AN EFFICIENT IMPLEMENTATION OF REAL TIME FACE-MASK VIOLATION DETECTION SYSTEM USING DEEP LEARNING

A Project Report Submitted to

the APJ Abdul Kalam Technological University

in partial Fulfilment of the Requirements for the Award of the Degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

BY

Mr. SANAL P	**VVT17CS007**
Mr. ABHISHEK KVK	**VVT16CS003**
Ms. VRINDA K	**VVT17CS009**

Under the Guidance of

Mrs. AMRITHA K DAS

(Assistant Professor, Dept. of CSE)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VEDAVYASA INSTITUTE OF TECHNOLOGY

KARADPARAMBA, MALAPPURAM

JUNE 2021

AN EFFICIENT IMPLEMENTATION OF REAL TIME FACE-MASK VIOLATION DETECTION SYSTEM USING DEEP LEARNING

A Project Report Submitted to

the APJ Abdul Kalam Technological University

in Partial Fulfilment of the Requirements for the Award of the Degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

BY

Mr. ABHISHEK KVK

Reg.No. VVT16CS003

Under the Guidance of

Mrs. Amritha K Das

(Assistant Professor, Dept. of CSE)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VEDAVYASA INSTITUTE OF TECHNOLOGY

KARADPARAMBA, MALAPPURAM

JUNE 2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VEDAVYASA INSTITUTE OF TECHNOLOGY, KARADPARAMBA

CERTIFICATE

This is to certify that the report entitled **'AN EFFICIENT IMPLEMENTATION OF REAL TIME FACE-MASK VIOLATION DETECTION SYSTEM USING DEEP LEARNING'** submitted by '**ABHISHEK KVK**' Reg.No.**VVT16CS003** to the APJ Abdul Kalam Technological University in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering is a bonafide record of the project work carried out by him/her under my/our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Internal Supervisor External Supervisor

HEAD OF THE DEPT

DECLARATION

I undersigned hereby declare that the project report **“AN EFFICIENT IMPLEMENTATION OF REL TIME FACE-MASK VIOLATION DETECTION SYSTEM USING DEEP LEARNING”**, submitted for partial fulfilment of the requirements for the award of degree of Master of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of **Amritha K Das**. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Place : Kozhikode Signature:

Date : Name of the student:

ACKNOWLEDGEMENT

We thank god, the almighty for the blessing us in making the Project a successful one. The satisfaction that accompanies that the successful completion of any task would be incomplete without the mention of people whose ceaseless cooperation made it possible, whose constant guidance and encouragement crown all efforts with success.

We would like to thank our Director ***Dr. Arun Korath****, for providing the best facilities and atmosphere for the Project completion and presentation.*

We express our heartfelt gratitude towards ***Dr.Sangheethaa.S****, Principal, Vedavyasa Institute of Technology, for extending all the facilities required for doing this work.*

We express our sincere gratitude to ***Prof. Mrs. S. Kavitha Murugesan****, Head of Department, Dept. of Computer Science & Engineering, Vedavyasa Institute of Technology, Malappuram for her guidance and support.*

We are grateful to our project guide ***Mrs. Amritha K Das*** *& Class tutor* ***Mrs. Anjana Nikhil****, Asst. Professor, Dept. of Computer Science & Engineering. For the guidance, inspiration and constructive suggestions that helpful us in the preparation of this work and all lecturer of department of computer science and engineering for their valuable suggestions.*

Last but not the least; we extend heartfelt gratitude to our parents and friends for their support and assistance.

TABLE OF CONTENTS

ABSTRACT

With the recent outbreak and rapid transmission of the COVID-19 pandemic, the need for the public to follow social distancing norms and wear masks in public is only increasing. According to the World Health Organization, to follow proper social distancing, people in public places must maintain at least 3ft or 1m distance between each other. This paper focuses on a solution to help enforce proper social distancing and wearing masks in public using YOLO object detection on video footage and images in real time. The experimental results shown in this paper infer that the detection of masked faces and human subjects based on YOLO has stronger robustness and faster detection speed as compared to its competitors. Our proposed object detection model achieved a mean average precision score of 94.75% with an inference speed of 38 FPS on video. The network ensures inference speed capable of delivering real-time results without compromising on accuracy, even in complex setups. The social distancing method proposed also yields promising results in several variable scenarios.

Keywords— *COVID-19, Social Distancing, Masks, YOLO, Real-time .*

LIST OF FIGURES

LIST OF ABBREVATIONS

WHO	WORLD HEALTH ORGANIZATION
YOLO	YOU ONLY LOOK ONCE
LBP	LOCAL BINARY PATTERNS
CNN	CONVOLUTIONAL NEURAL NETWORK
HOG	HISTOGRAM OF ORIENTED GRADIENT
CDC	CENTERS FOR DISEASE CONTROL AND PREVENTION

CHAPTER 1

INTRODUCTION

Since the COVID-19 pandemic took the world by storm, tough but necessary measures were taken by governments throughout the world to control its spread. This resulted in bringing normal day-to-day activities to a complete standstill. Months into lock down, when we see the curve flattening in several countries, the community grows restless. Relevant authorities like WHO have laid down certain guidelines to minimise people's exposure to the virus. Some safety measures people are encouraged to follow include wearing masks and maintaining a distance of 3 ft, which is approximately 1m, from another individual. Fig 1 shows two test cases of our violation detector. Both figures include evaluations of people of varying heights, standing at different angles. There are several countries in the world that have actually made mask wearing mandatory by law, and it has been observed that certain private organisations in the other countries have also been following in their footsteps. In vast establishments, it's hard to ensure that people are adhering to these crucial social distancing rules. To allow for easy tracking of such violators, an automated system is an absolute need of the hour.

We have recognized this need and have developed a model particularly suited to detect certain violations in real time. The first use of our model is to actually detect people's faces to determine whether or not they're wearing an acceptable mask. The second use is to determine whether or not social distancing is being maintained between 2 individuals, in the most efficient, accurate and simple manner, hence requiring overseeing authorities to take minimum effort.

To implement the above model, we have used object detection to detect exactly 3 classes: masked faces, unmasked faces, and people. While other models that have attempted to differentiate between masked and unmasked faces have favoured object detection networks like Single Shot Detector method, etc. to train on, we found that these were not efficient enough to help communities deal with potential risks in real-time.

