Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 132 (2018) 2–10

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2018)

Gender Recognition Through Face Using Deep Learning

Amit Dhomne[1], Ranjit Kumar[2], Vijay Bhan[3]
Department of Computer Science and Engineering
National Institute of Technology, Rourkela, Odisha, India

Abstract

Automatic gender recognition has now pertinent to an extension of its usage in various software and hardware, particularly because of the growth of online social networking websites and social media. However the performance of already exist system with the physical world face pictures, images are somewhat not excellent, particularly in comparison with the result of task related to face recognition. Within this paper, we have explored that by doing learn and classification method and with the utilization of Deep Convolutional Neural Networks (D-CNN) technique, a satisfied growth in performance can be achieved on such gender classification tasks that is a reason why we decided to propose an efficient convolutional network VGGnet architecture which can be used in extreme case when the amount of training data used to learn D-CNN based on VGGNet architecture is limited. We examine our related work on the current unfiltered image of the face for gender recognition and display it to dramatics outplay current advance updated methods.

Keywords: Gender Recognition,VGGNet,D-CNN,Soft-max,MLP , Deep Learning;

1. Introduction

In this paper, we have used VGGNet architecture of Deep Convolution Neural Network (D-CNN) for gender recognition so in order to improvise the previously used method and to obtain much accurate result.figure:1 depicts how to process face image through D-CNN and find the pattern, extract feature to recognize gender from image accurately. The advantage of using D-CNN is it automatically extract the feature from an image and give output, we don't require to use feature descriptors like Histogram Oriented Gradient(HOG) and Support vector machine(SVM), eigenvector to extract the feature from image manually so as to do further recognition task or classification task. Gender Recognition was started with the problem in psychophysical studies to classify gender from human face; it concentrates on the e orts of perceiving human visual processing and recognizing relevant features that can be used to distinguish between female and male individuals. Exploration has proved that the discrepancy between a female face and male face can be used effectively to improvised the result of face recognition software in bio-metrics devices

