International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Learning Based Real-Time COVID Norms Violation Detection System

Sasithradevi Anbalagan*[1], Saurav Gupta[2], P. Nirmala[3], S. Mohamed Mansoor Roomi[4]

Submitted: 29/01/2022 Accepted : 18/03/2022

Abstract: Corona virus disease-2019 (COVID-2019) has impacted on many social behaviours and has put forth some cautiousness in day-to-today life. Therefore, to remove the barrier of fearful life, it is essential to monitor the preventive guidelines suggested by the world health organization. The very first guideline to be followed is to wear a mask and maintain social distance. In order to implement this in a super populous country like India, the administration used very coercive steps. To aid the administration, this paper provides a simple and easy to implement deep learning technique for the detection and recognition of COVID norm violators. Given an unconstrained/ constrained real-time video, the proposed framework uses YOLOv4 model for person localization, height-width comparison for evaluating social distance, and a customized YOLOv4 model for face mask detection. Once the proposed algorithm localizes the violators, it identifies them using convolutional neural network-based face recognition library. The evaluation metrics on benchmark datasets as well as real-time data are obtained. The proposed framework outperforms existing solutions with mAP (mAP @ 0.50 i.e. Mean Average Precision) of 0.9395 on YOLOv4. Comparison of proposed technique with the existing literature illustrates the better trade-off between accuracy and complexity.

Keywords: *COVID-19, CNN, YOLOv4, social distance analyzer (SDA), height centroid index*

1. Introduction

With the outbreak of novel corona virus disease 2019 (COVID-19), the world has been suffering from valuable losses such as human lives, economy, etc. The count of casualties owing to COVID-19 is 4,742,994 and 4.46 Lakhs throughout the world and India, respectively. The count is persistently increasing. In order to return to a normal life, governments throughout the world are encouraging people to get vaccinated. But nobody knows how effective the vaccine works as of now. Though fully vaccinated people are less likely to be infected, vaccine breakthrough infections are still epidemic. Therefore, there is still a need to follow safety protocols suggested by the world health organization (WHO). It has been pointed out that there are two critical ways of spreading COVID-19 i.e., through respiratory droplets and any kind of physical contact [1]. The close contact can be alleviated by maintaining a social distance of at least six feet among individuals. The probability of being infected by COVID while staying 1 meter from an infected person is about 3 % while less than 1 meter increases the probability to 13%. The farther the individual stood away from others, the lower their spread. Apart from common safety measures like sanitation and handwashing, wearing a face mask and maintaining social distance are the significant preventive ways to reduce the spread of COVID. This article focuses on the development of an integrated and robust multimedia (video) based framework for the detection of COVID norm violators. The detection of violators is essential to curb the pandemic. The latter can be achieved by the incorporation of deep learning strategy which will aid the future multimedia world to flourish in different aspects. The prime goal is to transform multimedia to an actionable intelligence by incorporating situational analysis of video streams and alert the officials and public.

1.1. Literature Review

Research is still ongoing in this field and has gained importance due to the advancements of artificial intelligence in the field of multimedia. Some of the significant works related to this problem are discussed as follows: Techniques on People Localization: Object detectors like – You look only once (YOLO) [2] and single shot detector (SSD) [3] came out to be excellent deep learning approaches for object detection in real- time. SSD is a method for detecting objects using a single deep neural network. On the other hand, YOLO uses architecture having two fully connected layers. Other object detection methods like faster recursive convolutional neural network (RCNN) [4] are mainly classifiers which use regional proposal methods to generate bounding box on images. This method is complex and difficult to optimize and fails to give desired output in real-time compared to YOLO and SSD. YOLO treats and utilizes a regression problem instead of a classification problem and uses a single convolutional neural network. This shows that attaining an optimal [5-7] framework is the solution for the issues in RCNN Techniques on Camera Calibration: In [8], the author suggested a bird-eye view (top-down perspective approach) for calculating the distance between two pedestrians in the interested region. Here, four point input is marked by the user for mapping the region of interest which is then distorted into a square

[1] *Vellore Institute of Technology, Chennai-600127, India
ORCID ID: 0000-0001-5198-6648*
[2] *Vellore Institute of Technology, Chennai-600127, India
ORCID ID:0000-0001-8028-547X*
[3] *Vellore Institute of Technology, Chennai-600127, India
ORCID ID: 0000-0001-8010-7040*
[4] *Thiagarajar College of Engineering, Madurai, India
ORCID ID:0000-0001-5806-278X*
* *Corresponding Author email: sasithradevi.a@vit.ac.in*

shape for bird eye view perspective.

