Detection of Face Mask in Real-Time Using Deep Convolutional Networks with Alert System

Renugadevi A S (renugadevi.ece@kongu.ac.in)
Kongu Engineering College https://orcid.org/0000-0003-0619-3088
Nihila B
Kongu Engineering College
Rakesh A
Kongu Engineering College
priyadharshini M
Kongu Engineering College

Research Article

Keywords: covid-19, face mask detection, crowdest areas, YOLO

Posted Date: January 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1221329/v1

Abstract

Owing to the spread of covid-19 in and around all the areas, the people are advised to wear masks regularly, maintain social distancing and to sanitize the hands frequently. But most of the people are not properly following the wearing of masks in all the places including the crowdest areas. While the causes are complicated, there is a widespread belief that there is insufficient evidence to justify the use of face masks, particularly among the general people in a community environment. Detection of face mask is a difficult computer vision research subject due to the tiny size of the face cover region. The availability of appropriate datasets for this problem is rare so the way of finding solution is difficult. Considering all the drawbacks in the existing methodologies, The proposed method uses a innovative dataset that contains the images of masks and also the dataset includes 5,821 images of both the persons with masks and without masks that helps to classify in the different labels such as having mask and does not have mask. The box is bounded over the persons with and without masks. For detection, the system employs the YOLO v4 model, which has been shown to outperform prior versions of the YOLO and RESNET50 models.

1. Introduction

People all across the world have been wearing face masks to preserve themselves from Pandemic COVID-19.Not only COVID, in order to get relief from pollution and other infectious diseases for the past year, they used to wear face masks before that. Face recognition systems have found it difficult to detect the person beyond the mask in areas such as banks, ATMs, facial-biometric attendance systems and airport security checks because of the existence of face masks over the face. Face recognition algorithms recognize facial traits namely the shape and size of the eyes, cheekbones, nose and jaw, which might be difficult to detect in an atmosphere when everyone is having face masks. The systems can be trained with the above traits but their effectiveness is impeded by the use of face masks, which conceal the bulk of the key features for completing recognition tasks. The well equipped face mask detection system is needed to detect the persons with mask even in all the crowded areas because the size of the mask is very small, so it is tedious to detect with the normal detection systems. COVID-19, the largest pandemic in recent history, swept the globe in 2020. There have already been 152 million cases and 3 million deaths worldwide as of May 1st, 2021. These figures are significantly undercounted in many areas. Aside from that, many parts of the world have slowed or stopped because of the human economic and social consequences of distancing and protection measures. For the purposes of the current pandemic and future pandemic predictions, the proposed work aims to develop a mask detection system capable of determining whether people in surveillance-type video streams are wearing their masks correctly.

1.1Convolutional Neural Networks (CNNs)

Artificial intelligence has grown dramatically in its ability to bridge the gap between human and machine capabilities.[12] Enthusiasts and Researchers work on many elements of the subject to reach incredible outcomes. Machine vision systems are one of the main applications of convolutional neural networks.

The main theme of the CNNs is to imitate the people in various activities using the machines and also make the machines to perform tasks like as image and video recognition, media reconstruction, intelligent retrieval, image analysis and classification, recommendation systems, and so on. Convolutional Neural Network has brought greater advancements in the computer vision[11] and the architecture of the network is given in the below figure:

A CNN takes the picture or image as a input and produces the output as a identification of different objects with the help of biases or weights applied in various angles[13]. When compared to other classification methods, a ConvNet requires substantially less pre-processing. While basic approaches need hand-engineering of filters, ConvNets can learn these characteristics with enough training.

The architecture of Convnet is encouraged by the organization of the cortical region and is analogous to the connection arrangement of neurons in brain of humans. Separate neurons only reacted to Receptive Field’s stimuli, a small part of the visual field. If a group of similar fields overlap, they will encompass the whole visual region.

