International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol.11, No.1, 2024, pp.229-236.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Identifying People Wearing Masks in the Wild by Yolov7 Algorithm

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

[a] *Imam Al-Kadhim College, Iraq. E-mail: easeel18@yahoo.com*
[b] *Imam Al-Kadhim College, Iraq. Email: easeel18@gmail.com*

Abstract

Rapid face detection is an important matter at the present time, as face detection while wearing a mask has today become an important security matter in institutions, as well as protecting society from the spread of diseases, especially after the outbreak of infection with the Corona virus. Covid-19 virus. Due to the rapid spread of the Covid-19 pandemic, there is a need for society to adhere to wearing a mask in all public institutions, to prevent the spread of this disease. Researchers worked to find solutions to recognize faces and distinguish a person's identity. There was a problem in detecting faces and recognizing them easily, as researchers found many solutions to detect faces. So far, detecting faces while wearing a mask has problems with accuracy. In this research, a model of the deep learning algorithm, YOLOv7, will be used. It is a YOLO model that is characterized by accuracy and speed compared to previous YOLO models, and compared to a deep learning algorithm that performs two-stage detection, such as the CNN algorithm. Here the YOLOv7 model of the YOLO algorithm is proposed for face detection and recognition with and without mask. It is a model in which the structure of the algorithm has been modified, as it is distinguished by its speed in detecting faces compared to its predecessors. also reviewed most of the experiments with algorithm YOLO (You Only Look Once) detection algorithms and CNN (Convolutional Neural Network), and noticed that YOLOv7 is a better model than previous YOLO models in detecting faces while wearing a mask in terms of speed and accuracy. Face detection and discrimination has become very important at the present time from a security standpoint in all public places and requires accuracy and speed in detection.

Keywords: YOLOv7, Object Detection, Detection Algorithm, Wear Masks, Deep Learning.

DOI:10.71086/IAJSE/V11I1/IAJSE1126
Received: 15 April 2024/Revised: 29 April 2024/Accepted: 08 May 2024/Published: 14 May 2024

Introduction

Nowadays, the pursuit of safety and security in organizations is an important matter. The world is now moving towards rapid development in artificial intelligence and deep learning, and discovering things has become an important research trend in many areas of life. With the importance of keeping people safe in light of the emergence of the Covid-19 virus, it has become necessary to find algorithms in detecting targets, and research in the field of face detection is extremely important [1] [2]. However, there are many deep learning algorithm models for object detection such as the YOLO detector (You Only Look Once), the region-based (SSD) convolutional neural network (CNN), and the fast region-based convolutional neural network (Fast-RCNN) [3] The results of the YOLO algorithm had excellent specifications, as it outperformed the main algorithms in terms of speed and accuracy to discover and determine object coordinates. To address object detection problems, deep neural network deep learning methods are more effective in correcting object detection. With the presence of Covid-19 and everyone urging people to wear a mask and prevent the spread of the virus, there are many algorithms that detect faces, and it was necessary to search for a mechanism to distinguish people who refrain from wearing a mask or do not wear a mask in public places. There are two types of object detection algorithms: the CNN algorithm, which is good, and the other algorithm model that is better in terms of accuracy and speed is the YOLO algorithm. The main goal of this research is to find a technique that is more accurate than its predecessor in face recognition using deep learning, which is the YOLO algorithm. While the main goal of inventing YOLO was to achieve higher accuracy in a short time for detecting multiple objects [4], the YOLO (You Only Look Once) algorithm proposed by Joseph Redmon and Ross Girshick. The detection speed of this method is one of the fastest algorithms at present, but there is a shortcoming in the detection accuracy of small targets, so many models of the YOLO algorithm have emerged and been constantly improved. Therefore, many applications use the YOLO model for higher accuracy object detection. It turns out that the YOLO algorithm has been improved. YOLOv7 is the seventh version of the YOLO model. YOLOv7 will be used to detect faces.

