Open access • Proceedings Article • DOI:10.1109/IJCNN.2006.247311

A Gender Recognition System using Shunting Inhibitory Convolutional Neural Networks — Source link

Fok Hing Chi Tivive, Abdesselam Bouzerdoum

Institutions: University of Wollongong

Published on: 30 Oct 2006 - International Joint Conference on Neural Network

Topics: Face detection, FERET database, Convolutional neural network, Facial recognition system and Contextual image classification

16-7-2006

A gender recognition system using shunting inhibitory convolutional neural networks

Fok Hing Chi Tivive
University of Wollongong, tivive@uow.edu.au

Abdesselam Bouzerdoum
University of Wollongong, bouzer@uow.edu.au

A gender recognition system using shunting inhibitory convolutional neural networks

Abstract

In this paper, we employ shunting inhibitory convolutional neural networks to develop an automatic gender recognition system. The system comprises two modules: a face detector and a gender classifier. The human faces are first detected and localized in the input image. Each detected face is then passed to the gender classifier to determine whether it is a male or female. Both the face detection and gender classification modules employ the same neural network architecture; however, the two modules are trained separately to extract different features for face detection and gender classification. Tested on two different databases, Web and BioID database, the face detector has an average detection accuracy of 97.9%. The gender classifier, on the other hand, achieves 97.2% classification accuracy on the FERET database. The combined system achieves a recognition rate of 85.7% when tested on a large set of digital images collected from the Web and BioID face databases.

Disciplines

Physical Sciences and Mathematics

Publication Details

This article was originally published as: Tivive, FHC & Bouzerdoum, A, A gender recognition system using shunting inhibitory convolutional neural networks, International Joint Conference on Neural Networks (IJCNN '06), 16-21 July 2006, 5336 - 5341. Copyright IEEE 2006.

2006 International Joint Conference on Neural Networks
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

A Gender Recognition System using Shunting Inhibitory Convolutional Neural Networks

Fok Hing Chi Tivive, and Abdesselam Bouzerdoum, *Senior Member, IEEE*

***Abstract*— In this paper, we employ *shunting inhibitory convolutional neural networks* to develop an automatic gender recognition system. The system comprises two modules: a face detector and a gender classifier. The human faces are first detected and localized in the input image. Each detected face is then passed to the gender classifier to determine whether it is a male or female. Both the face detection and gender classification modules employ the same neural network architecture; however, the two modules are trained separately to extract different features for face detection and gender classification. Tested on two different databases, Web and BioID database, the face detector has an average detection accuracy of 97.9%. The gender classifier, on the other hand, achieves 97.2% classification accuracy on the FERET database. The combined system achieves a recognition rate of 85.7% when tested on a large set of digital images collected from the Web and BioID face databases.**

I. Introduction

Gender recognition is an important cue for many potential applications, such as biometric authentication, passive demographic data collection, and intelligent human-computer interfaces. Nevertheless, automatic gender recognition remains a difficult perceptual task because of variations in facial appearance and other extrinsic changes that occur in images, e.g. illumination conditions, face size, face pose and image quality. Over the past fifteen years, several studies have been conducted in gender recognition; however, most of the studies were concerned with developing different types of classifiers. Golomb *et al.* [1] developed a fully-connected two layer *multilayer perceptron* (MLP) to classify the gender of face images of size 30×30. Tested on a small set of 90 images, their neural network gender classifier achieved an accuracy of 91.9%. Brunelli and Poggio [2], on the other hand, used the radial basis function neural network for gender classification. They trained two competing networks, one network for male and the other one for female, and used 16 geometric features (e.g. pupil to eye brow separation, nose width, eye brow thickness, etc.) per face image as input. Evaluated on a test set of 168 face images, the classification rate of their gender classifier was 79%. To reduce the number of inputs to the neural network, Sun *et al.* [3] applied principle component analysis (PCA) on the face image to compress the input space into low dimensional features (or eigen-vectors), and used a genetic algorithm (GA) for selecting a set of eigen-vectors to represent the two classes. For classification, they used a two layer neural network with 400 neurons in the input layer and a single neuron as output. Using all the features generated by PCA, they achieved a classification accuracy of 82.3%; however, on a reduced set of features selected with GAs, the classification accuracy was increased to 88.7%. In addition to PCA for feature extraction, Balci and Atalay [4] developed a pruning scheme to improve the generalization ability of the MLP by deleting those connections that least affect the performance of the network. Based on a subset of 560 clean face images, i.e. images with no facial features or glasses, taken from the FERET database, the highest classification accuracy was 91.54%

