International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT206553

Facial Mask Detection Using Stacked CNN Model

Anushka G. Sandesara[1], Dhyey D. Joshi[2], Shashank D. Joshi[3]

[1]*Chandubhai S Patel Institute of Technology*, Charusat University, Gujarat, India

[2]*Devang Patel* Institute of Advance Technology and Research, Charusat University, Gujarat, India

[3]Vellore Institute of Technology, VIT University, Tamil Nadu, India

Article Info
Volume 6, Issue 5
Page Number: 264-270
Publication Issue :
September-October-2020

Article History
Accepted : 15 Oct 2020
Published : 23 Oct 2020

ABSTRACT

The coronavirus outbreak has affected the whole world critically. Amongst all other things, wearing a mask nowadays is mandatory to avoid the spread of the virus according to the World Health Organization. All the people in the country prefer to live a salubrious life by wearing a mask in public gatherings to avoid contracting the deadly virus. Recognizing faces wearing a mask is often a tedious job as there are no substantial datasets available comprising of masked as well as unmasked images. In this paper, we propose a stacked Conv2D model that is highly efficient for the detection of facial masks. Such convolutional neural networks work effectively as they can deduce even minute pixels of the images. The proposed model is a stack of 2-D convolutional layers with relu activations as well as Max Pooling and we implemented this model by using Gradient Descent for training and binary cross-entropy as a loss function. We trained our model on an amalgam of two datasets that are RMFD (Real World Masked Face Dataset) and Kaggle Datasets. Overall, we achieved a validation/testing accuracy of 95% and a training accuracy of 97%. In addition to this, we also developed an email notification system that sends an email whenever a person is entering without a mask and it will also prompt the user to wear the mask before entering into the system. Such a system is beneficial to large multinational companies and can be deployed there as the spread of viruses there is high because employees are from different regions.

Keywords : Facial Mask, Stacked CNN, Object Detection

I. INTRODUCTION

Detecting faces has developed as a fascinating branch of refining multiple images as well as computer vision. It possesses a myriad of applications such as detecting faces for automation related work, unlocking the phone, preventing crimes throughout the world, and many others. This branch is more practical nowadays because it is emerging not only for real-time scrutiny but it is also useful in video applications and video games. Highly precise models for facial detection are

developed only after the advancements in Convolutional Networks and Neural Networks.

Ordinary depictions of some images and videos mainly have people as their foreground. Moreover, according to [1] there are almost 35 % of pixels noted in movies hosted online and videos streaming on YouTube and 25% of pixels in portraits dedicated to people. This robust interest is growing parallelly with the augmentation of videos and photographs available online and due to this, there is an increasing desire for detection of public faces in the models.

The reports conducted by the WHO clearly show that the spread of the COVID-19 is increasing at an alarming rate. According to recent reports, there are 35.5 million infected people and 1.04 million deaths have already happened [2]**.** The population globally is decreasing and due to this there is a rise in various other problems too such as poverty, unemployment and all this leads to a financial crisis. There are various other respiratory diseases too which cause acute breathing problems and also reduce the oxygen rate of the body such as MERS and SARS [3]. But amongst these diseases, COVID-19 is deleterious and has affected all the countries critically. In such hard times, WHO advises all the people to wear a mask to avoid the spread of the harmful virus. In addition to this, all the public gatherings are restricted to a certain level and many shops have made the masks mandatory before reaching out. [4]

Figure (1) Examples of Masked Images in our dataset

Detection of facial mask works by detecting whether a person is wearing the mask. This issue resembles the detection of general objects to recognize the classes of the objects. As represented in Figure (1) masked faces have a variety of orientations and to accurately predict that if a person is wearing a mask is exacting.

We proposed a novel architecture called the Stacked Conv-2D model that is trained and tested on a dataset that is a combination of RMFD and Kaggle. The model works accurately not only for masked images but also for unmasked images. We have tested this model by taking a stack of different Conv-2D layers to attain more accuracy.

