FACE MASK DETECTION

Praveen Pandi[1], Sai Sree Depa[1], Vaishnavi Mamilla[1], and Jnaneswar Reddy Sabbella[1]

UBIT: ppandi,saisreed,vmamilla,jnaneswa
[1]Department of Computer Science, University at Buffalo
ppandi@buffalo.edu, saisreed@buffalo.edu, vmamilla@buffalo.edu, jnaneswa@buffalo.edu

Abstract.

The COVID-19 outbreak has changed society's rules, where keeping distance and wearing masks are necessary to stop the virus. However, a significant proportion of people are not adhering to these protocols when in public spaces, resulting in a significant increase in the probability of the virus spreading. If people continue keep ignoring the rules, the virus could become a bigger issue. Therefore, to prevent such circumstances, we need to closely examine people wearing face masks. To address the challenge of manually monitoring large groups of people for mask-wearing compliance in real-time, face mask detection has emerged as an automated solution that uses deep learning algorithms. Deep learning-based object detection systems have achieved significant success in detecting complex objects in images and have demonstrated promise in various real-world applications. In our project, we create a personalized YOLOV5 model that can differentiate between individuals who are wearing face masks, those who are not and improper masks.YOLO is fast, simple, accurate, and precise algorithm, to identify and locate individuals wearing face masks. This technology has become increasingly important in promoting public health and safety by ensuring compliance with mask-wearing policies in various settings, including hospitals, airports, public transportation, and workplaces.

Keywords: YOLOv3, YOLOv5, Face mask, Object Detection, COVID-19, Faster R-CNN, SSD, CNN, Deep Learning

1 Introduction

The COVID-19 pandemic has had a profound impact on various global industries, such as transportation, agriculture, and manufacturing. These industries have been compelled to halt their operations, and strict regulations were implemented to ensure social distancing and the wearing of face masks. As a consequence, businesses worldwide have suffered significant financial losses, with many companies struggling to survive. Moreover, this pandemic has also exposed the vulnerabilities of our global supply chains and emphasized the need for more resilient systems to mitigate the impact of such crises. It is challenging to overstate the significance of face mask detection, particularly given the present global COVID-19 outbreak. Face mask detection technology can also safeguard front line personnel, keep track of adherence to face mask regulations, and encourage people to wear masks by giving them immediate feedback on their actions. We can encourage the use of masks by using face mask detection technology, which will help stop the spread of COVID-19 and other infectious diseases and foster a safer and better society for all. One of the simplest and most reasonable solutions to this problem is to employ an object detection model, like YOLOv5, which can recognize objects by training on bounding boxes with corresponding labels.In the past, detection frameworks relied on image classification methods to locate objects in images by repeatedly scanning various areas of the image at different scales. However, this approach was inefficient and time-consuming. YOLOv5, on the other hand, employs a unique approach that involves examining the entire image at once and scanning the network only once in the computer vision pipeline to identify objects. Other popular object detection frameworks, such as Faster R-CNN and SSD, are also commonly used and employ their own unique techniques for object detection. Nonetheless, the YOLOv5 algorithm has gained particular attention for its speed and accuracy, making it a valuable tool for various computer vision applications. The faces will be the bounding boxes, and the labels will indicate whether the person is wearing a mask, an improper mask, or no mask. This approach would not only aid in the identification of individuals who are not following mask-wearing protocols, but it would also provide a means to monitor and enforce mask-wearing compliance in real-time, reducing the spread of COVID-19. Additionally, the use of object detection models could have broader implications for public safety and security, enabling organizations to identify and address potential threats and hazards with greater accuracy and efficiency.

2 Motivation

The COVID-19 pandemic continues to impact the world, and a reliable vaccine is yet to be developed. The motivation for this is safety always comes first in circumstances like Covid, where a single sneeze might

injure a large number of people. A technology that could independently check whether or not a face mask is used is necessary to assure everyone's safety. However, while there has been extensive research on face detection and recognition algorithms, there is still a significant difference between detecting a face with a mask and detecting a mask on a face. Most previous research has focused on detecting uncovered faces, and there is limited literature available on detecting masks over faces. Therefore, our aim is to develop a technique that can accurately and effectively identify face masks in public areas. By doing so, we hope to contribute to public health by preventing the spread of COVID-19.

