International Journal of Advanced Culture Technology Vol.10 No.1 294-301 (2022)
DOI https://doi.org/10.17703/IJACT.2022.10.1.294

IJACT 22-3-36

Study On Masked Face Detection And Recognition using transfer learning

[1]NaeJoung Kwak, [2]DongJu Kim*

[1]Prof., Dept. of Cyber and Security , Baejae Univ., Korea
[2]Professor, Postech Institute of Artificial Intelligence, POSTECH
knj0125@pcu.ac.kr, kkb0320@postech.ac.kr

Abstract

COVID-19 is a crisis with numerous casualties. The World Health Organization (WHO) has declared the use of masks as an essential safety measure during the COVID-19 pandemic. Therefore, whether or not to wear a mask is an important issue when entering and exiting public places and institutions. However, this makes face recognition a very difficult task because certain parts of the face are hidden. As a result, face identification and identity verification in the access system became difficult. In this paper, we propose a system that can detect masked face using transfer learning of Yolov5s and recognize the user using transfer learning of Facenet. Transfer learning preforms by changing the learning rate, epoch, and batch size, their results are evaluated, and the best model is selected as representative model. It has been confirmed that the proposed model is good at detecting masked face and masked face recognition.

Keywords: *Masked Face Detection, Mask Detection, Object Detection, Yolo, Masked Face Recognition*

1. INTRODUCTION

Recently, COVID-19 has spread around the world. COVID-19 can be spread through coughing, sneezing, or respiratory droplets while talking to an infected person[1]. It can be also spread by touching a surface or object that has the virus on it and then touching your mouth, nose, or eyes[2]. The two main ways to prevent COVID-19 are avoiding unnecessary contact and wearing a mask. Therefore, the use of face masks has become an important part of our lives. However, since most people identify non-masked people with their own eyes, manpower is required, and sometimes there may be parts that cannot be checked. Therefore, automatic detection of people of masked face is important. However, masked face makes serious problems for facial recognition systems used to unlock phones[3][4] and confirm attendance in public places and schools or offices.

Current deep learning-based face recognition systems have demonstrated excellent accuracy[5][6][7]. The accuracy of these systems depends on the characteristics of the training images available, and conventional systems learn important facial features such as eyes, nose, lips, face edges, etc. However, these systems cannot identify a person if they have a face mask. Therefore, with conventional face recognition systems, if you are wearing a mask, you have to take it off for identification. This not only makes the authentication process cumbersome for users, but also increases the risk of virus transmission. Therefore, it is necessary to develop a system that can verify masked face identification.

Manuscript received: February 23, 2022 / revised: March 1, 2022 / accepted: March 8, 2022
Corresponding Author: kkb0320@postech.ac.kr
Tel:*** - **** - ****
Professor, Postech Institute of Artificial Intelligence, POSTECH, korea

In this paper, we implement a system that determines whether a mask is worn or not and checks the identity of the person who enters. To determine whether a mask is worn or not, we derive the model to detect masked face. To confirm the identity of people wearing mask, we use face recognition. Therefore, we implement a system that detects masked face of people by transferring the S model of YOLOv5[8] among deep learning techniques for real-time object detection. In addition, by transfer learning of facenet[5] among face recognition models, we implement a system that can confirm the identity of people of masked face.

2. RELATED WORKS

2.1 YOLOv5

The algorithm applied in this study is CNN-based YOLO[9][10], which is the latest deep learning-based object detection algorithm that released version 1 in 2016, and is a representative model of a one-stage detector[11]. As the name suggests, one-stage detector is a method to detect objects by looking at an image only once and calculating the probability of an object to be found for each grid cell of the convolution layer and the class of the corresponding grid cell, and performing classification and localization at the same time. This method has a fast detection speed because the existing detection process is replaced with a single regression problem.

YOLOv5[8] shows excellent performance in terms of FPS and mAP by using a different backbone from the previous YOLO series. There are 4 types of backbone of YOLOv5, from the smallest and lightest Yolov5s(small) to Yolov5m(medium), YOlov5l(large), and Yolov5x(xlarge). For each model, the accuracy gradually improves from s to x, while the speed slows down.

2.2 Facenet[5]

Face recognition is performed by extracting a face region, extracting a feature from the face region, and recognition or verification through matching. Figure 1 shows the flow of the face recognition system.

