International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol.11, No.1, 2024, pp.01-08.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Comparison Between the Yolov4 and Yolov5 Models in Detecting Faces while Wearing a Mask

Aseil Nahum Kadhum [a], Aseel Nahum Kadhum [b]

[a] *Department of Computer Technology Engineering, Imam Al-Kadhim College -Iraq.*
E-mail: easeel18@yahoo.com
[b] *Department of Computer Technology Engineering, Imam Al-Kadhim College -Iraq.*
E-mail:easeel18@gmail.com

Abstract

Object detection based on deep learning has shown good results ever since the Coronavirus or Covid-19 started sweeping the entire world affecting and killing many people. One of the easiest and simplest ways to protect oneself from this virus is to wear a mask. In order to detect whether a person is wearing a mask or not, we propose here two models for detecting face masks. Facial recognition has been difficult, but with the development of deep learning, it has tremendous ability to detect objects, especially in public places. Therefore, it has become necessary for accurate diagnosis to protect people from Covid-19. In order to discover whether a person is wearing a mask or not, we propose a model to detect face masks whether the mask is worn or not, so it was proposed in this research to use two deep learning algorithms, which are YOLOv5 and YOLOv4, which are among the YOLO models that are characterized by accuracy and speed. Compare learning algorithms Deep and finding the difference between them in performance and accuracy, and there is a CNN algorithm that is also important in discovering things and achieved satisfactory results, but we will use the YOLO model. YOLO is an advanced algorithm with fast, real-time detection. As most of the results were reviewed with previous variations of YOLO and CNN, it is worth noting that the YOLO model is the best model in face detection. Face detection is of great importance in various fields, especially in public places, and requires security accuracy in detection. It is known that investigators' statements about images of masks are not very easy. Therefore, training and evaluation on the dataset available on Google Colab for YOLOv5 and YOLOv4 algorithm are conducted in this paper.

Keywords: YOLOv5, YOLOv4, Object Detection, Detection Algorithm, Wear Masks, Deep Learning.

DOI:10.71086/IAJSE/V11I1/IAJSE1101
Received:28 November 2023/Accepted: 23 December 2023/Published: 06 January 2024

Introduction

Object detection is crucial in many applications, such as security surveillance, disease detection, vehicle detection, and others. Much research has been done on object detection in images and videos, the most important of which is face detection and recognition, as there have been different techniques used to detect objects accurately and efficiently for different applications [1]. However, these proposed methods still suffer from problems related to inaccuracy and efficiency. To address object detection problems, deep neural network deep learning methods are more effective in correcting object detection. With the presence of the Covid-19 virus and requiring all people to wear a mask and prevent the spread of the virus, it was necessary to search for a mechanism to distinguish people from wearing a mask or not wearing a mask in public places such as airports, stations, etc. It was found that the algorithm model is better in terms of accuracy and speed than the YOLO algorithm models [2] [3]. The main goal of this research is to use models that are more accurate than previous ones in distinguishing faces using deep learning, namely the YOLOv5 and YOLOv4 models in distinguishing faces while wearing a mask, which is one of the YOLO models. Although the primary objective of the program's design was to gain improved accuracy for numerous item recognition in a short amount of time [4] [5], Joseph Redmon and Ross Girshick [6] developed the YOLO (You Only Look Once) approach, which considers object identification as a regression issue. A party can acquire the object's categorization and position on the network in a single step. This method has one of the fastest detection rates on the market right now, however it is not able to accurately identify pedestrian items in complex environments, and it is quite imprecise at recognizing microscopic targets. As a result, several YOLO algorithm models have been developed and are still being improved upon [7]. Because of this, apps that aim to recognize objects more precisely frequently employ the YOLO paradigm. These models were chosen and the images were trained on these two models from YOLOv5 and YOLOv4. We will notice the difference between them, as we noticed in some research papers that YOLOv4 is effective [8], while others claim that YOLOv5 is effective [9] [10]. To fill this gap, we compared these algorithms using data collected from the Internet, without changing any hyperparameters to compare the speed of YOLO modeling algorithms Comparison is done based on mean average precision (mAP) and frames per second (FPS).

