Science in Information Technology Letters
Vol. 4., No. 1, May 2023, pp. 22-39
ISSN 2722-4139
http://pubs2.ascee.org/index.php/sitech

YOLOv3 and YOLOv5-based automated facial mask detection and recognition systems to prevent COVID-19 outbreaks

Md Asifuzzaman Jishan [a,1,*], Ananna Islam Bedushe [b,2], Md Ataullah Khan Rifat [a,3], Bijan Paul [c,3], Khan Raqib Mahmud [c,5]

[a] Faculty of Statistics, Technische Universität Dortmund, Germany
[b] Department of Computer Science, Hofstra University, United States of America
[c] Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, Bangladesh
[1] md-asifuzzaman.jishan@tu-dortmund.de; [2] abedushe1@pride.hofstra.edu; [3] khan.rifat@tu-dortmund.de; [4] bijan.paul@ulab.edu.bd; [5] raqib.mahmud@ulab.edu.bd
* Corresponding Author

ARTICLE INFO

Article history
Received April 09, 2023
Revised April 12, 2023
Accepted May 08, 2023

Keywords
Facial mask detection
Classification of image
YOLOv3
YOLOv5
COVID-19 pandemic

ABSTRACT

Object detection system in light of deep learning have been monstrously effective in complex item identification task images and have shown likely in an extensive variety of genuine applications counting the Coronavirus pandemic. Ensuring and enforcing the proper use of face masks is one of the main obstacles in containing and reducing the spread of the infection among the population. This paper aims to find out how the urban population of a megacity uses facial masks correctly. Using YOLOv3 and YOLOv5, we trained and validated a brand-new dataset to identify images as "with mask", "without mask", and "mask not in position". In the YOLOv3 we carried out three pre-trained models which are: YOLOv3, YOLOv3-tiny, and SPP-YOLOv3. In addition, we utilized five pre-trained models in the YOLOv5: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The dataset is included 6550 pictures with three classes. On mAP, the dataset achieved a commendable 95% performance accuracy. This research can be used to monitor the proper use of face masks in various public spaces through automated scanning.

1. Introduction

In 2019, the COVID-19 Coronavirus outbreak was discovered for the first time in Wuhan, China. Since then, it has spread across the globe and mutated into the fifth pandemic recorded since the 1918 influenza pandemic. Almost 200 million people were infected with the COVID-19 virus, resulting in over 4.6 million deaths. This happened approximately two years after the initial discovery of the COVID-19 virus. The actions taken to get rid of the virus were directly responsible for how quickly things got worse in Wuhan. China was documenting dozens of new cases every day at the start of the outbreak, but by the end of March, that number had dropped to a few dozen. On the other hand, the number of reported cases in Europe increased significantly each day, with Italy reporting more than 250 fatalities daily. Due to this, the World Health Organization (WHO) determined on March 13 that Europe had become the epicentre of the pandemic. On the same day, the United States of America announced an emergency declaration in their state.

As a result of the pandemic, many countries and regions around the world have put in place strict safety measures. Among the measures adopted [1] were the re-establishment of social distance, the usage of face masks, tour regulations, and hand-washing instructions. Researchers concluded that a vaccine would be necessary to prevent the pandemic because such tactics were only expected to slow the virus's spread. People maintained their social distancing strategies by remaining at home, avoiding crowded venues, employing no-touch greetings, and physically isolating themselves from others [2]. During the epidemic, a number of countries made it illegal to be alone and put limits on it. On the other hand, just wearing a face mask wasn't enough to protect against the virus. As the disease spread through droplets, however, people started to use masks correctly [3].

Machine learning and artificial intelligence have proven to be invaluable tools and resources in the fight against the virus [4]. Identifying the patients and comprehending the nature of the virus posed one of the greatest obstacles in the fight against the virus's spread. Here is where machine learning and artificial intelligence stepped in. In addition, machine learning algorithms had a substantial impact on the tracking down of infected individuals' contacts, the prediction and forecasting of the infection rate, and the investigation into the development of novel treatments for the Covid-19 pandemic [5]. Even if some of the necessary systems were already in place, the development of additional systems was still necessary. Creating models that employ a person or object detection has never been straightforward. Using YOLO to automatically detect humans in thermal photographs was already a reality [6]. By utilising Cascade R-CNN [7], it was already possible to execute high-quality object detection on both generic and specialised datasets. YOLO is superior in its ability to recognise both facemasks and human faces, and it is also advantageous in detecting whether Social Distancing rules have been broken [8]. However, since 2020, when the epidemic first appeared, the amount of effort spent conducting research on mask detection using machine learning algorithms has expanded significantly. A technique known as YOLOASC has been introduced, which enables the detection of objects in real-time with improved precision even in the lack of a background [9]. YOLO could also be improved in terms of its capacity to predict the absolute distance of objects using just information acquired from a monocular camera running at 45 frames per second [10]. The Viola-Jones system was also proficient at identifying human faces. The distance between individuals was measured [11] in order to assess whether or not people maintain their social distance. Edge-AI algorithm was proposed to provide technology-based solutions for detecting mask wear in moving human objects, which might be applied not only during pandemics but also to societal and healthcare system-wide issues [12]. Edge-AI algorithm was presented to develop technology-based methods for identifying the use of masks by moving human objects. The Cascade VGGCOV19-NET design not only enhanced the previously known automatic detection of COVID-19 instances from X-ray images but also recognised the use of masks and social distancing [13]. Six versions of the YOLO object identification method (YOLOv3, YOLOv3-tiny, YOLOv4, YOLOv4-tiny, YOLOv5x, and YOLOv5s) have been tried for real-time bunch detection and counting in grapes [14]. In thermographic photographs of the human face, YOLO was also used to identify the areas of interest (ROIs) with the greatest temperatures [15]. This was performed in order to determine the types of human febrile states.

However, one of the primary issues of working with machine learning algorithms is that in order to train the machine, an enormous amount of data is required. This is one of the major drawbacks of working with machine learning algorithms. In this particular scenario, the data may consist of video footage of people wearing or not wearing masks. Putting together the dataset was the most difficult aspect of our research. We collected hundreds of real-world data in order to compile our dataset. We focused more emphasis on the quality of the dataset than on constructing new machine learning models because these models already exist, and the quality of the dataset is what determines the accuracy of the end outcome. Our research is unique because we did not only identify masked and unmasked individuals, but also the accurate application of face masks. This means that our technology can also determine if a person is wearing a mask properly. YOLOv3 has been utilised with three pre-trained models: YOLOv3-SPP, YOLOv3, and YOLOv3-tiny. In addition, the YOLOv5 model is employed alongside its five pre-trained variants: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x.

The introduction comprises the first part of the paper and provides a concise outline of the history of Covid-19, its severity, and the challenges that the globe is experiencing as a result. Following this is a summary of our work and an explanation of its significance. It also addresses the solutions and processes that were implemented to overcome the situation, as well as the impact that machine learning and artificial intelligence, which were utilized for this project, had on it. Moreover, it describes how the incident was addressed. The literature review part contains essential information regarding prior research on this topic. Following the part in the literature review, you will discover information on the dataset and method, followed by the results and conclusions, which will summarise the entire study. The references utilized for this research endeavor are listed at the end of this article.