Other types of gender classifiers have also been proposed. For instance, Wu *et al.* [5] employed the Adaboost learning algorithm to select a series of look-up table (LUT) weak classifiers for gender classification. They trained their Adaboost-classifier on over 11,000 face images, and achieved a classification rate of 88% with 200 LUTs. Jain and Huang [6] used independent component analysis and linear discriminant analysis to develop a gender classifier. To test their method, they used 500 face images of size 64×96 from the FERET database, where each face image was aligned to maintain constant distance between the two eyes using an eye detection system. Moreover, the face image was masked to remove the hair information, and a histogram equalization method was applied to account for variations in lighting conditions. The accuracy reported in their work was 99.3%. On the other hand, Moghaddam and Yang [7] investigated the use of support vector machines (SVMs) to classify gender from face images. They used 1755 face images of size 21×12 to evaluate their SVM-classifiers. The face images were preprocessed for contrast and geometric shape variations, and a mask was applied to extract the facial region. Based on five-fold cross validation and using 20% of the training set as support vectors, they obtained a classification rate of 95.12% with a cubic polynomial kernel and 96.62% with the Gaussian RBF kernel.

In this paper, we present an automatic gender recognition system that can detect faces in images of arbitrary size, and recognize their gender. The system consists of two modules: a face detector and a gender classifier. Both face detector and gender classifier are based on the recently developed *shunting inhibitory convolutional neural network*, named SICoNNet for short [8]. The motivation for using this type of network is twofold. First, these networks can learn to perform feature extraction and classification with the same network architecture. Second, a certain degree of invariance to distortions and variations in two dimensional shapes is

Fok Hing Chi Tivive and Abdesselam Bouzerdoum are with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Northfields Avenue, Wollongong, NSW 2522, AUSTRALIA (emails: [fhct243@uow.edu.au][a.bouzerdoum@ieee.org]).

0-7803-9490-9/06/$20.00/©2006 IEEE

introduced within the network while preserving the spatial topology of the input data. The built-in invariance of the network allows the face and the gender classifiers to not only process face images in an up-right frontal position but also slightly rotated ones. Moreover, the preprocessing stage between the face detection and gender recognition is eliminated.

The outline of the paper is as follows. Section II gives a detailed description of the proposed convolutional neural network structure, which forms the basic building block of the automatic gender recognition system. Section III presents the face detection system, followed by the gender classifier in Section IV. Section V presents the experimental results. Finally, some concluding remarks are given in Section VI.

II. SICoNNet Architecture

SICoNNet is a new class of convolutional neural networks in which the feature extraction is performed by a special processing unit, known as the shunting inhibitory neuron [8]. This neuron model is based on the physiologically plausible mechanism of shunting inhibition, which plays an important role in visual information processing in the cortex [9]–[11]. The input layer of the network is an arbitrary two-dimensional (2D) array, used by the network to receive images from the environment. The input layer is followed by a number of processing layers, called the *hidden layers*. Each hidden layer consists of several planes of shunting inhibitory neurons, known as *feature maps*. Each neuron in a feature map is connected locally to a small neighborhood in the preceding layer, its *receptive field* (Fig. 1). The size of the receptive field is fixed for each hidden layer. However, different feature maps in the same hidden layer use different set of weights, but the same set of weights are used across a single feature map. That is, the feature map is constrained to have one set of weights where each neuron in that feature map performs the same operation on different parts of the input image. As a result, the same elementary visual feature is extracted from different positions in the input plane. In other words, the set of weights behaves as a convolution filter

