Second International Conference on Smart Systems and Inventive Technology (ICSSIT 2019)
IEEE Xplore Part Number: CFP19P17-ART; ISBN:978-1-7281-2119-2

Face Recognition with Convolutional Neural Network and Transfer Learning

R.Meena Prakash[1], N.Thenmoezhi[2], M.Gayathri[3]
Department of Electronics and Communication Engineering, AAA College of Engineering and Technology, Amathur, India

***Abstract*— Face recognition finds its major applications in biometrical security, health care and marketing. Among many biometrics like finger print, iris, voice, hand geometry, signature etc., face recognition has attained the most popular research orientation because of high recognition rate, uniqueness and large number of features. Some of the challenges in face recognition task include the differences in illumination, poses, expressions and background. Now-a-days deep learning methods are widely employed and they ensure promising results for image recognition and classification problems. Convolutional Neural Networks automatically learn features at multiple levels of abstraction through back propagation with convolution layers, pooling layers and fully connected layers. In this work, an automated face recognition method using Convolutional Neural Network (CNN) with transfer learning approach is proposed. The CNN with weights learned from pre-trained model VGG-16 on huge ImageNet database is used to train the images from the face database. The extracted features are fed as input to the Fully connected layer and softmax activation for classification. Two publicly available databases of face images – Yale and AT&T are used to test the performance of the proposed method. Experimental results verify that the method gives better recognition results compared to other methods.**

Keywords— ***Face recognition, Convolutional Neural Network, transfer learning, classification***

I. INTRODUCTION

Face recognition is one of the active research areas which has its extensive range of applications such as identity management and security. The face recognition system is currently employed for numerous applications. In healthcare technology, emotion recognition facilitates automated diagnosis of diseases such as Down syndrome. It is also being employed by the advertising media to understand the customer's facial expressions. Face recognition poses certain challenges due to conditions such as poor illumination, differences in poses and background, occlusion etc., The design of face recognition systems focuses on improving the speed and accuracy. The face recognition task in general consists of the following steps – pre-processing of the input face images, followed by extraction of features and then classification. The pre-processing step is used to remove any noise content, background illumination, etc., and also normalization. From the pre-processed face images, features are extracted for which different methods are adapted including Histogram of Oriented Gradients (HOG), Principal Component Analysis (PCA), Local Binary Pattern (LBP), Scale Invariant Feature Transform (SIFT), Linear Discriminant Analysis (LDA), deep learning and others [13,14,15]. Classification is done using classifiers such as Support Vector Machine (SVM), Neural Networks, Radial Basis Function (RBF) network, Convolutional Neural Network (CNN) etc., The deep learning features give better classification results than the hand crafted features.

Nowadays CNN is used for major classification problems in which the features are automatically learnt from low level to high level on successive increasing layers of the network. Transfer learning is the process of transfer of knowledge gained in solving one problem to solve another related problem. Transfer learning is advantageous when there is lack of sufficient training data and also it reduces computational complexity. It offers good classification accuracy on smaller database. In the proposed work, VGG16 architecture which is pre-trained on huge ImageNet database with more than 1 million images belonging to 1000 different categories is used to train the input face images. Then, classification is done using Fully Connected Layer and Softmax activation. The contribution of the article is implementation of an automated face recognition method using CNN with transfer learning. The rest of the paper is organized as follows. Review of Literature is given in Section II. The proposed method is explained in Section III. Experiments are results are discussed in Section IV with a conclusion part in Section V.

II. LITERATURE SURVEY

An age - invariant face recognition method using discriminative model with deep feature training is proposed [1]. In this work, AlexNet is used as the transfer learning CNN model to learn high level deep features. These features are then encoded using a code book into a code word with higher dimension for image representation. The encoding framework ensures similar code word for the same person's face images photographed during different time scale. Linear regression based classifier is used for face recognition and the method is tested on three datasets including FGNET which are publicly available. A face recognition system which incorporates the Convolutional neural network, auto encoder and denoising is proposed which is called Deep Stacked Denoising Sparse Autoencoders (DS-DSA). Multiclass Support Vector Machine (SVM) and Softmax classifiers are used for the classification process. The method is tested on four publicly available datasets including ORL, Yale and Pubfig. Cross-modality face recognition using deep local descriptor learning frame work is proposed in which both the compact local information and discriminant features are learnt from raw facial patches directly [3]. Deep local descriptors are extracted with Convolutional Neural Network (CNN). The method is tested on six extensively used face recognition datasets of different modes.

