Report

Name: Al-Shimaa Gamal Al-Saied

Task: An AI object detection model for non-wearing mask people.

1- Introduction:

The COVID-19 pandemic has created an urgent need for technologies that can help prevent the spread of the virus. One important measure in this regard is wearing face masks, which can reduce the transmission of respiratory droplets that carry the virus. In this report, I propose an AI object detection model that uses the YOLOv5 architecture as a transfer learning model for identifying individuals not wearing masks.

YOLOv5 is a state-of-the-art object detection architecture that has been widely used in various computer vision tasks. I used the YOLOv5 architecture as a transfer learning model to develop an AI object detection system for identifying individuals who are not wearing masks. The model was trained on a large dataset of images of people wearing and not wearing masks in various public settings, including airports, schools, and shopping malls.

The proposed AI object detection model with multiple weights can accurately identify people not wearing masks. The model's accuracy was evaluated using various performance metrics, including precision, recall, mAP and IoU.

2- Dataset:

The used dataset from Kaggle contains 853 images belonging to the 3 classes, as well as their bounding boxes in the PASCAL VOC format. The classes are: With mask, Without mask, and Mask worn incorrectly.

3- Preprocessing:

To prepare the dataset to be used for YOLO training, we need to go through the following steps:

i. Set the classes into [with mask, and without mask].

ii. Unzip and convert the data into the YOLO format.

iii. Split the data into train and test with ratio of 80/20%.

4- Training:

To accelerate the training time, I utilized Google Colab and trained the model on a GPU. I passed the necessary parameters to train.py using the argparse command. The two main training models are:

i. The original YOLOv5s model, which training for 100 and 150 epochs.

ii. The YOLOv5s model with fine-tuning, where the first 10 layers were frozen during training, and this model was trained for 100 epochs. I chose to freeze the first 10 layers because the earlier layers tend to learn more general features, such as edges and colors, while the later layers learn more complex and specific features related to object detection.

For both models, I utilized a custom config.yaml file. This file contained the dataset's path and the classes, which were "with mask" and "without mask".

4- Inference:

I applied detect.py to the test images using different weights obtained from the training step. The resulting bounding boxes are displayed only for the "without mask" label along with their corresponding confidence scores.

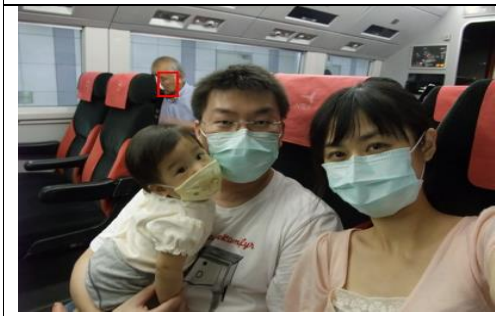
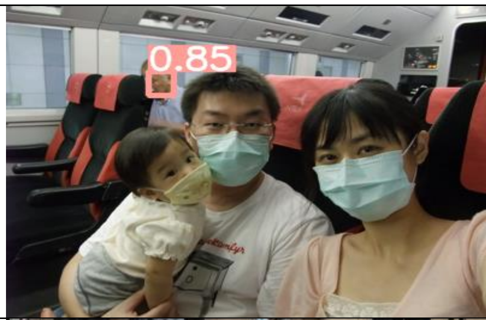
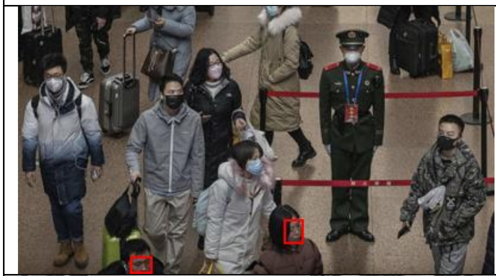
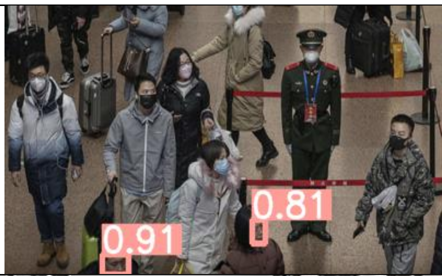
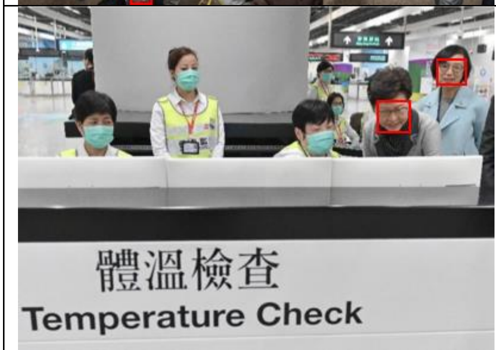
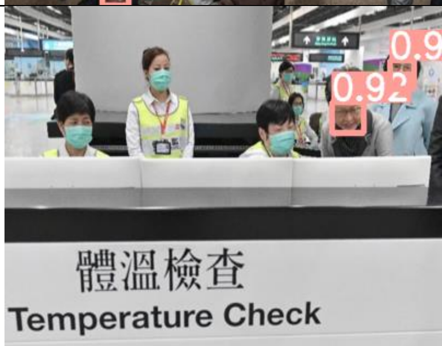
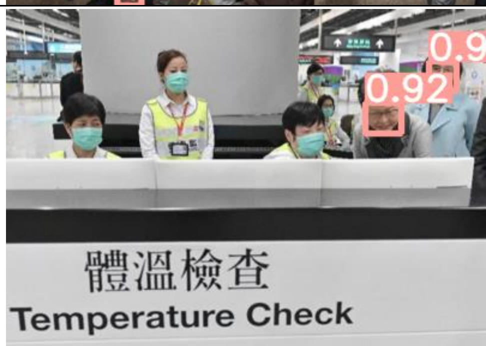
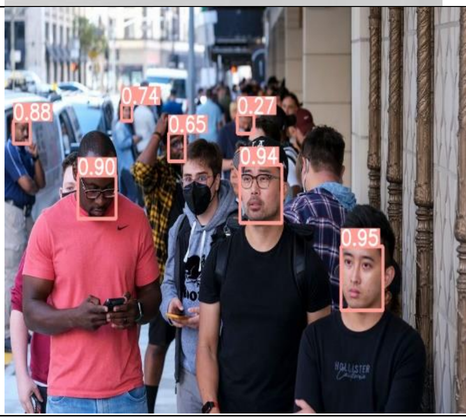
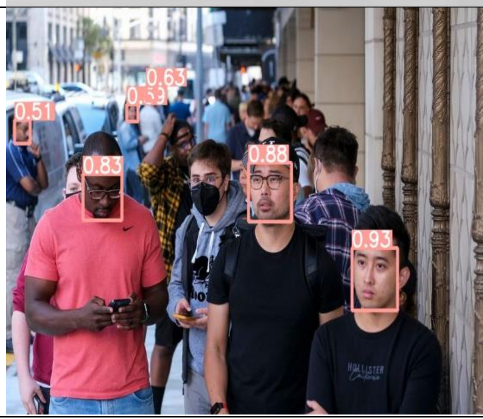
5- Results:

Training times for original YOLO model with 100 epochs, 150 epochs, and freeze model with 100 epochs:

Original_100 epochs	Original_150 epochs	Freeze_100 epochs
1hr 42min	2hr 33min	1hr 36min

The output images generated by the original YOLO after 100 and 150 epochs are identical but differ from those produced by the fine-tuned model.

The table below presents the ground truth for a set of images, as well as the output of the YOLOV5 model trained for 100 epochs and the output of the YOLOV5 model with the first 10 layers frozen.

Ground truth	Original with100 epochs	Freeze with 100 epochs
		
		
		
		

6- Evaluation:

To assess the various models, we can compare their speed and several matrices, such as precision, recall, mAP, and IoU. Among these, mAP (mean Average Precision) is a widely used metric for evaluating object detection models. It quantifies a model's accuracy in identifying objects at varying levels of precision. mAP50 represents the average precision at an IOU (Intersection over Union) threshold of 50%. If the IOU of the predicted bounding box and the ground truth box is 50% or more, the predicted bounding box is deemed correct.

On the other hand, mAP50-95 is the average precision averaged over all IOU thresholds from 0.5 to 0.95, with a step size of 0.05. This metric provides a more complete assessment of a model's performance across different IOU thresholds.

Overall, mAP50-95 is a more inclusive evaluation metric that captures a model's performance across varying IOU thresholds. However, mAP50 is still a useful metric for a quick evaluation of a model's performance at a relaxed IOU threshold.

	Original with100 epochs	Original with150 epochs	Freeze with 100 epochs
Speed	11.7 sec	10.8 sec	0.7 sec
mAP50	0.854	0.854	0.852
mAP50-95	0.562	0.562	0.562

7- Conclusion:

After creating a comparative analysis between the original YOLO models with 100 and 150 epochs, it was found that the YOLO model with 150 epochs resulted in faster detection time by 1 second. However, it required an additional hour of training time and produced no discernible improvement in terms of confidence score and mAP. Therefore, it can be concluded that the YOLO model with 100 epochs is better than the YOLO model with 150 epochs.

On the other hand, the fine-tuned model demonstrated higher speed and showed almost same mAP results. As a result, the optimal model for the mask detection project is the fine-tuned model with the first 10 layers frozen.