ISSN [Online]: 2583-2654

Face Mask Detection using CNN and Temperature Screening

[1]Sharmila Kumari N, [2]Charls Reynold J, [3]Jananey B, [4]Poovarasi S, [5]Preethi N.
Dept. of Computer Science and Engineering,
Dr. T. Thimmaiah Institute of Technology,
Karnataka

Abstract: Face Mask Detection has evolved as a very popular problem in Image Processing and Computer Vision. Many new algorithms are being devised using Convolutional Architectures have made it possible to extract even that pixel details. We aim to design a Binary Face Classifier which can detect any Face Mask present in the Frame irrespective of its alignment. We present a method to generate accurate Face Segmentation Mask from any arbitrary size input image. Beginning from the RGB image of any size, the Method uses Predefined Training Weights of MobileNetV2 Architecture for Feature Extraction. Training is performed through Fully Convolutional Networks to Semantically Segment out the faces present in the image. Data Augmentation is a technique to artificially create new training data from existing training data. Gradient Descent is used for training while Binomial Cross Entropy is used as a Loss Function. Further the output image from the FCN is processed to remove the Unwanted Noise and False Prediction if any and make Bounding Box around the Faces.

The Ultrasonic Sensor sends out 8 pulses of Ultrasonic sound when you pull the trigger line high these Sound Waves travel with the speed of sound. When the waves hit an obstacle, they bounce back and the sensor receives the waves. The sensor then pulls the echo pin high for few milliseconds. When connecting this sensor to an Raspberry pi, it is possible to measure the time between sending and receiving the pulses. Once we detect a person, then with the use of temperature sensor will detect the temperature of that person. This project describes an efficient and economic approach of using Machine Learning to create safe environment in a manufacturing setup.

Keywords: Binary Face Classifier, Semantic Segmentation, CNN, FCN, MobileNetV2, Data Augmentation, Hyperparameter Optimization, Ultrasonic Sensor, Temperature Sensor, Raspberry pi.

I. INTRODUCTION

Face mask detection has emerged as a very interesting problem in image processing and computer vision. It has a range of applications from facial motion capture to face recognition which at the start needs the face to be detected with a very good accuracy. Face detection is more relevant today because it not only used on images but also in video applications like real time surveillance and face detection in videos. High accuracy image classification is possible now with the advancements of Convolutional networks. Pixel level information is often required after face detection which most face detection methods fail to provide. Obtaining pixel level details has been a challenging part in semantic segmentation. Semantic segmentation is the process of assigning a label to each pixel of the image. In our case the labels are either face or non-face. Semantic segmentation is thus used to separate out the face by classifying each pixel of the image as face or background. Also, most of the widely used face

ISSN [Online]: 2583-2654

detection algorithms tend to focus on the detection of frontal faces.We propose a model for face detection using semantic segmentation in an image by classifying each pixel as face and non-face i.e. effectively creating a binary classifier and then detecting that segmented area. The model works very well not only for images having frontal faces but also for non-frontal faces. The paper also focuses on removing the erroneous predictions which are bound to occur. Semantic segmentation of human face is performed with the help of a fully convolutional network [10] [11] [12]. The ultrasonic sensor sends out 8 pulses of ultrasonic sound when you pull the trigger line high, these sound waves travel with the speed of sound.

When the waves hit an obstacle, they bounce back, and the sensor receives the waves. The sensor then pulls the echo pin high for a few milliseconds. When connecting this sensor to a Raspberry pi, it is possible to measure the time between sending and receiving the pulses. Once we detect a person, with the use of Infrared sensor will detect the temperature of that person.

MobileNetV1 is a family of general-purpose computer vision neural networks designed with mobile devices in mind to support classification, detection and more. The ability to run deep networks on personal mobile devices improves user experience, offering anytime, anywhere access, with additional benefits for security, privacy, and energy consumption. As new applications emerge allowing users to interact with the real world in real time, so does the need for ever more efficient neural networks. Today the availability of MobileNetV2 to power the next generation of mobile vision applications. MobileNetV2 is a significant improvement over MobileNetV1 and pushes the state of the art for mobile visual recognition including classification, object detection and semantic segmentation.

MobileNetV2 builds upon the ideas from MobileNetV1, using depthwise separable convolution as efficient building blocks. However, V2 introduces two features to the architecture:

- Linear bottlenecks between the layers.
- Shortcut connections between the bottlenecks.

