A study of machine learning algorithms and their performances on face mask detection and social distancing to prevent COVID-19

Abstract- **The global pandemic of COVID-19 has severely devastated the world, according to statistics collected by the World Health Organization (WHO) and has now infected more than eight million people globally. We suggested an efficient computer vision-based solution focused on the real-time automated monitoring of people to detect safe social violations through cameras. Wearing face masks and following safe social separation are two of the enhancements that contribute to public safety. Modern deep learning algorithms have been combined with geometric approaches to create a robust modal that covers three aspects of detection, tracking, and validation in this suggested system. As a result, the suggested approach benefits society by saving time and reducing the transmission of the coronavirus. It can be used efficiently in the current circumstances, where lockdown is being loosened in order to inspect people at public meetings, retail malls, and other places. Automated inspection saves time and money by reducing the number of people needed to inspect the public. It may also be utilized everywhere.**

Key Words—COVID-19, face mask, social distancing, CNN

I. INTRODUCTION

Since COVID-19 has become a pandemic, people across the world are brainstorming ways to stop it from spreading. Maintaining social distance and wearing a mask while out are the fundamental rules for preventing the spread of the virus. The Corona Virus infects others, but it coordinates real contact with afflicted individuals and through the air. The virus directly hit the lung cells through the patient's respiratory system, allowing it to replicate the infection and creating an extremely irresistible condition in breathing and very limited ability to focus. Our major goal, Face mask detection with Social Distancing, is to determine if a previously detected object is wearing a mask or not while walking with a Social Distance between them. The implementation of this policy cannot be tracked manually. The key to this is technology. We provide a Deep Learning-based system for detecting instances of improper use of face masks.

Our system uses a dual-stage Convolutional Neural Network (CNN) architecture that can recognize both masked and unmasked faces and is compatible with pre-installed CCTV cameras. This will make it easier to track safety infractions, promote the usage of face masks, and establish a safe working environment. The research is carried out by examining the required technologies used in prior studies and developing an effective model that assists people in real-time. Because application development utilizing Keras and Tensor flow are the most popular in current trends, we have adopted Python-based image processing and machine learning techniques to produce a stable structure

II. LITERATURE REVIEW

A cumulative article was presented, containing the various categories of COVID datasets that are available. These are open-source datasets that are freely available. X-rays and CT scan images are included in the listed sequence of data, and in a few circumstances where health backgrounds are narrowed, MRI images are also employed. Textual data based on debates, medical suggestions posted on social media, and other sources are another group of datasets used in COVID research. Many dataset windows that are used for diagnostic procedure analysis also consider clinical test findings and reports. [1] A virtual social distancing concept was evaluated to assist people who are being warned in public areas. Intimate space, personal space, social space, and public space were graphically depicted as four different forms of spacing. The spaces are measured using the distance measuring rule. Scene comprehension and geometrical measurement, homograph estimation, metric references, and density estimation, among other things, are all covered by this technique. Two-dimensional people detection and multiple angle people detection are included in the secondary analysis.[2] In a pandemic crisis, a human detection framework was evaluated for monitoring social distancing and safety misuse. To identify the distinct models, they used a pre-trained recurrent CNN model The human detection method adopts blob segmentation. The distance between these blobs is calculated by tracking their positions in relation to one another. They had trouble detecting the person's body parts in the outdoor area due to the correlations of other items nearby. They discovered that more research is needed to solve this problem. [3] Using principal component analysis (PCA) and a convolutional neural network, a unique face recognition system was developed. Discriminant algorithms, multi-layered perceptions, nave Bayes models, and support vector machines are used to test the experiment. The obstacles to identifying the multi-face model are discussed in this publication on face recognition, as well as how to overcome them in the future study. [4] A deep learning technique to face identification and categorization was tested. A pre-trained facial dataset is used to cluster various faces. The suggested model is trained and tested using the FDDB dataset. To achieve originality and good prediction performance, the proposed model is modified. The accuracy of the convolution neural network is stated to be high, and future challenges include real-time facial photos and live video capture. [5]. Multi-face identification approach based on machine learning algorithms Ada-Boost, SVM (support vector machine), and gradient boost models is described in depth. These strategies are taught in detail to attain high accuracy when used in conjunction with one another. The researcher's main challenge is an increase in the rate of false positives in particular facial data. [6]

