IMPLEMENTATION OF REAL-TIME FACE MASK DETECTION USING CONVNET

Narendra Ilaya Pallavan.P , Abhirame.V.S, Arvindh.J.J , Karthic Kumar.K

Department of Bio Medical Engineering,

Bannari Amman Institute of Technology,

Sathyamangalam, India.

Abstract

World-level medical emergency in healthcare is caused by this second wave pandemic. This virus mainly spreads through salivary droplets which come from a person with corona virus while sneezing and causes the risk to others. The rate of risk of transmission is very high in public places. Its negative impact was felt by almost all commercial establishments, education, economy, religion, transport, tourism, employment, entertainment, food security, and other industries. In this project, we propose a way which uses TensorFlow and OpenCV to detect whether people are wearing Mask or not. An abounding box will come over the face of the person which describes whether the person is wearing a mask or not. The increased requirement for an efficient system that detects whether a person wearing face masks in densely populated areas, transportation means, large-scale manufacturers, residential districts, and other enterprises to ensure maximum safety. This project uses a machine-learning Algorithm called Convolution neural network using OpenCV, Numpy, MatlibPlot, Keras modules on Jupyter Notebook to detect facemasks on people.

Keywords:Keras, OpenCV, TensorFlow and Anaconda Software.

1. INTRODUCTION

By the use of cascade classifier and a deep learning algorithm, real time face mask detection implementation is described in this proposal. Nowadays, we make ourselves comfortable to live with pervasive COVID-19. As per Government guidelines, many precautionary measures are instructed for this cause whereas wearing a mask is one among them. However, since people aren't following it correctly, monitoring is necessary. It is designed for people who are all intended to track people for wearing masks. They can now sit anywhere and monitor easily thereby clearly giving guidance.

Anaconda Software is installed and appropriate libraries are installed. Data preprocessing, Training the model and detecting faces and masks are the three steps to be followed to implement this proposed idea. The required portion from the live images/Video Stream were detected using a cascade classifier [1]. ROI (Region of Interest) image is converted as a grey image and resized into a constant size in order to enhance the image prediction. The Convolution Neural Network algorithm is a model pre-trained by collection of pre-proceeded training images with labels that displays the result (with or without mask that ROI image has). We can capture live images and videos with OpenCV. This proposed method would be indispensable in hospitals, shopping malls, theatres, colleges, schools, and banks, among other locations.

This introduced face mask detection model is based on machine learning and image processing techniques. The planned model will detect masks with testing images/videos as well as detect people wearing masks and not wearing masks in real time images and videos. With Oven CV, Tensor Flow, and Keras, the model blends deep learning and conventional machine learning techniques[2]. We compared various algorithms for object detection to find out the most suitable algorithm, that should achieve the highest accuracy as well as consumption of least amount of time to execute and detection part.

The rest of the paper is formatted as given below. Literature survey on various papers is discussed in Section II. The related studies taken and algorithm analysis are mentioned in Section III. In Section IV, methodology is discussed briefly. In Section V, the implemented algorithms are discussed [3]. In Section VI, experimental procedure is explained in detailed manner. In Section VII, Result of our proposed project is discussed and brings our work to a close, as well as the prospect of future work.

2. RELATED STUDIES

2.1 ARTIFICIAL INTELLIGENCE

Artificial intelligence is used in a variety of ways in almost all the fields. It is becoming increasingly important to the world, the particular reason for this circumstance is that, it efficiently perform complex problems in numerous sectors consists of healthcare, entertainment, banking, and education. Human lives are turning more peaceful and productive as a benefit of artificial intelligence. Artificial Intelligence can be classified in a variety of ways. The most common categorizations are based on skills as well as it is based on its functionality as shown in Fig 3.1.

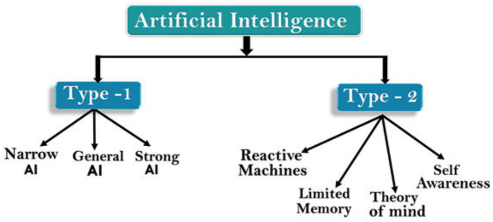

Fig 2.1 – Classification of AI

2.2 MACHINE LEARNING

Machine Learning algorithms are programme that itself learn hidden patterns in data, estimate output, and boost performance based on previous experiments. In machine learning, various algorithms can be used for multiple tasks, such as basic linear regression for prediction tasks like stock market prediction whereas KNN algorithm can be used for the classification problem [10]. On the basis of functionality, it is classified into three types.