Problem Background

There is a great need for automatic understanding and examination of data by intelligent systems, especially in detecting faces in photos and videos. Face detection has become a major challenge with its increasing use in a number of applications. It is the first step for face recognition, face analysis and detection of other facial features. Especially in light of the circumstances that the world has gone through, the Covid-19 virus disease, and with people being obligated to wear a mask, it has become necessary to find new algorithmic models to recognize faces by wearing a mask in all applications in public places, including stations and airports, from a security and health standpoint. Two models of the YOLO algorithm will be used, which are YOLOv4 and YOLOv5. Masks are still worn in all public places, which requires accurate and rapid detection technology. In this research, the YOLOv5 and YOLOv4 algorithms will be proposed as deep learning algorithms. Note that there is a difference in the performance of each , as some researchers describe YOLOv4 as the fastest, and some describe YOLOv5 as the fastest, so we will compare the performance with them in this research paper, as these are very easy algorithms to train. Inference on individual images, portability and use of weights is faster and lighter than the rest of the previous deep learning algorithm models.

Problem Statement

The Covid-19 pandemic came suddenly and affected all aspects of life and caused the death of millions of people in all countries of the world. Therefore, it was necessary to wear a mask to limit the spread of the epidemic by adhering to it and reducing infection with it. The focus on distinguishing faces from those wearing a mask or without a mask has become very important in recent years, and one of the important problems that we must find ways to detect with high accuracy using convolutional neural networks. The goal of this research is to find the difference between two algorithms for distinguishing whether a face is wearing a mask or not, in terms of accuracy and speed. There are problems in proving which is better. There are some research papers proving that the YOLOv4 model is the best [8], and some research papers proving that the YOLOv5 model is the best [9], and they are models of YOLO target detection algorithms. This research proposes to use these algorithms to train face recognition while wearing a mask to detect faces in photos and videos. Through deep learning, Python programming, and using both models, we can find the best way to train images.

Research Goal

The main goal of the research is to discover which type of YOLO model is the best, and it will be to explore current techniques for distinguishing faces whether wearing a mask or not, using deep learning and finding the difference in terms of accuracy and speed in face detection, which is the YOLOv5 and YOLOv4 algorithm.

Scope the Study

The importance of this research study lies in distinguishing faces in public places, especially when wearing a mask. There have been many previous studies on detecting faces, not by wearing a mask. But we need an improved algorithm for detection while wearing a mask. This can be attributed to its ability to find the best in performance, accuracy, and speed in detecting faces based on the use and testing of the YOLOv5 and YOLOv4 algorithm, which is the best.

Literature Review

Due to the Covid-19 epidemic, automatic identification of people has become more important than wearing a mask. In order to prevent the spread of this disease in public gatherings, it has become necessary to use a face mask, especially in public places, as wearing a mask has become compulsory in some countries, so it was necessary to find techniques to detect people who wear a mask or not.

We will explain in a brief literary review the most important basics of detecting and identifying objects.

Object Detection

1. Classification, which is to structure the image into a certain type of information, with a predefined category) String) or instance ID to describe the image.
2. Detection, the classification task is concerned with the whole. Compared with classification, detection gives the understanding of picture foreground and background. The output of the detection model is a list, and each item of the list used a data group to give the category and position of the detected target (commonly used coordinate representation of rectangular detection box.
3. Segmentation, which is a pixel-level description of an image, which gives meaning to each pixel category instance and is suitable for scenes requiring high understanding.

