High performance of CNN to Detect Face Masks Using Deep Learning

D. Aruna Kumari[1], N.Aparna[2], B.Jyothirmai[3], P.Rajendra[4],k.Bhaska Reddy[5],P .Krishna Rao[6]

[1,2,3,4,5,6] Assistant Professor, Department of ECE, Sri Indu Institute of Engineering & Technology, Hyderabad

Abstract: ***The COVID-19 pandemic redefined many aspects of our definition of normal life in which social distancing and wearing of face masks plays a vital role in controlling the spread of the virus. But many people are not adhering to the social precautions that were once essential of wearing face masks in public places which increases the spread of viruses. Hence to avoid such situations we must scrutinize and make people aware of wearing face masks. Manual real-time monitoring of face mask wearing for a large group of people is becoming difficult. Humans cannot be overly involved in this process, due to the chance of getting affected by the coronavirus. This can be combated using artificial intelligence and deep learning, which is the main theme of our study. The proposed approach is based on two steps. One step aims to create a DL model that detects facemasks and whether they are correctly worn. An online step that deploys the DL model at edge computing in order to detect masks in real-time. We have used a pre-trained CNN model of the VGG19 class for image classification and identification. A dataset from Kaggle will be used for both training and testing purposes. Overall training accuracy of over 95% is targeted.***

Keywords: Facemask detection, Real-time, COVID-19, Deep Learning, Convolutional Neural Network

1. Introduction

The novel coronavirus COVID-19 introduced a new standard of living. India is battling to recover from this virus outbreak, and the government has imposed a lengthy lockdown. Lockdown exerted strain on the international economy. Therefore, the administration eased the lockdown. WHO has declared that a prospective speech must be delivered while maintaining a distance and wearing a mask. The greatest help the government requires following a period of relaxation is social distance and the use of masks. However, many individuals are leaving their homes without a face mask, which may accelerate the spread of COVID-19. The Economic Times India said, "A Survey Shows That 90% of Indians are Aware, Yet Only 44% Wear a Mask" This poll demonstrates that although people are informed, they do not wear masks owing to discomfort and carelessness. This could facilitate the spread of COVID-19 in public spaces.

Monitoring the facemask use of a big population in real time is getting increasingly difficult. Due to the labor required to safeguard crowded locations and verify the persons is covering his nose and mouth properly, manual monitoring is generally difficult to implement. Aside from financial and management issues, the main challenge would be the health element, as the subset of individuals will interact with countless individuals regularly, posing the possibility of them functioning inflection point; hence, we strive towards remove human component interaction.

Lately, deep learning (DL) is indeed utilized in a variety of disciplines to address numerous complex issues, yielding substantial achievements. DL enables the analysis and interpretation of enormous amounts of data with speed and precision. In this project, we propose a method that would aid in enforcing the face-mask ban and facilitating its monitoring using real-time recordings. This advised technology would assist institutions and business spaces in efficiently monitoring masks.

The proposed method offers precise ability to spot a mask use and if it is used in an acceptable manner is examined either way in actual time. To achieve it, comprehensive data set has been utilized via both community and our own datasets. Moreover, VGG19 is employed as a Deep Learning system for facial mask identification. The strategy does have the upper hand of being quick as well as suitable for devices, it gives good performance outcomes of object recognition. Recommended remedy may be applied at public CCTVs in the real world. Checkpoints to determine if people are obeying the regulations and wearing the proper attire. With this, the approach may be executed with minimal effort and resources.

2. Literature Survey

This section contains the most recent and pertinent scholarly papers on face-mask detection using DL models. In [1], earlier days face detection models are implemented using edge, line and center near features and patterns are recognized from those features. These approaches are used to find binary patterns locally. These approaches are very effective to deal with Gray-scale images and the computation effort required also very less.

This paper observes and introduces the Dense Convolutional Network (DenseNet), which connects each layer to every other layer in a feed-forward fashion. Whereas traditional convolutional networks with L layers have L connections- one between each layer and its subsequent layer-our network has L(L+1)/2 direct connections. For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers. DenseNets have several compelling advantages: they alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters. The researchers of [3], using Viola Jones Detector, proposed a real time object model used to detect different classes of objects. It uses 24x24 base window size to evaluate any image with edge, line and four rectangular features. Harr-like features are like convolutions to check weather given feature is available in the image or not. This model fails to work in when image brightness varies even it exhibits poor performance when images are in different orientations. segmentation method was used to detect facial mask in this paper they have used VGG net for training and FCN is used to semantically segment the faces available in the image performed experiments on multi parsing human data sets and achieved higher accuracy [4]. Object detection becomes the important area in the field of image investigation there are various techniques for image analysis in [5]. In this paper author introduced a wavelet based neural network for feature extraction and learning which is working efficiently in object detection. The researchers in [6] gave a deep learning based assistive System to classify COVID-19 Face Mask which is implemented in rasbperrypi-3 using YOLOv3. Added with face mask detection performed well by the YOLOv3 where it measures real time performance regarding a powerful GPU. Whereas computation power with low memory YOLO darknet command sufficient for real time manner.

