# Image Processing for the Prevention of Infectious Diseases

## Determination of Mask Wearing, Measurement of Hand Washing Time, and Disinfection Support System

Sho Higuchi[1] · Shunichi Taniguchi[1,2] · Yuta Kawasaki[1] · Atom Sonoda[1,2]

## Abstract

The new coronavirus infection (COVID-19) has spread to numerous countries around the world since several cases of the disease were first reported in late December 2019 in China. Currently, the WHO strongly recommends infection prevention measures such as wearing masks, hand washing, and frequent disinfection of high-touch surfaces, but there were many arguments against infection prevention policy in March 2020. For example, the WHO did not recommend the use of masks for the healthy general public. In Japan, wearing a mask was required before the habit of wearing a mask was established, which gave additional works of checking whether customer was wearing a mask to the employees. To reduce the workloads of employee and ensure mask-wearing, we started providing the mask-wearing system free of charge on March 5, 2020. We also developed hand washing time estimation and disinfection support system. It is useful to accumulate data on the status of implementation of countermeasures by our application, which leads to gain useful knowledge regarding countermeasures against COVID-19 as well as other infectious diseases. In this paper, we describe the development and introduction impacts of these systems in a pandemic emergencies. In addition, because of the security and privacy issues in running these image analysis applications, we discuss the delivery methods suitable for each service.

✉ Sho Higuchi
  higuchi@lightblue-tech.com

  Shunichi Taniguchi
  taniguchi@lightblue-tech.com

  Yuta Kawasaki
  kawasaki@lightblue-tech.com

  Atom Sonoda
  atom@lightblue-tech.com

[1]  Lightblue Technology Inc., Tokyo, Japan

[2]  University of Tokyo, Tokyo, Japan

Ohmsha  Springer

## Introduction

The new coronavirus disease (COVID-19), which broke out in Wuhan, China (Hubei Province) around December 2019, has spread to numerous countries around the world. On March 11, 2020, the WHO declared COVID-19 a pandemic, and many countries took measures such as restrictions on traveling outside the country [1]. In March 2020, there were many arguments against the wearing of a mask during the pandemic. For example, while WHO did not recommend mask wearing for healthy members of the general population, the Japan's Ministry of Health, Labour and Welfare (MHLW) encouraged wearing a mask in public place. In Japan, although the habit of wearing a mask was not established, it was recommended to prevent the spread of the virus, which imposed additional works of checking whether customer was wearing a mask to the employees. Then it was worthwhile to implement immediately the mask-wearing check system to replace this new work with unattended, computerized work. Introducing this was expected to reduce the workload of employees.

There are two approaches to determine whether a mask is worn or not: object detection and face detection combined with classification. The former requires time and cost for collecting annotation data. If you ask a private company, it will take a month and cost about two to three dollars per image. In the latter option, we only need to apply a face detection model to the collected images and divide them into two categories: with and without masks. For the low cost and quickness of the development, we adopted the second and collected 1600 images of faces with mask by crawling and using an existing face detection model. We started providing the system free of charge on March 5, 2020, and have introduced it to 14 offices until April 2020.

We have also developed a hand washing time measurement system and a handrail contact counting system aiming to promote efficient handrail disinfection. These image analysis systems are useful to accumulate data on the status of implementation of countermeasures because the confusion was caused by the lack of evidence at the beginning of the infection. Since there are security and privacy problems in implementation depending on the situation, we discuss the forms of providing image analysis applications. These systems and data are expected to be useful for building new social systems in the post-COVID-19 era.

The rest part of this paper is organized as follows. In the next section, we review some of the relevant datasets and studies. In the three subsequent sections, we describe the mechanism and effect of assessing mask wearing, hand washing time and disinfection support system, respectively. The final section is devoted to discuss the advantages and challenges for each method of image analysis providing services.

