

A Method for Detecting Non-Mask Wearers Based on Regression Analysis

Dokyung Hwang¹, Hyeonmin Ro¹, Naejong Kwak², Jinsang Hwang³ and Dongju Kim^{1,*}

¹POSTECH Institute of Artificial Intelligence, Pohang, 24257, Korea

²Department of Information Security at Pai Chai University, Daejeon, 35345, Korea

³Department of Safety and Emergency Management at Kyungwoon University, Gumi, 39160, Korea

*Corresponding Author: Dongju Kim. Email: kkb0320@postech.ac.kr

Received: 22 November 2021; Accepted: 21 February 2022

Abstract: A novel practical and universal method of mask-wearing detection has been proposed to prevent viral respiratory infections. The proposed method quickly and accurately detects mask and facial regions using well-trained You Only Look Once (YOLO) detector, then applies image coordinates of the detected bounding box (bbox). First, the data that is used to train our model is collected under various circumstances such as light disturbances, distances, time variations, and different climate conditions. It also contains various mask types to detect in general and universal application of the model. To detect mask-wearing status, it is important to detect facial and mask region accurately and we created our own dataset by taking picture of images. Furthermore, the Convolutional Neural Network (CNN) model is trained with both our own dataset and open dataset to detect under heavy foot-traffic (Indoors). To make the model robust and reliable in various environment and situations, we collected various sample data in different distances. And through the experiment, we found out that there is a particular gradient according to the mask-wearing status. The proposed method searches the point where the distance between the gradient for each state and the coordinate information of the detected object is the minimum. Then it carry out the classification of mask-wearing status of detected object. Lastly, we defined and classified three different mask-wearing states according to the mask's position (With mask, Wear a mask around chin and Without mask). The gradient according to the mask-wearing status, is analyzed through linear regression. The regression interpretation is based on coordinate information of mask-wearing status and the sample data collected in simulated environment that considering distances between objects and the camera in the World Coordinate System. Through the experiments, we found out that linear regression analysis is more suitable than logistic regression analysis for classification of people wearing masks in general-purpose environments. And the proposed method, through linear regression analysis, classifies in a very concise way than the others.

Keywords: Automatic quarantine process; detection of improper mask wearers; facial image coordinates; convolution neural network



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

The Coronavirus disease 2019 (COVID-19) pandemic has changed our lives very quickly and extensively. As a matter of fact, we have been constantly confronted with respiratory viruses such as Middle East Respiratory Syndrome (MERS), Influenza A virus subtype, Hemagglutinin1 Neuraminidase1 (H1N1), Severe Acute Respiratory Syndrome (SARS), etcetera. And there are views of experts that humanity will become extinct due to virus based on these experiences [1–9]. At the national level, various disinfection policies are being announced to prevent these respiratory infections, and the most representative way among them is wearing a mask. Mask-wearing guidance is the simplest and the most effective way to prevent respiratory viral invasions [10–13]. However, as can be seen from the pandemic caused by COVID-19, many people do not follow the quarantine guidelines due to the prolonged respiratory infection virus. Operating on-site monitoring personnel to comply with the guidelines causes several cost problems [9,14]. To reduce costs and prevent viral infections, a lot of Artificial Intelligence (AI)-based public health guidance violator identification technology and kiosk-type access management systems are being developed [15–19]. Though the developed techniques and systems are used, however, there are limitations that the current systems cannot be used in universal situations and are difficult to measure numerous people at the same time [15–24]. Since the current kiosk-type measurement system is installed at the indoor entrance to measure body temperature or identify the mask-wearing status one at a time, it cannot measure multiple people at once and used in places where foot traffics are heavy. Regarding mask-wearing status identification technology using CNN, it is challenging to get versatility because it cannot detect people who are wearing mask incorrectly such as exposing their nose or mouth, and wearing mask under their chin. There is another limitation that the recognition rates get lowered in indoor lighting disturbances and when it comes to small object detection [15–24].

This paper proposes a new method for detecting public health guidance violators to solve the existing limitations and contribute to an effective and reliable automatic prevention system. The proposed method adopted YOLO model which is a 1-stage CNN detector for fast and accurate detection of objects, and the model was trained with open dataset and our own image dataset that is collected for robust recognition in external environments such as lightings and distances. Regarding our own created image data, it is collected from various places such as crowded lecture halls, indoor building entrances with a lot of foot traffic, and auditoriums. To derive accurate recognition performances from various external parameters of the learning model, it is separately collected by external parameters, such as illumination controls, weather conditions of sunny and cloudy days, distance controls for small object detection, and day/night environments. It uses YOLOv4 detector for fast object detection, which detects face and mask objects, respectively. Then, using the image coordination information of the detected face and mask, we constructed sample data for each distances so that the model can distinguish between the threshold value for correct mask-wearing status and the threshold value of incorrect mask-wearing status. The composition of the sample data is divided into three states: wearing a mask, around the chin, and not wearing a mask. Specific non-linear gradient distributions are shown by the sample data for each state, and as a result, we finally performed the classification of the mask-wearing state according to the non-linear gradients and the distributions of the data bounding box coordinate information. The classified object presents the Region of Interest (ROI) using detected facial box coordinate information, and it is divided into two states: “With Mask (denoted as WM)” and “Without Mask (WO)”. The case of “Wear a mask Around Chin (AC),” is also classified as WM. In additions, the proposed coordinate information comparison method pre-defines the linear regression slope information of the sample data constructed in the simulation environment considering the three mask-wearing states of the subject and the distance between the cameras, so that

it provides fast mask-wearing status result without frame delays even in a real time environment. Here is a summary of the significant contributions of the proposed method:

- It provides a guide to the system of automatic guideline that can effectively prevent viral respiratory infections.
- It proposes a method that effectively detect wearers and status (e.g., wear a mask around one's chin) using the image coordinate information of the object obtained from a CNN architecture.
- Image data sets improves performances of the trained model and are collected for robust object detection from various external disturbances.
- A study of the linear gradient according to the mask-wearing status of people using the image coordinate information of the detected objects and fast and accurate detection of multiple mask wearer using 1-stage CNN detector.

2 Related Work

This chapter discusses the prior models used to detect mask-wearing status.

2.1 Object Detection Model

The newly proposed model focuses on recognizing multiple objects simultaneously to suggest automatic prevention guidelines in densely foot traffic areas. Therefore, it is designed to detect multiple things at once by adopting an object detection model rather than a classification. Though high object detection accuracy is required to recognize the correct mask-wearing status, the accuracy can be complemented in recurrent neural network. Therefore, the object detection model aimed to detect pedestrian quickly by using 1-Stage series model. Among 1-Stage detectors, YOLO and Single-Shot Refinement Neural Network for Object Detection (SSD) are the most famous types, and YOLO has improved a lot in accuracy and speed compared to SSD from YOLOv2 [25–31]. In addition, YOLOv4 effectively improved in accuracy by introducing Distance Intersection Over Union (DIOU) and Complete Intersection Over Union (CIIOU) as new loss functions [25]. However, when detecting an object by using 1-Stage detector only, it is challenging get robust and reliable performances in object detection [19–24,32–34]. For example, assuming that 1-Stage detectors are used alone to detect objects who incorrectly wear a mask, object classification information may vary depending on the subject's rotation, distance, ambient lighting, and surrounding background. Furthermore, an accurate detection of the facial region are essential and required to identify mask-wearing status with using of tremendous amount of collected image datasets and elaborate hyper parameter settings. Even if it is assumed that a good model trained with sophisticatedly adjusted hyper parameter and a large amount of image data, it cannot guarantee the reliability of object classification using a detector alone. The purpose of detector is on object detection and it is specialized for region detection of the target object, thus sophisticated state classification such as AC state requires additional image processing works.