In figure1 the blue blocks represent composite convolutional building blocks. The intuition is that the bottlenecks encode the model's intermediate inputs and outputs while the inner layer encapsulates the model's ability to transform from lower-level concepts such as pixels to higher level descriptors such as image categories. Finally, as with traditional residual connections, shortcuts enable faster training and better accuracy. Overall, the MobileNetV2 models are faster for the same accuracy across the entire latency spectrum. In particular, the new models use 2x fewer operations, need 30% fewer parameters and are about 30-40% faster on a google pixel phone than MobileNetV1 models, all while achieving higher accuracy.

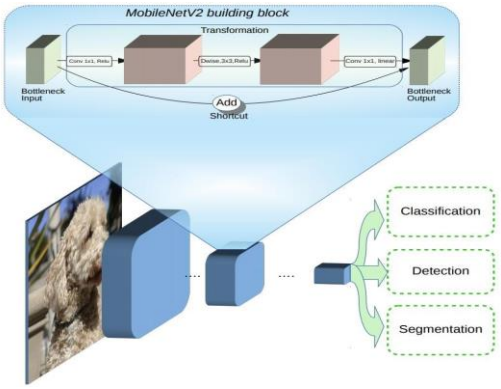

Fig 1.1 MobileNetV2 Architecture

ISSN [Online]: 2583-2654

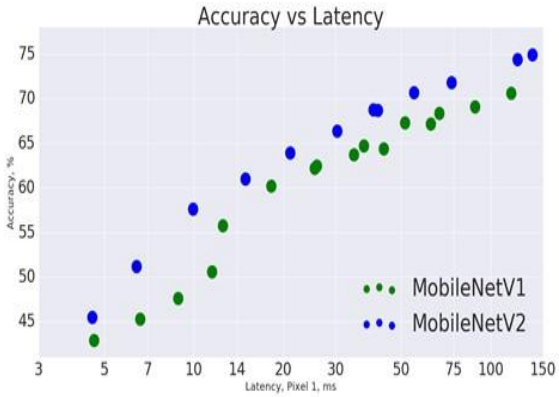

Fig 1.2 Accuracy graph of MobileNetV2

The figure 2 illustrates that the MobileNetV2 improves speed (reduces latency) and increased ImageNet top 1 accuracy. MobileNetV2 is a very effective feature extractor for object detection and segmentation. For example, for detection when paired with newly introduced SSDLite the new model is about 35% faster with the same accuracy than MobileNetV1. To enable on-device semantic segmentation, we employ MobileNetV2 as a feature extractor in a reduced form of DeepLabv3, that was announced recently. On the semantic segmentation benchmark, PASCAL VOC 2012, our resulting model attains a similar performance as employing MobileNetV1 as feature extractor but require 5.3 times fewer parameters and 5.3 times fewer parameter operations in terms of Multiply-Adds. MobileNetV2 provides aa very efficient mobile-oriented model that can be used as a base for many visual recognition tasks.

II. LITERATURE SURVEY

A literature review is brief summary of previous research on a topic. The literature review surveys scholarly articles and other resources relevant to a particular area of research. The work should enumerate, summarize and objectively evaluate the previous research.

T. Ojala, M. Pietikainen, and T. Maenppa, [1], published "Multiresolution grayscale and rotation invariant texture classification with local binary patterns," IEEE transactions a Pattern Analysis and Machine Intelligence, July 2002. This paper presents a theoretically very simple, yet efficient, multiresolution approach to gray-scale and rotation invariant texture classification based on local binary patterns and nonparametric discrimination of sample and prototype distributions. T. H. Kim, D. C. Park, D. M. Woo, T. Jeong, and S. y. Min, [2] published "Multiclass classifier-based adaboost algorithm," in Proceedings of the Second Sino foreign interchange Conference on Intelligent Science and Intelligent Data Engineering, ser. IScIDE'11. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 122–127. A multi-class classifier based AdaBoost algorithm for the efficient classification of multi-class data is proposed in this paper. The proposed AdaBoost architecture can save its training time drastically and obtain more stable and more accurate classification results than a typical multi-class AdaBoost architecture based on binary weak classifiers. P. Viola and M. J. Jones, [3], "Robust real-time face detection," Int. J. Computer. Vision, vol. 57, no. 2, pp. 137–154, May 2004. This face detection system is most clearly distinguished from previous approaches in its ability to detect faces extremely rapidly. Operating on 384 by 288-pixel images, faces are detected at 15 frames per second on a conventional 700 MHz Intel Pentium III. In other face detection systems, auxiliary information, such as image differences in video sequences, or pixel color in color images, have been used to achieve high frame rates. P. Viola and M. Jones, [4], published "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, This paper describes a

ISSN [Online]: 2583-2654

machine learning approach for visual object detection which is capable of processing images extremely rapidly and achieving high detection rates. This work is distinguished by three key contributions. Li, J. Zhao, Y. Wei, C. Lang, Y. Li, and J. Feng, [5] published "Towards real world human parsing: Multiple-human parsing in the wild," CoRR, vol. abs/1705.07206. In this paper, we design a novel semantic neural tree for human parsing, which uses a tree architecture to encode physiological structure of human body and designs a course to fine process in a cascade manner to generate accurate results.

