International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 29 Issue 04 – MAY 2022.

A Unified multi-scale deep convolutional neural network for Face Mask detection

Lokesh [1], Shiva Rajan [2], Sridhar [3], Vishnu [4*], Dr. D.M.D. Preethi [5]

Students - Department of CSE - PSNA, Dindigul [1 2 3 4]
Associate Professor - Department of CSE - PSNA, Dindigul [5]

Abstract - The purpose of our project is to develop an application for detection of face mask using Deep Learning. Our project tries to minimize the exposure to diseases by promoting people to wear face mask while maintaining distance from others. The pretrained models such as the MobileNet, Convolutional Neural Network, and Visual Geometry Group are used in our context. People who try to disobey and causing ruckus among the public without wearing masks are detected using this application. As the number of COVID-19 cases increased, all nations tried to control the epidemic cautiously. The main cause of the spread of the disease in due to the transmission of droplet through the air which made the situation more deadly. So to prevent this from happening, people were asked to wear mask and maintain physical distance from each other. The system performance is calculated in terms of precision, specificity, F1-score, sensitivity, support, recall and accuracy that shows the real-time applicability.

Keywords: Convolutional Neural Network, Deep Learning, F1-score, MobileNet, Visual Geometry Group.

I. Introduction

The Covid pandemic caused disastrous effects on daily lives and financial aspect of the people around the world. As the number of COVID-19 cases grew in many numbers, it showed its true devastating face to the society as it was much deadlier than the other diseases human kind faced in this era. However, this could be prevented if the droplet transmission is controlled while people come closer with eachother and maintaining the social distance between people. This has proven to be a good prevention from falling prey for the pandemic and also helped people combat against the COVID-19 epidemic. After accomplishing and training the models, the selected one proved to establish a confidence score of nearly 100%.

Face mask detection is termed in identifying whether a person is wearing a mask or not. The problem is reverse engineering of face detection where the face is identified using several machine learning algorithms in the terms of authentication, surveillance, and security[4]. Face detection is a key area in the field of Computer Vision and Pattern Recognition. A significant group of researches has contributed revolutionary experience to algorithms for face detection in past[4][5]. The primary research on face detection was done in 2001 using the design of handcraft characteristics and administration of traditional machine learning algorithms to train effective classifiers for recognition and detection[8]. The problems faced with this approach include high complexity in feature design and low detection accuracy[1]. In these years, face detection methods have been extensively developed to improve detection performance using deep convolutional neural networks (CNN)[2].

The system performance is calculated in terms of precision, specificity, F1-score, sensitivity, support, recall and accuracy that shows the real-time applicability. Hence, this helps tracking people with or without masks in a real-time scenario and helps maintaining physical distancing[3]. This can be used with the existing embedded camera infrastructure which can be applied to various verticals, as well as in an office building or at outstations/gates of an building.

II. Literature Review

The world is still in recovery path from the impacts of COVID-19, without an effective sure for the disease. Several countries have come across with different kind of strategies to prevent the widespread. They have restricted various activities in the recent times which had a great impact both in the economical aspect and daily lives of the people. In these modern times, we cant focus on preventive measures to stay healthy without getting in serious trouble as the pandemic is set loose on everyone. The healthcare system is in a state of emergency. Wearing masks is one of the preventative measures to control the spreading of this disease.

The environment is monitored for the scenario where a person is detected whether he/she is wearing a mask or not. If the person tries to violate the rule, his/her face would be captured. This monitoring is done with the help of Closed-Circuit-Television(CCTV)cameras[3]. We try to reduce the widespread of the COVID-19 by identifying the person who are not wearing the masks.A large number of studies have been conducted on recognizing faces in a variety of situations, such as shifting stance or light, degraded photos, and so on. Nonetheless, the challenges posed by masks are sometimes overlooked[6]. The main focus of this research is on facial masks, specifically how to improve the recognition accuracy of various masked faces[13].

For the training, we have gathered a dataset consisting of dataset consisting the images of people with mask and people without the mask. In this Face detection method, we have labelled the dataset(with_mask, without_mask) and try to train the machine with greater accuracy and less response

time. Then we augment the existing data like changing the attributes slightly to develop more dataset with the existing image data.

