VGG FaceNet Based Sketch to Face Recognition with Morphable Model

Ms. Ajita A. Patil, Research Scholar,
Department of Electronics and Telecommunication Engg. SCOE,Kopargaon.
Affiliated to Savitribai Phule Pune University,Pune
ajita24patil@gmail.com

Dr. B. S. Agarkar, Professor
Department of Electronics and Telecommunication Engg. , SCOE,Kopargaon
Affiliated to Savitribai Phule Pune University,Pune
bsagarkar977@gmail.com

Abstract— **Sketch to face recognition automation can play important role in forensic operations. The forensic departments can generate sketches with the help of drawing artists. The resulting sketch images may have difference compared to actual faces in terms of facial parts and expressions. The convolutional neural network (CNN) based method proposed in this paper shows augmentation based sketch and facial expression dataset generation by modifying the public dataset. The generated dataset is thus used to train the VGGFaceNet CNN model and performance is evaluated. The performance of VGGFaceNet model is tested with reference to parameters like accuracy, specificity and sensitivity. The proposed system indicates accuracy of 88% over to other conventional methods such as Local Binary Pattern, Support Vector Machine.**

Keywords— Face recognition, Morphable model, VGGFaceNet, Sensitivity.

I. INTRODUCTION

In biometric domain, Face Recognition has significance. There has been considerable evolution in face recognition systems in recent years. This has given direction and considerable proofs to the scientists working in the field of face recognition. In biometric domain, various applications of face recognition can be seen which includes indexing and compression mechanisms. The end user may have different interests of face recognition algorithms which includes classification of multimedia contents and search based applications. In forensics and surveillance systems face recognition plays considerable role. Face recognition is useful in Banking System, Law Enforcement and Authentication Mechanism. Authorized users have been provided access to such datasets along with controls. Face Recognition shows considerable need in forensic work for identification of people as there is rise in the criminal acts. Along with enhanced security, fast identification of people is the main objective in face recognition systems.

As face recognition is concerned with forensic field work, during primary evidence acts, artists can develop the face sketches with the help of information from eye witnesses. These sketches may not have perfect appearance of face and may deviate in terms of characteristic features. The characteristic features during face recognition algorithm working are shape and dimension based parameters. In case of hand drawn sketches, these parameters vary beyond control and actual face may have variations in terms of changes in eyes, nose, and mouth region along with chick and chin characteristics. Also, changes in these face parts due to emotional variations may also vary. The systems which work on sketch to face linkage and face recognition task, may show degradation in performance.

The sketch to face linkage with respect to various changes in face parts and generation of new sketches for face recognition dataset generation is first phase of work presented in this paper. The generated dataset is then used to train convolutional neural network (CNN) in second phase of work. Work done in two phases is explained in proposed work section after addressing existing techniques in the literature review section. The trained model is then evaluated for recognition of original face using sketch image input. The performance evaluation is shown in results and analysis section. This paper contributes in sketch to face detection work domain with novel deep learning approach and better accuracy.

II. LITERATURE SURVEY

Texture descriptors and edges, as well as landmarks, are frequently used in traditional face recognition algorithms [1]. Machine learning techniques such as Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), and Principal Component Analysis (PCA) are used to combine these distinguishing qualities. Scientists have turned their focus to other approaches such as age-invariant methods [2], [3], pose-invariant methods [4] and illumination-invariant methods [5], [6] due to the difficulty of these methods.

Apoorva et al [7] developed a Haar classifier for facial identification using a security camera. There were four steps in the system: (1) real-time picture training (2) Haar-classifier face recognition (3) comparing real-time photographs with photos taken with the camera; (4) generating a result based on the comparison Haar is used in real-time applications to recognize faces with high accuracy. The Open CV Library platform can be used for Haar cascading is a face detection technique. The library supports for both the operations that is tracking of faces and recognition. In this circumstance, the accuracy levels are considerably good. India has adopted the Aadhar system for citizen identification. A clear identification between a local citizen and a foreigner can be done with the use of this as a citizen database. This data can be used to assess whether or not the individual is a criminal.

Raj [8] demonstrated a real-time facial recognition PCA. The system works efficiently and experimentation is carried out using C++ and OpenCV. The system uses a number of distance classifiers to extract features. The three major distance classifiers utilized here are, Euclidian distance, Manhattan distance and Mahalanobis distance. The optimal

978-1-6654-9291-1/22/$31.00 ©2022 IEEE

identification rate is estimated to be around 92.3 percent using Mahalanobis distance. This is a huge improvement over the 73.1 percent achieved with ordinary PCA. Face recognition questions are answered in an estimated 0.2 seconds, which is quite fast.

