ROBUST FACE MASK DETECTION USING DEEP LEARNING CNN: AN APPLICATION OF COVID-19

B. Sathwika[1], A. Tejaswini[1], A. Varsha[1], N. Baby Rani[2]

[1]UG Student, [2]Assistant Professor, [1,2]Department of Information Technology

[1,2]Malla Reddy Engineering College for Women (UGC-Autonomous), Maisammaguda, Secunderabad, Telangana, India

ABSTRACT

COVID-19 epidemic has swiftly disrupted our day-to-day lives affecting the international trade and movements. Wearing a face mask to protect one's face has become the new normal. Soon, many public service providers will expect the clients to wear masks appropriately to partake of their services. Therefore, face mask detection has become a critical duty to aid worldwide civilization. This paper provides a simple way to achieve this objective utilising some fundamental Machine Learning tools as TensorFlow, Keras, OpenCV and Scikit-Learn. The suggested technique successfully recognises the face in the image or video and then determines whether it has a mask on it. As a surveillance job performer, it can also recognise a face together with a mask in motion as well as in a video. The technique attains excellent accuracy. We investigate optimal parameter values for the Convolutional Neural Network model (CNN) to identify the existence of masks accurately without generating over-fitting.

Keywords: Covid-19, face mask detection, deep learning, OpenCV

1. INTRODUCTION

1.1 Overview

The spread of COVID-19 has resulted in more than 1,841,000 global deaths and more than 351,000 deaths in the US by Dec. 31, 2020. The spread of virus can be avoided by mitigating the effect of the virus in the environment or preventing the virus transfer from person to person by practicing physical distance and wearing face masks. WHO defined physical distancing as keeping at least 6-ft. or 2-m distance from others and recommended that keeping the physical distance and wearing a face mask can significantly reduce transmission of the COVID-19 virus. Like other sectors, the construction industry has been affected, where unnecessary projects have been suspended or mitigated people’s interaction. However, many infrastructure projects cannot be suspended due to their crucial role in people’s life. Therefore, bridge maintenance, street widening, highway rehabilitation, and other essential infrastructure projects have been activated again to keep the transportation system’s serviceability. Although infrastructure projects are activated, the safety of construction workers cannot be overlooked. Due to the high density of workers in construction projects, there is a high risk of the infection spread in construction sites. Therefore, systematic safety monitoring in infrastructure projects that ensure maintaining the physical distance and wearing face masks can enhance construction workers’ safety.

Safety agents are sometimes deployed to infrastructure projects to inspect workers to see whether they are complying with social distancing or wearing face masks. However, once there are so many workers on a construction site, it is difficult for the officers to determine hazardous situations. In addition, assigning safety officers increases the number of people on-site, raising the chance of transmission even more, and putting workers and officers in a more dangerous situation. Recently, online video capturing in construction sites has become very common. Drones are used in construction projects to record online videos to manage worksites more efficiently. The current system of online video capturing can be used for safety purposes. An automatic system that uses computer vision techniques to capture real-time safety violations from online videos can enhance infrastructure project workers' safety. This study develops a model using Faster R-CNN to detect workers who either do not wear a face mask or do not maintain the physical distance in road projects. Once a safety violation occurs, the model highlights who violates the safety rules by a red box in the video.

Taneja et. al [1] proposed and implemented a face mask detection model that can accurately detect whether a person is wearing a mask or not. The model architecture uses MobileNetV2, which is a lightweight convolutional neural network, therefore requires less computational power and can be easily embedded in computer vision systems and mobile. As a result, it can create a low-cost mask detector system that can help to identify whether a person is wearing a mask or not and act as a surveillance system as it works for both real-time images and videos. The face detector model achieved high accuracy of 99.98% on training data, 99.56% on validation data, and 99.75% on testing data. Mohamed Loey et. al [2] presented a hybrid model using deep and classical machine learning for face mask detection. The proposed model consists of two components. The first component is designed for feature extraction using Resnet50. While the second component is designed for the classification process of face masks using decision trees, Support Vector Machine (SVM), and ensemble algorithm. Three face masked datasets have been selected for investigation. The Three datasets are the Real-World Masked Face Dataset (RMFD), the Simulated Masked Face Dataset (SMFD), and the Labeled Faces in the Wild (LFW). The SVM classifier achieved 99.64% testing accuracy in RMFD. In SMFD, it achieved 99.49%, while in LFW, it achieved 100% testing accuracy.