1.1.1 Model: Yolo V4

The YOLOv4 model evolved from the YOLOv3 model and is an object detection method. Hong-Yuan Mark Liao ,Chien-Yao Wang and Alexey Bochkovskiy and invented the YOLOv4 technology. The performance of YOLO-4 is double times faster when compared with the others[8]. The YOLO method is supported by a only one Convolutional Neural Network (CNN). The CNN split an image into zones before calculating the adjoining boxes and probabilities for each area. At the same time, it predicts several probabilities for those classes and also adjoining boxes for classes. V4 is given in the below diagram:

2. Related Works

In paper[1], the author said that Automatic License plate recognition is implemented using the cutting-edge object detector in YOLO algorithm. The networks are trained and also tuned at each stage of the ALPR. The efficiency of the system is ensured by varying the camera, lighting effects and the background settings.

In 2nd paper, the usage of the YOLO model is depicted as a little YOLO-SPP and the model is used to predict the theft of the cars. The performance of the recognition of cars is increased by extracting the features related to the vehicles and also the spatial pyramid pooling is applied along with the extraction of features helps to improve accuracy of learning in networks.

A unique flower detection application anchor-based approach is provided in [3]rd research paper, which is paired with an attention mechanism to recognise the flowers in an AIoT smart garden more correctly and quickly. While many academics have focused on flower categorization in previous studies, the topic of flower detection has largely gone unnoticed. The topic we've stated is primarily concerned with the investigation of a novel design and application of flower detecting. To begin, a novel end-to-end flower detecting anchor-based approach is inserted into the network's architecture to make it more valuable and quick, and the loss function and attention mechanism are incorporated into our model to suppress uninteresting characteristics. Second, we can incorporate our flower detecting algorithms into the mobile device. Through a series of experiments, we discovered that our flower detecting technology is quite effective. Our method's detection accuracy is comparable to that of the state-of-the-art, but its detection speed is quicker. It contributes significantly to flower detection in computer vision.

Author in his[4] study describes how to use Principal Component Analysis (PCA) on masked and unmasked face recognition. In today's world, security is a must-have concept. Face recognition is commonly used in biometric technology to safeguard any system since it is more trustworthy than other traditional approaches such as PIN, password, fingerprint, and so on in identifying or verifying a person effectively. Face recognition has been a particularly difficult process in recent years due to various occlusion or masks such as the presence of scarves,sunglasses, hats, and many forms of disguise or make up materials. These masks have an effect on the rate of accuracy of face recognition. Many algorithms for non-masked face recognition have recently been created that are extensively utilised and provide superior performance. Few contributions have been made in the field of masked face recognition. As a result, in this study, a statistical strategy that is used in both non-masked face identification and masked face recognition has been chosen. PCA is a more effective and extensively used statistical approach. As a result, the PCA method was used for this task. Finally, a comparison research was conducted in order to have a better understanding.

In [5],the fast spread of COVID-19 has caused significant harm and infected tens of millions of individuals throughout the world. Because there is no particular treatment, wearing masks has been a successful technique of preventing COVID-19 transmission and is mandatory in most public spaces, resulting in a rising need for automatic real-time mask detection systems to replace human detection. We suggested the Properly Wearing Masked Face Detection Dataset (PWMFD) in this study, which includes 9205 photos of mask-wearing samples divided into three categories. Furthermore, we suggested Squeeze and Excitation (SE)-YOLOv3, a mask detector with a reasonably balanced efficacy and efficiency. We added the attention mechanism into Darknet53 by introducing the SE block to acquire the relationships between channels so that the network could focus more on the relevant feature. To increase the stability of bounding box regression, we used GIoUloss, which can better express the spatial difference between predicted and ground truth boxes. For resolving the excessive foreground-background class imbalance, focal loss was used. In addition, we used image augmentation techniques to boost the model's resilience on the particular challenge. SE-YOLOv3 surpassed YOLOv3 and other state-of-the-art detectors on PWMFD, achieving a greater 8.6 percent mAP compared to YOLOv3 while having a comparable detection rate.