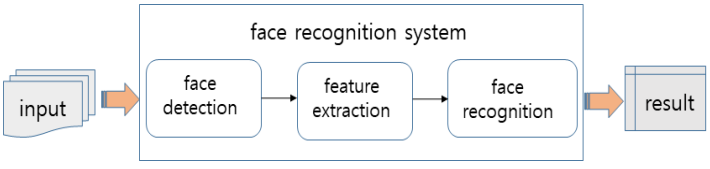
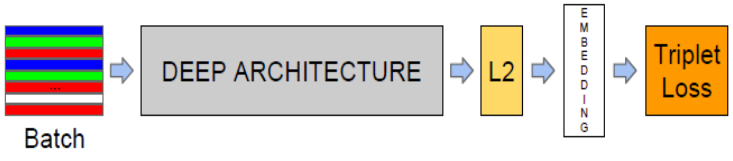
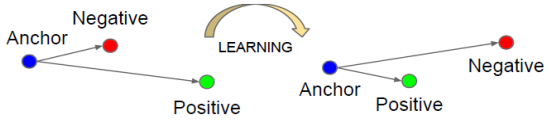


Figure 1. Flow of face recognition system



(a) Facenet model structure



(b) learning through triplet Loss

Figure 2. Facenet[5]

Face detection is being developed from feature-based face detection to deep learning-based face detection. As for the deep learning-based face detection algorithm, models using the Multi-Task Learning [12][13] strategy show good performance in the face detection field. MTCNN[13] is a representative face detector that aligns and rotates face regions to detect five face landmark positions.

In the face recognition technique, various face recognition methods have been developed along with the development of object recognition deep learning techniques[5][6][7]. The face recognition technique learns the features of face with CNN and embeds it so that the distance between the same faces is close and the distances between different faces are far in vector space.

Facenet is a representative method of face recognition. Given a face image, it extracts facial features from the face and creates a face embedding vector with 128 elements expressing these features. This embedding vector is mapped to the Euclidean space representing the face similarity. Figure 2(a) shows the structure of the Facenet model, which consists of an input layer, a deep CNN layer, and an L2 regularization to output embeddings. The weights are updated using treplet loss, and embedded feature vectors are input to the classifier to perform face verification and face identification. The basic deep learning network structure uses the 22-layer Inception network, and is trained with 500M face image datasets collected internally.

3. METHODOLGY

In this study, we implement a system that determines whether a mask is worn or not and checks the identity of the person who enters. To determine whether a mask is worn or not, we derive the model to detect masked face. To derive the model, transfer learning was conducted using the s-model, which is the fastest among YOLOv5. To confirm the identity of people wearing mask, we use face recognition. The masked face recognition model was derived by transfer learning of Facenet by inputting masked face images. Figure 3 shows the structure of the proposed method.

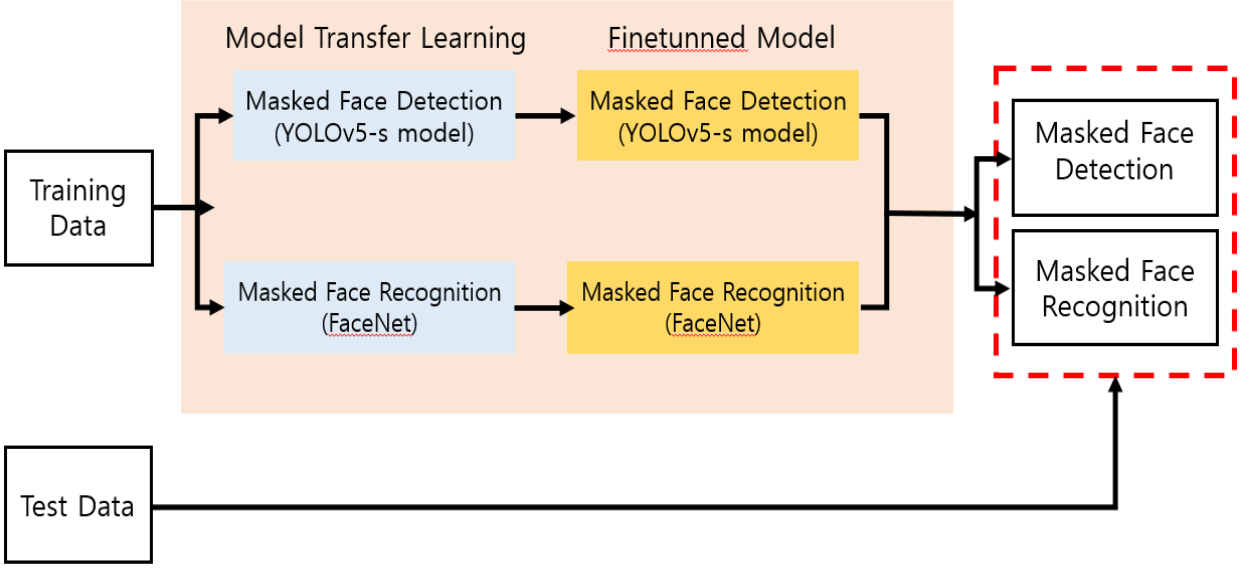


Figure 3. Structure of the proposed method

Experiments were conducted on Ubuntu 16.04.7 LTS in the following environment.

