

Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection

Mohamed Loey^{a,*}, Gunasekaran Manogaran^{b,c}, Mohamed Hamed N. Taha^d, Nour Eldeen M. Khalifa^d

^a Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha 13518, Egypt

^b University of California, Davis, USA

^c College of Information and Electrical Engineering, Asia University, Taiwan

^d Department of Information Technology, Faculty of Computers & Artificial Intelligence, Cairo University, Cairo 12613, Egypt



ARTICLE INFO

Keywords:
COVID-19
Medical masked face
YOLO
ResNet
Deep learning

ABSTRACT

Deep learning has shown tremendous potential in many real-life applications in different domains. One of these potentials is object detection. Recent object detection which is based on deep learning models has achieved promising results concerning the finding of an object in images. The objective of this paper is to annotate and localize the medical face mask objects in real-life images. Wearing a medical face mask in public areas, protect people from COVID-19 transmission among them. The proposed model consists of two components. The first component is designed for the feature extraction process based on the ResNet-50 deep transfer learning model. While the second component is designed for the detection of medical face masks based on YOLO v2. Two medical face masks datasets have been combined in one dataset to be investigated through this research. To improve the object detection process, mean IoU has been used to estimate the best number of anchor boxes. The achieved results concluded that the adam optimizer achieved the highest average precision percentage of 81% as a detector. Finally, a comparative result with related work has been presented at the end of the research. The proposed detector achieved higher accuracy and precision than the related work.

1. Introduction

The outbreak of coronavirus (COVID-19) has forced many countries to initiate new rules for face mask-wearing. Governments have started working on new strategies to manage spaces, social distancing, and supplies for medical staff and normal people. Also, the government has forced hospitals and other organizations to apply new infection prevention measures to stop the spreading of COVID-19. The COVID-19 transmission rate is about 2.4 (Ferguson et al., 2020; Sun & Zhai, 2020). However, the transmission rate may vary according to the measurement and policies applied by the governments. As COVID-19 is transmitted through airdrops and close contact, Governments have started applying new rules forcing people to wear face masks. The goal of wearing face masks is to reduce the transmission and spreading rate. The World Health Organization (WHO) has recommended the usage of personal protective equipment (PPE) among people and in medical care. However, the ability of most of the countries to expand the production of

PPE is very limited (PPE, 2020).

Today, COVID-19 is a significant public health and economy issue due to the detrimental effects of the virus on people's quality of life, contributing to acute respiratory infections, mortality and financial crises worldwide (Rahmani & Mirmahaleh, 2020). According to (WHO, 2020), more than six million cases were infected by COVID-19 in more than 180 countries with death-rate of 3%. The COVID-19 spreads easily in crowded environments and close contact. Governments are facing extraordinary challenges and risks to protect people from coronavirus in many countries (Altmann, Douek, & Boyton, 2020). As people are forced by laws to wear face masks in public in many countries, masked face detection is a key to face applications, such as object detection (Wu, Sahoo, & Hoi, 2020). To fight and win in the battle against COVID-19 pandemic, Governments need guidance and surveillance on people in public areas, especially the crowded to ensure that wearing face masks laws are applied. This could be applied through the integration between surveillance systems and Artificial Intelligence models.

* Corresponding author.

E-mail addresses: mloey@fci.bu.edu.eg (M. Loey), gmanogaran@ieee.org (G. Manogaran), mnasrta@cu.edu.eg (M.H.N. Taha), nourmahmoud@cu.edu.eg (N.E.M. Khalifa).

The main objective of this research is to detect and locate a medical face mask in an image as illustrated in Fig. 1. In this paper, the medical masked face is the main focus of research to reduce the spreading and transmission of Coronavirus specially COVID-19. Given an image, a region of the medical masked face on the input image based on YOLO-v2 with ResNet-50 will be illustrated in the output image.

All object detection methods classify machine learning-based and deep learning. Machine learning detectors such as Scale-invariant feature transform (SIFT) (Lowe, 1999), and Histogram of oriented gradients (HOG) (Lowe, 2004). Deep learning detectors such as region convolutional neural network (R-CNN) (Girshick, Donahue, Darrell, & Malik, 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren, He, Girshick, & Sun, 2017), You Only Look Once (YOLO v1 (Redmon, Divvala, Girshick, & Farhadi, 2016), v2 (Redmon & Farhadi, 2017), v3 (Lee, Lee, Lee, & Kim, 2019), and Single Shot Multi-Box Detector (SSD) (Liu et al., 2016). YOLO family rather than region convolutional neural network family designed for high speed and performance. The main contributions of this paper are conducted as follows:

- 1) A novel deep learning detector model that automatically finds and localize medical masked face on an image.
- 2) A New masked face dataset using two public masked face datasets to get rid of the dataset's scarcity problem.
- 3) The proposed model improves detection performance by introducing mean IoU to estimate the best number of anchor boxes.
- 4) Two optimizers are used in training to get the highest performance possible.
- 5) YOLO-v2 detector is an effective model to find a masked face on input image based on ResNet-50.

