Towards smart surveillance as an aftereffect of COVID-19 outbreak for recognition of face masked individuals using YOLOv3 algorithm

Saurav Kumar[1] · Drishti Yadav[2] · Himanshu Gupta[3] · Mohit Kumar[4] ·
Om Prakash Verma[3]

Abstract

The eruption of COVID-19 pandemic has led to the blossoming usage of face masks among individuals in the communal settings. To prevent the transmission of the virus, a mandatory mask-wearing rule in public areas has been enforced. Owing to the use of face masks in communities at different workplaces, an effective surveillance seems essential because several security analyses indicate that face masks may be used as a tool to hide the identity. Therefore, this work proposes a framework for the development of a smart surveillance system as an aftereffect of COVID-19 for recognition of individuals behind the face mask. For this purpose, transfer learning approach has been employed to train the custom dataset by YOLOv3 algorithm in the Darknet neural network framework. Moreover, to demonstrate the competence of YOLOv3 algorithm, a comparative analysis with YOLOv3-tiny has been presented. The simulated results verify the robustness of YOLOv3 algorithm in the recognition of individuals behind the face mask. Also, YOLOv3 algorithm achieves a mAP of 98.73% on custom dataset, outperforming YOLOv3-tiny by approximately 62%. Moreover, YOLOv3 algorithm provides adequate speed and accuracy on small faces.

Keywords COVID-19 · Facemask detection · Object detection · Surveillance system · YOLOv3 algorithm · Convolutional neural network · Deep neural networks

1 Introduction

After the Second World War, the novel COVID-19 pandemic has proved to be the greatest momentous emergency that the world has encountered [38]. COVID-19 has posed a central threat to humanity and society, and triggered a global economic crisis [20]. It has nearly affected almost all the nations across the globe resulting in devastating health

✉ Om Prakash Verma
vermaop@nitj.ac.in

Extended author information available on the last page of the article

consequences and a shattering death toll [58]. Moreover, total number of deaths due to the COVID-19 pandemic is regrettably destined to face an exponential rise in future [59]. In addition, unparalleled outcomes and disturbances in the society have been observed. This has led to severe restrictions in the whereabouts and social behavior of a significant global populace. In the absence of any scientifically proven solution, the policymakers across the globe employed complete lockdown for an unstipulated time and other preventive measures as an immediate solution [12]. Since the virus is airborne, an important safety precaution to break the infection chain of COVID-19 is the use of face masks [7, 37]. As per WHO, wearing a mask and social distancing are considered as the key preventive measures that may limit the spread of COVID-19 through the infectious droplets released into the atmosphere when an infected person coughs or sneezes or even talks [30, 57]. Further, the routine usage of face masks may significantly reduce the risk of infection by reducing the exposure to the pathogens, balances the high filtration, adequate breathability and optionally, fluid penetration resistance [32, 56].

Moreover, after the completion of lockdown period and restoration of normalcy at various organizations and institutions across the globe, the use of face masks will continue to exist as a preventive measure against airborne viruses including COVID-19 to mitigate its spread [57]. Therefore, owing to the use of face masks, the identification of face masked individuals associated with a certain place, for instance, office of national security, students at an educational institute, public places such as railway station, bus station, metro rail, and employees of various research labs and organizations etc., becomes essential in many aspects [15, 26]. These aspects include monitoring and surveillance in public places, educational institutions, organizations, and other work environments for the detection of interpersonal crimes and threats, etc. [14]. Therefore, there is an urgent need to develop a highly secure smart surveillance and security system for identifying the face masked individuals and upgrading the existing surveillance scenario.

Previous reported literature reveals the employment of several deep learning (DL) approaches for security and surveillance systems [8, 21, 39]. Various object detection techniques have been proposed pertaining to the unprecedented growth in the field of computer vision [19, 27, 28]. Yet, the acquisition of acceptable accuracy in detection and recognition tasks has not been achieved. Due to the drawbacks of conventional approaches for object detection (based on Haar cascade classifier [54], SVMs [13] and sliding window methods [18]), novel DL-based models have been proposed [60]. These DL methods include CNN [11, 24, 33, 45, 46, 50], R-CNN family [6, 49], SSD family [16, 34], YOLO (You Only Look Once) [22, 36], etc. and other classification methods (MLP, SVM) [10, 51], which have proven to be robust and competent in the context of real-time complex object detection tasks [5, 43, 62]. Moreover, for the detection of face mask, various DL-based methods have been proposed and utilized in the literature. However, they are only limited to detect either the face or face mask, but no significant work has been carried to identify the individuals behind the face mask. Some of the novel approaches employed to detect the face or face mask have been tabulated in Table 1.

Despite the efficient performance of these approaches in object detection, their low computational capability and slow response time (due to their pipeline architecture) serve as their major limitations [48]. These limitations have been drastically overcome by YOLOv3, a variant of YOLO algorithm [47], which is a promising approach for the real-time sophisticated object detection tasks [2, 29, 52, 55]. The superiority of YOLOv3 is evident from Fig. 1 which displays the speed and accuracy tradeoff on the mAP (Mean Average Precision) at 0.5 IoU (Intersection over Union) metric [47]. As seen from Fig. 1, YOLOv3 outperforms other approaches (including SSD, RetinaNet,

Table 1 Novel approaches for face and face mask detection

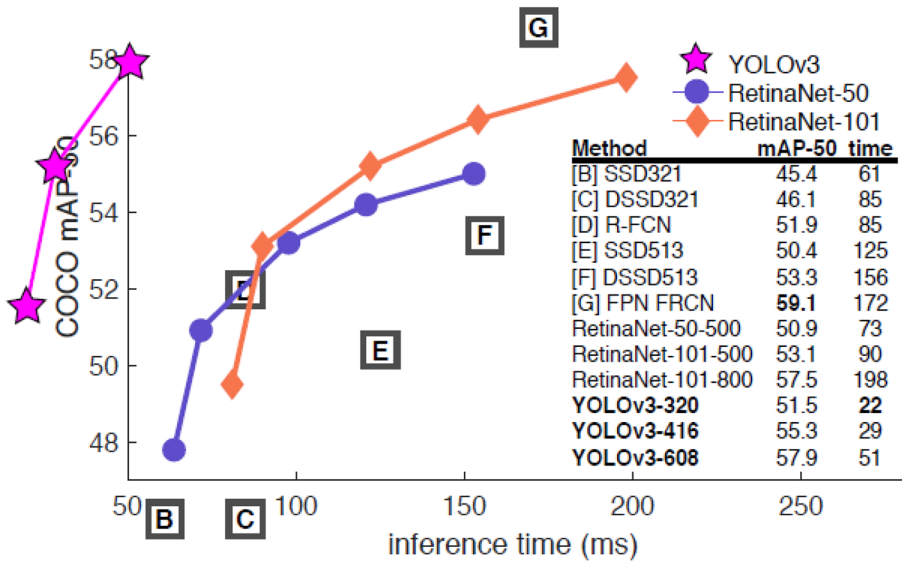| Methods	Reported work	Performance	Highlights
CNN based	Ge et al. [17]	Average precision: 74.6%	Face occlusion detection techniques using LLE-CNN
	Bu et al. [3]	Accuracy: 86.6%, Recall: 87.8%	Cascade framework of CNN for classifying masked face
	Qin et al. [44]	Accuracy: 98.70%	Combination of Image super-resolution network and classification network to identify facemask wearing conditions
	Inamdar et al. [23]	Accuracy: 98.60%	CNN (with 8 layers) is used to detect faces, extract ROI and detect facemask
	Chavda et al. [4]	Precision: 98.28%, Recall: 100%, F1-score: 99.13%	Two stage detection algorithm: First face detection by RetinaFace and then facemask detection by NASNetMobile
	Yadav [61]	Precision: 91.7%	CNNs, FPN, and a context attention module for facemask detection
	Militante et al. [40]	Accuracy: 96%	VGG16 is used to detect facemask
	Khandelwal et al. [25]	Accuracy: 97.6%	Detect social distancing and mask using MobileNetV2
	Loey et al. [35]	Average precision: 81%	Facemask classification algorithm using YOLOv2 and ResNet-50
Hybrid	Nieto-Rodr'iguez et al. [42]	Recall: >95%, False positive rate: <5%	It uses colour filter for classifying face and facemask using skin texture in HSV colour space
	Vinitha et al. [53]	Not specified	Real-time facemask detection using MobileNetV2 and PyTorch

Method	mAP-50	time
[B] SSD321	45.4	61
[C] DSSD321	46.1	85
[D] R-FCN	51.9	85
[E] SSD513	50.4	125
[F] DSSD513	53.3	156
[G] FPN FRCN	**59.1**	172
RetinaNet-50-500	50.9	73
RetinaNet-101-500	53.1	90
RetinaNet-101-800	57.5	198
YOLOv3-320	51.5	**22**
YOLOv3-416	55.3	29
YOLOv3-608	57.9	51

Fig. 1 Speed and accuracy tradeoff on the mAP at 0.5 IoU metric [47]

etc.) on the COCO dataset at an IoU of 0.5. Moreover, YOLOv3 provides a considerably high mAP (> 50) and significantly less inference time compared to other methods.