Convolutional networks play a significant role in tasks that are related to image processing and thus we implemented it. We divide the significant contributions of this paper into three bunches. Firstly, we demonstrate a mixture of two datasets that can be used for further development in the detection of the facial masks. Secondly, we proposed the Stacked Conv-2D model which outperforms some common models existing and lastly, we thoroughly analyzed the vital challenges faced by certain models of face mask detection.

II. RELATED WORK

The chronicle of detecting images dates back to the 80-90s as derived from [5] and these developments keep on climbing high walls of success with the advancements in neural networks. In the initial phase, researchers most prominently concentrated on the Gray value of all the images related to the face and not the coloured images. Ada-boost [6] was a well-known classifier used for training during those times. The

major breakthrough in this field happened with the emergence of the optimization of Haar attributes by eminent Viola-Jones [7]. But this discovery did not work properly under dark conditions and different orientation of images.

Instead of implementing this model by traditional approaches of machine learning and neural networks, there are other models developed that have displayed efficient performance. There are normally two well-known categories of detection namely one and two-stage object detectors. The two-phase detection produces proposals of the region in the fundamental step, then later proceeding towards tuning of these proposals. The primary work on R-CNN introduced in [8] works together with the Support Vector Machine (SVM) by extracting the features to distinguish different faces. But, the R-CNN model is high-budget and time-consuming as it involves the detection of proposals one-by-one. So, the solution is solved by the Faster R-CNN model [9] as it introduces a novel pooling layer called ROI. This model blends together the extractor of feature, discrete parts for detection, a special detector for an end to end architecture of neurons. One-phase detectors implement the model by using only one neural network and this in turn loosens the performance slightly compared to the two-phase detectors. The YOLO model [10] works by splitting the images into various cells but this model lacks the performance for minute objects. After some time, researchers concluded that one-phase detectors in neural networks do not function appropriately because in this mechanism the final map has limited fields that are unable to detect some vital things in actual images. In order to outperform certain models with the best accuracy multi-scale detection was brought into the light. Such as Retina Face Mask Detector in [11] is a one stage detector that is a pyramid model comprising various features to detect face mask. For better performance in this model, there is an attention mechanism and object removal feature to intensify accuracy.

III. PROPOSED METHODOLOGY

The main motive behind proposing this methodology is to develop a face mask detection system that can detect faces of different orientation regardless of the alignment. The primary function of the model is to extract the necessary features by pre-processing the data. The output produced from the network is a feature vector that is enhanced using Gradient Descent and the function used to calculate the loss in Binary Cross Entropy.

3.1 Proposed Work Flow

We decided to implement this through convolutional networks because it can extract features that are of high quality and generally higher accuracy is obtained by using deep learning models for image classification. We used Kaggle+ RMFD dataset and at the fundamental stage, we pre-processed the data by following the steps given in Figure (2). So, the images of the dataset consisting of random sizes are converted to a fixed size of 150 * 150, and then the images of the dataset are trained and tested on the stacked Conv-2D layer. Moreover, we used the dropout to ignore certain images of the dataset during the training phase to gain better accuracy.

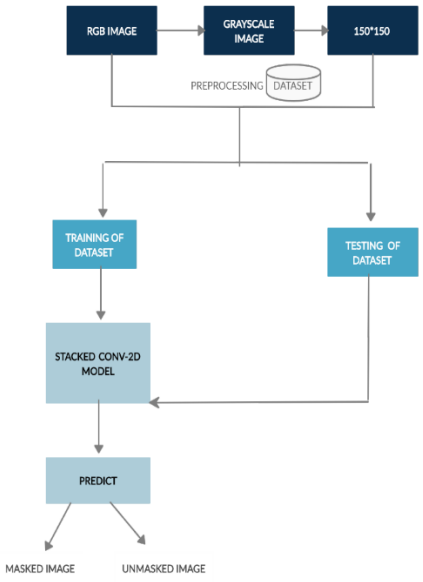


Figure (2) Flowchart of Proposed Model

3.2 Architecture

The Architecture of the Proposed Stacked CNN model is shown in Figure (3). We have trained our model by taking into consideration the crucial stages of feature extraction. The CNN layer is convenient for image classification tasks as it convolutes the image and the max-pooling stage makes sure that the dimensions of the feature vector produced after every stacked model are reduced to decrease the number of parameters. If the parameters of the images in the dataset used are not reduced then it would become time-consuming to anticipate the classes of all the pixels in a completely connected CNN model.

