ISSN No:-2456-2165

Social Distancing Technique for Covid-19 Using Yolo V5 & CNN for Fast object Detection and better Accuracy

[1]Baibhav Pathy

Abstract:- The world has fallen in the crunches of a deadly virus that has caused pandemics and recessions all over the world. It has forced even the superpower country like USA and RUSSIA to go into lockdown and thus decrease the Gross domestic product (GDP) of the economy. So to prevent the further spread of the virus, Awareness was required until the vaccine with full functionality of giving us immunity against any variant of covid. One way to stop the further spreading of the virus is social distancing. In this paper, we are implementing a deep-learning algorithm along with Yolo V5. This project will use OpenCV, Deep Learning, Computer Vision, and YOLO V5 to work together and through surveillance cameras to create social distance between people by constantly analyzing video input that will be fed into the designed system through the surveillance cameras and will notify the authorities if any social distancing violations occur. The proposed method can assist save money and save the authorities who are required to keep people maintaining social distance from getting infected with covid-19. It can also substantially reduce covid19 deaths.

I. INTRODUCTION

USA has utilized 2.58 million CCTV covering 15.35 million individuals 131 (2020a) to keep a record of people and make illegal pursuit easier. As a result, there are six people allocated to each camera. The cameras are used to monitor the facial feature of individuals. All of this is achievable because of the recurrent neural network Lecun et al. (2015). Deep learning is the process of extracting many levels of abstraction from data to learn attributes. Since its inception, this computational model has been employed in a wide range of applications, from recognizing production process faults to accurately identifying celestial objects that would take a long time or be inconceivable to discover with artificial cognition.

COVID-19 has caused a pandemic in the year 2019 [17] to date, killing approximately 5.43 million individuals and infecting 283.20 million people worldmeter(2021)[18] 131 (2020b). Due to the lack of a vaccine, the World health organization (WHO) recommends using hand sanitizer and maintaining safe social distancing to reduce the virus's spread throughout the globe.

The goal of this study is to utilize a deep convolutional neural network to recognize individuals in photos or video feeds and then use that information to estimate the distance between them. Although much research has already been done in this field, we are revisiting it with the help of a new object identification framework, YoloV5 [5]. YoloV5, a cutting-edge object detection algorithm, is extremely powerful and rapid, making it suitable for use in surveillance cameras.

II. ABOUT DATASET

Most of the Images and videos that we have used are taken by our student volunteers. Other sources for the collection of data are googled and YouTube videos with open license. Search on google has a large number of photos from many sources, making it swift & easier to complete image-gathering tasks.

We have also used Ms Coco's data set. Microsoft released the MS COCO dataset[19], which is large-scale object identification, classification, and labelling dataset. picture collection was built with the objective of improving image identification, therefore COCO stands for Common Objects in Context. The COCO dataset offers demanding, significant visual datasets for object recognition, with the majority of the datasets containing state-of-the-art neural networks. COCO, for example, is frequently used to test the efficiency of the context of real-world identification systems. Sophisticated artificial neural packages automatically comprehend the COCO dataset's format.

III. METHODOLOGY

Individual identification requires a deep learning model to be trained with images of numerous individuals in various scenarios. The detection procedure is divided into four phases: a collection of data, data categorization, model training, and validation of the model with testing as shown fig 1.