The COVID - 19 epidemic is wreaking havoc on humanity, regardless of caste, creed, gender, or religion. Until a vaccine is developed, we should all do our part to limit the corona-spread. virus's Using a face mask can surely aid in the control of the virus's spread. COVID - 19 face mask detector employs or owns Facemasknet, deep learning algorithms to successfully determine whether or not a person is wearing a face mask. The document has three classifications: individual wearing a mask, incorrectly worn masks, and no mask identified. Facemasknet, our deep learning approach, yielded an accuracy of 98.6 percent in [6]. The Facemasknet can function with both still photos and a live video stream. When the nose and mouth are partially covered, the mask is worn incorrectly. Our face mask identifier has a simple structure and produces rapid results, so it may be used in CCTV video to determine if a person is wearing a mask correctly so that he does not endanger others. Because mass screening is feasible, it may be employed in congested areas such as railway stations, bus stops, marketplaces, streets, mall entrances, schools, colleges, and so on. We can ensure that an individual wears the face mask correctly by monitoring how it is placed on the face. This helps to limit the spread of the virus.

In [9], the prevention of COVID-19 is done in the CCTV footages with the help of the embedded devices.The drawback is that the memory capacity of the embedded devices is low and also the power of computing is limited.

The Resnet is used instead of the embedded devices in paper [10] and also the face mask wearing is classified based on the decision tree algorithm.

3. Existing Method

Convolutional neural networks (CNNs)-based image identification algorithms can learn and extract complicated visual information automatically. This type of algorithm has sparked a lot of interest since it performs well in visual search, automated driving, medical diagnosis, and other areas. As a result, some researchers apply CNNs to the field of picture face mask recognition, resulting in the development of a self-learned method for collecting face mask image characteristics.

The goal of this study is to find the best strategy for recognizing the face mask among several deep CNN models. The evaluation measures were used to predict the accuracy of the various models, and the best-suited method was identified based on these indicators. The background concepts and associated works portion of the materials and techniques section covers the background concepts and relevant works on this topic, as well as a quick introduction to deep CNN models. The implementation section goes through the hardware and software requirements, the dataset used, and the parameter settings for the different deep CNN models. The performance comparisons of several deep CNN models are then shown in the results and comments section.

Finally, in the conclusion section, this work is summarized, along with a brief discussion of potential future improvements. This type of technique, which generates proposal areas by computing each one separately, does not employ CNNs in the global detection phase, resulting in a vast amount of computation and sluggish detection time.

The self-built fire picture dataset is used to create and train the image face identification system Retina Net face model in this work. Finally, the proposed approach uses the Exception model to identify the best detection performance for whether or not the individual is wearing a mask.. The findings of the study might be beneficial in modifying detecting algorithms to prevent the spread of Covid-19.

4. Proposed Method

4.1Proposed dataset:

For any study in computer vision to progress, it is critical to have fascinating and demanding datasets. Using the Image Net, MS COCO, and PASCAL VOC datasets, the overall assessment of object recognition algorithms is enhanced, and a new dimension for multi-class object identification is obtained.[7] Face mask detection datasets with one step ahead benchmark must be developed using identical procedures, and this will aid in recognizing persons hiding behind face masks. As a result, the attention on this challenging study field will be increased.

For the dataset, github and google were used to collect roughly 5821 photographs from the internet and scale them all to 416x416 pixels in order to input them into the utilized face mask detection technique. Tight bounding boxes were drawn over each of the 5821 unique photographs under various settings to extract distinct and rich information utilizing the Labeling Annotation tool for such a big labeling effect.

The data is then supplemented in order to increase the dataset's size. That is, operations like flipping, rotation, shearing, and HSV shift were performed on the datasets for augmentation in order to provide the object detector more characteristics to train with. Using data augmentation, 6985 picture samples were created from the initial 5821 data samples, resulting in a 20 percent increase in the dataset.

To describe the strategy for assessment, the dataset is divided into training, validation, and testing groups. This information is then passed on to the face mask detection method, which generates difficult results. The dataset is divided into four categories: with masks, without masks, masks that are wrongly applied, and mask area. The dataset is divided into three sections for training and testing: 4615 pictures are used for training the model, 427 pictures are used for validating the model, and 779 pictures are used for model testing.

4.2Dataset Comparison with existing datasets:

The interpretation along with bounded boxing is missed in the existing techniques used for the classification of masked and unmasked persons. The interpretations along with bounding boxes were not needed since the datasets that were available were mostly useful for face classification. Hence, richly annotated dataset is obtained in this paper to help progress the face mask detection research.