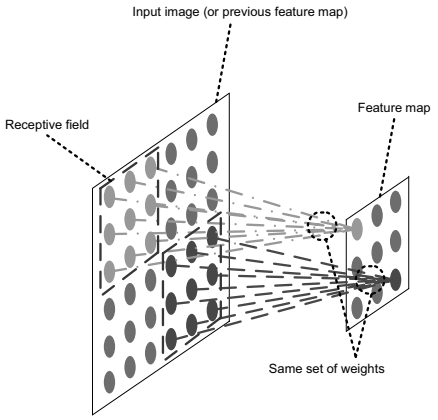

Fig. 1. A schematic diagram illustrates the application of local receptive fields in the input image.

where the outputs are stored at the corresponds neurons in the feature map. Different feature maps in the same hidden layer extract different types of local features; higher-order features are extracted by feature maps in higher processing layers. Another structural concept is implemented in the network so as to reduce the spatial resolution of the 2D input along the network and subsequently introduce a certain degree of shift and distortion invariance. This is done within each hidden layer by shifting the centers of receptive fields of adjacent neurons by two positions, horizontally and vertically. Thus, the size of the feature maps in each hidden layer is reduced by one quarter. The first hidden layer can have an arbitrary number of feature maps so long as the next hidden layer has twice the number of feature maps, and each feature map is connected to two feature maps in the succeeding layer, similar to a binary tree. Instead of using the traditional sigmoid neurons for feature extraction, a special type of processing elements, namely the shunting inhibitory neurons, is used in the feature map, and its neural response at location (i, j) in the k^{th} feature map of the L^{th} layer is computed using the following expression

$$Z_{L,k}(i,j) = \frac{X_{L,k}(i,j)}{a_{L,k}(i,j) + Y_{L,k}(i,j)}, \tag{1}$$

where

$$X_{L,k}(i,j) = g_L\Big(\sum_{m=1}^{S_{L-1}} [C_{L,k} * Z_{L-1,m}]_{(2i)(2j)} + b_{L,k}(i,j)\Big),$$

and

$$Y_{L,k}(i,j) = f_L\Big(\sum_{m=1}^{S_{L-1}} [D_{L,k} * Z_{L-1,m}]_{(2i)(2j)} + d_{L,k}(i,j)\Big).$$

$$\forall \;\; i,j = 1, ..., F_L$$

The parameters $C_{L,k}$ and $D_{L,k}$ are the set of excitatory and "inhibitory" weights, respectively, $b_{L,k}$ and $d_{L,k}$ are scalar parameters called the biases of the neuron, $a_{L,k}$ is the passive decay term, g_L and f_L are the activation functions, and F_L is the size of the feature map at the L^{th} layer. Within the same feature map, all the neurons share the same set of weights, $C_{L,k}$ and $D_{L,k}$, biases and the passive decay rate parameters, which implies that a feature map is formed by replicating a single shunting inhibitory neuron in a 2D plane. To prevent division by zero in (1), the passive decay rate, $a_{L,k}$, is constrained as follows:

$$a_{L,k}(i,j) + Y_{L,k}(i,j) \geq \varepsilon > 0, \tag{2}$$

where ε is a small positive constant.