Most existing facemask detection approaches focus on whether the person is wearing a facemask or not. Our suggested system detects if a facemask is worn properly or not and achieves outstanding accuracy when compared to existing systems.

3. Methodology

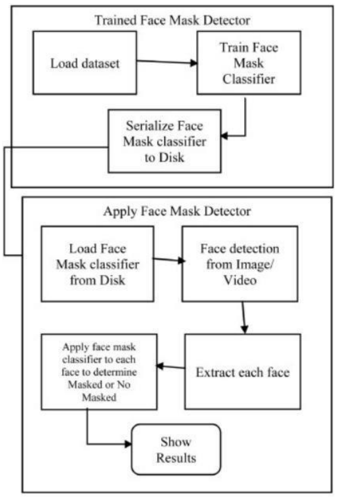

Figure 1: Proposed Approach

The proposed model aims to create a DL model that can detect and locate facemasks and whether they are appropriately worn. Figure 1 depicts the proposed approach.

a) CNN architecture

In this paper, we suggest using the VGG19 architecture to achieve accurate face-mask detection; the proposed CNN design is shown in Figure 2. VGG19 is chosen because it offers numerous advantages such as 1) faster training speed, 2) fewer training samples per time, and 3) higher accuracy [5]

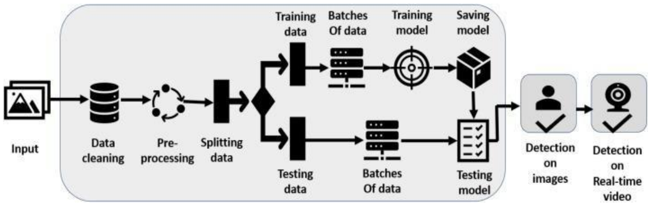

Figure 2: CNN Architecture

b) Data Augmentation

To increase the diversity of our dataset, we employed data augmentation. Multiple transformations, including geometric transformations, inversion, colour modification, width weight compensation, and rotation, are modes.

4. Experiment Setup

This section will begin by introducing the dataset utilized for this study. Then, numerous tests are undertaken to demonstrate that the examined CNN model is effective at detecting facemasks in image frames.

a) Dataset Description and Implementation Details

In order to build a complete and diversified dataset related to facemask, we combined public datasets with our own dataset. The public datasets used in this study are: 1) Face Mask Detection ~12K Images Dataset from Kaggle. The obtained dataset contains both; people wearing masks, without masks, wearing masks in an inappropriate way. The final dataset has an initial size of 328.92 MB and contains ~12K images.

The images present in the dataset underwent augmentation techniques like rescaling, adding sheer range and changing the zoom range. Said techniques were applied to introduce variation in the dataset to eliminate potential overfitting issues. Upon the completion of pre-processing steps, the images were fed into VGG19 model. Hyperparameters were fine-tuned to prioritize high accuracy metrics and improve efficiency thereby reducing the training time.

In this study, the experiments are carried out using a PC with the following configuration properties: an x64-based processor, an Intel Core i5-1135G7 CPU @2.40GHz, and a 8 GB RAM running on Windows 11 with NVIDIA GeForce GTX 1650. The CNN architectures are programmed using Jupyter notebook under python 3.10 programming language. We used both the Keras library and TensorFlow backend. For faster computation, we used Collaboratory pro with 25.51 GB RAM and 225.89 GB Disk, connected to python 3 Google Compute Engine backend.

b) Results

Several performance criteria, notably accuracy, recall, precision and F1-score, are employed in order to assess performance of the proposed CNN architecture. Four variables are used to calculate these measures: true negative (TN), true positive (TP), false negative (FN), and false positive (FP) . True Positive reflects a result for which the model successfully predicted the class positive. The TN variable represents a result for which the model accurately predicted the negative class. FP represents an outcome in which the model forecasts the positive class wrongly. Lastly, FN represents a result where model forecasts the negative class wrongly. Accuracy, precision, recall, and F1-score are calculated using the equation 1, 2, 3, and 4 respectively.

