Facemask detection and classification using you only look once version 7

Gareebah Al-Rasheedi, Rehan Ullah Khan
Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

Article Info

Article history:

Received Jun 13, 2023
Revised Oct 8, 2023
Accepted Feb 10, 2024

Keywords:

COVID-19
Deep learning
Face recognition
Image classification
Machine learning
Object detection

ABSTRACT

World Health Organization (WHO) suggests that wearing masks and keeping social distancing are the best ways to avoid infection transmission of communicable diseases. Consequently, most governments have forced people to wear masks in public areas to prevent communicable diseases such as COVID-19. Manual monitoring and surveillance are time-consuming and not always possible in crowded areas. Hence, object detection deep learning models can effectively handle these challenges. Therefore, this work aims to investigate the efficiency of different versions of the you only look once version 7 (YOLOv7) model in facemask detection and classification over the privately balanced dataset. The dataset comprises of 1,300 images with four novel classes; including no occlusion, correct mask, incorrect mask, and other use cases. Furthermore, the model's performance was evaluated based on mean average precision (mAP), recall, precision, and inference time. Finally, a comparative result analysis has been reported to determine the best model for facemask detection and classification. YOLOv7 model versions exhibit widely various performances ranging from 20.7% mAP for YOLOv7-D6 to 95.5% for YOLOv7-tiny. In contrast, the inference time for all YOLOv7 versions covers a narrow range of 3 ms. In conclusion, the YOLOv7-tiny version outperforms other models, achieving a high detection performance and acceptable detection speed.

Corresponding Author:

Gareebah Al-Rasheedi
Department of Information Technology, College of Computer, Qassim University
Buraydah, Saudi Arabia
Email: grob083@hotmail.com

1. INTRODUCTION

The communicable diseases spread rapidly, such as the recent COVID-19. These diseases pose severe health and economic threats to individuals [1]. According to the World Health Organization (WHO), wearing face masks and practicing social distancing are the ideal ways to protect people and prevent COVID-19 virus transmission [2], [3]. However, the traditional manual monitoring of whether people follow the facemask-wearing rules is a challenging time-consuming task. Consequently, the need for surveillance systems that rapidly detect and automatically classify facemasks is vital in fighting of the COVID-19 epidemic. The automatic detection and classification challenges can be handled by integrating machine learning techniques and surveillance systems.

Deep learning (DL) has been introduced as a powerful subfield of machine learning that enables machines to learn and make decisions from unstructured data. DL models have achieved remarkable breakthroughs in number of autonomous systems [4]. Moreover, DL models enable rapid and accurate analysis of massive amounts of information. Computer vision is an artificial intelligence (AI) subfield that allows machines to understand and analyze visual data, including images or videos [5]. Object detection, image classification and segmentation are the main computer vision tasks [6]. Object detection task involves locating and identifying objects of interest within an image or a video [7]. Consequently, facemask detection and classification can be treated as an object detection problem.

Facemask detection algorithms are based on different approaches classified into two main categories: traditional techniques and DL-based ones [8]. Traditional techniques include Viola jones detector, histogram of oriented gradients (HoG-based) detector, scale-invariant feature transform (SIFT) detectors, and deformable parts model (DPM). On the other hand, DL techniques based on convolutional neural network (CNN) include two-stage detectors such as region with convolutional neural network (R-CNN), Fast R-CNN, Faster R-CNN, and feature pyramid network (FPN), and one-stage detectors such as you only look once (YOLO), single shot detector (SSD), and RetinaNet [9]. In recent years, object detection algorithms using DL models have become more efficient than conventional models in addressing complicated jobs.