Table 1. Details of existing works

S. No	Methods		Dataset	Evaluation metrics	References
1	People Localization	YOLOv3	COCO	Average Precision-96.56%	1
		SSD	PASCAL VOC 2007	Average Precision-74.3%	
		RCNN	PASCAL VOC 2007	mAP -73.2%	
2	Camera Calibration	top-down perspective approach	nuScenes	IoU-92.4%	2
		triangular similarity	Real time dataset	RMSE-10.32	
3	Face Recognition	Local Linear Embedding	MAFA	Average Precision-76.4%	3
		Adversarial occlusion aware face detector	MAFA	Average Precision-81.3%	
		Google Brains	MAFA	Accuracy-99.86%	

Also, in [9], the authors used a distance tracking algorithm (triangular similarity), where the distance between camera and image is calculated by using focal length of the camera, width of the image and pixel of the image. Another approach [18] using the lens equation is used to calculate social distance between two pedestrians. The sensor's dimensions and focal length of the camera have to be predefined in order to proceed with the mathematical calculations. Several works have been proposed so far to detect the face masks like [21], [22]. Techniques on face recognition: In [10], the authors proposed the MAFA dataset, yet additionally use the utilization of locally linear embedding (LLE) CNNs. This proved to be an advanced model as it beats the previous existing models by accomplishing an Average Precision (AP) of 76.4 on the MAFA test set for face recognition. In [11], the authors proposed the idea of detecting occluded faces using adversarial occlusion aware face detector (AOFD) and it achieves the average precision (AP) of 81.3 on the MAFA dataset and recall of 97.88 on FDDB (Face Detection Data Set and Benchmark). In [12] 2015, the authors proposed a very geometry group (VGG)-face model based on the same VGG16 architecture. It comprises various architectures ranging from VGGFace1 and VGGFace2 developed by researchers at Oxford. They depict the way towards preparing a face classifier that utilizes a softmax activation function in the output layer to order faces as humans. This is accomplished using triplet loss which uses fine-tuning of the model where the Euclidean distance between vectors created for a similar identity is made more modest. Also, in [13], the authors proposed FaceNet which uses 128 vectors embedding for the extraction of facial features. It is a one-shot learning model, preferably used where there is a shortage of data. In 2017, [14] the team at google brains proposed the idea of face recognition in the form of a python library. It is easier to use and implement because of its simplest architecture. The authors use various regularization techniques to improve the model. The library out of 1400 images fetched to it, correctly predicted 1398 faces having an accuracy of 99.86 which is significantly high. Also, one of the very important features of the library is that it does not need a large number of images to be trained on for registering person. The examination proposed by Khandelwal et al. [15] was centered on utilizing vision-based object recognition models to screen covered countenances and social distance monitoring infringement utilizing film from reconnaissance cameras. This arrangement is explicitly implied for manufacturing plant arrangements. A two-phase arrangement was carried out for distinguishing veiled appearances (mask faces). acquired images will undergo face identification model utilizing a MobileNetV2 model [16]. The faces obtained are then classified to wearing mask or not wearing mask using a binary mask classifier. For social distance monitoring, the authors use SSD [17] for the recognition of individual class. The authors have carried out a technique of choosing 4 points that form a rectangle and have performed perspective transformations so that the given distances can be estimated on a solitary plane. The examination of these distances against the limit requires the total distance between 2 points to be given. This model is suitable for a production line or a shut room, however for each open region, each street, this would be expensive, and tedious. It should be noticed that these two models are independent entities and a coordinated arrangement has not been introduced. Table 1 shows the additional details about the existing works. The main contributions of this work are

– Localizing the individuals in the video stream for detecting the violators.

– Evaluating the social distance between two individuals using HCI

– Verifying the person for wearing mask using customized YOLOv4 model

– Identifying the unmasked person and informing the authorities – monitoring real-time data of COVID norms violators in the dashboard.

