Comparative Analysis of Convolutional Neural Network Architectures for Real Time COVID-19 Facial Mask Detection

Akash Manish Lad¹, Anubhav Mishra² and Arul Rajagopalan³, *

¹,²,³Vellore Institute of Technology, Chennai Tamil Nadu, India

*Corresponding author: arulphd@yahoo.co.in

Abstract. The late 2019 outbreak of Coronavirus Disease (COVID-19) had an indelible imprint on the humanity. The world is recovering from the outbreak but there is danger of a second wave of the outbreak. To get rid of the outbreak it is necessary to prevent the viral transmission and it is need of the hour to maintain social distancing and wear masks in public areas. The governments are providing strict guidelines to wear masks in public places. It is not manually feasible to check if people are wearing masks or not. In this paper, process of detecting face masks in public places is automated using Convolutional Neural Networks by performing comparative analysis on Sequential bi-layered CNN, VGG-16 CNN and MobileNetV2 CNN architectures. Among these three architectures MobileNetV2 outperformed with a performance accuracy of 99.2%. The efficient Deep Learning architecture of detecting face masks can be achieved with the help of IoT (Internet of Things) devices and cameras, of those who are not following guidelines in public places. Such a system is very useful in post outbreak period and can be installed in public places such as Railway Stations, Airports, Parks, Schools, colleges, offices etc. to track and ensure wearing of masks by people. The contribution of this paper is not to reel-off the finding from the original paper on Face Mask detection with various architectures rather to provide results on the efficiency of using the MobileNetV2 architecture in comparison with Sequential CNN and VGG-16 architectures for crowd analysis mask detection.

Keywords: COVID-19, Sequential, VGG-16, MobileNetV2, Face masks, Convolutional, Neural Networks, Comparative analysis, deep learning

1. Introduction

The COVID-19 pandemic has witnessed significant number of cases in past 7-8 months. This trend is increasing with days to come and every possible measure is taken to control of the spread the virus. Governments have imposed norms to maintain social distancing and increased usage of face masks in public places with simultaneous measures to sanitize workplaces and surroundings. Earlier, face masks were used to prevent ourselves from air pollution and dust in the surroundings. This trend has increased in the past 6 months due to the viral face-face spread.

It becomes important to develop a strong health infrastructure to combat the outbreak of this virus. An increased attempt is being done in the field of medicine to facilitate vaccine in the public domain after numerous trails of conduction. Ventilators, infrared thermometers, masks, sanitizers, hand-washing soaps and Personal Protective Equipment (PPE) kits is the ultimate need of the hour. Doctors and nurses
are highly indulged in taking precautionary measures of the patient and providing optimal treatment becomes of vital importance.

Developing tools and techniques to combat this virus can help significantly reduce the numbers of the patients suffering from COVID-19. This can be achieved with the help of technology by providing assistance for controlling the spread of the viral transmission. In this paper, an efficient approach is provided for controlling the outbreak using Deep Learning techniques like Convolutional Neural Networks for the detection of face mask in public places by analysis various CNN architectures and employing the best neural network for the detection of mask. This would help in tracking the people not wearing masks, thereby protecting the spread among other people by enforcing guideline measures for safety.

The system is built using an IoT (Internet of Things) device, Raspberry Pi model 3B+ along with camera for detection of face mask. The mask detection is achieved using OpenCV, tensorflow, keras libraries of python over which the neural networks were developed by providing training data and test data for image detection. The training and testing accuracies were analyzed among Sequential bi-layered CNN, VGG-16 CNN and MobileNetV2 CNN architectures using graphical libraries of python such as Matplotlib and plotly.

2. Literature Survey

The importance of Convolutional Neural Networks is discussed in many contexts with respect to image recognition, object detection, image segmentation etc. There are significant numbers of Deep Convolutional Neural Networks available, but it depends on the application of implementation. In case of Mask detection systems, there are plenty of neural networks-based papers which have been published. But most of the works rule out the comparative study on the efficiency of various architectures of CNN and often perform superficial work on the performance of Convolutional Neural Network. In this paper effective mask detection and data analysis on various architectures is performed to obtain the insight of hidden parameters of performance (training accuracy, training loss, execution time etc.) among the sequential, VGG-16 and MobileNetV2 architectures.