10.1016/j.procs.2018.05.053

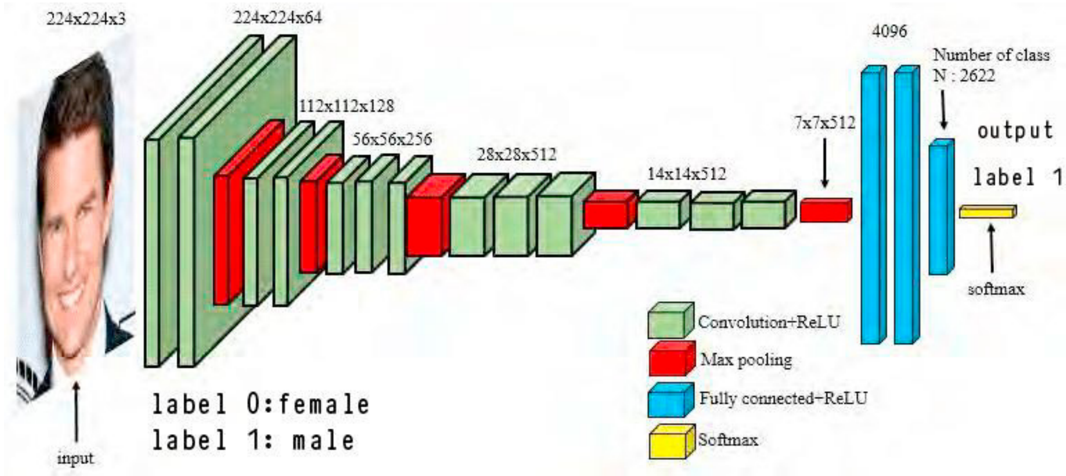

Fig. 1. Gender recognition

with following recognition of trait-like gender, age, human expression, facial disease etc. With this human-computer intercommunication, supervision, vigilance device, and digital vision system and much more that will work on whole human presence. Nevertheless, in a physical world scenario, the challenge is how to do work with the face image which influences various factors like illumination, pose facial expression, age estimation, occlusion, and background instruction data, and noise, error. It is a kind of motivation to do something new in the evolution of a boisterous face-based gen-der recognition application that has extreme detection accuracy. The Deep Conventional Neural Network(D-CNN) technique used in face recognition, involving face dependent gender recognition,age-based recognition, comprises the phases of accepting the image as input and then transforming input images for further processing, dimension reduction, feature extraction, feature procurement, and classification, in this sequence. Initial knowledge of these technique realms is needed to find out the finest extractor of feature for design. In extension to which, recognition method performance is highly vulnerable to the specified classifier used, which completely relies on the pattern retrieval technique applied to the method which we have used in the research work related to this paper. It is most di cult to find such a classifier that aggregate the finest among the chosen feature extractor so excellent recognition result can be obtained. The profound Deep Convolutions neural system (D-CNN) is a neural system varies with the number about convolutional layers utilized to be compatible with sub-sampling layers and end with you quit offering on that one or additional completely joined layers(Fully connected layer) in the calibre multilayer perceptron. A convincing gain of the D-CNN over another traditional method in feature recognition technique is its capability to simultaneously perform following tasks like features extraction, reducing data dimension, and classification in the particular organize network structure. this kind of model is described in Figure:2 can speed up recognition process and provide the result with high accuracy and minimum cost. The D-CNN performs both the work of feature classification and feature ex-traction and inside a single network structure through training a neural network on the collection of huge known data which is called training data Celebrity face dataset.

Feature selection is used in training procedure of neural network specially VGGNet by allowing a network to learn the weights responsible for extraction of features [2].The D-CNN has the capability of extraction of numerous different properties from a un-processed input image that requires either no or little pre-processing needed [2].

The D-CNN gives halfway resistance and boosts to geometric transformations and deformation and 2-dimensional changes in shapes and figure [2]. Hence, the D-CNN is specially made to overcome, lacking the other existing feature extractor that is described by having static behavior.The benefit obtains with the use of D-CNN is that they are comparatively easy to assist the network layer(input, hidden, output) in learning parameter, weight,(loss through BPNN).They have less number of parameters in comparison to fully connected multiple layer perceptron neural networks with the similar count of hidden layers used between input and output layer[3].Therefore the D-CNN has shown an excellent successful result in a huge range of applications such as tracking human in mob i.e human track--ing system(HTS),surveillance system that deals with object/article/human, traffc signal recognition(TSR),optical character recognition (OCR),face recognition(FR),and many others application of D-CNN and obviously computer vision numerous application. This paper mainly deals with an unaccustomed way for real-time gender recognition by using a D-CNN method. The fundamental worth of e ort carried by me ahead to this research are as follows.