Table 1. Experimental environment

OS	Ubuntu 18.04.5 LTS
GPU	Tesla V100-SXM2
CPU	16 core Intel(R) Xeon(R) Gold 5120 CPU @ 2.20GHz
RAM	177GB
CUDA	10.1
cuDNN	7.6.0
software	python/pytorch
Deep Learnig model	YOLOv5 S model

3.1 Masked face detection experiment and result analysis

As the dataset for model to determine whether a mask is worn or not, we used the Kaggle dataset Face Mask Dataset (YOLO Format)[14] and data collected from the web and labeled with mask/no_mask by a labeling tool[15]. A total of 2352 data were used, with 1679 training data, 398 verification data, and 280 test data.

First, we transfer learning the Yolov5s model with training data. In this study, mAP was obtained and compared according to batch size and epochs at a learning rate of 0.01 to find a model with optimal masked face detection performance. As a result, the model with the highest mAP is selected as the masked face detection model.

Table 2 shows the experimental results of different epochs when the batch size is 32. It shows the best performance with 0.95 mAP at 50 epochs, and shows that the mAP decreases as the epoch increases.

Table 2. mAP according to epochs at batch size 32

epochs	mask	no_mask	total
50	0.959	0.941	0.95
100	0.936	0.938	0.937
150	0.909	0.934	0.921
200	0.907	0.912	0.909

Table 3 shows the results of different batch sizes for 50 epochs and 100 epochs, which showed good performance among the results of Table 2. At 50 epochs, batch sizes 16 and 32 show the best results at 0.95. This is because the model used is already pre-learned because the training data is small, it converges quickly in small epochs.

Table 3. mAP according to batch size at 50/100 epochs

		mAP		
batch-size	epochs	mask	no_mask	total
16	50	0.962	0.939	0.95
	100	0.937	0.932	0.934
32	50	0.959	0.941	0.95
	100	0.936	0.938	0.937
64	50	0.948	0.945	0.946
	100	0.941	0.936	0.938
128	50	0.931	0.941	0.936
	100	0.939	0.936	0.938

Figure 4 shows a masked face detection results. It detects masked face in a single image, multiple images well and even in a small image. And a masked face is detected, but only a mask is not detected.

