2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), 24-25 December, Dhaka

MASKED FACE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK

Md. Sabbir Ejaz*, and Md. Rabiul Islam†
Dept. of Computer Science and Engineering
Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh
cse.sabbir7@gmail.com*, rabiul_cse@yahoo.com†

***Abstract*— Recognition from faces is a popular and significant technology in recent years. Face alterations and the presence of different masks make it too much challenging. In the real-world, when a person is uncooperative with the systems such as in video surveillance then masking is further common scenarios. For these masks, current face recognition performance degrades. An abundant number of researches work has been performed for recognizing faces under different conditions like changing pose or illumination, degraded images, etc. Still, difficulties created by masks are usually disregarded. The primary concern to this work is about facial masks, and especially to enhance the recognition accuracy of different masked faces. A feasible approach has been proposed that consists of first detecting the facial regions. The occluded face detection problem has been approached using Multi-Task Cascaded Convolutional Neural Network (MTCNN). Then facial features extraction is performed using the Google FaceNet embedding model. And finally, the classification task has been performed by Support Vector Machine (SVM). Experiments signify that this mentioned approach gives a remarkable performance on masked face recognition. Besides, its performance has been also evaluated within excessive facial masks and found attractive outcomes. Finally, a correlative study also made here for a better understanding.**

Keywords*— *Face detection, MTCNN, FaceNet, Face embedding, Masked face, SVM, Face recognition.

I. INTRODUCTION

Face recognition is a promising area of applied computer vision [1]. This technique is used to recognize a face or identify a person automatically from given images. In our daily life activates like, in a passport checking, smart door, access control, voter verification, criminal investigation, and many other purposes face recognition is widely used to authenticate a person correctly and automatically. Face recognition has gained much attention as a unique, reliable biometric recognition technology that makes it most popular than any other biometric technique likes password, pin, fingerprint, etc. Many of the governments across the world also interested in the face recognition system to secure public places such as parks, airports, bus stations, and railway stations, etc. Face recognition is one of the well-studied real-life problems. Excellent progress has been done against face recognition technology throughout the last years.

By the fast improvement and expansion in machine learning techniques, the dilemma of face recognition appears to be fully addressed. During current years, deep learning has obtained numerous breakthroughs in several computer vision areas, such as object detection, object classification, object segmentation and of course, face detection and verification [2, 3]. The previous face detection and verification algorithms need to manually design features, where deep learning methods do not require manual design. From training images, CNN can learn valuable features automatically. Recently, due to the success in detection and recognition problems, CNNs gained its popularity. CNNs successively apply convolution filters and they are accompanied by many non-linear activation functions. Before into this area uses a few convolutions filters supported by the average or sum pooling at the image [4]. Recently a more substantial quantity of filters is used which are pre-trained at huge datasets [5, 6, 7]. These methods are useful for detecting faces in various adjustments and poses [8]. The recognition accuracy of the traditional Eigenface algorithm at LFW is barely 60%. Where the recognition accuracy of the most advanced deep learning algorithm holds 99.63%. The acceptance rate is higher than the normal human eye because the human eye acceptance rate is 99.25% [9]. Multiple international projects have been effectively applied deep learning methods like FaceNet, DeepFace and many more for face recognition. Within those algorithms, the leading accuracy holds FaceNet and at the LFW dataset, it reached a 99.63% accuracy rate, which is higher than the normal human eye [9].

Although while handling unrestrained circumstances into which image degradation, changes in facial pose, occlusions, and other constraints normally occurs, popular systems still suffer. An individual can be disguised his identity by face alterations or using different altered physical attributes. For several deliberate or unintentional reasons, facial occlusions may have occurred. As an example, dacoit, hooligans, and offenders use sunglasses or scarves to block their faces for being unrecognized. Many people use a hood, make beards as religious faith or social tradition. Other origins regarding occlusions involve medical masks, caps, mustaches, makeup, etc. Sometimes faces are not clear, because many obstacles can be in front of the face. These obstacles can be a different object, glasses, scarves, caps or some occlusion on the face. Face occlusion problems can be classified into three categories in real-world: occlusion of facial landmarks, occluded by different objects and occluded by faces [6]. Facial landmark occlusion involves using masks and sunglasses. Occluded by faces is a complicated circumstance where miss-recognized several faces towards one face or hardly detect a portion of a face region. When occlusion caused by any objects, normally most of the faces will remain hidden.