Many researchers introduced various approaches for facemask detection and classification based on DL algorithms. Zhang *et al*. [10] introduced the R-CNN model for a facemask detection task the context-attention R-CNN model is trained and evaluated on the dataset with 4,672 images of three labels without the mask, mask correct, and mask incorrectly. The result shows that the proposed model achieves a mean average precision (mAP) of 84.1%. Loey *et al.* [11] proposed a novel model for medical mask detection using YOLOv2 with ResNet50. The dataset was collected from two datasets which are the medical mask dataset (MMD) and the face mask dataset (FMD) [12] with a total of 1,415 images consisting of two labels mask and no mask. The proposed model achieved an average precision of 81%. According to Jiang *et al.* [13], a real-time mask detection model with a modification on YOLOv3 was trained using a dataset named properly wearing masked face detection (PWMFD) [14]; that is collected from different datasets and sources such as masked faces (MAFA) [15], WIDER [16], real-world masked face dataset (RMFD) [17], and the internet. The result showed that the proposed model achieves an average precision of 73.3%. Bhuiyan *et al.* [18] trained the YOLOv3 model on a dataset containing 600 images collected from Google and consists of two labels mask and no mask. The proposed model achieves a mAP is 96%. While the Mokeddem *et al.* [19] used YOLOv4 for real-time facemask detection and classification. The YOLOv4 model was trained using a dataset of 14,409 images that belonging to three labels: masked faces, incorrectly masked and unmasked faces. The result shows that the proposed YOLOv4 model achieved a mAP of 88.82%. Yu and Zhang [20] introduce an improved facemask detection and classification method based on the YOLOv4. The model was trained using dataset of 10,855 images that were collected from two publicly available datasets RMFD [17] and MasedFace-Net [21], [22] that include three labels no mask is worn, face wears a mask, wearing masks irregularly. The result shows that the proposed work achieves a mean average of 98.3%. Kumar *et al.* [23] introduced an enhancement method for tiny YOLOv4-spatial pyramid pooling (SPP) to obtain higher detection accuracy. The first dataset is collected from Google and Bing application programming interface (API) with a total of 52,635 images belonging to four labels, including with mask, without mask, mask worn incorrectly, and mask area. Then the proposed model is trained on a collected dataset and evaluated on the MOXA dataset [24]. The result showed that Tiny YOLOv4-SPP model achieves an average precision of 84.42%. However, Nagrath *et al.* [25] using single shot multibox detector and mobilenetV2 (SSDMNV2) for facemask detection. The model was trained using dataset of 11,042 images consists of two labels with mask and without mask. The result shows the proposed work achieves an accuracy of 92.64%. On the other hand, Roy *et al*. [26] evaluated the performance of various models for medical mask detection. First, the Moxa3K dataset is collected from various sources and focuses on medical masks with a total of 3,000 images consisting of labels mask and no mask [24]. Then, different algorithms were trained on the dataset, including YOLOv3-YOLOv3 Tiny-Faster RCNN and SSD. As a result, YOLOv3Tiny outperforms the other models achieving 56.27% mAP and frames per second (FPS) of 138 which make it suitable for real-time application.

According to previous studies, improvements can still be made to improve the accuracy models using DL models. Besides, many datasets are available for facemask detection and classification, but some have limited images and no balance between classes, especially in incorrect mask cases. In addition, no studies have been done comparing the YOLOv7 model. Therefore, this study focused on facemask detection and classification using different versions of YOLOv7. Compared to the previous studies, this study used a custom dataset that includes no occlusion, correct mask, incorrect mask, and other occlusions with a balance in the number of objects in each class.

2. PROPOSED APPROACH

This work aims to develop the automatic facemask detection task using the YOLOv7 model and privately balanced dataset. Therefore, the paper introduces a comprehensive comparative study of the YOLOv7 model versions performance in facemask detection tasks. the model's performance was evaluated based on mAP, recall, precision, and inference time. Finally, a comparative study has been reported to determine the best model for facemask detection and classification. The methodology employed in this work consists of four main steps; DL model determination, dataset creation, training models on the dataset, and finally the models' performance evaluation as shown in Figure 1.

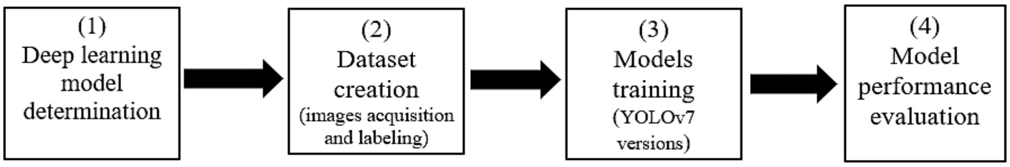

Figure 1. Methodology main steps

2.1. YOLOv7 model

YOLOv7 model is one of the fastest and most accurate models for computer vision tasks. YOLOv7 model architecture is based on previous YOLOv4, scaled YOLOv4, and YOLO-R model architectures. The authors of YOLOv7 introduced a couple of architectural changes and a group of bag-of-freebies to enhance the model accuracy. Extended efficient layer aggregation network (E-ELAN) is the computational block in YOLOv7 architecture. Furthermore, YOLOv7 introduces a new scaling technique where scaling of depth and width preserves optimal model structure. Apart from the YOLO models, the YOLOv7 model has two head types, a lead head responsible for final output and an auxiliary head for middle layers training. Additionally, YOLOv7 removes identity from re-parameterized convolutions (RepConv). YOLOv7 architecture is discussed in details in the YOLOv7 paper [27]. In this work, YOLOv7 different versions are trained on our dataset, and their performance is evaluated.

