Ilkogretim Online - Elementary Education Online, 2020; Vol 19 (Issue 2): pp. 2158-2165
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2020.02.696800

Proper Face Mask Detection Using Deep Learning

Shanmughapriya M[1] , Brindha Devi V[2] , Josephine Ruth Fenitha , Sanchana R[4]

[1,3,4]Assistant Professor , [2]Professor

[1,2,3,4] Department of Information Technology, Sri Sai Ram Institute of Technology

Abstract

The COVID19 pandemic and the variants of COVID19 have affected millions of people globally. With no effective cure, prevention is the only solution to slow down the rapid transmission of the virus. Face masks play a vital role in prevention of the transmission. Real time detection of proper face masks is crucial in this time. The proposed system helps in identifying proper face mask automatically compared to the manual detection. We propose a Localized YOLO algorithm in this paper which identifies the proper face masks using following two criteria, 1) Nose, Mouth and Chin is fully covered 2) Wearing any other object to cover face is detected. The DarkNET is used to focus more on the important features of the image. The spatial difference between the prediction and ground truth boxes are lowered in this proposed method for increased precision in prediction by adopting the GIoU loss method. We used the Kaggle Face Mask detection dataset containing 853 images, the data set is processed for three different categories namely 1)Without Mask, 2)With Proper Mask, 3)With Incorrect Mask. The proposed Localized YOLO has good prediction performance. The proposed algorithm can be used in real time over a video to detect people with incorrect face masks.

1. Introduction

COVID19 break out in the year 2019 has affected tens of millions of people globally. As of December 12, 2021 268,549,864 People have been infected worldwide and the total death due to the virus is total deaths is 5,287,534. The vaccinations help in slowing down the transmission of the virus, nearly 8,324,380,164 people have been vaccinated at least with one dose [1]. With new Omicron Variant of COVID19 [2] the threats remain challenging to control, thetransmissibility rate of the new variant pose a major threat for re infection and new infection. WHO recommends proven public health measures such as well fitted face masks, hand hygiene, avoiding crowd spaces and getting vaccinated to reduce the risk of transmission and infection. At present people have been advised to wear masks properly in public places through advertisements and each organization recommends employees to wear masks, specially educational institutions needs to be vigilant in ensuring students to wear masks properly at all times, manual intervention to check proper wearing of mask is in effect. We propose a automated method to detect proper wearing of face mask.

Computer Vision is a inter disciplinary approach that enables a system to extract meaningful information from Digital Images, Video and other forms of Visual Input. Object Detection plays a vital role in computer vision by ensuring the scene and context understanding of the digital image [3]. Face Mask is an object which can be detected by computer vision techniques. Mask detection in a crowded environment plays a vital role, computer vision paves a way to do it automatically.

2. Object Detection Techniques – A Review

Object detection is a major field of research in computer vision, Face Detection, Pedestrian Detection, Vehicle Detection, Sign Detection has been widely researched. Face Mask Detection isthe new hotspot since the outbreak of the pandemic. Wang et al[4] surveys the comprehensive review of new developments in Face Detection.

Object Detection involves taking a digital image as input and locating the presence of the object by drawing a bounding box around the object, then come the process of classifying the objects in the bounding box. The following deep learning models excel in Object Detection.

YOLO – You Only Look Once

Joseph Redmon et al[5] proposed seeing object detection as a regression problem and illustrated a new model YOLO – You Only Look Once which predict class probabilities and bounding boxes by learning the general representation of objects in a single neural network there by increasing speed and precision. YOLO has been evolved over the years since its first publication in 2016, YOLO9000 in 2016 can detect upto 9000 objects with efficient precision. In

[6] Joseph Redmon et al, claimed that YOLOv2 that is YOLO9000 gets 78.6 mAP, outperforming state-of-the-art methods like Faster RCNN with ResNet and SSD.