ISSN No:-2456-2165

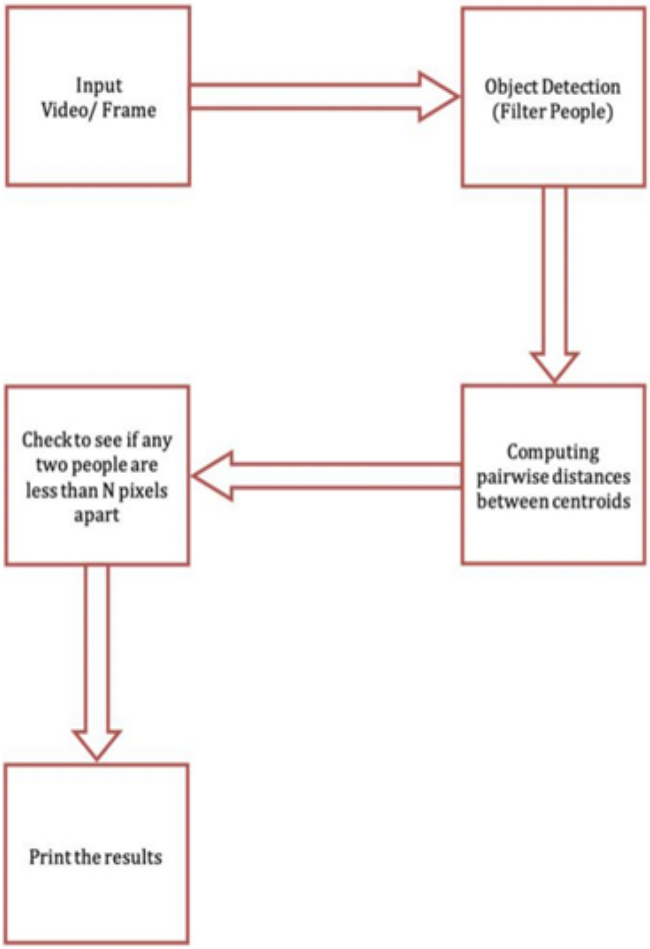

Fig 1 Flowchart of the Method to be followed while Executing the Model

➢ *Data Collection*

Data collection consisted of the collection of prerequisite data from external open license sources. Data used are videos and photos of the author taken from his smartphone and DLSR during various activities in college and in his nearby localities.

➢ *Data Categorization*

We must supply the predicted model parameters of the dataset while training and testing because we are utilizing supervised deep learning. Convolution layers are used to categories pictures in Deep learning for Object recognition. Below are some examples of how the photographs are labelled. To categories our training example, we utilized Labeling Bradski (2000) for mac.

➢ *Model Training and Validation of the Model with Testing*

The classifier is then trained once the data has been labelled. The classifier for people identification was trained using Yolov5. Configuring the hyper-parameters that play a big influence in the detecting mechanism's performance & reliability is a big part of fitting the network. A machine was used to run the training for thousand epochs. We perform the evaluation with the CNN model and show the results in an appropriate graph as shown in fig 2.

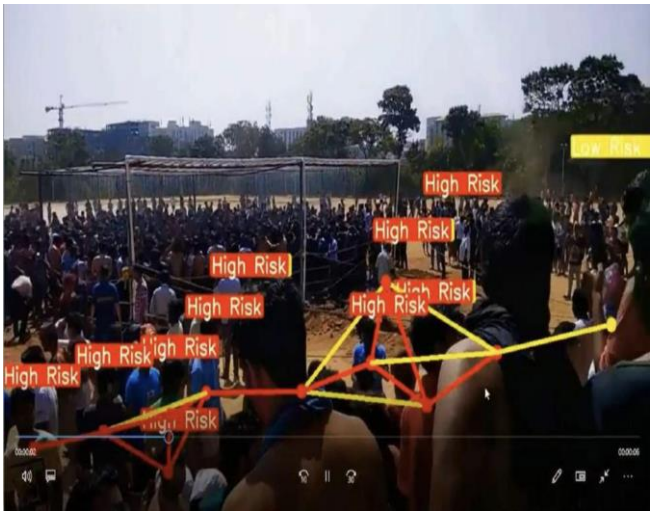

Fig 2 Output After Running the Model

IV. ABOUT YOLO V5

To locate the item inside the picture, all the previous object detection techniques have utilized areas. The system does not examine the entire picture. Rather, portions of the image with a high likelihood of containing the item. You Only Look Once, or YOLO is an object detection framework that differs significantly from the region-based techniques discussed previously. The bounding boxes and class probabilities for these boxes are predicted by a single neural network in YOLO.

YOLO works by splitting a picture into an S x S grid as shown in fig 3 and creating m bounding boxes inside each matrix. The model outputs a class probability and offset values for each frame for each bounding box. The bounding boxes with a class probability greater than or equal to a threshold value are chosen and utilized to find the item inside the picture. YOLO is folded higher quicker than any other object detection technique (45 frames per second). The YOLO algorithm's drawback is that it has trouble detecting tiny things in images; for example, it could have trouble group of people as shown in fig 4. This is owing to the algorithm's spatial restrictions.

