Contactless Surveillance for Preventing Wind-Borne Disease using Deep Learning Approach

Md Mania Ahmed Joy[1]
Department of CSE
East West University,
Dhaka, Bangladesh

Samira Hasan[4]
Department of CSE
East West University,
Dhaka, Bangladesh

Prof.Dr. Omar Farrok[7]
Department of EEE
Ahsanullah University of
Science and Technology, Dhaka,Bangladesh

Israt Jaben Bushra[2]
Department of CSE
East West University,
Dhaka, Bangladesh

Samia Binta Hassan[5]
Department of CSE
East West University,
Dhaka, Bangladesh

Mohammad Rifat Ahmmad Rashid[8]
Department of CSE
East West University,
Dhaka, Bangladesh

Razoana Ayshee[3]
Department of CSE
East West University,
Dhaka, Bangladesh

Md. Sawkat Ali[6]
Department of CSE
East West University,
Dhaka, Bangladesh

Maheen Islam[9]
Department of CSE
East West University,
Dhaka, Bangladesh

***Abstract*—Covid-19 has been marked as a pandemic world-wide caused by the SARS-CoV-2 virus. Different studies are being conducted with a view to preventing and lessening the infections caused by covid-19. In future, many other wind-borne diseases may also appear and even emerge as "pandemic". To prevent this, various measures should be an integral part of our daily life such as wearing face masks. It is tough to manually ensure individuals safety. The goal of this paper is to automate the process of contactless surveillance so that substantial prevention can be ensured against all kinds of wind-borne diseases. For automating the process, real time analysis and object detection is a must for which deep learning is the most efficient approach. In this paper, a deep learning model is used to check if a person takes any preventive measures. In our experimental analysis, we considered real time face mask detection as a preventive measure. We proposed a new face mask detection dataset. The accuracy of detecting a face mask along with the identity of a person achieved accuracy of 99.5%. The proposed model decreases time consumption as no human intervention is needed to check an individual person. This model helps to decrease infection risk by using a contactless automation system.**

***Keywords*—Computer vision; convolution neural network; COVID-19; deep learning; face mask detection; identity detection; object detection**

I. Introduction

Whenever the world faced any pandemic, history witnessed a pessimistic effect on economics, health, and national security both socially and globally [1]. Before the COVID-19, H1N1 or influenza was marked as a pandemic in the year 2009 [2] [3]. The COVID stands for "CoronaVirus Disease", it is referred to as "2019 novel coronavirus" or "2019-nCoV" as it was started in 2019 [4]. Form the beginning of this pandemic, around 228 countries have been affected by the COVID-19 to date [4]. The rapid growth of the COVID-19 infected cases has put the national healthcare capacity, modern ICU diagnostic methods, and public healthcare infrastructure to a test.

For instance, Bangladesh has less than 7,000 spaces in isolation units and therefore only 1622 health workers, including only 595 physicians to treat COVID-19 patients, whereas it has a population of about 165 million people [5]. On the other hand, the United States (US), which like Europe, spends about double as much per person on healthcare insurance as those other high-income counties, has drawn special criticism for its "maximum-possible-test-per-day" policy and use of its 96,596 ICU beds [5]. Overall, it became a chaotic situation which completely caught the whole world off guard. To prevent this disease, many effective vaccines have been invented [6] but these vaccines are not sufficient to give full protection to the people to prevent the COVID-19. So, people had to follow some non-therapeutic prohibitive measures like maintaining social distance, travel bans, remote office activities, country lockdown, wearing masks, etc. despite being vaccinated

Since 2019 almost all the nations of the world are struggling to get out of this misery. This kind of prohibition rules put a financial crisis as lockdown remains on all kinds of institutions [7]. In [8], according to the "Socio-Economic Aspects" about 60.5% of the respondents agreed that most of the low incoming people lost their jobs and around 54.8% of people also shifted to different places by leaving the city for livelihood. To ease the survival challenge of the people who lived from hand to mouth, the lockdown had to be lifted. Then there comes another challenge of ensuring that every individual wear masks in their working places. To ensure security in institutions or industries, many automation systems exist that can identify the registered person

One of them is a biometric system that works with measurements and by analyzing someone's unique features. Physical unique features like face, irises, veins, fingerprints, and behavioral characteristics such as voice, typing rhythm, or handwriting, is used for identification and authentication [9]. These systems are not contactless, and these require removing

the mask which can be one of the fatal reasons for any windborne disease [10]. For this, there is an increasing necessity to build a contactless surveillance system to detect facemask and person identification. So, there are a few systems that can perform both the operations. For designing such contact less automated systems, deep learning-based approaches are integrated.

The researchers tried to find out the best possible object detection model with the highest accuracy. Some of the models can even detect objects from video streams. In some cases, general convulation neural network (CNN) models are used; it cannot identify a specific object if a lot of background objects exist in a frame. Boundary box techniques and means Intersection over Union (IoU) is mostly used for object detection.