In the initial stage, we pre-processed the images of the dataset into size 150*150 so that all the images are of the same size. Then we followed the Sequential model for stacking the CNN models so it passes the output from one model as input to another model. The pre-processed images are passed to the Conv-2D layer of 32 filters. After this stage, Max Pooling is applied to make sure that over-fitting does not happen. In our model, we have applied Max Pooling after every layer to ensure that it achieves the best possible accuracy. Furthermore, other Conv2D layers that are used consist of 64,128 and 256 filters respectively. After stacking the model with four convoluted layers, in the last stage, we used one dense layer for output consisting of 100 neurons and we used a sigmoid activation function in this layer. Using activation function in the neural network is essential as it helps the neural network to understand the intricate and complex patterns of the data used.

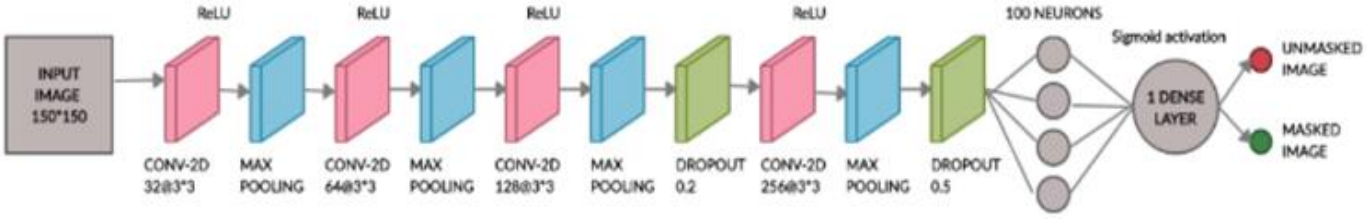


Figure (3) Model Architecture of Stacked Conv-2D Model

After pre-processing the images of the dataset, all images are converted into 150*150. These images are passed to the Conv-2D layer of 32 filters and after Max-Pooling, the size becomes 149*149. Similarly, after another Conv-2D layer of 64 filters and Max-Pooling, the size becomes 148*148. After these two stacks of the Conv-2D layer, there are other two stacks where the size becomes 147*147 and after the final layer, it becomes 146*146. Before passing the images to the final output layer we have flattened the images to impose the precise size of the image going into the dense layer.

3.3 Stacked Convolutional Neural Network

Stacked CNN is an ensemble model of different Convolutional Layers consisting of different filters. The presented novel method uses stacked CNN to learn distinguishable features and interpret images at different levels. So, the model is accurately trained for image processing and produce higher accuracy compared to traditional methods. The Pseudo Code for the Proposed network in our model is shown in Figure (4). The List3 shown in the algorithm is an amalgamation of List1 and List2 (masked and unmasked images preprocessed). We have used 4 stacked Conv-2D layers in our model to produce high results. We have then divided our training and testing data into 90 and 10% so to increase the probability of getting accurate predictions. If the dataset is not divided into training and testing data, then there are high chances of getting improper predictions.

Algorithm 1: Proposed Stacked Convolution Neural Network

```
Input: RMFD + KAGGLE DATASET
Output: Classification Results
initialization;
for i=1 to length(Masked Images) do
    Read Image
    Convert 150*150(GrayScale)
    Resizing
    Append 2D array to List1
end
for i=1 to length(Unmasked Images) do
    Read Image
    Convert 150*150(GrayScale)
    Resizing
    Append 2D array to List2
end
List3= List1+ List2
train , test = split ( 90 % Train, 10 % Test )
stacked_model= fit.classification(training data, training_label)
Classify masked and unmasked images
pred_label = classify(stacked_model,test_image)
return pred_label
```

