Classification - Fully Connected Layer (FC Layer).

2.2 ArcFace (Addtive Angular Margin Loss)

ArcFace Is a Loss function used in face recognition tasks. SoftMax is traditionally used in this task. ArcFace can be used to obtain highly discriminative features for face recognition. The proposed ArcFace has a clear geometric interpretation due to the exact correspondence to the geodesic distance on the hypersphere[7]. The most widely used classification is SoftMax loss, presented as the formula (1) below:

$$L_1 = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{W_{yi}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}} \qquad (1)$$

where xi ∈ R d indicate the deep feature of the i-th sample, belongs to the yi-th class [11]. The embedding feature dimension d is set to 512, Wj ∈ R d denotes the j-th column of the weight W ∈ R d×n and bj ∈ R n is the bias term[7]. The batch size and the class number are N and n, respectively. Traditional SoftMax loss is widely used in deep face recognition[11].

2.3 Retina Face

Pixel-wise face localization method is used to utilize multi-task learning. The strategy is to simultaneously predict the face score, face box, five facial landmarks, and 3D position and correspondence of each facial pixel [2]. As a result, retinaFace surpasses the accuracy point of the state-of-the-art two-stage method. In addition, RetinaFace can improve ArcFace's verification accuracy (with TAR equal to 89.59% when FAR=1e-6). This accuracy indicates that better face localization can significantly improve face recognition [2].

2.4 FaceNet

FaceNet is a Face Recognition system developed in 2015 that achieved state-of-the-art results on a face recognition benchmark dataset. The FaceNet system can be used to extract high-quality features from faces that are provided in images, called face embeddings. The face embedding is then used to train a face identification system. In addition, FaceNet directly learns a mapping from face images to a compact Euclidian space, where distances directly correspond to a measure of face similarity [8].

The difference between FaceNet and other methods is that FaceNet learns the mapping from the images or faces and creates embeddings rather than using any bottleneck layer for recognition or verification tasks. Once the embeddings are created, all the other tasks like verification and recognition can be performed using standard techniques [8]. FaceNet uses Triplet loss, which will directly reflect what we want to achieve in face verification, recognition, and clustering[8].

3 Guest Event Authorization

Guest Event Authorization (Presentik) is an attendance application using the latest technology by applying facial recognition as a biometric detection to reduce fraud during attendance. This Presentik has also implemented a system for calculating the distance between the event venue and the guest attendance. This Face recognition feature applies the CNN method to get a facial representation from guest data that has been filled in and then compared with the latest attendance face data so that the guest can be detected and who is registered or not.

3.1 Face Dataset Encoding Flowchart

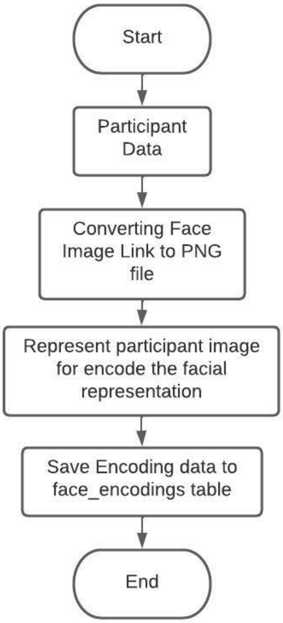

Start
Participant Data
Converting Face Image Link to PNG file
Represent participant image for encode the facial representation
Save Encoding data to face_encodings table
End

Fig. 1. Face Dataset Encoding Flowchart

In face dataset encoding is a flow where the computer performs training for a given dataset. Fig 2 above shows the Face dataset encoding flowchart. Encoding intends to perform feature embedding, which is carried out by the CNN method to obtain face classifications and provide results in the form of embeddings. Retina Face will assist this embedding feature in performing face detection to take facial characteristics in the feature extraction process. These results are in the form of an array that is only understood by the computer. In the flowchart Fig 1 above, the representation face is a feature embedding done in the CNN method. The results of the facial representation will be compared with the latest facial data that will be sent at the time of attendance.

