Improving Performance of Yolov5n v6.0 for Face Mask Detection

1st Muna Jaffer Al-Shamdeen
Computer Science Department
College of Computer Science and Mathematics /Mosul University
Mosul, Iraq
muna.jaffer@uomosul.edu.iq

2nd Fawziya Mahmood Ramo
Computer Science Department
College of Computer Science and Mathematics /Mosul University
Mosul, Iraq
fawziyaramo@uomosul.edu.iq

***Abstract*—The COVID-19 coronavirus pandemic has generated a global health crisis in all Worldwide. According to the World Health Organization (WHO), protection against COVID-19 infection is an essential countermeasure. one of the most effective countermeasures is wearing a facial mask which is imperative in our everyday activities, particularly in communal settings, to mitigate the transmission of the illness. In this study, we have enhanced the architectural design of YOLOv5n v6.0 for face mask detection by constructing a modified model known as Proposal YOLOv5n v6.0 model. The primary objective of this modification is to enhance the feature extraction and prediction capabilities of the YOLOv5n v6.0 model. In our proposal, we outline the integration of a residual network (ResNet) backbone into the YOLOv5n v6.0 architecture by replacing the first three layer of YOLOv5n v6.0with ResNet_Stem module and ResNet_Block module to enhance the feature extraction capabilities of the model and replace Spatial Pyramid Pooling Fast (SPPF) module in original model with Spatial Pyramid Pooling-Cross Stage Partial (SPPCSP) modules which combines SPP and CSP to create a network that is both effective and efficient. In our proposal model we have carefully curated a set of anchor configurations tailored to the specific requirements of small object mask detection. MJFR dataset was used for testing and evaluation of the original and proposed model which consist of 23,621 images and collected by the authors of this paper. The performance of both models is evaluated using the following metrics: mean average precision (mAP50), mAP50-95, recall (R) and precision (P). We conclude that the proposed model outperforms the original model in terms of accuracy for mAP50, mAP50-95 which is the metric used to measure the performance of the object detection model.**

Keywords— YOLOv5, YOLOv5n v6.0, Object Detection, Resnet, Deep Learning

I. Introduction

The emergence of global health challenges, such as the COVID-19 pandemic, has underscored the critical role of technology in safeguarding public health. The study of authors Tian and et al. provided a comprehensive quantitative analysis of the impact of mask-wearing has shown that one of the pivotal measures in curbing the spread of respiratory diseases is through the adoption of face masks as it stands as a primary line of defense. This author stated clearly that masks have remained an important mitigation strategy in the fight against COVID-19 due to their ability to prevent the transmission of respiratory droplets between individuals. Thus, humans are expected to keep using their face mask especially outdoor or in public places [1].Therefore, it is essential for automation of face mask detection system. However, it is well known that one quick option for automation is the adoption of computer systems in doing tasks that are tedious or repetitive for humans to do. Thus, the use of deep learning and computer vision is far more import for this task [2]. Object detection in computer vision refers to the task of identifying and localizing multiple objects within an image or a video frame[3]. Unlike object classification, which simply assigns a label to an entire image, object detection provides information about the spatial location of each detected object[4]. Additionally, the object detection it is commonly perceived as more complex than image classification primarily due to five distinct challenges: the presence of competing priorities, the need for high speed, handling multiple scales, dealing with limited data, and addressing class imbalance. Extensive research has been directed towards mitigating these obstacles, often leading to impressive outcomes. Nonetheless, substantial hurdles remain to be addressed in this field[5]. This task has gained significant attention in recent years due to its applications in fields such as autonomous driving, surveillance, robotics, and more. Modern approaches to object detection are primarily driven by deep learning techniques, particularly convolutional neural networks (CNNs). The advent of deep learning has revolutionized object detection by enabling algorithms to learn features directly from raw pixel[6]. It's so much important to emphasize the fact, that deep learning (DL) is only a subfield of Artificial Intelligence (AI) which has emerged as a transformative force reshaping the landscape of various industries and scientific domains [7]. Rooted in the aspiration to bestow machines with human-like cognitive abilities, AI encompasses a plethora of techniques, including machine learning (ML), neural networks, and deep learning, all united by the common goal of enabling machines to mimic intelligent behavior. AI has transitioned from a theoretical concept to a reality with profound implications for technology and society. Machine learning, a subfield of AI, has gained prominence due to its ability to enable machines to learn from data and make predictions based on what they have learned [8].