Nowrin et. al [3] presented a narrative and meta-analytic review covering all the existing facemask detection algorithms, considering the context of Covid-19. The procedure of the existing algorithms, their considerations, effectiveness, evaluation process, and outcomes were presented. Moreover, the datasets used in those algorithms were discussed briefly. The shortcomings of the existing algorithms were reviewed, and the future challenges were outlined. Although a significant amount of research has been focused on developing an efficient facemask detection algorithm, they mainly concentrated on the same set of problems neglecting some other significant issues. This paper highlighted those shortcomings, such as, maintaining image-resolution during detection process, scarcity of rich dataset, categorical classifications, and others. Also, it specified the future scopes which includes diversity in datasets and facemask types, different facemask wearing conditions, reconstruction of the masked face, and so on. This comprehensive review will pave the way for the research community to understand the current facemask detection algorithms. By analyzing the shortcomings and future challenges in this field, researchers will develop novel approaches to fill those gaps.

1.2 Objective of the Project

Wearing a protective face mask has become a new normal. In the near future, many public service providers will ask the customers to wear masks correctly to avail of their services. Therefore, face mask detection has become a crucial task to help global society. This work presents a simplified approach to achieve this purpose using deep learning convolutional neural networks (DL-CNNs). The proposed method detects the face from the image correctly and then identifies if it has a mask on it or not.

2. LITERATURE SURVEY

Peishu Wu et. al [4] proposed an efficient automatic face mask recognition and detection framework FMD-Yolo and corresponding algorithm. A modified Res2Net structure Im-Res2Net-101 serves as the backbone to extract features with rich receptive fields from the input. Subsequently there follows a feature fusion component En-PAN, which is a novel path aggregation network and primarily consists of Yolo Residual Block, SPP, Coord Conv, SE Layer blocks. In En-PAN, high-level semantic information and low-level details can be sufficiently merged so that highly distinctive feature with strong robustness can be constructed. In addition, the localization loss function including IoU and IoU-aware loss is applied for training FMD-Yolo, and parallel computing method Matrix NMS is applied in the post-processing stage, which greatly enhances the model inference efficiency. Benchmark evaluations on two publicly available datasets and comparison with other eight state-of-the-art object detection algorithms have demonstrated the effectiveness and superiority of FMD-Yolo.

Singh et. al [5] proposed an efficient real-time deep learning-based technique to automate the process of detecting masked faces, where each masked face is identified in real-time with the help of bounding boxes. The extensive trials were conducted with popular models, namely, Faster RCNN and YOLO v3. F-RCNN has better precision, but for applying this in real-world surveillance cameras, it would be preferred to use the model with YOLO algorithm as it performs single-shot detection and has a much higher frame rate than Faster-RCNN or any other state-of-the-art object detection algorithms. If they look at the speed/accuracy tradeoff on the mAP at .5 IOU metric, one can tell YOLOv3 is better than faster R-CNN. Which model to use, also depends on the resources available. If high-end GPUs are available on the deployed devices, faster R-CNN must be used. YOLOv3 can be deployed on mobile phones also. Since this approach is highly sensitive to the spatial location of the camera, the same approach can be fine-tuned to better adjust with the corresponding field of view. These models can be used along with surveillance cameras in offices, metros, railway stations and crowded public areas to check if people are following rules and wearing marks. The trained weights provided by the authors can be further improved by training on larger datasets and can then be used in real-world applications.

Mohamed Loey et. al [6] introduced a novel model for medical masked face detection, focusing on medical mask object to prevent COVID-19 spreads from human to human. For image detection, they have employed the YOLO v2 based ResNet-50 model to produce high-performance outcomes. The proposed model improves detection performance by introducing mean IoU to estimate the best number of anchor boxes. To train and validate our detector in a supervised state, they design a new dataset based on two public masked face datasets. Furthermore, performance metrics such as AP and log-average miss rates score had been studied for SGDM and Adam optimizer experiments. They have shown that the proposed model scheme of YOLOv2 with ResNet-50 is an effective model to detect a medical masked face. As a future study, they plan to detect a kind of masked face in image and video-based on deep learning models.

Zhang et. al [7] first propose a practical and challenging dataset, which aims to reflect the conditions of wearing face mask in the era of COVID-19. Then they analyze the main challenging points in this task. Based on these analyses, they further develop Context-Attention R-CNN, a framework to detect conditions of wearing face mask, which contains three novel points: multiple context feature extractor, decoupling branches, and attention module. With these components, the Context-Attention R-CNN brings significant improvements for region-based detector. The extensive experiments show that Context-Attention R-CNN outperforms many state-of-the-art detectors, including two-stage detectors and single-stage detectors. They believe that this dataset and Context-Attention R-CNN can make contributions for preventing the COVID-19 virus from spreading. Besides the coronavirus, our work is applicable to protect against other infectious diseases which can be spread by such things as coughing, sneezing, or even speaking at close range.