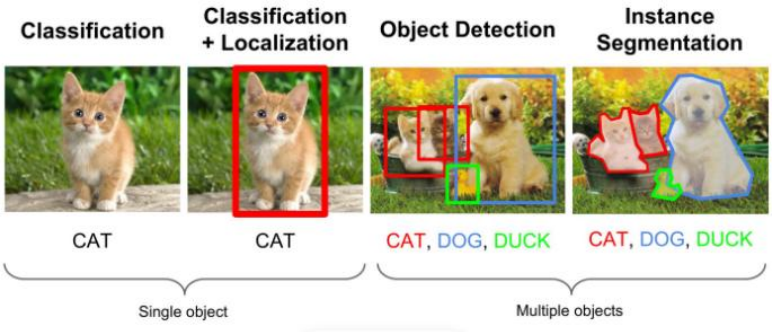


Fig. 1: Comparison between Image Classification, Object Detection and Instance Segmentation

One-Stage and Two-Stage

The CNN convolutional neural network is one of the first networks used to discover and identify objects based on the region and it is Two-Stage, where the CNN algorithm has been improved many times, including R-CNN, Fast-RCNN, and Faster-R. CNN where good performance was achieved, and with this it had many disadvantages, including (1) the training takes place in several stages, (2) it consumes a large volume and takes a long time, the discovery of objects is very slow compared to other algorithms , so you need To improve detection speed. After that, an algorithm was found. The one-stage network is represented by the YOLO chain network, where the one-stage object detection algorithm is fast and can accomplish the task in real time, which has achieved high efficiency for object detection.

YOLO

The acronym for "You Only Look Once" is "YOLO." The YOLO paradigm was initially explained by Joseph Redmon and colleagues. Using either a picture or a video input, this technology operates in real time to detect the type of item and its position. The idea of a neural network is used by the YOLO algorithm to identify objects. Bounding boxes are applied, and a CNN (Convolutional Neural Network) is utilized to decide how to arrange things. The phrase "You Only Look Once" refers to the capacity to identify whole visual objects in a single run. A unique neural network approach called YOLO (You Only Look Once) aims to recognize objects in real-time environments YOLO algorithms are quite popular because of their better effectiveness and quicker detection times. Among YOLO's many uses are the detection of objects, people, animals, parking spaces, traffic signs, and more. Traffic light, parking meter, animal, and human detection are only a few of its extensively used applications.

In recent years, the YOLO algorithm series has shown remarkable results in multiple subfields of object detection . Many researchers have used whole and small variants of the YOLO algorithm to solve detection problems such as the general object. Where the YOLO algorithm and many versions of models were used, including the updated YOLO versions such as YOLO v1, YOLO v2, YOLO v3, YOLO v4, YOLO v5, where the mask was detected whether it was correct or not. Many scientific researchers have been accessed and viewed, as the majority of research studies only detect faces, and few of them distinguish from faces with or without wearing a mask. YOLO has become the fastest algorithm but still lags behind in accuracy compared to other similar technologies. A deep learning algorithm has been used to detect objects in everything in robots, in drones, and in traffic videos even in different driving conditions by training the network on different object classes.

Advantages of YOLO Algorithm

There are following advantages of YOLO algorithm as below:

a) The Aptitude to Understand Object Representation

The YOLO model is better at detecting objects in real-time scenarios due to its enhanced capacity to learn object representation. .Both object location and object class identification can be done on the same network

b) Increased Precision

Better real-time item detection accuracy is provided by the Yolo algorithm. YOLO models provide more accurate microscopic item identification.

c) Faster Object Recognition in Real-Time Situations

Quick object recognition is the primary benefit of the YOLO method. Smaller items may be detected more rapidly, .like traffic lights, etc

d) A Single Network

Object classes are recognized and their distinct locations may be found using a single network. Thus, it is feasible to improve forecast accuracy by modifying just one network.

e) The Larger Framework

YOLO is the most widely used model because it offers the best accuracy for item detection in both creative and original photos.

f) Change Rate

Frame rate is highly suitable for real-time applications, as each frame processes at a pace of 150 frames per second for smaller networks and 45 frames per second for larger networks.