2. Literature Review

Regarding the scientific fields of image processing and computer vision, datasets are always an essential factor to consider. Due to the fact that the dataset, model, modification, and other factors have a substantial impact on precision and loss. However, data is unquestionably more important if one desires to develop a new dataset that reflects a different perspective. Additionally, it is essential to consider how the dataset can be deployed and employed for machine-level processing. This is a very significant factor to consider. In addition to the dataset, it is vitally essential to have a suitable and efficient model. The model should be capable of learning from the data and displaying the dataset's accuracy, loss, and eventual inference result.

When it comes to image classification, the ability of a model to identify objects is more important than its ability to identify things in a dataset [16]. The YOLO model is the most popular option among the numerous other picture classification systems available today. The authors of the paper [17] implement a YOLOv5 model to measure social distance and social restrictions. Because maintaining social distance has a significant impact on reducing the infection rate, which is not unexpected. In addition, they utilised the Common Objects in Context (COCO) collection, which included 200,000 photographs and 80 different categories, for their implementation. The authors of the paper [18] propose using DNN techniques to correct the Singapore Maritime Dataset (SMD) dataset's annotations in order to generate the SMD-Plus dataset. They integrated the YOLOv5 model and the SMD-Plus dataset, both of which are used for object detection and classification, in order to facilitate growth.

The dataset [19] used for medical face mask detection included both WIDER-FACE and MAFA. This data set contains 7,959 photographs divided into two categories: those in which the subject's face is covered and those in which it is not. Using the YOLOv4 framework, they were successful 99.98% of the time during the training and testing phases of this study. Moreover, the authors of the study [20] demonstrated a real-time face mask identification system that implemented the Tiny-YOLOv4 model. Throughout their research, they applied the Kaggle face mask dataset. The author used the Pascal VOC standard dataset, which included 7,959 images and 16,635 annotations, to perform face mask identification using transfer learning and PP-YOLO [21]. They obtained a mask detection mAP score of 86.69 percent for this model. These articles define the performance standard for this issue, and we used them as inspiration for our investigation [22].

At the Politeknik Negeri Batam, an apparatus known as a face mask detector has been installed to detect face masks in real-time using the YOLO V4 deep learning algorithm [23]. The method that was used to identify face masks in this work [24] reached a high level of accuracy. It was accurate in both categorization and detection to a 96% degree. In the paper [25], they decided to use instead of Mask-R-CNN. You only look once, sometimes known as YOLO, is a strategy for increasing real-time processing speed while maintaining accuracy [26]. YOLO is also a more efficient model. The research presented in paper 24 demonstrates that the algorithm's minimum achievable performance (mAP) on the self-made verification set is 86.92 percent [27]. Furthermore, the study demonstrates that the preset process can be achieved in the actual identification process by operating the robot. The accuracy described in the research was accomplished through the utilization of two cutting-edge object detection models, namely YOLOv3 and faster R-CNN [28]. It presented a method that kept a record of the ratio of mask usage and drew bounding boxes around the faces based on whether or not the mask was present.

Due to COVID-19, the paper focuses on improving public health through computer vision education. Combining classification and detection makes it more difficult to detect microscopic items during an autopsy [29]. Deep neural networks are capable of detecting face masks. YOLOv3 is a restricted topic that can only be accessed through natural illness. Face mask detection in YOLOv3 analyses real-time GPU performance. According to the test, the average loss after 4000 training cycles is 0.0730. After 4000 epochs, mAP is 0.96. This innovative technique for photographing face masks yielded 96% accurate results [30]. Moreover, in the paper [31], to prevent the spread of COVID-19, the government requires public mask use. Masks are manually detected by security personnel, which requires time. Comparative Study of CNN and YOLOv3 in Public employs the well-known CNN and YOLOv3 deep learning models. The system utilised object classification, image, and object tracking to identify masking in camera images and videos. YOLOv3 can detect more accurately and faster than CNN in training and deployment, with 4.8 FPS. Furthermore in the paper [32], masks can be worn to reduce the number of SARS-CoV2 droplets an infected person exhales. The YOLOv3 model is used to create a mask detection model in this research article. The performance of the model makes it ideal for video surveillance. The proposed method employs image filtering to improve a 300-image dataset. The mean average precision of the Data augmentation-based mask detection model was 89.8% during training and 100% during overall testing, with per-frame accuracy ranging between 40.03 and 65.03 per cent. The paper [33] also suggests that object detection is crucial in the era of deep learning. They used two face mask datasets containing 680 and 1400 images, FMY3 with the YOLOv3 Algorithm and FMNMobile with the NASNetMobile and Resnet SSD300 Algorithms. The FMY3 Model's mAP was 34% and its recall rate was 91.7%, whereas the FMNMobile Model's mAP was 98% and its recall rate was 99%. Both models' face mask detection results are displayed.

3. Method

3.1. Dataset

Data is the most crucial part of our life. For a piece of good research, the dataset always plays a crucial vital role. Depending on the data a few times we can get more expected results from research and for that reason, a good and well-trained dataset is important. Moreover, in the research sector, good data priority is almost 60% whereas in the industry it is almost 90%. Aside from that, we can comprehend the significance of data in our actual problems and their solutions. We can use data from a variety of sources to solve real-world problems.

In light of the COVID-19 outbreak, the most important thing we can do right now is make necessary changes to our lifestyles to reduce the spread and severity of the disease. When the proper use of a mask is not observed, judicial intervention is often necessary because mask-wearing in public settings has become an essential civil practice. A mask is required to enter many public places, such as shopping malls, public transportation, offices, educational institute, and so on. As a result, monitoring face mask usage on public property is now an essential civil practice. Covid Face-Mask Monitoring Dataset is a brand-new dataset that we describe in our paper [34]. Our primary goal is to determine whether or not people in crowded public spaces and on the streets are wearing face masks. In addition, it is noted that some individuals do not properly wear masks, which is just as bad as wearing no masks at all and contributes to the spread of infection to others. As a result, we expect to also detect the proper use of face masks. In addition, from the perspective of a nation with a high population density, a significant portion of the population tries to circumvent the COVID regulations and has no intention of wearing facial masks [34]. Walking streets, bus stops, street tea stalls, foot-over bridges, and other similar public spaces in the country are common places to observe this violation regarding aversion to wearing face masks. Fig. 1 shows a representation of our dataset's sample photos.

Fig. 1. Sample image of our proposed dataset

The 6,550 images in our proposed dataset were gathered from the walking streets, bus stops, street tea shops, foot-over bridges, and many other locations. To gather frames and include them in our final dataset, we used personal cell phones and DSLRs. We also have plans to use an action camera or CCTV security camera to get photographs from similar public locations. However, Bangladesh's public spaces rarely have CCTV coverage. Most places where wearing a mask is required, such as offices, hospitals, and shopping centers, use CCTV surveillance cameras. However, the scope of our inquiry includes various viewpoints and vantage points of monitoring while donning a face mask. There are three classifications in our dataset: Mask, No Mask, and Mask not in position. The entire dataset was divided into three classes and scaled in a fixed image dimension format. The photos have a 1080 × 720 pixel resolution. 800 images were chosen for validation reasons and 5,750 images were chosen for training. All frame types are set to JPG, and Table 1 shows the data formats, image sizes, resolutions, and bit depth. For the tagging of the entire self-created dataset, we used LabelImg tools. We implemented classes 0 for mask, 1 for no mask, and 2 for mask not in position while labeling the dataset.

