Robust Face Recognition based on VGGNet

Piyush Ingale[1], Divyashri Desale[2], and , Rajni Kanse[3]
1Pimpri Chinchwad college of engineering and Research, Ravet, Pune.
2Department of Electronics and Telecommunication, PCCOER Ravet

Abstract - Face recognition has recently received significant attention as one of most successful applications of image analysis and understanding. There are many reasons for growing interest in face recognition has a rising concern for public security, identity verification. There is a need for face analysis and modelling techniques in the data management and digital entertainment. In the face recognition we are facing challenges like pose variation, variation in illumination, occlusion problem in the face. To deal with all these problems we present face recognition based on VGG16 deep learning model. For the experimentation we have use real time database of images. The performance of proposed method is evaluated on the basis of cross validation accuracy and it has shown better performance compared to previous one.

Keywords: **Face Recognition, VGG, Face Detection, Rectified Linear Unit**

I. Introduction

Face recognition is a technique in which we identify and verify the identity of an individual. In face recognition we compare the features of face of subject with all features of database images. Face recognition basically works in two modes face verification and face identification. The face verification is consisting of one to one match that compare subject's face with template face images. Face identification contains one to many matches [1]

We need face recognition because in this rapidly evolving world speed and automation is an important aspect which can be introduced in real life using face recognition. Face recognition used in , access control, security, surveillance, multimedia management, law enforcement and many others.

There are many challenges in face recognition such as pose variation, change in the lighting conditions, different backgrounds, change in face expression and colour complexity. There are also many limitations to face recognition such as the occlusion and low resolution. In these robust conditions the accuracy drops drastically.

In this research paper for making the model robust to the problems we are using vggnet which is a pretrained model with millions of data. We are using multitask cascaded convolutional networks for face detection followed by vggnet for training of our model using real time database. To deal with many problems like pose variation etc we are using real time database with images which have different pose, illumination, background. We are getting accuracy around 96% even in the robust conditions.

.

II. Methodology

The proposed methodology encompasses face detection and recognition using MTCNN and VGG Net deep learning architecture as shown in Fig 1.

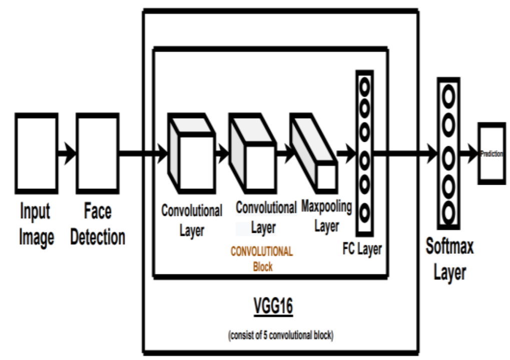

Fig. 1 Flow diagram of proposed system

III. Face detection using MTCNN:

The input images are split into various scaled images to form the image pyramid. This image pyramid is given as input to the MTCNN face detector that includes Proposal Network (P-Net) followed by Refine Network (R-Net) followed by Output Network (O-Net) stage which is almost the same as R-Net. P-Net is CNN based network that selects the possible candidates along with bounding boxes. RNet suppresses the false candidate regions, merges the dense overlapping regions, and calibrates the bounding boxes. O-Net detects facial landmarks such as eye, nose, mouth, etc, and classifies the facial region. Training involves face classification and bounding box regression. The cost function for the face classification is given in equation 1 and the regression problem is defined in equation 2 [16] [17].

B. Face detection using VGG16:

VGG is a pre-trained model with millions of data. It is based on CNN. CNN is very effective in the object & face recognition area. VGG is a feed-forward neural network with multiple convolution layers. The VGG-16 contains 13 layers of convolutional, 5 layers of max polling, and one fully connected layer. In VGG, convolutional uses a 3×3 filter with a stride of 1 while max-pooling uses a 2x2 filter with a stride of 2. The structure of VGG contains a convolution layer, max pooling layer, fully connected layer, and rectified linear unit activation function

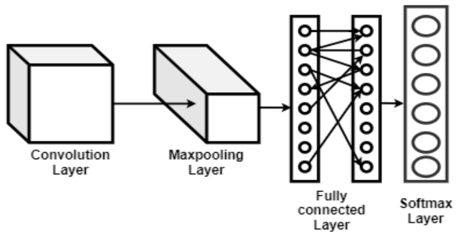

Fig. 2 VGG16 Network architecture

Convolution Layer.
Convolution layer is the layer that convolves the input image (I) with a convolution filter (K) as shown in equation 3. The convolution layer learns features from the input image using a small square of the input image. It gives better feature connectivity, correlation, and representation of the face image. (Equ 3-)

$$conv(I,K)_{x,y} = \Sigma_{i=1}^{nH} \Sigma_{j=1}^{nW} \Sigma_{k=1}^{nC} K_{i,j,k} I_{x+i-1,y+j-1,k}$$

Where, nH, nW and ncstands for height, width, and several filters.