If mask analysis does not particularly take into consideration, then it can change the performance seriously most of the sophisticated face recognition methods. The key features to identify a person are decreasing by using various sorts of masks or occlusions. Fewer numbers of facial features in the masked face cause difficulties than other normal face recognition techniques [10]. Therefore, the accuracy rate of recognition is decreasing. For disguising identities, terrorists and criminals are covered their faces with the mask. That's why the masked face is being one of the majors concerned

978-1-7281-6099-3/19/$31.00 ©2019 IEEE

factors within the domain of face recognition. On the other hand, the usage of a deep learning network is more challenging because the quantity of training data is not sufficient to train the deep learning networks for this application which forces to use of transfer learning [11]. Usually, transfer learning performs great but sometimes does not give satisfactory results when training data is not enough for fine-tuning those pre-trained deep learning networks. The focus of this work rests upon to develop the face recognition accuracy within different types of masks.

This writing is arranged as follows: Section 1 shows our face detection methodology. Section 2 describes about face embedding process. Section 3 narrates the classification procedure. Section 4 signifies the dataset. Section 5 presents test results analysis and section 6 draws concludes.

II. Convolution Neural Network

Convolutional Neural Network (CNN) is comprised of various convolutional layers, several pooling layers (e.g. min, max or average), non-linear layer (e.g. sigmoid, ReLU) and classification layer (e.g. Softmax) units. Deep CNN networks are typically trained on large labeled datasets like ImageNet to pick out general features which are appropriate into several detections and recognition jobs like image classification and verification, object detection, segmentation to texture identification [4]. When CNN architecture combines with multiple detectors which can confine different portion of an object. Into the area of fine-grained recognition like identifying a dog class, kinds of a bird, or a car model, these structures can also have delivered state-of-the-art outcomes [12]. CNN can obtain different local essential features from the data, can select global training components, and have been successfully implemented to many disciplines of pattern recognition applications. The deep CNN has sufficient robustness on scaling, shifting and transposing [13]. To fulfill our work, different forms of CNN architecture is used here. Figure 1 shows a classic CNN architecture.

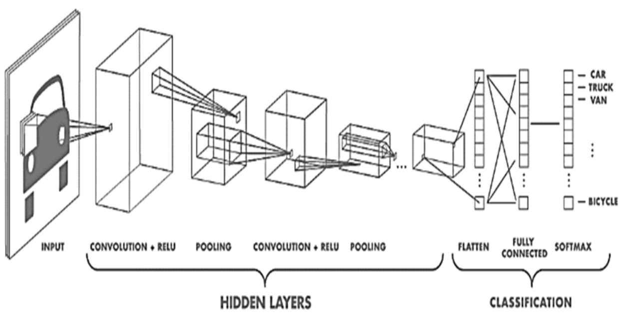


Fig. 1. Classic CNN architecture [13].

III. Methodology

Our approach consists of three principal modules: Detecting face from a given image, extract features and finally recognition. Figure 2 demonstrates our methodology to the problem of face recognition within masked conditions.

A. Facial Image Acquisition

Image acquisition implies the first key stage regarding the face recognition method. We collected masked and non-masked face images from AR, IIIT-Delhi face database. The scarcity of a large number of images, we apply the data augmentation process on masked and non-masked face images available in the database to enlarge the dataset images, so that our work is more reliable and efficient.

