Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 171 (2020) 1651–1659

Procedia Computer Science

www.elsevier.com/locate/procedia

Third International Conference on Computing and Network Communications (CoCoNet'19)

Design and Evaluation of a Real-Time Face Recognition System using Convolutional Neural Networks

Pranav KB and Manikandan J*

Crucible of Research and Innovation (CORI) and Department of ECE,
PES University, 100-Feet Ring Road, BSK Stage III, Bangalore 560085, Karnataka, India

Abstract

The advent of high speed processors and high resolution cameras has spearheaded the research towards design of face recognition systems for various applications. Face recognition systems use either offline data or real-time input, based on the application. In this paper, design and evaluation of a real-time face recognition system using Convolutional Neural Network (CNN) is proposed. The initial evaluation of the proposed design is carried out using standard AT&T datasets and the same is later extended towards the design of a real-time system. Details about the tuning of CNN parameters to assess and enhance the recognition accuracy of the proposed system are also reported. A systematic approach to tune the parameters is also proposed to enhance the performance of the system. Maximum recognition accuracies of 98.75% and 98.00% are obtained on using the proposed system with standard datasets and real-time inputs respectively.

Keywords: Convolutional Neural Network; Deep Learning; Face Recognition; Real-Time.

* Corresponding author. Tel.: +91−814−777−8114.
E-mail address: manikandanj@pes.edu

10.1016/j.procs.2020.04.177

1. Introduction

Face recognition is a method to identify or verify the identity of an individual using their face. Face recognition has been used for various applications such as automatic classroom attendance management system in [1], surveillance of access restricted zones including living spaces for intruder detection in [2], recognition of celebrities in public space [3], recognition of house inmates by networked home automation system in [4] and many more. Most of the face recognition systems designed consist of two major modules: feature extraction and classifier. Various combinations of feature extraction and classifier algorithms have been employed for the design of face recognition systems such as Histogram of Gradients (HOG) and Support Vector Machine Classifier(SVM) in [5], HOG and Relevance Vector Machine (RVM) Classifier in [6], Principal Component Analysis (PCA) and SVM in [7]. Convolutional Neural Network (CNN) is a deep learning algorithm, most commonly recommended for applications using images, because it performs the combined task of feature extraction and classification. A detailed literature review on face recognition using various algorithms, datasets, their pros and cons are reported in [8].

In this paper, design of a real-time face recognition using CNN is proposed, followed by the evaluation of the system on varying the CNN parameters to enhance the recognition accuracy of the system. An overview of proposed real-time face recognition system using CNN is shown in Fig. 1. The organization of the paper is as follows. Section 2 gives an overview of Convolutional Neural Network, followed by the experimental results of evaluation, comparison of the results with results reported in literature, conclusion and references.

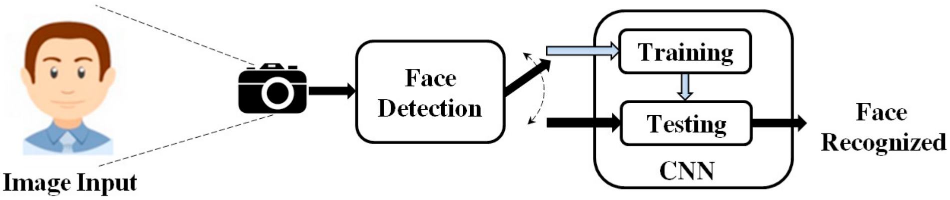


Fig. 1. Block diagram of proposed real-time face recognition system

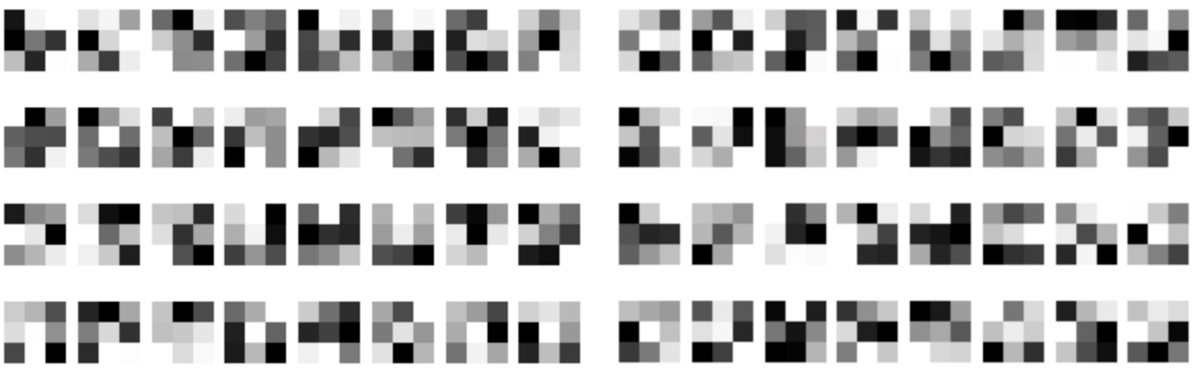
2. Proposed Convolutional Neural Network Architecture

Convolutional Neural Networks are similar to ordinary neural networks, but with an explicit assumption that the inputs are images, allowing designers to encode certain properties into the architecture. CNN architecture comprises of a sequence of layers with the simplest architecture being [INPUT−CONV−RELU−POOL−FC]. INPUT layer holds the raw pixel values of the images, CONV layer consists of a kernel or filter of a fixed size which slides in a window fashion to perform the convolution operation on the windowed image to extract features. Padding is applied onto the size of input image to overcome uneven mapping with filter size. RELU stands for rectified linear units, which is an element wise activation function that assigns zero value to hidden units. POOL denotes the pooling layer, which is responsible for down sampling and dimensionality reduction that in turn reduces the computational power required to process data. Pooling layer also has a kernel or function which slides like a window onto the input to extract dominant features that are rotational and positional invariant. Max pooling and Average pooling are the two common functions used. FC is the fully connected layer where each neuron in the input is connected to each neuron in the output and this layer is responsible in computing the score of a particular class, resulting in N outputs where N denotes the number of classes/categories to be classified. The class with maximum score is decided as the predicted class of the CNN architecture. FC layer is also referred to as DENSE layer. It may be noted that the CNN architecture can be modified based on the design requirements and performance of the system. Some of the other layers that are used in CNN architecture include DROPOUT and FLATTEN. DROPOUT layer is a regularization technique to prevent over fitting of CNN, wherein a fraction of inputs (referred to as dropout rate) are dropped out by setting their values to 0 at each update during training. The values of inputs that are retained are scaled up, so that their sum is unchanged during training. FLATTEN layers are introduced before FC layer to convert the two dimensional features into one dimension.

CNN architectures vary from designer to designer and the sequence of layers can be modified based on repeated evaluations to attain maximum recognition accuracy. The CNN architecture considered for proposed work is given in Fig. 2, after evaluating various combinations of sequence layers. The proposed CNN architecture is designed using an Open Source Neural Network library called Keras running on top of Tensorflow. The CONV layer mentioned in Fig. 2 includes CONV and RELU layers. The real-time input image captured from camera is first fed to Viola Jones algorithm for face detection. The cropped face image is then converted into gray scale, resized to 120×120 pixels and fed to first convolution layer comprising of 32 filters of size 3×3 pixels as shown in Fig. 3. It may be noted that the weights of these filters are initialized to random numbers and they get updated using back propagation algorithm over a set of few epochs to yield final weights as shown in Fig. 3 for these filters. These final weights are later used during classification phase. The output of first CONV+RELU layer with 32 filters mentioned above is shown in Fig. 4, which is in turn fed to second CONV+RELU layer with a different set of 32 filters of size 3×3 pixels to yield an output as shown in Fig. 4. The output from second CONV+RELU layer is fed to POOL layer with max pooling function using a window size of 4×4 pixels. An illustration of POOL layer output using max pooling and average pooling is shown in Fig. 5. It is observed during evaluation that max pooling gave better accuracy over average pooling for the proposed work and hence max pooling is employed in this work.