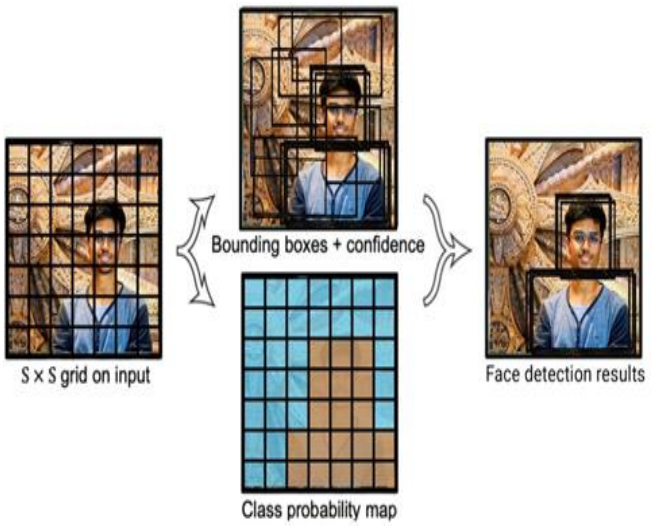

Fig 3 Class Probability Map

ISSN No:-2456-2165

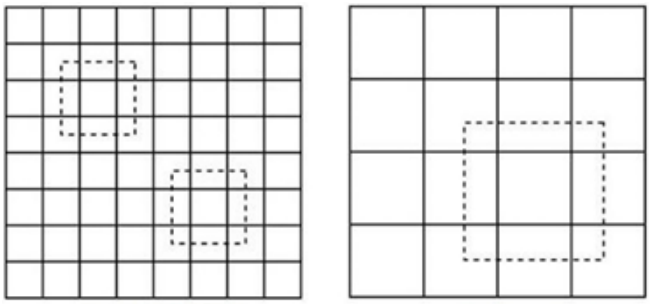

Fig 4 Feature Selection

V. PROPOSED MODEL

As shown Figure 5 depicts a high-level perspective of the suggested social distancing surveillance system. A dense framework was built on top of the YOLOv5 model for effective feature utilization and visualization. The Bbox can better match the form of a tomato, allowing for more exact localisation. Furthermore, the Bbox can calculate a more accurate IoU between predictions, which is critical in the NMS process, and therefore increase detection outcomes. YOLO- social distancing is the name of the suggested model. A flowchart of the YOLO- social distancing training and detection procedure is shown in Figure 5.

