Deep Learning Model based Face Recognition with Mask

Anushree N S
Department of Computer Science and Engineering
C Byre Gowda Institute Of Technology
Kolar, India
anushree.ranikumar@gmail.com

Mr. Vasudeva R
Asst.Prof Department of Computer Science and Engineering
C Byre Gowda Institute Of Technology
Kolar,India

***Abstract*-** **Face detection has evolved as a popular problem in image processing and Computer Vision. Many new algorithms are developed using convolutional constructions to make the algorithm as accurate as possible. These convolutional structures have made it easier to extract pixel details. We aim to design a binary face divider that can find any face present in the frame no matter how aligned. We present how to make an accurate face mask from any image of incorrect size. Starting from an RGB image of any size, the method uses VGG Pre-defined Training Weights - 16 Architecture rendering feature. The training was done through Fully Convolutional Networks to differentiate the face and face present in the image. Gradient Decline is used for training while Binomial Cross Entropy is used as a loss function. Progressively the output image from FCN is processed to remove unwanted noise and to avoid false guesses if any and to create a binding box around the face. In addition, the proposed model also shows good results in visual facial recognition. In line with this it is also able to get many face masks in one frame. Tests were performed on the Multi Parsing Human Dataset to detect an estimated pixel density of 93.884% on split face masks.**

Keywords-VGG, Face detection, FCN, binary face, visual facial recognition, RGB.

I. INTRODUCTION

Existing biometric systems that rely on passwords or fingerprints are no longer secure since the COVID-19 virus can be disseminated through contaminated contact areas. Because it does not require any touch with the device, face recognition is incredibly safe. According to recent coronavirus studies, wearing a face mask by both healthy and sick people dramatically reduces virus transmission. Masking, on the other hand, has the following disadvantages:

1) Fraudsters and burglars are using the mask to steal and commit crimes while remaining anonymous.
2) Whenever the significant amount of the face is concealed by a mask, public access control and facial verification have become more challenging duties.
3) Existing face recognition methods do not work well if you wear a mask that can give the whole face a definition.
4) In the facial recognition function, highlighting the nose area is critical since it is used to acquaint the face [1], set the correction [2], and match the face [3]. Face masks severely opposed to current facial expressions as a result of these issues. To deal with these problems, we separated two functions: facial recognition and hidden face. The first determines whether or not the person is wearing a mask. This can be worn in public locations where wearing a mask is required. Face recognition with a mask based on the eyes and forehead areas, on the other hand, seeks to see faces with a mask based on the eyes and forehead regions. We are conducting a second activity in this paper that is based on in-depth reading. We use a low-level model of learning to extract features in unmarked facial regions (excluding the mask region).

COVID-19 is now attacking the planet. COVID-19 is an infectious disease caused by SARS-CoV-2, a dangerous respiratory infection [1]. Close contact with an infected person by respiratory drops during coughing, sneezing, and/or talking can cause infection. Furthermore, the virus can be spread by touching a virus-infected surface or object and then touching the mouth, nose, or eyes. In the meantime, we may keep ourselves safe by avoiding viral exposure. The best approach to avoid the spread of the disease, according to the CDC, is to isolate yourself from society and wear masks when out in public [2]. Two main ways to prevent it are to avoid unnecessary contact and to wear a face mask. Implementing these guidelines, has a significant impact on current face-based security systems that have already been put in place by a number of local companies and government agencies. Fingerprint or login credentials security systems that entail finger-to-finger touch are so ineffective in preventing the transmission of illnesses that render them dangerous. Face recognition-based security that minimizes needless communication, making it more secure than before. However, such techniques presuppose that a full-face image can be captured for better recognition. As a result, widespread use of face masks renders existing vision systems ineffective, and the infrastructure surrounding the face mask may degenerate. The accuracy of modern face-to-face learning programs has been demonstrated [3]. The quality of the training photos supplied determines the accuracy of these systems. For recognition, most of these applications use anonymous faces.

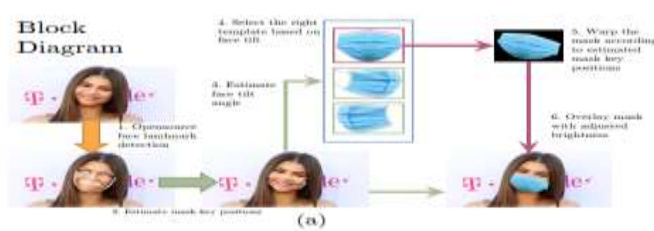

Fig. 1: MaskTheFace offers up to 100 mask variations to choose from.