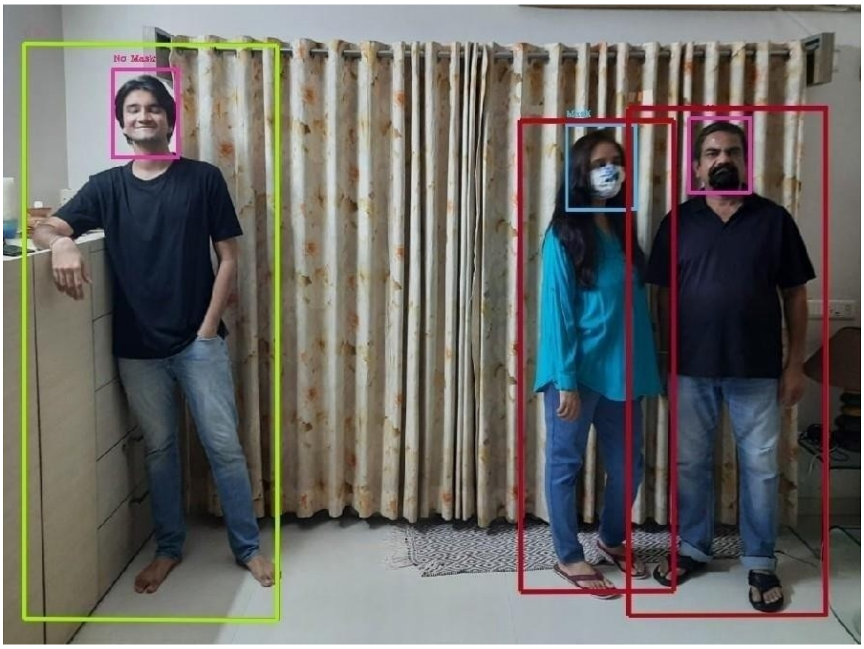

Figure 1. Social distancing, face mask norms – not followed by some subjects

CHAPTER II

LITERATURE SURVEY

[1] DC-SPP-YOLO: Dense Connection andSpatialPyramid Pooling Based YOLO for Object Detection

-Zhanchao Huang, Jianlin Wang

The object detection approaches based on computer vision have been widely used in security monitoring, automatic driving, medical diagnosis, and other fields. Early vision-based object detection approaches, which had low detection accuracy and narrow application range, rely on object features such as edges, key points or templates. In this regard, Haarlike features, HOG (Histogram of Oriented Gradient), LBP (Local Binary Patterns) and other feature extraction approaches with better object expression ability were proposed and used for object detection task together with machine learning approaches. In 2007, Felzenszwalb et al. introduced the DPM (Deformable Parts Models) approach creatively, which got a higher detection accuracy than other approaches at that time by a new object detection pipeline based on handcrafted futures and machine learning. After that, a variety of object detection approaches based on handcrafted features and machine learning were proposed one after another and performed well in successive PASCAL VOC object detection challenges. However, most of these methods scanned through the entire image to detect the object regions by a sliding window, which hadlow detection efficiency. Also, the accuracy of object detection that was restricted by the expression ability of handcrafted futures was challenging to improve further. With the improvement of computing performance and the abundance of data resources, the AlexNet approach based on CNN (Convolutional Neural Network) was proposed by Krizhevsky et al. in 2012, which proved that using the features extracted by CNN could classify the object images more accuracy than using the handcrafted futures. The AlexNet provided a new idea forthe object detection task that was regarded as the extension of image classification. In 2014, the R-CNN proposed by Girshick et al. employed CNN to extract features in object detection task for the first time and got the object detection accuracy which was superior to the state-of-the-art approaches at that time. Since then, the object detection approaches based on deep learning have gradually replaced the object detection approaches based on handcraft features and machine learning, and have become a research hotspot in this field. Generally, there are two main categories of CNN-based object detection approaches: the object proposal-based approaches and the regression-based approaches.

The object proposal-based approaches are mostly improved and developed from the R-CNN. For the problem of slow detection speed of R-CNN, Fast R-CNN and Faster R-CNN, which adopted Selective Search and RPN (Regional Proposal Network) instead of sliding window search respectively, were proposed to simplify the region proposal generation and improve the object detection speed. In 2016, Dai et al.proposed an R-FCN (Region-based Fully Convolutional Networks) to solve the problem that the ROI-wise sub network of Faster R-CNN did not share calculations in different region proposals. In the past two years, based on the Faster R-CNN and RFCN, RRPN (Rotation Region Proposal Networks), R-FCN- 3000 and other object proposal-based approaches of which the detection accuracy was further improved were presented. However, the frameworks of proposal-based approaches that had two stages, the region proposal generation and the subsequent feature resampling, were much more complex in comparison with the regression-based approaches; which resulted in low speed and difficulty in real-time performance.

In 2016, Redmon et al.presented for the first time a regression-basedapproach, YOLO (You Only Look Once), for object detection where a single convolutional network that simultaneously predicted bounding box coordinates and class probabilities was trained end-to-end directly. Even if YOLO opened the door to achieve real-time object detection, it was difficult to detect small-sized objects in the image, and the error of bounding box coordinates was large. In this regard, Liu et al. proposed an SSD (Single Shot Multi-Box Detector) that introduced reference boxes and detected the object on multi-scale feature maps to improve the accuracy of object detection. In 2017, Redmon and Farhadi proposed the YOLOv2 approach, and its accuracy and speed of object detectionwere significantly ameliorated compared with the YOLO approach; however, this method still used the Darknet19 with low ability on feature extraction as the backbone network and did not fully utilize the multi-scale local region features of the object, which constrained the further improvement of detection accuracy. Subsequently, the deep residual network was employed as the backbone network to get the detection accuracy that was further superior to state-of-the-art approaches in DSSD (Deconvolutional Single Shot Detector) and YOLOv3; on the other hand, the detection speed of these approaches was severely degraded due to the excessive number of network layers. In 2018, Zhou et al. introduced DenseNet-169, a dense convolutional network with better performance than the deep residual network, as the backbone network of SSD and proposed an STDN (Scale- Transferrable Detection Network) approach which achieved the detection accuracy close to the DSSD while improving the detection speed. Also, Jeong et al., Lee et al., Cao et al. and Zheng et al.proposed other improved SSD approaches for object detection, but the existing research on the improvements of YOLO series approaches are stillless.