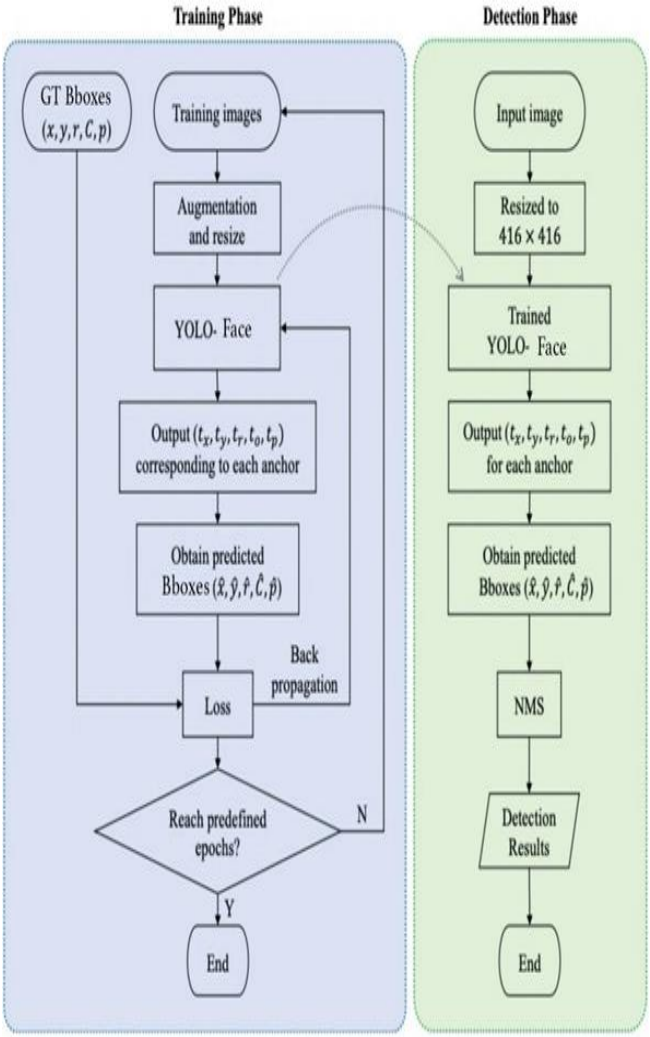

Fig 5 A flowchart of training and detection process of YOLO-Face

VI. IMPLEMENTATION

The method of calculating the isometric perspective of images collected from a certain angle is known as Inverse Perspective Mapping. The OpenCV Bradski (2000) module makes finding the inverse perspective mapping of any picture simple. We can acquire the projection of the entire picture to the top view by using four features of the image and translating them to actual points in an isometric perspective. The detailed structure of the picture acquired using this mapping approach is shown in figure 6.

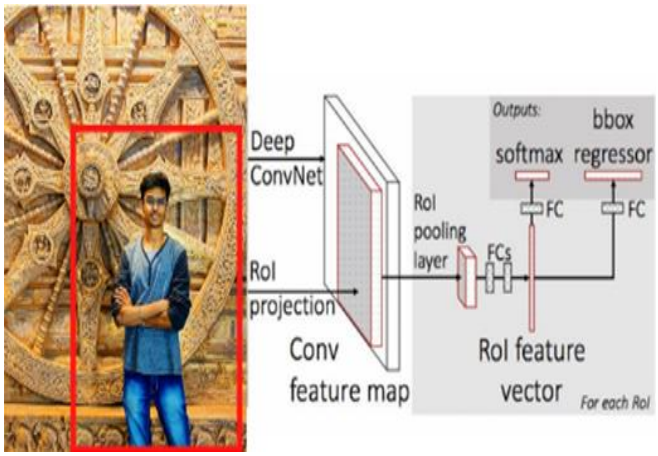

Fig 6 Flowchart of Implementation Method Layer by Layer

Fig 7 Final Output Footage

The 4 dots on both flanks of the frame reflect the visual references used to get the object's isometric perspective. This assessment was completed once, and we used that linear transformation matrix to calculate path length as shown in fig 7. The graph's resolution is essential to determine the true proximity between the items. To compute the difference between items in this research, we employed an estimated metric.

ISSN No:-2456-2165

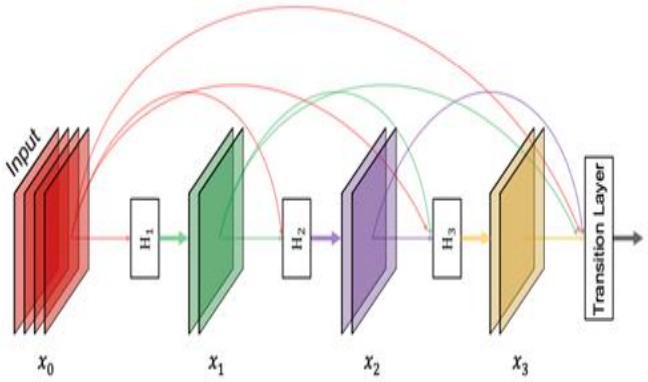

Fig 8 Dense four-layer block All previous feature maps are used as input for every tier, which in turn makes a significant contribution for all future levels. H_i denotes the operation BN-ReLU-Conv$_{1\times1}$-BN-ReLU-Conv$_{3\times3}$.