3 Related Work

Various studies have been conducted on face mask detection and classification using different deep learning techniques.

3.1 OpenCV Haar Cascades:

Haar Cascades, based on machine learning, is an approach used for detecting objects. It is commonly used for detecting faces and can also be utilized for identifying face masks. This model uses a set of features to determine whether a face or mask is present in the image.

3.2 Convolutional Neural Networks:

Using convolutional neural networks (CNNs), it is possible to identify objects and images.CNNs are capable of finding objects using the features and patterns they have learned from photos. A model for the detection of face masks can be developed by training CNNs on a collection of images of persons wearing and not wearing masks.

3.3 Faster R-CNN:

It is another object detection technique that uses a region proposal network (RPN) to propose object regions and a fast R-CNN network to classify and refine the proposals[2]. Faster R-CNN has several advantages over other object detection techniques. It is excellent for object detection in a variety of applications because it can detect things with diverse scales and aspect ratios. Finally, even in busy and obstructed surroundings, it is able to recognize things with great precision[3]. Due to its two-step architecture, even though it is efficient, the training time for this system is prolonged and it is unable to perform real-time detection effectively[6].

3.4 RetinaNet:

RetinaNet is a one-stage object detection method that addresses the class imbalance issue in object detection by using a feature pyramid network. By training RetinaNet on a collection of pictures of people wearing and not wearing masks, it is possible to use it for face mask detection. A top-down pathway and a bottom-up pathway make up the FPN. The input image is processed by a number

of convolutional layers in the bottom-up pathway to create feature maps at various scales[4]. These maps are combined and created into a pyramid of feature maps using the top-down pathway and a series of lateral connections. After that, the bounding boxes and class probabilities for each object are predicted using the feature maps. When input images contain a larger face, there is a tendency for the system to fail[6].

3.5 SSD (Single Shot Multi-Box Detector):

It is an object detection technique that uses a single neural network for object detection. In order for SSD to function, the input image is divided into a number of fixed-size grids, with numerous bounding boxes and class probabilities predicted for each grid cell. However. to achieve better accuracy with SSD, a larger amount of training data is needed[6].

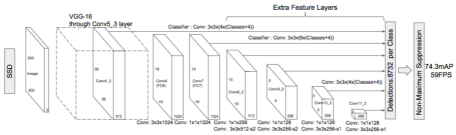

Figure 2: SSD architecture [6]

3.6 YOLO:

YOLO, which stands for You Only Look Once, is a real-time object detection model that has gained widespread popularity due to its ability to detect multiple objects in an image or video quickly. One of the advantages of YOLO is that it can be easily customized for specific applications by fine-tuning the pre-trained model on a new dataset.Furthermore, it surpasses other models due to its ability to learn from generalized object images and make accurate predictions [1]. Bounding boxes are identified by YOLO as a regression problem while simultaneously classifying. YOLO performs object detection in a single pass. The input image goes through several layers of the network, which produces a prediction as the output. The object identification algorithm YOLOv3 predicts object boundaries while accurately detecting small objects. Bounding boxes are predicted by YOLOv3 using three distinct sizes and a feature extraction network built on the DarkNet-53 architecture. It also introduces the idea of anchor boxes, which facilitates improved object location.YOLOv3 utilizes DarkNet-53 to detect features, which is a 53-layer convolutional neural network (CNN) trained on ImageNet[6]. This allows for the development of highly accurate and specific models for face mask detection in various environments, such as schools, hospitals, and public spaces.

4 YOLOv5

A convolutional neural network (CNN) architecture is used by YOLOv5 to carry out simultaneous object detection and classification tasks. The model does a single pass processing of the input image, predicting the bounding boxes and class probabilities of each object in the picture. The weights assigned to the predicted bounding boxes are then determined by the

corresponding probabilities, resulting in the identification of the image's most likely items. As a result, the YOLOv5 model is a reliable and effective choice for face mask identification. After making the predictions, non-max suppression is used to get rid of duplicate detections of the same object. This makes it possible for the algorithm to recognize each object just once, even if it appears more than once in the image. Our proposed facial mask detection system employed YOLOv5, which consists of three main architecture blocks: **backbone, neck, and head**. The backbone of YOLOv5 uses CSPDarknet to extract features from the image dataset via a partial network. This architecture allows for efficient and accurate feature extraction.In addition to the backbone, the neck of YOLOv5 utilizes PANet to generate a pyramids feature network.