Therefore, the proposed method uses YOLOv4 to detect the subject's facial area and mask area. The trained weight file contains a diverse image dataset considering lighting, subject distances, rotations, and surrounding environments. It is also contains classification function using the coordinates of the detected bbox and configured to ensure reliability for sophisticated detection.

2.2 A Study on Mask-Wearing Detection Technologies

To build an automatic prevention system against viral respiratory infection, it is important to find a person's facial area at first. The most common method for the detection is to starts from finding an object using a CNN-structured detector [15,32–34]. Facial and mask area detection use

image datasets such as Detecting Masked Faces in the Wild (MAFA), MaskedFace-Net, and Properly Wearing Masked Detect Dataset (PWMDD). This paper used these open datasets and an own custom image dataset to train the YOLO detector [35–42]. To build an automatic remedial system against respiratory viruses, it is first to find a person's facial and mask region. The most common method for it starts from seeing an object using a CNN-structured detector [15,32–34]. The next task is to find the facial and mask's region and determine the mask-wearing state based on the detected part. Technologies released before COVID-19, however, show poor versatility, accuracy, and processing speed [43–45]. And the technologies released after the COVID-19 pandemic was mostly made into a WM and WO binary classification or focused on effective detection of facial and mask region [20–25,46]. Bucju et al. attempted to develop automatic mask detection using RGB color information of the input image, and Mohamed et al. combined YOLOv2 and ResNet50 to detect the facial region and mask region, but neither technique mentions the AC status [22,47]. Jiang et al. adopted ResNet and MobileNet as backbones and used Feature Pyramid Network (FPN) structure, but it does not consider AC state detection, and the accuracy is 89.6% [48–50]. Mohamed et al. introduced transfer learning to recognize non-masked users with a combination of Support Vector Machine (SVM) and ResNet50, but the proposed technique requires a lot of computations. Therefore, it requires high performance Graphic Process Unit (GPU), which incurs a lot of costs [19]. Additionally, Preeti et al. proposed a real-time mask free recognition system using a combination of SSD and MobileNetV2; still, there is a drawback in that the detection accuracy is lower than that of using the YOLO detector [23]. Hammoudi et al. effectively designed an application for WM, WO, and AC state recognition by introducing an internal algorithm similar to Haar feature point detection in obtained bbox using the CNN detector [50]. Xinbei et al. proposed the Squeeze and Excitation-YOLOv3 model that improved the backbone structure of YOLOv3 and constructed a self-constructed image data set for effective non-masked user detection, but Darknet53 is basically optimized by Cross-Stage-Partial-Connections (CSPDarknet) [25,42]. Jimin and Wei improved the speed and accuracy of object recognition flow by selecting the YOLOv4 detector for effective detection of non-masked user and introducing the CSPn_X Net concept that modified the backbone and neck structure of the model. It, further, enhanced the practicality by focusing on the awareness of WM, WO, and AC [32–34].

3 Proposed Method

The newly proposed method in this paper can recognize the WM, WO, and AC states the same way as the most recently introduced method for detecting the unmasked person. This can be used in various environments and circumstances, and we built our own image dataset. The created dataset is constructed to consider various mask types, lighting disturbances, small objects, multiple time periods (day to night), weather, races, and many faces simultaneously. In the paper, we will denote it as AMD: Advanced Mask Dataset. Existing technologies for detecting unmasked people have mainly focused on datasets or neural net models. The proposed method concentrates on the dataset as well but does not modify the neural net model itself. We first collected image dataset by assuming two separate situations which a person wears a mask correctly and which a person wears a mask incorrectly. And using bbox coordinates of the face and mask region obtained by YOLO detector, we regressively analyzed the properties of the data according to the state of people wearing the mask. And we finally classify the mask-wearing state by using the regression line of each state and the distribution of the data we want to measure. In other words, the proposed technology does not depend on the detector due to the regression consideration of the data for each state in which people wear masks, and this method can be used together with other technologies to improve detection accuracy. The overall flowchart of the proposed method is shown in Fig. 1.

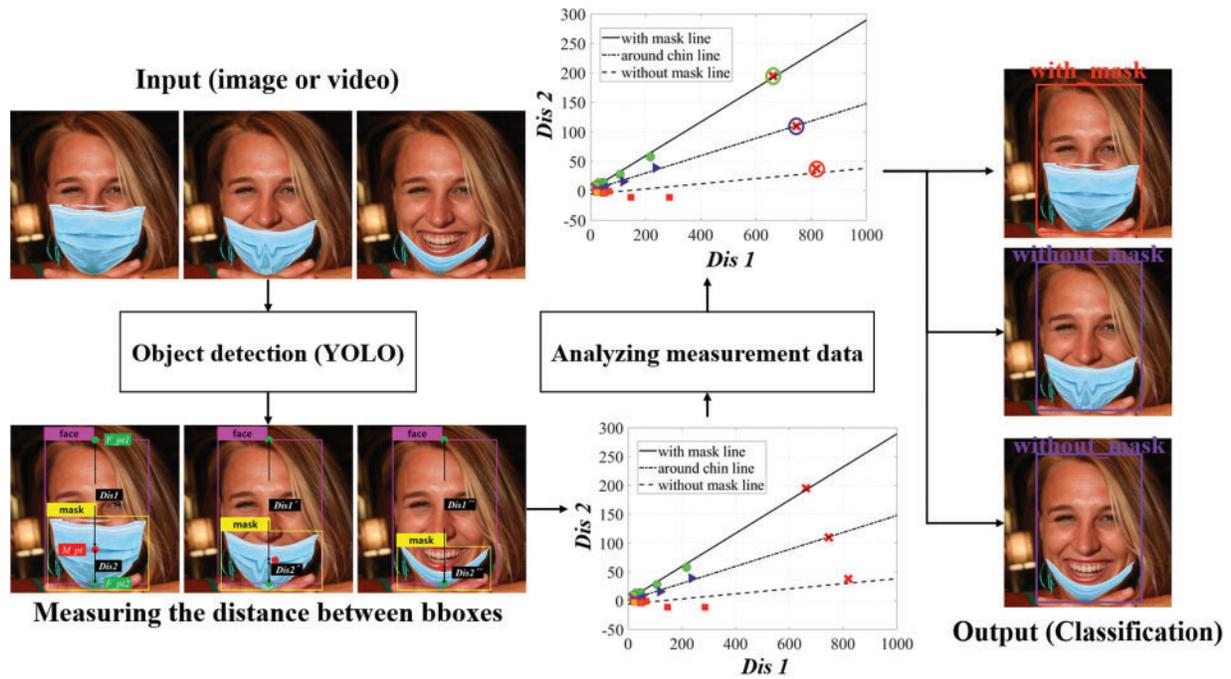


Figure 1: Overview of the proposed non-mask wearer detection algorithm architecture

As it is mentioned in the previous chapter, accurate finding of facial and mask regions needs to precede establishing a non-mask wearer detection system. Facial image datasets collected in various environments and mask-wearing images such as MaskFaceNet are opened to establish a prevention system after outbreak of COVID-19 pandemic [40,41]. There are open image sets that changes the small-large space, rotates the image, contains occlusion facial image, creates intentional blurring, and reflect weather effects [35–39,51]. Image datasets for non-mask wearer recognition that emerge after COVID-19 are divided into two types; Dataset that classifies WO and WM [37], and datasets that classify AC situations as well [40–42]. There is a mask dataset that also includes occluded objects or blurred images in it [37].