III. EXPERIMENTAL METHODS

A. Faces with and without mask:

The dataset is collected from KAGGLE which is a publicly available dataset containing up to 11,000 faces where up to 5,500 of them are masked and 5,500 are unmasked.[7]

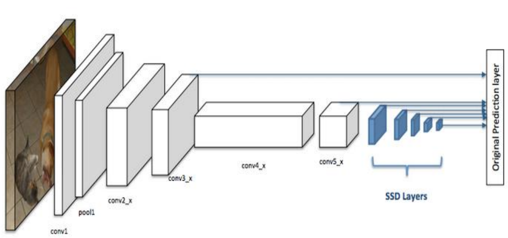
B. Detection of People:

You simply have to look once. (YOLO) is a real-time object detection system that is state-of-the-art. It processes images at 30 frames per second on a Pascal Titan X and has an mAP of 57.9% on the COCO test-dev. Prior detection methods perform detection by reusing classifiers or localizers. They apply the model to an image at different scales and places. The image's high-scoring sections are referred to as detections. We take a completely unique approach. We use a single neural network to process the entire image. The image is divided into regions by this network, which predicts bounding boxes and probabilities for each region. The projected probabilities are used to weight these bounding boxes. Compared to classifier-based systems, our model has significant advantages. At test time, it examines the entire image, thus its predictions are influenced by the image's overall context. It delivers predictions with just one network evaluation, unlike systems like R-CNN, which require thousands of network evaluations for a single image. It's a thousand times faster than R-CNN and a hundred times quicker than Fast R-CNN because of this. Multi-scale predictions, a better backbone classifier, and other techniques are used by YOLOv3 to improve training and performance.[8]

C. Detection of faces:

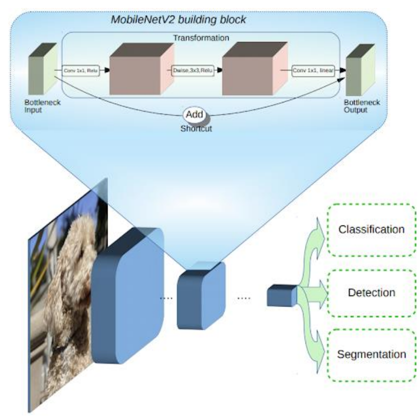
In this model, face detection is done by pretrained SSD. An SSD is made up of two parts: a backbone model and an SSD head. As a feature extractor, the backbone model is commonly a pre-trained image classification network. This is often a ResNet network trained on ImageNet that has had the final fully linked classification layer removed. As a result, we have a deep neural network that can extract deeper information from an input image while keeping the image's geometric features, although at a lesser resolution. For an input image, ResNet34's backbone generates 256 7x7 feature maps. The SSD head consists of one or more convolutional layers added to this backbone, with the outputs interpreted as bounding boxes and classes of objects in the spatial position of the final layers' activations.[9]

The backbone is represented by the first few layers (white boxes) in the diagram below, while the SSD head is represented by the last few layers (blue boxes).



D. Face Mask Classifier:

We are using MobilenetV2 as a face mask classifier. MobileNetv2 is a 53-layer deep convolutional neural network. It is a powerful feature extractor for detecting and segmenting objects. When used in conjunction with an SSD, the new model is around 35 percent faster while maintaining the same level of accuracy as the MobileNetV1. The model is available under the Tensorflow Object Detection API. We use MobileNetV2 as a feature extractor in a simplified edition of DeepLabv3 [10] to enable on-device semantic segmentation. Our model achieves identical performance on the PASCAL VOC 2012 semantic segmentation benchmark as using MobileNetV1 as a feature extractor, but with 5.3 times fewer parameters and 5.2 times fewer Multiply-Add operations. Below is the basic structure of MbileNetV2:



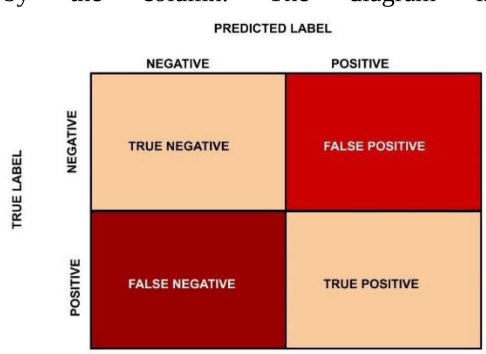
Euclidean Distance:

The distance between two or more people will be calculated by the Euclidean distance formula. The formula is:

$$D_e = \left(\sum_{i=1}^{n} (p_i - q_i)^2 \right)^{1/2} \qquad (1)$$

F. Confusion Matrix:

A confusion matrix is a table structure that is often used in statistical classification and machine learning to visualize an algorithm's performance. The projected class instance is represented by the row, and the actual class instance is represented by the column. The diagram is shown below.