2.2. Dataset

The finalized dataset contains 1,300 images and 1,853 faces. Our collected dataset consisted of 500 images from the PWMFD dataset [14] and 800 images from the internet. Table 1 provides an overview of the dataset composition. The images were classified into four categories; no occlusion, other occlusion, correct mask and incorrect mask. The no occlusion label indicates that the mask area is not covered by any occlusions. Moreover, the other occlusion label means that the mask area is covered by any occlusions rather than facemask; such as a hand, scarf or handkerchief. The correct mask label means that a mask covers the mask area, and the location mask is correct. Finally, incorrect mask label indicates that mask covers the mask area, but not in proper way. The images were labeled using LabelImg annotation tool then saved annotations in YOLO format [28]. A sample of images with all classes in our dataset are shown in Figure 2. The dataset is arranged into 80% train, 10% test, and 10% validation arrangements.

Table 1. Summary of sources images in our dataset

Source	Total images	Total faces	Description
PWMFD [14]	500	670	This dataset comprises three categories: without mask, with mask, and incorrect mask. The images include both individual persons and multiple persons. The images from categories without masks, with masks, and incorrect masks were used for this project. The dataset is taken from the Ethancvaa GitHub account.
Internet	800	1,183	These images were taken from the internet. The images include both individual persons and multiple persons. The images from categories no occlusion, correct mask, incorrect mask, and other occlusions were used for this project.

2.3. Model training

In the training process, all the seven versions of YOLOv7 model are trained in Google Colab Pro using NVIDIA A100-SXM4-40 GB GPU. All models were trained using stochastic gradient descent (SGD) optimizer for 50 epochs. Furthermore, models were trained with the hyperparameters set as input size to 640×640, batch size to 16, learning rate to 0.01, object confidence threshold for detection is 0.5, and intersection over union (IoU) threshold for NMS is 0.6. The same dataset contained 1,040 images was used through training all models.

Figure 2. Sample of images in our dataset

2.4. Model performance evaluation

Based on IoU, the model will learn and evaluate itself during training [29]. IoU threshold was set before training to determine true true positive (TP), false positive (FP), and false negative (FN). The precision and recall parameters calculations are based on numbers of TP, FP, and FN for each labeled class [30], [31]. Precision is calculated by an in (1). Recall is calculated by an in (2).

$$P = \frac{TP}{TP+FP} = \frac{TP}{all\ detections} \tag{1}$$

$$R = \frac{TP}{TP+FN} = \frac{TP}{all\ ground\ truths} \tag{2}$$

The precision determines the model's reliability in identifying relevant objects but recall measures the model's ability to detect all ground truth. Due to recall-precision trad-off, the average precision average precision (AP) metric was developed. This metric measures the precision averaged across all recall values at each threshold. The range of AP is between 0 to 1 [32]. The mAP is calculated by taking the average of AP across all classes [30] and is given by in (3).

$$mAP = \frac{1}{k}\sum_{i}^{k} APi \tag{3}$$

In addition to assessing the mAP of each model, the evaluation of object detection models also considered detection speed as a critical criterion. The inference time is the best performance metric for detection speed evaluation. The superior-performing model was identified by analyzing the speed and accuracy of each model.

3. RESULTS AND DISCUSSION

The YOLOv7 versions models were tested on 130 images from testing dataset, and the results are displayed in Table 2. Among the all-tested models, the highest precision value was achieved by YOLOv7-tiny version at 96.8% indicating the fewest false positives predictions. Furthermore, The YOLOv7-tiny version also has the highest recall value at 95.3%. A higher recall value indicates that the YOLOv7-tiny model has a higher ability in detecting and identifying objects in the image. Based on mAP@50, YOLOv7-tiny outperforms the other models with mAP@50 at 95.5 % (with all classes). On the other hand, YOLOv7 version achieved acceptable mAP@50 at 92.7 % (with all classes). From detections results, the top three model for facemask detection are YOLOv7-tiny followed by YOLOv7 then YOLOv7-X.

Detection speed is a vital parameter especially in real time applications. the inference time for all tested models were investigated and results are depicted in Figure 3. The YOLOv7-W6 model is the fastest model in detection achieving 5.8 ms as inference time. The YOLOv7-W6 model has poor detection performance so it can be neglected in our comparative study. Consequently, the YOLOv7 model has the smallest inference time followed by YOLOv7-X and then YOLOv7-tiny. Figure 2 shows that the difference in inference time between the top 3 models (YOLOv7, YOLOv7-X, and YOLOv7-tiny) is just 1 ms. Although YOLOv7-tiny model has the fewest trainable parameters, it has inference time larger than YOLOv7 and YOLOv7-X.

