



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 6, June 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.512**



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

# Face Mask Detection System

Joby K. J.<sup>1</sup>, Priyanga K K<sup>2</sup>

P.G. Student, Department of Computer Science, Christ College (Autonomous), Irinjalakuda, Kerala, India<sup>1</sup>

Assistant Professor, Department of Computer Science, Christ College (Autonomous), Irinjalakuda, Kerala, India<sup>2</sup>

**ABSTRACT:** Coronavirus disease 2019 has affected the world seriously. One major protection method for people is to wear masks in public areas. Furthermore, many public service providers require customers to use the service only if they wear masks correctly. However, there are only a few research studies about face mask detection based on image analysis. In this paper, we propose RetinaFaceMask, which is a high-accuracy and efficient face mask detector. The proposed RetinaFaceMask is a one-stage detector, which consists of a feature pyramid network to fuse high-level semantic information with multiple feature maps, and a novel context attention module to focus on detecting face masks. In addition, we also propose a novel cross-class object removal algorithm to reject predictions with low confidences and the high intersection of union. Besides, we also explore the possibility of implementing RetinaFaceMask with a light-weighted neural network MobileNet for embedded or mobile devices. In this project, we propose a mask detector using image processing, the manual system automated. An IoT Component is connected to Display the Real time update of the Image. The technique used for segmentation and classification of images is Convolutional Neural Network.

**KEYWORDS:** Face Mask detection, Image Processing, Convolutional Neural Network, RetinaFaceMask.

## I. INTRODUCTION

The situation report 96 of world health organization (WHO) [1] presented that coronavirus disease 2019 (COVID-19) has globally infected over 2.7 million people and caused over 180,000 deaths. In addition, there are several similar largescale serious respiratory diseases, such as severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS), which occurred in the past few years [2] reported that the reproductive number of COVID-19 is higher compared to the SARS. Therefore, more and more people are concerned about their health, and public health is considered as the top priority for governments. Fortunately, studies showed that the surgical face masks could cut the spread of coronavirus. At the moment, WHO recommends that people should wear face masks if they have respiratory symptoms, or they are taking care of the people with symptoms. Furthermore, many public service providers require customers to use the service only if they wear masks [3]. Therefore, face mask detection has become a crucial computer vision task to help the global society, but research related to face mask detection is limited.

Face mask detection refers to detect whether a person wearing a mask or not and what is the location of the face [4]. The problem is closely related to general object detection to detect the classes of objects and face detection is to detect a particular class of objects, i.e. face. Applications of object and face detection can be found in many areas, such as autonomous driving, education, surveillance and so on. Traditional object detectors are usually based on handcrafted feature extractors, such as histogram of oriented gradients (HOG), scale-invariant feature transform (SIFT) and so on [5]. Recently, deep learning-based object detectors demonstrated excellent performance and dominate the development of modern object detectors. Without using prior knowledge for forming feature extractors, deep learning allows neural networks to learn features with an end-to-end manner. There are one-stage and two-stage deep learning based object detectors. One-stage detectors use a single neural network to detect objects, such as single shot detector (SSD) [6] and you only look once (YOLO). In contrast, two-stage detectors utilize two networks to perform a coarse-to-fine detection, such as region-based convolutional neural network (R-CNN) and faster R-CNN. Similarly, face detection adopts similar architecture as general object detector, but adds more face-related features, such as facial landmarks in RetinaFace [7], to improve face detection accuracy. However, there is rare research focusing on face mask detection.



## II. RELATED WORK

Traditional object detection uses a multi-step process [8]. A well-known detector is the Viola-Jones detector, which is able to achieve real-time detection [9]. The algorithm extracts feature by Haar feature descriptor with an integral image method, selects useful features, and detects objects through a cascaded detector. Although it utilizes integral image to facilitate the algorithm, it is still very computationally expensive. In human detection, an effective feature extractor called HOG is proposed, which computes the directions and magnitudes of oriented gradients over image cells. Later on, deformable part-based model (DPM) detects objects parts and then connects them to judge classes that objects belong to. Rather than using handcrafted features, deep learning based detector demonstrated excellent performance recently, due to its robustness and high feature extraction capability. There are two popular categories, one-stage object detectors and two-stage object detectors. Two-stage detector generates region proposals in the first stage and then fine-tune these proposals in the second stage. The two-stage detector can provide high detection performance but with low speed. The seminal work R-CNN is proposed by R. Girshick et al. [10]. R-CNN uses selective search to propose some candidate regions which may contain objects. After that, the proposals are fed into a CNN model to extract features, and a support vector machine (SVM) is used to recognize classes of objects. Finally, a region proposal network (RPN) is proposed in faster R-CNN to take the place of selective search, which limits the speed of such detectors. Faster R-CNN integrates each individual detection components, such as region proposal, feature extractor, detector into an end-to-end neural network architecture. One-stage detector utilizes only a single neural network to detect objects. In order to achieve this, some anchor boxes which specifies the ratio of width and heights of objects should be predefined. Rather than the two-stage detector, one-stage detectors sacrifice the performance slightly to improve the detection speed significantly. In order to achieve the goal, YOLO divided the image into several cells and then tried to match the anchor boxes to objects for each cell, but this approach is not good for small objects [11]. The researchers found that one-stage detector does not perform well by using the last feature output only, because the last feature map has fixed receptive fields, which can only observe certain areas on original images. Therefore, multi-scale detection has been introduced in SSD, which conducts detection on several feature maps to allow to detect faces in different sizes. Later on, in order to improve detection accuracy, Retina Network (RetinaNet) by combining an SSD and FPN architecture, which also include a novel focal loss function to mitigate class imbalance problem.