The primary contribution of this study is compilation of large dataset consist of 23,621 images of various face-mask users, image qualities and light intensity, which will help future

The authors are very grateful to the university of Mosul/ college of Computer Science and Mathematics for their provided facilities, which helped to show this work in this format.

researchers to skip the stress of data collection and improved accuracy of YOLOv5n v6.0 algorithm by constructing a proposed model of YOLOv5n v6.0 which is meticulously tailored to efficiently and accurately identify individuals wearing or not wearing face masks and outperforms the original model in terms of accuracy.

The rest of this paper is structured as follows. The lecture review is presented in section 2. The architecture of the YOLOv5n v6.0 model is presented in section 3.The proposed YOLOv5n v6.0 is covered in section 4, the results analysis is covered in section 5, and the conclusions are covered in section 6.

II. Lecture Review

In 2021, Singh and et al. used YOLOv3 and Faster R-CNN to do face mask detection on MAFA WIDER FACE dataset and random images from website which is a total of approximately 7500. They trained their model to draw bounding boxes (red or green) around the faces of people, based on whether a person is wearing a mask or not. Their study provided two major contributions which using YOLOv3 and faster R-CNN models for face mask detection and secondly a comprehensive survey on the key difficulties in face mask detection. Some of the notable challenges they face in their studies include the fact that there is diversity of camera angles in images and people used different mask types. They train for approximately 50 epoch and obtained training loss of about 0.04 and total validation loss was 0.15 for the faster R-CNN and 0.1 and 0.25 for YOLOv3 respectively. Having compared the result from both architectures, they conclude that the accuracy of faster R-CNN is better than YOLOv3, while YOLOv3 algorithm is faster for prediction [9]. In 2023, Ramadhan and et al. use explored various computer vision-based architectures for mask detection, including ResNet50, VGG11, InceptionV3, EfficientNetB4, and YOLO (You Only Look Once) so as to assess these architectures in the context of face-mask detection and determine the most effective approach. They used Masked Face-Net dataset which consist of 1000 images and found that EfficientNetB4 architecture exhibited the highest accuracy at 95.77% which was more than the accuracy of YOLOv4 which achieved an accuracy of 93.40%, InceptionV3 87.30%, YOLOv3 86.35%, ResNet50 84.41%, VGG11 84.38%, and YOLOv2 78.75%. On the other hand, they stated that ResNet50 exhibits the swiftest architecture in the model training phase, completing in 25 minutes trailed by VGG11 at 31 minutes, EfficientNetB4 at 52 minutes, YOLOv2 at 3 hours and 24 minutes, YOLOv4 at 3 hours and 45 minutes, YOLOv3 at 4 hours and 17 minutes, and InceptionV3 at 4 hours and 41 minutes. Their study indicated mainly why there is need to improve on previous YOLO algorithm version which led to the introduction of newer YOLO models like (YOLOv5, YOLOv7 and YOLOv8) which are faster and accurate[10]. In 2023, Olorunshola and et al. conducted a comparative analysis between YOLOv5 and YOLOv7, using an independent training approach on a customized Remote Weapon Station dataset which comprises 9,779 images for four classes: Persons, Handguns, Rifles, and Knives. Their result shows that YOLOv7 achieves a precision score of 52.8%, recall value of 56.4%, mAP@0.5 of 51.5%, and mAP@0.5:0.95 of 31.5% while YOLOv5 demonstrates results of 62.6% precision, 53.4% recall, 55.3% mAP@0.5, and 34.2% mAP@0.5:0.95. They concluded that YOLOv5 excels over YOLOv7 in precision and mAP@0.5:0.95, while YOLOv7 exhibits higher recall and that the YOLOv5's accuracy improves by 4.0% compared to YOLOv7 [11]. In 2023, Joodi and et al. introduces a proposed model for face mask detection consisting of two-stage, The first stage employs a Haar cascade detector to locate faces, while the second stage employs a novel CNN model built from scratch for classification. Evaluated on the MAFA dataset consist of 5902 images. Notably, the model achieves validation accuracy ranging from 97.55% to 98.43%, varying with learning rates and the features vector in the dense neural network layer. This proposed approach not only improves performance metrics like precision and recall but also contributes to advancing accurate and efficient face mask detection. Their contribution also includes the use of different values of features vector thereby influencing the performance metrics of classification, as demonstrated by the reduction of the loss between the validation and training datasets[12] .