A wide range of images taken from the internet of people with masks is used in the Masked face detection dataset and is considered as a advanced dataset compared with the existing datasets. This dataset focuses only on no mask category and improper wearing of mask.

Real masked Face recognition dataset contains the largest dataset with classes of faces without masks and faces with masks. The mask area and improper mask wearing is not taken into consideration in this dataset.

Retina Net facemask contains face having mask and without having mask is placed in face mask dataset. The dataset not contains face mask area and face wearing masks incorrectly. Masked Faces (MAFA) dataset contains dataset of people with masked faces annotated and misses classes face wearing masks incorrectly and face without masks. MOXA dataset contains only two classes with and without masks of people collected over internet and publicly available datasets.

And thus, with the comparison of previous datasets, the created data collection is better and contains variations since it includes all the four classes, people wearing facemasks, people wearing incorrectly face masks, persons not wearing face masks and particularly area with masks in pictures having people faces with and people wearing masks incorrectly. This helps the face detection algorithm and the detection systems identify if the face mask is present on any area of the face to prevent occurring of faults in detection of a face owing to hiding of important features in face.

4.3Proposed dataset Evaluation:

YOLOv4 model has an advantage of fast detection and generalization over any other algorithm and with the proposed dataset, it is implemented on system having intel i5 configuration with 8GB memory and processor as NVIDIA with input size 416x416, 0.001 rate of learning, 0.9 momentum with the 0.0005 decay.

4.4Performance comparison with other YOLO models and similar works:

With dataset classified as 80 percent for training, 7 percent for validation and 13 percent for testing, the accuracy obtained for each of the three YOLO variation models are as follows:

The YOLOv2 model has obtained 83.83% accuracy along with 74.50% accuracy for validation set and 78.95% accuracy for Testing set. It was evaluated to have 45 Frames per second (FPS), high variance and high bias and so performance is very poor. The YOLOv3 model has obtained 99.75% accuracy along with 87.16% accuracy for validation set and 90.18% accuracy for Testing set. It was evaluated to have 23 Frames per second, medium variance and low bias, so performance is good. The YOLOv4 model has obtained 99.65% accuracy along with 88.38% accuracy for validation set and 98.95% accuracy for Testing set. It was evaluated to have 22 Frames per second, medium variance, and low bias and thus, performance is better than others. And hence, it is proved for YOLOv4 model to be superior than other variants of its model.

The results of the YOLOv4 approach with existing YOLOv3 and YOLOv2 are taken for consideration hence prove that the result of the YOLOv4 variants are better on the proposed dataset and thereby justifying its importance.

The output obtained when the recorded CCTV footage is given as a input to the YOLOv4 model is given in the below figure:

The mask count is shown as 42 which means that they are having masks in the crowd and no mask count as 1 so the status is showing as warning.The real time video is captured using the web camera and the mask count is updated in the following two figures:

5. Conclusion

People have suffered both physically and mentally since outbreak due to the causalities caused by covid-19. The main thing we should note that all over the world every country have implemented wearing of face mask as a mandatory one. The main reason for implementing this is covid-19 which is caused by corona virus spreads mainly through air which acts as a transfer medium. In this modern era each and everything has been monitored through cameras for safety purpose. So this project follows the idea of capturing the video of people and monitoring them to detect the people wearing mask and people who are not wearing mask in public environment. It is done in real-time which is a wonderful thing to be noted. The main idea is not only to detect the persons with or without masks. It also says the live count status of masks wearing persons and no mask wearing persons. The person who is in-charge will be alerted with danger signal in real-time for humans who are not wearing masks. The accuracy and time taken for detection and processing has been great in general with yoloV4 model used here. The first and foremost thing is wearing a facemask these days to prevent the disease. We as a whole should stand united and control and defeat the disease. Further the idea of maintain social distance to this has been under planning.

Declarations

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Funding

No funding is provided for the preparation of manuscript.

Contributions

All the authors have equal contribution in this article.

References

(1)Jonathan S. Talahua 1,2,Jorge Buele 1ORCID,P. Calvopiña 3 andJosé Varela-Aldás, Realtime license plate detection for non-helmeted motorcyclist using YOLO, Ict Express (2020), https://doi.org/ 10.1016/j.icte.2020.07.008.