As the output layer is used as the classification layer, all the extracted features from the last hidden layer are fed to the output layer for classification. There are two ways of organizing the inputs from the last hidden layer to the output layer: (i) all neuron outputs from the last hidden layer are directly fed to the output layer; (ii) a local averaging operation is applied to the feature maps before sending the signals to the output layer. Here, a small 2×2 non-overlapping mask is applied to all feature maps of the last hidden layer, and the average signals are fed to the output

neurons. The activity of an output neuron is given by

$$y = h\Big(\sum_{i=1}^{S_N} w_i z_i + b\Big), \tag{3}$$

where y is the output of the neuron, h is the output activation function, w_i's are the connection weights, z_i's are the input signals, S_N is the number of input signals, and b is the bias term

III. Face Detection and Localization

Before recognizing the gender from face images, the faces need to be detected and localized in the input image. Thus, a face detector is integrated into the gender recognition system. The face detector is built from a SICoNNet comprising three processing layers: two hidden layers and one output layer. The input layer is a 2D plane of size 32×32. The first hidden layer has three feature maps of size 16×16 and the second layer has six feature maps of size of 8×8. The output layer has one sigmoid neuron whose output is used to classify the 2D input pattern into a face or a non-face. Figure 2 presents a schematic diagram of a three layer SICoNNet. The activation functions, g_L and f_L, for the first layer are chosen to be the hyperbolic tangent and exponential functions, respectively, whereas in the second layer, g_L is set to the logarithmic sigmoid function. At the output layer, a hyperbolic tangent function is used as the output activation function. A 2×2 local averaging operation is applied to each feature map in the second layer to generate 97 inputs for the output neuron, including the bias constant. The size of the receptive field used throughout the network is 5×5. Overall, the network has 574 free parameters that need to be adapted.

A. Training Face Detector

Before training the network, the weights of the receptive fields are initialized with random values using a uniform distribution between $-1/w$ and $1/w$, where w is the width of the receptive fields. The bias parameters $b_{L,k}$ and $d_{L,k}$ of the neurons in the feature maps are initialized similarly with $w = 1$. As the passive decay rate parameter $a_{L,k}$ is constrained to be a positive constant, it is initialized in the range $[0, 1]$, subject to the condition given in (2). The training data was taken from the face database created by Phung *et al.* [12], which contains a large number of segmented face images. These face images have people of different ages and ethnic backgrounds, and they are varied in terms of the background, the person, the lighting condition, the facial expression and the face pose, as shown in Fig. 3. The desired outputs corresponding to the face and non-face patterns are set to 1 and -1, respectively. To avoid manually collecting non-face patterns from scenery images, a bootstrap training is used to train the network. First, 100 face and 100 non-face patterns are used as an initial training set from which the network parameters are adapted by a variant of Levenberg-Marquardt (LM) algorithm developed by Ampazis and Perantonis [13], where the Hessian matrix and the gradient information are computed in terms of Jacobian

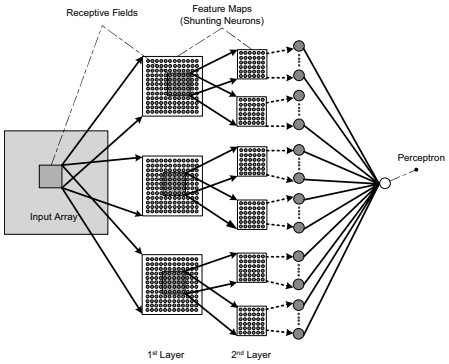

Fig. 2. A three layer binary-connected SICoNNet.

matrix using a modified error-backpropagation proposed by Hagan [14]. This training technique is chosen due to its superior convergence rate than the original LM algorithm; it includes an adaptive momentum term which reduces the risk of getting stuck in local minima. To avoid failure in the training process, the constraint in (2) is also maintained. Another separate set of equal number of face and non-face images, called the validation set (total 4,000 patterns), is used to validate and select the network with the lowest validation error for the next bootstrap session. Moreover, the threshold used at each bootstrap session to differentiate between the face and non-face patterns is computed as the point at which the total number of false detections and false dismissals is minimized over the validation set. The selected network is then applied as a face filter to scan scenery images. Those scanned windows that are falsely detected as faces (false detections) are collected as non-face patterns. To avoid adding to the training set a large number of well-trained patterns (i.e. face and non-face patterns that are close to the desired values), a small number of non-face patterns (100 samples) are added in the training set at each bootstrap session. Then the trained network is applied to a large set of face patterns; those face patterns that are misclassified as non-face (false dismissals) are collected and used to balance the number of face and non-face patterns in the training set. Therefore, at each bootstrap session, 100 false detections and 100 false dismissals are added to the existing training set. The bootstrap procedure is repeated with the same network until the size of the training set reaches 25,000 samples.