III. YOLOv5n v6.0 Architecture

YOLO (You Only Look Once) stands as a cutting-edge, real-time object detection technique developed by "***Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi in 2015***". The model was initially trained on the COCO dataset [13].YOLO employs a singular neural network to analyze complete images. These images are partitioned into segments, and the algorithm forecasts probabilities and bounding boxes for each of these segments. YOLOv5 is a single-stage object detector whose architecture is composed of three components which include the Backbone, Neck and a Head to make dense predictions as shown in Fig1[14].

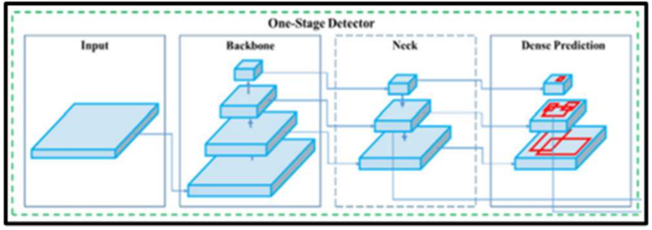


Fig. 1. Architecture of Single Stage Object Detector [14].

The backbone refers to a pre-trained network that is utilized to extract comprehensive features from images. This process involves reducing the spatial resolution of the image while increasing its feature resolution. The model neck is responsible for extracting feature pyramids, which enables the model to effectively handle objects of various sizes and scales. Lastly, the model head is responsible for executing the final stage operations. The anchor boxes on feature maps are utilized to generate the ultimate result, which includes classes, objectness scores, and bounding boxes. YOLOv5 was introduced with a range of five distinct sizes, encompassing the n variant denoting the nano size model, s representing the small size model, m indicating the medium size model, l signifying the large size model, and lastly, x denoting the extra-large size model.[15].YOLOv5 v6.0 is a more recent and stable edition. The architecture of YOLOv5n v6.0 is illustrated in fig 2[16].

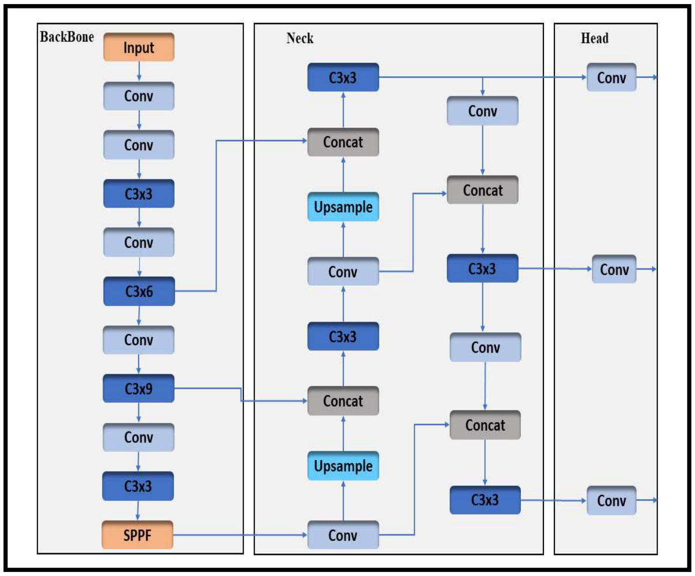


Fig. 2. The Architecture of YOLOv5n v6.0 [16]

The backbone component primarily consists of Convolutional (Conv), C3, Spatial Pyramid Pooling Fusion (SPPF) modules. The Conv module performs the output layer by applying the input features to the Conv, Batch Normalization, and activation function as shown in figure (3a). The C3 module is made up of three Convolutional and Multiple Bottleneck modules which divided the feature map of the base layer into two partitions. The first part moves via a Conv and multiple Bottleneck modules, while the second part moves just through a Conv module. After then, the two parts are joined together, and the third Conv module is then used to get the result as seen in figure (3 b). This approach aids in diminishing the quantity of parameters and therefore reduces a significant amount of computational workload, specifically in terms of floating-point operations per second (FLOPS). Consequently, this contributes to enhancing the inference speed, a critical metric for real time object detection methods.[17]. Fig 3 shows the architecture of Conv and C3 modules [18].

