An AI-based Mask-Wearing Status Recognition and Person Identification System

Minxuan Wen[1] Kentaro Yokoo[1], Xuebin Yue[1], and Lin Meng[1]

Dept.of Electronic and Computer Engineering, Ritsumeikan University,
Kusatsu, Shiga, Japan.
{gr0518eh@ed,ri0086ps@ed,gr0468xp@ed,menglin@fc}.ritsumei.ac.jp

Abstract

In the tough COVID-19 pandemic, wearing a mask in daily life becomes an important habit. However, sometimes people forget to wear a mask or wear a mask incorrectly in careless. Hence, alarming the problem of protecting ourselves from COVID-19 becomes a key challenge. Unfortunately, at home security or office security systems, wearing a mask lets the person identification lost its function. Hence, masked-person identification becomes an essential issue. This paper proposes an AI-based mask-wearing status recognition and person identification system for solving the above problems. The system consists of three stages, face detection based on MTCNN, mask-wearing status recognition, and person identification using MobileNetV2. Masked-person identification is one of the functions of the proposed system. The experimental results show that the face detector reaches almost 100% accuracy among 3000 images. The mask-wearing status recognition has a 96.1% test accuracy in 300 test images, and person identification achieves a 98% recognition rate. In summary, the effectiveness of the proposed system is proved by the high accuracy recognition rate.

1 Introduction

In 2021, the largest pandemic in recent history spread through the world: COVID-19. As of May 15, 2021, there have already been 156 million cases and 3 million deaths around the world [1]. Many people in many countries and regions are unwilling to wear a mask[7]. People who wear a mask incorrectly are under the same risk as people who do not wear the mask. As deep learning developing, utilizing neural networks to recognize the mask-wearing status is a challenge to alarm people to wear masks correctly.

However, person identification becomes difficult when people wear a mask because feature points on the lower half of the face are untouchable[5]. Especially at home security or office security systems [16], person identification may lose its function when people wear a mask. For solving these problems, we propose an AI-based system for recognizing three mask-wearing statuses and identifying masked-person.

The system consists of three stages, face region detection, mask-wearing status recognition, and person identification. As the first stage, face region detection aims to detect and crop the face region in the image accurately. MTCNN (Multi-task Cascaded Convolutional Networks)[17] is an effective method for face region detection, which is employed for detecting face regions in images. The second stage is mask-wearing status recognition which aims to recognize the mask-wearing status. The mask-wearing status consists of three classes, wearing the mask correctly, wearing the mask incorrectly, and not wearing the mask. Each class uses 1500 images from Real-World Masked Face Dataset(RMFD)[15] ,and Adnane Cabani's Incorrectly Masked Face Dataset(IMFD)[2]. For realizing the system in small tissues such as the laboratory, community, or the company, we apply MobileNetV2 that is a slight, high accuracy model [8] for mask

CEUR Workshop Proceedings (CEUR-WS.org)

wearing-status recognition. Person identification is the third stage for identifying the person. For realizing the masked-person identification, the proposal uses the top-half face merely. Hence we only need to crop the top-half face precisely by image processing. In terms of MobileNetV2 selection, the default output of MobileNetV2 is 1000 classes which fits for a larger number of people identification. The contributions of this research are shown as follows:

- In the authors' opinion, a few papers introduce a deep learning method for detecting the face, recognizing the mask-wearing status of the mask, and identifying a person (including the masked person) in the same time. Hence, it is a significant challenge to using the deep learning-based method for person identification.
- A few papers pay attention to incorrect mask-wearing status. Hence, it is meaningful to do this mask-wearing status recognition in this project.
- The proposal contributes to inhibiting the pandemic by applying this system. When people are used to wearing a mask correctly, the transmission capacity of the virus would be under control.

This paper composes of 5 sections. Section 2 is related works that introduce the material. Section 3 describes the system flow entirely and explains the algorithm which we used in the processing. Section 4 explains the experimental condition, dataset, and analysis of the experimental results. Section 5 details the conclusion.

2 Related works

Mask-wearing status recognition is a general project due to the Covid-19 pandemic. The paper "Covid-19 Face Mask Detection Using TensorFlow, Keras, and OpenCV" which is created by Arjya Das from Jadavpur University[3], raises a method that uses Haar-cascade to detect face and trains CNN model to classify mask-wearing status. Chandrikadeb purposes another approach in their GitHub project[4], which similar to our proposed system. They utilize a Caffe-based face detector in conjunction with a fine-tuned MobileNetV2 for mask-wearing status recognition. They can achieve a decent 0.93 F1 score on the classification. Nevertheless, they only set two classes which are the status of mask-wearing and not mask-wearing.

3 System flow

Figure 1 shows the system utilizes three neural networks, including face region detection using MTCNN, mask-wearing status recognition using MobileNetV2, and person identification using another MobileNetV2. MTCNN detects the face in the image and takes the facial image out, firstly. The facial image is transferred to the second stage where MoblieNetV2 is applied for recognizing the mask-wearing status. Then, the facial image is also rotated and cropped for creating the top half face by image processing. After the image processing, the top half face image is sent to the last stage for person identification by another MobileNetV2.

3.1 Face region detector and image processing

Based on our survey, MTCNN is a well pre-trained network for face region detection with high accuracy, becomes one of the most popular face detection tools today. Hence, the face region recognition package of MTCNN is used for building on-top of Tensorflow. MTCNN is a neural

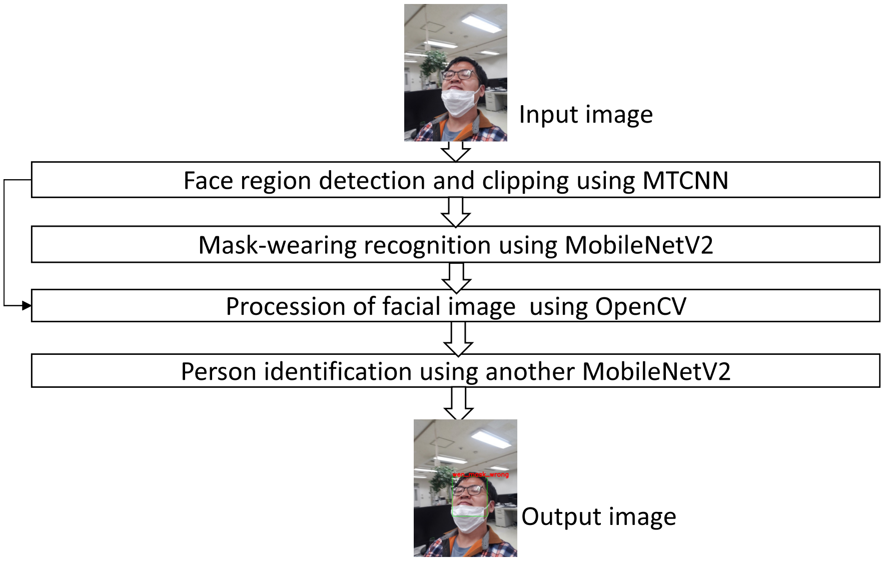

Input image
Face region detection and clipping using MTCNN
Mask-wearing recognition using MobileNetV2
Procession of facial image using OpenCV
Person identification using another MobileNetV2
Output image

Figure 1: System Flow

network for detecting faces and facial landmarks on images, which consists of 3 neural networks connected in a cascade, P-Net, R-Net, and O-Net.

In the first stage, the MTCNN creates multiple frames which scans through the entire image starting from the top left corner and eventually progressing towards the bottom right corner. The information retrieval process is called P-Net(Proposal Net), is a shallow, fully connected CNN. In the second stage, all the information from P-Net is used as an input for the next neural network called R-Net(Refinement Network), a fully connected, complex CNN which rejects a majority of the frames that do not contain faces. In the final stage, a powerful and complex neural network, known as O-Net(Output Network), as the name suggests, outputs the facial landmark position detecting a face from the given image.

A person with a mask is difficult to be recognized, the reasons are listed as follows.

- The facial features, such as the nose and mouth are blocked. The features that can be utilized to identify the face will be reduced apparently.

- The physical distribution of distinguishable information such as face contour also changes. Hence, the accuracy of the person identification model trained according to the traditional methods will decrease.

We propose that using solely the top half face to do person identification to solve the problems. For properly cropping the top half face, an efficient method building on-top of OpenCV is employed. [14] adopted that method to rotate images. However, our system calculates the angle between the eyes and the horizontal line to rotate the facial images. The calculation of angle is based on the key points recorded by MTCNN. By obtaining the coordinates of the eyes, we get the angle between the eyes and horizontal line. Result of one rotation in our images is shown in figure 2.

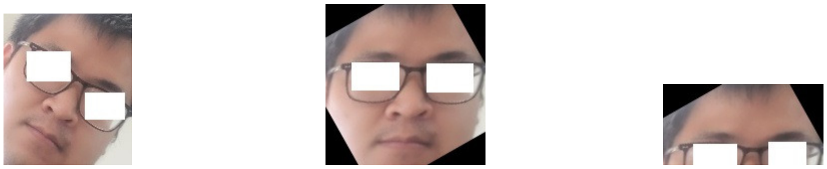


Figure 2: Utilizing key points to rotate and crop the facial image

Networks	macs(million)	parameters (million)
MobileNetV2	314.13	3.50
Resnet18	1819.00	11.69
VGG16	15484.00	138.35
Densenet121	2.86	7.97
GoogLeNet	1505.00	6.62

Table 1: The numbers of floating-point operations and parameters

3.2 Mask-wearing status recognition and person identification

It is important to compare the accuracy and networks size. We test the networks inculding ResNet18[6],VGG16[12], MobileNetV2[8],DenseNet121[9], and GoogLeNet[13]. Due to the MobileNetV2 has less parameteres (see table1) and relatively high inferences speed, we choose MobileNetV2 as the mask-detecter and person identifier.

MobileNet is a CNN architecture model for Mobile Vision. What makes MobileNet special is that it can run and apply transfer learning with less computation power. Therefore, it has a perfect fit for Mobile devices, embedded systems, and computers without GPU or low computational efficiency with relatively high accuracy. MobileNet is based on a streamlined architecture that adopts depth-wise separable convolutions to build light weight deep neural networks. Due to this special architecture, MobileNet significantly reduces the number of parameters.

The mask-wearing dataset and the mask-not-wearing dataset we used originate from the Real-world Masked Face Dataset(RMFD) from Wuhan University[15]. The dataset that people don't wear masks properly comes from Adnane Cabani's Incorrectly Masked Face Dataset(IMFD)[2]. Each dataset has 2,000 images, 60% of the images used as training dataset, 35% of images utilized as validation dataset, and 5% of images applied as test dataset. Because of the great dataset, this mask detector has a steady accuracy.

Similarly, we use MobileNetV2 continuously as our person identification model. The difference is that we train this model using our dataset. In this stage, we create a dataset, each of the classes puts in 500 pre-processed images. Pre-processing has two steps, firstly, take the photos and crop the face in the image. As much as possible to take photos from different distances and different angles. Then, cropping the top half of the face and enhancing the data.

4 Experimentation

The current system including three stages is built in the Intel(R) Core(TM) i7-9700 CPU and Intel(R) UHD Graphics 630. The image size is set to 320 × 320 in the experimentation. This section shows the experimental results of each stage.

Detector	accuracy	FPS
MTCNN	0.99	4.5
Haar-cascade	0.20	46

Table 2: Performance Comparison between MTCNN and Haar-cascade

4.1 Experimental results of face region detection

Because the MTCNN is pre-trained in Tensorflow very well, the training dataset and training processing of MTCNN are omitted. For comparing the performance of MTCNN with current research, the additional experimentation of Harr-cascade is done in this research.

2000 test images are selected from Karras, T's Github project[10] randomly, including front faces and side faces.

The comparing results of Haar-cascade and MTCNN are shown in table 2. MTCNN achieves 99% accuracy with the speed of 4.5 FPS in detecting the face region. Haar-cascade has only 20% accuracy and a speed of 45 FPS. Due to the high accuracy of MTCNN, the MTCNN is applied as the face region detector.

4.2 Experimental results of mask-wearing status recognition

MobileNetV2 is applied for recognizing the mask-wearing status. The training processing is done on the CPU intel Xeon e5 1650-v3 and GPU Geforce GTX TITAN X. In detail training parameter, the batch size, the epochs number, learning rate, and the weight decay are set to 4, 20, 0.0001, and 0.1, respectively. The dataset of RMFD and IMFD are used in mask-wearing status recognition. We divide the dataset into the training dataset, validation dataset, and test dataset. Three statuses are defined as classes, including wearing a mask, not wearing a mask, and wearing a mask incorrectly. 6000 images are employed in the experimentation, each class has 1200 training images, 700 validation images, and 100 testing images.

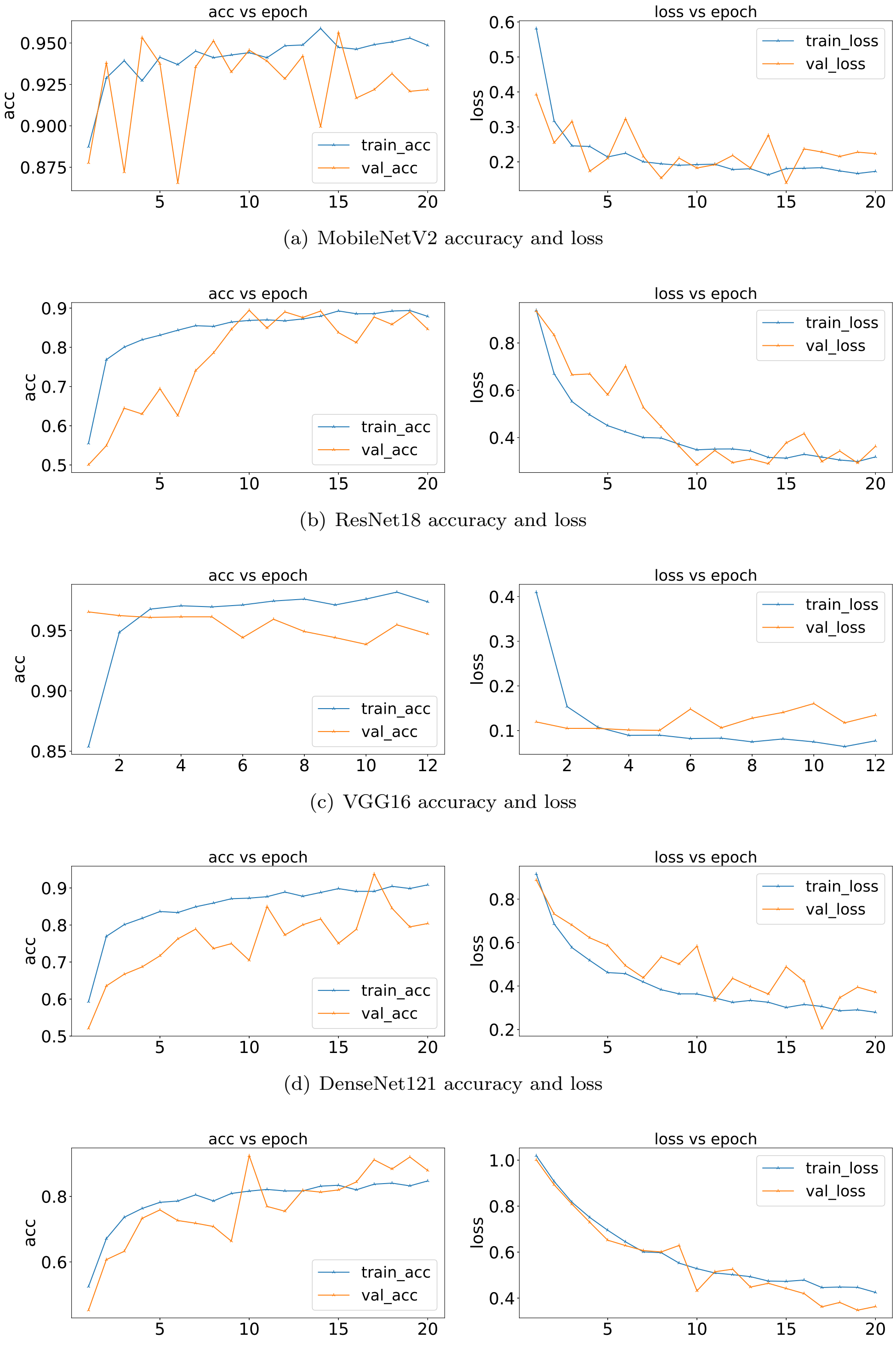
Figure 3 (a) shows the accuracy and loss of mask-wearing status recognition. MobileNetV2 achieves the 0.17 validation loss and the 95.3% validation accuracy.

Four kinds of state-of-the-art deep learning models, VGG16, ResNet18, DenseNet121, and GoogLeNet, are equipped for proving the effectiveness of our proposal. The accuracy and loss of VGG16, ResNet18, DenseNet, and GoogLeNet are shown in figure3 (b) (c) (d) and (e), respectively.

The accuracy of MobileNetV2, ResNet18, VGG16, DenseNet121, GoogLeNet is convergent at the epoch of 11th, 11th, 5th, 12th, and 20th, respectively. ResNet18 achieves the 0.28 validation loss and the 89.3% validation accuracy. VGG16 achieves the 0.10 validation loss and the 96.1%. validation accuracy, which is the best accuracy in these state-of-the-art models. DenseNet121 achieves the 0.28 validation loss and the 89.1% validation accuracy. GoogLeNet achieves the 0.43 validation loss and the 92.4% validation accuracy. The test accuracy of the models is shown in table 3. In conclusion, MobileNetV2 achieved similar accuracy with the four kinds of state-of-the-art models. However, considering the model size, MobileNetV2 is about one percent of VGG16, the MobileNetV2 is selected in the research [11].

4.3 Experimental results of person identification

Another MobileNetV2 is also applied for recognizing the mask-wearing status. The results of test accuracy are shown in table 3. MobileNetV2 achieved similar accuracy with others.



(a) MobileNetV2 accuracy and loss

(b) ResNet18 accuracy and loss

(c) VGG16 accuracy and loss

(d) DenseNet121 accuracy and loss

(e) GoogLeNet accuracy and loss

Figure 3: CNN models comparison on mask-wearing status recognition

Models	Acc. of mask-wearing status recognition	Acc. of person identification
MobileNetV2	0.961	0.989
ResNet18	0.932	0.942
VGG16	0.967	0.981
DenseNet121	0.928	0.951
GoogLeNet	0.915	0.931

Table 3: Acc. of mask-wearing status recognition and person identification

The training environment and hyperparametersis is similar to the second stage. Four people are defined as classes, and 2000 images are employed in the experimentation, each class has 350 training images, 100 validation images, and 50 testing images. Data augmentation is utilized for increasing the images in the dataset. 5 main methods that built-in PyTorch transformations is applied, including ColorJitter, Random Rotation, RandomResizedCrop, GaussianBlur, and RandomErasing.

4.4 System time consumption and discussion

Loading models coss $0.28s$. MTCNN takes $0.22s$ to detect the face regions. Mask-wearing status recognition utilizes $0.19s$. Rotating and cropping the image costs $0.0019s$. Identifying a person from the rotated image costs $0.20s$. In total, the system takes at least $0.69s$, including mask-Wearing status recognition and person identification, which may be used in practical application scenarios.

By using the depth-wise separable convolution, MobileNetV2 only utilizes shallow CNN, but guaranteeing the relatively high accuracy.

In terms of system design, masked-person identification may be easier to design on training the masked-person dataset, without doing mask-wearing status recognition and person identification step by step. It may be effective for time reduction.

However, in this case, the class number increases 3X when adding one person, causing the class number to become larger.

5 Conclusion

As the world pandemic does not change, we want to help to control the disease spreads. Our program helps people of communities get into the habit of wearing masks correctly. We want to build a slight, robust, and efficient system. This mask-wearing status recognition and person identification system have advantages that the high accuracy and applicability. The total accuracy is 97.5%(average of mask-wearing status recognition accuracy and person identification accuracy). In future work, applying this system in the security project would be a meaningful choice. In future works, utilizing YOLO networks to build a new objection detection stage for mask detection would be a constructive reference.

References

[1] google covid19 map. news.google.com/covid19/map.

[2] Adnane Cabani, Karim Hammoudi, Halim Benhabiles, and Mahmoud Melkemi. Maskedface-net–a dataset of correctly/incorrectly masked face images in the context of covid-19. *Smart Health*, 19:100144, 2021.

[3] Arjya Das, Mohammad Wasif Ansari, and Rohini Basak. Covid-19 face mask detection using tensorflow, keras and opencv. In *2020 IEEE 17th India Council International Conference (INDICON)*, pages 1–5. IEEE, 2020.

[4] Chandrika Deb. Face-Mask-Detection. `github.com/chandrikadeb7/Face-Mask-Detection`.

[5] Md Sabbir Ejaz, Md Rabiul Islam, Md Sifatullah, and Ananya Sarker. Implementation of principal component analysis on masked and non-masked face recognition. In *2019 1st international conference on advances in science, engineering and robotics technology (ICASERT)*, pages 1–5. IEEE, 2019.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] Lu He, Changyang He, Tera L Reynolds, Qiushi Bai, Yicong Huang, Chen Li, Kai Zheng, and Yunan Chen. Why do people oppose mask wearing? a comprehensive analysis of us tweets during the covid-19 pandemic. 2021.

[8] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[9] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[10] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *CoRR*, abs/1812.04948, 2018.

[11] Henyi Li, Zhichen Wang, Xuebing Yue, Wenwen Wang, and Hiroyuki Tomiyama. A comprehensive analysis of low-impact computations in deep learning workloads. In *in Proceedings of the Great Lakes Symposium on VLSI 2021 (the 31st GLSVLSI)*. ACM, 2021.

[12] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[13] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[14] Jiaxi Wang and Junzo Watada. Panoramic image mosaic based on surf algorithm using opencv. In *2015 IEEE 9th International Symposium on Intelligent Signal Processing (WISP) Proceedings*, pages 1–6. IEEE, 2015.

[15] Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, et al. Masked face recognition dataset and application. *arXiv preprint arXiv:2003.09093*, 2020.

[16] Kentaro Yokoo, Masahiko Atsumi, Kei Tanaka, Haoqing Wang, and Lin Meng. Deep learning based emotion recognition iot system. In *The 2020 International Conference on Advanced Mechatronic Systems (ICAMechS 2020)*, 2020.

[17] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.