The accuracy of modern face-to-face learning programs has been demonstrated [3]. The quality of the training photos supplied determines the accuracy of these systems. For recognition, most of these applications use anonymous faces. This case is ideal if you are confident that the system can access a known person's entire unencrypted face. The algorithm that has been trained on such photographs learns to focus on crucial face features such as the eyes, nose, lips, and facial margins, among others. When these applications are supplied with a face mask, however, the system is unable to identify the person who is giving the unusable application. We're working on resolving this security issue in order to improve the credibility of the face-based system.

II. THEORY AND CONCEPTS

The use of Deep Convolutional Neural Networks (DCNNs) for large-scale recognition of the construction of meaningful loss functions that provide discriminatory power is one of the most difficult issues in feature learning. Major softmax loss methods, such as SphereFace's loss of angular softmax, CosFace's significant loss of cosine margin, and ArcFace's loss of angular margin line, have recently demonstrated outstanding performance in deep facial recognition. During the coronavirus outbreak, practically everyone wears a mask to effectively prevent the spread of the COVID-19 virus. This renders general facial recognition technology ineffectual in a variety of situations, including public access control, face access control, face-to-face security checks at train stations, and so on. In scenarios where, incomplete face images may be captured due to conditions, sightings, and vast viewing angles, such as video surveillance and mobile devices, low-resolution (PFR) recognition in an unlimited region is a critical task. A mask-learning technique to detect and dispose of damaged features so that they can be detected, is more sensitive to the impacts and focuses primarily on the missing facial areas. Using the meticulously built Pairwise Differential Siamese Network, was first developed by exploiting the differences between the two integrated pairs of closed and non-closed face pairs (PDSN). The second is based on in-depth study and especially Convolutional Neural Networks (CNN). We employ a CNN model called "AlexNet" that has been trained to perform well in the ImageNet database.

The proposed system:

1)Preparing and cropping images at 240 * 240 pixels.

2)Deep feature extraction using VGG16 and feature extraction from 3 * 3 feature map maps and feature vets are transmitted by RBF layer and quantization layer and MLP separator is used for labelling.

3)The extraction of all RBF neurons is accumulated in the Quantization layer, which also includes the vector histogram of the relevant earth element required in the classification algorithm.

4)We were able to construct a convolutional neural network-based artificial network that can detect images. Check the database after separating it from the train database. We'll eventually use a training database to develop and train the model.

5)Once the model has been trained, a test model can be created. At this stage a collection of test data has been uploaded. Because this data set has never been seen by a model, its accuracy will be verified. Finally, the model that has been saved can be used in the actual world.

III. DESIGN AND CONFIGURATION OF VGG-16 MODEL

A Methodology

we describe an effective method for performing quantization-based facial recognition pooling using a pre-trained VGG-16 model in this study. Using the Bag-of-Features (BoF) paradigm, we only look at feature maps in the final layer of resolution (also known as channels). The core idea behind the standard BoF paradigm is to portray images as loose collections of local characteristics. The initial step in finding these sets is to eliminate the local features from the training photos, with each element representing a region in the image. Then, and only then, are all features limited to calculating the codebook. The test image's features are given to the closest code in the codbook, which is represented by a histogram. The BoF paradigm has been employed in the literature to perform image separation tasks, especially for the fabrication of a handcrafted element [19, 20].

In ImageNet, a dataset with over 14 million images per 1000 classes, VGG16's proposed convolutional neural network model achieves 92.7 percent top-5 test accuracy. One of the most popular models on ILSVRC-2014 was this one. By replacing big kernel-sized filters (11 and 5 in the first and second convolutional layers, respectively) with many filters the size of a 3 3 kernel, it outperforms AlexNet. NVIDIA Titan Black GPUs were utilized to train IVGG16 for weeks.

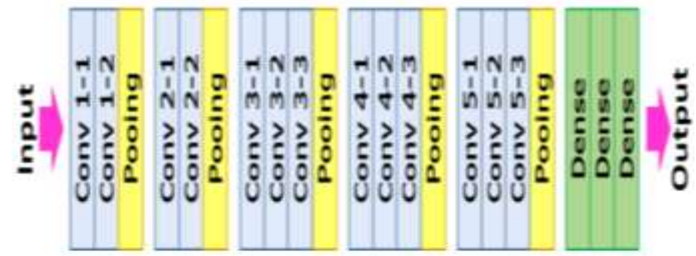

Fig 2: VGG16 Proposed convolutional network model

Multiple filters of size 3 * 3 kernel and layer integration make up the convolutional network model, which has a maximum output size of 3 * 3. The nearest code in the conv represented by the histogram receives the test image features. The BoF paradigm is commonly utilized in the literature.