SSD – Single Shot Detector

Wei Liu et al [7] proposed a single deep neural network which discretizes the output space and combines multiple feature maps with varying resolution to produce precise bounding boxes. The SSD completely combines all computations in a single network, thereby making SSD easy to implement. SSD300 achieves 74.3% mAP at 59 FPS while SSD500 achieves 76.9% mAP at 22 FPS. SSD also helps in detecting smaller objects, SSD has the ability to handle lower aspect ratio precisely compared to YOLO.

RCNN - Region Based Convolutional Neural Networks

Ross Girshick[8] proposed a two stage object detection technique, which involves selective Search to extract Region of Interest(ROI) which is then fed to the neural network for object detection. The RCNN detects objects present in the rectangular ROI. RCNN has evolved over the years with variations such as Fast RCNN, Mesh RCNN and Mask RCNN. ROI Pooling is used in Fast RCNN which uses single neural network to ROI identification and object detection. Google Lens[9] uses the RCNN for Object Detection.

3. Proposed System

The proposed system uses the underlying yolov3 model to find the face masks, but after detecting the face masks the localized features of the bounding box is in turn taken into account for detecting the three categories of mask wearers 1)Without Mask, 2)With Proper Mask, 3)With Incorrect Mask

Implementation

YOLOv3 uses DarkNET53 which uses 53 convolution layers of network trained on Image NET.Image NET [10] is an image data base containing thousands of images organized in a word net hierarchy containing only the nouns. Millions of images are annotated by hand for object and is categorized in Image NET. The proposed method uses a K Means clustering algorithm to generate anchor boxes which reduces the burden on learning difficulty. The 9 anchor nodes are generated P0 to P8 which constitutes of 10X10, 20X20 and 40X40 each with three anchor nodes grouped. The features inside the bounding box play a vital role in determining the correctness of the mask in the face. Thus after detecting the mask the information of the bounding box is then applied to detect presence of nose, mouth and chin in the bounding box area. The visible presence of these features indicates that the mask is not worn properly.

The LOSS function in YOLO is formulated (1) as below

$$LOSS_{total} = LOSS_{Bounding\ Box} + LOSS_{Confidence} + LOSS_{Clas} \quad (1)$$

The loss due to bounding box will affect the scale of targets; the outlying features may affect theprediction model. GloU Loss function is used; it is defined as equation (2)