The proposed method additionally created an image dataset to derive the accurate result of facial and mask areas against any surrounding disturbance. Detectors adopting supervised learning methods such as YOLO have severe model performance differences depending on the quantity and quality of images used in training procedure. By providing images that are familiar with disturbances in training procedure, the detector makes better object recognition performances under surrounding disorders [52]. Therefore, the previous released dataset also contains images with scale changes, blurring, and various climate environments. The created image dataset basically considered the rotation of the face, scale changes, and various climatic conditions, as existing released facial image sets. Moreover, our own dataset contains images with multiple people for detection in heavy foot-traffic area, images with artificial lighting changes to derive strong recognition results under lighting disturbance, and 1760 of masks images with 16 different types of patterns. We denoted our own created dataset as Advanced Mask Dataset (AMD), and Fig. 2 shows the significant considerations of AMD.



Figure 2: Essential considerations of creating AMD datasets; we purchased and used 1760 masks of 16 types for AMD, which considered more realistic parts than the other existing dataset released to recognize the face and mask

Unlike other open datasets, AMD focused much on these five factors as shown in Fig. 2. To recognize multiple objects at the heavy foot-traffic area, multiple people's image that are taken at the entrance of a building and open dataset images are used together. The subjects (facial images) are taken by rotating up, down, left and right, 0–30 degrees each. The scale changes of the object for small object recognition is divided into two types: intentional (manual) situation and fully automatic deployment. The intentional conditions are obtained by adjusting the space between the camera and the subject as 10 cm interval each. Fully automatic deployment situation is constructed by automatically configured image data when a person is detected to a camera using a trained weight file that can recognize people. To perform a robust detection from light disturbance, we collected images that adjust the brightness of illuminations into 200–400 Lux using artificial lightning and images that is taken under natural light by dividing the time from day to night. For lighting brightness, we referred to the interior illumination standard of International Commission on Illumination (CIE) [51]. We also constructed the types of masks as white one, which is commonly used, black masks, medical masks, various pattern masks, and industrial masks. To consider the effect on image of climate condition, we adopted Unconstrained Face Detection Dataset (UFDD)'s open dataset. And for the robust detection of the facial area, we

used Blur and occlude image of Wider Face and Wildest Face. Moreover, there are several other open datasets for detecting facial and mask regions that are not used in this paper. The proposed method detects the facial and mask areas using open datasets and AMD to robustly detect the facial and mask regions. The proposed AMD is more realistic than the other open datasets considering disorders since it is taken using a camera. [Tab. 1](#) is a comparison and summary of datasets to detect facial and mask regions.

Table 1: Comparison of detection datasets

Datasets	Images	Source	Purpose	Considering the AC	Considering the disturbances
Face detection dataset and benchmark (FDDB) [35]	2,845	Yahoo	Face		
Wider face [36]	32,203	Google, Bing	Face		✓
Wildest faces [53]	67,889	YouTube	Face		✓
UFDD [38]	6,424	World Wide Web (WWW)	Face		
Multi-attribute labelled faces (MALF) [39]	5,250	Flickr, Baidu	Face		
MAFA [37]	35,806	www	Face & mask	✓	✓
MaskedFace-Net [40]	133,783	Flickr-Faces-HQ (FFHQ) [45]	Face & mask	✓	
PWMDD [41]	16,702	www	Face & mask	✓	
AMD (Ours)	22,702	camera	Face & mask	✓	✓

4 Experiment

The proposed method in this paper classifies the detected object whether wearing a mask or not using YOLO algorithm. Coordinate information of the bounding box obtained from YOLO is used to classify the object's state, and [Fig. 3](#) shows the coordinate information using the proposed method. Those coordinate information utilize the next three coordinates: the center coordinates (Facial Point1:F_pt1 and Facial Point2:F_pt2) of each of the top and bottom of the facial region and the center coordinate (Mask Point:M_pt) of the detected mask bounding box. Furthermore, we define the distance from F_pt1 to M_pt as Distance1, and the distance from M_pt to F_pt2 as Distance2.

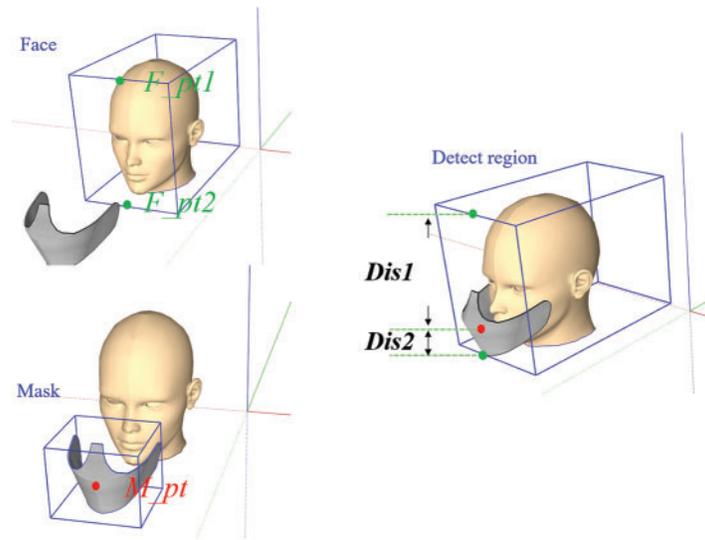


Figure 3: Coordinate information of the bounding box using the proposed method: The proposed method uses the distance values between each coordinate using two coordinates: the top and bottom ends of the facial bounding box and the central coordinates of the mask bounding box. Besides, the two distance values are used to classify the object state whether wearing a mask or not

4.1 A Study on Data by Mask-Wearing Conditions

To detect facial and mask areas using CNN detectors and recognize a state of correctly wearing masks, we inspected the coordinate data for the three conditions defined in this paper (WM: With Mask, WO: Without Mask, and AC: Around Chin). The simulation environment was set up as Fig. 4. The people were positioned from 1 to 6.5 M in the indoor building entrance at 0.5 M intervals. And, distance data (i.e., Distance1 and Distance2) were collected for WM, WO, and AC conditions as shown in Figs. 5 and 6 shows the results of distance data collected for all cases.

In Fig. 5, whole image data were acquired while gazing at the front for fundamental consideration of the data. Here, the measurement distance is limited to 6.5 M, and the proposed method uses the bounding box coordinates obtained from the CNN detector. Its size is inversely proportional to the distance of the subject. In other words, the bounding box becomes smaller as the distance increases. If the distance is closer to 6.5 M, the bounding box size is tiny, so we did not consider an additional space greater than 6.5 m in the simulation situation. And looking at the measured results in Fig. 6, the distance data are distributed with a constant slope depending on the condition of wearing the mask. When the mask is worn correctly, the length of Distance1 decreases, and the length of Distance2 increases; conversely, if the cover is worn incorrectly or on the around chin, Distance1 increases, and Distance2 decreases. Based on these results, we suggest that the classification of mask-wearing status can be interpreted regressively based on the Distance1 and Distance2 information specified in this paper. Therefore, in this paper, the condition of wearing a mask is finally classified through regression analysis of the measurement target.

