Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 7, Special Issue "COVID-19: Emerging Research", 2023

Edge Deep Learning and Computer Vision-Based Physical Distance and Face Mask Detection System Using Jetson Xavier NX

Ahmad Aljaafreh [1], Ahmad Abadleh [2*], Saqer S. Alja'Afreh [3], Khaled Alawasa [3], Eqab Almajali [4], Hossam Faris [5]

[1] *Electrical Engineering Department, Tafila Technical University, Tafila, Jordan.*

[2] *Computer Science Department, Mutah University, Karak, Jordan*

[3] *Electrical Engineering Department, Mutah University, Karak, Jordan.*

[4] *Electrical Engineering Department, University of Sharjah, Sharjah, UAE.*

[5] *Information Technology Department, The University of Jordan, Amman, Jordan.*

Abstract

This paper proposes a fully automated vision-based system for real-time COVID-19 personal protective equipment detection and monitoring. Through this paper, we aim to enhance the capability of on-edge real-time face mask detection as well as improve social distancing monitoring from real-live digital videos. Using deep neural networks, researchers have developed a state-of-the-art object detector called "You Only Look Once Version Five" (YOLO5). On real images of people wearing COVID19 masks collected from Google Dataset Search, YOLOv5s, the smallest variant of the object detection model, is trained and implemented. It was found that the Yolov5s model is capable of extracting rich features from images and detecting the face mask with a high precision of better than 0.88 mAP_0.5. This model is combined with the Density-Based Spatial Clustering of Applications with Noise method in order to detect patterns in the data to monitor social distances between people. The system is programmed in Python and implemented on the NVIDIA Jetson Xavier board. It achieved a speed of more than 12 frames per second.

Keywords:

COVID-19; Mask Detection; Social Distancing; YOLOv5s.

Article History:

Received:	17	April	2022
Revised:	30	July	2022
Accepted:	16	August	2022
Published:	26	September	2022

1- Introduction

In machine learning and artificial intelligence, deep neural networks have become powerful tools used for various applications. Learning from existing examples enables neural networks to approximate functions and dynamics. Numerous previous works on image processing enabled by deep learning for various applications [1-13] have been proposed in this regard. Different objects have been detected using image processing, for instance, face recognition, user identity, skin recognition, human recognition, etc. Automated face recognition has been one of the most researched topics in computer vision and biometrics since the 1970s [1, 2]. Recently, deep neural networks have received great attention for face recognition [3, 4]. Generic object category detection is defined as determining whether there are instances of objects from predefined categories or not. Once it exists, the spatial location and extent of each instance are returned [5]. Many approaches have been proposed to detect faces, parts of faces, and features of faces [6–12]. A novel reverse engineering approach is used to detect which facial features are critical for users' identity and to use these

* **CONTACT**: ahmad_a@mutah.edu.jo

DOI: http://dx.doi.org/10.28991/ESJ-2023-SPER-05

features as eye prints [6]. Recently, deep learning techniques have developed as incredible strategies for gaining highlight portrayals naturally from information [7, 8]. Specifically, these systems have given significant upgrades in object identification. Profound learning has changed a wide scope of the AI era, from picture order and video handling to discourse acknowledgment and common language understanding [9, 10].

The Coronavirus Disease 2019 (COVID-19) emerged suddenly in 2019 and has had an impact around the world. As of December 25, 2021, approximately 279 million people have been infected and over 5.4 million people have died [11, 12]. According to studies and health experts, wearing preventive tools like face masks and social distancing can help control COVID-19 virus spread [13]. However, people normally break these preventive tools and protocols. This in-turns reflected in national lockdowns in countries from time to time to avoid virus spikes and accelerate infection rates. To manage this issue, AI technology provides great solutions through computer vision [12, 14]. In this regard, computer vision has been used to detect the use of face masks [15, 16, 17]. However, most of these studies addressed the detection of the mask without paying attention to the correct use of the mask [12]. Moreover, most of the proposed models in the literature are either used for mask detection or person detection for social distancing [18-21]. Therefore, the proposed solution in this paper addresses the gaps in the literature where a single deep model is used to detect the people and the face masks. Moreover, the distance between people is computed in order to produce a warning whether the people are safe or not due to the coronavirus. To the best of our knowledge, we are the first to train a model to detect four objects, namely; " face with mask", "face without mask", "Person", and "face with incorrect mask", which enabled us to deploy the model on the real-time AI Jetson board.

In this work, a dataset of images that have four different objects (face with mask, face without mask, incorrect mask, and persons) is collected to train and evaluate deep models. This research aims to provide on-edge technology for face mask detection in digital videos. Deep neural network research has recently led to the creation of a state-of-the-art object detector called You Only Look Once version five (YOLO5) [22]. This paper builds a dataset of 2.4K source images of the four classes (face with mask, face without mask, incorrect mask, and persons) and evaluates the latest object detection algorithms, focusing on YOLOV5s. The results of the Yolov5s model show that the Yolov5s new network models can extract rich face mask features from images and detect the face mask with a high precision of better than 0.88 mAP. The two key preventive measures against the spread of COVID-19 are: (1) use of protective equipment (masks) and (2) social distancing.

In particular, being able to identify places/areas with a high concentration of people not following the recommendations could help governments predict potential risk/outbreak places and act more precisely/locally and timely—instead of constantly warning the whole country via the mainstream media (which causes negative side effects, including panic). The solution is designed as a series of successive modules, which identify people on images obtained from surveillance/ip/usb cameras and then inspect if that person wears a mask, gloves, etc. For these purposes, advanced deep learning procedures and architecture have been proposed and tested:

The main contributions of this paper are:

- Utilizing the state-of-art deep object detection model (YOLOv5s) to detect four objects: face with mask, face without mask, incorrect masking, and persons in digital real-time video;
- Detecting the social distancing based on DBSCAN method;
- Deploying the models on real-time AI Jetson board.