Deep learning with transfer learning CNN architecture is proposed which consists of three modules: base convolution module, transfer module and linear modules [4]. The method is tested on SCFace, ChokePoint, FSV and TIP datasets. A face recognition system based on individual component is proposed in which knowledge is gathered from whole images of face and transfer learning is used to classify the face components using a CNN architecture [5]. The method is tested to classify face images from Computer Vision Research Projects Faces94 database. A face recognition method consisting of the steps of face enhancement, extraction of features and finally classification into one of the known faces is proposed in which Scale Invariant Feature Transform (SIFT) descriptors are calculated from detected face regions' patches [6]. Multiclass SVM is used as the classifier and the method is tested on popular face databases. An automated face recognition method based on CNN architecture with pre-trained VGG model for transfer learning is proposed [7]. The cosine similarity and the linear SVM are used for classification and the method is tested on two public datasets.

Livestock identification method using CNN is proposed in which pre-trained CNN model is trained using artificially augmented dataset [8]. Face recognition using a recurrent regression neural network (RRNN) framework is proposed and is applied to capture video frames and still images [9]. For still images, sequential poses of the images are predicted from one static image and for face recognition in videos, one entire video sequence is used as input. A shallow CNN with cascade classifier is proposed for face recognition [10]. The method is tested on publicly available datasets –Wild dataset and the AT&T database. A face recognition method which employs an architecture consisting of the stages of image enhancement with noise removal, histogram equalization, face region detection, segmentation of faces, feature extraction and finally classification is proposed [11]. A face benchmark dataset with ID photos of 100 celebrities is also constructed and used for training and testing.

CNN architecture with two different loss functions based on Compact Discriminative loss is proposed for face recognition and three CNN's –CNN-M, LeNet, and ResNet-50 are employed to test the method [12]. The implementation of facial recognition framework in smart glasses is proposed which employs CNN with transfer learning [16]. A survey on face recognition methods, face emotion recognition and gender classification techniques is done in which the performance of different methods such as CNN, transfer leaning and traditional methods of HOG, Random Decision Forests is compared [17]. Based on the literature survey, it is inferred that CNN gives promising results for classification problems and can be used for face recognition application. To reduce the computational complexity, transfer learning can be employed in which knowledge gained by the pre-trained models on huge image database can be transferred to learn on new database of relatively smaller size. In this article, VGG16 architecture is used for transfer learning and to train on the face images from the database. The extracted features are given to the Fully Connected Layer for classification followed by face recognition in the testing phase.

III. PROPOSED METHOD

A. Convolutional Neural Network (CNN) and Transfer Learning

Deep learning is one of the areas of machine learning which learns representations from data with emphasis on learning successive layers of increasingly meaningful representations. The number of layers which contribute to the information is called the depth of the model. The neural network basically comprises of three layers - input layer, multiple hidden layers and output layer. The number of layers which contribute to the data is called the depth of the model. Convolutional Neural Networks are neural networks that employ convolution operation in place of matrix multiplication in the convolutional layers. The convolution operation is denoted as given in equation (1)

$$s(t) = (x * w)(t) \tag{1}$$

In equation (1), x is the input, w is filter kernel and s is the output called feature map.

