JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DEEP NEURAL NETWORK FOR HUMAN FACE RECOGNITION

[1]D. GAYATHRY, [2] R. LATHA
[1]Research Scholar, [2] Professor & Head
Department of Computer Science
St. Peter's Institute of Higher Education and Research
Chennai - 600 054, Tamil Nadu, INDIA

Abstract

Face recognition (FR), the process of identifying people through facial images, has numerous practical applications in the area of biometrics, information security, access control, law enforcement, smart cards and surveillance system. Convolutional Neural Networks (CovNets), a type of deep networks has been proved to besuccessful for FR. For real-time systems, some preprocessing steps like sampling needs to be done before usingto CovNets. But then also complete images (all the pixel values) are passed as input to CovNets and all thesteps (feature selection, feature extraction, training) are performed by the network. This is the reason that implementing CovNets are sometimes complex and time consuming. CovNets are at the nascent stage and the accuracies obtained are very high, so they have a long way to go. The paper proposes a new way of using a deep neural network (another type of deep network) for face recognition. In this approach, instead of providing raw pixel values as input, only the extracted facial features are provided. This lowers the complexity of while providing the accuracy of 97.05% on Yale faces dataset.

Index Terms: Face recognition, haar cascade, deep neural networks, convolutional neural networks, softmax.

Introduction

Face recognition (FR) system identifies a face by matching it with the facial database. It has gained great progress in the recent years due to improvement in design and learning of features and face recognition models. As humans have an exceptional ability to recognize people irrespective of their age, lighting conditions andvarying expressions. The aim of researchers is to design an FR system which can match or even surpass the human recognition rate which is nearly 97.5%.

The techniques used in best facial recognition systems may depend on the application of system. Face recognition systems may be divided into two broad categories:

- Find a person from his image in a large database of facial images (eg. a police database). These systems returns the details of the person being searched for. Often only one image is available per person. It is usually not necessary for recognition to be done in real time.
- Identify a person in real time. These are used in systems which allow access to a certain group of people and deny access to others. Multiple images per person are often available for training and real time recognition is required. The proposed idea is for the second type of systems with varying facial details, expressions, and angles. It remains an open problem to find an ideal facial feature which is robust for FRin unconstrained environments.

The conventional face recognition pipeline consists of four stages: face detection, face alignment, face representation (or feature extraction), and classification . The proposed method extracts facial features from input images and feeds them to deep neural networks for training and classification (softmax layer is used). Thearchitecture of network is very flexible and layers can be added or removed to get best results. In recent times there are numerous libraries, functions and platforms to create and modify a network.

CovNets are a specialized kind of neural networks for processing data that has a known, grid-like topology. These networks have been tremendously successful in practical applications that include time-series data, which can be thought of as a 1D grid taking samples at regular time intervals, and image data, which can be thought of as a 2D grid of pixels. Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers. The name “convolutional neural network” indicates that the network employs a mathematical operation called convolution. Convolution is a specialized kind of linear operation .

1.1. Literature Work

Recently, multiple CovNets or deep CovNets have shown good results for face verification. According to Yi Sun et.al , existing methods generally address the problem of FR in two steps: feature extraction (design or learn features from each individual face image separately to acquire a better representation) and recognition (calculate similarity score between two compared faces using feature representation of each face . For face recognition (FR), many approaches have been implemented earlier, like the use of neural networks , geometrical features, Eigen faces, template matching, and graph matching. CovNets has shown many promising results for FR . Automatic feature extraction method using ratios of distances, presented by Kanade used geometrical features and reported a recognition rate between 45-75% with a database of 20 people.

The approaches like self-organizing maps (SOM) and Karhunen-Loeve (KL) transform both can be used for dimensionality reduction, from which SOM proved to be an efficient algorithm . Principal Component Analysis (PCA) has also been successfully implemented for same purpose Though CovNets have shown promising results for FR, it remains still ambiguous to design a good CovNet architecture for a specific classification task due to the lack of theoretical guidance. According to , CovNet-Restricted Boltzmann Machine (RBM) has shown 97.08% accuracy for matching two images of same person in unconstrained environment.

Brunelli and Poggio computed a set of geometrical features such as nose width and length, mouthposition and chin shape. They reported a recognition rate of 90% on a database of 47 people. However, they showed that a simple template matching scheme shows 100% recognition for the same database. Cox, *et.al* have introduced a mixture-distance technique which achieved a recognition rate of 95% using a query database of 95 images, where each face was represented by 30 manually extracted features.