All configurations adhere to the architecture's basic configurations and differ only in depth: from 11 layers in network A (8 and 3 layers FC) to 19 layers in layer E. (16 layers and 3 layers FC). The number of channels in the conv. layers is minimal, starting at 64 in the first layer and increasing by two elements after each max-pooling layer, up to 512.

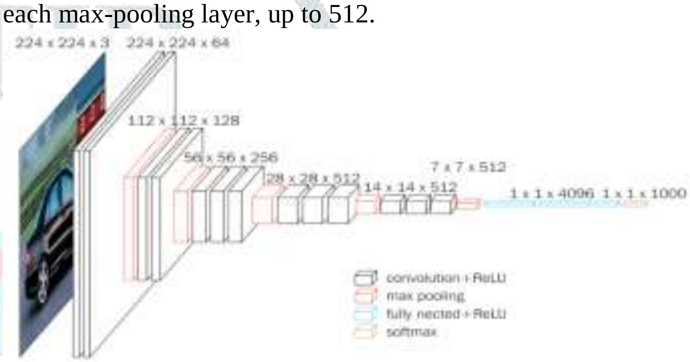

Fig 3 :Maxpooling and fully nected Relu model

B System Architecture

Inter-design interface - defines user design and layout. Includes a screenshot representation, a description of the interaction methods, and a description of the movement modes. Interaction Control Methods- using navigation options, the designer selects one of several communication modes. Database flowchart with face mask data and input predefined data provided with bound data and trainset is used for model training and verified using a set of test and prediction results.

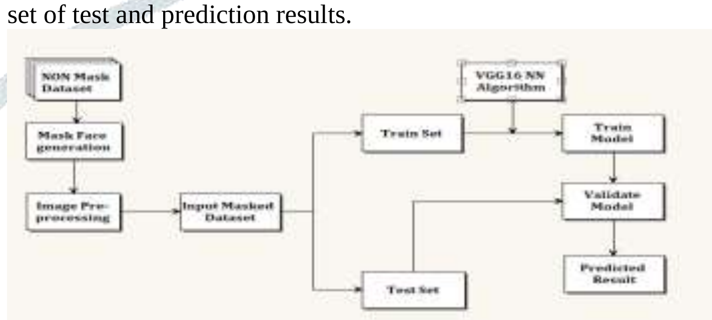

Fig 4: System Architecture

1)Train Set: The set of preprocessed images which is used for training the model with VGG-16 NN algorithm and the output is

Compared with the test set for predicting the results.

2)Test Set: The set of images which is used to compare with the train dataset and validate the model with train set and predict the results of the model.

3)Predicted Result: The result is obtained with the label of the image of the person after validating the trained set.

IV. SYSTEM IMPLEMENTATION

A Pre-Processing and cropping filter

We employ a 2D rotation to make them horizontal depending on the position of the eye. The next step is to use a crop filter to eliminate only the parts of the image that aren't covered by the mask. To accomplish so, we must first adapt all facial photos to a resolution of 240 240 pixels. The partition with blocks is then applied. The purpose of this technique is to divide the image into 100 square kilometers (for us, that's 24 x 24 pixels). After that, only blocks with a transparent circuit are removed (blocks from number 1 to 50).

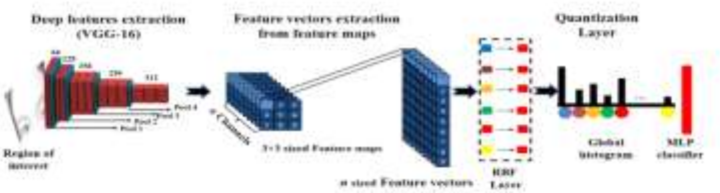

Fig 5: Overview of the proposed method

B Feature Extraction Layer

In 2D images, we extract in-depth features using the VGG-16 face CNN descriptor [24]. ImageNet, a database containing over 14 million images and 1000 categories, was used to train. The term VGG-16 stems from the fact that it has 16 layers. Convolutional layers, Max Pooling layers, starting layers, and fully integrated layers are among the layers included. There are 13 solution layers, 5 Max Pooling levels, and 3 thick layers totaling 21 layers, but only 16 layers of weight. The creation of VGG-16 is depicted in Figure 4. We only look at the feature maps (FM) in the final resolution layer, also known as channels, in this function. These features will be used in the rating area in the future phase.

