LONTAR KOMPUTER VOL. 14, NO. 1 APRIL 2023 **p-ISSN 2088-1541**
DOI : 10.24843/LKJITI.2023.v14.i01.p06 **e-ISSN 2541-5832**
Accredited Sinta 2 by RISTEKDIKTI Decree No. 158/E/KPT/2021

Real-time Face Recognition System Using Deep Learning Method

Ayu Wirdiani[a1], I Ketut Gede Darma Putra[a2], Made Sudarma[a3], Rukmi Sari Hartati[a4], Lennia Savitri Azzahra Lofiana[a5]

[a]Information Technology, Udayana University
Jimbaran, Bali, Indonesia
[1]ayuwirdiani@unud.ac.id (Corresponding author)
[2]ikgdarmaputra@unud.ac.id
[5]savitriazzahra@student.unud.ac.id

[b]Electrical Engineering, Udayana University
Jimbaran, Bali, Indonesia
[3]msudarma@unud.ac.id
[4]rukmisari@unud.ac.id

Abstract

Face recognition is one of the most popular methods currently used for biometric systems. The selection of a suitable method greatly affects the reliability of the biometrics system. This research will use Deep learning to improve the reliability of the biometric system and will compare it with the SVM method. The Deep Learning method will be adopted using the Siamese Network with the YoloV5 detection method as a real-time face detector. There are two stages in this research: the registration process and the recognition process. The registration process is image acquisition using YoloV5. The image result will be saved in the storage folder, and the preprocessing and training process will use the Siamese Network. The face feature model will be stored in the database. The recognition process is the same as the registration, but the feature extraction result will be embedded and compared with the already trained models. The accuracy rate using the Siamese model was 94%.

Keywords: *Biometrics, Face recognition, Deep Learning, Siamese Network, YoloV5.*

1. Introduction

The identification system generally uses conventional techniques, such as recognition based on passwords, ID cards, and PINs. The use of conventional methods still has some weaknesses, such as using an ID card is easily damaged or lost, while using a PIN or password is possible for the user to forget it. Based on these problems, a biometrics system is needed. Biometric technology is an identity recognition system using human physical characteristics, where these characteristics generally cannot be separated from humans. The human physical features are relatively stable, such as the face, palmprint, voice, fingerprints, iris, and hand geometry [1][2]. One of the unique body parts in humans is the face because every human has a different face. Face recognition applications have been widely applied in life, such as office attendance, cellphone security systems, and tools for tracking criminals [3] [4].

The process for marking attendance using a biometric system is known as AAS (Automated Attendance System); the system can monitor and track students, for example, in the environment around the classroom if learning is offline [5]. An attendance system in the manual way by using paper often has problems, such as the manipulation of attendance data, loss of attendance books, and difficulty in recapitulating attendance data. An attendance system with face recognition can be a solution for a manual attendance system because it only requires a camera to take pictures. In an attendance system, the accuracy and reliability of the system are essential, especially if the system is real-time based so that there are no recognition errors.

This research develops a recognition real-time system using human facial characteristics that can be used as an attendance system. This system can be a reference for improving the reliability of facial recognition methods in real-time. The methods used in this research combine two of the best current methods, YoloV5 for face detection and the Deep Learning method for recognition. The deep learning method used is Siamese Network because this method does not require a large number of datasets. This method can also analyze faces even though it uses a few images of facial data as a dataset. YoloV5 is the latest real-time object detection with better efficiency than the previous version [6]. The selection of YoloV5 is expected to improve the system's performance. The face classification process with the Siamese Network is the latest method for recognition. This method also can get accurate results with minimum data.

