IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IMPROVE ACCURACY IN FACE MASK DETECTION USING DEEP LEARNING

[1]Pratyush P. Urade, [2]Dr. C. M . Raut

[1]Student , [2]Asst. Professor,

[1]Computer Engineering ,

[1]Datta Meghe College of Engineering , Navi Mumbai, India

Abstract: Coronavirus illness 2019 has affected the glob seriously. One major protection technique for individuals is to wear masks publicly areas. what is more, several public service suppliers need customers to use the service provided that they wear masks properly. However, there square measure solely some analysis studies regarding mask detection supported image analysis. during this paper, we tend to propose mask, that may be a high-accuracy and economical mask detector. The projected mask may be a one-stage detector, that consists of a feature pyramid network to fuse high-level linguistics info with multiple feature maps, and a completely unique context attention module to target police work face masks. additionally, we tend to conjointly propose a completely unique cross-class object removal formula to reject predictions with low confidences and therefore the high intersection of union. Besides, we tend to conjointly explore the chance of implementing mask with a light-weighted neural network for embedded.

Keywords:- Feature Extraction , light-weighted neural network, Images classification.

I. INTRODUCTION

The situation report ninety six of world health organization (WHO) [1] conferred that corona-virus sickness a pair of covid19 has globally infected over 2.7 million individuals and caused over million of deaths.Additionally, there area unit many similar massive scale serious metabolic process diseases, like severe acute metabolic process syndrome (SARS) and therefore the Middle East metabolic process syndrome (MERS), that occurred within the past few years [2, 3]. Liu et al. [4] report able that the generative variety of COVID-19 is higher compared to the respiratory illness. Therefore,additional individuals area unit involved regarding their health, and public health is taken into account because the prime priority for governments [5]. luckily, Leung et al. [6] showed that the surgical face masks might cut the unfold of corona-virus. At the instant, UN agency recommends that folks ought to wear face masks if they need metabolic process symptoms, or they're taking care of the individuals with symptoms [7]. what is more, several public service suppliers need customers to use the service as long as they wear masks [5]. Therefore, mask detection has become an important digital vision task to assist the worldwide society, however analysis associated with mask detection is restricted.

The methodology of developing computer science systems area unit correlative to full stack developments approaches through completely different aspects like, offer the finish,front,forepart,side,face. Applications to the end-users and host the AI models on the rear end that area unit engineered to be running on a server. In general, atomic number 87 systems use several software system development methodologies to boost the system's performance, like victimization the agile approach to deploy identity verification systems in simply manageable environments, with the agile approach the atomic number systems result delivered in fast, constant management. The agile software system developments area unit acceptable to digital vision systems as a result of the end-users got to be concerned earlier to any examination, changes, enhancements, and mask observation refers to detect whether or not someone carrying a mask or not and what's the placement of the face [9]. the matter is closely associated with general object observation to observe the categories of objects and face detection is to detect a specific category of objects, i.e. face [10, 11]. Applications of object and face detection are often found in several areas, like autonomous driving [12], education [13], police work so on

[10]. ancient object detectors area unit typically supported handcrafted feature extractors. Viola Jones detector uses Haar feature with integral image methodology [14], whereas alternative works adopt completely different feature extractors, like bar graph of oriented gradients (HOG), scale-invariant feature remodel (SIFT) so on [15]. Recently, deep learning based mostly object detectors give consistently glorious performance and dominate the event of recent object detectors.

In this paper, we tend to projected a unique mask detector, Face Mask, that is ready to observe face masks and contribute to public care.

II. REVIEW OF LITERATURE

Different examples are explained with relevancy the involved domain, their perspective is clearly seen within the following review.

Face Mask Recognition victimization varied algorithms is complicated task. mask Recognition has variety of applications and creating it promising analysis topic. This chapter offers an outline of general methods in visual concept-based video retrieval, specializing in strategies used for detection , classification of objects , together with Region of Interest ,Boundary detection, key frame extraction and extraction of options and idea detection.

Review of idea and Theories

The work carried and revealed by completely different researchers and authors ar studied and analyzed as below.