B. Face Localization Procedure

Once the network has been trained, it can be deployed for detecting and locating human faces. In general, faces have different sizes, and to locate them, the input image is sub-sampled at different scales (with a factor of 1.2) to form an image pyramid. The image at each scale of the pyramid

Fig. 3. Sample face patterns used to train the face detector.

is then processed by the network to generate an image of network responses. A common strategy to construct this output image from the scaled image is to extract individual input window at every position in the scaled image and pass it to the network to compute a network response. However, this is a time-consuming operation. Therefore, as the receptive fields in a CoNN behave as convolutional masks, the entire scaled image can be sent to the network, and at each hidden layer, the image is convolved with the receptive fields and down-sampled by a factor of two in both dimensions. This results in an output image which is $1/16^{th}$ the size of the scaled input image. The network responses from the output image are compared to a threshold T_{net}, and those network responses that are greater than T_{net} are considered face candidates; their positions are mapped back to an image having the same size as the original input image. This process is repeated for every scaled image.

It is unavoidable that during the face detection process a certain number of background windows will generate high network responses, and hence will be misclassified as face candidates. To circumvent this problem, the following post-processing step is performed. Considering that the human face is symmetric, the detected face candidate is folded along the Y-axis (mirror face image) and passed back to the network. The average of both network responses is taken as the final score of the detected face candidate. If the final score is less than T_{net}, the network response at that location is set to zero. Moreover, a number of overlapping detections usually occurs around the true face position. To fuse these overlapping detections into a single representative face candidate, a grouping technique similar to that used by Delakis and Garcia [15] is employed. First, all the face candidates from the series of output images are stored in a list and sorted in descending order according to their network responses. Suppose that S_{max} is the size of the top face candidate in the list, i.e. the face candidate with the highest network response. All face candidates whose centers are within a neighborhood of $0.25S_{max}$ from the center of the top face candidate and whose sizes are between $0.6S_{max}$ and $1.4S_{max}$ are grouped into a single face candidate cluster, and the cluster is removed from the list of face candidates. The process is repeated until all face candidates in the list are clustered. For each cluster, the center of the representative face is taken as the centroid of the cluster, and its confidence score is computed as the sum of all face candidate scores in the cluster, i.e. the volume of positive network responses in the cluster. The number of face candidates in the cluster is used to verify the corresponding representative face candidate by comparing it to a threshold T_{gp}. If the number of face candidates in the cluster is greater than T_{gp}, the corresponding representative face is mapped to the final output map.

To precisely locate the facial region for gender classification, two fine searches, both in location and size of the face candidate, are performed. The final position of the face is sought in a region of eight pixels around the center of the representative face. The location of all detected face candidates (positive detections) within the search grid are averaged to give the final position of the face. The number of positive detections in the search grid is compared to a threshold T_p, and if it is below the threshold, the representative face candidate is discarded; otherwise, the representative face is tested at 11 scales, ranging from 0.4 to 1.6 of the detected size. The number of positive detections is also recorded across all scales, and if it is greater than a threshold T_{sz}, the representative face candidate is considered as a true face candidate. Then, the size of the representative face is computed as the average size of the positive detections, and its confidence score is the volume of positive network responses.

Even though the representative face candidates have passed through all the above post-processing steps, there may still be some multiple detections around the true face. All remaining face candidates, which have a high confidence score, are stored in a list and sorted in descending order. Then a search for overlapping face candidates is performed, starting around the center of the face with the highest confidence score. Those overlapping face candidates whose centers are within a search region of size $0.5S_{max}$ are rejected. Furthermore, if the intersecting area of the overlapping face candidate is greater than $0.2S_{max}^2$, the face candidate is also removed. Finally, the remaining representative faces are passed to the network for final verification.

