Active Face Mask Detection for Social Distancing Mobile Robot

*Sumiya EJAZ, Ayanori YOROZU, Akihisa OHYA
Graduate School of Science and Technology, University of Tsukuba, Ibaraki, Japan

I. Introduction

In the context of the transmission of coronavirus infection, the World Health Organization (WHO) advised different nations to guarantee that their citizens are wearing face masks in public places to prevent the person-to-person spread of disease. They also recommend keeping a social distance should be maintained between individuals. Before COVID-19, some people wore masks to protect their health from air pollution, pollen allergy, and only a few health professionals wore masks while working in medical clinics. Now, it becomes part of everyone's daily life. According to the World Health Organization (WHO, as of July 09, 2021), 185,038,806 confirmed cases of COVID-19, including 4,006,882 deaths and a total of 3,032,217,959 vaccine doses have been administered[1]. Most of the positive instances are observe in overcrowding and congested areas. As a result, scientists advised that wearing a mask in public settings can prevent us. Therefore, detecting face masks and maintaining acceptable social distances has become a critical computer vision challenge for supporting the world population. The French government has launched an operation in metro stations to identify people who are not wearing masks by creating AI (Artificial Intelligence) software integrated with security cameras in Paris metro stations [2]. We can use Artificial intelligence (AI) techniques such as machine learning (ML) and deep learning (DL) to inhibit the spread of COVID-19 in a variety of ways to develop an alert system to monitor the spread of disease [3].

This work aims to design a robot based on detecting people and face masks in a real environment using an omnidirectional camera. The challenging issue is recognizing face masks and calculating social distance when people's faces are not in front of the camera. In this situation, the robot moves towards those people and checks the facemask. Suppose people who do not wear a facemask, robot display a warning on the screen. Similarly, for social distancing, if people do not keep an appropriate distance, a robot alerts them to move away and maintain social distance. The proposed method uses the YOLOv5 to identify people who are not wearing a mask in public places. We also presented a technique to find people's distance from the camera on an image-based approach.

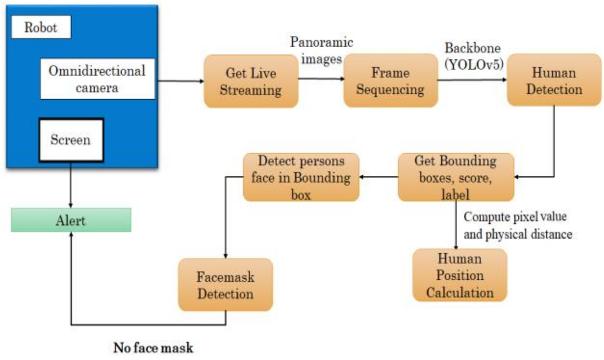

Figure 1: System Architecture of (i) human and face mask detection, (ii) Human position calculation

The rest of the paper is organized into four sections: Section II: deals with reviewing previous works in the past. The proposed methodology describes in Section III . The experimental results present in Section IV:. Finally, in Section V: the conclusion and the future work are delivered.

II. Related Work

In recent years, object identification approaches utilizing deep models have made phenomenal development in computer vision, and they are potentially more capable than shallow models in tackling complicated problems. Feature learning, contextual information learning, and occlusion handling are important aspects of deep models for person detection. In [4], a quadruped robot was designed specifically promoting social distancing in urban environments. F-RVO and YOLO were employed to extract detected objects and keep track of people, while multiple cameras and 3D Lidar sensor mounted on the robot.

Several facemask detection techniques were presented when the world began to take preventive measures against the coronavirus. A method was presented based on computer vision techniques to calculate social distance and identify whether a face mask. The method adopted a combination MobileNetV2 and Single Shot Detector (SSD) neural network with a transfer learning technique to balance resource limitations and recognition accuracy and implemented the model on raspberry pi4 to monitor activity and detect violations through the camera [5]. Recently researchers proposed a two-stage CNN architecture [6], with the first stage detecting human faces.

