

Face Mask Detection and Social Distance Monitoring System for COVID-19 Pandemic

Iram Javed ¹, Muhammad Atif Butt ², and Samina Khalid ¹

¹Affiliation not available

²Information Technology University

October 30, 2023

Abstract

In this paper, a novel pipeline for developing an end-to-end masked/non-masked face detection method is proposed to improve the effectiveness of real-time surveillance systems at public places.

Face Mask Detection and Social Distance Monitoring System for COVID-19 Pandemic

Iram Javed¹, Muhammad Atif Butt², Samina Khalid^{1*}, Tehmina Shehryar³, Rashid Amin⁴, Adeel Muzaffar Syed⁵, and Marium Sadiq¹

¹ Department of Computer Science and Information Technology, Mirpur University of Science and Technology, Azad Jammu and Kashmir, Pakistan.

² Information Technology University, Punjab, Lahore, Pakistan.

³ Department of Software Engineering, Mirpur University of Science and Technology, Azad Jammu and Kashmir, Pakistan.

⁴ Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.

⁵ Department of Software Engineering, Bahria University, Islamabad, Pakistan.

*Corresponding Author: samina.csit@must.edu.pk

Abstract—Coronavirus triggers several respiratory infections such as sneezing, coughing, and pneumonia, which transmit humans to humans through airborne droplets. According to the guidelines of the World Health Organization, the spread of COVID-19 can be mitigated by avoiding public interactions in proximity and following standard operating procedures (SOPs) including wearing a face mask and maintaining social distancing in schools, shopping malls, and crowded areas. However, enforcing the adaptation of these SOPs on a larger scale is still a challenging task. With the emergence of deep learning-based visual object detection networks, numerous methods have been proposed to perform face mask detection on public spots. However, these methods require a huge amount of data to ensure robustness in real-time applications. Also, to the best of our knowledge, there is no standard outdoor surveillance-based dataset available to ensure the efficacy of face mask detection and social distancing methods in public spots. To this end, we present a large-scale dataset comprising of 10,000 outdoor images categorized into a binary class labeling i.e., face mask, and non-face masked people to accelerate the development of automated face mask detection and social distance measurement on public spots. Alongside, we also present an end-to-end pipeline to perform real-time face mask detection and social distance measurement in an outdoor environment. Initially, existing state-of-the-art single and multi-stage object detection networks are fine-tuned on the proposed dataset to evaluate their performance in terms of accuracy and inference time. Based on better performance, YOLO-v3 architecture is further optimized by tuning its feature extraction and region proposal generation layers to improve the performance in real-time applications. Our results indicate that the presented pipeline performed better than the baseline version, showing an improvement of 5.3% in terms of accuracy ¹.

Index Terms—Face Mask Detection, Social Distance Measurement, Single and multi-stage Detectors, Coronavirus.

¹The proposed dataset along with implementation is available at <https://github.com/iram1994/Face-Mask-Detection/>

I. INTRODUCTION

Coronavirus broke out at the end of 2019, and it is still devastating havoc on the livelihood and businesses of millions of people around the world [1]. Since the world has started recovering from the pandemic, people intend to return to a state of regularity, the same as before the pandemic. However, there is an upsurge of uneasiness among the people in getting back to their normal routine because this virus spreads through droplets of saliva from an infected person which can affect the people within the range of approximately 6 feet. The main symptoms of this infection are fever, headache, cough, respiratory difficulties, loss of taste, and smell ability which leads to the death of the infected person [2]. The incidence rate of COVID-19 is higher than other acute respiratory problems like severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS).

To prevent this deadly virus, World Health Organization (WHO) [3] issued guidelines and SOPs such as wearing a face mask and maintaining social distance in public spots. In this regard, several research studies also reported that maintaining the distance while physical interaction between people can prevent the spread of most respiratory diseases [4]. Tangana et. al [5] presented a mathematical model to demonstrate the impact of physical distance while interaction on transmission possibilities of virus among the people. In another study [6], it is demonstrated that wearing a face mask is highly effective in mitigating the reproduction of coronavirus. However, manual monitoring and enforcement of the aforementioned SOPs in public places such as schools, universities, shopping malls, and parks is a quite challenging task.

In step with the rapid advancement in Artificial Intelligence (AI), Deep Learning in particular, the computer vision

community has contributed various state-of-the-art methods for intelligent surveillance [7], object detection [8] and recognition [9], [10], and scene understanding [11]. These methods can be employed to develop an intelligent monitoring system for face mask detection and social distance measurement in public places. However, there are two main challenges in this direction. Firstly, to the best of our knowledge, there is no South Asian standard benchmark available to evaluate facial mask detection and social distance measurement methods. Secondly, there is no pipeline available for the development of an end-to-end real-time intelligent monitoring system for facial mask detection and social distance measurement. It is important to mention that several research studies have employed standard single- and multi-stage object detectors such as Faster-RCNN, SSD, and Retina-Net to perform face mask detection [12]. However, these methods do not consider the impact of social distance measurement, which make these methods insufficient for deployment in actual public places.

To address the aforementioned short-comings of existing state-of-the-art methods, in this paper, we have made the following contributions.

- 1) A local dataset containing 10,000 images based on two classes (i.e., masked face and unmasked face) has been collected from public places. It is worth noting that these classes are unique in orientation and dress codes, which are not covered in the existing datasets.
- 2) Existing state-of-the-art single and multi-stage object detectors are fine-tuned on the proposed dataset. Based on the analysis, an improved YOLO-v3 based object detection architecture is presented to enhance robustness of real-time surveillance systems.
- 3) Alongside, a machine-vision based distance measurement method has been proposed to ensure social distancing on public places.
- 4) Lastly, an extensive comparative study has been carried out between state-of-the-art Face mask detection methods and the proposed method to demonstrate the effectiveness of our proposed method in terms of higher detection and recognition accuracy, and inference time.

The rest of the paper is organized as follow. In Section II, we briefly discuss existing state-of-the-art facial mask detection and social distance measurement methods, along with the available datasets. In Section III, we present a detailed overview of our proposed end-to-end pipeline for face mask detection and social distance measurement. The experimental results have been presented in Section IV. Finally, the paper is concluded in Section V.