Therefore, we propose a Dense Connection and Spatial Pyramid Pooling Based YOLO object detection approach for improving YOLOv2 by optimizing the connection structure of the backbone network and introducing the multi-scale local region feature extraction. This approach is more accurate than YOLOv2 while keeping the detection speed close to YOLOv2, higher than DSSD, YOLOv3, and STDN. The main contributions of this paper are as follows:

1. We employ the dense connection structure of the convolutional layers to improve the backbone network of YOLOv2 for strengthening feature extraction and ensuring maximum information flow between layers in the network.
2. An improved spatial pyramid pooling is introduced to collect and concatenate the local region features on different scales in the same convolutional layer for learning multi-scale object features more comprehensively.
3. The improvements above are introduced in YOLOv2, and the cross-entropy which can effectively alleviate the vanishing-gradient problem isutilized

instead of the mean squared error to represent object classification loss; a DC-SPP-YOLO approach is presented for ameliorating the detection accuracy with a fast detection speed.

This paper is organized as the following. Section 2 gives a brief review of the related works. In Section 3, we explain the proposed approaches in detail. Section 4 presents a series of experimental results and discussion. Finally, we make conclusions in Section 5.

[2] Detecting Masked Faces in the WildwithLLE-CNNs

- Shiming Ge1, Jia Li2, Qiting Ye1,3, Zhao Luo1,3

With the rapid development of machine learning methods, the problem of face detection seems to be well addressed yet. For example, the face detector proposed in achieves an average precision of 98.0% on the public image benchmark AFW by using the cascaded Convolutional Neural Networks, while the speed of some face detectors can reach up to 35 FPS or even 400 FPS.

Figure 2. Masked faces may have diversified orientations, degrees of occlusion and mask types, making their detection an extremely challenging task for existing face detectors.

Due to the great success of these face detectors, some of them have been integrated into applications so as to facilitate auto-focusing , human computer interaction and image database management.

Beyond the remarkable success achieved by existing works, there is increasing concern that the development of better face detectors is now becoming more and more difficult. In particular, the detection of masked faces, which can be very helpful for applications like video surveillance and event analysis, is still a challenging task for many existing models. As shown in Fig. 1, masked faces may have different orientations, degrees of occlusion and diversified types of masks, which make the accurate detection of masked faces a really challenging task even for the state-of-the-art face detectors. Compared with the classic task of normal face detection, existing models often have a sharp performance drop in detecting masked faces, which may be mainly caused by two reasons. First, there lacks a large dataset with massive masked faces for exploring the key attributes shared by various masked faces and identifying the models with the state-of-the-art performance. Second, facial features from the occluded parts are no longer available in the detection process, while the existence of masks inevitably bring in certain kinds of noise. With insufficient training and testing data as well as incomplete and inaccurate features, masked face detection has been becoming a widely recognized challenging task in the area of face detection. Although this issue has been tentatively studied in some recent works such as , it is still necessary to construct large datasets and develop effective and efficient models for masked face detection. Toward this end, this paper presents a data set for masked face detection, which is denoted as MAFA. The data set consists of 30,811 Internet images, in which 35,806 masked human faces are manually annotated. In the annotation process, we ensure that each image contains at least one face occluded by various types of masks, while the six main attributes of each masked face, including locations of faces, eyes and masks, face orientation, occlusion degree and mask type, are manually annotated and cross-checked by nine subjects. The data set will be released soon on the Internet, which we believe can facilitate the development of new face detectors in the future.

By inspecting the main characteristics of masked faces in MAFA, we find that most facial attributes can be lost in heavily occluded faces (e.g., faces with only eyes un occluded by masks), while the highly diversified masks can bring in various types of noises. Inspired by this fact, we propose LLE-CNNs for masked face detection by recovering missing facial cues and suppressing non-facial cues in the feature subspace. The proposed approach consists of a proposal module, an embedding module and a verification module. The proposal module first extracts a set of face proposals and characterize each proposal with a 4096d descriptor with a pre-trained VGG-Face model . Considering that the descriptor of a masked face can be incomplete or noisy, we further embed it into a feature subspace formed by two dictionaries that consist of the descriptors from representative normal faces and non-faces. Note that such dictionaries are learned from a large pool of normal faces, masked faces and non-faces from previous datasets and the training set of MAFA. With an approximate locally linear embedding, a candidate region can be characterized by the similarity scores to representative normal faces and non faces. Finally, such similarity-based descriptor is fed into the Verification module that consists of a Deep Neural Networks with only Fully-Connected (FC) layers so as to identify the real faces. Experimental results on the proposed MAFA data set show that the proposed LLE-CNNs significantly outperform 6 state-of-the- arts in detecting masked faces.

The main contributions of this paper are three folds.

1. We present a data set of masked faces that can be used as an additional training source for developing new face detectors;
2. We propose LLE-CNNs for masked face detection, which outperforms 6 state-of-the-art face detectors in detecting masked faces; and
3. We conduct a comprehensive analysis on the key challenges in masked face detection, which may be helpful for developing new face detectors in the future.

3 YOLOv3:An IncrementalImprovement

-Joseph Redmon Ali Farhadi

Bounding Box Prediction

Following YOLO9000 our system predicts bounding boxes using dimension clusters as anchor boxes. The network predicts 4 coordinates for each bounding box, tx, ty, tw, th. If the cell is offset from the top left corner of the image by (cx,cy) and the bounding box prior has width and height pw, ph, then the predictions correspond to:

- bx = σ(tx) +cx
- by = σ(ty) +cy
- bw = pwetw
- bh = pheth

During training we use sum of squared error loss. If the ground truth for some coordinate prediction is ˆ t* our gradient is the ground truth value (computed from the ground truth box) minus our prediction: ˆ t*−t*. This ground truth value can be easily computed by inverting the equations above. YOLOv3 predicts an objectness score for each bounding box using logistic regression. Thisshouldbe1ifthebounding box prior overlaps a ground truth object by more than any other bounding box prior. If the bounding box prior is not the best but does overlap a ground truth object by more than some threshold we ignore the prediction, following. We use the threshold of.5. Unlike our system only assigns one bounding box prior for each ground truth object. If a bounding box prior is not assigned to a ground truth object it incurs no loss for coordinate or class predictions, only objectness.