By Pentland *et al.* good results are reported on a large database (95% recognition of 200 people out of 3000). It is difficult to draw broad conclusions as many images of the same people looked very similar . In Lades *et al.* presented a dynamic link architecture for distortion invariant object recognition which employs elastic graph matching to find the closest stored graph. Sparse graphs whose vertices are labeled witha multi-resolution description in terms of a local power spectrum, and whose edges are labeled with geometrical distances. They presented good results with a database of 87 people and test images composed of different expressions and faces turned 15^0. The matching process is computationally expensive, taking roughly 25s to compare with 87 stored objects when a parallel machine with 23 transputers is used. Thus, Eigen faces isa fast, simple, and practical algorithm. However, it may be limited because optimal performance requires a highdegree of correlation between the pixel intensities of the training and test images . Graph matching is another approach to face recognition.

Wikott *et al.* [12] used an updated version of the technique and compare 300 faces against 300 different facesof the same people taken from the Face Recognition Technology (FERET) database. They report a recognition rate of 97.3%.

In constrained environments, hand-crafted features such as Local Binary Patterns (LBP) and Local Phase Quantization (LPQ) have received respectable performance in FR. However, the performance degrades dramatically when applied on images taken in unconstrained environments such as varying facial alignment, expression and illumination.

High-level recognition is typically modeled with many stages of processing as in Marr paradigm of processing from images to surfaces to three-dimensional (3D) models to matched models [10]. However, Turk and Pentland [18] argue that there is also a recognition process based on two-dimensional (2D) image processing. They presented a face recognition scheme in which face images are projected onto the principal components of the original set of training images. The resulting Eigen faces are classified by comparison with known individuals .

None of the previous methods have used the idea of feeding only the extracted features into deep neural networks to accomplish the task of FR. The paper proposes the use of haar cascade (frontal face) for pre- processing the images which are then fed as input to deep neural networks for face recognition rather than directly passing the pixel values to CovNets.

2. Deep Neural Network

A Neural Network is human brain inspired algorithm designed to recognize pattern in numerical datasets.The real world data for example image, text audio, video etc; needs to be transformed into numerical vectors to use neural nets. A neural network is composed of different layers and a layer is made up of multiple nodes. Based on the type of pattern the neural network is trying to learn each input data fed into a node is assigned some weight. These weights determine the importance of the input data in producing the end result. The weighted sum of input data is calculated and depending on some threshold biases the output for the node is determined. The mapping of input to output is performed by some activation function.

The goal of a neural network is to approximate some function 'f'. Task of a simple classifier function y = f(x), is to map the input data x to a class y, while the neural network identifies the parameter β, that results in best approximation function, y = f(x,).

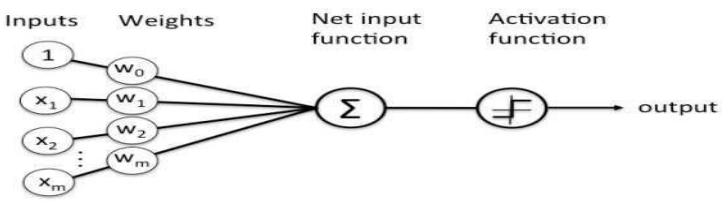

Fig.1. A neural node

A simple neural network is a network of such functions, that may be defined as $f(x) = f^3(f^2(f^1(x)))$. In thechain, f^1 is called the first layer, similarly f^2 is the second layer and so on. The length of this chain determines the depth of the neural network. Final layer is called the output layer. A schematic representation of a neural network is depicted in fig2. While training the desired output of each layer is not visible therefore the middle layers are called the hidden layer. A Deep Neural Network (DNN) is a feed forward Artificial Neural network (ANN), with multiple hidden layers and higher level of abstraction.

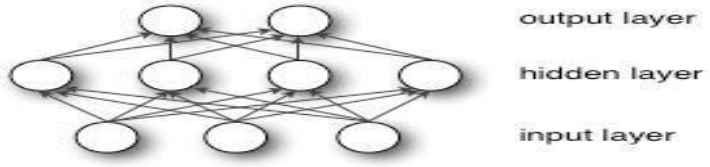

Fig.2. A Small Neural Network

The width of the DNN is determined from the dimensionality of the hidden layer. The hidden layer valuesare calculated through activation function. Learning in deep neural networks requires minimizing the cost function, like in case of classification cost function is the difference between actual label and the predictedlabel. Generally gradient descent is used for this purpose. In modern neural network, it is recommended to use Rectilinear Unit or relu as activation function. A single hidden unit $h^{(i)}$activation is given by

$$h^{(i)}=\sigma(w^{(i)T}x) \quad (1)$$

Where, () is the tanh function $^{(i)}$, with the weight vector for the i^{th} hidden unit, and x is the input. It givesa nonlinear transformation still it remains very close to the linearity making linear models to be easilyoptimized by Gradient Descent .