2- YOLO

You Only Look Once (YOLO) is a new technique for detecting objects [23]. When using YOLO, you can locate objects instantly [24]. YOLO divides the input image into N grids, each with an SxS dimension. Detection and localization of objects are handled by each grid. Based on the convolutional feature extractor, Yolo predicts bounding box coordinates directly. To enhance YOLO's performance, several versions, such as 2015's YOLO, 2016's YOLOv2 and YOLO9000, 2018's YOLOv3, 2020's YOLOv4, and YOLOv5, have been developed [22]. The model architecture remains like that of YOLOv4. It primarily relies upon PyTorch training procedures to improve its performance. YOLOv5 support is easier, and deployment is simpler. One of the major improvements in YOLOv5 is the augmentation of mosaic data with bounding boxes [25, 26]. Yolov5 is available in a variety of sizes: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. It has been written in PyTorch first rather than Darknet, so it is easier to support and deploy. YOLOv5 has better performance than YOLOv4 in terms of speed, accuracy, and size [27].

3- Hardware

This section presents the main components of the hardware used in this study, which are:

- NVIDIA® Jetson Xavier;
- A digital camera;
- MicroSD card (64GB).

The main component is the NVIDIA Jetson Xavier, shown in Figure 1. NVIDIA® Jetson is the world's leading platform for AI at the edge. Its high-performance, low-power computing for deep learning and computer vision makes it the ideal platform for compute-intensive studies. It is possible to develop multi-modal AI applications using the NVIDIA Jetson Xavier NX Developer Kit in as little as 10 W. Additionally, cloud-native support is available [28].

Figure 1. NVIDIA Jetson Xavier [28]

4- Face Mask Detection Model Implementation

Figure 2, shown below, exhibits the main steps of the proposed detection model implementation procedure. The three main steps in the process are (1) data preparation: for gathering, preprocessing, annotating, and augmenting photos to provide a dataset for training, assessing, and testing the YOLO model; (2) model implementation: used for choosing a deep learning model (Yolov5s), training it on both the training and validation datasets, and evaluating it on the test dataset; and (3) model inference: for detection models that are implemented in the real world.

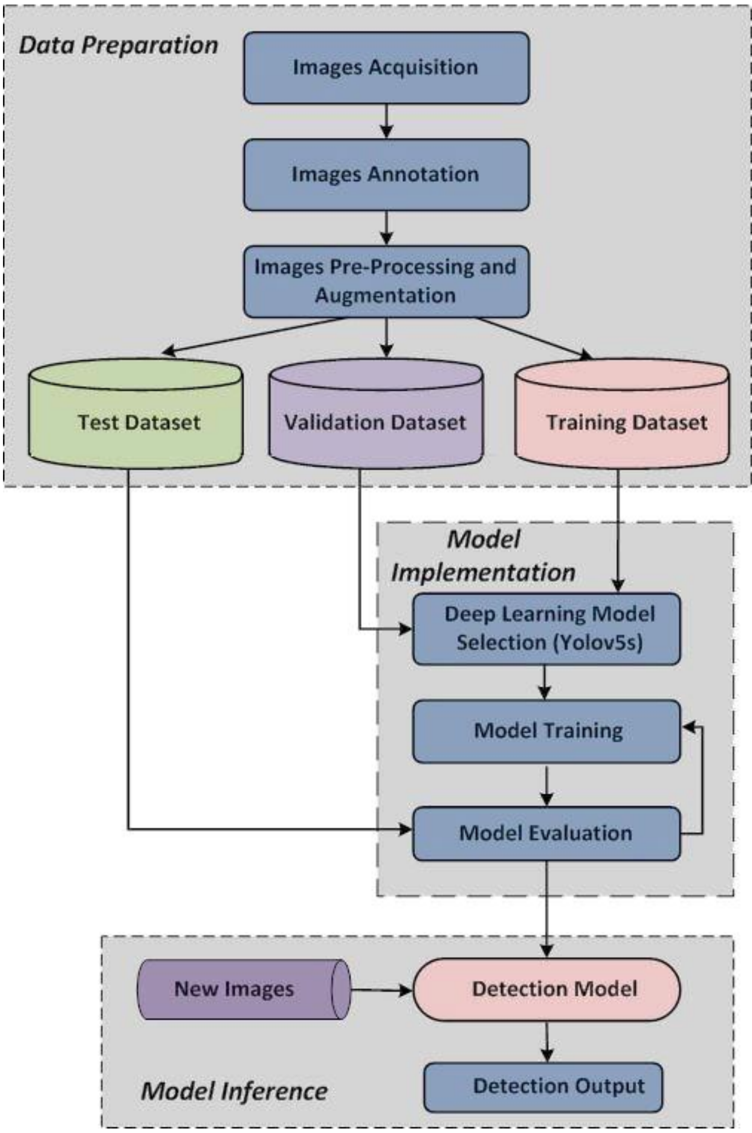

Figure 2. Block diagram of the face mask detection model implementation

4-1- Data Preparation

The proposed detection model aims to detect the face mask and social distancing in real-time. Yolov5s is implemented and trained on real-face mask people's images that are collected from Google Dataset Search. Thus, it can be deployed on-edge to detect face-masked people on the streets using digital cameras. The preparation of the data phase includes the processes of image acquisition, annotation, and augmentation.

