Indian Institute of Information Technology Vadodara

Machine Learning Project

on

Face-Mask Detector

Submitted by

Nikhil Rana **201951103**

Vishal Singh Rajput **201951171**

Shashank Jaiswal **201951140**

Tanmay **201952237**

Under The Supervision of

Dr.Jignesh Bhatt

Abstract—**In the current times, the fear and danger of COVID-19 virus still stands large. Manual monitoring of social distancing norms is impractical with a large population moving about and with insufficient task force and resources to administer them. There is a need for a lightweight, robust and 24X7 video-monitoring system that automates this process.With the help of this project we hoped to find a comprehensive ans effective solution to perform face detection and further face mask detection using Colvolution Neural Network.For this, YOLOv3, Density-based spatial clustering of applications with noise (DBSCAN), Dual Shot Face Detector (DSFD) have been employed on images and then on random videos.The system performance is evaluated in terms of accuracy, F1 score as well as the prediction time, which has to be low for practical applicability. The system performs with an accuracy of 98.2%.**

I. Introduction

THe use of face masks is very necessary to prevent and limit the spread of certain respiratory viral diseases, including COVID-19 . Face mask can be used to protect either healthy people or prevent infection by infected persons. However, wearing face mask correctly is very important to reduce risks of contamination. The World Health Organization recommends to use face mask at crowded spaces such as station, office, school, etc. In order to check whether a person wears a face mask or not, we need a computer vision system that is able to perform this type of detection. The problem of face mask detection in computer vision research is a special case of general object detection. Object detection has many uses, such as in smart home systems, surveillance systems, autonomous vehicles, etc. Object detectors are usually based on Haar feature extractors, support vector machines (SVMs), or Bayesian networks. Recent advances in deep learning have made object detection a widely used technique. The Yolov3 algorithm is currently one of the most promising algorithms for detection and classification problems.

II. Literature Survey

Propossed Algorithm The model proposed by [3] is an integration between deep transfer learning (ResNet-50) and classic machine learning algorithms. The last layer in ResNet-50 was removed and replaced with three traditional machine learning classifiers layers to improve their model performance. Among the four types of datasets they used, one dataset contained the largest number of images among the datasets, consisting of real face masks and fake face masks, and consumed more time compared to the others during the training process. There is also no reported accuracy according to related works for this type of dataset. On the training over dataset having real face masks, the decision trees classifier wasn't able to achieve a good classification accuracy (68%) on fake face masks.

Another approach is a self-developed model named SocialdistancingNet-19 by Rinkal Keniya [2] for detecting the frame of a person and displaying labels for deciding if they are safe or unsafe if the distance is less than a certain value. If a webcam is to be used, it is necessary to have people moving continuously else the detection goes incorrect. This may happen due to the detection method used by the network where the entire frame is detected, and the distance is calculated between the people using centroids (brute force approach). Shashi Yadav proposed a deep learning approach with Single Shot object Detection (SSD) using MobileNet V2 and OpenCV for social distancing and mask detection[4]. Challenges faced using this approach was that it categorizes people with hand over their faces or occluded with objects

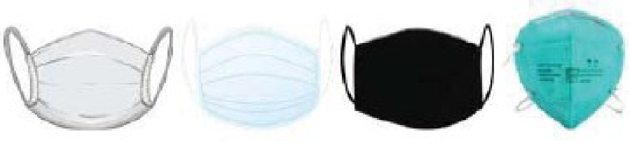

Fig. 1. Masks used for augmentation

as masked. These scenarios are not suited for this model. Here, although an SSD is capable of detecting multiple objects in a frame, it is limited to the detection of a single person in this system. Most of the papers have tackled either the issue of social distancing monitoring, or face mask detection. And where both were implemented, there is still scope for using better models to achieve better accuracy. Our paper mentions the importance of prediction time, a feature missing in other papers, with prediction time as an evaluation measure, which is necessary for practical applicability of the system. For person detection, the model proposed by the paper is YOLOv3, a state-of-the-art object detection model, followed by DBSCAN to calculate the distances between people and perform clustering to identify if they are far apart or not, which is effectively better than other clustering methods like or brute force distance calculations or k-means, which requires that number of clusters be decided before performing clustering. For face detection we have DSFD which is a powerful feature extractor with a good accuracy for detecting faces. Finally, a mask dataset was created using data augmentation techniques and a labelled video dataset was also created for testing the system by labelling the frames of the video.