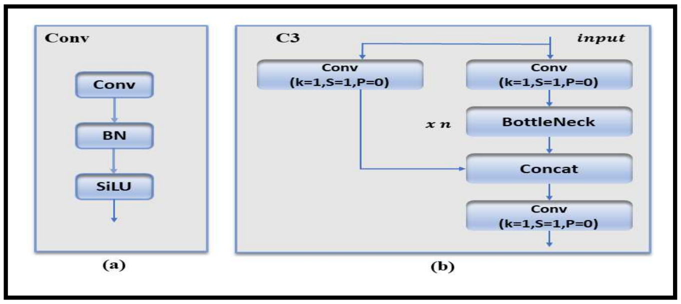


Fig. 3. The architecture of (a) The Convolutional module. (b) The C3 module[18] .

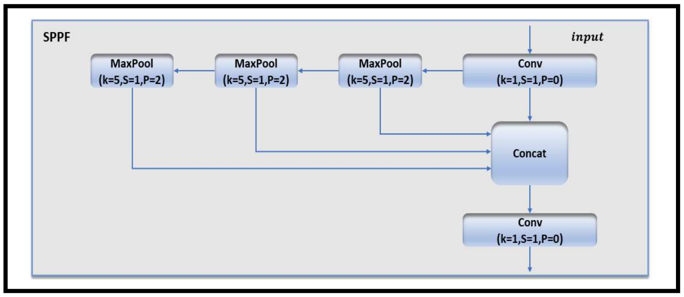
The backbone network architecture of YOLOv5n v6.0 with it details are presented in table 1.

TABLE 1. Backbone Network Architecture of YOLOv5n v6.0[16]

No.of Layer	From	Repeat	Module type	Filter	Filter Size	Stride	Padding
0	-1	1	Conv	64	6 x 6	2	2
1	-1	1	Conv	128	3 x 3	2	1
2	-1	3	C3	---	------	----	----
3	-1	1	Conv	256	3 x 3	2	1
4	-1	6	C3	---	------	----	----
5	-1	1	Conv	512	3 x 3	2	1
6	-1	9	C3	---	------	----	----
7	-1	1	Conv	1024	3 x 3	2	1
8	-1	3	C3	---	------	----	----

YOLOv5n v6.0 backbone adopts the use of SPPF to optimize network speed. The SPPF consists of two convolution modules, three maxpooling layers and the concatenation module as shown in fig 4[19]

Fig. 4 .The architecture of SPPF[19].



For the Neck section, YOLOv5 integrates the utilization of multiple of convolutional, c3, concat, upsample modules as shown in figure 2. The upsampling operation increases the spatial resolution of the feature maps by a factor of 2.and the Neck network architecture of YOLOv5n v6.0 with it details are presented in table 2.

TABLE 2. Neck Network Architecture of YOLOv5n v6.0.[16]

No.of Layer	From	Repeat	Module type	Filter	Filter Size	Stride	Padding
10	-1	1	Conv	512	1 x 1	1	0
11	-1	1	Upsample	---	---	--	--
12	[-1,6]	1	Concat	---	---	--	--
13	-1	3	C3	---	---	--	--
14	-1	1	Conv	256	1 x 1	1	0
15	-1	1	Upsample	---	---	--	--
16	[-1,4]	1	Concat	---	---	--	--
17	-1	3	C3	---	---	--	--
18	-1	1	Conv	256	3x3	2	1
19	[-1,14]	1	Concat	---	---	--	--
20	-1	3	C3	---	---	--	--
21	-1	1	Conv	512	3x3	2	1
22	[-1,10]	1	Concat	---	---	--	--
23	-1	3	C3	---	---	--	--
	[17 ,20 ,23]	--	Detect	---	---	--	--

The head segment of YOLOv5 pick up from where the neck stops, the output of the neck serves as the input to the head part In a more succinct manner, we can say that the head of the YOLOv5 architecture plays a crucial role in accurately predicting object locations and classes. It collaborates with the backbone to process feature maps extracted from the input image. C3 layers enhance these feature maps by integrating context information, aiding in capturing complex features. Convolutional layers, followed by upsampling, extract and refine features of different scales. The concatenated feature maps from various paths create a feature pyramid, enabling the network to detect objects of varying sizes. The Detect layer integrates predictions from multiple scales, adjusting anchor box parameters to fit object boundaries accurately. Non-Maximum Suppression removes duplicate predictions. This coordinated approach combines multi-scale features, anchor boxes, and efficient processing to achieve accurate real-time object detection[20, 21].