III. DATASET

One dataset consisting of with_mask, without_mask have been used for training the model. It contains 1,915 image data for with_mask label and 1,918 images for without_mask label. Every data in the dataset is unique and has several facial properties and several augmented functions like attributes with different angles, lightings, coloured masks, face perspectives and facial emotions.

Fig.1. Samples from with_mask dataset

Fig.1. above, represents the samples from the dataset which is used to train the system to detect masked people.Similarly Fig.2. depicts the samples to detect the people who are not wearing any masks.

Fig.2. Samples from without_mask dataset

IV. PROPOSED SYSTEM

We propose to use the Convolutional Neural Network (CNN) algorithm to detect masks and to monitor social distancing. The proposed training model for mask detection. The testing of this model is performed on complex images including face turning, wearing glasses, beard faces, and scarf images[7]. This system is able to train the datasets of both person without mask and person with mask and can predict whether he/she is wearing a mask or not with greater accuracy. It can access the webcam to process the result accurately.

A. DEEP LEARNING:

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to “learn” from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy.

Deep learning drives many artificial intelligence (AI) applications and services that improve automation, performing analytical and physical tasks without human intervention. Deep learning technology lies behind everyday products and services (such as digital assistants, voice-enabled TV remotes, and credit card fraud detection) as well as emerging technologies (such as self-driving cars).

Face mask detection is done using Convolution Neural Networks, a Deep Learning approach (CNN). The connectivity pattern between neurons in convolutional networks is similar to the organisation of the visual cortex, which was inspired by biological processes. CNN is a type of multilayer neural network that is applied to 2-dimensional arrays (typically pictures) and is based on spatially localised neural input. CNN For pattern recognition, create "patterns of patterns[12]."

Fig.3. below explains the process of convolution neural network’s working with the feature map representation.

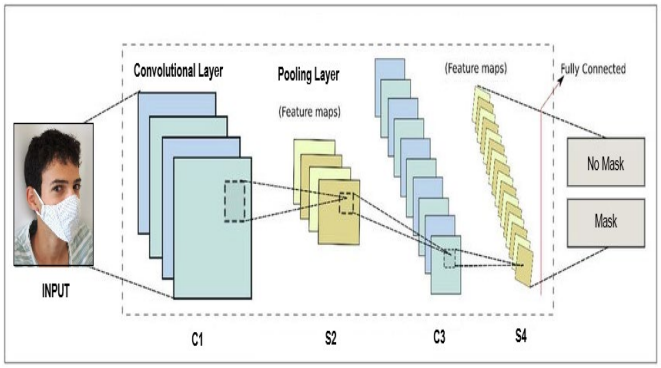

Fig.3. Feature extraction with convolution and pooling layer

Convolutional networks are multistage topologies with numerous stages that can be trained. Enter a Each stage produces feature maps, which are collections of arrays. Each feature map on the output represents a single feature taken from all input locations.

A filter bank layer, a non-linearity layer, and a feature pooling layer make up each stage. Basic structure of CNN, where it consists of two C1,C3 are convolution layers and two S2,S4 are pooled layers.

In a filter bank, a trainable filter (kernel) connects the input feature map to the output feature map.Convolutional layers perform a convolution on the input before forwarding the output to the next layer. The convolution simulates a single neuron's reaction to visual input.

B. CONVOLUTIONAL LAYER:

The CNN is a special type of neural network model designed to work on images data that can be one-dimensional, two-dimensional, and sometimes three-dimensional. Their application ranges from image and video recognition, image classification, medical image analysis, computer vision and natural language processing.

This part of the process receives the image (a matrix containing the pixel values). We optimize the image data by selecting the smaller matrix which is called “Filter”. The filter is used to multiply the original pixel value by its value. Then the values are added into a single number. In simple terms, we are trying to distinguish the visual boundaries and simple colours. Similarly several convolutional networks are blended with pooling layers in the network.