III. Dataset

A. Dataset Creation

The dataset collected from existing sources consisting of unmasked and masked faces proved to be insufficient, hence new and two data augmentation techniques were employed to add masks on unmasked faces and to add blurred images. Data augmentation for unmasked faces was performed on datasets with four types of masks as shown in Fig.1.

B. Data Cleaning and Preprocessing

Fig 2 illustrates the algorithm used here. First, the algorithm identifies the defining points of a face's outline. In order to identify the top, left, and right parts of the mask, locate the nose bridge, while the bottom, right, and left parts of the mask are found by finding the chin points. Right and left halves of the mask are prepared according to their size. The mask's angle of rotation is calculated based on the face's orientation. The mask is placed on the face after calculating the coordinates for superimposing it on the face.

Due to the nature of the problem, the images were taken from a surveillance camera, which resulted in very blurry or occluded faces. However, the dataset displayed clear faces. In Fig 3., to help the model to adapt to the surveillance quality faces, blurring filters such as motion blur, gaussian blur and average

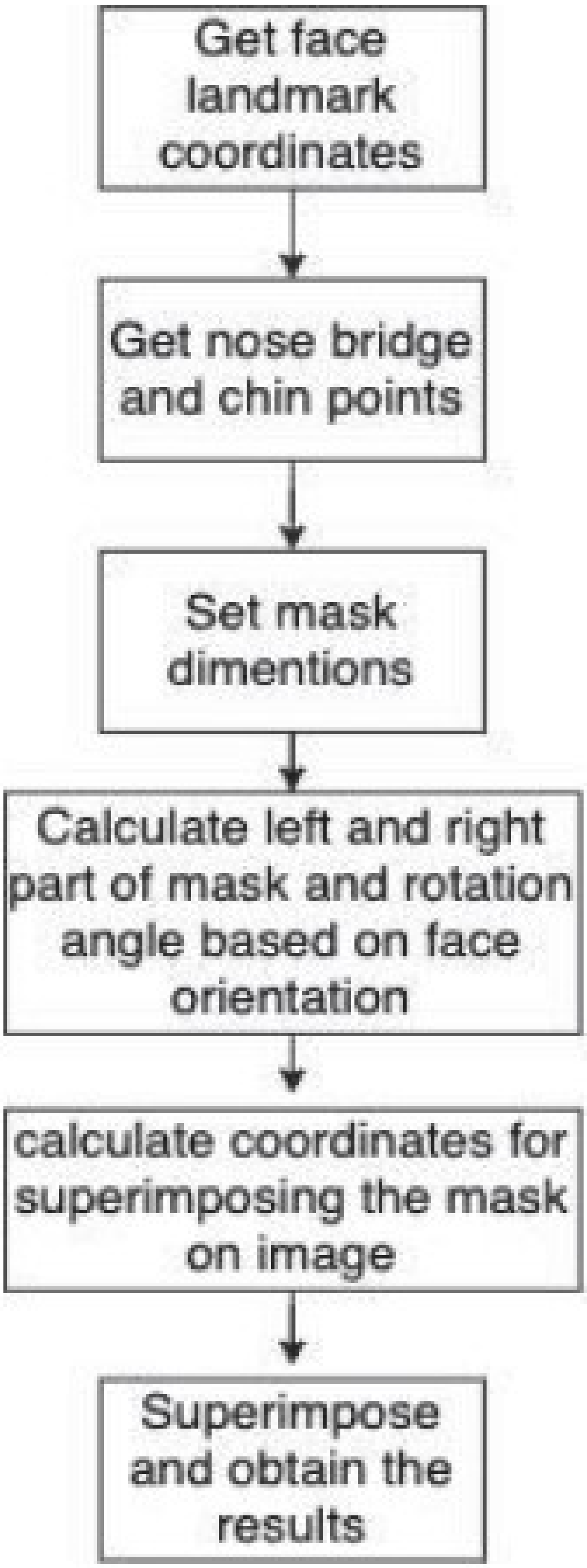
Get face landmark coordinates
Get nose bridge and chin points
Set mask dimentions
Calculate left and right part of mask and rotation angle based on face orientation
calculate coordinates for superimposing the mask on image
Superimpose and obtain the results

Fig. 2. Steps for dataset generation

blur are used for the second data augmentation.
Blurring options include gaussian blur with kernel sizes ranging from 6 to 10, average blur with kernel sizes ranging from 3 to 9, motion blur with kernel sizes ranging from 3 to 10, and no blur option. An image is blurred with a gaussian function [14]. A kernel that is close to the center is given more weight than one that is far from the center. Using a kernel to mesh the image with, motion blur is the apparent streaking of moving objects in a photograph or sequence of frames by choosing a random direction from vertical, horizontal, main diagonal, and anti-diagonal. An image is blurred by convolving it with a box filter (normalized). An average of all pixels in the kernel area is used to replace the central element of the image.

