Comparative Analysis of Masked and Unmasked for Face Recognition Using VGG Face and MTCNN

Hanif Naufal Arif Sunarko[1], Risanuri Hidayat[2], Rudy Hartanto[3]

{ hanif.n.a@mail.ugm.ac.id[1] , risanuri@ugm.ac.id[2], rudy@ugm.ac.id[3]}

Department of Electrical Engineering and Information Engineering of Engineering Gadjah Mada University, Komplek Fakultas Teknik UGM, Grafika street No.2 Yogyakarta 55281 INDONESIA[1,2,3]

Abstract. Face recognition is a system that is widely used in various fields such as security, attendance system, and other fields. Currently Covid-19 is still a major problem around the world and almost everyone is protecting themselves with masks. This is a problem for the face recognition system. This happen because most of the faces are covered by masks so that face recognition system will be difficult to recognize the face. This paper will do a comparison between a dataset without a mask and a mixed dataset. This study was conducted to find out how the effect of the dataset used on the accuracy of face recognition system either with masks or without masks and to find out how well the performance of face recognition with different dataset. VGG Face and MTCNN are used to detect and recognize faces based on landmarks. This study compares the level of accuracy, level of precision and level of sensitivity. The result shows that using a mixed dataset containing masked and unmasked faces will increase the accuracy rate from 86.7% to 93.3%. For the level of precision increased from 87.7% to 93.5%. And the Sensitivity level increased from 86.7% to 93.3%.

Keywords: face recognition, covid-19, mask, dataset, VGG-Face, MTCNN

1 Introduction

Face recognition is a system that can detect and then identify a person face. Face recognition is a system that is most widely used for security. This happens because face recognition is a system that easy to use to recognize a person because the person does not need to do anything, the system will detect his face. While in fingerprint recognition, the person needs to put his finger on the detector. Face recognition is also widely used in forensics, access control, and attendance systems. Face recognition has gained many interests nowadays, and it is expected to replace some biometric applications [1]. Currently, the development of face recognition has advanced through a lot of research, the accuracy and speed of recognition are getting better and faster. However, the face recognition system has drawbacks because it still depends on the light level, the pose of the person being detected and the person's expression. These things can reduce the accuracy of face recognition system. The face recognition system has several steps in doing the

ICAE 2022, October 05, Batam, Indonesia

DOI 10.4108/eai.5-10-2022.2327473

recognition, first face detection, where the system will detect the presence of a face, then face alignment, where the detected face will then be straightened out which then will be detected several special face landmark for example eyes, nose, and mouth. Then the system will compare these points with the face points in the system. Currently, where Covid-19 is still a scourge in various parts of the world, people are wearing masks in public areas. This causes the face recognition system to have difficulty recognizing faces because most of the faces are covered, so there are fewer facial features that can be used by face recognition. In Fig. 1 we can see that some face landmark are covered with mask for example is the mouth and nose. In addition, the use of various forms of masks is also disturbing in face detection because the shape of the face will change because it is covered by the shape of the mask. An example is the mask-covered chin and cheekbones reducing the detected landmarks.

Fig. 1. Masked Person Reduce Detected Face Landmark

There are many models that can be used for face recognition. In the development of face recognition, CNN is one of the most frequently used. Besides being used for face recognition, CNN is also used for object detection, image recognition, image classification, and of course for face identification. Convolutional Neural Network (CNN) is a Deep Learning algorithm that can train large data sets with millions of parameters and take the form of 2D images as input, and combine it with filter to produce the desired output. CNN is a type of neural network that dominates many computer vision tasks and has attracted attention in many fields. CNN aims to automatically and adaptively study the spatial hierarchical structure of elements by using backward propagation of several building blocks. CNN has several layers namely convolutional layer, connection layer, and fully connected layer. Convolutional neural networks also contribute to facial recognition by providing powerful classifiers [2]. Besides being able to be used on images, CNN can also be used for speech recognition. In this study one of the convolutional neural networks (CNN) models will be used, namely VGG Face because it has high level of precision [3][10][11]. The use of VGG Face combined with face detection can increase the accuracy to recognize faces [13] .And for the face identification, Multitask Cascade Convolutional Neural Network (MTCNN) will be use.

The analysis was carried out by collecting datasets in the form of photos of people's faces using masks and without using masks. Then create datasets, namely training datasets without masks, mixed datasets with masks and without masks, and test datasets. Furthermore, the face is

detected using MTCNN and VGG Face to get embeddings. Furthermore, the test dataset will be used as a prediction, the results obtained are then entered into the confusion matrix and then analyzed for accuracy, precision and sensitivity. After getting the results for the two datasets, they are compared to find out the difference in the level of accuracy between the datasets.

2 Methodology

In this study, there are 3 processes carried out to obtain data that can be compared. The first step is to collect photos for dataset, the second step is to perform face detection from the given image and process image to use in the next process. The last is to extract facial landmarks such as eyes, nose and mouth to use facial recognition by comparing the extract results to existing data.

2.1 Image Collection for Dataset

The first step is to collect images to be used as a dataset. The images collected are photos of people wearing masks and without masks, which are taken by taking photos of people from the front and facing forward. Public datasets were not used because the datasets available at the time of the research were not found to be suitable. Where the face is facing forward and not tilted.

2.2 Face Detection using MTCNN and Image Processing

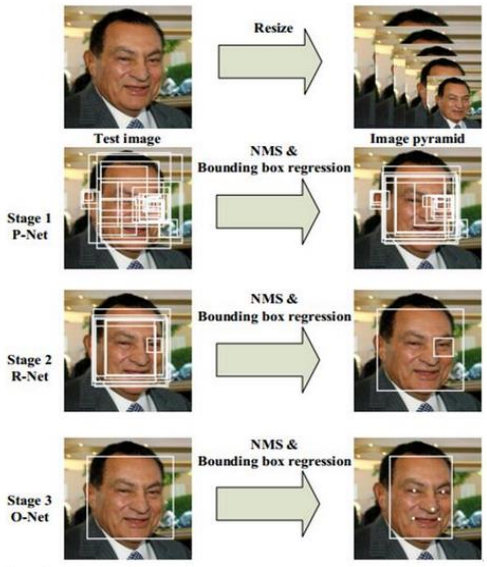

Fig. 2. Pipeline of the Multi-Task Cascaded Convolutional Neural Network [7]

Face detection is important when doing face recognition. If there is no face detected, the face recognition system cannot continue its recognition. In this study the Multi-task Cascaded Convolutional Neural Network (MTCNN) is used to perform face detection. MTCNN is a model for detecting faces. MTCNN is used to detect faces from a given image and then generates a high dimensional facial descriptor. There are 3 stages of neural network detector on MTCNN.

Before the first stage starts, the given image is resized many times so that faces of different sizes can be detected. In the first stage, called P-Net, the first detection is carried out. This first detection has a low level of accuracy so that there are many wrong bounding boxes. In the second stage, which is called R-Net, a selection is made on the results from the first stage, so as to get a more accurate bounding box. In the last stage called O-Net, the results from the previous stage are refined again so that it gets a bounding box that is more accurate than before. The three stages can perform facial classification, bounding box regression and detect facial landmarks such as eyes, nose, and mouth. All the steps in MTCNN can be seen in Fig. 2.

MTCNN was chosen because it has a high level of accuracy for face detection [4][9] and efficiency in pipeline-based image recognition [5]. In addition, MTCNN can also detect faces even if the face is partially covered with hands or other objects [6]. This is very useful for detecting and recognizing faces using masks. In this study, the points extracted from the face between the mask and without the mask were the same. With MTCNN will detect the eyebrows on both eyes, mouth on both ends and nose. MTCNN will predict the layout of the points of the mouth and nose as seen in Fig 3.

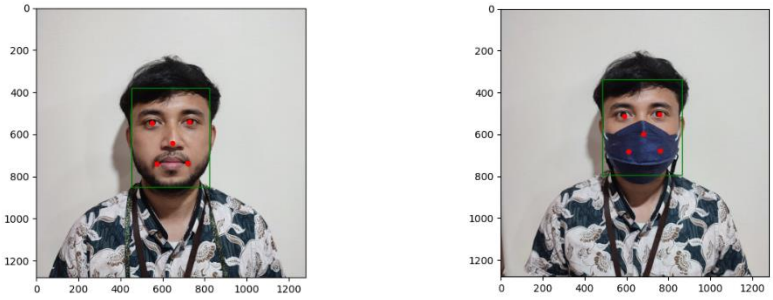

Fig. 3. MTCNN face detection on unmasked face and masked face.

After face detection is complete, the next step is to cut and adjust the image size from the MTCNN results. The bounding box generated from MTCNN is used to crop the image face from the entire image. Before image can be used to next step which is Face Recognition image need to be resize first. To be able to be used on VGGFace, an image with a size of 224 x 224 pixels is required. For this reason, the image is cropped then resized according to these needs.