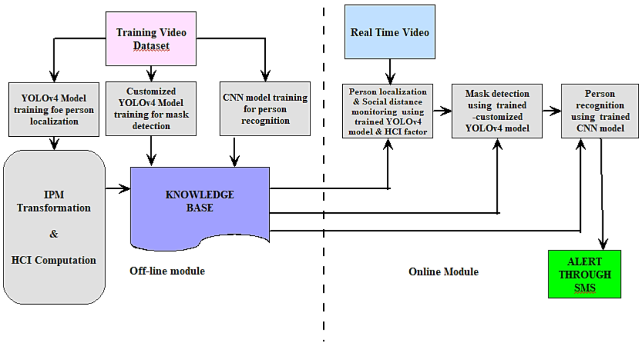

Training Video Dataset
Real Time Video
YOLOv4 Model training foe person localization
Customized YOLOv4 Model training for mask detection
CNN model training for person recognition
Person localization & Social distance monitoring using trained YOLOv4 model & HCI factor
Mask detection using trained -customized YOLOv4 model
Person recognition using trained CNN model
IPM Transformation & HCI Computation
KNOWLEDGE BASE
Off-line module
Online Module
ALERT THROUGH SMS

Figure 1. Proposed framework for COVID norms monitoring

2. Design Approach and Methodology

This Section elaborates the framework for the proposed technique to detect the COVID norms violators.

2.1. Design Approach

Fig. 1 shows the block diagram for COVID norms monitoring in public/crowd/ corporate sectors. The proposed system comprises (i) surveillance camera, (ii) person localization module, (iii) social distance analyzer, (iv) mask detector, and (v) person recognizer. Surveillance camera captures the real time raw (video) data that is being sent to the person localization module. The later use the YOLOv4 model to localize the person in the real-time videos. Once the persons are localized, the social distance among them is computed using social distance analyzer (SDA). It consists of

Inverse Perspective Mapping (IPM) and height-centroid index (HCI) to perform camera calibration and social distance calculation, respectively. If the SDA module identifies any person as not following the social distance, customized YOLOv4 mask detection technique inspects the person for wearing mask. To penalize the violator, a CNN-based face recognition technique is employed in the proposed design approach. The pipeline is evaluated on Benchmarks like COCO, MAFA, and YALE dataset. The metrics like precision, recall and F1score are used to validate the pipeline.

2.2. Hypothesis

In the proposed method, the assumption is that the data available in the video stream contains both violators and law-abiders. The proposed system can be mathematically represented as equation 1,

$$V(x,y,t) = \alpha T(x,y,t) + \beta T'(x,y,t) + \varepsilon \qquad (1)$$

Where $V(x,y,t)$ represents the video stream data, $T(x,y,t)$ and $T'(x,y,t)$ denotes violators and law abiders in the video sequence respectively, α and β are the corresponding weight factors. Variable ε represents the background of the video stream which is not essential for the proposed methodology. The higher value of α represents the highest probability of occurrence of violators in the video stream. The null hypothesis of the proposed framework denotes that the video stream contains only law abiders and can be written as eq.2. Here, H0: hypothesis does not require any further actions since it depicts no risk factor.

$$H0{:}\, \widetilde{V}\,(x,y,t) = \alpha_0\, \widetilde{T}(x,y,t) + \beta_0\, \widetilde{T}'(x,y,t) + \tilde{\varepsilon}. \qquad (2)$$

The alternate Hypothesis H1: and H2: of the proposed framework can be represented mathematically as

$$H1{:}\, \widetilde{V}\,(x,y,t) = \alpha_1\, \widetilde{T}(x,y,t) + \beta_1\, \widetilde{T}'(x,y,t) + \tilde{\varepsilon} \qquad (3)$$

$$H2{:}\, \widetilde{V}\,(x,y,t) = \beta_2\, \widetilde{T}'(x,y,t) + \tilde{\varepsilon} \qquad (4)$$

where $H1$ and $H2$ in eqs. 3 and 4 respectively illustrates the hypothesis for low and high-risk factor. Consider the training and testing domain as T_{rd} and T'_{rd} respectively. $V \in T_{rd}$ where V is a set of video data. V is projected on the defined hypothesis ($H0$, $H1$, and $H2$) to detect the violators in the defined training domain T_{rd}. In the testing domain T'_{rd}, the hypothesis is generated and analyzed to categorize the scenario in the video stream as no risk, low risk or very risk. Therefore, the prime goal of the proposed framework is to localize and identify the violators and inform their details to the authorities for necessary action.