Conclusively, the outstanding performance of YOLOv3 for detection in terms of real-time response rate has proved its preeminence over other algorithms belonging to the YOLO family and classical approaches [9]. However, till the reporting of this investigation, to the best of the knowledge of the authors, the implementation of YOLOv3 algorithm has not been reported earlier for the surveillance applications, especially in the identification of face masked individuals. Hence, under the above backdrop, it appears that there is an enormous scope to ensure high prediction probability with acceptable computational speed and accuracy by YOLOv3 in the identification of face masked individuals. This may assist in the design of an effective surveillance system for monitoring and identification tasks. Based on the above discussion, the main contributions of this work can be sketched as follows:

- Firstly, an attempt has been made to propose a framework to assist in the development of an effective smart security and surveillance system as an aftereffect of COVID-19 pandemic for facial recognition. The main focus is to identify and recognize the people with partially covered faces (for instance, people who are wearing a mask).
- Secondly, an effort has been made to examine the efficiency of state-of-the-art YOLOv3 algorithm in precise and automatic identification and recognition of face masked individuals in real-time. Contrary to the existing works which focus on the detection of people who are wearing masks (wearing mask or not), this work focuses on identifying a face masked person.
- Thirdly, this work reflects the competence of YOLOv3 algorithm over YOLOv3-tiny towards the recognition of face masked individuals in an institutional background.

The rest of this paper is structured as follows. Section 2 describes the dataset utilized in the present investigation. Thereafter, Section 3 provides a brief sketch of YOLOv3 algorithm and YOLOv3-tiny algorithm. In Section 4, the experimental platform and parameter settings are discussed. Section 5 presents a comprehensive discussion of the experimental results. Finally, the concluding remarks are presented in Section 6. Comprehensive conclusions have been drawn with directions of future research.

2 Dataset

2.1 Data acquisition

In the current investigation, every individual is considered as a specific class of object. The images were captured using Apple iPhone XR having 12-megapixel camera with an f/1.8 aperture with 1280×960 pixels resolution. All the images were taken under several disturbances like illumination variation, occlusion, and overlap. A total of 900 images were captured and divided into a training set, validation set, and a test set. The training set considered of 720 (80%) images, validation set having 90 (10%) images, and the remaining 90 (10%) images made up the test set. Figure 2 shows some samples from the dataset under different conditions.
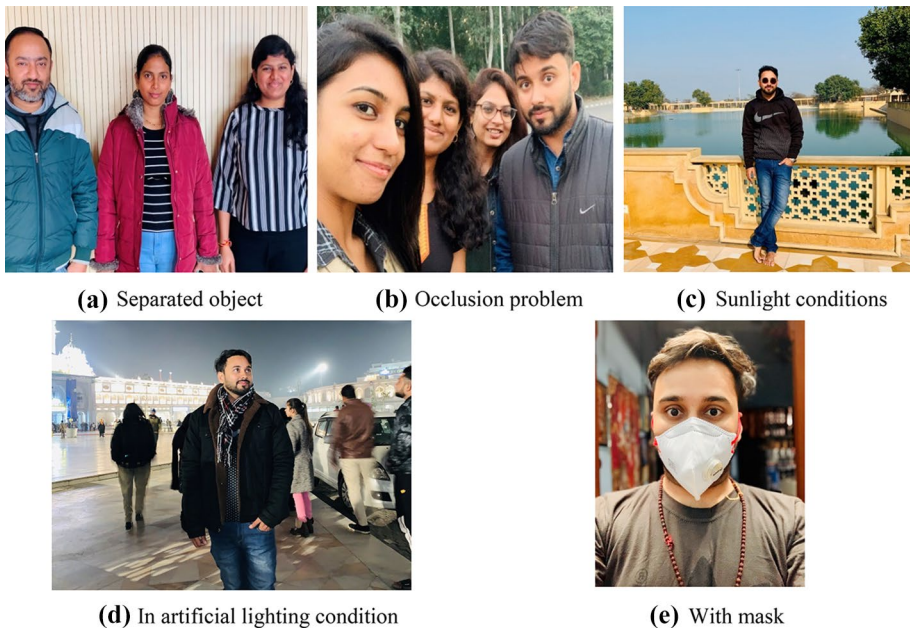

(a) Separated object   **(b)** Occlusion problem   **(c)** Sunlight conditions

(d) In artificial lighting condition   **(e)** With mask

Fig. 2 Samples from the considered dataset

2.2 Image augmentation

To increase the dataset size in the present investigation, the image augmentation technique has been employed. Image augmentation is used to create variants of existing original images and magnify the original dataset. This improves the model performance with improved generalizability, thereby avoiding overfitting. In this work, following augmentation techniques have been utilized to make the dataset: Flipping, Rotation, Cropping, Zoom in/out, Varying brightness or contrast, etc. (Fig. 3).

3 Transfer learning-based object detection

Over the years, YOLO has gained popularity as an efficient deep learning approach for real-time object detection [47]. In general, in the field of computer vision, any object detection problem consists of two tasks: (1) Identification of an object, and (2) Localization (estimating the location of the object in the image). Previous object detection techniques, like CNN and its family, executed object detection task in numerous steps on account of their pipeline architecture for execution. This pipeline architecture results in slow speed and increased complexity in optimization [48]. These limitations of traditional object detection methods are overcome by YOLO which involves transformation of object detection problem into a single regression model. YOLO accomplishes concurrent estimation of several bounding boxes (BBs) along with their class probabilities. In contrast to region-proposal based approaches and sliding window techniques for object detection, YOLO performs training on complete images and therefore, results in an optimized detection performance. Nonetheless, YOLOv3 approach (based on Darknet-53 architecture) has demonstrated its efficient performance in complex object detection on account of its near real-time speed, high generalization skill and high average precision.

3.1 YOLOv3 algorithm

The schematic of YOLOv3 algorithm has been illustrated in Fig. 4. Usually, a single image is extracted by YOLOv3 which after suitable resizing (416×416) is fed to the neural network (NN) of YOLOv3. Generally, the YOLOv3 algorithm initiates with the acquisition of an input image and builds a CNN network to predict a tensor. Further, it performs a linear regression using two fully connected layers to make BB predictions. For final predictions, only the BB predictions with high confidence are considered.
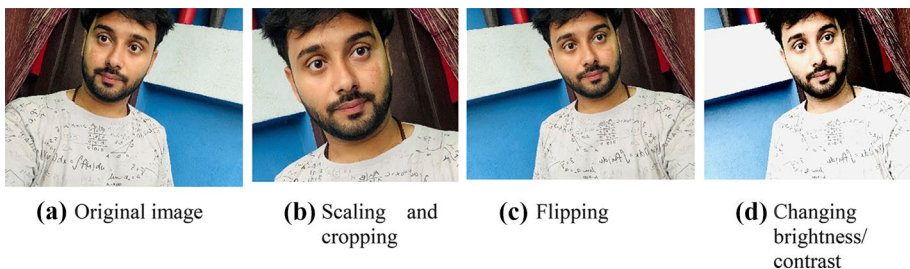

(a) Original image   (b) Scaling and cropping   (c) Flipping   (d) Changing brightness/ contrast

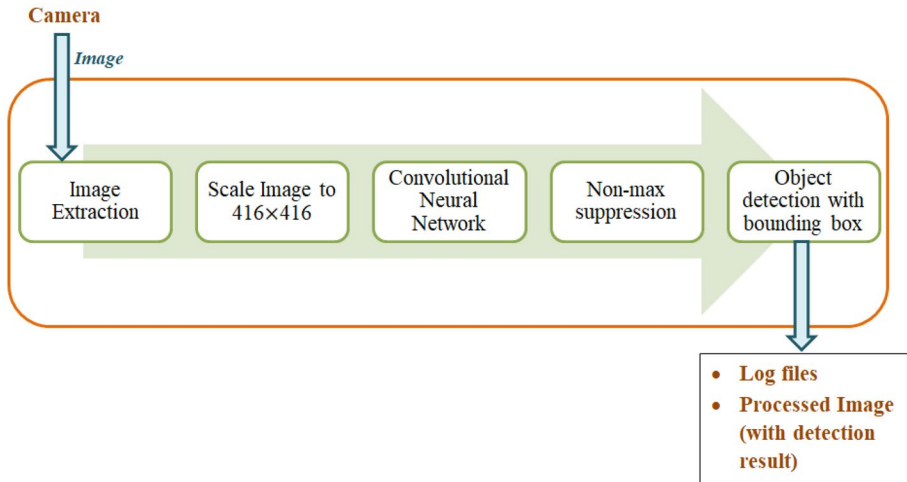**Fig. 3** Some sample images showing augmentation techniques

Fig. 4 Schematic of YOLOv3 algorithm

Thereafter, YOLOv3 produces an output vector with BB properties, prediction probability and class probabilities as the vector elements.