## Related Work

After we started providing our application on March 2020, there has been a lot of progress in datasets and methods related to mask-wearing detection since then. For example, Wang et al. [2] provide the following three masked face datasets; Masked-Face Detection Dataset (MFDD), the Real-world Masked-Face Recognition Dataset (RMFRD), and the Simulated Masked-Face Recognition Dataset (SMFRD). MaskedFace-Net [3] is a dataset of faces with mask generated artificially with two labels, mask is correctly worn and incorrectly worn. On the Kaggle portal there are three-label datasets with mask, without mask, and incorrectly worn mask [4, 5]. The author [6] proposed Face-Mask Label Dataset (FMLD) and gave comparisons of ethnic, pose, and regional proportions with other datasets.

There are following two main approaches for determining mask-wearing. The first one leverages a pre-trained face detector in combination with a mask-wearing image classifier [6, 7] and the second is to perform detection and classification of faces in one shot [8–12]. The advantage of the first method is that it is less expensive to prepare the training data, and the disadvantage is that the pre-trained face detection model does not take occluded faces into account and may fail when facial features are hidden by a mask or hands. The advantage of the second is that the detection accuracy is relatively high because it is trained on faces with masks, and the disadvantage is that it is expensive to prepare the bounding box dataset [11]. When we started the development, we did not have enough datasets, and the first method was chosen due to the low cost and the quickness.

In the early stage of our development, the authors did not consider the case of incorrectly wearing of masks [7–9, 13], but recent papers now take into account such cases [6, 10–12]. As of March 2020, since it was difficult to collect images where the mask was not worn correctly, our system does not take into account it. We would like to improve it in the future by developing the above training datasets.

The feature of our system is that we provided it in JavaScript format to introduce it quickly. Users only need to access a browser to use it, and there are no privacy concerns because there is no need for network communication except for downloading the model. In addition, we did not need to prepare any computational resources and could provide it free of charge. For these reasons, the application was launched on March 5, 2020, and it had been installed in 14 offices by April 2020, when the state of emergency was declared for the first time in Japan.

## Mask-Wearing Classification Application

### Background

The MHLW and the WHO disagreed on the efficiency of wearing masks for the prevention of COVID-19 infection when the outbreak began, but the wearing of

**Table 1** Comparison of face detection models. Detection results for 105 people with masks

| Model | Good detection | False positive | Average processing time (s/image) |
|---|---|---|---|
| OpenCV (Cascade) | 56 | 56 | 0.023 |
| OpenCV (DNN) | 80 | 3 | 0.35 |
| Dlib (HOG) | 46 | 1 | 0.71 |
| Dlib (DNN) | 81 | 0 | 17 |

The OpenCV DNN model has the best balance between speed and accuracy

masks had become widely recommended later [14]. Signs, security guards, and staff members encouraged people to wear masks in the early stage of COVID-19 disease. However, these call and confirmation of wearing a mask were additional tasks for the employees so the workloads of personnel were increasing and their vigilance and frequency of alerts were limited. This fact gave us the great motivations to develop a mask-wearing classification application using image analysis to reduce the workload while still ensuring mask-wearing.

## System Development

We considered the following two methods for determining mask-wearing:

1. After cropping faces from an image utilizing a face detection model, the image classification model was applied.
2. Creating an object detection model capable of detecting the entire masked face.

Because the first method of creating teaching data using annotation work was low in cost, it proved better for early development and implementation. This method requires that the face can be detected even if it is masked. Then, the detection performance on masked faces was first verified.

The 105 people wearing masks in photos of existing face-detection models were compared in the following four methods:

1. Face detection using Haar Cascades with Open Source Computer Vision Library (OpenCV) [15]
2. Face detection model using Deep Neural Network (DNN) with OpenCV
3. Face detection model with Dlib [16] using Histogram of Oriented Gradients feature combined with a linear classifier
4. Face detection model using DNN with Dlib

Table 1 shows that the OpenCV DNN model has the best balance between speed and accuracy for masked face detection. We also performed a face detection experiment setting a web camera near a door. It was confirmed that the system can successfully detect people who are close to the camera as shown in Fig. 1. In this experiment, we

**Fig. 1** The mask-wearing classification application demo screen. The numbers represent the confidence of the mask classification. The green and red bounding boxes represent the results of face detection. Green one indicates that the face was classified to be wearing a mask, and red one indicates that it was classified not to be wearing a mask

**Table 2** Confusion matrix of mask-wearing classification model

|  | Predicted positive | Predicted negative |
| --- | --- | --- |
| Positive class | 99 | 1 |
| Negative class | 0 | 100 |
| Accuracy |  | 99.5% |
| Precision |  | 99.0% |
| Recall |  | 100% |
| F-measure |  | 99.2% |

Classification results for 200 faces with mask and without mask

used the MacBook Pro 2019 model (Specs: 1.4 GHz quad-core Intel Core i5, 16 GB 2133 MHz LPDDR3 memory).