Class Prediction

Each box predicts the classes the bounding box may contain using multi label classification. We do not use a soft max as we have found it is unnecessary for good performance, instead we simply use independent logistic classifiers. During training we use binary cross-entropy loss for the class predictions. This formulation helps when we move to more complex domains like the Open Images Data set . In this data set there are many over lapping labels(i.e. Woman and Person). Using a soft max imposes the assumption that each box has exactly one class which is often not the case. A multi label approach better models the data.

Predictions Across Scales

YOLOv3 predicts boxes at 3 different scales. Our system extracts features from those scales using a similar concept to feature pyramid networks . From our base feature extractor we add several convolutional layers. The last of these predicts a 3-d tensor encoding bounding box, objectness, and class predictions. In our experiments with COCO[10]wepredict3boxesateachscalesothetensoris N ×N

×[3∗ (4+1+80)] for the 4 bounding box offsets, 1 objectness prediction, and 80 class

predictions. Next we take the feature map from 2 layers previous and upsample it by 2×. We also take a feature map from earlier in the network and merge it with our upsampled features using concatenation. This method allows us to get more meaningful semantic information from the upsampled features and finer-grained information from the earlier feature map. We then add a few more convolutional layers to process this combined feature map, and eventually predict a similar tensor, although now twice the size. We perform the same design one more time to predict boxes for the final scale. Thus our predictions for the 3rd scale benefit from all the prior computation as well as fine grained features from early on in the network. We still use k-means clustering to determine our bounding box priors. We just sort of chose 9 clusters and 3 scales arbitrarily and then divide up the clusters

evenly across scales. On the COCO data set the 9 clusters were: (10×13),(16×30),(33×23),(30×61),(62×45),(59×119),(116×90),(156×198),(373×36)

.

[4] Monitoring COVID-19 social distancing with person detection and tracking via fine-tunedYOLO v3 and Deepsort techniques

-Narinder Singh Punn, Sanjay Kumar Sonbhadra and Sonali Agarwal

COVID-19 belongs to the family of corona virus caused diseases, initially reported at Wuhan, China, during late December 2020. On March 11, it spread over 114 countries with 118,000 active cases and 4000 deaths, WHO declared this a pandemic . On May 4, 2020, over 3,519,901 cases and 247,630 deaths had been reported worldwide. Several healthcare organizations, medical experts and scientists are trying to develop proper medicines and vaccines for this deadly virus, but till date, no success is reported. This situation forces the global community to look for alternate ways to stop the spread of this infectious virus. Social distancing is claimed as the best spread stopper in the present scenario, and all affected countries are locked-down to implement social distancing. This research is aimed to support and mitigate the corona virus pandemic along with minimum loss of economic endeavours, and propose a solution to detect the social distancing among people gathered at any public place.

The word social distancing is best practice in the direction of efforts through a variety of means, aiming to minimize or interrupt the transmission of COVID-19. It aims at reducing the physical contact between possibly infected individuals and healthy persons. As per the WHO norms it is prescribed that people should maintain at least 6 feet of distance among each other in order to follow social distancing.

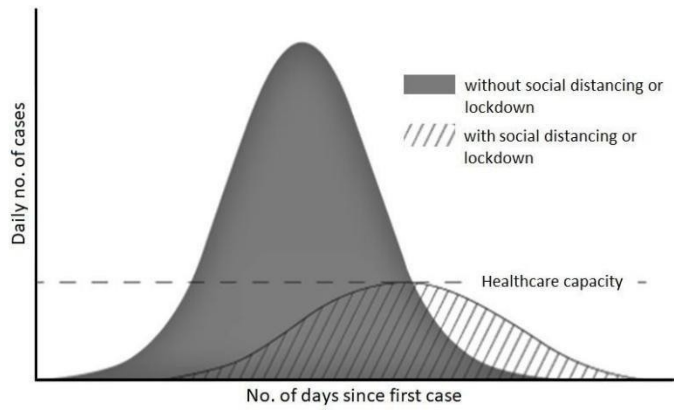

Figure 3: An outcome of social distancing as the reduced peak of the epidemic and matching with available healthcare capacity.

A recent study indicates that social distancing is an important containment measure and essential to prevent SARSCoV-2, because people with mild or no symptoms may fortuitously carry corona infection and can infect others . Fig. 1 indicates that proper social distancing is the best way to reduce infectious physical contact, hence reduces the infection rate . This reduced peak may surely match with the available healthcare infrastructure and help to offer better facilities to the patients battling against the corona virus pandemic. Epidemiology is the study of factors and reasons for the spread of infectious diseases. To study epidemiological phenomena, mathematical models are always the most preferred choice. Almost all models descend from the classical SIR model of Kermack and McKendrick established in 1927 . Various research works have been done on the SIR model and its extensions by the deterministic system , and consequently, many researchers studied stochastic biological systems and epidemic models . Respiratory diseases are infectious where the rate and mode of transmission of the causing virus are the most critical factors to be considered for the treatment or ways to stop the spread of the virus in the community. Several medicine organizations and pandemic researchers are trying to develop vaccines for COVID-19, but still, there is no well-known medicine available for treatment. Hence, precautionary steps

are taken by the whole world to restrict the spread of infection. Recently, Eksin et al. proposed a modified SIR model with the inclusion of a social distancing parameter, a(I,R) which can be determined with the help of the number of infected

and recovered persons represented as I and R, respectively.

$$dS/dt = -\beta S\ I/N(a(I,N))$$

$$dI/dt = -\delta I + \beta I\ (I/N)(a(I,N))$$

$$dR/dt = \delta I$$

where β represents the infection rate and δ represents recovery rate. The population size is computed as N = S + I + R. Here the social distancing term (a(I,R) : R^2 [0,1]) maps the transition rate from a susceptible state (S) to an infected state (I), which is calculated by aβSI N . The social distancing models are of two types, where the first model is known as “long-term awareness” in which the occurrence of interaction of an individual with other is reduced proportionally with the cumulative percentage of affected (infectious and recovered) individuals

$$a = (1 - (I + R) / N)^k$$

Meanwhile, the second model is known as “short-term awareness”, where the reduction in interaction is directly proportional to the proportion of infectious individuals at a given instance

$$a = (1 - (I / N))^k$$

where k is behavior parameter defined as, $k \geq 0$. Higher value of k implies that individuals are becoming sensitive to the disease prevalence.