Related Workers

Facial recognition has become extremely important from a security standpoint in all public institutions, especially when wearing a mask. In light of the difficult circumstances that the general public in the world is experiencing due to the Covid-19 virus disease, and with the imposition of wearing masks in all public institutions, all researchers' work has been directed towards finding a solution with new and improved facial recognition techniques through wearing a mask in all public places. Including stations and airports from a security and health perspective. YOLO algorithm models were applied ,YOLOv5 YOLOv4 YOLOv3.Through research, it was noted that there is not enough accuracy in these models [4] [5]. Due to the continued wearing of masks in all public places, this has become a matter that requires accurate detection technology. There are attempts to reveal the identity of the person wearing the mask and who is not wearing the mask using deep learning techniques to recognize faces with different types of masks. There are problems worth investigating [6]. There are many algorithms through which objects are detected and then classification is performed [7]. Some region-based object identification algorithms, such as Fast-RCNN, use a selective search method to extract proposals while still achieving CNN classification. The process of extracting proposals takes some time. Convolutional Neural Network (CNN) can be used by humanoid robots to find and classify objects [8]. The original network is specifically optimized for object detection using a faster region-based CNN (Faster R-CNN). To detect and recognize faces with and without a mask, training and evaluation on the current version of the YOLO network is proposed here. A highly effective real-time face detection system [9]. For each YOLO variant, several tests and evaluations have been performed. YOLO is very fast and accurate . Fast and accurate algorithms allow for better face detection and recognition in public places. This study is based on convolutional neural networks as well as a deep learning model. The comparison that will be presented in this research helps the YOLO algorithm and its importance in terms of speed and accuracy in detecting and distinguishing faces while wearing a mask in public institutions. In this research, the YOLO version 7 algorithm, one of the deep learning algorithms, will be used, as it has been proven to be the most powerful face detection algorithm model at present. Because the YOLOv7 algorithm is easy to train and infer on single images and video feeds, it has the advantage of portability and uses faster and lighter weights than previous YOLO models.

A. Face Detection and Recognition

Deep learning algorithms have become an important role in face detection and recognition. In the field of computer vision, there is often confusion between image concepts in terms of 1) Classification 2) Object Localization 3) Object Detection. Deep learning can recognize a specific pattern through a huge amount of data and applications in social media, security control systems, and in the aspect of health control in terms of the spread of diseases. Image

classification refers to determining the category of the object, while spatial detection includes designing a bounding box around each object that needs to be distinguished. As for the term object detection, the object class is defined and a bounding box is created for each object. Finding facial patterns is one of the most difficult tasks. Deep learning mainly uses CNN (convolutional neural network) for semantic segmentation and pattern recognition. Face detection and recognition may require face grouping, face pattern recognition, face verification, etc. Traditional deep learning mechanism needs better performance for face detection in real-time environment. Therefore, need to identify the characteristics of traditional deep learning in this network. Its main goal is to quickly recognize objects, so the YOLO model was invented, which mainly focuses on detecting and identifying objects only once. The YOLO model uses the entire image during the training and testing process, solving the performance in real-time environment of face detection and recognition as well as solving complex training models by reducing the time and complexity of face detection training.