IV. PROPOSAL YOLOV5N V6.0

In our proposal, we outline the integration of a ResNet backbone into the YOLOv5 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 YOLOv5 backbone, we remove the first three layers and replace them with ResNet_Stem stem and ResNet_ Block.
- Replace the SPPF modules with Spatial Pyramid Pooling-Cross Stage Partial Networks (SPPCSP).
- In our enhanced YOLOv5n v6.0 implementation, we have carefully curated a set of anchor configurations tailored to the specific requirements of small object mask detection. These anchor configurations have been designed to improve object localization and enhance the model's ability to detect small masks accurately. Below are the anchor configurations for our proposed model compared to the original YOLOv5n v6.0 anchors.

  [10, 13, 16, 30, 33, 23] (unchanged from the original)
  [20, 35, 40, 65, 60, 100]
  [80, 120, 160, 200, 240, 320].

ResNet is a type of convolutional neural network (CNN) that is known for its ability to learn deep representations from data. ResNets have been shown 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.

* ResNet_Stem: This module initiates the backbone with a convolutional layer followed by batch normalization, ReLU activation, and max-pooling. It increases the number of filters and reduces the resolution by a factor of 2.

* ResNet_Downsample: These modules perform down sampling operations using 1x1 convolutions, batch normalization, and ReLU activation. They also increase the number of filters and reduce the resolution by a factor of 2.

* ResNet_Block: These blocks contain 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. The following figure show the architecture of Proposed YOLOv5n v6.0. In addition to the previously mentioned modification in YOLOv5n v6.0 architecture, also the number of layers and concation between modules is different from the original model as shown in following figure compare with the fig 2.

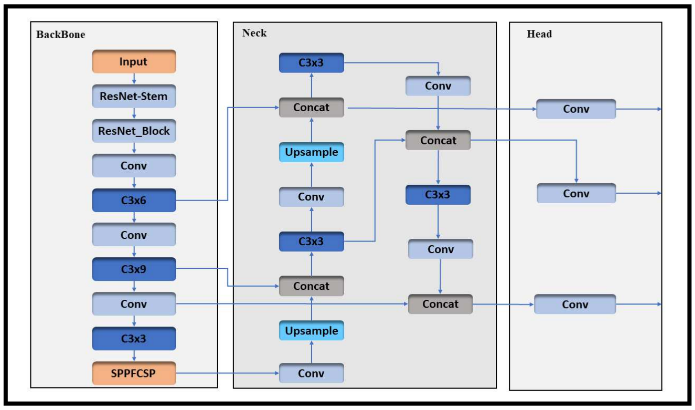


Fig. 5. The Architecture of the Proposal YOLOv5n v6.0.

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 extracts features at multiple scales from an image, and the CSP layer enables the network to learn more complex features which divides an input into multiple stages and then partially connects the stages. The following figure shows SPPCSP architecture.

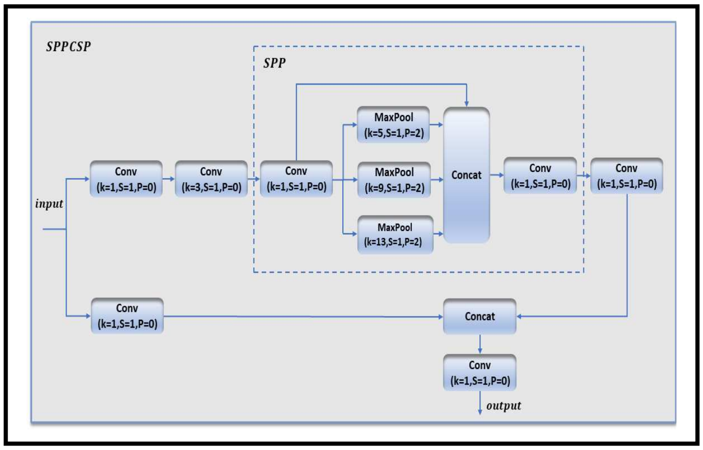


Fig. 6. Architecture of SPPCSP modules**.**

The tabel 3 and 4 shows the backbone network architecture of proposal YOLOv5n v6.0 with it details and Neck Network architecture of proposal YOLOv5n v6.0 with it details respectlively .