Table 1. Covid Face-Mask Monitoring Dataset's technical characteristics

No of frames	Resizing and Formation of Dataset		
	File Format	*Image Dimension*	*Class Size*
6550	JPG	1080 × 720	3

3.2. Proposed Method

One of the quickest and most well-liked object detection techniques is called You Only Look Once (YOLO). Due to its extraordinary capacity for learning, accuracy, and training speed, YOLO is more well-liked than other apps. It has been applied in many different kinds of applications, including the detection of traffic lights, objects, masks, and license plates. Remaining blocks, bounding boxes, and intersection over union are the fundamental components of YOLO's operation [15]. From YOLO to YOLOv5, many improvements have been made. It was primarily to blame for the detection of objects and the loss of confidence in YOLO. However, in YOLOv2, anchors are added to K-Means. In addition, YOLOv2 included two training stages and a fully convolutional neural network system. On the other hand, FPN-based multi-scale detection was added to YOLOv3. The YOLOv4 also included new features like the loss function, SPP, the MISH activation function, data enhancements with Mosaics, and so forth. At long last, YOLOv5 accompanied adaptable control of the model, information upgrade, and use of the Hardswish enactment capability [26].

The YOLOv3 feature extractor, known as Darknet-53 (it has 52 convolutions), was inspired by ResNet and FPN (Feature-Pyramid Network) architectures. It has skip connections (like ResNet) and 3 prediction heads (like FPN) and processes the image at various spatial compressions. Yolov3, like its predecessor, performs well at a wide range of input resolutions. Several checkpoints can be found in the model zoo of GluonCV: each with a different input resolution, but identical network parameters are stored in those checkpoints. Yolov3 received a mean average precision (mAP) of 37 when tested on the COCO-2017 validation set with an input resolution of 608 × 608. This score is 17 times faster than the trained version of GluonCV's Faster-RCNN-ResNet50, which is a faster RCNN architecture that uses ResNet-50 as its backbone. In that model zoo, the only detectors with mAP scores below 30 that were fast enough to compete with Yolov3 (Mobilenet-SSD architectures). YOLOv3 model design in Fig. 2.

Fig. 2. YOLOv3's model architecture

For the proposed dataset that was utilized by the facial mask detection system, we implemented YOLOv5. The backbone, neck, and head are the three distinct architectural blocks in the YOLOv5 family. YOLOv5 spine for the most part, utilizes CSPDarknet as the foundation of extraction highlights from the picture dataset utilizing a halfway organization. Additionally, if we focus on the YOLOv5 neck, we discover that the PANet is utilized to generate pyramid feature networks [19]. In addition, it aggregates the feature before sending it to the head for predictions. Additionally, the YOLOv5 head has layers that are able to generate predictions for the detection objects based on the anchor boxes. Leaky ReLU and Sigmoid activation functions are primarily utilized in YOLOv5's activation function and optimization strategies. Additionally, it employs SGD and ADAM optimization methods as the optimization [17]. We addressed the YOLOv3 and YOLOv5 model design in Fig. 3.

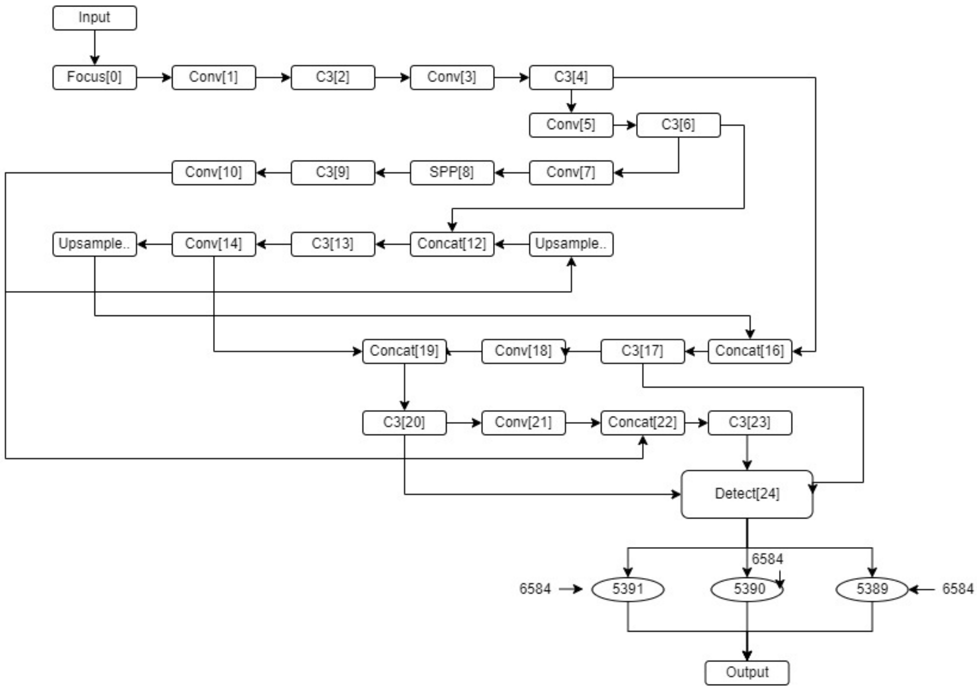

Fig. 3. YOLOv5's model architecture

4. Results and Discussion

We started by implementing the YOLOv3 model as well as the YOLOv3-SPP, YOLOv3, and YOLOv3-tiny pre-trained models for face mask detection. In addition, we used the YOLOv5 model as well as the pre-trained models YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The proposed model has a batch size of 4, a learning rate of 0.001, and 75 training epochs for each pre-trained model during the training period. The primary purpose of precision, recall, the F-score, and mAP is to evaluate the object's exact detection accuracy. The accurate prediction and recall measures that the positive classes easily detect were the primary focus of precision. The following formulas are used to determine precision and recall. Here, Pos stood for "positive," Neg stood for "negative," T stood for "true," and F stood for "false

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (1)$$

$$Recall = \frac{T\_Positive}{T\_Positive + F\_Negative} \quad (2)$$

Accuracy is primarily used as an additional metric for evaluating the classification's performance. T stood for true, and F stood for false as well. The equation between accuracy and error is depicted in the following :

$$Accuracy = \frac{T\_P\,os + T\_Neg}{T\_P\,os + T\_Neg + F\_P\,os + F\_Neg} \tag{3}$$

$$Error = 1 - Accuracy \tag{4}$$

The mean average of precision (mAP) is the average of the AP calculated for all the classes. The equation of mAP is given below :

$$Mean\ Average\ Precision\ (mAP\ ) = \frac{\sum_{q=0}^{Q} AveP\ (q)}{Q} \tag{5}$$

4.1. Implementation of the YOLOv3 model

When compared to the data and image quality, our results are remarkable. because, from Bangladesh's point of view, the lack of camera facilities makes it extremely difficult to properly collect data. In addition, a variety of image types exist, including noisy images, light capacity, sunlight, shadow, and others. As a result, getting good results can be challenging.

In this section, we represented the outcomes of the YOLOv3 model along with three pre-trained models using our proposed dataset for mask detections. In Fig. 4, we represented the confusion matrix for the YOLOv3-tiny pre-trained model where the result is 0.94 for mask detection, 0.96 for no mask detection, and 0.81 for mask not in position.

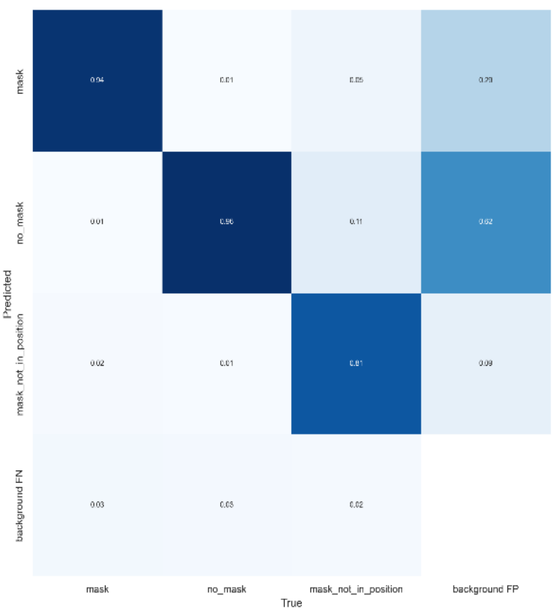

Fig. 4. Confusion matrix for YOLOv3-tiny using proposed dataset

Moreover, in Fig. 5, the confusion matrix of the YOLOv3 pre-trained model is presented whereas the result is 0.97 for mask detection, 0.97 for no mask detection, and 0.83 for mask not in position. After that, we are concerned about the proposed dataset's precision, recall, and mAP score. Recall scores can easily measure positive classes, whereas precision scores are primarily dependent on accurate prediction. 75 epochs were used to train the entire dataset with a batch size of 4.

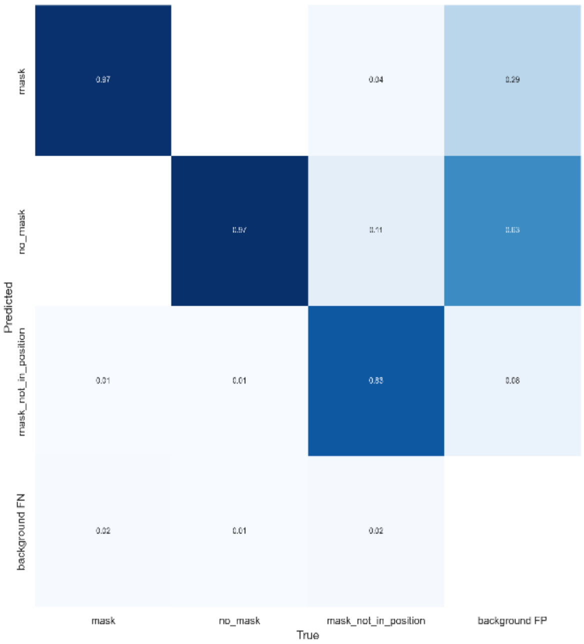

Fig. 5. Confusion matrix for YOLOv3 using proposed dataset

Furthermore, we presented a confusion matrix for YOLOv3-SPP pre-trained model in Fig. 6 where the accuracy is 0.95 for mask, 0.97 for no mask, and 0.88 for mask not in position class. Compared to the previous two pre-trained models, in YOLOv3-SPP accuracy is increased.

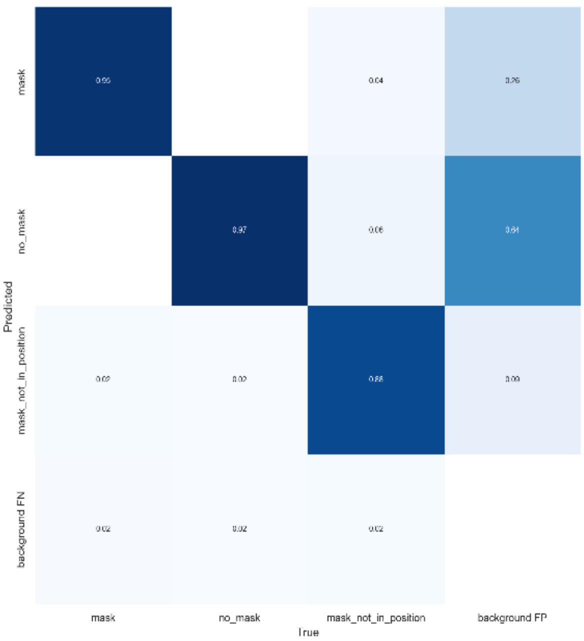

Fig. 6. Confusion matrix for YOLOv3-SPP using proposed dataset

The best precision score during training was 0.9253 for the YOLOv3-tiny pre-trained model, 0.9381 for the YOLOv3 pre-trained model, and 0.9454 for the YOLOv3-SPP, which is the benchmark score for our proposed dataset. Fig. 7 is a graphic representation of their score.

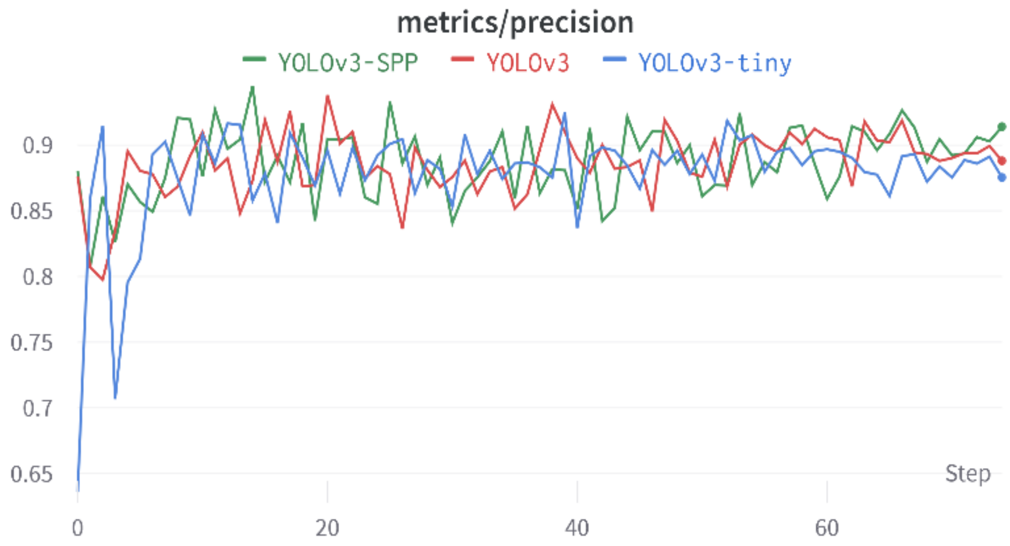

Fig. 7. Precision score for YOLOv3 using proposed dataset

In addition, the best recall scores for the same pre-trained models—tiny YOLO, YOLOv3, and SPP-YOLO—are 0.8872, 0.9285, and 0.9325, respectively—all of which serve as a benchmark score for the same dataset. The recall scores are shown in Fig. 8.

Fig. 8. Recall score for YOLOv3 using proposed dataset

We used a scale of 0.5 to represent the mean Average Precision (mAP) score in Fig. 9. We obtained three distinct types of outcomes for three pre-trained models for our proposed dataset. We achieved a score of 0.9292 for the tiny YOLOv3 pre-trained model, which serves as the benchmark for our proposed dataset. Besides, when we dealt with the YOLOv3 pre-prepared model around then we see that the pre-prepared model accomplished a 0.9413 score which is the benchmark result for our proposed dataset. To add, that score was 0.9395 for the SPP-YOLOv3 pre-trained model.

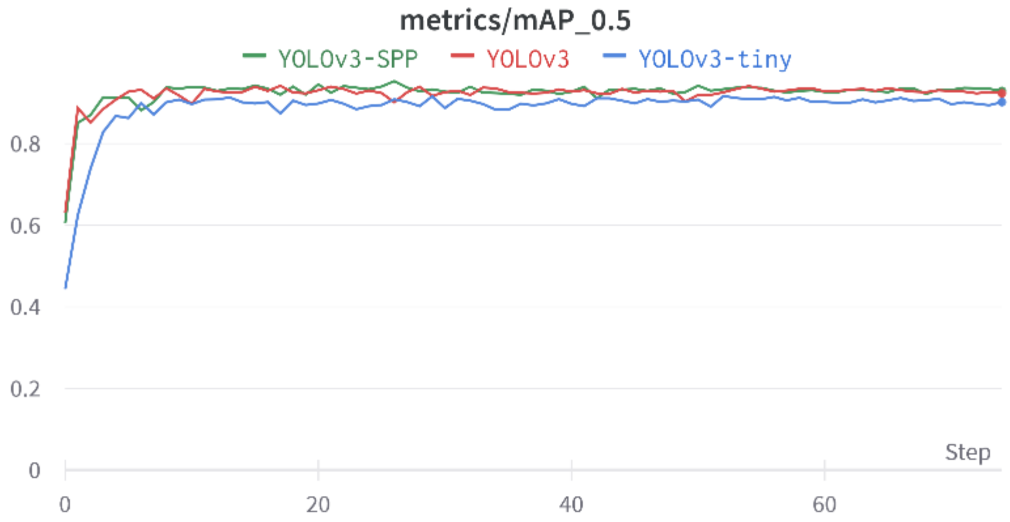

Fig. 9. Mean average precision score for YOLOv3 using proposed dataset

On a scale of 0.5:0.95, we depicted the mean Average Precision (mAP) score in Fig. 10. For the tiny YOLOv3 pre-prepared model, we accomplished a 0.6772 score on the size of Guide 0.5:0.95 which is the benchmark result for our proposed dataset. Besides, during the preparation time of the YOLOv3 pre-prepared model we see that model accomplished a 0.7136 score on the size of Guide 0.5:0.95 which is the benchmark result for our proposed dataset. To add, that score was 0.7266 for the SPP-YOLOv3 pre-trained model.

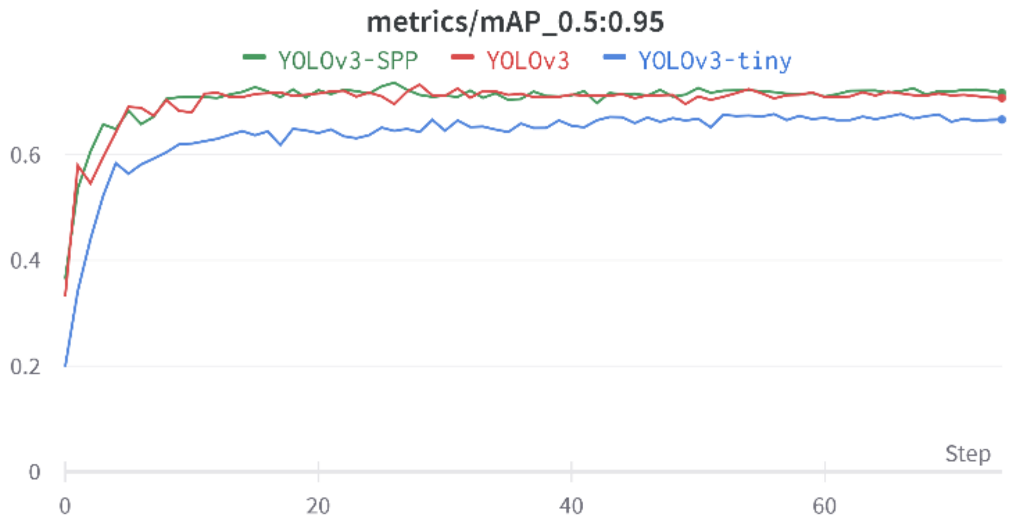

Fig. 10. Mean average precision score in the scale of 0.5:0.95 for YOLOv3 using proposed dataset

4.2. Implementation of the YOLOv5 model

Using our proposed dataset, we presented the results of the YOLOv5 model as well as five pre-trained mask detection models in this section. The confusion matrix for the pre-trained YOLOv5n model is depicted in Fig. 11, with values of 0.95 for mask detection, 0.97 for no mask detection, and 0.85 for mask not in position.

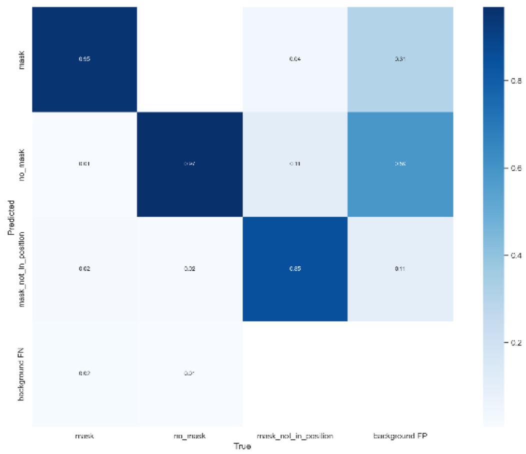

Fig. 11. Confusion matrix for YOLOv5n using proposed dataset

In addition, the YOLOv5s pre-trained model's confusion matrix is depicted in Fig. 12, with values of 0.97 for mask detection, 0.97 for no mask detection, and 0.81 for mask not in position.

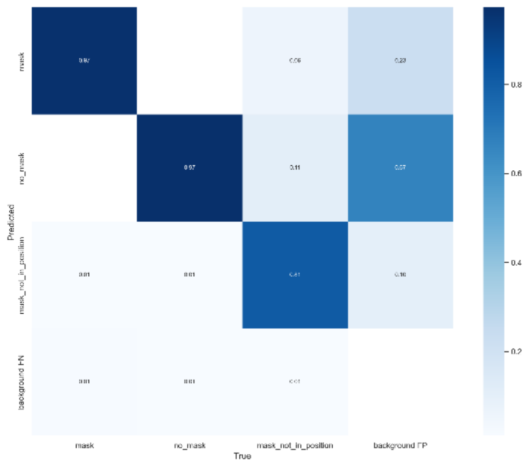

Fig. 12. Confusion matrix for YOLOv5s using proposed dataset

In addition, the confusion matrix for the YOLOv5m pre-trained model is shown in Fig. 13, with values of 0.96 for mask detection, 0.95 for no mask detection, and 0.84 for a mask that is not in position.

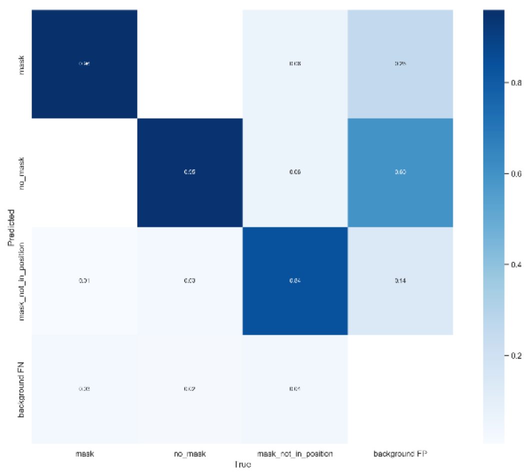

Fig. 13. Confusion matrix for YOLOv5m using proposed dataset

In addition, the confusion matrix of the YOLOv5l pre-trained model is shown in Fig. 14, with values of 0.96 for mask detection, 0.97 for no mask detection, and 0.88 for a mask that is not in position.

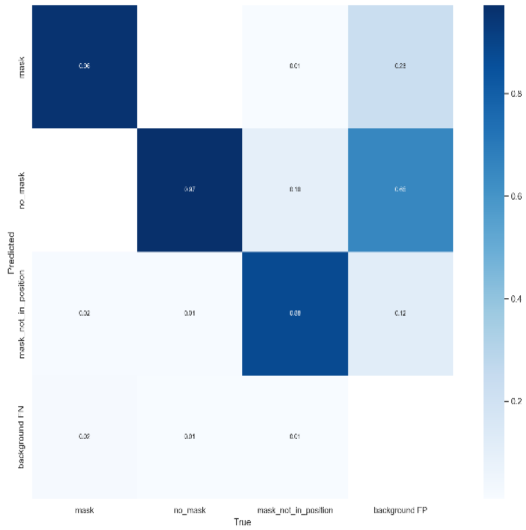

Fig. 14. Confusion matrix for YOLOv5l using proposed dataset

In addition, in Fig. 15, we displayed a confusion matrix for the YOLOv5x pre-trained model with accuracy values of 0.97 for mask, 0.97 for no mask, and 0.84 for mask not in position class.

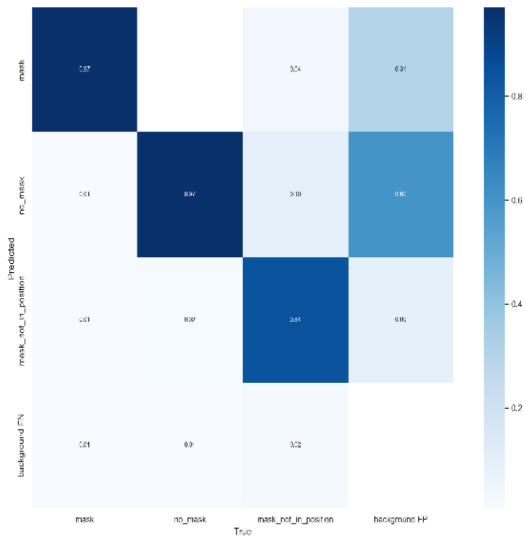

Fig. 15. Confusion matrix for YOLOv5x using proposed dataset

After that, the precision, recall, and mAP score of the proposed dataset concern us. Precision scores are primarily dependent on precise prediction, whereas recall scores can easily measure positive classes. The entire dataset was trained over 75 epochs with a batch size of 4 each.

The best precision score during training was 0.9205 for the YOLOv5n pre-trained model, 0.9094 for the YOLOv5s pre-trained model, 0.9209 for the YOLOv5m pre-trained model, 0.9037 for YOLOv5l pre-trained model, and 0.9032 for the YOLOv5x pre-trained model, which is the benchmark score for our proposed dataset. Fig. 16 is a graphic representation of their score.

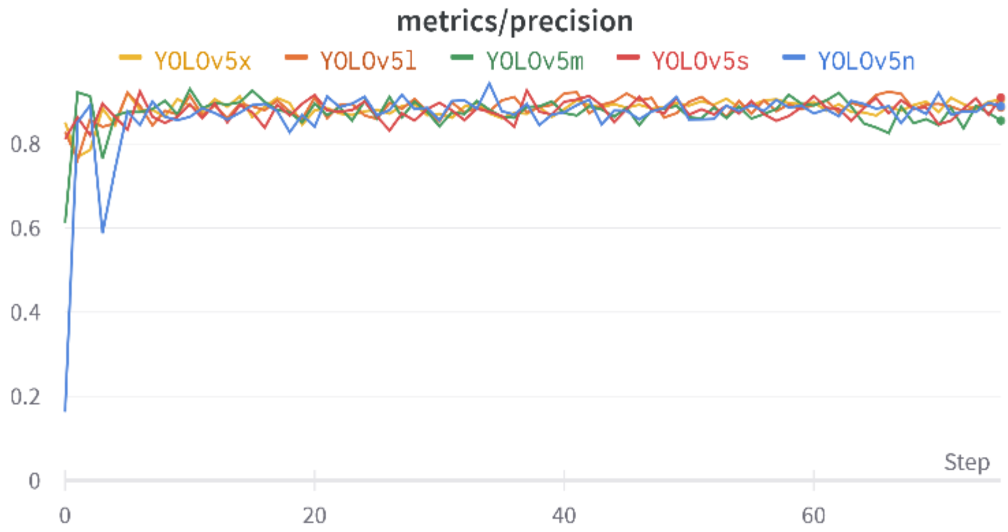

Fig. 16. Precision score for YOLOv5 using proposed dataset

In addition, the best recall scores for the same pre-trained models YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x —are 0.9037, 0.8959, 0.8756, 0.9445, and 0.9286 respectively—all of which serve as a benchmark score for the same dataset. The recall scores are shown in Fig. 17.

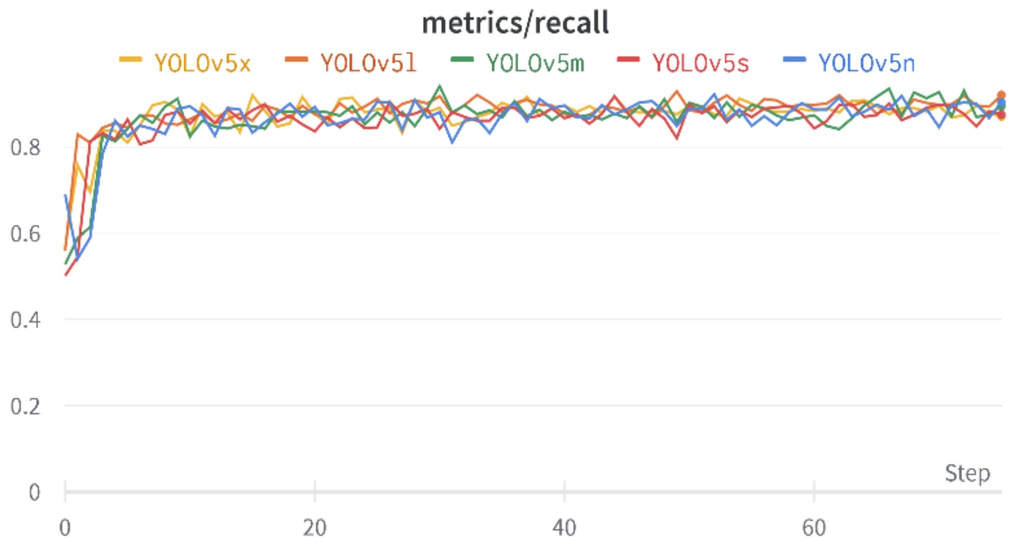

Fig. 17. Recall score for YOLOv5 using proposed dataset

We used a scale of 0.5 to represent the mean Average Precision (mAP) score in Fig. 18. We obtained five distinct types of outcomes for five pre-trained models for our proposed dataset. We achieved a score of 0.9299 for the YOLOv5n pre-trained model, which serves as the benchmark for our proposed dataset. Besides, when we dealt with the YOLOv5s pre-trained model around then we see that the pre-trained model accomplished a 0.9472 score which is the benchmark result for our proposed dataset. To add, that score was 0.9347 for the YOLOv5m pre-trained model. Moreover, 0.9447 and 0.9417 scores for YOLOv5l and YOLOv5x respectively.

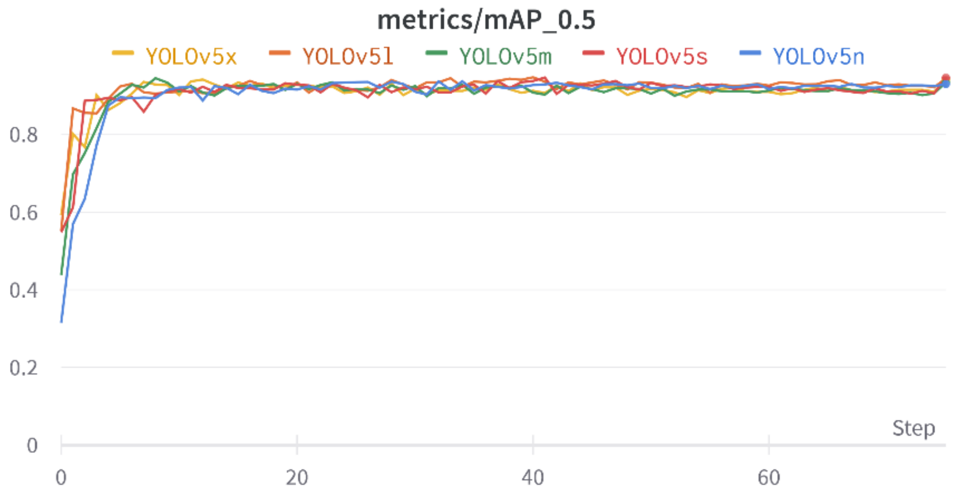

Fig. 18. Mean average precision score for YOLOv5 using proposed dataset

Mean average precision score on the scale of 0.5:0.95 for YOLOv5 using proposed dataset show as Fig. 19.

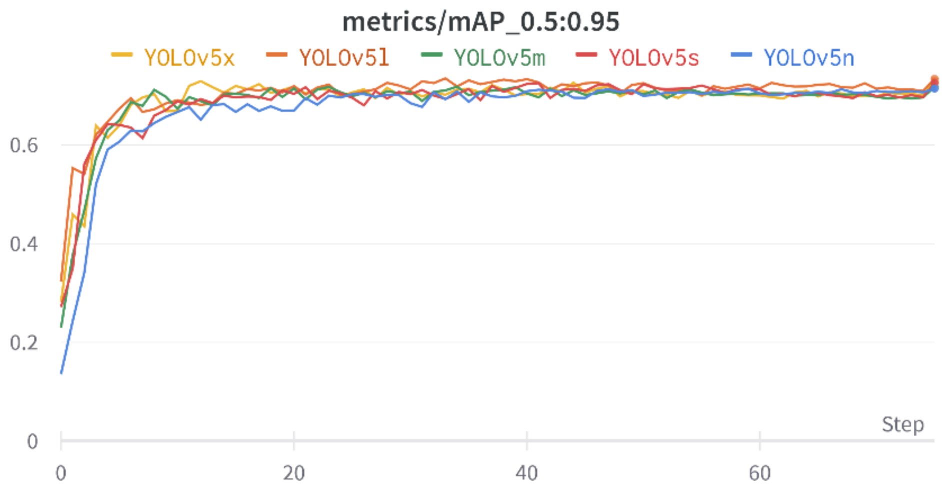

Fig. 19. Mean average precision score on the scale of 0.5:0.95 for YOLOv5 using proposed dataset

4.3. Discussion

Based on Bangladeshi perspective images, we represented the Covid Face-Mask Monitoring Dataset for determining whether a mask is worn correctly, whether it is missing, and whether it is not. Using YOLOv3, YOLOv5, and a variety of pre-trained models, we were able to achieve remarkable precision, recall, and mAP results for the dataset we proposed. Fig. 4, Fig. 5, and Fig. 6 show, respectively, the confusion matrix for the YOLOv3 model and three distinct types of pre-trained models. In addition, as the benchmark score for our proposed dataset, we represented the precision score, recall score, and mAP score of YOLOv3 for three pre-trained models in Fig. 7, Fig. 8, and Fig. 10. Fig.11, Fig. 12, Fig. 13, and Fig. 15 show, in addition, the confusion matrix for the YOLOv5 model and five distinct types of pre-trained models, respectively. In addition, as the benchmark score for our proposed dataset, we represented the precision score, recall score, and mAP score of YOLOv5 for five pre-trained models in

Fig. 16, Fig. 17, Fig. 18, and Fig. 19. Finally, we presented the test results and a comparison results in Table 2.

Table 2. Covid Face-Mask Monitoring Dataset Model Comparison

Model	Evaluation score				
	Pre-trained models	*Precision*	*Recall*	*mAP_0.5*	*mAP_0.5:0.95*
YOLOv3	YOLOv3-tiny	0.9253	0.8872	0.9292	0.6772
	YOLOv3	0.9381	0.9285	0.9413	0.7136
	YOLOv3-SPP	0.9454	0.9325	0.9395	0.7266
YOLOv5	YOLOv5n	0.9205	0.9037	0.9299	0.7143
	YOLOv5s	0.9094	0.8959	0.9472	0.7178
	YOLOv5m	0.9209	0.8756	0.9347	0.7256
	YOLOv5l	0.9037	0.9445	0.9447	0.7344

5. Conclusion

Due to the fact that proper face mask use is an essential means of preventing COVID-19 infection from person to person, we have presented a comprehensive study based on a novel dataset for the detection of proper face mask use in this work. We used YOLOv3 and YOLOv5 and their various pre-trained models for mask detection to achieve higher accuracy and better performance. We created a dataset with tens of thousands of face images, both masked and unmasked, in order to train and validate our detection system. We were able to demonstrate that YOLOv3 and YOLOv5 are efficient and effective models for the use case by focusing not only on the presence of masks but also on their proper use and positioning. Our proposed solution can be used in a wide range of situations where CCTV observations alone are not sufficient to monitor the proper use of a face mask in a crowded setting.

Acknowledgment

The authors are thankful to the Faculty of Statistics at the Technische Universität Dortmund for their research lab and other crucial technical support along with the Department of Computer Science and Engineering at the Hofstra University and the University of Liberal Arts Bangladesh. They are immensely grateful to everyone who provided insight and expertise that greatly assisted the research. The authors are also thankful to the reviewers for their insights.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.
Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper

References

[1] N.-C. Chiu *et al.*, "Impact of Wearing Masks, Hand Hygiene, and Social Distancing on Influenza, Enterovirus, and All-Cause Pneumonia During the Coronavirus Pandemic: Retrospective National Epidemiological Surveillance Study," *J. Med. Internet Res.*, vol. 22, no. 8, p. e21257, Aug. 2020, doi: 10.2196/21257.

[2] Y. Kebede, Y. Yitayih, Z. Birhanu, S. Mekonen, and A. Ambelu, "Knowledge, perceptions and preventive practices towards COVID-19 early in the outbreak among Jimma university medical center visitors, Southwest Ethiopia," *PLoS One*, vol. 15, no. 5, p. e0233744, May 2020, doi: 10.1371/journal.pone.0233744.

[3] L. Martinelli *et al.*, "Face Masks During the COVID-19 Pandemic: A Simple Protection Tool With Many Meanings," *Front. Public Heal.*, vol. 8, p. 606635, Jan. 2021, doi: 10.3389/fpubh.2020.606635.

[4] S. Kushwaha *et al.*, "Significant Applications of Machine Learning for COVID-19 Pandemic," *J. Ind. Integr. Manag.*, vol. 05, no. 04, pp. 453–479, Dec. 2020, doi: 10.1142/S2424862220500268.

[5] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review," *Chaos, Solitons & Fractals*, vol. 139, p. 110059, Oct. 2020, doi: 10.1016/j.chaos.2020.110059.

[6] M. Ivašić-Kos, M. Krišto, and M. Pobar, "Human Detection in Thermal Imaging Using YOLO," in *Proceedings of the 2019 5th International Conference on Computer and Technology Applications*, Apr. 2019, vol. Part F1482, pp. 20–24, doi: 10.1145/3323933.3324076.

[7] Z. Cai and N. Vasconcelos, "Cascade R-CNN: High Quality Object Detection and Instance Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 5, pp. 1483–1498, May 2021, doi: 10.1109/TPAMI.2019.2956516.

[8] K. Bhambani, T. Jain, and K. A. Sultanpure, "Real-time Face Mask and Social Distancing Violation Detection System using YOLO," in *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)*, Oct. 2020, pp. 1–6, doi: 10.1109/B-HTC50970.2020.9297902.

[9] P. Hurtik, V. Molek, and P. Vlasanek, "YOLO-ASC: You Only Look Once And See Contours," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, pp. 1–7, doi: 10.1109/IJCNN48605.2020.9207223.

[10] M. Vajgl, P. Hurtik, and T. Nejezchleba, "Dist-YOLO: Fast Object Detection with Distance Estimation," *Appl. Sci.*, vol. 12, no. 3, p. 1354, Jan. 2022, doi: 10.3390/app12031354.

[11] F. Hardan and A. R. J. Almusawi, "Developing an Automated Vision System for Maintaing Social Distancing to Cure the Pandemic," *Al-Khwarizmi Eng. J.*, vol. 18, no. 1, pp. 38–50, Mar. 2022, doi: 10.22153/kej.2022.03.002.

[12] K. Sengupta and P. R. Srivastava, "HRNET: AI-on-Edge for Mask Detection and Social Distancing Calculation," *SN Comput. Sci.*, vol. 3, no. 2, p. 157, Mar. 2022, doi: 10.1007/s42979-022-01023-1.

[13] A. Karacı, "VGGCOV19-NET: automatic detection of COVID-19 cases from X-ray images using modified VGG19 CNN architecture and YOLO algorithm," *Neural Comput. Appl.*, vol. 34, no. 10, pp. 8253–8274, May 2022, doi: 10.1007/s00521-022-06918-x.

[14] M. Sozzi, S. Cantalamessa, A. Cogato, A. Kayad, and F. Marinello, "Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms," *Agronomy*, vol. 12, no. 2, p. 319, Jan. 2022, doi: 10.3390/agronomy12020319.

[15] J. R. da Silva *et al.*, "Recognition of Human Face Regions under Adverse Conditions—Face Masks and Glasses—In Thermographic Sanitary Barriers through Learning Transfer from an Object Detector," *Machines*, vol. 10, no. 1, p. 43, Jan. 2022, doi: 10.3390/machines10010043.

[16] I. H. Al Amin and F. H. Arby, "Implementation of YOLO-v5 for a Real Time Social Distancing Detection," *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 01–06, Jul. 2022, doi: 10.30871/jaic.v6i1.3484.

[17] J.-H. Kim, N. Kim, Y. W. Park, and C. S. Won, "Object Detection and Classification Based on YOLO-V5 with Improved Maritime Dataset," *J. Mar. Sci. Eng.*, vol. 10, no. 3, p. 377, Mar. 2022, doi: 10.3390/jmse10030377.

[18] 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)*, Mar. 2021, pp. 209–213, doi: 10.1109/ICAIS50930.2021.9395857.

[19] K. V. Sathyamurthy, A. R. Shri Rajmohan, A. Ram Tejaswar, K. V, and G. Manimala, "Realtime Face Mask Detection Using TINY-YOLO V4," in *2021 4th International Conference on Computing and Communications Technologies (ICCCT)*, Dec. 2021, pp. 169–174, doi: 10.1109/ICCCT53315.2021.9711838.

[20] W. Jian and L. Lang, "Face mask detection based on Transfer learning and PP-YOLO," in *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, Mar. 2021, pp. 106–109, doi: 10.1109/ICBAIE52039.2021.9389953.

[21] S. Susanto, F. A. Putra, R. Analia, and I. K. L. N. Suciningtyas, "The Face Mask Detection For Preventing the Spread of COVID-19 at Politeknik Negeri Batam," in *2020 3rd International Conference on Applied Engineering (ICAE)*, Oct. 2020, pp. 1–5, doi: 10.1109/ICAE50557.2020.9350556.

[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)*, Jul. 2020, pp. 1–5, doi: 10.1109/ICCCNT49239.2020.9225384.

[23] R. Liu and Z. Ren, "Application of Yolo on Mask Detection Task," *2021 IEEE 13th Int. Conf. Comput. Res. Dev. ICCRD 2021*, pp. 130–136, Jan. 2021, doi: 10.1109/ICCRD51685.2021.9386366.

[24] Y. Li, J. Yan, and B. Hu, "Mask Detection Based On Efficient-YOLO," in *2021 40th Chinese Control Conference (CCC)*, Jul. 2021, vol. 2021-July, pp. 4056–4061, doi: 10.23919/CCC52363.2021.9549903.

[25] S. Singh, U. Ahuja, M. Kumar, K. Kumar, and M. Sachdeva, "Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment," *Multimed. Tools Appl.*, vol. 80, no. 13, pp. 19753–19768, May 2021, doi: 10.1007/s11042-021-10711-8.

[26] 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, Feb. 2021, doi: 10.1016/j.scs.2020.102600.

[27] A. Kumar, A. Kalia, A. Sharma, and M. Kaushal, "A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 6, pp. 6783–6796, Jun. 2023, doi: 10.1007/s12652-021-03541-x.

[28] J. Yu and W. Zhang, "Face Mask Wearing Detection Algorithm Based on Improved YOLO-v4," *Sensors*, vol. 21, no. 9, p. 3263, May 2021, doi: 10.3390/s21093263.

[29] S. Balaji, B. Balamurugan, T. A. Kumar, R. Rajmohan, and P. P. Kumar, "A Brief Survey on AI Based Face Mask Detection System for Public Places," *Irish Interdisciplinary Journal of Science & Research (IIJSR) 2021*, p. 10, 2021. [Online]. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3814341.

[30] D. Danu Prasetia, A. Yuswanto, and B. Wibowo, "Design Of Machine Learning Detection Mask Using Yolo And Darknet On Nvidia Jetson Nano," *TEKNOKOM*, vol. 5, no. 1, pp. 88–95, Mar. 2022, doi: 10.31943/teknokom.v5i1.69.

[31] Z. Qin, Z. Guo, and Y. Lin, "An Implementation of Face Mask Detection System Based on YOLOv4 Architecture," in *2022 14th International Conference on Computer Research and Development (ICCRD)*, Jan. 2022, pp. 207–213, doi: 10.1109/ICCRD54409.2022.9730470.

[32] J. Ha, M. Kambe, and J. Pe, "Data Mining," *Elsevier*, p. 703, 2012, doi: 10.1016/C2009-0-61819-5.

[33] M. J. Zaki and W. Meira, Jr, "Data Mining and Analysis,". *Cambridge University Press*, 2014, doi: 10.1017/CBO9780511810114.

[34] M. I. Masum, M. A. Jishan, K. R. Mahmud, and A. K. Al Azad, "Covid Face-Mask Monitoring Dataset," *Harvard Datavers*, Apr. 13, 2022. [Online]. Available at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BKPMHT.