Object Detection in Video

Submitted By

Akshay P. Patel

20MCEC09

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2021

Object Detection in Video

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering

Submitted By

Akshay P. Patel

(20MCEC09)

Guided By

Dr Jaiprakash Verma

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2021

Certificate

This is to certify that the major project part 2 entitled **"Object Detection in Video"** submitted by **Akshay P. Patel (20MCEC09)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this Major Project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Dr Jaiprakash Varma
Internal Guide & Associate Professor
CSE Department
Institute of Technology
Nirma University, Ahmedabad

Dr Sudeep Tanwar
Professor & PG Coordinator (M.Tech - CSE)
CSE Department
Institute of Technology
Nirma University, Ahmedabad

Dr Madhuri Bhavsar
Professor & Head
CSE Department
Institute of Technology
Nirma University, Ahmedabad

Dr Rajesh Patel
Director
Institute of Technology
Nirma University, Ahmedabad

Acknowledgements

I wanted to make this project useful and accessible to everyone. And as the great things aren't done alone, this project would be totally impossible without faculty guide Dr. Jai Prakash Verma and Project In-charge Dr. Priyank Thakkar who kept me motivated and dedicated for this project and helped me in every situation. And, without family and friends, this would be nothing but a dream. The surprising fact was that even I asked for help from professors even any faculty of the college and they offered their advice and helped me without any concerns. So, Kudos to all the people who encouraged me for this project. This project came to its very existence by getting to know reasons behind importance of time. I also wanted this project to be a source of inspiration for our fellow juniors who have an incredible opportunity to make something that really does matter for project completion. This is dedicated to everyone who's been with me.

- Akshay P. Patel

20MCEC09

Abstract

A recent advances in the discovery of face mask have made the object detection a hot topic for research. In the fields of computer vision object detection is the basic but a very tough and difficult task. With its ability to extract the powerful features, essential for the accretion to detect an object. A Deep learning methods has discovered it's areas of practice in the field over the past few years after that it get the success. We have powerful methods, but the conditions for obtaining a face and decide whether that face has mask or not has their different and complex challenges, it is like real-time detection, changing climate, and complex lighting conditions. In this research project, We have focus Single Object Detection in Video, which aims to confirm whether the person wear the mask or not by using object detection method. In year 2020 a dangerous virus disease named COVID-19 has affected our daily life and negatively affected the communal health, world trade and global economy. As Covid-19 virus spreads through the population, mutations and symptoms will be changed. In march 2021 new delta variant detected. As per the latest news recently new variant named Omicron is detected in Botswana. To contribute to public health, the aims of this paper is to explore a more accurate and real-time approach that can better visualize a non-mask face in public. Therefore, the discovery of a face mask has become an important task of helping the international community. There are extraordinary styles of algorithms to be had, YOLOV4 stands out from all the different gift currently. The custom dataset have been used to understand face masks and have been skilled on those dataset for detection and monitoring. For evaluation of the skilled model, Precision, keep in mind and Accuracy turned into calculated, it really works by using comparing the ground-fact bounding field vs the detected box and, in the long run, returns the accuracy rating. The higher the Accuracy score could be, the better version is within the detection of items.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

In this chapter we discuss about the Object Detection in Video for face mask detection and how it plays roles in the Domain of real life. The chapter discusses in brief about why face mask detection which play an important role in Object Detection in Video.

1.1 Problem Summary and Introduction

The unfold of COVID-19 is more and more worrying for everybody within the world.The virus can infect a person to another person by airborne and droplets. In the absence of effective anti retroviral and limited medical resources, WHO recommended many measures to control the rate of infection and to avoid eliminating limited medical resources. From this pandemic wearing a masks is one of the non-pharmacological interventions that can be used to reduce the main source of SARS-CoV2 droplets that are expelled through an inflamed individual. With the exception of the talk about medical services and the variety of masks, all countries approve public nose and mouth coverings. According to instructions from WHO, to reduce the spread of COVID-19, anybody desires to wear a face mask, avoid crowded places and always hold the immune gadget. consequently, to shield every other, every body ought to wear a face mask well whilst they may be going out or meeting others. But, a few non-respondents refuse to wear a face masks for so many excuses. in addition, to enhance the face mask the detector is very essential in this example. This paper aims to improve face masks detector that could hit upon any type of face masks.

Object acquisition can be done using traditional techniques or morden techniques. Traditional techniques incorporate image processing techniques and morden techniques in-

corporate deep learning networks. For this research we have choose deep learing methods because Deep learning based object detection is significantly more robust for a complex scenes and a challenging illumination. Deep Learning methods often rely on supervised training. Performance is limited by the ability to integrate rapidly growing GPUs each year. There are sorts of algorithms for item detection primarily based on Deep Learning.

1. *Two stage based object detection :* Two stage based object detection algorithms are Fast RCNN and Faster RCNN, RCNN and SPPNet , Mask R-CNN, Pyramid Networks/FPN and G-RCNN (2021).

2. *One Stage based object detection :* One stage based object detection algorithms includes YOLO, SSD , RetinaNet, YOLOv3,YOLOv4 and YOLOR.

Object detection using deep learning method extract the features from the input image or from the video frame.
Two subsequent tasks are solved by this object detection method:
Task 1: Find the arbitrary number of objects (It also can be zero)
Task 2: Classify all the objects and estimate the size of an object along with bounding box.
To make the process easier, we can break these tasks into two stages. While other methods include both the tasks into the one step. One stage detectors predict bounding boxes over this images without the Object region proposal with conventional computer vision method or deep networks.
The process of One stage detectors is time consuming too so it is used in real time programs.Great advantage of single stage that those algorithms are usually faster there are multistage detectors and they are simple in structure. Thus, to detect the face mask, we have chosen deep learning based One Stage object detection algorithm YOLO V4.

The unfold of COVID-19 is more and more traumatic for everyone inside the global. The virus can be inflamed in a person to individual via airborne and droplets. In the skive of effective anti retrival and limited medical resources, WHO recommended many measures to control the rate of infection and to avoid eliminating limited medical resources. From this pandemic wearing a masks is one of the non-pharmacological interventions that may be used to reduce the primary source of SARS-CoV2 droplets which are expelled by

an infected individual.

To overcome the problem i came up with the solution implemented in the form of Object Detection in video using deep learning techniques. To perform this method (You Only Look Once) YOLO v4 is used. This model will give the output that whether the person wear the mask or not.

1.2 Problem Specification

The spread of COVID-19 is increasingly worrying for everyone in the world. The virus can be infected in a person to person by airborne and droplets. In the absence of effective anti retroviral and limited medical resources, WHO recommended many measures to control the rate of infection and to avoid eliminating limited medical resources. From this pandemic wearing a mask is one of the non-pharmacological interventions that can be used to cut the main source of SARS-CoV2 droplets that are expelled by an infected person.

1.2.1 Need Analysis

The Need Analysis of our study are listed below:

- Demonstrating the technical effectiveness of a deep learning approach to accelerate deployment of safeguards against the global Covid19 pandemic.
- Create an innovative and flexible face mask identification paradigm that recognizes and prevents Covid19 through deep learning.
- Advanced Deep Convolution YOLOv4 neural network architecture to discover saliency features and classify it.
- To provide quality and the quantitative analysis using the log loss price,consider,accuracy and f1 score.