2. Related works

Object detection from an image is probably the deepest aspect of computer vision due to widely used in many cases. There has been supervised or unsupervised based learn in the field of computer vision to outfit the work of object detection in an image. This section conducts the recent academic papers for applying representative works related to object detection based on deep learning for the medical face mask. Most of mask face detection focus on face construction and face recognition based on traditional machine learning techniques. In this paper, our focus is on detecting and find the human who is wearing a face mask to help in lessening the spreading of the COVID-19. The authors of this paper also designed a lot of deep learning architecture based on solving image classification problems in many scientific fields (El-Sawy, EL-Bakry, & Loey, 2017; El-Sawy, Loey, & EL-Bakry, 2017; Khalifa, Taha, & Hassanien, 2018; Khalifa, Taha, Hassanien, & Hemedan, 2019; Khalifa, Taha, Ezzat Ali, Slowik, & Hassanien, 2020; Loey, Smarandache, & Khalifa, 2020).

In (Ejaz et al., 2019), authors have implemented a classical machine learning method to recognize masked and unmasked face using

Principal Component Analysis (PCA). The paper concluded that face without a mask gives a better recognition rate in PCA. It is found that extracting features from a masked face is less than an unmasked face. They found that the accuracy of mask face classification using the PCA is related to wear masks. When wearing a mask, the accuracy decreases to 70%. In (Ud Din, Javed, Bae, & Yi, 2020) proposed a novel GAN-based network to remove mask objects in facial images. The proposed GAN used two discriminators: The first one extracts the global structure of the masked face, The Second one extract the missing region from the masked face. In the training process, they used paired synthetic datasets. The outcome of the introduced model producing high-quality results for removing masked from the face. In (Loey, Manogaran, Taha, & Khalifa, 2021), the authors proposed a hybrid deep transfer learning model with machine learning methods for face mask classification. The proposed model consisted two phases 1) feature extraction base on ResNet50 2) classification based on SVM, decision tree, and ensemble. Three datasets were used as benchmarks to evaluate the proposed methodology. The SVM classifier achieved the highest accuracy with 99.64%. In (Ge, Li, Ye, & Luo, 2017) The authors proposed a model and dataset to find the normal and masked face in the wild. They introduced a large dataset Masked Faces (MAFA), which has 35, 806 masked faces. The proposed model based on a convolutional neural network called LLE-CNNs, which consists of three modules (a proposal, embedding, and verification). The works showed that LLE-CNNs using MAFA achieved the average precision equal to 76.1%.

3. Datasets Characteristics

This paper conducted its experiments based on two public medical face mask datasets. The first dataset is Medical Masks Dataset (MMD) published by Mikolaj Witkowski (<https://www.kaggle.com/vtech6/medical-masks-dataset>). The MMD dataset consists of 682 pictures with over 3k medical masked faces wearing masks. Fig. 2 illustrates samples of images in MMD.

The second public masked face dataset is a Face Mask Dataset (FMD) in (<https://www.kaggle.com/andrewmvd/face-mask-detection>). The FMD dataset consists of 853 images. Some samples of the FMD are introduced in Fig. 3. We created a new dataset by combining MMD and FMD. The merged dataset contains 1415 images by removing bad quality images and redundancy.

4. The Proposed Detector Model

Fig. 4 presents the architecture diagram of the proposed detector model. The introduced model includes three main components: the first component is the number of anchor boxes, the second component is the data augmentation, the final main component is the detector. Fig. 4 illustrates the proposed detector model. Mainly, the detector used YOLOv2 with ResNet-50 for the feature extraction and detection in the training, validation, and testing phase.



Fig. 1. The outcome of the proposed masked face detector.



Fig. 2. Samples of MMD dataset.



Fig. 3. Samples of FMD dataset images.

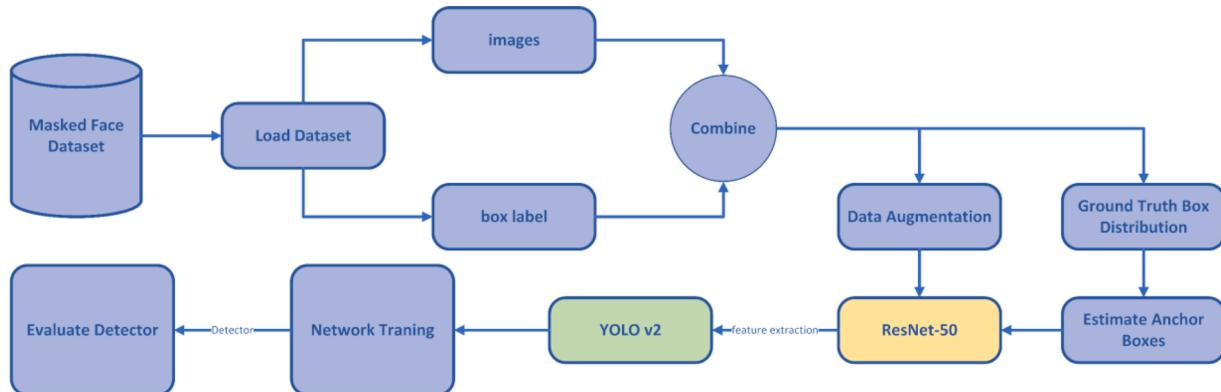


Fig. 4. The proposed detector model.