2.3. Methodology:

In this work, a complete solution starting from person localization to violators recognition has been developed. It employs the renowned YOLOv4 model, HCI calculator, customized YOLOv4 model, and CNN-based face recognizer. The integration of all four modules contributes towards the accomplishment of the proposed work. Description of all the individual modules is as follows.

2.3.1. YOLOv4 model for people localization

YOLOv4 is an enhanced version of YOLOv3 [17] as it has reported improved mean average precision (mAP) [15]. The architecture is trained for two classes namely Person and objects. The architecture of YOLOv4 model has three blocks namely backbone, neck, and head are the baseline. Cross stage partial Darknet53 (CSPD53) acts as a backbone for YOLOv4, this uses 53 neural network layers. Spatial pyramid pooling (SPP) is used as a neck of YOLOv4, which separates out vital features and increases the receptive field. Path aggregation Network (PANet) is used for feature aggregation. The head of YOLOv4 is the same as YOLOv3 and it is the prime block for performing localization of people in the video stream using feature mapping. All the three blocks combined are responsible for the localization of people in the video stream.

2.3.2. Social Distance Monitoring using HCI calculator

IPM Transformation: To find the distance between two persons in the video, camera calibration is highly needed. A CCTV camera installed in the monitoring region fetches the video which converts the three dimensional (3D) coordinates to two dimensional (2D) frames which suffers from perspective effect [18].Thus the camera is calibrated using inverse perspective mapping transformation. The latter is used in IPM to transform 2D points (x, y) to 3D parameters using the equation 5,

$$[x\ y\ 1]^T = K\ R\ T\ [X\ Y\ Z\ 1]^T \qquad (5)$$

Where $K, R\ and\ T$ are camera parameters, rotation matrix and translation matrix respectively.

2.3.3. HCI (γ) Computation:

Once the person is localized using YOLOv4 model, the proposed method computes the social distance between them. This method first localizes the person in the frame and marks them with bounding boxes. Their heights are represented by h1 and h2 (assume there are two persons in the scene). The Euclidean centroid distance D is computed between the two bounding boxes. The HCI-based risk factor [20] computation is described in Algorithm 1. The process flow of the proposed framework is well depicted in Fig.2. The complexity is reduced based on the basic idea that the spread of COVID is not accountable if social distance is maintained. The complexity of the overall framework is reduced by running mask detection algorithm only on the bounding boxes that violates social Customized YOLOv4 Model for Mask Detection. The complexity of the algorithm has been scaled down by running the customized YOLOv4 model only on positive samples of social distance analyzer (persons violating social distance). To achieve the defined goal of mask detection, custom model is derived from early YOLOv4 model.

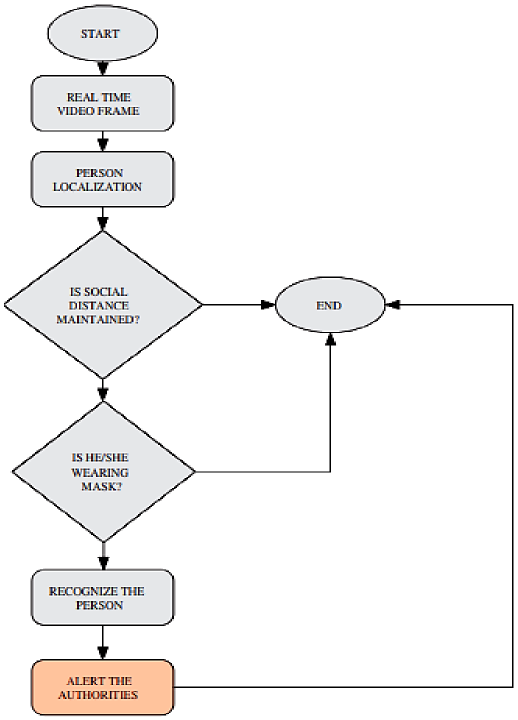


Figure 2. Proposed workflow for COVID norms monitoring

Here, a modified 6 fully connected layer is introduced at the end of the hidden layer. The empirical values for the hyper parameters (trail and error) used in the customized model are set as listed in Table 2. The model trained with the samples of two varieties namely

mask and no mask. Therefore, YOLOv4 model is customized in a sense that it can discriminate persons wearing masks from the violators distancing guideline