Some researchers have done real-time face recognition, including the CNN, DCT (Discrete Cosine Transform), Haar Cascade, and KNN (K-Nearest Neighbour) methods. In a real-time face recognition system using Convolutional Neural Networks proposed by Prannav KB, the CNN architecture is evaluated by setting various parameters to improve the recognition accuracy of the system. The maximum accuracy obtained is 98.75%, and 98.00% was obtained from the proposed system to use AT&T and real-time input. This research can be applied to the proposed work [7]. Another study for real-time face recognition is based on facial descriptors applied for door locks. The method used to process face descriptors uses the DCT method and predictive linear discriminant analysis (PDLDA) for face classification using KNN. The accuracy obtained is 98.30% [8]. Another real-time face identification is using the Haar cascade and KNN classifier machine learning algorithms. The implementation of face identification is for security authentication systems [9]. This face recognition system uses the YoloV5 detection method and Siamese Network, a learning method that only requires one or more steps. It aims to learn information about an object with only a few samples. This method differs from other Deep Learning algorithms for object categorization that require hundreds or even thousands of data for training. The following is research that has previously been related to this research topic. This study considers the use of Siamese Network as a training method for face recognition because of its superiority in performing training with one-shot.

Among the studies using the Siamese network are studies that develop mobile applications for facial recognition. This method gets the best accuracy of 98% [10]. To solve the problem of low accuracy and efficiency in terms of uneven and image quality occlusion due to sudden changes in light in face recognition, This study uses LBP and Perception Frequency Features and the Siamese Network for classification. The tests were conducted on the CASIA-WebFace, Yale-B, and LFW standard face data sets. Compared with other network models, they show that the proposed SN-LF network can improve the recognition accuracy of the algorithm and get better recognition [11].

Studies that use Yolo for face recognition, such as research conducted by Yuanzhang Zhao, made a face recognition application that can distinguish between wearing a mask and using YoloV5. The experiment results show that the YOLOv5 algorithm can improve the effect of object detection for various types of facial occlusion. The average accuracy has been improved by about 6%, the average draw rate has been increased by 2%, and the average training loss value has been reduced by about 0.01 [12]. Research conducted by Falah Hikamudin Arby is implementing YoloV5 for real-time social distancing. This system aims to detect people who violate health protocols during social distancing and then issue a sound warning to keep a distance from others. The accuracy of human detection using YOLO-v5 reaches 83.28%, and the accuracy rate for detecting social distance reaches 90.8%. From the results of these percentages, it can be concluded that the system can work appropriately for social distancing detection [13]. Another face recognition research was conducted in real-time by testing sixteen videos consisting of eight videos with one different student in each video and the other eight videos containing two different students. The best YOLOv5 model training results were obtained with parameters epoch=250, batch size=16, and learning rate=0.01. The best accuracy result obtained is 99.88% [14]. The following study aims to design and evaluate the recognition in real-time using CNN. This study uses a real-time facial recognition system using offline data—this face recognition system is used in real-time using the Convolutional Neural Network (CNN). The results of this research get a precision value of 98.4%, a recall of 98%, and an accuracy of 99.84% [15]. The other research implements the CNN method into facial recognition using the Tensorflow library. This method is used because the learning process is in-depth (deep learning).

The CNN method has several layers in the training process: the Conv2D, MaxPooling2d, Flatten, and Dense layers. The result of this application is face detection using Haar Cascade with the help of the OpenCV library. The number of face images used in this study was 90,000 from 36 sets of images, producing an accuracy rate of 97% [16]. The other journal makes face recognition applications using the Haar cascade method to obtain optimum accuracy for analyzing faces. The databases used are FERET and LFW and use Xception (Depth Wise Separable). The feature extraction method used is CNN. The accuracy obtained using the FERET database is 96.73%, and the LFW database is around 98.45% [17].

The research from Khatina Sari is the system that uses a Convolutional Neural Network algorithm for face detection and calculation of the Eye Aspect Ratio (EAR) as a Liveness Detector. The test results in this study are the system can recognize all faces originating from a database with an accuracy rate of 100%. In contrast, facial recognition from random faces obtained from the internet has an accuracy rate of 73.33% [18]. The other research from Aanchal Singh makes face recognition using the VGG-16 Algorithm with the transfer learning principle. This paper uses 15 subjects, and each topic consists of 11 images. The experiment compares two configurations, one with standard VGG architecture and the other with a customized Fully Connected Layer in VGG architecture. The methodology recorded a % test accuracy of 83.11% using the first customized configuration and 92.80% accuracy with the second custom configuration [19].

Based on research conducted by researchers, this research developed a real-time face recognition system using the Deep Learning method, especially using YoloV5 for face detection and the Deep Learning method for classification with the Siamese Network.