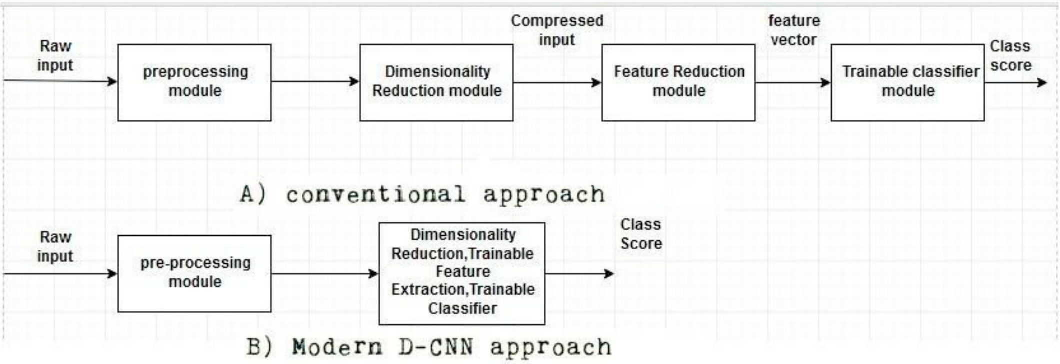

Fig. 2. Pattern Recognition Approach to CNN

First, a profound4-layer D-CNN for facial-based gender recognition is proposed. The architecture of the network is such that it involves the reduction in complexity of design with a very few count of layers, neurons, training parameters, and the connection between them in comparison to existing methods.
The erstwhile is a general model trained for viewable recognition of object, face etc., and whereas, the new method is a face based recognition system. These type of transferred system in contrast to state-of-the-art gender recognition system, which completely relies on a Deep Convolution Neural Network (D-CNN) models, utilized by other researcher and mathematician. Nevertheless, this model is learned particularly for gender recognition. The experiments have been conducted on the celebrity-face data-set and LFW face dataset which contains a number of face images. The beneficence of this research paper can be outlined as follows:
In the beginning, it is shown that general and domain-oriented Deep Convolution Neural Network (D-CNN) model is used for gender recognition tasks successfully. The second one is, by using accurate transfer learning method, an already trained Convolution neural network (CNN) model can have potential to show good result than the training of a new problem-related D-CNN model from the beginning. With this method, we have obtained up to 7% and 4.5% excellent improvement in performance in comparison to a problem-related method for gender recognition process. The third one is, transferring a domain-specific Deep Convolution neural network (D-CNN) method instead of generic Deep Convolution neural network(D-CNN) method has proved extremely useful, for achieving much better gender recognition results. If we see overall then, we have highlighted that when amount of data available is limited to performing a specific work or any experiment related to computer vision, in that case it is good to take advantage from a already trained Deep Convolution Neural Network (D-CNN) method and applied it to problem in hand, in spite of building and train the network from beginning a experiment-oriented Convolution Neural Network (D-CNN)model with relatively fewer layers.

2. Proposed Work

We are going to tune fine the already trained VGG-Face Deep Convolution Neural Network (CNN) method for the task of recognizing the gender of a human being from his/her face image. The VGGNet-face network has been trained to recognize more than 10000 celebrity IDs. Also, we can ignore/chop-o the classification hidden layer used in the network and perceive the output generated from the fully connected layer that is used as a representation for the input face. We are going to learn a new Softmax classification layer on top of these FC layer features and train it to classify the gender of the input face image into one of the two classes ”male” and ”female”.
The architecture is quite straightforward. We have one Linear(4096, 2) layer (FC-layer) which takes a 4096-d input and computes a vector that has two elements.Pre-trained VGG-Face descriptors are 4096-d in size. So, our FC-layer can take the pre-trained

descriptors as input and return a pair of values which would be the (un-normalized) likelihoods of the input image belonging to each of the two classes ”male” and ”female”. The FC-layer is followed by a Log SoftMax() layer which converts the un-normalized likelihoods returned by the FC-layer into log-probabilities which would then be sent to our loss module. By training the parameters of the FC-layer to minimize the loss, our network will learn to map the pre-trained face descriptors to the correct gender class. Architecture VGG-Face Deep Convolution Neural Network (CNN) model (PVZ15) is the similar to the VGGNet architecture(SZ15), consist of 16 number of layers and is build up for recognition and detection of object, articles, and difficulty mainly arises in training phase of convolution neural network based on VGGNet architecture network process. VGG-Face deep Convolution neural network (CNN) model (PVZ15) was trained on 2.6M of facial images out of 10000 identities of the complete facial image, whereas VGGNet Deep Convolution Neural Network (D-CNN) model, was framed upon ImageNet ILSVRC algorithm challenge by training of 1.3M images of 10000 objects of the class. In VGG-Face deep Convolution Neural Network (CNN) model (PVZ15), there are five convolutional layers. Out of two convolution layer, one layer is superseded by another one and whatever the output obtained from each layer is activated by an activation function named ReLU(Rectified Linear Unit).At the last output obtained from the previous layer, is required to be max-pooled again so as to do the reduction in the size of images. This process follows till the last layer of Deep Convolution Neural Network (D-CNN).Then every convolution layer in the block is come into action by an activation function ReLU(Rectified Linear Unit) and at last final output is required to be max-pooled. The last convolution layer is then immediately followed by back to back three fully connected layers (Fc6, Fc7, and Fc8).Among all fully connected layer, 6 and 7 that is having an output size of 4096 and dropout ratio of with respect to it is nearly 0.5. Rest of the fully connected layer number 8 is doing the task of classification and has an output size of 10000 which is equal to the number of identities defined in the training set. VGGNet-Face deep Convolution Neural Network model with a three channel input image, whose size is 224x224 pixels. Cropping and flipping are also applying on the number of training images for the convenience learning the neural network in order to perform a task. Here we are using 64 as batch size for each feed forward pass (p).