TABLE 3. Backbone network architecture of proposal YOLOv5n v6.0

No.of Layer	From	Repeat	Module type	Filter	Filter Size	Stride	Padding
0	-1	1	RestNet_Stem	---	----	--	--
1	-1	3	ResNet_Block	---	----	--	--
2	-1	1	Conv	256	3 x 3	2	1
3	-1	6	C3	---	----	--	--
4	-1	1	Conv	512	3 x 3	2	1
5	-1	9	C3	---	----	--	--
6	-1	1	Conv	1024	3 x 3	2	1
7	-1	3	C3	---	----	--	--
8	-1	1	SPPCSP	---	----	--	--

TABLE 4. Neck network architecture of proposal YOLOv5n v6.0

No.of Layer	From	Repeat	Module type	Filter	Filter Size	Stride	Padding
9	-1	1	Conv	1024	1 x 1	1	0
10	-1	1	Upsample	---	---	--	---
11	[-1,5]	1	Concat	---	---	--	---
12	-1	3	C3	---	---	--	---
13	-1	1	Conv	256	1 x 1	1	0
14	-1	1	Upsample	---	---	--	---
15	[-1,3]	1	Concat	---	---	--	---
16	-1	3	C3	---	---	--	---
17	-1	1	Conv	256	3x3	2	1
18	[-1,12]	1	Concat	---	---	--	---
19	-1	3	C3	---	---	--	---
20	-1	1	Conv	512	3x3	2	1
21	[-1,6]	1	Concat	---	---	--	---
	[15,18,21]	--	Detect	---	---	--	---

V. Results and Discussion

The evaluation of the performance of the original YOLOv5n and Proposal YOLOv5n v6.0 is conducted using the following measures.

- Mean average precision (mAP): The area under the precision-recall curve for a given class is called average precision (AP) which measures the model's performance in detecting a specific class. The average of the average precision for all classes is called mAP which measures the overall performance of the model and indicate the accuracy of the model[22].
- Precision (P): The proportion of predicted positive objects that are actually positive.
- Recall (R): The proportion of actual positive objects that are predicted positive[23]. The MJFR dataset were cloned from various repositories on Roboflow computer vision platform. The total number of images are 23,621 with 20,658 of the images used as train set, 1,952 of images used as validation set and the remaining used as test set.

Table 5 presents the evaluation outcomes of the YOLOv5n v6.0 original model on the MJFR dataset. This dataset comprises 23,621 images that were collected by the authors of this research study from four distinct datasets, all of which were used for facemask detection purposes.

TABLE 5. Evaluation Results of Original YOLOv5n v6.0 on MJFR dataset

Validation Results					
Classes	**Images**	**Precision**	**Recall**	**MAP50**	**MAP50-95**
All	1952	0.945	0.862	0.905	0.565
Mask	1952	0.977	0.971	0.975	0.626
No Mask	1952	0.913	0.752	0.834	0.503
Testing Results					
Classes	**Images**	**Precision**	**Recall**	**MAP50**	**MAP50-95**
All	1011	0.879	0.862	0.901	0.497
Mask	1011	0.928	0.924	0.958	0.549
No Mask	1011	0.829	0.8	0.844	0.446

Table 6 shows the evaluation results of a proposed YOLOv5n v6.0 model on MJFR dataset.

TABLE 6. Evaluation Results of Proposed YOLOv5n v6.0 on MJFR dataset

Validation Results					
Classes	**Images**	**Precision**	**Recall**	**MAP50**	**MAP50-95**
All	1952	0.943	0.879	0.916	0.59
Mask	1952	0.985	0.977	0.985	0.664
No Mask	1952	0.902	0.78	0.848	0.517
Testing Results					
Classes	**Images**	**Precision**	**Recall**	**MAP50**	**MAP50-95**
All	1011	0.899	0.863	0.907	0.5
Mask	1011	0.939	0.924	0.962	0.554
No Mask	1011	0.859	0.802	0.852	0.445