4.1. Estimate Anchor Boxes

Estimating the number of anchor boxes is an important stage to produce a high-performance detector. To compute the number anchor boxes, we visualize the labeled boxes of images as shown in Fig. 5. The figure illustrates a similar size of objects. The proposed model used mean Intersection over Union (IoU) as illustrated in equation (1) (Barthakur & Sarma, 2019) distance metric to estimate the number of anchor boxes. IoU in object detection is a method to calculate the distance of similarity between the bounding box of target and predicted output. In the training data, the mean IoU guarantee that the anchor boxes overlap with the boxes. The best number of anchors is 23 (mean IoU = 0.8634) to increase detector performance as shown in Fig. 6.

$$\text{IoU} = \frac{\text{Overlap Area between bounding box of target and predicted}}{\text{Combined Area between bounding box of target and predicted}} \quad (1)$$

4.2. Data Augmentation

Data augmentation is a method that can be used to artificially increase the diversity of datasets for training detectors. By transforming the original masked face images during training. Data augmentation improves the performance of the detector in training (AbdElNabi, Wajeeh Jasim, EL-Bakry, Taha, & Khalifa, 2020; Loey, ElSawy, & Afify, 2020; Loey, Naman, & Zayed, 2020). Data and their box labels were flipped horizontally to increase the masked face dataset as shown in Fig. 7.

4.3. YOLO v2 with ResNet-50 Detector

In 2016, YOLO v2 and YOLO 9000 was proposed by J. Redmon and A. Farhadi (Redmon & Farhadi, 2017). YOLO v2 object detection deep network is composed of feature extraction network and detection network as shown in Fig. 8. The feature extraction network (ResNet-50) is a deep transfer learning model. In the proposed model, ResNet-50 used as a deep transfer model for feature extraction. A residual neural

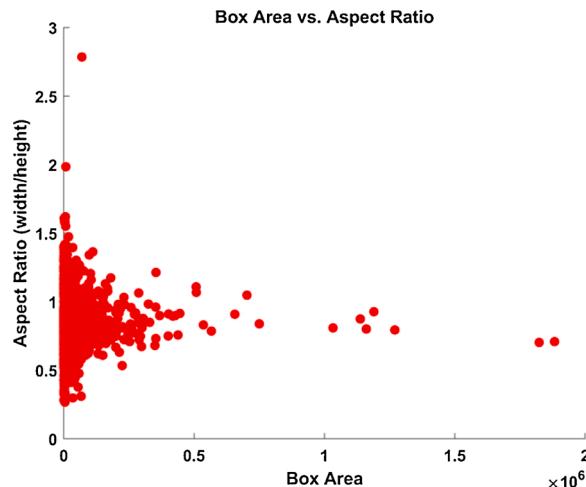


Fig. 5. Visualize the labeled boxes of Medical Masked faces images.

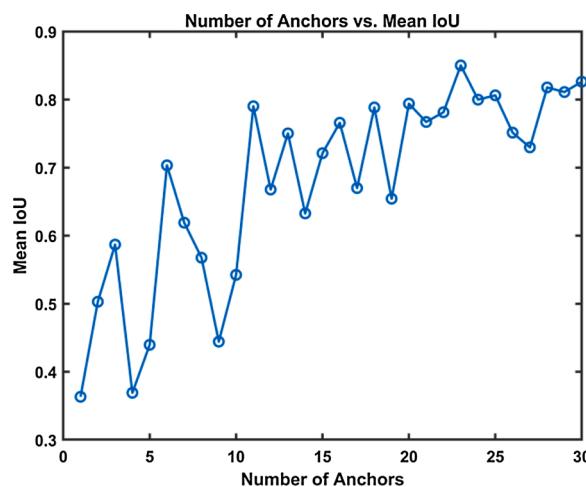


Fig. 6. Mean IoU of number of anchors for Medical Masked faces images.

network (ResNet) is a class of deep transfer learning based on a residual network (He, Zhang, Ren, & Sun, 2016). ResNet-50 has 16 residual bottleneck blocks each block has convolution size 1x1, 3x3, and 1x1 with feature maps (64, 128, 256, 512, 1024) as shown in Fig. 8 (ResNet part). The detection network (YOLO v2) is a convolutional neural network contain few convolutional layers, transform layer, and finally output layer. The transform layer extracts activations of convolutional layer and improves the steadiness of the deep neural network. The transform layer converts the bounding box forecast to be in outlines of the target box. The locations of pure bounding box of the goal is produced by the output layer.