As a dataset for creating a mask classifier, 1600 images of faces wearing masks and 1600 images of people not wearing masks was prepared. Images of faces wearing masks were obtained by applying the face detection model to these images and extracting faces wearing masks manually from the cropped images. Images of faces not wearing masks were created by adopting the face detection to a database of face photographs, Labeled Faces in the Wild [17, 18]. We classified images with and without masks using transfer learning from VGG16 [19] trained on ImageNet [20]. We added an additional, fully-connected layer of dimension 256 and used cross entropy as a loss function. The training data included 1500 images with and without masks; the remaining 100 images were used as test data. The resultant accuracy was 98.8% for the training data and 99.5% for the test data as shown in Table 2.

It efficiently takes less than 0.1 s to judge an image, so this application can be executed on a PC without a GPU. The alert system is designed to play a sound when someone is not wearing a mask for three consecutive frames.

## Installation Effects and Discussion

Facilities and companies that have implemented the system have reported reduced administrative burdens and increased rates of mask use. In terms of installation, it is

a lightweight model that works well on commercial laptops and does not require a network environment after downloading the software. The web application version does not require downloading. In any case, it is available for use right away. When the system was provided, it was before the COVID-19 crisis/pandemic was declared in Japan; thus, there were individual differences regarding the importance of wearing masks. Under those circumstances, this application contributed to the improvement of etiquette.

However, problems were found in terms of the accuracy of detection in practical use. The challenge of the decision accuracy of the model is due to training dataset bias as follows:

– Since most of the images are front-facing images, profile or slumped faces are identified as masked.
– Because the faces wearing masks are predominantly Asian while the faces not wearing masks are predominantly Western, Western people are more likely determined to be not wearing masks.

In early development, the model with a limited data set was applied, as discussed in Sect. 3.2. This is because we aimed to introduce the system as soon as possible during the aggressive spread of COVID-19. We received post-publication feedback on these issues and attempted to update them by re-gathering images. The latest version currently offered is the third generation. With these accuracy-improving measures, we are now able to achieve sufficient detection accuracy.

## Hand Washing Time Estimation

### Background

The WHO recommends hand washing under running water for more than a certain amount of time as an effective means of preventing the spread of infections [21]. To achieve this, timers have been installed in hand washing stations. Hand washing is considered to be more individualized than the wearing of masks, and a sufficient hand washing system is required in certain places. We have provided motion analysis systems using image analysis technology for safety management on construction sites and hygiene management in factories. It was expected that our technologies could be applied to a hand-washing detection, then we developed a hand-washing time-measurement system, aiming to ensure thorough hand washing as shown in Fig. 2.

### System Development

First, this system makes a decision on whether or not a person washed his/her hand in each frame. Then the results are analyzed chronologically to detect the start and end hand-washing. Finally, the total hand-washing time is calculated. It judges

**Fig. 2** The hand-washing time estimation application screen. The red box on the right represents the locations of vanity. The time above the red box represents the measured hand-washing time
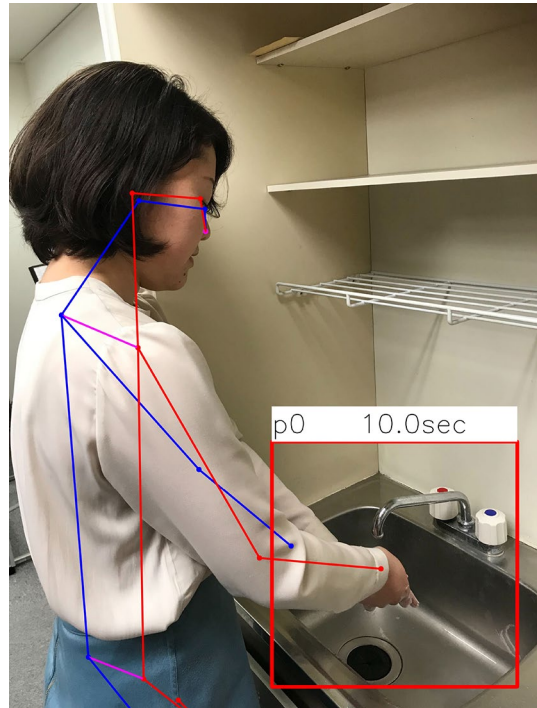