The focus of our work is to develop an automated contactless preventive measurement system using a deep learning approach. More specifically, our system will identify if a person is wearing a mask or not, and at the same time detect the person's identity. For object detection task using face mask dataset, there are many models like CNN [12], NASNet [31], R-CNN [11], FAREC [27], CSPResNeXt-50 [23], etc., and the most recent model is YOLOv5 [19] which shows moderate inference speed and accuracy. The followings are the significant contributions of this paper:

- A new dataset is presented that consists of 177 pictures with 6 labels in total. There will be 97 photos of random people wearing masks and without masks, and 80 images of four persons, 20 images of each person wearing a mask. For each image, it contains a text file with information about the boundary boxes.
- A comparative analysis is performed using the new dataset using the YOLOv5 model. The YOLOv5 model is used for training the dataset. After 100 epochs with batch size of 8, the accuracy of 99.5% is obtained.
- This approach helps to identify any person along with detecting if he/she is wearing a mask or not in a real-life scenario. It works best for both images and video streams.
- A dedicated application is developed for contactless preventive measures using our proposed approach.

The rest of this paper is arranged in the following manner: related work that provides an overview of the relevant papers in condensed form, proposed work provides an initial overall view of the work process, materials and methods section provides an explanation of the preparation, preprocessing of the dataset, methodology, and training, result section provides the experimental analysis and conclusion provides the conclusive observations and future works.

II. Related Work

There are usually two types of object detection methods based on the deep learning models: one-stage algorithms and two-stage algorithms. Two-stage methods such as R-CNN [11], Fast R-CNN [12], and Faster R-CNN [13] create bounding boxes of objects and then classify these objects. But these methods are comparatively slower and hard to implement in real-time. So, for the efficient results, some one-stage algorithms have been developed such as You Only Look Once (YOLO) [14], RetinaNet [15], Single Shot Detector (SSD) [16] which are based on Anchor and ATSS (Adaptive Training Sample Selection) [17].

In the article [18], the authors used YOLOv5 to detect face masks, which is the most powerful object detection model at present. Their experiment has achieved a success rate of about 97.9% by using 7,959 images for training and testing. There are different forms of the YOLOv5 model. The YOLOv5x model is much heavier than the most used YOLOv5s [19].

Face Mask Recognition System has also been developed by using different versions of the YOLO algorithm [18] [20] [21]. In the article [20], the authors used improved YOLOv4 which is based on CSPDarkNet53. They used adaptive image scaling algorithms and PANet structure to get more information on the feature layer. The authors have shown a comparison between different models like YOLOv3, YOLOv4, SSD, and Faster R-CNN and have got the prominent outcome using their proposed model. Yet, there are some problems with feature extraction and false detection cases as well as low light performance has not been considered. In another the article [21], the authors have used YOLOv2 with ResNet-50 to detect medical face masks. They have used mean Intersection over Union (IoU) to estimate the best number of anchor boxes and have achieved 81% precision as a detector. For their model, they have used 1415 images as datasets. They proposed a detector model using YOLOv2 with Resnet-50 for extracting the feature and detecting different phases. Their model works better to detect medical face masks. They have also planned to use video-based deep learning models in the future. The article [22] presents the use of the YOLOv4 model which is the fourth generation of the YOLO model proposed by the authors of [23] in 2020. It is relatively faster and has better accuracy, and it is incredibly quick, easy to train, robust and also reliable. It produces promising results even for detecting small objects compared to the previous versions. It recognizes items from three types of input image/frame: unmasked faces, masked faces, and humans.

Face mask detection using YOLOv4 was introduced in [24]. As the YOLOv4 is built on the Darknet Framework and features an NVIDIA Graphics Processing Unit (GPU) with 16 gigabytes of RAM and a 2.30GHz Intel i5 CPU, its framework has created two models, one with three classes of the face veil, face concealed face, and face cover, and the other with only two classes of face cover and face cover. These models perform better for the video which creates a picture (or frame) with localized faces and a bounding box indicating whether the face is masked or not.

Multi-Stage CNN Architecture has been proposed in [25] for face mask detection. A pre-trained model that has been trained on a large dataset in order to achieve rapid generalization and reliable detection using RetinaFace [26], was selected as a single stage-1 model. Its' performance was evaluated to two other prominent face detection algorithms models, Dlib, and MTCNN [27] [28]. In stage 2, a CNN classifier was trained using three picture classification models: MobileNetV2 [29], DenseNet121 [30], and NASNet [31]. Recently, face identification along with face mask recognition using CNN approach was presented in [32]. To process a sequence of

TABLE I. COMPARISON OF RELATED WORK

Author/ref	Year	Approach/Model	Detection	Model Accuracy
K. Zhang [28]	2016	CNN	Face	95.4%
S. Sharma [27]	2016	FAREC	Face	96%
B. Zoph [31]	2018	NASNet	Scalable Image	96.2%
Kaihan [34]	2019	R-CNN (G-mask)	Face	95.97%
M. Loey [21]	2020	YOLO-v2 with ResNet-50	Face Mask	81%
A. Bochkovskiy [23]	2020	CSPResNeXt-50	Object	94.8%
G.Yang [18]	2020	YOLOV5	Face Mask	97.90%
K. Bhambani [22]	2021	YOLOv4	Real-time Face Mask	95%
S. Degadwala [24]	2021	Yolo-v4	Face Mask	96%
A. Chavda [25]	2021	Multi-Stage CNN	Face Mask	99.49%
M.S. Mazli Shahar [32]	2021	CNN, Procrustes Analysis	Face Mask, Face Identity	97.14%

pictures, the system uses the OpenCV computer vision library as well as a deep learning method for face mask identification. The receiver operating characteristic (ROC) and the calculated mean accuracy are used to estimate the system's performance (ACC).