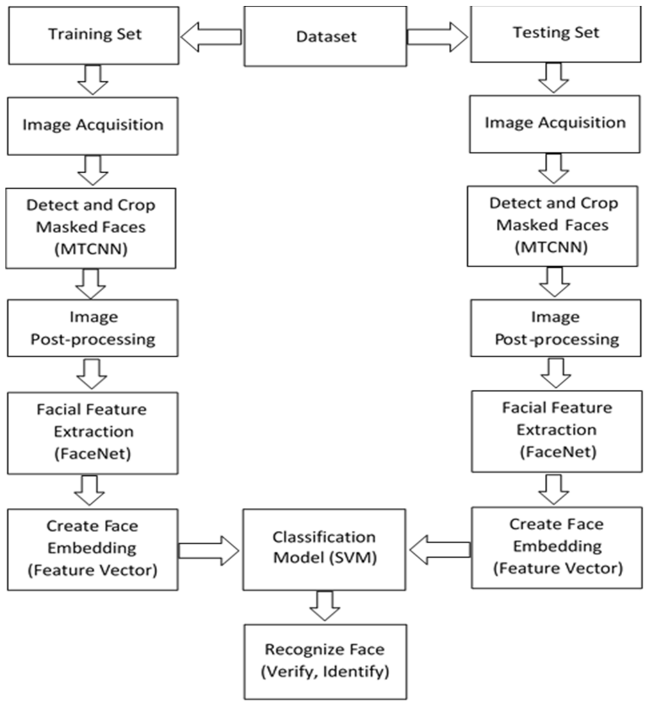


Fig. 2. Masked face recognition flow chart.

B. Masked Face Detection Using MTCNN

Detecting the faces from an input image is an essential job during face recognition. Without faces, the work can't go further. K. Zhang, Z. Li, Z. Zhang, and Y. Qiao proposed Multi-task Cascaded Convolutional Neural Network (MTCNN) to detect and align face from an input image which is able to outperform many face-detection benchmarks while holding real-time performance. So we use a pre-trained MTCNN model to detect candidate masked and non-masked face portion of the given image and interpret them into high dimensional facial descriptors. There are three networks cascading in this model. The model first rescaled the image to a certain extent of sizes. It is called an image pyramid. Then the candidate facial regions are introduced by the first network, P-Net or Proposal Network. The second network known as R-Net or Refine Network refines the bounding boxes. And finally, the third network, O-Net or Output Network determines facial landmarks from the image. Those networks P-Net, R-Net, and O-Net can perform classification of face, regression of bounding box and localization of facial landmarks. That's why this model is known as a multi-task network [15]. These three models are not directly connected. The next stage is fed with the previous stage outputs. As a result, different stages are accommodated with additional processing. Non-Maximum Suppression (NMS) is applied in MTCNN model to refine the candidate bounding boxes suggested by the first stage P-Net, the second stage R-Net and the third stage O-Net network prior to providing output. This face detection method has benefits over different lighting conditions, poses and visual variations of the face [15]. All steps involved in MTCNN method shown in figure 3.

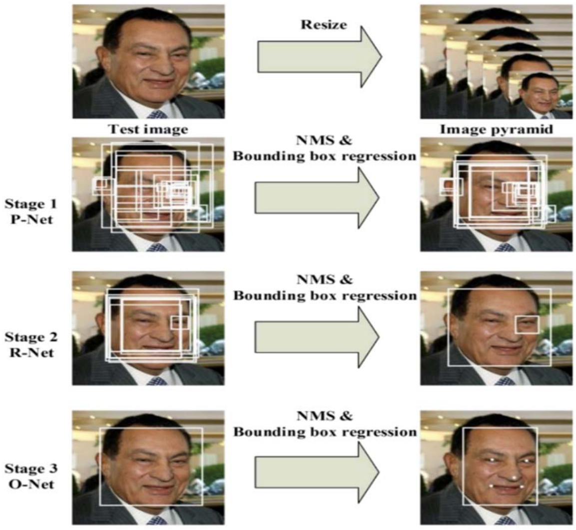


Fig. 3. Pipeline of the Multi-Task Cascaded Convolutional Neural Network [15].

C. Image Post-processing:

After face detection, crop and resizing methods are applied to input images. The bounding box founded in the face detection level is produced by the MTCNN model. The bounding box then used to crop the face portion from the input images. A particular model normalized parameters are specified in the model architecture details. According to the FaceNet architecture details all the cropped images are resized to 160x160.