A. Krizhevsky, I. Sutskever, and G. E. Hinton, [6] published "Image net classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. J. Simonyan and A. Zisserman, [7] published "Very deep convolutional networks for large-scale image recognition," CoRR, vol. abs/1409.1556, 2014.In this work we investigate the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting. Our main contribution is a thorough evaluation of networks of increasing depth using an architecture with very small (3x3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16-19 weight layers. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, [8] published "Going deeper with convolutions," 2015.We propose a deep convolutional neural network architecture codenamed "Inception", which was responsible for setting the new state of the art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC 2014). The main hallmark of this architecture is the improved utilization of the computing resources inside the network. This was achieved by a carefully crafted design that allows for increasing the depth and width of the network while keeping the computational budget constant. K. He, X. Zhang, S. Ren, and J. Sun [9], published "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016. Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. K. Li, G. Ding, and H. Wang, [10], published "L-fcn: A lightweight fully convolutional network for biomedical semantic segmentation," in 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Dec 2018, pp. 2363–2367. For the past few years, deep learning-based methods have been widely used in the field of biomedical imaging. In biomedical image processing, the typical application of deep learning is semantic segmentation. However, the classical deep learning methods require higher hardware consumption and computational costs.

III. SYSTEM ARCHITECTURE

The Figure 3 shown below illustrates the architecture of face mask detection. The input image samples are collected and given as input to the system. The input image will be split into two classes by applying the MobileNetV2 architecture and the extracted feature from the input image will be edge which is the subset of image. Training

ISSN [Online]: 2583-2654

Process will be done using Classification Technique.

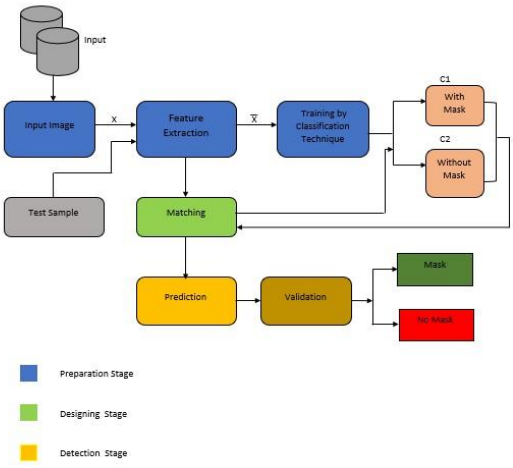

Fig 3 System Architecture

IV. DATA FLOW DIAGRAM

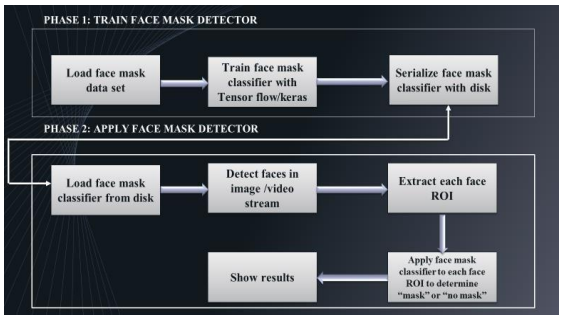

Fig 4 Data Flow diagram

The above diagram 4 illustrates the phases and individual steps for building COVID19 face mask detector with computer vision and deep learning using Python, OpenCV, and TensorFlow/Keras.: To train a custom face mask detector, we need to break our project into two distinct phases, each with its own respective sub-steps:

- **Training:** here we'll focus on loading our face mask detection dataset from disk, training a model (using Keras/TensorFlow) on this dataset, and then serializing the face mask detector to disk.
- **Deployment:** once the face mask detector is trained, we can then move on to loading the mask detector, performing face detection, and then classifying each face as with mask or without mask.