The YOLOv3 algorithm employs NN architecture in the Darknet-53 framework which is diagrammatically shown in Fig. 5. Clearly, the network architecture of YOLOv3 involves several layers including convolutional layers, residual layers and upsampling layers along with shortcut or skip connections. Inclusive information pertaining to YOLOv3 architecture can be found in previous literature [41].

After receiving an input image, the YOLOv3 NN yields an output consisting of various parameters (as shown in Fig. 6). The vector elements, *i.e.,* parameters of the output vector are as follows:

- P_c (Prediction Probability): Probability that each BB contains detectable object.
- BB properties: B_w (Width), B_h (Height) and (B_x, B_y) *i.e.*, Cartesian position (*x* and *y*) of the box inside the image.
- C_1, C_2, C_3, C_4, C_5, C_6 (Class probabilities): Probabilities that each object within its BB is associated with a specific class.
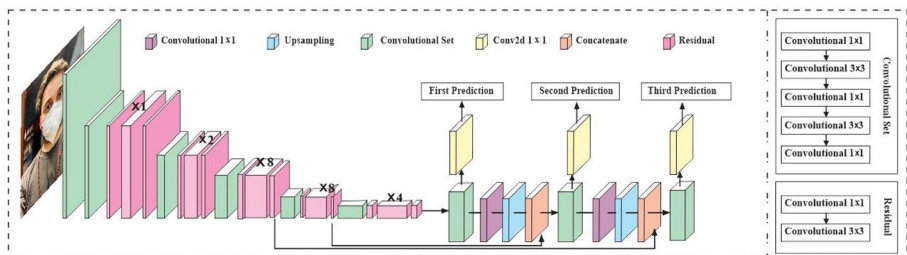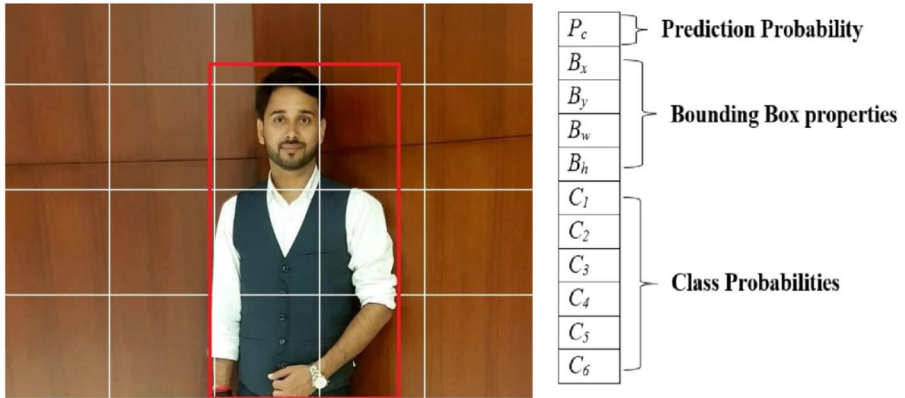

Fig. 5 YOLOv3 architecture

Fig. 6 Bounding box prediction

For predicting the BBs, YOLOv3 make use of the dimension clusters [31] as anchor boxes. For each BB, the NN of YOLOv3 algorithm predicts four coordinates (i.e., t_x, t_y, t_w and t_h) resulting in the BB properties (B_x, B_y, B_w and B_h) as shown in Fig. 7. (C_x, C_y) represent the cell offset from the top left corner of the image while (P_w, P_h) denote the width and height of the BB prior. In addition, if $\hat{t^*}$ and t^* respectively represent the ground truth corresponding to a certain coordinate prediction and the estimated
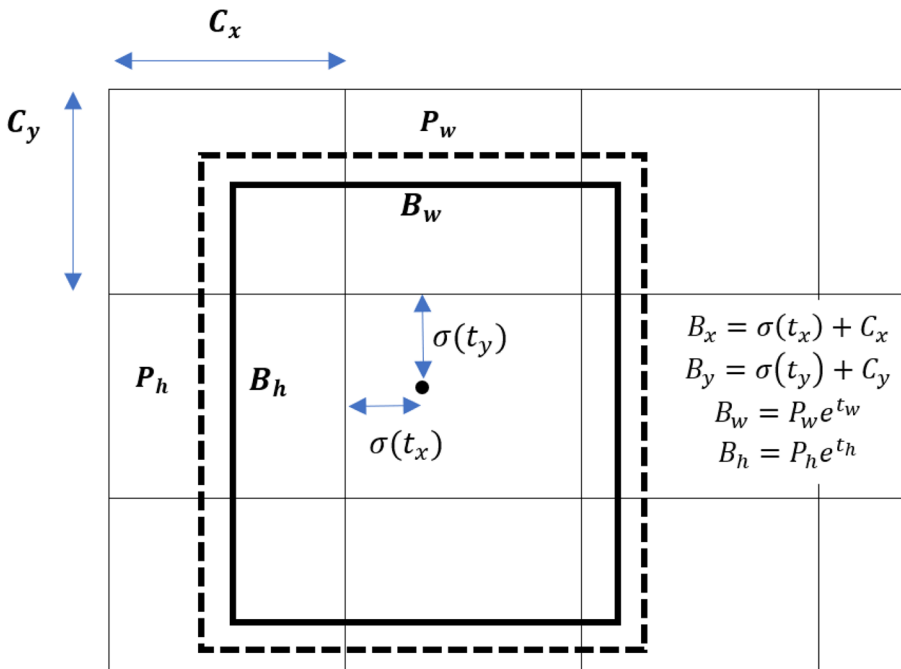

Fig. 7 Bounding box for location prediction

prediction, then their difference serves as the gradient and is expressed as $\widehat{t^*} - t^*$. The equations describing the relationship between the BB coordinates and BB properties (represented in Fig. 7) can be inverted to compute the ground truth value. Moreover, for predicting the score of an object associated with a BB, YOLOv3 utilizes the concept of logistic regression. If the overlapping of the BB prior is the highest among others with reference to the ground truth object, then the object's score becomes 1. Even if the overlapping of the BB priors (except the best one) exceeds the threshold (taken as 0.5 in this study), they are disregarded from prediction. Thus, in YOLOv3, corresponding to each ground truth object, only one BB is allotted.

The pseudo-code of the detection method by YOLOv3 algorithm for the identification of face masked individuals is given below in Algorithm 1. The present investigation uses 9 clusters and 3 scales. The selected clusters are consistently distributed among scales as

$(76 \times 149), (80 \times 268), (125 \times 245), (92 \times 340), (125 \times 352), (187 \times 302), (161 \times 369), (308 \times 358), (1002 \times 1820).$

Algorithm 1: Pseudo-code of YOLOv3

---

begin
Divide the image into S×S grids
while (*i* <= S)
  Predict confidence of BBs for each grid
  Compute the probability of having an object in each BB
  Ignore the BBs with low probabilities
  Estimate good priors // using *k*-means clustering on the training BBs
end while
Run the YOLOv3 on clustering boxes
Determine the prediction of those boxes
Output: {Labeled class}
end

3.1.1 Evaluation metrics

The performance of developed algorithms has been evaluated based on the following metrics: Precision, Recall, IoU, Average Precision (AP), mAP, Loss and F1 Score. More details about these fundamental key values can be found in [27, 41].