Wang et. al [8] proposed a hybrid deep transfer learning and BLS for facial mask detection. It is designed to contain two stages: predetection and verification. The predetection is implemented by the Faster_RCNN framework through a transfer learning technique. The detection model is fine-tuned from a multiple-class detection model. The verification is implemented by a classifier of BLS. With a low score setting in predetection, more candidate regions are used for verification. This strategy is able to reach a tradeoff between Recall and Precision. Notably, they build a wearing mask data set containing 17 654 train masks, 1936 val masks, and 6813 test masks. The test set encompasses three sets varying from easy to hard. Experimental results shed light on our approach's effectiveness with a Recall of 93.54% and a Precision of 94.84% and advantages over the compared methods. The proposed method is expected to detect wearing masks to help realize the functions, such as noncontact temperature measurement and monitoring crowd in the pandemic era and other situations. Hopefully, our work can provide some help in the fighting against COVID-19.

Fan et. al [9] proposed a novel SL-FMDet, which is efficient and has low hardware requirements. To overcome the lower feature extraction capability caused by its light-weight backbone, they proposed RCAM and SGHR. RCAM can extract rich context information and focus on crucial face mask related areas. By using SGHR as an auxiliary task, the model is able to learn more discriminating features for faces with and without masks. The model with SGHR yielded a better attention map, which qualitatively supports the effectiveness of this auxiliary task. The proposed model achieved state-of-the-art results on two public face mask datasets, AIZOO and Moxa3K. Compared with another light-weight model, YOLOv3-tiny, the mAP of our detector is 1.7% higher on AIZOO and 10.47% higher on Moxa3K. Experimentally, they have shown that light-weight models can achieve similar or even better performance than heavy models by using RCAM and SGHR. The qualitative results also show the model is capable of tackling the challenges present in face mask detection. Therefore, the proposed face mask detector has a high potential to contribute to public health care to control the spread of COVID-19. One drawback of the method is the extra computation required for generating heatmaps and, due to limitations of the datasets, the method cannot distinguish between correct and incorrect mask wearing.

Agarwal et. al [10] proposed an intelligent face mask detector framework based on deep learning concept which can classify the person who wear mask from those who are not wearing mask. In the proposed work, a hybrid model of convolution neural network with support vector machine is used for designing the mask detector. The performance of the proposed method is evaluated on real-world masked face recognition dataset (RMFD) and medical mask dataset (MDD). When implemented, it has been found that the proposed method can achieve high accuracy (99.11%). The excellent performance of the proposed model is very suitable for video surveillance equipment also.

Razavi et. al [11] developed a computer vision system to automatically detect the violation of face mask wearing and physical distancing among construction workers to assure their safety on infrastructure projects during the pandemic. For the face mask detection, they collected and annotated

1000 images, including different types of face mask wearing, and added them to a pre-existing face mask dataset to develop a dataset of 1853 images and increased the dataset to 3300 images by data augmentation. Then, they trained and tested multiple Tensorflow state-of-the-art object detection models on the face mask dataset and chose the Faster R-CNN Inception ResNet V2 network that yielded the accuracy of 99.8%. For physical distance detection, they employed the Faster R-CNN Inception V2 to detect people. A transformation matrix was used to eliminate the camera angle's effect on the object distances on the image. The Euclidian distance used the pixels of the transformed image to compute the actual distance between people. A threshold of six feet was considered to capture physical distance violation. They also used transfer learning for training the model. The final model was applied on four videos of road maintenance projects in Houston, TX, that effectively detected the face mask and physical distance. They recommend that construction owners use the proposed system to enhance construction workers' safety in the pandemic situation.

3. PROPOSED SYSTEM

3.1 Face Mask Dataset

Masks play a crucial role in protecting the health of individuals against respiratory diseases, as is one of the few precautions available for COVID-19 in the absence of immunization. With this dataset, it is possible to create a model to detect people wearing masks, not wearing them, or wearing masks improperly. This dataset contains 853 images belonging to the 3 classes, as well as their bounding boxes in the PASCAL VOC format.