B. Detection Algorithms

Previous research works focused on applying several deep learning algorithms mainly for object detection and face detection using multiple techniques, the most important of which is the YOLO algorithm. The current study mainly focuses on detecting faces with or without wearing a mask. Face detection is widely applied in many applications. Faces are previously detected by traditional methods after image preprocessing, region selection, and pixel-based image segmentation. Later with the introduction of deep learning based learning, many researchers have invented many models to detect objects. There are two types of deep learning algorithm which are mainly divided into two-stage and single-stage detection: One of the earliest networks to recognize and classify objects based on region was the neural network. Convolutional CNN, which operates in two stages. However, it has a number of disadvantages and one of the most important disadvantages is the slow speed and phased training. The main goal is to invent the YOLO model of a convolutional neural network that simultaneously addresses the object problem and detects objects with a single forward pass. YOLO model able to recognize objects with high performance and a very fast method compared to the deep learning algorithm CNN, two-stage detection, where in the YOLO model the entire image is entered. YOLO detector is fast and effective. The YOLO model has many advantages, so the YOLO model for object detection in applications has higher accuracy. Each model has some advantages and limitations that are discussed. The YOLO model is more accurate and requires less time to discover things. sIn this research, will use the YOLOv7 model, the seventh version of the YOLO algorithm, where the model is designed as a lightweight model. It is found that the training process is very easy in YOLOv7 compared to other YOLO face detection models by considering faces as different objects, as in the above diagram Figure 1.

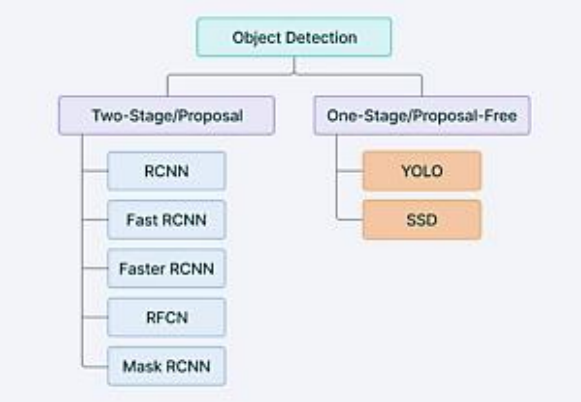


Figure 1: Detection Algorithms

Research Goal

The primary goal of the study is to detecting and distinguishing faces while wearing a mask using the YOLO model. The YOLOv7 deep learning algorithm will be used in this research paper.

Methodology

Study of Yolo algorithm models for face recognition and detection. The goal of this work is to find the best algorithm for detecting the human face, where the images were collected and converted into images of a certain scale, and then the face was extracted from the image. There are two main parts to it: The Yolo technology proposed

in the first section is used to quickly crop people's faces. Binary classification is then used to classify the face, which is similar to the task of detecting a face with or without wearing a mask. Face classification involves three basic stages: (1) preprocessing, (2) feature extraction, and (3) classification. Evaluation, The metrics used for face detection were accuracy rate and absolute error. Then the results are obtained and the model's work in this field is evaluated. In this research, were able to take pictures of university students. These custom data images were trained using Google COLAB below and Robo flow.

A. Data Collection

Data Collection is the process of collecting data from Babylon University students who took original photos. Data collection is the process of collecting live primary data using live photography method cameras on campus for a group of students. Images were taken in bright, partly cloudy natural light in the morning, midday, and afternoon. A total of 219 images were acquired.

- **Single Photography:** A method of collecting images using a personal camera.
- **Image Dataset:** A large-scale image dataset has been prepared for training. Available to the public.

B. Image Preprocessing and Dataset Partitioning

A test set of 219 images was randomly selected to evaluate the generality of the detection model. By defining the category label to which the bounding box belongs and creating a bounding box around the object, facial image data annotation software was used. For each image in the training batch to finish the image classification guide. To ensure that the background area of the rectangle is as minimal as possible, images are classified using the rectangle surrounding the face. Figure 2 provides examples of images of people's faces.