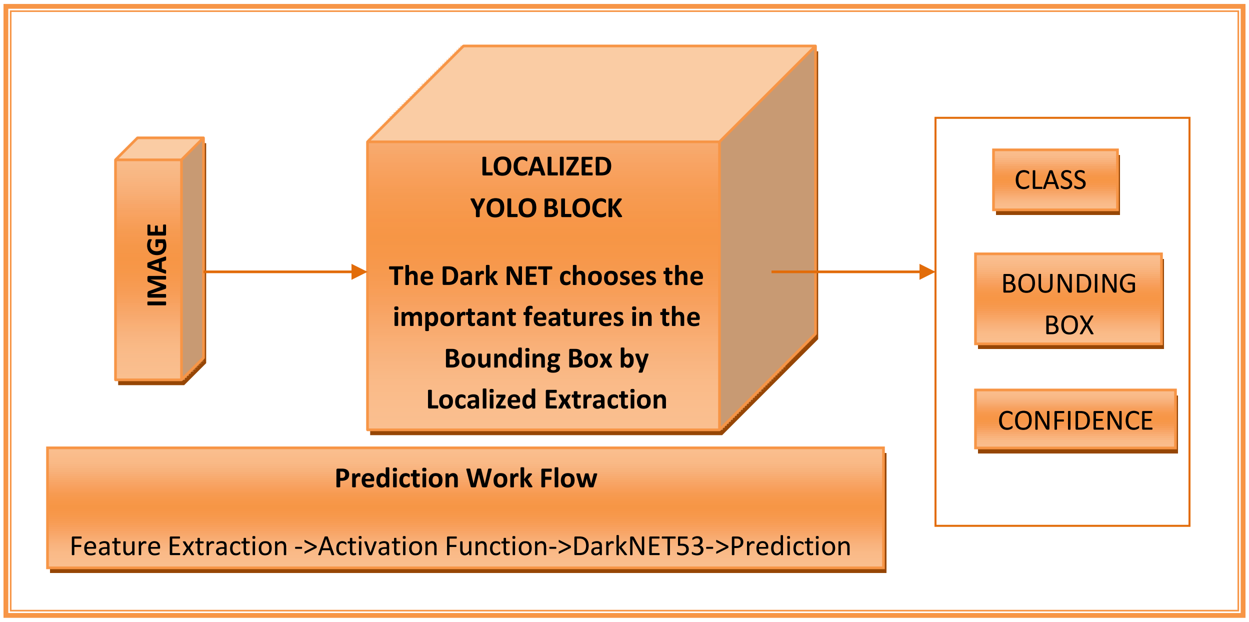
$GloU\ Loss$

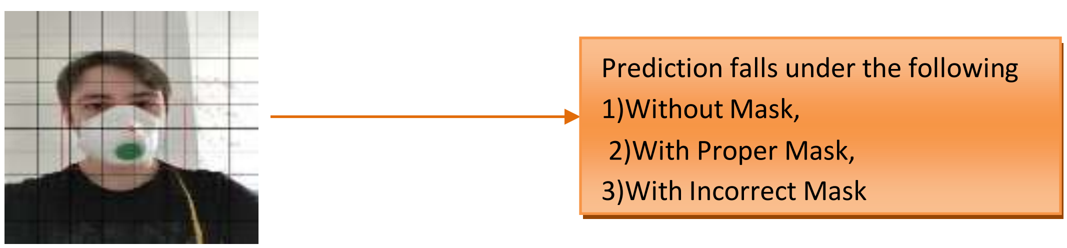
$area\ (predicted\ \cap labeled\ value)$

$$= \frac{}{area\,(predicted \cup labled\ value)}$$

$$- \frac{area(Minumum\ Bounding\ Box\ of\ Predicted\ and\ Labeled) - area\ (predicted\ \cap labeled\ value)}{area(Minumum\ Bounding\ Box\ of\ Predicted\ and\ Labeled)}$$

The GloU loss function helps in achieving maximum precision by eliminating outliers.





Grid Cells and Ground Truth Bounding Box Red dot in picture indicates the Object Center

Figure 1: Implementation Architecture of Proper Face Mask Detection

The distribution of labels in the data set is shown in figure 2. Three colored bounding boxes are used for differentiating three different classes of detection, red is used for No Mask, Green is used for Proper Mask and yellow is used for Incorrect Mask. The anchor nodes helps in selecting the random images for each training epoch.

As the layers of the network increases, the convergence occuring in the network will cause degradation. The precision gets exhausted. Training error occur in deep neural models, degradation and exhaustion indicates that system is not easily optimized. In DARKNet53, the layers are explicitly made to fit the anchor nodes found using K Means, thereby decreasing the precision errors considerably.