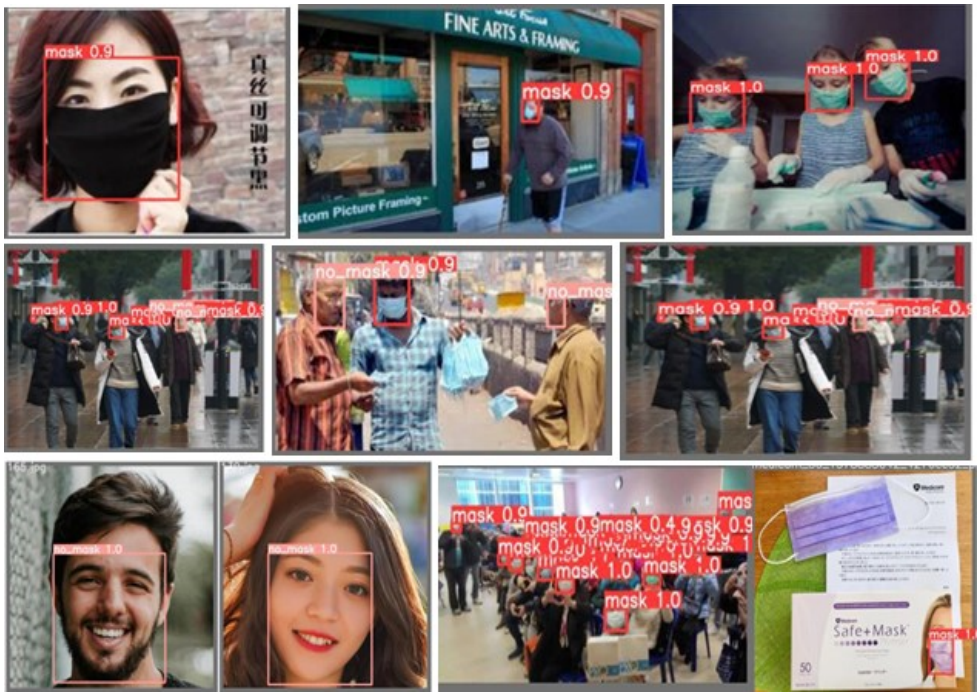


Figure 4. Mask wearing detection result

3.2 Masked face recognition experiment and result analysis

In this paper, transfer learning is performed using the pre-trained Facenet model with the 'vggface2' face dataset. Therefore, to get model of masked face recognition, transfer learning uses masked face dataset. The dataset for training use the dataset provided in [16] and is a ‘Masked LFW’ dataset created by covering the LFW face dataset with a mask. Some of the data were extracted to make the training data, and to measure the performance of the model, data that did not overlap with the training data was selected to make the test data. In addition, in order to evaluate whether the trained model can be recognized even non-masked face, it was tested using the Real faces dataset[11], in which a face with a mask and a face without a mask are mixed. Table 4 shows the training and test datasets. In table 4, ‘class’ entry is the number of people and ‘file’ entry is the number of data.

Table 4. Dataset for face identification

	train	test	
	MASKED LFW	MASKED LFW	Real_faces
class	446	23	75
file	5237	137	434

The experiment was performed at a learning rate of 0.001, and a batch size of 32. 20% of the total training data was used as data for verification.

Table 5. Accuracy by epoch

	100	500	1000	1500	2000	2500
train	0.469	0.659	0.926	0.929	0.932	0.933
vaild	0.413	0.553	0.885	0.889	0.929	0.912

Table 5 show that the results of the 2000 epoch and the 2500 epoch are the best. Therefore, the model with the best performance is selected as the representative model of this study by comparing the performance of the models of the 2000 epochs (called M1 model) and the 2500 epochs(called M2 model).

After making an embedding vector with face data as input, recognition is confirmed by the distance between the two embedding vectors. In this study, Euclidean distance was used.

$$d=e1-e2 \tag{1}$$

$$R(e1,e2)=\begin{cases}\text{same}, & d \le Th\\ \text{diff}, & \text{otherwise}\end{cases} \tag{2}$$

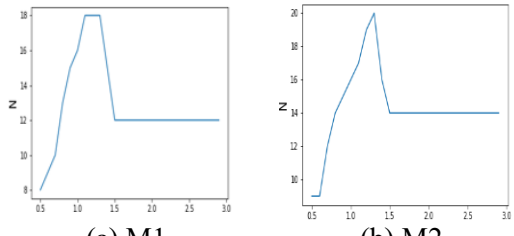
In eq(1), e1 and e2 are the embedding vectors of the input image, and d is the distance between e1 and e2. In eq(2), Th is the distance threshold for determining whether two images are the same or different people, and R(e1, e2) is the result of determining whether the two images are the same.

Figure 5 shows that the identification result is different when the distance threshold is set to 1.3. and 1.4.

(a) d=1.3 (b) d=1.4

Figure 5. Identification results according to distance (M2 model)

Figure 6 shows the number of pairs correctly discriminated according to the distance for each model when the threshold values are different from 0.5 to 20 for 20 pairs of inputs composed of the same person and different people with the 'MASKED LFW' test set.



(a) M1 (b) M2

Figure 6. Identification result according to distance

The performance of each model was compared by making a pair of input images. Table 6 shows the results

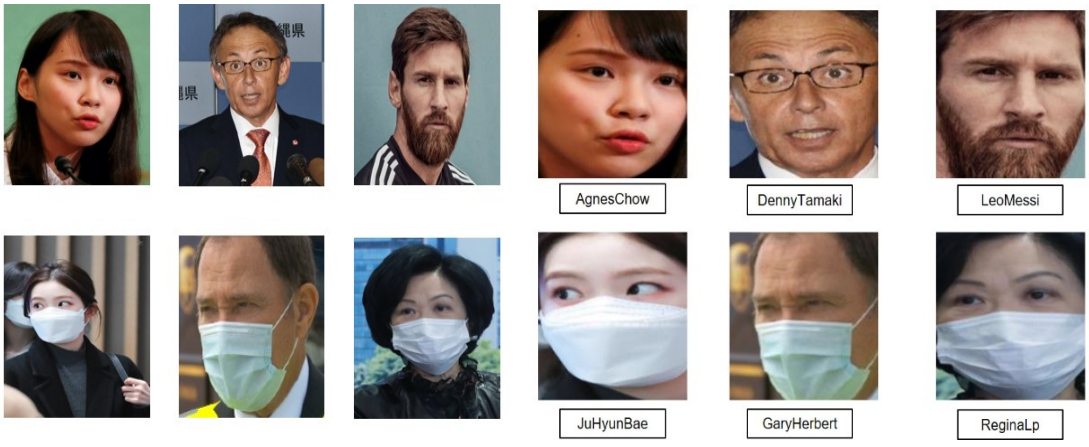
of applying the M1 and M2 models to the test data composed of 'MASKED LFW', forming 500 pairs and 1000 pairs for the same person pair and another person pair.