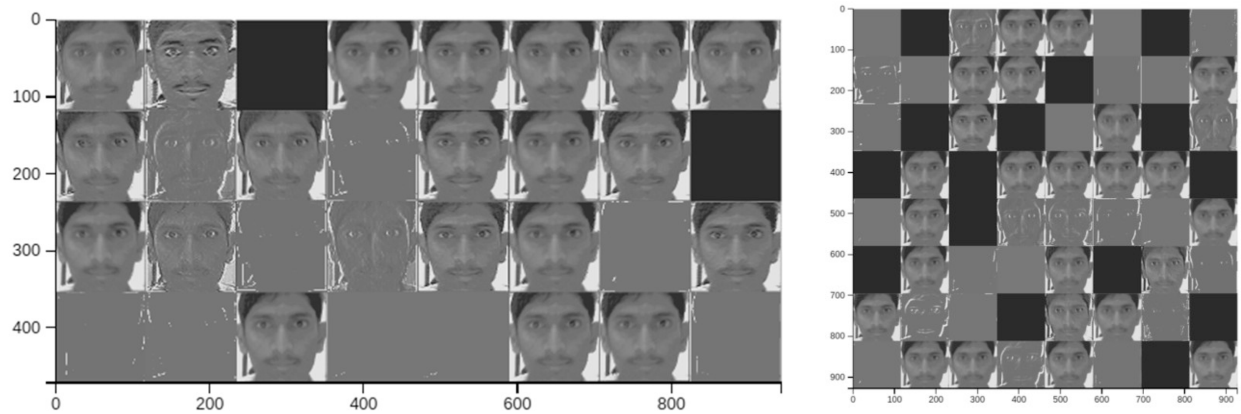
The output from POOL layer is fed to DROPOUT layer. An illustration of dropout layer output for three different dropout rates is given in Fig. 6. It is observed during evaluation that a drop rate of 0.5 yielded maximum accuracy for proposed application and hence the same is employed in this work. The output from subsequent stages of CONV+RELU, POOL and DROPOUT is shown in Fig. 7. It is observed from DROPOUT layer output that there is not much information left and hence additional stages of CONV+RELU, POOL and DROPOUT are not included. The output is then flattened and fed to DENSE/FC layer for classification. The size of final DENSE layer in Fig. 2 is 5×1, because the proposed real-time system is currently designed to classify faces of five individuals, whereas the system designed with AT&T datasets has a DENSE layer of size 40×1 to classify faces of 40 individuals.

Fig. 2. CNN Architecture for proposed face recognition system



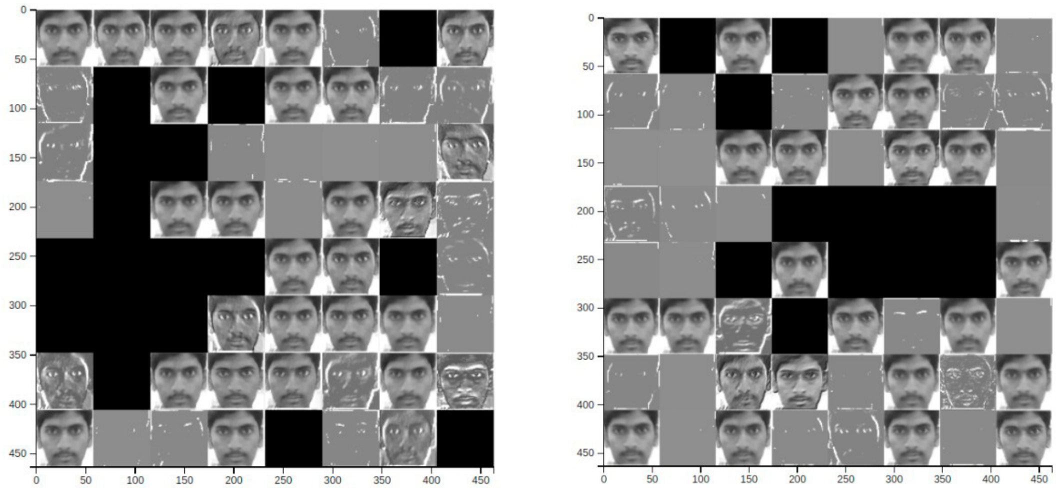
Randomly Initialized Weights Updated Weights after pre-defined epochs

Fig. 3. Filter weights of the 32 3×3 filters in Convolution Layer 1



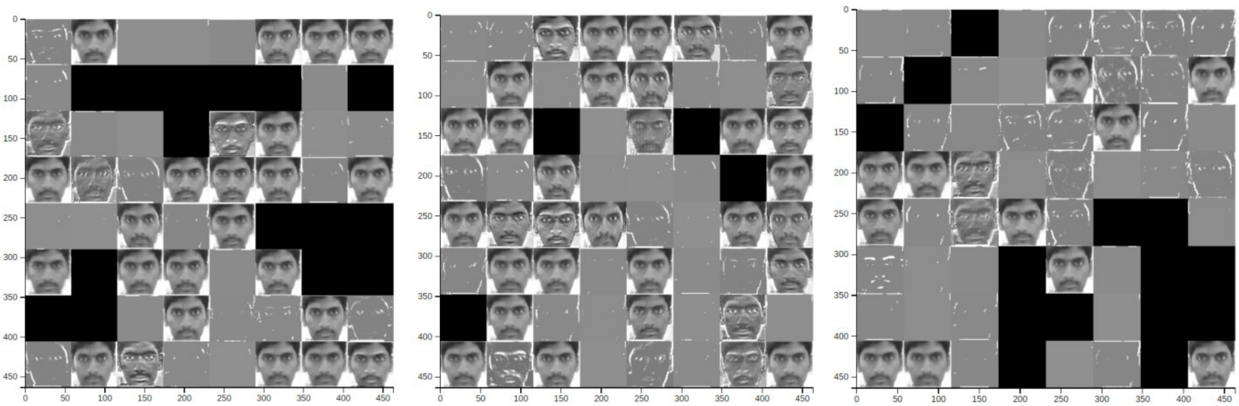
First CONV + RELU layer output Second CONV + RELU layer output

Fig. 4. Output of CONV + RELU layer with 32 filters and window size 3×3



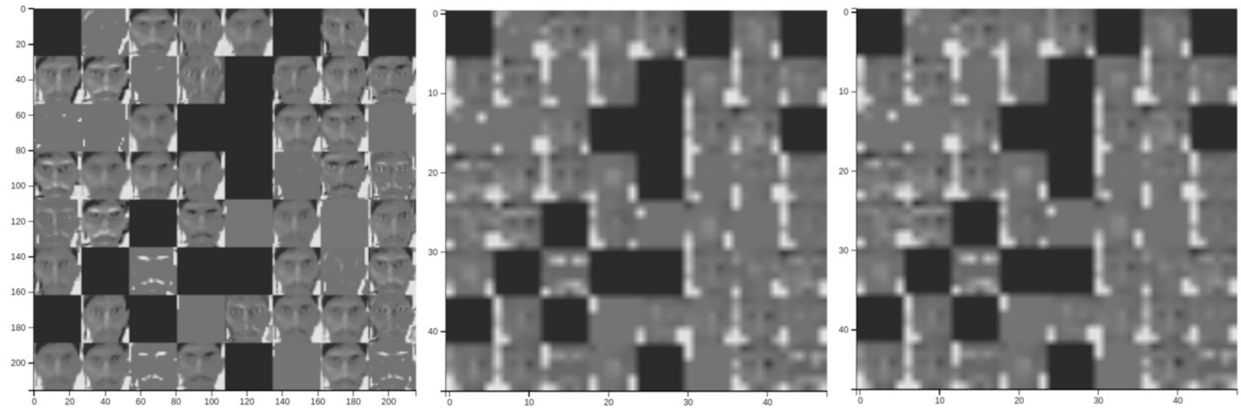
Using max pooling Using Average pooling

Fig. 5. An illustration of CNN Architecture Output from POOL layer



Dropout Rate = 0.1 Dropout Rate = 0.5 Dropout Rate = 0.8

Fig. 6. An illustration of CNN Architecture Output from DROPOUT layer



Output of CONV+RELU layer | Output of POOL layer | Output of DROPOUT layer

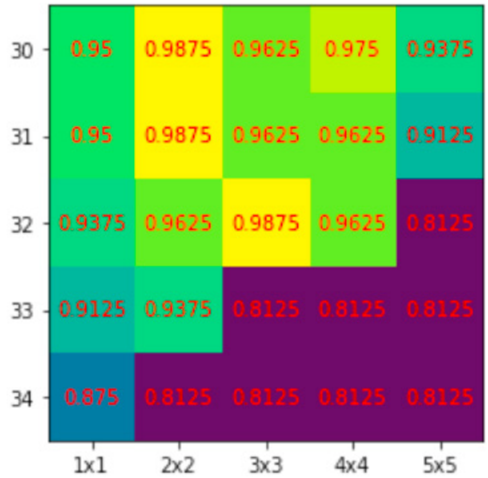
Fig. 7. CNN Architecture Output of Second Stage

3. Experimental Results

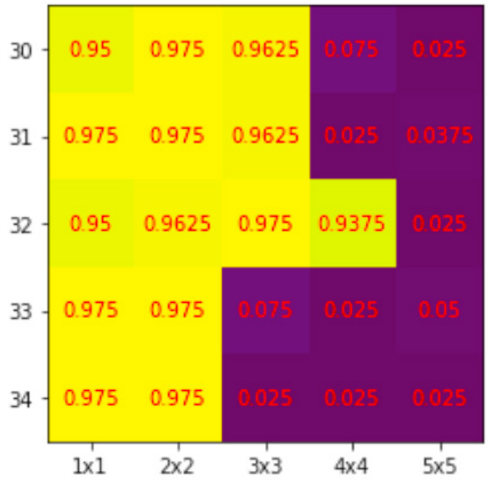
The performance evaluation of proposed face recognition system was initially carried out using standard AT&T database [9], comprising of 10 images from 40 individuals leading to a total of 400 images. The samples of 40 individuals from AT&T database are shown in Fig. 8. Out of 400 images, 320 images (8 images from 40 individuals) were used for training and remaining 80 images for testing.

Fig. 8. Samples from AT&T database employed for the evaluation of proposed system

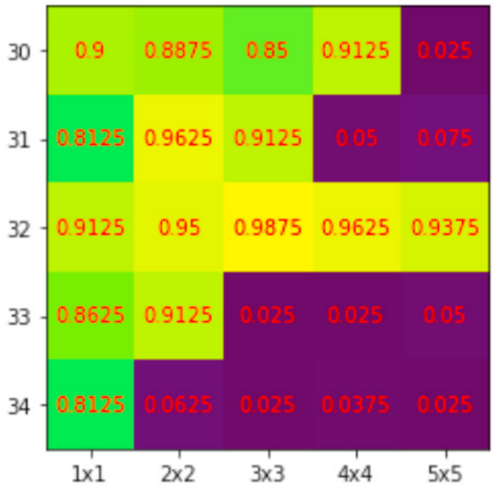
The performance evaluation of proposed system is carried out by varying the number of filters in convolution layer and the window size of convolution filter for different pooling window sizes. The results of this evaluation along with the recognition accuracy of the system are plotted in Fig. 9 with x−axis representing the window size for convolution filter and y−axis representing the number of filters in convolution layer. It is observed from Fig. 9 that convolution filter of size 3×3 pixels with 32 filters yielded a maximum recognition accuracy of 98.75% for proposed system on using a pooling window size of 2×2 and 4×4 pixels. The performance evaluation of proposed work is compared with the results reported in literature on using the same dataset for face recognition in Table 1. It is observed that the proposed approach and CNN architecture can be considered in par with the work reported in literature. The enhancement in recognition accuracy of proposed work is obtained by optimizing the number of convolution filters, window size for convolution filter and pooling.



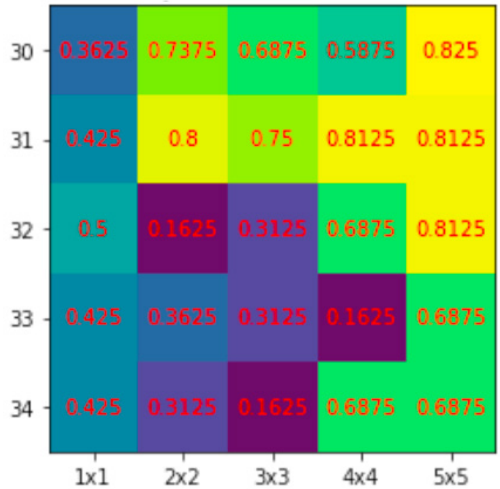
2×2 pooling window



3×3 pooling window



4×4 pooling window



5×5 pooling window

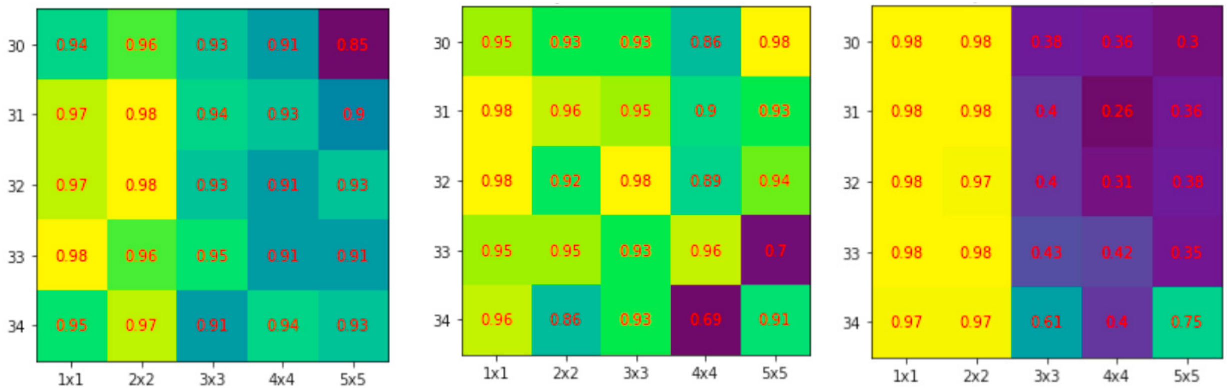
Fig. 9. Recognition Accuracy of proposed CNN Architecture for different combinations (AT & T Dataset)

Table 1. Comparison of Face recognition results reported in literature for AT&T datasets

Reference	Method	Recognition Accuracy
[6]	HOG and RVM	97.00%
[10]	PCA and MDC	96.70%
[11]	PCA and MDC	95.63%
[12]	PCA and KNN	92.00%
[13]	ICA and MDC	85.00%
[14]	Markov Random Fields	86.95%
[15]	Eigen face	92.87%
[16]	Statistical	92.60%
[17]	CNN	98.30%
[18]	CNN	95.00%
Prop. Work	CNN	98.75%

After the successful evaluation and testing of proposed system using standard AT&T dataset, the performance of proposed system is evaluated for real-time inputs through camera. The samples of 5 individuals including authors and family members are considered for evaluation of proposed real-time system are shown in Fig. 10. 40 images of each individual are captured totalling to 200 images. In order to find the recognition accuracy of proposed system for real-time input, out of 200 images, 100 images (20 images from 5 individuals) were used for training and remaining 100 images for testing. Experiments were carried out for real-time system also to identify the optimum number of convolution filters and window size of filters for convolution and pooling layer. The results of the evaluation are shown in Fig. 11 with x−axis representing the window size for convolution filter and y−axis representing the number of filters in convolution layer. It is observed from Fig. 11 that a maximum recognition accuracy of 98.00% was obtained for real−time system on using 32 convolution filters with pooling window size of 2×2, 3×3 and 4×4 pixels and different window size of convolution filter. The snapshot of output results obtained during the live demonstration of proposed real-time face recognition system are shown in Fig 12, with author images as inputs and their identities displayed on the top left by proposed system. It may be observed from Fig. 12 that the proposed system first detects a face in the image and once detected, it recognizes the face and displays the identity of the person.

Fig. 10. Samples employed for the evaluation of proposed real-time face recognition system



2×2 pooling window 3×3 pooling window 4×4 pooling window

Fig. 11. Recognition Accuracy of proposed CNN Architecture for different combinations (Real Time Dataset)

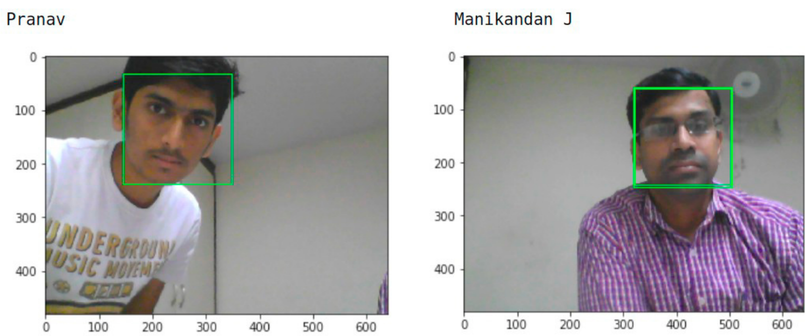


Fig. 12. Snapshot of results obtained during evaluation of the proposed real-time face recognition system

Details about the execution time of proposed system running on a Lenovo make laptop with Intel I5-7200 processor operating at 2.5 GHz with NVIDIA Geforce 940MX Graphics Processor and 8GB RAM is listed in Table 2. It is observed from Table 2 that 74% of time is consumed in capturing image, followed by 25% for pre-processing (includes face detection, RGB to gray scale conversion and resizing) and only 1% for CNN classification by the proposed real-time face recognition system. It is also observed from Table 2 that 80.65% of time is consumed in reading an image from a file and only 19.35% of time is consumed for CNN classification on using AT&T datasets. It may be noted that pre-processing isn't required for AT&T dataset, as all images are in gray scale, equally sized and have only faces.

Table 2. Execution time of proposed system

Real-time System			Using AT&T dataset		
Function	Execution Time (sec.)	% of Time	Function	Execution Time (sec.)	% of Time
Capturing Image	1.0120	74.06%	Reading from file	0.0200	80.65%
Pre-processing	0.3408	24.94%	Pre-processing	0.0000	0.00%
CNN Classification	0.0136	1.00%	CNN Classification	0.0048	19.35%
Total	1.3664	100.0%	**Total**	0.0248	100.0%

4. Conclusion

In this paper, design and evaluation of a real-time face recognition system using Convolutional Neural Networks is proposed. The performance of proposed system and CNN architecture is evaluated by tuning various parameters of CNN to enhance the recognition accuracy of the system designed. Maximum recognition accuracy of 98.75% and 98.00% is obtained from the proposed system on using AT&T and real-time inputs respectively. The proposed work can be easily adapted for various consumer applications such as face detection based home automation, device control, attendance system, intruder detection etc.

References

[1] R Samet, and M Tanriverdi. (2017) "Face Recognition based Mobile Automatic Classroom Attendance Management System." *International Conference on Cyberworlds*, Chester, United Kingdom, 20−22 September, IEEE Computer Society, pp. 253−256.

[2] Fahad P, Md. Mahmudul, Md. Atiqur, Susan M, Moslehuddin M, and Pandian V. (2017) "Face recognition based real time system for surveillance." *Intelligent Decision Technologies,* IOS Press, **11 (2017):** 79−92.

[3] Ouanan H, Ouanan M, and Aksasse B. (2018) "Pubface: Celebrity face identification based on deep learning." *IOP Conference Series: Materials Science and Engineering,* IOP Publishing Ltd., **353 (1):** 1−6.

[4] Fei Z, and N de With. (2005) "Real-time face recognition for smart home applications." *International Conference on Consumer Electronics*, Las Vegas, USA, 8−12 January, IEEE Press, pp. 35−36.

[5] D Cherifi, R Kaddari, H Zair, and A Nait Ali. (2019) "Infrared Face Recognition Using Neural Networks and HOG-SVM." *Third International Conference on Bio-engineering for Smart Technologies*, Paris, France, 24−26 April, IEEE Press, pp. 1−5.

[6] Karthik HS, and Manikandan J. (2017) "Evaluation of relevance vector machine classifier for a real-time face recognition system." *International Conference on Consumer Electronics – Asia*, Bangalore, India, 5−7 October, IEEE Press, pp. 26−30.

[7] Faruqe M, and M Hasan. (2009) "Face recognition using PCA and SVM." *Third International Conference on Anti-counterfeiting, Security, and Identification in Communication*, Hong Kong, 20−22 August, IEEE Press, pp. 97−101.

[8] Tolba AS, El-Baz AH and El-Harby AA (2008) "Face Recognition : A Literature Review." *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, **2(7):**2556-2571.

[9] AT&T Database of Faces. (2002) AT&T Laboratories Cambridge. https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[10] Soumen B, and Goutam S. (2011) "An efficient face recognition approach using PCA and Minimum Distance Classifier." *International Conference on Image Information Processing,* Shimla, India, 3−5 November, IEEE Press, pp. 1−6.

[11] Shalmoly M, and Soumen B. (2016) "Face Recognition using PCA and Minimum Distance Classifier." *Fifth International Conference on Frontiers in Intelligent Computing: Theory and Applications*, 16−17 September, Bhubaneswar, India, Springer, pp. 397−405.

[12] Kukreja S, and Rekha G. (2011) "Comparative study of different face recognition techniques." *International Conference on Computational Intelligence and Communication Networks*, 7−9 October, Gwalior, India, pp. 271−273.

[13] Yang J, Zhang D, Frangi F, and Yang Jing. (2004) "Two Dimensional PCA: A New approach to appearance based Face representation and Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26(1):** 131−137.

[14] Huang R, Pavlovic V, and Metaxas D. (2004) "A hybrid face recognition method using Markov Random Fields." *International Conference on Pattern Recognition*, 26 August, Cambridge, UK, pp. 157−160.

[15] Yan Ma, and ShunBao Li. (2008) "The modified Eigenface Method using Two thresholds." *International Journal of Computer and Information Engineering*, **2(9):** 3233−3236.

[16] MA Rabbani, and Chellappan C. (2007) "A different approach to appearance based statistical method for face recognition using median." *International Journal of Computer Science and Network Security*, **7(4):** 262−267.
[17] PatriK K, Miroslav B, Tomas M, and Roman R. (2017) "A New Method for face recognition using Convolutional Neural Network." *Digital Image Processing and Computer Graphics*, **15(4):** 663−672.
[18] Hu H, Shah A, Bennamoun M, and M Molton. (2017) "2D and 3D face recognition using convolutional neural network," *IEEE Region 10 Conference*, 5−8 November, Penang, Malaysia, pp. 133−132.