whether the hand coordinates obtained from human pose estimation fall within the specified coordinates. For human pose estimation, we use CenterNet [22] with Deep Layer Aggregation (DLA) [23]. CenterNet ascertains the coordinates of the wrists, but not the palms of the hands, so the coordinates of the palm were calculated as follows. Based on the coordinates of the elbow and wrist obtained from the estimation, the coordinates of the palm were obtained by extending a quarter of the distance from the elbow to the wrist toward the wrist. If both the coordinates of the right and left hands were within the predetermined coordinates which represent the locations of vanity, the person was considered to be washing his or her hands. Our previous system used multiple frames to assess motion, but this system only evaluated the position of the hands and washbasin in each frame. This is because we initially aimed for early development and easy implementation by reducing development man-hours as well as the computational complexity as in the case of the mask-wearing system. The mean average precision is 58.9 for the MS COCO dataset [24]. In addition, we conducted an experiment using personal data to determine the hand washing behavior. Assuming that there are multiple washbasins, we detected people at different distances of 2 m and 4 m, checking if the accuracy does not change with distance. The detection rate of hand positions was 100% in 300 frames at both 2 m and 4 m for 10 s.

Next, we discuss start and end detection. First of all, for each of the hand washing locations in the image, there are two states: hand washing and no hand washing. The start of hand washing is when the state of hand washing continues for more than

$th_1$ seconds, changing from no hand washing. In the same way, the end judgment is made when the state of no hand washing lasts more than $th_2$ seconds, changing from hand washing. The two thresholds, $th_1$ and $th_2$, can be set depending on the situation. These systems require a GPU for processing. We provided using a laptop (Galleria GCR2070RGF-E) with an NVIDIA GeForce RTX2070 MAX-Q. You can also use the cloud environment. In this case, it is possible to construct a relatively inexpensive system by running only for the necessary time and using a spot instance if the possibility of service outage is acceptable.

## Installation Effects and Discussion

Facilities and companies that have implemented this system have reported reduced administrative burdens. Some said that the voice alerts were too strong warning due to people's reactions. We are planning a new type of warning system using LEDs or screens. In addition, depending on the camera's position and the overlap of several people, human detection failed in some cases. We need to develop robust algorithms for object detection and pose estimation, even when the whole body is not visible.

In this system, we used CenterNet, which is a DLA backbone network, for pose estimation. This is because the DLA was found to be a relatively small network structure that allowed for efficient pose estimation. It provides a good balance in terms of computational complexity and accuracy when there are about 1–6 people in the image, such as in our system, in which there are many sideways and oblique images.

## Disinfection Support System

### Background

The WHO recommends appropriate disinfection of public places as well as hands to control the spread of the COVID-19 infection [25]. Thorough and frequent disinfection is required in public facilities and stores where people frequently come and go. In real-world, it is difficult to keep everything always clean because of the lack of sufficient personnel. There are no strict guidelines for disinfection and it is limited to disinfecting 2–3 times a day or more in areas that are likely to be touched by hands. We believe that the locations and frequencies of disinfection should be adjusted according to the number of times they come into human contact and length of stay. This allows to reduce the workloads and perform critical disinfection.

We have been providing a motion analysis system using image analysis technology for the management of safety behaviors in workplaces. One of these systems we have already developed detects safe and unsafe ascents and descents of stairs. It detects and warns of unsafe behaviors such as not holding the handrails, fast ascents and descents, looking away, and skipping steps. We applied this technology to our data acquisition, such as the number of times a person touched the handrail and

the number of people who took the stairs. Hence, we developed a handrail contact counting system, aiming to promote efficient handrail disinfection.