CNN consists of three stages. In the first stage, multiple convolutions are performed using filters simultaneously to produce a set of linear activations. In the second stage, a non-linear activation, for example, rectified linear activation function is performed which will directly output the input if positive and zero otherwise. In the third stage, pooling operation is used to reduce size of the feature map. Some of the pooling functions are max pooling which finds and replaces with the maximum value in a rectangular window, average pooling which replaces with the average of all the values in the rectangular window, weighted average pooling and others. At the end of CNN, the output of last pooling layer is given as input to the fully connected (FC) layer. Every node in FC layer is connected to all nodes in the preceding layer. The FC layer performs classification into different classes.

The CNNs are trained through a process called back propagation which consists of 4 stages – forward pass, loss function, backward pass and weight update. The filter weights are initialized randomly. During the forward pass, the training images are passed to the network. The error rate is calculated using loss function which compares the network output with the desired output. Then back pass and weight update stages take place based on the error rate. The back propagation process is repeated for several iterations till convergence occurs.

Transfer learning is a machine learning method where the knowledge gained from a particular task is transferred to improve the process of learning in another related task. The CNN architectures like VGG, ResNet, AlexNet etc., are already trained on the huge image database of ImageNet consisting of more than 1 million labelled high resolution images belonging to 1000 classes. In this way, the knowledge gained already from one task is transferred to learn new task. Transfer learning is especially used where there is lack of sufficient training data. It shows good performance in classification and also the computational complexity is greatly reduced as the process need not start from the scratch.

B. CNN Architecture

In the proposed method, VGG16 CNN architecture is used for transfer learning and is trained on the input images from the faces dataset. The input images to the CNN are of size 224x224x3. The input images are passed through stack of 13 convolutional layers with filter size 3x3 and different depths of 64,528,256 and 512. Following the convolutional layers, three fully connected layers (FC) are used. The first two FC layers have 4096 units and the last FC layer performs classification into 1000 classes. All the hidden layers have non-linear rectification (ReLU). Softmax layer is used as the last layer of the CNN which performs activation function. Including the input and output layer, there are totally 41 layers in the architecture. The architecture of VGG16 CNN model is shown in Figure 1.