2. Research Methods

An overview of the Real-time Face Recognition system based on the Deep Learning Method can be seen in Figure 1.

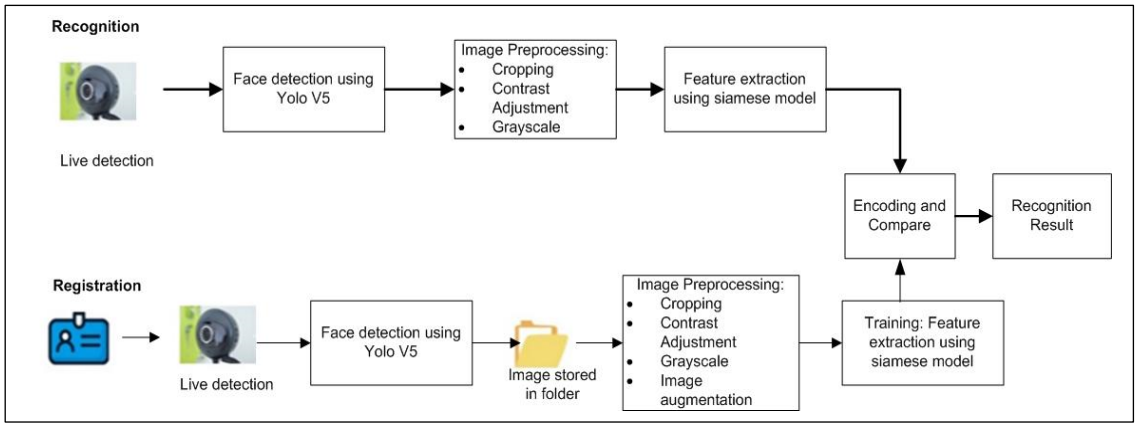


Figure 1. Overview Face Recognition System

In this section, we will outline the methods used in this research. In general, this research uses a combination of Yolo V5 to detect faces and the Siamese model to extract features, as well as Euclidean distance and triplet loss for the matching method. The facial recognition process begins with the registration process first. In this process, face detection is carried out using YoloV5. The resulting facial images that have been detected will be captured and saved as many as 20 images in a folder with the name that has been registered. The image will be pre-processed by cropping it to a size of 350x450, then adjusting the contrast and changing it to grayscale. Image augmentation was held to increase variations in the facial dataset. The augmentation process is carried out by flipping and varying the image rotation angle. The stored image dataset will be divided 80:20, where 80% is used for training and 20% for validation data. In the training process, image data will be input into the CNN architecture, where the result is a 1-dimensional matrix measuring 128 feature vectors, which is used as primary data for the matching process. In the validation process, feature vector matching will be done using Euclidean distance. The triplet loss function is carried out after getting the distance between the two features. Triplet loss is used to

learn the embedding of the face image. If the embeddings are close, they will be identified as the same face image, whereas if they are far apart, they are different images. This Siamese model will be used during the testing process.

2.1. Data Acquisition

The process begins with the image acquisition process carried out by the user. This provides input images originating from a live webcam detected using YoloV5. Yolo takes a very different approach from the previous algorithm. It is applying a single neural network to the entire image. This network will divide the image into regional areas and then predict the bounding and probability boxes. Each bounding area box's probability is weighted to classify it as an object. Dataset collection from 50 users, of which each user will capture 20 images in the registration process.

2.2. Image Preprocessing

The pre-processing stage begins by cropping the input image to 350x450, adjusting the contrast, and changing the image to a grayscale image. Image augmentation was held to increase variations in the facial dataset. The augmentation process is carried out by flipping and varying the image rotation angle.