2.1. VGGNet Architecture Detail

The VGGNet architecture was popularized by Simonyan and Zisserman in their 2014 research paper [11]. VGGNet stands for Visual Geometric Group Convolution Neural Network (CNN). fig.3 shows the architecture of VGGNet CNN used in our work. This network is known by its simplicity, by using 33 convolution layer stacked on the top of each other on the order of increasing depth.By reducing the dimension size which is deal by max pooling technique. Then at last two fully-connected layers(FC) number 7 and 8 each having 4,096 nodes are followed by a function called softmax classifier. .

2.2 Deep convolutional neural networks

fig.4 ConvNet configurations(shown in columns).The depth of the configurations increases from the left (A) to the right (E), as more number of layers are added (the added layers are shown in bold). The convolutional layer parameters are denoted as conv (receptive field size)-(number of channels).The ReLU activation function is not shown for brevity [11].

Deep Convolutional Neural Networks(D-CNN) is popularized with the LeNet-5 architecture used for recognition and detection of an object. In contrast to modern Deep Convolutional Neural Networks (D-CNN) like VGGNet, LeNet-5 network architecture was comparatively moderate because of the limited number of computational resources such as time, memory, processing power, and the various algorithmic challenges for training the huge networks. Yet much potential exists in Deep Convolutional Neural Networks (D-CNN) architectures (a neural network with the huge number of neuron layers), recently they have become vogue, because of an unexpected rise in both the computational power with the usage of Graphical Processing Units(GPU), and the amount of dataset that is easily available on the Internet or prepares by a researcher in order to do practically. One of the biggest examples used in the physical world with the use of Deep Convolution Neural Network(D-CNN) is for image classification, Recognition on various facial database of million number of raw unfiltered face image like LFW, Celebrity face dataset,ImageNet,adience,3DMAD,AT

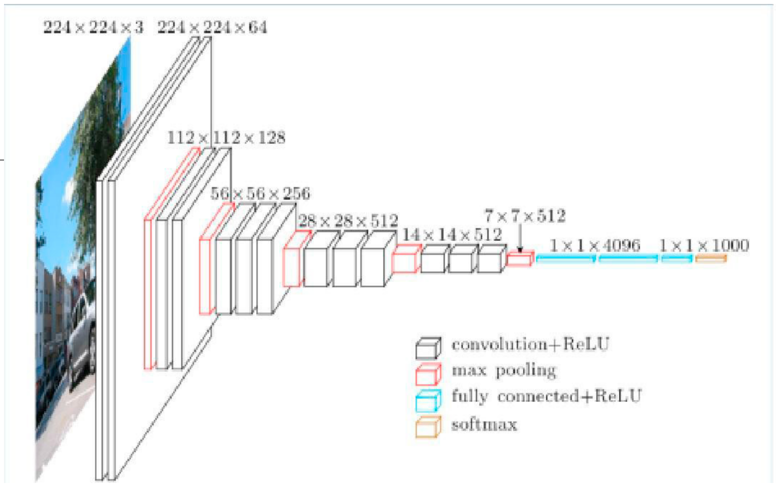

Fig. 3. Architecture of VGGNet CNN[11]

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Fig. 4. ConvNet Configuration [11]

Deep CNN(D-CNN) is using in this area also including Articulated pose estimation, Body configuration parsing, face parsing, Face recognition, object detection, path detection, plant disease estimation through the image of plant leaves, age and expression recognition through the face of human being, facial key-point detection, speech recognition.

2.3. Composition of Deep Convolutional and Subsample layer

The number of layers involves the work of Deep Convolution Neural Network (D-CNN) is scaled down by the convolutional layer and a subsample layer in a single layer. This concept was popularized by Simard, which was after known by Mamalet and Garcia.In this task, we change back to back sub-sample layers and convolutional the single convolutional layer using two strides. A pattern on an image can be extracted by the following expression:

$$p_i^{(t)}(q,p) = F\Big(\sum_{i=0}^{B}\sum_{u=0}^{R_q^{(t)}}\sum_{v=0}^{R_p^{(t)}} p_i^{(t-e)}(s_q^{(t)}q+u,\, s_p^{(t)}p+v)\, m_{ef}^{(t)}(u,v) + \theta_j^{(t)}\Big) \tag{1}$$