## System Development

The system uses the position of the handrails and the positions of the hands.

Firstly, there are two possible methods for detecting a hand on a handrail:

1. Automatic segmentation model detecting handrails
2. Manual segmentation of handrails following camera installation

Assuming that the position of the camera does not change frequently, we adopted the second method in this system because it was more advantageous in terms of development man-hours and accuracy, as well as the computational complexity. For widespread use as a general-purpose system, the system users need to be able to perform segmentation easily. Thus, we developed a segmentation GUI, which allows users to segment the handrails. Once users completed the segmentation on the GUI, the configuration information is exported for our system, and the location information on the handrails is imported by it.

This system determines whether or not a touch of the handrail occurred in each frame. The number of people and the amount of times touched are calculated in chronological order. The hands detector determines whether the wrist coordinates obtained from human pose estimation fall within the segmented handrail coordinates. For human pose estimation, we used CenterNet with DLA as well as hand washing time estimation.

In addition to the handrails, the system can also pre-set the stair positions by the segmentation GUI. It detects a handrail contact if the ankle coordinates are on the stairs. Even if there are several people on the stairs, we can record how many people touch the handrail by calculating and tracking the Intersection over Union of the rectangle surrounding the person in the previous frame.

Due to the difference in the angle of view depending on the location of the camera, we checked whether the frontal and left-to-right angles of view in relation to the stairs affected the detection accuracy of handrail contacts. The detection rate of the hand position in front of the camera was 100%, whereas in the left and right angles of view, the position of the wrist could not be detected because of the blind spot at the back of the handrail when climbing the stairs. To prevent the omission of touches, we designed the railing-view counting to be performed unconditionally from the time the above action was detected by the person's position on the stairs and direction of movement. The accuracy of this system was investigated for the assessment of handrail contact from the following angles of view, as shown in Fig. 3. A total of 287 people climbing up and down the stairs were studied. Table 3 shows that the assessment accuracy is good for 137 people who touched the handrails and 150 people who did not touch the handrails. These systems require a GPU for processing, and we utilized a Galleria GCR2070RGF-E laptop with an Nvidia
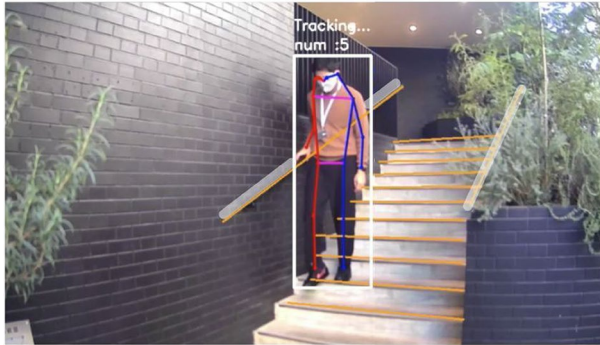
**Fig. 3** The appearance of the handrails contact assessment application. The orange lines represent the locations of the stairs and handrails specified in the GUI. The white border is the tracking result, and the skeleton inside the box represents the human pose estimation result

**Table 3** Confusion matrix of handrails contact assessment models

|  | Predicted positive | Predicted negative |
|---|---|---|
| Positive class | 127 | 10 |
| Negative class | 5 | 145 |
| Accuracy |  | 94.8% |
| Precision |  | 92.7% |
| Recall |  | 96.2% |
| F-measure |  | 94.4% |

Detection results for 287 people climbing up and down the stairs

GeForce RTX2070 Max-Q, also used for our hand washing time determination. This system can also used in the cloud environment.

## Installation Effects and Discussion

Using this system in real time, it is possible to display the number of people who touched the handrail since it was disinfected, which can be used as an indicator for disinfection frequency. In terms of accuracy, the skeletal structure of the wrist was not well detected when the shape and color of the handrail was similar to the person's clothing, or when the hand was in a blind spot, depending on the camera's angle. In addition, tracking sometimes failed when multiple persons were vertically aligned, as this system acquires two-dimensional results from an image and determines whether or not a person's hand is on the handrail based on the resulting two-dimensional skeletal coordinate information.