3.1.2 Anchor boxes

The detection of several objects whose centers lie on the same grid cell becomes difficult. This difficulty is resolved by assigning an anchor box to each object associated with the same grid. If three anchor boxes are taken, then it will result in three predictions in a single grid cell. The anchor box with the highest IoU is allotted to the corresponding object. If the threshold (taken as 0.5 in this study) is more than the IoU, then the associated object will be disregarded from detection. Clearly, by exploiting the exclusive concept of anchor boxes (demonstrated in Fig. 8), multiple objects in a single grid cell are detected easily by YOLOv3 algorithm.
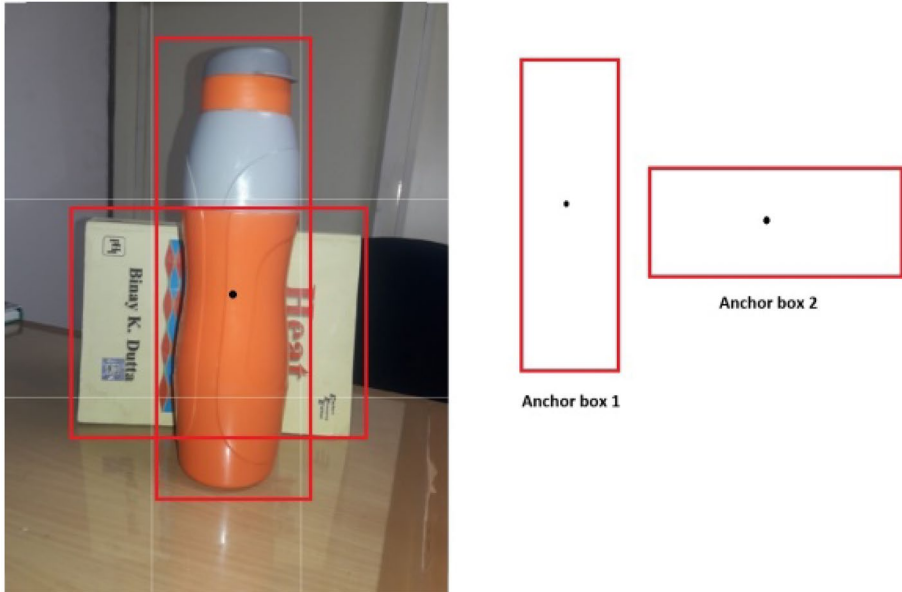
Fig. 8 Illustration of Anchor boxes

3.1.3 Non-max suppression (NMS)

Instead of detecting an object only one time, object detection techniques using YOLO perform multiple detections of the same object, which is an additional difficulty associated with them. This problem is addressed by using non-max suppression (NMS) to remove duplications with lower confidence which helps in one-time detection of an object. The basic concept of NMS algorithm is the sequential comparison of BB having maximum P_c with all other boxes which intersect with it. As illustrated in Fig. 9, all those BBs associated with the distinct object are suppressed which have comparatively
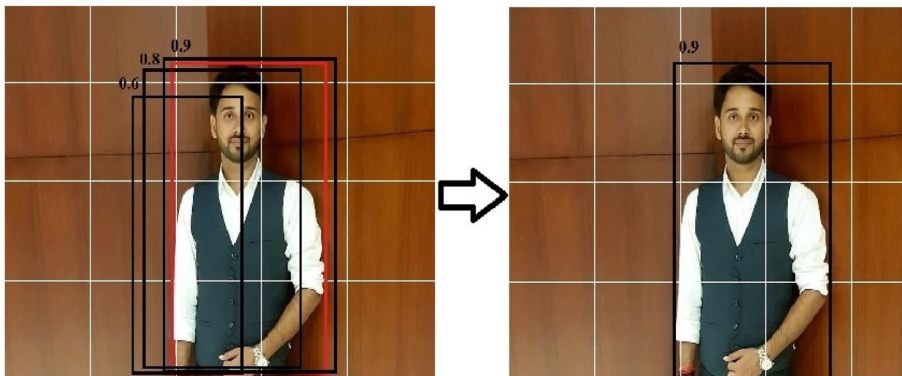

Fig. 9 Non-max suppression for filtering multiple detections

lower P_c, except the one with maximum P_c. Non-maximal suppression adds 2–3% in mAP.

3.2 YOLOv3-tiny

YOLOv3-tiny is a variant of YOLOv3 algorithm with the decreased depth of the convolutional layer [1]. The detailed architecture is shown in Fig. 10. YOLOv3-tiny offers significantly increased running speed (approximately 440% faster than the former variants of YOLO), but with reduced detection accuracy. The Darknet-53 architecture of YOLOv3 employs many 1×1 and 3×3 convolution layers for feature extraction. However, the architecture of YOLOv3-tiny involves pooling layer and less convolution layers. YOLOv3-tiny predicts a 3-D tensor containing confidence score, BB, and class predictions at two different scales. YOLOv3-tiny ignores the BBs with poor confidence score for final detections. For feature extraction, YOLOv3-tiny utilizes a feed forward arrangement of convolution layers and max-pooling layers. BB prediction occurs at two different feature map scales (1 3×13, and 26×26 merged with an upsampled 13×13 feature map).

4 Experimental platform and parameter settings

Experiments are conducted on a PC with Windows 10 Operating System, Intel(R) Core (TM) i7-9700F CPU @ 3.00 GHz, and NVIDIA MSI Gaming GeForce GTX 1650 with 8 GB RAM. Further, this work utilized Visual Studio 2017, and CUDA10.0 and CUDNN7.4 GPU acceleration libraries. For the initial weight assignment, the pre-trained weights for the convolutional layers darknet53.conv.74 for YOLOv3 and yolov3-tiny. conv.15 for YOLOv3-tiny have been employed.

The parameter settings being considered in the training phase are as follows: the momentum is 0.9, decay is set to 0.0005, and the batch size is 64. For both YOLOv3 and YOLOv3-tiny, the training process runs up to 8000 iterations. The learning rate (LR) at the start of the training has been set to 0.001. However, after 6400 and 7200 iterations, the LR has been reduced by a factor of 10. On the aforementioned experimental platform, the training took approximately 26 h and 10 h for YOLOv3 and YOLOv3-tiny, respectively. The methodology used in the training and detection phase, which is self-explanatory, has been illustrated in Fig. 11.

5 Experimental results and discussion

Table 2 elaborate on the detailed quantitative results. The table presents the evaluation metrics (namely mAP, Average IoU, Precision, Recall and F1 score) for YOLOv3 and YOLOv3-tiny algorithms. From Table 2, it is obvious that mAP value of approximately 97% is attained by YOLOv3 after 6000th iteration compared to YOLOv3-tiny which attains a mAP value of 60.12% after the same number of iterations. The mAP settles at a value of 98.73% for YOLOv3 algorithm after the completion of 8000 iterations. On the other hand, the YOLOv3-tiny algorithm does not attain a settled value of mAP even after 8000 iterations and the value in this case is noted to be 61.04%.

For better visualization of results, the trends in the loss function and mAP values are illustrated in Fig. 12a and b for YOLOv3 and YOLOv3-tiny algorithms, respectively. It

is evident that after the training completion (i.e., after 8000 iterations), YOLOv3 and YOLOv3-tiny algorithms show an average loss of 0.5016 and 0.9908, respectively. Also, the training results reveal that the mAP for the YOLOv3 in the custom dataset is 98.73%, while the same value for the YOLOv3-tiny is 61.04%. Clearly, the YOLOv3 offers an improvement of approximately 62% in the best mAP value than YOLOv3-tiny. Figure 13 shows the variations of mAP and IoU during the training. After training, the trained model has been validated via test images. Figure 14 shows some snapshots of the implemented codes during testing.

Figure 15 present a visual comparison between YOLOv3 and YOLOv3-tiny algorithms. Moreover, for quantitative assessment of the simulation results, the performance of YOLOv3 and YOLOv3-tiny algorithms have been compared in Table 3. From Fig. 15 and Table 3, it has been observed that YOLOv3 correctly recognizes all the test with reasonable detection accuracy. On the other hand, YOLOv3-tiny is unable to recognize some of the test images. A comparative assessment indicates that for test images (without mask) 1, 2, 3, 5 and 6 respectively, the prediction probability of YOLOv3 is 51.51%, 2.04%, 4.17%, 14.53%, and 10.59% higher than YOLOv3-tiny. However, YOLOv3-tiny did not detect the test image 4 (without mask) correctly. Moreover, in case of images with mask (for test images 1, 3 and 6), the prediction probability of YOLOv3 is 19.28%, 42.39%, and 175% higher than YOLOv3-tiny. But, YOLOv3-tiny wrongly detects the test images 4 and 5 (with mask) and is unable to detect the test image 2 (with mask) at all.

The comparative analysis in terms of detection time has been tabulated in Table 4. The results confirm the poor computational speed of YOLOv3 as compared to YOLOv3-tiny. However, the significantly enhanced and robust performance of YOLOv3 over YOLOv3-tiny is validated. The detection time of YOLOv3-tiny is noteworthy; however, the detection capability is poor as evident from Table 3. Moreover, Table 4 shows that the detection time required for GPU is around 230 ms (for YOLOv3) and 52 ms (for YOLOv3-tiny). On the other hand, for CPU, the detection time for YOLOv3 and YOLOv3-tiny is around 4 s and 320 ms, respectively.

