Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 13(1), 22-32, 2023
Journal of the Institute of Science and Technology, 13(1), 22-32, 2023
ISSN: 2146-0574, eISSN: 2536-4618
Bilgisayar Mühendisliği / *Computer Engineering* **DOI: 10.21597/jist.1243233**
Araştırma Makalesi / *Research Article*
Geliş tarihi / *Received*: 27.01.2023 **Kabul tarihi / *Accepted*: 10.02.2023**

Atıf İçin: Eği, Y. (2023). Kalabalık Kamu Alanları için YOLO V7 ve Bilgisayar Görmesi Temelli Maske Giyim Uyarı Sistemi. *Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 13(1), 22-32.

To Cite: Eği, Y. (2023). YOLO V7 and Computer Vision-Based Mask-Wearing Warning System for Congested Public Areas. *Journal of the Institute of Science and Technology*, 13(1), 22-32.

Kalabalık Kamu Alanları için YOLO V7 ve Bilgisayar Görmesi Temelli Maske Giyim Uyarı Sistemi

Yunus EĞİ[1*]

Öne Çıkanlar:
- YOLO V7
- Derin Öğrenme
- Maske Tanımlama

Anahtar Kelimeler:
- Derin Öğrenme
- YOLO V7
- Covid 19
- Maske Tanımlaması
- Uyarı Sistemi

ÖZET:

Hastane, okul, alışveriş merkezi gibi insanların bir arada olması gereken kalabalık alanlarda sosyal mesafe ve maske takma kurallarına uyulmaması nedeniyle dünya genelinde Covid 19 vakalarının etkisi artıyor. Yetkililer her ne kadar maske takılmamasını engellemek için çeşitli önlemler alsalar da kalabalık ortamlarda maske denetlemesi güç olmaktadır. İnsan eli ile yapılan denetimlerde maske takmayan kişiler gözden kaçabilmekte olup bu durum salgının artışında önemli bir etken olmaktadır. Bu çalışmanın amacı yoğun insan trafiğinin olduğu kalabalık ortamlarda insanların Covid-19 salgınından korunmalarını sağlamak için son teknolojik algoritma olan YOLO V7 derin öğrenme yöntemi ile Yapay Zeka (YZ) destekli maske denetleme sistemi oluşturmaktır.

YOLO V7 and Computer Vision-Based Mask-Wearing Warning System for Congested Public Areas

Highlights:
- YOLO V7
- Deep Learning
- Mask Detection

Keywords:
- Deep Learning
- YOLO V7
- Covid-19
- Mask Detection
- Warning System

ABSTRACT:

The impact of Covid 19 cases is increasing worldwide due to not complying with social distancing and mask-wearing rules in congested areas such as hospitals, schools, and malls where people have to be together. Although the authorities have taken various precautions to prevent not wearing masks, it is challenging to inspect masks in crowded areas. People who do not wear masks can be unnoticed by visual inspections, which is a critical factor in the increase of the epidemic. This study aims to create an Artificial Intelligence (AI) based mask inspection system with the YOLO V7 deep learning method to ensure that overcrowded public areas are protected from the Covid-19 epidemic.

[1] Yunus EĞİ (**Orcid ID**: 0000-0001-5185-8443), American University of Middle East, College of Engineering and Technology, Egaila, 54200, Kuwait

***Sorumlu Yazar / Corresponding Author:** Yunus EĞİ, e-mail: yunus.egi@aum.edu.kw

INTRODUCTION

One way of spreading Covid-19 is the transmission from the infected person through droplets scattered in the air. These droplets are released into the air when people talk, cough, and sneeze. Therefore, people who have been sick can transmit the disease to other people without realizing it. For this reason, it is vital to wear a mask in crowded areas (Liao et al., 2021). Today, face recognition models have an essential role in various scientific areas. Face recognition models are developed by computer vision and deep learning algorithms. These developed models can detect facial features such as wearing a mask, not wearing a mask, and incorrect mask-wearing. We can list some of the algorithms that can be used to detect facial features in the environment where Covid-19 or other types of pandemics are widespread.

EP (Evolutionary Pursuit) is an adaptive algorithm that can classify and define a person's evolutionary characteristics while analyzing images through encoding and classification(Liu et al., 2000). EP can be considered an optimal basis learner that increases machine learning by narrowing the Confidence level. Elastic Bunch Graph Matching (EBGM) is a computer vision method that detects and classifies critical points on the face using graphic templates (Chen et al., 2013). EBGM is constrained by the object, which common landmarks such as frontal face pose and nose shape. In case of missing the landmarks, the algorithm becomes dynamic in terms of attributes and shape. ICA (Independent Component Analysis) is based on the principle of detecting a fundamental component in the image and extracting the function of the other components. Rather than processing the image precisely, it allows processing only specific functions, extracting approximate values, and performing operations on them (Shihab et al., 2022). Linear Discriminant Analysis (LDA) is an algorithm aiming to select distinctive data classification features by eliminating non-discriminatory features. Thus, it is a method that can analyze images according to their attributes, not their content (Jana et al., 2022). Principal Component Analysis (PCA) is a method based on compressing and comparing the image to extract the specific parts in the image (Wang et al., 2022). Haar Cascade algorithm is also used on a trained model for face reading operations. This approach utilizes Haar features obtained by the Haar wavelet transform. Objects are divided into different sub-particles using different color distribution and density information of their sub-particles. The properties of the parts are expressed with different property sets to define the whole object. The face-finding process starts with scanning the whole picture with a classifier that can find facial features of different sizes. It repeats the same search process for different scale parameters to find facial features in different scales in the image that comes as a parameter (Kalangi et al. 2022). Artificial Intelligence (AI) is the ability of a computer or computer-controlled machine to make decisions with mechanisms similar to humans (Feuerriegel et al., 2022). In other words, AI makes the computer think like humans. Recently, AI has been used to detect faces with deep learning algorithms and parallel processing.

Deep Learning is a field of study that covers Artificial Neural Networks (ANN) and similar machine learning algorithms with one or more hidden layers (Eyceyurt et al., 2022). These layers comprise the features' weights that need to be optimized through feed-forward and backward propagation (Srinivas et al, 2012). The feature extraction can be performed via deep Convolutional Neural Networks (CNN) which are a special type of ANN. CNNs are very useful for image classification and segmentation due to their ability to extract features and cope with significant variations (Egi et al., 2022). Deep learning applications such as colonic polyp detection for colorectal cancer, flying object detection, high throughput prediction, and signal power estimation are employed by health, defense, and communication systems(Karaman et al.; Sharma et al.,2018; Murshid et al., 2017; Egi & Eyceurt, 2022).

There are many algorithms used in the field of Deep Learning. One of these algorithms, YOLO (You Only Look Once), is an algorithm that detects objects using convolutional neural networks (CNN)(Pacal et. al. 2022). YOLO algorithm provides high FPS (Frame per Second) due to its high processing speed and gives more precise results. The state-of-the-art version of this algorithm is called YOLO V7 (Wang et al., 2022). So far, the different sizes of the obtained models run in the range of 36 to 161 FPS with high accuracy, which is incredibly remarkable. This high performance makes YOLO a better suit for real-time automatic pandemic mask detection and warning deep learning models for crowded areas.

This research uses a real-time automatic face detection model for pandemic cases such as Covid-19. The rest sections are organized as follows: Material and Methods; Results and Discussion; Conclusion

MATERIALS AND METHODS

You Only Look Once (YOLO)

YOLO is one of the states of art object detection algorithms. It splits the given image into NxN grids, then draws the bounding boxes surrounding the objects in each image and calculates the probability of finding objects in each region within a confidence interval (Redmon et al., 2016). This score reveals how likely the object is the desired object or not. The YOLO algorithm applies non-maximum suppression to objects inside bounding boxes that disables low-confidence objects from evaluation and searches for a high-confidence bounding box in the same region. If the object's midpoint, height, and width are found, then a bounding box is drawn based on this data. A prediction vector is generated for each region, and a confidence score is calculated within these vectors. As a result, if the confidence score is 0, the object does not exist; if the confidence score is 1, the object exists.

YOLO V7 and Its Architecture

YOLO V7 is the latest YOLO version with the highest performance and efficiency in terms of real-time application, including instant segmentation and pose estimation. Researchers consider the number of parameters, computation volume, and processing density to increase performance. Unlike other object detection algorithms, YOLO V7 has many novelties in its architecture, such as Extended Efficient Layer Aggregation Network (E-ELAN) and Compound Model Scaling (Wang et al., 2022).

Extended Efficient Layer Aggregation Network (E-ELAN)

YOLO V7 is built on traditional YOLO models such as YOLO V4 and YOLOR. For previous Efficient Layer Aggregation Network (ELAN) architecture, directing gradient path with a deeper will be converged in the most optimized way.