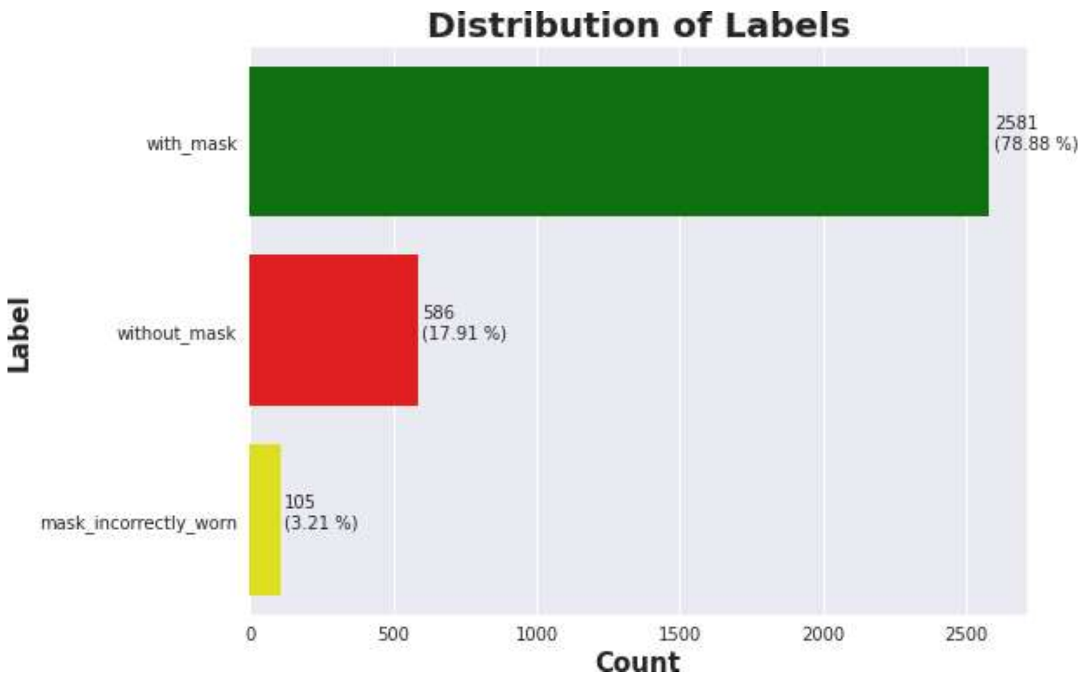


Figure 2: Distribution of Labels in the Data Set

Experimental Results and Discussion

The experiments were conducted in a tensor flow deep learning framework, in google colab environment. The initial learning rate is set at $4 \text{ X } e^{-5}$ gradually the learning rate is increased to $1 \text{ X } e^{-1}$. The Kaggle data set[11] has 853 images with all three categories of images 1)Without Mask, 2)With Proper Mask, 3)With Incorrect mask. The dataset is split into training data containing 550 images and to test the

remaining 303 images are used. The figure shows the3 different outputs.

Figure 3: Output with all three classes

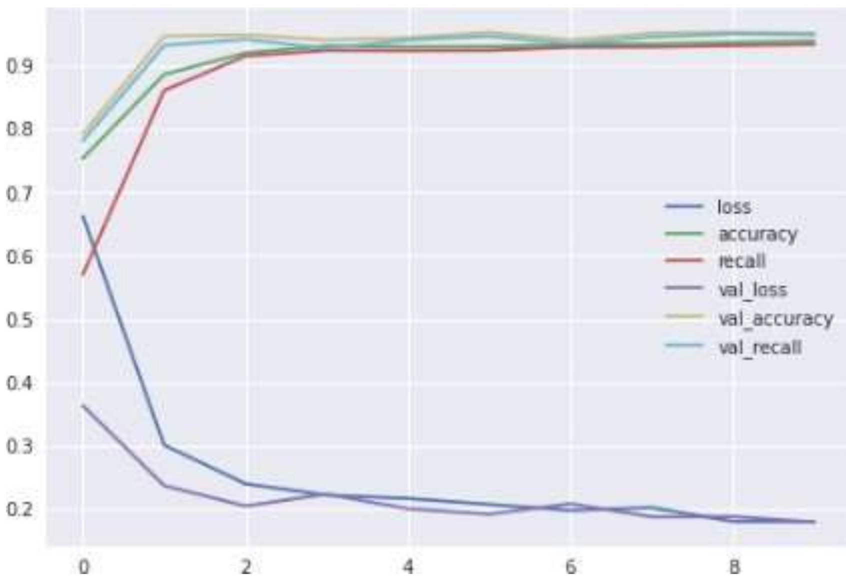


Figure 4: Loss and Accuracy plot in Anchor Nodes 0 to 9

Future Work and Applications

Ranjith et al [12] suggesting a way to recognize action from surveillance videos, the similar approach can be used for surveillance of incorrect mask wearers in a public domain. Jiang at al [13] recommends a real time system where entry can be restricted on accounts of incorrect or no mask classes. Figure 5 represents an application architecture based on [13],which uses a camera and a Raspberry pi, Arduino controlled door. The camera captures the video and from the video automated mask detection is done, if the mask is properly worn then the door is opened else a notification is shown. This application is theoretically concluded for implementation in schools and colleges.