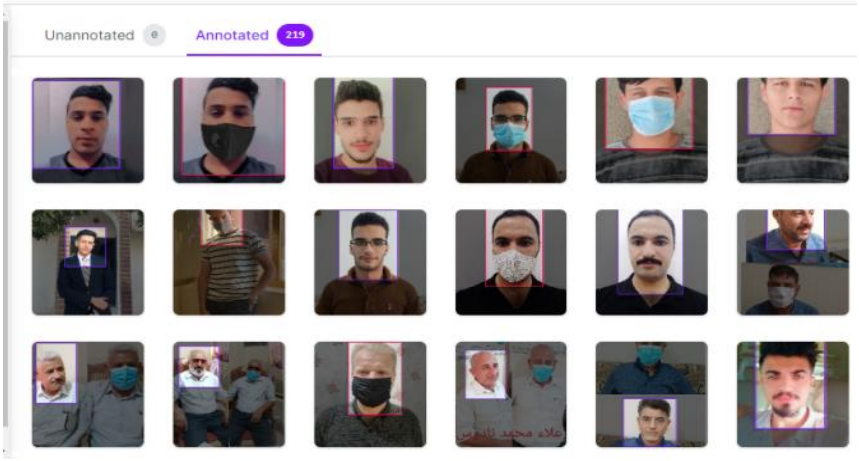


Figure 2: Image Preprocessing and Dataset Partitioning

Here, to detect faces in each image, a set of steps had to be implemented. The datasets were collected from 219 photographs of university students, and these datasets were downloaded and pre-processed for further analysis. All images were resized to 640 × 640 pixels and converted to grayscale for face detection and recognition. The pixels were evaluated and adjusted until they were within the range from 0 to 1. Next, the data set was divided into specific sets: training set 318 images, validation set 29 images, and test set 14 images. The seventh trained YOLO model was evaluated on the validation set to detect faces wearing or not wearing a mask.

C. Data Augmentation

This large, high-quality visual dataset can help in developing a deep learning-based object detection model to detect faces with or without wearing a mask, as well as in model training and prediction accuracy. As a result, a series of photographs was obtained, and it was necessary to increase the number of photographs. For the set of 219 training images, a number of image sampling techniques were applied in order to enhance generalizability and prevent over-detection of the model. Since image grayscale enables a deep learning model to understand more about object characteristics in different lighting situations during training, the aforementioned image grayscale methods apply to 25% of images. The Hue degree ranges between +25 degrees and -25 degrees, the image brightness ranges between +25% and -25%, and the Blur reaches 2.5 pixels, through which the image brightness is improved and reduced. It can be made in complex lighting conditions, and image mirroring (horizontal and vertical mirroring) can increase the viewing angle of people's image. Static Crop function is used: use 25-75% horizontal area and 25-75% vertical

area to flip the original image. The image is divided into upper and lower halves to achieve vertical inversion and change the center of the image symmetrically on the axis that runs horizontally. Therefore, fine detection accuracy of the detection model is obtained by providing all these images under different scenarios. There are 318 images in the final augmented training set, which includes 219 original images and 318 enhanced images. Figure 3 displays the exact distribution of the expanded training set.