Table 1. Summary of trained different versions of YOLOv7 performance

Used-model Class	mAP (50 %) all	mAP (50-95 %) all	P all	R all	Inference time(ms) all
YOLOv7-tiny	95.5	64.7	96.8	95.3	7.1
YOLOv7	92.7	59.7	95.4	92.5	6.1
YOLOv7-X	89.3	58.6	93.9	89.4	6.6
YOLOv7-W6	53.3	34.3	92.8	53.7	5.8
YOLOv7-E6	62.2	38.8	88	63.7	7.3
YOLOv7-D6	20.7	13.8	73	20.5	8.2
YOLOv7-E6E	81.3	53.5	93.5	81.3	8.9

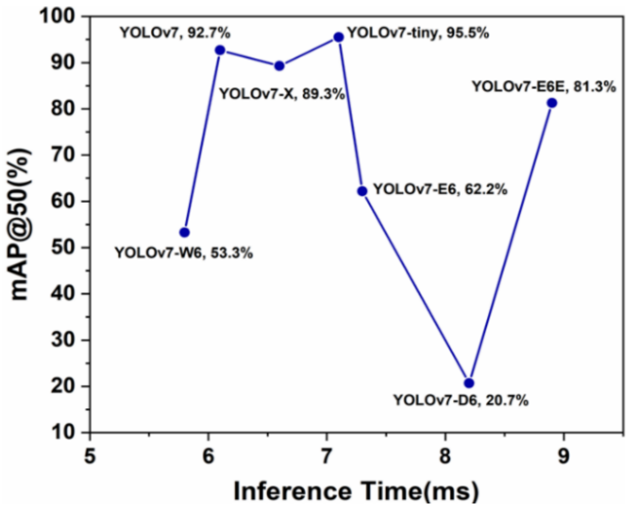

Figure 3. Inference time and mAP@50(%) for models

Based on the detection performance and speed, the top three models are YOLOv7-tiny, YOLOv7, and YOLOv7-X. Thus, the confusion matrixes for the top three were investigated. The confusion matrix for YOLOv7-tiny shown in Figure 4 indicates that the model can detect all people with other occlusion correctly. Furthermore, the YOLOv7-tiny model misclassifies about 4% of people with incorrect masks and 2% of people with no occlusion as ones with correct masks. Although the YOLOv7 model can catch all people with incorrect masks, it couldn't detect about 7% of people with no occlusion as shown in Figure 5. Additionally, YOLOv7 has 12% misclassification for people with other occlusion, but only 3% of these errors will attract our concerns corresponding to people with other occlusion who are classified as ones with the correct mask. YOLOv7-X has about 13 % misclassification of people with incorrect masks and 12% misclassification of people with other occlusion as shown in Figure 6. The examples of the output images from the detection of different YOLOv7 models are shown in Figure 7. It is obvious that the YOLOv7-tiny as shown in Figure 7(a), YOLOv7 as shown in Figure 7(b), YOLOv7-X as shown in Figure 7(c), YOLOv7-W6 as shown in Figure 7(d), YOLOv7-E6 as shown in Figure 7(e), on the other hand, and YOLOv7-D6 as shown in Figure 7(f) models only detect and classify one person out of two people in image, and YOLOv7-E6E6 as shown in Figure 7(g) models can detect and classify all persons properly. Although the five models out of seven could detect all persons successfully but the YOLOv7-tiny models showed the highest confidence ratio. The higher confidence indicates the best detection performance. Based on confusion matrix and misclassifications results, the YOLOv7-tiny model is the best one for facemask detection applications. Also, the detection time speed for YOLOv7-tiny allow it to be applicable in real time applications.