For the face detection process, the training has been done by CNN model to recognize the owner's identity. The similarity identification approach utilizing the 68-landmark plotting and the face identification method were used for the process. To capture each frame of the video stream, a 50-neuron fully connected dense layer is flattened and fed from the output. It provides the probability for a person wearing a mask or not into 2-neuron layers in the convolutional neural network training. The CNN with Keras and Tensorflow is then implemented to train the face mask classifier and for face mask detection, the first 20 epochs, a model checkpoint trains and validates where using facial 68-landmarks is a technique that locates a human face in a picture and delivers values on the face Region of Interest (ROI) as dots. All the work was done with OpenCV, Keras and TensorFlow. Standard face recognition performance metrics were used for person identification at the time of using a face mask [33].

As it has been discussed earlier, some of the works as like [18], [20], [21] have used YOLO algorithms that function using IoU. The authors of [21] using YOLOv2 with ResNet-50 model achieved nearly 81% accuracy, in the paper [22] achieved nearly 95% accuracy. The authors of [18] achieved nearly 97.90% accuracy using YOLOv5 but all of these works only determined whether a person was wearing a mask or not. In [25], the authors have achieved more than 99% accuracy using Multi-Stage CNN. In the case of face detection, the authors of [27] and [28] have achieved up to 96% percent accuracy but they did not train their model for identifying people when they are using a face mask.

Table I presents a comparative analysis on all the related works that have been done. Although different types of object detection have been done so far, there is not a single work that matches with the objective of this paper.

III. EXPERIMENT

This section describes how the proposed methodology has been designed. At first, a dataset of images with specific formats has been generated and then classified into labels and classes. Furthermore, preprocessing on images has been done according to the need of the research model. Then the YOLO algorithm and how YOLOv5 incorporates this work is elaborated. In conclusion, all these steps lead to training that is the final step to get the desired outcome. Fig. 1 illustrates the proposed framework for using face mask dataset for predicting contactless surveillance having a machine learning backbone using deep learning.

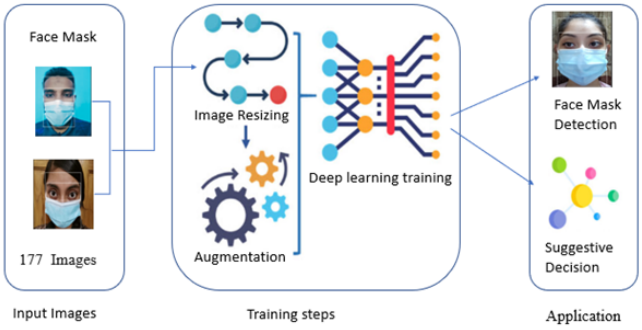

Fig. 1. Example of the Proposed Framework for Face Mask Detection and Contactless Surveillance System.

The YOLOv5 model is used to detect faces with masks and identification of that person who is wearing a mask. To train the model, at first, the images are collected. After preprocessing the images, every picture is labeled with the necessary class name. Then a dataset.YAML file is created. Then after training the model, it has been implemented on hardware. Finally, the performance has been evaluated.

A. Dataset

The images are in jpg, jpeg, png format, and the labels are in txt format. The images are of people of different ages, gender, wearing face masks and without masks. Fig. 2 shows a sample dataset with masked faces.

Fig. 2. Sample Dataset Consists of Images for a few Type Labels.

Table II shows the labels depending on the types of images and presents the number of images. A total of 177 images are used for this work and among them 97 images are of random people wearing a mask and without a mask, and among 177 images, 80 of them are of 4 people, 20 images of each person wearing a mask along with their identification. So, there are a total of 6 labels in this dataset. They are mask, no_mask, bushra, ayshee, joy, samira.

TABLE II. DETAILS OF THE IMAGES AND THEIR LABELS

Types of Images	Labels	Number of Images
People wearing masks (including random people)	Mask	129
Random people wearing no mask (including random people)	No_mask	48
Person 1 wearing a mask with an identification	Bushra	20
Person 2 wearing a mask with an identification	Ayshee	20
Person 3 wearing a mask with an identification	Joy	20
Person 4 wearing a mask with an identification	Samira	20

When a dataset is created by collecting different pictures, all pictures cannot be of the same size. A model trains faster with smaller-sized images. If an image is three times as large, training the model will take six times as long, which will mount up over time. As all pictures vary in size, so if the picture is too small, it may appear unclear and if it is too big then the training time gets longer. That is why, all pictures have been resized. From the pictures, specific labeling is also needed to make classes which is the necessary step for training the model. Therefore, images are preprocessed before using the pictures as datasets.

At first, all images are resized to 1000×1000 pixels. Then all images are labeled by a software called "makesense.ai". In this software, all the labels are created first. Then on every image, the boundary box is drawn like Fig. 3 and then the labels are selected for the boundary boxes. A boundary box is drawn for each label and there are six labels used depending on the class. The class or label values are shown in Table III.

TABLE III. CLASS (LABELS) AND VALUES

Class (Labels)	Values
Bushra	0
Joy	1
Ayshee	2
Samira	3
Mask	4
no_mask	5

We organized boundary boxes in the following format: {Bx,By} = Coordinates of the boundary box's center

{Bh,Bw} = Width, height of the boundary box as a percent of the cell's width or height

Fig. 3. An Example of the Values from a Sample Image.