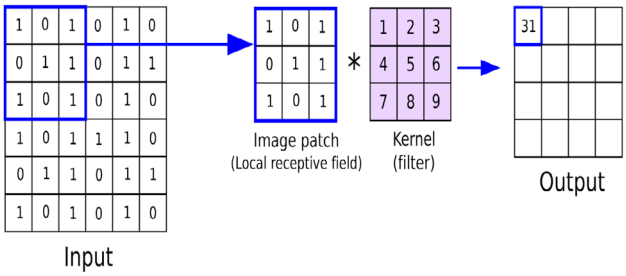

Fig.4. Convolution layer with a filter

Fig.4. describes a 6 x 6 whose image pixel values are 0, 1 and filter matrix 3 x 3 as shown above. Then the convolution of 6 x 6 image matrix multiplies with 3 x 3 filter matrix which is called “Feature Map” as output shown above.

C. POOLING LAYER:

Pooling involves selecting a pooling operation, much like a filter to be applied to feature maps. The size of the pooling operation or filter is smaller than the size of the feature map; specifically, it is almost always 2×2 pixels applied with a stride of 2 pixels[13].This means that the pooling layer will always reduce the size of each feature map by a factor of 2, e.g. each dimension is halved, reducing the number of pixels or values in each feature map to one quarter the size.

For example, a pooling layer applied to a feature map of 4×4 (16 pixels) will result in an output pooled feature map of 2×2 (4 pixels). This results in the reduced size of image(width and height).

i. **Average Pooling**: Calculate the average value for each patch on the feature map. Fig.5. explains the working of the average pooling concept.

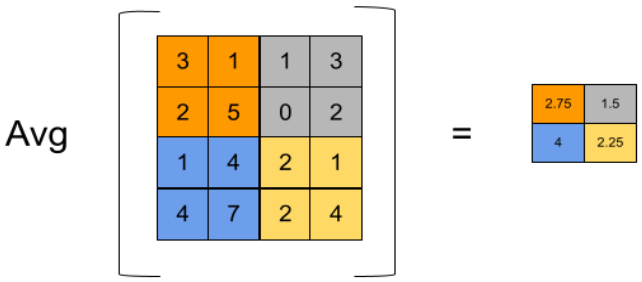

Fig.5. Average pooling

ii. **Maximum Pooling (or Max Pooling)**: Calculate the maximum value for each patch of the feature map. Fig.6. explains the working of the max pooling concept.

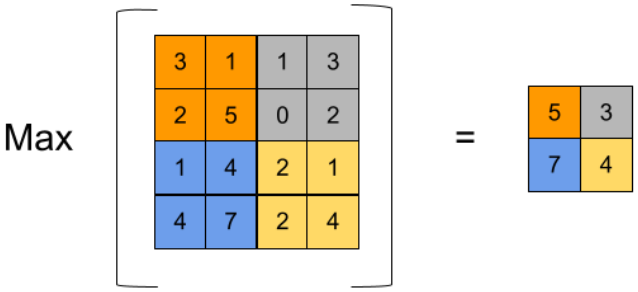

Fig.6. Max pooling

D. FULLY CONNECTED LAYER:

Fully Connected layers in a neural networks are those layers where all the inputs from one layer are connected to every activation unit of the next layer. In most popular machine learning models, the last few layers are full connected layers which compiles the data extracted by previous layers to form the final output. Fig.7. shows the fully connected layer.

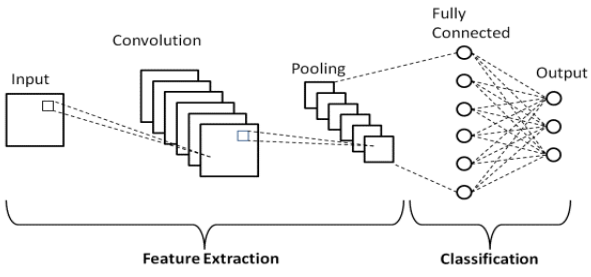

Fig.7. Fully connected layer

V. PACKAGES INCORPORATED

The Tensorflow, Keras framework and the OpenCV library were used to create this CNN model, which is widely utilised in real-time applications. This concept can also be used to create a full-fledged software that scans everyone entering a public meeting.