Traditional object detection uses a multi-step method a well known detector is that the Viola-Joins detector, that is in a position to attain period of time detection [14]. The formula extracts feature by Haar feature descriptor with associate degree integral image technique, selects helpful options, and detects objects through a cascaded detector. though it utilizes integral image to facilitate the formula, it's still terribly computationally expensive . In [7] for human detection, an efficient feature extractor known as HOG is planned, that computes the directions and magnitudes of directed gradients over image cells. Later on, deformable part-based model (DPM) detects objects components so connects them to guage categories that objects belong to [15].

Rather than victimization handcrafted options, deep learning primarily based detector incontestable glorious performance recently, thanks to its strength and high feature extraction capability. There are 2 fashionable classes, one-stage object detectors and two-stage object detectors. Two-stage detector generates region proposals within the initial stage so fine-tune these proposals within the second stage. The two-stage detector will give high detection performance however with low speed. The seminal work R-CNN is planned by R. Girshick et al. [15]. R-CNN uses selective search to propose some candidate regions which can contain objects. After that, the proposals ar fed into a CNN model to extract options, and a support vector machine (SVM) is employed to acknowledge categories of objects. However, the second-stage of R-CNN is computationally expensive , since the network must discover proposals on a one-by-one manner and uses a separate SVM for final classification. quick R-CNN solves this downside by introducing a part of interest (ROI) pooling layer to input all proposal regions promptly [8]. Finally, a part proposal network (RPN) is planned in quicker R-CNN to require the place of selective search, that limits the speed of such detectors. quicker R-CNN integrates every individual detection elements, like region proposal, feature extractor, detector into associate degree end-to-end neural specification. One-stage discoveror utilizes solely one neural network to detect objects. so as to attain this, some anchor boxes that specifies the quantitative relation of breadth and heights of objects ought to be predefined. instead of the two-stage detector, one-stage detectors scarify the performance slightly to boost the detection speed considerably. so as to attain the goal, YOLO divided the image into many cells so tried to match the anchor boxes to things for every cell, however this approach isn't smart for tiny objects. The researchers found that one-stage detector doesn't perform well by victimization the last feature output solely, as a result of the last feature map has fastened receptive fields, which may solely observe sure areas on original pictures. Therefore, multi-scale detection has been introduced in SSD, that conducts discoverion on many feature maps to permit to detect faces in numerous sizes. Later on, so as to boost detection accuracy, Lin et. al [9] proposes tissue layer Network (RetinaNet) by combining associate degree SSD and FPN design, that conjointly embody a unique focal loss operate to mitigate category imbalance downside.

CNN plays a vital role in laptop vision connected pattern recognition tasks, owing to its superior spatial feature extraction capability and fewer computation value [10]. CNN uses convolution kernels to convolute with the initial pictures or feature maps to extract higher-level options. However, a way to style higher convolutional neural network architectures still remains as a gap question. beginning network planned in [11] permits the network to find out the simplest combination of kernels. so as to coach abundant deeper neural networks, K. He et al. propose the Residual Network (ResNet) [11], which may learn associate degree identity mapping from the previous layer. As object detectors ar typically deployed on mobile or embedded devices, wherever the procedure resources ar terribly restricted, Mobile Network (MobileNet) [13] is planned. It uses depth-wise convolution to extract options and channel wised convolutions to regulate channel numbers, therefore the procedure value of MobileNet is far under networks victimization commonplace convolutions

.

Background Research:

Traditional Object Detection:

The problem of detective work multiple disguised and unmasked faces in pictures is resolved by a conventional object detection model. the method of object detection principally involves localizing the objects in pictures and classifying them (in case of multiple objects). ancient algorithms like Haar Cascade (Viola and Jones, 2001) and HOG (Dalal and Triggs, 2005) have well-tried to be effective for such tasks, however these algorithms ar heavily supported Feature Engineering. within the era of Deep learning, it's doable to coach Neural Networks that beat out these algorithms, and don't would like any further Feature Engineering.

Convolutional Neural Networks:

Convolutional Neural Networks (CNNs) (LeCun et al., 1998) could be a key side in fashionable laptop Vision tasks like pattern object detection, image classification, pattern recognition tasks, etc. A CNN uses convolution kernels to convolute with the first pictures or feature maps to extract higher-level options, so leading to a awfully powerful tool for laptop Vision tasks.