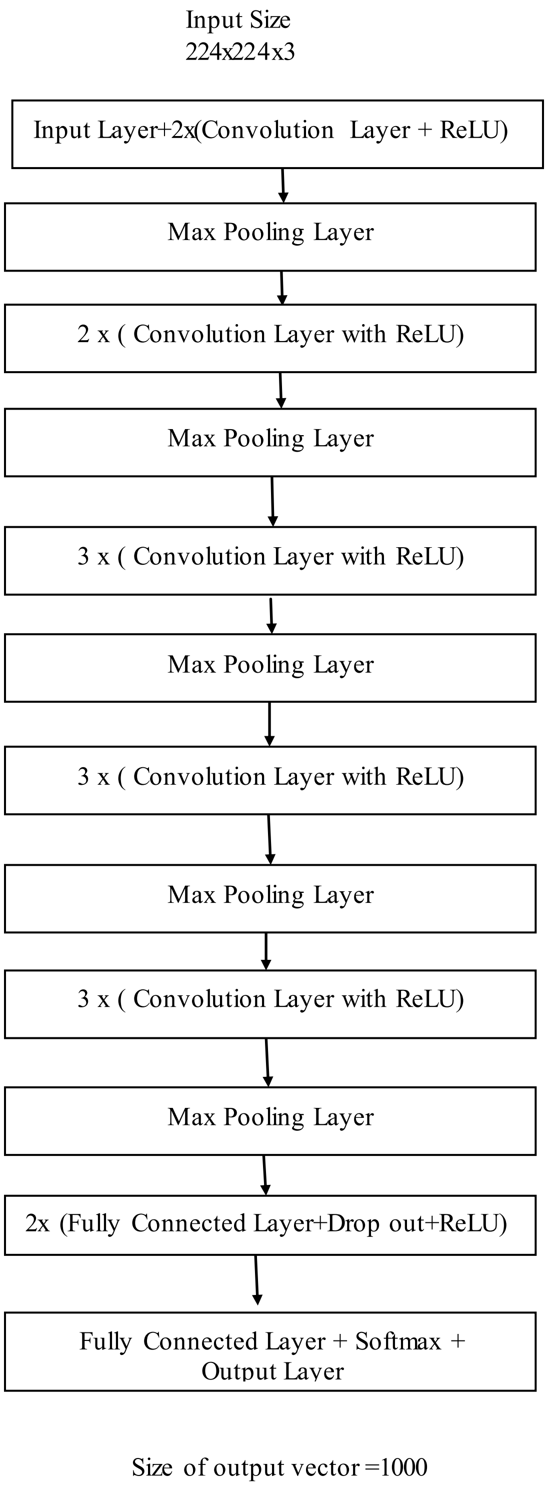

Fig. 1 VGG16 Architecture

The VGG16 was already trained on huge ImageNet database and the weights were learnt. These weights are used to initialize the model and trained on the input faces images training dataset. The extracted features are fed as input to the FC layer followed by softmax activation and classified into different faces.

IV. EXPERIMENTS AND RESULTS

A. Description of database

The performance of the method is tested on two popular publicly available face databases – Yale and AT&T. The Yale face database contains 165 gray scale images, where there are faces of 15 individual persons with different emotions. There are 11 images per subject with different facial expressions. The size of each image is 480x640. In the AT&T database, there are 10 different face images of 40 individual persons. The images have different facial expressions and were taken at different times. The images are of size 92x112. The dataset is formed with images for training and images for testing. The description of the dataset is given in Table 1.

TABLE 1
DESCRIPTION OF DATASET

S.No	Dataset	Total Images	Training Images	Testing Images	No. of classes
1	Yale	165	90	75	15
2	AT & T	400	240	160	40

B. Experimental Results

The sample images from AT&T and yale datasets are given in Figures 2 and 3 respectively. The proposed method is implemented with Keras using python. The number of epochs is set as 10; batch size is selected as 2. The learning rate of the network is set at 0.00001. The gradient descent optimization algorithm is used. The results obtained are given in Table 2 and are compared against results obtained with face recognition using Principal Component Analysis. The screen capture of the results obtained using training phase is given in Figure 4.

The experimental results for PCA and the proposed method using CNN with transfer learning in terms of classification accuracy are given in Table 2. It is inferred from the results that the proposed method of face recognition using CNN with transfer learning achieves better classification accuracy compared to the method using Principal Component Analysis. It achieves 100% accuracy for AT&T database face images and 98.7% accuracy for face images from Yale dataset.

Fig. 2 Sample Images from AT & T database

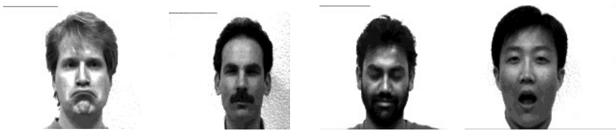

Fig. 3 Sample Images from Yale database

```
   ...: print("--- %s seconds ---" % (time.time() - start_time))
Train on 270 samples, validate on 75 samples
Epoch 1/20
270/270 [==============================] - 19s 71ms/step - loss: 0.0848 - acc: 0.9741
val_loss: 0.0577 - val_acc: 0.9733
Epoch 2/20
270/270 [==============================] - 21s 77ms/step - loss: 0.0926 - acc: 0.9630
val_loss: 0.0347 - val_acc: 0.9867
Epoch 3/20
270/270 [==============================] - 19s 71ms/step - loss: 0.0562 - acc: 0.9889
val_loss: 0.0400 - val_acc: 0.9733
Epoch 4/20
270/270 [==============================] - 17s 62ms/step - loss: 0.0792 - acc: 0.9741
val_loss: 0.0301 - val_acc: 1.0000
Epoch 5/20
270/270 [==============================] - 17s 61ms/step - loss: 0.0497 - acc: 0.9889
val_loss: 0.0210 - val_acc: 1.0000
Epoch 6/20
270/270 [==============================] - 18s 65ms/step - loss: 0.0504 - acc: 0.9852
val_loss: 0.0228 - val_acc: 0.9867
```

Figure 4 Screen capture of results obtained using training phase

TABLE 2
CLASSIFICATION RESULTS IN TERMS OF ACCURACY

S.No	Method	Dataset	Accuracy (%)
1	PCA	Yale	82.0
		AT & T	96.5
2	Proposed Method using CNN	Yale	98.7
		AT & T	100.0

V. CONCLUSION

An automated face recognition method has been proposed in this paper. The pre-trained CNN model VGG16 which was trained on huge database of ImageNet is used to initialize the weights and the model is trained on the input face images dataset. The features are extracted during the training face and fed to the fully connected layer with softmax activation function for classification. The method is tested on two publicly available face datasets of Yale and AT&T. Face recognition accuracy of 100% is achieved for AT&T database face images and 96.5% for Yale database face images. The experimental results show that face recognition using CNN with transfer learning gives better classification accuracy in comparison with other methods.

REFERENCES

[1] Shakeel, M. S., & Lam, K. (2019). Deep-feature encoding-based discriminative model for age-invariant face recognition. *Pattern Recognition, 93*, 442-457.

[2] Görgel, P., & Simsek, A. (2019). Face recognition via Deep Stacked Denoising Sparse Autoencoders (DSDSA). *Applied Mathematics and Computation, 355*, 325-342.

[3] Peng, C., Wang, N., Li, J., & Gao, X. (2019). DLFace: Deep local descriptor for cross-modality face recognition. *Pattern Recognition, 90*, 161-171.

[4] Banerjee, S., & Das, S. (2018). Mutual variation of information on transfer-CNN for face recognition with degraded probe samples. *Neurocomputing, 310*, 299-315.

[5] Kute, R. S., Vyas, V., & Anuse, A. (2019). Component-based face recognition under transfer learning for forensic applications. *Information Sciences, 476*, 176-191.

[6] Umer, S., Dhara, B. C., & Chanda, B. (2019). Face recognition using fusion of feature learning techniques. *Measurement, 146*, 43-54.

[7] Elmahmudi, A., & Ugail, H. (2019). Deep face recognition using imperfect facial data. *Future Generation Computer Systems, 99*, 213-225.

[8] Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M., & Grieve, B. (2018). Towards on-farm pig face recognition using convolutional neural networks. *Computers in Industry, 98*, 145-152.

[9] Li, Y., Zheng, W., Cui, Z., & Zhang, T. (2018). Face recognition based on recurrent regression neural network. *Neurocomputing, 297*, 50-58.

[10] Leng, B., Liu, Y., Yu, K., Xu, S., Yuan, Z., & Qin, J. (2016). Cascade shallow CNN structure for face verification and identification. *Neurocomputing, 215*, 232-240.

[11] Cui, D., Zhang, G., Hu, K., Han, W., & Huang, G. (2019). Face recognition using total loss function on face database with ID photos. *Optics & Laser Technology, 110*, 227-233.

[12] Zhang, M. M., Shang, K., & Wu, H. (2019). Deep compact discriminative representation for unconstrained face recognition. *Signal Processing: Image Communication, 75*, 118-127.

[13] A Survey on Face Recognition Techniques for Undersampled Data. (2017). *International Journal of Science and Research (IJSR), 6*(1), 1582-1585.

[14] Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. Determination Press.

[15] Xu, Y., Ma, H., Cao, L., Cao, H., Zhai, Y., Piuri, V., & Scotti, F. (2018). Robust Face Recognition Based on Convolutional Neural Network. *DEStech Transactions on Computer Science and Engineering,* (Icmsie)

[16] Khan, S., Javed, M. H., Ahmed, E., Shah, S. A. A., & Ali, S. U. (2019). Facial Recognition using Convolutional Neural Networks and Implementation on Smart Glasses. *2019 International Conference on Information Science and Communication Technology (ICISCT)*.

[17] Mane, S., & Shah, G. (2018). Facial Recognition, Expression Recognition, and Gender Identification. *Data Management, Analytics and Innovation Advances in Intelligent Systems and Computing*, 275–290.