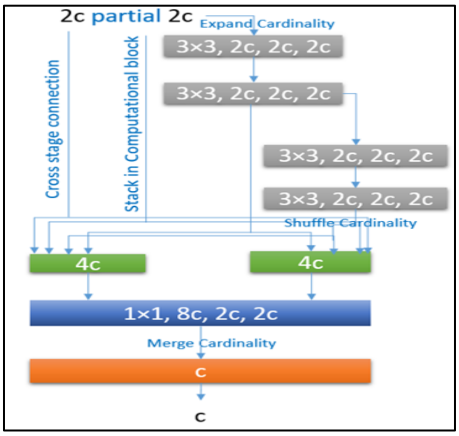

Figure 1. YOLO V7 E-ELAN Architecture

The problem with regular ELAN is that it cannot handle large computational blocks and be destroyed with its gradient path, leading to less parameter utilization. YOLO V7 seven solves this issue by using expand, shuffle, and merge cardinality, as seen in Figure 1(Hussain et al., 2022).

YOLOv7 compound model scaling

The model scaling is one of the essential key attributes that optimize the model with depth and resolution. The current scaling approaches face different problems in evaluating independently and using the different scaling factors. In other words, any scaling changes will affect the transition layer's input-output ratio. The YOLO V7 solves that problem by introducing compound model scaling without touching the initial design, as seen in Figure 2(Onyeogulu et al., 2022).

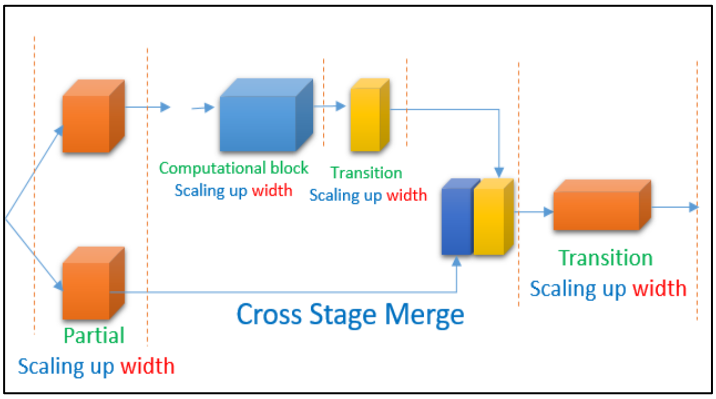

Figure 2. YOLO V7 compound scaling for concatenation-based model.

Trainable bag-of-freebies

In YOLO V7, there are two trainable bags of freebies named planned re-parameterized convolution (RepConv) and Coarse for auxiliary and fine for lead loss (CAFL). The main benefit of RepConv is the minimization of accuracy loss without identity connection. In addition, CAFL removes the single-head limitation of the architecture by using an assigner mechanism, as seen in Figure 3 (Wang et al., 2022).

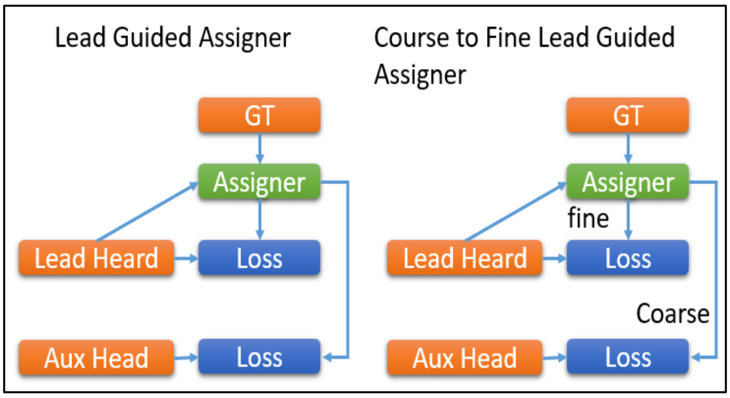

Figure 3. Re-parameterized convolution assigner mechanism

Data Acquisition

For this research, data containing 880 images(Make ML et al, 2022) are obtained through the Kaggle data platform and google creative common license images. This data was augmented through Roboflow online data preparation web tool, and 2640 images were obtained. The images were categorized into three classes incorrect mask-wearing, wearing the mask, and not wearing the mask, and represented as 0, 1, and 2, respectively. These categories are divided into training, test, and test data with a ratio of 70:20:10. An overview of the raw and augmented dataset is provided in Table 1.

Table 1 Overview of raw and augmented dataset

Dataset	Raw images	Row dataset annotations	Augmented dataset(x3)	Training dataset(70%)	Validation dataset(20%)	Test dataset(10%)
Images with 'Incorrect mask wearing' class	95	121	285	200	57	28
Images with 'wearing mask' class	795	3420	2385	1670	477	238
Images with 'not wearing mask' class	304	773	912	638	182	92
All images	880	4314	2640	1848	528	264

All the data is annotated by Roboflow online labeling tool. The batch of the training dataset is represented in Figure 4.

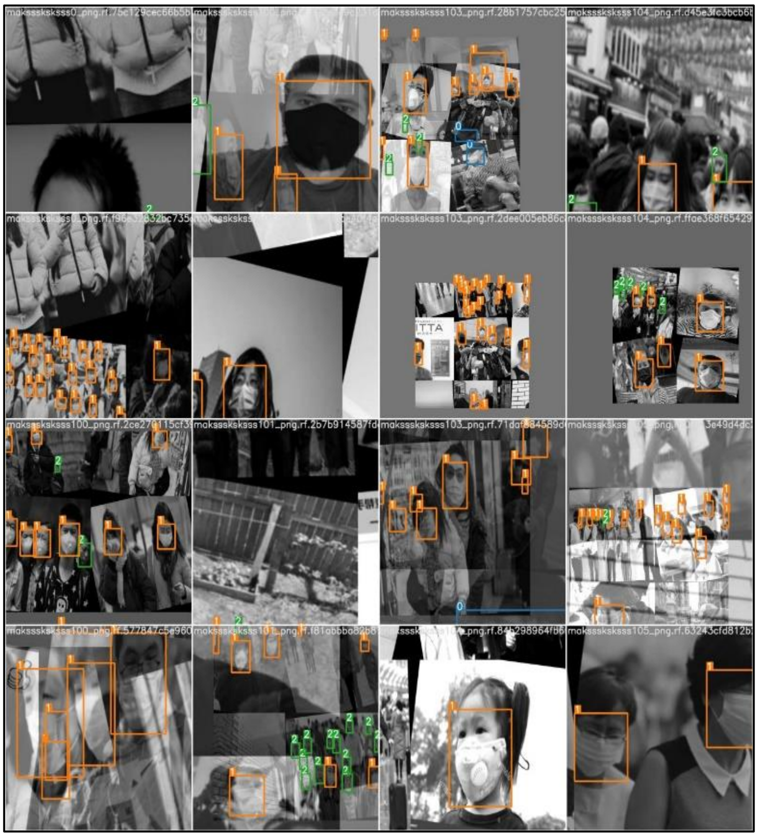

Figure 4. Training batch data with annotations

Performance Parameters

Performance parameters play an essential role in assessing the success of the deep learning model. In this research, Precision, Recall, F1 score, and mAP will be used to evaluate the accuracy of the YOLO V7 model. Here, the precision is calculated by dividing the true positive estimation by whole estimations. On the other hand, recall values only provide positive examples. The F1 score is critical since it provides which confidence level precision and recall values are evenly scattered. In addition, the PR curve (precision and recall) demonstrates the amount of loss among precision and recall for different thresholds. The mAP, which stands for Mean Average Precision, summarizes the model. The following equations represent how the performance parameters are computed.

$$\text{Precision} = \frac{\text{TruePositive}}{\text{True Positive} + \text{False Positive}} \quad (1)$$

$$\text{Recall} = \frac{\text{TruePositive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

$$\text{F1} = \frac{2\text{Precision}.\text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{mAP} = \frac{1}{n}\sum_{1}^{n} AP_k \quad (4)$$

Where:

AP= $\sum_0^n [\text{recalls}(i) - \text{Recalls}(i + 1)].\text{Precision}(i)$.

n and i are the number of classes and thresholds, respectively

The Training Process and Model Deployment

The training process is performed on GPU-aided Google Colab Notebook, which lets the user execute the provided python code or rich text. The used GPU is equipped with Tesla T4 and 15 GB Ram from Google Colab. The hyperparameters such as learning rate, batch size, and epoch numbers are selected as 0.01, 16, and 55, respectively. A deep learning model has been developed through the training process for real-time automatic pandemic mask detection and warning. The obtained model is deployed on an external computer and a connected speaker to warn people, as seen in Figure 5.

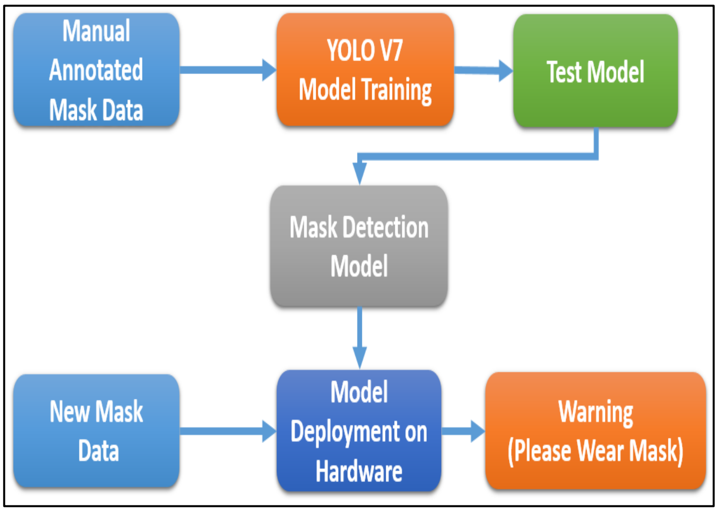

Figure 5. Proposed deployment model

RESULTS AND DISCUSSION

In this article, computer vision and YOLO V7 deep learning model are used to develop a real-time mask warning model for crowded areas. The YOLO V7 model is trained through Google Colab and Python with 1848 augmented images, corresponding to %70 of the data. The required TensorFlow deep learning and data libraries are imported. The training is initialized with a learning rate of 0.01 and an epoch of 55. The results are represented in Figure 6. After training is completed, all classes' best precision and recall values are obtained at 0.798 and 0.698, respectively. The mAP@.5 value 0.718 for all classes demonstrates a significant detection result.

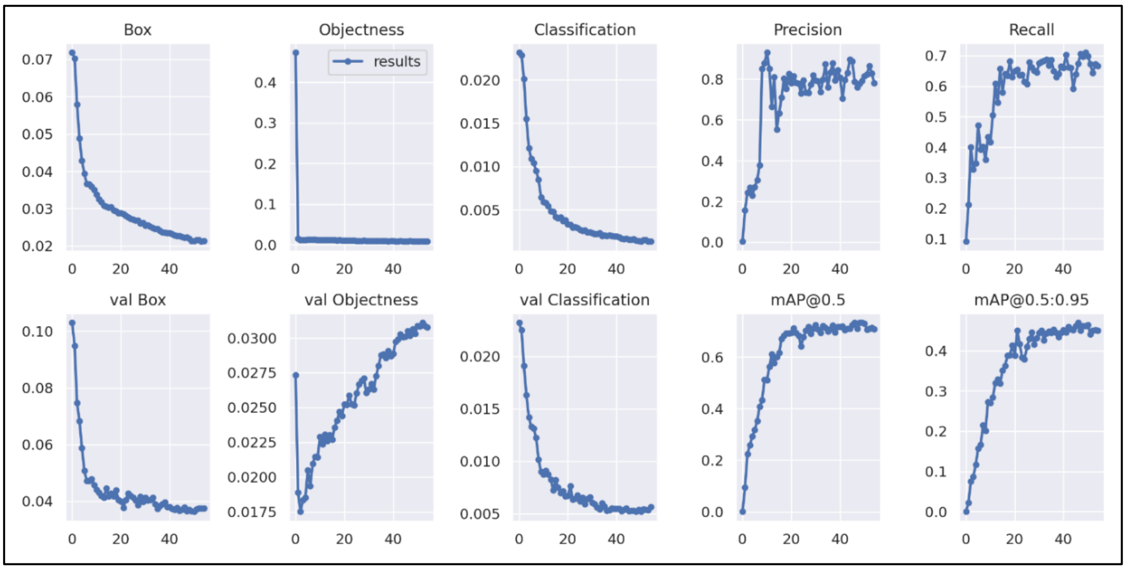

Figure 6. Training batch data with annotations

The highest mAP@.5 detection accuracy is obtained at 0.92 for the "wearing mask" class. YOLO V7 detection results are represented in Table 2.

Table 2. YOLO V7 detection results

Class	P	R	mAP@.5	mAP@.95
All	0.798	0.698	0.718	0.449
Incorrect Maks Wearing	0.704	0.476	0.464	0.33
Wearing Mask	0.925	0.881	0.9227	0.574
Not Wearing mask	0.765	0.735	0.763	0.444

The training performance of individual classes based on confidence level is also critical for a robust model evaluation. The F1 score and the PR graph are plotted. The best precision for all classes is obtained at the confidence level of 0.957, as indicated in Figure 7.

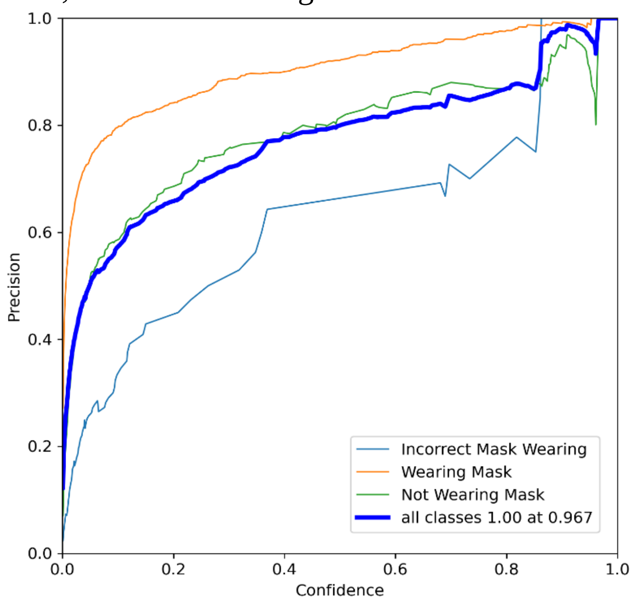

Figure 7. Precision-Confidence level graph for individual and all classes

However, this value is varied for other individual classes. This variability can be attributed to the different number of samples and image noise levels.

The recall values versus confidence level are plotted in Figure 8 to identify how accurately the YOLO V7 model explains the data. The recall value of 0.89 for all classes at 0 confidence shows that the model has a significant level of identification of the data. In addition, it is seen that the incorrect wearing mask approximately achieves the recall values of 0.62 due to the low number of samples in the data. The opposite case can also be seen with the class wearing the mask. The number of samples is relatively high to explain the uncertainty of the data.

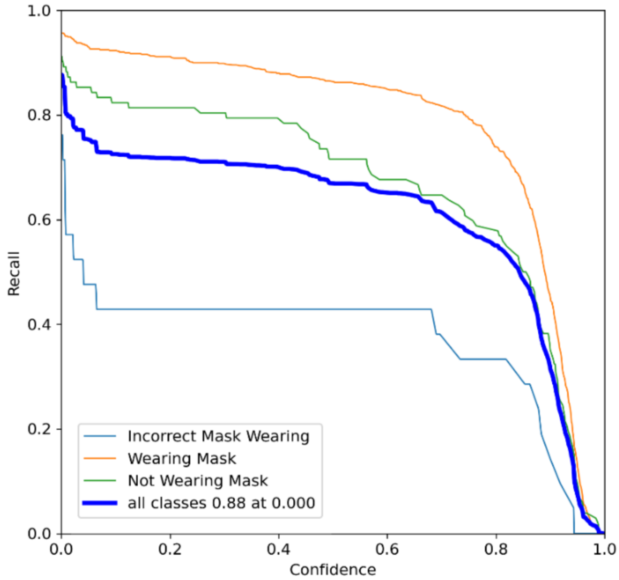

Figure 8. Recall-Confidence level graph for individual and all classes

To find which confidence level precision and recall values are evenly scattered, the F1 score is plotted as seen in Figure 9.

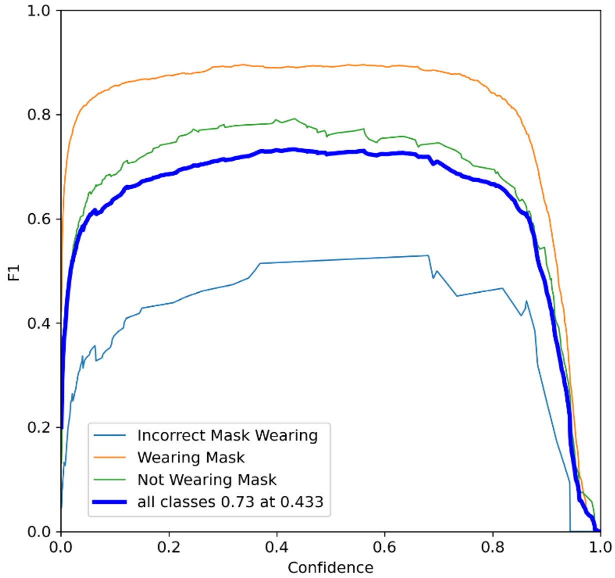

Figure 9. F1-Confidence level graph for individual and all classes

According to the results, an F1 score of 0.72 for all classes is obtained at a 0.516 confidence level, meaning there is a decent amount of scattering between classes. The F1 scores are higher for the classes wearing the mask and not wearing the mask, which means that there are low false positives and false negatives in those classes. In contrast, for incorrect mask-wearing, the F1 score is below the average, indicating high false positives and high false negatives. Finally, the PR graph is plotted to see the amount of loss among precision and recall values for different thresholds from 0-1, as seen in Figure 10.

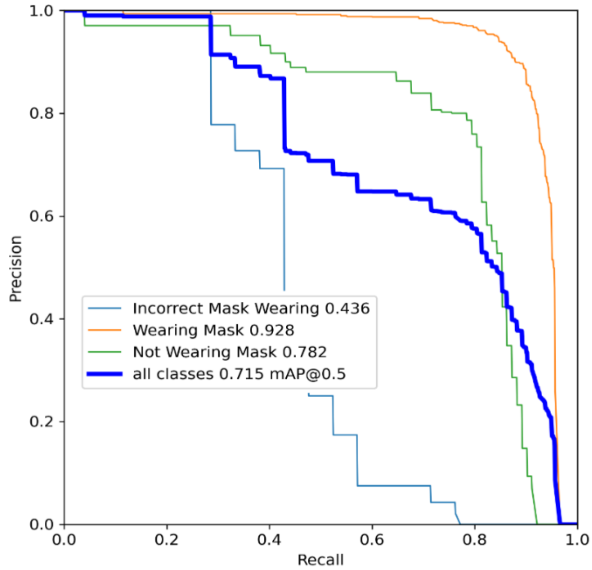

Figure 10. P-R graph for individual and all classes

It is seen that minimum loss is obtained from wearing the mask class. In addition, the average loss for all classes is worse than not wearing the mask when the recall value is lower than approximately 0.85. Taking all into account, the YOLO V7 model has a high training performance in the detection of people with masks and without masks. The best detection results with the final model are represented in Figure 11.

Figure 11. P-R graph for individual and all classes

After obtaining the best results with YOLO V7, the model is compared with previous models such as YOLO V6 and YOLO V5 as seen in Table 3. P, R, mAP@.5, and mAP@.95 are selected as performance parameters(Pacal et al., 2022). According to F1 scores, YOLO V7 shows the best performance among all models with a value of 0.73. The reason for having an F1 score less than 0.9 is the fact that the incorrect mask-wearing class has fewer samples than the other two classes. The MT-YOLO V6 and YOLO V5 have lesser F1 scores than YOLO V7 since the precision and recall values show insignificant results for the given data. Similar results can be obtained from mAP@.5 and mAP@.95 values as well. YOLO V7 has the highest mAP@.5 and mAP@.95 with values of 0.718 and

0.449, respectively. In summary, the YOLO V7 demonstrates the best performance comparing the previous YOLO models.

Table 3 Comparison of YOLO models based on performance parameters

Class	P	R	mAP@.5	mAP@.95	F1 Score
YOLO V7	0.798	0.698	0.718	0.449	0.73
MT-YOLO V6	0.435	0.644	0.435	0.246	0.51
YOLO V5	0.829	0.398	0.497	0.280	0.53

CONCLUSION

In this research, we built a YOLO V7 deep learning-based warning system that distinguishes the people wearing the mask, not wearing the mask, and incorrect mask-wearing as three different classes in real-time. The primary purpose is to create awareness of social distance in crowded areas. According to mAP@.5 results, there is a significant detection accuracy for all classes, wearing the mask, not wearing the mask, and incorrect mask wearing with 0.718, 0.464, 0.922, and 0.763. YOLO V7 also shows the best performance against MT-YOLO V6 and YOLO V5. In the future, this work can be implemented in congested areas such as malls, schools, and restaurants.

Conflict of Interest

The author declared that there is no conflict of interest.

REFERENCES

Chen X, Zhang C, Dong F, Zhou Z, (2013). Parallelization of elastic bunch graph matching (EBGM) algorithm for fast face recognition. In 2013 *IEEE China Summit and International Conference on Signal and Information Processing* (pp. 201-205). IEEE.

Egi, Y., Hajyzadeh, M., & Eyceyurt, E. (2022). Drone-Computer Communication Based Tomato Generative Organ Counting Model Using YOLO V5 and Deep-Sort. Agriculture, 12(9), 1290.

Egi, Y., Eyceyurt, E. (2022). Classified 3D mapping and deep learning-aided signal power estimation architecture for the deployment of wireless communication systems. *J Wireless Com Network* 2022, 107. https://doi.org/10.1186/s13638-022-02188-2

Eyceyurt E, Egi Y, Zec J, (2022). Machine-Learning-Based Uplink Throughput Prediction from Physical Layer Measurements. Electronics, 11(8), 1227.

Feuerriegel S, Shrestha YR, von Krogh G, Zhang C (2022). Bringing artificial intelligence to business management. Nature Machine Intelligence, 4(7), 611-613.

Hussain M, Al-Aqrabi H, Munawar M, Hill R, Alsboui T, (2022). Domain Feature Mapping with YOLOv7 for Automated Edge-Based Pallet Racking Inspections. Sensors, 22(18), 6927.

Jana S, Thangam S, Selvaganesan S, (2022). Gender Identification Using Ensemble Linear Discriminant Analysis Algorithm Based on Facial Features. In Machine Learning and Autonomous Systems (pp. 23-35). Springer, Singapore.

Kalangi RR, Maloji S, Sundar PS, Ahammad SH, (2022). Deployment of Haar Cascade Algorithm to Detect Real-Time Faces. In 2022 *4th International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 1676-1680). IEEE.

Karaman, A., Karaboga, D., Pacal, I. et al. (2022). Hyper-parameter optimization of deep learning architectures using artificial bee colony (ABC) algorithm for high performance real-time automatic colorectal cancer (CRC) polyp detection. Appl Intell. https://doi.org/10.1007/s10489-022-04299-1

Liao M, Liu H, Wang X, Hu X, Huang Y, Liu X, Lu JR, (2021). A technical review of face mask wearing in preventing respiratory COVID-19 transmission. *Current Opinion in Colloid & Interface Science*, 52, 101417.

Liu C, Wechsler H, (2000). Evolutionary pursuit and its application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6), 570-582.

Make ML, Mask Dataset, Link:https://www.kaggle.com/code/nageshsingh/mask-and-social-distancing-detection-using-vgg19/data Accessed: 10/10/2022.

Murshid SH, Chowdhury B, Eyceyurt E, Alanzi SF, (2017). Experimental verification of two spatially multiplexed 10Gbps channels in single core multimode fibers for data center applications. *In Frontiers in Optics* (pp. JW4A-59). Optical Society of America.

Onyeogulu T, Islam A, Khan S, Teeti I, Cuzzolin F, (2022). Situation Awareness for Automated Surgical Check-listing in AI-Assisted Operating Room. arXiv preprint arXiv:2209.05056.

Pacal, I., Karaman, A., Karaboga, D., Akay, B., Basturk, A., Nalbantoglu, U., & Coskun, S. (2022). An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets. *Computers in biology and medicine*, 141, 105031.

Pacal, İ. (2022). Deep Learning Approaches for Classification of Breast Cancer in Ultrasound (US) Images . *Journal of the Institute of Science and Technology* , 12 (4) , 1917-1927 . DOI: 10.21597/jist.1183679

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the *IEEE conference on computer vision and pattern recognition* (pp. 779-788).

Sharma, D. Y. K., & Pradeep, S. (2019). Deep Learning based Real Time Object Recognition for Security in Air Defense. *Proceedings of the 13th INDIACom, New Delhi, India*, 13-15.

Shihab MA, Al-Dabagh MZ N, jawad kadhim Alrubaie A, Al-Yoonus M, Ghazal M, (2022). Face Recognition System Using Independent Components Analysis and Support Vector Neural Network Classifier. *In ITM Web of Conferences* (Vol. 42). EDP Sciences.

Srinivas, Y., Raj, A. S., Oliver, D. H., Muthuraj, D., & Chandrasekar, N. (2012). A robust behavior of Feed Forward Back propagation algorithm of Artificial Neural Networks in the application of vertical electrical sounding data inversion. *Geoscience Frontiers*, 3(5), 729-736.

Wang CY, Bochkovskiy A, Liao HYM, (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696.

Wang Y, Wu Q, (2022). Research on Face Recognition Technology Based on PCA and SVM. In 2022 *7th International Conference on Big Data Analytics (ICBDA)* (pp. 248-252). IEEE.