1.3 Scope of Work

To overcome the problem i came up with the solution implemented in the form of Face mask detection using deep learning techniques. To perform this method (You Only Look

Once) YOLO v4 is used. This model will give the output that whether the person wear the mask or not.

Chapter 2

Literature Survey

In this literature phase, many procedures had been proposed for the item detection based totally on deep mastering techniques. So, I have labeled this section based totally at the specific deep gaining knowledge of strategies and additionally I've got covered the summary of paper in Table 1.

2.0.1 Two stage based Object Detection

The state of art of two stage object detection algorithm As presented in [2] Isunuri B Venkateswarlu, et al. had done research in 2020 to detect the masked faces using CNN. For this research he employs a global pooling layer to carry out a flatten of the feature vector. a completely connected dense layer associated with the softmax layer has been applied for the class. This approach gives result when I train it but when I have to apply it to CCTV footage then it may gives bad result which is not satisfactory. In [4] Yu Liyan, Sansan Zhou et al. presented the enhanced RCNN network in 2018 to discover and see more objects in the image. RCNN is the primary convolutional neural network implemented to object detection in a discovery model. They pick out candidate regions based on traditional selective search. Then i set up an RCNN to extract features from the photograph. item detection for RCNN models requires 4 steps. There are selective search, CNN feature extraction, category, and b box regression. In [5] Krishan Kumar, Alok Negi, R.S. Rajput and Prachi Chauhan were presented the VGG16 and a CNN models in 2021. Both the models are deep educational that recognizes a camouflage and exposed person to help keep track of security violations and maintaining a safe working atmosphere. This study uses the concept of data suppression, data suits, normalization and

transmission. This technology may require observations where observations are needed as shopping centers, hospitals, transportation centers, restaurants and other conferences in other communities. In order to implement CNN, filters 16, 16, 32, 32, 128 and 128 with 3 x 3 are used, respectively, and 10 layers of the convolution are implemented, and RELU is used as activation. The model applied up to 5 layers at a speed of 2 and then applied the plane layer. Then, at first, second and third density of the first, second and third density of 512, 128 and 64 hidden nodes, using RELU activation functions Density is generated. Using the softmax activation function, a fourth dense layer with two hidden nodes is used for the final output.

In [18] T. A. Gulliver et al. were applied the Faster RCNN method to detect the pedestrians by applying training and tuning VGGNet in 2016 [19] for pedestrian-only networks. Compared to the RCNN-based fast pedestrian detection mentioned in Ref. [20] J. Li, X et al. had provided an acceptable failure rate as well as a faster test speed in 2018. Before RPN Kmeans clustering algorithm is added in the Ref. [21] H. Zhang et al.had given the method to perform the pedestrian detection task in 2017. In this article authors had main focus to devoted the computational time issues. In the process of generating a sentence, it takes a long time to search the entire image, and the area occupied by pedestrians is small, and there are many unnecessary sentences, increasing the learning and learning complexity. As i have mentioned that the state network and our idea is kind of similar. Instead of using the RPN directly to generate some of the initial candidates Kmeans clustering algorithm is used to get the proposal, and it is passed to the RPN. Therefore, the calculation time is shortened. In [22] K. Saleh et al. had used the Faster RCNN infrastructure to detect the cyclists in 2017. Generation module is added in this article and that generates a composite depth image that is sent to the backbone network compared to the original structure. A depth image can represent the shape of an object, and DCNN can learn it more efficiently. Even if you use a computer generator for learning, performance is much better when testing a deep image converted from a real image. In 2016 [23] Fan Q et al. had done research on Product results The RCNN detected automobile is faster to purchase and see a series of large experiments and analysis. The creator [24] of the link uses the image created by the UAV as an item of the detector and improves the efficiency to detect the object of these special images. To use the faster RCNN image in the block images without doing any changes, the HYPERREGION OF-

FER Network is used additionally. In 2017 [25] Tang, T had provided method to improve the performance in small objects detection using the RCNN. For that authors uses the several enhanced classifiers to filter the interest participants to remove incorrect detections. That two modifications growth the accuracy of detection for small vehicles and reduce the number of false positives such as objects such as vehicles.

Previously, there were many implementation of traffic scenarios using the Faster RCNN. In [26] Q. Zhou, X. Wang and T. Li et al. had used the Faster RCNN architecture by doing small modification in the code to detect the full traffic sign. The introduced community performs a coarse seek to find possible areas of hobby, and the RPN is used to filter the areas of hobby, lowering computation because the RPN now not desires to swipe through the complete practical map to generate suggestions. After the CONV layer DeCONV layer is used to helps in boom the accuracy of detection for small gadgets. The magnified feature map generated from the high-meaning DeCONV layer is used together with the attention map, which contains information about small features, so both surface and deep information are used. The Local Contextbased Faster RCNN introduced in reference [28] which will add the local context layer. The reason behind that is to detect the traffic signs belonging from the small objects. For capturing the information from around to the target. The local context layer is used after the RPN transforms the each sentence into three sentences, extending horizontally and vertically respectively. It then combines all the sentences and stores the surrounding information for final discovery.

In [29], the face mask detection system uses Fast RCNN as its base model. Considering that the color of the facemask is often uniform and unique, the system implements the region suggestion task using Maximum Stable Extreme Regions (MSER) [30,31] instead of SS [32]. After development of Fast RCNN Pedestrian detection is also the benefit. In [20] Li et al. was proposed a architecture containing two subnets that would put in force pedestrian detection with the people at the one-of-a-kind spatial scales. offers generated by means of the SS set of rules are served from 2 subnets at simultaneously after several layers of CONV. To complete the final discovery two output feature maps were combined with the weighting strategy. After training with large pedestrian input, output of the large subnet scores higher, so this subnet largely determines the final detection result.

I have summarized the paper and organize based on frameworks in Table 1.

Reference	Model	Modification	Limitation
[29]	Fast R-CNN	Using MSERS instead of SS	Different implementation platforms for region proposal and detector
[20]	Fast R-CNN	Different spatial scales of pedestrians	Additional 2 FC layer parameters
[26]	Faster R-CNN	Adding attention network	Traffic lights should be treat as signs
[28]	Faster R-CNN	Adding a local context layer to focus on small objects	There is No distinction in between traffic signs
[24]	Faster R-CNN	Detection on UAV images	Faster R-CNN detector has no change
[25]	Faster R-CNN	Use boosted classifiers	Long computation time
[18]	Faster R-CNN	Training and Tuning the specific VGG	Not good for poor quality images
[33]	YOLO	Multi-functional YOLO	Powerless to color markings
[34]	YOLO	Multi-directional detection	Due to the camera oblique there is no consideration of plate deformations
[35]	YOLO	Add Average pooling layer instead of FC layer	Need to do pre-processing on Night scene images
[38]	YOLOv2 object tracking	Using Kalman filter and Hungarian algorithm	When people are overlap at that time detection is not enough stable
[39]	YOLOv2 people	Combining detection with re-identification	Weight factor is a determined parameter
[40]	YOLOv2 pedestrian	Simplifying with 9 CONV layers	Bad performance on overlapping objects
[41]	YOLOv2	Adding 3 3×3 and 1 1×1 CONV layers	Experimental article
[42]	YOLOv2	Using 1×1 filters	Low accuracy for small signs
[43]	YOLOv2 vehicle	Using K-means cluster and Grid of size 14×14	For distant objects accuracy is not high
[44]	YOLOv3 pedestrian	Using an image enhancement policy	Pre-processing is needed on input images
[45]	SSD on-road objects	Fine-tuning the SSD on KITTI [46] datasets	Real-time processing is slow
[46]	SSD pedestrian	Use small patches	Long time for segNet
[47]	SSD	Using text detector and a FCN and	Time-consuming with sliding window
[49]	SSD	Using Inceptionv3 instead of VGG	It is hard to detect error of target objects from mirror images

