2020 4th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)

Towards Facial Recognition Problem in COVID-19 Pandemic

Imran Qayyum Mundial[1], M. Sohaib Ul Hassan[2], M Islam Tiwana[4], Waqar Shahid Qureshi[5]
Department of Mechatronics Engineering, College of E&ME, National University of Sciences and Technology, H-12
Islamabad, Pakistán
[1]iqmundial@gmail.com , [2]mshassan51@gmail.com, [4]mohsintiwana@ceme.nust.edu.pk,[5]waqar.shahid@ceme.nust.edu.pk

Eisa Alanazi[3]
Department of Computer Science, Umm Al Qura University, Makkah
Saudi Arabia
[3]eaanazi@uqu.edu.sa

Abstract **– In epidemic situations such as the novel coronavirus (COVID-19) pandemic, face masks have become an essential part of daily routine life. The face mask is considered as a protective and preventive essential of everyday life against the coronavirus. Many organizations using a fingerprint or card-based attendance system had to switch towards a face-based attendance system to avoid direct contact with the attendance system. However, face mask adaptation brought a new challenge to already existing commercial biometric facial recognition techniques in applications such as facial recognition access control and facial security checks at public places. In this paper, we present a methodology that can enhance existing facial recognition technology capabilities with masked faces. We used a supervised learning approach to recognize masked faces together with in-depth neural network-based facial features. A dataset of masked faces was collected to train the Support Vector Machine classifier on state-of-the-art Facial Recognition Feature vector. Our proposed methodology gives recognition accuracy of up to 97% with masked faces. It performs better than exiting devices not trained to handle masked faces.**

Keywords - ***Face Mask, COVID-19, Corona Virus, Facial Recognition, Convolutional Neural Network, Support Vector Machine.***

I. INTRODUCTION

Face recognition has been one of the famous and challenging topics among computer vision researchers for decades. With the rapid development of artificial intelligence (AI) in recent years, face recognition technology also evolved on a fast track. The face recognition-based attendance system has various advantages compared to traditional card recognition, iris recognition, and fingerprint recognition. Facial recognition techniques include high concurrency, noncontact and user-friendly systems. The face recognition system has vast applications such as employment in the government sector, public facilities, security, e-commerce, retailing, and many other fields.

The development of deep learning architecture has played a vital role in the advancement of facial recognition techniques. Large datasets are available for the development of deep learning models for facial recognition. These datasets consist of many images that have been collected from various social networks and search engines [1][2]. The availability of these datasets and massive computational resources in modern computing systems like multi-core processors (PCs) and Graphical Processing Units (GPUs) has enabled deep learning algorithms. The most important contribution of these algorithms is Convolutional Neural Network (CNN) as a feature extractor. Fig. 1 shows a typical machine learning approach to solve the facial recognition problem. In Deep Face [3], face recognition is considered as a multiclass classification problem. To learn features on large multi identities datasets, CNN models were also introduced for face recognition, and traditional SoftMax loss is used with the remarkable performance [4]. Taigman et al. [5] model's accuracy is 97.35% on the Labeled Faces in the wild (LFW) dataset. Sun et al. [6] train CNN with the grouping of contrastive loss and SoftMax loss. It increases the inter-class separability and intra-class compactness. Face Net [7] model is the state-of-the-art face recognition model from Google. It builds face embeddings based on the triplet loss. It uses 128-dimensional representations from a very deep network. The model is trained on a 260 million image data, using a triplet loss at the final layer, with a margin, this loss separates a positive pair from a negative pair. The system achieved an accuracy of 99.63% on Labeled Faces in the wild (LFW) dataset.

Increased accuracy of these facial recognition algorithms has resulted in widespread commercial use in modern life. Governments and private firms worldwide are currently using facial recognition technology to recognize people at offices, airports, schools, and various other places. Security agencies worldwide are using this technology to eliminate the threats of terrorism and drug smuggling across borders. With the spread of the COVID-19 pandemic, Governments across the globe have made it compulsory for people to wear face masks in public places. In this scenario, the most important means of identification, face recognition techniques, almost failed. Which ultimately setback multiple authentication application that depends on facial recognition, such as face attendance, face access control, face authentication based mobile payment, face gates at