G. Accuracy:

This is one of the best measures as our dataset is balanced. It is described as the number of correctly classified data instances divided by the total number of data instances is known as accuracy.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \qquad (2)$$

H. Precision:

Precision is a metric that indicates how many of the positive forecasts were correct.

Precision= TP/(TP+FP) (3)

I. Recall and F1 score:

Recall which is also known as sensitivity or true positive state can be defined as

$$Recall = \frac{TP}{TP + FN} \qquad (4)$$

F1 score is a metric that considers both precision and recall into account and can be expressed as

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad (5)$$

VI. RESULTS AND DISCUSSION

A. *System Performance*

[INFO] evaluating network...

	precision	recall	f1-score	support
with_masko	0.90	0.89	0.89	387
without_masko	0.89	0.90	0.90	386
accuracy			0.90	773
macro avg	0.90	0.90	0.90	773
weighted avg	0.90	0.90	0.90	773

Figure(a): using stochastic gradient descent

[INFO] evaluating network...

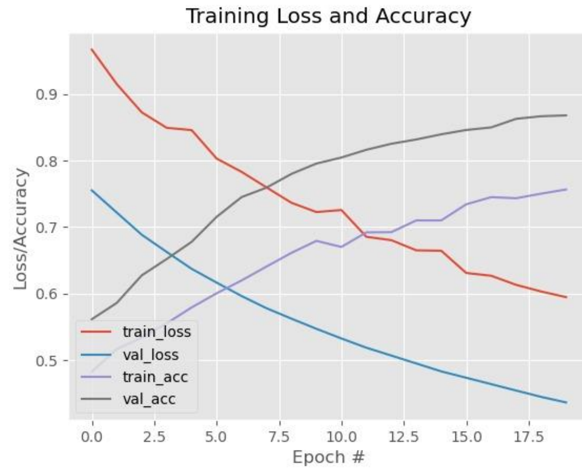
	precision	recall	f1-score	support
with_masko	0.97	0.98	0.98	387
without_masko	0.98	0.97	0.98	386
accuracy			0.98	773
macro avg	0.98	0.98	0.98	773
weighted avg	0.98	0.98	0.98	773

Figure(b): using Maxpooling, tanh,and Adam optimizer

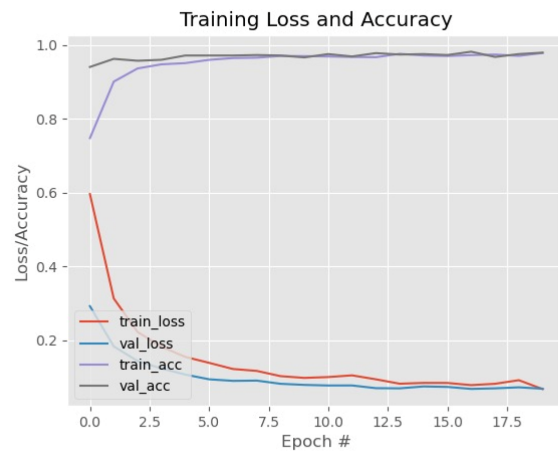
[INFO] evaluating network...

	precision	recall	f1-score	support
with_mask	0.99	0.98	0.99	387
without_mask	0.98	0.99	0.99	386
accuracy			0.99	773
macro avg	0.99	0.99	0.99	773
weighted avg	0.99	0.99	0.99	773