2.3 Face Landmark Extraction using VGGFace and Comparing Face Embedded

VGGFace is a model developed for face recognition. There are several models in VGGFace, namely VGG16, Resnet50, and Senet50. In this study, Resnet50 will be used to perform landmark extraction. Resnet50 was chosen because it has a higher level of accuracy than other models [12]. Then the face embedding is predicted from the given model in the form of a vector. Then the length of the vector is normalized using L2 vector normalization. These are referred as 'face descriptors'. Distances between face descriptors were calculated using the cosine equation [8]. Then the face embedding is saved to be used as a comparison.

After obtaining the face embedding of each dataset. Then a test is conducted to test the level of accuracy, sensitivity, and speed of identification. To carry out the test, an appropriate person face image is used that is not included in training dataset. The images used are 6 different images for each person, namely with a mask and without a mask. The image provided is then extracted with the face embedding to be matched with what is already in the data.

3 Dataset

In this study, data were obtained by taking photos of the front facing faces of several people. The photo that taken is the face without a mask and face with a mask. There were photos from 15 people, each of which had 13 photos without masks and 13 photos with masks, for a total of 390 photos. Then the dataset will be split into two dataset namely test dataset and training dataset. The test dataset consists of 3 photos without masks and 3 photos with masks for each person. For the training dataset, 2 different datasets will be created, where the first is a dataset containing photos of 10 faces without masks, while the second dataset contains a mixture of 5 face photos without masks and 5 using masks. In this study, we did not use a public dataset because there was no suitable dataset where the face was facing straight ahead when wearing or not wearing a mask. In addition, to ensure the results obtained from the balance dataset. After the dataset is obtained, training will be carried out to create an embedding model using each dataset. An example of the image used in dataset 1 can be seen in fig. 4 and an example of the image used in dataset 2 can be seen in fig. 5.

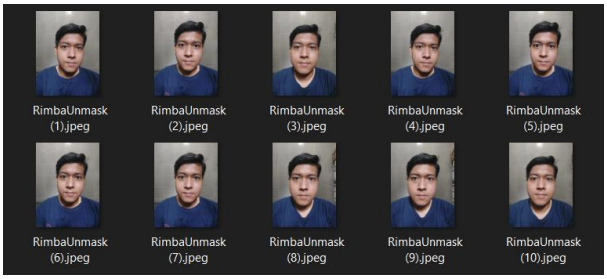

Fig. 4. Example of Image Used in Dataset 1

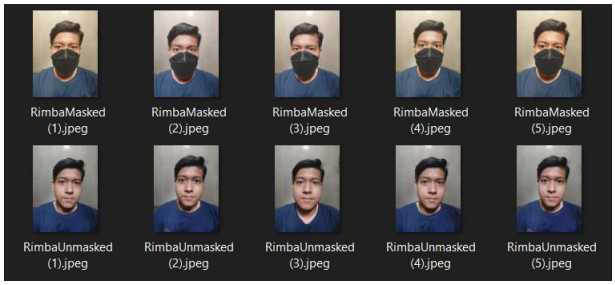

Fig. 5. Example of Image Used in Dataset 2

4 Result and Discussion

After obtaining the embedding model from the two datasets, which are multiclass classifications and the test dataset has been prepared, a test is carried out to obtain the level of accuracy, level of precision, and level of sensitivity. Both embedding models will be tested using the same test dataset which is a mixture of face photos without masks and using masks. To get the level of accuracy, level of precision and level of sensitivity used confusion matrix for multiclass classification with formulas (1), (2), and (3).

$$\text{Accuracy} = \frac{\text{TP}}{\text{Total Test Dataset}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP}+\text{FP})} \quad (2)$$

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP}+\text{FN})} \quad (3)$$

True Positive (TP) is obtained from the number of face predictions that are detected correctly. For False Positive (FP) it is obtained from the number of predictions made to detect the wrong face. Meanwhile, False Negative (FN) is obtained from the number of faces that are detected incorrectly. and the total test dataset is the number of photos in the test dataset or the number of predictions made.

From the tests that have been carried out on the embedding of the first model trained using a training dataset containing a face without a mask, the results are shown in Table 1. It was obtained from the test that the sensitivity level of all classes was 86.7% and the precision level of all classes was 87.7%. Furthermore, the results obtained for the accuracy level of the first model on the test dataset, which is 86.7%.