II. RELATED WORK

Real-time object detection and recognition methods can play an important role in developing intelligent monitoring methods for face mask detection and social distancing measurement to prevent coronavirus transmission. In this section,

we analyze the existing state-of-the-art methods employed in developing an intelligent monitoring system for face mask detection and social distancing measurement which includes: (i) single- and multi-stage detection methods—for face masked and non-masked face detection, (ii) Available Datasets—to develop generalized face detection systems and, (iii) social distance measurement methods.

A. Facial Mask Detection

In the majority of existing research works, the researchers focused on face construction and identity recognition while wearing face masks. However, the aim of this study is to identify the human face in both states—wearing the mask, or not wearing the mask in order to assist in reducing COVID-19 transmission and spread. In recent studies, researchers have demonstrated that wearing face masks minimizes the rate of COVID-19 spread as it can interrupt airborne germs effectively [31]. However, monitoring the people in public places is still a challenging task. In this regard, Zhang et al. [32] proposed a single shot refinement face detector namely Refine Face to detect people not wearing a face mask. In another research work, Jagadeeswari et al. [33] proposed SSD-based face mask detection method for an outdoor environment. Khandelwal et al. [34] presented a deep learning approach for classifying human face with and without mask. Onyema et al. [35] proposed method for facial expression recognition based convolutional neural network. Hussain et al. [15] proposed deep learning based IoT system to detect face mask using transfer learning approach.

Besides, aforementioned approaches achieved better accuracy on the respective test data. However, the real-time face mask detection is still a critical challenge for the system developers. In this regard, Snyder et al. [16] introduced deep learning based approach for mask detection to prevent COVID-19 transmission. Kodali et al. [17] presented custom CNN-based model to detect face wearing a mask in the public spots. Similarly, Sagayam et al. [18] proposed deep neural network based method for binary class (i.e., masked and non-masked) face state recognition. Degadwala et al. [19] proposed YOLO-v4 based face detection method which has been trained and tested over WIDER-FACE and MAFA datasets. Likewise, Taneja et al. [20] presented facial mask detection system with MobileNetV2 lightweight CNN and achieved 99.98% accuracy. On the other hand, Sethi et al. [36] aims to detect mask using ResNet-50. The model give 11.07% and 6.44% higher precision and recall and compared it with RetinaFaceMask detector model.

In another research work, Loey et. al [37] presented multi-stage detection method for face detection with wearing or not wearing mask. Alongside, ensemble method combined with deep learning model to detect face masks using real-world and synthetic data to improve the generalizability of machine learning models. These research works are discussed along with insightful strengths and limitations in the Table I. To this

TABLE I: An Overview of Existing Machine Learning Methods Used for Face Mask Detection and Recognition Tasks.

Author	Methods	Dataset	Accuracy	Limitation
Roy et al.[13]	MOXA included YOLO-v3, Tiny YOLO-v3, SSD and Faster R-CNN	Kaggle's medical masks dataset — 3000 images	YOLO-v3: 63.99% mAP, Tiny YOLOv3: 56.27% mAP, SSD: 46.52% mAP, and F-RCNN: 60.5% mAP	Unmanned approach MOXA requires improvement including more innovative object detectors
Nagrath et al.[14]	Single shot multibox object detection model and MobileNetV2	Kaggle's medical masks and PyImage search dataset contains 1,376 images	The SSDMNv2 model attained 92.64% accuracy	SSDMNV2 was trained on artificially produced images, still not tested in real situations as well as with real-time CCTV
Hussain et al.[15]	Transfer learning with CNN, VGG-16, MobileNetV2, ResNet-50, Inceptionv3	MAFA dataset, Masked Face-Net, and Bing dataset	Using VGG-16 achieved 99.81% accuracy and with MobileNetV2 attained 99.6% accuracy	Online accessible dataset contain noisy and construct by artificially, which is not suitable for real time system
Snyder et al.[16]	ResNet-50 with FPN and Multi-Task CNN	MCelebFaces Attributes, Microsoft Common Objects in Context, WIDER FACE dataset and Custom Mask Community Dataset	87.7% detection accuracy	Incorrectly identify faces with mask and without mask
Kodali et al.[17]	CNN model	Kaggle dataset with 853 images	96% detection accuracy	Incorrectly identify faces with mask and without mask
Sagayam et al.[18]	OpenCV and MobileNet-V2 used to detect face mask	Kaggle's medical masks and PyImage search	99% accuracy achieved by MobieNet-V2	Trained on limited dataset which is not perform well in real time situation
Degadwala et al.[19]	YOLO-v4	MAFA and WIDER-FACE dataset	99.98% accuracy obtained by YOLO-v4	Have need of more computational power and require 30FPS camera resolution rate
Taneja et al.[20]	MobileNet-V2 lightweight CNN model used to detect face mask	Medical Masks Dataset and the Face Mask Dataset	99.98% accuracy	Performance of MobileNet-V2 is not accurate as compared to Faster R-CNN and Inception-V2
Chadav et al.[21]	Multi-stage CNN model	Kaggle with 853 images	98% accuracy	Dual-stage CNN model do not detect side views of the face
Bhuiyan et al.[22]	YOLO-v3	Google colab datasets having 650 images	96% accuracy	Limited dataset used and cannot test on real time condition
Ejaz et al.[23]	Using Principal Component Analysis (PCA) recognition faces with masks and without the mask	ORL face dataset is used for masked faces containing 500 images	Attain accuracy for face mask is 72% and without the mask is 95%	PCA gave poor results in mask face, only front side face images use for the dataset
Qin et al.[24]	Classification of facial image with SRCNet and automatic identifies faces wear with mask	Medical Masks dataset having 3835 images	Acquire 98.70% accuracy with image super resolution classification network (SRCNet)	Use of limited number of images
Jiang et al.[25]	Proposed SSD to classify face with FPN	7959 images collected from internet	Without mask: 89.6% precision, with mask: 91.9% precision	Do not differentiate between mask and unmask face properly
Rahman et al.[26]	Facial mask detection in smart city through CCTV	1539 images are collected from different sources	98.7% Achieve accuracy	Confuse system with a hand covered face
Punn et al.[27]	YOLO-v3 used to monitoring real time social distance	800 images taken from OID dataset	YOLO v3 with deep sort acquire better result as compared to FPS	Privacy issue, do not record violations
Yang et al.[28]	Faster R-CNN and YOLOv4 detects real time social distance and critical density	Taken 12300 images from MS-COCO dataset	Accuracy and performance are good to monitor social distance	Do not record data, crowd analysis still a challenge
Militante et al.[29]	Single shot detector used to detect face mask and physical distance with alarm system	20000 images collected from web	accuracy rate of 97%	Do not detect face mask and distance at the same time
Yadav et al.[30]	face mask and social distance detection and generate an alert signal with SSD	used custom dataset of 3165 images	obtain accuracy 85% and 95%	N/A

end, we conclude that the deployment of the above-discussed face mask detection systems encounter several constraints at development and deployment level such as diverse types of face masks, face orientation, and illuminance conditions [13]. Furthermore, stabilizing object detection model accurateness and real time condition, placement of detector on system with limited computing capacity. In the circumstance of the epidemic, facial mask detection is not still explored in images, videos as well as closed circuit television (CCTV) to control transmission chain of virus [14].

B. Available Datasets

In the context of COVID-19, the face datasets have an essential role in training deep models for face mask and non-masked face detection. Recently, several datasets have been proposed to accelerate research in this direction. In this regard, Ge et al. [38] proposed MAFA dataset contains 30811 images which are collected from the Internet. These images have distinct types of masks, several occlusion degree and orientations. Furthermore, Laxel [39] introduced Face Mask Dataset (FMA) holds 853 images with three classes collected from Kaggle. Another extent version of kaggle dataset proposed by Wobot [40] denoted as FMA containing 6024 images having 20 classes. Rahmani et al. [26] proposed Medical Mask Dataset (MMD). The MMD dataset consist of 9067 images with three classes use to detect only medical mask. On the other hand, Wang et al. [41] proposed a large-scale dataset of masked faces for detection and recognition Masked-Face Detection Dataset (MFDD), the Real-world Masked-Face Recognition Dataset (RMFRD) and the Simulated Masked-Face Recognition Dataset (SMFRD). MFDD contain 24771 mask face image that were collected from the internet. The RMFRD have 2203 mask face image and 90,000 without mask images. SMFRD includes 50000 images. These detectors achieve 95% accuracy with the multi granularity model. We do not collect any images from the existing available dataset. Instead, we build a challenging dataset to perform experiments on existing object detector.

C. Social Distance Measurement

Social distancing is a significant safety measure to control the spread of COVID-19. Computer vision application has shown better applicability in detection [42] and emotion enable cognition task in real time environment [43]. In this regard, computer vision play an important role to dimensionality reduction with Matrix Factorization (MF) has valuable framework to treat against COVID-19 [44]. Additionally, Feature Selection and Prognosis Classification used to develop machine learning based intelligent system for COVID-19 disease [45]. The spectral clustering [46] and gene selection technique [47] has been presented to map to a low-dimensional space by merging node centrality

and community detection. Due to increase spread of COVID-19 outbreak cause serious condition to the global education systems. During the school closure, Computer Science innovation technologies have been useful and comfortable for teaching as well as learning [48]. Prem et al. [49] used susceptible-exposed-infected-removed (SEIR) method to study the special effects of social distancing on the spread of the virus. Levchev et al. [50] aimed to study a database configuration in multiple sensor technologies similar to cameras, LiDAR, inertial gyroscopes, wireless sensors and additional sensors used as data acquisition stages. Liang et al. [51] utilized various sensors to get image information and geographic location information at the same time build an indoor 3D chart using geographic coordinates. Niu et al. [52] highlighted social distancing problem in 3D view by using monocular cameras pedestrian 3D localization. Furthermore, Magoo et al. [53] setting bird eye view framework with YOLO v3 model to monitor social distance in public area. Though, the research community has contributed several social distance measurement methods, however, deployment of such systems in real-world environment is still a challenging task.

III. THE METHOD

To address the above-mentioned issues, we propose a novel pipeline for developing an end-to-end face mask detection methods to monitor the public spots in order to mitigate the COVID-19 spread, as shown in the Figure 1. Firstly, we present a large-scale **MUST Face Dataset (MFD)**—containing 10,000 images along with binary class bounding box annotations i.e., Face wearing mask, and Face not wearing mask. Alongside, we analyzed the existing state-of-the-art single stage and multi-stage object detector over our proposed dataset. Specifically, we fine-tuned the existing YOLO-v3 [54], SSD [55], RetinaNet-50 [56], Fast-RCNN [57], Faster R-CNN (FPN) [58], Faster-RCNN (ResNet-50) [59] and Faster-RCNN (ResNet-101) [60] on our proposed dataset through transfer learning. Based on the better performance, we further improved the YOLO-v3 architecture to robustify its performance in outdoor environment. On the basis of our face detector, we employed our self-proposed social distance measurement method—which takes input from the face detector and computes the distance between the two human beings to mitigate the COVID-19 spread in public spots.

A. MUST Face Dataset

To this end, we collect and release **MUST Face Dataset (MFD)**—a large-scale dataset to accelerate the development of generalized methods for end-to-end face mask detection in public places. Our MFD contains 10,000 images along with binary class (i.e., masked face, non-masked face) bounding box annotations. The proposed dataset is generated from

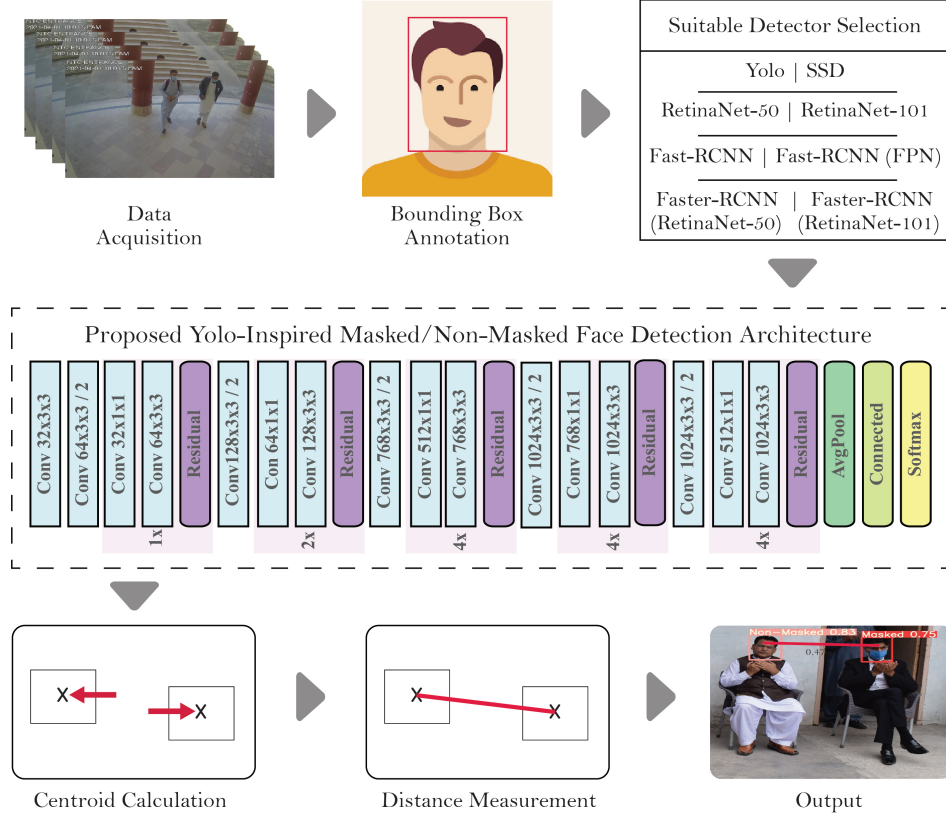


Fig. 1: The Proposed Pipeline For Developing Face Mask Detection And Social Distance Measurement in Public Places

the video sequences captured by the surveillance cameras installed at the outdoors of the departmental buildings. The average height of the installed cameras is in the range of 12 feet to 15 feet from the ground. After successful video sequence collection, the crowded frames are manually extracted while ensuring the quality control parameters such as positioning of the people and the clarity of the images. It is important to mention that we comply with the regulatory bodies and collected the data from the permitted areas. To protect the privacy, we do not disclose or release the personal identities, Geo-location, incoming and outgoing pattern based information of the people.

After completing frame extraction, considering the use-case of our proposed method, we defined two classes for annotations i.e., masked face, and non-masked face. For this purpose, we employed **LabelImg** annotation tool to label the human faces according to the aforementioned defined classes. One of the reasons of manual annotations instead of automated labeling is to maintain the accuracy of the coordinates of ground truth which plays an important role in training a robust face detection model. All the annotations are cross-validated by a team of experts to ensure the quality of ground truth. Some of the samples of our dataset are shown in the Figure 2.

B. Suitable Face Detection Method Selection

Till recently, deep object detection methods have demonstrated better applicability in various real-time object detection and recognition tasks [61]. To select the suitable deep learning object detector, firstly, we fine-tuned the existing state-of-the-art single-stage and multi-stage detection methods including YOLO-v3 [54], SSD [55], RetinaNet-50 [56], Fast-RCNN [57], Faster R-CNN (FPN) [58], Faster-RCNN (ResNet-50) [59] and Faster-RCNN (ResNet-101) [60] on our proposed MFD through transfer learning. The results show that existing YOLO-v3 outperformed aforementioned employed detection methods in terms of inference time and accuracy. Based on the better performance, we further improved the YOLO-v3 architecture to robustify its performance in outdoor environment.

C. Proposed Facial Mask Detection Architecture

In the proposed framework, we have employed YOLO-v3 architecture to perform facial mask detection in real-time, one of the most outstanding deep learning object detectors proposed by Joseph Redmon and Ali Farhadi in 2018 [62], which demonstrated consistent performance for object detection and recognition tasks. One of the main issues in existing detection network was the vanishing gradient



Fig. 2: Sample Images From Our MUST Face Dataset.

problem, which commonly occurs by increasing network layers. Therefore, multi-scale YOLO-v3 has been proposed which hold residual connections—which join the input from the previous layer to output of next layer similar to ResNet architecture. Resultantly, Yolo-v3 achieved good performance even over low resolution images due to inclusion of multi-scale feature extraction property. To this end, we employed the existing YOLO-v3 architecture and inserted k-means anchoring to 9 anchor boxes and then isolate them into three locations to get more more bounding boxes per image than baseline version.

The input layer takes an RGB image with a size of 416x416 pixels. As a backbone network, we employed DarkNet-53 to accomplish the maximum calculated floating-point procedure per second. The internal structure of the model includes fully connected network that does not contain max-pooling layer. As depicted in Figure 1 the network contains convolution block, residual block, and scale output layers. In convolution block, convolution functions of the kernel size hold strides instead of max pooling to reduce size of input images; each monitored by batch normalization and ReLU activation. On the other hand, residual block having different kernel size of two convolution block named as mega-block. In existing YOLO-v3 architecture, the convolution blocks iterates by 1x, 2x, 4x, and 8x. However, considering the use-case of our application, we reduced the iterations of convolution blocks to 1x, 2x, 4x in order to improve the learning performance and inference time. In the bottom of the architecture, an average pool, followed by a fully connected layer and softmax activation is employed to down-sample the feature map and get binary class output probability, respectively. To improve the learning process, we applied the concept of transfer learning to utilize the storing

knowledge of a neural network to do new tasks by simply learning new weights. The ultimate aim of employing this technique is to increase the learning process.

D. Social Distance Measurement Methods

With the recent advancement in the field of AI, computer vision based applications have demonstrated better applicability in several applications such as scene understanding, object recognition, speed, and distance estimation [63]. Some research used proportional-integral-derivative (PID) [64] due to it's simplicity and non-optimal performance. Since, it is suitable for distance measurement as well as will consume less power and memory. Zhang et al. [65], proposed distance estimation method to localization of an object in the camera coordinate frame. Their method contain three steps. The first step is regarding camera calibration and the second step is concerned with constitute a model for distance measurement between camera coordinate frame with their projection frame and third step is representing absolute distance estimation.

The distance is computed with respect to the pivot point of bounding box known as centroid—which is calculated using equation 1, mentioned below.

$$C(x, y) = \frac{\hat{x}_{min} + \hat{x}_{max}}{2}, \frac{\hat{y}_{min} + \hat{y}_{max}}{2} \quad (1)$$

It can be seen from equation 1, C means centroid—means that minimum and maximum width of the bounding box whereas y_{min} , y_{max} means that minimum and maximum height of the bounding box. Calculated centroid and then use Euclidean distance formula to measure distance between centroids, as shown in the 2 and then compared the distance with ground truth value.

$$D(C_2(x,y), C_1(x,y)) = \sqrt{(x_{min} - x_{max})^2 + (y_{min} - y_{max})^2} \quad (2)$$

After calculating centroid of bounding box, a unique ID is assigned to each centroid. In the next step, the distance between every detected centroid is computed using Euclidean distance. To validate the correctness, Root Mean Square Error (RMSE) (mentioned in the equation 3) to estimate the error between actual value and predicted value of the model.

$$RMSE = \sqrt{\sum_{i=1}^N \frac{(Predictedvalue - Actualvalue)^2}{N}} \quad (3)$$

E. Proposed Algorithm For Real-time Face Mask Detection

Here we present a novel algorithm, depicted in Algorithm 1, for developing and deploying an end-to-end face mask detection and social distance monitoring system in the public spots.

In the first step, the real-time stream of the camera get the visual frames—which is passed to our developed face mask detection method for inference. Our proposed method analyzes the frames, if there is no face detected, our network returns null. If face is detected, face detect and also compute distance between faces by using our proposed method. To find out the precautionary measure according to the facial mask and measure social distance, a discussion performed in the section 2. Following scenarios has been performed: if person wear a mask and distance is greater than 6 feet then no action performed. But when person not wearing a mask and social distance is greater than 6 feet then alert is high. On the other hand, when person wear mask and social distance is less than 6 feet again alarm generated. The masked person and not maintain social distance, then generated warning.

IV. EXPERIMENTS AND RESULTS

In this section, we evaluate the effectiveness of the proposed mask/non-mask face detection method and present the comparison study with current cutting-edge techniques. The studies are conducted on a powerful computer running a 64-bit version of Windows 10 that has an RTX 2080TI graphics card, an 11 GB DDR5 GPU, a core i9- 9900k CPU, and 32 GB of RAM.

A. Training Setup

The training process of the proposed pipeline is divided into three fundamental steps: data pre-processing, model training, and model evaluation. Firstly, the whole dataset is randomly split into training, validation, and test set with 80:10:10 percent ratio and normalized the input size to 416x416 pixel resolution. In the next step, Pytorch library is used for the implementation of the proposed

Algorithm 1:	Real-time surveillance Pro
1)	Proc FMSD ()
2)	While True
3)	Do
4)	Initiate VisualStream (Camera)
5)	detect-faces $\leftarrow$ <i>Proposed</i> - FM - <i>Detector</i> (VisualStream)
6)	if (detect-face.count $\leq$ 0)
7)	return
8)	else
9)	for i in range (detect-face.count):
10)	social-distance $\leftarrow$ ComputeDistance (x, y
	coordinates of i)
11)	if (i == 'masked' & social-distance $\geq$ 6ft)
12)	return
13)	else if (i == 'non-masked' & social-distance
	$\geq$ 6ft)
14)	Set Alarm = High
15)	else if (i == 'non-masked' & social-distance
	$\leq$ 6ft)
16)	Set Alarm = High
17)	else if (i == 'masked' & social-distance $\leq$ 6ft)
18)	Set warning = True
19)	else
20)	return
21)	endif
22)	end for
23)	end if
24)	end do
25)	End While
26)	End Proc

pipeline. Moreover, the experiments are categorized into three phases i.e. (i) evaluation of the existing state-of-the-art object detection networks on proposed dataset, and (ii) evaluation of improved Yolo-v3 network on proposed dataset, and (iii) evaluation of proposed distance measurement method.

B. Evaluation of Existing State-Of-The-Art Object Detection Networks On Proposed Dataset

To evaluate the existing state-of-the-art deep object detection models—YOLO-v3, SSD, RetinaNet-50, RetinaNet-101, Fast-RCNN, Faster R-CNN (FPN), Faster-RCNN (ResNet-50) and Faster-RCNN (ResNet-101) are fine-tuned on the proposed face mask detection dataset. Pytorch 1.4.0 library and cuda 11.0 version are used to configure the training runs. The hyper-parameters such as learning rate, batch size and epochs are set to 0.0001, 32, and 100 with the stochastic gradient descent optimizer to update model weights, respectively. The performance matrices of the employed models are shown in Table II.

It can be seen from Table II that single-stage detectors demonstrated better applicability in term of low inference time due to their less parametric architectures. Whereas, the multi-stage object detectors have been computationally expensive while achieving significantly higher inference time. It is also important to mention that Yolo-v3 with 53 layers demonstrated better accuracy than the SSD, RetinaNet-50,



Fig. 3: Qualitative examples of our masked/non-masked face detection method on our face mask dataset.

RetinaNet-101, Fast-RCNN, Faster R-CNN (FPN), Faster-RCNN (ResNet-50) and Faster-RCNN (ResNet-101). For instance, YOLO-v3 achieved 64.1% mean accuracy, 59.6% mAP, 53.1% mAP @ 0.95 and 28ms inference time. Similarly, SSD achieved 61.8% mean accuracy, mAP 56.2%, mAP @ 0.95 is 48.6% and take 34 prediction time. Also, RetinaNet-50 demonstrate 55.2% mean accuracy, mAP 51.9%, mAP @ 0.95 is 44.7% with the inference time of 37ms on the test set. Whereas, RetinaNet-101 achieved 51.0% mean accuracy, mAP 46.3%, and 44.7% mAP @ 0.95 with 39ms inference time which is comparatively higher than RetinaNet-50. On the other hand, We next analyze the multi-stage object detector i.e., Fast R-CNN which demonstrated 41.7% mean accuracy, 39.4% mAP, and 37.1% mAP @ 0.95 with 132ms inference time on the our test set which is significantly higher than the employed single shot detectors. In another experiment, Faster R-CNN based on FPN 119 achieved mean accuracy of 47.3%, 44% mAP, and 41.5% mAP @ 0.95. Whereas, the sample Faster R-CNN with ResNet-50 feature extraction network achieved mean accuracy of 59.0%, mAP 44%, and 57.4% mAP @ 0.95 with inference time of 108ms. However, with ResNet-101 as a backbone feature extraction network, Faster-RCNN shows mean accuracy of 62.7%, mAP 61.3%, and 59.0% mAP @ 0.95 with inference time of 98ms. Consequently, it can be assumed that YOLO-v3 with DarkNet-53 can achieve better accuracy after further architectural fine-tuning.

C. Evaluation of Improved YOLO-V3 Architecture On Proposed Dataset

Based on the above discussed analysis, the architecture of the YOLO-v3 is further improved by trimming the less con-

TABLE II: Evaluation of existing state-of-the-art object detection networks on proposed dataset

Method	Mean Accuracy	mAP	mAP @ 0.95	Inf.Time (ms)
YOLO-V3	64.1%	59.6%	53.1%	28
SSD	61.8%	56.2%	48.6%	34
RETINA-NET 50	55.2%	51.9%	44.7%	37
RETINA-NET 101	51.0%	46.3%	41.8%	39
FAST-RCNN	41.7%	39.4%	37.1%	132
FASTER-RCNN (FPN)	47.3%	44.0%	41.5%	119
FASTER-RCNN (ResNet-50)	59.0%	57.4%	55.6%	108
FASTER-RCNN (ResNet-101)	62.7%	61.3%	59.0%	98

tributing convolutional layers and residual connections. The improved feature extractor—DarkNet has been evaluated on the proposed dataset. In order to train the network faster, we employed transfer learning to learn the high level features from the proposed dataset. In the training setup, we employed SGD optimization algorithm with momentum to train and evaluate the improved network on our proposed dataset for mask/non-mask face detection tasks. The re-known performance metrics such as mean accuracy, mAP, mAP @ 0.95 and inference time are used to evaluate the performance of our improved face mask/non-mask face detection on our dataset. The mean accuracy refers to the sum of correct predictions divided by the sum of total data samples. Whereas, mAP denotes mean average precision, and AP @ 0.95 shows the average precision with 0.95 intersection over union. Furthermore, inference time refers to the total time taken from getting an input to producing an output.

TABLE III: Evaluation of improved YOLO-v3 on proposed dataset

Method	Mean Accuracy	mAP	mAP @ 0.95	Inf. Time
Existing				
YOLO-V3	64.1%	59.6%	53.1%	28 ms
Proposed				
YOLO-V3	69.4%	64.7%	62.0%	25 ms

TABLE IV: Results of proposed distance measurement methods

Sr. No.	Ground Truth (ft)	Predictions (ft)	RMSE
Distance 1	2.44	2.37	0.035
Distance 2	2.99	2.95	0.020
Distance 3	3.16	3.10	0.030

It can be seen from the Table III that our improved Yolo-V3 based detection network outperformed the baseline Yolo-v3 in mask/non-mask face detection tasks on our proposed dataset. One of the main reasons behind the increase of accuracy in our model is the trimming of less contributing residual connections with accelerated the performance of our model as compared to the baseline model. Some of sample results are demonstrated in figure 3 to show the effectiveness of our proposed masked/non-masked face detection method.

D. Evaluation of Proposed Distance Measurement Method

After evaluating our proposed mask/non-mask face detection, in next step, we evaluated our proposed machine-vision based distance measurement method to ensure social distancing on public places. Following the standard performance metrics, we employed root mean square error to analyze the correctness of our method as compared to the ground truth. Some of the quantitative analysis is shown in Table IV.

The vision based system detect faces of the person and give the bounding boxes information. Later on, detect the central point of the bounding boxes around the face and then measure distance between two central point (centroid) using the standard equation of euclidean distance. The error rate is computed using RMSE which computes the difference between ground truth value and predicted value of the model. For instance, in the Distance 1 sample, the actual distance (ground truth) between two persons is 2.44 feet, whereas our proposed vision-based distance measurement method predicts 2.37 with quite lesser error rate i.e., 0.035 RMSE. In the next data sample i.e., Distance 2, the actual distance is 2.99 whereas, our model inferred 2.95 with the RMSE of 0.020. Similarly, in Distance 3 sample, the ground truth value is 3.16 whereas, the proposed method predicts 3.10 holding the error rate of 0.030 which is quite effective performance on our test set.

V. CONCLUSION

In this paper, a novel pipeline for developing an end-to-end masked/non-masked face detection method is proposed to improve the effectiveness of real-time surveillance systems at public places. Alongside, a new dataset containing 10,000 images of two classes (masked face, non-masked face) is constructed to develop a generalized masked/non-masked face detection and social distance measurement in outdoor public places. While fine-tuning existing state-of-the-art single-stage and multi-stage detection methods, it is observed that Yolo-v3 outperformed the other networks in terms of accuracy and inference time. Based on analysis, we further improved the baseline Yolo-v3 by eliminating the less contributing residual connections in the network. Consequently, the results indicate that our customized YOLO-v3 performed better than baseline version, showing an improvement of 5.3% in terms of accuracy. In the future, we are aiming to extend our work to develop an image segmentation-based system that can provide accurate level information and gives greater clarity to detect face mask.

DATA AVAILABILITY

All the data used to support the findings of the study are available in the manuscript.

DISCLOSURE

Iram Javed and Muhammad Atif Butt are the joint first authors to this work.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTIONS

The research conceptualization and methodology were done by Iram Javed, Muhammad Atif Butt, and Samina Khalid. The technical and theoretical framework was prepared by Iram Javed and Muhammad Atif Butt. The technical review and improvement were performed by Tehmina Shehryar, Rashid Amin, Adeel Muzaffar Syed, and Marium Sadiq. The overall technical support, guidance, and project administration were done by Iram Javed, Muhammad Atif Butt, and Samina Khalid.

CODE AVAILABILITY

The proposed dataset along with implementation is available at <https://github.com/iram1994/Face-Mask-Detection/>

REFERENCES

- [1] X.-Y. Ge, Y. Pu, C.-H. Liao, W.-F. Huang, Q. Zeng, H. Zhou, B. Yi, A.-M. Wang, Q.-Y. Dou, P.-C. Zhou *et al.*, "Evaluation of the exposure risk of sars-cov-2 in different hospital environment," *Sustainable cities and society*, vol. 61, p. 102413, 2020.
- [2] K. A. Prather, C. C. Wang, and R. T. Schooley, "Reducing transmission of sars-cov-2," *Science*, vol. 368, no. 6498, pp. 1422–1424, 2020.
- [3] J. Middleton, J. M. Martin-Moreno, H. Barros, L. Chambaud, and C. Signorelli, "Aspher statement on the novel coronavirus disease (covid-19) outbreak emergency," pp. 237–238, 2020.
- [4] R. Katz, A. Vaught, and S. J. Simmens, "Local decision making for implementing social distancing in response to outbreaks," *Public Health Reports*, vol. 134, no. 2, pp. 150–154, 2019.
- [5] A. Atangana, "Modelling the spread of covid-19 with new fractal-fractional operators: can the lockdown save mankind before vaccination?" *Chaos, Solitons & Fractals*, vol. 136, p. 109860, 2020.
- [6] M. H. Goldberg, A. Gustafson, E. W. Maibach, M. T. Ballew, P. Bergquist, J. E. Kotcher, J. R. Marlon, S. A. Rosenthal, and A. Leiserowitz, "Mask-wearing increased after a government recommendation: A natural experiment in the us during the covid-19 pandemic," *Frontiers in Communication*, vol. 5, p. 44, 2020.
- [7] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.
- [8] M. A. Butt and F. Riaz, "Carl-d: a vision benchmark suite and large scale dataset for vehicle detection and scene segmentation," *Signal Processing: Image Communication*, vol. 104, p. 116667, 2022.
- [9] M. A. Butt, A. M. Khattak, S. Shafique, B. Hayat, S. Abid, K.-I. Kim, M. W. Ayub, A. Sajid, and A. Adnan, "Convolutional neural network based vehicle classification in adverse illuminous conditions for intelligent transportation systems," *Complexity*, vol. 2021, 2021.
- [10] M. A. Butt, A. Ul-Hasan, and F. Shafait, "Traffsign: Multilingual traffic signboard text detection and recognition for urdu and english," in *International Workshop on Document Analysis Systems*. Springer, 2022, pp. 741–755.
- [11] M. Rasib, M. A. Butt, S. Khalid, S. Abid, F. Raiz, S. Jabbar, and K. Han, "Are self-driving vehicles ready to launch? an insight into steering control in autonomous self-driving vehicles," *Mathematical Problems in Engineering*, vol. 2021, 2021.
- [12] A. Iftikhar *et al.*, "An insight into facial mask and social distance monitoring system based on deep learning object detector to prevent covid-19 transmission," in *Sintez 2021-International Scientific Conference on Information Technology and Data Related Research*. Singidunum University, 2021, pp. 120–127.
- [13] B. Roy, S. Nandy, D. Ghosh, D. Dutta, P. Biswas, and T. Das, "Moxa: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks," *Transactions of the Indian National Academy of Engineering*, vol. 5, no. 3, pp. 509–518, 2020.
- [14] P. Nagraath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, "Ssdmmv2: A real time dnn-based face mask detection system using single shot multibox detector and mobilenetv2," *Sustainable cities and society*, vol. 66, p. 102692, 2021.
- [15] S. Hussain, Y. Yu, M. Ayoub, A. Khan, R. Rehman, J. A. Wahid, and W. Hou, "Iot and deep learning based approach for rapid screening and face mask detection for infection spread control of covid-19," *Applied Sciences*, vol. 11, no. 8, p. 3495, 2021.
- [16] S. E. Snyder and G. Husari, "Thor: A deep learning approach for face mask detection to prevent the covid-19 pandemic," in *SoutheastCon 2021*. IEEE, 2021, pp. 1–8.
- [17] R. K. Kodali and R. Dhanekula, "Face mask detection using deep learning," in *2021 International Conference on Computer Communication and Informatics (ICCCI)*. IEEE, 2021, pp. 1–5.
- [18] K. M. Sagayam *et al.*, "Cnn-based mask detection system using opencv and mobilenetv2," in *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*. IEEE, 2021, pp. 115–119.
- [19] S. Degadwala, D. Vyas, U. Chakraborty, A. R. Dider, and H. Biswas, "Yolo-v4 deep learning model for medical face mask detection," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. IEEE, 2021, pp. 209–213.
- [20] S. Taneja, A. Nayyar, P. Nagraath *et al.*, "Face mask detection using deep learning during covid-19," in *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security*. Springer, 2021, pp. 39–51.
- [21] A. Chavda, J. Dsouza, S. Badgajar, and A. Damani, "Multi-stage cnn architecture for face mask detection," in *2021 6th International Conference for Convergence in Technology (I2CT)*. IEEE, 2021, pp. 1–8.
- [22] M. R. Bhuiyan, S. A. Khushbu, and M. S. Islam, "A deep learning based assistive system to classify covid-19 face mask for human safety with yolov3," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2020, pp. 1–5.
- [23] M. S. Ejaz, M. R. Islam, M. Sifatullah, and A. Sarker, "Implementation of principal component analysis on masked and non-masked face recognition," in *2019 1st international conference on advances in science, engineering and robotics technology (ICASERT)*. IEEE, 2019, pp. 1–5.
- [24] B. Qin and D. Li, "Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19," *Sensors*, vol. 20, no. 18, p. 5236, 2020.
- [25] N. Jiang, Y. Lu, S. Tang, and S. Goto, "Rapid face detection using a multi-mode cascade and separate haar feature," in *2010 International Symposium on Intelligent Signal Processing and Communication Systems*. IEEE, 2010, pp. 1–4.
- [26] M. K. I. Rahmani, F. Taranum, R. Nikhat, M. R. Farooqi, and M. A. Khan, "Automatic real-time medical mask detection using deep learning to fight covid-19," *COMPUTER SYSTEMS SCIENCE AND ENGINEERING*, vol. 42, no. 3, pp. 1181–1198, 2022.
- [27] N. S. Pun, S. K. Sonbhadra, S. Agarwal, and G. Rai, "Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques," *arXiv preprint arXiv:2005.01385*, 2020.
- [28] D. Yang, E. Yurtsever, V. Renganathan, K. A. Redmill, and Ü. Özgüner, "A vision-based social distancing and critical density detection system for covid-19," *Sensors*, vol. 21, no. 13, p. 4608, 2021.
- [29] S. V. Militante and N. V. Dionisio, "Deep learning implementation of facemask and physical distancing detection with alarm systems," in *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*. IEEE, 2020, pp. 1–5.
- [30] S. Yadav, "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*, vol. 8, no. 7, pp. 1368–1375, 2020.
- [31] W. Naudé, "Artificial intelligence vs covid-19: limitations, constraints and pitfalls," *AI & society*, vol. 35, no. 3, pp. 761–765, 2020.
- [32] S. Zhang, C. Chi, Z. Lei, and S. Z. Li, "Refineface: Refinement neural network for high performance face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 4008–4020, 2020.
- [33] C. Jagadeeswari and M. U. Theja, "Performance evaluation of intelligent face mask detection system with various deep learning classifiers," *International Journal of Advanced Science and Technology*, vol. 29, no. 11s, pp. 3074–3082, 2020.
- [34] P. Khandelwal, A. Khandelwal, S. Agarwal, D. Thomas, N. Xavier, and A. Raghuraman, "Using computer vision to enhance safety of workforce in manufacturing in a post covid world," *arXiv preprint arXiv:2005.05287*, 2020.
- [35] E. M. Onyema, P. K. Shukla, S. Dalal, M. N. Mathur, M. Zakariah, and B. Tiwari, "Enhancement of patient facial recognition through deep learning algorithm: Convnet," *Journal of Healthcare Engineering*, vol. 2021, 2021.
- [36] S. Sethi, M. Kathuria, and T. Kaushik, "Face mask detection using deep learning: An approach to reduce risk of coronavirus spread," *Journal of biomedical informatics*, vol. 120, p. 103848, 2021.
- [37] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. 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, p. 102600, 2021.
- [38] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting masked faces in the wild

- with lle-cnns,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2682–2690.
- [39] “Mask dataset.” [Online]. Available: <https://makeml.app/datasets/mask>
- [40] W. Intelligence, “Face mask detection dataset,” 2021.
- [41] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei *et al.*, “Masked face recognition dataset and application,” *arXiv preprint arXiv:2003.09093*, 2020.
- [42] M. Rasib, M. A. Butt, F. Riaz, A. Sulaiman, and M. Akram, “Pixel level segmentation based drivable road region detection and steering angle estimation method for autonomous driving on unstructured roads,” *IEEE Access*, vol. 9, pp. 167 855–167 867, 2021.
- [43] M. A. Butt, F. Riaz, Y. Mehmood, and S. Akram, “Reec-agent: human driver cognition and emotions-inspired rear-end collision avoidance method for autonomous vehicles,” *Simulation*, vol. 97, no. 9, pp. 601–617, 2021.
- [44] A. Mehrpooya, F. Saberi-Movahed, N. Azizizadeh, M. Rezaei-Ravari, F. Saberi-Movahed, M. Eftekhari, and I. Tavassoly, “High dimensionality reduction by matrix factorization for systems pharmacology,” *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab410, 2022.
- [45] F. Saberi-Movahed, M. Mohammadifard, A. Mehrpooya, M. Rezaei-Ravari, K. Berahmand, M. Rostami, S. Karami, M. Najafzadeh, D. Hajinezhad, M. Jamshidi *et al.*, “Decoding clinical biomarker space of covid-19: Exploring matrix factorization-based feature selection methods,” *Computers in biology and medicine*, vol. 146, p. 105426, 2022.
- [46] K. Berahmand, E. Nasiri, Y. Li *et al.*, “Spectral clustering on protein-protein interaction networks via constructing affinity matrix using attributed graph embedding,” *Computers in Biology and Medicine*, vol. 138, p. 104933, 2021.
- [47] M. Rostami, S. Forouzandeh, K. Berahmand, M. Soltani, M. Shahsavari, and M. Oussalah, “Gene selection for microarray data classification via multi-objective graph theoretic-based method,” *Artificial Intelligence in Medicine*, vol. 123, p. 102228, 2022.
- [48] V. Edmond, E. Onyema, A. Osijirin, and O. Oka, “Application of innovative technologies in computer science education during covid-19 school closure in enugu,” vol. 12, pp. 5129–5139, 01 2022.
- [49] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday *et al.*, “The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study,” *The Lancet Public Health*, vol. 5, no. 5, pp. e261–e270, 2020.
- [50] P. Levchev, M. N. Krishnan, C. Yu, J. Menke, and A. Zakhor, “Simultaneous fingerprinting and mapping for multimodal image and wifi indoor positioning,” in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2014, pp. 442–450.
- [51] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, “Image based localization in indoor environments,” in *2013 Fourth International Conference on Computing for Geospatial Research and Application*. IEEE, 2013, pp. 70–75.
- [52] Y. Niu, Z. Xu, E. Xu, G. Li, Y. Huo, and W. Sun, “Monocular pedestrian 3d localization for social distance monitoring,” *Sensors*, vol. 21, no. 17, p. 5908, 2021.
- [53] R. Magoo, H. Singh, N. Jindal, N. Hooda, and P. S. Rana, “Deep learning-based bird eye view social distancing monitoring using surveillance video for curbing the covid-19 spread,” *Neural Computing and Applications*, vol. 33, no. 22, pp. 15 807–15 814, 2021.
- [54] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” *arXiv preprint arXiv:1804.02767*, 2018.
- [55] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, “Single-shot refinement neural network for object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4203–4212.
- [56] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [57] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *Advances in neural information processing systems*, vol. 28, 2015.
- [58] X. Mai, H. Zhang, X. Jia, and M. Q.-H. Meng, “Faster r-cnn with classifier fusion for automatic detection of small fruits,” *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 3, pp. 1555–1569, 2020.
- [59] C. Lee, H. J. Kim, and K. W. Oh, “Comparison of faster r-cnn models for object detection,” in *2016 16th international conference on control, automation and systems (iccas)*. IEEE, 2016, pp. 107–110.
- [60] B. Liu, W. Zhao, and Q. Sun, “Study of object detection based on faster r-cnn,” in *2017 Chinese Automation Congress (CAC)*. IEEE, 2017, pp. 6233–6236.
- [61] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [62] J. Redmon and A. Farhadi, “Yolo9000: better, faster, stronger,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.
- [63] I. Ghodgaonkar, S. Chakraborty, V. Banna, S. Allcroft, M. Metwally, F. Bordwell, K. Kimura, X. Zhao, A. Goel, C. Tung *et al.*, “Analyzing worldwide social distancing through large-scale computer vision,” *arXiv preprint arXiv:2008.12363*, 2020.
- [64] I. Suwarno, A. Ma’arif, N. M. Raharja, T. K. Hariadi, and M. A. Shomad, “Using a combination of pid control and kalman filter to design of iot-based telepresence self-balancing robots during covid-19 pandemic,” 2020.
- [65] Z. Zhang, Y. Han, Y. Zhou, and M. Dai, “A novel absolute localization estimation of a target with monocular vision,” *Optik*, vol. 124, no. 12, pp. 1218–1223, 2013.