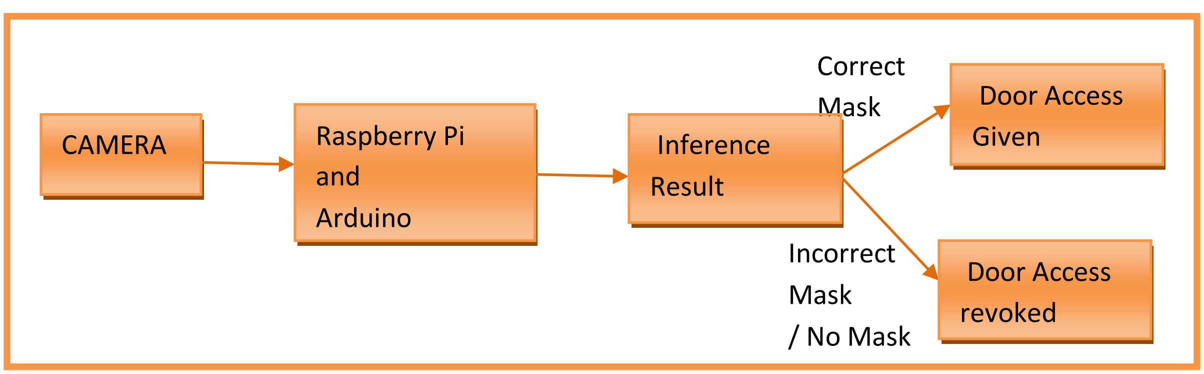


Figure5: Proposed Application Architecture

4. Conclusion

The COVID19 pandemic is challenging to control, the recommended precautions [14] suggests that the mask plays a vital role in controlling the transmission. A system which detects mask automatically will help in reducing the manual labor. The proposed application architecture figure 5 can be implemented in organizations. There is always a room for optimizations in terms of data set annotations and prediction precision, which can contribute to safety measures.

5. References

[1]. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Available online: https://coronavirus.jhu.edu/map.html (accessed on 1 March 2021).

[2]. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern at WHO, World Health Organization, Available online https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern

[3]. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. ComputIntell. Neurosci. **2018**.

[4]. Wang, M.; Deng,W. Deep Face Recognition: A Survey. arXiv 2018, arXiv:1804.06655.

[5]. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real- Time Object Detection", Computer Vision and Pattern Recognition (cs.CV) https://arxiv.org/abs/1506.02640

[6]. Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger", Computer Vision and Pattern Recognition (cs.CV) https://arxiv.org/abs/1612.08242v1

[7]. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, "SSD: Single Shot MultiBox Detector", Computer Vision and Pattern Recognition (cs.CV) arXiv:1512.02325, https://arxiv.org/abs/1512.02325

[8]. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", Computer Vision and Pattern Recognition (cs.CV),arXiv:1311.2524, https://arxiv.org/abs/1311.2524

[9]. Sagar, Ram (Sep 9, 2019). "These machine learning methods make google lens a success".

Analytics India. Retrieved Mar 28, 2020.

[10]. Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015. https://arxiv.org/abs/1409.0575

[11]. Kaggle Face Mask Detection data set available onlinehttps://www.kaggle.com/andrewmvd/face-mask-detection

[12]. Ranjith Balakrishnan, S. M. (2017). Anticipating Human Activities from Surveillance Videos. International Journal of Engineering and Computer Science, 6(4). Retrieved from https://www.ijecs.in/index.php/ijecs/article/view/3784

[13]. Jiang, X.; Gao, T.; Zhu, Z.; Zhao, Y. Real-Time Face Mask Detection Method Based on

YOLOv3.
Electronics 2021, 10, 837. https://doi.org/10.3390/electronics10070837
[14]. Peijie Chen, Lijuan Mao,George P. Nassis,Peter Harmer,Barbara E. Ainsworth,and Fuzhong Lif " Coronavirus disease (COVID-19): The need to maintain regular physical activity while taking precautions", Journal of Sport and Health Science Volume 9, Issue 2, March 2020, Pages 103- 104, https://doi.org/10.1016/j.jshs.2020.02.001