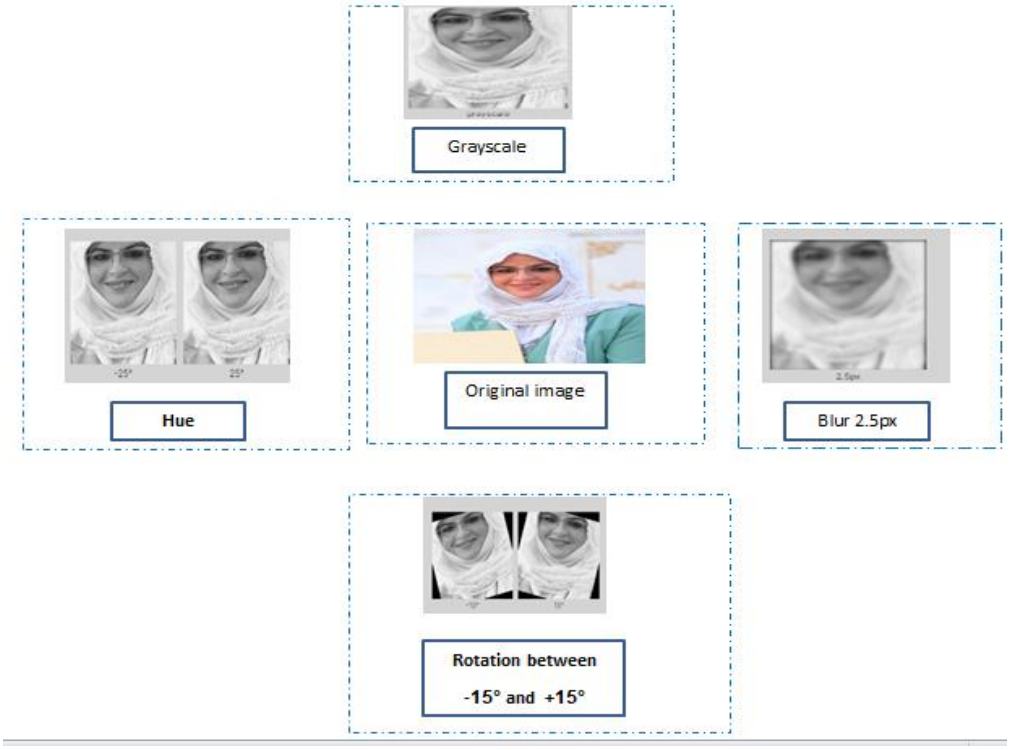


Figure 3: Distribution of the Augmented Training Set

D. YOLO v7 Network Architecture

- YOLOv7 is the latest release in the YOLO family (You Only Look Once) models, which are single-stage object detectors. It aims to provide faster and more accurate object detection performance compared to its predecessors. The YOLOv7 network consists of three main parts: the input network, and the backbone network, and main network. The network design of YOLOv7 is shown in Figure 4, Structure.
- **Input Network:** Preprocesses the image and resizes it to 640 × 640 × 3. Passes the preprocessed image to the backbone network.
- **Backbone Network:** Composed of the CBS composite module, efficient layer aggregation networks (ELAN) module, and MP (maxpooling) module. Reduces the length and width of the feature map by 1/2 while increasing the number of output channels. CBS composite module performs convolution + BN (batch normalization) + activation function on the input feature map. ELAN module improves the learning ability of the network without destroying the original gradient path. MP module halves the length and width of the feature map and channels through maxpooling and convolution operations.
- **Head Network:** Receives the three-layer output from the backbone network. Outputs three layers of feature maps of different sizes. Utilizes the Repconv module to adjust the number of output channels. Performs convolution operations with kernel_size = 1 for objectness, class, and box prediction tasks. Includes the SPPCSPC module, a series of CBS modules, MP module, Catconv module, and subsequent Repconv modules. SPPCSPC module increases the receptive field of the network by performing maxpooling operations with different kernel sizes. Repconv structure improves training time and inference effect by splitting a whole module into multiple branches and re-parameterizing them.

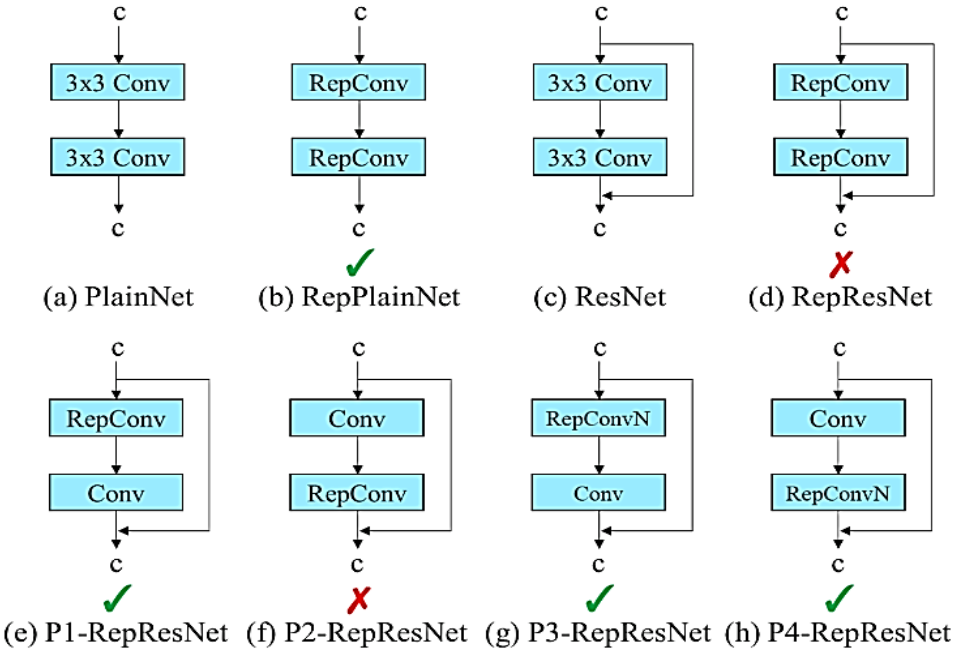


Figure 4: Architecture of Yolov9 Network

E. Parameter Settings

In this work, transfer learning was used using YOLOv7 networks to build a model to detect faces with or without wearing a mask. There were 250 training sessions. What is the batch size of the training model? It is set to 8, which means that there are eight data samples in each stage. The input dimensions are 640 x 640. YOLOv7 incorporates advanced deep learning techniques and architectural design to achieve superior performance in object detection tasks.

Results and Discussion

An example of our complete pipeline can be seen in Figure 5 which shows the classification of faces with and without wearing a mask. The red color represents the designated category without a mask, and the blue color represents the designated category with wearing a mask. When testing images on models of the YOLO algorithm, it was found that detection using the YOLOv7 algorithm obtained a higher degree of accuracy than the previous detection models that had been achieved.

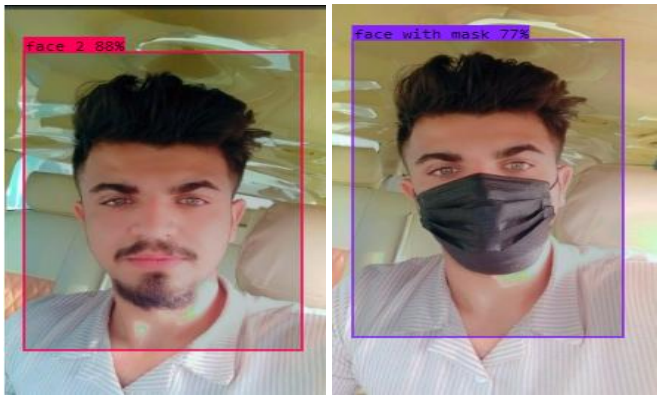


Figure 5: Results of the Real-time Classification of Faces with and Without Wearing a Mask