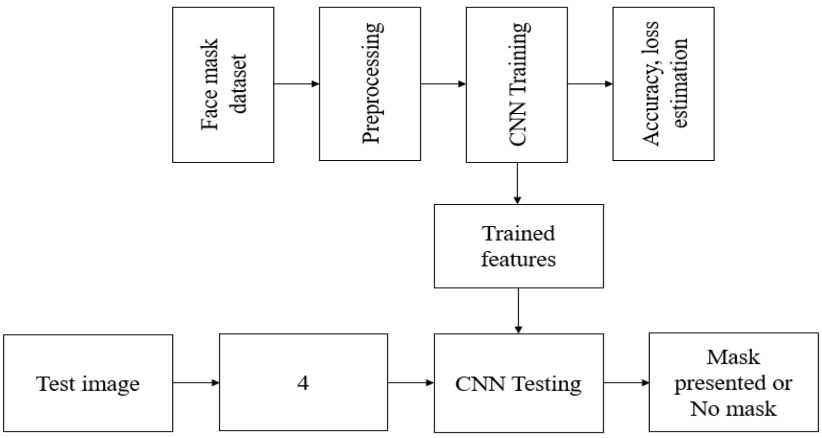


Fig.1: Block diagram of proposed system.

3.2 Pre-processing

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task.

3.2.2 Splitting the Dataset into the Training set and Test set

In machine learning data pre-processing, we divide our dataset into a training set and test set. This is one of the crucial steps of data pre-processing as by doing this, we can enhance the performance of our machine learning model. Suppose if we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models. If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset. Here, we can define these datasets as:

Training Set: A subset of dataset to train the machine learning model, and we already know the output.

Test set: A subset of dataset to test the machine learning model, and by using the test set, model predicts the output.

3.3 Proposed ResNet-CNN

Deep neural network is gradually applied to the identification of crop diseases and insect pests. Deep neural network is designed by imitating the structure of biological neural network, an artificial neural network to imitate the brain, using learnable parameters to replace the links between neurons.

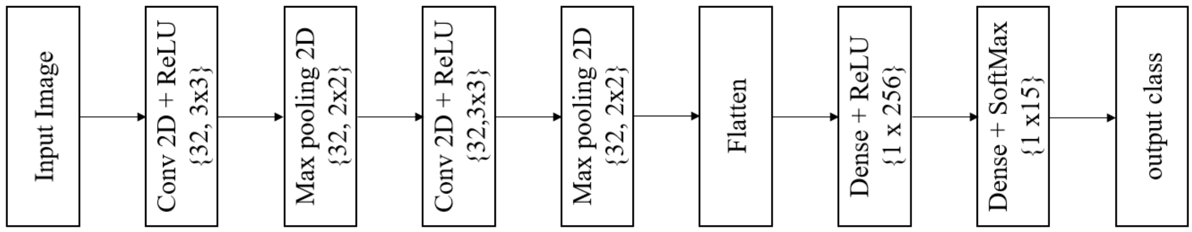


Fig. 2: Proposed ResNet-CNN.

Table. 1: Layers description.

Layer Names	No. of filters	Kernel size	Feature size
Conv 2D +ReLU	32	3 x 3	62x62x32
Max pooling 2D	-	3 x 3	31x31x32
Conv 2D+ReLU	32	3 x 3	29x29x32
Max pooling 2D	-	3 x 3	14x14x32
Flatten	-	1x6272	1x6272

Dense +ReLU		1 x 256	1 x 256
Dense + SoftMax		1 x 15	1 x 15

Convolutional neural network is one of the most widely used deep neural network structures, which is a branch of feed forward neural network. The success of AlexNet network model also confirms the importance of convolutional neural network model. Since then, convolutional neural networks have developed vigorously and have been widely used in financial supervision, text and speech recognition, smart home, medical diagnosis, and other fields.

Convolutional neural networks are generally composed of three parts. Convolution layer for feature extraction. The convergence layer, also known as the pooling layer, is mainly used for feature selection. The number of parameters is reduced by reducing the number of features. The full connection layer carries out the summary and output of the characteristics. A convolution layer is consisting of a convolution process and a nonlinear activation function ReLU. A typical architecture of CNN model for crop disease recognition is shown in Fig. 11.

The leftmost image is the input layer, which the computer understands as the input of several matrices. Next is the convolution layer, the activation function of which uses ReLU. The pooling layer has no activation function. The combination of convolution and pooling layers can be constructed many times. The combination of convolution layer and convolution layer or convolution layer and pool layer can be very flexibly, which is not limited when constructing the model. But the most common CNN is a combination of several convolution layers and pooling layers. Finally, there is a full connection layer, which acts as a classifier and maps the learned feature representation to the sample label space.

Convolutional neural network mainly solves the following two problems.