V. METHODOLOGY

The Face Mask Recognition is developed with a machine learning algorithm through image classification method MobilenetV2. MobileNetV2 is a method based on Convolutional Neural Network that developed by Google with Improved performance and enhancement to be more efficient. This study conducted its experiments on two original datasets. The first dataset was taken from the Kaggle dataset and the real-world masked face dataset; used for the training, validation and testing phase. The model can be produced by following some steps which are as follows:

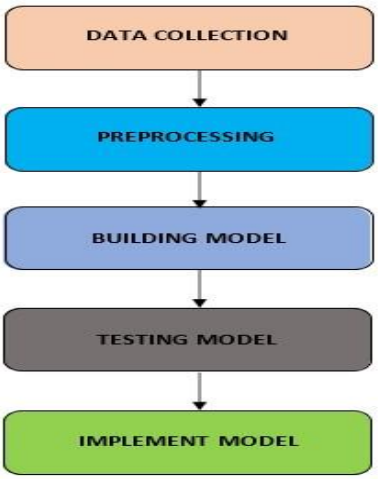

Fig 5 Flow Chart of face mask detection

5.1 Object Detection

An ultrasonic sensor is an electronic device that measures the distance of a target object by emitting ultrasonic sound waves and converts the reflected sound into an electrical signal. Ultrasonic waves

ISSN [Online]: 2583-2654

travel faster than the speed of audible Ultrasonic sensors have two main components: the transmitter and the receiver (which encounters the sound after it has travelled to and from the target).

5.2 Data Collection

The data collection consists of "with mask" & "without mask" images. The development of the face mask recognition model begins with collecting the data. The model will differentiate between people wearing masks and not. For building the model, this study uses 70% of data with mask and 30% data without a mask.

5.3 Pre- Processing

The pre- processing phase is a phase before the training and testing of the data. There are four steps in the pre-processing which are

• Resizing image: The Resizing image is a critical pre-processing step in computer vision due to the effectiveness of training models..

• Converting Image to array: The image is converted into array for calling them by the loop function. After that, the image will be used to Pre-process input using MobileNetV2.

• Performing Hot Encoding on Labels: The last step in this phase is performing hot encoding on labels because many machine learning algorithms cannot operate on data labelling directly. They require all input variables and output variables to be numeric, including this algorithm.

5.3 Building The Model

- Constructing the training image generator for augmentation, Data Augmentation is a technique to artificially create new training data from existing training data.
- Training the model: Facial landmarks allow us to automatically infer the location of facial structures.

5.4 Temperature Screening

The DHT11 is a basic, ultra low-cost digital temperature and humidity sensor. It uses capacitive humidity sensor and a thermistor to measure the surrounding air, and spits out a digital signal on the data pin (no analog input pins needed). Its fairly simple to use, but requires careful timing to grab data. The only real downside of this sensor is you can only get new data from it once every 2 seconds, so when using our library, sensor readings can be up to 2 seconds old.

Compared to the DHT22, this sensor is less precise, less accurate and works in a smaller range of temperature/humidity, but its smaller and less expensive comes with a 4.7k or 10k resistor.

VI. RESULTS

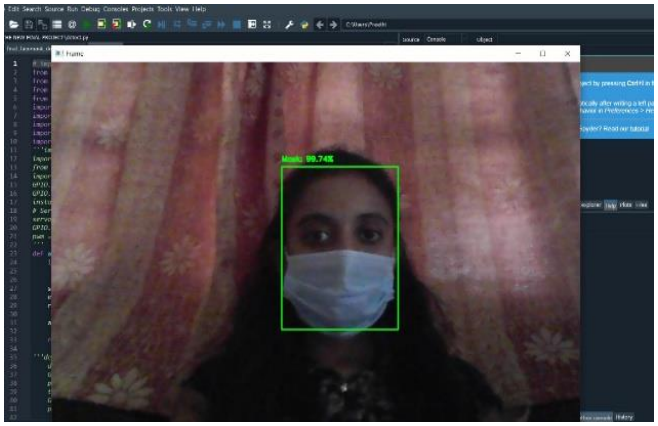

Fig 6.1 Image Representing with Mask.

ISSN [Online]: 2583-2654

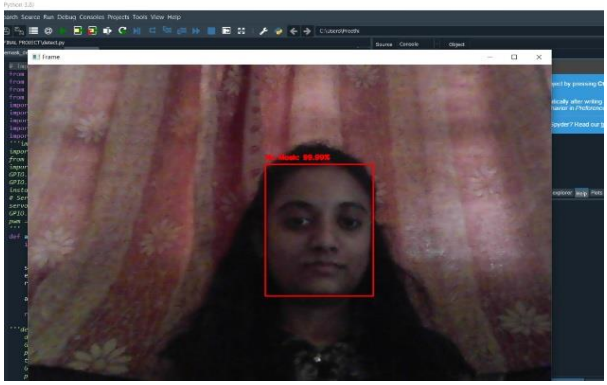

Fig 6.2 Image Representing without Mask.