Modern Object Detection Algorithms:

CNN primarily based object detection algorithms is classified into a pair of categories: Multi-Stage Detectors and Single-Stage Detectors.
Multi-Stage Detectors: in an exceedingly multi-stage detector, the method of detection is split into multiple steps. A two-stage detector like RCNN (Girshick et al., 2014) initial estimates and proposes a collection of regions of interest mistreatment selective search. The CNN feature vectors ar then extracted from every region severally. Multiple algorithms supported Regional Proposal Network like quick RCNN (Girshick, 2015) and quicker RCNN (Ren et al., 2015) have achieved higher accuracy and higher results than most single stage detectors.
Single-Stage Detectors: A single-stage detector performs detections in one step, directly over a dense sampling of doable locations. These algorithms skip the region proposal stage employed in multi-stage detectors and ar so thought of to be usually quicker, at the price of some loss of accuracy. one in all the foremost standard single-stage algorithms, you merely Look Once (YOLO) (Redmon et al., 2016), was introduced in 2015 and achieved on the brink of period of time performance. Single Shot Detector (SSD) (Liu et al., 2016) is another standard algorithmic rule used for object detection, which supplies wonderful results. RetinaNet (Lin et al., 2017b), one in all the simplest detectors, relies on Feature Pyramid Networks (Lin et al., 2017a), and uses focal loss.

Face Mask Detection

As the world began implementing preventative measures against the Coronavirus, varied implementations of mask Detection systems came forth. (Ejaz et al., 2019) have performed face recognition on cloaked and unmasked faces exploitation Principal part Analysis (PCA). However, the popularity accuracy drops to but seventieth once the recognized face is cloaked. (Qin and Li, 2020) introduced a technique to spot mask sporting conditions. They divided the facemask sporting conditions into 3 categories: correct mask sporting, incorrect mask sporting, and no mask sporting. Their system takes a picture, detects and crops faces, and so uses SRCNet (Dong et al., 2016) to perform image super-resolution and classify them. The work by (Nieto-Rodríguez et al., 2015) given a technique that detects the presence or absence of a medical mask. the first

objective of this approach was to trigger associate alert just for medical employees WHO don't wear a surgical mask, by minimizing as several false positive face detection as attainable, while not missing any medical mask detection. (Loey et al., 2021) planned a model that consists of 2 parts. the primary part performs uses ResNet50 (He et al., 2016) for feature extraction. succeeding part could be a mask classifier, supported associate ensemble of classical Machine Learning algorithms. The authors evaluated their system and calculable that Deep Transfer Learning approaches would win higher results since the building, comparing, and choosing the simplest model among a group of classical Machine Learning models could be a time overwhelming method.

III. PROBLEM DEFINITION

In this project, we proposed a singular mask detector, Peeper mask, which is prepared to detect face masks and contribute to public healthcare. To the foremost effective of our knowledge, Peeper mask is one in every of the first dedicated mask detectors. In terms of the spec,Peeper mask uses multiple feature maps so utilizes feature pyramid network (FPN) to fuse the high-level semantic information. to comprehend better detection, we propose a context attention detection head and a cross-class object removal algorithm to spice up the detection ability. Furthermore, since the mask data-set may well be a comparatively small data-set where features could even be hard to extract, we use transfer learning to transfer the learned kernels from networks trained for a similar face detection task on an exhaustive data-set . The proposed method is tested on masked and unmasked dataset.Experiment results shows that PeeperFaceMask achieves state-of-the art results, which is 2:3 percentage and 1:5 percentage beyond the baseline end in face and mask detection precision respectively, and 11:0 percentage and 5:9 percentage above baseline for recall. Objective of the Project To propose a way which is able to help the authorities and officials to acknowledge persons who aren't wearing a mask and take necessary action the system that will work far more easy to control the mob and avoid spread of disease.