To compute mean squared error loss between training predicted bounding boxes and the target in YOLO v2, the loss function of YOLO v2 is calculated as (Redmon et al., 2016), (Redmon & Farhadi, 2017):

$$\text{YOLO v2 loss} = \text{Localization loss} + \text{Confidence loss} + \text{Classification loss} \quad (2)$$

Localization loss measures error between the target and the predicted bounding box. The coefficients for calculating the localization loss include the width (w) and height (h) of the grid cell of the bounding box. The localization loss coefficients are calculated as follows (equation 3).

$$\begin{aligned} \text{Localization loss} = & q_1 \sum_{a=0}^{g^2} \sum_{b=0}^v 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\ & + q_1 \sum_{a=0}^{g^2} \sum_{b=0}^v 1_{ij}^{obj} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \end{aligned} \quad (3)$$

Where q_1 is a weight, g is a count of grid cells, v is a count of bounding boxes in each g , (x_i, y_i) is a center of v in g , (w_i, h_i) is a width and height of v in g , (\hat{x}_i, \hat{y}_i) is a center of the target in g , (\hat{w}_i, \hat{h}_i) is a center of the target in g , 1_{ij}^{obj} is 1 when there is an object in v in each g otherwise 0.

Confidence loss computes the confidence score of error when an object is detected in the v bounding box of g . The coefficients of confidence loss are calculated in equation 4.

$$\text{Confidence loss} = q_2 \sum_{a=0}^{g^2} \sum_{b=0}^v 1_{ij}^{obj} (cs_i - \hat{cs}_i)^2 + q_3 \sum_{i=0}^{g^2} \sum_{j=0}^v 1_{ij}^{noobj} (cs_i - \hat{cs}_i)^2 \quad (4)$$

Where q_2, q_3 is a weight of confidence error, cs_i is the confidence score of v in g , \hat{cs}_i is the confidence score of the target in g , 1_{ij}^{obj} is 1 when there is an object in v in each g otherwise 0, 1_{ij}^{noobj} is 1 when there is no object in v in each g otherwise 0.

The classification loss measures the error between the class conditional probabilities for each class in grid cell i . The parameters for computing the classification loss are defined in equation 5.

$$\text{Classification loss} = q_4 \sum_{a=0}^{g^2} 1_i^{obj} \sum_{z \in \text{classes}} (p_i(z) - \hat{p}_i(z))^2 \quad (5)$$

Where q_4 is a weight of Classification error, $p_i(z)$ and $\hat{p}_i(z)$ is the probability of object estimated and actual conditional class in grid cell a .

5. Experimental Results

To evaluate the YOLO v2 with ResNet-50 performance to find and localize the medical masked face, different experiments have been conducted throughout this research. The proposed model was implemented on the system having the following specifications: The GPU used NVIDIA RTX with the CUDA with Tensorflow, MATLAB, and Deep



Fig. 7. Sample of data augmentation in Medical Masked faces images.