1) Problem of too many parameters: It is assumed that the size of the input picture is $50 * 50 * 3$. If placed in a fully connected feedforward network, there are 7500 mutually independent links to the hidden layer. And each link also corresponds to its unique weight parameter. With the increase of the number of layers, the size of the parameters also increases significantly. On the one hand, it will easily lead to the occurrence of over-fitting phenomenon. On the other hand, the neural network is too complex, which will seriously affect the training efficiency. In convolutional neural networks, the parameter sharing mechanism makes the same parameters used in multiple functions of a model, and each element of the convolutional kernel will act on a specific position of each local input. The neural network only needs to learn a set of parameters and does not need to optimize learning for each parameter of each position.

2) Image stability: Image stability is the local invariant feature, which means that the natural image will not be affected by the scaling, translation, and rotation of the image size. Because in deep learning, data enhancement is generally needed to improve performance, and fully connected feedforward neural is difficult to ensure the local invariance of the image. This problem can be solved by convolution operation in convolutional neural network.

3.3.1 ResNet-CNN

According to the facts, training and testing of ResNet-CNN involves in allowing every source image via a succession of convolution layers by a kernel or filter, rectified linear unit (ReLU), max pooling, fully connected layer and utilize SoftMax layer with classification layer to categorize the objects with probabilistic values ranging from [0,1].

Convolution layer as depicted is the primary layer to extract the features from a source image and maintains the relationship between pixels by learning the features of image by employing tiny blocks of source data. It's a mathematical function which considers two inputs like source image $I(x, y, d)$ where x and y denotes the spatial coordinates i.e., number of rows and columns. d is denoted as dimension of an image (here $d = 3$, since the source image is RGB) and a filter or kernel with similar size of input image and can be denoted as $F(k_x, k_y, d)$.

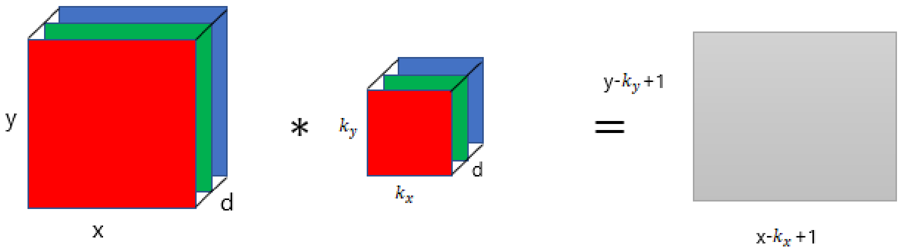


Fig. 3: Representation of convolution layer process.

The output obtained from convolution process of input image and filter has a size of $C\left((x - k_x + 1), (y - k_y + 1), 1\right)$, which is referred as feature map. Let us assume an input image with a size of 5×5 and the filter having the size of 3×3. The feature map of input image is obtained by multiplying the input image values with the filter values.

1	1	1	0	0
0	0	1	1	1
1	1	0	0	1
0	0	0	1	1
1	1	1	0	0

5x5 image

*

1	0	1
0	1	0
1	0	1

3x3 kernel

(a)

1	1	1	0	0
0	0	1	1	1
1	1	0	0	1
0	0	0	1	1
1	1	1	0	0

5x5 image

*

1	0	1
0	1	0
1	0	1

3x3 kernel

=

3	3	3
2	2	3
3	2	3

Feature map

(b)

Fig. 4: Example of convolution layer process (a) an image with size $\mathbf{5} \times \mathbf{5}$ is convolving with $\mathbf{3} \times \mathbf{3}$ kernel (b) Convolved feature map

3.3.2 ReLU layer

Networks those utilizes the rectifier operation for the hidden layers are cited as rectified linear unit (ReLU). This ReLU function $\mathcal{G}(\cdot)$ is a simple computation that returns the value given as input directly if the value of input is greater than zero else returns zero. This can be represented as mathematically using the function $max(\cdot)$ over the set of 0 and the input x as follows:

$$\mathcal{G}(x) = \max\{0, x\}$$

3.3.3 Max pooing layer

This layer mitigates the number of parameters when there are larger size images. This can be called as subsampling or down sampling that mitigates the dimensionality of every feature map by preserving the important information. Max pooling considers the maximum element form the rectified feature map.

Softmax classifier

Generally, as seen in the above picture softmax function is added at the end of the output since it is the place where the nodes are meet finally and thus, they can be classified. Here, X is the input of all the models and the layers between X and Y are the hidden layers and the data is passed from X to all the layers and Received by Y. Suppose, we have 10 classes, and we predict for which class the given input belongs to. So, for this what we do is allot each class with a particular predicted output. Which means that we have 10 outputs corresponding to 10 different class and predict the class by the highest probability it has.