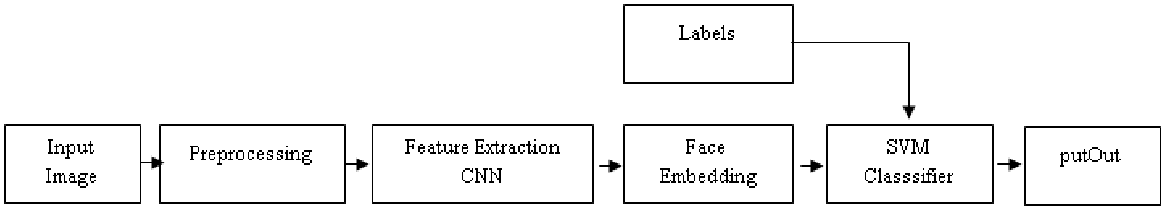

Fig. 1. Machine learning based approach for face recognition.

metro/ train stations, face recognition based social security investigation, community entry and exit points, etc [8]. Since the COVID-19 coronavirus spread through contact, the previously used unlocking system based on fingerprint or password is also not safe.

Due to the COVID-19 coronavirus, masked face recognition poses a challenge at public places, while removing face mask will increase the risk of novel virus infection. In this paper, we present a scheme to recognize masked faces with higher accuracy. State-of-the-art deep learning algorithms inspire our work. First, we trained a CNN model to generate an embedding of features of an image. Then we focus on a dataset which helps in building classifier for masked faces. This dataset consists of three images of a person, two masked face images, and one without a face mask. In the end, the Support Vector Machine (SVM) is used for classification.

This paper is organized as follows. Section 2 presents our proposed method to recognize masked faces, followed by experimental results in Section 3.

II. METHODOLOGY

This section provides an in-depth study of modeling of an approach that helps in the recognition of masked faces. As previously stated, the proposed methodology was inspired by ultra-modern facial recognition techniques. Each procedural step is discussed in detail in the following paras.

A. Feature Extractor CNN

Unlike conventional CNNs where the output layer (usually SoftMax layer) gives the probability of a certain output class, to whom input image belongs, this CNN feature extractor gives compact features of an input image as an output. The feature vector obtained from CNN is called Face Embeddings. These Face Embeddings are a result of a trained CNN [7] (trained via Triplet Loss Function). Triplet loss function is carried out in such a way that the Euclidean distance between two images of the same category is small and Euclidean distance between two faces of a different category is large. Fig. 2 gives a pictorial description of training with a triplet loss function. Training of CNN Feature extractor with triplet loss is an end to end learning process like any other CNN training. Triplet loss function is motivated by the nearest-neighbor classification method [9]. If there is an anchor image X_a^i, positive image X_p^i and negative image X_n^i. Where anchor image and positive image are of the same category and anchor and negative image are of different categories. Triplet loss function would bring positive image X_p^i closer to anchor image X_a^i and it would take negative image X_n^i away from anchor X_a^i. Equation (1) describes the Triplet loss L with respect to images X_a^i, X_p^i and X_n^i.

$$L = \sum [\|f(X_a^i) - f(X_p^i)\|^2 - \|f(X_a^i) - f(X_n^i)\|^2 + \alpha] \quad (1)$$

Where f(Xa), f(Xp) and f(Xn) represents the embeddings of anchor image, positive image and negative image respectively and α represents margin value.

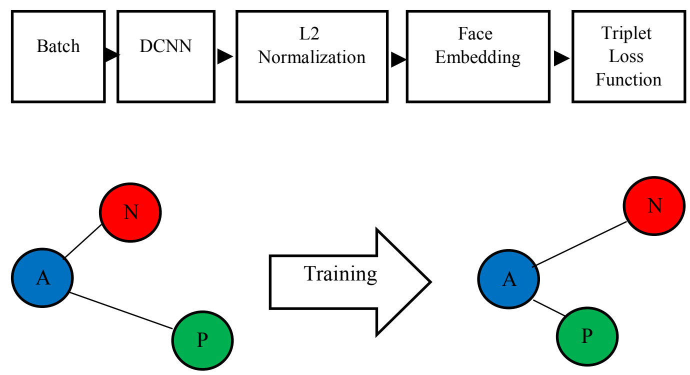

Fig. 2. Training model for CNN feature extractor for facial recognition [7].

There is various CNN architecture that is currently in use, more or less all these architectures perform the same. In these architectures, the feature extraction part is carried out by the Convolutional and Pooling layer, classification is carried out using fully connected layers and a SoftMax layer. FaceNet [7] used Zeiler and Fergus style network [10]. This network is shown in Fig. 4. It has 22 layers deep network, originally designed for ImageNet ILSRVC dataset with 1000 outcomes. The output layer is modified from 1000 outcomes of the SoftMax layer to 128D features in L2 normalized space. Our proposed Methodology used this network as a features extractor. This network generates a face embedding of dimension 128D. There are two approaches to train the network with triplet loss function, online triplet selection and offline triplet selection. In this work, an online triplet selection method was used during the training phase.