A. TensorFlow:

TensorFlow, an interface for expressing machine learning algorithms, is utilized for implementing ML systems into fabrication over a bunch of areas of computer science, including sentiment analysis, voice recognition, geographic information extraction, computer vision, text summarization, information retrieval, computational drug discovery and flaw detection to pursue research. It is also used to reshape the data (image) in the data processing.

B. Keras:

Keras gives fundamental reflflections and building units for creation and transportation of ML arrangements with high iteration velocity. It takes full advantage of the scalability and cross-platform capabilities of TensorFlow. The core data structures of Keras are layers and models. All the layers used in the CNN model are implemented using Keras.

C. OpenCV:

OpenCV (Open Source Computer Vision Library), an open□source computer vision and ML software library, is utilized to differentiate and recognize faces, recognize objects, group movements in recordings, trace progressive modules, follow eye gesture, track camera actions, expel red eyes from pictures taken utilizing flflash, fifind comparative pictures from an image database, perceive landscape and set up markers to overlay it with increased reality and so forth. The proposed method makes use of these features of OpenCV in resizing and color conversion of data images.

Fig.8. Packages used in the Proposed System

Fig.8. represents the packages used in the proposed system for image classification and detection. These frameworks are makes our machine learning to be efficient and easy to achieve the desired output.

VI. METHODOLOGY

Various steps involved in the proposed system is given below. These steps consists of training the machine using the data provided and building a pre-trained classification model inorder to develop a CNN model which is finally deployed to monitor the real-time scenario to detect people whether they are wearing masks or not.

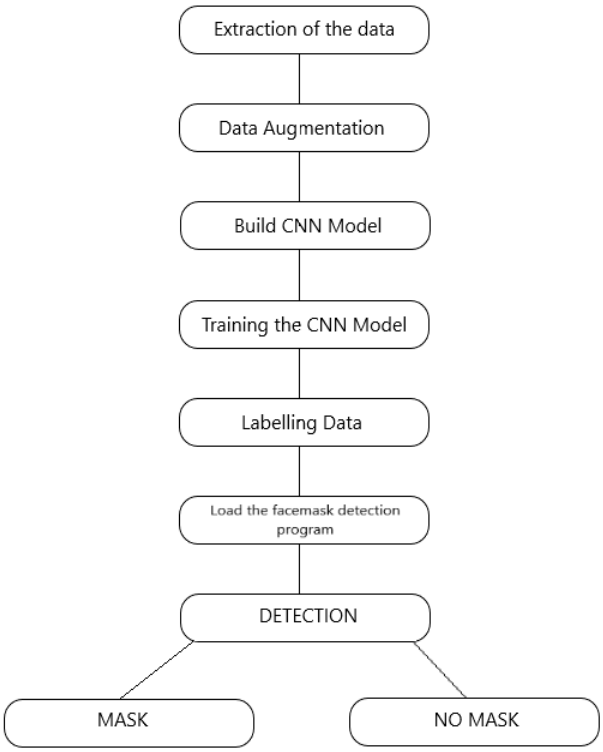

Fig.9. Methodology for the proposed system

VII. RESULT AND ANALYSIS

The proposed system is trained for 30 epochs to increase accuracy in detecting the frames. For high accuracy, we have to use more number of epochs in its training but there it occurs over-fitting.

Fig.10. Real-time detection for no mask

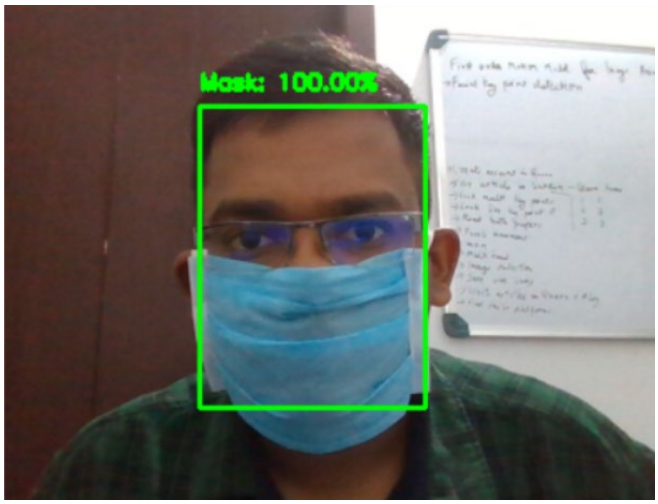

Fig.11. Real-time detection for mask

Fig.10. shows the real-time detection of a person not wearing any masks.Fig.11. shows the result of a person wearing a mark. The proposed system is effective in detecting people with higher accuracy.