(2) Shilpa Sethi,a Mamta Kathuria,a, and Trilok Kaushikb: A delicate real-time vehicle detection algorithm, Optik 255 (2021), https://doi.org/10.1016/j.ijleo.2020.165818.

(3) G K Jakir Hussain 1, R Priya2, S Rajarajeswari2, P Prasanth 2, N Niyazuddeen2, Flower end-to-end detection based on YOLOv4 using a mobile device, Wirel. Commun. Mob. Comput. (2020), https://doi.org/10.1155/ 2020/887064.

(4) M.S. Ejaz, M.R. Islam, M. Sifatullah, A. Sarker, Implementation of principal component analysis on masked and non-masked face recognition, 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) (2019), https://doi.org/10.1109/ICASERT.2019.8934543.

(5) G. Jignesh Chowdary, Narinder Singh Punn, Sanjay Kumar Sonbha dr Sonali Agarwal, Detect faces and determine whether people are wearing mask. FaceMaskDetection, 2020. Accessed on 22-May-2020, https://github.com/ AIZOOTech/.

(6) Jesús Tomás, Albert RegoORCID,Sandra Viciano-Tudela andJaime LloretORCID, RealTime face mask identification using face masknet deep learning network, Ssrn Electron. J. (2020), https://doi.org/10.2139/ ssrn.3663305.

(7) Bing API, 2020. Accessed on 27-May-, https://pypi.org/project/bing-image-downloader/.

(8) Safa Teboulbi, Seifeddine Messaoud, Mohamed Ali Hajjaji , and Abdellatif Mtibaa: Optimal Speed and Accuracy of Object Detection, arXiv, 2020, https://arxiv.org/abs/2004.10934.

(9) B. Roy, S. Nandy, D. Ghosh, MOXA: A deep learning based unmanned approach for realtime monitoring of people wearing medical masks, Trans. Indian Nat. Acad. Eng. 5 (2020) 509–518, https://doi.org/10.1007/s41403-020-00157-z.

(10) M. Loey, G. Manogaran, M.H. N.Taha, N.E. M.Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, Measurement 167 (1) (2021), https://doi.org/10.1016/j.measurement.2020.108288.

(11) K.Venu and Natesan.P., Classification of Myocardial Infarction using Convolution Neural Network,International Journal of Recent Technology and Engineering (IJRTE),Vol.8,No.,pp.12763-12768, Nov-2019.

(12)Renugadevi A.S., Ananthi B., Dr.Jayavadivel R., Pradeep S.,Aravind M,Dharanesh S,"Classification of Custard Apple Leaves Using Deep Convolutional Networks",Turkish Journal of Computer and Mathematics Education,Vol.12 No.6(2021), 3288-3292.

(13)N.T.Renukadevi, K.Saraswathi, P. Prabu, Venkatachalam "Brain Image Classification Using Time Frequency Extraction with Histogram Intensity Similarity", Computer Systems Science & Engineering, Vol. 41, No. 03, 2021.

Tables

Table 1: Comparision of accuracy with three models

Name of the model	Training Accuracy	Testing Accuracy
YOLOV2	83.83%	78.95%
YOLOv3	99.75%	90.18%
YOLOv4	99.65%	98.95%