- Supervised learning algorithms
- Un-supervised learning algorithms
- Reinforcement learning algorithm

2.2.1 SUPERVISED LEARNING ALGORITHM

In supervised learning, the machine learns under some external supervision. The labelled dataset is used to train the supervised learning models. After the model has been trained and processed, it is checked by presenting a sample of test data to detect if it correctly predicts the performance.In simple terms, the aim of the supervised algorithm is to map input data to output data.

The topic of supervised learning can be further divided into two categories:

- Classification
- Regression

Simple Linear Regression, Decision Tree, Logistic Regression, KNN algorithm, are some of the examples for supervised learning algorithms.

2.2.2 UNSUPERVISED LEARNING ALGORITHM

In unsupervised learning, the machine learns from data without the assistance of a human. Unsupervised models can be trained with an unlabeled dataset that is neither graded nor categorized, and the algorithm must operate on the data without supervision. The model in unsupervised learning doesn't have a predefined output and instead attempts to extract useful information from a large amount of data. These are used to solve the problems of association and clustering. As a result, it can be divided into two types

- Clustering
- Association

K-means Clustering, Apriori Algorithm, Eclat and few like this are the best examples for the unsupervised learning algorithms.

2.2.3 REINFORCEMENT LEARNING

In reinforcement learning, an entity interacts with its surroundings by producing actions and learns from the input it receives. The feedback is provided to the agent in the form of incentives. The agent does not receive any supervision. In reinforcement, Q-learning algorithm can be used. Some of the most common machine learning algorithms are as follows. Linear Regression Algorithm, Logistic Regression Algorithm, Decision Tree, PCA (Principal Component Analysis) etc.,

2.3IMAGE PROCESSING

Image processing is a technique for performing operations on a picture in attempt to optimize it or obtain useful information from it. "It is the analysis and multiplication of a digitized image, particularly to enhance its quality" is the central theme of image processing.

2.4OPENCV

OpenCV is a large open-source python library for computer vision, machine learning, and image processing, so it acts as a catalyst in real-time applications, which are critical in today's modern environment. It can be used to recognize objects, faces, and even person handwriting in photographs and videos. We can use vector space and it undertakes mathematical computations over the parameters to identify pictures, patterns as well as its different features.

2.5DEEP LEARNING

It is currently one among the most common subjects in data science, artificial intelligence etc. It is a subfield of ml that consists series of algorithms for learning data representations [8]. It has a structure and functions that are identical to those of the human brain is represented in Fig 3.2. In deep learning, we mainly use neural networks. We use this method to train networks to identify text, data, figures, sound, and other types of data [5]. It is used in some of today's most groundbreaking technical advances, such as robotics, computer vision, image recognition, and a slew of other stuff.

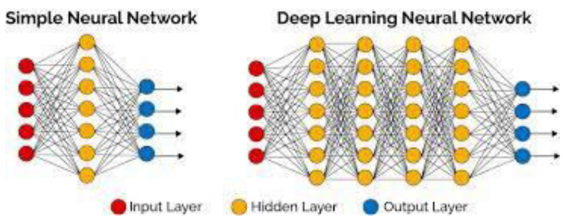

Fig 2.2 - Deep Neural Network

There are several deep learning libraries available, but TensorFlow, Keras, and PyTorch are the most common among all.

2.6ALGORITHMS FOR OBJECT DETECTION

It is one among computer vision that permits to find our required region getting from either real-time image/video or by sample image. The algorithm analysis for the above specific application is given as follows.

2.6.1 FAST R-CNN

Fast Region-Based Convolutional Network System, or Fast R-CNN, is a training algorithm for object detection written in two languages Python and C++ (Caffe) [6]. This algorithm primarily tackles the shortcomings of simple R-CNN while also concern.

2.6.2 HISTOGRAM OF ORIENTED GRADIENT (HOG)

The histogram of oriented gradients (HOG) is literally a efficient classifier that is used to predict objects. The Histogram of oriented gradients descriptor technique captures gradient orientation in specific areas in image like detection window and the region of interest (ROI).

2.6.3 REGION-BASED CONVOLUTION NEURAL NETWORKS (R-CNN)

The Region Convolutional Network system (RCNN) combines region proposals with Convolution Neural Networks to form the Region-based Convolutional Network method (CNNs). With only minimal amount of annotated detection data, R-CNN can help localize objects using a network and train a high-performing model [9]. It achieves great object detection accuracy without rely on approximate techniques like hashing.

2.6.4 REGION-BASED FULLY CONVOLUTION NETWORK (R-FCN)

R-FCN (Region-based Fully Convolutional Networks) is a segment-based object detector. Lest make a look over other region-based detectors like Fast R-CNN , which use a high amount per-region sub network, this segment-based detector is completely convolutional, with almost all computation shared across the overall image.

2.6.5 SINGLE SHOT DETECTOR (SSD)

The Single Shot Detector (SSD) is a neural network method to detect objects in images. Over a range of aspect ratios, the SSD concept discretizes the output space of bounding boxes into a set of default boxes. The method counts per feature map location after discretization. The Single Shot Detector network combines assumptions from multiple feature maps of varying orientation into a single network.

2.6.6 SPATIAL PYRAMID POOLING

SPP-net (Spatial Pyramid Pooling) is a network structure that produces a fixed-length representation irrespective of image size or scale. Pyramid pooling is said to be insensitive to deformations of objects, it outperforms all Convolution based image classification methods. Researchers may use SPP-net to create standard-length representations to train the detectors by computing feature maps from the entire image once and then pooling features in negative regions (background-images). This approach prevents having to compute the convolution features several times.

3. ALGORITHMS USED

3.1 CASCADE CLASSIFIER

Cascading is a type of ensemble learning that involves concatenating the outputs of many classifiers and using all of the information gathered as additional information for the next classifier in the cascade. Cascading is a multistage system, in contrast to casting a ballot or stacking groups, which are multiexpert systems. Several hundred "positive" example perspectives on a single object and arbitrary "negative" images of the same size are used to train cascading classifiers [4]. After the classifier has been educated, it can be used to detect the object in concern by applying it to a region of an image. The search window can be shifted around the image to check any position for the classifier to find the object in the entire picture. The cascade classifier is made up of stages, each of which is made up of a group of weak learners. The weak learners are decision stumps, which are basic classifiers. Boosting is a technique used to train each level [14]. By taking a weighted average of the decisions taken by the weak learners, boosting allows you to train a highly accurate classifier. Each stage of the classifier assigns a positive or negative label to the region identified by the sliding window's current position. A positive value indicates that an object was discovered, while a negative value indicates that no objects were discovered. If the label is negative, the detector moves the window to the next position after completing the classification of this area. The classifier passes if the mark is positive. When the final stage classifies the region as positive, the detector reports an object located at the current window spot. The stages are set up in such a way that negative samples are rejected as soon as possible. The premise is that the point of interest is not visible in the vast majority of windows.

1. True positives, on the other hand, are uncommon and worth investigating.
2. When a positive sample is correctly labelled, it is considered a true positive.
3. When a positive sample is wrongly labelled as negative, it is known as a false negative.

Each stage in the cascade must have a low false negative rate in order for it to function properly. If a stage incorrectly marks an object as negative, the classification process is halted, and you are unable to correct the error. Each point, however, has a high rate of false positives [11]. When a negative sample is wrongly labelled as positive, it is considered a false positive. And if the detector wrongly marks a nonobject as positive, the error can be corrected later.

3.2 CONVOLUTION NEURAL NETWORKS

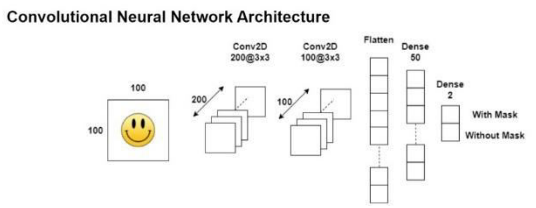

Fig 3.1 - CNN architecture

Convolutional neural organizations (CNNs, or ConvNets) shown in Fig 5.1 are a type of profound neural organization used to examine visual symbolism in profound learning. Focused on the common weight engineering of the convolution parts that move over input includes and give interpretation equivariant reactions, they are otherwise called shift invariant or space invariant fake neural organizations (SIANN). Image and video recognition, recommender systems, image detection, image segmentation, medical image interpretation, natural language processing, brain-computer interfaces, and financial time series are only some of the fields where they can be used [12]. Multilayer perceptions are regularized variants of CNNs. Multi-facet discernments are normally totally associated networks, implying that every neuron in one layer is connected to all neurons in the following layer. These organizations' "full availability" makes them powerless against information overfitting. Regularization, or avoiding overfitting, can be accomplished in a variety of ways, including penalizing parameters during preparation (such as weight loss) or trimming compatibility (skipped connections, dropout, etc.) CNNs take a different approach to regularization. As a consequence, CNNs are at the lower end of the connectivity and complexity scale. In the way that the communication pattern between neurons resembles the organisation of the animal visual cortex, convolutional networks were inspired by biological processes. Individual cortical neurons respond to improvements simply in the open field, which is a little bit of the visual field [7]. Various neurons' responsive fields mostly cover, permitting them to possess the whole visual field. This infers that the framework figures out how to redo the channels (or bits) through programmed learning, rather than hand-designed channels in regular calculations. This absence of dependence on earlier aptitude or human inclusion in highlight extraction is a critical increase. In contrast with other picture grouping calculations, CNNs need almost no pre-handling.

4. METHOD OF EXPERIMENTAL STUDIES

Our project explains the design, implementation steps, and executing methodology of the face mask detector in the following steps. Convolution Neural Network is one among the neural network architecture highly effective in object detection applications. Cascade classifier is the second efficient object detection algorithm. The live image from the webcam/CCTV is sent to the cascade classifier. The Cascade classifier's sliding window moves and locates the face (ROI) from the image. The ROI is pre-processed into equal size and sent to a pre-trained CNN mask classifier. CNN has two convolution layers. One is 200 conv2D and another one is 100 conv2D. Both convolution layers having

3*3-pixel kernels. Various platforms are available like google colab, anaconda, etc to train the model. On account of capturing live images and videos, we used anaconda software because google colab has an absence of webcam functionality. Anaconda software is installed for Windows operating system 64 GB. Once it is installed, the Jupyter notebook is installed in anaconda command prompt using the command "PIP INSTALL JUPYTER NOTEBOOK". Jupyter notebook is an interactive interface to code and run their codes. Let's see our project execution in three phases[13]. Phase 1 includes the data pre-processing, phase 2 includes training the model and the final phase includes detection and output presentation. It is important to concern that all the images given for training (datasets) must be pre-processed. Similarly, testing images were also pre-processed for the fast response. Required libraries like OpenCV, matplotlib, Keras, TensorFlow, NumPy, etc are installed in anaconda software with anaconda command prompt using PIP install command.

4.1 PHASE 1 – DATA PREPROCESSING

Sample data image is taken. Colour is one of the major factorsthat may cause some deviation in the output. To neglect this, all the images are converted into grey using RGB to GREY converter so that all the images are maintained in a constant way. Region of Interest portion may differ from size. To make it as constant size, all images are re-sized into 100*100 (pixel). This pre-processing is done for all the training images provided by the user (data set) as well as for the testing images. Image pre-processing steps is shown in Fig 4.1 for visual understanding.

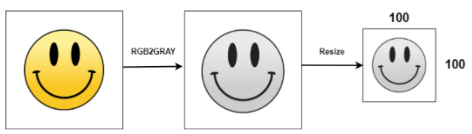

Fig 4.1 – Data pre-processing

4.2 PHASE 2 - TRAINING THE MODEL

Cascade classifier and Convolution Neural Network both algorithms require training. We need to provide hundreds and thousands of sample images with labels to train the model and make itself understand the object. In order to ensure whether mask is in the face, extract the face from the image must be done first. Cascade classifier moves in the entire frame through its search window [15]. If the result is negative, it moves to the next set of pixels in the same frame. If the result is positive, it goes to the next stage and if all the stage in the set of pixels is positive, then cascade classifier confirmed that the image is found. This is how a cascade classifier works. CNN is made up of many numbers of convolution layers. Here, we used two convolution layer-based CNN. First convolution layer is 200 Conv 2D which has 200 kernels with size 3*3. It extracts the feature in each kernel and produce the result. Second convolution layer is 100 Conv 2D which has 100 kernels with 3*3. It extracts the feature in each kernel and produce the result. Max pooling is done on the extracted feature and the layer is flattened and dense and end up with displaying the output.

4.3 PHASE 3 - TESTING

For data pre-processing, training and detection, python coding is done and compilation and running was done on Jupyter notebook. Many Epochs were created with varying accuracy and loss while training. We'll be having high accuracy low loss epoch among other epochs.Live images are captured with webcam and our trained model monitored the whole frame, set the identifier and predicted the correct output. The executed methodology is shown in Fig 4.2 for great understanding.

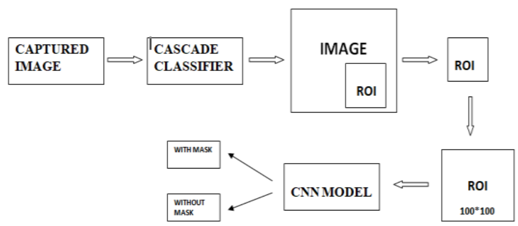

Fig 4.2 – Executed methodology of face mask detector

5. RESULT AND FUTURE ADVANCEMENT

The model is trained, validated and tested upon two datasets. Corresponding to dataset 1, the method attains accuracy of 95.77% optimized accuracy that overcomes the percentage of error. The main cause behind in achieving this accuracy lies in MaxPooling concept. From the Fig 5.1 and Fig5.2, we able to graphically visualize the accuracy and loss while training the model. X axis denotes the epochs (number of samples we used for training) and Y axis represents the accuracy (in Fig 5.2) and loss (in Fig 5.1) with respect to the individual epoch performance.

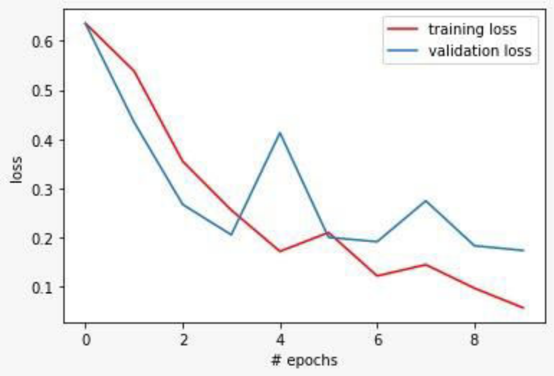

Fig 5.1 - Epochs Vs. Loss

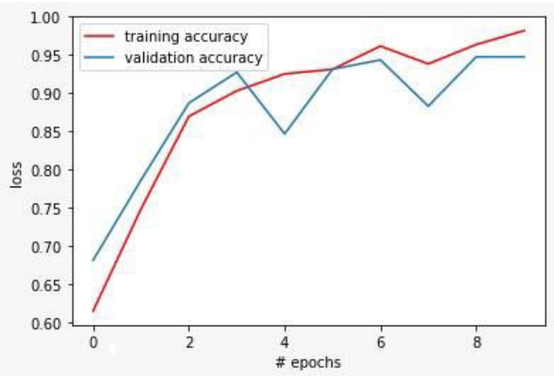

Fig 5.2 - Epochs Vs. Accuracy

The model is tested with real time live images and video streams and the output is shown in the Fig 8.3. It is going to play a crucial role in almost all the fields in future.

Fig 5.3 - Real-Time face mask detection (with Mask)

Fig 5.4 - Real-Time face mask detection (without Mask)

The Existing Projects detects whether a person wears a mask or not which is a Boolean value. We have Certain features to this project to make this even more powerful and Useful. We all know that every citizen in India has Aadhar card and it has every details of that person like Photo, Biometrics, Phone number and Address of that particular Individual. So,what we can do is, if possible, the government can give the software access to this database like it is given access to banks for doing KYC, and this program can be ran in public CCTV cameras. Now the camera can identify the individuals passing in that area and since it has access to the Aadhar Database, it can check whether the person is wearing mask or not and if He / She is not wearing the mask , by the use of the DB, it can access the data of the Individuals and it can check the phone number and it can send the penalty bill to the individual person and send a copy to the respective police station, Even in case if the person is not paying the penalty, we still have his address and the police men can tract him and collect fine from him. By this way the people will start to wear masks even if there are no police men in those surroundings. It also employs almost no people for monitoring and it is very feasible method for monitoring a country with a huge population like India. The further development will be to develop such mask detector with features containing whether the mask is virus efficient or not i.e., the type of the mask is medically approved one like N95 etc.

REFERENCES

[1]Imran Ahmed, Misbah Ahmad,Joel J.P.C. Rodrigues and Gwanggil Jeon, (2021) '*A deep learning based covid distance monitoring framework for COVID-19*', Journal of Sustainable Cities and Society, Vol. 65 pp. 113-127.

[2] P. Viola and M. Jones (2001) '*Rapid object detection using a boosted cascade of simple feautures*', IEEE Computer Society conference on Computer Vision and Pattern Recognition, Pp. 511-518.

[3] T. Ojala, M. Pietikainen, and D. Harwood(1996) '*A Comparative Study of Texture Measures with Classification Based on Feature Distribution*', Pattern Recognition, Vol. 29, Pp. 51–59.

[4] I. Gangopadhyay, A. Chatteijee, and I.Das (2019) '*Face Detection and Expression Recognition Using Haar Cascade Classifier and Fisherface Algorithm*', Recent Trends in Signal and Image Processing. Advances in Intelligent Systems and Computing, Vol. 922, Springer

[5] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, H. Chen, Y. Miao, Z. Huang, and J. Liang (2020) '*Masked Face Recognition Dataset and Application*',arXiv preprint arXiv:2003.09093, 2020.

[6]Dmitriy Anisimov, Tatiana Khanova (2017) '*Towards lightweight convolutional neural networks for object detection*', arXivpreprintarXiv:1707.01395 .

[7] M.O. Oloyede, G.P. Hancke and H.C. Myburgh(2018) '*Improving face recognition systems using a new image enhancement technique hybrid features and the convolution neural network*', IEEE Access, Vol. 6, pp. 75181-75191.

[8] Md ZahangirAlom, Tarek M. Taha, Christopher Yakopcic, Stefan Westberg, PahedingSidike, MstShamima Nasrin, Brian C Van Esesn, Abdul A S. Awwal, Vijayan K. Asari(2018) '*The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches'*,arXiv preprintarXiv:1803.01164.

[9]Shiming Ge, Jia Li, QitingYel, Zhao Luo (2010) '*Detecting masked faces in the wild with LLE-CNNs*', IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[10] Changhao Piao, Xianhao Wang, Mingjie Liu (2019) '*Pedestrian Detection using optimized YOLOv3 in UAV scenario*', International Conference on Internet Computing, Automation and Systems (ICICAS),2019.

[11] Romulus Lungu, Mihai Lungu, Lucian and TeodorGrigorie, (2013*) 'Thermal image human detection using Haar-cascade classifier'*,7th International Annual Engineering Seminar (InAES), Yogyakarta, pp. 1-6.

[12] A. Mohamed, A. Issam, B. Mohamed, and B. Abdellatif(2015) ''Real-time Detection of Vehicles Using the Haar-like Features and Artificial Neuron Networks', Procedia Computer Science,Vol. 73, pp. 24-31.
[13] B.S.B. Dewantara and J. Miura(2016)*'Estimating Head Orientation using a Combination of Multiple Cues'*, IEICE Trans, on Information and Systems, vol. E99-D, no. 6, pp. 1603-1614.

[14] B.S.B. Dewantara and J. Miura(2015), *'The AISL Head Orientation Database and Preliminary Evaluation*',IEEE International Electronic Symposium (IES), Pp. 146-150.

[15]B. Jun, I. Choi, D. Kim (2012) '*Local transform features and hybridization for accurate face and human detection*',IEEE Transactions on Pattern Analysis and Machine Learning, Vol. 35, No. 6, pp. 1423-1436.