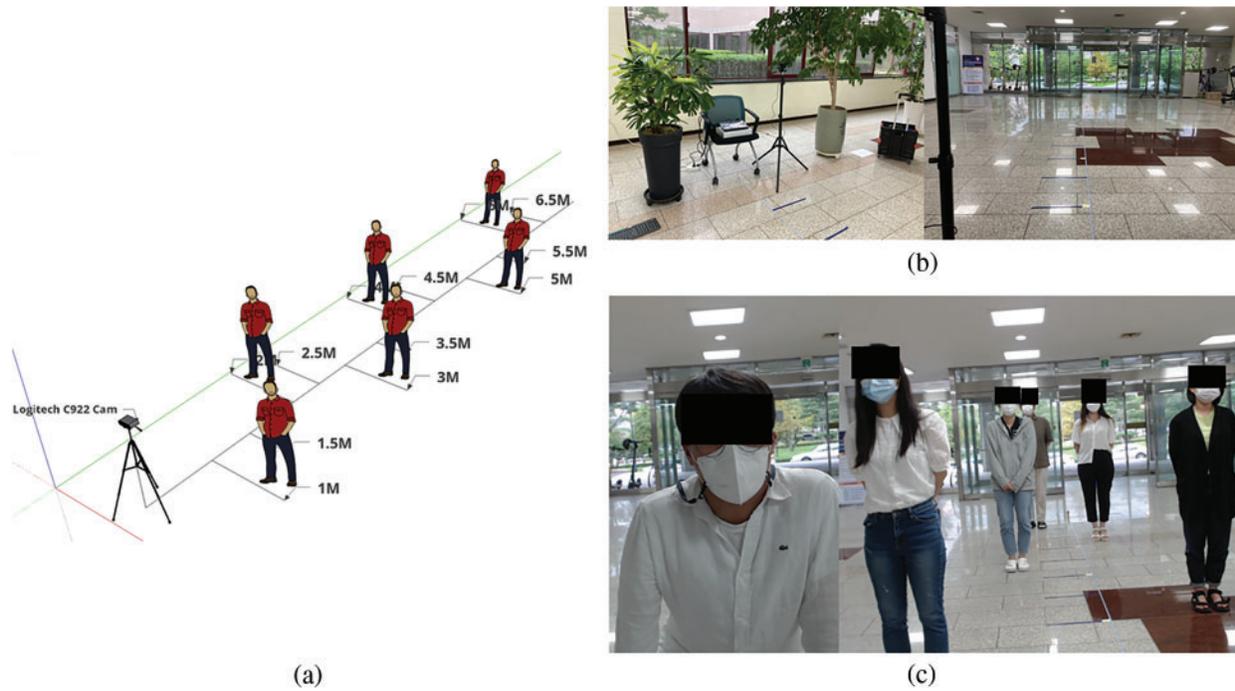


Figure 4: Simulation environment for examining data according to the mask-wearing state: (a) Schematic of the simulation environment; (b) Simulation place; and (c) People setup in simulation place (people were positioned from 1 to 6.5 M at 0.5 M intervals)



Figure 5: (Continued)



Figure 5: Specific situations for collecting WM, WO, and AC data for each distance: (a) state: WM, distance: 1–6 M; (b) state: AC, distance: 1–6 M; (c) state: WO, distance: 1–6 M; (d) state: WM, distance: 1.5–6.5 M; (e) state: AC, distance: 1.5–6.5 M; (f) state: WO, distance: 1.5–6.5 M

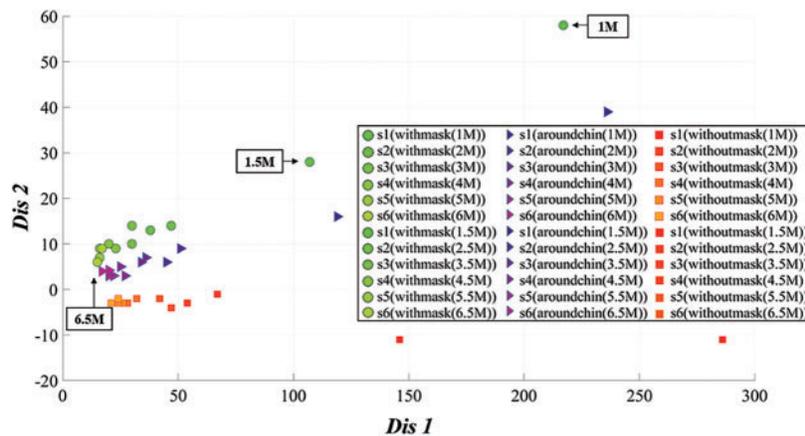


Figure 6: Measurement results of Distance1 and Distance2 for each state

4.2 Linear Regression Model According to the Mask-Wearing Status

To classify the mask-wearing status, we performed linear regression with the measured data in Fig. 6, and Eq. (1) represents a linear regression method for the measured data.

$$Y = wX + b \tag{1}$$

where, Y is a vector corresponding to the response variable, i.e., Distance2 data, and X represents an independent variable corresponding to Distance1. And b is a variable corresponding to the y-intercept, and w represents the gradient of the regression line. Since the linear regression model is a problem finding the gradient of a straight line that minimizes the cost of squared error between the returned observed value and the sampled true value, a cost function that minimizes the error can be defined as follows:

$$Cost(w, b) = \min \sum_{i=1}^n (wx_i + b - y_i)^2 \tag{2}$$

The purpose of the regression analysis results in obtaining w and b values in which $Cost(w, b)$ of Eq. (2) has the closest value to 0. Therefore, the independent variable X and the response variable Y

are treated as constants, and Eq. (2) is solved and summarized to partial derivative w and b as follows:

$$Cost(w, b) = \frac{1}{n} \sum_{i=1}^n (w^2 x_i^2 + 2x_i b w - 2b y_i - 2x_i y_i w + b^2 + y_i^2) \tag{3}$$

Eq. (3) is expressed as follows by partial derivative for w and b .

$$w = \frac{\partial Cost(w, b)}{\partial w} = \frac{1}{n} \sum_{i=1}^n (2x_i^2 w + 2x_i b - 2x_i y_i)$$

$$w = \frac{\partial Cost(w, b)}{\partial w} = \frac{1}{n} \sum_i (x_i w + b - y_i) 2x_i \tag{4}$$

$$b = \frac{\partial Cost(w, b)}{\partial b} = \frac{2}{n} \sum_i (x_i w + b - y_i) \tag{5}$$

The goal is to obtain w and b , which causes $Cost(w, b)$ to converge to zero; so Eqs. (4) and (5) are summarized as follows:

$$w = \sum_{i=1}^n x_i y_i = \sum_{i=1}^n x_i^2 w + \sum_{i=1}^n b x_i \tag{6}$$

$$b = \sum_{i=1}^n y_i = \sum_{i=1}^n x_i w + \sum_{i=1}^n b \tag{7}$$

To apply the simultaneous equation, Eqs. (6) and (7) are expressed in a determinant as follows.