Pooling Layer.
VGG16 uses max polling for polling operations. Max polling increases the conversion rate by decreasing the dimension of the input image. It helps to improve the generalization performance [20].

Rectified Linear Unit.
It performs the non-linear operation by rounding the nonzero values to zero. The mathematical representation of ReLU is:

$$Y = \max(0, x) \tag{5}$$

Where x is the neuron input of the ReLU layer and Y is the output of the ReLU activation function. After the convolution and pooling layer, we get high-level features in form of data. It needs to classify that data into various classes for that purpose. A fully connected layer followed by a Softmax classifier is used to learn a non-linear combination of these features. Finally, the model is trained with Adam's optimization algorithm using a learning rate of 0.001.

IV. Database:

This database contains images of 5000 images of 100 subjects that consist of 50 images per person captured using the mobile camera in the various pose, angles, view, illumination condition, and background environmental conditions. All these images are in RGB format every image has a different resolution and we convert all these images into 312×312×3.

The images are captured by varying the face position in between -65 to +65 degree orientationThe complete database is split into a train and test database in ratio 70% (3500 images) and 30% (1500 images) respectively. Additionally, two standard face image datasets such as ORL [21] and FERET [22] are considered for the evaluation of the performance of the proposed system. The ORL and FERET consist of 400 (40 subjects) and 14000 (1199 subjects) images respectively.

Fig. 3 Sample dataset images

V. Experimental Result

We are designing our model using VGG so the input image has a size of 312×312×3. After passing this image as input, the size changes with every block as 156×156×64, 78×78×128, 39×39×256, 19×19×512, and 9×9×512. The network is trained using a Mini-batch gradient descent algorithm for 15 epochs with a batch size of 50. The feature representation of VGGNet is given in table 1.

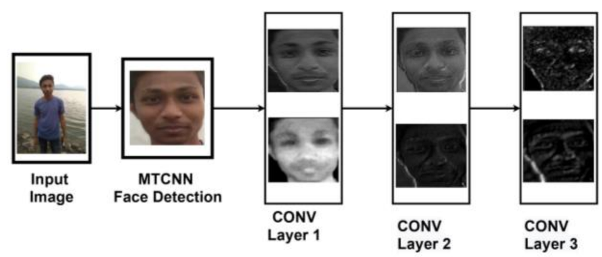

Fig. 4 MTCNN flow diagram

Total trainable parameters are 2073650 and while non-trainable parameters are 14714688. The performance of the proposed system is evaluated on the ORL and FERET dataset and it is noticed that the proposed system gives enhanced performance compared with existing face recognition systems as given in table 2. The proposed system has shown improved performance because of the ability of VGGNet to better represent the face images under various pose, occlusion, illumination, and environmental conditions. The proposed system results in 95.80%, 77.50%, and 98.20% accuracy for ORL, FERET, and proposed real-time face image datasets respectively. For the FERET dataset face recognition rate is lowerbecause some of the images are having larger occlusion.

TABLE 1 NETWORK PARAMTERS AND SPECIFICATION

SR. NO	LAYERS	IMAGE SIZE	NO OF FILTERS	STRIDE	ACTIVATION	PARAMETERS
1	conv[2]	312x312x64	64	1	ReLU	36928
	max pooling	156x156x64	64	2	ReLU	0
2	conv[2]	156x156x128	128	1	ReLU	147584
	max pooling	78x78x128	128	2	ReLU	0
3	conv[3]	78x78x256	256	1	ReLU	590080
	max pooling	39x39x256	256	2	ReLU	0
4	conv[3]	39x39x512	512	1	ReLU	2359808
	max pooling	19x19x512	512	2	ReLU	0
5	conv[3]	9x9x512	512	1	ReLU	2359808
	max pooling	9x9x512	512	2	ReLU	0
6	fully connected	1x41472	-	-	ReLU	0
7	softmax	-	-	-	softmax	2073650

VII. CONCLUSION

In this paper, we have implemented deep learning-based face recognition that can cope up with illumination problems, pose variation, view variation, and occlusion problems. It used two deep learning structures such as MTCNN for face detection and VGGNet for face recognition. It has shown improved performance over the previous techniques for face recognition and results in 95.80%, 77.50%, and 98.20% accuracy under extreme uncontrolled conditions for ORL, FERET, and real-time face database. VGGNet is slower to train and needs larger space which may limit its application for the extremely higher dataset. In the future, the performance of the network can be improved for the larger database and highly occluded face images.