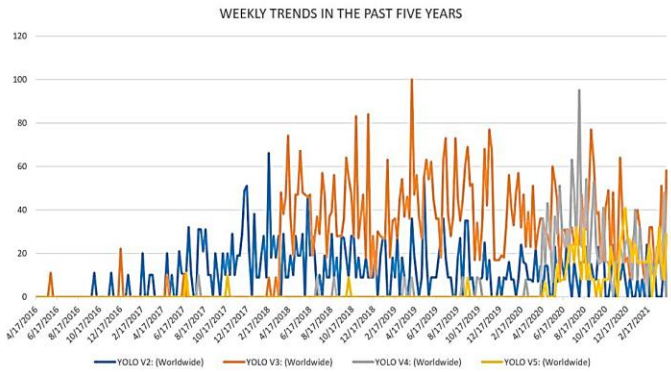


Fig. 2: Trend of YOLO models in last few years[2]

The author compared the design and implementation aspects of many YOLO implementations. Prior to employing publicly available data to assess the data outputs in a tabular and graphical manner, the authors expounded upon the primary concept based on the version. The compact form factor and rapid object detection of the YOLO model are well known. YOLO uses regression analysis to determine an object's identity. Based on size, things may be divided into three main classes: massive, medium, and microscopic. Thus, the creation of a generalized model that can categorize all three kinds of objects is necessary. The multilayer YOLO model with different sample values can be used for this type of detection. There are 26 layers in all that make up the YOLO primary structure. Two levels remain as FC (Fully Connected) layers, and the other twenty-four levels are convolution layers. The Yolo model's incorrect placement and worse recall rate are its two main shortcomings.