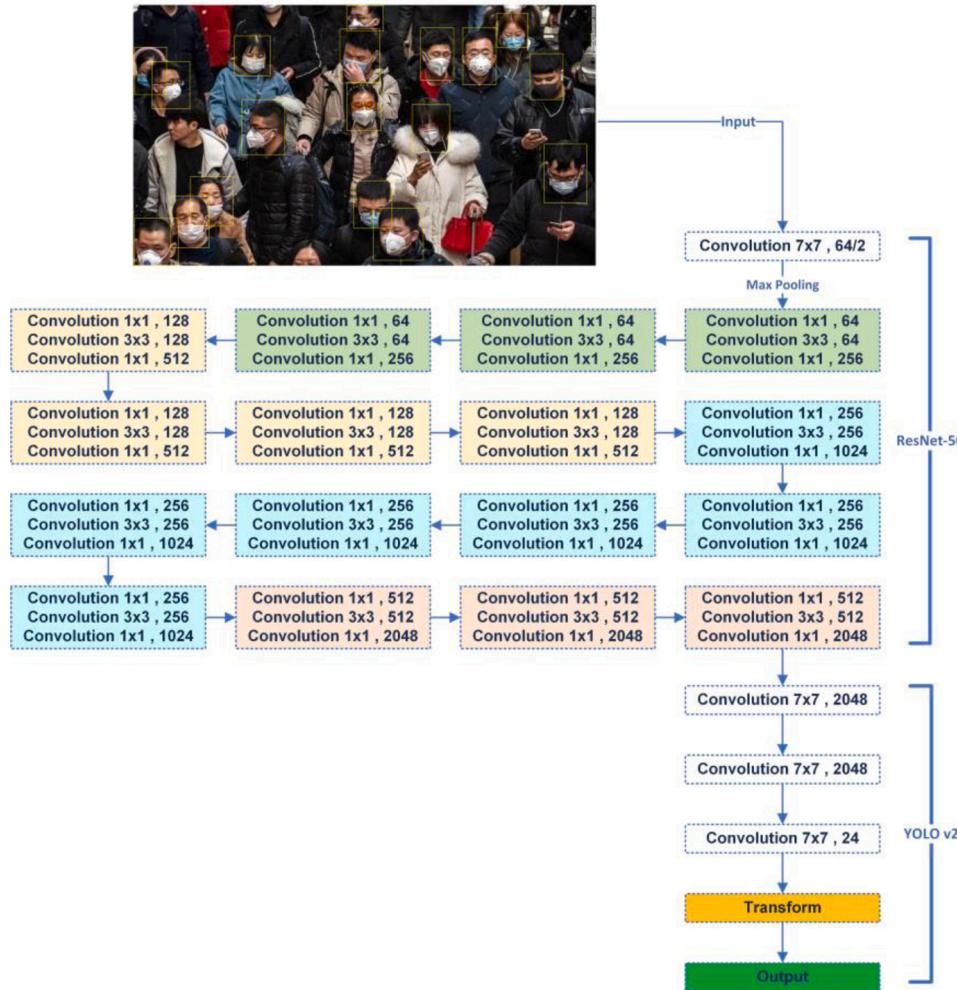


Fig. 8. Proposed detector based on YOLO v2 with ResNet-50.

Neural Network library (CuDNN) for GPU learning. The experiment configuration is presented in Table 1.

Dataset split up to 70% training images, 10% validation images, and 20% testing images. The configuration of YOLO v2 with ResNet-50 with initial learning rate (σ) and the number of epochs equal to 60 as illustrated in Table 1. The mini-batch size of the detector is set to 64. In terms of optimizer technique, Stochastic Gradient Descent with momentum (SGDM) (Sutskever, Martens, Dahl, & Hinton, 2013) and Adam (Kingma, 2015) is chosen to be our optimizer technique to improve detector performance.

Tables 2 and 3 illustrate the training process using SGDM and Adam. SGDM achieves less time consumption in training than Adam. From Tables 1 and 2, we conclude that SGDM is better than Adam on time and validation Root Mean Square Error (RMSE) and validation Loss. But Adam is better than SGDM in Mini-batch RMSE and Loss.

5.1. Detector Performance Evaluation

The performance measurement of the proposed detector model was quantitatively evaluating. The most common performance metrics in the

field of deep learning detection are average precision and log-average miss rates. The average precision (AP) (Padilla, Netto, & da Silva, 2020) combines recall and precision as shown in equation (6) to evaluate the ability of the detector to find all relevant objects and the ability of the detector to detect objects correctly. At all recall levels, the ideal precision is one. log-average miss rates are a mean of nine False Positives Per Image (FPPI) miss rate in the range of 10^{-3} to 10^0 to give stable performance (Dollar, Wojek, Schiele, & Perona, 2012; Dollar, Appel, Belongie, & Perona, 2014). As shown in Figs. 9 and 10, Adam optimizer AP = 0.81 is better than SGDM AP = 0.61 in all recall levels. Also, it can be clearly seen that Adam optimizer log-average miss rates = 0.4 is better than SGDM log-average miss rates = 0.6.

$$\text{average precision} = \frac{\sum \text{precision}}{\text{recall}} \quad (6)$$

Table 2
The training and validation process based on SGDM.

Epoch	Iteration	Time Elapsed	Mini-batch (RMSE)	Validation (RMSE)	Mini-batch Loss	Validation Loss
10	150	0:15:03	0.90	0.88	0.8187	0.7818
20	300	0:30:18	0.74	0.82	0.5484	0.6661
30	450	0:45:11	0.65	0.79	0.4240	0.6229
40	600	1:00:12	0.63	0.77	0.3967	0.5908
50	750	1:14:58	0.56	0.78	0.3083	0.6051
60	900	1:29:41	0.53	0.78	0.2808	0.6066

Table 3

The training and validation process based on Adam.