IV. PROPOSED SYSTEM

we have mentioned strategies used for recognition and classification of object in image . conjointly we have got mentioned System styles that helps to understand the event of the system. Methodology Used The algorithms used for the event of the thesis square measure CNN at the side of transfer learning. What is CNN? CNN might be a range of deep learning model for process information that has a grid pattern, like pictures, that is galvanized by the organization of animal visual cortex and designed to mechanically and adaptively learn abstraction hierarchies of options, from low- to high-level patterns. CNN might be a mathematical construct that is usually composed of 3 kinds of layers (or building blocks): convolution, pooling, and totally connected layers. the first 2, convolution and pooling layers, perform feature extraction, whereas the third, a completely connected layer, maps the extracted options into final output, like classification. A convolution layer plays a key role in CNN, that consists of a stack of mathematical operations, like convolution, a specialised variety of linear operation. In digital pictures, picture element values square measure keep during a} very two-dimensional (2D) grid, i.e., Associate in Nursing array of numbers (Fig. 2), and alittle grid of parameters known as kernel, Associate in Nursing optimizable feature extractor, is applied at every image position, that makes CNNs extremely economical for image process, since a feature might occur anyplace among the image. reciprocally layer feeds its output into the following layer, extracted options will hierarchically and additional} become more advanced. the tactic of optimizing parameters like kernels is known as coaching, that is performed thus on minimize the distinction between outputs Associate in Nursingd ground truth labels through an improvement algorithmic rule known as back-propagation and gradient descent, among others. What is Transfer Learning? In the supervised learning position of machine learning, if we are going to train a model for some task and domain AA, we have a tendency to assume that we have a tendency to square measure given labeled information for the identical task and domain. we are going to see this clearly in Figure one, wherever the task and domain of the coaching and take a look at information of our model AA is that the identical. we'll later outline in further detail what specifically a task and an internet site are). For the moment, permit United States to assume that a task is that the target our model aims to perform, e.g. acknowledge objects in pictures, and an internet site is wherever our information is returning from, e.g. pictures taken in purpose of entry occasional outlets.

Figure 1: The traditional supervised learning setup in ML

We can currently train a model AA on this information-set and expect it to perform well on unseen data of constant task and domain. On another occasion, once given information for a few different task or domain BB, we have a tendency to need once more tagged information of constant task or domain that we will use to coach a brand new model BB so that we will expect it to perform well on this information. The traditional supervised learning paradigm breaks down when we have a tendency to|once we|after we} don't have decent tagged information for the task or domain we care close to train a reliable model. If we wish to coach a model to find pedestrians on night-time pictures, we have a tendency to might apply a model that has been trained on an identical domain, e.g. on day-time pictures. In observe, however, we frequently expertise a deterioration or collapse in performance because the model has genetic the bias of its coaching information and doesn't skills to generalize to the new domain. If we wish to coach a model to perform a brand new task, like police investigation bicyclists, we have a tendency to cannot even utilize Associate in Nursing existing model, because the labels between the tasks disagree. Transfer learning permits United States to trot out these eventualities by investment the already existing tagged information of some connected task or domain. we have a tendency to try and store this data gained in resolution the supply task within the supply domain and apply it to our downside of interest as are often seen in Figure two.

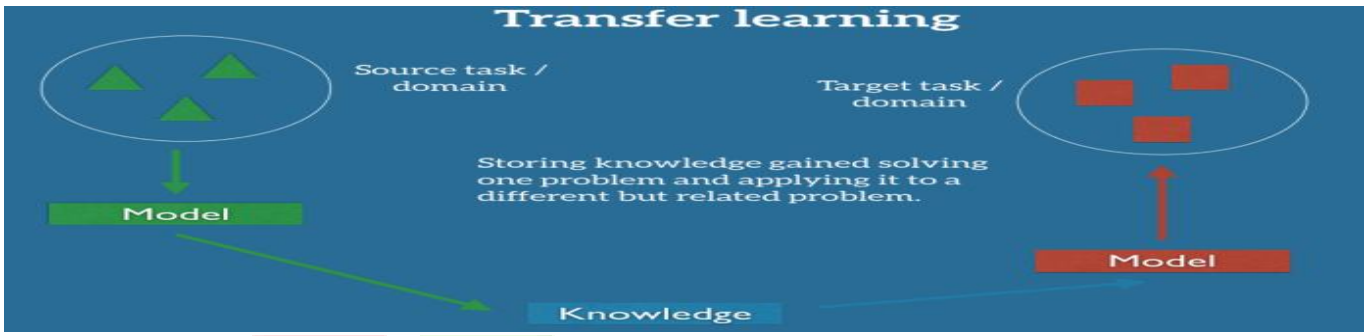

Figure 2: The transfer learning setup

In follow, we have a tendency to ask for to transfer the maximum amount data as we will from the supply setting to our target task or domain. this data will fight numerous forms counting on the data: it will pertain to however objects area unit composed to permit North American nation to a lot of simply establish novel objects; it are often with relation to the final words individuals use to specific their opinions, etc.