where $p_i^{(t-e)}$ and $p_j^{(t)}$ are the input and output pattern map respectively, F() is function i.e known by activation function which we have used in our work,$m_{ef}^{(t)}$ is the convolutional kernel weight $_j^{(t)}$ represents bias denotes total number of input feature mapping, $s_q^{(t)}q$ represents horizontal convolution step size, $s_p^{(t)}p$ represents vertical convolution step size, and $R_q^{(t)}$ and $R_p^{(t)}$ are width and height of convolutional kernels, respectively. where $M^{(t-e)}$and $A^{(t-e)}$ and height and width of input feature mapping.

$$A^{(t)} = (A^{(t-e)} - R_p^{(t)})/s_p^{(t)} + 1 \tag{2}$$

$$M^{(t)} = (M^{(t-e)} - R_q^{(t)})/s_q^{(t)} + 1 \tag{3}$$

figure:5 depicts how CNN work with an image for Gender Recognition and give final unique output on the basis of internal processing of pixels, patterns of image.

2.4. A CNN for gender estimation

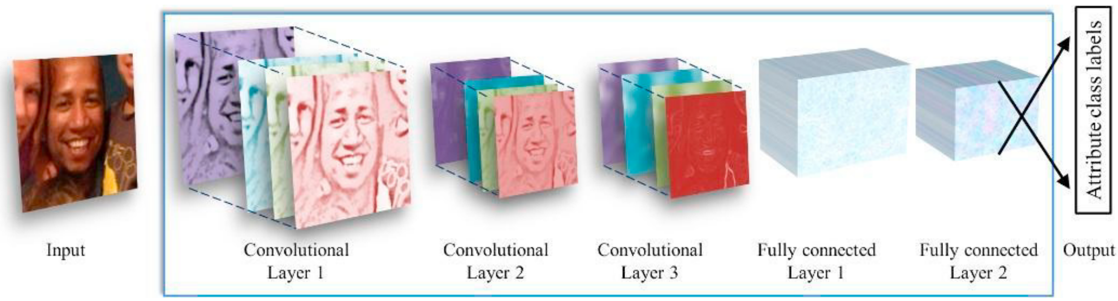

Fig. 5. How CNN work with image for Gender Recognition

Collecting a huge, labelled image of training dataset, gender estimation from collection of social picture, images that don’t needs an access to personal details of the subjects that are not displaying in the images for example their birth date and traditional method that involves collection of other details of a person and on the basis of which we find gender on manually handle labelled data-sets for gender recognition. That is the reason why we are using D-CNN which directly work on an image and helps to estimate gender accurately.Overfitting is a trivial problem usually Comes when machine learning or deep learning based methods have come into action on such a limited collection of face images of our dataset.

2.5. Network architecture

Fig. 6. Network Architecture for Gender classification

Images are scaled again to 256 x 256 size image and a then perform cropping operation on the image of size 227 x 227 which is passed into the network. The three consecutive convolutional layers are then described as
The following fully connected layers are then described as follow:
The first step is, FC layer that gets the output from the third convolutional layer and which exhibit neurons equal to 512 and superseded by an activation function Rectified Linear Unit(Relu) and a dropout layer.
The second step is, FC layer that gets the 512-dimensional output from the first FC layer and same procedure follow like in the first layer.
The third step is, absolutely affiliated fully connected layer which maps to the final classes for gender classification.At last, the output obtained from last one absolutely fully connected layer is forward to a soft-max layer function then it assigns a probability for each label in gender detection.The anticipation is fabricated by using the label that is having high probability from the rest of the test image used in gender recognition.

3. Dataset

We are going to use a subset of the Celebrity face data-set for our experiment. Celebrity face is a large-scale celebrity face attributes data-set. It consists of more than 100k celebrity face images, each having 40 binary face attribute annotations, with gender being one of them. We have selected a random subset of 200 face images in which 100 male and female faces each for the purpose of this experiment. Let us take a look at the distribution of the training and test splits. First, we will run our code on training-set consist of 160 images and test consist of 40 images. Then label the image in data-set with label=0 stands for"female" whereas label=1 stands for a male. This same procedure we follow with LFW face dataset also.