C. *Final dataset*

The final dataset consists of 1410+ labeled images with the classes 0 for no mask and 1 for mask. Due to the problem being a recent one, there were no video datasets available. We have used validation dataset where a sample of data held back from training your model that is used to give an estimate of model skill while tuning model's hyperparameters.

IV. CNN and Model Training

After creating the dataset with the augmented images we shifted our focus to build our model.

A. *DSFD*

For that first we had to again shrink the image taking into account only the face in each image and for that we used the module face detection using DSFD algorithm an open-source face detection algorithm that offers higher accuracy over Haar Cascades, Dlib, MTCNN, and DNN.It addresses three key areas of facial detection that includes feature learning, progressive loss design, and anchor assign based data augmentation. It also ranked 1st across the board in the WIDER FACE Face Detection Benchmark.

B. *Neural Network*

Now we had our fully processed data to be fed to our Neural network.Face mask classification is implemented using CNN binary image classification architecture to check the presence of a mask on the faces detected. A number of models were developed using CNN for mask classification on 224x224 images and their performance in terms of accuracy.But the accuracy achieved was only 48%. So we made changes to the predefined model first freezed all the layers.

C. *Maxpooling*

Max Pooling is a pooling operation that calculates the maximum value for patches of a feature map, and uses it to create a downsampled (pooled) feature map. It is usually used after a convolutional layer. It adds a small amount of translation invariance - meaning translating the image by a small amount does not significantly affect the values of most pooled outputs.We added a maxpooling to shrik the image and consider only the relevant information necessary for our classification , flatten to convert our shrinked 2-d data to a single dimension array to put it into the next layer.

D. *Dense Layer and Activation Function*

In any neural network, a dense layer is a layer that is deeply connected with its preceding layer which means the neurons of the layer are connected to every neuron of its preceding layer. This layer is the most commonly used layer in artificial neural network networks. Activation function decides, whether a neuron should be activated or not by calculating weighted sum and further adding bias with it. The purpose of the activation function is to introduce non-linearity into the output of a neuron.The ReLU function is another non-linear activation function that has gained popularity in the deep learning domain. ReLU stands for Rectified Linear Unit. The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time thus passing on the most important information to the next layer and helping us make a non linear function piece wise linear.Thus we used a hidden dense layer and with ReLu activation .

E. *Dropout*

Now we used a dropout layer which randomly sets input units to 0 with a frequency of rate at each step during training time, which helps prevent overfitting. Inputs not set to 0 are scaled up by 1/(1 - rate) such that the sum over all inputs is unchanged.

F. *Output*

Finally an output layer using sigmoid activation.This function takes any real value as input and outputs values in the range of 0 to 1. The larger the input (more positive), the closer the output value will be to 1.0, whereas the smaller the input (more negative), the closer the output will be to 0.It is commonly used for models where we have to predict the probability as an output. Since probability of anything exists only between the range of 0 and 1, sigmoid is the right choice because of its range. In our model also we are giving the image of the person and predicting if the person is wearing a mask or not which is a binary result.On training the new model we got an accuracy of 98%. Then we tested the model with the validation data using new images.

V. Video mask detection

A. *Person Detection*

YOLOv3 model was used for person detection [16]. It consists of 53 layers of Darknet-53 trained on Imagenet that acts as a powerful feature extractor and an additional 53 layers for detection giving a total of 106 layered fully convolutional neural network. Fig 5. depicts the YOLOv3 architecture. Anchor box with 3 scales: 13x13, 26x26 and 52x52 are used. These three anchor boxes are used to predict the presence of a person as shown in figure. The output of this model after prediction is a list of bounding boxes along with the confidence of the person class detected.