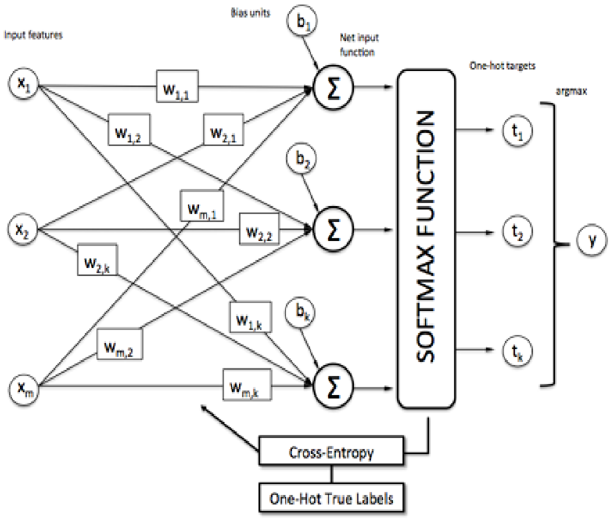


Fig. 5: Crop disease prediction using SoftMax classifier.

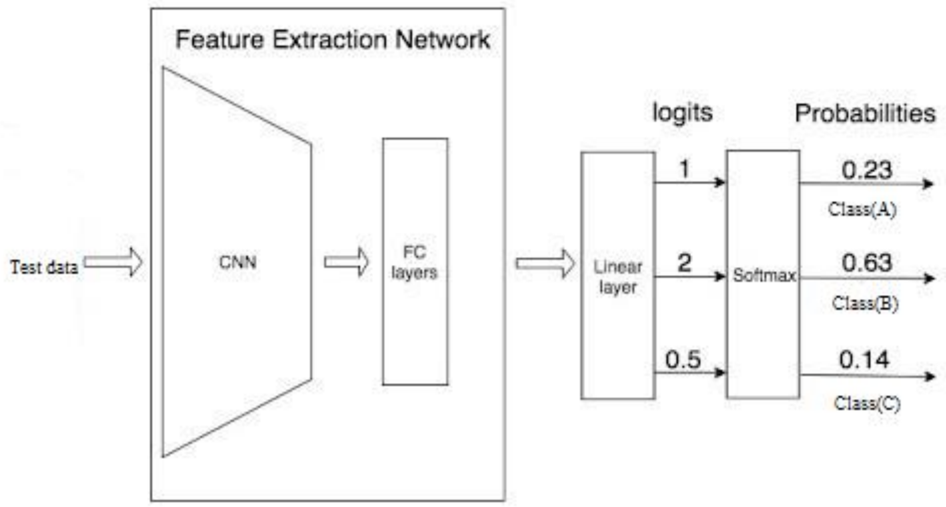


Fig. 6: Example of SoftMax classifier.

In Fig. 6, and we must predict what is the object that is present in the picture. In the normal case, we predict whether the crop is A. But in this case, we must predict what is the object that is present in the picture. This is the place where softmax comes in handy. As the model is already trained on some data. So, as soon as the picture is given, the model processes the pictures, send it to the hidden layers and then finally send to softmax for classifying the picture. The softmax uses a One-Hot encoding Technique to calculate the cross-entropy loss and get the max. One-Hot Encoding is the technique that is used to categorize the data. In the previous example, if softmax predicts that the object is class A then the One-Hot Encoding for:

Class A will be [1 0 0]

Class B will be [0 1 0]

Class C will be [0 0 1]

From the diagram, we see that the predictions are occurred. But generally, we don't know the predictions. But the machine must choose the correct predicted object. So, for machine to identify an object correctly, it uses a function called cross-entropy function.

So, we choose more similar value by using the below cross-entropy formula.