4. Training and testing

The weights that we have used as a part of all layers are introduced with stochastic esteems from a mean Gaussian with value zero and a standard deviation around 0.01.We don't utilize pre-prepared profound Convolutional Neural Network models for introducing the system; this system framework is trained, from a root, without utilizing any in-formation outward of the pictures and the names accessible by the benchmark. This is again, should be contrasted and profound Convolutional Neural Network(CNN) executions utilized for confronting acknowledgment with

gender, age, facial articulation, where a huge number of pictures are utilized for preparing.
We have used specific network architecture as well as the dataset. We are going to use a combination of LogSoftMax + NLL-Loss in PyTorch to train the network.
Next, we load the pre-trained VGG-Face model and dataset. We have also initialized our network architecture, loss module and certain other training parameters such as the number of epochs to train and the batch size. At first, we fix the seed of the various random number generators that our code required to use. Why do we require to do this. In

Table 1. Comparison between accuracy of existing method and our proposed work

S.no	Method	Method Description	Accuracy
1	VGGNet arch	Proposed Work	0.95
2	ALEXNet [Gkhan zbulak et al.,2016]	existing[8]	0.9330
3	High-dim-LBP[Cao et al.,2013]	existing[5]	0.9317
4	Discrete wavelet transform [Sajid Ali Khan et al.,2013]	existing[4]	0.91
5	GOOGLENet arch [Sebastian Lapuschkin et al.,2017]	existing[12]	0.89

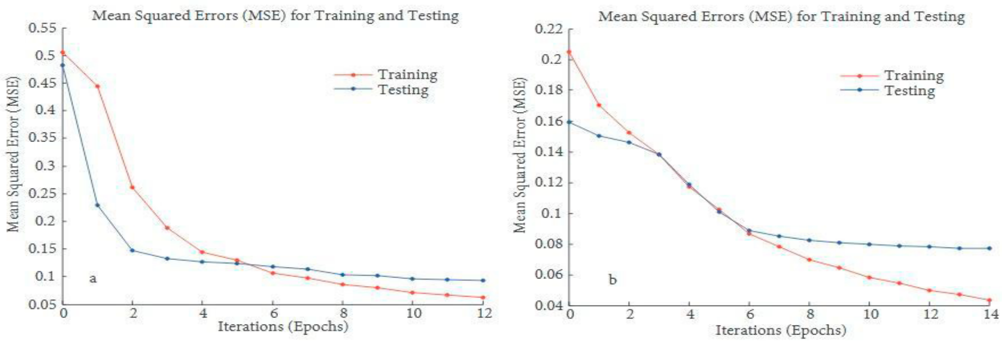

Fig. 7. Training and testing errors for Celebrity face data-set and LFW data-set respectively