IV. Gender Classifier

Once the face is located and segmented into a face image, the next process is to determine the gender of the face. To perform this task, we apply the same network structure, except that no local averaging operation is applied to the feature maps of the last hidden layer, i.e. all the neurons in the feature maps of the second layer are fully connected to the output neuron. To extract and learn the features that differentiate between a male and a female, the network is trained on segmented male and female face images; see Fig. 4 for some sample images. The training of the network is based on a set of 11,000 face patterns including the mirror image versions, and the aforementioned training algorithm is used to train the network. The desired responses for male and female face images are set to 1 and -1, respectively.

A detected face is passed to the gender classifier where both the face image and its mirror version are processed by the trained gender classifier. The average of both network responses is taken as the gender score of the face image, which is compared to a threshold T_{gd}; if the score is greater than T_{gd}, the face is considered a male, otherwise it is a female. In certain cases, the face detector does not detect the exact size of the face. Therefore, the face image is evaluated at seven scales of its detected size, ranging from 0.7 to 1.3; a majority voting scheme is applied on the set of scores to verify the gender of the face image.

V. Experimental Results

In this section we present the experimental results of the gender recognition system. However, before evaluating the

Fig. 4. Face patterns from the gender database: (a) male patterns and (b) female patterns.

performance of the overall system, the performances of the face detector and gender classifier are first assessed separately. Then, the two systems are integrated and evaluated together as an automatic gender recognition system.

A. Performance of the Face Detector

The face detector was tested on two large sets of images. The first set consists of 1,000 images collected from the Web with people of different ages and gender. Each image contains one person whose face is not necessarily in full frontal position, but slightly rotated and/or tilted. Some of these images have people wearing spectacles or having moustache and/or beard. The other set of images is from the BioID face database [16], which has been used as a benchmark data set for face detection. It consists of 1,521 images of 23 distinct persons (i.e 1,521 faces). The thresholds used in both set of images were $T_{net} = 0.15$, $T_{gp} = 10$, $T_p = 10$ and $T_{sz} = 3$. To evaluate the performance of the face detector, the number of faces that have been correctly detected is recorded. A correct detection is defined by marking tightly the face region with a square box that covers the facial features such as the eyes, the nose and the mouth. False detections are also recorded for statistical analysis. Table I summarizes the detection results obtained from the two test sets, together with results from other face detectors based on the BioID dataset. Tested on images collected from the Web, the face detector achieves a classification rate of 95.7% with 109 false detections, whereas on the BioID database, the detection accuracy is 99.3% with 178 false detections. Overall, the face detector has an average detection accuracy of 97.9%. Compared with other existing face detection systems tested on the BioID database, our face detector achieves the best detection rate.

TABLE I

DETECTION PERFORMANCES OF THE FACE DETECTOR AND A COMPARISON OF OUR RESULT WITH OTHER FACE DETECTION SYSTEMS BASED ON THE BIOID DATABASE.

Database	Face Detector	F. Detections	C. Detection rate (%)
Web	our face detector	109	95.7
BioID	our face detector	178	99.3
BioID	Jesorsky *et al.* [16]	Not reported	91.8
BioID	Hamouz *et al.* [17]	25	97.6
BioID	Ramírez and Fuentes [18]	5240	95.0