Generally limited data causes problem of over fitting in DNN. To avoid this dropout is used [6]. It randomly drops some nodes from the layers based on their probability. "dropping out" indicates temporarily removing units along with its incoming and outgoing edges. this is depicted in figure 3.

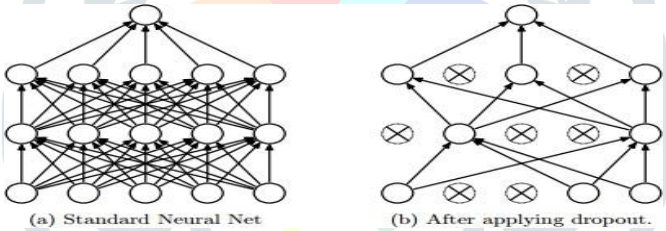

Fig.3. Dropout Neural Net Model

3. Proposed Work

The following subsections explore the components of the new system. The approach proposes the use of frontal face in haar cascade (defined in opencv) for preprocessing the input images and feeding only the facial features to network for learning and classification. The following steps are performed for FR on yalefaces dataset.

3.1. Preprocessing

Feature selection is the process of selecting the useful features and leaving the extra features. Feature extraction is the process of making combining more than one feature to a single feature. In the proposed approach uses frontal face, for both feature selection and extraction, which takes the image as input and returns only the values of key features of the face present in image. It first detects the face in the image (selection of features representing face in complete image) and then extract facial features (like making a set of features that denote eyes, nose, lips in the image), as shown in figure 4.

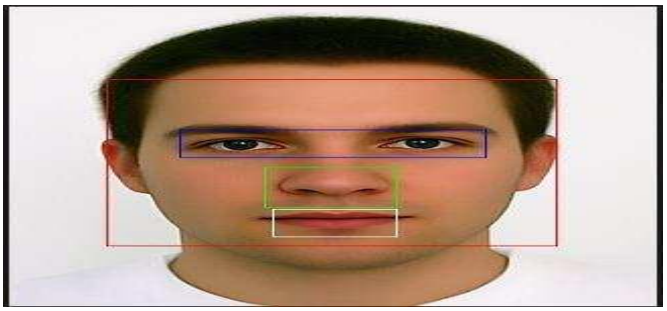

Fig.4. Facial Features in Image

3.2. Learning

A multi-layer feed-forward deep neural network is designed to learn the simplified features from theprevious step. Multiple sets of activation, dense (fully-connected) and dropout layers are added in order tomake learning efficient After each set the number of features reduces thus using only the important features for classification in the last layer. The architecture of the network is shown in figure 5. This network is trained in50 epoches and training and testing accuracies are plotted and shown in figure 6, and losses in figure 7. As itcan be seen, the accuracies are increased as the number of iterations proceeds. Further increasing the iterations or number of epoches to 75 or 100 results in same final accuracy, so 50 is the most efficient value.

3.3. Classification

The last layer of the network is softmax is used for classification. This is because the extracted facial featuresare simply stored for training images. Each authorized person will have a class, and the input image mustbelong to either one of the classes. If it matches to any of the images present in database, the system will allow the person to enter the secured place or to access private information.

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 512)	11520512
activation_1 (Activation)	(None, 512)	0
dropout_1 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 512)	262656
activation_2 (Activation)	(None, 512)	0
dropout_2 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 256)	131328
activation_3 (Activation)	(None, 256)	0
dropout_3 (Dropout)	(None, 256)	0
dense_4 (Dense)	(None, 15)	3855
activation_4 (Activation)	(None, 15)	0

Fig.5. Architecture of DNN used

3.4. Algorithm

The method for proposed FR system is defined as follows. The dataset used is Yalefaces A.

- Load the pixel values of all images from dataset
- Detect the facial features of all images using haar cascade
- Crop the face according to the output of previous step
- Split the data for cross validation in the ratio 9:1
- Design the following Neural Network

1. Model consists four layers of Neural Network
2. First layer is dense layer giving 512 outputs with relu activation and dropout of 0.2.
3. Second layer is dense layer giving 512 outputs with relu activation and dropout of 0.2.
4. Third layer is dense layer giving 256 outputs with relu activation and dropout of 0.2.
5. Fourth layer (output layer) is dense layer giving 15 outputs with softmax activation and dropout of 0.2.

- Train the Neural Network with epoch=50.
- Plot the train and test accuracies.
- Calculate the final average accuracy.