Convolutional Neural Network plays an important role in computer vision related pattern recognition tasks, because of its superior spatial feature extraction capability and fewer computation cost [12]. CNN uses convolution kernels to convolve with the original images or feature maps to extract higher-level features. However, how to design better convolutional neural network architectures still remains as an opening question. Inception network the network to learn the best combination of kernels.

## III. METHODOLOGY

In this paper, we proposed a novel face mask detector, RetinaFaceMask, which is able to detect face masks and contribute to public healthcare. To the best of our knowledge, RetinaFaceMask is one of the first dedicated face mask detectors. In terms of the network architecture, RetinaFaceMask uses multiple feature maps and then utilizes feature pyramid network (FPN) to fuse the high-level semantic information. To achieve better detection, we propose a context attention detection head and a cross-class object removal algorithm to enhance the detection ability. Furthermore, since the face mask dataset is a relatively small dataset where features may be hard to extract, we use transfer learning to transfer the learned kernels from networks trained for a similar face detection task on an extensive dataset.

Deep learning Technique is used in our system with CNN. Convolutional Neural networks is used for image extraction and an IoT device is connected to send the messages.

- Convolutional Neural Network
- Arduino (microcontroller (Nodemce)) - Display Unit.

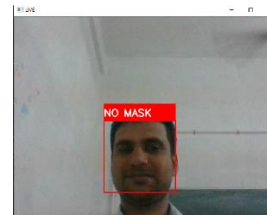
## IV. EXPERIMENTAL RESULTS

Figures show the results of face mask detection from a video and the message is sent to Arduino connected mobile device. As a result a message is sent to the mobile device using Arduino device whether the person in the video has a

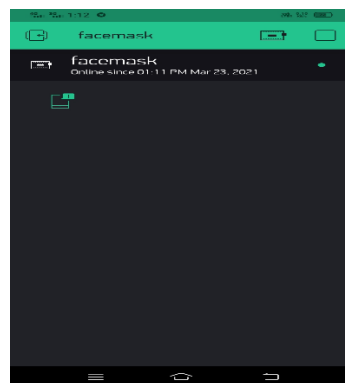
mask on his or her face or not. (a) shows the live image with mask. (b) live image without mask which is indicated that there is no mask. In (c) message display unit in the mobile and in (d) show that there is a message saying that non masked person detected.



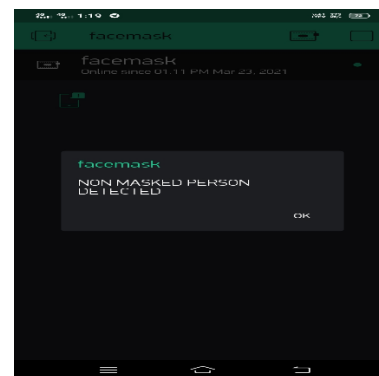
(a)



(b)



(c)



(d)

## V. CONCLUSION

In this paper, we have proposed a novel face mask detector, namely Facemask Detection System, which can possibly contribute to public healthcare. The architecture of Facemask Detection System consists of CNN as backbone, FPN as the neck, and context attention modules as the heads. The strong backbone ResNet and light backbone MobileNet can be used for high and low computation scenarios, respectively. In order to extract more robust features, we utilize transfer learning to adopt weights from a similar task face detection, which is trained on a very large dataset. Furthermore, we have proposed a novel context attention head module to focus on the face and mask features, and a novel algorithm object removal cross class, i.e. ORCC, to remove objects with lower confidence and higher IoU. The proposed method achieves state-of-the-art results on a public face mask dataset. An IoT component is added to the System to display the messages in a hardware. The data from the program is collected by the IoT hardware and it will be displayed as the warning signal to the concerned authority.

## REFERENCES

- [1] W. H. Organization et al., "Coronavirus disease 2019 (covid-19): situation report, 96," 2020.
- [2] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of covid-19 is higher compared to sars coronavirus," Journal of travel medicine, 2020.
- [3] Y. Nie, and M. Penny, "Transmission dynamics of the covid-19 outbreak and effectiveness of government interventions: A data-driven analysis," Journal of medical virology, vol. 92, no. 6, pp. 645–659, 2020.
- [4] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei et al., "Masked face recognition dataset and application," arXiv preprint arXiv:2003.09093, 2020.





- [5] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008, pp. 1–8.
- [6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in European conference on computer vision. Springer, 2016, pp. 21–37.
- [7] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-stage dense face localisation in the wild," arXiv preprint arXiv:1905.00641, 2019.
- [8] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," arXiv preprint arXiv:1905.05055, 2019.
- [9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol. 1. IEEE, 2001, pp. I–I.
- [10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.
- [11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
- [12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009, pp. 248–255.



**Impact Factor:  
7.512**



# **INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH**

**IN SCIENCE | ENGINEERING | TECHNOLOGY**

**9940 572 462 6381 907 438 ijirset@gmail.com**



**www.ijirset.com**

Scan to save the contact details