Table 2.1: Summary of Framework-based Object Detection

2.0.2 One stage based Object Detection

One stage based detection algorithm's performance if very good in terms of speed compared to the two stage detection based algorithms series. For that reason, in this phase i have include the nation of artwork the usage of unmarried level detector for static photograph detection, and additionally it is object detection in video streams. As presented in [1], Susanto, et al, introduced YOLO based on a deep learning approach in 2017 to differentiate the goal from the white ball associated with a soccer ball. The NVIDIA JETSON TX1 control board is designed for this algorithm. As presented in [3] Apoorva Raghunandan et al. had done research on the Object detection algorithm for face detection, colour detection, skin detection, Targeted detection and shape detection is performed and manipulated using the MATLAB 2017b for the various types of video surveillance tools in 2018. In this process various face components were detected using the Viola Jones Algorithms. Output of this algorithm shows the various parts of the face with respect to Indicates nose detection, Indicates eye detection and also shows the detection for all the features like nose,eyes and mouth. In [6] Liu, et al implemented the work using YOLO in 2018. on this studies paintings, they completed the conventional photograph processing of shooting blur, sound and clear out rotation in the actual international. to enhance the visitors signal detection they had used the YOLO set of rules to educate a sturdy model. In 2018 [7] Jan et al. had used the YOLO set of rules for face detection in actual-time programs with brief detection times. In [11] Jiang et al., have developed a face mask detector called Retina Face Mask in 2020. The evolved version is a unmarried-degree detector and includes a function pyramid community to mix high-degree semantic records with a couple of feature maps, and an interest module to hit upon facemasks.

In year 2018 [39] Van Lanst et al. combines the YOLOv2 model with a re identity network into a one framework which could speedy recognize and re pick out the people in other photos additionally. The YOLOREID framework [39] supports the "mixed-end" or the "split-end" architecture with 128-value embedded output in each cell instead of using classified output. Now Darknet19 weights can be shared with the discovery network and the reidentification network when combined. Consequently, YOLOREID can perform each tasks with a negligible boom in complexity compared to YOLOv2. In 2018 [40] Heo, D et al. proposed the YOLOv2 is simplified in that it contains only 9 CONVs, 6 maximum poolings, and 2 Fully Connected layers. Compare to the original YOLO v2

model where the image is input directly into the network. A small YOLOv2 model is combined with the feature map from the Adaptive Boolean Mapbased Saliency kernel for this evening's pedestrian detection task. This can show that pedestrian prominence has higher performance than the background. In 2017 [41], Jensen's studies that last CONV layer of the YOLOv2 model has been removed. This is for the purpose of improving traffic light detection performance, the author has added the four CONVOLUTION layers, from that four layers three of had the kernel sizes of 3x3 and 11x1. When it comes to detecting small objects like traffic signs, by using YOLOv2 and based on the original YOLOv2 Zhangetal. In [42] Zhang, J et al. is getting three different models. The change occurred in the middle tier and a new CONV tier with a kernel size of 1x1 was inserted to hold the cross-channel information. In this article, I propose a deadlock detection path using YOLO-inspired ConvDet to compute bounding containers and perform type. The ConvDet elegance removes the final FC layer from YOLO and includes the idea of a binding container from RPN. As a end result, fewer model parameters are used in queeDet and sentences may be generated for the same number of levels in that model domain compared to YOLO.

YOLOv2 is actually superior to native YOLO and researchers are implementing this popular framework to solve various object detection problems. Blended with the Kalman clear out and Hungarian set of rules, YOLOv2 creates a real-time device that can music more than one special classes of objects simultaneously [38]. in this system, YOLOv2 acts as a trigger that prevents detecting items in the first frame. In the next frame with YOLOv2 detection of body t 1, the Kalman filter is liable for generating predictions for body t. It then makes use of this prediction to healthy the detection consequences over the tframe to decide if this clear out is correct. Van Ranst et al. In reference [39], YOLOv2 is combined with a re-identification community into a unmarried framework which can hastily detect and re-discover people in exceptional photos. In place of using categorical heuristics, the YOLOREID platform [39] uses default heuristics of 128 values in step with mobile to support "split" or "mixed" architectures. consequently, Darknet19 weights may be shared and re-identified across seek networks whilst merged. So with a moderate increase in complexity in comparison to YOLOv2, YOLOREID can do each. For reference [40], YOLOv2 is simplified as it contains only 9 CONVs, 6 max aggregates and 2 FC classes. Compare the image to the original YOLOv2, taken directly from

the network. In this nocturnal pedestrian detection mission, a small YOLOv2 combined with a BooleanMap-based adaptive salinity kernel (ABMS) functional map derived from a BooleanMap-based adaptive salinity kernel (ABMS) could reveal pedestrian serviceability. over the background. Jensen [41] skipped the last CONV layer of the YOLOv2 structure. To improve the traffic light detection performance, the author added 4 CONV layers, of which 3 kernel sizes are 3×3, and 1 layer has a kernel size of 1×1.et al. [42] There are three models based on the original Yolov2. The fertilization occurs in the intermediate layer and a new convection layer is inserted with a core size of 1 x 1 to gain information about the course. Because deep networks lose properties of small items which include avenue signs, a few calling layers are removed from the small layers. The consequences display that the proposed model improves each accuracy and velocity. To in addition improve vehicle detection, a 7×7 grid can be changed with a REF[43] to a 14×14 grid that could do away with larger functions from the same community. consequently, extra statistics can be conveyed to the graph, which improves the popularity accuracy of the machine. for the reason that the YOLOv2 anchors have been originally created from a shared dataset in place of a car, the authors used the Kmeans algorithm to institution bounding boxes to determine the range and size of anchors for factors. that is for a selected purpose.

YOLOv3 itself isn't always a entire framework as it's far the author's interim research report. consequently, there are not many packages based totally on YOLOv3, specifically in positive scenes, which includes traffic scenarios. The work developed with the aid of Q[44] is based on Yolov3 for pedestrian detection provided on this paragraph. I progressed item detection using deep studying techniques in the 30:13 reversal scenario in article with pre-processed training patterns to minimize environmental influences such as changes in lighting conditions or people density. Then, these preprocessed images were injected into Yolov3 to detect pedestrians and increase their accuracy. In 2016, several months after the first SSD proposal, Kim et al. If SuperRoom [45] shows improved performance, apply this structure. very own model with SSD reference model built on Pascal VOC. on this statistics set, small objects including pedestrians and cyclists perform worse than automobiles. So a skinny SSD bezel is used, plus an extra element ratio to reduce the a whole lot large thing ratio due to photograph enlargement. The effects of this text show that combining an improved SSD version with a data growth strategy can

improve performance. In fused, the DNN fusion method and DNN segmentation division are performed in the DNN fusion method and DNN segmentation division with the DNN fusion network (FDNN) [46] to perform a pedestrian detection operation in the DNN fusion network (FDNN) [46]. SSD FDNN is used to generate pedestrian candidates for source pics. these applicants are then blanketed in different DNNs, together with more than one DNN classifiers to clarify proposals. After a small object in a large image, I have developed a frame that uses a small patch generated in a large image generated in a large image. The kernel of this network consists of many small objects that are used to detect objects in each patch used. Based on SSDS SOSCNN, it is characterized by the fact that the first fourth layer layer is stored in the VGG16 network. So, I achieve an image pyramid by performing one prediction only on SOSCNN and generating an image pyramid instead of a function prediction of a function generated in another transform layer. In other words, i take different magnified images from SOSCNN over the same network and extract different sized object maps. SSDs have the advantage of being able to detect small objects as they predict gadgets on exceptional function maps. for that reason, detection of avenue signs and symptoms, which are considered small gadgets as compared to motors and pedestrians, could benefit from the use of SSD chassis. hyperlink [48] contains a signature detector that detects supply photograph markers through FCN, and a text detector that extracts text and determines precisely what the text is. This article inherits a text detection mechanism called TextBoxes from SSD, which has the disadvantage of extracting the object map from the last CONV class.