REFERENCES

1. Jain, A. K., & Li, S. Z. Handbook of face recognition (Vol. 1). New York: springer (2011).
2. Bhangale, K. B., Jadhav, K. M., & Shirke, Y. R. Robust Pose Invariant Face Recognition using DCP and LBP. Int. Journal of Mgmt., Tech. and Engg, 8(11), 1026-1034 (2018).
3. Kortli, Y., Jridi, M., Al Falou, A., &Atri, M. Face recognition systems: A Survey. Sensors, 20(2), 342 (2020).
4. Payal, P., &Goyani, M. M. A comprehensive study on face recognition: methods and challenges. The Imaging Science Journal, 68(2), 114-127 (2020).
5. Pan, J., Wang, X. S., & Cheng, Y. H. Single-sample face recognition based on LPP feature transfer. IEEE Access, 4, 2873-2884 (2016).
6. Vigneau, G. H., Verdugo, J. L., Castro, G. F., Pizarro, F., & Vera, E. Thermal face recognition under temporal variation conditions. IEEE access, 5, 9663-9672 (2017).
7. Sonawane, M. U. Inamdar and K. B. Bhangale, "Sound based human emotion recognition using MFCC & multiple SVM," *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC)*, 2017, pp. 1-4, doi: 10.1109/ICOMICON.2017.8279046
8. Bhangale, Kishor, and K. Mohanaprasad. "Speech Emotion Recognition Using Mel Frequency Log Spectrogram and Deep Convolutional Neural Network." In *Futuristic Communication and Network Technologies*, pp. 241-250. Springer, Singapore, 2022.
9. Bhangale, Kishor Barasu, and K. Mohanaprasad. "A review on speech processing using machine learning paradigm." *International Journal of Speech Technology* 24, no. 2 (2021): 367-388.
10. Bhangale, Kishor & Kothandaraman, Mohanaprasad "Survey of Deep Learning Paradigms for Speech Processing" Wireless Personal Communications. 1-37. 10.1007/s11277-022-09640-y, 2022.
11. Biradar, Priya, Priyanka Kolsure, Sujata Khodaskar, and Kishor B. Bhangale. "IoT based smart bracelet for women security." *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)* 8, no. 11 (2020): 688-691.
12. Bhangale, Kishor B., Pranoti Desai, Saloni Banne, and Utkarsh Rajput. "Neural Style Transfer: Reliving art through Artificial Intelligence." In *2022 3rd International Conference for Emerging Technology (INCET)*, pp. 1-6. IEEE, 2022.
13. Sarraf, Rajang, Shalini Ojha, Damini Biraris, and Kishor B. Bhangale. "IoT based smart quality water management system." *International Journal* Of Advance Scientific Research And Engineering Trends5, no. 3 (2020).
14. Mapari, Rahul, Kishor Bhangale, Laukik Deshmukh, Prashant Gode, and Ankit Gaikwad. "Agriculture Protection from Animals Using Smart Scarecrow System." In *Soft Computing for Security Applications*, pp. 539-551. Springer, Singapore, 2022.
15. Anand, Tejveer, Sourabh Upare, Siddhant Jain, Maithili Andhare, and Kishor Bhangale. "Deployment of Real-Time Energy Monitoring System Using IoT."

In *2022 3rd International Conference for Emerging Technology (INCET)*, pp. 1-4. IEEE, 2022.

16. Mapari, Rahul G., Kishor B. Bhangale, Pranjal Patil, Harish Tiwari, Shivani Khot, and Sanjana Rane. "An IoT based Automated Hydroponics Farming and Real Time Crop Monitoring." In *2022 2nd International Conference on Intelligent Technologies (CONIT)*, pp. 1-5. IEEE, 2022.

Mapari, Rahul G., Harish Tiwari, Kishor B. Bhangale, Nikhil Jagtap, Kunal Gujar, Yash Sarode, and Akash Mahajan. "IOT Based Vertical Farming Using Hydroponics for Spectrum Management & Crop Quality Control." In *2022 2nd International Conference on Intelligent Technologies (CONIT)*, pp. 1-5. IEEE, 2022.