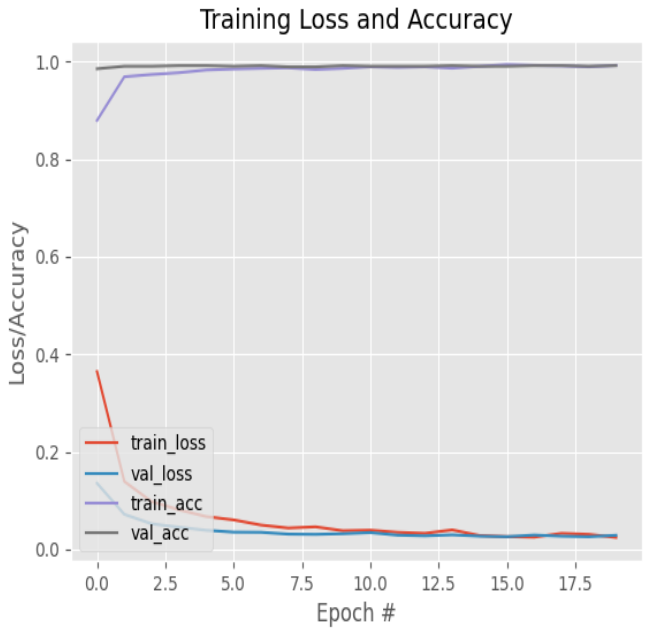

Fig.12. Training Loss and Accuracy of the proposed system

Fig.12. depicts a plotted graph with the training accuracy and loss after training a dataset model. Number of Epochs are given in X-axis while Loss/Accuracy is given in Y-axis.

VIII. CONCLUSION

The proposed system has an efficient accuracy of 99% in training under MobileNetV2. It can detect approximately the person wearing a mask or not in a real time scenario using the webcam[1][2]. This algorithm is providing contactless facial authentication and is reducing the widespread of the COVID-19 outbreak[10]. We have contributed in scientific and technological aspects inorder to improve the lifestyle of the people.

IX. REFERENCES

[1] M. S. Ejaz and M. R. Islam, "Masked Face Recognition Using Convolutional Neural Network," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), 2019, pp. 1-6, doi: 10.1109/STI47673.2019.9068044.

[2] M. R. Bhuiyan, S. A. Khushbu and M. S. Islam, "A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT).

[3] M. M. Rahman, M. M. H. Manik, M. M. Islam, S. Mahmud and J. -H. Kim, "An Automated System to Limit COVID-19 Using Facial Mask Detection in Smart City Network," 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2020.

[4] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification- verification," in Advances in neural information processing systems, 2014, pp. 198hy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," 2014.

[5] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on, pp. 138–142, IEEE, 1994.

[6] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," CoRR abs/1411.7923, 2014.

[7] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun, "A practical transfer learning algorithm for face verification," in Computer Vision (ICCV), 2013 IEEE International Conference on, pp. 3208–3215, IE8– 1996.

[8] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," ACM computing surveys (CSUR), vol. 35, no. 4, pp. 399–458, 2003.

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. KarpatEE, 2013.

[10] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," Pattern Analysis and Machine Intelligence, IEEE Transactions on 19(7), pp. 711– 720, 1997.

[11] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun, "A practical transfer learning algorithm for face verification," in Computer Vision (ICCV), 2013 IEEE International Conference on, pp. 3208–3215, IEEE, 2013.

[12] Y. Sun, X. Wang, and X. Tang. "Deep learning face representation by joint identification-verification". CoRR, abs/1406.4773, 2014.

[13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors". Nature, 1986.