In real time application YOLO model series is very popular to detect the object because it's keep the balance between detection speed and accuracy. But for real time object detection in video still needs power consumption plateform and high performance. Now, the problem is how to choose appropriate model for object detection. For that In this paper I proposed the YOLOv4 model and compare with other state of art model and will evaluate them based on best accuracy as well best performance in terms of time.

Chapter 3

Dataset Description

Flickr-Faces-HQ (FFHQ) dataset have been used for experimenting the face mask detection. In this dataset have total 4000 images. These 4000 images are divided into two class one is "With mask" and second is "Without mask". Around 2000 images having proper mask is included in "With mask" class and other 2000 images are without mask or with improper mask are included in "Without mask" class.Total size of this dataset is 1.16 GB.

Figure 3.1: Dataset Images

3.1 Characteristics of Dataset

- FFHQ dataset include almost 4k images for the Object detection.
- This is a dataset for the classification of binary classes with labled "With mask" and "With out mask" and it containing the substantially more data than the previous benchmark datasets.

Chapter 4

Proposed Method

4.1 Proposed Research Work

In year 2020-21 a dangerous virus disease named COVID-19 affected our daily life by infecting people health which became cause of many life losses. To break the chain of virus spread WHO recommended many measures to control the rate of infection and to avoid eliminating limited medical resources. Wearing a mask is one of the measure which is most important to control spread of COVID-19 due to this virus.

Now in day to day life there are many places where government body needs to manually check in camera that how many people/employees actually wear the mask. Thus, to reduce the human efforts I proposed the face mask detection technique. I implementaded this work the usage of a deep neural network version named YOLOv4. YOLOv4 can run two times as fast as different deep neural community strategies used for object detection. The performance of YOLOv4 version will be improveed by 10% than YOLOv3 AP and about 12% FPS. Reflecting these results, it is well suited to implement the method in interval mask detectors wherever high detection accuracy is required.

Objectives of this research are as follows:

- Design and Development of Face mask detection using You Only Look Once (YOLO) v4 model.
- Comparison of proposed method with state of art research work.

Bochkovskiy et al [56] 2020 proposed YOLOv4 for certain big modifications from its archetype YOLOv3, which has endorsed full-size upgrades regularly in terms of pace

as well accuracy. YOLOv4 may be very speedy, easy to fix, durable, stable, and gives promising effects even within the smallest detail, which is why I have selected it as our 1 workspace issue. With a photo / enter body, you get things in it less than 3 classes - blank face and man or woman faces. This means that the same model is used in the singular recognition following community deviations and identification covered exterior. This improves dramatically standard productivity and job specificity.

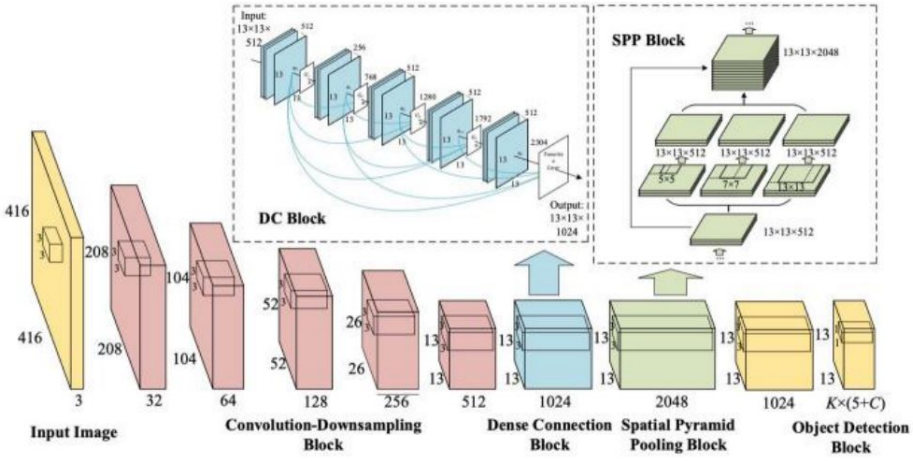

Figure 4.1: Yolo v4 Architecture

The cylinder has three sections, backbone, neck, and head. I enterprise awaits RGB photo or frame. spine to respond by using removing highlights from the image. This move-level-Partial-Darknet (CSPDarknet53) join [56] at the end it has been a very good selection. In [14], the harvest is inside the foundation layer divided into parts. One is going to Dense Block and the other goes without delay to the next development layer as exhibition show. 2b. Thick squares cowl the layers and every the layer includes Batch Normalization and ReLU followed by using a layer of convolution. every layer of Dense Block takes partial tips for all previous layers as hooked up. This expands the spinal cord accessible and assists visibility highlights of the image. Integrating the local pyramid (SPP) used as a neck band containing squares expand the acquisition field and the interface of the visual interface from the variety spinal levels. The combination of SPP and YOLO pipe is approx presented in Fig. 1. In this organisation, the idea of the Fund's Freebies (BoF) were found inside the loose references within the preparation techniques stepped forward the agency's presentation externally including to its dimension costs. there are numerous loose styles of that browse, plus, deliver Cutmix, Mosaic ex-

tensions data, DropBlock. related to the usage of decided on magnificence marker spine. techniques, as an example, SelfAdversarial training, CIoU-misfortune, cross Mini-Batch Normalization (CBN), mosaic information development, framework the involvement, and the homicide of DropBlock changed into acknowledged as one of the benefits of identity. unique provide (BoS) the add-on approach is customary while 'special' is noted systems enhance community performance even as expanding viewing costs at low fee.

YOLO model treats the item detection as a proprietary regression trouble, without delay from the photo pixels to the bounding field coordination and item chances. An integrated network predicts the a couple of sure packing containers and the chances of those boxes. YOLO implements full-size image detection as well as inversion improving the receiver performance. For detection of the objects integrated model has an advantages over traditional methods.