Figures

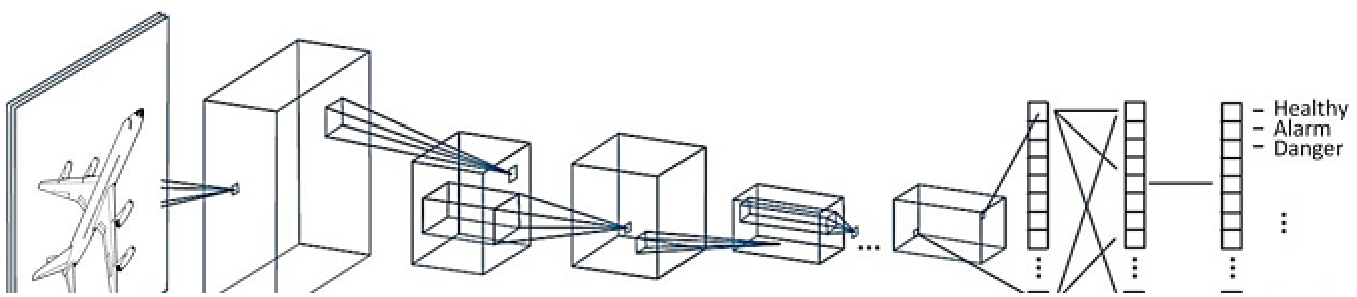

Figure 1

Convolutional Neural Network

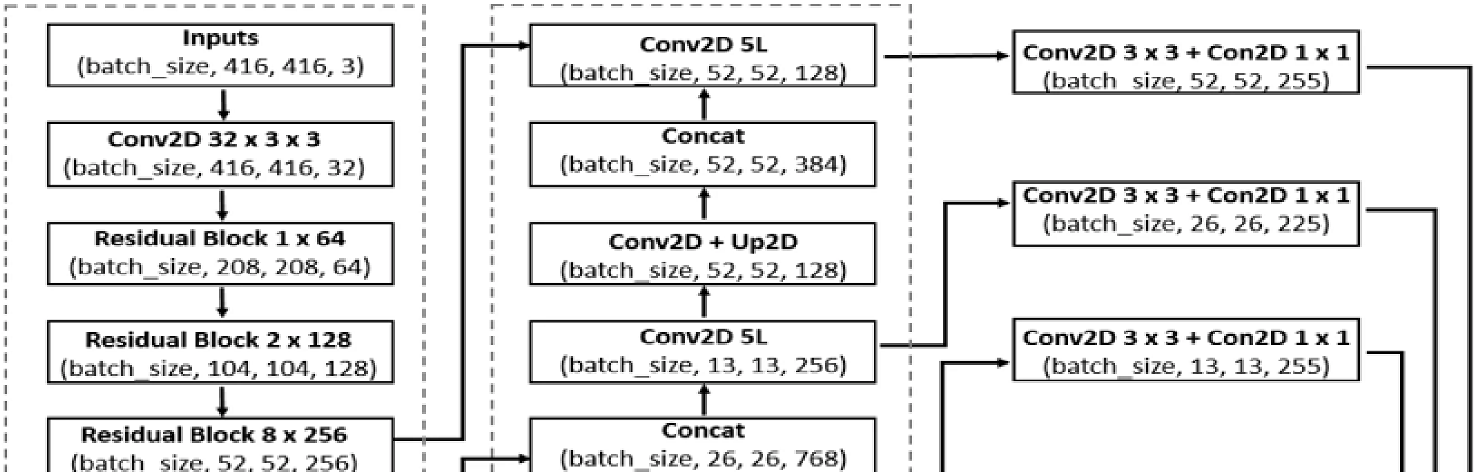

Figure 2

YOLOV4 Architecture

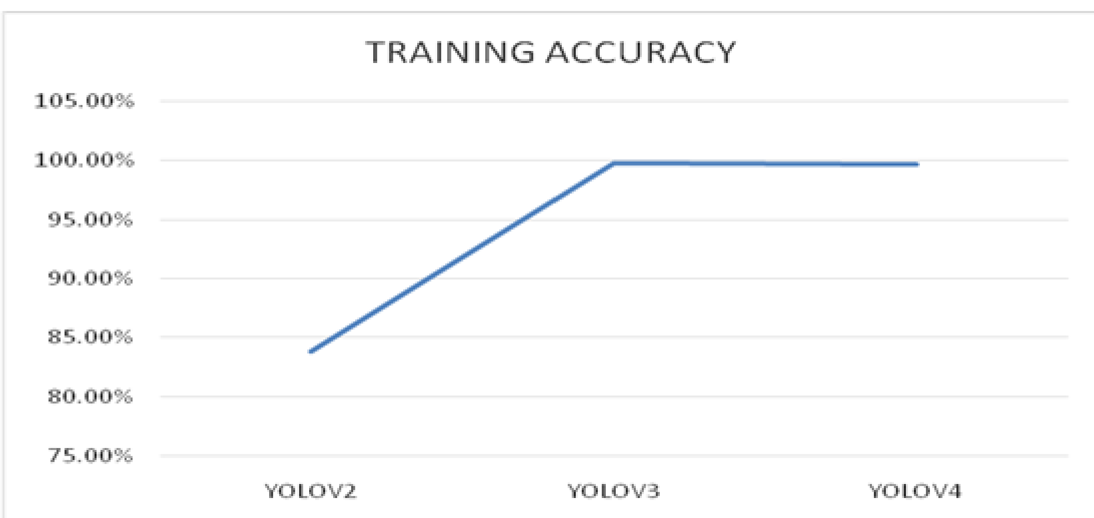