C Deep Bag Of Features Layer

Using the output element layer described above, we extract the feature maps from the image. We employed the RBF kernel as the metrics matrix to quantify the similarity between the output vectors and code words, also known as the term vector, as proposed in [2]. As a result, the first sub layer will be made up of RBF neurons, each of which will be assigned to a code. The number of feature carriers to be employed in the BoF layer is determined by the size of the retrieved map. configuration and

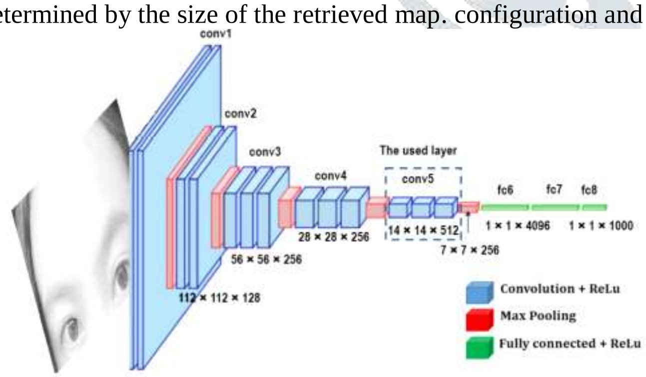

Fig 6: VGG16 network architecture

1) RBF Layer: The similarity of the probe faces' input features to the RBF centers is measured. The jth RBF neuron (Xj) is defined as follows:

$$Ø(X_j) = \exp(||x-c_j||2/\sigma_j),$$

Where x is a feature vector and cj is the center of the jth RBF neuron.

2) Quantization Layer: This layer collects the RBF neuron's output, which includes a histogram of the global quantized feature vector that will be used in the classification process. The following is the definition of the final histogram:

$$h_i = V_j \sum^{N_k} Ø(V_{jk})$$

k

D Model Training

The whole database is divided into three parts - training data, verification data and test data. The model is trained in training data over and over again with many repetitions. This is also known as Epochs. After each period, the model is tested using verification data. Finally, the most efficient model in verification data is loaded.

E Model Testing and Evaluation

To date, 200 frames per video have been released and are provided as per models. But the method of removing these structures is not very suitable. What was done was that for each video, 200 frames were released (8 seconds). We know that the human body performs these functions (running, boxing, etc.) at a certain speed. In one second, the human body does not make many movements. Therefore, we do not need to collect all the frames per second of the video we are shooting. A different approach can be used, where only a certain number of frames are released per second. Now, we will release 5 frames per second (the first 5 frames per second). So, let's say we have a 10-second video, we're going to get 10 x 5 = 50 frames. I have set this number to 40. Therefore, these 40 frames will be pre-selected from the released frames.

V. RESULTS AND ANALYSIS

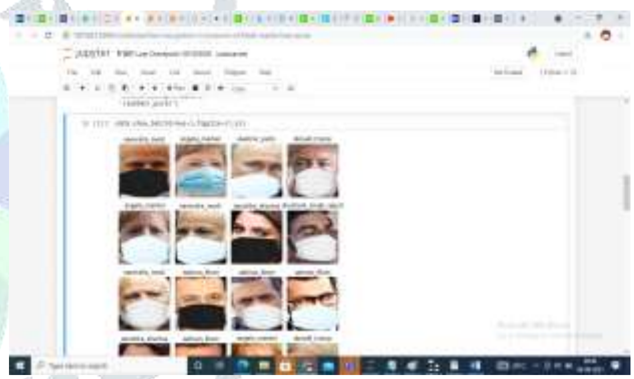

Fig 7: Masked Dataset

The masked dataset which is used as the image preprocessed dataset and the dataset which is used for the VGG-16 algorithm for training of the model. And the dataset which is given as the input to the model and training of the model to obtained the trained model for the feature extraction of features by RBF layer.

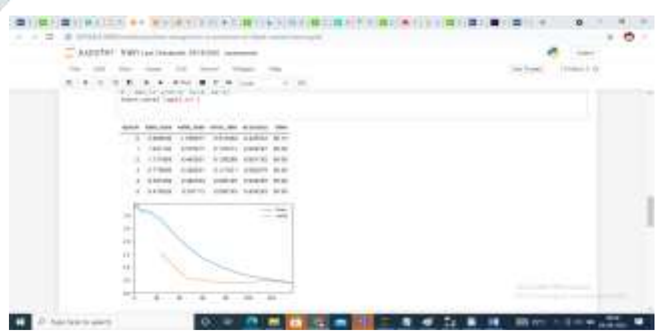

Fig 8: Model Training with Parameter

The Graph which is used for the comparison of the train dataset with the test dataset with the accuracy of the 69.9% with the time lapse of 0.08. The error rate of the train model will be 0.0075% the valid loss will be 0.3075 and the train loss will be 0.4123% the final accuracy with 70% than the baseline model and more accuracy will be obtained with time.