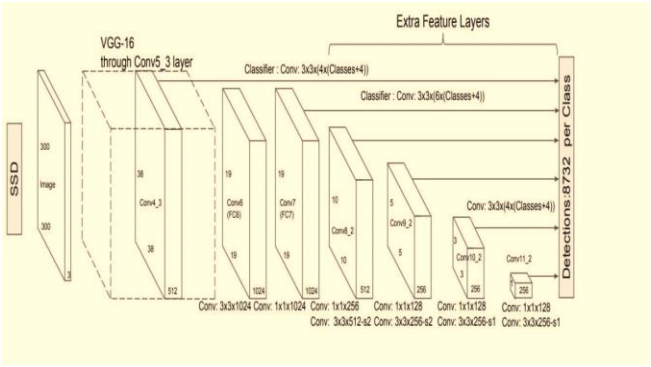

Fig 9 Depth Analysis of Each Layer

Extraction image features and implementing convolution are the two elements of SSD object recognition. Detection is performed by the SSD from a single layer. In fact, it detects objects separately using many layers (cross-function mappings) as shown in fig 8. The precision of the extracted features decreases when the spatial dimension of CNN is increasingly reduced. SSD detects larger-scale objects with lower resolution layers. The 4x4 extracted features, for instance, are employed for the massive object. After VGG16, SSD adds 6 extra auxiliary convolution layers to the picture as shown in fig 9. For object recognition, five of these levels would be implemented. We produce six forecasts instead of four in three of those levels. SSD uses 6 pooling layers to create 8732 forecasts in total.

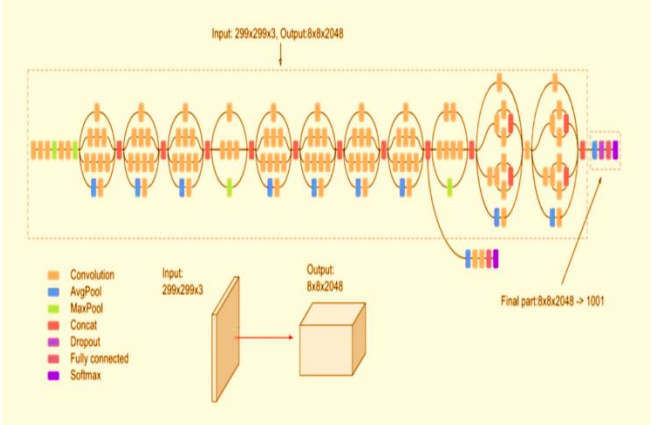

Fig 10 Neural Network Overall Representation

We then use the R - CNN neural network for model evaluation. To get around the challenge of picking a large couple of aspects in the image, Ross Girshick et al. devised an approach in which the limited selection is used to extract just that regions from the image, which he calls segmentation proposal region [9]. As a result, rather than attempting to identify a large number of locations. The selective search technique described below was used to create these region ideas. These proposed potential areas are twisted into a square and input into a CNN (convolutional neural network), which outputs a 4096-dimensional feature vector as shown in fig 10. The CNN acts as a feature extractor, and the resultant dense layer contains the characteristics collected from the picture, which are input into an SVM to evaluate the existence of the item inside the region proposals suggestion. The method anticipates values, which are offset features for raising the accuracy of the bounding box, in order to forecast the presence of an object inside the zone suggested.

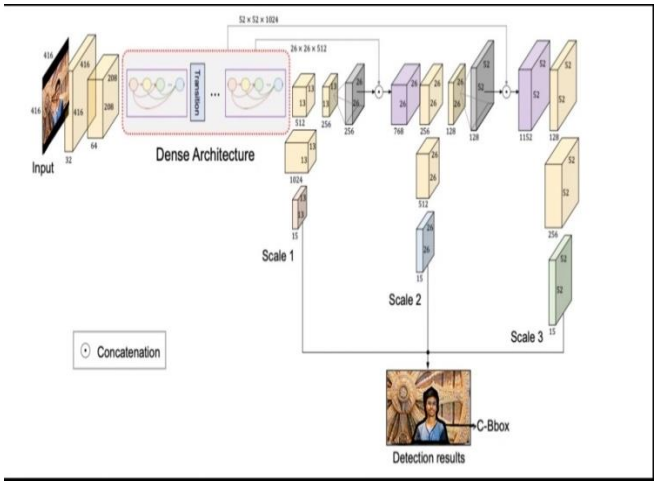

Fig 11 Overall Representation of Bounding Box Suggesting Presence of an Object Inside the Zone

➢ *Algorithm 1 The Pseudo-Code*

- *Input*:

✓ = {b_1, • • • , b_N}, ς = {C_1, • • , C_N}, λ_{nms}
✓ is the list of initial detection boxes
✓ contains corresponding detection confidences
✓ λ_{nms} is the NMS threshold