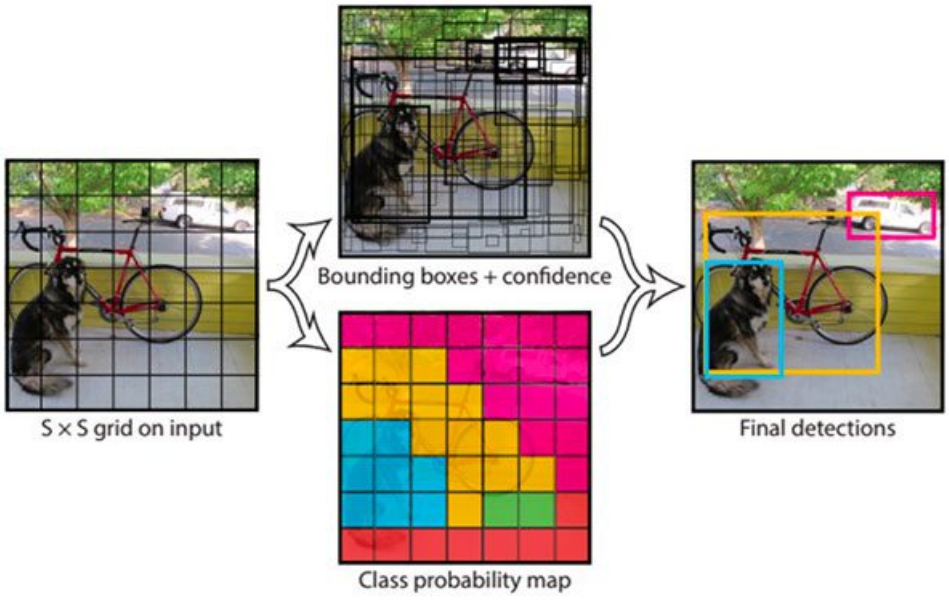

Figure 4.2: YOLO Model

In all the grid cells it divides the image into an MxM grid, and predicts B-binding boxes, the probability of the object and the confidence level of the predictive binding boxes. Each cell grid predicts B-binding boxes and certainty in these boxes. Those confidence figures show how confident the model is in how the box contains the object and the way appropriately it thinks the field and the anticipated items. clearly I am defining the self-confidence as Pr(gadgets) IOU. If the mobile is empty then it should be zero as an

conceitedness score. otherwise, the vanity score identical to the intersection over union (IOU) among the predicted box and as a consequence the lowest reality each bounding container includes the self assurance and 5 predictions: x, y, w, h. Where co-ordinates (x, y) is representing the center of the box compare to the grid cell of boundaries. While the height and width prediction is depending on the whole image/picture. Finally, the prediction of arrogance represents the IOU between the expected box and the basic truth box. The conditional class probabilities Pr(Classi % | Object) is predicted by every grid. These opportunities are converted to the grid cell that holds by the object. I predict only one set of complex objects that can be complex per cell, regardless of the number of boxes B.

The YOLO interface has 24 bendy layers followed by 2 completely connected layers. It definitely makes use of 1×1 reduction layers observed by way of 3×3 rapid YOLO convolutional layers that exercise the neural community with the nine convolutional layers in place of 24 and a few filters between those layers. aside from the quantity of community, all of the education and the trying out parameters are equal for the YOLO version and the fast YOLO.

A YOLO is optimized for squared sum blunders in our version because the output. It does a total mistakes of as it's clean to optimize, although it would not align to maximize average precision. It has a uniform weighting of placement errors with a non-prototype classification error. In addition, in each image, many cells of the grid do not contain any objects. This will push the "confidence" of many of these cells to zero, often reducing the slope of the cells containing the objects. This will make the model unstable, leading to early training divergence. To change this, YOLO enhances the lack of bounding box coordinate predictions and it reduces the lack of self assurance prediction for the bins that do not incorporates any object. YOLO is uses two parameters, wire and nobj to acquire this. YOLO defines coord = 5 and noobj = zero.5. The sum squared errors is likewise weighted inside the large and small bins similarly. Its mistakes metric ought to replicate that small deviation inside the huge containers, however the small boxes. managing that is being able to predict a bit the bottom of the box certain width and the peak instead of the exact width and the period.

YOLO predicts more than one bound cells according to grid cellular. At schooling time, I want best man or woman bounding box predictors responsible for each item. I assign one

predictor responsible for predicting one supported item whose prediction has the cutting-edge pleasant IOU with lower truth. This ends in a special of between the bounding container predictors.

4.1.1 Overview of YOLOv4

Yolov4 is the improved version of Yolov1, Yolov2 and Yolov3. As illustrated in Figure 2, it uses the CSPDarkNet53 as the main structure of the network. When I compare the cspresnext050 and the efficaceenetb3, the Yolov4 structure introduces Pyramid group (SPP) to the CSPDarknet53, which helps the YOLOv4 to improve the size of Reception field significantly. It can get more parameters as an input without reducing in the operation speed. Because of this feature it is works better in the classification tasks. When we calk about the data development, Yolov4 is using mosaic to create four images in one, eventually it will increases the size of the small section and adds an automatic training (SAT) to create an update of nerve cells and interfere with the image in pictures. In addition, YOLOv4 is also useing the PAN to consolidate the multi-channel feature to avoid information loss.

From the above figure it can be observed that the One-Shot Detector has 4 main components, those 4 main components are:

1. *Input:* The input to the detector can be an image or video based on the use cases specified in the research.

2. *Backbone:* The backbone of the object detector contains models, these models can be ResNet, DenseNet, VGG.

3. *Neck:* The neck in the detector acts as an extra layer, which goes in parallel to the backbone the head.

4. *Head:* The head is the network that is in charge of the detection of objects based on bounding boxes.

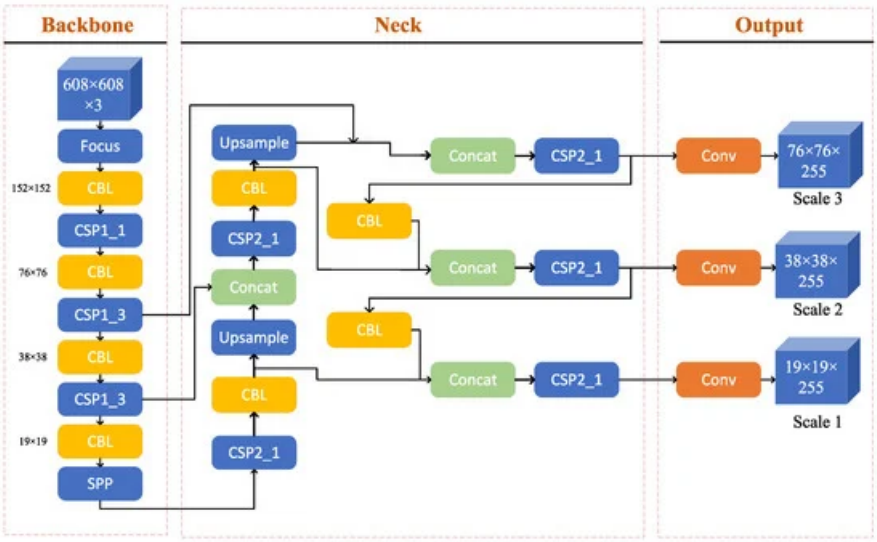

Figure 4.3: YOLOv4 Network

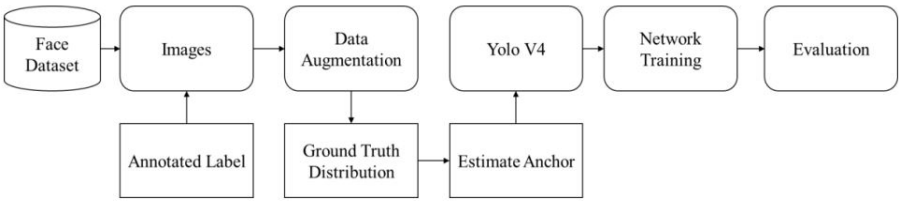

Figure 4.4: Proposed System Block diagram

4.1.2 Advancement in YOLOv4 in comparison to prior YOLO models

- It is a proficient and authoritative object detection model that allows individuals with a 1080 Ti or 2080 Ti GPU to training a very fast and accurate object detector.
- The consequences of state-of-the-art "Bag-of-Freebies" and "Bag-of-Specials" object detection procedures all the while detector training was confirmed.
- The born-again progressive ways covering CBN (Cross-iteration batch normalization), PAN (Path aggregation network), that area unit larger masterful and applicable for single GPU coaching.