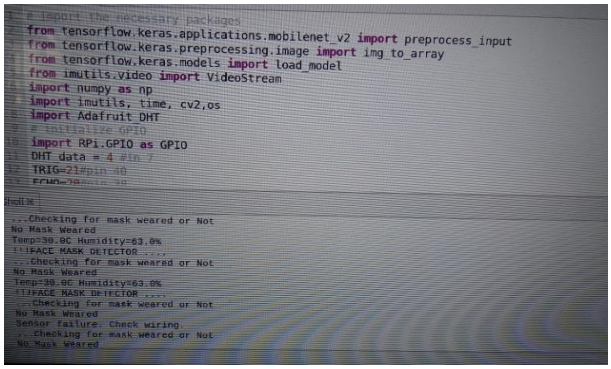

Fig 6.3 Image Representing the Screened Temperature.

VII. CONCLUSION

Covid-19 has become a pandemic and it is now spreading rapidly through direct and indirect contacts among individuals. Our proposed model is a practical approach for rapid screening of people with an automated system. The modules of our proposed system are for detecting the person wearing mask or not in two classes, which can play a vital role in controlling and tracing the person who may suspect of covid-19. The system consists of module for temperature screening which uses DHT11 sensor in measuring temperature of a person. Our Face Mask classification and detection gave us accuracy of 99.74% , by using MobileNetV2, CNN architecture for Face Mask detection.

In this work, the Face Net pre-trained model has been used for improving masked face recognition. We have benchmarked this approach with two well-known datasets and our dataset. Our approach tested on these datasets shows better recognition rates. So Face Net model trained on masked and non-masked images gives better accuracy for simple masked face recognition. mustache, and medical mask, our methodology can still be extended to more complex and many other sources of occlusion.

Obviously, this method may not be appeasement for all types of masks. Further, the more accurate and sophisticated approach may than be needed. In later work, it is our importance to enhance and enlarge our work to address different extreme masks condition of face recognition.

REFERENCES

[1] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and 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] T.-H. Kim, D.-C. Park, D.-M. Woo, T. Jeong, and S.-Y. Min, "Multi-class classifierbased adaboost algorithm," in Proceedings of the Second Sinoforeign-interchange Conference on Intelligent Science and Intelligent Data Engineering, ser. IScIDE'11. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 122–127.

[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] 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, Dec 2001, pp. I–I.

[5] J. Li, J. Zhao, Y. Wei, C. Lang, Y. Li, and J. Feng, "Towards real world human parsing: Multiple-human parsing in the wild," CoRR, vol. abs/1705.07206.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," CoRR, vol. abs/1409.1556, 2014.
[8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2015.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770– 778, 2016.

[10] K. Li, G. Ding, and H. Wang, "L-fcn: A lightweight fully convolutional network for biomedical semantic segmentation," in 2018 IEEE International

ISSN [Online]: 2583-2654

Conference on Bioinformatics and Biomedicine (BIBM), Dec 2018, pp. 2363–2367

[11] X. Fu and H. Qu, "Research on semantic segmentation of high-resolution remote sensing image based on full convolutional neural network," in 2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE), Dec 2018, pp. 1–4.

[12] S. Kumar, A. Negi, J. N. Singh, and H. Verma, "A deep learning for brain tumor mri images semantic segmentation using fcn," in 2018 4th International Conference on Computing Communication and Automation (ICCCA), Dec 2018, pp. 1–4.

[13] Samrat Kumar Dey, Arpita Howlader, "MobileNet Mask: A Multiphase Face Mask Detection Model to Prevent Person- To- Person Transmission of SARS- Conv2," in 2020 17th December International Conference Paper.

[14] Ming Yui Cheng, Lung S Chan, I J Lauder, Cyrus Rustam Kumana, "Detection of Body Temperature with Infrared Thermography: accuracy in detection of fever," in 2012 Hong Kong Medical Journal.

[15] Arun Francis G, Arulselvan M, ElangKumaran P, Keerthivarman S, Vijaya Kumar J, "Object Detection Using Ultrasonic Sensor," in 2019 International Journal of Innovative and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue- 6S, April 2019.