Figure (4) Pseudocode of proposed model

IV.EXPERIMENTS

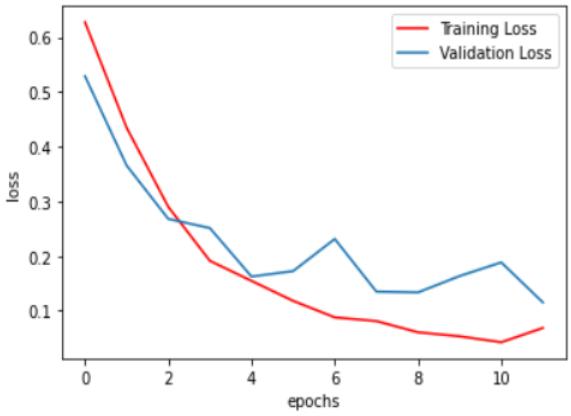
We conducted a comprehensive study on how the system works accurately considering different datasets as input. By analysing these results, we gained the understanding of how different parameters have an impact on training and on which terms the model will gain more accuracy. We used Keras framework to train the Stacked-CNN Model. Moreover, we trained our model with the help of Adam optimizer and binary cross entropy as the loss function. The technique proposed here works in two different ways. Firstly, if a person is caught without a mask, then he would be restricted to enter the place and an alert system will notify the person through mail to wear a mask. Secondly, when images of the datasets are considered the model is trained sequentially and achieves a high accuracy of 95%.

4.1 Experiment Results

In this proposed work, we not only considered the enclosed box of the images formed but also worked on splitting up the information related to faces from images that include disparate backgrounds and angles. We carried the work on a combination of two datasets that is RMFD and Kaggle. We divided data into training and testing that is 90% and 10%, used 12 epochs with batch size of 108. Figure (5) and Figure (6) demonstrates the relationship between training accuracy and loss with respect to epochs. As the number of epochs increases the training loss gradually decreases. And when the loss increases the accuracy of the model increases and effective results are obtained.

The accuracy of the model in the beginning was only 0.65 but as the epochs increased the accuracy after 12 epochs became 0.97.

Figure (5) Model Training and Validation Loss



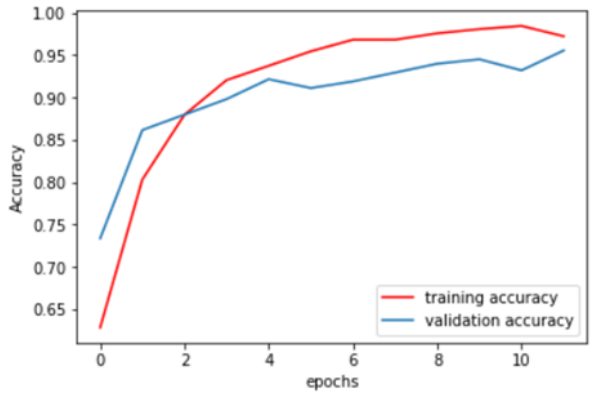


Figure (6) Model Training and Validation Accuracy

4.2 Results and Analysis

The dataset which we considered for training the model consists of images from various different scenes and backgrounds so it is challenging to receive high accuracy. We compared our model with models such as Mask R-CNN [12], Retina-Face-Mask [11], Faster R-CNN [9].

We further considered MAFA and WIDER datasets to check the accuracy of our trained model. To further show our face mask detection accuracy we tested our Stacked CNN model on MAFA+ WIDER dataset but we obtained accuracy of only 66.56%. Furthermore, we trained the sequential model with MAFA+ WIDER dataset and tested on the same. But it displayed very low accuracy as shown in Figure (7) and Figure (8). The accuracy obtained was only 47.32%. So, as less accuracy was obtained, we worked on the Stacked Conv-2D Model to train and test the model efficiently.