2.3. Feature Extraction

The method used to extract facial features is modeling using Siamese Network. This network has two identical, fully connected CNNs of equal weight and receiving two images. Normal CNN uses softmax to get its classification, but here, the output from the fully connected layer is considered a 128-dimensional encoding of the input image. The first network issues the first input image encoding, and the second network outputs the second. This method is a good representation of this input image. Siamese Network can be accomplished directly by developing the specific features of a domain; that inference procedure has a very discriminatory nature towards the assignment target [20]. The Siamese architectural model used is shown in Figure 2:

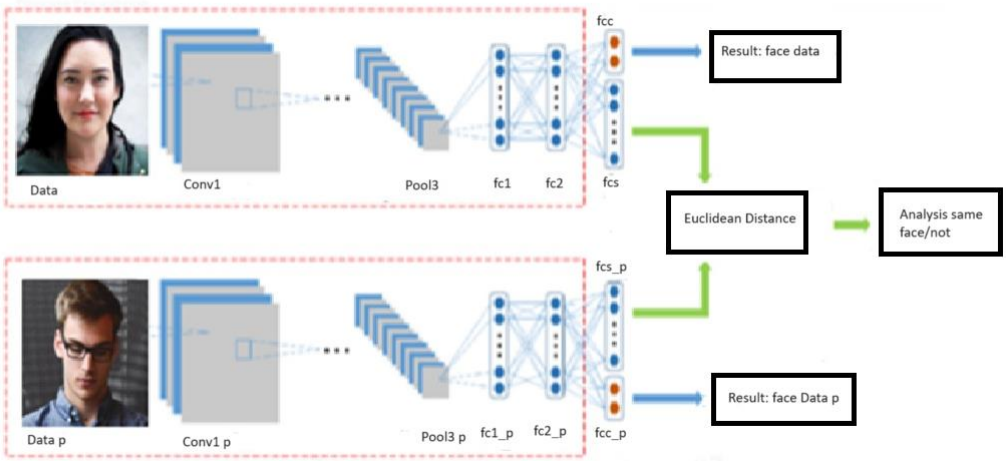


Figure 2. Siamese Network

The Siamese network method, as shown above, shows the data to be compared, and what will do the matching is the input fc1, which CNN will then process, flatten, and produce embedding in the form of fc1_p. The rest of the data in the form of fc2 will be carried out the same process as that of fc1, but in the end, it will be compared with fc1_p, whether there is fc1_p that has a significant degree of similarity or weakness then the classification process will be carried out. Siamese Networks eliminate fully-connected layers to lower the computational amount, and only five full convolution layers are implemented to train end-to-end Siamese networks to learn similarity functions [21]. The significant similarity between class and intra-class variation can be reduced by using conjoined networks [22].

3. Result and Discussion

3.1. Registration Process

The registration process is done by inputting the user's face data into the application system. The face data will be stored in the folder. Figure 3 displays the registration process and the image stored in the folder. First, fill in the Name in the space provided, then click the Take Face Data button to take face data using a camera or webcam. The system will automatically retrieve face data from as many as 20 images.

Table 1. Face Data

Registration Process	The image is stored in the folder.