B. Feature Extractor CNN Training

VGGFACE2 [2] dataset was used to train the model. This dataset has images of about 9500 different classes. Each class having approximately 330 images. These images have been collected from different internet sources and social media platforms.

C. Face Detection and Pre-Processing

Both in academia and industry, several algorithms are performing well for face detection. DLIB facial detector, Haar-Cascade Face detector and MTCNN facial detector are few of them. MTCNN is the most popular of all these facial detectors. But due to its computational requirements, it is preferably not used on a live stream. For the pre-processing of raw static images, MTCNN is an ideal choice. Haar-Cascade Face Detector is used to develop a classifier for live stream images. Both these face detectors can detect masked faces. In the proposed methodology, MTCNN was used as a facial detector.

Detected Face images were resized to 220x220x3. The resized image was normalized using mean and standard deviation. Normalize image was fed to Feature extractor CNN to generate face embeddings of dimension 128D. The initial steps of the face detection pipeline are shown in Fig. 3.

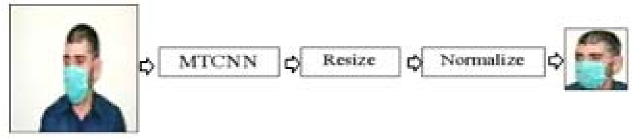

Fig.3. Face detection steps.

D. Face Classifier

Face classifier is an essential part of this paper and presents our contribution to masked faces recognition. A masked face classifier was built on feature extractor CNN discussed in the previous section. For every image, feature extractor CNN generates 128D embeddings, which will be stored in memory. Support Vector Machine (SVM) classifier was trained with these feature vector embeddings. Since the goal is to recognize the faces wearing a mask, therefore a small database with masked faces was established. Fig. 5 Shows the training images for the datasets.

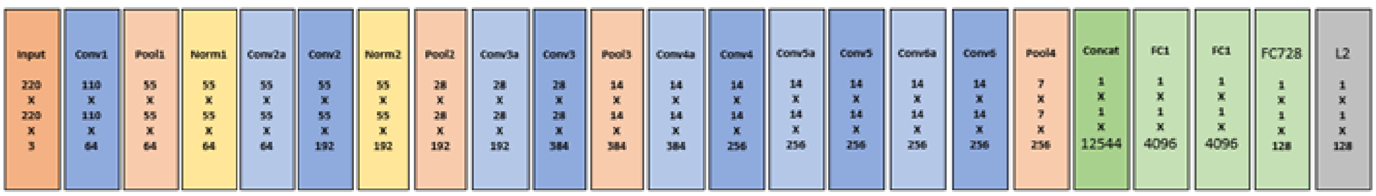

Fig. 4. Zeiler & Fergus style network is used as feature extractor and output is 128D compact feature vector or face embedding.

Fig. 5. SVM classifier dataset of masked faces containing two masked faces with side pose and one unmasked front pose image.

After deliberate research, two masked faces and one without mask face images were chosen to build a classifier. Training images were collected so that critical parts of the face are exposed to feature extractor CNN to learn masked faces. Masked and Unmasked faces address the associated problems such as occlusion and non-rigid transformation.

A mask, face image taken from the front, and masked face images with slight angles from both sides (left and right) are collected. These three images (two masked faces and one without mask) enable-feature extractor CNN to generate features vector embeddings close to each other. Using this approach, the SVM classifier distinguishes between two faces of different categories. An unmasked look helps to correlate features of masked faces. Three images were found to be enough to build the SVM Classifier. Images of the training set are shown in Fig. 5, where three images in each row (training images portion) belong to one category.

E. Implementation

This network's training was carried out on the core I9 9th Generation PC with NVIDIA GeForce GTX 1660 GPU. GeForce GTX 1660 is the most recent development by NVIDIA. The training of the network took about five days. Caffe Deep Learning library was used to train this model. In the training step, a batch of sample input images is passed through the model via the forward pass and compute the batch's triplet loss using online triplet selection. At the end of each epoch, validation was performed to find the accuracy of the model. Caffe gives update on loss function and accuracy with log files. The model is saturated in performance at about 6000 iterations. The trained model achieved a testing accuracy of 99% on the training dataset [2]. The trained network showed an accuracy of about 99% on the LFW dataset [1].