Table 1. Level of Precision and Sensitivity Masked Dataset

Class	Sensitivity	Precision
Hanif	0,67	0,57
Azka	0,83	1,00
Kresna	0,67	1,00
Rizky	0,50	0,60
Taufik	1,00	1,00
Idos	1,00	0,75
Friska	1,00	1,00
Ulfa	0,50	0,75
Yusak	1,00	1,00
Grace	1,00	1,00

Class	Sensitivity	Precision
Yuda	1,00	1,00
Rimba	1,00	0,86
Adhi	1,00	1,00
Siska	1,00	1,00
Hesa	0,83	0,63

The next step is to test the second embedding model that trained using a mixed dataset between masked and unmasked faces. The results of the test can be seen in Table 2. The sensitivity level of this model for all classes is 93.3% and the precision level is 93.5%. Furthermore, for the results of the test the accuracy of this model is obtained at 93.3%.

Table 2. Level of Precision and Sensitivity Mixed Dataset

Class	Sensitivity	Precision
Hanif	0,83	0,71
Azka	1,00	1,00
Kresna	1,00	1,00
Rizky	0,67	0,80
Taufik	1,00	1,00
Idos	1,00	1,00
Friska	1,00	1,00
Ulfa	0,67	0,80
Yusak	1,00	1,00
Grace	1,00	1,00
Yuda	1,00	1,00
Rimba	1,00	1,00
Adhi	1,00	1,00
Siska	1,00	1,00
Hesa	0,83	0,71

From the tests that have been carried out and the results that have been processed, the increase in the level of accuracy, precision and sensitivity on the dataset containing a mixture of masked and unmasked faces can occur because MTCNN provides additional prediction points for faces with masks to be processed at VGG Face. So that by using a mixed dataset, the face recognition system can be more accurate in recognizing the face of someone who is wearing a mask.

5 Conclusion

This study aims to determine whether the use of the image of a masked person in the dataset can affect the performance of the facial recognition system. We tested with two different datasets where the first dataset is people without masks and the second dataset is a mixture of images of people with masks and without masks. It was found that using a dataset with a mix of people wearing masks can increase the accuracy of facial recognition to recognize people wearing masks. In addition, it is also obtained by using a mixed dataset of people with masks, also increasing the level of precision and level of sensitivity. By using the second dataset, masked faces can be recognized better with an increase from 86.7% using the first dataset to 93.3% using the second dataset. However, with a mixed dataset of masked and unmasked faces, prediction errors still occur on faces that have similar landmarks. We conclude that the use of a mixed dataset between people with masks and without masks can increase the accuracy level of face recognition system to recognize people even though they are wearing masks.

References

[1] Puthea, Khem, Rudy Hartanto, and Risanuri Hidayat. : The Attendance Marking System based on Eigenface Recognition using OpenCV and Python. Journal of Physics: Conference Series. Vol. 1551. No. 1. IOP Publishing (2020)

[2] Puthea, Khem, Rudy Hartanto, and Risanuri Hidayat. : A review paper on attendance marking system based on face recognition. 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE, (017)

[3] Y. B. Chandra and G. K. Reddy. : A Comparative Analysis Of Face Recognition Models On Masked Faces. International Journal of Scientific & Technology Research Vol 9, 2020

[4] Mool, Akshay, J. Panda, and Kapil Sharma. : Optimizable face detection and tracking model with occlusion resolution for high quality videos. Multimedia Tools and Applications 81.8 (2022)

[5] An, Xiangjing, Wensen Chang, and Xiangdong Chen. : Multi-layer template correlation neural network for recognition of lane mark based on pipelined image processing structure. Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C). Vol. 3. IEEE, (1999)

[6] Lindner, Tymoteusz, et al. : Face recognition system based on a single-board computer. 2020 International Conference Mechatronic Systems and Materials (MSM). IEEE (2020)

[7] K. Zhang, Z. Zhang., Z. Li, and Y. Qiao, : Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters 23, no.10, (2016)

[8] Cao, Qiong, et al. : Vggface2: A dataset for recognising faces across pose and age. 13th IEEE international conference on automatic face & gesture recognition (FG 2018). IEEE, (2018)

[9] Wu, Chunming, and Ying Zhang. : MTCNN and FACENET based access control system for face detection and recognition. Automatic Control and Computer Sciences 55.1 (2021)

[10] Gyawali, Dipesh, et al. : Age Range Estimation Using MTCNN and VGG-Face Model. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, (2020).

[11] Acien, Alejandro, et al. : Measuring the gender and ethnicity bias in deep models for face recognition. Iberoamerican Congress on Pattern Recognition. Springer, Cham, (2018).

[12] S. Mhadgut : Masked Face Detection and Recognition System in Real Time using YOLOv3 to combat COVID-19. 12th International Conference on Computing Communication and Networking Technologies. (2021)
[13] F. Firdaus and R. Munir : Masked Face Recognition using Deep Learning based on Unmasked Area. Second International Conference on Power, Control and Computing Technologies. (2022)