4-1-1- Dataset Collection

A set of more than 1K source images from a well-known dataset on kaggale.com to build our training model has been collected. Among the selected images, there were 112 images used for validation and 69 images used for testing. All images are high quality with a dimension of 1094 1459.

4-1-2- Annotation, Pre-processing, Augmentation

Face mask images in the original images were labeled using Roboflow. 1000 different face mask images were annotated with four different classes, namely “with mask”, “without mask”, “Person”, and “incorrect”. A computer vision platform (Roboflow) is used to construct the dataset, which is divided into training, validation, and testing datasets. All images were resized to 800x600 and have been augmented to generate enough images for the detection model.

4-2- YOLOv5 Model Implementation

Object detection is one of the most important problems in computer vision as it helps in recognizing where things are presented in an image as well as classifying them. The YOLO (You Only Look Once) algorithm was introduced in 2015 with a novel method of object reframing. YOLO has been upgraded into five versions and is rated as one of the best object identification algorithms by merging several of the most original ideas from the computer vision research field. YOLOv5s, the latest version of YOLO, is the fifth generation of YOLO. YOLOv5s was created after YOLOv4. However, the YOLOv5 outperforms the YOLOv4 in terms of speed and accuracy [27]. The proposed approach uses YOLOv5s as the object detection model. Figure 3 depicts the flowchart of the paper.

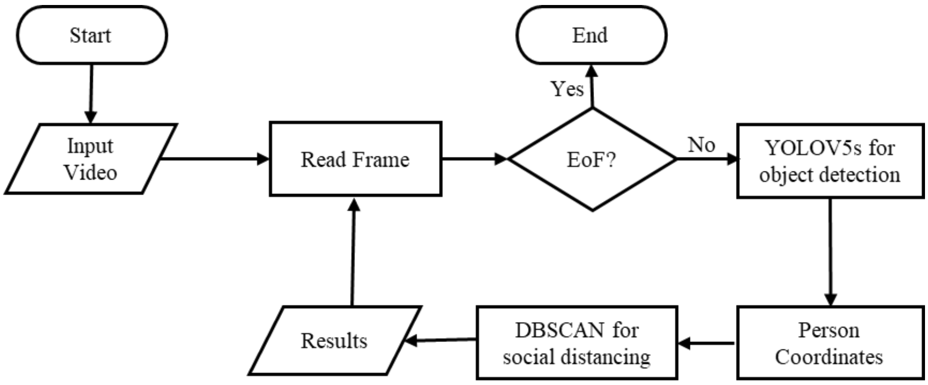

Figure 3. Proposed flowchart for face mask and distance detection

Figure 3 illustrates the flowchart of the face mask and distance detection. The first step of the system is capturing a frame from the input video, and then the YOLOV5s model detects the bounding boxes for four different objects. In this paper, the YOLOV5s model was trained to detect four different objects, namely “face with mask”, “face without mask”, “person”, and “face with incorrect mask”. After detecting these objects, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is used to determine the distance between persons to detect social distancing. Then it is used to cluster people where the minimum sample is 2 and the distance threshold is 1.5 meters.

4-3- Model Evaluation

4-3-1- Experimental Platform

YOLOv5s model was trained and evaluated in this work for detecting the four classes on Tesla K80 GPU that has Linux-5.4.104+-x86_64-with-Ubuntu-18.04-bionic operating system where a Python version 3.7.11 is installed.

4-3-2- Model Evaluation Indicators

The following are the evaluation metrics that were utilized to assess the model:

(a) Precision is a metric that measures a network's ability to recognize targets accurately at a single threshold [1]:

$$\text{Precision} = \frac{\text{True positi ve}}{\text{Actual Results}}\ or \frac{\text{Tp}}{\text{Tp+Fp}} \tag{1}$$

(b) Recall is a metric for how well a network can detect its target [1]:

$$\text{Recall} = \frac{\text{Face mask positive}}{\text{Predicted Results}} \; or \; \frac{\text{Tp}}{\text{Tp+Fn}} \tag{2}$$

(c) Intersection over Union (IoU) is a technique for comparing two arbitrary shapes, such as object widths, heights, and location in the original region, using two boxes. This will test the object detector's precision on a specific data set [1]. In this study, we use mAP_0.5 to evaluate the model where 0.5 means the IoU threshold.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{3}$$

(d) Average Precision is a method that combines recall and precision for the whole ranking [2]:

$$AP = \frac{1}{|Class|} \sum_{c \in class} \frac{TP(c)}{TP(c)+FP(c)} \tag{4}$$

(e) Mean Average Precision (mAP) it is determined by calculating the mean AP across all courses and/or overall IoU thresholds [3]:

$$mAP = \frac{AP}{Total\ number\ of\ class} \tag{5}$$

where, Tp are the Bounding Boxes (BB) having an intersection over union (IoU) greater than 0.5 with the ground truth (GT), Fp, two cases (a) BBs where the IoU with GT is less than 0.5; (b) BBs where the IoU with a GT has already been discovered, Tn, there are not true negative, the image are expected to contain at least one object, Fn, those images were the method failed to produce a BB.

5- Results and Discussion

5-1- Performance Metrices

Yolov5s model has achieved a mAP_0.5 of 0.88. The model is trained for 100 epochs as in Figure 4. As shown in the figure the mAP_0.5 became steady after epoch number 40. Such a result indicates a fast-training speed capability of the proposed model. Each epoch has 2.5k images.