3.2 Recognizing Face Flowchart

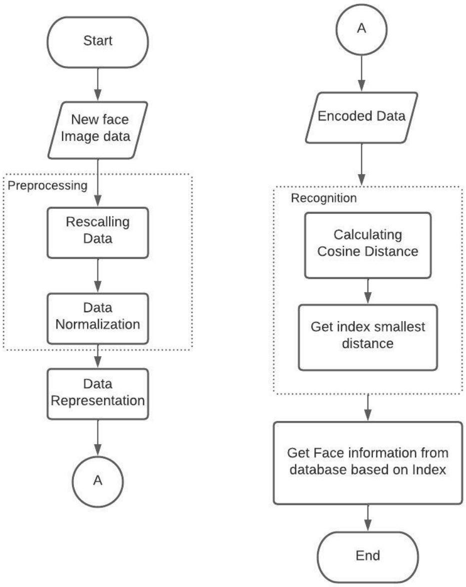
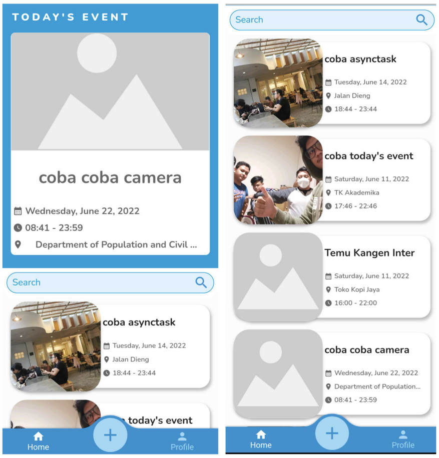


Fig. 2. Face Recognition Flowchart

In Face Recognition, the system will use a deepface library. Deepface library is a face recognition using CNN method that will help the system recognize the face. The face recognition process starts from getting the new uploaded attending image, which will then be preprocessed by scaling the data and performing data normalization. This data normalization aims to normalize the light intensity in the image obtained. After preprocessing, the system will perform data representation using the deepface library. This representation data will produce data that is already in the encoding and can only be understood by the computer only. Fig 2 above is the Face Recgnition phase flowchart used in the application.

This app is run by someone who has a role as an event organizer. The event organizer can manage events that have been made, create events, do attendance, and manage attendance logs.

3.3 User Interface Implementation



coba coba camera
Wednesday, June 22, 2022
08:41 - 23:59
Department of Population and Civil ...

coba asynctask
Tuesday, June 14, 2022
Jalan Dieng
18:44 - 23:44

coba today's event
Saturday, June 11, 2022
TK Akademika
17:46 - 22:46

Temu Kangen Inter
Saturday, June 11, 2022
Toko Kopi Jaya
16:00 - 22:00

coba coba camera
Wednesday, June 22, 2022
Department of Population...
08:41 - 23:59

Fig. 3. Mainpage UI of Guest Event Authorization Application

The development of the front-end system is done with Flutter 3.0.1 and Dart 2.17.1. The UI is developed based on the mockups that already been designed using Figma in the system design phase. The UI implementation of the application is shown in the Fig 3 above.

4 Experiments

The experiment will be aimed at finding the best combination between the face recognition model and the backend detector. After finding the best combination, it will try to be implemented in the Guest Event Authorization (Presentik) application.

4.1 Looking for the best combination

We conduct the experiment with 200 face image samples to look for the best combination. Each person only has one face close-up image in the dataset. In the experiments, several data were measured, such as encoding time and accuracy in each combination. Fig 4 below shows the encoding time or face representation result from FaceNet and ArcFace model.

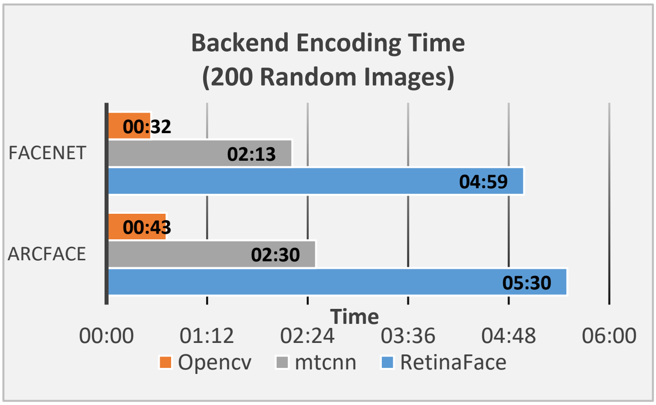


Fig. 4. Encoding (Face Representation) Time Charts

Table 1 below shows the average accuracy obtained in each combination test for each face recognition model. ArcFace has a significant result compare to the FaceNet in normal lighting condition using rear smartphone camera. This result is giving us one conclusion that we should use ArcFace rather than FaceNet.