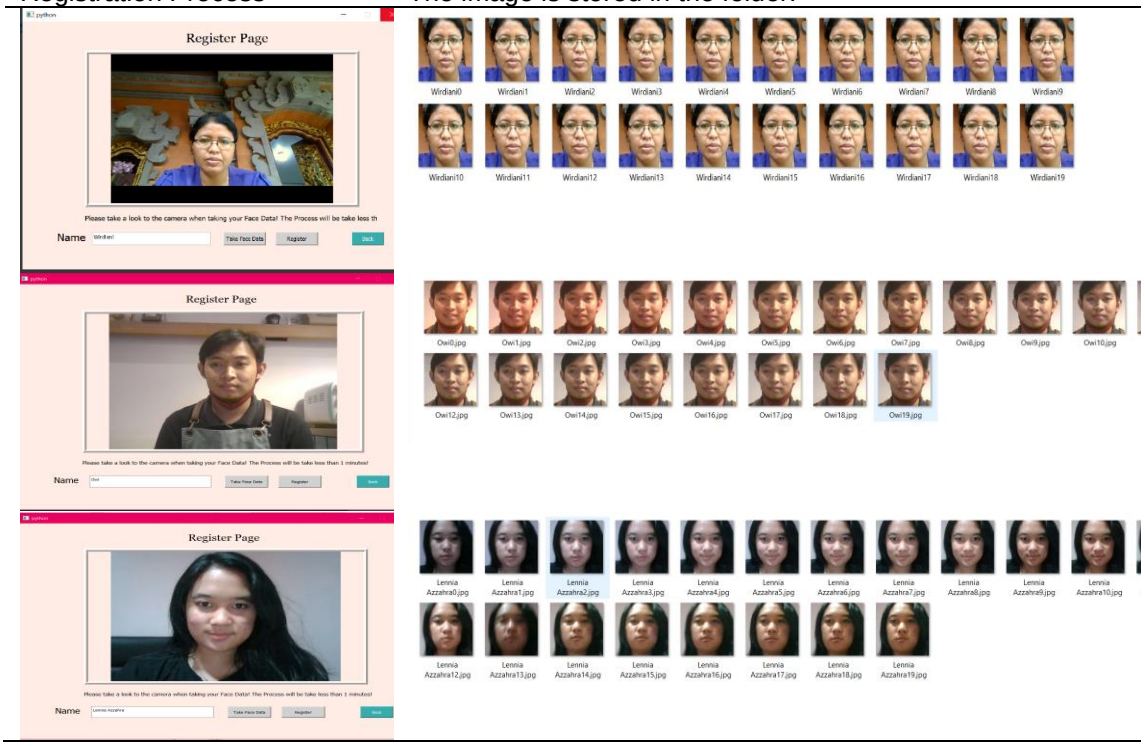


Figure 3. Registration Process

3.2. Recognition Process

The recognition process will automatically detect faces and recognize the user class. The figure below shows the recognition process of the system.

Table 2. Recognition Process

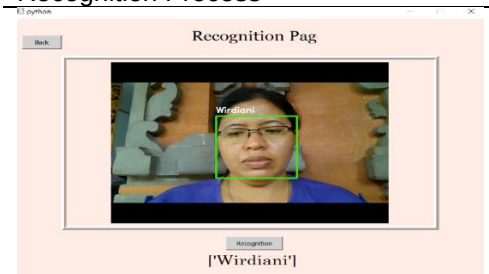
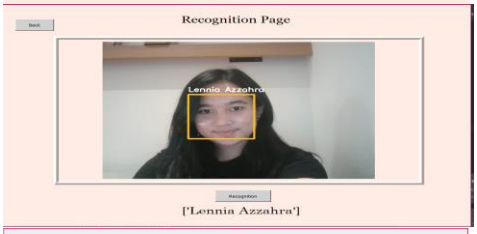
Recognition Process	Condition	Image recognition
	Face with glasses	Face detected 1, result correct.
	Normal	Face detected 1, result correct.
	Face with mask	detected 1, result correct
	Multiple Face	detected: 2, result correct
	Multiple Face mask	detected 2, result correct
	Face rotated to the left	detected 1, result correct
	The face rotated to the right.	Face detected 1, result correct.

Table 2 shows the result of the recognition process, where the system can recognize faces properly according to the name in the registration process. The system can also identify the user's

face when using glasses, with a mask, multiple users, multiple users with masks, and when faces rotate to the left and right. The system can still recognize users with glasses even though, in the registration process, the user is not wearing glasses. This is because of the model Siamese use. The system takes the face feature in the recognition process and compares it with the encoding code in the registration process using Euclidean distance. Then, the triplet loss function is carried out after the distance between the two features is determined. Triplet loss is used to learn the embedding of the face image. If the embeddings are close, they will be identified as the same face image. If the embeddings are close, they will be identified as the same face image. The system can also recognize faces that rotate to the left and the right because the dataset is already augmented so that the system can learn the image with variations in positions.

3.3. Result Analysis

The test results of Real-time Face Recognition using the Deep Learning method get a good result in the recognition process from 50 class tests. The accurate result is 94%, where three classes are recognized as others. The application can also recognize faces using masks and without masks, with glasses and multiperson. The system also can detect the face at a maximum of 90 degrees when looking to the right and left. The eye object is essential as the primary point must be visible. Compared to the previous testing with the SVM algorithm, Deep Learning has a good performance. The result of the live face recognition using SVM is shown in Figure 4 below:

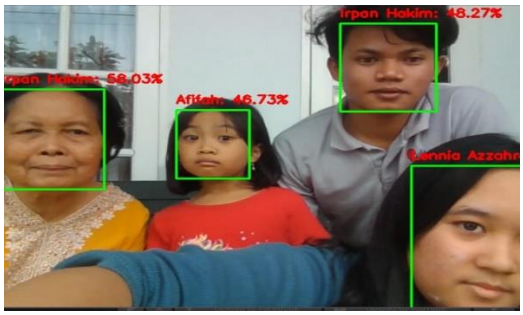


Figure 4. Face Recognition using the SVM method

The SVM method is a supervised learning algorithm whose concept is a simple attempt to find the best hyperplane functions to separate two classes in the input space. Unlike neural networks that search for hyperplanes that separate between layers, SVM tries to find the best hyperplane from the input space. Linear classification is the basic principle of SVM [23]. Compared with the Siamese Network, the SVM method has a relatively high error rate, and the training process requires a lot of images. The following table compares the results of SVM and Siamese Network:

Table 3. Comparison of SVM And Siamese Network

	SVM	Siamese Network
Advantages	• Can detect multiple faces in one frame	• It can detect multiple faces in one frame. • Minimum error rate • Can recognize faces accurately • Can recognize a face with many conditions (using eyeglasses, mask)
Weakness	• Highly error rate • Slower and need large dataset for training • Cannot recognize a face with many conditions (using eyeglasses, mask)	• The quality of the registered image for encoding should be good because it will affect the accuracy of the result
Usage of Data	Large (in the training process using 70 images for every person)	Small (in the training process using 20 images for every person)
Accuracy	The middle range (40%-60%)	94%

Table 3 shows that Deep Learning has higher efficiency and advantages than the SVM method. The range of accuracy results also indicates that the Deep Learning method gives a higher accuracy.

4. Conclusion

Face recognition by using the YoloV5 detection and Siamese Network gives high accuracy in recognizing a face in real-time. The system can also detect faces with glasses, masks, and a multi-person in one frame with minimum error recognition. Siamese Network is the recommended method to carry out the recognition process with good performance, as evidenced by the results of comparison with the previous SVM method.

This research still has shortcomings due to time and other limitations, so there is a suggestion that the following study can be better. The face recognition system does not have a notification in the registration process (training); the time for encoding data while registering still took a long time for extensive data.

References

[1] Darma Putra, *Sistem Biometrika: Konsep Dasar, Teknik Analisis Citra, dan Tahapan Membengun Aplikasi Sistem Biometrika*. 2009.
[2] S. H. Moi *et al.*, "An Improved Approach to Iris Biometric Authentication Performance and Security with Cryptography and Error," *International Journal on Informatics Visualization*, vol. 6, no. August, pp. 531–539, 2022, doi: http://dx.doi.org/10.30630/joiv.6.2-2.1091.
[3] V. M. Arun Ross, Sudipta Banerjee, Cunjian Chen, Anurag Chowdhury, "Some Research Problems in Biometrics: The Future Beckons," *Computer Vision and Pattern Recognition*, 2 019.
[4] R. Blanco-Gonzalo *et al.*, "Biometric Systems Interaction Assessment: The State of the Art," *IEEE Transaction on Human-Machine Systems*, vol. 49, no. 5, pp. 397–410, 2019, doi: 10.1109/THMS.2019.2913672.
[5] G. Rajkumar, V. Garg, A. Anand, and M. Eshasree, "Face Recognition Attendance Marking

using YOLOv3," *Internasional Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 2806–2811, 2020.

[6] A. Malta, M. Mendes, and T. Farinha, "Augmented Reality Maintenance Assistant Using YOLOv5," *Applied Sciences*, vol. 11, pp. 1–14, 2021, doi: 10.3390/app11114758.

[7] K. B. Pranav and J. Manikandan, "Design and Evaluation of a Real-Time Face Recognition System using Convolutional Neural Networks," *Procedia Computer Science*, vol. 171, no. 2019, pp. 1651–1659, 2020, doi: 10.1016/j.procs.2020.04.177.

[8] I. G. P. S. Wijaya, A. Y. Husodo, and I. W. A. Arimbawa, "Real Time Face Recognition Based on Face Descriptor and Its Application," *Telkomnika*, vol. 16, no. 2, pp. 739–746, 2018, doi: 10.12928/telkomnika.v16.i2.7418.