The comparison analysis of validation and testing results for both the original and proposed YOLOv5nv6.0 can be seen in table 7. The proposal YOLOv5n v6.0 performed better in terms of accuracy, as evidenced by the mAP50 and mAP50-95 metrics for all classes for both validation and testing results as show in Tabel 7, the proposal model had 0.916 for mAP50 and 0.59 for mAP50-95 in validation results compared to the original model which have 0.905 and 0.565 for mAP50 and mAP50-95 respectively by a difference of 1.1% and 2.5% .The testing results of proposal model had 0.901 for mAP50 and 0.497 for mAP50-95 compared to original model which have 0.901 and 0.497 for mAP50 and mAP50-95 respectively by a difference of 0.6% and 0.3%

TABLE 7. Comparative analysis between proposed model and original model

Validation Results						
Model	**P**	**R**	**MAP 50**	**MAP 50-95**	**No. of Par-meter (M)**	**GFLOP S**
Original	0.945	0.862	0.905	0.565	1.76	4.1
Proposed	0.943	0.879	0.916	0.59	3.42	19.0
Testing Results						
Model	**P**	**R**	**MAP 50**	**MAP 50-95**	**No. of Par-meter (M)**	**GFLOP S**
Original	0.879	0.862	0.901	0.497	1.76	4.1
Proposed	0.899	0.863	0.907	0.5	3.42	19.0

It can be concluded that the proposed model architecture is better than the original model in terms of accuracy that can be used in detection of face mask. The comparison results of accuracy (mAp50) of original and proposed YOLOv5n v6.0 models are made on different size of image (256 ,348 ,512 ,640, 896 ,1024 ,1152, 1280 ,1408 and 1536). As seen in figure 8, the proposed models outperform the original model in all image sizes used in comparison.

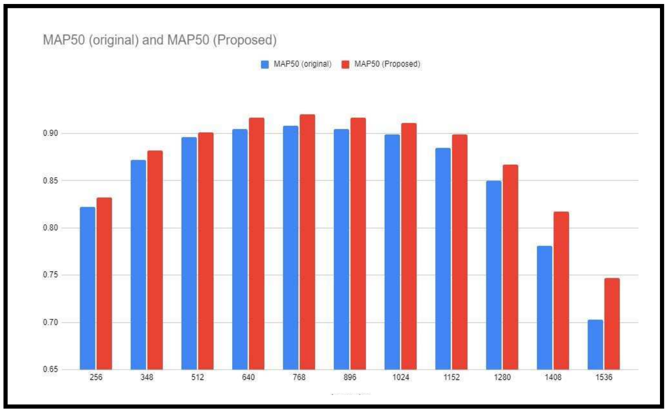


Fig. 7. Comparative analysis of original and Proposed YOLOv5n v6.0 of mAp50 for different size of images

VI. Conclusion

During the convid-19 pandemic the detection of face masks has emerged as a crucial undertaking for security organizations across many establishments, including buildings, government offices, and other public spaces. This study aims to improve the YOLOv5n v6.0 model for face mask detection. The proposed modifications enable the model to accurately detect both masked and unmasked faces. The performance of the proposed YOLOv5n v6.0 model on MJFR dataset outperforms the YOLOv5n v6.0 raw model in terms of accuracy on both validation and testing evaluations. Based on the comparison of detection results, the YOLOv5n v6.0 model achieved 0.905 and 0.566 for mAP50 and mAP50-95 respectively in validation result while the proposed model achieved 0.916 and 0.59 for mAP50 and mAP50-95 respectively. in the testing result the original model achieved 0.901 for mAP50 and 0.49 for mAP50-95 while the proposed model achieved 0.907 and 0.50 for mAP50 and mAp50-95 respectively .it can be concluded that proposed YOLOv5n v6.0 outperform the original model in both testing and validation result and for different size of image

References

[1] Tian, Y., Sridhar, A., Wu, C. W., Levin, S. A., Carley, K. M., Poor, & . et al., The Role of Masks in Mitigating Viral Spread on Networks (preprint). 2021.

[2] Goyal, H., Sidana, K., Singh, C., Jain, A., & Jindal, S., A real time face mask detection system using convolutional neural network. Multimedia Tools and Applications, 2022. 81(11): p. 14999-15015. https://doi.org/10.1007/s11042-022-12166-x.

[3] Xiao, C., Q. Shi, and H. Zhang, 34. Artificial intelligence in project organizing. Research Handbook on Complex Project Organizing, 2023: p. 344..