Sumiya Ejaz (Student at Master's Program in Computer Science, University of Tsukuba, Ibaraki, Japan)
Ayanori YOROZU (Assistant Professor at Department of Computer Science, University of Tsukuba, Ibaraki, Japan)
Akihisa OHYA (Professor at Department of Computer Science, University of Tsukuba, Ibaraki, Japan)

The second stage employed a lightweight image classifier to label the observed faces as 'Mask' or 'No Mask. ' Then, draw bounding boxes around them with the detected class name. For calculating object distance in [7], a new approach proposed, using a single image to train a system that identifies a mapping between an object's pixel height and physical distance for computing an object's distance—then utilized this mapping to calculate the physical distance between test objects and the image's pixel height. We also utilized this method for human position calculation.

Equirectangular panorama (ERA) has quickly become the primary format to store and transmit videos. ERA images create new challenges for computer vision and image processing as i) lack of annotated 360 datasets for many problems, ii) imagery are often of high-resolution to cover the viewing sphere with the reasonable resolution and iii) equirectangular projection creates severe geometric distortions for objects away from the central horizontal line [8]. Advanced techniques such as machine learning and deep learning strategies permit estimation of COVID-19 and accommodate to design an early prediction framework that can observe the assist spread of infection.

Based on the above context, various research publications have been published on facemask detection and calculating social distance to date. However, the present methodologies still need to be improved. As a result, a mobile robot designed based on a deep neural network technique employs the YOLOv5 model that perform face mask detection as well as calculate social distance in the fight against COVID-19 to contribute to further advances.

III. Methodology

In this section, we first discuss how our method effectively performs face mask detection and refer to people who are not wearing a face mask. Secondly, we stated a method for calculating distance between people on an image-based approach. Our overall system architecture is shown is shown in Figure 1.

A: Human and Face Mask Detection

From the perspective of machine vision, human detection is a massive challenge since it is affected by an extensive range of possible appearances due to changes in posture, clothing, lighting, and background. However, prior knowledge of these constraints can improve detection performance, and an intelligent framework recognizes and collects data of moving objects for accurate object classification. The proposed method uses an advanced deep learning model YOLOv5, OpenCV, TensorFlow to automatically detect people with or without facemasks in an indoor and outdoor environment with a camera integrated with a mobile robot. The YOLOv4 and YOLOv5 have a Cross Stage Partial (CSP) backbone and Path Aggregation Network (PA-NET) neck, but YOLOv5 differs from all previous releases because this is a PyTorch implementation rather than a fork from the original Darknet. The significant improvements include mosaic data augmentation and auto-learning bounding box anchors. For human detection, we identified two classes: The first class indicates the number of people in front of the camera and the second class for those in the opposite direction. For face mask detection, we identified two classes: the first one for those who wear a mask, the second one for those who do not wear a mask. In this process, we first get a live video streaming of panoramic images from an omnidirectional camera, read all input image frames, and then perform feature extraction with the help of deep neural network model YOLOv5 and obtained bounded boxes and classes. To evaluate the performance of YOLOv5, we trained the model for the various number of iterations on different epochs.

Figure 2: Hardware Platform

B: Human Position Calculation

A common rule is that when an object on the floor moves away from the camera, the object's footprint appears at a higher position in the image. We proposed a method to placed people in different locations and determined their footprint pixel value and physical distance. For this purpose, we performed an experiment on high resolution (1920*960), calculated the number of footprint pixels by knowing the exact distance, and made a reference table.

Figure 3: Sample images for training

We can obtain the physical distance in a testing environment and compare the actual results with estimation results and compute error differences. This work is still part of our future work.

IV: Experiment and Results

In this section, we elaborate on how our system was implemented on a robot, explain the model's accuracy, and proposed method of human position calculation.

