Using Convolutional Neural Networks to discover cogntively validated features for Gender Classification

Ankit Verma, Lovekesh Vig
School of Computational & Integrative Sciences
Jawaharlal Nehru University
New Delhi-110067, India
e-mail: ankit78_sit@jnu.ac.in, lovekesh@jnu.ac.in

Abstract— **The human visual cortex is extremely adept at distinguishing between male and female faces, or performing "Gender Classification". While the subject of face detection and recognition has received a lot of focus, research into the features or cognitive processes that are useful for identifying gender have received relatively little attention. Researchers have attempted to extract hand crafted features like wavelet coefficients, histograms etc. on the basis of which to generate a model to classify the male and female faces. However, these models tend to compress the image into a vector and disregard the two dimensional spatial correlations between the pixels in an image. Additionally, these features have to hand crafted and may or may not be ideal for the classification at hand. Ideally, the system should be able to generate specific features from the input face image which would help in classification of male faces from female faces. In this paper and a Deep Convolution Neural Network (CNN) model is presented for gender classification. The features generated by the CNN appear to agree with known results from the cognitive science community indicating that these models may be closer to biological neuronal processes governing gender classification. The classification results are compared with different regularization techniques and other standard classifiers, and the CNN models yield higher accuracy than both svms and random forest classifiers.**

Keywords-convolutional neural network, backpropagation algorithm, gender classification, horizontal flipping, L2 Regularization.

I. INTRODUCTION

Image recognition has been a particularly important challenge for machine learning scientists. Computational vision researchers have devised many different features to extract from and classify images utilizing variance in color, luminance, texture and relative positions of features. However, a lot of time and effort must be spent in determining which features would be most appropriate for a particular application and this still remains somewhat of an art. In this paper we utilize a convolutional network in order to discover novel features for the gender classification problem. Convolutional networks have their roots in neuroscience and we attempt to study whether these networks generate features that have been indicated by the cognitive neuroscience community as being useful for gender classification. We also wish to compare the performance of convolutional networks against standard classifiers like SVMs (Support Vector Machines) and random forests. On seeing a human face through our eyes, there are some specific features for example eyebrows, eyes, skin, lips, texture, color etc. on the basis of which our brain classifies male and female faces [1]. Instead of attempting to code the features by hand, we want to allow the network to learn features relevant to the gender classification task.

Although a lot of work has been done in the area of gender classification and models that yield a reasonable accuracy has been proposed in this field, yet most of these models depend exclusively on hand coded features which are very often the same features utilized for face detection or recognition. Additionally these methods often ignore the two dimensional spatial correlations between pixels because they fold the image into a one dimensional vector. In this paper we take a 6 layer convolution neural network and train it using Resilient Back propagation as the training algorithm. The resulting network is robust to orientation and other lighting effects as compared to the other general classifiers. The accuracy of these networks also exceeds the accuracy obtained by the other methods.

In the input a human face image of size 32x32 is input to a multilayer CNN. Which then goes for multiple sequential passes through several intermediate hidden layers one after another for processing and finally on the last layer i.e. on the output layer we get the result whether it's a male face or female face. These hidden layers are just alternative combinations of Convolution and Sub-sampling sub-layers.

A. Introduction to CNN

CNN are biologically inspired models inspired from the experiments performed by Hubel and Wiesel on the cat cortex in 1959[2] to understand the mechanism of visual system. CNNs are multilayer feed forward neural network models and have been used for visual pattern recognition. Although CNNs are biologically inspired and are a variation of multilayer perceptrons but these are more quick, accurate, robust and reliable than standard multi layer perceptrons. Because they accommodate two dimensional filters, they provide more accuracy than primitive classification tools like SVM or PCA. CNN are running successfully in various practical applications from OCR to video surveillance [3][4][5][6][7]. One main feature of CNNs is that these allow for weight sharing within layers, thereby greatly reducing the

978-1-4673-6751-6/14 $31.00 © 2014 IEEE
DOI 10.1109/ISCMI.2014.17

number of parameters in the network. This allows the networks to reduce overfitting and generalize well to new examples. Additionally, minimal preprocessing is required and no hand coded features are necessary for a CNN as these are able to extract features Some features of CNN which makes it a robust approach are-Its feature extraction, learning capability and results doesn't effect from disturbances of geometric transformations like image scaling, shifting, translating, rotating, contrast, brightness, gamma etc. Additionally, It is the single integrated architecture of CNN that is adaptive for feature extraction, learning and classification. Finally, CNN extracts the features at a very fast rate as compared to other approaches.

II. Related Work

A. Convolutional Neural Networks

Convolutional Neural Networks have been proved robust and reliable assets for visual pattern recognition over the years. There is renewed interest in CNN's as modern hardware now permits many more layers than were previously possible. CNNs are running successfully in various practical applications from OCR to video surveillance [3][4][5][6][7]. There are many hundreds of usage and researches worldwide that have improved the CNN periodically, customized it architecturally according to their problem and proved the capability of CNN.

In 2003 CNNs were established as a robust tool in the field of face detection with a good generalization [8]. CNN also yielded a better result and performance in classification of facial expressions like smiling face recognition. In 2011 Ciresan et al. refined the architecture of CNN and make it possible to implement on GPU machines, which in turn made it possible to train CNN's with many more layers and therefore learn more complex non-linearities [9]. In 2012 Ciresan et al. make it possible to prove the performance and highly impressive accuracy of CNN for multiple object image databases like CIFAR10, NORB, GTSRB and handwritten digit database MNIST [10]

CNN's have been deployed in industrial applications in many the organizations. CNNs have been deployed by Microsoft for OCR and handwriting recognition systems supporting Chinese and Arabic alphabets also [3][4][5][6]. To protect privacy CNN is deployed by Google to detect faces and license plate in StreetView images [11]. CNN has been implemented by NEC in supermarket of Japan to recognize the gender and age of customers [7]. Using CNN Vidient Technologies has implemented video surveillance system on many airports in the US [7]. CNN is being used for face detection in video conferencing system by France T´el´ecom [12]. CNN have also been used for vision based obstacle avoidance for off-road mobile robots [13].

B. Gender Classification

Cognitive scientists have for long been focused on determining the criteria human observers use for gender discrimination ([14][15][16][17][18]). Researchers have systematically manipulated facial elements or features to observe the effect of these features on human performance in the gender classification task. Two prominent approaches include geometric features based on distances between facial landmarks [17], and looking at each facial component separately [19]. However, the results demonstrate that gender discriminative features may be more complex than simple geometrically defined landmark distances [18] and have emphasized the significance of a wide range of shape and intensity cues[18][19]. The work with discrete features also showed that isolated discrete features the nose[20], jaw[21] ,eyebrows[22] and facial outline also play a role in gender classification. It has also been observed that there may be variation of the discriminative features across faces of different races[20]. One of the objectives of this paper is to examine the features that the convolutional network learns and attempt to find correlations between the features learnt by the network, and the features indicated by the cognitive neuroscience community.

Gender classification has also been the focus of machine learning researchers. Early systems used a PCA based mapping into eigenspace to classify male and female faces. The mapping captured information useful for discrimination of faces [23][24]. When combined with a linear classifier these models yielded good classification accuracies. However, the errors made by these were easily classified by human subjects thus indicating that the processes in the brain do something different than a PCA. Recently, CNNs have also been applied to the problem of gender classification: Ng et al. [25] have utilized a convolutional neural network for gender classification using full body images. Tivive et al [26] implemented a shunting inhibitory CNN and applied it to the Feret dataset with very high accuracy. Duffner [27] has applied CNNs to face detection and gender recognition.

In this paper we show that a much simpler CNN with fewer parameters is capable of performing gender classification on a harder dataset of images collected from the internet, simply by utilizing some regularization techniques like image flipping and weight decay. Additionally the dataset used in this paper differs from earlier datasets in that there is no overlap between the test and training set images, i.e. images of the same person is not presented in both training and test sets. Additionally, the images are not from an artificially generated set of images, but are randomly downloaded from the internet. The intermediate features generated appear to correlate well with the discriminative features observed by cognitive scientists.

III. Experiments

Data Preparation:

In the experiment total 4700 face images (2350 male and 2350 female) of different age group are used at 32x32 resolution. The images are extracted from various sources over the web to cover different illumination, intensity, brightness and contrast such that the network would train to

accommodate these effects more robustly. Before preparing datasets, the images are cropped in such a way that only the facial area should be visible and hair and background are cropped. The image size after cropping is kept 32x32. The experiment is performed on two versions of the dataset separately. In the initial approach the train set contains 1700 male and 1700 female faces, and in the test 650 male and 650 female faces are taken. To generate additional training data another version of train data set is prepared such that, all the original training set images are flipped and added in training set. After this augmentation training set gets doubled i.e. it contains 3400 male and 3400 female faces. The test set is kept same and was not modified and also there is no intersection between train set and test set images i.e. both have their different set of images. Some of the faces from database are shown below in figure 1 -

Figure 1. Some faces from the database

Architecture Used:

CNN Architecture used here is consists of 6 layers. Out of six, three layers are convolution layers, two are sub-sampling and one is the output layer as shown in Figure 2.

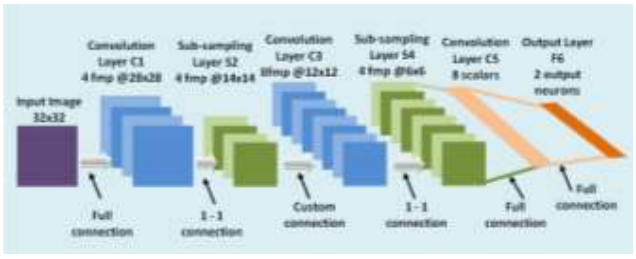

Figure 2. Network Architecture used for the experiment

Experiments Performed:

The CNN architecture described in previous section was designed. A training set of 1700 male and 1700 female faces, and test set of 650 male and 650 female faces is taken, the weights of filters are initialized randomly with values 0 to 1 using standard normal distribution. The network was implemented in MATLAB. Using this architecture after 25,000 epochs the classification accuracy rate of **83.46%** is achieved on test set. This accuracy rate was somehow much better to an extent as compared to general classifiers like SVM or PCA. But for the purpose to increase in the accuracy rate 'FLIPPING' technique is applied, i.e. all the face images in training set made flipped about a vertical axis and augmented to the training set. Flipping make the training more robust regarding positional changes in features. After flipping and augmentation there were 3400 male and 3400 female face images in training set. The test set was not modified. Again the network is trained afresh. After running for 25,000 epochs the accuracy get stabilized and improvement in accuracy is noticed, the test classification accuracy rate is reached to 86.15% i.e. flipping give us an improvement of 2.69%. To improve it more, along with flipping, L2 regularization is also applied with the weight decay of 0.01. In regularization, an additive term is added to the error value proportional to the sum of squared weights, thereby ensuring that weight values are not allowed to grow to very large values, thereby preventing the network from getting stuck in local minima resulting in less over-fitting. On implementing L2 Regularization the network is again put up on training freshly starting from epoch 1 with same training and test sets that were obtained after flipping. After 2600 epochs at accuracy stability the test classification accuracy rate of 88.46% is observed, i.e. L2 Regularization give us an improvement of 2.31%. This is significantly much better classification accuracy than the other general classifier as shown in Table 1.

Now, let us see that how the features are extracted at different layers in CNN. In figures given below it is tried to display the features maps at different convolution and sub-sampling layers. The features extracted by the filters are shown for layer 1 to layer 4 only because after layer 4 each feature map gets decoded into a single scalar value, which can't be seen in the form of picture.

The feature extraction for an image of a female face at different layers is shown in the figure 3. Examining the feature maps obtained in the first convolution layer gives us clues about which features the network is encoding. It is clear from the figure that the different feature maps encode different attributes of the facial image such as the variations in the luminance between the eyes and the mouth, the nose, eyes, jaw and upper mouth regions and the outer contours of the face. These features are then combined in a non-linear fashion in the subsequent convolution layers. The features tend to correlate well with the findings of cognitive researchers who also observed that these features were useful for gender recognition. Furthermore, when we conducted tests with human observers, 80.4 % of the errors made by the network were also recorded as "unsure" by the human subjects. This indicates that CNN's are closer to the way humans recognize gender than say PCA, SVMs or Eigenspace based models.

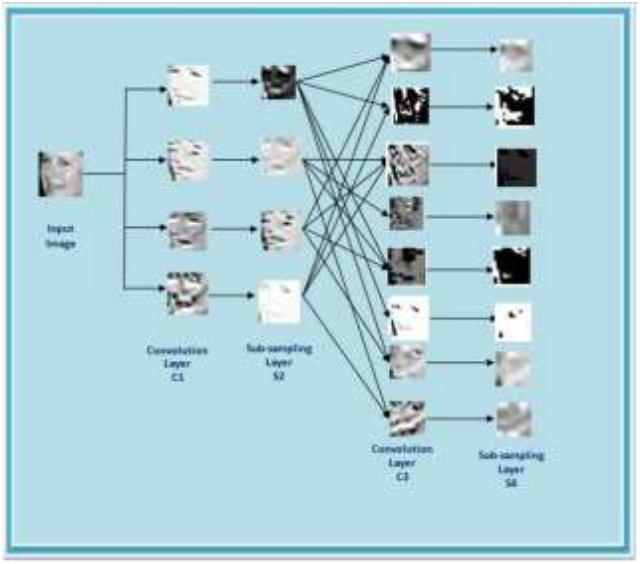

Figure 3. Feature maps generated at different layers

IV. RESULTS

In the results it is found that in gender classification CNNs with average pooling give an accuracy of 83.46%. If the training set is augmented with its horizontally flipped images then this accuracy gets increased by 2.69% i.e. it becomes 86.15%. But if L2 Regularization is also applied with a weight decay of 0.01 times then the accuracy rate improves to 88.46% which is a good classification accuracy rate.

The following table summarizes the results of experiment:

TABLE I. RESULT COMPARISON OF CNN WITH OTHER TECHNIQUES

System	Classification Accuracy
Random Forest (100 trees)	58.00%
SVM + Linear Kernel	79.31%
CNN	83.46%
CNN+ Horizontal Flipping	86.15%
CNN+Flipping+L2Regularization	**88.46%**

After training the CNN we get a well trained and robust network that can classify the male and female faces. As you know that network will use the filters at various convolution layers to extract the features relevant for classification. In the following figure, the filters from the first layer of our trained CNN having 88.46% classification accuracy are being displayed in Figure 4.

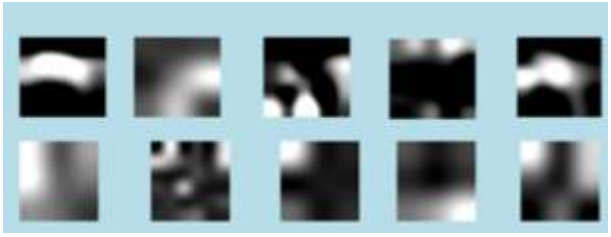

Figure 4. Some filters of the trained CNN

TABLE II. Classification Results for different CNNs

Architectures/ Methods	No. of weights	No. of Feature Maps						Classification rate (%)		
		L1	L2	L3	L4	L5	L6	Male	Female	Average
CNN	10608	8	8	16	16	16	1	78.71%	77.28 %	78.00 %
CNN	6012	6	6	12	12	12	1	84.57 %	77.28 %	80.92 %
CNN	2712	4	4	8	8	8	1	85.00 %	81.92 %	83.46 %
CNN+L2 Regularization	2712	4	4	8	8	8	1	89.38 %	83.53 %	86.46 %
CNN + Hor. Flipping	2712	4	4	8	8	8	1	89.38 %	82.92 %	86.15 %
CNN + Hor.Flipping + L2 Regularization	2712	4	4	8	8	8	1	**90.15%**	**86.76 %**	**88.46%**

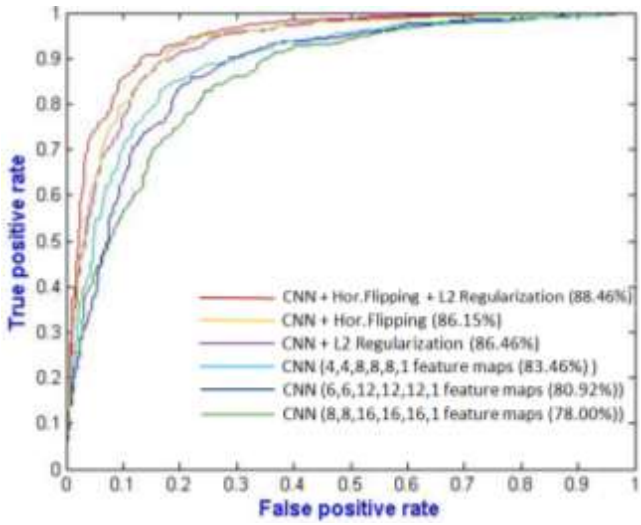

Figure 5. The ROC (Receiver Operating Characteristic) curves of the trained CNNs

V. DISCUSSION & FURTHER WORK

The results show that CNN is much better and robust tool than previous general classifiers like SVM and Random Forest etc. Although it is geometrically invariant but if a little bit dependency is found due to positional variance in features extracted from the images then this dependency can be completely removed by using the flipping technique. So, CNN performs better when the flipping is used and it becomes superior if Regularization is also applied along with flipping. The network takes a very long time in training, so if implement and run it over GPU machine then it will

give surprisingly very fast training as compared to a normal system/server.

REFERENCES

[1] Russel, Richard (2003), Sex, beauty and the relative luminance of facial features, Perception, 32, 1093–1107

[2] Hubel, D. H.; Wiesel, T. N. (1959). "Receptive fields of single neurones in the cat's striate cortex". The Journal of physiology 148 (3): 574–591. PMC 1363130. PMID 14403679

[3] Y. Simard, Patrice, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in ICDAR'03

[4] K. Chellapilla, M. Shilman, and P. Simard, "Optimally combining a cascade of classifiers," in Proc. of Document Recognition and Retrieval 13, Electronic Imaging, 6067, 2006.

[5] A. Abdulkader, "A two-tier approach for arabic offline handwriting recognition," in IWFHR'06

[6] Chellapilla and P. Simard, "A new radical based approach to offline handwritten east-asian character recognition," in IWFHR'06

[7] Garcia and M. Delakis, "Convolutional face finder: A neural architecture for fast and robust face detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004.

[8] Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). "Subject independent facial expression recognition with robust face detection using a convolutional neural network". Neural Networks **16** (5): 555–559. doi:10.1016/S0893-6080(03)00115-1

[9] Ciresan, Dan; Ueli Meier; Jonathan Masci; Luca M. Gambardella; Jurgen Schmidhuber (2011). "Flexible, High Performance Convolutional Neural Networks for Image Classification". Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Two **2**: 1237–1242

[10] Ciresan, Dan; Meier, Ueli; Schmidhuber, Jürgen (June 2012). "Multi-column deep neural networks for image classification". 2012 IEEE Conference on Computer Vision and Pattern Recognition (New York, NY: Institute of Electrical and Electronics Engineers (IEEE)): 3642–3649.

[11] Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L.Vincent, "Large-scale privacy protection in street-level imagery," in ICCV'09

[12] LeCun, Yann; Léon Bottou, Yoshua Bengio, and Patrick Haffner (1998). "Gradient-based learning applied to document recognition". Proceedings of the IEEE 86 (11): 2278–2324. doi:10.1109/5.726791

[13] LeCun, U. Muller, J. Ben, E. Cosatto, and B. Flepp, "Off-road obstacle avoidance through end-to-end learning," in Advances in Neural Information Processing Systems (NIPS 2005) MIT Press, 2005

[14] Bruce V., Burton, A. M., Dench, N., Hanna, E., Healey, P., Mason, 0., Coombes, A., Fright, R., & Linney, A. (1993). Sex discrimination: How do we tell the difference between male and female faces? Perception, 22, 131-152.

[15] Bruce, v., Ellis, H., Gibling, E, & Young, A. (1987). Parallel processing of the gender and familiarity of faces. Canadian Journal of Psychology, 41, 510-520.

[16] Bruce, v., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. Perception, 23, 803-822.

[17] Bruce, v., & Young, A. W. (1986). Understanding face recognition.British Journal ofPsychology, 77, 305-327.

[18] Burton, A. M., Bruce, v.. & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. Perception, 22,153-176.

[19] Richard Russell, Sex, beauty, and the relative luminance of facial features, Perception, 2003, 32, pages 1093 -1107

[20] Chronicle, E. P., Chan, M., Hawkings, C., Mason, K., Smethurst,K., Stallybrass, K., Westrope, K., & Wright, K. (1995). You can tell by the nose-Judging sex from an isolated facial feature. *Perception,* 24, 969-973.

[21] Brpwn, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception,*22, 829-840.

[22] Yamaguchi, M. K., Hirukawa, T.,& Kanazawa, S. (1995). Judgment of gender through facial parts. *Perception,* 24, 563-575.

[23] A. B. A. Graf and F. A. Wichmann. Gender classification of human faces. Biologically Motivated Computer Vision, pages 491–501, 2002.

[24] A. J. O'Toole, T. Vetter, N. F. Troje, and H. H. Bulthoff. Sex classification is better with three-dimensional structure than with image intensity information. Perception, 26:75–84, 1997.

[25] C. Boon Ng, Y. Haur Tay, and B. M. Goi, "Vision-based Human Gender Recognition: A Survey, " ArXiv e-prints, Apr. 2012.

[26] Fok Hing Chi Tivive , A shunting inhibitory convolutional neural network for Gender Classification, Faculty of Engineering and Information Sciences, University of Wollongong, 2006

[27] Stefan Duffner, Face Image Analysis With Convolutional Neural Networks, Albert-Ludwigs-University, Freiburg Breisgau