Fig. 3 shows the dot is the center of the boundary box. In Table IV, the boundary box values are shown for 1 and 4 number classes.

TABLE IV. CLASS (LABELS) AND SAMPLE BOUNDING BOX VALUES

Class	Bx	By	Bh	Bw
1	0.439	0.387	0.559	0.157
4	0.435	0.509	0.575	0.677

B. Deep Learning Framework for Training and Inference

To begin with, a dataset consisting of 177 images has been prepared where a total of 6 labels named mask, no_mask, bushra, ayshee, joy, samira are used. 97 images of random people wearing masks and without masks, and 80 images of 4 people, 20 images of each person wearing mask along with their identification are used to detect wearing mask and identify people.

After training the model, the images have split into training and validation labels, and validation prediction with confidence for the labels has been done. For training, Mosaic Data Loader [36] was used. Every training batch contains 8 mosaics, and each mosaic is a combination of four images.

The YOLOv5 [36] model works on a custom dataset that includes three cases i.e., 1st one to detect the identification of the person wearing a mask, 2nd one to detect face with a mask, and 3rd one to detect face without mask. The model is trained for predicting the label of classes for each input image.

YOLO stands for "You Only Look Once". Researcher Joseph Redmon and colleagues introduced YOLO in 2015 [35]. The General YOLO model works in three steps. In the

first step, it divides a picture into many cells with an m × m grid. Then in the second step, "s" bounding boxes are predicted in each cell. YOLO returns to the bounding boxes that exceed the minimum confidence threshold in the third step. The YOLO algorithm has various versions. In this paper, the YOLOv5 model is used. It is written in Python programming language's using Pytorch framework. This framework is user-friendly and a community driven github repository [36].

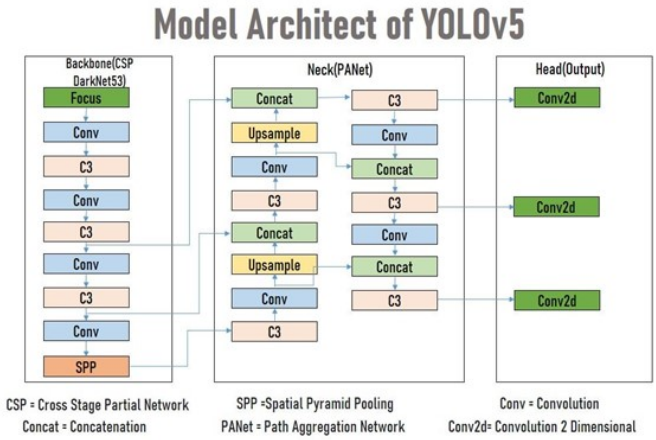

Fig. 4. YOLOv5 Model Architecture.

YOLOV5 consists of three parts: Backbone CSP darknet 53, Neck PANet, and Head Yolo V5 output layer depicted in Fig. 4. The backbone extracts feature from the inputs, Neck works to fusion features from backbone and Head generates the final results.

First part of the YOLOV5 is the Backbone CSP darknet or cross stage partial network. It focuses on the structure (Modified from Yolo V3) as well as on the CSP darknet 53 of Pytorch. CSP network has advantages like it solves repeated gradient problems, reduces model size, cuda memory layers, and increases forward, backward propagation. The CONV denotes convolution layers and C3 is composed of three convolution layers that are cascaded by various bottlenecks to find out the best fit in the network. And spatial pyramid pooling (Spp-Net) (Fig. 5) is a pooling layer that is used to extract features from the feature maps passed by previous layers. SPP works to generate fixed-size output for any type of input size. It is also used to extract necessary features by pooling that it's a multi-scale version of itself. Feature maps of inputs are duplicated to "n" numbers of versions. Using kernels of various sizes, each version is conducted by max pooling. This is how the SPP block extracts "n" number of various types of required and important features simultaneously [36].

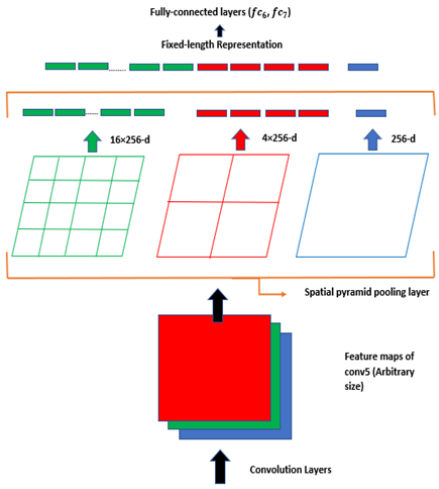

Fig. 5. Representation of SPP-NET.

Second part of this model is the Neck path aggregation network (PA-Net) [37]. It uses a modified version of feature pyramid network (FPN) structure that utilizes the information passed by the backbone. It employs bottom up (up sampling) augmentation path along with top-down It is used in FPN to directly connect with the fine-grained or learned features from low-level layers to top-level layers from backbone. For boosting information flow, prevention of missing information, and duplicate prediction, this concatenation operation is applied. It improves accuracy of the location of the object.