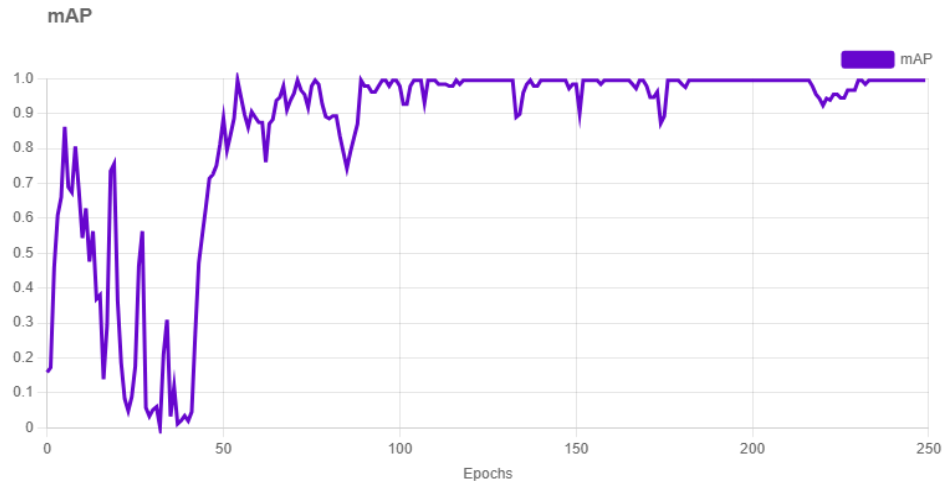


Figure 6: mAP Accuracy of YOLOv7 Model for Classification of Faces with and Without Wearing a Mask

The accuracy rates achieved for detecting faces with and without wearing a mask through the deep learning models of the YOLOv7 algorithm are estimated at 97%, and the precision, recall, and mAP (mean absolute precision) values of the YOLOv7 face detection model are shown in Figure 6. YOLOv7 generally outperforms all Other models of the YOLO algorithm preceded it in terms of speed, and scalability. The improvements in performance are due to advances in model engineering.



Figure 7: Loss Curve of Yolov7 Model Training

In above YOLOv7 graph x-axis represents Epochs & Y-axis represents LOSS/Precision values and blue line represents LOSS values, for the Yolov7 validation at 250 epochs, the box loss decreased to 2.47 and the object loss to 2.48 at 250 epochs. The value of the validation loss peaked when the class loss reached 2.47, at 250 epochs. Figure 7. Loss curve for training the yolov7 model.

Conclusion

In this research, the YOLOv7 algorithm was tested and trained on the data. The highest accuracy rates achieved for face detection by deep learning model are YOLOv7 algorithm mAP (absolute precision) 97%, and accuracy of 87.7% and recall of 87. 1%. YOLOv7 generally outperforms previous YOLO models in terms of speed and scalability but also needs to improve accuracy. Performance improvements are due to advances in model architecture, training techniques, and integration of new features.

References

Chen, W., Huang, H., Peng, S., Zhou, C., & Zhang, C. (2021). YOLO-face: a real-time face detector. *The Visual Computer, 37(4)*, 805-813.

Kumar, A., Kaur, A., & Kumar, M. (2019). Face detection techniques: a review. *Artificial Intelligence Review, 52*(2), 927-948.

Jia, W., Xu, S., Liang, Z., Zhao, Y., Min, H., Li, S., & Yu, Y. (2021). Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector. *IET Image Processing, 15*(14), 3623-3637.

Zhao, Z., Hao, K., Ma, X., Liu, X., Zheng, T., Xu, J., & Cui, S. (2021). SAI-YOLO: a lightweight network for real-time detection of driver mask-wearing specification on resource-constrained devices. *Computational Intelligence and Neuroscience*, *2021(1),* 4529107.

Wan, J., Chen, B., & Yu, Y. (2021). Polyp detection from colorectum images by using attentive YOLOv5. *Diagnostics, 11*(12), 2264. https://doi.org/10.3390/diagnostics11122264

Kim, N., Kim, J.H., & Won, C.S. (2022). FAFD: Fast and Accurate Face Detector. *Electronics, 11*(6), 875.

Hassan, N.F.A., Abed, A.A., & Abdalla, T.Y. (2022). Surveillance system of mask detection with infrared temperature sensor on Jetson Nano Kit. *Bulletin of Electrical Engineering and Informatics, 11*(2), 1047-1055.

Kuznetsova, A., Maleva, T., & Soloviev, V. (2020). Using YOLOv3 algorithm with pre-and post-processing for apple detection in fruit-harvesting robot. *Agronomy, 10*(7), 1016. https://doi.org/10.3390/agronomy10071016

Bai, Q., Li, S., Yang, J., Song, Q., Li, Z., & Zhang, X. (2020). Object detection recognition and robot grasping based on machine learning: A survey. *IEEE access, 8*, 181855-181879.