Figure 3

Training Accuracy with three YOLO model

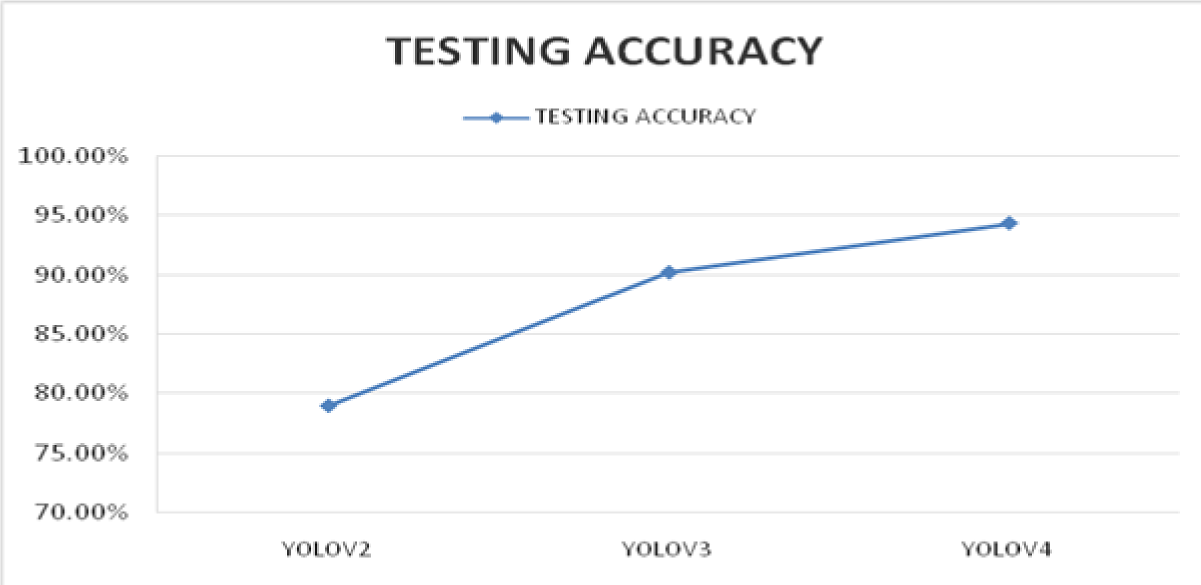

Figure 4

Testing Accuracy with three YOLO models

Figure 5