$$Precision = \frac{tp}{tp+fp} \quad (1)$$

$$Recall = \frac{tp}{tp+fn} \quad (2)$$

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} \quad (3)$$

$$F1 - score = 2 * \frac{precision * recall}{precision+recall} \quad (4)$$

In Figure 3, the classification report of the proposed CNN architecture is depicted. We can conclude that our model achieves excellent values in detecting facemask with over 92% accuracy, precision, recall, and F1-score.

	precision	recall	f1-score
with_mask	0.97	0.90	0.94
without_mask	0.99	1.00	0.99
accuracy			0.98
macro avg	0.98	0.95	0.96
weighted avg	0.98	0.98	0.98

Figure 3

c) Evaluation of the proposed approach

The considered VGG19 CNN model is compared to state-of-the-art DL models namely DenseNet, ResNet50, and MobileNetV2. They are trained with 32 and 20 epochs to compare their validation accuracy and loss on the same dataset.

	precision	recall	f1-score
with_mask	0.93	0.71	0.80
without_mask	0.94	0.99	0.96
accuracy			0.94
macro avg	0.93	0.85	0.88
weighted avg	0.94	0.94	0.94

ResNet50

	precision	recall	f1-score
with_mask	0.97	0.90	0.94
without_mask	0.99	1.00	0.99
accuracy			0.98
macro avg	0.98	0.95	0.96
weighted avg	0.98	0.98	0.98

VGG16

	precision	recall	f1-score
with_mask	0.98	0.97	0.98
without_mask	1.00	1.00	1.00
accuracy			0.99
macro avg	0.99	0.98	0.99
weighted avg	0.99	0.99	0.99

DenseNet

	precision	recall	f1-score
with_mask	0.98	0.98	0.98
without_mask	1.00	1.00	1.00
accuracy			0.99
macro avg	0.99	0.99	0.99
weighted avg	0.99	0.99	0.99

MobileNetV2

Figure 4: Comparison of performance metrics between ResNet50, VGG19, DenseNet, and MobileNetV2

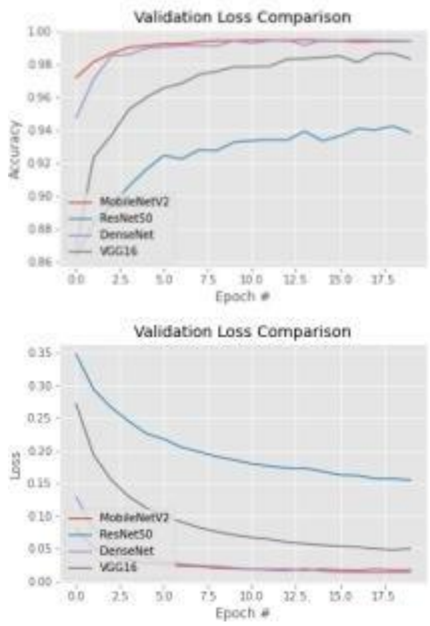

Figure 5: Comparison of validation accuracy and loss between MobileNetV2, ResNet50, DenseNet, and VGG19

Figure 5 depicts a comparison of the validation accuracy and loss results for MobileNetV2, ResNet50, DenseNet, and VGG19. The VGG19 model produces the greatest results, which justifies our usage of this model in this instance.

5. Conclusion

This study proposes a VGG19-based efficient CNN model for Real-time Facemask Detection. In training and testing, the suggested strategy achieved an accuracy of over 95%. Extensive testing indicates that the VGG19 model is capable of detecting facemasks in real time footage. Additionally, various comparison with numerous cutting-edge models, like ResNet50, DenseNet, and MobileNetV2, reveal the VGG19's superior training time and accuracy performance.

In the future, additional tests will be undertaken to assess the performance of the supposed method. We would like to carry out the planned strategy utilizing CCTVs in public locations to guarantee that persons are abiding by the laws and wearing the proper masks.

References

[1] (T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scaleand rotation invariant texture classification with local binary patterns,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987, July 2002.)
[2] (G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.)
[3] (P. Viola and M. J. Jones, “Robust real-time face detection,” Int. J.Comput. Vision, vol. 57, no. 2, pp. 137–154, May 2004.)
[4] (S. Kumar, A. Negi, J. N. Singh, and H. Verma, “A deep learning for brain tumor mri images semantic segmentation using fcn,” in2018 4th International Conference on Computing Communication and Automation (ICCCA), Dec 2018, pp. 1–4.)
[5] (Satapathy, Sandeep Kumar, et al. "Deep learning-based image recognition for vehicle number information." International Journal of Innovative Technology and Exploring Engineering 8 (2019): 52-55)
[6] (Bhuiyan, Md & Khushbu, Sharun & Islam, Md. Sanzidul. (2020). A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3. 10.1109/ICCCNT49239.2020.9225384.)