International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Performance Evaluation for Face Mask Detection Based on Mult Modification of Yolov8 Architecture

Muna Jaffer AI-Shamdeen[1] and Fawziya Mahmood Ramo[1]

[1] Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Iraq

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: This work aims to engineer a robust system capable of real-time detection, accurately discerning individuals who are either adhering to or neglecting face mask mandates, across a diverse range of scenarios encompassing images, videos, and live camera streams. This study improved the architecture of YOLOv8n for face mask detection by building a new two-modification version of Yolov8n model to improve feature extraction and prediction network for YOLOv8n. In Proposed yolov8n-v1, we outline the integration of a residual Network backbone into the Yolov8n architecture by replacing the first two layers of yolov8n with ResNet_Stem and ResNet_Block modules to improve the model's ability for feature extraction and replace Spatial Pyramid Pooling Fast (SPPF) module with Spatial Pyramid Pooling-Cross Stage Partial (SPPCSP) modules which combine SPP and CSP to create a network that is both effective and efficient. The proposed yolov8n-v2 is built by integration Ghostconv and ResNet_Downsampling modules into the Proposed yolov8n-v1 backbone. All models have been tested and evaluated on two datasets, The first one is MJFR dataset, which contains 23,621 images. and collected by the authors of this paper from four distinct datasets, all of which were used for facemask detection purposes. The second one is MSFM object detection dataset has been collected from groups of videos in real life and images based on the curriculum learning technology. The model's performance is assessed by using the following metrics: mean average precision (mAP50), mAP50-95, recall (R) and precision (P). We conclude that both versions of Proposed yolov8n outperform the original model in terms of accuracy for both datasets. Finally, the system was successfully implemented in one of the medical clinics affiliated with a medical complex, where the results of its application showed high efficiency in various aspects of work, and it effectively contributed to improving the public health and safety.

Keywords: Yolov8, Object detection, Detection Algorithm, The Residual network

1. Introduction

Our daily lives have become embedded into object detection applications including people counting, face detection, text detection, face mask detection, pose detection, animal detection, vehicle detection, and more. Object detection is an important and powerful computer vision method that focuses on classifying(identification) and localizing multiple objects in an image, videos, and even real-time video. The process of identifying the correct location of one or multiple objects using bounding boxes, which correspond to rectangular shapes around the objects are called image Localization [1]. Typically, object detectors consist of three basic parts. 1) The part for features extraction from the provided image which is called Backbone .2) The feature network, which receives input from the backbone at various feature levels and generates a list of fused features that reflect the key aspects of the image.3) The last part which is the final class/box network predicts the class and position of each object using the fused features [2][3]. For everyone in the world, the dissemination of the 2019 coronavirus illness, commonly referred to as COVID-19, is a major apprehension. It is an infectious illness that has had an impact on human life all around the world. According to medical professionals, the virus may spread by either direct or indirect contact with a person who is infected [4]. The world was greatly affected by the 2019 Coronavirus Disease (Covid-19) pandemic. Globally, Covid-19's contagious spread has impacted nearly 172 million people as of May 2021[5][6]. Wearing a face mask is among the greatest strategies to avoid the transmission of COVID-19 in accordance with the world health organization (WHO). Face masks are required in many nations, especially in public areas. Also face masks were once worn by people to safeguard their health from air pollution and in the medical domain [7]. Deep learning is a new part of machine learning methods that has recently grown in prominence, which is based on artificial neural networks to model and solve complex problems [8]. To learn various features with various levels of abstraction, deep learning refers to architectures that uses several layers among the input and output layers with higher level learnt features expressed in terms of lower-level characteristics [9][10]. In the deep learning the key characteristic is that the layers of features are gradually learned from data using

E-mail:author's email

a general-purpose learning technique instead of being handcrafted and created by human engineers [11][12].

In 2023, Yunus EĞİ presented YOLO V7 deep learning-based warning system that makes a distinction in real time among those who are and aren't wearing masks and individual who wear masks improperly. The detection accuracy in term of mAP@.5 results of for all classes, wearing the mask, not wearing the mask, and improper mask wearing are 0.718, 0.464, 0.922, and 0.763 respectively, the class of not wearing the mask have a much higher accuracy according to mAP@.5 among the other classes [13] .In 2022, Vrigkas and et al. presented three versions of the Yolo model, namely Yolov3, Yolov4, and Yolov4-tiny, to recognize persons wearing masks using the Facemask picture dataset that consist of 4866 images for mask and no-mask classes , carefully chosen to correlate to real-world settings and the result of experiment shows that YOLOv4 presents the best performance than yolov3 and YOLOv4-tiny[14]. In 2022, Kumar and et al. suggested ETL-YOLO v4 for detection of face mask that is a modified version of tiny YOLO v4 to improve feature extraction and prediction network. A dense SPP network is initially added to the feature extraction network followed by adding two additional detection layers are added to ETL-YOLO v4. The suggested ETL-YOLO v4 acquired 9.93% higher mAP, 5.75% higher average precision (AP) for faces with masks, and 16.6% higher average precision (AP) for the face mask region in comparison to its original base-line form [15]. In 2021, Loey et al presented a hybrid model for the detection of face mask using deep transfer learning and classical machine learning, the hybrid model consists of two parts. The first part uses Residual Neural Network (Resnet50) feature extraction technology. While the second part is intended for classifying the face masks by employing support vector machines (SVM), decision trees and ensemble algorithms. for the experiment three different datasets of face masked have been used. the Real-World Masked Face Dataset (RMFD), the Simulated Masked Face Dataset (SMFD), and the Labeled Faces in the Wild (LFW). During the testing experiment, the SVM obtained the highest detection accuracies as compared to decision trees and ensemble algorithm classifiers. The SVM attained an accuracy of 99.64% in RMFD and an accuracy of 99.49% in SMFD, while in LFW achieved 100% accuracy [16].

After studying research's related to the subject of the research and to increase the efficiency of the results to detect face mask, we used yolov8n raw model, studied and analyses its results then make a modification to the head and backbone network architecture of yolov8n raw model. The results were evaluated using international standards metrics (P, R, mAP@.5 and mAP@.5:95) and very efficient results were obtained compared to the original yolov8n. The contributions of this paper are as follows.

1) Impact on Public Health and Safety: this work addresses a critical societal issue by contributing to the development of automated systems that promote public health and safety. The real-time face mask detection system has the potential to assist authorities, businesses, and institutions in enforcing mask-wearing protocols during public health emergencies, thereby reducing the risk of disease transmission and enhancing community well-being.
2) Building and compilation of large dataset: Created two datasets for face mask detection and addressing the challenges posed by varying environmental conditions.
3) Proposed a two new hybrid models using a YOLOv8n deep learning and a residual Network backbone tailored for face mask detection, optimizing accuracy for both proposed models more than the original models.
4) Real-Time Accuracy: achieving real-time face mask detection without compromising the accuracy of the predictions. Real-time applications require swift decision-making, but maintaining a high level of precision in distinguishing between masked and unmasked individuals is imperative.

The remaining parts of this work is arranged in a specific style: The original yolov8n architecture is displayed in section two, section3 covers the proposed Yolov8n, section four covers the description of datasets section5 covers the analysis of results, finally section 6 covers the conclusions.

2. YOLOv8 MODEL

YOLOv8 is a newer version that builds on the success of YOLOv7. Which is an object detection neural network model that able to identify objects in images and videos. The model is made up of a backbone and a head. Features from the given image are extracted by the backbone while the head is responsible for detecting objects based on these features [17]. It uses a modified version of the ResNet architecture as its backbone network and introduces several new techniques, such as multi-scale prediction and cross-scale connections, to further improve object detection performance [18][19]. The backbone network architecture of yolov8n is presented in table1.

Table 1. Backbone network architecture of yolov8n [20].

No.of Layer	From	Repeat	Module type	Filter	Stride	Padding
0	-1	1	Conv	64	2	1
1	-1	1	Conv	128	2	1
2	-1	3	C2f	-----	--	---
3	-1	1	Conv	256	2	1
4	-1	6	C2f	-----	--	---
5	-1	1	Conv	512	2	1
6	-1	6	C2f	-----	--	---
7	-1	1	Conv	1024	2	1
8	-1	3	C2f	-----	--	---
9	-1	1	SPPF	-----	--	---

1.1. Explanation of the Backbone network

The backbone network plays an important role in extracting features of YOLOv8 architecture, which processes the given image to generate a collection of feature maps that are then used by the detection head to predict the final object detections. A succession of convolutional layers makes up the backbone that are arranged in a specific pattern to capture features at different spatial scales. In this architecture, the backbone consists of several types of modules that are repeated multiple times with varying numbers of filters, kernel sizes, and stride values [18]. Here is a brief explanation of each module:

Convolutional layer (Conv): This module performs a 2D convolution operation on the input feature map using a set of learned filters. The filters are learned through the process of training to capture specific patterns in the data. The number of filters, kernel size, and stride values can be adjusted for each convolutional layer in the architecture [21].

Cross Stage Partial Network (C2f): This module is a modified version of the ResNet block, which uses skip connections to help gradients flow through the network during training. Two convolutional layers that have an identical number of filters and kernel sizes make up the C2f module, which is then followed by a skip connection that add the input feature map to the second convolutional layer output [21]. Fig 1 shows the structure of C2f, BottleNeck and CBS respectively

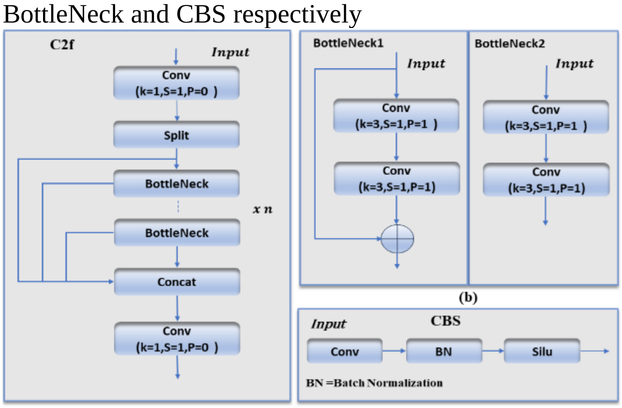

Figure 1. The Structure of (a) C2f. (b) BottleNeck (c) CBS.

Spatial Pyramid Pooling (SPPF): This module performs a pooling operation on the input feature map at multiple scales to capture features at different spatial resolutions. The pooled features are merging and passing them through a convolutional layer to generate a single feature map [20]. The backbone network is designed to capture features at different spatial scales, which are necessary for detecting objects at different sizes and distances from the camera. The C2f modules and SPPF module help improve the flow of gradients through the network and capture features at different spatial resolutions [22][23]. The SPPF architecture is depicted in figure 2.

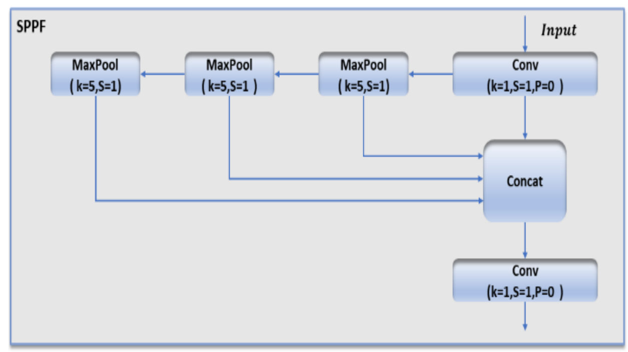

Figure 2. Architecture of SPPF [20].

1.2. Explanation of the head network

The head network consists of a series of layers that progressively refine the feature maps generated by the backbone. First, an upsampling layer is used to double the resolution of the feature maps, which helps to recover spatial information that may have been lost during the downsampling performed by the backbone. Then, the head network concatenates the upsampled feature maps with feature maps from the backbone network that have been selected to have a compatible resolution. This permits the head to incorporate both low and high-level features from input image [22]. Next, the concatenated feature maps are passed through a sequence of convolutional layers and other operations that minimize the dimensions of the feature maps while increasing their depth. This is done to prepare the feature maps for the final detection layer. Table 2 shows the head network architecture of yolov8n [20].

Table 2. Head network architecture of yolov8n [22].

No.of Layer	From	Repeat	Module type	Filter	Stride	Padding
10	-1	1	Upsample	---	---	--
11	[-1,6]	1	Concat	---	---	--
12	-1	3	C2f	---	---	--
13	-1	1	Upsample	---	---	--
14	[-1,4]	1	Concat	---	---	--
15	-1	3	C2f	---	---	--
16	-1	1	Conv	256	3x3	2
17	[-1,12]	1	Concat	---	---	--
18	-1	3	C2f	---	---	--
19	-1	1	Conv	512	3x3	2
20	[-1,9]	1	Concat	---	---	--
21	-1	3	C2f	---	---	--
	[15 ,18 ,21]	--	Detect	---	---	--

Finally, the head network passes the processed feature maps through a detection layer that predicts bounding boxes and class probabilities for the objects presented in the given image. The detection layer uses anchor boxes, which are predefined boxes of various sizes and aspect ratios, to predict the location and size of objects in the image [22][24]. The structure of Yolov8n has been displayed in fig 3 [25].

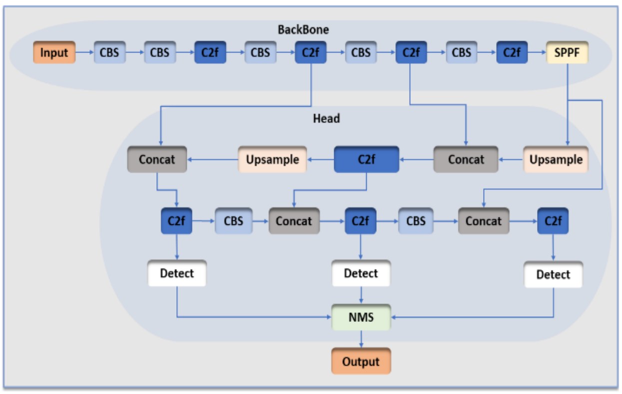

Figure 3. Yolov8n network architecture [13].

3. PROPOSED YOLOv8n

In Proposed yolov8n-v1, we built a modified architecture of yolov8n model which outline the integration of a ResNet backbone into the YOLOv8n architecture to create a hybrid model. Key changes and features include:

- ❖ We've seamlessly integrated the ResNet_Stem module and ResNet_Block modules with the YOLOv8n backbone, we remove the first two layers and replace them with ResNet_Stem and ResNet_ Block. ResNet is a form of convolutional neural network (CNN) that is known for its ability to learn deep representations from data. ResNets have showed to achieve state-of-the-art results on a variety of computer vision tasks, including object detection. The ResNet backbone consists of multiple essential components, each specifically tailored to enhance the feature representation. The ResNet_Stem module initiates the backbone with a convolutional layer then performs batch normalization, ReLU and max pooling. while the ResNet_Block contains a pair of 3x3 convolutional layers, each followed by batch normalization and ReLU activation. They also incorporate a skip connection with a 1x1 convolution to match the dimensions when necessary. Fig 4 shows the architecture of ResNet_Stem and ResNet_Block.

- ❖ Replace the SPPF modules with SPPCSP. The SPPCSP architecture combines spatial pyramid pooling (SPP) and cross stage partial networks (CSP) to create a network that is both effective and efficient. The SPP layer extract features at multiple scales from an image, and the CSP layer enables the network to learn more complex features which divide an input into multiple stages and then partially connects the stages. Figure 5 shows SPPCSP architecture.

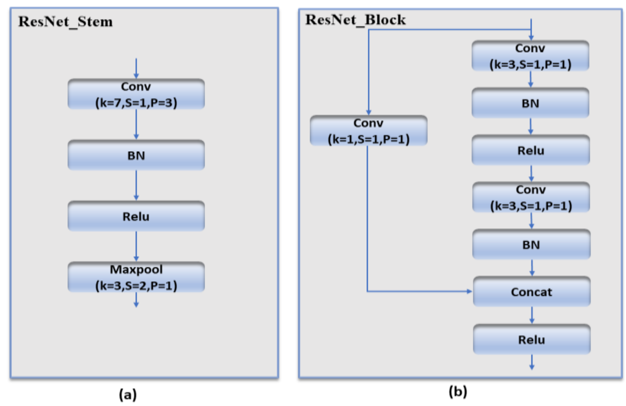

Figure 4. The architecture of (a) ResNet_Stem module (b) The ResNet_Block module.

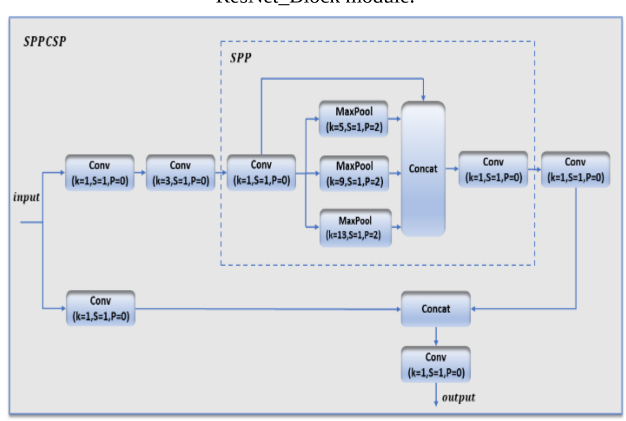

Figure 5. SPPCSP modules Architecture.

The configuration of Proposed yolov8n-v1 has been presented in figure6.

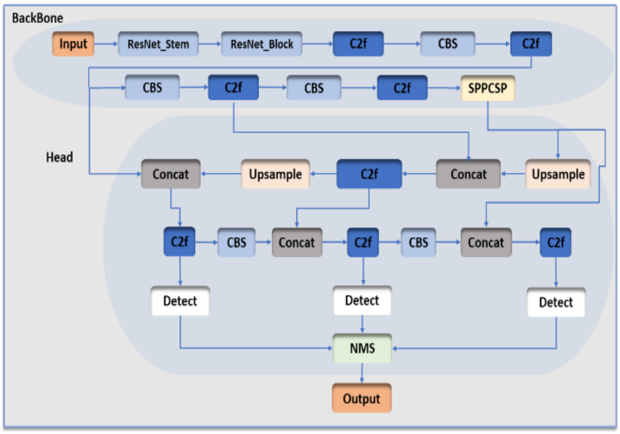

Figure 6. The Architecture of the Proposed yolov8n-v1.

The details of the proposed yolov8n-v1 backbone have been presented in table 3.

Table 3. Backbone network architecture of Proposed YOLOv8n-v1.

No.of Layer	From	Repeat	Module type	Filter	Stride	Padding
0	-1	1	RestNet_Stem	---	----	--
1	-1	3	ResNet_Block	---	----	--
2	-1	4	C2f	---	----	--
3	-1	1	Conv	256	3 x 3	2
4	-1	5	C2f	---	----	--
5	-1	1	Conv	512	3 x 3	2
6	-1	5	C2f	---	----	--
7	-1	1	Conv	1024	3 x 3	2
8	-1	5	C2f	---	----	--
9	-1	1	SPPCSP	---	----	--

In Proposed yolov8n-v2, the ResNet_Stem module, ResNet_Block modules, ResNet_downsample modules, GhostConv modules, SPPCSP modules are intergrated with the YOLOv8n model backbone to enhance the feature extraction for the model. we remove the first two layers and replace them with ResNet_Stem and ResNet_ Block. The fourth and sixth layer which is conv module is replaced with GhostConv module. The architecture of GhostConv module is shown in fig 7. The eighth and ninth layers are replaced by ResNet_Downsample and ResNet_Block. The ResNet_Downsample modules perform down sampling operations using 1x1 convolutions, batch normalization, and ReLU activation. The last layer is replaced by SPPCSP module. Table4 shows the architecture of the Backbone Proposed yolov8n-v2.

Table 4. Backbone network architecture of Proposed YOLOv8n-v2

No.of Layer	From	Repeat	Module type	Filter	Stride	Padding
0	-1	1	RestNet_Stem	---	------	--
1	-1	3	ResNet_Block	---	------	--
2	-1	3	C2f	---	------	--
3	-1	1	GhostConv	245	3 x 3	2
4	-1	6	C2f	---	------	--
5	-1	1	GhostConv	512	3 x 3	2
6	-1	6	C2f	---	------	--
7	-1	1	ResNet_Downsample	1024	3 x 3	2
8	-1	6	ResNet_Block	---	------	--
9	-1	1	SPPCSP	---	------	--

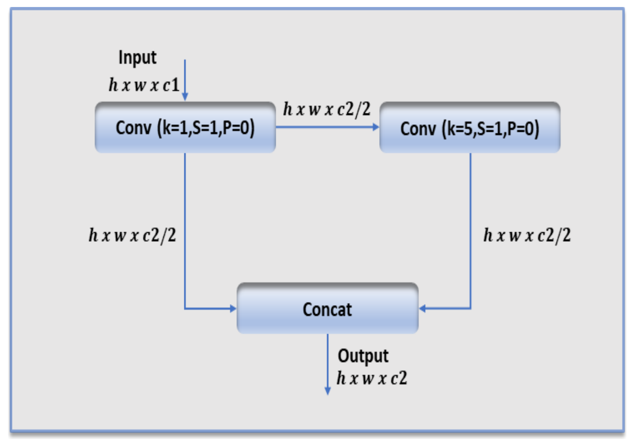

Figure 7. The Ghostconv Module Structure.

Figure 8 shows the architecture of Proposed Yolov8-v2.

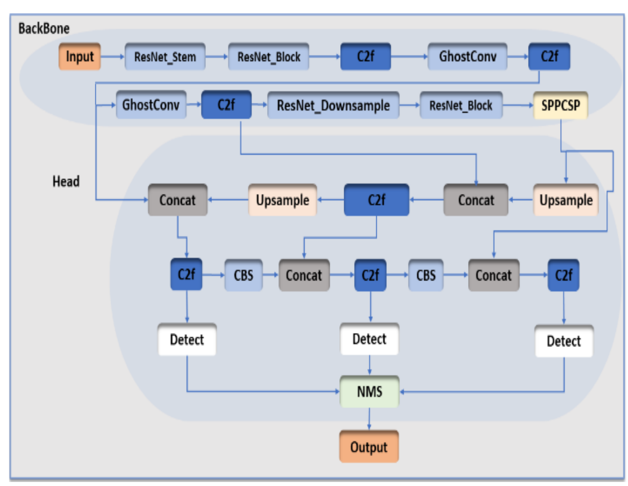

Figure 8. Architecture of the Proposed yolov8n-v2.

4. DATASETS

In this paper, two types of object detection datasets have been used namely The MJFR and MSFM dataset. The MJFR is object detection dataset collected by the authors available on Roboflow platform cloned from four different repositories on roboflow computer vision platform. The MJFR contains 23621 images in total, including 20658 images for training and 1952 for validation and 1011 images for testing. The dataset consisted of images containing individuals with and without Face masks. Special attention was given to ensure accurate ground truth annotations and check the class label for each object in dataset corresponding to ('mask' or 'no-mask'). The MSFM object detection dataset has been created from groups of images and videos in real life taken by the webcam and mobile camera based on the curriculum learning technology and annotated by the authors of this paper using yolov8n model. Curriculum learning relies on the idea of arranging data to categories or levels from easy to hard in a way that makes the model learn better and more effectively. The training of dataset is started by presenting the easiest data first and then gradually increase the complexity. The MSFM dataset contains 19601 images in total, including 14293 images for training, 4204 for validation and 1104 images for testing.

5. RESULTS AND DISCUSSIONS

Experiments were conducted by training both datasets with YOLOv8n raw model and both version of Proposed Yolov8n. To discover which one of all models are better in terms of P, R, mAP@0.5 and mAP@0.5:0.95 since these metrics define which one performs better in terms of overall detection. The metrics employed for quantitative examination of the models are described as following.

- ❖ The ratio of correctly classified positive samples (True Positives) to all positively classified samples (True Positives plus False Positives), whether they were correctly classified or not, is known as precision [26].
- ❖ The ratio of true positive to all positive samples (True Positive + False Negative) is used to compute the recall value. It gauges how well the model can locate positive samples [26]
- ❖ By comparing the detected box to the ground-truth box bounding box at an IoU threshold of 0.5, the mAP@0.5 determines a score. The higher the score, the more accurate the model's detections are [27].
- ❖ The phrase "mAP@0.5:0.95" denotes the average mAP over different thresholds, from 0.5 to 0.95, in steps of 0.05 [16][27].

To apply the system in medical clinic, two external Full HD 1080p web camera has been used with a cable of five meters length and a graphics user interfaces (GUIs) for the proposed system has been built using the qt designer application and connect it with python language using the package pyqt5.

Colaboratory by Google (also known as Google Colab) is a product of Google Research which is a runtime environment based on Jupyter notebook that enables us to train our deep learning and machine learning models on CPUs, GPUs, and TPUs. makes it the ideal tool for data analytics and deep learning enthusiasts because of computing limits of local devices [8]. When you create Your personal Colab notebooks they are kept in your Google Drive account. The training, validation, and testing results were saved on Google Drive and are accessible for further use. The platform also gives users access to Google Drive, which is crucial for importing data and saving files. Collaboratory can also be defined as a data analysis platform that integrates text, code, and code outputs into one document, enables anyone to create and run arbitrary python code over the web, and is particularly well suited to machine learning, data analysis [3]

We are training the yolov8n raw model, proposed yolov8n-v1 and proposed yolov8n-v2 in 100 epochs for MJFR and MSFM datasets. The output values of the performance results got from the validation and testing of YOLOv8n raw model on both dataset is shown in Table 5.

Validation Results of MJFR					
Classes	Images	P	R	MAP50	MAP50-95
All	1952	0.948	0.865	0.913	0.607
Mask	1952	0.975	0.979	0.989	0.676
No Mask	1952	0.921	0.752	0.836	0.538
Testing Results of MJFR					
Classes	Images	P	R	MAP50	MAP50-95
All	1011	0.896	0.86	0.909	0.529
Mask	1011	0.942	0.931	0.97	0.599
No Mask	1011	0.85	0.789	0.848	0.458
Validation Results of MSFM					
Classes	Images	P	R	MAP50	MAP50-95
All	4204	0.933	0.851	0.907	0.642
Mask	4204	0.941	0.942	0.966	0.701
No Mask	4204	0.925	0.761	0.848	0.582
Testing Results of MSFM					
Classes	Images	P	R	MAP50	MAP50-95
All	1104	0.948	0.881	0.939	0.668
Mask	1104	0.938	0.86	0.929	0.564
No Mask	1104	0.958	0.894	0.949	0.772

Table 6 shows evaluation outcomes of the proposed YOLOv8n-v1 model on MJFR and MSFM datasets.

Table 6: Evaluation Results of Proposed YOLOv8n-v1on MJFR dataset and MSFM *dataset.*

Validation Results of MJFR					
Classes	Images	P	R	MAP50	MAP50-95
All	1952	0.953	0.882	0.932	0.629
Mask	1952	0.972	0.973	0.982	0.685
No Mask	1952	0.935	0.792	0.881	0.574
Testing Results of MJFR					
Classes	Images	P	R	MAP50	MAP50-95
All	1011	0.909	0.875	0.922	0.529
Mask	1011	0.948	0.929	0.968	0.58
No Mask	1011	0.87	0.821	0.876	0.478
Validation Results of MSFM					
Classes	Images	P	R	MAP50	MAP50-95
All	4204	0.94	0.892	0.935	0.681
Mask	4204	0.939	0.961	0.972	0.724
No Mask	4204	0.941	0.823	0.898	0.638
Testing Results of MSFM					
Classes	Images	P	R	MAP50	MAP50-95
All	1104	0.943	0.911	0.95	0.689
Mask	1104	0.937	0.903	0.938	0.576
No Mask	1104	0.948	0.92	0.961	0.802

Table 7 shows evaluation outcomes of the Proposed YOLOv8nv2 model on MJFR and MSFM dataset.

Table 7. Evaluation Results of Proposed YOLOv8n-v2on MJFR and MSFM dataset

Validation Results of MJFR					
Classes	Images	P	R	MAP50	MAP50-95
All	1952	0.961	0.881	0.929	0.619
Mask	1952	0.981	0.974	0.985	0.674
No Mask	1952	0.941	0.787	0.874	0.565
Testing Results of MJFR					
Classes	Images	P	R	MAP50	MAP50-95
All	1011	0.909	0.863	0.920	0.523
Mask	1011	0.95	0.925	0.968	0.574
No Mask	1011	0.867	0.80	0.871	0.471
Validation Results of MSFM					
Classes	Images	P	R	MAP50	MAP50-95
All	4204	0.936	0.901	0.94	0.687
Mask	4204	0.94	0.962	0.975	0.727
No Mask	4204	0.933	0.841	0.905	0.647
Testing Results of MSFM					
Classes	Images	P	R	MAP50	MAP50-95
All	1104	0.943	0.918	0953	0.686
Mask	1104	0.938	0.899	0.941	0.574
No Mask	1104	0.948	0.937	0.964	0.798

The comparison analysis of validation and testing results for the original model, proposed yolov8n-v1 and proposed yolov8n-v2 can be seen in table 8.

In terms of P, as indicated in table 8, From the detection comparison in MJFR and MSFM datasets, both versions of proposed Yolov8n achieves better performance and has more true positives to total number of detected objects compared to raw model in the testing and validation of both dataset except in the testing result of MSFM, the value of P in proposed yolov8n_v1 model has 0.943 while the original model has 0.948.

In terms of R, as indicated in table 8, comparing the results of YOLOv8 raw model with both versions of proposed Yolov8n, it has been shown that both versions of proposed model outperform raw model in all cases for both validation and testing results and for both datasets.

In terms of mAP@0.5, the results of mAP@0.5 seen that both versions of proposed model have a better results and performance in term of accuracy as evidenced by the mAP50 in both datasets and for both validation and testing.

The proposed yolov8n-v1 had 0.929 for mAP50 and 0.932 for proposed yolov8n-v2 in validation results of MJFR dataset compared to the original model which have 0.913 for mAP50 by a difference of 1.6% and 1.9% for the proposed yolov8n-v1 and proposed yolov8n-v2 respectively. The mAP50 metrics for testing results had 0.920 and 0.922 of proposed yolov8n-v1 and proposed yolov8n-v2 respectively compared to original model which have 0.909 for mAP50 by a difference of 1.1 % and 1.3 % for the proposed yolov8n-v1 and proposed yolov8n-v2 respectively.

Based on the comparison of detection results, the Yolov8n model achieved 0.939 for mAP50 in MSFM testing result while the proposed Yolov8n-v1 model achieved 0.953 for mAP50 and 0.95 for mAp50 in proposed Yolov8n-v2 model by difference of 1.4% and 1.1% for v1 and v2 respectively. While the value of mAP50 in MSFM validation results have 0.94 and 0.935 for the proposed Yolov8b-v1 and Yolov8n-v2 respectively compared to the base model which have 0.907 by a difference of 3.3% and 2.8% for v1 and v2 respectively.

In terms of Map50-95, as seen in table 8, the results of mAP@0.5-0.95 shown that both versions of proposed model have a better result than original model in both datasets and for both validation and testing except in the testing result of MJFR, the value of mAP@0.5-0.95 in proposed yolov8n_v1 model has 0.523 while the original model has 0.529.

Table 8. Comparative analysis among Yolov8n model, Proposed YOLOv8n-v1 model Proposed YOLOv8n-v2 in MJFR and MSFM dataset

Validation Results of MJFR				
Classes	P	R	MAP50	MAP50-95
Original	0.948	0.865	0.913	0.607
Proposed yolov8n_v1	0.961	0.881	0.929	0.619
Proposed yolov8n_v2	0.953	0.882	0.932	0.629
Testing Results of MJFR				
Classes	P	P	MAP50	MAP50-95
Original	0.896	0.86	0.909	0.529
Proposed yolov8n_v1	0.909	0.863	0.920	0.523
Proposed yolov8n_v2	0.909	0.875	0.922	0.529
Validation Results of MSFM				
Classes	P	R	MAP50	MAP50-95
Original	0.933	0.851	0.907	0.642
Proposed yolov8n_v1	0.936	0.901	0.94	0.687
Proposed yolov8n_v2	0.94	0.892	0.935	0.681
Testing Results of MSFM				
Original	0.948	0.881	0.939	0.668
Proposed yolov8n_v1	0.943	0.918	0.953	0.686
Proposed yolov8n_v2	0.948	0.911	0.95	0.689

As summary , It can be concluded that both versions of proposed model architecture (Proposed YOLOv8n-v1 and Proposed YOLOv8n-v2) is better than the original model in accuracy that can be used in detection of face mask.

The comparison results of accuracy (mAp50) of original, proposed Yolov8n_v1 and proposed Yolov8n_v2 models are made on different sizes of image for both the validation and testing. As seen in fig 9, the proposed Yolov8n_v1 and proposed Yolov8n_v2 models outperform the original model in all image sizes used in comparison for the validation results in MJFR dataset.

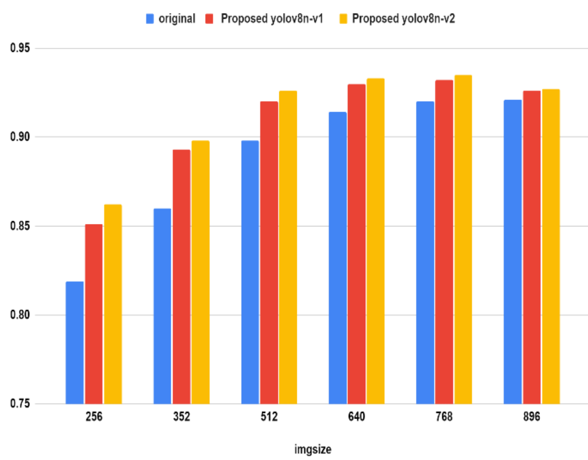

Figure 9. Comparative analysis of original Yolov8n, proposed Yolov8n_v1 and proposed Yolov8n_v2 of mAp50 for different size of images in the validation Results of MJFR Dataset.

Figure 10 shows that the proposed Yolov8n_v1 and proposed Yolov8n_v2 models outperform the original model in all image sizes used in comparison for validation results in MSFM dataset

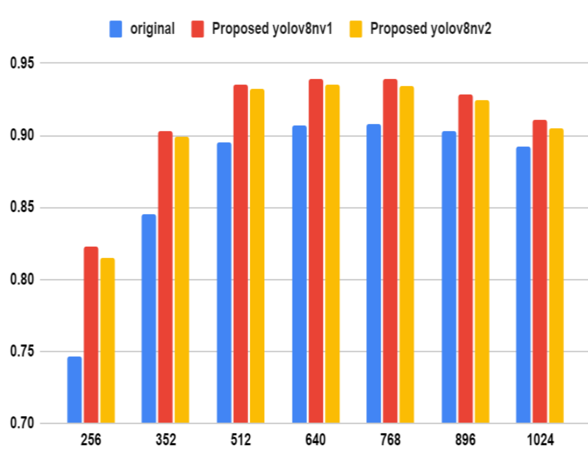

Figure 10. Comparative analysis of original Yolov8n, proposed Yolov8n_v1 and proposed Yolov8n_v2 of mAp50 for different size of images in the validation Results of MSFM dataset.

We can see some detection results in MJFR test images for the yolov8n model, proposed yolov8n-v1 and proposed yolov8n-v2 in figures 11, 12 and 13 respectively. Red box denotes the masked face; pink box denotes the non-masked face. As we see in below figures that both versions of proposed models can detect some objects in images that is not detected using original model and the confidence score for both versions of proposed model are better than original model.

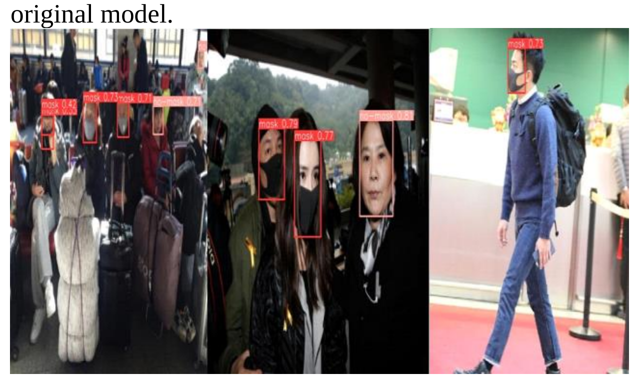

Figure 11. Images from the test dataset evaluated by original yolov8n.

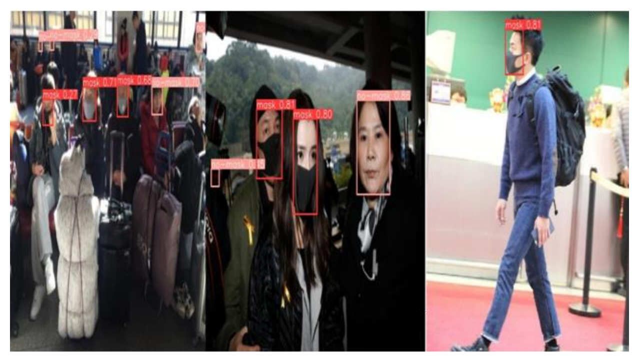

Figure 12. Images from the test dataset evaluated by Proposed yolov8n_v1.

Figure 13. Images from the test dataset evaluated by Proposed yolov8n_v2.

6. Conclusion

Based on the COVID-19 pandemic's fast spread, we must wear a face mask in our daily lives particularly in public areas to avoid transmission of this disease. The present work aims to build an intelligent system that achieves a high accuracy to detect the persons wearing a mask or not and give a sound alert to the person who is not wearing the mask across a wide range of scenarios and improve the yolov8n model for face mask.

Both versions of proposed Yolov8n have been applied on two datasets which are MJFR and MSFM that are collected and built from the authors of this paper, and it enable the models to accurately detect both masked and unmasked faces. our proposed YOLOv8n_v1 and YOLOv8n_v1 models achieved significant improvements in object detection accuracy compared to the original model on the MSFM, MJFR dataset and real time. The performance of both versions of the proposed Yolov8n model outperforms the yolov8n original model in terms of accuracy in validation and testing evaluations.

The experiments results shown that both version of proposed Yolov8n outperform the original model in both testing and validation results for mAP50 metrics which is a metric of object detection for both datasets and the proposed Yolov8n-v2 model outperform the performance of proposed Yolov8n-v1 in both testing and validation results for Map50 in MJFR datasets while in MSFM, the performance of proposed Yolov8n-v1 outperform the performance of proposed Yolov8n-v2 in both testing and validation results for Map50. It is shown that the performance of both proposed models depends on dataset.

As a future work we can extending the system's capabilities to include real-time social distancing monitoring can be valuable for enforcing physical distancing measures. By detecting and notifying instances of proximity between individuals, the system can aid in maintaining safe distancing guidelines in crowded areas.

Acknowledgements

The authors are very grateful to the college of computer science and mathematics at the university of mosul for their provision of facilities that enabled the presentation of this work in its current format.

Reference

[1]. Diwan, T., G. Anirudh, and J.V. Tembhurne, Object detection using YOLO: Challenges, architectural successors, datasets and applications. multimedia Tools and Applications, 2023. 82(6): p. 9243-9275.

[2]. Elharrouss, O., et al., Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. arXiv preprint arXiv:2206.08016, 2022.

[3]. Rajeshwari, P., et al., Object detection: an overview. Int. J. Trend Sci. Res. Dev.(IJTSRD), 2019. 3(1): p. 1663-1665.

[4]. Vibhuti, et al., Face mask detection in COVID-19: a strategic review. Multimedia Tools and Applications, 2022. 81(28): p. 40013-40042.

[5]. Nowrin, A., et al., Comprehensive review on facemask detection techniques in the context of covid-19. IEEE access, 2021. 9: p. 106839-106864.

[6]. Tian, Y., et al., Role of masks in mitigating viral spread on networks. Physical Review E, 2023. 108(1): p. 014306.

[7]. Mohammed Ali, F.A. and M.S. Al-Tamimi, Face mask detection methods and techniques: A review. International Journal of Nonlinear Analysis and Applications, 2022. 13(1): p. 3811-3823.

[8]. Ahuja, A.S., et al., Artificial intelligence in ophthalmology: A multidisciplinary approach. Integrative Medicine Research, 2022. 11(4): p. 100888.

[9]. Mbunge, E., et al., Application of deep learning and machine learning models to detect COVID-19 face masks-A review. Sustainable Operations and Computers, 2021. 2: p. 235-245.

[10]. Wani, M.A., et al., Advances in deep learning. 2020: Springer.

[11]. Kelleher, J.D., Deep learning. 2019: MIT press.

[12]. Yunus, E., YOLO V7 and Computer Vision-Based Mask-Wearing Warning System for Congested Public Areas. Journal of the Institute of Science and Technology, 2023. 13(1): p. 22-32.

[13]. Deng, J., et al., Retinaface: Single-stage dense face localisation in the wild. arXiv preprint arXiv:1905.00641, 2019.

[14]. Vrigkas, M., et al., Facemask: A new image dataset for the automated identification of people wearing masks in the wild. Sensors, 2022. 22(3): p. 896.

[15]. Kumar, A., A. Kalia, and A. Kalia, ETL-YOLO v4: A face mask detection algorithm in era of COVID-19 pandemic. Optik, 2022. 259: p. 169051.

[16]. Loey, M., et al., A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement, 2021. 167: p. 108288.

[17]. Phan, Q.B. and T. Nguyen, A Novel Approach for PV Cell Fault Detection using YOLOv8 and Particle Swarm Optimization. 2023.

[18]. Wu, W., et al., Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image. PloS one, 2021. 16(10): p. e0259283.

[19]. Chabi Adjobo, E., et al., Automatic Localization of Five Relevant Dermoscopic Structures Based on YOLOv8 for Diagnosis Improvement. Journal of Imaging, 2023. 9(7): p. 148.

[20]. Talaat, F.M. and H. ZainEldin, An improved fire detection approach based on YOLO-v8 for smart cities. Neural Computing and Applications, 2023: p. 1-16.

[21]. Terven, J. and D. Cordova-Esparza, A comprehensive review of YOLO: From YOLOv1 and beyond. arXiv 2023. arXiv preprint arXiv:2304.00501.

[22]. Ju, R.-Y. and W. Cai, Fracture Detection in Pediatric Wrist Trauma X-ray Images Using YOLOv8 Algorithm. arXiv preprint arXiv:2304.05071, 2023.

[23]. Solawetz, J. (2023). What is YOLOv8? The Ultimate Guide. https://blog.roboflow.com/whats-new-in-yolov8/

[24]. Reis, D., et al., Real-Time Flying Object Detection with YOLOv8. arXiv preprint arXiv:2305.09972, 2023.

[25]. Lou, H., et al., DC-YOLOv8: Small-Size Object Detection Algorithm Based on Camera Sensor. Electronics, 2023. 12(10): p. 2323.

[26]. Al-Shamdeen, M.J., A.N. Younis, and H.A. Younis, Metaheuristic algorithm for capital letters images recognition. Computer Science, 2020. 16(2): p. 577-588.

[27]. Padilla, R., S.L. Netto, and E.A. Da Silva. A survey on performance metrics for object-rences detection algorithms. in 2020 international conference on systems, signals and image processing (IWSSIP). 2020. IEEE.

[28]. Gunawan, T.S., et al., Development of video-based emotion recognition using deep learning with Google Colab. TELKOMNIKA (Telecommunication Computing Electronics and Control), 2020. 18(5): p. 2463-2471.

[29] Bhujel, S. and S. Shakya, Rice Leaf Diseases Classification Using Discriminative Fine Tuning and CLR on EfficientNet. Journal of Soft Computing Paradigm, 2022. 4(3): p. 172-187.

M.Sc Muna Jaffer AL_Shamdeen
e-mail: muna.jaffer@uomsul.edu.iq

She has BSc degree in Computer Science from the Iraq /university of Mosul, collage of computer sciences & mathematics, department of computer science at 2005, she received her MSc degrees in computer Science form the same university & department at 2011, she worked since 2005 until this time in same department

Research interests: digital image processing, computer vision, remote sensing, pattern recognition.

https://orcid.org/0000-0002-2806-532X

Prof: Dr fawziya ramo
e-mail: fawziyaramo@uomosul.edu.iq

Obtained BA degree in Computer Science in 1992, then obtained MA degree in Computer Architecture in 2001 and PhD in Artificial Intelligence in 2007, obtained Assistant Professor in 2013,

Research interests: computer science , artificial intelligence and machine learning

https://orcid.org/0000-0002-7510-0482