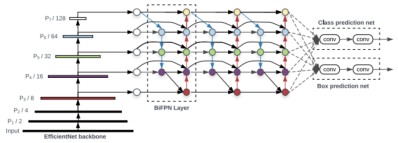

Figure 3: YOLOv5 Architecture

The feature aggregation is then passed on to the head for predictions. The head of YOLOv5 consists of layers that generate predictions from anchor boxes for object detection. Leaky ReLU and Sigmoid activation functions, which are effective and frequently utilized in object identification models, are used in YOLOv5. To further increase the model's accuracy, YOLOv5 makes use of two optimization techniques: Stochastic Gradient Descent (SGD) and ADAM optimization. These methods assist in optimizing the model's parameters and enhancing its precision in identifying face masks.

5 Methodology

This section of the paper gives a brief overview of the methodology used in this paper. Here the developed methodology includes dataset, data pre-processing, visualization, model development, model training, and the last stage is testing the developed model. We have used several tools and technologies such as NumPy, PyTorch, and Open CV.

5.1 Dataset:

The Face Mask Detection dataset available on Kaggle is an open-source dataset that was used in this project. The dataset includes 853 images, which are in PNG format, and annotations for each image, it also contains bounding boxes that follow the PASCAL VOC format. This dataset is augmented for training by various factors to produce 6k images. The dataset is labeled into three different classes: with mask, without mask, and mask worn incorrectly. This means that each image is labeled according to whether the person in the image is wearing a mask correctly, wearing a mask incorrectly, or not wearing a mask at all.

5.2 Data preprocessing:

Data preprocessing is done on the dataset. In the dataset, the class mask incorrectly worn has significantly less amount of the data.

These images have been enhanced by factors such as Image resizing(Normalization), Flipping(Left to Right) and reshaping, transpose, rotation at angles of 90, 180, and 270. The purpose of normalization is to ensure that the data have a uniform distribution and to make computations more efficient. Additionally, normalization promotes quicker convergence. As each image must have the same dimension when entering it into the CNN model, the photos are modified to ensure this. The ratio of the data's train and test sets is 80:20. To obtain a dataset consisting of individual faces with their respective labels, we need to crop the faces from each original image using the given coordinates and place them in the corresponding folder according to the person's label. This process is necessary since each original image may contain multiple faces. Once this step is completed, we will have a dataset with images consisting of only one face along with its corresponding label. To enhance the model performance, we use data augmentation to create more images artifically. The dataset has annotations for each image and each annotation depicts the boundaries and the class of the object. This XML annotations file is converted into Txt file.

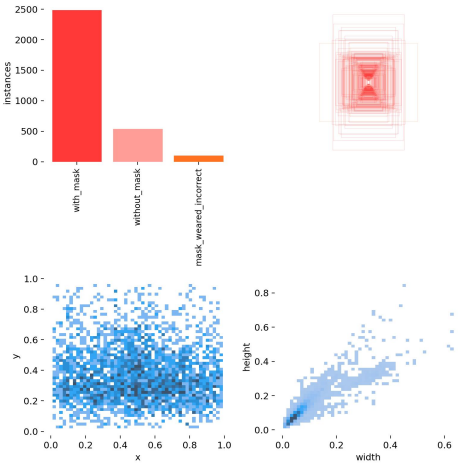

Figure 4: Data Visualization Graphs

5.3 Proposed Method:

To begin with, we utilized the YOLOv3 model to identify face masks. However, the model did not perform well. Therefore, we incorporated the YOLOv5 model to our dataset. PyTorch framework is used for data pre-processing, data loading, creating YOLO model, training and evaluation. For visualizing the results, Seaborn and Matplotlib libraries are used. For logging the experiments, Tensorboard is used. The YOLOv5 model is employed for detecting face masks in the proposed dataset. Ultralytics released the first official version of YOLOv5, a computer vision model for object detection, on June 25th, 2020. YOLOv5 is a member of the You Only Look Once (YOLO) family of computer vision models and has a similar architecture to YOLOv4 but with improved performance due to PyTorch training procedures.