$$\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix} \tag{8}$$

Eq. (8) is expressed as follows according to Cramer’s rule.

$$w = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \sum x_i \sum x_i} \tag{9}$$

where, if we define the expected value for X as follows:

$$E(X) = \frac{1}{m} \sum X \tag{10}$$

Then, Eq. (9) is expressed in the form of expected values for X and Y .

$$w = \frac{E(XY) - E(X)E(Y)}{E(X^2) - E(X)E(X)} \tag{11}$$

Therefore, w can be expressed as follows in the form of variance over X and covariances over X and Y .

$$w = \frac{Cov(X, Y)}{Var(X)} \tag{12}$$

Furthermore, since we should obtain the intercept b converging to 0, Eq. (5) can be summarized.

$$b = \frac{1}{n} \sum_{i=1}^n b = \frac{1}{n} \sum_{i=1}^n y_i - \sum_{i=1}^n x_i w = \frac{\sum y_i - \sum x_i w}{n} \quad (13)$$

Eq. (13) is expressed in the form of an expected value as follows:

$$b = E(Y) - E(X)w \quad (14)$$

In other words, since the purpose is to find the measured linear regression for Distance1 and Distance2, it can be interpreted as finding the Global Optimum for w and b as in Eqs. (11) and (14). And Fig. 7 shows the linear regression results for Fig. 6.

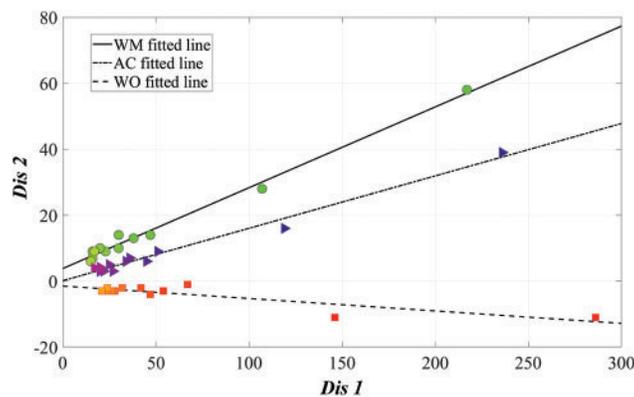


Figure 7: Linear regression results for Fig. 6

4.3 Classification of Status Using Regression Gradient

The proposed method performs mask-wearing status classification using the gradient and intercept of the regressed line. To conduct further experiments on the classification operation using the linear equation of the regression line, a single image containing the three states WM, AC, and WO was used, as shown in Fig. 8. In the case of single image, Distance1 and Distance2 for each state, and the data represent in Fig. 8 with the same way as Fig. 6.

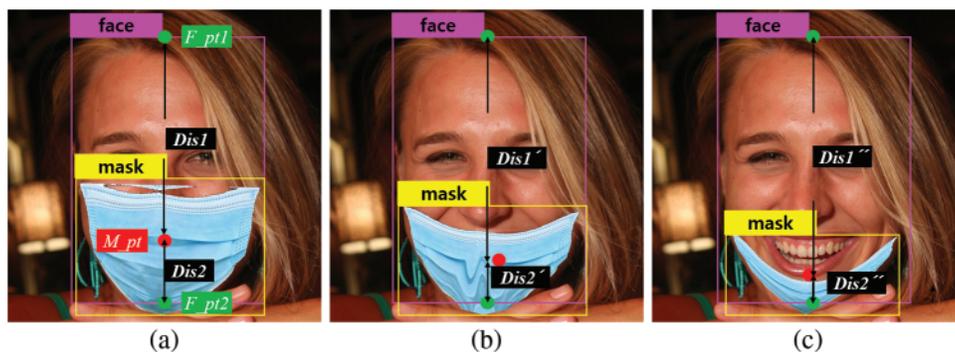


Figure 8: Data acquisition example for a single image: (a) WM: With Mask; (b) AC: Around chin and (c) WO: Without Mask

And, Fig. 9 shows the results of Distance1 and Distance2 measured in the state of Fig. 8.

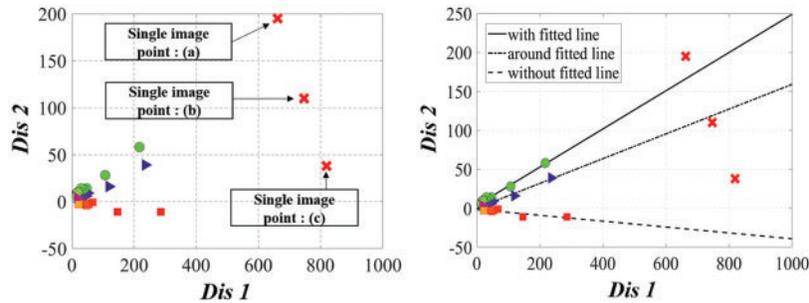


Figure 9: Measurement results of Distance1 and Distance2 for a single image

When the regression lines for each states are organized in the form of $ax + by + c = 0$, the proposed method uses a and b corresponding to the coefficients of x and y and c corresponding to the intercepts. As shown in Fig. 10, it is the way that finding the one has closest distance from each regression lines, and the calculation of the distance from the regression line can be defined as follows:

$$d_s = \frac{|a_s x_i + b_s y_i + c_s|}{\sqrt{a_s^2 + b_s^2}} \tag{15}$$

s is a variable representing three states, and i is a target data for measurement in here. It is to calculate distances between each values (d_{wm} , d_{AC} , and d_{wo}) to $data_i(x_i, y_i)$, and find out d_s with the smallest value in those calculated distances. The final result of mask-wearing status classification can be confirmed as shown in Fig. 11, using proposed method of Fig. 10. In the results of Fig. 11 shows that the results are correctly classified, comparing the final mask-wearing status classification result with the status of Fig. 8, which is the Ground Truth (GT). The main purpose of the proposed method which is classification method of mask-wearing status is to contribute an efficient and automatic monitoring system for public health guidelines that preventing viral respiratory infections such as COVID-19. Therefore, both AC and WO states are classified as without mask. If d_{wm} and d_{AC} have the same distance value for the classification target data $data_i(x_i, y_i)$, then it is classified as d_{AC} by conservative decision.

5 Validation

5.1 Validation of Regression Model

When the new measurement data in Fig. 8 are added, the gradient of the regression line that are already fitted to existing data is modified by Eqs. (11) and (14). Fig. 12 shows the revised regression results considering the measurement data in Fig. 8.

The gradient result of the regression line, reflecting the data additionally measured in Fig. 12 shows that all obtained data in the simulation environment in Fig. 5 could be commonly classified using the proposed method. For more precise and reliable gradient acquisitions, we additionally measured the gradient of 120 situations of rotations, distance adjustments, and additional single images to classify mask-wearing status.