Epoch	Iteration	Time Elapsed	Mini-batch (RMSE)	Validation (RMSE)	Mini-batch Loss	Validation Loss
10	150	00:15:07	0.60	0.70	0.3584	0.4918
20	300	00:30:05	0.53	0.71	0.2817	0.4989
30	450	00:45:11	0.48	0.72	0.2265	0.5223
40	600	01:00:35	0.46	0.71	0.2147	0.5046
50	750	01:15:53	0.43	0.72	0.1878	0.5214
60	900	01:31:06	0.38	0.81	0.1416	0.6573

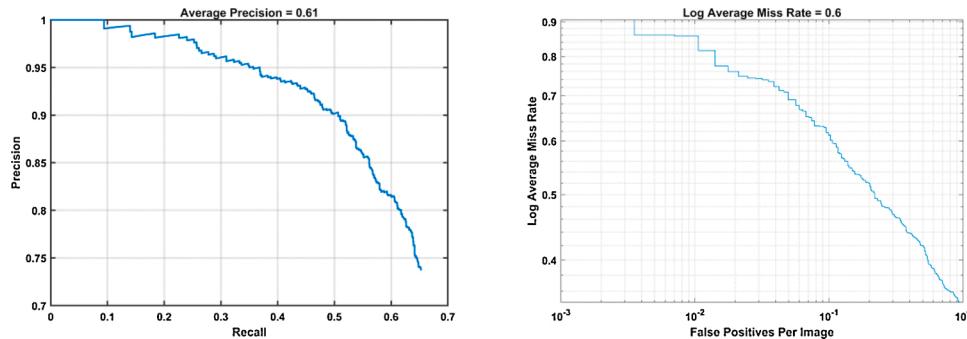


Fig. 9. Detector average precision and log-average miss rates based on SGDM.

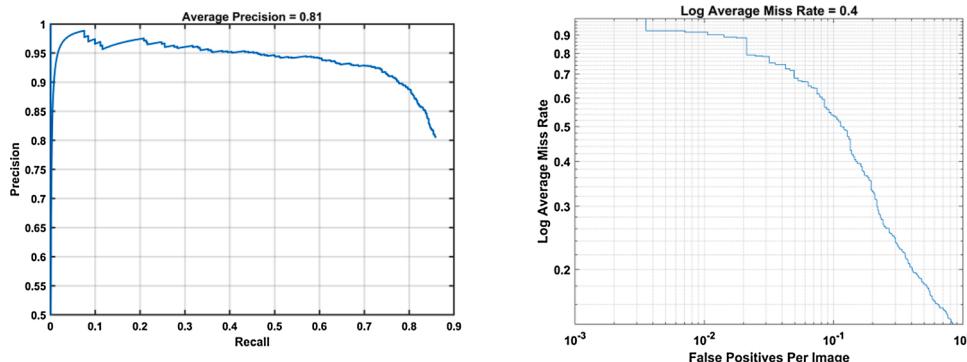


Fig. 10. Detector average precision and log-average miss rates based on Adam.

5.2. Comparison with related works and Discussion

The proposed detector model outcomes of using YOLO v2 with ResNet-50 models in MMD and FMD datasets are shown in Fig. 11. The YOLO v2 with the ResNet-50 model performs better in medical masked face detection, which introduced the effectiveness of the proposed model in medical masked face detection. We improve our model by using Adam optimizer. Most of related work focus on classification of mask face only. Performance comparison of different methods in term of Accuracy (AC), and Average Precision (AP) shown in Table 4. The related work presented in (Ge et al., 2017) used the MAFA dataset which includes the real masked face dataset. The authors of (Ge et al., 2017) achieved a testing average precision equal to 76.1% using LLE-CNN. The drawback of their MAFA dataset is not specified in the medical mask, it also detects any mask on the face as a medical mask. In the presented work, the average precision equal to 81% using YOLOv2 with ResNet-50 based on Adam optimizer. By analyzing the performance of YOLO v2 based on ResNet-50 in handling medical masked faces with the different optimizers, we find that the performance measurement of all masked face detectors increases sharply on masked faces with strong occlusions. Although our detector performs the best, its AP only reaches 81% based on Adam optimizer.

6. Conclusion and Future Works

In this work, we have introduced a novel model for medical masked face detection, focusing on medical mask object to prevent COVID-19 spreads from human to human. For image detection, we have employed the YOLO v2 based ResNet-50 model to produce high-performance outcomes. The proposed model improves detection performance by introducing mean IoU to estimate the best number of anchor boxes. To train and validate our detector in a supervised state, we design a new dataset based on two public masked face datasets. Furthermore, performance metrics such as AP and log-average miss rates score had been studied for SGDM and Adam optimizer experiments. We have shown that the proposed model scheme of YOLOv2 with ResNet-50 is an effective model to detect a medical masked face. As a future study, we plan to detect a kind of masked face in image and video-based on deep learning models.

Funding

This research received no external funding.