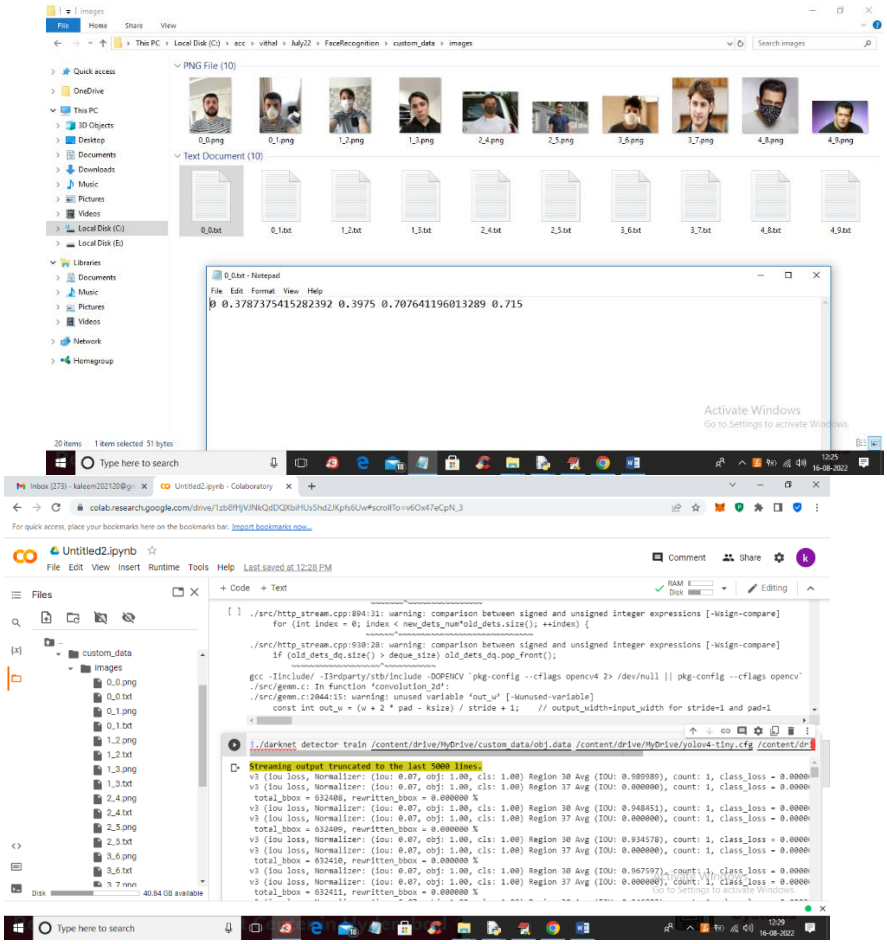
Data Acquisition

Data acquisition is the process of data collection that includes a set of public data that can be downloaded from the Internet with hard-to-find photos with a mask on. Our database images were collected through extensive research. But to train this optimized algorithm we need 1000 images with bounding boxes for each person, so we trained the algorithm using images by downloading from the internet.

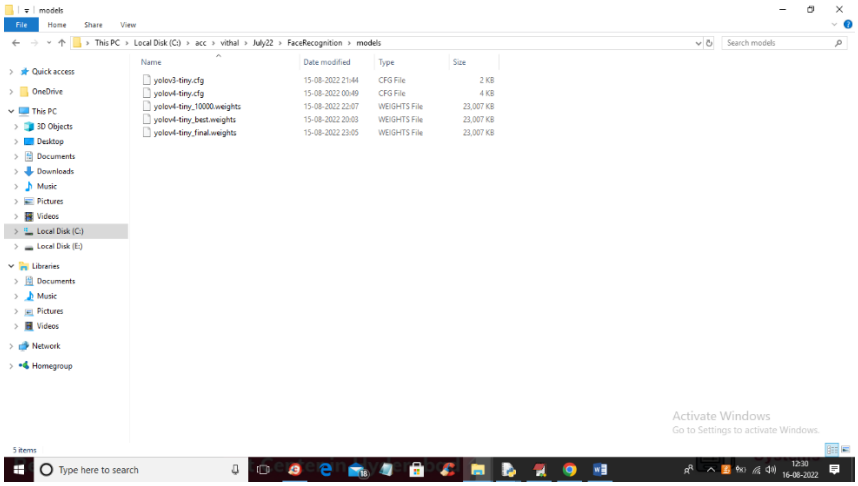
Methodology

Experiments were performed by training the model on custom datasets with both YOLOv5 and YOLOv4 independently. In order to consider which, one performs better in terms of accuracy, Google COLAB is a platform that provides a free coding notebook and a cloud virtual machine with storage and GPU. A Tensor Processing Unit (TPU) was used to run the long and complex computation of the experiment on the detection models. All experiments were conducted on Google COLAB using Google Chrome with Roboflow. The same samples will be for data obtained from Google COLAB, and the same data was trained on both the YOLOv5 algorithm and the YOLOv4 algorithm. This method can achieve a comparison in quickly and accurately discovering with high efficiency which is better. Models are used effectively and straightforwardly to achieve accurate and fast real-time object detection. The main contributions of this research are as follows:

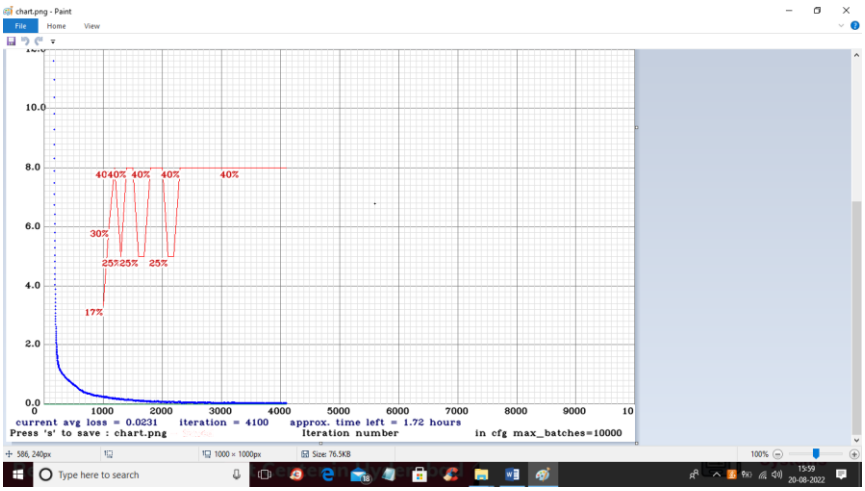
In this project, face recognition was implemented using YOLOv4 and YOLOv5 with face and face mask, but to train these algorithms, we need 1000 images with bounding boxes for each person but we don't have any such datasets, so trained YOLOv4 using some images by Download from the Internet Below are those images In above screen you can see for each person took only two images with mask and without mask and for each face I gave bounding box file and then we trained this custom_data images using below Google COLAB images