Bah and Ming [9] devised a Local Binary Pattern (LBP) for recognizing faces. To improve the system's overall accuracy, this was integrated with other image processing techniques such as Contrast Adjustment, Histogram Equalization and Bilateral Filter. Here, the LBP codes have been updated, resulting in enhanced system performance. According to the experimental results, the technique is dependable, durable, and accurate. In a real-world context, it can be utilized as an attendance management system.

Kumar et al [10] proposed a method which uses Haar cascades and AdaBoost for real-time face recognition system. This system collapses the majority of the variance. Haar cascade is combined with AdaBoost to distinguish human faces. LDA and a fast PCA make face recognition easier. Attendance is recorded in the laboratory using matched faces. It is gaining popularity as a real-time attendance system that uses rapid and efficient algorithms because it is a biometric system. This process has a high level of accuracy as well.

Shieh et al [11] built real-time facial recognition systems using a PCA combination with SVM-Particle Swarm Optimization. Because of its capacity to reduce dimensionality, a PCA-based technique is used in the majority of human-robot interaction applications. SVMs are used to implement PSO's fitness functions for classification tasks, while PSO is used to implement feature selection. The results reveal that the proposed technique lowers features completely while maintaining high classification accuracy.

An LBP for real-time face recognition was presented by Shubha and Meenakshi [12]. To depict it, information on the texture and shape of the face is used. To portray the face in detail, the face is separated into many portions. The histograms of LBP are then extracted and concatenated to form a single histogram. The Nearest Neighbor classifier is then used to recognize faces. A prototype model built using the Raspberry Pi single-board computer and MATLAB is used to test the concept. When compared to other approaches, the face recognition rate of the LBP algorithm is relatively higher.

In a demanding context, Zhang et al [13] suggested an efficient and robust facial recognition system. The signal processing modalities used in its implementation include Ada Boost, cascade classifier, PCA, Haar like feature and LBP. This method uses a cascade classifier to train precision eye and face detectors. The LBP descriptor, which can distinguish faces quickly, is used to extract facial features. The algorithm is used to detect eyes, lowering the rate of incorrect face detection. Face recognition is done using the PCA method. The face recognition system is trained using large datasets of pictures of faces and non-faces. The algorithms are extremely accurate for facial recognition.

Machine learning-based face recognition systems are effective. These algorithms, however, take a long time to process and train. Deep learning algorithms have mostly supplanted facial recognition algorithms in recent years. Deep learning has been demonstrated to be more effective when dealing with large data sets. When working with smaller datasets, traditional machine learning, on the other hand, performs well. A task must be split down into discrete steps in traditional machine learning. Then you must solve each step separately. One algorithm should be used for feature extraction and the other for facial detection when it comes to facial recognition. Deep learning could be able to assist. The acquisition of random face datasets was facilitated by Faces in the Wild, a vast online archive of faces [14-18]. Variations from real life are presented. On big datasets, face recognition algorithms based on CNN [19] have been trained with excellent results. The growing usage of deep learning has expedited face recognition research. Computer vision applications that use CNNs include age estimation, facial expression analysis, segmentation, object recognition, and detection.

For real-time facial detection, Rekha and Kurian [20] presented a HOG descriptor. Using the HOG of an image, the approach determines the weights associated with facial traits. Features like the lips, nose, and eyes are assigned positive weights. This will aid in imagining a face in its entirety. The algorithm can recognize faces from a variety of angles, including obstructed faces. These systems rely exclusively on human-determined attributes and use a typical computer vision algorithm. Once the system generates the findings, the strength of the features is determined. In deep learning, the algorithm finds the best attributes that are unique to the situation. CNN has made significant advancements in the field of computer vision throughout the years. While utilizing CNN, the accuracy of facial recognition has improved significantly. The presence of a large-scale dataset is one of the main reasons behind this.

Almabdy and Elrefaei [21] proposed a solution for facial recognition that integrated SVM and CNN. The study takes into account the CNN architecture, which has performed well in the ILSVRC in recent years. According to the findings, the model outperformed other modern models in terms of accuracy. The accuracy was found to range between 94 and 100 percent. Furthermore, recognition improved dramatically, reaching 39 percent.