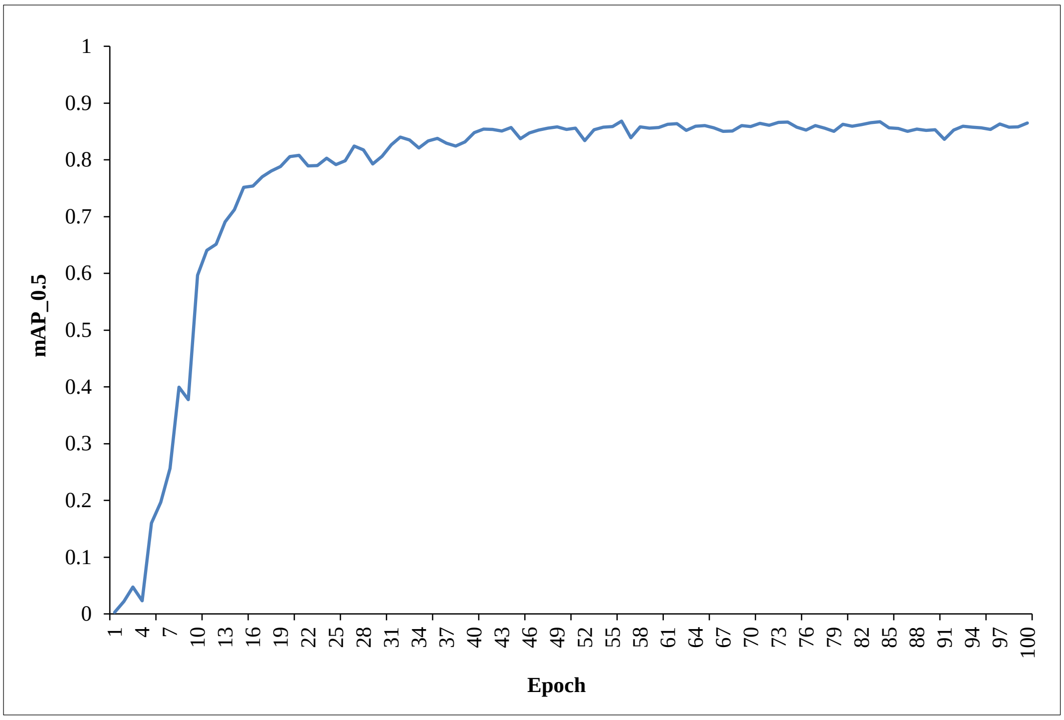

Figure 4. Yolov5s model mAP_0.5 vs epochs

Figure 5 shows the overall precision of Yolov5s model. As shown, the precision is about 0.84 which is acceptable for real-time use. The precision becomes steady after epoch 40. This indicates that the training was quite enough to achieve accurate results.

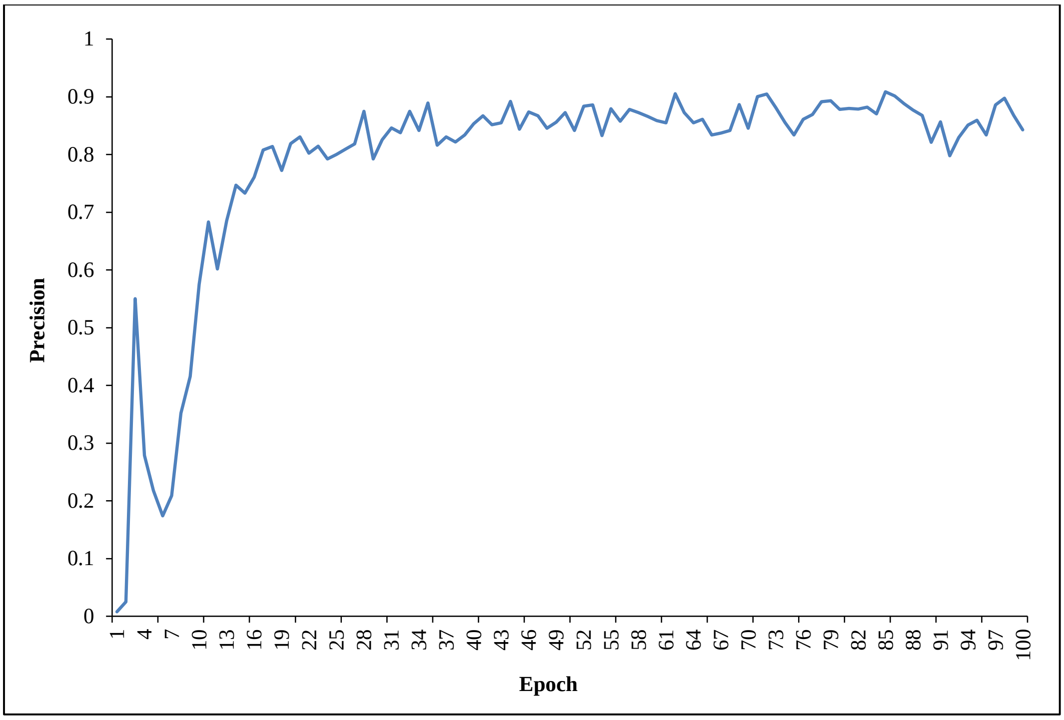

Figure 5. **Yolov5s model precision vs epochs**

Figure 6 shows that Yolov5s model achieved a recall of 0.89. This means that the number of true positives classes is high. The recall became steady after epoch 45, which indicates that the training succeeded to find a high number of true positives results using fewer epochs.

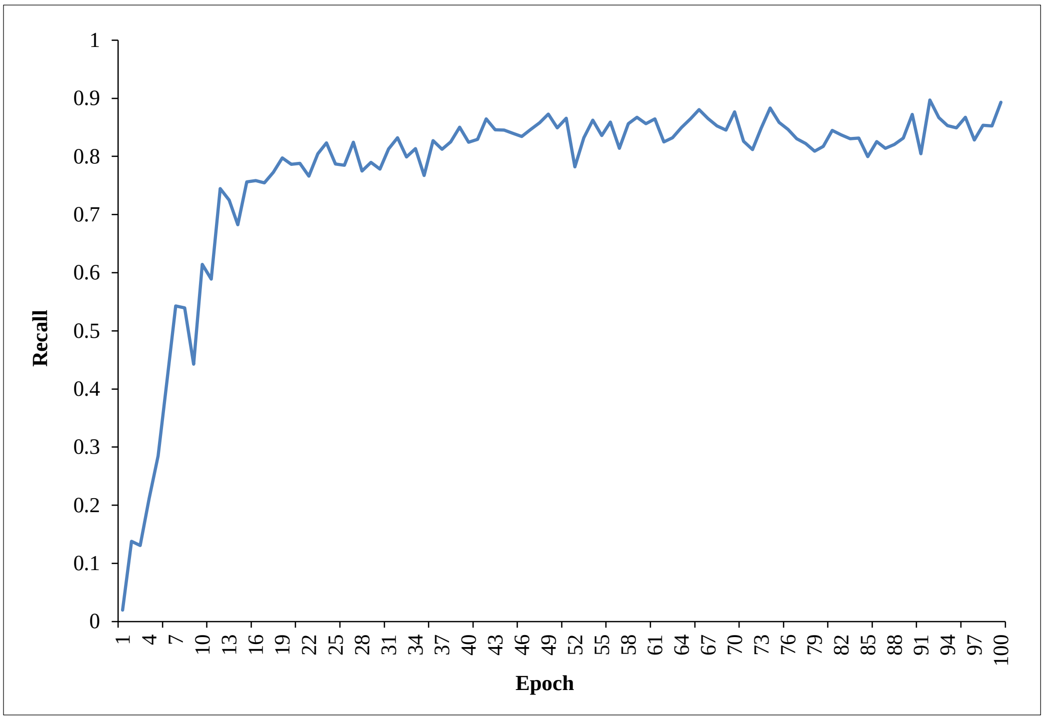

Figure 6. **Yolov5s model Recall vs epochs**

Figure 7 shows the percentage of class loss during the training process. As shown from the figure, there is an inverse correlation between the number of epochs and the class loss. It is an evident that the train class loss is below 0.01 at epoch 40 and has a well convergence as low as 0.005 at the end of training.

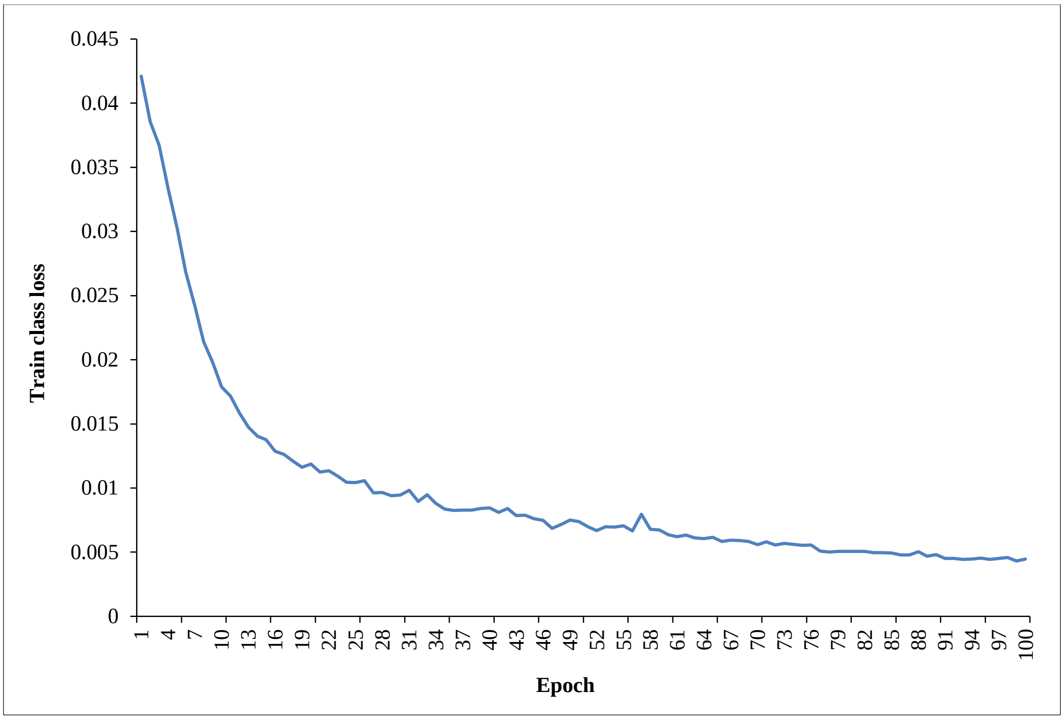

Figure 7. **Yolov5s model training class loss vs epochs**

Figure 8 shows the percentage of object loss during the training process. Train object loss represents the number of incorrectly detected objects. As observed from the figure, the train object loss is inversely proportional to the number of epochs. Eventually, the object loss achieved was about 0.005.