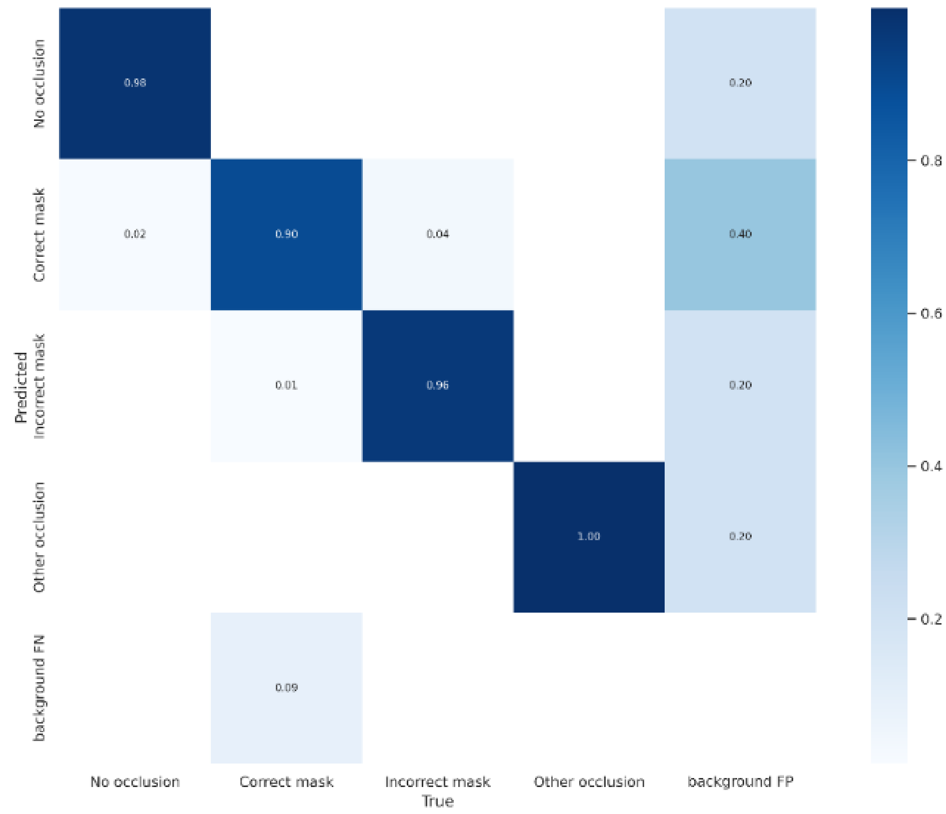

Figure 4. Confusion matrix for YOLOv7-tiny

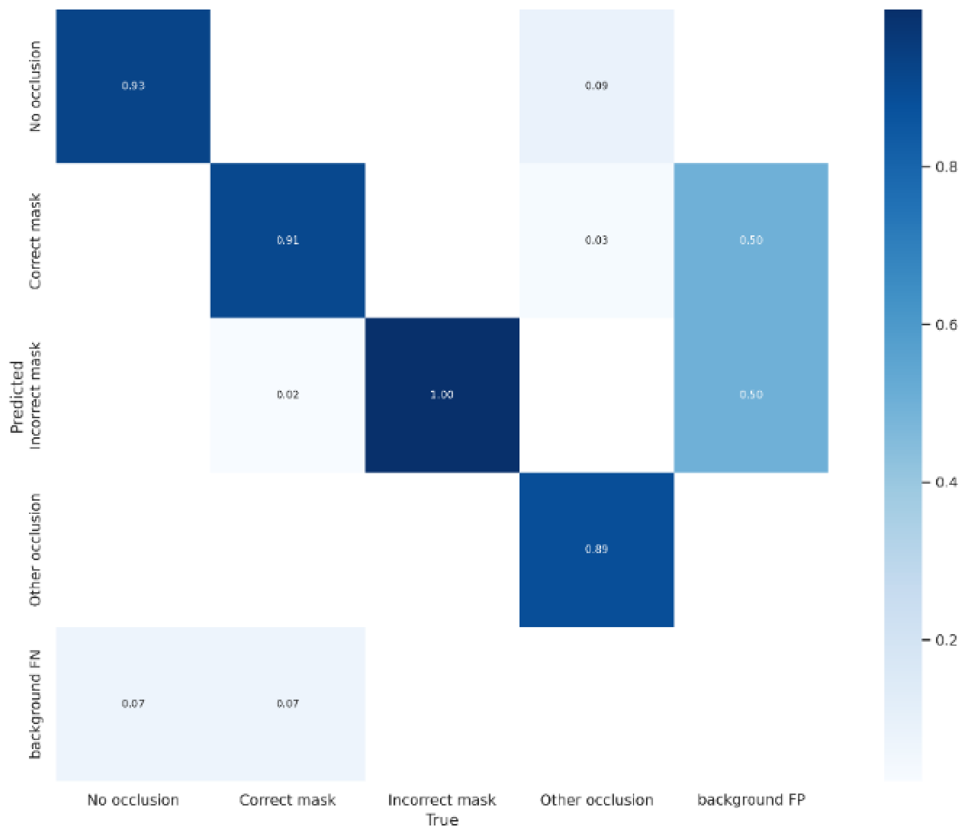

Figure 5. Confusion matrix for YOLOv7-tiny

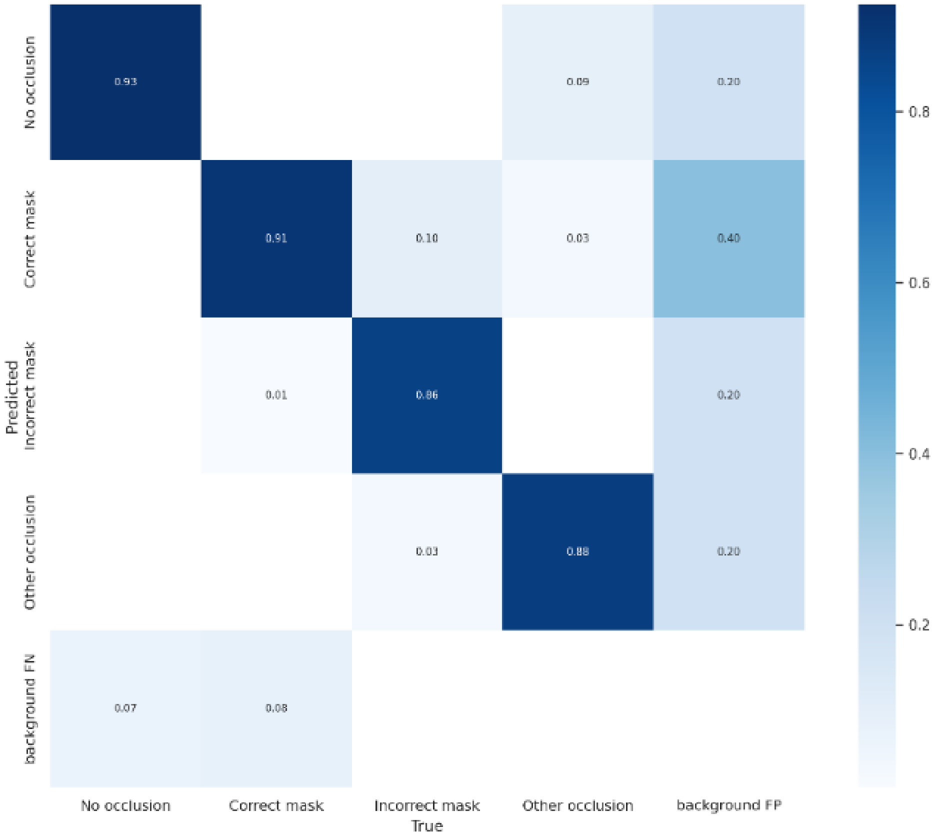

Figure 6. Confusion matrix for YOLOv7-X