In above screen you can see by using those images we are training with Google COLAB and then get below YOLOv4 weights

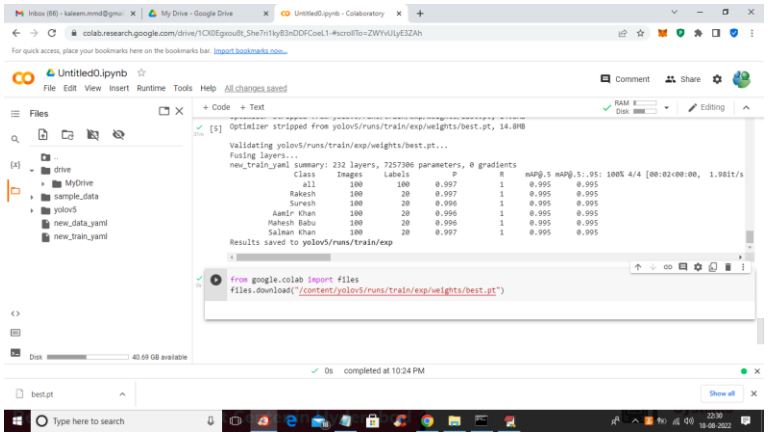


In above screen we got weights files from YOLOv4 and below is the YOLOv4 loss graph.

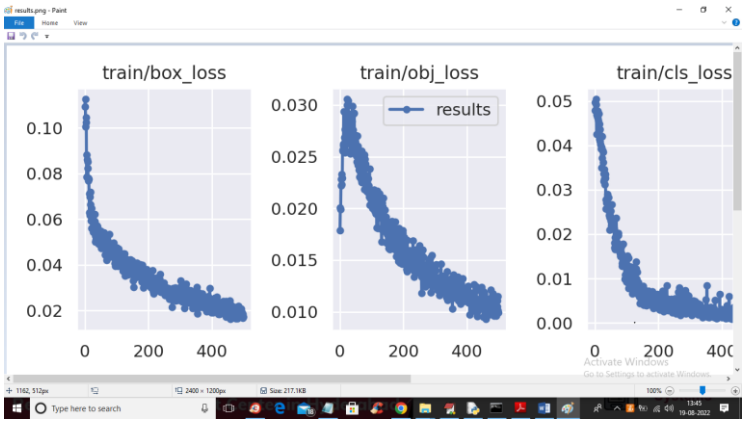


In above YOLOv4 graph x-axis represents Epochs & Y-axis represents LOSS/Precision values and blue line represents LOSS values and with each increasing epoch loss values get down and reached closer to 0 and red line represents Precision values which starts from 17 and reached to 40 so with YOLOv4 we got precision as 40%.