- *Output*: List of final detection boxes θ

✓ θ ← {}
✓ **while** B ≠ φ **do**
✓ m ← argmax C
✓ θ ← θ ∪ b_m; β ← β - b_m; ς ← ς - C_m
✓ **for** b_i ∈ β **do**
✓ **if** IoU (b_m, b_i) ≥ λ_{nms} **then**
✓ β ← β - b_i ; ς ← ς - C_i
✓ end if
✓ end for
✓ end while

ISSN No:-2456-2165

VII. MATHEMATICAL MODEL

A. *Network Architecture :*

➢ *Yolo Backbone*

The entire design of improved\sYOLOv5s is represented in Figure 3 which comprises the foundation, recognition collar, and recognition face. Firstly, a newly designed backbone termed CSPNet is employed. We update it with a new element called CBS consisting of a Convolution layer, a SILU and a BN layer[10]. Furthermore, a stem file is utilised to replace the centre layer in YOLOv5s. a C3 square is being utilized to recreate the preceding CSP block with two pieces. One is transported through a CBS block, numerous bottleneck frames, and a Conv layer, while the other one consists of a Convolution layer. After the two paths with a Concatination and a CBS block proceeded, we also change the SPP frame [11] to boost the face recognition efficiency. In this block, the dimension of the tri kernels is modified to relatively small kernels.

➢ *Recognition Collar*

The architecture of the detecting neck is also depicted in Figure 3 which comprises a conventional feature pyramid network (FPN) [12] and path aggregation network (PAN) [3]. However, we adjust the specifics of several sections, such as the CS module and the CBS block as shown in fig 12.

➢ *Recognition Face*

Through multilayer perceptron architecture and path consolidation [13] network, the front segment of the network achieves the full fusion of low-level features and high-level features to build rich feature maps, which can identify the most happening more often instances. However, for low-resolution photographs, feature fusion cannot improve the original information of the image, and after layers of iteration, the prior knowledge of small faces is still lacking. To boost the recognition rate of small faces in low-resolution pictures, SR is fused in the detection head component of the system. For the grid to be computed, the area data is entered into SRGAN to carry out high-level functional reconstruction and face detection again through its coordinate information. Finally, the output of the two-stage.

➢ *Loss of Functionality*

In detection systems, the IOU index is commonly utilized. It is utilized in most alignment [14] approaches not only to evaluate the favorable and unfavourable specimen but also to measure the difference between the projected box's position and the classification algorithm. The study suggests that the following factors be taken into account: intertwining area, convergence point proximity, and image resolution, all of which have sparked consternation. More scholars are proposing superior performance techniques, such as DIOU, IOU, CIOU and GIOU at the moment. In this study, we suggest replacing GIOU with CIOU and nonmaximal reduction in YOLOv5s (NMS) Our bounding box regression loss function is defined **as.**

$$l_{box}^{\prime} = 1 - IOU + \frac{\rho^2\left(b,\hat{b}\right)}{c^2} + \frac{16}{\pi^4}\frac{\left(\arctan\left(w\wedge/h\wedge\right) - \arctan\left(w/h\right)\right)^4}{1 - IOU + \left(4/\pi^2\right)\left(\arctan\left(w\wedge/h\wedge\right) - \arctan\left(w/h\right)\right)^2}, \quad \text{Eq no. 1}$$

Where b, b′ is the box's hotspot, symbolizes the Euclidean distance, c is the diagonal separation of the smallest encompassing rectangular box, and w, h is the target's physical size. Facial subjects are not only abundant but also layered in security footage pictures [15], resulting in several targets for each sector. However, basing decisions on a single criterion frequently results in low precision and recall [16]. As a result of the ring structure created by combining CIOU and NMS, the applicant box in the same matrix may be assessed and inspected many rounds, successfully overcoming the risk of skipped identification.

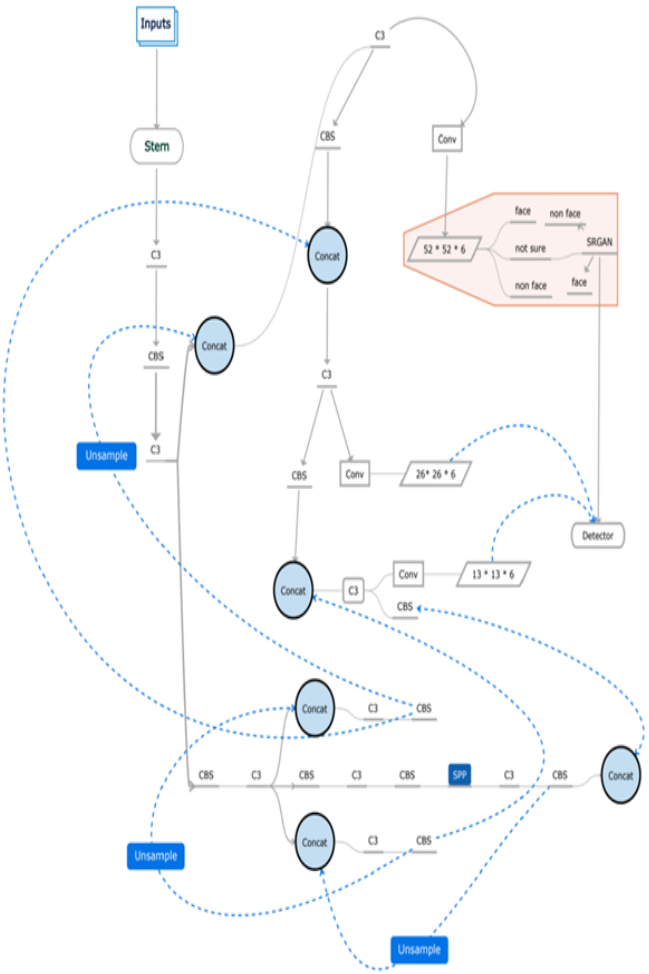

Fig 12 Block System of how the Model Takes Data and Processes in Different Layer of the Model

VIII. DISCUSSION

Given the amount of information we were working with, the findings appeared to be sufficient. The covid detection algorithm's results are evaluated using the Accuracy, Reliability, Precision, and mAP indicators. The graph below depicts the evolution of numerous metrics throughout several training rounds.

ISSN No:-2456-2165

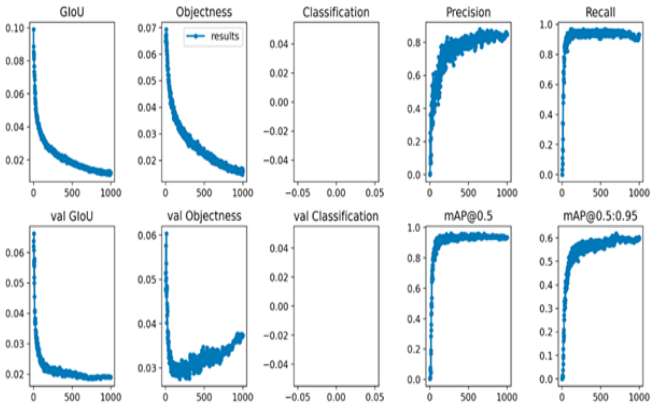

Fig 13 Graph of Accuracy , Precision & Recall for Yolo V5 Face Mask Detection

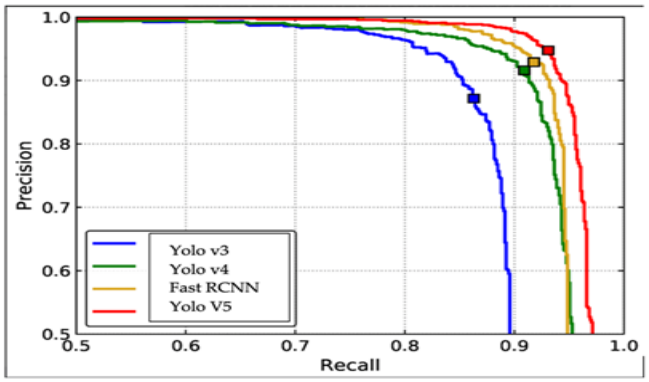

Fig 14 Comparision in Different Yolo Version

Now with the public dataset, we acquired a Recall of 98 per cent and a Precision of 92 per cent after thousand epochs. The mAP@0.6 and mAP@0.6-0.95 are respectively 0.96 and 0.59 from graph in fig 13 and fig 14.

Our algorithm recognises things based on their categories and assigns each entity a tag as well as percentages on the detected image, as expected. With the particular location of a pixel item in the image in the x,y-axis, we may discover things more correctly and identify them independently, based on the study findings of the trials. This investigation also numerous utilizations on different ways for item detection and characterisation, as well as an assessment of each approach's efficiency.

IX. CONCLUSION

Yolov5, as a unified framework that is faster than earlier two-stage detectors, was able to accurately recognise humans. We were also able to use image manipulation to turn the image into a bird's-eye viewpoint and calculate the distance between two spots between people.

Visual recognition systems may help with security cameras, face recognition, defect condition monitoring, text classification, and other applications. The purpose of this thesis is to develop object recognition software that can differentiate between people and determine their relative distances. The performance of an object recognition system is determined by the attributes employed and the recognition algorithm used. The goal of this research is to offer a novel feature extraction method for obtaining global features. features and getting local features from the study area. In addition, the study endeavour aims to combine classical classifiers to recognise the item.

The algorithm was able to recognise persons in the video stream and estimate proximity between them as a whole. We were also capable to use red boundary lines to alert persons who did not acknowledge social separation.

FUTURE SCOPE

A conventional webcam was used to do item recognition and target tracking. The principle may be used in a multitude of scenarios, including Artificial Robots, Automatic Assisted Automobiles, Network Security Improvement to recognize suspect behaviour as well as weaponry, detect anomalous enemy movements on the border with the aid of target acquisition cams, and many more.

REFERENCES

[1]. The Guardian. Big brother is watching. https://www.theguardian.com/cities/ 2019/dec/02/big-brother-is-watching-chinese-city-with-26m-cameras-is-worlds-most-hea Online; accessed 11 Nov 2020.
[2]. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. Nature, 521(7553):436–444, May 2015.
[3]. Worldometer. Coronavirus update. https://www.worldometers.info/coronavirus/.Online; accessed 29 Nov 2021.
[4]. https://covid19.who.int/
[5]. Ultralytics. Yolov5. https://https://github.com/ ultralytics. Online; accessed 11 Nov 2020.
[6]. https://cocodataset.org/#home
[7]. https://doi.org/10.1038/nature14539
[8]. G. Bradski. The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.
[9]. R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580-587, doi: 10.1109/CVPR. 2014.81.
[10]. S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," Neural Networks, vol. 107, pp. 3–11, 2018.
[11]. K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, pp. 1904–1916, 2015.
[12]. T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125, Honolulu, HI, United States, 2017.

[13]. H. Bai, J. Cheng, X. Huang, S. Liu, and C. Deng, "HCANet: a hierarchical context aggregation network for semantic segmentation of high-resolution remote sensing images," IEEE Geoscience and Remote Sensing Letters., pp. 1–5, 2021.
[14]. B. Yu and D. Tao, "Anchor cascade for efficient face detection," IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2490–2501, 2019
[15]. C. Ding and D. Tao, "Trunk-branch ensemble convolutional neural networks for video-based face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 1002–1014, 2018.
[16]. Z. Tang, G. Zhao, and T. Ouyang, "Two-phase deep learning model for short-term wind direction forecasting," Renewable Energy, vol. 173, pp. 1005–1016, 2021.
[17]. Qian M, Jiang J. COVID-19 and social distancing. Z Gesundh Wiss. 2022;30(1):259-261. doi: 10.1007/s10389-020-01321-z. Epub 2020 May 25. PMID: 32837835; PMCID: PMC7247774.
[18]. Kumar P, Sah AK, Tripathi G, Kashyap A, Tripathi A, Rao R, Mishra PC, Mallick K, Husain A, Kashyap MK. Role of ACE2 receptor and the landscape of treatment options from convalescent plasma therapy to the drug repurposing in COVID-19. Mol Cell Biochem. 2021 Feb;476(2):553-574. doi: 10.1007/s11010-020-03924-2. Epub 2020 Oct 7. PMID: 33029696; PMCID: PMC7539757.
[19]. arXiv:1405.0312