III. RESULTS

We have built the face recognition system for masked faces. Initially, our trained model showed almost 99% accuracy on the LFW dataset, confirming that the model is accurately trained on the VGGFACE2 dataset. Later, this trained model was deployed for the masked faces recognition pipeline. The masked dataset contained 800 images of 200 individuals. These 800 images comprise of 200 unmasked and 600 masked images (frontal and size pose for classifier building purposes). Once the system is trained with unmasked classifiers only and tested with masked faces, its accuracy was 79%. However, when the side poses masked faces are included with unmasked face classifiers, later training is carried out with the same classifier. System performance increased remarkably and accuracy improved up to 98% on the local dataset. The confusion matrix for masked face testing is as shown in Fig. 6.

N = 100		Predicted	
		Yes	No
Actual	Yes	73	2
	No	0	25

Fig. 6. Confusion matrix

Moreover, the proposed system achieved an accuracy of 97% on the RWMFD dataset [8] of masked faces.

TABLE I
PROPOSED MODEL ACCURACY ON VARIOUS DATASETS

Dataset	Accuracy
Trained model tested on LFW dataset	98%
Proposed system without masked dataset classifier	79%
Proposed system with masked dataset classifier	97%
Proposed system on RWMFD dataset with masked faces	97%

IV. DISCUSSION

For the last three decades, researchers work a lot on faces detection/recognition, however, masked face recognition was less explored area and even masked faces dataset was also not publicly available. But since the emergence of COVID-19 coronavirus plenty of work has been carried out in this field. Wang et al. [8] presented masked faces dataset, that is freely available for academia and industry. Jiang et al. [11] proposed a face mask detector using ResNet or MobileNet, feature pyramid network (FPN), and context attention modules. Hariri et al. [12] applied a deep learning-based method and quantization based technique to recognize masked faces. Our proposed work is simple and can be easily applied to a stand-alone system for independent application. The building of the classifier (with three images, as discussed) is very important in our work. If an individual masked image is more tilted towards left or right as shown in Fig.6, the system might generate false results.

Fig. 6. Unrecognisable pose for masked classifier.

V. CONCLUSION

In this paper, a machine learning methodology was presented to efficiently recognize the masked faces, inspired by the state of the art algorithms; the proposed method achieved 99% accuracy on the classifier, which was built on the masked faces dataset. Due to COVID 19, masked faces have created a huge challenge for facial recognition. However, this simple and innovative approach effectively solves the problem and address the security and social concerns. In the future, accurate masked face registration can be accomplished by the employment of some algorithms such as Iterative Closest Point (ICP). Moreover, this method can also be extended to higher-level applications such as violent video retrieval and video surveillance

REFERENCES

[1] G. B. Huang, M. Ramesh, T. Berg, and E. “Learned-miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments”. In ECCV Workshop on Faces in Real-life Images, 2008.

[2] VGGFACE2: https://zeus.robots.ox.ac.uk

[3] M. Parkhi, Andrea Vedaldi and Andrew Zisserman: “Deep Face Recognition”, In Proc. CVPR, 2015.

[4] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”, arXiv preprint arXiv:1409.1556, 2014.

[5] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. “Deep-Face: Closing the gap to human-level performance in face verification”, In Proc. CVPR, 2014.

[6] Y. Sun, X. Wang, and X. Tang. “Deeply learned face representations are sparse, selective, and robust”, CoRR, abs/1412.1265, 2014.

[7] F. Schroff, D. Kalenichenko, and J. Philbin. “Facenet: A unified embedding for face recognition and clustering”, In Proc. CVPR, 2015.

[8] Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, Heling Chen, Yu Miao, Zhibing Huang and Jinbi Liang.” Masked Face Recognition Dataset and Application” unpublished.

[9] K. Q.Weinberger, J. Blitzer, and L. K. Saul. “Distance metric learning for large margin nearest neighbor classification”, In NIPS. MIT Press, 2006.

[10] M. D. Zeiler and R. Fergus. “Visualizing and understanding convolutional networks”, CoRR, abs/1311.2901, 2013.

[11] MingjieJiang, Xinqi Fan, Hong Yan “RetinaMask: A Face Mask detector” unpublished.

[12] Hariri, Walid. "Efficient Masked Face Recognition Method during the COVID-19 Pandemic." unpublished.