Algorithm 1: Social Distance Monitoring

```
Input: Position of two bounding boxes.
Output: → γ and Risk factor
1: Compute h1 and h2 using equation 8.
2: Compute ED between two bounding boxes.
3: Compute γ=minimum(h1, h2)
if ED ⩽ γ/2
    RiskFactor ← High-Risk
elseif ED > γ/2 || ED < γ
    RiskFactor← Low-Risk
else RiskFactor←No-Risk
end
```

Table 2. Hyper-parameter settings for custom YOLOv4 model

HYPERPARAMETER	EMPIRICAL VALUE
Learning Rate	0.001
Batch	64
Subdivisions	32
Steps	4800
Max Batches	6000
Epochs	548
Momentum	0.9
Decay	0.0006

2.3.4. Face recognition of violators using CNN

After running the data over mask detection algorithm, a face recognition algorithm needs to be implemented to find the violators. The aim of this module is to recognize the person detail who have violated the COVID norm guidelines. The latter can be achieved by running a CNN based face recognition algorithm. This algorithm is simple in terms of computation and power [16, 19]. The working architecture of CNN consists of three pairs of convolutional and pooling layers which forms the building block of the architecture. The last layer is the fully connected layer which recognizes the violators using softmax function [20].

3. Simulation results and discussion

This section elaborates the simulation results on the proposed framework using benchmark and real time data. The pipeline of proposed work is trained using different benchmark like MS-COCO [23], MAFA [24], YALE dataset [25] and tested on both offline and online videos. We have used Python as the programming platform to carry-out simulations. YOLOv4 (pre-trained and custom trained both) models run in concatenated fashion. The unprocessed input frame (unconstrained) given as input to YOLOv4 model for person localization. Pre-trained YOLOv4 was used for people (human) detection which is the first step of modules in four folds, followed by social distance analyzer. Fig. 3 depicts the output of the social distance analyzer module (IPM transformation + HCI computation). The parameter namely Angle factor has to be pre-defined for IPM transformation. Angle factor specifies the angle at which the camera is pointing the ground plane with respect to the vertical plane and it ranges between 0 and 1. It reaches the maximum value 1 when the camera becomes vertical. The camera was calibrated at 3m high and angle factor was kept at 0.5 (trial and error) in our case. In this framework, the threshold γ of HCI is chosen between 0.7 and 1.3 (trial and error) Fig. 4 shows the visual results in unconstrained offline video proving the astounding performance of the proposed framework. Our pipeline works well in detecting the person not wearing mask. To reduce the overall complexity, the mask detection model was applied only where social distance is not maintained. The customized YOLOv4 model was trained and the loss after training YOLOv4 was 0.3. The proposed algorithm is also tested on unconstrained real time videos (with authors in the scene) at different instances, Our HCI module is able to compute the social distance accurately. The Green colour window is used for indicating the positive scenario for maintaining social distance whereas the orange and red colour window shows the medium and high risk instances (Fig.4). Fig.5 - Fig.7 shows the simulated visual results of the proposed framework on real time videos. Fig.7. shows the performance of proposed framework in real time indoor scenario. Once the social distance is analyzed, the mask detection model was trained on both YOLOv4 and YOLOv3. The loss after training was 0.355 for YOLOv4 model and 1.19 for YOLOv3 model. Table 3 shows the mAP value for YOLOv3 and YOLOv4.After training process, we achieved shows the performance of proposed framework in real time indoor scenario.

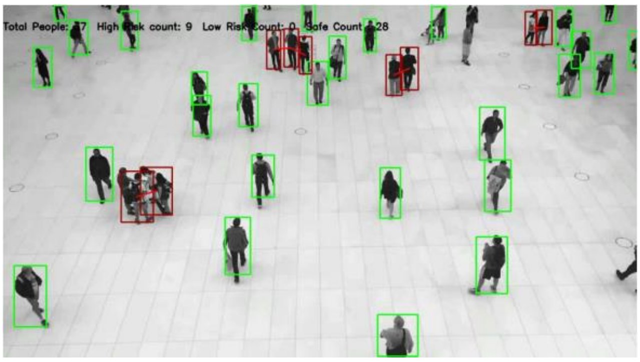


Figure 3. Output of people localizer and social distance analyser

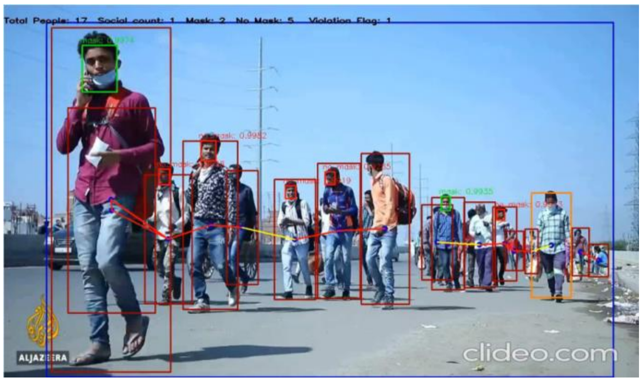

Total People: 17 Social count: 1 Mask: 2 No Mask: 5 Violation Flag: 1
ALJAZEERA
clideo.com

Figure 4. Performance of YOLOV4 and Social distance analyzer in unconstrained offline video

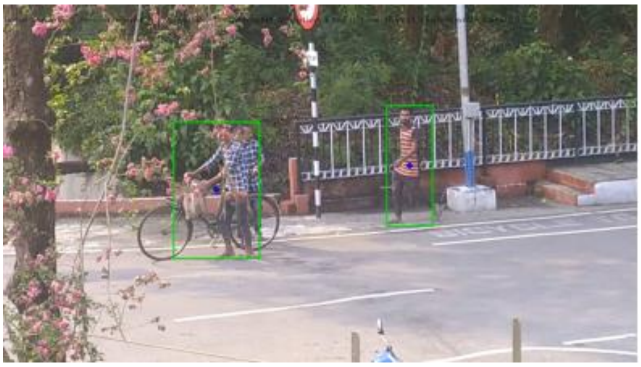


Figure 5. Performance of YOLOv4 and Social distance analyzer in unconstrained real time video

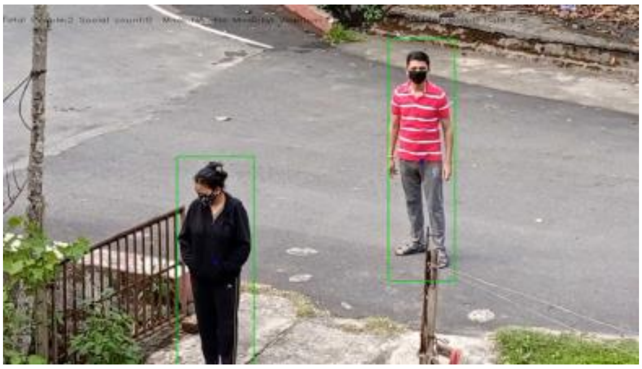


Figure 6. Sample of performance of proposed framework in outdoor environment at instance-1.

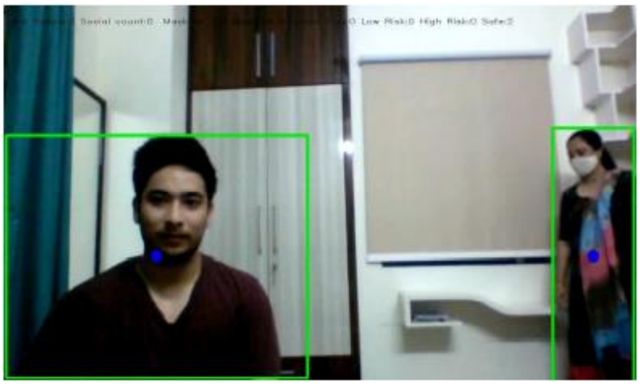


Figure 7. Sample of performance of proposed framework in outdoor environment at instance-2

Table 3. Performance Comparison of Mask Detection Model

Metrics	**Method**	**Train**	**Validate**	**Test**
Precision	YOLOv3	0.89	0.91	0.92
	YOLOv4	0.98	0.88	0.92
Recall	YOLOv3	1	0.92	0.91
	YOLOv4	0.99	0.90	0.95
F1-Score	YOLOv3	0.99	0.91	0.91
	YOLOv4	0.99	0.89	0.93
Average-.Intersection Over Union(%)	YOLOv3	86.08	70.59	73.65
	YOLOv4	87.55	69.37	73.90
Average Precision-Mask(%)	YOLOv3	99.82	93.48	96.56
	YOLOv4	99.57	93.38	97.8
Average Precision -No mask(%)	YOLOv3	99.69	80.84	83.8
	YOLOv4	99.73	83.39	90.1