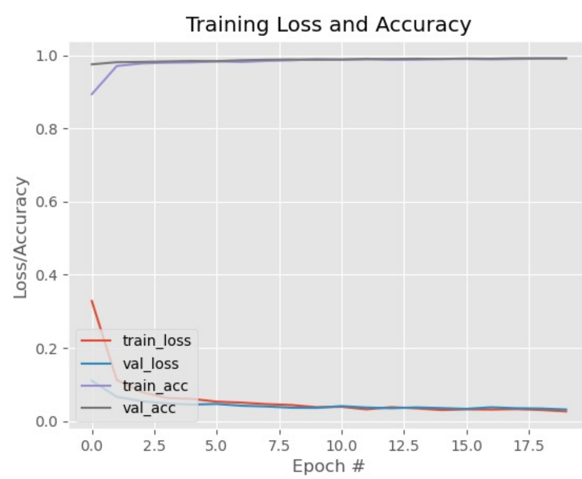
Figure(c): using Averagepooling, Relu,and Adam optimizer



Figure(d): plot of Stochastic gradient descent



Figure(e): Using maxpooling, tanh adam optimiser



Figure(f): plot using average pooling Relu and Adam optimizer

B. *Discussion*

The face mask detection using this method is proven to be highly accurate however we have used mobileNetV2 to train the model which resulted in accuracy with which we realized to change the activation functions and optimization technique

Although this method had a descent accuracy after trying different pooling methods, we got desired and efficient results.

V. CONCLUSION

In the face of the epidemic, corporate heavyweights from a variety of industries are resorting to AI and machine learning, setting technology to work in the service of mankind. Mask detection API services are being developed by digital product development businesses, allowing developers to swiftly build a face mask identification system to help the community during an emergency. We presented an approach that employs computer vision and Mobile Net V2, SSD architecture to help maintain a safe environment and assure individual protection by automatically monitoring public places to prevent the spread of the COVID-19 virus and assisting police by reducing their physical surveillance effort in containment zones and public areas where surveillance is required via camera feeds. The system ensures that people using masks are detected accurately and in real-time Furthermore, the solution is simple to integrate into any current company system while maintaining the security and privacy of users' data. This proposed approach will work effectively in the current scenario when the lockdown is being lifted and will aid in the automated tracking of public locations. We've gone through the tracking of social distancing and identifying face masks that aid in human health in detail. As a result, the Social Distancing and Face Mask Detection system will be the leading digital solution for a wide range of industries, including retail, healthcare, temples, shopping malls, transportation hubs, airports, and commercial entities.

VII. REFERENCE

[1] Sohan, M. (2020). So you need datasets for your COVID-19 Detection Research using Machine Learning. ArXiv. doi:abs/2008.05906

[2] M. Cristani, A. D. Bue, V. Murino, F. Setti and A. Vinciarelli, "The Visual Social Distancing Problem," in IEEE Access, vol. 8, pp. 126876- 126886, 2020, doi: 10.1109/ACCESS.2020.3008370

[3] A. H. Ahamad, N. Zaini and M. F. A. Latip, "Person Detection for Social Distancing and Safety Violation Alert based on Segmented ROI," 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2020, pp. 113-118, doi: 10.1109/ICCSCE50387.2020.9204934.

[4] S. Sharma, M. Bhatt, and P. Sharma, "Face Recognition System Using Machine Learning Algorithm," 2020 5th International Conference on Communication and Electronics Systems (ICCES), COIMBATORE, India, 2020, pp. 1162-1168, doi: 10.1109/ICCES48766.2020.9137850

[5] D. Garg, P. Goel, S. Pandya, A. Ganatra and K. Kotecha, "A Deep Learning Approach for Face Detection using YOLO," 2018 IEEE Punecon, Pune, India, 2018, pp. 1- 4, doi: 10.1109/PUNECON.2018.8745376.

[6] H. Filali, J. Riffi, A. M. Mahraz and H. Tairi, "Multiple face detection based on machine learning," 2018 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, 2018, pp. 1-8, doi: 10.1109/ISACV.2018.8354058

[7] https://www.kaggle.com/datasets/omkargurav/face-mask-dataset

[8] Joseph Redmon, Ali Farhadi, "YOLOv3: An Incremental Improvement", University of Washington

[9] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu: "SSD: Single Shot MultiBox Detector", 2016;

[10] Rethinking Atrous Convolution for Semantic Image Segmentation, Chen LC, Papandreou G, Schroff F, Adam H. arXiv:1706.05587, 2017.

VI. ACKNOWLEDGMENT

We would like to thank Dr. Hien V. Nguyen for aiding in the development of the project through all the material we learned in class.