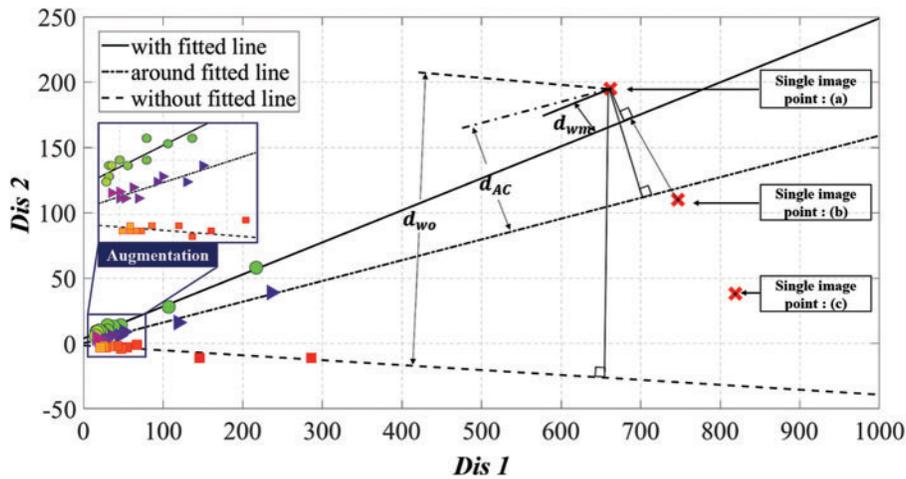


Figure 10: An example of the classifying mask-wearing status process for newly measured data: The proposed method classifies the final mask-wearing status by figuring out d_i that has the smallest distance value for each gradient straight line corresponding to WM, AC, and WO in the classification target data (x_i, y_i)

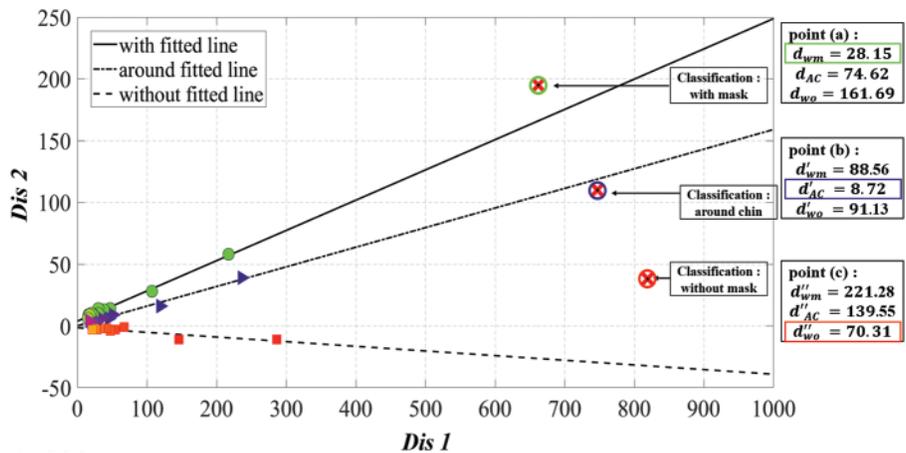


Figure 11: Results of final classification of mask-wearing status in the proposed method for Fig. 10

It can be seen that the non-updated linear slope in Fig. 13a can commonly classify not only additionally measured situations but also data taken from long distances (see the augmentation spot). On the other hand, the updated gradient in Fig. 13b shows that WM and AC gradients intersect each other or that data in the WO state close to WM or AC gradients in case of data taken from long distances (see the magenta line). In other words, Fig. 13b shows that the updated gradients that considering all data cannot classify correctly in the proposed method. In addition, the logistic regression results rather than that of linear in Fig. 13b have lower Root Mean Square Error (RMSE) values for all data, which means having higher fitting compatibility but still locating close to WM data for long-distance objects. Therefore, it can confirm that object classification cannot be performed normally in this case (see the cyan line). In a real-time usage of the proposed method, it is impossible to update the regression line gradient by every time the additional data measured. Therefore, this paper

pre-defines the linear regression result according to each mask-wearing status and use the regression result that is updated in Fig. 12 as a reference line. And for each states, the $ax + by + c = 0$ of pre-defined values presented in this paper are shown in Tab. 2.

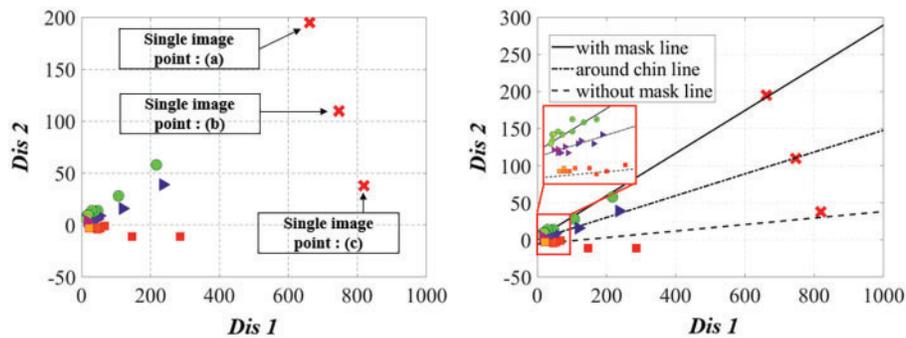


Figure 12: Regression results revised in consideration of the measurement data in Fig. 8

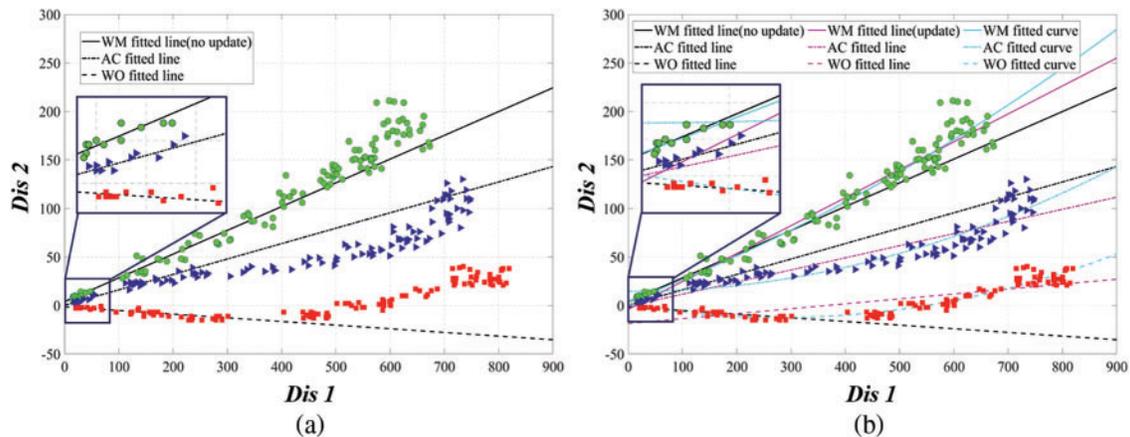


Figure 13: Additional experimental results to obtain detection baseline in more precise situations: (a) Additional 120 situations and non-updated linear gradients in Fig. 12; (b) Results including updated linear and non-linear gradients of additional 120 data

Table 2: Definitions of the coefficients of the pre-defined $ax + by + c$ equations for classifying mask-wearing conditions

Status	a	b	c
WM	0.245	1	3.822
AC	0.1589	1	0.1342
WO	-0.03769	1	-1.496

5.2 Validation of Status Classification Results

The object detection is performed using open datasets considering various disturbances and 22,702 AMD datasets that are directly photographed with cameras. For the mask-wearing status defined in three situations (WM, AC, and WO), the mask-wearing group is classified by comparing the measurement data with status gradients which is previously described. After classify the mask-wearing status, the proposed method uses bounding box of the facial area that are detected by YOLO detector for final result decision. Since both AC and WO states are classified as without masks, the result classes of final object detection are divided into two types: with mask and without mask. Fig. 14 shows basic YOLO only result and the results of mask-wearing status classification using the proposed method, respectively. The detected situation was constructed, focusing on the reliable object detections and classifications under various lighting environments with a large populated area (Illuminations & darkroom, Multiple objects).



Figure 14: Detection result of wearing a mask: (a) only using the YOLO detector; (b) YOLO with Ours

As can be seen from the results of comparative experiment in Fig. 14, the results with YOLO detector alone classifies mask-wearing condition incorrectly and detects the bounding box with overlaps. The proposed method, however, has higher accuracy in detections and classifications by analyzing the coordinate information after the YOLO detection additionally. The detector even has high detection performances since the model is trained with various datasets to detect strongly in dazzling or dark environments. The weight file can be used universally by considering various scale changes and rotations of objects, blur, and multiple objects during training phase. The performance of the supervised learning model varies greatly depending on the quantity and quality of the dataset configuration of training [52]. Since the AMD dataset assumed various disturbances in indoor environments that the proposed method will be applied, the dataset is constructed to enable the successful detection of people far from the camera in a large populated area. Fig. 15 shows a comparison between results using AMD dataset and results without AMD dataset. The essence of the proposed method is coordinates analysis of bounding box in the detected facial and mask areas. Though YOLO detector is used in this paper, it can be used with other detectors as necessary. Tab. 3 shows comparison of performance comparison of various detectors; the proposed methods, and various situations using AMD.



Figure 15: Detection result of wearing a mask: (a) without AMD; (b) with AMD

Table 3: Comparative results of different mask wearer detection methods

Detector	Objects	AMD	Proposed method	mAP@.50	Aver. FPS
YOLOv4-416 [25]	20			0.759	22.2
YOLOv4-416	20	✓		0.889	22.2
YOLOv4-416	20		✓	0.826	19.8
YOLOv4-416	20	✓	✓	0.952	19.8
Faster-RCNN [54]	20			0.785	12.1
Faster-RCNN	20	✓		0.901	12.1
Faster-RCNN	20		✓	0.836	10.9
Faster-RCNN	20	✓	✓	0.941	10.9

(Continued)

Table 3: Continued

Detector	Objects	AMD	Proposed method	mAP@.50	Aver. FPS
SSD-513 [26,28]	20			0.731	16.7
SSD-513	20	✓		0.871	16.7
SSD-513	20		✓	0.795	13.5
SSD-513	20	✓	✓	0.885	13.5

6 Conclusion

In this paper, an effective and reliable facial mask-wearing status detection method is proposed to prevent viral respiratory infections. Various existing open datasets and additional Advanced Mask Dataset are introduced and used for model learning to carry out proposed method, and the AMD is collected considering various external environments and disturbances; blurred image, different climate conditions, occlusion disturbances, dazzling and darkroom, various scale changes of objects, mask types, rotation, and multiple objects. This method can also applied in general circumstances especially in indoor situations by detecting facial and mask area using trained weight files and AMD. The mask-wearing status are divided into three types which are WM, AC, and WO. The gradients of each regression line is deeply considered through the experiment, and by calculating the orthogonal distances of newly measured data, proposed method, detection system, classifies the object's mask-wearing status. Lastly, this paper presents the results of comparative experiment under different situations using AMD and the proposed method, and as a result, we confirmed that this proposed method is not restricted to a specific detector but can create synergy with multiple other detector and the AMD had verified validity. That is, both AMD data and the proposed method are expected to be used as the cornerstone of technology to prevent pandemic situations from viral respiratory infections such as COVID-19.

Acknowledgement: This research was supported by a grant (2019-MOIS32-027) of Regional Specialized Disaster-Safety Research Support Program funded by the Ministry of Interior and Safety (MOIS, Korea), and This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-01972).

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] P. Zhou, X. Yang, X. Wang, B. Hu, L. Zhang *et al.*, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, no. 7798, pp. 270–273, 2020.
- [2] F. Wu, S. Zhao, B. Yu, Y. Chen, W. Wang *et al.*, "A new coronavirus associated with human respiratory disease in China," *Nature*, vol. 579, no. 7798, pp. 265–269, 2020.
- [3] World Health Organization (WHO), Coronavirus disease 2019 (COVID-19): Situation report 46. 2020. [Online]. Available: <https://www.who.int/docs/default-source/coronavir-use/situation-reports/20200426-sitrep-97-covid-19.pdf>.
- [4] C. Wang, P. Horby, F. Hayden and G. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*, vol. 395, no. 10223, pp. 470–473, 2020.

- [5] D. Hui and A. Zumla, "Severe acute respiratory syndrome: Historical, epidemiologic, and clinical features," *Infectious Disease Clinics*, vol. 33, no. 4, pp. 869–889, 2019.
- [6] E. Azhar, D. Hui, Z. Memish, C. Drosten and A. Zumla, "The Middle East respiratory syndrome (MERS)," *Infectious Disease Clinics*, vol. 33, no. 4, pp. 891–905, 2019.
- [7] V. Corman, D. Muth, D. Niemeyer and C. Drosten, "Hosts and sources of endemic human coronaviruses," *Advances in Virus Research*, vol. 100, pp. 163–188, 2018.
- [8] K. Andersen, A. Rambaut, W. Lipkin, E. Holmes and R. Garry, "The proximal origin of SARS-CoV-2," *Nature Medicine*, vol. 26, no. 4, pp. 450–452, 2020.
- [9] M. Yamin, "Counting the cost of COVID-19," *International Journal of Information Technology*, vol. 12, no. 2, pp. 311–317, 2020.
- [10] W. Parmet and M. Sinha, "Covid-19—the law and limits of quarantine," *The New England Journal of Medicine*, vol. 382, no. 5, pp. e28, 2020.
- [11] J. Singh, "COVID-19 and its impact on society," *Electronic Research Journal of Social Sciences and Humanities*, vol. 2, no. 1, pp. 168–172, 2020.
- [12] J. Ae Ri and H. Eun Joo, "A study on anxiety, knowledge, infection possibility, preventive possibility and preventive behavior level of COVID-19 in general public," *Journal of Convergence for Information Technology*, vol. 10, no. 8, pp. 87–98, 2020.
- [13] G. Alberca, I. Fernandes, M. Sato and R. Alberca, "What is COVID-19," *Frontiers Young Minds*, vol. 74, no. 8, pp. 1–8, 2020.
- [14] M. Ciotti, M. Ciccozzi, A. Terrinoni, J. Wen Can, W. Cheng Bin *et al.*, "The COVID-19 pandemic," *Critical Reviews in Clinical Laboratory Sciences*, vol. 57, no. 6, pp. 365–388, 2020.
- [15] T. Safa, M. Seifeddine, A. Mohamed and M. Abdellatif, "Real-time implementation of AI-based face mask detection and social distancing measuring system for COVID-19 prevention," *Scientific Programming*, vol. 2021, no. 2, pp. 1–21, 2021.
- [16] K. Aishwarya, K. Puneet Gupta and S. Ankita, "A review of modern technologies for tackling COVID-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, pp. 569–573, 2020.
- [17] S. Halgurd Maghded, K. Zrar Ghafoor, A. Safa Sadiq, C. Kevin, R. Danda *et al.*, "A novel AI-enabled framework to diagnose coronavirus COVID-19 using smartphone embedded sensors: Design study," in *IEEE 21st Int. Conf. on Information Reuse and Integration for Data Science (IRI)*, Ohio, USA, pp. 180–187, 2020.
- [18] B. Joseph, L. Alexandra, P. Katherine Hoffman, L. Cynthia Sin Nga and M. Luengo Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *Computers and Society*, vol. 69, pp. 807–845, 2020.
- [19] M. Loey, M. Gunasekaran, M. N. Taha and N. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," *Measurement*, vol. 167, pp. 108288, 2021.
- [20] J. Mingjie, F. Xinqi and Y. Hong, "RetinaMask: A face mask detector," *ArXiv*, abs/2005.03950v2, 2020.
- [21] M. Loey, M. Gunasekaran, M. N. Taha and N. M. Khalifa, "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," *Sustainable Cities and Society*, vol. 65, pp. 102600, 2021.
- [22] K. Xiangjie, W. Kailai, W. Shupeng, W. Xiaojie, J. Xin *et al.*, "Real-time mask identification for COVID-19: An edge computing-based deep learning framework," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 15929–15938, 2021.
- [23] N. Preeti, J. Rachna, M. Agam, A. Rohan, K. Piyush *et al.*, "SSDMNV2: A real time DNN-based face mask detection system using single shot multi box detector and MobileNetV2," *Sustainable Cities and Society*, vol. 66, pp. 102692, 2021.
- [24] G. Jignesh Chowdary, P. Narinder Singh, S. Sanjay Kumar and A. Sonali, "Face mask detection using transfer learning of inceptionv3," in *Int. Conf. on Big Data Analytics*, Sonapat, India, pp. 81–90, 2020.
- [25] B. Alexey, W. Chien Yao and L. Hong Yuan Mark, "Yolov4: Optimal speed and accuracy of object detection," *ArXiv*, abs/2004.10934v1, 2020.