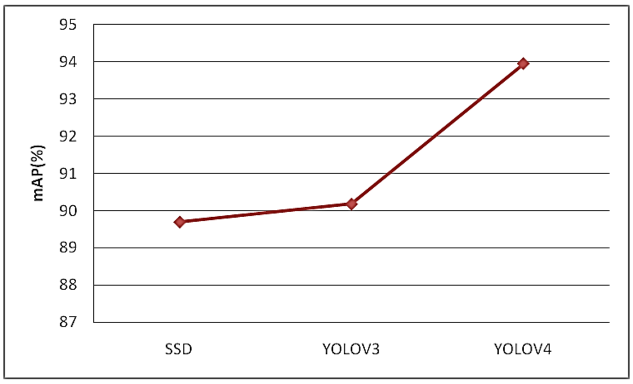
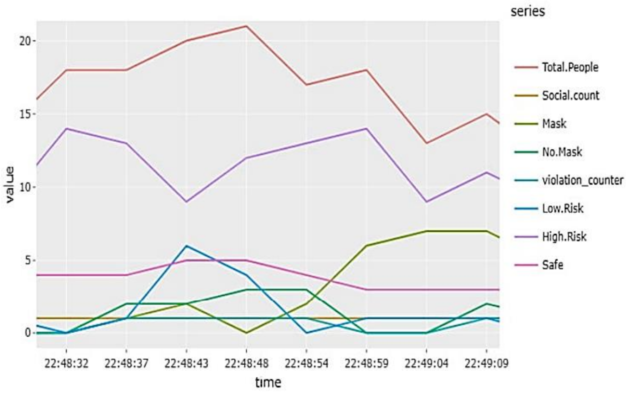


Figure 8. Comparison of mAP

Once the social distance is analyzed, the mask detection model was trained both on YOLOv4 and YOLOv3. The loss after training was 0.355 for YOLOv4 model and 1.19 for YOLOv3 model. Table 2 shows the mAP value for YOLOv3 and YOLOv4. After training process, we achieved the mAP of 0.9018 and 0.9395 for YOLOv3 and YOLOv4 models respectively (Metric is mAP@0.5 i.e Mean Average Precision) which is significantly accurate as compared to 0.897 mAP@ 0.50 intersection over union (IoU) by khandelwal. Comparison of mAP values of proposed framework with the existing solutions is shown in Fig. 8. Corporate Social Responsibility Dashboard (CSRD) is a dashboard used to analyze real-time data generated by back end deep learning models of social distancing, mask detection and face recognition. CSRD also includes features like alerting people(employees in companies) via SMS and email to different regions. The dashboard analysis is shown in Fig. 12 These screenshots shows the effectiveness both in indoor and outdoor environment. The video frames that were analyzed are shown along with the bounding boxes in different colors. It is evident that our algorithm detects the violators and can be visualized in red color bounding boxes. Besides the outstanding performance of our algorithm in detecting the violators, it has its own limitations too. One of the major drawbacks is the night scenario that has many lighting and illumination issues. Also, localizing people during night mode is not analyzed, that will be a fine extension for future work. If the height and width of the bounding box of person is not comparable then it is degrading mAP.

Figure 9. Dashboard displaying violators' information



4. Conclusion

An efficient framework for the COVID norms violators' detection system is proposed in this article. Given a constrained/unconstrained video, the proposed framework detects the violators based on IPM transformation, HCI factor and inspects the person for wearing a mask using customized YOLOv4 model. Once it detects the violator, the face recognition module identifies the person. In addition to that, the proposed framework aids the user to get information about the distribution of violators in a particular place. Hence, the system is suitable for the unconstrained area and can be integrated on convoluted system where CCTV is placed at a specific stature. The proposed deep learning-based framework outperforms the existing works in detecting the violators of COVID norms. In future, the violator details can be communicated to the law enforcement personnel using communication protocols. Also, parallel algorithms will be included to reduce the complexity introduced due to the use of YOLOv4 for people localization and mask detection.

References

[1] D. Yang, E. Yurtsever, V. Renganathan, K. A. Redmill, and U.Ozguner, "A vision-based social distancing and critical density detection system for covid19," *Sensors*, vol. 21, no. 13, pp. 4608, 2021.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You look only once: unified real-time object detection," *arXiv preprint arXiv:1506.02640*, 2015.