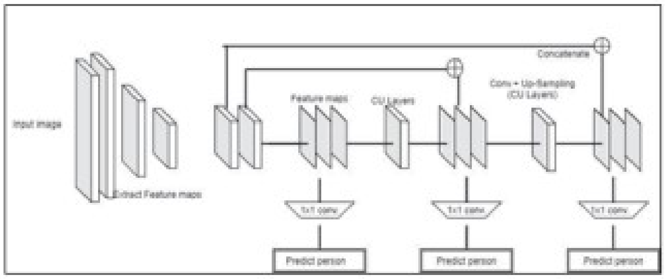

Fig. 3. YOLOv3 architecture

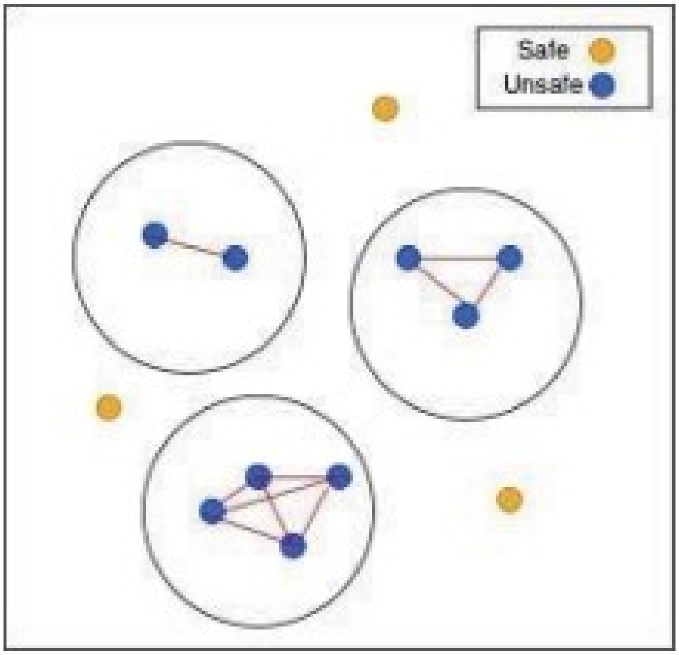

Fig. 4. DBScan for social distancing

B. Distance Detection

DBSCAN algorithm was used to check if social distancing is maintained between the persons detected. It is an unsupervised learning algorithm which groups similar points together. DBSCAN, unlike the k-means algorithm, does not require the number of clusters to be set prior training. It also ignores the noisy or outlier points while forming the clusters.

VI. Conclusion

This paper has presented a face mask detection system which uses the YOLOv3 algorithm and DFDS algorithm classifier. The proposed algorithm employs image enhancement technique to improve the accuracy of the system. Thanks to the advantages of the YOLOv3 network, the system can work in real-time with 30fps. This system can be applied effectively for practical applications to reduce the spread of infectious diseases such as Covid-19.

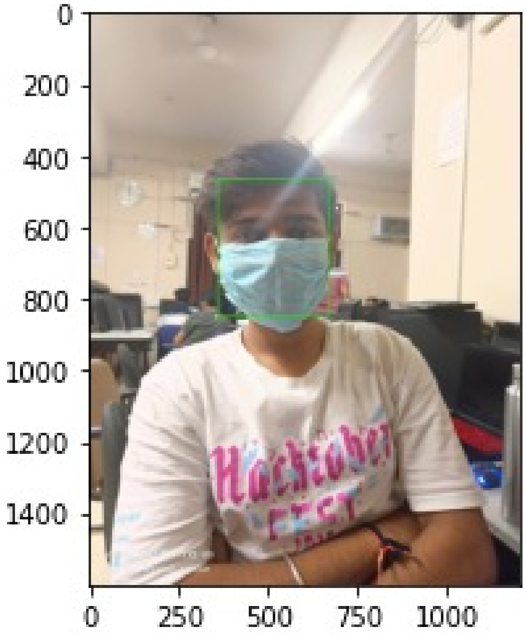

Fig. 5. Detected face mask correctly.

VII. Acknowledgement

We would like to express our sincere gratitude to Dr. Jignesh Bhatt, for his valuable guidance and support in completing our project.

References

[1] Arjya Das, Mohammad Wasif Ansari, and Rohini Basak. “Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV”. In: *2020 IEEE 17th India Council International Conference (INDICON)*. 2020, pp. 1–5. DOI: 10.1109/INDICON49873.2020.9342585.

[2] Rinkal Keniya and Ninad Mehendale. “Real-time social distancing detector using socialdistancingnet-19 deep learning network”. In: *Available at SSRN 3669311* (2020).

[3] Mohamed Loey et al. “A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic”. In: *Measurement* 167 (2021), p. 108288. ISSN: 0263-2241. DOI: https://doi.org/10.1016/j.measurement.2020.108288. URL: https://www.sciencedirect.com/science/article/pii/S0263224120308289.

[4] Shashi Yadav. “Deep learning based safe social distancing and face mask detection in public areas for covid-19 safety guidelines adherence”. In: *International Journal for Research in Applied Science and Engineering Technology* 8.7 (2020), pp. 1368–1375.