A: Hardware Platform

We deploy a three-wheel robot of size 200*30 cm, as shown in Figure 2. To effectively detect people and perform facemask detection, the Theta V camera mounted on the robot. For the computational platform, we use GPU Nvidia GeForce Gtx 1080 Ti that supports the CUDA toolkit to accelerate the processing of images during the training process. We also deploy Intel(R) Core (TM) i5-8250U CPU and Robotic Operating System (ROS) for detection in a real-time environment.

B: Dataset Collection

We created our dataset in this study by selecting 1920*960, and 1024*512 resolutions and captured videos consisting of people with and without face masks. A small portion of public [9] dataset is also used. Makesense.ai tool is utilized to annotate objects. The dataset used to train our model consists of 1065 images. From this dataset, 745 images are used for training, 213 images are used for testing, while 107 images are used for validation. The images in the training data collection are classified into four categories: person, person' (this class identified those persons whose faces are not in front of the camera) mask, and no mask. Few sample images from the dataset are shown in Figure 3.

C: Human and Facemask detection

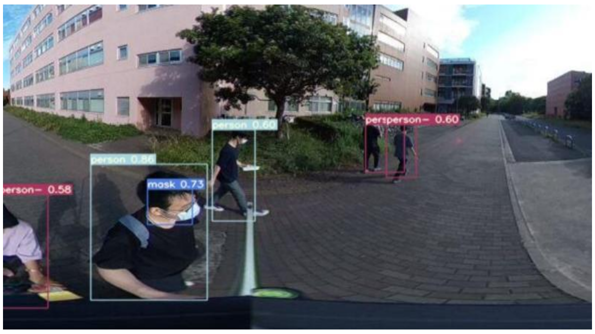

Figure 4: Result

i. Model Training

For training purposes, the proposed deep neural network model YOLOv5 is used in backend. The model is trained on GPU with labeled images when the data set is imported into the project directory. The images are downsized to 1024*512 pixels with a frame rate of 30 in pre-processing phases. Several performance metrics are used to evaluate the model's performance including Accuracy, Precision, and Intersection over Union (IoU). The performance metrics are formulated in terms of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The TP, FP, TN, and FN are represented in a grid-like structure called the confusion matrix which indicates the precision and recall accuracy of all the classes.

ii. Model Testing

We check the model's accuracy on the test dataset by showing the bounding box with the tag's name and the confidence score at the top of the box after trained the model with the custom data set. The proposed model detects all people in the camera's range by displaying a grey bounding box and showing a red bounding box for those in the opposite direction. Face mask detection is performed simultaneously by displaying bounding boxes on the identified person's face with mask or no-mask labels and confidence scores. Figure 4 shows the detection result. The performance for facemask detection with a precision and recall score is 80.2% of all the classes, with a confidence score of 0.5 is achieved which is shown in Figure 5.

iii. Model Implementation

The proposed system employs a mobile robot with an omnidirectional camera that monitors public places in a real environment. ROS is used to operate the trained model with the custom data set. The camera feeds real-time videos of public spaces to the model in the ROS, which extracts frames from the video and detects whether people are wearing masks. If a person is detected without a facemask, the robot will approach them and warn them by displaying a message on its screen.

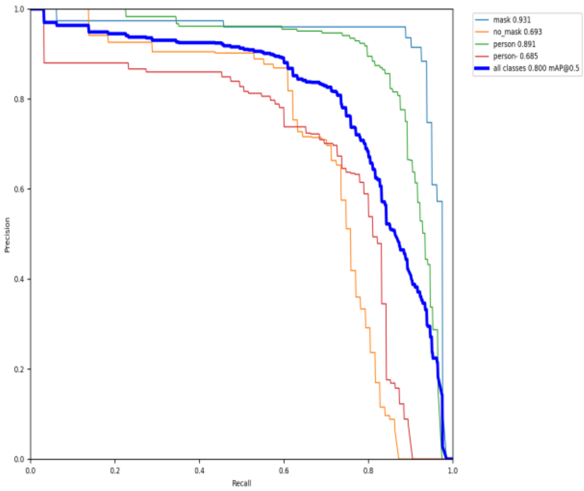

Figure 5: Precision and Recall score of all classes

D: Human Position Calculation

For human position calculation, we obtained number of images to calculate the footprint pixel value, for noise elimination and edge detection from the images, sobel filter is applied. Figure 6 illustrates the experiment and fitting of 2nd polynomial, representing the relationship between object depth and pixel value.