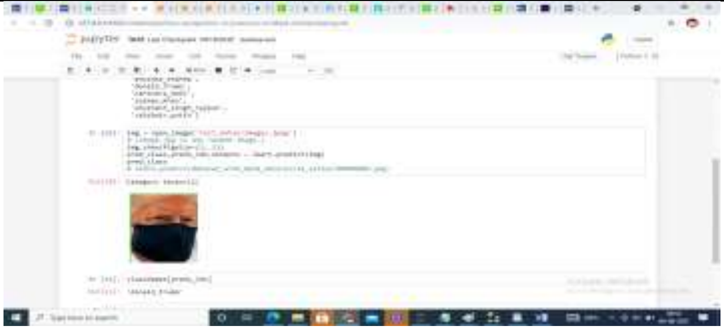

Fig 9: Final Result

The final result obtained will be the name of the particular person which is labelled in the dataset and the label is the name of the particular individual obtained after validation.

VI. CONCLUSION AND FUTURE ENHANCEMENT

A Conclusion

Corporate giants from a variety of verticals are turning to AI and ML, using technology in human use in the midst of this epidemic. Digital product development companies are introducing API masking services that enable developers to create a face mask application quickly so they can work for the public during a crisis. The technology ensures reliable face detection and real-time for users wearing masks. Besides, the system is easy to install on any existing business system while maintaining the security and privacy of user information. The face mask application system will therefore be the leading digital solution in many industries, especially the marketing, health and corporate sectors.

B Future Enhancement

1)A plan to replace the laptop with an active model / a device like the Raspberry Pi that needs to be installed on checkpoint / location. Raspberry Pi variants are included in size. This will reduce complex installation by 90%.

2)Comparable device by credit card size with camera attached and loaded with the required algorithms / code can meet requirements. There will be no professional obligation / trained people stop using the commands. Normally the sender / operator must be able to use the model.

3)As such a need for a laptop loaded with the camera is a necessity for making software. Therefore, the laptop should be installed during self-examination again another skilled operator should be present to type and perform appropriate order.

VII. REFERENCES

[1] Xiaoming Peng, Mohammed Bennamoun, and Ajmal S Mian. A training-free nose tip detection method from face range images. Pattern Recognition, 44(3):544–558, 2011.

[2] Xiaoguang Lu, Anil K Jain, and Dirk Colbry. Matching 2.5 d face scans to 3d models. IEEE transactions on pattern analysis and machine intelligence, 28(1):31–43, 2005.

[3] Melissa L Koudelka, Mark W Koch, and Trina D Russ. A prescreener for 3d face recognition using radial symmetry and the hausdorff fraction. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops, pages 168–168. IEEE, 2005.

[4] Aleix M Martínez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Transactions on Pattern analysis and machine intelligence, 24(6):748–763, 2002.

[5] Renliang Weng, Jiwen Lu, and Yap-Peng Tan. Robust point set matching for partial face recognition. IEEE transactions on image processing, 25(3):1163–1176, 2016.

[6] Yueqi Duan, Jiwen Lu, Jianjiang Feng, and Jie Zhou. Topology preserving structural matching for automatic partial face recognition. IEEE Transactions on Information Forensics and Security, 13(7):1823–1837, 2018.

[7] Niall McLaughlin, Ji Ming, and Danny Crookes. Largest matching areas for illumination and occlusion robust face recognition. IEEE transactions on cybernetics, 47(3):796–808, 2016.

[8] Parama Bagchi, Debotosh Bhattacharjee, and Mita Nasipuri. Robust 3d face recognition in presence of pose and partial occlusions or missing parts. arXiv preprint arXiv:1408.3709, 2014.

[9] Hassen Drira, Boulbaba Ben Amor, Anuj Srivastava, Mohamed Daoudi, and Rim Slama. 3d face recognition under expressions, occlusions, and pose variations. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(9):2270–2283, 2013.

[10] Ashwini S Gawali and Ratnadeep R Deshmukh. 3d face recognition using geodesic facial curves to handle expression, occlusion and pose variations. International Journal of Computer Science and Information Technologies, 5(3):4284–4287, 2014.