Table 1. Average accuracy percentage for each face recognition Model

Model	w/o Flashlight	with Flashlight
ArcFace	78%	83%
Facenet	19%	63%

4.2 Results

According to table I result, it has been concluded that the best combination is obtained by using the face recognition model ArcFace. To get a higher accuracy result, we use the Retina Face backend detector as an advanced encoding algorithm option and MTCNN as the basic encoding algorithm. Each person was tested with several conditions, as many as 15 poses.

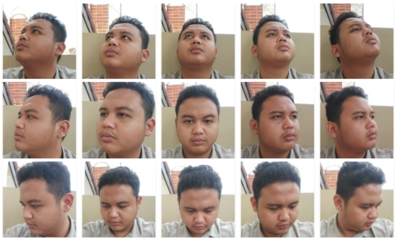


Fig. 5. Example of 15 Various poses

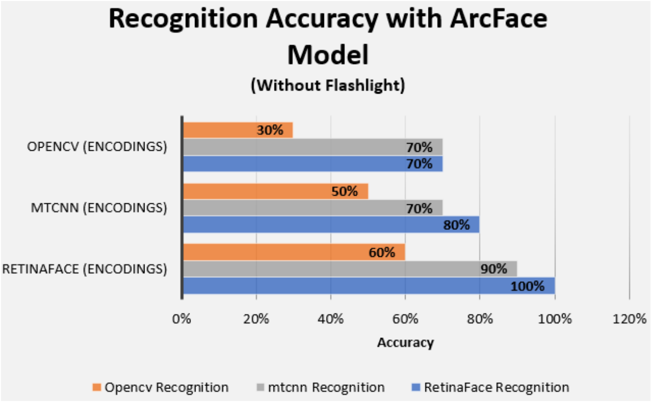


Fig. 6. Recognition Accuracy percentage with ArcFace in 200 Image samples

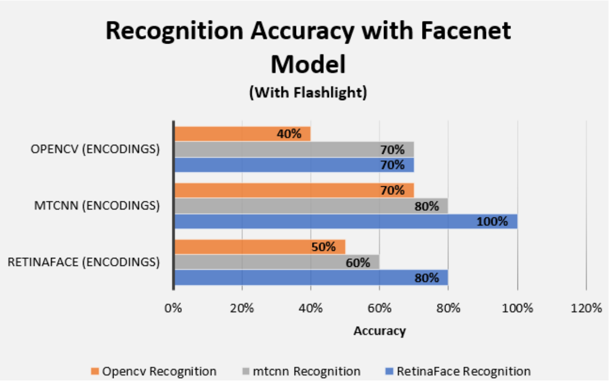


Fig. 7. Recognition Accuracy percentage with Facenet in 200 Image samples

We conduct real-time facial recognition experiment using a photo dataset of 200 photos and where each person only has one face close-up image. The experiment resulted in a longer and varied encoding time, while the results for some had poor accuracy. On the other hand, some had the same or higher accuracy than before. From the test in the second scenario, regarding the combination of the face recognition model and the backend detector, it has been concluded that the ArcFace face recognition model is more robust than the facenet face recognition model. Furthermore, the Retinaface backend detector helps the recognition be more accurate than MTCNN and OpenCV. Therefore, the combination which will be used is ArcFace with Retinaface as the advanced method and ArcFace with MTCNN for the basic method. Fig 6 and 7 above shows the result of Recognition accuracy percentage using ArcFace and FaceNet in 200 image samples.

The Fig 8 below shows the result of the volunteer's face recognition that has been entered into guest data in an event.