D. Feature Extraction using FaceNet

FaceNet is the start-of-art for face recognition [9], identification, verification and clustering of neural networks. This pre-trained FaceNet model is used here as a baseline for a deep network. The FaceNet is combined with a batch layer and a very deep CNN network. The deep CNN is supported by L2 normalization. Then face embedding is the result of this normalization. During training, the face embedding is pursued with triplet loss. When the identities are the same then the triplet loss has minimum distance value between an anchor and a positive. But the distance value is higher between a negative and the anchor when the identities are different. Figure 4 shows the pipeline of FaceNet model.

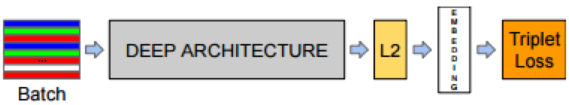


Fig. 4. The Pipeline of FaceNet model [9].

FaceNet developed on 22 deep convolutional network layers. Its output is directly trained on these deep layers to obtain a compact 128-dimensional embedding. After rectification the fully connected layer to be used as the face descriptor. These descriptors turn into a similarity-based descriptor using the embedding module. To prepare a unique feature vector from a template, max operator has been applied to the features.

For the particular task of face recognition and verification, the network must be fine-tuned for expecting a significant boost. A very large amount of masked and non-masked face images are used to this work to re-train the FaceNet model.

E. Face Verification Using SVM

Lastly, this verification process is consolidated to recognize candidates face by performing the classification task within a unified Support Vector Machine (SVM). The SVM algorithm is effectively applied in different classification related problems since it was introduced [16]. SVM finds out a hyperplane to perform classification tasks of an optimization problem. It maximizes the border among the two classes of a given input-target pair. The classifier is the outcome with a particular level of robustness to overfitting. The margin represents the class separation efficiency.

In this segment, we examine a test face to other train faces using SVM. The classification result is considered as correct if the distance among the test image and the train image of the identical person is minimum. A masked face similarity is measured upon the masked and non-masked face by estimating an L2 normalization within the features key points collected from the net structure.

IV. DATASET CONSTRUCTION

To train the deep learning networks requires a large number of images. The scarcity of large masked face datasets and lack of facial makes masked face recognition more challenging. To alleviate the shortage of large datasets, we proposed our Masked Face Database (MFD). A different level of occlusions and orientations are available in this database. We added more complicated masked face images in our MFD Database. Therefore, our Masked Face Dataset (MFD) has 45 subjects including various disguises, comprising of different simple and complex backgrounds. The offered dataset has been made of 990 images with male and female subjects aged between 18 to 26 years. The images of masked face datasets have been collected from 45 subjects with different backgrounds and 7 different levels of disguises. The disguises are (i) sunglass (ii) scarf (iii) medical mask (iv) beard (v) sunglass with scarf (vi) sunglass with a beard and (vii) sunglass with a medical mask.

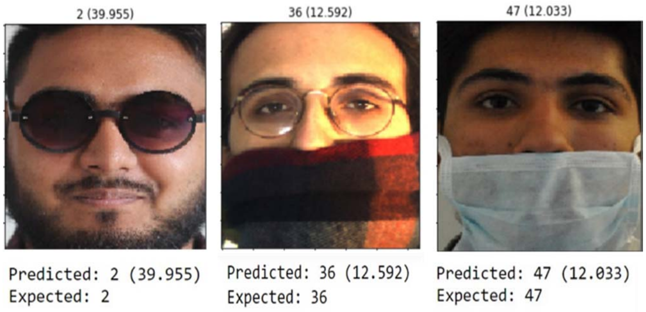
We also use two existing databases AR Face Database [17] and IIIT-Delhi Disguise Version 1 Face Database [18] to validate our work. The example images are shown in Figure 5 from each database.

Fig. 5. Example images from (a) AR [17] (b) IIIT-Disguise V1 [18] (c) MFD face database.

We divide our datasets images into test and train set. Each train set contains an average of 70% of total images and the test set contains an average of 30% of the total images of each dataset.

V. EXPERIMENTAL RESULTS

We evaluate our approach on the AR face database, IIIT Disguise Face Database, and our MFD database. Figure 6 shows the testing output of different database images.