Passos et al [22] proposed a deep neural network based approach. It uses Multi-Layer Perceptron (MLP) with CNN for facial recognition. It's a deep-learning-based open-source solution for facial identification. To extract fiducial points and embedding, deep learning algorithms are used. SVM is used for classification jobs since it is fast for both training and inference. The system achieves an error rate of 0.12103 for facial characteristics detection, which is comparable to state-of-the-art approaches, and 0.05 for face recognition. It is also capable of running in real-time.

Saypadith and Aramvith [23] developed a CNN model for facial recognition on the embedded GPU system. Face tracking and a deep CNN facial recognition algorithm are used in the system, as well as CNN-based facial recognition. According to the findings, the system is capable of distinguishing various faces. Within 0.23 seconds, an estimated 8 faces can be detected concurrently. The recognition rate exceeded 83.67 percent.

Schroff et al. [24] developed the CNN based model. The model named FaceNet system, uses many pictures to directly map a face as Euclidean space. Face recognition tasks, grouping, and verification are implemented after the space is created using various methodologies. In contrast to the bottleneck layer utilized in earlier approaches, the method uses a Convolutional network. The representational efficiency is the key benefit here. 128 bytes per face can be used for facial recognition.

To recognize faces, Sun et al [26] proposed using a DeepID method. DeepID's generalization capabilities might be used to increase the number of faces in training. The training data set tested 10,000 different facial identities in order to minimize the number of neurons. The characteristics are usually retrieved using over-complete representations from various face areas.

The facial landmark detection problem was studied by Wu et al [27]. According to the findings, facial photos can be classified into several subsets. A CNN architecture, according to the findings, recognizes faces in specific appearances and positions. Training examples could be developed to address the issue of training data scarcity. Traditional landmark detection algorithms may be outperformed by the Tweaked CNN.

III. PROPOSED WORK

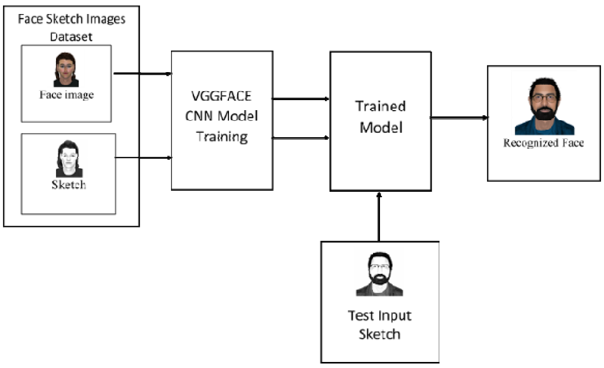


Figure 1: Proposed Face Recognition system using VGGFace Model

Preparation of Dataset:

3D morphable model is used to generate the image dataset from face images and sketches with facial expression changing mechanism. The 3D morphable model makes use of morphable model to construct the facial expression images from original input.

The steps in 3D morphable model for changing the facial expressions are done as described in next part:-

1. Estimate facial region
2. Segmentation of morphable face region using face analyzer
3. Estimate important landmarks of face for facial expressions
4. Apply warping and shifts to change the facial expressions

The respective paired changes in original image and in sketch image are done to prepare the dataset for training. The generated images are then stored in folders with numbers. Different facial expressions of the same person are stored in same folder to establish class label structure. The input image and its respective facial expression generated image is shown in figure 2.

(a) (b)

Figure 2: (a) Input Original Face Image (b) Output Face Image with Changes in Expression

VGGFaceNet: The VGGFaceNet model was designed by the Visual Geometry Group (VGG) at Oxford University. It is one of the most widely used and widely recognized facial recognition technologies. It was trained on 2.6 million images of over 2600 people and contains 38 layers. Each of the thirteen convolutional layers in the VGGFaceNet has its own set of hybrid parameters. Each group of convolutional layers has max pooling layers as well as 15 corrected linear units (ReLUs). The FC6, FC7 and FC8 layers, which are all fully connected, follow these layers [15]. Each of the first two has 4096 channels, while the IDs are classified using FC8, which has 2622 channels. The classifier, a softmax layer used to categories a picture and decide which face class each individual belongs to, is the final layer [4].

IV. RESULTS AND ANALYSIS

Database:

The proposed model is evaluated using publicly available UoM-SGFS database [28]. The dataset contains the face images. Total number of 600 face images are available. The preprocessing is performed using warping mechanism to generate new facial expression along with variations in face parts appearance. Around 5 different attempts are done to modify the facial expressions. Original five face images and new 5 images of facial expression gives total 10 face images for each face from 600 face classes. The generated and original images are then converted into sketches using various steps. The facial expression changes are done using facial model based operation.