In the similar background, on April 16, 2020, a company Landing AI under the leadership of most recognizable names in AI, Dr. Andrew Ng announced the creation of an AI tool to monitor social distancing at the workplace. In a brief article, the company claimed that the upcoming tool could detect if people are maintaining the safe physical distance from each other by analyzing real-time video streams from the camera. It is also claimed that this tool can easily get integrated with existing security cameras available at different workplaces to maintain a safe distance among all workers. A brief demo was released that shows three steps: calibration, detection and measurement to monitor the social distancing. On April 21, 2020, Gartner, Inc. identified Landing AI as Cool Vendors in AI Core Technologies to appreciate their timely initiative in this revolutionary area to support the fight against the COVID -19.

[5] Using Computer Visionto enhance Safety of Workforce in Manufacturing in a Post COVIDWorld

Prateek Khandelwal1, Anuj Khandelwal1, Snigdha Agarwal1, Deep Thomas1, Naveen Xavier1,Arun Raghuraman1

The spread of COVID-19 virus and the ensuing large scale lockdowns across the globe has given rise to an alarming situation. The resumption of production in manufacturing setups across all sectors is a key pre-requisite for kick starting economic activity of a nation. While there is an urgent need to resume operations at these plants, the safety of the workforce operating these plants cannot be compromised. Accordingly, processes are being put in place to educate the workforce regarding new safety regulations at the workplace which helps reduce the risk of virus transmission. However, to help the workforce transition into a post COVID world, there was a need for us to build solutions that help monitor and alert individuals once a safety violation occurs.

All plants have CCTV installation, with at least a few hundred cameras as part of their security system setup. It is however not practical to monitor all these feeds concurrently due to the manual nature of the task. We have built a system that takes in these feeds and analyzes frames using deep learning models to detect whether violations have occurred or not. Once detected, a real time voice alert is triggered in the area of the violation. This feedback helps reduce the violations and thus contributes to the overall safety at the plant. In addition, these alerts are stored in a central repository that helps the management analyze the trends and take suitable actions to curb the violations. An overview of the solution is provided in Section III of this paper.

Given the context of COVID-19, we focused on building features that help reduce the risk of virus transmission. Research indicated that maintaining social distance between co-workers as well as wearing face masks were effective means of reducing this risk. We hence built solutions that could monitor these actions through video feeds.

World Health Organization (WHO) has recommended that a social distance of at least 2m be maintained between individuals. While the requirement is simple, monitoring this aspect through video feeds that provide a perspective view makes it difficult to ascertain the exact distance on ground. We provide the details of the model and its aspects in Section IV of the paper.

WHO has also recommended that personnel are encouraged to wear face masks to avoid the risk of virus entering the body through the nasal / oral cavity . During the lockdown, it was encouraged by the Indian government for the people to come up with mask substitutes as most countries saw a scarcity of required PPE. Hence, the face masks worn are not of standard type and come in different colors, shapes and sizes. The lack of such diversified data for training purposes makes mask detection a challenging task. The approach taken to overcome this problem, the model and other details are discussed in Section V of the paper.

[6] SSD: Single Shot MultiBox Detector

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed,
Cheng-Yang Fu, Alexander C. Berg

Current state-of-the-art object detection systems are variants of the following approach: hypothesize bounding boxes, resample pixels or features for each box, and apply a high quality classifier. This pipeline has prevailed on detection benchmarks since the Selective Search work through the current leading results on PASCAL VOC, COCO, and ILSVRC detection all based on Faster R-CNN albeit with deeper features such as . While accurate, these approaches have been too computationally intensive for embedded systems and, even with high-end hardware, too slow for real-time applications.

Often detection speed for these approaches is measured in seconds per frame (SPF), and even the fastest high-accuracy detector, Faster R-CNN, operates at only 7 frames per second (FPS). There have been many attempts to build faster detectors by attacking each stage of the detection pipeline, but so far, significantly increased speed comes only at the cost of significantly decreased detectionaccuracy.

This paper presents the first deep network based object detector that does not re sample pixels or features for bounding box hypotheses and is as accurate as approaches that do. This results in a significant improvement in speed for high-accuracy detection (59 FPS with mAP 74.3% on VOC2007test, vs. Faster R-CNN 7 FPS with mAP 73.2% or YOLO 45 FPS with mAP 63.4%). The fundamental improvement in speed comes from eliminating bounding box proposals and the subsequent pixel or feature re sampling stage. We are not the first to do this (cf), but by adding a series of improvements, we manage to increase the accuracy significantly over previous at tempts. Our improvements include using a small convolutional filter to predict object categories and offsets in bounding box locations, using separate predictors (filters) for different aspect ratio detections, and applying these filters to multiple feature maps from the later stages of a network in order to perform detection at multiple scales. With these modifications—especially using multiple layers for prediction at different scales—we can achieve high-accuracy using relatively low resolution input, further increasing detection speed. While these contributions may seem small independently, we note that the resulting system improves accuracy on real-time detection for PASCAL VOC from 63.4% mAP for YOLO to 74.3% mAP for our SSD. This is a larger relative improve ment in detection accuracy than that from the recent, very high-profile work on residual networks . Furthermore, significantly improving the speed of high-quality detection can broaden the range of settings where computer vision is useful.

We summarize our contributions as follows:

- We introduce SSD, a single-shot detector for multiple categories that is faster than the previous state-of-the-art for single shot detectors (YOLO), and significantly more accurate, in fact as accurate as slower techniques that perform explicit region proposals and pooling (including FasterR-CNN).

- The core of SSD is predicting category scores and box offsets for a fixed set of default bounding boxes using small convolutional filters applied to feature maps

- To achieve high detection accuracy we produce predictions of different scales from feature maps of different scales, and explicitly separate predictions by aspect ratio.

- These design features lead to simple end-to-end training and high accuracy, even on low resolution input images, further improving the speed vs accuracy trade-off.