(a) (b) (c)

Fig. 6. Test images from (a) MFD (b) AR (c) IIIT Disguise V1 face database

Accuracy of masked face recognition has been measured on these datasets. In this work different combinations of masked and non-masked face images are made to find out the best recognition accuracy. Table-I shows train and testing scenarios with different combinations of masked and non-masked face images. And table-II shows the recognition accuracy of all the scenarios of table-I.

TABLE II. MASKED FACE RECOGNITION ACCURACY

Scenario	**Train Image**	**Test Image**	**Train Accuracy (%)**	**Test Accuracy (%)**
Scenario 1	Non-masked Faces	Masked Faces	100.00	90.40
Scenario 2	Non-masked Faces + Masked Faces	Masked Faces	99.96	98.50
Scenario 3	Non-masked Faces	Masked Faces	100	89.49
Scenario 4	Non-masked Faces + Masked Faces	Masked Faces	99.37	82.21
Scenario 5	Non-masked Faces	Masked Faces	100.00	63.52
Scenario 6	Non-masked Faces + Masked Faces	Masked Faces	99.91	98.10
Scenario 7	Non-masked Faces	Masked Faces (Complex)	100.00	47.43
Scenario 8	Non-masked Faces + Masked Faces (Complex)	Masked Faces (Complex)	99.05	90.24

From the above table-I, accuracy of masked face image recognition using FaceNet is higher when we mixed masked and non-masked images for training rather than when we only use non-masked images for training. In the meantime, it has been found that when the complexity of masked increased then the recognition accuracy decreased. A graphical representation of these recognitions accuracy also presented in figure 7 for a better data visualization.

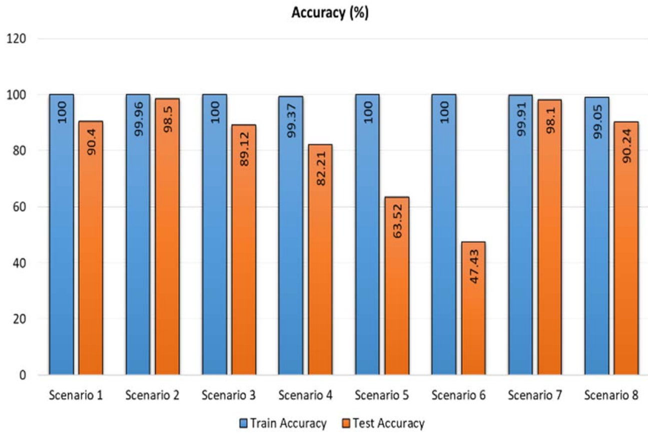
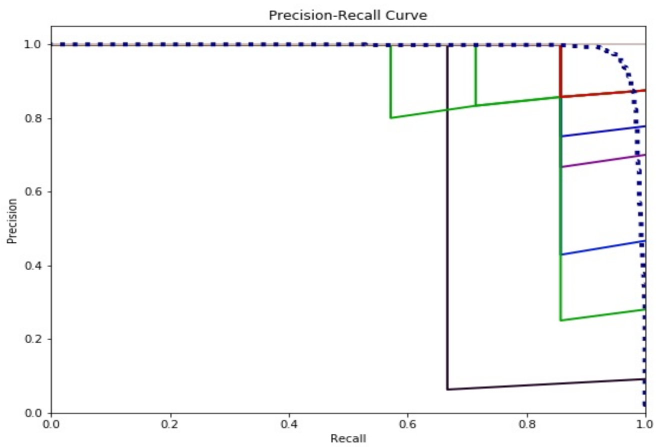


Fig. 7. Training and testing accuracy of masked face recognition

To show the performance and correctness of a classification algorithm, the Precision-Recall curve is another visualization technique. For various thresholds, the precision-recall curve has been shown different changes among precision and recall. The larger area under the curve shows both high precision rate and high recall rate. High precision defines a lower false-positive rate. At the same time, high recall defines a lower false-negative rate. Both large values indicate that the classifier gives correct outputs.