Steps of sketch generation:

1. *Take input face image*
2. *Apply facial model to extract face region*
3. *Apply warping to change facial expression*
4. *Apply edge detection*
5. *Apply morphological operations such as opening and closing*
6. *Extract structural regions from regions other than edge regions*
7. *Mix the structural features and edge outputs to generate final sketch images*

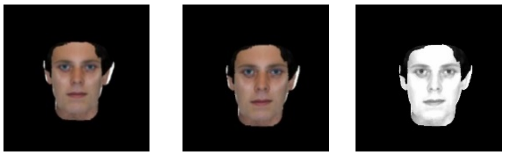
For all 10 face images of each face from face dataset, sketch images are generated. This gives total 20 images for each face image in which 10 images are color face images with different facial expressions and 10 respective sketches. Total 6000 images are seen after preprocessing the dataset.

Figure 3 shows the facial model output for extracting landmark based face region and applying facial expressions.

(a) (b) (c)

Figure 3: Face region extraction (a) Face model recognized, (b) face model fitted in face image (c) Masked image output

Following Figure 4 shows input original face, its output facial expression change and its sketch converted output.



(a) (b) (c)

Figure 4: (a) Input Original Face Image (b) Change in facial expression (c) Output Face Sketch Image

The dataset is split into training and testing parts in 80% and 20% respectively. The training and validation graphical output is also considered in performance calculation.

The performance evaluation is done for different face for detecting the accuracy. The set of 50 images is used for recognition of faces from sketch input. The respective output of original faces is considered for evaluation. The accuracy, specificity and sensitivity parameters are used to evaluate the results. Table 1 shows the formulae for these parameters. For calculation of these parameters we need True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) of the respective image input.

Graphical representation of parameters evaluated is shown in figure 5.

Table 1: Formulae for performance parameters

Accuracy	(TP+TN)/(TP+TN+FP+FN)
Specificity	TN/(TN+FP)
Sensitivity	TP/(TP+FN)

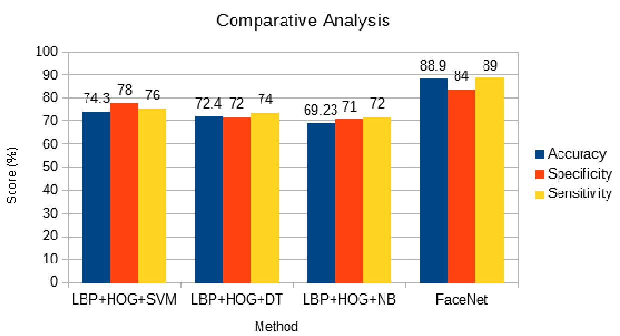


Figure 5: Analysis of methods developed for Face recognition

The performance evaluation is carried out by using conventional feature extraction and conventional classifiers. The feature extraction is done using local binary pattern (LBP) [12] and histogram of Gradient (HOG) [20] techniques. The conventional classifiers such as support vector machine (SVM), Naïve Bays (NB) and decision tree (DT) are used. The performance evaluation shows, VGG FaceNet model outperforms other methods in terms of accuracy, specificity and sensitivity.

V. CONCLUSION

This paper contributes the work in face recognition domain with sketch to face recognition system. The synthetic face dataset is used based on which new augmented dataset is generated. The facial expressions modification and sketch generation is performed to compose a dataset which is used to train VGGFaceNet model. The trained model is evaluated which shows improved performance with respect to sensitivity over other methods. The accuracy, specificity and sensitivity of 89%, 84% and 88.9% is seen.

The model accuracy further can be improved with use of right candidate feature extraction approach which may involve attention layer addition in the model.

REFERENCES

[1] Zhu, X. and Ramanan, D. (2012) Face Detection, Pose Estimation and Landmark Localization in the Wild. IEEE Conference on Computer Vision and Pattern Recognition, Providence, 16-21 June 2012, 2879-2886.

[2] Park U., Y. Tong, and A. K. Jain, “Age-invariant face recognition,” IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 5, pp. 947–954, 2010.

[3] Li Z., U. Park, and A. K. Jain, “A discriminative model for age invariant face recognition,” IEEE transactions on information forensics and security, vol. 6, no. 3, pp. 1028–1037, 2011.

[4] Ding C. and D. Tao, “A comprehensive survey on pose-invariant face recognition,” ACM Transactions on intelligent systems and technology (TIST), vol. 7, no. 3, p. 37, 2016.

[5] Liu D.-H., K.-M. Lam, and L.-S. Shen, “Illumination invariant face recognition,” Pattern Recognition, vol. 38, no. 10, pp. 1705–1716, 2005.

[6] Tan X. and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” IEEE transactions on image processing, vol. 19, no. 6, pp. 1635–1650, 2010.

[7] Apoorva P., H.C. Impana, S.L. Siri, M.R. Varshitha and B. Ramesh, Automated Criminal Identification by Face Recognition using Open Computer Vision Classifiers, IEEE 3rd International Conference on Computing Methodologies and Communication (ICCMC), Pp.775-778, 2019.

[8] Raj D., A realtime face recognition system using pca and various distance classifiers, CS676: Computer Vision and Image Processing, Pp.1-11, 2011.

[9] Bah S.M. and F. Ming, An improved face recognition algorithm and its application in attendance management system, Array, 2020.

[10] Kumar K.S., V.B. Semwal and R.C. Tripathi, Real time face recognition using adaboost improved fast PCA algorithm. arXiv preprint arXiv:1108.1353, 2011.

[11] Shieh M.Y., J.S. Chiou, Y.C. Hu and K.Y. Wang, Applications of PCA and SVM-PSO based real-time face recognition system, Mathematical Problems in Engineering, 2014.

[12] Shubha P. and M. Meenakshi, Human Face Recognition Using Local Binary Pattern Algorithm-Real Time Validation, International Conference On Computational Vision and Bio Inspired Computing, Pp.240-246, 2019.

[13] Zhang X., T. Gonnot and J. Saniie, Real-time face detection and recognition in complex background, Journal of Signal and Information Processing, Vol.8, No.2, Pp.99-112, 2017.

[14] Sun Y., X. Wang, and X. Tang, “Deep learning face representation from predicting 10,000 classes,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1891–1898, 2014.

[15] Parkhi O. M., A. Vedaldi, A. Zisserman, et al., “Deep face recognition.,” in BMVC, vol. 1, p. 6, 2015.

[16] Guo Y., L. Zhang, Y. Hu, X. He, and J. Gao, “Ms-celeb-1m: A dataset and benchmark for large-scale face recognition,” in European Conference on Computer Vision, pp. 87–102, Springer, 2016.

[17] Bansal, A. Nanduri, C. D. Castillo, R. Ranjan, and R. Chellappa, “Umdfaces: An annotated face dataset for training deep networks,” 2017 IEEE International Joint Conference on Biometrics (IJCB), ,pp. 464–473, 2017.

[18] Cao Q., L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “Vggface2:A dataset for recognising faces across pose and age,” arXiv preprint arXiv:1710.08092, 2017.

[19] Trigueros, D.S., Meng, L. and Hartnett, M., 2018. Face recognition: from traditional to deep learning methods. arXiv preprint arXiv:1811.00116.

[20] Rekha N. and M.Z. Kurian, Face detection in real time based on HOG, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol.3, No.4, Pp.1345-1352, 2014.

[21] Almabdy S. and L. Elrefaei, Deep Convolutional Neural Network-Based Approaches for Face Recognition, Applied Sciences, Vol.9, No.20, Pp.1-21, 2019.

[22] Passos W.L., I.M. Quintanilha and G.M. Araujo, Real-Time Deep-Learning-Based System for Facial Recognition, Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT), Pp.895-899, 2018.

[23] Saypadith S. and S. Aramvith, Real-Time Multiple Face Recognition using Deep Learning on Embedded GPU System, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Pp.1318-1324, 2018.

[24] Schroff F., D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering", Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 815-823, Jun. 2015.

[25] Sun Y., D. Liang, X. Wang and X. Tang, "DeepID3: Face recognition with very deep neural networks" in arXiv:1502.0087, Feb. 2015.

[26] Sun Y., X. Wang and X. Tang, "Deep learning face representation from predicting 10000 classes", Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1891-1898, Jun. 2014.

[27] Wu Y., T. Hassner, K. Kim, G. Medioni and P. Natarajan, "Facial landmark detection with tweaked convolutional neural networks", Proc. IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 12, pp. 3067-3074, Dec. 2018.

[28] Face Sketch Database: https://cnpgs.wordpress.com/profession/main projects/fpsr/uom-sgfs/