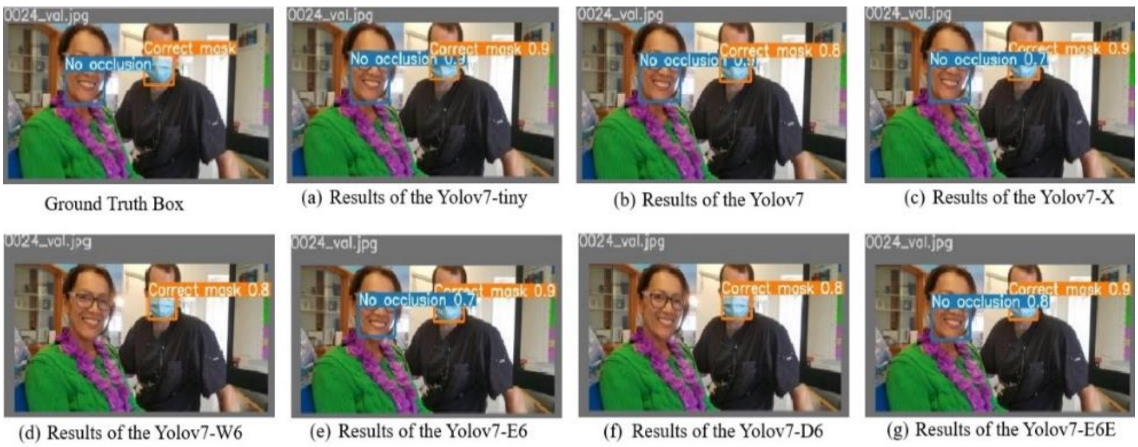

Figure 7. Detection results in different models (a) YOLOv7-tiny detections, (b) YOLOv7 detections, (c) YOLOv7-X detections, (d) YOLOv7-W6 detections, (e) YOLOv7-E6 detections, (f) YOLOv7-D6 model, and (g) YOLOv7-E6E detections