Receiver Operating Characteristic (ROC) curve is another visualization method which shows the actual performance of a classification model. The ROC curve presents the changes of recall vs precision correlation with the variation of the threshold value for defining a positive in our model. In the ROC curve, the X-coordinate defines the false positive rate and the Y- coordinate defines the true positive rate.

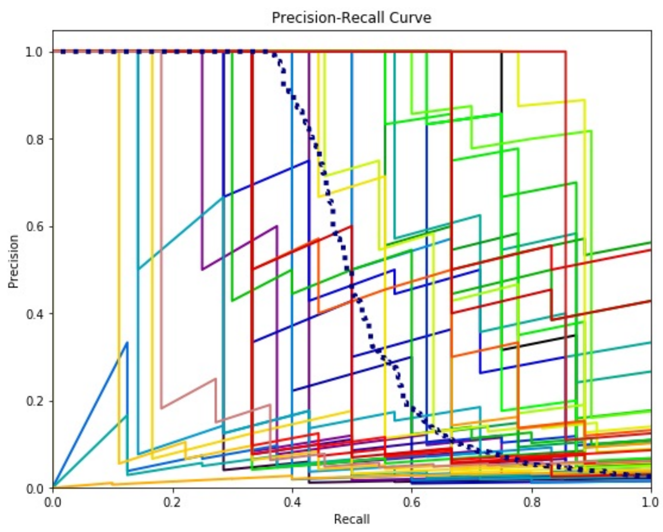
Therefore we determine both the Precision-Recall curve and ROC curve for each scenario of table-I. Figure 8 shows only Scenario 2 (the highest accuracy) and Scenario 7 (the lowest accuracy) Precision-Recall curve and ROC curve of this work.



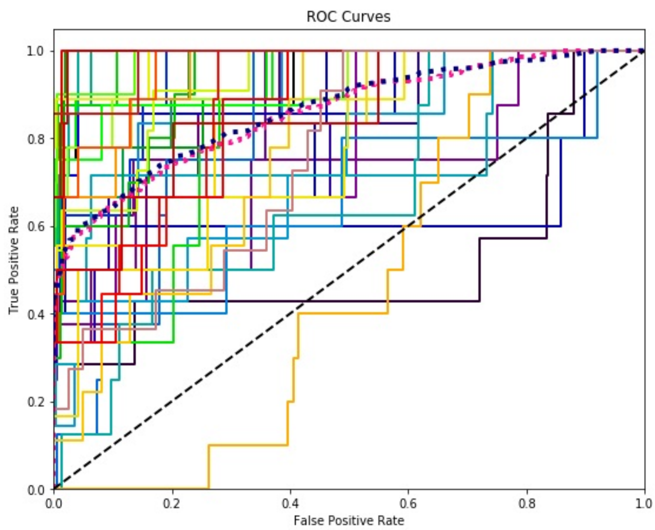
(a)

TABLE I. MASKED FACE RECOGNITION TRAINING AND TESTING SCENARIO

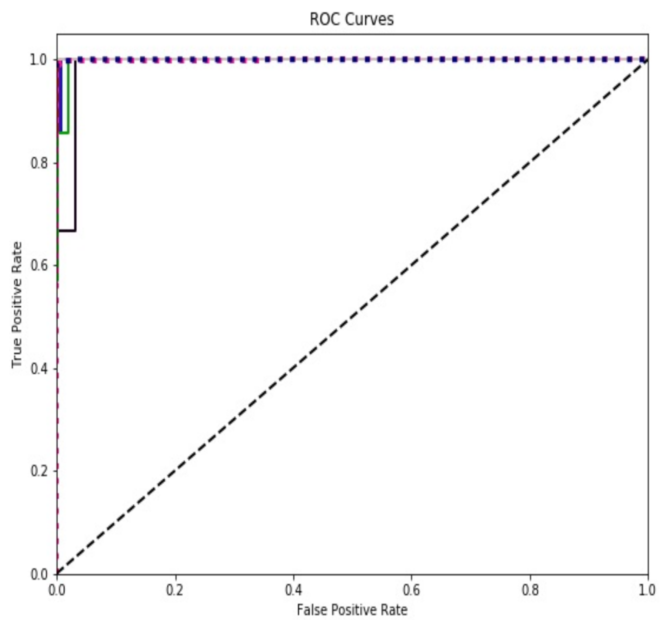
Dataset	Scenario	Subjects			Train			Total Test Images (All Masked)	Total Images (Train + Test)
		Male	*Female*	*Total*	*Masked*	*Non-Masked*	*Total*		
AR	Scenario 1	76	59	135	0	1666	1666	760	2426
	Scenario 2	76	59	135	704	1666	2370	935	3305
IIIT Version 1	Scenario 3	58	17	75	0	1273	1273	533	1806
	Scenario 4	58	17	75	325	148	473	208	681
Proposed MFD	Scenario 5	31	14	45	0	900	900	233	1133
	Scenario 6	31	14	45	178	895	1073	315	1388
	Scenario 7	31	14	45	0	900	900	350	1250
	Scenario 8	31	14	45	253	899	1152	461	1613