V: Conclusion

This work aims to design a robot based on detecting people and face masks in a real environment by using an omnidirectional camera to identify people who do not wear masks in public places. The results show that the system can accurately detect people and per-forms face mask detection from equirectangular panoramic images. We found that all the classes achieve good precision and recall accuracy but can be improved further. These results can be helpful in efforts to detect target persons based on YOLOv5 and suggested in what ways the detection performance can be better by increasing the training data and understanding of how the robot performs by utilizing the proposed method. Our proposed method has a few limitations; in the low intensity of light and distortions, it could not efficiently detect a social breach.

In future work, several of the currently under development features calculate the social distance between people and evaluate the robot performance in crowded situations.

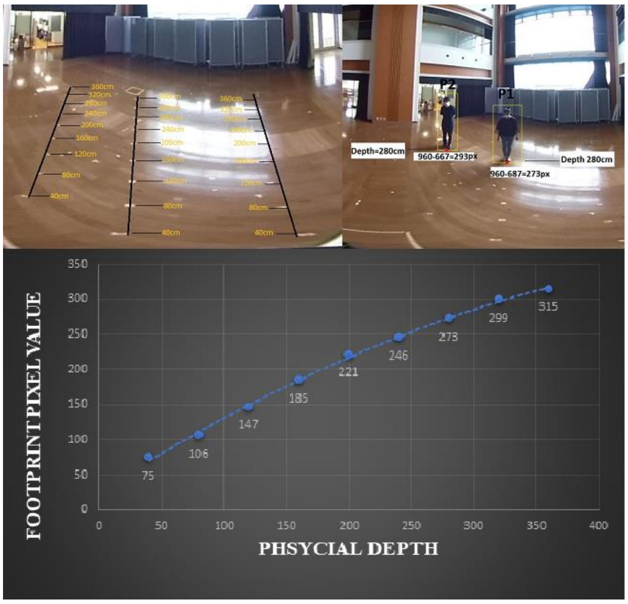

Figure 6: Human Position Calculation

REFERENCES

[1] Who coronavirus disease (covid-19) dashboard, https://covid19.who.int/, on- line accessed June 16, 2021.

[2] Paris tests face-mask recognition software on metro riders, http://bloomberg. com/, online accessed May 25, 2020

[3] Agarwal, S., Punn, N.S., Sonbhadra, S.K., Nagabhushan, P., Pandian, K., Saxena, P.: Unleashing the power of disruptive and emerging technologies amid covid 2019: A detailed review. arXiv preprint arXiv:2005.11507 (2020).

[4] T. Fan et al.,"Autonomous Social Distancing in Urban Environments using a Quadruped Robot," pp. 1–10, 2020.

[5] Yadav, S. (2020). Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID-19 Safety Guidelines Adherence. International Journal for Research in Applied Science and Engineering Technology, 8(7), 1368–1375.

[6] Chavda, A., Dsouza, J., Badgujar, S., & Damani, A. (2021). Multi-Stage CNN Architecture for Face Mask Detection. 2021 6th International Conference for Convergence in Technology, I2CT 2021.

[7] Rahman, A., Salam, A., Islam, M., & Sarker, P. (2008). An image-based approach to compute object distance. International Journal of Computational Intelligence Systems, 1(4), 304–312

[8] Yang, W., Qian, Y., Kamarainen, J. K., Cricri, F., & Fan, L. (2018). Object Detection in Equirectangular Panorama. Proceedings - International Conference on Pattern Recognition, 2018-August 2190–2195.

[9] https://github.com/uenian33/360_object_detection_dataset.