B. Performance of the Gender Classifier

To evaluate the gender classifier, 9100 face images with 4820 males and 4280 females from Phung *et al.* database [12] were used in a five-fold cross validation procedure, where in each fold, 7280 face images were gathered for training and 1820 face images for testing. To increase the number of samples in the training set, all the 7280 face images were folded along the Y-axis. For comparison with other published results, the gender classifier was also tested on 1,762 face images (1,150 males and 612 females) from the FERET database, using five-fold cross validation. These images were histogram equalized so as to normalize the contrast of the face image. The experimental results based on these two databases are listed in Table II. Tested on the segmented face images from the Phung *et al.* database, the gender classifier achieves a classification rate of 87.7%. This result shows that not only can SICoNNets be used to localize faces but also to perform gender classification on misaligned face images. On frontal face images obtained from the FERET database, the classification rates of the gender classifier are 97.5% for male and 93.3% for female using a small network with six feature maps (Net-A): two feature maps in the first layer and four in the second layer. However, when using a network structure with four feature maps in the first hidden layer and eight feature maps in the following layer (Net-B), the overall classification accuracy on the FERET images increases to 97.2%.

TABLE II

CLASSIFICATION PERFORMANCE OF THE GENDER CLASSIFIER.

Face database	Classifier ID	Classification rates (%)		
		Male	Female	Total
Phung [12]	SICoNNet Net-A	86.2	89.3	87.7
FERET [19]	SICoNNet Net-A	97.5	93.3	96.0
FERET [19]	SICoNNet Net-B	97.6	96.4	97.2
FERET	SVM [7] Moghaddam-Yang	97.95	95.21	96.62

C. Performance of the Gender Recognition System

In this experiment, the gender recognition system is tested on digital images collected from the Web and BioID database. Here, the ability of the system to automatically detect and recognize the gender of the face is tested. The recognition performance of the system is given in Table III; some sample output images are shown in Fig. 5. From a set of 502 images collected from the Web, the gender recognition system detects 485 faces (115 males and 370 females) and achieves recognition rates of 80.9% for male and 84.6% for female. On the BioID database with 968 male and 543 female faces detected, the recognition rates for male and female are 89.2% and 81.4%, respectively. These experiments demonstrate that the proposed gender classification system, based on a unified framework of convolutional neural networks, can perform face detection and gender recognition.

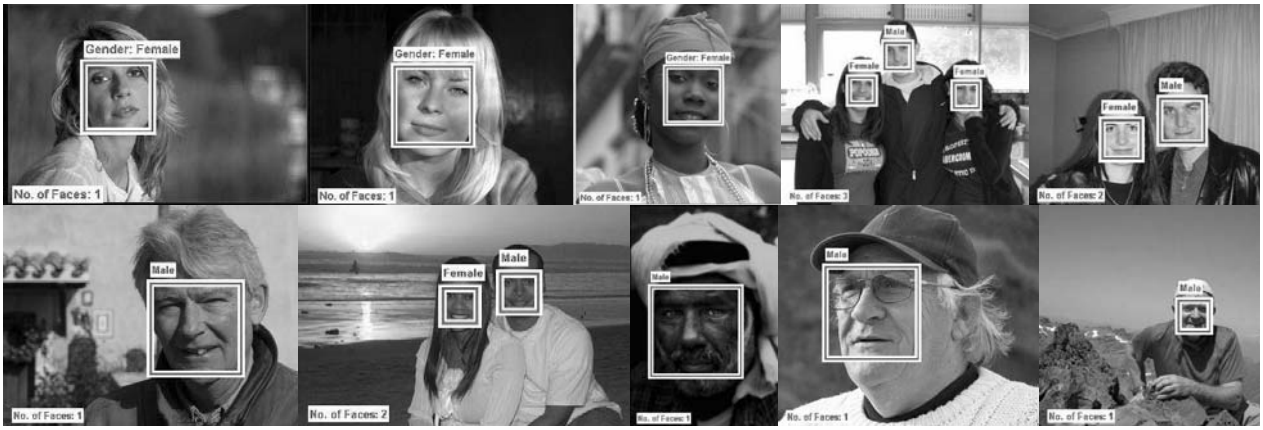

Fig. 5. Samples of output images from the gender recognition system.

TABLE III
PERFORMANCE OF THE GENDER RECOGNITION SYSTEM.