Last part is of the YOLOv5 model is the Head Yolo V5 output layer that uses information passed from the neck layer. The head of Yolo V5 known as Yolo layer generates three different sizes which are 18 × 18, 36 × 36, 72 × 72 of feature maps using convolutional 2-dimensional networks to achieve multi scale prediction. It enables the model to handle small, medium, and big objects and the output is a boundary box of the predicted object along with accuracy and class of the object.

C. Performance Measure

In this approach, Generalized Intersection Over Union (GIoU) loss has been used to determine the loss of the bounding box. The formula and algorithm for GIoU loss [38] is shown in Algorithm 1.

Algorithm 1 Generalized Intersection Over Union(GIoU)

Input: *Two arbitrary convex shapes*; $C,\ D \subseteq S \in R^n$
Output: *GloU*
1: *For* C *and* D, *find the smallest enclosing convex object* E, *where* $E \subseteq S \in R^n$
2: $loU = \frac{C \cap D}{C \cup D}$
3: $GloU = loU - \frac{|E \backslash (C \cup D)|}{|E|}$

Here in algorithm 1, to compare similarity IoU (Intersection over Union) between two arbitrary convex shapes $C, D \subseteq S \in R^n$ is determined by, IoU = $\frac{|C \cap D|}{|C \cup D|}$. IoU has a big defect that if $|C \cap D| = 0$ and then IoU (C, D) =

0. This proves that IoU does not determine that both shapes are in proximity or not from each other. That is why GIoU (Generalized Intersection over Union) is used.

In GIoU, for both arbitrary shapes (volumes), $D \subseteq S \in R^n$, firstly the smallest arbitrary convex shapes are found $E \subseteq S \in R^n$ enclosing both C and D. E come from a similar type of geometric shape while comparing two geometric shapes. When both arbitrary ellipsoids are compared, E is smaller. They are encircled by ellipsoids. Then a ratio is computed between the area (volume) which is occupied by E subtracted from the area (volume) occupied by C and D and then divided by E's entire area (volume). This is a normalized metric that is focused on the vacant space area (volume) between C and D. Finally, GIoU has been determined. This ratio is subtracted from the value of IoU. Algorithm 1 summarizes the computations of GIoU.

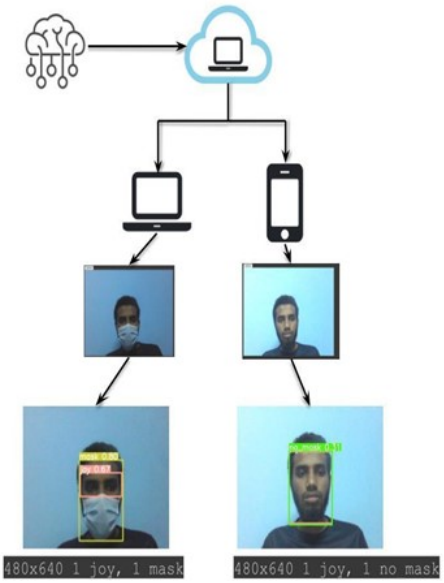

Fig. 6. Implementation Diagram.

IV. Results and Evaluation

A. Experimental Results

The model has been implemented practically in an enclosed environment to see how it works in a real-life situation. Hardware implementation is needed to analyze the efficiency of the model. Implementation of this system provides a contactless process which ensures safety. It mitigates the spread of respiratory transmittable diseases. Implementation of this system will compel people to use masks because the system will mark people without face masks as any organization may use this system at their main entrance.

The model's dataset consists of 177 images and corresponding labels. For labeling, a total of six classes are used which are mask, no_mask, bushra, ayshee, joy and samira. A custom yaml file is created that contains the path of the dataset, number of classes and class names that are used. To train the dataset and yaml file, the YOLOv5x model is used. After training, two weights best.pt and last.pt have been obtained. YOLOv5x is a large model for processing and the weights were also large, so it needs a powerful processor. Hence, Google Colab, a cloud computing system is used to run the model. It has 12GB free graphics support and between GPU or CPU any of these can be used. Real time analysis has been performed by using the 12GB GPU of "Google Colab". For experimental purposes, the mobile and laptop devices have been used to run the model in "Google Colab". The configuration for mobile is f/2.45 aperture, SONY IMX471 sensor, and for laptop CMOS sensor technology, RGB HD Camera configuration was used. This experiment yields to a successful result to detect a person's identity wearing a face mask. Fig. 6 shows the implementation diagram of the model.

The proposed model has used GPU runtime for training the YOLOv5x model using Pytorch libraries and Stochastic Gradient Descent (SGD) optimizer. Batch size of 8 is used for each epoch and there are a total of 50 epochs for training the model, thus the experimental configuration has been obtained. The optimizer that is implemented for training the model is SGD.

SGD is used because the training time is 4 seconds which is faster than any other optimizers. SGD can be implemented on the convex problem and simple linear problems and the learning rate is determined by, $\eta/(1 + \lambda_0 t)$. Where, λ_0 = regularization constant and η_0 is achieved by conducting initial experiments on any subsample of a dataset [39]. When the epoch increases with time, every epoch's precision (P), Recall (R), and Mean Average Precision (mAP) increases, and the training and validation of the model is also improved.