Table 6. 500/1000 pairs of accuracy

	M1	M2
500	0.751	0.841
1000	0.717	0.846

From the results in Table 6, it can be seen that the performance of the M2 model is better than that of the M1 model. Therefore, this paper selects the M2 model as the optimal model.

Figure 7 is the result of checking the identity using the Real faces test dataset for the selected model. Figure 7 shows that the identity of the input image can be verified well.



(a) input images (b) result images

Figure 7. Result image using Real_faces test dataset

4. CONCLUSION

Masks are inspected in public places and indoors and outdoors, and masks must be worn inside as well. However, since most people identify non-masked people with their own eyes, manpower is required, and sometimes there may be parts that cannot be checked. Therefore, an automatic detection algorithm of whether a mask is worn is required. In addition, a system that confirms the identity of the person entering and leaving is required for security. Most automated systems for identification use face recognition techniques, but if you are wearing a mask with the existing face recognition system, you may have to take off the mask for identification. This needs improvement because it not only makes the authentication process cumbersome for users, but also increases the risk of virus transmission.

To improve these points, we use transfer learning of Yolov5s model to detect masked face and use transfer learning of Facenet to derive an optimal deep learning model that can automatically recognize person of masked face and verify its performance. The dataset for masked face detection were used data collected from the web and the data set provided by Kaggle. Part of the MASKED LFW dataset was used for training of Facenet, and some data that did not overlap with the training data among the MASKED LFW datasets and the Real Faces dataset were used for test. The proposed model detected masked face of people well and performed face recognition of masked face.

These results show that the research in this paper can increase work efficiency by determining whether or not a mask is worn, and that it can be used in terms of security because identification is possible even when wearing a mask.

ACKNOWLEDGEMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2021-0-01972)

REFERENCES

[1] C. I. Paules, H. D. Marston, and A. S. Fauci, "Coronavirus infections more than just the common cold," Jama, Vol. 323, No. 8, pp. 707–708, 2020.

[2] E. Y. Cho, J. G. Kim, "Analysis of Factors Affecting the Knowledge with COVID-19," IPACT, Vol.9, No. 4, pp.219-225, 2021.

[3] K. H. Sung, G. H. Ryu, and D. Y. Yun, "Sasang Constitution Analysis and Wine Recommendation App suggestion through Mobile Face Recognition," IJIBC, Vol.13, No.34, pp.155-162, 2021.

[4] S. G. Chae, "A survey on the use of mobile phones due to COVID-19," IJIBC, Vol.12, No.3, pp.233-243, 2020.

[5] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815–823, 2015.

[6] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface:Deep hypersphere embedding for face recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 212–220, 2017.

[7] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4690–4699, 2019.

[8] ultralytics/yolov5. https://github.com/ultralytics/yolov5

[9] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.,"You Only Look Once: Unified, Real-Time Object Detection," Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.779-788, 2016.

[10] N.J. Kwak, D.J. Kim, "Object detection technology trend and development direction using deep learning," IJACT, Vol.8, No. 4, pp.119-128, 2020.

[11] G. Liu, S. H. Lee, "Municipal waste classification system design based on Faster-RCNN and YoloV4 mixed model," IJACT, Vol.9, No. 3, pp.305-314, 2021.

[12] Chen, D., Ren, S., Wei, Y., Cao, X., & Sun, J., "Joint cascade face detection and alignment", In European conference on computer vision, pp.109-122, 2014.

[13] Kaipeng Zhang, Zhanpeng Zhang and Zhifeng Li, "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks," arXiv:1604.02878 [cs.CV], 11 Apr 2016.

[14] Kaggle Dataset:https://www.kaggle.com/aditya276/face-mask-dataset-yolo-format

[15] Labeling Tools (labelmg). https://github.com/tzutalin/labelImg.

[16] Masked dataset :https://github.com/SamYuen101234/Masked_Face_Recognition.