6 Conclusions

This article presented a maiden attempt towards the recognition of face masked individuals. This work paves way for researchers by proposing a framework to design an effective surveillance system to identify people who are wearing masks. To identify faces with masks, a novel application of YOLOv3 algorithm as a realistic solution to surveillance systems has been presented. The model has been trained on a self-made dataset and the efficiency of YOLOv3 algorithm has been analyzed. The results demonstrated that YOLOv3 outperforms YOLOv3-tiny algorithm for the recognition of face masked individuals. In terms of mAP, YOLOv3 shows a substantial improvement in mAP (approximately 62%) over YOLOv3-tiny algorithm. Further, it is observed that the prediction probability is relatively high for YOLOv3 algorithm. YOLOv3 offered a maximum mAP of nearly 99% which is highly promising. Conclusively, the results clearly endorse the design and implementation of YOLOv3-based smart surveillance schemes for recognition of face masked individuals which is necessary after the outbreak of COVID-19 pandemic. Several research directions can be recommended for
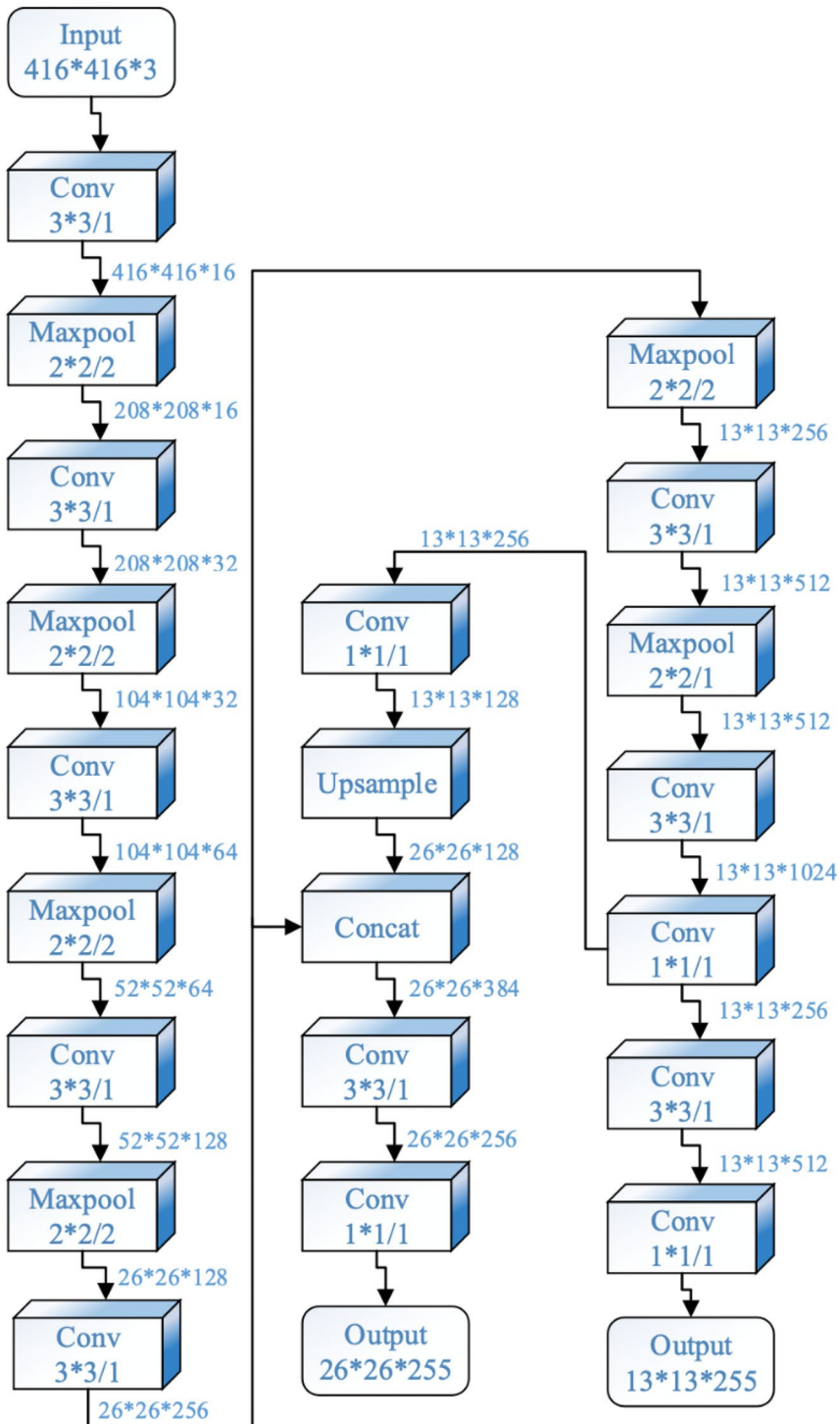
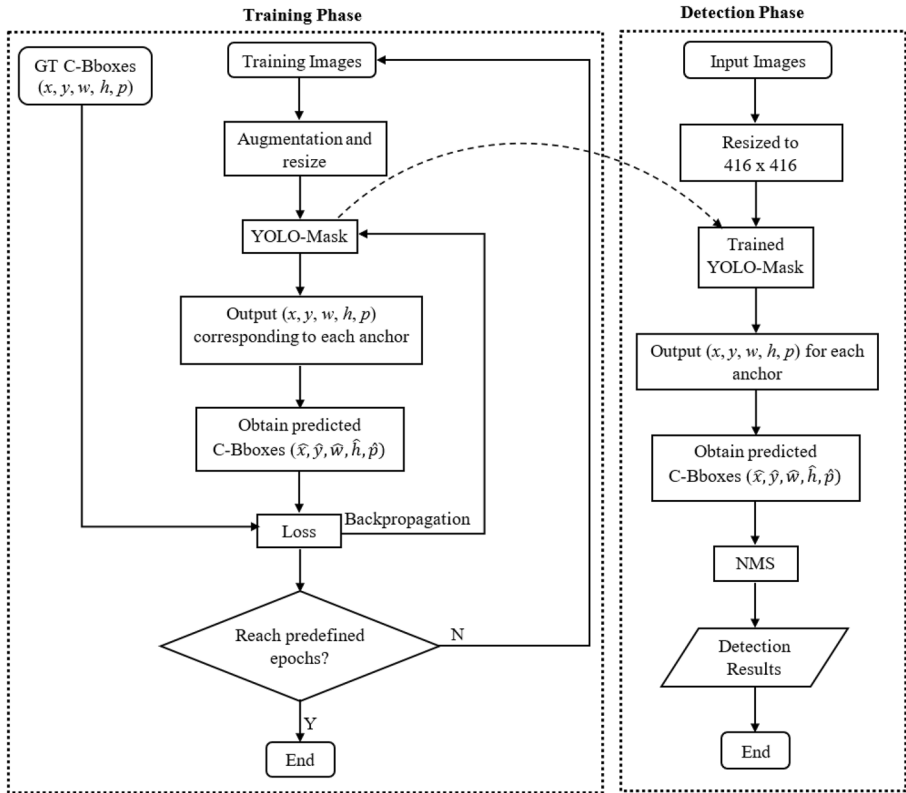**Fig. 10** The architecture of YOLOv3-tiny

Fig. 11 The Training and Detection phases

Table 2 Comparative study of training results of YOLOv3 and YOLOv3-tiny algorithms

Iterations	YOLO version	mAP (%)	Average IoU (%)	Precision	Recall	F1 score
1000	v3	75.04	55.10	0.76	0.29	0.42
	v3-tiny	12.67	17.27	0.27	0.19	0.22
2000	v3	97.03	74.70	0.95	0.80	0.87
	v3-tiny	31.54	27.00	0.40	0.38	0.39
3000	v3	88.58	61.07	0.85	0.90	0.87
	v3-tiny	34.35	27.40	0.40	0.50	0.45
4000	v3	96.92	78.07	0.97	0.88	0.92
	v3-tiny	53.02	55.05	0.75	0.36	0.49
5000	v3	94.35	75.80	0.96	0.75	0.85
	v3-tiny	60.37	57.27	0.77	0.45	0.57
6000	v3	97.74	71.53	0.94	0.97	0.95
	v3-tiny	60.12	58.81	0.78	0.43	0.56
7000	v3	97.77	78.84	0.97	0.94	0.95
	v3-tiny	58.64	50.54	0.69	0.53	0.60
8000	v3	98.73	72.84	0.95	0.97	0.96
	v3-tiny	61.04	54.95	0.74	0.51	0.61

Fig. 12 Trends of loss function and mAP in the training for **(a)** YOLOv3 **(b)** YOLOv3-tiny
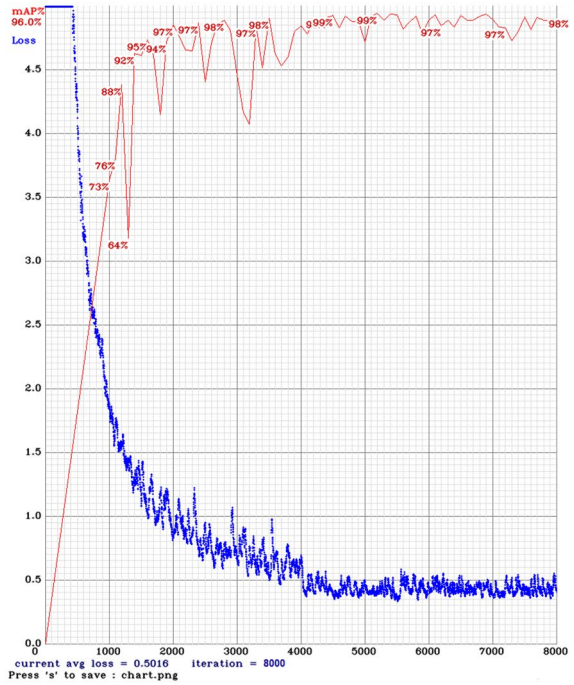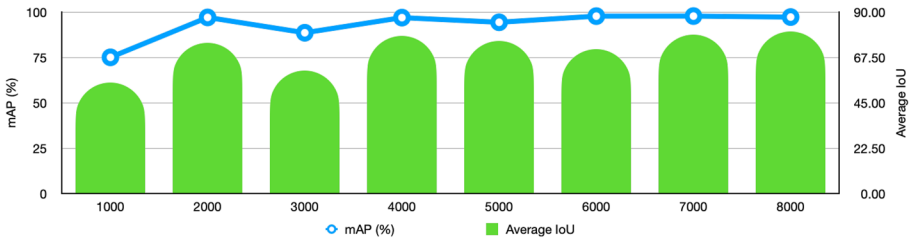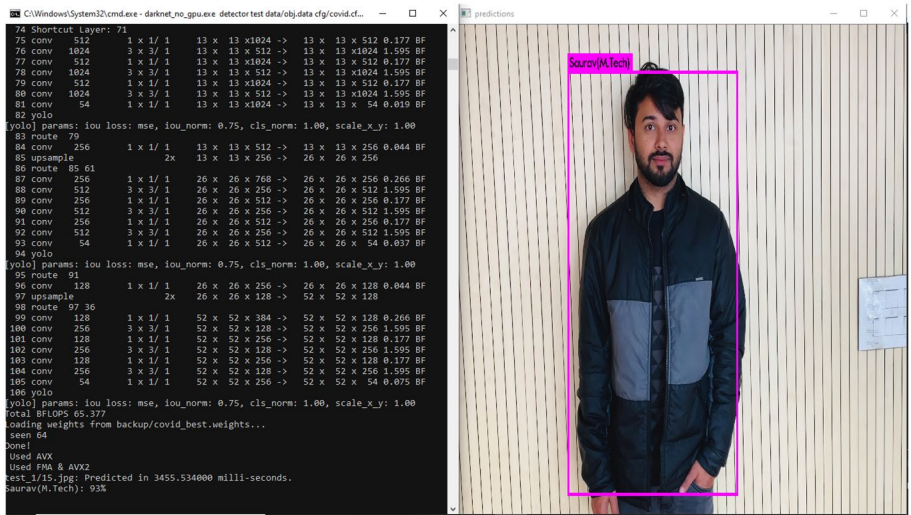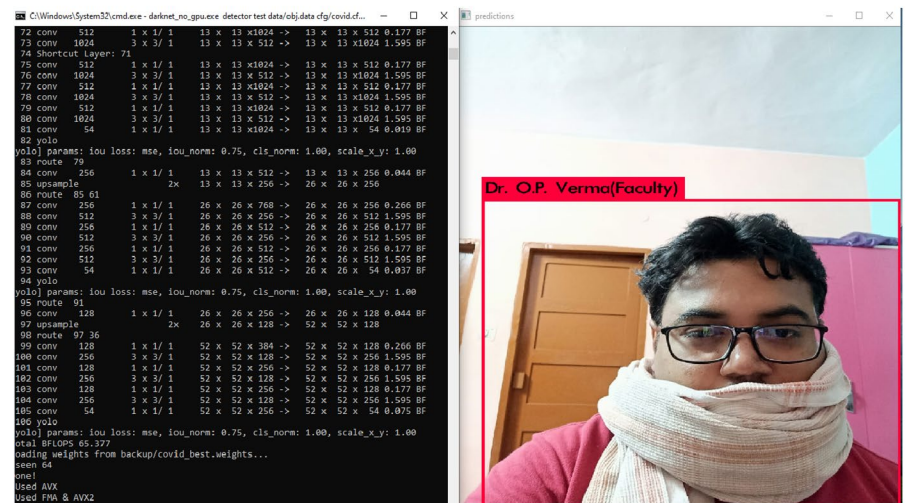

(a)
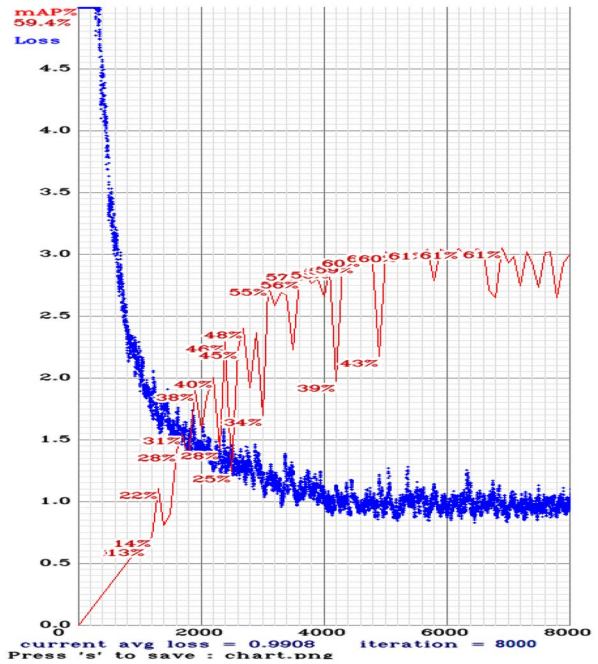
(b)

Fig. 13 mAP and IoU trends over the training phase for YOLOv3 algorithm

(a)

(b)

Fig. 14 Snapshots taken during testing phase

Fig. 15 Experimental results on some sample images **a** Original image **b** Detection results by YOLOv3 **c** Detection results by YOLOv3-tiny

future work. The development of new algorithm with optimized detection time and exceptionally high prediction probability is an interesting research work. Further, the
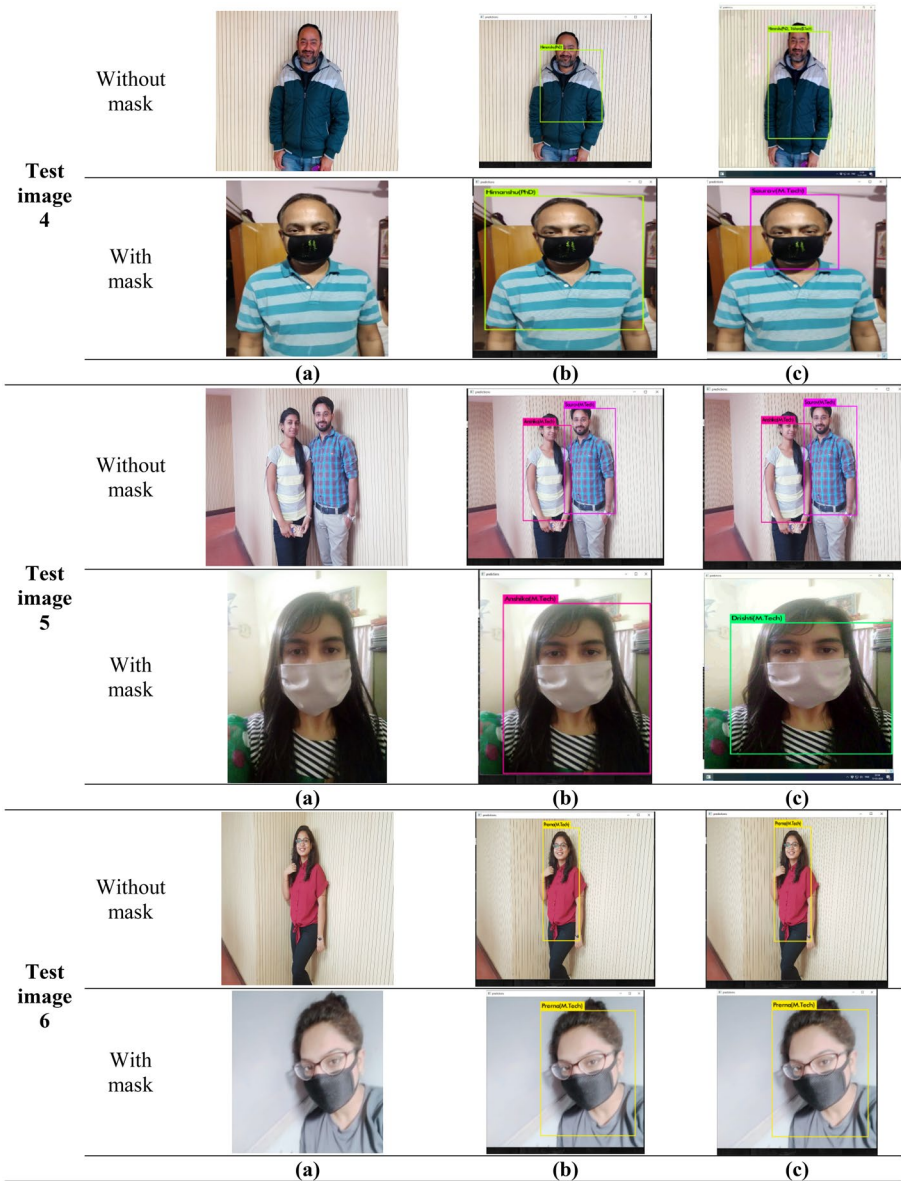
Fig. 15 (continued)

focus of future research would be to investigate the efficiency of the proposed framework on an extensive and diverse dataset to examine its robustness. Also, integration of other machine learning techniques to obtain more realistic results can be addressed in future works.

Table 3 Quantitative comparison of YOLOv3 and YOLOv3-tiny algorithms

Test image		Name	Detection accuracy (in %)	
			YOLOv3	YOLOv3-tiny
1	Without mask	Saurav(M.Tech)	100	66
	With mask	Saurav(M.Tech)	99	83
2	Without mask	Drishti(M.Tech)	100	98
	With mask	Drishti(M.Tech)	99	No detection
3	Without mask	Dr. O.P. Verma(Faculty)	100	96
	With mask	Dr. O.P. Verma(Faculty)	92	53
4	Without mask	Himanshu(PhD)	99	Wrong detection
	With mask	Himanshu(PhD)	98	Wrong detection
5	Without mask	Anshika(M.Tech)	97	89
		Saurav(M.Tech)	98	85
	With mask	Anshika(M.Tech)	99	Wrong detection
6	Without mask	Prerna(M.Tech)	94	85
	With mask	Prerna(M.Tech)	88	32

Table 4 Quantitative comparison of detection time

Test Image		Image size (Pixels)	Prediction time (in milliseconds)			
			Platform 1[a]		Platform 2[b]	
			YOLOv3	YOLOv3-tiny	YOLOv3	YOLOv3-tiny
1	Without mask	744 × 558	224.14	51.51	4381.05	366.90
	With mask	2316 × 2941	261.36	51.84	3890.63	323.59
2	Without mask	3017 × 2264	232.29	51.86	3988.60	311.72
	With mask	960 × 1280	234.75	51.85	4732.87	312.70
3	Without mask	4623 × 3461	261.25	52.31	4554.85	373.05
	With mask	960 × 1280	246.07	51.79	4055.01	368.66
4	Without mask	4032 × 3024	240.91	51.89	4082.82	389.19
	With mask	960 × 1280	236..36	51.80	4435.33	307.05
5	Without mask	3017 × 2264	257.73	51.89	4359.05	323.80
	With mask	960 × 1280	229.57	51.75	4407.87	368.54
6	Without mask	3017 × 2267	228.22	51.90	3903.71	320.87
	With mask	967 × 760	228.88	51.41	4170.70	317.90

[a]Platform 1: the experimental platform specified in Sect. 4
[b]Platform 2: Intel(R) Core (TM) i5-4200U CPU @ 1.60 GHz 2.30 GHz

Declarations

Conflict of interest All the authors declared that they have no conflict of interest.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent to publish The authors affirm that human research participants provided informed consent for the publication of the images in Fig. 2, Fig. 3, Fig. 5, Fig. 6, Fig. 9, Fig. 14, and Fig. 15. Additional informed consent was obtained from all individual participants for whom identifying information is included in this article.

References

1. Adarsh P, Rathi P, Kumar M (2020) YOLO v3-tiny: Object detection and recognition using one stage improved model. In: Proceedings 6th international conference on advanced computing and communication systems, ICACCS 2020 pp 687–694. https://doi.org/10.1109/ICACCS48705.2020.9074315
2. Al-masni MA, Al-antari MA, Park JM et al (2018) Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Comput Methods Program Biomed 157:85–94. https://doi.org/10.1016/j.cmpb.2018.01.017
3. Bu W, Xiao J, Zhou C, et al (2017) A cascade framework for masked face detection. In: Proceedings IEEE international conference on cybernetics and intelligent systems, CIS 2017 and IEEE conference on robotics, automation and mechatronics, RAM 2017. Institute of Electrical and Electronics Engineers Inc., pp 458–462
4. Chavda A, Dsouza J, Badgujar S, Damani A (2021) Multi-stage CNN architecture for face mask detection. In: 6th international conference for convergence in technology, I2CT 2021. Institute of Electrical and Electronics Engineers Inc., pp 1–8
5. Chen Q, Sang L (2018) Face-mask recognition for fraud prevention using Gaussian mixture model. J Vis Commun Image Repres 55:795–801. https://doi.org/10.1016/j.jvcir.2018.08.016
6. Chen C, Liu MY, Tuzel O, Xiao J (2017) R-CNN for small object detection. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp 214–230
7. Cheng Y, Ma N, Witt C et al (2021) Face masks effectively limit the probability of SARS-CoV-2 transmission. Science (New York, NY) 372:1339–1343. https://doi.org/10.1126/SCIENCE.ABG6296
8. Chowdhury AE, Chowdhury OK, Samrat MA, et al (2020) HFDCM: a low-cost machine learning based class attendance monitoring system. Institute of Electrical and Electronics Engineers (IEEE), pp 1348–1353
9. Corovic A, Ilic V, Duric S, Marijan M, Pavković B (2018) The real-time detection of traffic participants using YOLO algorithm. In: Proceedings of 2018 26th Telecommunications Forum, (TELFOR) 2018. Institute of Electrical and Electronics Engineers Inc., pp 1–4. https://doi.org/10.1109/TELFOR.2018.8611986
10. Dollár P, Wojek C, Schiele B, Perona P (2012) Pedestrian detection: an evaluation of the state of the art. IEEE Trans Pattern Anal Mach Intell 34:743–761. https://doi.org/10.1109/TPAMI.2011.155
11. Dundar A, Jin J, Martini B, Culurciello E (2017) Embedded streaming deep neural networks accelerator with applications. IEEE Trans Neural Netw Learn Syst 28:1572–1583. https://doi.org/10.1109/TNNLS.2016.2545298
12. Fang Y, Nie Y, Penny M (2020) Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: a data-driven analysis. J Med Virol 92:645–659. https://doi.org/10.1002/JMV.25750
13. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32:1627–1645. https://doi.org/10.1109/TPAMI.2009.167
14. Feng S, Shen C, Xia N et al (2020) Rational use of face masks in the COVID-19 pandemic. Lancet Respir Med 8:434–436. https://doi.org/10.1016/S2213-2600(20)30134-X
15. Fischer EP, Fischer MC, Grass D et al (2020) Low-cost measurement of face mask efficacy for filtering expelled droplets during speech. Sci Adv. https://doi.org/10.1126/SCIADV.ABD3083
16. Fu C-Y, Liu W, Ranga A, et al (2017) DSSD : deconvolutional single shot detector. arXiv preprint arXiv:1701.06659

17. Ge S, Li J, Ye Q, Luo Z (2017) Detecting masked faces in the wild with LLE-CNNs. In: Proceedings 30th IEEE conference on computer vision and pattern recognition, CVPR 2017 2017-January, pp 426–434. https://doi.org/10.1109/CVPR.2017.53

18. Glumov NI, Kolomiyetz EI, Sergeyev VV (1995) Detection of objects on the image using a sliding window mode. Opt Laser Technol 27:241–249. https://doi.org/10.1016/0030-3992(95)93752-D

19. Gupta H, Verma OP (2021) Monitoring and surveillance of urban road traffic using low altitude drone images: a deep learning approach. Multimed Tools Appl. https://doi.org/10.1007/S11042-021-11146-X

20. Gupta H, Kumar S, Yadav D et al (2021) Data analytics and mathematical modeling for simulating the dynamics of COVID-19 epidemic: a case study of India. Electronics 10:127. https://doi.org/10.3390/electronics10020127

21. Harikrishnan J, Sudarsan A, Sadashiv A, Remya Ajai AS (2019) Vision-face recognition attendance monitoring system for surveillance using deep learning technology and computer vision. In: Proceedings - international conference on vision towards emerging trends in communication and networking (ViTECoN) 2019. Institute of Electrical and Electronics Engineers Inc

22. Hendry CRC (2019) Automatic license plate recognition via sliding-window darknet-YOLO deep learning. Image Vis Comput 87:47–56. https://doi.org/10.1016/j.imavis.2019.04.007

23. Inamdar M, Mehendale N (2020) Real-time face mask identification using facemasknet deep learning network. SSRN Electron J. https://doi.org/10.2139/SSRN.3663305

24. Karthik R, Hariharan M, Anand S et al (2020) Attention embedded residual CNN for disease detection in tomato leaves. Appl Soft Comput J 86:105933. https://doi.org/10.1016/j.asoc.2019.105933

25. Khandelwal P, Khandelwal A, Agarwal S, et al (2020) Using computer vision to enhance safety of workforce in manufacturing in a post COVID world. http://arxiv.org/abs/200505287v2

26. Klompas M, Morris CA, Sinclair J et al (2020) Universal masking in hospitals in the Covid-19 era. N Engl J Med 382:e63. https://doi.org/10.1056/NEJMP2006372

27. Kumar S, Yadav D, Gupta H et al (2020) A novel YOLOv3 algorithm-based deep learning approach for waste segregation: towards smart waste management. Electronics 10:14. https://doi.org/10.3390/electronics10010014

28. Kumar S, Gupta H, Yadav D et al (2021) YOLOv4 algorithm for the real-time detection of fire and personal protective equipments at construction sites. Multimed Tools Appl. https://doi.org/10.1007/S11042-021-11280-6

29. Lan W, Dang J, Wang Y, Wang S (2018) Pedestrian detection based on yolo network model. In: Proceedings of 2018 IEEE International Conference on Mechatronics and Automation, ICMA 2018. Institute of Electrical and Electronics Engineers Inc., pp 1547–1551

30. Leung NHL, Chu DKW, Shiu EYC et al (2020) Respiratory virus shedding in exhaled breath and efficacy of face masks. Nat Med 26:676–680. https://doi.org/10.1038/S41591-020-0843-2

31. Li C, Chang E, Garcia-Molina H, Wiederhold G (2002) Clustering for approximate similarity search high-dimensional spaces. IEEE Trans Knowl Data Eng 14:792–808. https://doi.org/10.1109/TKDE.2002.1019214

32. Liang WH, Huang J, Zhang CJP et al (2020) Facemask shortage and the novel coronavirus disease (COVID-19) outbreak: reflections on public health measures. EClinicalMedicine 21:100329. https://doi.org/10.1016/j.eclinm.2020.100329

33. Liu N, Han J, Zhang D, et al (2015) Predicting eye fixations using convolutional neural networks. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. IEEE Computer Society, pp 362–370

34. Liu W, Anguelov D, Erhan D, et al (2016) SSD: Single shot multibox detector. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Springer Verlag, pp 21–37

35. Loey M, Manogaran G, Taha MHN, Khalifa NEM (2021) Fighting against COVID-19: a novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustain Cities Soc 65:102600. https://doi.org/10.1016/J.SCS.2020.102600

36. Lu S, Wang B, Wang H et al (2019) A real-time object detection algorithm for video. Comput Electr Eng 77:398–408. https://doi.org/10.1016/j.compeleceng.2019.05.009

37. Ma Q, Shan H, Zhang H et al (2020) Potential utilities of mask-wearing and instant hand hygiene for fighting SARS-CoV-2. J Med Virol 92:1567–1571. https://doi.org/10.1002/JMV.25805

38. Mahapatra R, Samanta S, Pal M et al (2021) Colouring of COVID-19 affected region based on fuzzy directed graphs. Comput Mater Contin 68:1219–1233. https://doi.org/10.32604/CMC.2021.015590

39. Mery D, Mackenney I, Villalobos E (2019) Student attendance system in crowded classrooms using a smartphone camera. In: Proceedings - 2019 IEEE Winter Conference on Applications of Computer Vision, WACV 2019. Institute of Electrical and Electronics Engineers Inc., pp 857–866

40. Militante S, Dionisio NV (2020) Real-time facemask recognition with alarm system using deep learning. In: Proceedings of11th IEEE Control and System Graduate Research Colloquium, ICSGRC 2020. pp 106–110. https://doi.org/10.1109/ICSGRC49013.2020.9232610

41. Nie X, Yang M, Liu RW (2019) Deep neural network-based robust ship detection under different weather conditions. In: 2019 IEEE intelligent transportation systems conference, ITSC 2019. Institute of Electrical and Electronics Engineers Inc., pp 47–52

42. Nieto-Rodríguez A, Mucientes M, Brea VM (2015) System for medical mask detection in the operating room through facial attributes. Lecture Notes Comput Sci 9117:138–145. https://doi.org/10.1007/978-3-319-19390-8_16

43. Pathak AR, Pandey M, Rautaray S (2018) Application of deep learning for object detection. Procedia computer science. Elsevier B.V, Amsterdam, pp 1706–1717

44. Qin B, Li D (2020) Identifying Facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. Sensors 20:5236. https://doi.org/10.3390/S20185236

45. Rangel JC, Martínez-Gómez J, Romero-González C et al (2018) Semi-supervised 3D object recognition through CNN labeling. Appl Soft Comput J 65:603–613. https://doi.org/10.1016/j.asoc.2018.02.005

46. Redmon J, Angelova A (2015) Real-time grasp detection using convolutional neural networks. In: Proceedings IEEE international conference on robotics and automation. Institute of Electrical and Electronics Engineers Inc., pp 1316–1322

47. Redmon J, Farhadi A (2018) YOLOv3: an incremental improvement. arXiv preprint arXiv:1804.02767

48. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. IEEE Computer Society, pp 779–788

49. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst 28:91–99

50. Silva SM, Jung CR (2020) Real-time license plate detection and recognition using deep convolutional neural networks. J Vis Commun Image Represent. https://doi.org/10.1016/j.jvcir.2020.102773

51. Sung KK, Poggio T (1998) Example-based learning for view-based human face detection. IEEE Trans Pattern Anal Mach Intell 20:39–51. https://doi.org/10.1109/34.655648

52. Tian Y, Yang G, Wang Z et al (2019) Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Comput Electron Agric 157:417–426. https://doi.org/10.1016/j.compag.2019.01.012

53. Vinitha V, Velantina V (2020) COVID-19 facemask detection with deep learning and computer vision. Int Res J Eng Technol 7:3127

54. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. Proc IEEE Comput Soc Conf Comput Vis Pattern Recogn 1:511–518. https://doi.org/10.1109/CVPR.2001.990517

55. Wang S, Niu L, Li N (2019) Research on image recognition of insulators based on YOLO algorithm. In: 2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings. Institute of Electrical and Electronics Engineers Inc., pp 3871–3874

56. World Health Organization (2020) Advice on the use of masks in the community, during home care and in healthcare settings in the context of the novel coronavirus (2019-nCoV) outbreak: Interim Guidance

57. World Health Organization (2020) Advice on the use of masks in the context of COVID-19: interim guidance-2, No. WHO/2019-nCoV/IPC_Masks/2020.4. https://doi.org/10.1093/jiaa077

58. World Health Organization (2020) WHO Coronavirus Disease (COVID-19) Dashboard | WHO coronavirus disease (COVID-19) Dashboard. https://covid19.who.int/. Accessed 29 Jun 2020

59. World Health Organization (2021) Weekly epidemiological update on COVID-19 - 31 August 2021. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---31-august-2021. Accessed 5 Sep 2021

60. Wu X, Sahoo D, Hoi SCH (2020) Recent advances in deep learning for object detection. Neurocomputing. https://doi.org/10.1016/j.neucom.2020.01.085

61. Yadav S (2020) Deep learning based safe social distancing and face mask detection in public areas for COVID-19 safety guidelines adherence. Int J Res Appl Sci Eng Technol (IJRASET) 8:1368–1375. https://doi.org/10.22214/ijraset.2020.30560

62. Zhao ZQ, Zheng P, Xu ST, Wu X (2019) Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst 30:3212–3232

Authors and Affiliations

**Saurav Kumar[1] · Drishti Yadav[2] · Himanshu Gupta[3] · Mohit Kumar[4] ·
Om Prakash Verma[3]**

Saurav Kumar
saurav_k@ee.iitr.ac.in

Drishti Yadav
drish131196@gmail.com

Himanshu Gupta
guptah.nitj@gmail.com

Mohit Kumar
kumarmohit@nitj.ac.in

1　Department of Electrical Engineering, Indian Institute of Technology, Roorkee, India

2　Faculty of Informatics, Technische Universität Wien, 1040 Vienna, Austria

3　Department of Instrumentation and Control Engineering, Dr. B. R. Ambedkar National Institute of Technology Jalandhar, Jalandhar 144011, India

4　Department of Information & Technology, Dr. B. R. Ambedkar National Institute of Technology Jalandhar, Jalandhar 144011, India