V. IMPLEMENTATIONENVIRONMENT

Table 1. Hardware Components

Type	Description
Machine Processor	Intel(R) Core(TM) i3-8145U CPU @ 2.10GHz 2.30 GHz or higher
Memory	8 GB DDR 4 RAM
GPU	Graphic Processing unit (we are using Google colab for the same)

Table 1 shows the setup of the framework and apparatuses utilized for the usage. The machine utilized is a Think Pad XPS having Intel i3 processor and came pre-production with 8 GB RAM alongside Windows 10 master working framework.

Table 2. Software Components

Type	Description
Python	**Python 3.6.3 bundled with Anaconda and Jupyter Notebook**
Machine OS	**Windows 10 Pro, 64 bit**

Table 2 shows the setup of the framework and apparatuses utilized for the usage. The software used python 3.6.3 , Operating system Windows 10 Pro, 64 bit.

Google Colab is a free cloud service and it supports free GPU.we can improve your Python programming language coding skills. develop deep learning applications using popular libraries such as Keras, TensorFlow, PyTorch, and OpenCV.We use Google colab for developing the training model.

VI. EXPERIMENTAL EVALUATION

The aim of the chapter is probably the results of mask detection using both the models and analyse how well each method performs and output of the program. This section presents the results of the methodologies.
Data-set of images is stored in two folders ie with mask and without mask we've divided the knowledge set into 80% train data and 20 %test or validation data are used for performance computing . Train data are used for both trainning and validation(test) of the model which the test data are used for performance computation .It has been observed that the proposed model attains maximum performance after 11 epoches and maintains consistent learning.

This part explains the results of per-processing the information then training on data.we are using data for images and these images are converted to arrays and stored in array called data, at this stage labels are still in alphabet so we would like to convert them additionally in to array.we give path of our data set of mask and without mask . A function pre-processing is defined, which takes the folder to the dataset as input, then loads all the files from the folder and resizes the pictures according to the model. Then after the record is sorted, the pictures are converted into tensors are geometric objects that show linear relationship between vectors, scalers and domains. Then the list is transfigured to NumPy array for faster computation. After implementing pre-processing, the perfection of our model increased remarkably. After this, the tactic of information augmentation is applied to increase the accuracy after training the model.