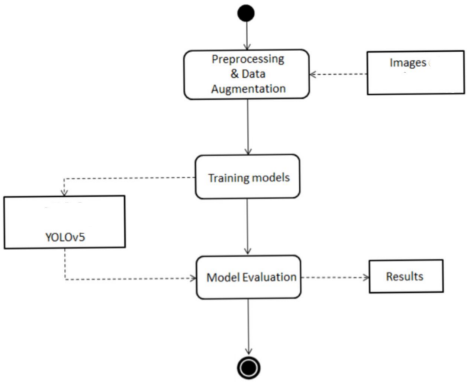

Figure 8: System Flow

The YOLOv5 model is fitted with the preprocessed dataset for training. The model is modified, and it is adjusted by tuning its hyperparameters. To obtain accurate findings, the batch size was varied to 4, 8, and 16 values, and the image size to 224, 320, and 640 accordingly with epochs set as 500 each time. YOLOv5 generates multiple predictions for each location in an image at varying scales. This approach helps in accurately detecting small objects. The predictions generated include information such as object details, boundary box regression, and classification scores. But with various changes in the hyperparamaters, the model failed to give desired outputs. We then stick to the default YOLOv5 model to accurately classify the images as face with mask, no mask and the improper mask. The Mean Average Precison, Recall and precision was compared with the various hyperparameters and with modifications and changes, the default model accuracy was obtained as 90%.

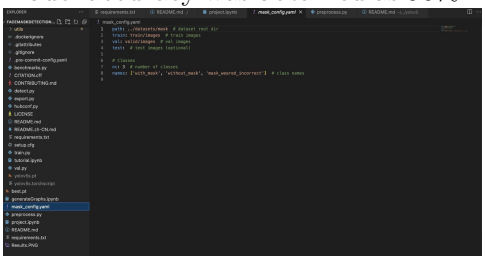

Figure 9: parameters for YOLOv5 model

YOLOv5 model can take upto 80 parameters, and based on our requirements we customised YOLOv5 model to take in three parameters, which are as shown in the figure 9. It can be seen that the three parameters are with-mask, without-mask, and mask-weared-incorrect which are stored in the maskconfig.yaml.

The model is tested using the testing set and it performed well. To understand the training and validation performance, the losses are plotted. Additionally demonstrating that there is no overfitting, the bounding box loss, the abjectness loss, and the classification loss are all decreasing for both the train and test sets. Additionally, the mAP is rising for both the train and validation sets.

5.4 Results:

The graphs below represents the testing and training loss, precision, recall, f1, mAP as shown in the below:

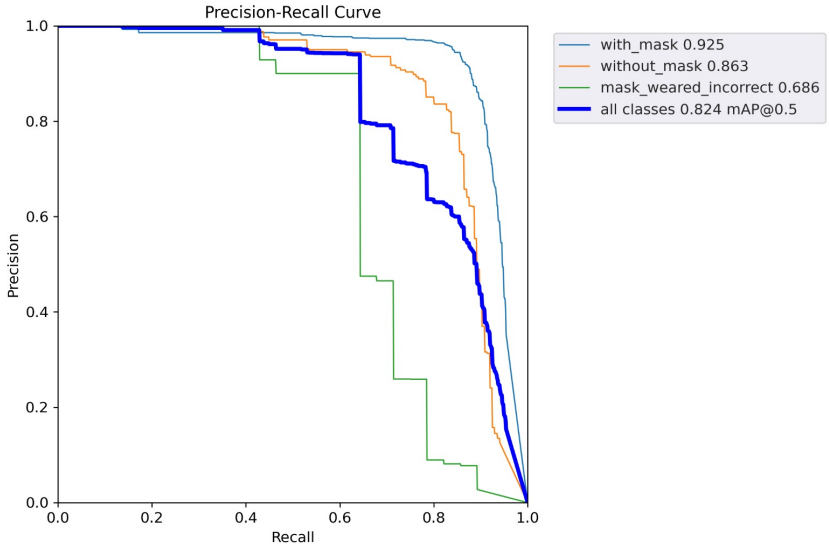

Figure 10: Precision and Recall

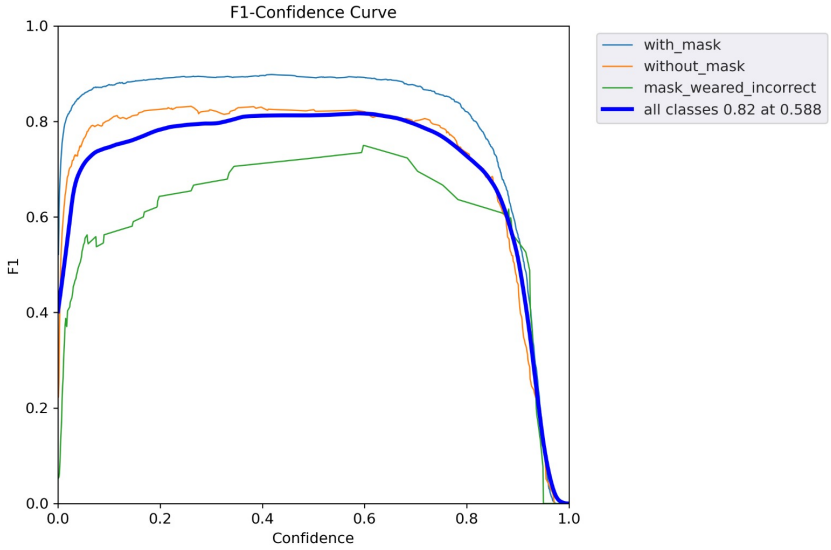

Figure 11: F1 Curve

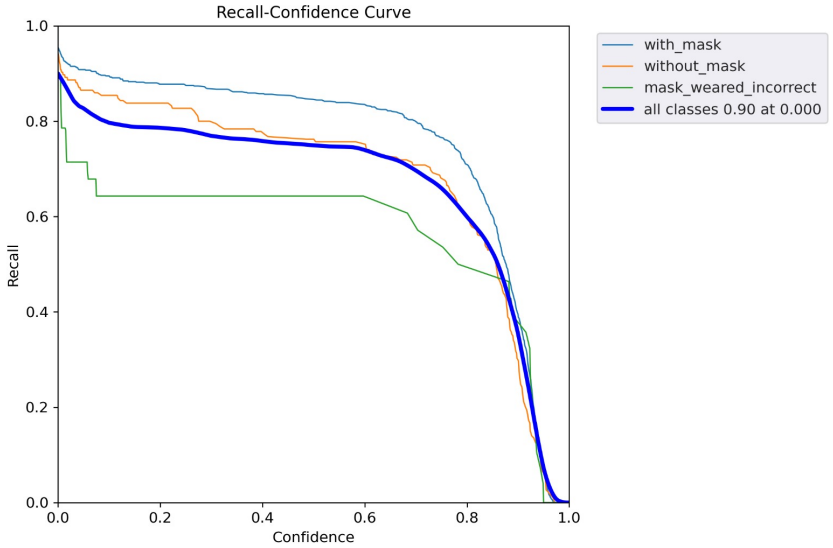

Figure 12: Recall-Confidence Curve

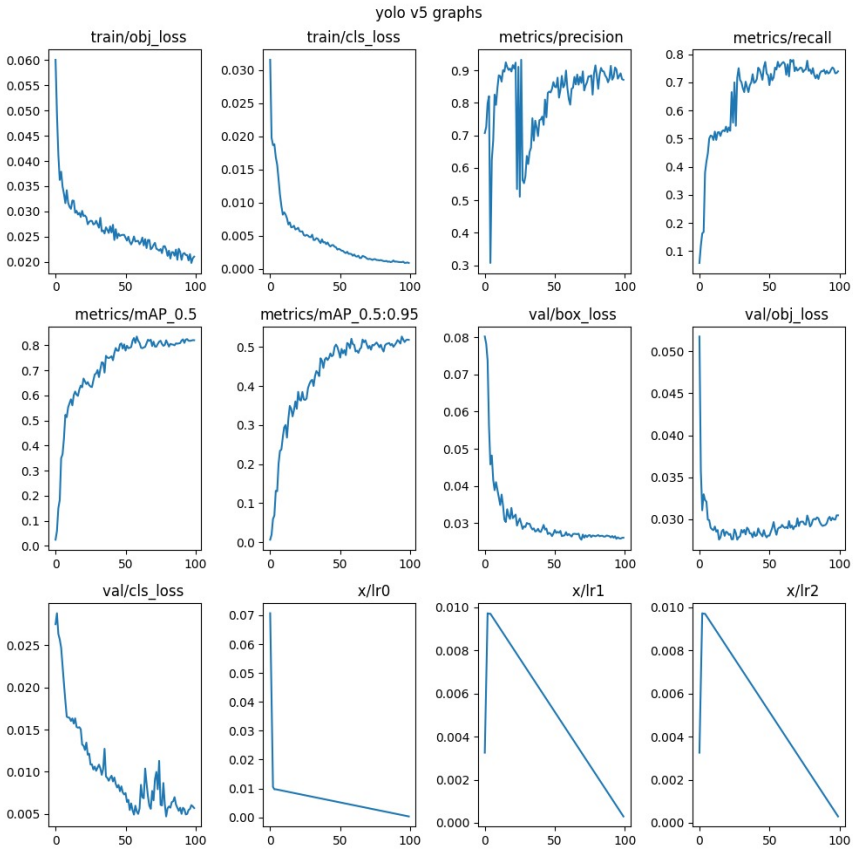

Figure 13: Loss and mAP vs Epoch

The YOLOv5 model provides accurate results as shown in the below fig-

ures.

Figure 14

Figure 15

Figure 16

6 Conclusion:

For the Face Mask Detection project, we utilized the YOLO (You Only Look Once) object detection method to identify whether individuals are wearing masks correctly or not. Our YOLO algorithm was trained on a dataset containing various mask-related images, including those depicting individuals wearing, not wearing, or wearing masks improperly. We used a customised trained YOLOv5 model, and the model ran with a frequency of 90%. We can enhance the project's functionality by introducing additional features, such as determining if the mask is worn upside down, detecting the correct mask position, or recognizing whether an individual wears multiple masks. This can aid in providing more precise information and enhancing public health monitoring during pandemics or in places where wearing masks is mandatory. Furthermore, we can utilize this project's approach in other comparable scenarios, such as detecting helmets worn by bikers, safety glasses worn in factories, and monitoring social distancing in public spaces. The possibilities for this project are vast, and it can be applied to a wide range of real-world situations that require object detection and tracking.