In the future, we plan to perform both pose estimation and registration of the position coordinates of the handrails in 3D to reduce the influence of the angle of view and the number of individuals. Unfortunately, the computational load is too large to handle in real time on the laptop PC used in this system because of the

increased computational complexity of 3D analysis. To address this issue, we are developing a lightweight 3D analysis model.

## Image Analysis Applications

This section discusses the three forms of providing image analysis applications: the download format (DL format), the JS format, which runs on a web browser with JavaScript, and the cloud format which runs on a cloud server.

### Offered as a Stand-Alone Application

The DL format was developed as a stand-alone application and is available via download. Both the mask-wearing and hand-washing detection systems were provided in this format at the beginning of the service. Users only need to download it once and run it easily. It is also advantageous in terms of privacy and security, as it does not require a connection to the Internet except for downloading. There is little concern about the video being obtained externally. The service provider does not need to provide computational resources, and it is easy to provide the service free of charge because the expenditure does not increase as the number of users increases. Although there are concerns about the copying and reverse engineering of models, the constraints imposed by the terms and conditions are sufficient in cases of high public interest and urgency. Because the execution environment is dependent on the performance of each user's computer, the execution speed cannot be guaranteed, but this point was resolved, to some extent, through the recommended computer's performance.

### Provided on a Web Browser

The JS format is designed to allow deploying deep learning models on a web browser. Now, the current mask-wearing detection system is provided in this format. The system was created using the framework tensorflow.js [26]. Users can use it by simply accessing the web page from their browsers, which is the most convenient way to use it. Additionally, since the image processing takes place on a computer provided by the user, it is as secure in terms of privacy and security as the DL format. This method is also easy to provide free of charge because the service provider does not need computational resources and the expenditure does not increase even if the number of users increases. In fact, service providers need to be careful due to the fact that, in the case of web applications, users may perceive the images as being sent to the outside.

### Provided on a Cloud Server

The cloud format allows users to send images from their computers to a cloud machine for image analysis. Users need to have a permanent connection to the

Internet to send images. Since the image processing is performed on a cloud server, it is expected to produce satisfactory results even with low-performance computers. Also, the service provider supplies the image processing results, which is advantageous in terms of intellectual property protection. On the other hand, there is a disadvantage that the service provider needs computational resources and the fee for the service has to increase. However, in cases such as our system, where service outage is not considered a serious accident, it is possible to construct services at relatively low costs by utilizing spot instances. Since the video is transmitted externally, user communication and system construction are necessary to resolve security concerns. Because of these features, our system utilizes the cloud when individually requested.

## Summary

In this paper, we present a case study on the detection of mask-wearing, the measurement of necessary hand washing time and the disinfection support system by image analysis.

For the mask-wearing detection, the user only needs a computer and a web camera. It judges whether people on the web camera are wearing masks, and if not, it prompts them to wear masks via text on the screen and voice from a speaker. The application can be downloaded as a standalone application for Mac OS[1] and Windows.[2] The web application with JavaScript is also available free of charge.[3]

For hand-washing time measurement, the user only requires a computer and a web camera. The system determines whether a person is washing his or her hands in the video taken by the web camera. If the hand-washing time is insufficient, text on the screen and a voice over the speakers will encourage the user to wash their hands sufficiently. The application is available free of charge as a stand-alone application for Mac OS and Windows.[4] These two applications have been available free of charge since March 5, 2020, and were implemented in 14 offices before the first declaration of the state of emergency in Japan. To the best of our knowledge, it was one of the earliest similar efforts in Japan and was unique in that it was provided free of charge.

These applications were easy to implement in the office due to the following four advantages:

1. Versatility ensured by image analysis utilizing deep learning
2. Ability to be used in an environment disconnected from the Internet, except for installation
3. Less security concerns as no recording is done

---

[1] https://mask.lightblue-tech.comdetect_mask_mac.

[2] https://mask.lightblue-tech.comdetect_mask_win.exe.

[3] https://mask-app.lightblue-tech.com.

[4] contact: https://www.lightblue-tech.com/contact/.

4. Free of charge

They can be introduced in a wide range of offices, regardless of size, and are considered to be very useful in preventing the spread of infections.