- Experiments include timing and accuracy analysis on models with varying input size evaluated on PASCAL VOC, COCO, and ILSVRC and are compared to a range of recent state-of-the-art approaches.

[7]WIDER FACE:AFace DetectionBenchmark

Shuo Yang ,Ping Luo ,Chen Change Loy, Xiaoou Tang

Face detection is a critical step to all facial analysis algorithms, including face alignment, face recognition, face verification, and face parsing. Given an arbitrary image, the goal of face detection is to determine the presence of faces in the image and, if present, return the image location and extent of each face . While this appears as an effort less task for human, it is a very difficult task for computers. The challenges associated with face detection can be attributed to variations in pose, scale, facial expression, occlusion, and lighting condition, as shown in Fig. 4. Face detection has made significant progress after the seminal work by Viola and Jones. Modern face detectors can easily detect near frontal faces and are widely used in real world applications, such as digital camera and electronic photo.

Figure 4. WIDER FACE dataset for face detection,

In the above figure , WIDER FACE dataset for face detection, which has a high degree of variability in scale, pose, occlusion, expression, appearance and illumination. We show example images (cropped) and annotations. The annotated face bounding box is denoted in green color. The WIDER FACE dataset consists of 393, 703 labeled face bounding boxes in 32, 203 images (Best view in color) bum. Recent research in this area focuses on the unconstrained scenario, where a number of intricate factors such as extreme pose, exaggerated expressions, and large portion of occlusion can lead to large visual variations in face appearance.

Publicly available benchmarks such as FDDB ,AFW , PASCAL FACE , have contributed to spurring interest and progress in face detection research. However, as algorithm performance improves, more challenging datasets are needed to trigger progress and to inspire novel ideas. Current face detection datasets typically contain a few thousand faces, with limited variations in pose, scale, facial expression, occlusion, and background clutter, making it difficult to assess for real world performance. As we will demonstrate, the limitations of datasets have partially contributed to the failure of some algorithms in coping with heavy occlusion, small scale, and atypical pose

In this work, we make three contributions.

1. We introduce a large-scale face detection data set called WIDER FACE. It consists of 32, 203 images with 393, 703 labeled faces, which is 10 times larger than the current largest face detection data set . The faces vary largely in appearance, pose, and scale, as shown in Fig. 4. In order to quantify different types of errors, we annotate multiple attributes: occlusion, pose, and event categories, which allows in depth analysis of existing algorithms.

2. We show an example of using WIDER FACE through proposing a multi-scale two-stage cascade framework, which uses divide and conquer strategy to deal with large scale variations. Within this framework, a set of convolutional networks with various size of input are trained to deal with faces with a specific range of scale.

3. We benchmark four representative algorithms, either obtained directly from the original authors or re implemented using open-source codes. We evaluate these algorithms on different settings and analyze conditions in which existing methods fail.

CHAPTER III

COVID-19

Corona virus disease (COVID-19) is an infectious disease caused by a newly discovered corona virus.Most people who fall sick with COVID-19 will experience mild to moderate symptoms and recover without special treatment.

3.1 HOW IT SPREADS

The virus that causes COVID-19 is mainly transmitted through droplets generated when an infected person coughs, sneezes, or exhales. These droplets are too heavy to hang in the air, and quickly fall on floors or surfaces.

You can be infected by breathing in the virus if you are within close proximity of someone who has COVID-19, or by touching a contaminated surface and then your eyes, nose or mouth.

Corona viruses (CoV) are a large family of viruses that cause illness ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV). A novel corona virus (nCoV) is a new strain that has not been previously identified in humans.

Corona viruses are zoonotic, meaning they are transmitted between animals and people. Detailed investigations found that SARS-CoV was transmitted from civet cats to humans and MERS-CoV from dromedary camels to humans. Several known corona viruses are circulating in animals that have not yet infected humans.

Common signs of infection include respiratory symptoms, fever, cough, shortness of breath and breathing difficulties. In more severe cases, infection can cause pneumonia, severe acute respiratory syndrome, kidney failure and even death.

Standard recommendations to prevent infection spread include regular hand washing, covering mouth and nose when coughing and sneezing, thoroughly cooking meat and eggs. Avoid close contact with anyone showing symptoms of respiratory illness such as coughing and sneezing

3.2 WHAT IS MASK

A surgical mask, also known as a medical face mask, is a personal protective equipment worn by health professionals during medical procedures. It prevents airborne transmission of infections between patients and/or treating personnel by blocking the movement of pathogens (primarily bacteria and viruses) shed in respiratory droplets and aerosols into and from the wearer's mouth and nose.

- Surgical mask
- Surgical face mask.jpg
- Asurgicalmask

Other names

- Procedure mask
- medical mask
- isolation mask
- laser mask
- fluid-resistant masks
- face mask

As seen here, a correct use of a surgical mask, a cord is visible behind the ears, and the other under the chin

Typically quite impermeable to moisture, the mask act as an additional barrier for the airway and are not usually designed (unless N95-rated) to completely prevent the wearer from inhaling smaller airborne pathogens, but could be still protective by filtering out and trapping most of the droplets that carry them. There is a predominance of evidence that surgical masks protect both the wearer (by filtering the inhaled air) and bystanders (by blocking down forceful exhalations from the wearer that can spread pathogens afar).

3.3 WHY MASK IS NECCESSARY

When you wear a mask, you protect others as well as yourself. Masks work best when everyone wears one.

A mask is NOT a substitute for social distancing. Masks should still be worn in addition to staying at least 6 feet apart, especially when indoors around people who don't live in your household.

Masks should completely cover the nose and mouth and fit snugly against the sides of face without gaps.

Masks should be worn any time you are traveling on a plane, bus, train, or other form of public transportation traveling into, within, or out of the United States and in U.S. transportation hubs such as airports and stations.

People age 2 and older should wear masks in public settings and when around people who don't live in their household.

Wear a mask inside your home if someone you live with is sick with symptoms of COVID-19 or has tested positive for COVID-19.

Wash your hands with soap and water for at least 20 seconds or use hand sanitizer with at least 60% alcohol after touching or removing your mask.

Masks may not be necessary when you are outside by yourself away from others, or with people who live in your household. However, some areas may have mask mandates while out in public, so please check the rules in your local area (such as in your city, county, or state). Additionally, check whether any federal mask mandates apply to where you will be going.