[3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," *in European conference on computer vision*. Springer, pp. 21–37, 2016.

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, pp. 91–99, 2015.

[5] S. Gupta, A. K. Sahoo, and U. K. Sahoo, "Wireless sensor network-based distributed approach to identify spatio-temporal volterra model for industrial distributed parameter systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7671–7681, 2020.

[6] Sasithradevi and S. M. M. Roomi, "Video classification and retrieval through spatio-temporal radon features," *Pattern recognition*, vol. 99, pp. 107099, 2020.

[7] R. Kumar, A. Kumar, and G. K. Singh, "Electrocardiogram signal compression using singular coefficient truncation and wavelet coefficient coding," *IET Science, Measurement & Technology*, vol. 10, no. 4, pp. 266–274, 2016.

[8] A. Loukkal, Y. Grandvalet, T. Drummond, and Y. Li, "Driving among flatmobiles: Bird-eye-view occupancy grids from a monocular camera for holistic trajectory planning," *in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 51–60, 2021.

[9] Xu, Jiaojiao and Gong, Chen and Xu, Zhengyuan, "Experimental indoor visible light positioning systems with centimeter accuracy based on a commercial smartphone camera", *IEEE Photonics Journal*, Vol.10, No.6, pp.1-17, 2018.

[10] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting masked faces in the wild with lle-cnns," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2682–2690, 2017.

[11] Y. Chen, L. Song, Y. Hu, and R. He, "Adversarial occlusion-aware face detection," *in IEEE 9th International Conference on Biometrics Theory*, Applications and Systems (BTAS). IEEE, pp. 1–9, 2018.

[12] Alonso-Fernandez, Fernando and Diaz, Kevin Hernandez and Ramis, Silvia and Perales, Francisco J and Bigun, Josef, "Soft-biometrics estimation in the era of facial masks", *2020 International Conference of the Biometrics Special Interest Group* (BIOSIG), pp.1-6, 2020.

[13] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, 2015.

[14] L. Blanger and A. Panisson, "A face recognition library using convolutional neural networks," *International Journal of Engineering Research and Science*, vol. 3, no. 8, pp. 84–92, 2017.

[15] P. Khandelwal, A. Khandelwal, S. Agarwal, D. Thomas, N. Xavier, and A. Raghuraman, "*Using computer vision to enhance safety of workforce in manufacturing in a post covid world*," arXiv preprint arXiv:2005.05287, 2020.

[16] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[17] N. S. Punn, S. K. Sonbhadra, S. Agarwal, and G. Rai, "Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques," *arXiv preprint arXiv:2005.01385*, 2020.

[18] O. Arandjelovi´c, D.-S. Pham, and S. Venkatesh, "Cctv scene perspective distortion estimation from low-level motion features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 939–949, 2015.

[19] D. T. Nguyen, T. N. Nguyen, H. Kim, and H.-J. Lee, "A high-throughput and power-efficient fpga implementation of yolo cnn for object detection," *IEEE Transactions on Very Large-Scale Integration (VLSI) Systems*, vol. 27, no. 8, pp. 1861–1873, 2019.

[20] Velazquez-Pupo, R.; Sierra-Romero, A.; Torres-Roman, D.; Shkvarko, Y.V.; Santiago-Paz, J.; Gómez-Gutiérrez, D.; Robles-Valdez, D.; Hermosillo-Reynoso, F.; Romero-Delgado, M. "Vehicle Detection with Occlusion Handling, Tracking, and OC-SVM Classification: A High-Performance Vision-Based System". *Sensors* 2018, *18*, 374. https://doi.org/10.3390/s18020374

[21] Koklu M, Cinar I, Taspinar YS. "CNN-based bi-directional and directional long-short term memory network for determination of face mask", *Biomed Signal Process Control*, Vol. 71, pp. 103216, 2022.

[22] Cabani A, Hammoudi K, Benhabiles H, Melkemi M., "MaskedFace-Net - A dataset of correctly/incorrectly masked face images in the context of COVID-19", *Smart Health*, Vol.19, pp.100144,2021.

[23] https://cocodataset.org/, last access: 4.01.2022

[24] https://www.kaggle.com/rahulmangalampalli/mafa-data, last access: 4.01.2022

[25] http://vision.ucsd.edu/content/yale-face-database, last access: 4.01.2022