Test Images	Recognition rates		
	Male	Female	Total
Web	93 (80.9%)	313 (84.6%)	83.7%
BioID	863 (89.2%)	442 (81.4%)	86.4%
Average classification accuracy			85.7%

VI. CONCLUSION

In this paper, we developed an automatic gender recognition system based on shunting inhibitory convolutional neural networks. The system consists of a face detector and a gender classifier. The face detection system processes the entire input image and generates a location map of the detected faces. The detected faces are then passed to the gender classifier, which has the same network structure as the face detector. Tested on the BioID face database, the accuracy of the face detector was 99.3%. On face images taken from FERET database, the gender classifier achieved an accuracy of 97.2%. Tested on images collected from the Web and the BioID face database, the gender recognition system achieves an overall average recognition rate of 85.7%.

REFERENCES

[1] B. Golomb, D. Lawrence, and T. Sejnowski, "Sexnet: a neural network identifies sex from human faces," *Advances in Neural Information Processing Systems*, pp. 572–577, 1991.

[2] R. Brunelli and T. Poggio, "Hyberbf networks for gender classification," in *Proc. DARPA Image Understanding Workshop*, 1992, pp. 311–314.

[3] Z. Sun, X. Yuan, G. Bebis, and S. Louis, "Neural-network-based gender classification using genetic eigen-feature extraction," in *Proc. of the International Joint Conference on Neural Networks*, vol. 3, 2002, pp. 2433 – 2438.

[4] K. Balci and V. Atalay, "PCA for gender estimation: which eigenvector contribute?" in *Proc. of the Sixteenth International Conference on Pattern Recognition*, vol. 3, 2002, pp. 363–366.

[5] B. Wu, H. Ai, and C. Huang, "LUT-based adaboost for gender classification," in *Proc. of the Fourth International Conference on Audio- and Video-Based Biometrie Person Authentication*, 2003, pp. 104–110.

[6] A. Jain and J. Huang, "Integrating independent components and linear discriminant analysis for gender classification," in *Proc. of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 159–163.

[7] B. Moghaddam and M.-H. Yang, "Learning gender with support faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 707–711, 2002.

[8] F. H. C. Tivive and A. Bouzerdoum, "Efficient training algorithms for a class of shunting inhibitory convolutional neural networks," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 541– 556, 2005.

[9] L. J. Borg-Graham, C. Monier, and Y. Fregnac, "Visual input evokes transient and strong shunting inhibition in visual cortical neurons," *Nature*, vol. 393, no. 6683, pp. 369–373, 1998.

[10] J. S. Anderson, M. Carandini, and D. Ferster, "Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex," *Journal of Neurophysiology*, vol. 84, pp. 909–926, 2000.

[11] Y. Frégnac, C. Monier, F. Chavane, P. Baudot, and L. Graham, "Shunting inhibition, a silent step in visual computation," *Journal of Physiology*, vol. Paris 97, pp. 441–451, 2003.

[12] S. L. Phung, A. Bouzerdoum, and D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 148 – 154, 2005.

[13] N. Ampazis and S. J. Perantonis, "Two highly efficient second-order algorithms for training feedforward networks," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1064–1074, 2002.

[14] M. T. Hagan and M. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, pp. 989–993, 1994.

[15] M. Delakis and C. Garcia, "Robust face detection based on convolutional neural networks," in *Proc. of the Second Hellenic Conference on Artificial Intelligence*, Thessalonique, Greece, 2002, pp. 367–378.

[16] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, "Robust face detection using the hausdorff distance," in *Proc. of the Third International Conference Audio- and Video-Based Biometric Person Authentication*, Halmstad, Sweden, 2001, pp. 90–95.

[17] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paa;anen, and H. Kälviäinen, "Affine-invaraint face detection and localization using gmm-based feature detector and enhanced appareance model," in *Proc. of the Sixth International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 67–72.

[18] G. A. Ramírez and O. Fuentes, "Face detection using combinations of classifiers," in *Proc. of the Second Canadian Conference on Computer and Robot Vision*, 2005, pp. 610 – 615.

[19] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.