References

[1] Mahurkar RR, Gadge NG (2021) Real-time Covid-19 face mask detection with YOLOv4 .2021 second international conference on electronics and sustainable communication systems (ICESC). 10.1109/ICESC51422.2021.9533008
[2] Girshick R. 2015 IEEE International Conference on Computer Vision (ICCV) 2015. Fast R-CNN.
[3] Ren S, He K, Girshick R, & Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2015;39(6):1137–1149. doi: 10.1109/TPAMI.2016.2577031
[4] "COVID-19 Face Mask Detection using RetinaNet" by H. Khan, H. Aslam, M. I. Khan, and M. A. Mahmood.
[5] Chandan G, Jain A, Jain H, Mohana (2018) Real time object detection and tracking using deep learning and OpenCV. In:2018 international conference on inventive research in computing applications (ICIRCA).
[6] Vibhuti, Jindal N, Singh H, Rana PS. Face mask detection in COVID-19: a strategic review. Multimed Tools Appl. 2022;81(28):40013-40042. doi: 10.1007/s11042-022-12999-6. Epub 2022 May 5. PMID: 35528282; PMCID: PMC9069221.
[7] Hongyu Ding, Muhammad Ahsan Latif, Zain Zia, Muhammad Asif Habib, Muhammad Abdul Qayum, Quancai Jiang, "Facial Mask Detection Using Image Processing with Deep Learning", Mathematical Problems in Engineering, vol. 2022, Article ID 8220677, 10 pages, 2022. https://doi.org/10.1155/2022/8220677
[8]" You Only Look Once: Unified, Real-Time Object Detection" by Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi.
[9]"YOLO9000: Better, Faster, Stronger" by Joseph Redmon, Ali Farhadi