Output screen of CCTV footage input

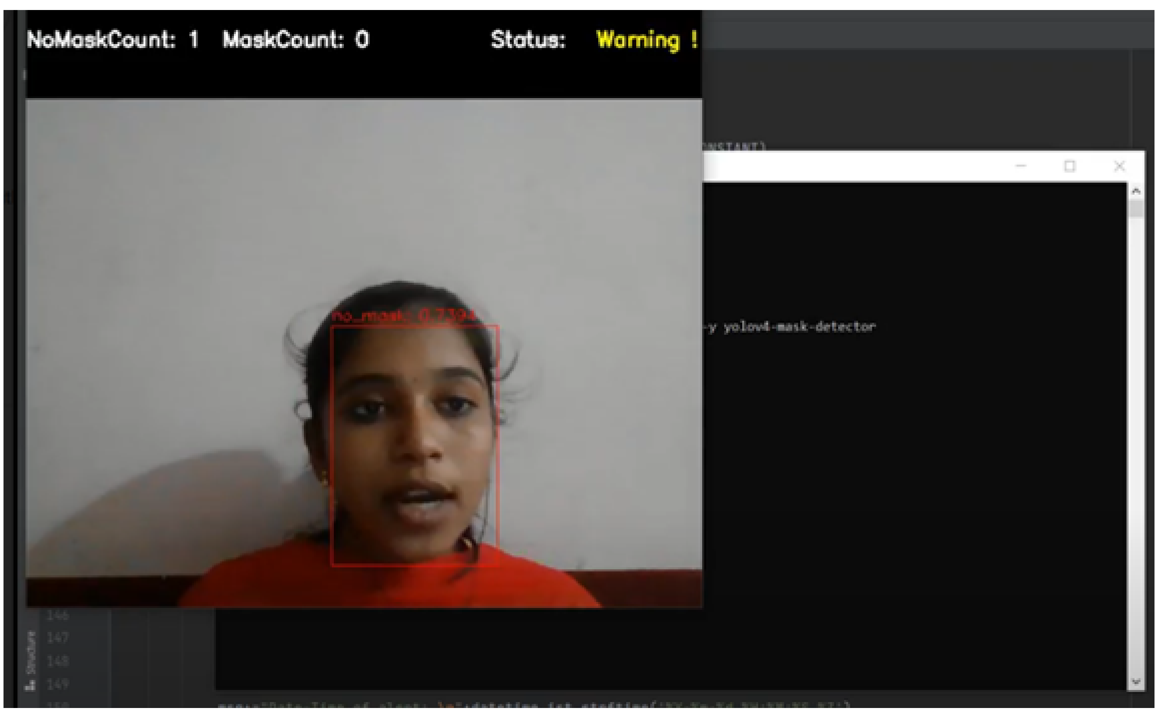

Figure 6

Output of Person having mask

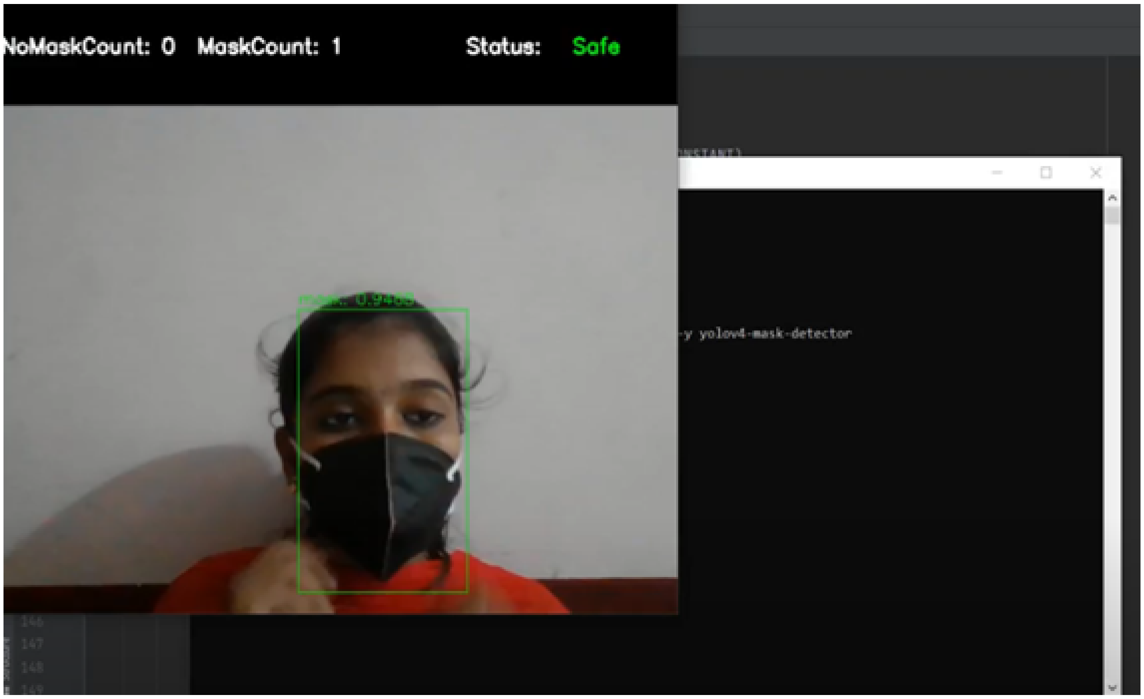

Figure 7

Output of Person not having mask