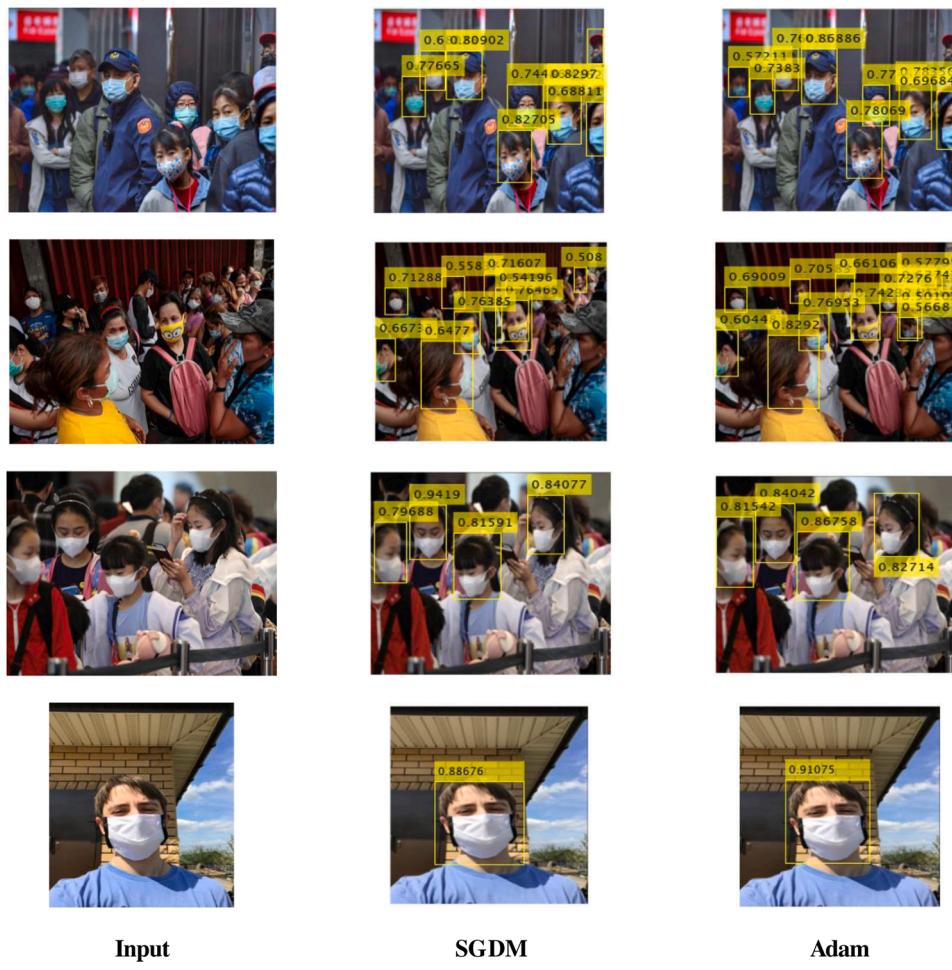


Fig. 11. Representative detector outcomes of our YOLO v2 with ResNet-50 model.

Table 4

Performance comparison of different methods in term of Accuracy (AC), and Average Precision (AP).

Reference	Methodology	Classification	Detection	Result
(Ejaz et al., 2019)	PCA	Yes	No	AC = 70%
(Ud Din et al., 2020)	GAN	Yes	Yes	—
(Loey et al., 2021)	hybrid	Yes	No	AC = 99.64%
(Ge et al., 2017)	LLE-CNNs	Yes	Yes	AP = 76.1%
Proposed	YOLOv2 with ResNet	Yes	Yes	AP = 81%

Declaration of Competing Interest

The authors report no declarations of interest.

References

- AbdElNabi, M. L. R., Wajeeh Jasim, M., EL-Bakry, H. M., Taha, M. Hamed N., & Khalifa, N. E. M. (2020). Breast and Colon Cancer Classification from Gene Expression Profiles Using Data Mining Techniques. *Symmetry*, 12(3). <https://doi.org/10.3390/sym12030408>. Art. no. 3, Mar.
- Altmann, D. M., Douek, D. C., & Boyton, R. J. (2020). What policy makers need to know about COVID-19 protective immunity. *The Lancet*, 395(10236), 1527–1529. [https://doi.org/10.1016/S0140-6736\(20\)30985-5](https://doi.org/10.1016/S0140-6736(20)30985-5)
- Barthakur, M., & Sarma, K. K. (2019). Semantic Segmentation using K-means Clustering and Deep Learning in Satellite Image. *2019 2nd International Conference on Innovations in Electronics, Signal Processing and Communication (IESC)*, 192–196. <https://doi.org/10.1109/IESPC.2019.8902391>
- Dollar, P., Wojek, C., Schiele, B., & Perona, P. (2012). Pedestrian Detection: An Evaluation of the State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4), 743–761. <https://doi.org/10.1109/TPAMI.2011.155>
- Dollar, P., Appel, R., Belongie, S., & Perona, P. (2014). Fast Feature Pyramids for Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(8), 1532–1545. <https://doi.org/10.1109/TPAMI.2014.2300479>
- Ejaz, Md. S., Islam, Md. R., Sifatullah, M., & Sarker, A. (2019). Implementation of Principal Component Analysis on Masked and Non-masked Face Recognition. *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 1–5. <https://doi.org/10.1109/ICASERT.2019.8934543>
- El-Sawy, A., EL-Bakry, H., & Loey, M. (2017). *CNN for Handwritten Arabic Digits Recognition Based on LeNet-5 BT - Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016*, Cham (pp. 566–575).
- El-Sawy, A., Loey, M., & EL-Bakry, H. (2017). Arabic Handwritten Characters Recognition Using Convolutional Neural Network Accessed: Apr. 01, 2020. [Online]. Available: <http://www.wseas.org/multimed/ia/journals/computerresearch/2017/a045818-075.php>.
- Ferguson, N., et al. (2020). “Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand,” Report, Mar <https://doi.org/10.25561/77482>
- Ge, S., Li, J., Ye, Q., & Luo, Z. (2017). Detecting Masked Faces in the Wild with LLE-CNNs. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 426–434. <https://doi.org/10.1109/CVPR.2017.53>
- Girshick, R. (2015). Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 580–587. <https://doi.org/10.1109/CVPR.2014.81>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Khalifa, N. E. M., Taha, M. H. N., & Hassanien, A. E. (2018). Aquarium Family Fish Species Identification System Using Deep Neural Networks. *Proceedings of the*