(b)



(d)



(c)

Fig. 8. (a) (b) Precision-Recall Curve and ROC Curve for Scenario 1
(c) (d) Precision-Recall Curve and ROC Curve for Scenario 7

VI. CONCLUSION

In this work, the FaceNet pre-trained model has been used for improving masked face recognition. We have benchmarked this approach with two well-known datasets and our dataset. Our approach tested on these datasets shows better recognition rates. So FaceNet model trained on masked and non-masked images gives better accuracy for simple masked face recognition. Although we concentrated on masks induced by a hat, sunglasses, beard, long hairs, mustache, and medical mask, our methodology can still be extended to more complex and many other sources of occlusion. Obviously, this method may not be appeasement for all types of masks. Further, the more accurate and sophisticated approach may than be needed. In later work, it is our importance to enhance and enlarge our work to address different extreme masks condition of face recognition.

At the same time Industry 4.0 and/or Sustainable Technology will try to enhance the computer adoption and mechanization with autonomous and intelligent systems fed by data and machine learning. Working with data, security is essential. So our work can help these smart and autonomous industries to be more self-governing, secure, accurate and efficient which helps more production and less waste.

VII. References

[1] Roomi, Mansoor, Beham, M.Parisa, "A Review Of Face Recognition Methods," in *International Journal of Pattern Recognition and Artificial Intelligence*, 2013, 27(04), p.1356005

[2] A. S. Syed navaz, t. Dhevi sri, Pratap mazumder, " Face recognition using principal component analysis and neural network," in *International Journal of Computer Networking, Wireless and Mobile Communications (IJCNWMC)*, vol. 3, pp. 245-256, Mar. 2013.

[3] Turk, Matthew A., and Alex P. Pentland. "Face recognition using eigenfaces," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 1991, pp. 586-591.

[4] H. Li, Z. Lin, X. Shen, J. Brandit, and G. Hua, "A convolutional neural network cascade for face detection," in *IEEE CVPR*, 2015, pp.5325-5334.

[5] Wei Bu, Jiangjian Xiao, Chuanhong Zhou, Minmin Yang, Chengbin Peng, "A Cascade Framework for Masked Face Detection," in *IEEE International Conference on CIS & RAM*, Ningbo, China, 2017, pp.458-462.

[6] Shiming Ge, Jia Li, Qiting Ye, Zhao Luo, "Detecting Masked Faces in the Wild with LLE-CNNs," in *IEEE Conference on Computer Vision and Pattern Recognition*, China, 2017, pp. 2682--2690.

[7] M. Opitz, G. Waltner, G. Poier, and et al, "Grid Loss: Detecting Occluded Faces," in *ECCV*, 2016, pp. 386-402.

[8] X. Zhu and D. Ramanan, "Face Detection, pose estimation and landmark localization in the wild," in *IEEE CVPR*, 2012, pp.2879-2886.

[9] Florian Schroff, Dmitry Kalenichenko, James Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815-823

[10] Ankan Bansal Rajeev Ranjan Carlos D. Castillo Rama Chellappa, "Deep Features for Recognizing Disguised Faces in the Wild," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2018.