3.5. Implementation Details

The proposed method is tested on Yaleface database which consists of black and white images of 15samples, each having 11 images in different expressions making a total of 165 images. A sample is shown in figure 7. Images for one subject in different facial expression or configuration: center-light, with glasses,happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink. The categories are defined as subjects along with labels, so total classes are 15. The complete dataset is splitted in two parts: 148 images for training, and 17 for testing. Final average accuracy is calculated at softmax layer by checking the number of test samples which identified correctly. The idea is implemented on Python 3.5.3 (64 bit) system. Opencv package is used for pre-processing using frontal face feature of haar cascade. Creation and training of neural network is done using keras, theano, and tensorflow (packages available in python). Final average accuracy achieved in proposed system is 97.05%, which is close to human face recognition accuracy of nearly 97.5%. The comparative analysis from some of the previous FR systems is provided in tables.

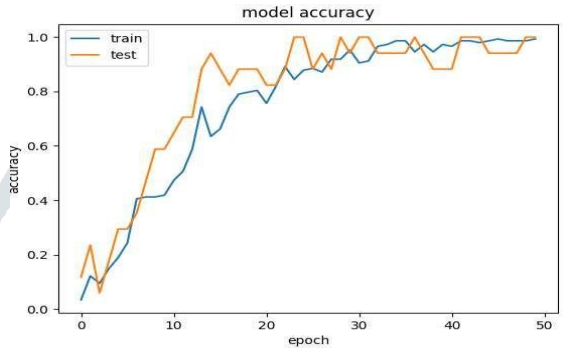

Fig.6. Graph Showing Testing and Training Accuracies

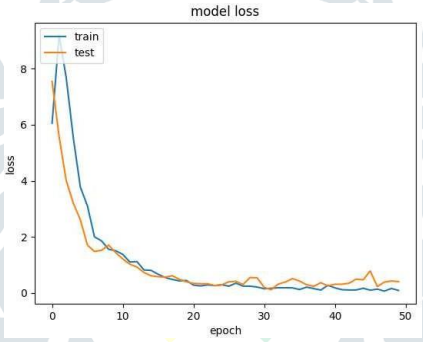

Fig.7. Losses

Fig.8. Yalefaces Dataset

Table 1. Accuracies of FR system

Method	Accuracy(%)
Geomerical features	90
Mixture-distance [15]	95
Eigenfaces [27]	95
PCA [26]	Lighting 96, orientation 85, scale variation 64
CV_DNN	97.05

Table 2. Details of above FR system

Method	No. of features	Dataset
Geometrical features	38,400	47 people
Mixture-distance	30 for each face	685 individuals
Eigenfaces	-	15 images for 200 people
PCA	-	16 subjects
CV_DNN	22,500	Yalefaces

4. Conclusion

The use of haar cascade for extracting facial features and feeding them instead of raw pixel values helps in decreasing the complexity of neural network based recognition framework as the number of redundant input features has been decreased. Also the use of DNN instead of CovNets makes the process lighter and faster. Also, the accuracy is not compromised in the proposed method as average accuracy obtained is 97.05%.Though one additional step of extraction of facial features from each image is added, still the process is better for small datasets.

References

[1] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Hybrid Deep Learning for Face Verification." IEEE Transactions on Pattern Analysis and Machine Intelligence 38.10 (2016): 1997-2009.

[2] Hu, Guosheng, et al. "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015.

[3] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT Press, 2016.

[4] Zhang, Tong, et al. "A deep neural network driven feature learning method for multi-view facialexpression recognition." IEEE Trans. Multimed 99 (2016): 1.

[5] Lawrence, Steve, et al. "Face recognition: A convolutional neural-network approach." IEEE transactionson neural networks 8.1 (1997): 98-113.

[6] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." Journal ofMachine Learning Research 15.1 (2014): 1929-1958.

[7] Kanade, Takeo. "Picture processing system by computer complex and recognition of human faces" Doctoral dissertation, Kyoto University 3952 (1973): 83-97.

[8] Brunelli, Roberto, and Tomaso Poggio. "Face recognition: Features versus templates." IEEEtransactions on pattern analysis and machine intelligence 15.10 (1993): 1042-1052.

[9] Cox, Ingemar J., Joumana Ghosn, and Peter N. Yianilos. "Feature-based face recognition using mixture-distance." Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on. IEEE, 1996.

[10] Ahonen, Timo, et al. "Recognition of blurred faces using local phase quantization." Pattern Recognition,2008. ICPR 2008. 19th International Conference on. IEEE, 2008.