- International Conference on Advanced Intelligent Systems and Informatics 2018*, 347–356.
- Khalifa, N. E. M., Taha, M. H. N., Hassani, A. E., & Hemedan, A. A. (2019). Deep bacteria: robust deep learning data augmentation design for limited bacterial colony dataset. *International Journal of Reasoning-based Intelligent Systems*. <https://doi.org/10.1504/IJRIIS.2019.102610>
- Khalifa, N. E. M., Taha, M. H. N., Ezzat Ali, D., Slowik, A., & Hassani, A. E. (2020). Artificial Intelligence Technique for Gene Expression by Tumor RNA-Seq Data: A Novel Optimized Deep Learning Approach. *IEEE Access*. <https://doi.org/10.1109/access.2020.2970210>
- Kingma, D. P. (2015). *Adam: A Method for Stochastic Optimization* [Online]. Available: <http://arxiv.org/abs/1412.6980>.
- Lee, Y., Lee, C., Lee, H.-J., & Kim, J.-S. (2019). Fast Detection of Objects Using a YOLOv3 Network for a Vending Machine. *2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 132–136. <https://doi.org/10.1109/AICAS.2019.8771517>
- Liu, W., et al. (2016). SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016, Cham* (pp. 21–37). https://doi.org/10.1007/978-3-319-46448-0_2
- Loey, M., Smarandache, F., & Khalifa, N. E. M. (2020). Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning. *Symmetry*, 12(4). <https://doi.org/10.3390/sym12040651>. Art. no. 4, Apr.
- Loey, M., ElSawy, A., & Afify, M. (2020). Deep Learning in Plant Diseases Detection for Agricultural Crops: A Survey. *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)* (accessed April 11, 2020) www.igi-global.com/article/deep-learning-in-plant-diseases-detection-for-agricultural-crops/248499.
- Loey, M., Naman, M., & Zayed, H. (2020). Deep Transfer Learning in Diagnosing Leukemia in Blood Cells. *Computers*, 9(2). <https://doi.org/10.3390/computers9020029>. Art. no. 2, Jun.
- Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*, 167, 108288. <https://doi.org/10.1016/j.measurement.2020.108288>
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1150–1157. <https://doi.org/10.1109/ICCV.1999.790410>. vol.2.
- Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 237–242. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>
- Rational use of personal protective equipment (PPE) for coronavirus disease (COVID-19). Accessed: Jun. 02, 2020 [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/331498/WHO-2019-nCoV-IPC_PPE_use-2020-2-eng.pdf.
- Rahmani, A. M., & Mirmahaleh, S. Y. H. (2020). Coronavirus disease (COVID-19) prevention and treatment methods and effective parameters: A systematic literature review. *Sustainable Cities and Society*, 102568. <https://doi.org/10.1016/j.scs.2020.102568>
- Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6517–6525. <https://doi.org/10.1109/CVPR.2017.690>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- Sun, C., & Zhai, Z. (2020). The efficacy of social distance and ventilation effectiveness in preventing COVID-19 transmission. *Sustainable Cities and Society*, 62(Nov), 102390. <https://doi.org/10.1016/j.scs.2020.102390>
- Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). On the Importance of Initialization and Momentum in Deep Learning. *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28, Atlanta, GA, USA. III-1139-III-1147*.
- Ud Din, N., Javed, K., Bae, S., & Yi, J. (2020). A Novel GAN-Based Network for Unmasking of Masked Face. *IEEE Access*, 8, 44276–44287. <https://doi.org/10.1109/ACCESS.2020.2977386>
- WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int/> (accessed Jun. 02, 2020).
- Wu, X., Sahoo, D., & Hoi, S. C. H. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, 396, 39–64. <https://doi.org/10.1016/j.neucom.2020.01.085>