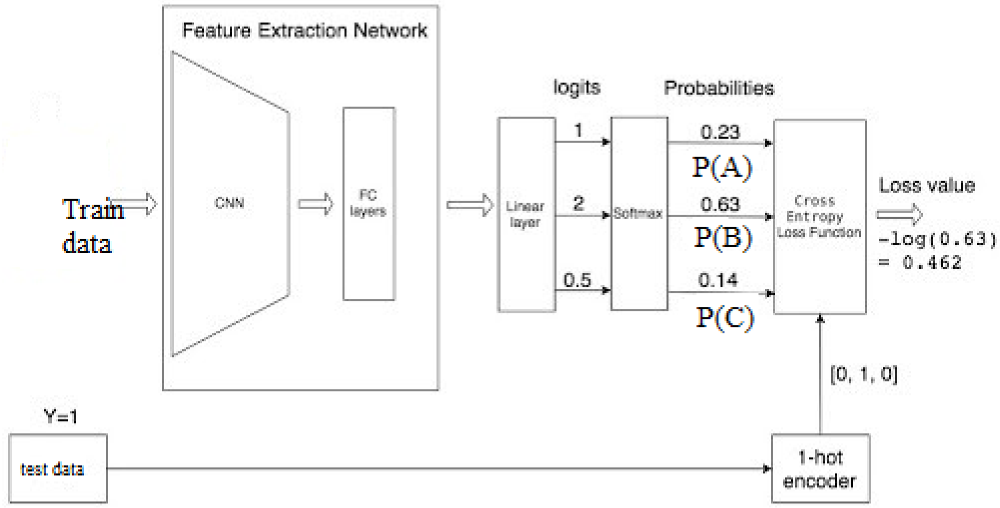


Fig. 7: Example of SoftMax classifier with test data.

In the above example we see that 0.462 is the loss of the function for class specific classifier. In the same way, we find loss for remaining classifiers. The lowest the loss function, the better the prediction is. The mathematical representation for loss function can be represented as: -

$$LOSS = np.sum(-Y \; * \; np.log(Y\_pred))$$

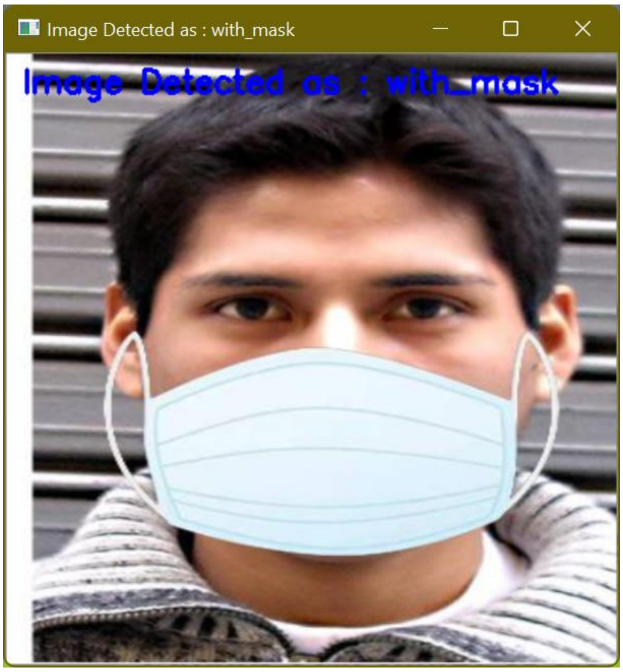
4. RESULTS AND DISCUSSION

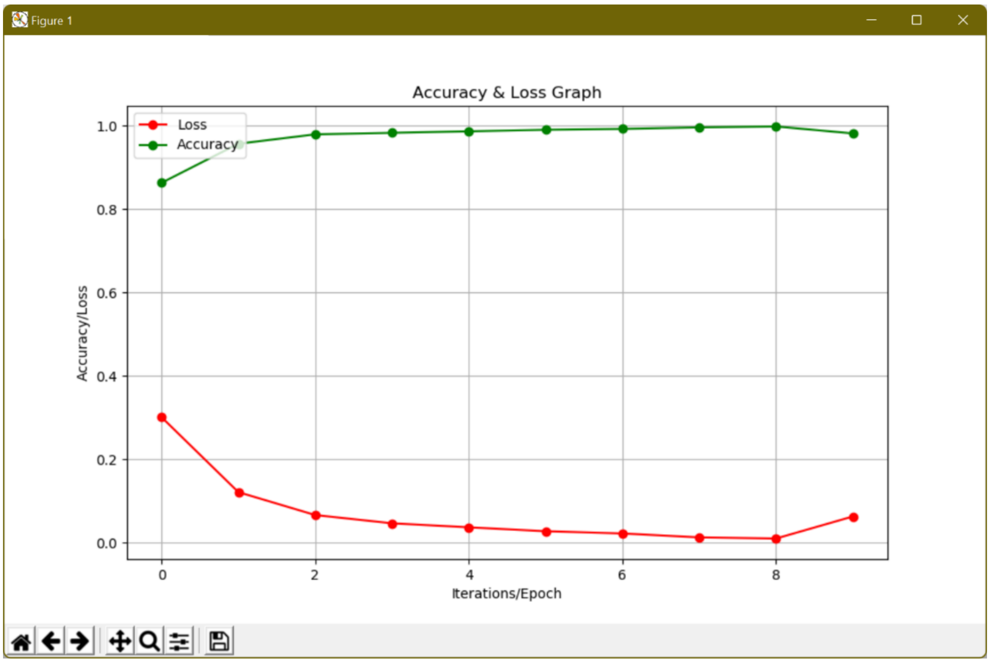
Modules

- Upload Mask Detection Dataset: This user will upload the Mask Detection Dataset and then the dataset may load.
- Train Mask Images using CNN: This is the second module in our project, after loading the dataset we can train with CNN.
- Upload Test Image & Detect Mask: Then we must upload our test image to detect mask.
- Accuracy & Loss Graph: In graph x-axis represents EPOCH and y-axis represents accuracy and loss values.

Here, CNN is trained on 1376 images and its prediction accuracy is 98% ad now model is ready and now 'Upload Test Images & Detect Mask' button to upload test image and then application predict whether person wear mask or not

In above screen application saying image is without mask and you can see predicted result in image title bar also. Now test with other images





In above screen red refers to loss and green line refers to accuracy and x-axis represents EPOCH and y-axis represents accuracy and loss value. In above graph with each increasing epoch accuracy get better and loss get decrease. Below screen showing CNN model with 200, 100 layers for input and 64.

5. Conclusion

Measures must be taken to control the spread of the COVID19 pandemic. This face mask recognition system is a very good and efficient way to do so. The system will separate the people from the crowd who are not wearing mask. The identification of people, violating the COVID norms increases the adaptability of the face mask detection system for the public sake. If applied in a correct way, the face mask detection system could be used to make sure our safety and for others too. This approach gives not only helps in achieving high precision but also enhance the face detection tempo considerably. The system can be applied in many areas like metro stations, markets, schools, railway stations and many other crowded places to monitor the crowd and to ensure that everyone is wearing mask.

REFERENCES

[1] Taneja, S., Nayyar, A., Vividha, Nagrath, P. (2021). Face Mask Detection Using Deep Learning During COVID-19. In: Singh, P.K., Wierzchoń, S.T., Tanwar, S., Ganzha, M., Rodrigues, J.J.P.C. (eds) Proceedings of Second International Conference on Computing, Communications, and Cyber-Security. Lecture Notes in Networks and Systems, vol 203. Springer, Singapore. https://doi.org/10.1007/978-981-16-0733-2_3

[2] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M. Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, Measurement, Volume 167, 2021, 108288, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2020.108288.

[3] A. Nowrin, S. Afroz, M. S. Rahman, I. Mahmud and Y. -Z. Cho, "Comprehensive Review on Facemask Detection Techniques in the Context of Covid-19," in IEEE Access, vol. 9, pp. 106839-106864, 2021, doi: 10.1109/ACCESS.2021.3100070.

[4] Peishu Wu, Han Li, Nianyin Zeng, Fengping Li, FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public, Image and Vision Computing, Volume 117, 2022, 104341, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2021.104341.

[5] Singh, S., Ahuja, U., Kumar, M. et al. Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. Multimed Tools Appl 80, 19753–19768 (2021). https://doi.org/10.1007/s11042-021-10711-8

[6] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M. Khalifa, Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection, Sustainable Cities and Society, Volume 65, 2021, 102600, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2020.102600.

[7] J. Zhang, F. Han, Y. Chun and W. Chen, "A Novel Detection Framework About Conditions of Wearing Face Mask for Helping Control the Spread of COVID-19," in IEEE Access, vol. 9, pp. 42975-42984, 2021, doi: 10.1109/ACCESS.2021.3066538.

[8] B. Wang, Y. Zhao and C. L. P. Chen, "Hybrid Transfer Learning and Broad Learning System for Wearing Mask Detection in the COVID-19 Era," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2021, Art no. 5009612, doi: 10.1109/TIM.2021.3069844.

[9] X. Fan, M. Jiang and H. Yan, "A Deep Learning Based Light-Weight Face Mask Detector With Residual Context Attention and Gaussian Heatmap to Fight Against COVID-19," in IEEE Access, vol. 9, pp. 96964-96974, 2021, doi: 10.1109/ACCESS.2021.3095191.

[10] Agarwal, C., Kaur, I., Yadav, S. (2023). Hybrid CNN-SVM Model for Face Mask Detector to Protect from COVID-19. In: Gupta, M., Ghatak, S., Gupta, A., Mukherjee, A.L. (eds) Artificial Intelligence on Medical Data. Lecture Notes in Computational Vision and Biomechanics, vol 37. Springer, Singapore. https://doi.org/10.1007/978-981-19-0151-5_35

[11] Razavi, M., Alikhani, H., Janfaza, V. et al. An Automatic System to Monitor the Physical Distance and Face Mask Wearing of Construction Workers in COVID-19 Pandemic. SN COMPUT. SCI. 3, 27 (2022). https://doi.org/10.1007/s42979-021-00894-0.