- [26] F. Cheng Yang, L. Wei, R. Ananth, T. Ambrish and B. Alexander, "DSSD: Deconvolutional single shot detector," *ArXiv*, abs/1701.06659v1, 2017.
- [27] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 779–788, 2016.
- [28] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed *et al.*, "SSD: Single shot multi box detector," in *European Conf. on Computer Vision (ECCV)*, Amsterdam, Netherlands, pp. 21–37, 2016.
- [29] D. Gordon, A. Kembhavi, M. Rastegari, J. Redmon, D. Fox *et al.*, "IQA: Visual question answering in interactive environments," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, USA, pp. 4089–4098, 2017.
- [30] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, USA, pp. 6517–6525, 2017.
- [31] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA, pp. 6848–6856, 2018.
- [32] T. Jesus, R. Albert, T. Sandra Viciano and L. Jaime, "Incorrect facemask wearing detection using convolutional neural networks with transfer learning," *Healthcare*, vol. 9, no. 8, pp. 524–540, 2021.
- [33] D. Yuchen, L. Zichen and Y. David, "Real-time face mask detection in video data," *ArXiv*, abs/2105.01816v1, 2021.
- [34] Y. Ji Min and Z. Wei, "Face mask wearing detection algorithm based on improved YOLO-v4," *Sensors*, vol. 21, no. 9, pp. 756–776, 2021.
- [35] V. Jain and E. Learned Miller, "FDDB: A benchmark for face detection in unconstrained settings," *UMass Amherst Technical Report*, vol. 2, no. 6, pp. 1–11, 2010.
- [36] Y. Shuo, L. Ping, L. Chen Change and T. Xiaoou, "Wider face: A face detection benchmark," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 5525–5533, 2016.
- [37] G. Shiming, L. Jia, Y. Qiting and L. Zhao, "Detecting masked faces in the wild with LLE-CNNs," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, USA, pp. 2682–2690, 2017.
- [38] N. Hajime, S. Vishwanath, Z. He and P. Vishal, "Pushing the limits of unconstrained face detection: A challenge dataset and baseline results," in *IEEE 9th Int. Conf. on Biometrics Theory, Applications and Systems (BTAS)*, Redondo Beach, CA, USA, pp. 1–10, 2018.
- [39] Y. Bin, Y. Junjie, L. Zhen and L. Stan, "Fine-grained evaluation on face detection in the wild," in *11th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, Ljubljana, Slovenia, vol. 1, pp. 1–7, 2015.
- [40] A. Cabani, K. Hammoudi, B. Halim and M. Mahmoud, "MaskedFace-Net—A dataset of correctly/incorrectly masked face images in the context of COVID-19," *Smart Health*, vol. 19, pp. 1–5, 2021.
- [41] K. Tero, L. Samuli and A. Timo, "A Style-based generator architecture for generative adversarial networks," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, California, USA, pp. 4401–4410, 2019.
- [42] J. Xinbei, G. Tianhan, Z. Zichen and Z. Yukang, "Real-time face mask detection method based on YOLOv3," *Journal of Electronics*, vol. 10, no. 7, pp. 837–853, 2021.
- [43] J. Uijlings, K. Van De Sande, T. Gevers and A. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision (IJCV)*, vol. 104, no. 2, pp. 154–171, 2013.
- [44] M. Gunasekaran, N. Mohamed and M. Nour Eldeen, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," *Measurement*, vol. 167, pp. 108288, 2021.
- [45] W. Jianfeng, Y. Ye and Y. Gang, "Face attention network: An effective face detector for the occluded faces," *ArXiv*, abs/1711.07246v2, 2017.
- [46] B. Ioan, "Color quotient based mask detection," in *Int. Symp. on Electronics and Telecommunications (ISETC)*, Timisoara, Rumania, pp. 1–4, 2020.
- [47] J. Mingjie, F. Xinqi and Y. Hong, "RETINAFACEMASK: A face mask detector," *ArXiv*, abs/2005.03950v2, 2020.

- [48] H. Kaiming, Z. Xiangyu, R. Shaoqing and S. Jian, "Deep residual learning for image recognition," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 770–778, 2016.
- [49] G. Andrew Howard, Z. Menglong, C. Bo, K. Dmitry, W. Weijun *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *ArXiv*, abs/1704.04861v1, 2017.
- [50] H. Karim, C. Adnane, B. Halim and M. Mahmoud, "Validating the correct wearing of protection mask by taking a selfie: Design of a mobile application check your mask to limit the spread of COVID-19," *Computer Modeling in Engineering & Sciences*, vol. 124, no. 3, pp. 1049–1059, 2020.
- [51] CIE International Commission on Illumination, ISO/CIE 19476: Characterization of the performance of illuminance meters and luminance meters, 2014.
- [52] W. Steven Euijong and L. Jae Gil, "Data collection and quality challenges for deep learning," in *Proc. of the VLDB Endowment*, Tokyo, Japan, vol. 13, no. 12, pp. 3429–3432, 2020.
- [53] Y. Mehmet Kerim, B. Yunus Can, O. Oguzhan, N. Ikizler Cinbis, D. Pinar *et al.*, "Wildest faces: Face detection and recognition in violent settings," *ArXiv*, abs/1805.07566v1, 2018.
- [54] R. Shaoqing, H. Kaiming, G. Ross and S. Jian, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.