[4] Zhao, L., T. Tohti, and A. Hamdulla, BDC-YOLOv5: a helmet detection model employs improved YOLOv5. Signal, Image and Video Processing, 2023: p. 1-11.

[5] Fessel, K., Significant Object Detection Challenges and Solutions. Medium. https://towardsdatascience. com/5-significant-object-detection.

[6] Feng, X., Jiang, Y., Yang, X., Du, M., & Li, X,Computer vision algorithms and hardware implementations: A survey. Integration, 2019. 69: p. 309-320. https://doi.org/10.1016/j.vlsi.2019.07.005.

[7] Ahuja, A. S., Wagner, I. V., Dorairaj, S., Checo, L., & Ten Hulzen, R, Artificial intelligence in ophthalmology: A multidisciplinary approach. Integrative Medicine Research, 2022. 11(4): p. 100888. https://doi.org/10.1016/j.imr.2022.100888.

[8] Salehi, H. and R. Burgueño, Emerging artificial intelligence methods in structural engineering. Engineering structures, 2018. 171: p. 170-189. https://doi.org/10.1016/j.engstruct.2018.05.084

[9] Singh, S., Ahuja, U., Kumar, M., Kumar, K., & Sachdeva, M., Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. Multimedia Tools and Applications, 2021. 80: p. 19753-19768.

[10] Ramadhan, M. V., Muchtar, K., Nurdin, Y., Oktiana, M., Fitria, M., Maulina, N. & et al., Comparative analysis of deep learning models for detecting face mask. Procedia Computer Science, 2023. 216: p. 48-56. https://doi.org/10.1016/j.procs.2022.12.110 .

[11] Olorunshola, O.E., M.E. Irhebhude, and A.E. Evwiekpaefe, A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms. Journal of Computing and Social Informatics, 2023. 2(1): p. 1-12. https://doi.org/10.33736/jcsi.5070.2023.

[12] Joodi, M.A., M.H. Saleh, and D.J. Kadhim, Increasing validation accuracy of a face mask detection by new deep learning model-based classification. Indonesian Journal of Electrical Engineering and Computer Science, 2023. 29(1): p. 304-314. DOI: 10.11591/ijeecs.v29.i1.pp304-314

[13] Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. A Review of Yolo algorithm developments. Procedia Computer Science, 2022. 199: p. 1066-1073. https://doi.org/10.1016/j.procs.2022.01.135.

[14] Bochkovskiy, A., C.-Y. Wang, and H.-Y.M. Liao, Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020. https://doi.org/10.48550/arXiv.2004.10934.

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

[16] Jocher, G., Stoken, A., Chaurasia, A., Borovec, J., Kwon, Y., Michael, K., ... & Thanh Minh, M. ultralytics/yolov5: v6. 0-YOLOv5n'Nano'models, Roboflow integration, TensorFlow export, OpenCV DNN support. Zenodo, 2021.

[17] Chen, H., Liu, H., Sun, T., Lou, H., Duan, X., Bi, L., & et al., MC-YOLOv5: A Multi-Class Small Object Detection Algorithm. Biomimetics, 2023. 8(4): p. 342. https://doi.org/10.3390/biomimetics8040342

[18] Li, Z., Tian, X., Liu, X., Liu, Y., & Shi, X., A two-stage industrial defect detection framework based on improved-yolov5 and optimized-inception-resnetv2 models. Applied Sciences, 2022. 12(2): p. 834. https://doi.org/10.3390/app12020834.

[19] Bai, J., Dai, J., Wang, Z., & Yang, S., . A detection method of the rescue targets in the marine casualty based on improved YOLOv5s. Frontiers in Neurorobotics, 2022. 16: p. 1053124.

[20] Wang, J., Yang, P., Liu, Y., Shang, D., Hui, X., Song, J., & Chen, X. Research on Improved YOLOv5 for Low-Light Environment Object Detection. Electronics, 2023. 12(14): p. 3089. https://doi.org/10.3390/electronics12143089

[21] Li, Z. Road aerial object detection based on improved YOLOv5. in Journal of Physics: Conference Series. 2022. IOP Publishing. doi:10.1088/1742-6596/2171/1/012039.

[22] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L . Microsoft COCO: common objects in context. Computer Vision–ECCV 2014: 13th European Conference. 2014, Springer International Publishing.

[23] 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.