4. CONCLUSION

This work introduces the comprehensive investigation and comparative study of the efficiency of different versions of the YOLOv7 model in facemask detection tasks. Moreover, it proposed the best YOLOv7 version model that can be integrated with surveillance systems to detect and classify face masks effectively. Facemask detection is a vital first step in the COVID-19 epidemic fighting process. The dataset of 1,300 images with 1,853 faces was collected, and images were classified into four novel categories; no occlusion, other occlusion, correct mask, and incorrect mask. To avoid biasing, the collected dataset has an equal number of objects in each class. Then, the pre-trained YOLOv7 versions trained and their performance were investigated in terms of precision, recall, mAP, and inference time. Based on results, YOLOv7-tiny model achieved the highest detection performance with mAP@50 of 95.5%. Although the YOLOv7-tiny model couldn’t achieve the best inference time, it has an acceptable value of about 7.1 ms. Thus, it can be applied for mask detection in real-time applications. This model can be improved by increasing the dataset. Moreover, future studies can employ the model in any specific area, such as schools, airports, and hospitals, to monitor people who wear or are not wearing masks properly. In addition, future studies can tune the hyper-parameters to optimize the detection performance.

ACKNOWLEDGEMENTS

The researchers would like to thank the Deanship of Scientific Research, Qassim University for funding publication of this project.

REFERENCES

[1] S. Saadat, D. Rawtani, and C. M. Hussain, “Environmental perspective of COVID-19,” *Science of the Total Environment*, vol. 728, 2020, doi: 10.1016/j.scitotenv.2020.138870.

[2] WHO, “Coronavirus disease (COVID-19) advice for the public: When and how to use masks,” Corona desease 2019. Accessed: Sep. 10, 2023. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks

[3] J. M. D. -Watson, “Centers for disease control and prevention,” *Women’s Health: Understanding Issues and Influences*, vol. 1. pp. 119–121, 2022, doi: 10.1097/01.phm.0000189841.18419.16.

[4] A. R. Pathak, M. Pandey, and S. Rautaray, “Application of deep learning for object detection,” *Procedia Computer Science*, vol. 132, pp. 1706–1717, 2018, doi: 10.1016/j.procs.2018.05.144.

[5] X. Feng, Y. Jiang, X. Yang, M. Du, and X. Li, “Computer vision algorithms and hardware implementations: A survey,” *Integration*, vol. 69, pp. 309–320, 2019, doi: 10.1016/j.vlsi.2019.07.005.

[6] V. K. Sharma and R. N. Mir, “A comprehensive and systematic look up into deep learning based object detection techniques: A review,” *Computer Science Review*, vol. 38, 2020, doi: 10.1016/j.cosrev.2020.100301.

[7] L. Jiao *et al.*, “A survey of deep learning-based object detection,” *IEEE Access*, vol. 7, pp. 128837–128868, 2019, doi: 10.1109/ACCESS.2019.2939201.

[8] A. Nowrin, S. Afroz, M. S. Rahman, I. Mahmud, and Y. Z. Cho, “Comprehensive review on facemask detection techniques in the context of Covid-19,” *IEEE Access*, vol. 9, pp. 106839–106864, 2021, doi: 10.1109/ACCESS.2021.3100070.

[9] L. Liu *et al.*, “Deep learning for generic object detection: A survey,” *International Journal of Computer Vision*, vol. 128, no. 2, pp. 261–318, Feb. 2020, doi: 10.1007/s11263-019-01247-4.

[10] J. Zhang, F. Han, Y. Chun, and W. Chen, “A novel detection framework about conditions of wearing face mask for helping control the spread of COVID-19,” *IEEE Access*, vol. 9, pp. 42975–42984, 2021, doi: 10.1109/ACCESS.2021.3066538.

[11] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, “Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection,” *Sustainable Cities and Society*, vol. 65, Feb. 2021, doi: 10.1016/j.scs.2020.102600.

[12] Larxel, “Face mask detection dataset,” *Kaggle*, 2020. Accessed: Sep. 09, 2023. [Online]. Available: https://www.kaggle.com/andrewmvd/face-mask-detection

[13] X. Jiang, T. Gao, Z. Zhu, and Y. Zhao, “Real-time face mask detection method based on yolov3,” *Electronics*, vol. 10, no. 7, pp. 1–17, 2021, doi: 10.3390/electronics10070837.

[14] Ethancvaa, “Properly-wearing-masked-detect-dataset,” *GitHub*. Accessed: Sep. 05, 2023. [Online]. Available: https://github.com/ethancvaa/Properly-Wearing-Masked-Detect-Dataset

[15] Revanth, “Mafa-dataset,” *Kaggle*. Accessed: Jun. 10, 2023. [Online]. Available: https://www.kaggle.com/datasets/revanthrex/mafadataset

[16] V. Shanawad, “Face mask segmentation-WIDER face dataset,” *Kaggle*. Accessed: Sep. 03, 2023. [Online]. Available: https://www.kaggle.com/datasets/vinayakshanawad/face-mask-segmentation-wider-face-dataset

[17] Z. Wang, B. Huang, G. Wang, P. Yi, and K. Jiang, “Masked face recognition dataset and application,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 5, no. 2, pp. 298–304, 2023, doi: 10.1109/TBIOM.2023.3242085.

[18] M. R. Bhuiyan, S. A. Khushbu, and M. S. Islam, “A deep learning based assistive system to classify COVID-19 face mask for human safety with YOLOv3,” in *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020*, 2020, doi: 10.1109/ICCCNT49239.2020.9225384.

[19] M. L. Mokeddem, M. Belahcene, and S. Bourennane, “Yolov4FaceMask: COVID-19 mask detector,” in *2021 1st International Conference On Cyber Management and Engineering, CyMaEn 2021*, IEEE, 2021, doi: 10.1109/CyMaEn50288.2021.9497305.

[20] J. Yu and W. Zhang, “Face mask wearing detection algorithm based on improved yolo-v4,” *Sensors*, vol. 21, no. 9, 2021, doi: 10.3390/s21093263.

[21] “cabani/MaskedFace-Net public dataset,” *GitHub*. Accessed: Sep. 07, 2023. [Online]. Available: https://github.com/cabani/MaskedFace-Net

[22] A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, “MaskedFace-Net-A dataset of correctly/incorrectly masked face images in the context of COVID-19,” *Smart Health*, vol. 19, 2021, doi: 10.1016/j.smhl.2020.100144.

[23] A. Kumar, A. Kalia, A. Sharma, and M. Kaushal, “A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 6, pp. 6783–6796, 2023, doi: 10.1007/s12652-021-03541-x.

[24] “MOXA: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks dataset,” *GitHub*. [Online]. Available: https://shitty-bots-inc.github.io/MOXA/index.html

[25] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, “SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2,” *Sustainable Cities and Society*, vol. 66, 2021, doi: 10.1016/j.scs.2020.102692.

[26] B. Roy, S. Nandy, D. Ghosh, D. Dutta, P. Biswas, and T. Das, "MOXA: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks," *Transactions of the Indian National Academy of Engineering*, vol. 5, no. 3, pp. 509–518, 2020, doi: 10.1007/s41403-020-00157-z.

[27] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 7464–7475, doi: 10.1109/cvpr52729.2023.00721.

[28] Tzutalin, "Heartexlabs/labelImg," *GitHub*. [Online]. Available: https://github.com/heartexlabs/labelImg

[29] R. Padilla, S. L. Netto, and E. A. B. D. Silva, "A survey on performance metrics for object-detection algorithms," in *International Conference on Systems, Signals, and Image Processing*, IEEE, Jul. 2020, pp. 237–242, doi: 10.1109/IWSSIP48289.2020.9145130.

[30] A. Aqeel, "What is average precision in object detection and localization algorithms and how to calculate it?," *Towards Data Science.* Accessed: Sep. 12, 2023. [Online]. Available: https://towardsdatascience.com/what-is-average-precision-in-object-detection-localization-algorithms-and-how-to-calculate-it-3f330efe697b

[31] K. E. Koech, "Confusion matrix for object detection," *Towards Data Science*. Accessed: Sep. 13, 2023. [Online]. Available: https://towardsdatascience.com/confusion-matrix-and-object-detection-f0cbcb634157

[32] V. Dubey, "Evaluation metrics for object detection algorithms," *Medium.* Accessed: Sep. 11, 2023. [Online]. Available: https://medium.com/@vijayshankerdubey550/evaluation-metrics-for-object-detection-algorithms-b0d6489879f3

BIOGRAPHIES OF AUTHORS

Gareebah Al-Rasheedi received a Bachelor of College of Computer in Information Technology from Qassim University (QU). She is currently a Master's student at Qassim University. Her research areas of interest include artificial intelligent, specifically in deep learning and computer vision. She can be contacted at email: grob083@hotmail.com.

Rehan Ullah Khan graduated from the University of Engineering and Technology Peshawar with a B.Sc. in Information Systems in 2004 and M.Sc. in Information Systems in 2006. He obtained his Ph.D. degree in 2011 from the Vienna University of Technology, Austria. He is currently an Associate Professor and Director of R&D at the College of Computer, Qassim University, KSA. His current research 9 include machine learning, pattern recognition, and computer vision. He is an author/co-author of more than 100 publications. He can be contacted at email: re.khan@qu.edu.sa.