Masked Face Recognition using Deep Learning based on Unmasked Area

Fauzan Firdaus
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
faoezanfirdaoes@gmail.com

Rinaldi Munir
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
rinaldi@informatika.org

2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T) | 978-1-6654-5858-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICPC2T53885.2022.9776651

***Abstract*—Researches related to face recognition find new problems, especially due to current lifestyle change by wearing a face mask. A masked face means that features or information contained in the face are missing. Hence, some modifications should be made to the existing face recognition system in this new situation. On the other hand, some studies indicate that face recognition using partial faces still gives high accuracy, especially on the top-half face (part of the face that contains the forehead to before the nose). Therefore, the proposed training technique uses only partial faces for masked face recognition. The used dataset in this research is our own dataset consisting of 125 subjects. This dataset contains videos taken of people's faces with different devices and backgrounds. Some methodologies from previous studies were used in this study, namely YOLOv4 as a face detection method, pre-trained VGGFace model as a feature extraction method, and artificial neural network as a classification method. Using the same training technique as the previous research, our masked face recognition system achieved a test accuracy of 79.58%. When using the proposed training technique, the obtained test accuracy was 99.53%. It shows that the performance is much better when using a more appropriate training technique in the case of masked face recognition.**

Keywords—Masked Face, Face Recognition, Unmasked Area Training, Deep Learning.

I. INTRODUCTION

The field of artificial intelligence (AI) is one of the rapidly growing computer fields. Computer vision, one of the AI branches, is an example of a topic that is widely studied by researchers. One of the case studies researched on this topic is face recognition. With the ability to recognize a person by their face, the face recognition system has many benefits for some applications. For instance, an organization (such as a school, company, etc.) can automate its attendance system. Face verification is an example of a face recognition system that can verify a person for certain security access, namely smartphone lock, payment, etc.

Face recognition systems in several studies have shown very satisfactory results in their performance. Parkhi et al. [1] have built a facial recognition system using a very large dataset. The used dataset contains approximately 2.6M images and more than 2.6K people. To compare the system performance, they used LFW and YTF datasets for benchmarking. By using an architecture of VGGFace, their system achieved 98.95% accuracy on the LFW dataset, and 97.3% accuracy on the YTF dataset. Angadi et al. [2] proposed a method to build their face recognition system and achieved 95.97% accuracy in the AR Face dataset with 120 people and 97.20% accuracy in the VTU-BEC-DB multimodal database. An example of a face recognition system that has been implemented into the attendance system has been carried out by Arsenovic et al.[3]. They were implementing RFID-based system for checking employee attendance. Based on the real-time environment, they achieved 95.02% of overall accuracy on their small dataset. Another research has been done by Sunaryono et al. [4] by building a face recognition system for the attendance system in the school environment. Their proposed attendance system managed to achieve an accuracy of 97.29%. Masi et al. [5], Wang et al. [6], Guo et al. [7], Ding et al. [8], Li et al. [9], and William et al. [10] have conducted surveys related to face recognition in the past few years. Based on their survey results, the observed face recognition system has obtained satisfactory performance with various proposed methodologies. Apparently, face recognition becomes one of the solved topics in AI development. However, this does not close the gap for researchers to conduct research related to face recognition systems with different cases and conditions.

The COVID-19, which began to emerge at the end of 2019 has become one of the most massive pandemics in human life. The effects of this pandemic are enormous. Every aspect of human life, including economic, social, academic, human lifestyle, and many more are affected. One of the direct effects is the use of a face mask. This aims to protect us from being infected with COVID-19 and it can minimize the spread of COVID-19. Therefore, there are impacts caused by the use of a face mask, especially the existing face recognition system performance in this new situation.

Mundial et al. [11] have built a masked face recognition system. Their proposed system achieved an accuracy of 97% by inserting masked face images into their training data. However, another performance was achieved by the accuracy of 79% without masked face images in their training data. A similar study has been carried out by Anwar et al. [12]. By adding masked face images into their training data, their system managed to increase the true positive rate by around 38%. Another research has been conducted by Li et al. [13] who simulated augmentation of face datasets by adding a non-physical face mask to their dataset with the aim of building a training model with a masked face dataset. With the examples of these studies, a new challenge arises, namely how to build a masked face recognition system that has higher performance without using masked face images inside the training model.

A masked face image means that it has only half of the facial features. The other half's features are covered by the face mask. From these conditions, there is an alleged statement that if the training model is focused on only the top-half face, hence the test performance is expected to increase. In order to check the feasibility of the statement before conducting the experiment, there are some studies that have been found to be supportive. Elmahmudi et al. [14] conducted an experiment of deep face recognition using partial faces. The top-half face has provided a high recognition rate with

978-1-6654-5858-0/22/$31.00 ©2022 IEEE

more than 90%. In the next year, Elmahmudi et al. [15] conducted similar research with the more advanced problems, such as data augmentation. The research conclusion is still the same that the top-half face has a high recognition rate. Another study has been carried out by [16]. By focusing the training data on top-half face images, Their masked face recognition system achieved 88.9% accuracy in the SMFRD dataset and 91.3% accuracy in the RMFRD dataset.

The practical application of this research is to build a masked face recognition using a dataset that has been built ourselves. As previously explained, the training model will only use the top-half of unmasked face images. The purpose of this system is to recognize masked face images from unmasked data while maintaining the performance. The remainder of this paper is structured as follows. In the second section, the proposed system will be explained, including the dataset and methodologies. Section 3 discusses the experimental results, followed by the research conclusion in Section 4.

II. Materials and Methodologies

This section describes our proposed masked face recognition system. All scopes involved in this system, including the used dataset and methodologies will be explained. Each explanation is discussed in detail in the following paras.

A. Dataset Construction

The dataset that has been built contains 250 videos with 125 Indonesian people. There is no age limit for the research target, people are recorded with the age range of children to elderly. Two videos were recorded for each individual, namely videos without and wearing a face mask. The average video duration is about 5 seconds with the fps rate of 30. The unmasked videos will be further used for training and validating steps whereas the masked videos will be used for the testing stage. Fig. 1 shows the frame samples of the dataset.

Fig. 1. Frame samples of the dataset.

B. System Flow

The proposed system flow is divided into two main different schemes. Fig. 2 shows the system flow in the training scheme. The video angle was taken from the left side to the right side of the corresponding person.

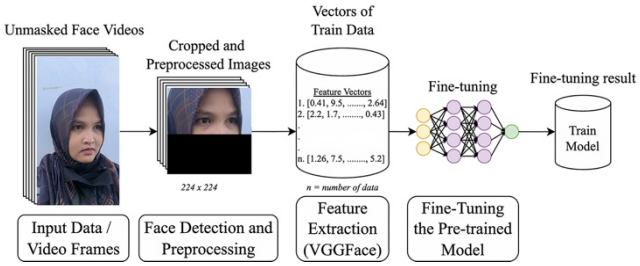

Fig. 2. System flow of training scheme.

By using unmasked face videos for the train data, the videos will be processed by framing process in order to convert the video into a set of images. Furthermore, all images will be processed for face detection using YOLOv4. After obtaining the face image localization, the images will be processed for a preprocessing step by cropping the image in half and concatenating the image with black pixels.

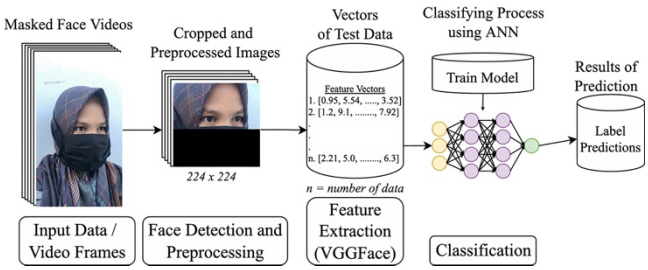

Fig. 3. System flow of testing scheme.

Fig. 3 shows the system flow in the testing scheme. The initial steps taken in the testing scheme are exactly the same as the training scheme, from the framing process up to feature extraction using VGGFace. The next step is doing a classifying process using an artificial neural network (ANN) for predicting people's names or labels. Besides, the cosine similarity method will be implemented as in [14] and [15] for comparison. Finally, the set of predictions will be used to evaluate the system's performance.

C. Face Detection and Preprocessing

There is important process in the face recognition system, namely face detection process. Detecting face intends to localize face on the image. By obtaining the localized face image, the classifier will focus on recognizing faces without considering other features including neck, limbs, and other features that interfere with the person identification.

To carry out the object detection process, especially face detection, there are currently many methodologies proposed by many researchers. One approach that is currently proven to have good performance is YOLO (You Only Look Once) developed by Redmon et al. [17]. In the past few years, YOLO evolved into YOLOv2 [18] and YOLOv3 [19] which managed to achieve better performance and computational time. On the other hand, Bochkovskiy et al. [20] have developed YOLOv4 which makes improvements by implementing state of the art from Bag of Freebies (BoF) and Bag of Specials (BoS). Of all the existing versions, the base

operation remains the same, namely using Convolutional Neural Network. Examples of research related to face detection using YOLO were conducted by Yang et al. [21], Garg et al. [22], Li et al. [23], Silva et al. [24], Chen et al. [25], and Yu et al. [26] which have achieved high detection performance.

Initially, the input image will be divided into S by S grid followed by four different coordinates contained in each cell. The first two coordinates x and y indicate the boundary box position on the center of detection, and the two other coordinates namely w and h indicate the width and the height of its boundary box. There are confidence scores that simply reflect how confident and accurate is the created model. The confidence score equation is as follows.

$$C(X) = \Pr(X) \,.\, IOU_{pred}^{truth} \tag{1}$$

From (1), X indicates the object, Pr is the class probabilities per grid cell, and IOU is the overlapping rate of the predicted box and the ground truth. After getting the confidence score for each prediction box, a lower confidence score against a threshold value will be removed and non-maximum suppression will be performed for removing the remaining bounding box.

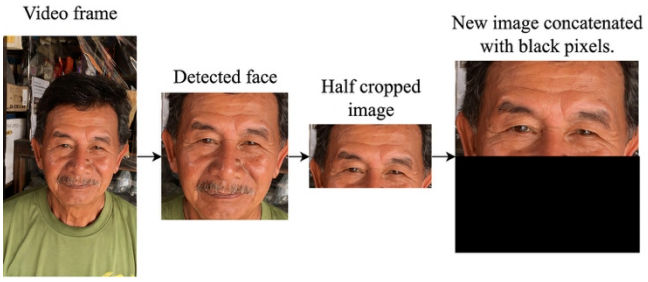

Fig. 4. An example of preprocessing on a single frame.

Fig. 4 shows the example of the preprocessing step on a single frame. From the image result of the face detection process by YOLO, the image will be preprocessed by cropping and resizing the image. Afterward, the image will be concatenated with black pixels. The half-cropped image will be resized into 224 by 112. Finally, the cropped image will be concatenated with black pixels in the same dimensions, which is 224 by 112. Hence, the final dimensions of the image will be 224 by 224. These dimensions are the requirements for the input image of the feature extraction method to be used, namely VGGFace.

D. Face Identification

Another important process in a face recognition system is face identification. Face identification is the vital step of a face recognition system that can provide predictions of people's labels or names from an image. There is one step called feature extraction which performs the process of extracting information from the image.

Referring to studies [14] and [15] which discussed the recognition rate of partial faces, the VGGFace is proven to provide high performance, especially recognition rate on the top-half face. Wang et al. [27] have developed face recognition in real-time (surveillance) application with an accuracy of 92.1% on the fine-tuned model. On the other hand, Ghazi et al. [28] in their research stated that VGGFace showed better performance than other observed deep learning methods. VGGFace itself is a pre-train model from [1] which was trained using a very large scale dataset. VGGFace as a feature extraction step can be used for the classification stage using various methods, such as artificial neural network (ANN) and other conventional classification methods. Fig. 5 shows the architecture of VGGFace. The VGGFace network architecture is equipped with zero padding to avoid dimension reduction.

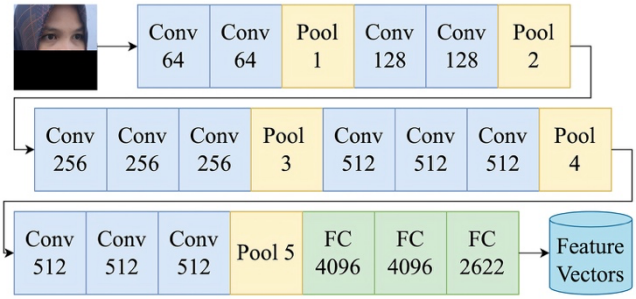

Fig. 5. VGGFace architecture

The obtained feature vectors will further be used for the classification stage using an artificial neural network (ANN) as in [29] with some modifications and softmax function. The cosine similarity method is also used for comparison. According to [14] and [15], cosine similarity has provided a better overall recognition rate compared to another observed classifier.

III. Experimental Results

The proposed masked face recognition system has been built using our dataset. As previously explained, the dataset contains 125 Indonesian people with different age ranges, gender, and video shooting conditions. A video can produce approximately 150 frames. In order to save computational time, 20% of total frames are taken from unmasked videos for the training scheme and another 20% of total frames of masked videos for the testing scheme. Therefore, the average number of observed images for this experiment is 30 per individual for each train and test data. By using the training technique and methodologies described previously, see Table 1 for the performance of the proposed system.

TABLE I. The Performance of Proposed Masked Face Recognition System

Schemes	Feature Extraction Method(s)	Classification Method(s)	
		ANN	Cosine Similarity
Training	VGGFace	100%	100%
Validation		100%	99.87%
Testing		99.53%	98.88%

From Table 1, the overall performance of the proposed system has excellent accuracy. Since the used feature extraction method was VGGFace, which is a pre-trained model from a very large scale dataset, the achieved performance is reasonable. In addition to comparing a masked face recognition system using the proposed training technique, a training technique in [11] has also been experimented on our dataset. The performance can be seen in Table 2. From Table 2, there is a big improvement in performance when using the

proposed training technique. The best accuracy of the previous training technique was achieved by the cosine similarity classification method by 79.58%. This accuracy shows that the proposed training technique is proven that it can increase accuracy up to approximately 99% with the same methodologies and dataset.

TABLE II. THE PERFORMANCE OF MASKED FACE RECOGNITION SYSTEM USING TRAINING TECHNIQUE IN [11]

Schemes	Feature Extraction Method(s)	Classification Method(s)	
		ANN	Cosine Similarity
Training	VGGFace	100%	100%
Validation		100%	100%
Testing		76.78%	79.58%

Table 3 shows the comparison of other evaluation metrics in the system using the proposed training technique and the training technique in [11].

TABLE III. CLASSIFICATION REPORT COMPARISON ON 2 DIFFERENT TRAINING TECHNIQUES

Training Technique	Classification Report	Classification Method(s)	
		ANN	Cosine Similarity
Proposed Training Technique	Precision	99.60%	99.20%
	Recall	99.53%	98.88%
	F1-Score	99.53%	98.80%
Training Technique in [11]	Precision	89.57%	81.03%
	Recall	76.78%	79.58%
	F1-Score	73.32%	76.20%

In addition to building a masked face recognition system and evaluating the performance, the built model was tested in real-time applications as well. First, the model was tested on a webcam. In this experiment, there were 3 different lighting conditions. The process in the system was to find a boundary box containing one or more faces. With a face detected by YOLO, the system will label the face with "without_mask" or "with_mask", along with each confidence score. After that, the system will predict the person's name on the corresponding face. If the face detection label is "with_mask", then the training model that will be used is the top-half face model. However, if the detected label is "without_mask" then the used training model is the full-face model.

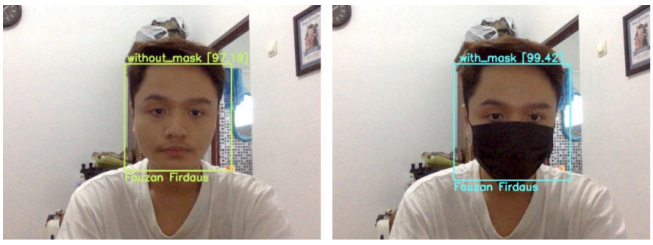

Fig. 6. Experimental results on a webcam with normal lighting.

An example of the labeling results on a sample frame can be seen in Fig. 6. The label on the top side of the boundary box is the face detection label followed by the confidence score, while the label on the bottom side of the boundary box is the predicted name of the corresponding face. The actual label or name for the person in this experiment is "Fauzan Firdaus". In this normal lighting condition, the system can label each frame correctly. Another lighting condition tested in this experiment was using a smartphone flashlight. Fig. 7 shows the experimental results.

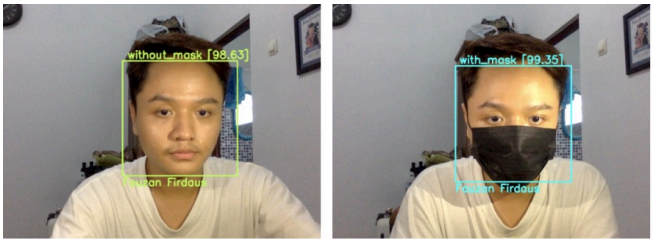

Fig. 7. Experimental results on a webcam with additional lighting

As can be seen in Fig. 7, the system can correctly label the boundary box of the detected image. Other than that, the low-light condition has also been experimented. Fig. 8 shows the experimental results. As can be seen, the system still labels correctly on each generated frame. Apparently, the proposed model is robust for different lighting conditions in the use of a webcam for this real-time application experiment.

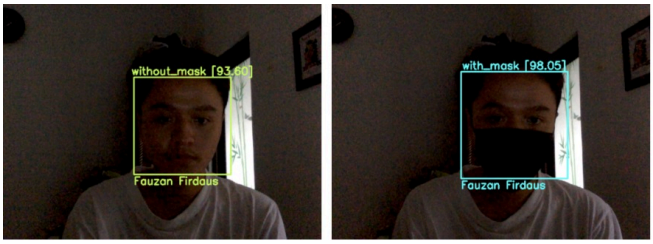

Fig. 8. Experimental results on a webcam with low-light condition.

The second experiment which is more advanced has been conducted in the form of a surveillance masked face recognition system. With the same mechanisms as the previous experiment, the system will detect the faces contained in each frame in the video. Finally, the label or name predictions will be given to each detected face. The input used in this experiment is a video containing 3 people with different distances. Fig. 9 shows a sample frame of a correctly labeled video output.

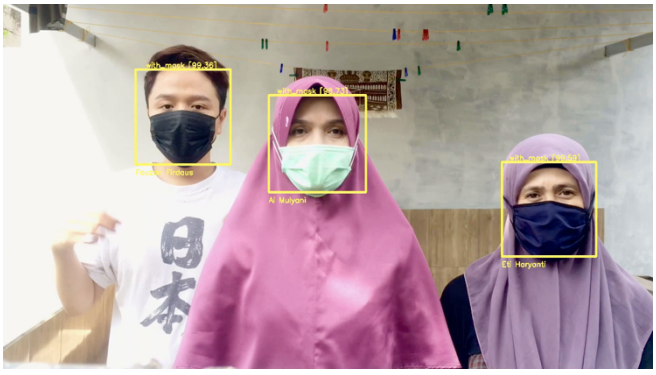

Fig. 9. A correct output sample on the second real-time experiment.

The actual labels for the people in the video from the left are "Fauzan Firdaus", "Ai Mulyani", and "Eti Haryanti". As can be seen in Fig. 9 that the example of an output frame was labeled correctly. However, if the distance of the people with the camera is too far, the system was still difficult to predict the correct label. Fig. 10 shows the sample output with a

longer distance and is labeled incorrectly. Although the person on the right side was labeled correctly, the other two people were still labeled incorrectly.

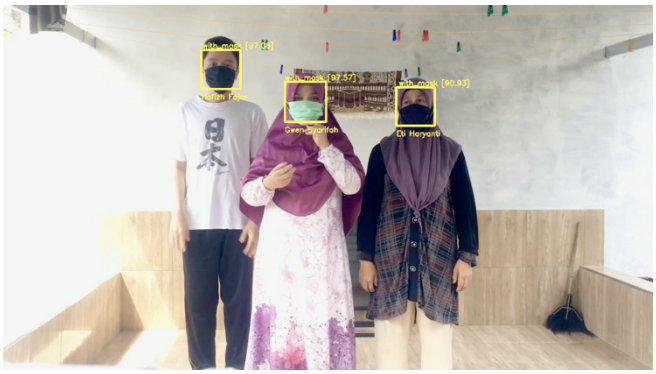

Fig. 10. An incorrect output sample with a longer distance.

With the same distance as in Fig. 10, the system could predict the labels of the unmasked face correctly. Fig. 11 shows an example of the correct output sample with unmasked faces from a longer distance. Based on the actual label in each person, the model that has been built is robust for a face that is not covered by a mask with a longer distance.

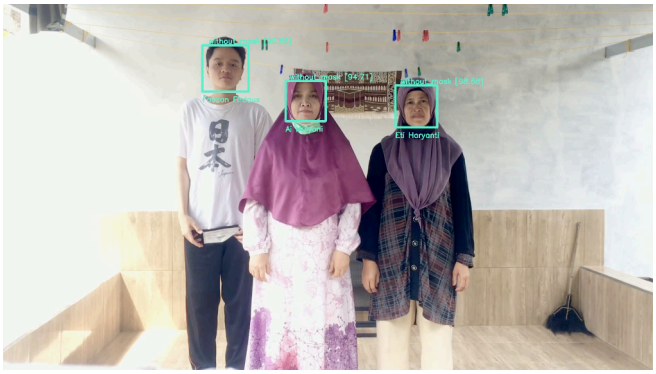

Fig. 11. A correct output sample with unmasked faces.

From the overall experimental results, the model still has a limitation in predicting frames on the masked face with lower resolutions. The farther a person is from the point of the camera, the smaller the resolution of the resulting face frame will be. Since the VGGFace was used for building the model, the optimal face image input resolution is approximately 224 by 224 and will be better if the resolution is more than that.

IV. Conclusion

In this paper, the proposed masked face recognition system performs training that only focused on the top-half part of the face, or the part that is not covered by the face mask. This mechanism is similar to human vision. When someone tries to recognize people they knew, usually the face mask is not considered as the person's identity. The performance of the system has been proven to have increased compared to using the previous training technique in [11]. On the other hand, the built model was also tested in real-time applications, namely using a webcam and surveillance system. Although the surveillance system that has been built still has an output with the wrong label, the model still predicts correctly on a webcam with different lighting conditions. In the future, the dataset used needs to be increased by the number of subjects. In addition, the masked face recognition model needs to be retrained with more perfect preprocessed data. For instance, histogram equalization for contrast adjustment, or conducting some data augmentation. The goal is to widen the variety of training model, hence the model will be more robust and accurate for predicting in more advanced systems.

References

[1] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in *British Machine Vision Conference*, 2015.

[2] S. A. Angadi and S. M. Hatture, "Face Recognition through Symbolic Modeling of Face Graphs and Texture," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 33, 2019.

[3] M. Arsenovic, S. Sladojevic, A. Anderla, and D. Stefanovic, "FaceTime - Deep learning based face recognition attendance system," *SISY 2017 - IEEE 15th Int. Symp. Intell. Syst. Informatics, Proc.*, pp. 53–57, 2017.

[4] D. Sunaryono, J. Siswantoro, and R. Anggoro, "An android based course attendance system using face recognition," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 3, pp. 304–312, 2021.

[5] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep Face Recognition: A Survey," *Proc. - 31st Conf. Graph. Patterns Images, SIBGRAPI 2018*, pp. 471–478, 2019.

[6] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021.

[7] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Comput. Vis. Image Underst.*, no. 189, p. 102805, 2019.

[8] C. Ding and D. Tao, "A comprehensive survey on Pose-Invariant Face Recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, 2016.

[9] P. Li, L. Prieto, D. Mery, and P. Flynn, "Face Recognition in Low Quality Images: A Survey," *arXiv:1805.11519*, 2018.

[10] I. William, D. R. Ignatius Moses Setiadi, E. H. Rachmawanto, H. A. Santoso, and C. A. Sari, "Face Recognition using FaceNet (Survey, Performance Test, and Comparison)," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, 2019.

[11] I. Q. Mundial, M. S. Ul Hassan, M. I. Tiwana, W. S. Qureshi, and E. Alanazi, "Towards facial recognition problem in COVID-19 pandemic," *2020 4th Int. Conf. Electr. Telecommun. Comput. Eng. ELTICOM 2020 - Proc.*, pp. 210–214, 2020.

[12] A. Anwar and A. Raychowdhury, "Masked Face Recognition for Secure Authentication," *arXiv:2008.11104*, 2020.

[13] Y. Li, K. Guo, Y. Lu, and L. Liu, "Cropping and attention based approach for masked face recognition," *Appl. Intell.*, vol. 51, no. 5, pp. 3012–3025, 2021.

[14] A. Elmahmudi and H. Ugail, "Experiments on deep face recognition using partial faces," *Proc. - 2018 Int. Conf. Cyberworlds, CW 2018*, pp. 357–362, 2018.

[15] A. Elmahmudi and H. Ugail, "Deep face recognition using imperfect facial data," *Futur. Gener. Comput. Syst.*, pp. 213–225, 2019.

[16] W. Hariri, "Efficient Masked Face Recognition Method during the COVID-19 Pandemic," *arXiv:2105.03026*, 2021.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 779–788, 2016.

[18] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, pp. 6517–6525, 2017.

[19] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv:1804.02767*, 2018.

[20] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv:2004.10934*, 2020.

[21] Y. Wang and J. Zheng, "Real-time face detection based on YOLO," *1st IEEE Int. Conf. Knowl. Innov. Invent. ICKII 2018*, vol. 2, pp. 221–224, 2018.

[22] D. Garg, P. Goel, S. Pandya, A. Ganatra, and K. Kotecha, "A Deep Learning Approach for Face Detection using YOLO," *2018 IEEE Punecon*, 2018.

[23] C. Li, R. Wang, J. Li, and L. Fei, "Face Detection Based on

YOLOv3," *Adv. Intell. Syst. Comput.*, vol. 1031, pp. 277–284, 2020.

[24] L. P. e Silva, J. C. Batista, O. R. P. Bellon, and L. Silva, "YOLO-FD: YOLO for Face Detection," in *Iberoamerican Congress on Pattern Recognition*, vol. 11896, 2019, pp. 209–218.

[25] W. Chen, H. Huang, S. Peng, C. Zhou, and C. Zhang, "YOLO-face: a real-time face detector," *Vis. Comput.*, vol. 37, pp. 805–813, 2021.

[26] J. Yu and W. Zhang, "Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4," *Sensors*, vol. 21, no. 3263, 2021.

[27] W. Ya, T. Bao, D. Chunhui, and M. Zhu, "Face recognition in real-world surveillance videos with deep learning method," *2017 2nd Int. Conf. Image, Vis. Comput. ICIVC 2017*, pp. 239–243, 2017.

[28] M. M. Ghazi and H. K. Ekenel, "A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 102–109, 2016.

[29] Y. Wu, J. Li, Y. Kong, and Y. Fu, "Deep convolutional neural network with independent softmax for large scale face recognition," *Proc. 24th ACM Int. Conf. Multimed.*, pp. 1063–1067, 2016.