So far, different handcrafted and deep learning feature-based approaches have been proposed. In recent days, deep learning approaches for Face-Mask detection is gaining popularity because of better accurate results of pre-trained models. Initially, the images from the camera-feed are extracted and segmented by marking with bounded boxes depicting confidence levels around person the wearing or not wearing a mask. There are many feature engineering techniques available for mask detection and are similar to the face detection systems. The mask detection is similar to face detection, the training dataset consists of images with masks and without masks. This is a simple classification problem that is addressed with the help of deep learning techniques. Shashi Yadav [1] proposed deep learning-based face mask detection for COVID-19 safety using MobileNetV2 which trained to precision of 91.7%. Bryan Knessis [2] highlights the features of Mask RCNN architecture for crowd analysis with labelled faces in the wild, however it does not highlight the importance if the facial features are manipulated, that is, in case of mask-detection system the region of interest is comparatively less and requires more knowledge of existing dataset. There are drawbacks in existing architectures, the need of the hour is to analyze important and efficient state-of-the-art architectures to bring out the best results with minimum error.
3. Proposed Approach

3.1 Object Detection:

Object detection models are generally being used so far for detecting objects in images. Same model can also be used for detecting multiple faces in images and later classifying as masked or unmasked. Traditional methods such as Viola and Jones (HAAR Cascade) and HOG is used to detect objects such as faces in images [3]. But these methods require a lot of feature engineering tasks. With the help of deep learning techniques such overload can be reduced and we do not require lots of feature engineering.

3.2 Convolutional Neural Networks:

CNN is very crucial these days for computer vision related tasks such as detecting patterns in image. It's possible due to excellent feature extraction ability. Convolution kernels are used to convolve images and extract features from it makes it highly accurate for pattern recognition tasks [4]. Fig 1. represents basic CNN architecture.

![Basic Convolutional Neural Network Architecture](image)

**Fig 1. Basic Convolutional Neural Network Architecture**

3.3 Data collection and pre-processing:

Three different CNN architectures are used to implement face mask detection. For this task dataset images having masked and unmasked faces were collected from the custom-built Face Mask detection dataset. In the data pre-processing step, each image is read and vectorized. After that each image is resized (100 X 100) and images are converted to greyscale. Fig 2. represents classification of images into ‘mask’ and ‘no mask’ respectively.
3.4 Training of models for classification:

For classification of masked and unmasked images three CNN classifiers were trained. Before training, the model data is filtered and augmented which reduces the size of the dataset by rotating and flipping each image. We split our dataset into training set and testing set. With training set, CNN model is trained upon to distinguish the masked and unmasked faces whereas the test set is used for testing the neural network. The dataset is split size in the ratio of 80:20, signifying that the training set consists of 80% of the dataset and the testing set contains 20% of the dataset.

3.5 Sequential CNN model:

The Sequential CNN model consists of layers such as Conv2D, MaxPooling2D, Flatten, Dropout and Dense. To get the probability of each class SoftMax function is used in the Dense layer [5]. The Sequential CNN model is trained with 20 epochs (iteration) which gives us training accuracy of 0.9899 with loss of 0.0346 beyond which the accuracy changes due to increased training time and based upon parameter values. Fig 3. represents sequential CNN architecture.

![Sequential CNN Architecture Diagram](image-url)
### 3.6 VGG-16 CNN architecture:

Before training with VGG16 CNN we used ImageDataGenerator for data augmentation. Later VGG16 CNN is trained which has 17 convolution layers and 5 Max Pooling layers [6]. The model is trained upon the training dataset and gives accuracy of 0.9427 with loss of 0.1436. Fig 4. represents VGG-16 architecture.

![VGG-16 Architecture Diagram](image)

**Fig 4.** VGG-16 Architecture Diagram

### 3.7 MobileNetV2 CNN architecture:

MobileNetV2 architecture has two blocks of stride 1 and 2 for residual block and downsizing [3]. Each block consists of 3 layers such as 1x1 convolution with ReLU6, depth wise convolution and 1x1 convolution without non-linearity [4]. Model is trained for 20 epochs and gives accuracy of 0.9922 with loss of 0.0282 for the training set. Fig. 5 represents the MobileNetV2 architecture. Fig 5. represents MobileNetV2 architecture.

![MobileNetV2 Architecture Diagram](image)

**Fig 5.** MobileNetV2 Architecture Diagram
4. Implementation

The Face mask detection system is designed using Raspberry pi (IoT device) with peripherals including HDMI cable, Keyboard, mouse, a 5MP Pi-Camera, 16 GB SD card for setting up Raspbian Operating System [7]. It has a 1 Giga-Byte of RAM, wireless LAN, Bluetooth and many more features.

The Python code for Image tracking and detection is executed successfully. The images closer to the camera gets detected early and on comparing with the training dataset, it segments a square boundary around the face [8] with red color if the mask is not detected otherwise with green color if the mask gets detected along with the percentage accuracy of detection. Fig 6. represents Raspberry-pi hardware.

4.1 Face Mask Detection:

The python code after successful execution, pops up the camera using the camera driver from the Raspberry Pi. It then performs computations for around 10-12 seconds and detects images with mask and without mask by providing appropriate levels. Fig. 7 is an output screenshot after running the MobileNetV2 architecture which performed better in comparison with Sequential CNN and VGG-16 model. Fig 7 a). represents mask detection for single person. Fig 7 b). represents mask detection for 3 people.

Fig 7 a). Mask Detection of a person in a frame using 2-MP Raspberry-pi Camera
Fig 7 b). Mask Detection of 3-people in a frame using 2-MP Raspberry-pi Camera

Fig. 8 is an output screenshot after running the video as input fed to the system for detection of mask. The video represents a busy public place, and the mask detection segmented closer-by images with accuracies greater than 91%.