TABLE V. Training and Validation Result based on SGD

Epoch	Precision (P)	Recall (R)	mAP@0.5	mAP@[0.5:0.95]
5	0.197	0.67	0.304	0.117
10	0.711	0.718	0.67	0.386
20	0.829	0.904	0.95	0.572
30	0.816	0.963	0.949	0.528
48	0.987	0.995	0.995	0.637

In Table V, it is visible that the epoch has a proportional relation with the precision and recall, that is when the epoch is increasing the precision and recall also gets improved. For detecting the identity of a person wearing a mask, "GIoU Loss" has been used. The test results of the model are shown in Fig. 7.

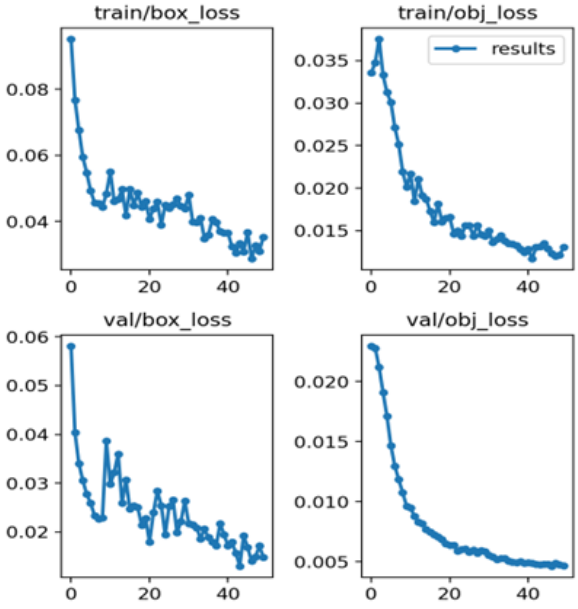

Fig. 7. The GIoU Loss Curve of Training where X Axis Represents Epoch.

After completion of the model training, the trained weights have been used for testing the model and then the performance is evaluated. The test results have classified into three types of categories.

True Positive (TP) indicates that in the test set the categories are equal to the results of the test. False Positive (FP) indicates that the sample number in the category of detected object is inconsistent. False Negative (FN) means that the resultant pattern has been detected as the inverse result or in the category of undetected. Precision refers to the ratio of TP to (TP+FP) where (TP+FP) is the number of all positive cases. Precision illustrates the sample proportions of the real positive cases from the samples detected by the model. The equation-1 shows the equation for precision.

$$Precision = \frac{\text{TP}}{\text{TP+FP}}; \tag{1}$$

Recall refers to the ratio of TP to (TP+FN) where (TP+FN) is the number of all positive test cases. The model's efficiency can be measured by recall which detects real positive cases of the test set out of the all-actual positive values. The equation that follows equation-2 shows the equation for recall.

$$Recall = \frac{\text{TP}}{\text{TP+FN}}; \tag{2}$$

Fig. 8 shows the graphs based on the precision and recall of the applied model.

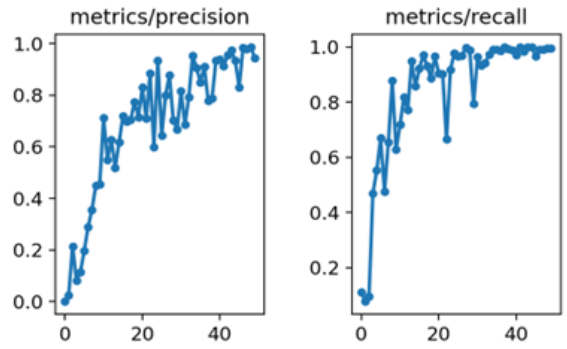

Fig. 8. Precision and Recall of the Model where X Axis Represents Epoch.

Then, to evaluate the accuracy of the model Mean Average Precision (mAP) is used. The following equation-3 shows hoe Map is determined:

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{\text{N}}; \tag{3}$$

The performance evaluation from the training is shown in Table VI.

TABLE VI. PERFORMANCE EVALUATION

Class	Precision	Recall	mAP for IoU 0.5	mAP for IoU [0.5:0.95]
All	0.944	0.994	0.995	0.615
Bushra	0.719	1	0.995	0.331
Joy	0.979	1	0.995	0.639
Ayshee	1	1	0.995	0.48
Samira	0.975	1	0.995	0.733
mask	1	0.966	0.995	0.772
no_mask	0.99	1	0.995	0.736

The graphs in Fig. 9 presents the performance evaluation of the model.

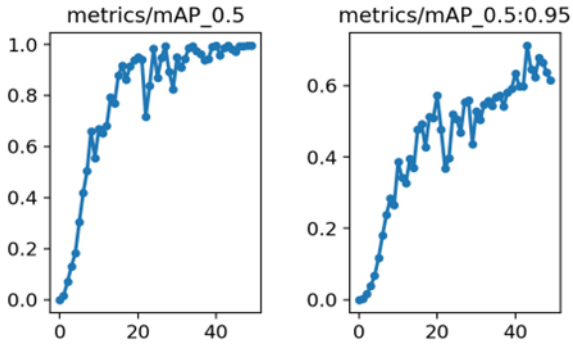

Fig. 9. mAP@0.5 and mAP@[0.5: 0.95] of the Model where X Axis Represents Epoch.

V. LIMITATIONS AND FUTURE WORK

This research includes the identification of individuals with face masks using the deep learning approach. The YOLOv5 model can detect whether any individual is maintaining the health protocols such as wearing a face mask or not. This

model is proven beneficial for preventing any airborne disease. Though it achieved the goal of the research, there are some limitations of the research. The dataset used for the model's implementation, contained only 177 images. Such small dataset does not ensure a significant accuracy if this research is applied for real-life events. Due to time constraints and inappropriate circumstances, it was not possible to conduct the research on a larger scale. A better accuracy could be expected from the model if a dataset containing more variety was used. Another limitation of the research is that the model is used only in detecting face masks, hence the user case of this model was confined into a specific object.

This research can be extended by overcoming the limitations and adding some insights to it. This model can be used while designing any other contactless surveillance systems. This model worked efficiently in face mask detection, but other object detection can also be included by implementing this model. Such systems significantly help to build any other contactless surveillance system that can highly be effective on controlling any contagious diseases as this ensures the social distancing. Another addition might be considered in this research that is, collecting enormous real-time data and implementing the model into that. Thus, ensuring the social distancing and detecting any kind of object with this model can be used to build further contactless system which can tend to prevent wind borne diseases. Enabling the unique biometric identification with the object detection system may prove as an enhanced version of this research.Such model can be utilized not only as contactless surveillance system, but also can keep the entrance record of any organization.

VI. Conclusion

This paper depicted the application of contact-less surveillance systems using a deep learning approach. For experimental analysis, we use YOLOv5 for object detection over a new dataset as mask detection. The ability of the YOLOv5 model to perform an efficient detection task was investigated in this research. The results demonstrate that YOLOv5 can achieve proficiency in detecting and spotting an item in a very short amount of time. In the case of surveillance of any organization at their entrance, the system not only identifies the person but also checks if he/she is wearing a mask or not. A masked face and a non-masked face dataset with no biased pictures have been developed and for the purpose of identification, the eye characteristics was examined. This system is unique in terms of operating in three different classes at the same time: a person with mask, without a mask, and the identification of a person. As the system identifies the person, it will be quite easier to detect if this specific person works in the organization or not. Using the system, if any individuals without masks are identified, they will be alerted and made conscious about wearing masks which is a preventive way to decrease COVID 19. The system also compels people to abide by the protocols of preventing this pandemic as well as helps to maintain hygiene for any wind-borne diseases. If the number of infected people decreases, the economic loss will also decrease. Manual identification of each people and checking if they are wearing masks or not is a difficult and time-consuming process. This process has been made easier and time-saving by this system. As it is a contact-less system, so social distancing can be highly maintained. This is a very useful technique for implementing in open spaces as well. An accuracy of 99.5% is achieved from this research. Some other models also have been implemented for this task but neither the accuracy was so high, nor the model could identify the features correctly in real-life scenarios whereas our chosen model successfully did the detections. Moreover, this trained model can perfectly identify features from the video streams along with the still images. This work will help to minimize the risk of air-borne viruses' spread during the time of the pandemic. Even if the COVID-19 pandemic fades away and life gets back to normal, this system can be quickly deployed to optimize the appropriate use of face masks in the workplaces so that a safe and secure work environment can be built.

References

[1] W. Qiu, S. Rutherford, A. Mao and C. Chu, "The Pandemic and its Impacts", Health, Culture and Society, 9, pp. 1-11. 2017. DOI: 10.5195/hcs.2017.221

[2] J. Watkins, "Preventing a covid-19 pandemic," BMJ, vol. no. 368, pp. 1–2, February 2020, doi: 10.1136/bmj.m810.

[3] D. M. Morens, G. K. Folkers, and A. S. Fauci, "What is a pandemic?," The Journal of infectious diseases, vol. 200, no. 7, pp. 1018–1021, 2009, doi: 10.1086/644537.

[4] WHO, IFRC, and unicef, "Key Messages and Actions for Prevention and Control in Schools," Key Messag. Actions COVID-19 Prev. Control Sch., no. March, p. 13, 2020, [Online]. Available: https://www.who.int/docs/default-source/coronaviruse.

[5] Khan JR, Awan N, Islam MM, Muurlink O. Healthcare capacity, health expenditure, and civil society as predictors of COVID-19 case fatalities: a global analysis. Frontiers in public health. 2020 Jul 3;8:347.

[6] E. Ong, M. U. Wong, A. Huffman, and Y. He, "COVID-19 Coronavirus Vaccine Design Using Reverse Vaccinology and Machine Learning," Front. Immunol., vol. 11, p. 1581, 2020, doi: 10.3389/fimmu.2020.01581.

[7] S. Gunay, "COVID-19 Pandemic Versus Global Financial Crisis: Evidence from Currency Market," SSRN Electron. J., pp. 1–15, 2020, doi: 10.2139/ssrn.3584249.

[8] M. Bodrud-Doza, M. Shammi, L. Bahlman, A. R. M. T. Islam, and M. M. Rahman, "Psychosocial and Socio-Economic Crisis in Bangladesh Due to COVID-19 Pandemic: A Perception-Based Assessment," Front. Public Heal., vol. 8, no. June 2020, doi: 10.3389/fpubh.2020.00341.

[9] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki, "An Introduction to evaluating biometric systems," Computer (Long. Beach. Calif)., vol. 33, no. 2, pp. 56–63, 2000, doi: 10.1109/2.820040.

[10] M. G. Garner, G. D. Hess, and X. Yang, "An integrated modelling approach to assess the risk of wind-borne spread of foot-and-mouth disease virus from infected premises," Environ. Model. Assess., vol. 11, no. 3, pp. 195–207, 2006. doi: 10.1007/s10666-005-9023-5

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[12] R. Girshick, "Fast R-CNN," Proc. IEEE Int. Conf. Comput. Vis., vol. 2015 Inter, pp. 1440–1448, 2015, doi: 10.1109/ICCV.2015.169.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., , pp. 779–788, december 2016, doi: 10.1109/CVPR.2016.91.

[15] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," Proc. IEEE Int. Conf. Comput. Vis., pp. 2999–3007, October 2017, doi: 10.1109/ICCV.2017.324.

[16] W. Liu et al., "SSD: Single shot multibox detector," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes

Bioinformatics), vol. 9905 LNCS, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.

[17] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 9756–9765, 2020, doi: 10.1109/CVPR42600.2020.00978.

[18] G. Yang et al., "Face Mask Recognition System with YOLOV5 Based on Image Recognition," 2020 IEEE 6th Int. Conf. Comput. Commun. ICCC 2020, vol. 1, no. January 2020, pp. 1398–1404, 2020, doi: 10.1109/ICCC51575.2020.9345042.

[19] . Glenn Jocher, Alex Stoken, Jirka Borovec, NanoCode012, Christopher-STAN, Liu Changyu, Laughing, tkianai, Adam Hogan, lorenzomammana, yxNONG, AlexWang1900, Laurentiu Diaconu, Marc, wanghaoyang0106, ml5ah, Doug, Francisco Ingham, Frederik, … Prashant Rai. (2020). ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements (v3.1). Zenodo. https://doi.org/10.5281/zenodo.4154370

[20] J. Yu and W. Zhang, "Face mask wearing detection algorithm based on improved yolo-v4," Sensors, vol. 21, no. 9, 2021, doi: 10.3390/s21093263.

[21] 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," Sustain. Cities Soc., vol. 65, p. 102600, 2021, doi: 10.1016/j.scs.2020.102600.

[22] K. Bhambani, T. Jain, and K. A. Sultanpure, "Real-Time Face Mask and Social Distancing Violation Detection System using YOLO," Proc. B-HTC 2020 - 1st IEEE Bangalore Humanit. Technol. Conf., 2020, doi: 10.1109/B-HTC50970.2020.9297902.

[23] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020, [Online]. Available: http://arxiv.org/abs/2004.10934.

[24] S. Degadwala, D. Vyas, U. Chakraborty, A. R. Dider, and H. Biswas, "Yolo-v4 Deep Learning Model for Medical Face Mask Detection," Proc. - Int. Conf. Artif. Intell. Smart Syst. ICAIS 2021, pp. 209–213, 2021, doi: 10.1109/ICAIS50930.2021.9395857.

[25] A. Chavda, J. Dsouza, S. Badgujar, and A. Damani, "Multi-Stage CNN Architecture for Face Mask Detection," 2021 6th Int. Conf. Converg. Technol. I2CT 2021, pp. 1–8, 2021, doi: 10.1109/I2CT51068.2021.9418207.

[26] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-shot multi-level face localisation in the wild," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 5202–5211, 2020, doi: 10.1109/CVPR42600.2020.00525.

[27] S. Sharma, K. Shanmugasundaram, and S. K. Ramasamy, "FAREC - CNN based efficient face recognition technique using Dlib," Proc. 2016 Int. Conf. Adv. Commun. Control Comput. Technol. ICACCCT 2016, no. 978, pp. 192–195, 2017, doi: 10.1109/ICACCCT.2016.7831628.

[28] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499–1503, 2016, doi: 10.1109/LSP.2016.2603342.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, pp. 2261–2269, January 2017, doi: 10.1109/CVPR.2017.243.

[31] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 8697–8710, 2018, doi: 10.1109/CVPR.2018.00907.

[32] M. S. Mazli Shahar and L. Mazalan, "Face identity for face mask recognition system," ISCAIE 2021 - IEEE 11th Symp. Comput. Appl. Ind. Electron., pp. 42–47, 2021, doi: 10.1109/ISCAIE51753.2021.9431791.

[33] Z. Wang et al., "Masked Face Recognition Dataset and Application," pp. 1–3, 2020, [Online]. Available: http://arxiv.org/abs/2003.09093.

[34] K. Lin et al., "Face Detection and Segmentation Based on Improved Mask R-CNN," Discret. Dyn. Nat. Soc., vol. 2020, 2020, doi: 10.1155/2020/9242917.

[35] R. Shukla, A.K. Mahapatra and J.S.P. Peter, "Social distancing tracker using yolo v5," Turkish J. Physiother. Rehabil., vol. 32, no. 2, pp. 1785–1793, 2021.

[36] D. Thuan, "Evolution of Yolo Algorithm and Yolov5: the State-of-the-Art Object Detection Algorithm," p. 61, 2021. Available online: https://www.theseus.fi/handle/10024/452552 (accessed on 12 November 2021).

[37] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 8759–8768, 2018, doi: 10.1109/CVPR.2018.00913.

[38] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 658-666, doi: 10.1109/CVPR.2019.00075.

[39] Ketkar, N. (2017). Stochastic Gradient Descent. In: Deep Learning with Python. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-2766-4_8