[11] Maneet Singh, Richa Singh, Mayank Vatsa, Nalini K. Ratha, and Rama Chellappa. "Recognizing disguised faces in the wild," in *IEEE Transactions on Biometrics, Behavior, and Identity Science* 1, no. 2, 2019, pp. 97-108.

[12] K. Yan, S. Huang, Y. Song, W. Liu and N. Fan, "Face recognition based on convolution neural network," in *36th Chinese Control Conference (CCC)*, Dalian, 2017, pp. 4077-4081.

[13] Amarjot Singh, Devendra Patil, G Meghana Reddy, Sn Omkar, "Disguised Face identification (DFI) with facial Key Points using Spatial Fusion Convolutional Network," in *IEEE International Conference on Computer Vision Workshops (ICCVW)*,Venice, Italy, 2017, pp. 1648 - 1655.

[14] Chandra Churh Chatterjee, "Towards Data Science," [Online]. Available:https://towardsdatascience.com/basics-of-the-classic-cnn- a3dce1225add. [Accessed 1 September 2019].

[15] K. Zhang, Z. Zhang., Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," in *IEEE Signal Processing Letters 23*, no.10, 2016, pp.1499-1503.

[16] Byun, Hyeran, and Seong-Whan Lee, "Applications of support vector machines for pattern recognition: A survey," in *International Workshop on Support Vector Machines*, pp. 213-236. Springer, Berlin, Heidelberg, 2002.

[17] Aleix M Martinez, "AR Face Database," [Online]. Available: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html. [Accessed 1 September 2019].

[18] "Image Analysis and Biometrics Lab @ IIIT Delhi," [Online]. Available: http://www.iab-rubric.org/resources/facedisguise.html. [Accessed 1 September 2019].

[19] N. Kohli, D. Yadav, and A. Noore, "Face Verification with Disguise Variations via Deep Disguise Recognizer," in *CVPR Workshop on Disguised Faces in the Wild*, 2018.

[20] S. V. Peri and A. Dhall, "DisguiseNet : A Contrastive Approach for Disguised Face Verification in the Wild," in *CVPR Workshop on Disguised Faces in the Wild*, 2018.

[21] K. Zhang, Y.-L. Chang, and W. Hsu, "Deep Disguised Faces Recognition," in *CVPR Workshop on Disguised Faces in the Wild*, 2018.

[22] T. Y. Wang and A. Kumar, "Recognizing human faces under disguise and makeup," in *IEEE International Conference on Identity, Security and Behavior Analysis*, 2016.

[23] Yingcheng Su, Zhenhua Guo, Yujiu Yang, Weiguo Yang, "Face Recognition with Occlusion," in *3rd IAPR Asian Conference on Pattern Recognition,* 2015.

[24] G. Righi, J. J. Peissig, and M. J. Tarr, "Recognizing disguised Faces," in *Visual Cognition*, pp. 143–169, 2012.

[25] R. Singh, M. Vatsa, and A. Noore, "Face recognition with disguise and single gallery images," in *Image and Vision Computing*, pp. 245–257, 2009.

[26] R. Min, A. Hadid and J. Dugelay, "Improving the recognition of faces occluded by facial accessories," in *Face and Gesture 2011(FG)*, Santa Barbara, CA USA, 2011, pp. 442-447.

[27] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa, "Disguised faces in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018.

[28] G. Apostolopoulos, V. Tzitzilonis, V. Kappatos and E. Dermatas, "Disguised face identification using multi-modal features in a quaternionic form," in *7th International Conference on Imaging for Crime Detection and Prevention (ICDP)*, Madrid, 2016, pp. 1-6.

[29] A. Matin, F. Mahmud, T. Ahmed and M. S. Ejaz, "Weighted score level fusion of iris and face to identify an individual," *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Feb. 2017, pp. 1-4.

[30] M. S. Ejaz, M. A. Hossain, A. Matin and M. T. Ahmed, "Performance Comparison of Partition Based Clustering Algorithms on Iris Image Preprocessing, " in *International Conference on Electrical & Electronic Engineering (ICEEE)*, Dec. 2017, pp. 1-4.