Fig 8. Mask Detection in a public place with video-input on MobileNetV2 architecture
5. Comparative Analysis of CNN architecture

Fig 9. represents comparative study for Sequential CNN, VGG-16 and MobileNetV2 architectures with performance measures.

|                | Sequential CNN | VGG16 CNN  | MobileNetV2 CNN |
|----------------|----------------|------------|-----------------|
| Epochs         | 20             | 20         | 20              |
| Training Accuracy | 98.9%         | 94.2%      | 99.2%           |
| Training Loss  | 3.4%           | 14.3%      | 2.8%            |
| Training Execution Time | 1566.5250945091248 seconds | 1485.85455252552 seconds | 433.84504747390747 seconds |
| Loss Vs Epochs |                |            |                 |
| Accuracy Vs Epochs |              |            |                 |

Fig 9. Comparison of Sequential, VGG-16 and MobileNetV2 CNN architecture

6. Conclusion

The Face Mask Detection have been carried out on three Convolutional Neural Network architectures which include the sequential CNN, VGG-16 CNN and MobileNetV2 CNN. The Face mask detection was performed on 1-GigaByte RAM, Cortex-A53(ARM-v8) 64 bit Soc @1.4GHz and 5GHz IEEE 802.11.b/g/n/ac wireless LAN Raspberry pi Model 3B+ (IoT device). Performing a comparative performance analysis of three architectures, it was inferred that MobileNetV2 architecture performed significantly better based on the accuracies, than other two architectures tested, with an accuracy of 99.2%. Whereas, the sequential CNN was trained to a maximum accuracy of 98.9% and the VGG-16 Convolutional Neural Network architecture achieved a maximum accuracy of 94.2%. These deep learning models perform complex training computations for mask detection between masked and unmasked face on the 1-GigaByte hardware infrastructure of Raspberry pi Model 3B+ with a total training time of 20 epochs till where the maximum accuracy was achieved, it was observed that
MobileNetV2 architecture outperformed sequential CNN and VGG-16 with a higher accuracy of 99.2%. The pre-trained MobileNetV2 architecture performance and training accuracies can be improved for faster and efficient Face-Mask detection using Neural Magic Inference Engine software on an Intel x86 CPU by increasing the batch size. In this paper we have selected MobileNetV2, VGG-16 along with simple sequential CNN model to compare their accuracy. MobileNetV2 performance was analyzed as the model is faster processing across whole latency spectrum. Such models while maintaining same accuracy need very small number of operations, it requires 30% less parameters and it is found to be 30-40% quicker in phones. MobileNetV2 also found to be very accurate in feature extraction for detecting objects in images which is essential for detecting face masks, also it performs image segmentation very effectively. VGG-16 is very effective in bench-marking for some jobs and also available pre-trained. VGG-16 has more receptive field of a neurons. Simple sequential CNN and VGG-16 is also available publicly in free and very commonly used, all this drives us to select these architectures for analysis and comparison.

7. Acknowledgement

We would like to express our gratitude towards our parents who have motivated us to pursue a research project with sincerity and determination and would like to specially thank our institution Vellore Institute of Technology, Chennai campus for encouraging research-based projects.

8. References

[1] Shashi Yadav, “Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID19 Safety Guidelines Adherence”, Volume 8 Issue VII July 2020, doi: 10.22214/ijraset.2020.30560

[2] Kneis B. (2018) Face Detection for Crowd Analysis Using Deep Convolutional Neural Networks. In: Pimenidis E., Jayne C. (eds) Engineering Applications of Neural Networks. EANN 2018. Communications in Computer and Information Science, vol 893. Springer, Cham. https://doi.org/10.1007/978-3-319-98204-5_6

[3] F. Saxen, P. Werner, S. Handrich, E. Othman, L. Dinges and A. Al-Hamadi, "Face Attribute Detection with MobileNetV2 and NasNet-Mobile," 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 2019, pp. 176-180, doi: 10.1109/ISPA.2019.8868585.

[4] D. Chiang, “Detect faces and determine whether people are wearing mask,” https://github.com/AIZOOTech/FaceMaskDetection, 2020.

[5] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, “Object detection with deep learning: A review,” IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212–3232, 2019.

[6] H. Qassim, A. Verma and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018, pp. 169-175, doi: 10.1109/CCWC.2018.8301729.

[7] J. Marot and S. Bourennane, "Raspberry Pi for image processing education," 2017 25th European Signal Processing Conference (EUSIPCO), 2017, pp. 2364-2366, doi: 10.23919/EUSIPCO.2017.8081633.

[8] T. Meenpal, A. Balakrishnan and A. Verma, "Facial Mask Detection using Semantic Segmentation," 2019 4th International Conference on Computing, Communications and Security (ICCCS), Rome, Italy, 2019, pp. 1-5, doi: 10.1109/CCCS.2019.8888092.