The disinfection support system can count the number of times a handrail is touched by combining motion analysis and object detection. Based on the information of the number of contacts, we can identify the areas to be disinfected intensively. This is also beginning to be tested and will be further improved with developments such as 3D pose estimation.

The introduction of our system was expected to be effective in preventing the spread of infections and reducing management costs. We confirmed the usefulness of image analysis technology in combating infectious diseases. Furthermore, by introducing these systems, it is possible to collect data such as the rate of mask use, hand-washing practices and the number of times a handrail is touched in public places in a privacy-sensitive manner. We expect that our systems and applications have great impacts on gathering previously unavailable data on measures of preventing the spread of infections, which will give us useful knowledge regarding countermeasures against COVID-19 as well as other infectious diseases, such as influenza.

# References

1. W.H. Organization: Who director-general's opening remarks at the media briefing on COVID-19—11 march 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (2020). Accessed: Dec 8 2020
2. Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., Pei, Y., et al.: (2020) arXiv preprint arXiv:2003.09093
3. Cabani, A., Hammoudi, K., Benhabiles, H., Melkemi, M.: MaskedFace-Net—a dataset of correctly/incorrectly masked face images in the context of COVID-19. Smart Health **19**, 100144 (2021). https://doi.org/10.1016/j.smhl.2020.100144
4. Larxel. Face mask detection. https://www.kaggle.com/andrewmvd/face-mask-detection. Accessed: Aug 26 2021
5. W. Intelligence. Face mask detection dataset. https://www.kaggle.com/wobotintelligence/face-mask-detection-dataset. Accessed: Aug 26 2021
6. Batagelj, B., Peer, P., Štruc, V., Dobrišek, S.: How to Correctly Detect Face-Masks for COVID-19 from Visual Information? Applied Sciences **11**(5) (2021). https://doi.org/10.3390/app11052070
7. Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P.: SSDMNV2: a real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. J. Hemanth. Sustain. Cities Soc. **66**, 102692 (2021)
8. Loey, M., Manogaran, G., Taha, M.H.N., Khalifa, N.E.M.: Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustain. Cities Soc. **65**, 102600 (2021)
9. Jiang, M., Fan, X., Yan, H.: (2020). arXiv preprint arXiv:2005.03950
10. Qin, B., Li, D.: Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. Sensors (2020). https://doi.org/10.3390/s20185236
11. Ding, Y., Li, Z., Yastremsky, D.: (2021). arXiv preprint arXiv:2105.01816
12. Yu, J., Zhang, W.: Face mask wearing detection algorithm based on improved YOLO-v4. Sensors (2021). https://doi.org/10.3390/s21093263
13. Loey, M., Manogaran, G., Taha, M.H.N., Khalifa, N.E.M.: A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement **167**, 108288 (2021)

14.  W.H. Organization: Coronavirus disease (COVID-19) advice for the public: Wwhen and how to use masks. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks (2020). Accessed: Dec 8 2020
15.  Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000)
16.  King, D.E.: Dlib-ml: a machine learning toolkit. J. Mach. Learn. Res. **10**, 1755 (2009)
17.  Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Tech. Rep. 07-49. University of Massachusetts, Amherst (2007)
18.  Learned-Miller, G.B.H.E.: labeled faces in the wild: updates and new reporting procedures. Tech. Rep. UM-CS-2014-003. University of Massachusetts, Amherst (2014)
19.  Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (2015)
20.  Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
21.  W.H. Organization: Hand hygiene: Why, how & when? https://www.who.int/gpsc/5may/Hand_Hygiene_Why_How_and_When_Brochure.pdf (2020). Accessed: Dec 8 2020
22.  Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., Tian, Q. (2019) arXiv preprint arXiv:1904.08189
23.  Yu, F., Wang, D., Shelhamer, E., Darrell, T.: Deep layer aggregation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2403–2412 (2018)
24.  Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: common objects in context. In: European Conference on Computer Vision, pp. 740–755. Springer, Heiderberg (2014)
25.  W.H. Organization: Cleaning and disinfection of environmental surfaces in the context of COVID-19: interim guidance, 15 may 2020. Technical documents (2020)
26.  Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/ (2015). (**Software available from tensorflow.org**)

**Publisher's Note**  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.