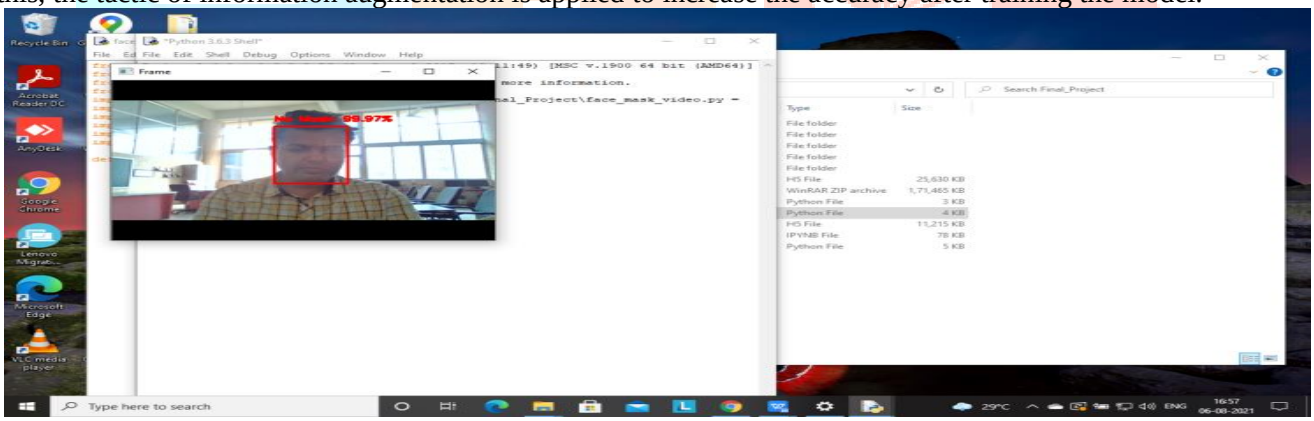

Figure3 : Prediction of without mask person on video screen shot

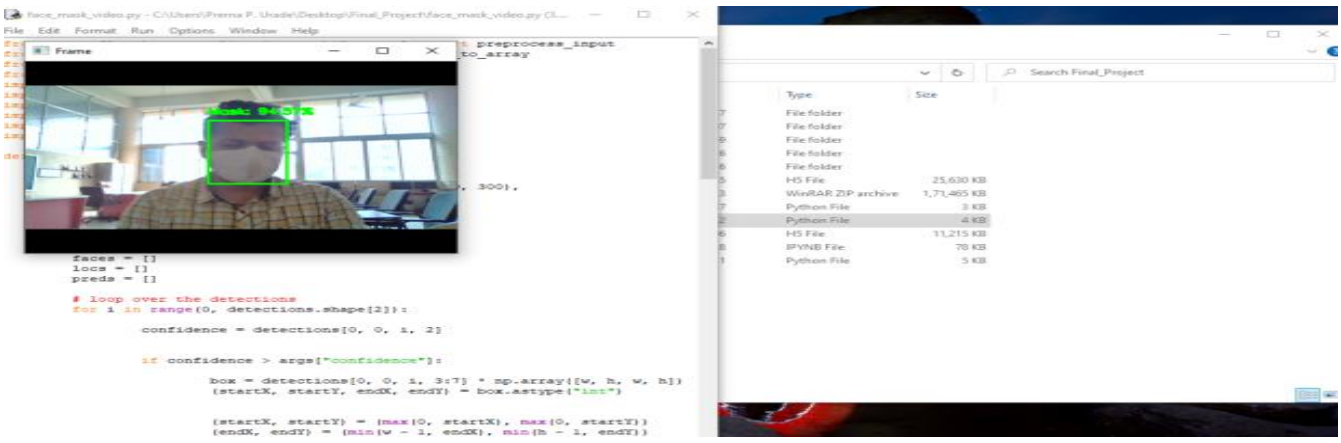

Fig4: Prediction of with mask

Comparison of Model :

There are several measures to be performed of results . we have considered accuracy on training data set which we have considered 80 percent of data set and validation or testing data set which is 20 percent of data set. The metrics have been computed using true positive, true negative , false positive , and false negative obtained from matrix Accuracy is sum of true positive and true negative over the total number of samples.

Accuracy = $\frac{TP+TN}{TP+FP+FN+TN}$S

Table3:- Face mask classification Training and validation statistics

Model	Training		Validation	
	Acc (%)	Loss	Acc(%)	Loss
CNN Model	98.17	.0255	98.31	.5943
Proposed Model	99.75	.0066	99.31	.0400