Chapter 5

Execution and Implementation

5.1 Work Flow

Here, the workflow of the YOLOv4 object detection algorithm is discussed in detail. Initially, an image dataset is collected and used for training through the use of YOLOv4. The dataset includes images of people with and without masks. Figure 4 shows the YOLOv4 workflow.

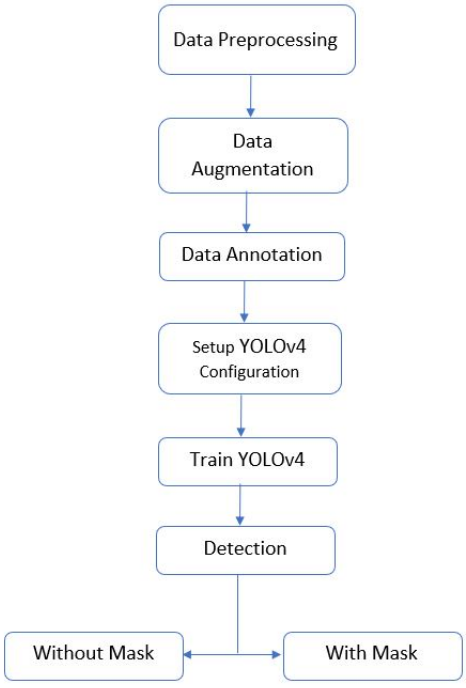

Figure 5.1: YOLOv4 Work Flow

5.1.1 Data Preprocessing

The hyper-parameters of YOLOv4 become used as follows: There are pix with diffent length so, that we've resized the enter length pix and set it to 608 × 608 pixels to facilitate detection. initial getting to know price became set as zero.001 and increased with a thing 0.1 at 5000 steps and 5500 steps, respectively. for you to perform multi-scale education all architectures used a single GPU with batch length sixty four and mini-batch length sixteen. The default momentum 0.949, IoU threshold zero.213 and loss normalizer 0.07 become done as proposed by way of authors of YOLOv4.

5.1.2 Data Augmentation

To increase the performance of our model, data augmentation is used. It is introducing the neural network with the large number of variations of inputs. In our case, there are not many such data types available. In our dataset, most of the images are masked and unmasked fronts. Therefore, it is necessary to increase the data to obtain the better result. In order to maximize data diversity on the face and prevent the model from learning unimportant patterns, we have used flip-flop data augmentation in our pre-processed data. Hidden and uncovered images, as shown in Figure 5. By default, the YOLOv4 architecture has some data enhancement techniques such as Cut Mix, Mosaic data augmentation, classroom label smoothing, Self Adversarial (SAT) training, which also helps our model to achieve accuracy. more accurate.

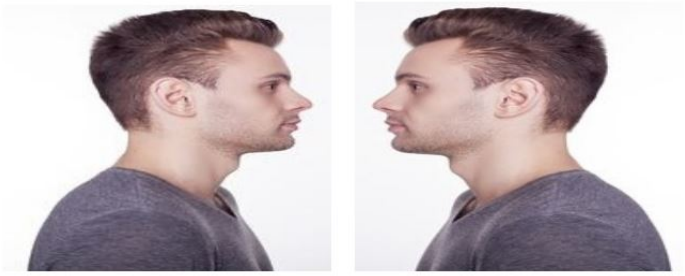

Figure 5.2: Horizontal Flip

5.1.3 Data Annotation

Data annotation is very important for our model. It is nothing but data or photo labeling. There are various types of data annotations, such as the picture captions, the text annotations, and the video captions. In our model, we have used the picture captions.

For this we have to draw the binding boxes in the pictures as a rectangle. There are many tools available for this purpose. We have used the Ybat tool to annoting the image. This is a long and difficult task, as the pictures have to be interpreted by hand.

5.1.4 Setup YOLOv4

We mainly need three files to configure the YOLOv4:

1. object.names file:

   It contains the names of classes. One in each line. In our case that is 2 classes name mask and no-mask

2. object.data file:

   It simply contains the number of the classes, path of the training and the testing data and location for the backup.

3. custom_yolov4.cfg file:

   It contains the information about the width, filters, height, steps, max_batches, burnout etc. In our case we set batch size to 64, split to 16 and learning rate to 0.001. Also set layer to 2 for three YOLO blocks and filter size to 21.

5.1.5 Training YOLOv4

The YOLOv4 model takes a captured photograph and divides it into grid cells, each accountable for a particular description of objects. In each mobile grid, the self belief college is calculated using its compound container. efficiently separates statistics into fits and identifies characteristic features from this mesh. traits seen with high self assurance in adjoining cells are grouped right into a unmarried location for version overall performance. Our model is now equipped for schooling once the setup is complete. We used 80 % information for training and 20 % for verification. We started education the usage of the Darknet framework, included into the Google research lab, to teach our real-time mask discovery model. Darknet develops an underlying community architecture and serves as the muse for YOLO architecture. Fig. 6, illustrates steps wherein the version split the photo. Following that, it demonstrates how the capabilities are discovered, and then, it displays the recognized object.

Figure 5.3: Workflow of YOLOv4

5.2 Evaluation Measures

The performance of the YOLOv4 method is compared using the recall, precision, F1 score,specificity and Mean Average Precision (mAP) as the evaluation metrics. Intersection over Union, also called the Jaccard index, is used to calculate the accuracy of an object detector in a given data set. Specificity is the measure of how many negative predictions made are true negatives (correct).

Precision is the measurement of how many of the predicted positives values are actually positive.

Recall is the measurement of how many of the true positives values are correctly classified. F1 score, also called F score, is a function of precision and recall and it is needed to maintain a balance between precision and recall.

Average Precision (AP), one of the most used metrics to measure the accuracy of object detectors, is used to find the area under the precision-recall curve.Mean of the Average Precision (mAP) is the average of AP calculated for all classes.

Chapter 6

Results and Discussion

In order to show the detection performance of YOLOV4 method, 800 test images were used and the results of the method presented in Table 3.

Class	With mask	Without mask
With mask	565	44
Without mask	27	164

Table 6.1: Average precision values of the YOLOv4 for each class

As we can see in Table 3 the precision, recall, F1 score and specificity of the proposed method were 92.77%, 895.44% and 93.58%, 78.84% respectively. In this study the average IoU value was observed as 79.48%. Furthermore, the performances of the YOLOv4 model for each class are shown in Table 2.

Method	Precision(%)	Recall(%)	F1-Score(%)	Specificity(%)
Yolov4	92.77%	95.44%	93.58%	78.84%

Table 6.2: Detection results of the YOLOv4 method

6.0.1 Analysis

As can be seen from Table 4, the model shows the best performance with accuracy 91.12% in identifying suitable mask wearers.