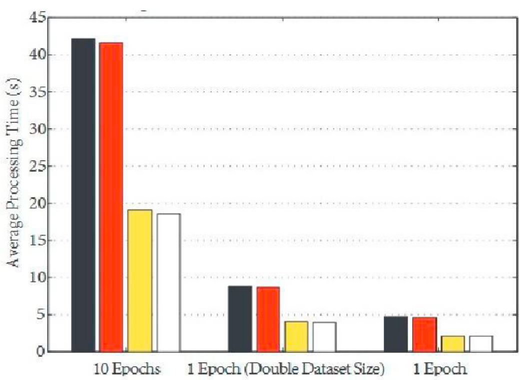

Fig. 8. Training performance for VGGNet D-CNN architectures running on GPU

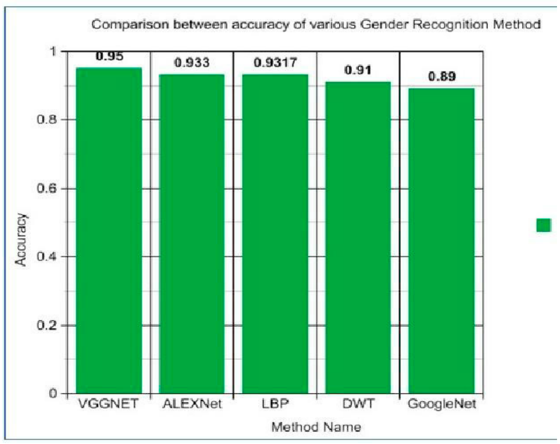

Fig. 9. Accuracy Comparison

5. Conclusion and Future work

Despite numerous past strategies have to manage the issue of gender recognition through face image, in this paper we have set some standard that relies on state-of-the-art VGGNet network architectures and tries to show the deter-mination of gender recognition through face image that can improve overall accuracy with VGGNet architecture of Deep Convolution Neural Network (D-CNN). Before my research, no one has used VGGNet for gender prediction with Celebrity face dataset and obtain such accuracy. I have used OpenCV, pytorch, tensorflow library with python language to implemented code with Graphical Processing unit(GPU) that has shown a great result on such huge number facial image dataset. On the other hand, if I had additional time, I would have devoted more e ort towards performing calibration of the parameters and the adjustment of architectures I explored di erent avenues regarding our experiment. I would wish to use the other Deep Convolution Neural Network architecture like, GoogleNet, Resnet, DenseNet in the same problem in place of VGGNet that I have used in this paper. I would have the wished to supplant the different completely associated fully connected layers at the end part of this architecture with just a single layer and rather than this moved those parameters over to extra convolutional layers.
By a wide margin, the most troublesome region of this undertaking was building up the preparation foundation with the appropriate division of the information into folds, prepare every classifier, performing cross-approval, what's more, join the different resulting classifiers into a test-prepared classifier. I am expecting that the future scope work in this is to involve using face age, human expression classification to aid face recognition, facial disease detection, improve experiences with images, pics of social media, and much more than this.

Reference

[1] Tivive FHC, A.Bouzerdoum(Sep 2006) ”A gender recognition system using shunting inhibitory convolutional neural networks” In: International Joint Conference on Neural Networks; Vancouver, Canada. New York, NY, USA: IEEE. pp.5336-5341.

[2] Eidinger, R. Enbar, and T. Hassner(Dec 2014)”Age and gender estimation of unfiltered faces”IEEE Transactions on Information Forensics and Security,9(12):21702179,.

[3] Fu,G.Guo, and T.S.Huang(Nov 2010)”Age synthesis and estimation via faces:A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence,32(11):19551976.

[4]Sajid Ali Khan, Muhammad Nazir, Naveed Riaz(2013)”Gender Classification using Multi-Level Wavelets on Real World Face Images”, Springer Acta Polytechnica Hungarica Vol. 10, No. 4.

[5] Dong Chen, Xudong Cao, Fang Wen, Jian Sun(2013),”Blessing of Dimensionality: High-dimensional Feature and Its E cient Compression for Face Verification”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3025-3032.

[6] Y. H. Kwon and N. da Vitoria Lobo(Jun 1994)” Age classification from facial images “In Computer Vision and Pattern Recognition,1994. Proceedings CVPR., 1994 IEEE Computer Society Conference on, pages 762767.

[7]Gil. Levi and T. Hassner(June 2015) ”Age and gender classification using convolutional neural networks.In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops.

[8] Gkhan zbulak, Yusuf Aytar, Hazm Kemal Ekenel(September, 2016.) ”How Transferable are CNN-based Fea-tures for Age and Gender Classification”,arXiv:1610.00134v1 [cs.CV].

[9]E. Makinen and R. Raisamo(September, 2016)”Evaluation of gender classification methods with automatically detected and aligned faces.IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(3):541547.

[10]. Baluja and H. A. Rowley(2007)”Boosting sex identification performance” Int. J. Comput. Vision, 71(1):111119.

[11]Karen Simonyan Andrew Zisserman(10 April,2015)”Very Deep Convolution Networks For Large ScaleImage Recognition”,,Published as a conference ICLR arXiv:1409.1556 [cs.CV].

[12] Alexander Binder, Sebastian Bach, Gregoire Montavon, Klaus-Robert ,Klaus-Robert Mller, Wojciech Samek, ”Layer-Wise Relevance Propagation for Deep Neural Network Architectures”, Springer Singapore, 913–922, 978-981-10-0557-2, 10:1007=978 981 10 0557 287: