Face Mask Detection Using YOLOv4

Conference Paper · August 2021

1 author:

Murat Ucar

Face Mask Detection Using YOLOv4

Murat UÇAR[1*]

[1]*İskenderun Technical University, Faculty of Business and Management Science, Department of Management Information Systems, Hatay, TURKEY*

Abstract

The coronavirus disease (COVID-19), which first appeared in Wuhan, China towards the end of 2019, spread rapidly all over the world within a few months and caused the death of a large number of people. One of the most important features of the disease is that it can spread from person to person through small droplets that come out of the mouth or nose as a result of coughing of the infected person. The use of masks is one of the effective measures taken to reduce the spread of COVID-19. Therefore, a face mask detection system based on image analysis is a crucial task to assist the community. In this study, the newly developed YOLOv4 algorithm has been used for face mask detection. This study makes a detection in three classes including not wearing a mask, the mask does not conforming to the rules and wearing the mask properly. The obtained findings as a result of experimental tests proved that the model is effective and usable for face mask detection with its 82.2% mean average precision value.

Keywords: *COVID-19, Face mask detection, YOLOv4, Object detection*

YOLOv4 ile Yüz Maskesi Algılama

Özet

İlk olarak 2019 yılının sonlarına doğru Çin'in Wuhan kentinde ortaya çıkan korona virüs hastalığı (COVID-19), birkaç ay içinde tüm dünyaya hızla yayılarak çok sayıda insanın ölümüne neden olmuştur. Hastalığın en önemli özelliklerinden biri de enfekte kişinin öksürmesi sonucu ağızdan veya burundan çıkan küçük damlacıklar yoluyla insandan insana yayılabilmesidir. Maske kullanımı, COVID-19'un yayılmasını azaltmak için alınan etkili önlemlerden biridir. Bu nedenle, görüntü analizine dayalı bir yüz maskesi algılama sistemi, topluluklara yardımcı olmak için çok önemli bir görevdir. Bu çalışmada yüz maskesi tespiti için yeni geliştirilen bir algoritma olan YOLOv4 algoritması kullanılmıştır. Bu çalışma, maske takmama, maskenin kurallara uymaması ve maskeyi düzgün takma olmak üzere üç sınıfta tespit yapmaktadır. Deneysel testler sonucunda elde edilen bulgular, modelin %82.2 ortalama hassasiyet değeri ile yüz maskesi tespiti için etkili ve kullanılabilir olduğunu kanıtlamıştır.

Anahtar Kelimeler: *COVID-19, Yüz maskesi algılama, YOLOv4, Nesne tanıma*

1 Introduction

Covid-19 disease, which occurred in Wuhan, China in December 2019, spread all over the world in a short time and was declared as a pandemic by the World Health Organization. It continues to spread rapidly, and there are about 43,341,451 total cases and 1,157,509 total deaths as on October 25, 2020, worldwide [1]. This is why public health has been considered a top priority for governments since the outbreak began. Leung et al. proved in their study that surgical face masks can prevent the spread of coronavirus [2]. For this reason, the use of masks, one of the biggest protection methods from the virus, has been widespread worldwide. In addition, curfews without a mask have been imposed in many cities in our country. Nowadays, face mask detection systems have become a very important task, but research on face mask detection in the literature is limited.

* Corresponding e-mail: murat.ucar@iste.edu.tr

In this paper, YOLOv4 object detection algorithm was used for face mask detection. The novelty of this article is that the newly developed YOLOv4 algorithm is used for face mask detection. YOLOv4 provides higher accuracy in real-time face mask detection. One of the contributions of this article to the literature is; while it has been determined in studies conducted to date whether only people are wearing masks or not, in this study it was also determined whether the mask was worn properly or not.

The remainder of this paper is organized as follows. Section II presents studies on face mask detection using deep neural networks. The used data set, proposed method, experimental settings and evaluation criteria is detailed in Section III. Section IV describes the results and comparisons. Finally, Section V concludes the article.

2 Literature Survey

Recent studies show that models developed based on deep learning have made extraordinary progress in object detection and computer vision compared to traditional models. Deep learning based object detection models can generally be divided into two categories. The first of these are single-stage models such as YOLO [3] and SSD [4], the second is two-stage models such as R-CNN [5], Fast R-CNN [6] and Faster R-CNN [7].

Various studies have been conducted on face mask detection and classification using different deep learning techniques. The first study on face mask classification is a hybrid model developed by Loey et al. They used the ResNet50 model, one of the deep learning architectures, for feature extraction, and the decision trees (DT) and support vector machines (SVM) from traditional machine learning models for classification. They tested the developed model on three datasets and reported that the SVM classifier achieved between 99% and 100% accuracy rate [8]. The other classification study was suggested by Qin and Li. They have developed a facial mask recognition system using a classification network (SRCNet) that incorporates image super resolution. They tested their proposed system on Medical Masks Dataset, which contains 3835 images. The proposed SRCNet model achieved 98.70% accuracy [9].

The studies on facemask detection were summarized below. Inamdar and Mehendale proposed a model called Facemasknet using deep learning techniques for mask detection. Using the proposed model, they made a triple classification: no mask, the mask has worn incorrectly and the mask has worn properly. They reported that the model, which can be used in both images and video streaming, achieved 98.6% accuracy rate [10]. Wang et al. created three public data sets: Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD) and Simulated Masked Face Recognition Dataset (SMFRD). They also proposed a multi-particle masked face recognition model to recognize masked faces. The obtained findings showed that the proposed model reached an accuracy of 95% [11]. In another study, Yadav has proposed a model that focuses on automatic real-time tracking of people to detect safe social distance and face masks in public places. The proposed model consists of combining the MobileNetV2 model with the Single Shot object Detection (SSD) framework. The obtained findings showed that the system could detect social distance and facial masks with a sensitivity score of 91.7% [12]. Jiang et al., have developed a facemask detector called RetinaFaceMask. The developed model is a single-stage detector and includes a feature pyramid network to combine high-level semantic information with multiple feature maps, and an attention module to detect facemasks. They reported that RetinaFaceMask achieved high accuracy results in a public facemask dataset [13].

3 Material and Method

3.1 Dataset

In this study, a dataset containing 853 images shared publicly on the Kaggle website was used [14]. The images in the dataset belong to three classes: with mask, without mask and mask has worn incorrectly. Sample images of the dataset are shown in Figure 1. Their bounding boxes in the PASCAL VOC format was converted to YOLO format and the Pseudo-code of this process was presented in Figure 2.

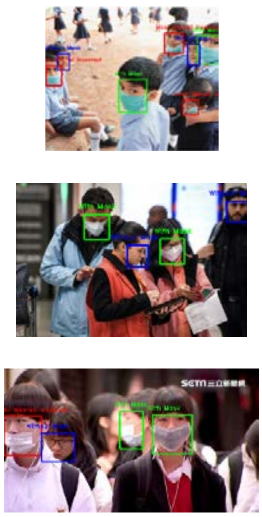

Figure 1. Sample images of dataset

```
Given SOURCE_FILES (List of Annonation files)
Initialize CLASSNAMES = {'without_mask':0,'with_mask':1,'mask_weared_incorrect':2}

for i = 0 to length of SOURCE_FILES
        replace .xml with .txt in SOURCE_FILES[i] and save yolofilename

        for j = 0 to object count in SOURCE_FILES[i]
                read name value from object[j] and save in name
                read width value from object[j] and save in width
                read height value from object[j] and save in height
                read xmin value from object[j] and save in xmin
                read ymin value from object[j] and save in ymin
                read xmax value from object[j] and save in xmax
                read ymax value from object[j] and save in ymax
                dw = 1 / width
                dh = 1 / height
                x = (xmin + xmax) / 2
                y = (ymin + ymax) / 2
                w = xmax - xmin
                h = ymax - ymin
                x = x * dw
                w = w * dw
                y = y * dh
                h = h * dh
                classcode = CLASSNAMES[name]
                write classcode, x, w, y, h into yolofilename
```

Figure 2. Pseudo-code of the converting process

3.2 Yolov4

You Only Look Once (YOLO) is a novel algorithm for real-time object detection, introduced by Joseph Redmon et al., in 2016 [3]. The main feature that distinguishes YOLO from other region-based algorithms is that a single convolutional network can predict classes and bounding boxes for the entire image in a single execution of the algorithm. Later, in December 2017, Redmond and Farhadi developed a higher version of YOLO, also known as YOLO9000 [15]. Then, a year later, Redmond and Farhadi introduced YOLOv3. YOLOv3 is an improved version of YOLOv2. It uses the multi-scale estimation method to detect the target and has a more complex network structure than YOLOv2. YOLOv3 detects small targets more effectively than YOLOv2, thanks to this multi-scale estimation method [16]. However, in April 2020, Bochkovskiy et al., have proposed the YOLOv4 algorithm, which is the most advanced version of YOLO and provides much better accuracy than previous versions [17]. YOLOv4 has a deeper and more complex network structure with Dense Blocks to achieve better accuracy. It contains multiple convolutional layers using the batch normalization and RELU activation function. YoloV4 algorithm uses CSPDarknet-53 framework, which uses Cross-Stage Partial connections (CSP) connections as backbone. The CSP technique carries out the separation of Dense Blocks feature maps into two parts. One of the split pieces goes directly to the next transition layer, while the other piece goes directly to Dense blocks. Furthermore, the YoloV4 architecture includes SPP additional module, PANet path-aggregation neck and YOLOv3 anchor-based head. The SPP block is tasked with increasing the receiving area and separating the most important context features with virtually no reduction in network processing speed. PANet method is used to collect parameters from different backbone levels for different detector levels.

3.3 Experimental setup

In this study, all experiments were carried out in Google's Jupyter notebook Colab with one GPU (1xTesla K80, compute 3.7, 2496 CUDA cores, 12GB GDDR5 VRAM). 80% of the data set (683 images) were used as training data for training the YOLOv4 algorithm, and the remaining 20% (170 images) were used as testing data to test the accuracy of the used model in detecting face masks.

The hyper-parameters of YOLOv4 was used as follows: Due to the differences in the size of the images in the used dataset, the input size of the images was set to 416 × 416 pixels to facilitate detection. Initial learning rate was set as 0.001 and multiplied with a factor 0.1 at 4800 steps and 5400 steps, respectively. In order to perform multi-scale training all architectures used a single GPU with batch size 64 and mini-batch size 16. The default momentum 0.949, IoU threshold 0.213 and loss normalizer 0.07 was executed as proposed by authors of YOLOv4.

Table 1. Parameters of YOLOv4 face mask detection model.

Parameters	**Value**
Input size	416 x 416
Initial learning rate	0.001
Batch size	64
Mini batch size	16
Iterations	6000

3.4 Evaluation metrics

The performance of YOLOv4 method was compared using the average Intersection over Union(IoU), precision, recall, F1 score, average precision and Mean Average Precision (mAP) as the evaluation metrics.

Intersection over Union, also called the Jaccard index, is used to calculate the accuracy of an object detector in a given data set. The IoU is calculated by the ratio of the intersection between the ground-truth box and the predicted box to the total area of the union of these two boxes. This metric is given in equation 1.

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ union} \tag{1}$$

Precision is a measure of how many of the predicted positives are actually positive. The mathematical definition of it is as given in equation 2.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall is a measure of how many of the true positives are correctly classified. The mathematical definition of it is as given in equation 3.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1 score, also called F score, is a function of precision and recall and it is needed to maintain a balance between precision and recall. This metric is given in equation 4.

$$F1\ Score = 2\ x\ \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

Average Precision (AP), one of the most used metrics to measure the accuracy of object detectors, is used to find the area under the precision-recall curve. Mean of the Average Precision (mAP) is the average of AP calculated for all classes.

4 Experimental Results and Discussion

In order to show the detection performance of YOLOV4 method, 170 test images were used and the results of the method presented in Table 2.

Table 2. Detection results of the YOLOv4 method.

Method	Precision(%)	Recall(%)	F1-Score(%)	Average IoU(%)	mAP(%)
Yolov4	86	89	88	69.48	82.20

As seen in Table 2 the precision, recall and F1 score of the proposed method were 86%, 89% and 88%, respectively. In this study the average IoU value was observed as 69.48% and the mAP(@.50) value was observed as 82.20%. Furthermore, the performances of the YOLOv4 model for each class are shown in Table 3.

Table 3. Average precision values of the YOLOv4 for each class.

Class	TP	FP	AP(%)
Without mask	127	31	82.62
With mask	526	70	94.75
Mask has worn incorrectly	15	5	69.22

As can be seen from Table 3, the model shows the best performance with 94.75% in identifying suitable mask wearers. This performance is followed by the detection of people who do not wearing masks with 82.62%. The performance of the model in detecting people who do not wear their masks properly is the lowest. The reason for this is that the images belonging to this class are more difficult to separate than the other two classes.

The detection results of the YOLOv4 model on these three classes is presented in Figure 3. As can be seen from figure 2, the recognition ability of the model on images is well. In addition, the trained model can be run by accessing the link: https://github.com/mrtucar/FaceMaskDetection.

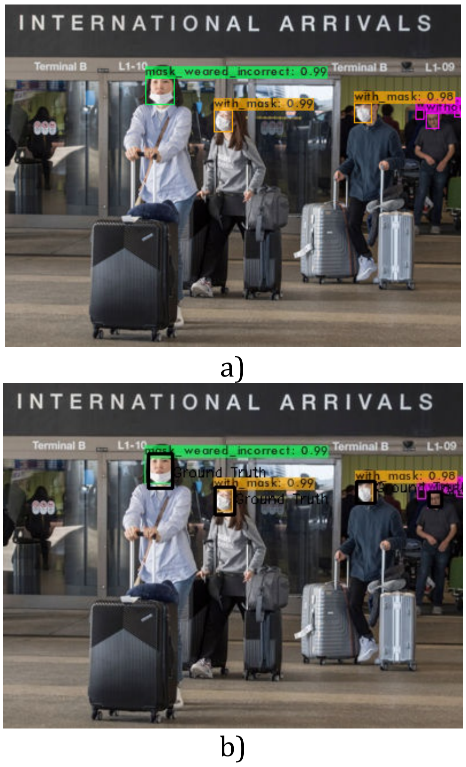

Figure 3. Detection results of YOLOv4 model on test images. a) Image with only predicted boxes b) Image with ground truth boxes and predicted boxes.

The accuracy results obtained from this study were compared with those reported in the literature (Table 4). As can be seen from Table 4, in studies where facemask detection was performed before, it was determined that only people were wearing masks or not. In this study, unlike the literature, a triple detection has made because it is also important to wear the mask properly in terms of controlling the spread of COVID-19 disease. As a result of the study, when the mAP score of 82.20% obtained with the YOLOv4 model was compared with the alternative solutions in the literature, it was observed that the detection performance of the solution was high.

Table 4. Accuracy results comparison with related works.

Author	Model	Training dataset	Detection	Accuracy (%)
Inamdar and Mehendale	FaceMaskNet	10 masked images 10 unmasked images 15 mask has worn incorrectly images	-With mask -Without mask	98.60 (Acc)
Wang et al.	Multi-granularity masked face recognition	RMFD- 5000 masked, 90000 unmasked images MFDD- 24771 masked images	-With mask -Without mask	95.00 (Acc)
Yadav et al.	SSD+MobileNetV2	3165 masked and unmasked images	-With mask -Without mask	91.70 (Acc)
Jiang et al.	RetinaFaceMask	7959 masked and unmasked images	-With mask -Without mask	91.90 (Precision for No Mask images) 93.40 (Precision for Masked images)
This study	YOLOv4	853 images	-With mask -Without mask -Mask has worn incorrectly	82.20 (mAP)

5 Conclusion

In this study, the state-of-art YOLOv4 detection model was used for detecting face masks effectively. The experimental results show that the YOLOv4 model has achieved high accuracy with the 82.2% mean average precision score. Therefore, it can be used for detecting cases of not wearing a mask, the mask does not conforming to the rules and wearing the mask properly.

This study will help reduce the spread of the coronavirus by facilitating the identification of people without masks or who do not wear masks properly in public places such as schools, shopping malls, subway stations and markets.

References

[1] WHO, "Report," 2020. https://www.who.int/publications/m/item/weekly-epidemiological-update---27-october-2020. Accessed: Oct. 30, 2020. [Online].

[2] N. H. L. Leung *et al.*, "Respiratory virus shedding in exhaled breath and efficacy of face masks," *Nat. Med.*, vol. 26, no. 5, pp. 676–680, May 2020, doi: 10.1038/s41591-020-0843-2.

[3] J. Redmon, S. Divvala, R. B. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 779–788, 2016.

[4] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9905 LNCS, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[5] K. L. Masita, A. N. Hasan, and S. Paul, "Pedestrian Detection Using R-CNN Object Detector," Jan. 2019, doi: 10.1109/LA-CCI.2018.8625210.

[6] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[8] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," *Meas. J. Int. Meas. Confed.*, vol. 167, p. 108288, Jan. 2021, doi: 10.1016/j.measurement.2020.108288.

[9] B. Qin and D. Li, "Identifying Facemask-wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID-19," 2020. Accessed: Oct. 30, 2020. [Online]. Available: https://www.kaggle.com/vtech6/medical-masks-dataset.

[10] M. Inamdar and N. Mehendale, "Real-Time Face Mask Identification Using Facemasknet Deep Learning Network," *SSRN Electron. J.*, Aug. 2020, doi: 10.2139/ssrn.3663305.

[11] Z. Wang *et al.*, "Masked Face Recognition Dataset and Application," Mar. 2020, Accessed: Oct. 30, 2020. [Online]. Available: http://arxiv.org/abs/2003.09093.

[12] S. Yadav, "Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID-19 Safety Guidelines Adherence," *academia.edu*, Accessed: Oct. 30, 2020. [Online]. Available: https://www.academia.edu/download/64029481/560.pdf.

[13] M. Jiang and X. Fan, "RetinaMask: A Face Mask detector," Accessed: Oct. 30, 2020. [Online]. *arXiv Prepr. arXiv2005.03950*, 2020.

[14] "Kaggle Dataset," 2020. https://www.kaggle.com/andrewmvd/face-mask-detection. Accessed: Aug. 28, 2020. [Online].

[15] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Nov. 2017, vol. 2017-January, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.

[16] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, Accessed: Oct. 30, 2020. [Online]. Available: http://arxiv.org/abs/1804.02767.

[17] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 2020, Accessed: Oct. 30, 2020. [Online]. Available: http://arxiv.org/abs/2004.10934.