[9] P. Nagaraj, R. Banala, and A. V. K. Prasad, "Real Time Face Recognition using Effective Supervised Machine Learning Algorithms Real Time Face Recognition using Effective Supervised Machine Learning Algorithms," in *Concilio 2021*, 2021, pp. 1–7, doi: 10.1088/1742-6596/1998/1/012007.

[10] Y. Aufar and I. S. Sitanggang, "Face recognition based on Siamese convolutional neural network using Kivy framework," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 2, pp. 764–772, 2022, doi: 10.11591/ijeecs.v26.i2.pp764-772.

[11] C. Song and S. Ji, "Face Recognition Method Based on Siamese Networks Under Non-Restricted Conditions," *IEEE Access*, vol. 10, no. 3, pp. 40432–40444, 2022, doi: 10.1109/ACCESS.2022.3167143.

[12] Y. Zhao and S. Geng, "Face Occlusion Detection Algorithm based on YoloV5," in *2nd Signal Processing and Computer Science*, 2021, pp. 1–6, doi: 10.1088/1742-6596/2031/1/012053.

[13] F. H. Arby, I. Husni, and A. Amin, "Implementation of YOLO-v5 for a real-time Social Distancing Detection," *Journal of Applied Informatics and Computing*, vol. 6, no. 1, pp. 1–6, 2022.

[14] N. Hidayat, S. Wahyudi, and A. A. Diaz, "Individual Recognition Through Face Identification Based On You Only Look Once (YOLOv5) Method)," in *Seminar Nasional Matematika, Geometri, Statistika, dan Komputasi SeNa-MaGeStiK 202*, 2022, pp. 85–98.

[15] N. Dewi and F. Ismawan, "Implementasi Deep Learning Menggunakan Convolutional Neural Network untuk Sistem Pengenalan Wajah," *Faktor Exacta*, vol. 14, no. 1, pp. 34–43, 2021, doi: 10.30998/faktorexacta.v14i1.8989.

[16] D. Setiawan, A. D. Putra, K. Stefani, and J. Felisa, "Implementasi Convolutional Neural Network untuk Facial Recognition," *Media Informatika*, vol. 20, no. 2, pp. 66–79, 2021.

[17] P. V. Lal, U. Srilakshmi, and D. Venkateswarlu, "Face Recognition Using Deep Learning Xception CNN Method," *Journal of Theoretical and Applied Information Technology.*, vol. 100, no. 2, pp. 531–542, 2022.

[18] K. Sari and Y. Arvita, "Perancangan Sistem Absensi Facial Recognition menggunakan CNN dan Liveness Detector pada BPR Central Dana Mandiri," *Jurnal Informatika Dan Rekayasa Komputer*, vol. 1, no. April, pp. 70–80, 2022.

[19] A. Singh, J. Kansari, and V. K. Sinha, "Face Recognition Using Transfer Learning by deep VGG16 model," *Journal of Emerging Technologies Innovative Research*, vol. 9, no. 4, pp. 121–127, 2022.

[20] G. Koch, R. Zemel, and R. Salakhudinov, "Siamese Neural Networks for One-shot Image Recognition," in *International Conference on Machine Learning*, p. 37, 2011.

[21] D. Li, Y. Yu, and X. Chen, "Object tracking framework with Siamese network and re-detection mechanism," *EURASIP Journal on Wireless Communications and Networking*, vol. 261, 2019, doi: https://doi.org/10.1186/s13638-019-1579-x.

[22] A. Nazareth, F. Radilla, K. Ruby, and P. Daniel, "Siamese Convolutional Neural Network for ASL Alphabet Recognition," *Computacion y Sistemas*, vol. 24, no. 3, pp. 1211–1218, 2020, doi: 10.13053/CyS-24-3-3481.

[23] A. Wirdiani, D. Putra, M. Sudarma, and R. S. Hartati, "Palmprint Identification using SVM and CNN Method," in *2021 International Conference on Smart-Green Technology in Electrical and Information Systems (ICSGTEIS)*, pp. 18–23, 2021,doi: 10.1109/ICSGTEIS53426.2021.9650406.