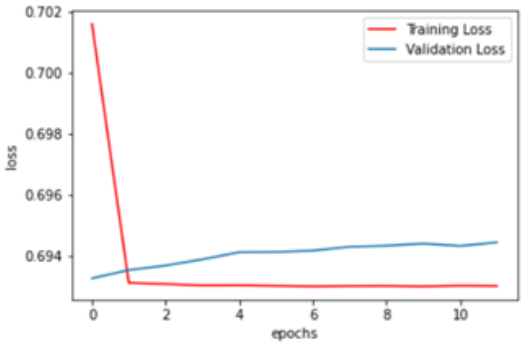


Figure (7) Model Training and Validation Loss

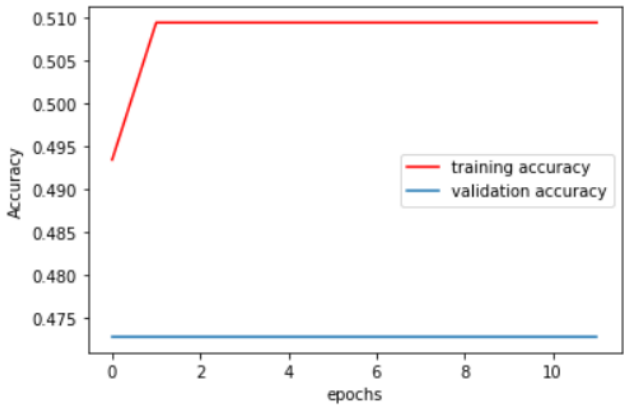


Figure (8) Model Training and Validation Accuracy

V. CONCLUSION

In this paper we developed a novel Stacked-CNN model to ensure the safety of individuals in surrounding and to decrease the spread of Coronavirus. The practical model can even be deployed in industrial and public areas where the risk of contracting the virus is extremely high. Thus, the proposed system will work effectively to ease the deleterious effects of the virus. The model is highly trained to achieve the maximum accuracy possible. This system can augment the safety of the public as the spread of virus is continuously rising above. For future work we would like to concentrate on how to detect the temperature by using ensemble model and detection of sneezing and coughing using a deep learning-based model.

VI. REFERENCES

[1]. Ivan Laptev. Modeling and visual recognition of human actions and interactions. Computer Vision and Pattern Recognition. Ecole Normale Superieure de Paris - ENS Paris, 2013.

[2]. Coronavirus Update Live(24/7) Online . Available at: www.worldometers.info/coronavirus/

[3]. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Peñaranda S,

Bankamp B, Maher K, Chen MH, Tong S, Tamin A, Lowe L, Frace M, DeRisi JL, Chen Q, Wang D, Erdman DD, Peret TC, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Günther S, Osterhaus AD, Drosten C, Pallansch MA, Anderson LJ, Bellini WJ. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science. 2003 May 30;300(5624):1394-9. doi: 10.1126/science.1085952. Epub 2003 May 1. PMID: 12730500.

[4]. Fang, Yaqing & Nie, Yiting & Penny, Marshare. (2020). Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis. Journal of Medical Virology. 92. 10.1002/jmv.25750.

[5]. Regis Vaillant, Christophe Monrocq, & Yann Le Cun. (1994). An Original Approach for the Localization of Objects in Images.

[6]. Wang L, Chu J (2011) Fused multi-sensor information image stitching. In: Proceedings of the international conference on intelligent science and intelligent data engineering, F. Springer

[7]. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I-I, doi: 10.1109/CVPR.2001.990517.

[8]. R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.

[9]. R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[10]. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[11]. Mingjie Jiang, Xinqi Fan, & Hong Yan. (2020). RetinaMask: A Face Mask detector.

[12]. Lin, R. 2020. Face Detection and Segmentation Based on Improved Mask R-CNN. Discrete Dynamics in Nature and Society, 2020, p.9242917.

Cite this article as :

Anushka G. Sandesara, Dhyey D. Joshi, Shashank D. Joshi, "Facial Mask Detection Using Stacked CNN Model", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 5, pp. 264-270, September-October 2020. Available at
doi : https://doi.org/10.32628/CSEIT206553
Journal URL : http://ijsrcseit.com/CSEIT206553