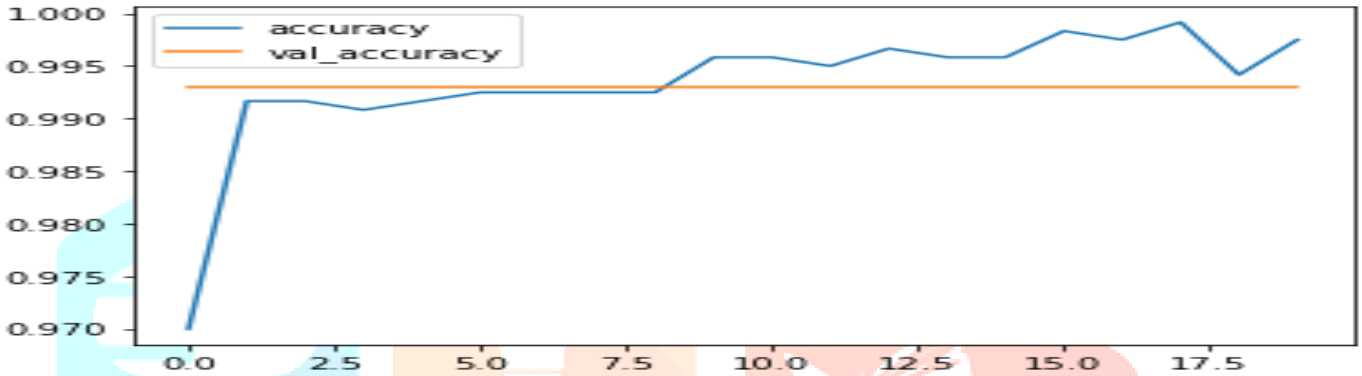

Figure 5 Training and validation accuracy

Training and Validation accuracy is graphically shown in above figure

Based on Table no , we can say that above both models have performed well and achieved good statistics . The combined model of CNN and Mobile net overall slightly better numbers com-paired to CNN model .

VII. CONCLUSION

In this project, we've projected a completely unique mask detector, particularly Peeper FaceMask Detection System, which may probably contribute to public attention. The design of Peeper FaceMask Detection System consists of CNN as backbone, FPN because the neck, and context attention modules because the heads. The robust backbone ResNet and lightweight backbone MobileNet will be used for prime and low computation situations, severally. so as to extract additional sturdy options, we tend to utilize transfer learning to adopt weights from the same task face detection, that is trained on a really giant dataset. moreover, we've projected a completely unique context attention head module to concentrate on the face and mask options, and a completely unique formula object removal cross category, i.e. ORCC, to get rid of objects with lower confidence and better note. The projected technique achieves progressive results on a public mask dataset. associate degree Iot part is supplemental to the System to show the messages during a hardware. the information from the program is collected by the Iot hardware and it'll dispalyed because the alert to the involved authority.

VIII REFERENCES

[1] W. H. Organization et al., "Coronavirus disease 2019 (covid-19): situation report, 96," 2020.

[2] P. A. Rota, M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M.-h. Chen et al., "Characterization of a novel coronavirus associated with severe acute respiratory syndrome," science, vol. 300, no. 5624, pp. 1394–1399, 2003.

[3] Z. A. Memish, A. I. Zumla, R. F. Al-Hakeem, A. A. Al-Rabeeah, and G. M. Stephens, "Family cluster of middle east respiratory syndrome coronavirus infections," New England Journal of Medicine, vol. 368, no. 26, pp. 2487–2494, 2013.

[4] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of covid-19 is higher compared to sars coronavirus," Journal of travel medicine, 2020.

[5] Y. Fang, Y. Nie, and M. Penny, "Transmission dynamics of the covid-19 outbreak and effectiveness of government interventions: A data-driven analysis," Journal of medical virology, vol. 92, no. 6, pp. 645–659, 2020.

[6] N. H. Leung, D. K. Chu, E. Y. Shiu, K.-H. Chan, J. J. McDevitt, B. J. Hau, H.-L. Yen, Y. Li, D. KM, J. Ip et al., "Respiratory virus shedding in exhaled breath and efficacy of face masks."

[7] S. Feng, C. Shen, N. Xia, W. Song, M. Fan, and B. J. Cowling, "Rational use of face masks in the covid-19 pandemic," The Lancet Respiratory Medicine, 2020.

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

[9] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei et al., "Masked face recognition dataset and application," arXiv preprint arXiv:2003.09093, 2020.

[10] 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.

[11] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: a review," Artificial Intelligence Review, vol. 52, no. 2, pp. 927–948, 2019.

[12] D.-H. Lee, K.-L. Chen, K.-H. Liou, C.-L. Liu, and J.-L. Liu, "Deep learning and control algorithms of direct perception for autonomous driving," arXiv preprint arXiv:1910.12031, 2019.

[13] K. Savita, N. A. Hasbullah, S. M. Taib, A. I. Z. Abidin, and M. Muniandy, "How's the turnout to the class? a face detection system for universities," in 2018 IEEE Conference on e-Learning, e-Management and e-Services (IC3e). IEEE, 2018, pp. 179–184.

[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol. 1. IEEE, 2001, pp. I–I.

[15] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in 2008 IEEE ConferenSce on Computer Vision and Pattern Recognition. IEEE, 2008, pp. 1–8.