The accuracy results obtained from this study were compared with those reported in the literature (Table 3). As can be seen from Table 3, in studies where face mask detection

Reference	Model	Training Dataset	Detection	Accuracy(%)
This study	YOLOv4	400 with mask, 400 without mask and with improper mask	- With mask - Without mask	91.12%
[3]	MATLAB 2017b	100 with mask, 150 without mask	- With mask - Without mask	86.24%
[6]	YOLO	100 with mask, 100 without mask	- Without mask - With mask	83.46%
[8]	KNN, SVM and MobileNet	161 with mask, 161 without mask	- Without mask - With mask	87.8%
[8]	SVM	161 with mask, 161 without mask	- Without mask - With mask	89.4%
[8]	MobileNet	161 with mask, 161 without mask	- Without mask - With mask	88.7%
[12]	Resnet50	511 with mask, 511 without mask	- Without mask - With mask	47.7%
[12]	MobilenetV2	511 with mask, 511 without mask	- Without mask - With mask	91.2%
[16]	CNN-based cascade framework	776 with mask, 776 without mask	- Without mask - With mask	86.6%

Table 6.3: Accuracy results comparison with related works

was performed before, it was determined that only people were wearing masks or not. As a result of the study, when the accuracy score of 91.12% obtained with the YOLOv4 model was compared with the alternative solutions in the literature, it was observed that the detection performance of the solution was high.

Chapter 7

Conclusion and Future Work

In this study, the YOLOv4 detection model was used for detecting face masks effectively. Test results show that the YOLOv4 model achieved a high accuracy of 91.12 % which is the average accuracy of the points. Therefore, it can be used to find cases of not wearing a mask, the mask does not comply with the rules and wearing the mask properly. This study will help reduce the spread of the coronavirus by facilitating the identification of people without masks or who do not wear masks properly in public places such as schools, shopping malls, railway stations and markets. In future I can expand this project and also check the social distance for public places like shopping malls, railway stations and markets. I also propose that in online examination we should use this face detection method so, that all students can give exam honestly. Also This system can be used at traffic signals if the driver breaks the traffic rules then this system will recognized the number plate of vehicle and send the e-memo.

Bibliography

[1] Rudiawan, E., Analia, R., Sutopo, P.D. and Soebakti, H., 2017, September. The deep learning development for real-time ball and goal detection of barelang-FC. In 2017 International Electronics Symposium on Engineering Technology and Applications (IES-ETA) (pp. 146-151). IEEE.

[2] I. B. Venkateswarlu, J. Kakarla and S. Prakash, "Face mask detection using MobileNet and Global Pooling Block," 2020 IEEE 4th Conference on Information Communication Technology (CICT), 2020, pp. 1-5, doi: 10.1109/CICT51604.2020.9312083.

[3] A. Raghunandan, Mohana, P. Raghav and H. V. R. Aradhya, "Object Detection Algorithms for Video Surveillance Applications," 2018 International Conference on Communication and Signal Processing (ICCSP), 2018, pp. 0563-0568, doi: 10.1109/ICCSP.2018.8524461.

[4] Yu, Liyan, Xianqiao Chen, and Sansan Zhou. "Research of image main objects detection algorithm based on deep learning." 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). IEEE, 2018.

[5] Negi A, Kumar K, Chauhan P, Rajput RS. Deep neural architecture for face mask detection on simulated masked face dataset against COVID-19 pandemic. In2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) 2021 Feb 19 (pp. 595-600). IEEE.

[6] Liu C, Tao Y, Liang J, Li K, Chen Y. Object detection based on YOLO network. In2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC) 2018 Dec 14 (pp. 799-803). IEEE.

[7] Yang W, Jiachun Z. Real-time face detection based on YOLO. In2018 1st IEEE international conference on knowledge innovation and invention (ICKII) 2018 Jul 23 (pp. 221-224). IEEE.

[8] W. Vijitkunsawat and P. Chantngarm, "Study of the Performance of Machine Learning Algorithms for Face Mask Detection," 2020 - 5th International Conference on Information Technology (InCIT), 2020, pp. 39-43, doi: 10.1109/InCIT50588.2020.9310963.

[9] B. N. K. Sai and T. Sasikala, "Object Detection and Count of Objects in Image using Tensor Flow Object Detection API," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), 2019, pp. 542-546, doi: 10.1109/ICSSIT46314.2019.8987942.

[10] I. Kilic and G. Aydin, "Traffic Sign Detection And Recognition Using TensorFlow' s Object Detection API With A New Benchmark Dataset," 2020 International Conference on Electrical Engineering (ICEE), 2020, pp. 1-5, doi: 10.1109/ICEE49691.2020.9249914.

[11] M.Jiang and X.Fan, "RetinaMask:A Face Mask detector," Accessed:Oct.30,2020.[Online].arXiv Prepr.arXiv2005.03950,2020.

[12] S. Srinivasan, R. Rujula Singh, R. R. Biradar and S. Revathi, "COVID-19 Monitoring System using Social Distancing and Face Mask Detection on Surveillance video datasets," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 449-455, doi: 10.1109/ESCI50559.2021.9396783.

[13] L. H. Jadhav and B. F. Momin, "Detection and identification of unattended/removed objects in video surveillance," 2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), 2016, pp. 1770-1773, doi: 10.1109/RTEICT.2016.7808138.

[14] S. W. Moon, J. Lee, J. Lee, D. Nam and W. Yoo, "A Comparative Study on the Maritime Object Detection Performance of Deep Learning Models," 2020 International Conference on Information and Communication Technology Convergence (ICTC), 2020, pp. 1155-1157, doi: 10.1109/ICTC49870.2020.9289620.

[15] Hu Q, Paisitkriangkrai S, Shen C, van den Hengel A, Porikli F. Fast detection of multiple objects in traffic scenes with a common detection framework. IEEE Transactions on Intelligent Transportation Systems. 2015 Dec 3;17(4):1002-14.

[16] Bu, Wei, Jiangjian Xiao, Chuanhong Zhou, Minmin Yang, and Chengbin Peng. "A cascade framework for masked face detection." In 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), pp. 458-462. IEEE, 2017.

[17] G. Deore, R. Bodhula, V. Udpikar and V. More, "Study of masked face detection approach in video analytics," 2016 Conference on Advances in Signal Processing (CASP), 2016, pp. 196-200, doi: 10.1109/CASP.2016.7746164.

[18] X. Zhao, W. Li, Y. Zhang, T. A. Gulliver, S. Chang and Z. Feng, "A Faster RCNN-Based Pedestrian Detection System," 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), 2016, pp. 1-5, doi: 10.1109/VTCFall.2016.7880852.

[19] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[20] J. Li, X. Liang, S. Shen, T. Xu, J. Feng and S. Yan, "Scale-Aware Fast R-CNN for Pedestrian Detection," in IEEE Transactions on Multimedia, vol. 20, no. 4, pp. 985-996, April 2018, doi: 10.1109/TMM.2017.2759508.

[21] H. Zhang, Y. Du, S. Ning, Y. Zhang, S. Yang and C. Du, "Pedestrian Detection Method Based on Faster R-CNN," 2017 13th International Conference on Computational Intelligence and Security (CIS), 2017, pp. 427-430, doi: 10.1109/CIS.2017.00099.

[22] K. Saleh, M. Hossny, A. Hossny and S. Nahavandi, "Cyclist detection in LIDAR scans using faster R-CNN and synthetic depth images," 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017, pp. 1-6, doi: 10.1109/ITSC.2017.8317599.

[23] Fan Q, Brown L, Smith J. A closer look at Faster R-CNN for vehicle detection. In2016 IEEE intelligent vehicles symposium (IV) 2016 Jun 19 (pp. 124-129). IEEE.

[24] Xu, Y., Yu, G., Wang, Y., Wu, X. and Ma, Y., 2017. Car detection from low-altitude UAV imagery with the faster R-CNN. Journal of Advanced Transportation, 2017.

[25] Tang, T., Zhou, S., Deng, Z., Zou, H. and Lei, L., 2017. Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining. Sensors, 17(2), p.336.

[26] Z. Zuo, K. Yu, Q. Zhou, X. Wang and T. Li, "Traffic Signs Detection Based on Faster R-CNN," 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), 2017, pp. 286-288, doi: 10.1109/ICDCSW.2017.34.

[27] Yang, T., Long, X., Sangaiah, A.K., Zheng, Z. and Tong, C., 2018. Deep detection network for real-life traffic sign in vehicular networks. Computer Networks, 136, pp.95-104.

[28] Cheng, P., Liu, W., Zhang, Y. and Ma, H., 2018, February. LOCO: local context based faster R-CNN for small traffic sign detection. In International conference on multimedia modeling (pp. 329-341). Springer, Cham.

[29] Qian, Rongqiang, Qianyu Liu, Yong Yue, Frans Coenen, and Bailing Zhang. "Road surface traffic sign detection with hybrid region proposal and fast R-CNN." In 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), pp. 555-559. IEEE, 2016.

[30] Greenhalgh, J. and Mirmehdi, M., 2012. Real-time detection and recognition of road traffic signs. IEEE transactions on intelligent transportation systems, 13(4), pp.1498-1506.

[31] Matas, J., Chum, O., Urban, M. and Pajdla, T., 2004. Robust wide-baseline stereo from maximally stable extremal regions. Image and vision computing, 22(10), pp.761-767.

[32] Uijlings, J.R., Van De Sande, K.E., Gevers, T. and Smeulders, A.W., 2013. Selective search for object recognition. International journal of computer vision, 104(2), pp.154-171.

[33] M. Kawano, K. Mikami, S. Yokoyama, T. Yonezawa and J. Nakazawa, "Road marking blur detection with drive recorder," 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 4092-4097, doi: 10.1109/BigData.2017.8258427.

[34] L. Xie, T. Ahmad, L. Jin, Y. Liu and S. Zhang, "A New CNN-Based Method for Multi-Directional Car License Plate Detection," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 2, pp. 507-517, Feb. 2018, doi: 10.1109/TITS.2017.2784093.

[35] J. Tao, H. Wang, X. Zhang, X. Li and H. Yang, "An object detection system based on YOLO in traffic scene," 2017 6th International Conference on Computer Science and Network Technology (ICCSNT), 2017, pp. 315-319, doi: 10.1109/ICCSNT.2017.8343709.

[36] Dai, J., Li, Y., He, K. and Sun, J., 2016. R-fcn: Object detection via region-based fully convolutional networks. Advances in neural information processing systems, 29.

[37] Wu, Bichen, Forrest Iandola, Peter H. Jin, and Kurt Keutzer. "Squeezedet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 129-137. 2017.

[38] K. Jo, J. Im, J. Kim and D. -S. Kim, "A real-time multi-class multi-object tracker using YOLOv2," 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2017, pp. 507-511, doi: 10.1109/ICSIPA.2017.8120665.

[39] W. V. Ranst, F. De Smedt, J. Berte and T. Goedemé, "Fast Simultaneous People Detection and Re-identification in a Single Shot Network," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018, pp. 1-6, doi: 10.1109/AVSS.2018.8639489.

[40] Heo, D., Lee, E. and Ko, B.C., 2018. Pedestrian detection at night using deep neural networks and saliency maps. Electronic Imaging, 2018(17), pp.060403-1.

[41] M. B. Jensen, K. Nasrollahi and T. B. Moeslund, "Evaluating State-of-the-Art Object Detector on Challenging Traffic Light Data," 2017 IEEE Conference on Com-

puter Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 882-888, doi: 10.1109/CVPRW.2017.122.

[42] Zhang, J., Huang, M., Jin, X. and Li, X., 2017. A real-time chinese traffic sign detection algorithm based on modified YOLOv2. Algorithms, 10(4), p.127.

[43] Yang, W., Zhang, J., Wang, H. and Zhang, Z., 2018, May. A vehicle real-time detection algorithm based on YOLOv2 framework. In Real-Time Image and Video Processing 2018 (Vol. 10670, p. 106700N). International Society for Optics and Photonics.

[44] H. Qu, T. Yuan, Z. Sheng and Y. Zhang, "A Pedestrian Detection Method Based on YOLOv3 Model and Image Enhanced by Retinex," 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2018, pp. 1-5, doi: 10.1109/CISP-BMEI.2018.8633119.

[45] H. Kim, Y. Lee, B. Yim, E. Park and H. Kim, "On-road object detection using deep neural network," 2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), 2016, pp. 1-4, doi: 10.1109/ICCE-Asia.2016.7804765.

[46] Du, Xianzhi, Mostafa El-Khamy, Jungwon Lee, and Larry Davis. "Fused DNN: A deep neural network fusion approach to fast and robust pedestrian detection." In 2017 IEEE winter conference on applications of computer vision (WACV), pp. 953-961. IEEE, 2017.

[47] Z. Meng, X. Fan, X. Chen, M. Chen and Y. Tong, "Detecting Small Signs from Large Images," 2017 IEEE International Conference on Information Reuse and Integration (IRI), 2017, pp. 217-224, doi: 10.1109/IRI.2017.57.

[48] Zhu Y, Liao M, Yang M, Liu W. Cascaded segmentation-detection networks for text-based traffic sign detection. IEEE transactions on intelligent transportation systems. 2017 Dec 25;19(1):209-19.

[49] J. Müller and K. Dietmayer, "Detecting Traffic Lights by Single Shot Detection," 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018, pp. 266-273, doi: 10.1109/ITSC.2018.8569683.

[50] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.

[51] Peng H, Guo S, Zuo X. A Vehicle Detection Method Based on YOLOV4 Model. In2021 2nd International Conference on Artificial Intelligence and Information Systems 2021 May 28 (pp. 1-4).

[52] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[53] Nayyar, A., Jain, R. and Upadhyay, Y., 2020, August. Object detection based approach for Automatic detection of Pneumonia. In 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD) (pp. 1-6). IEEE.

[54] Taneja, S., Nayyar, A. and Nagrath, P., 2021. Face Mask Detection Using Deep Learning During COVID-19. In Proceedings of Second International Conference on Computing, Communications, and Cyber-Security (pp. 39-51). Springer, Singapore.

[55] Mukhopadhyay, M., Pal, S., Nayyar, A., Pramanik, P.K.D., Dasgupta, N. and Choudhury, P., 2020, February. Facial emotion detection to assess Learner's State of mind in an online learning system. In Proceedings of the 2020 5th International Conference on Intelligent Information Technology (pp. 107-115).

Appendix A

- https://viso.ai/deep-learning/object-detection/
- https://stats.stackexchange.com/questions/230125/get-precision-and-recall\
  value-with-tensorflow-cnn-implementation
- https://www.analyticsvidhya.com/blog/2021/05/alleviation-of-covid-by-means-of-so
- https://www.analyticsvidhya.com/blog/2021/05/alleviation-of-covid-by-means-of-so