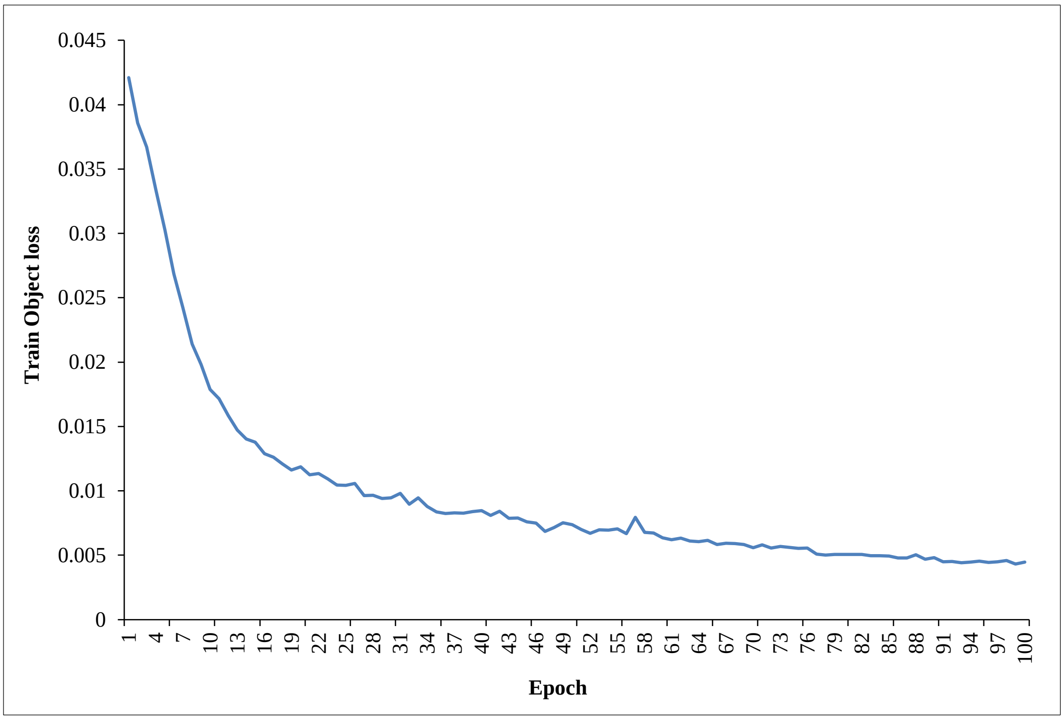

Figure 8. **Yolov5s model training object loss vs epochs**

Figure 9 shows the percentage of class losses during the validation process. As shown in the figure, there is an inverse correlation between the number of epochs and the class loss. Eventually, the number of class losses was about 0.005.

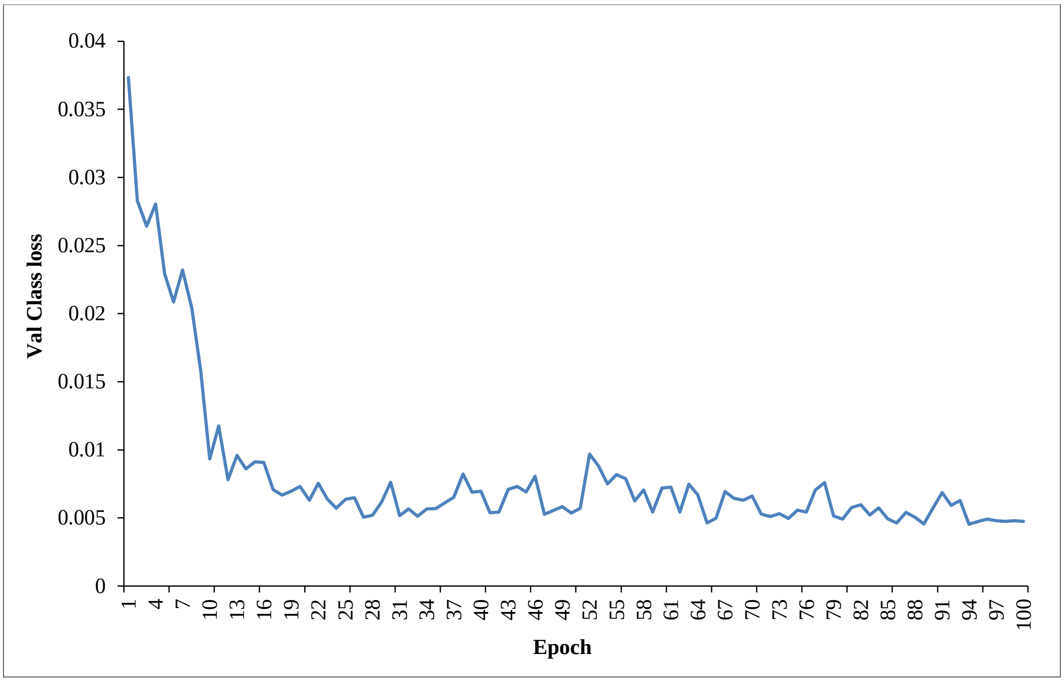

Figure 9. **Yolov5s model validation class loss vs epochs**

Regarding the class loss and object loss in both the training and validation processes, both are almost the same. This indicates there is no overfitting in the model as the results of the training and validation processes converged to same value at the end of the training.

The average results for all classes are displayed in Table 1. As an illustration, the class "with mask" attained 88 percent precision, which is adequate for observing people's commitment to safety. The mAP is calculated over different thresholds. Overall, the mAP averaged about 87 percent. In other words, the proposed approach has a high level of accuracy and can be applied in real-time.

Table 1. **Average results for all classes**

Class	Precision	Recall	mAP@.5
Incorrect	0.979	1.00	0.993
Person	0.816	0.793	0.82
With mask	0.879	0.854	0.893
Without mask	0.696	0.925	0.754
	AVG = 0.843	**AVG = 0.893**	**AVG = 0.865**

5-2- Yolo Model Latency

The model was tested on Tesla k80 GPU. The total time for all images was 1.913 s. The maximum time was 0.026s. The minimum time was 0.017s. The YoloV5s approach can therefore achieve 35 frames per second (FPS), making it suitable for real-time applications.

5-3- Social Distancing

Face mask detection and social distance monitoring were developed based on a deep learning model for the COVID-19 scenario. By combining YOLOv5s object identification with DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering, we were able to identify patterns in the data.

DBSCAN detects social distance between individuals and checks if a predefined "distance parameter" was maintained. This unsupervised learning technique groups comparable points together. There is no need to specify the number of clusters prior to training. It can also ignore noisy or outlier points while creating clusters. People are grouped into clusters based on the midpoints of their bounding boxes. The minimum number of points in the cluster is set to 2, and the distance parameter is set to 150cm, since social distancing is evaluated between a minimum of two individuals. Individuals are grouped together if the distance between them is smaller than the distance parameter. Cluster members are separated by red lines, and people who are too close together are bound by blue boxes if the spacing is less than the distance parameter. A green box will surround the image of a person if it does not belong to a cluster.

5-4- Comparison

To find out the validity and effectiveness of the proposed model for the face mask detection task, the results have been compared with those from YOLOv3, YOLOv4, and tiny YOLO v4 as shown in Table 2.

Table 2. Comparison between the proposed model with other models

Reference	Model	No. of Objects	Dataset size	mAP
The proposed	YOLOv5s	4	2.5 k	0.88
Jiang et al. (2021) [12]	YOLOv3	3	7.4 k	0.86
Kumar et al. (2022) [29]	YOLOv4	4	52 k	0.67
Takimoto et al. (2021) [30]	tiny YOLO v4	4	52 k	0.57

6- Conclusion

In this paper, a completely automated vision-based system for the real-time identification and monitoring of COVID-19 personal protection equipment is proposed. The purpose of this study is to enhance social distance monitoring from live digital recordings and to strengthen the capability of cutting-edge real-time face mask detection. You Only Look Once version five (YOLO5) is a cutting-edge object detector that was created as a result of recent deep learning research. The YOLOv5s object identification model, the simplest variation, is trained and used using actual images of people wearing COVID19 masks that were gathered through Google Dataset Search. A dataset is used in this work to train and evaluate Yolov5s. The Yolov5s model's findings demonstrate that the new network models from Yolov5s are capable of extracting detailed face mask characteristics from images and accurately detecting faces with an accuracy of more than 0.88 mAP 0.5. Then, using YOLOv5s and the Density-Based Spatial Clustering of Applications with Noise technique, patterns are discovered in the data to track social distance between individuals. The NVIDIA Jetson Xavier board is used to implement the Python-based system. More than 12 frames per second were accomplished.

7- Declarations

7-1- Author Contributions

Conceptualization, A.J., H.F., and A.A.; methodology, A.J.; software, A.A.; validation, A.J., A.A., and S.J.; formal analysis, A.J.; writing—original draft preparation, A.J., A.A., E.M. and S.J.; writing—review and editing, K.A., A.J and A.A.; visualization, A.J.; project administration, A.J.; funding acquisition, K.A. The final document was approved by all authors after they had read and approved it.

7-2- Data Availability Statement

The data that support the findings of this study are available in "*Roboflow*" at *https://app.roboflow.com/ds/RdfkjL7YAl?key=83WTKxq5Hs*.

7-3- Funding

This research was funded by Jordan Scientific Research Support Fund under Grant Number Cor-ICT/1/4/2020.

7-4- Institutional Review Board Statement

Not applicable.

7-5- Informed Consent Statement

Not applicable.

7-6- Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition. ACM Computing Surveys, 35(4), 399–458. doi:10.1145/954339.954342.

[2] Li, S. Z., & Jain, A. K. (2011). Handbook of Face Recognition. Springer-Verlag London Limited, London, United Kingdom. doi:10.1007/978-0-85729-932-1.

[3] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2015.7298682.

[4] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. Procedings of the British Machine Vision Conference 2015. doi:10.5244/c.29.41.

[5] Zhang, X., Yang, Y.-H., Han, Z., Wang, H., & Gao, C. (2013). Object class detection. ACM Computing Surveys, 46(1), 1–53. doi:10.1145/2522968.2522978.

[6] Abudarham, N., Shkiller, L., & Yovel, G. (2019). Critical features for face recognition. Cognition, 182, 73–83. doi:10.1016/j.cognition.2018.09.002.

[7] Rajani Kumari, L.V., Saher Fathima, S., Sai Praneeth, G., Mamatha, D., Pranitha, B. (2022). Dynamic Face Recognition System Using Histogram of Oriented Gradients and Deep Neural Network. Sustainable Communication Networks and Application. Lecture Notes on Data Engineering and Communications Technologies, 93, Springer, Singapore. doi:10.1007/978-981-16-6605-6_16.

[8] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning (ICML 2015), 448-456, 6-11 July, 2015, Lille, France.