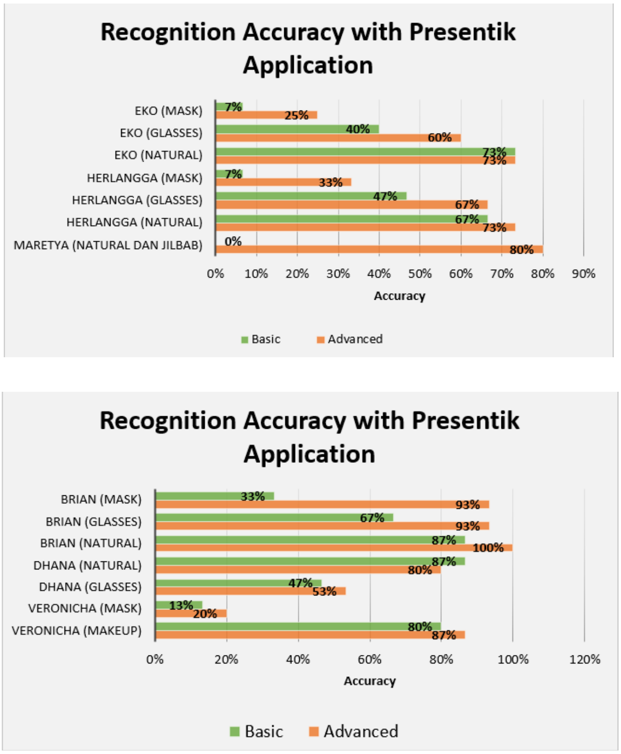


Fig. 8. Recognition Accuracy with Presentik Application on six different person and using several conditions (Natural, Masked, Glasses)

5 Conclusion

Based on the results of research and development from Face Recognition Attendance System Using CNN, we have come up with the conclusions: 1) The best combination of face recognition model and backend detector is ArcFace with Retina Face. 2) In the face recognition experiment using the Presentik system, the highest accuracy in the advanced algorithm was 100% in Brian (Natural), and the smallest was 20% in Veronicha (Mask). For the basic algorithm, the highest accuracy was 87% in Brian (Natural) and Dhana (Natural) and the smallest at 0% in Maretya (Natural and Hijab)
3. In an experiment with facial recognition accuracy using the presentik system, the average accuracy for advanced algorithms is 67%, and basic is 47%. This average is obtained without distinguishing the accessories worn on a person's face. 4). The use of the Deepface library helps in facilitating research and system development. This library also comes with data normalization for more accurate face reading or detection. 5) The use of Google Cloud Platform as a backend server using Cloud Run and Docker as its facilities can help the use of endpoints with very small server latency. 6) Docker is python:3.10-slim with running using gunicorn, which has one worker and

eight threads. 7) The cloud run has a CPU specification of 4 with 8 GB of RAM, with a request timeout of 3600 seconds. For instance, a minimum number is two, and the maximum number is ten.

Acknowledgment

The author would like to thank for the funding of master thesis scheme provided Research and Community Service Unit (P2M) Politeknik Negeri Malang.

References

1. S. Zhang, C. Chi, Z. Lei, and S. Z. Li, "RefineFace: Refinement Neural Network for High Performance Face Detection," Sep. 2019.
2. J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-stage Dense Face Localisation in the Wild," May 2019.
3. Ulla Delfana Rosiani, Rosa Andrie Asmara, and Nadhifatul Laeily, "PENERAPAN FACIAL LANDMARK POINT UNTUK KLASIFIKASI JENIS KELAMIN BERDASARKAN CITRA WAJAH," J. Inform. Polinema, vol. 6, no. 1, pp. 55–60, Jan. 2020, doi: 10.33795/jip.v6i1.328.
4. C. Rahmad, K. Arai, R. A. Asmara, E. Ekojono, and D. R. H. Putra, "Comparison of Geometric Features and Color Features for Face Recognition," Int. J. Intell. Eng. Syst., vol. 14, no. 1, pp. 541–551, Feb. 2021, doi: 10.22266/IJIES2021.0228.50.
5. R. Andrie Asmara, M. Ridwan, and G. Budiprasetyo, "Haar Cascade and Convolutional Neural Network Face Detection in Client-Side for Cloud Computing Face Recognition," in Proceedings - IEIT 2021: 1st International Conference on Electrical and Information Technology, Sep. 2021, pp. 1–5. doi: 10.1109/IEIT53149.2021.9587388.
6. Sumit Saha, "A Comprehensive Guide to Convolutional Neural Networks—The ELI5 way," Dec. 16, 2018. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
7. J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," Jan. 2018.
8. Florian Schroff, Dmitry Kalenichenko, James Philbin, "FaceNet: A unified embedding for face recognition and clustering", 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)