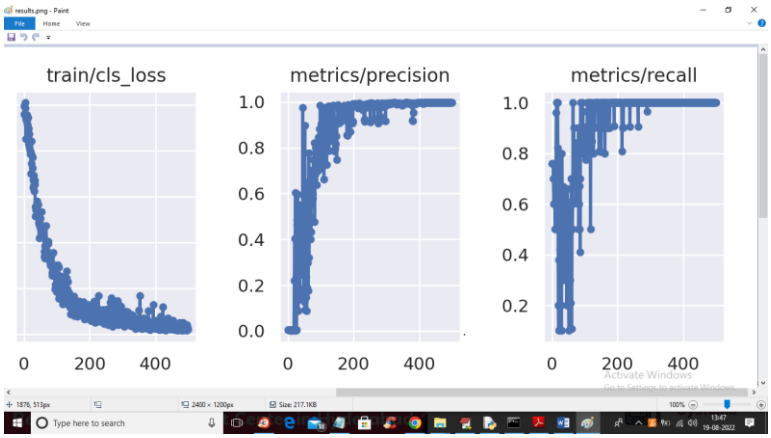
In below screen we are showing YOLOv5 precision, recall and MAP (mean absolute precision) values



In above screen we can see precision (P), recall (R) for each person in dataset and this values we will get after training model and in below screen I am showing YOLOv5 training loss graph



In above graph we can see training bounding box loss and object detection loss where x-axis represents Training EPOCHS and y-axis represents LOSS values and we can see with each increasing epoch loss values get decrease and reach closer to 0 and in below graph we can see precision and recall for YOLOv5.



In above graph we can see precision and recall starts from 0 and reach closer to 1 as we are getting precision values as 0.98.

Conclusion

Research Paper Results, for YOLO Editions. Here the following observations were recorded. First, the versions have a lot of differences. However, they still have some common features. Therefore, they are still similar. Again, by training a set of images on both algorithms, I found that YOLOv5 is the best, fastest, and most accurate compared to the YOLOv4 algorithm. The accuracy value of YOLOv5 is 0.98, while the accuracy of the YOLOv4 algorithm is 0.40.

Reference

Ahmad, T., Ma, Y., Yahya, M., Ahmad, B., Nazir, S., & Haq, A. U. (2020). Object detection through modified YOLO neural network. *Scientific Programming, 2020(1)*, 8403262.

Du, J. (2018). Understanding of object detection based on CNN family and YOLO. *In Journal of Physics: Conference Series*, *1004,* 1-8. IOP Publishing.

Fan, X., Jiang, M., & Yan, H. (2021). A deep learning-based light-weight face mask detector with residual context attention and Gaussian heatmap to fight against COVID-19. *IEEE Access*, *9*, 96964-96974.

Raj, R., Nagaraj, S. S., Ritesh, S., Thushar, T. A., & Aparanji, V. M. (2021, September). Fruit classification comparison based on cnn and yolo. *In IOP Conference Series: Materials Science and Engineering (Vol. 1187, No. 1, p. 012031).* IOP Publishing.

Tan, L., Huangfu, T., Wu, L., & Chen, W. (2021). Comparison of YOLO v3, faster R-CNN, and SSD for real-time pill identification.

Kadhum, A.N., & Kadhum, A.N. (2023). Literature Survey on YOLO Models for Face Recognition in Covid-19 Pandemic. *Journal of Image Processing and Intelligent Remote Sensing (JIPIRS)*, *3*(04), 27-35.

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo algorithm developments. *Procedia Computer Science*, *199*, 1066-1073.

Rahman, E.U., Zhang, Y., Ahmad, S., Ahmad, H.I., & Jobaer, S. (2021). Autonomous vision-based primary distribution systems porcelain insulators inspection using UAVs. *Sensors*, *21*(3), 974.

Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO series in 2021. *arXiv*. https://doi.org/10.48550/arXiv.2107.08430

Murthy, J.S., Siddesh, G.M., Lai, W.C., Parameshachari, B.D., Patil, S.N., & Hemalatha, K.L. (2022). Objectdetect: A real-time object detection framework for advanced driver assistant systems using yolov5. *Wireless Communications and Mobile Computing*, *2022(1),* 9444360.