CDC continues to study the effectiveness of different types of masks and update our recommendations as new scientific evidence becomes available. The most recent scientific brief is available here: Scientific Brief: Community Use of Cloth Masks to Control the Spread of SARS-CoV-2 | CDC

CDC recently conducted a study in a laboratory that tested the performance of different mask combinations.

There are several easy methods to improve the performance of your mask. Visit CDC's Improve the Fit and Filtration of Your Mask to Reduce the Spread of COVID-19 webpage to learn more.

3.4 IMPORTANCE OF WEARING MASK NOWADAYS

As COVID-19 spread all across the world, many of us became aware of how important face masks are. While face masks may cause a slight inconvenience, especially during the summer season, these tissues are the only barrier between us and the deadly SARS-CoV-2 (corona virus).

Healthcare officials from the World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), and local institutions from all around the world are urging people to wear face masks, as it's the only way to prevent the transmission of the virus

This article will focus on the importance of wearing face masks to keep yourself and your family safe.

Why is it important to wear face mask?

When we go back to the basics, the primary role of face masks is to prevent the transmission of SARS-CoV-2 via respiratory droplets, which can easily enter through the mouth and nose to infect new hosts.

This protection is provided thanks to the complete barrier between your nasal and oral cavities with the outside world.

However, if you wear the mask in a floppy manner that doesn't cover your mouth and nose properly, your risk of catching the virus will increase.

In fact, the CDC has released several statements and how-to infographics to emphasize the importance of properly wearing face masks, asking people to:

Put the mask over the nose and mouth, then secure it under the chin Try to fit

it snugly against the sides of the face

One major benefit of wearing face masks is to protect the people around you, especially if you have been exposed to someone carrying the virus or work in a crowded place (e.g., grocery stores, restaurants, airports).

Because of this risk/benefit analysis, many governments around the world released emergency laws that mandate the public to wear masks in an attempt to stop the further spread of COVID-19.

Face masks considerably reduce COVID-19 cases in Germany

Face masks against COVID-19

Takeaway message

Wearing masks is an extremely important habit that every one of us should be applying since this simple step could significantly reduce the risk of transmission.

Hopefully, this article helped you appreciate the role of masks.

*

World Medical Card is your personal medical profile controlled by you. It is always available via web, app or physical card.

It's understood by Healthcare professionals all over the world.

By sharing your profile with relatives or travel companions, you help the ones closest to you to assist when required.

– WMC -makes is easy to gather all critical information in one place

– WMC -makes it easy to communicate critical information digital and physical

– All information in the WMC profile is user defined and personal

– WMC profile information is stored and protected securely

CHAPTER IV

IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE

Covid-19 has built a completely new frequency, and the people have realized themselves moving into a new world. While currently our society speedily transforming, we need to be quick to be able to answer fresh specifications, which have encircled all of us. Making risk-free surroundings will be priority of every person's mind so that the life can be conductive just as before. Options need to be taken to secure all those going back to our workplace and to keep ourselves and our loved ones devoid of problems. Brand-new plans have methodized every day in order to meet policies and regulation. Although, face masks are getting to be a whole new implemented standard for daily life, yet, to build safe environment that contributes to public safety, it becomes necessary to be observant throughout day and to take action against those who have not wearing mask in public places or work places. Many parts of society seem to be accepting some Covid tracking tools for safety. One of the most important tools is face mask detector. This system enables to identify who is without a required face mask. These systems work with existing surveillance systems along with innovative neural network algorithms to check whether a person has worn a face mask or not. In this chapter, we will briefly discuss the artificial intelligence and its subsets namely machine learning and deep learning, deep learning frameworks followed by a simple implementation for face mask detection system.

CHAPTER V

COVID-19: FACE MASK DETECTOR

In this tutorial, we'll discuss our two-phase COVID-19 face mask detector, detailing how our computer vision/deep learning pipeline will be implemented.

From there, we'll review the dataset we'll be using to train our custom face mask detector.

I'll then show you how to implement a Python script to train a face mask detector on our dataset using Keras and TensorFlow.

We'll use this Python script to train a face mask detector and review the results.

Given the trained COVID-19 face mask detector, we'll proceed to implement two more additional Python scripts used to:

1. Detect COVID-19 face masks in images
2. Detect face masks in real-time video streams

5.1 Two-phase COVID-19 face mask detector

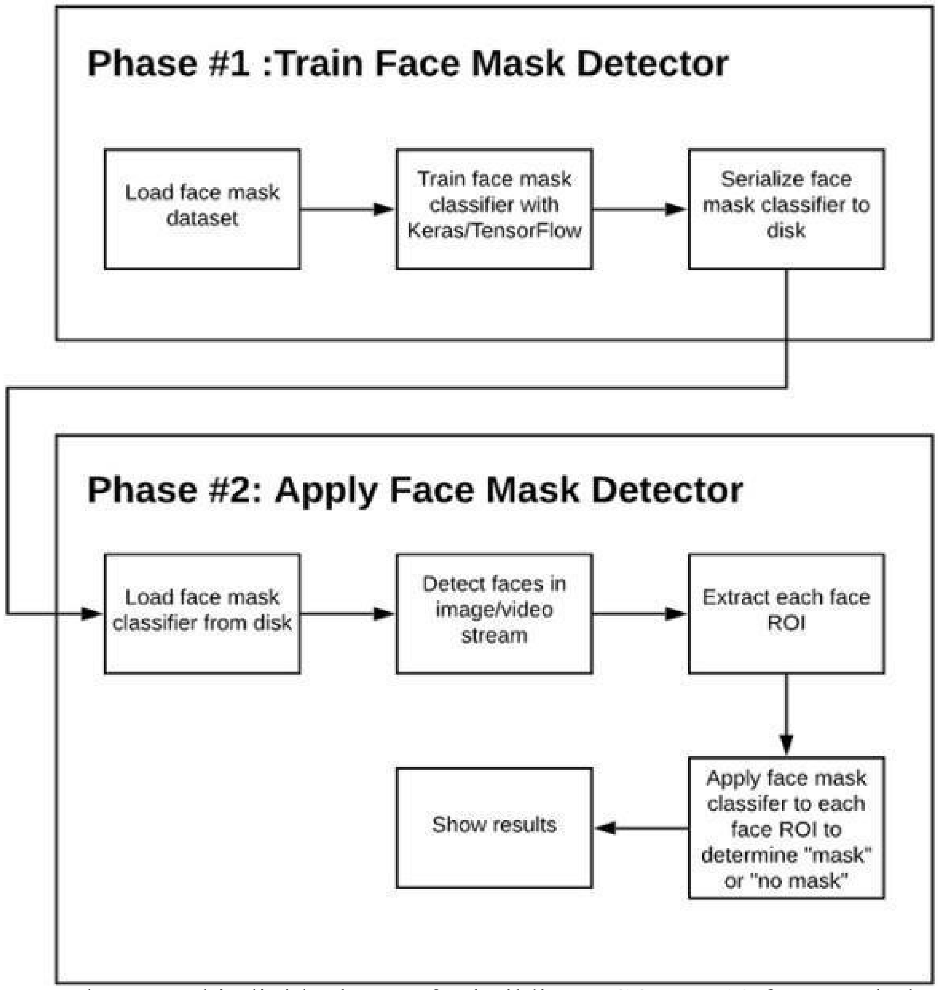

Figure 5: Phases and individual steps for building a COVID-19 face mask detector

In the above figure, Phases and individual steps for building a COVID-19 face mask detector with computer vision and deep learning using Python, OpenCV, **and TensorFlow/Keras.**

In order to train a custom face mask detector, we need to break our project into two distinct phases, each with its own respective sub-steps (as shown by Figure 1 above):

1. Training: Here we'll focus on loading our face mask detection dataset fromdisk, training a model (using Keras/TensorFlow) on this dataset, and then serializing the face mask detector to disk

2. Deployment: Once the face mask detector is trained, we can then move on to loading the mask detector, performing face detection, and then classifying each face as with_mask or without_mask

5.2 Our COVID-19 face mask detection dataset

Figure 6: A face mask detection dataset consists of “with mask” and “without mask” images.

In the above figure, A face mask detection dataset consists of “with mask” and “without mask” images.

We will use the dataset to build a COVID-19 face mask detector with computer vision and deep learning using Python, OpenCV, and TensorFlow/Keras.

This dataset consists of 1,376 images belonging to two classes:

with_mask: 690 images

without_mask: 686 images

Our goal is to train a custom deep learning model to detect whether a person is or is not wearing a mask.

5.3 How was our face mask dataset created?

To create this dataset,

1. Taking normal images of faces
2. Then creating a custom computer vision Python script to add facemasks to them, thereby creating an artificial (but still real-world applicable) dataset

This method is actually a lot easier than it sounds once you apply facial landmarks to the problem.

Facial landmarks allow us to automatically infer the location of facial structures, including:

1. Eyes
2. Eyebrows
3. Nose
4. Mouth
5. Jawline

To use facial landmarks to build a dataset of faces wearing face masks, we need to first start with an image of a person not wearing a face mask:

Figure 7: Person Without Mask

In the above figure, To build a COVID-19/Coronavirus pandemic face mask dataset, we’ll first start with a photograph of someone not wearing a face.

From there, we apply face detection to compute the bounding box location of the face in the image:

Figure 8: Face Detection

In the above figure, The next step is to apply face detection. Here we’ve used a deep learning method to perform face detection with OpenCV.

Once we know where in the image the face is, we can extract the face

Figure 9: Face Mask

In the above figure, An example of covid 19/corona virus facemask shield.

An example of a COVID-19/Coronavirus face mask/shield. This face mask will

on the This mask will be automatically applied to the face by using the facial landmarks (namely the points along the chin and nose) to compute where the mask will be placed.

face ROI automatically since we know the face landmark locations.

Figure 10: Properly Masked Face

In the above figure, Properly masked face of someone.

Figure 11: Face Mask Images

In the above figure, An artificial set of COVID-19 face mask images is shown. This set will be part of our "with mask" / "without mask" dataset for COVID-19 face mask detection with computer vision and deep learning using Python, OpenCV, and TensorFlow/Keras.

However there is a caveat you should be aware of when using this method to artificially create a dataset!

If you use a set of images to create an artificial dataset of people wearing masks, you cannot "re-use" the images without masks in your training set — you still need to gather non-face mask images that were not used in the artificial generation process!

If you include the original images used to generate face mask samples as non-face mask samples, your model will become heavily biased and fail to generalize well. Avoid that at all costs by taking the time to gather new examples of faces without masks.

CONCLUSION

We have hence created a well integrated real time face mask and social distancing violation detection system, where object detection takes place using YOLO v4. The three classes that are simultaneously detected are masked and unmasked faces, as well as whole people. Using the coordinates given by the detection of the class person, the relative distance between 2 individuals is hence estimated using the principles of optics. After rigorous testing, we observe that the model yields fairly accurate results for a wide field of view, which is an essential criteria for usage in public places. Without any addition of time consuming computations or image warping, this light weight model is easy to calibrate and can be well used in real time due to high FPS and good accuracy.

REFERENCES

1.Zhanchao Huang, Jianlin Wang "DC-SPP-YOLO: DenseConnectionand Spatial Pyramid Pooling Based YOLO for ObjectDetection"

2.Shiming Ge1, Jia Li2∗ , Qiting Ye1,3, Zhao Luo1,3 "Detecting Masked Faces in

the Wild with LLE-CNNs"

3.Joseph Redmon Ali Farhadi "YOLOv3:An Incremental Improvement"

4.Narinder Singh Punn, Sanjay Kumar Sonbhadra and Sonali Agarwal "Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques"

5.Prateek Khandelwal1, *Anuj Khandelwal1, *Snigdha Agarwal1, Deep Thomas1, Naveen Xavier1,Arun Raghuraman1 "Using Computer Vision to enhance Safety of Workforce in Manufacturing in a Post COVID World"

6.Wei Liu[1], Dragomir Anguelov[2], Dumitru Erhan[3], Christian Szegedy[3], Scott Reed[4], Cheng-Yang Fu[1], Alexander C. Berg[1]"SSD: Single Shot MultiBox Detector"

7.Shuo Yang[1] Ping Luo[2,1] Chen Change Loy[1,2] Xiaoou Tang[1,2]"WIDER FACE:
A Face Detection Benchmark"