[9] Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep High-Resolution Representation Learning for Human Pose Estimation. CVPR 2019. Available online: https://jingdongwang2017.github.io/Projects/HRNet/ObjectDetection.html (accessed on June 2022).

[10] Rouxel, F., Yauy, K., Boursier, G., Gatinois, V., Barat-Houari, M., Sanchez, E., Lacombe, D., Arpin, S., Giuliano, F., Haye, D., Rio, M., Toutain, A., Dieterich, K., Brischoux-Boucher, E., Julia, S., Nizon, M., Afenjar, A., Keren, B., Jacquette, A., … Genevieve, D. (2022). Using deep-neural-network-driven facial recognition to identify distinct Kabuki syndrome 1 and 2 gestalt. European Journal of Human Genetics, 30(6), 682–686. doi:10.1038/s41431-021-00994-8.

[11] John Hopkins Universoty (JHU). COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Available online: https://coronavirus.jhu.edu/map (accessed on April 2022).

[12] Jiang, X., Gao, T., Zhu, Z., & Zhao, Y. (2021). Real-time face mask detection method based on YOLOv3. Electronics (Switzerland), 10(7), 837. doi:10.3390/electronics10070837.

[13] Eikenberry, S. E., Mancuso, M., Iboi, E., Phan, T., Eikenberry, K., Kuang, Y., Kostelich, E., & Gumel, A. B. (2020). To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infectious Disease Modelling, 5, 293–308. doi:10.1016/j.idm.2020.04.001.

[14] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. Computational Intelligence and Neuroscience, 2018, 1–13. Doi:10.1155/2018/7068349.

[15] Jiao, L., Zhang, R., Liu, F., Yang, S., Hou, B., Li, L., & Tang, X. (2022). New Generation Deep Learning for Video Object Detection: A Survey. IEEE Transactions on Neural Networks and Learning Systems, 33(8), 3195–3215. doi:10.1109/tnnls.2021.3053249.

[16] Suresh, K., Palangappa, M., & Bhuvan, S. (2021). Face Mask Detection by using Optimistic Convolutional Neural Network. 2021 6th International Conference on Inventive Computation Technologies (ICICT). doi:10.1109/icict50816.2021.9358653.

[17] Crespo, F., Crespo, A., Sierra-Martínez, L. M., Peluffo-Ordóñez, D. H., & Morocho-Cayamcela, M. E. (2022). A Computer Vision Model to Identify the Incorrect Use of Face Masks for COVID-19 Awareness. Applied Sciences, 12(14), 6924. doi:10.3390/app12146924.

[18] Wu, P., Li, H., Zeng, N., & Li, F. (2022). FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public. Image and Vision Computing, 117, 104341. doi:10.1016/j.imavis.2021.104341.

[19] Vibhuti, Jindal, N., Singh, H., & Rana, P. S. (2022). Face mask detection in COVID-19: a strategic review. Multimedia Tools and Applications. doi:10.1007/s11042-022-12999-6.

[20] Shinde, R. K., Alam, M. S., Park, S. G., Park, S. M., & Kim, N. (2022). Intelligent IoT (IIoT) Device to Identifying Suspected COVID-19 Infections Using Sensor Fusion Algorithm and Real-Time Mask Detection Based on the Enhanced MobileNetV2 Model. Healthcare (Switzerland), 10(3). doi:10.3390/healthcare10030454.

[21] Razavi, M., Alikhani, H., Janfaza, V., Sadeghi, B., & Alikhani, E. (2022). An Automatic System to Monitor the Physical Distance and Face Mask Wearing of Construction Workers in COVID-19 Pandemic. SN Computer Science, 3(1), 1–8. doi:10.1007/s42979-021-00894-0.

[22] Github (2022). YOLOv5. GitHub, Inc. Available Online: https://github.com/ultralytics/YOLOv5 (accessed on June 2022).

[23] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition, 7263-7271, 21-26 July, 2017, Honolulu, United States.

[24] Du, J. (2018). Understanding of Object Detection Based on CNN Family and YOLO. Journal of Physics: Conference Series, 1004, 012029. doi:10.1088/1742-6596/1004/1/012029.

[25] Rajput, M. (2020). YOLO V5 is Here! Custom Object Detection Tutorial with YOLO V5. Available Online: https://towardsai.net/p/data-science/YOLO-v5-is-here-custom-object-detection-tutorial-with-YOLO-v5 (accessed on July 2022).

[26] Supeshala, C. (2020). YOLO v4 or YOLO v5 or PP-YOLO?. Available Online: https://towardsdatascience.com/YOLO-v4-or-YOLO-v5-or-pp-YOLO-dad8e40f7109 (accessed on June 2022).

[27] Nelson, J., & Solawetz, J. (2020). YOLOv5 is Here: State-of-the-Art Object Detection at 140 FPS. Roboflow, Available Online: https://blog.roboflow.com/YOLOv5-is-here/ (accessed on June 2022).

[28] NVIDIA (2022). Jeston Xavier NX Series Modules. NVIDIA Corporation. Available Online: https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-xavier-nx/ (accessed on April 2022).

[29] Kumar, A., Kalia, A., & Kalia, A. (2022). ETL-YOLO v4: A face mask detection algorithm in era of COVID-19 pandemic. Optik, 259, 169051. doi:10.1016/j.ijleo.2022.169051.

[30] Takimoto, H., Sato, Y., Nagano, A. J., Shimizu, K. K., & Kanagawa, A. (2021). Using a two-stage convolutional neural network to rapidly identify tiny herbivorous beetles in the field. Ecological Informatics, 66, 101466. doi:10.1016/j.ecoinf.2021.101466.