20-8-2006

A shunting inhibitory convolutional neural network for Gender Classification

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

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

A shunting inhibitory convolutional neural network for Gender Classification

Abstract

Demographic features, such as gender, are very important for human recognition and can be used to enhance social and biometric applications. In this paper, we propose to use a class of convolutional neural networks for gender classification. These networks are built upon the concepts of local receptive field processing and weight sharing, which makes them more tolerant to distortions and variations in two dimensional shapes. Tested on two separate data sets, the proposed networks achieve better classification accuracy than the conventional feedforward multilayer perceptron networks. On the Feret benchmark dataset, the proposed convolutional neural networks achieve a classification rate of 97.1%.

Disciplines

Physical Sciences and Mathematics

Publication Details

A Shunting Inhibitory Convolutional Neural Network for Gender Classification

Fok Hing Chi Tivive and Abdesselam Bouzerdoum, *Senior Member, IEEE*
School of Electrical, Computer and Telecommunications Engineering
University of Wollongong
Northfields Avenue, Wollongong, NSW 2522, AUSTRALIA.
Email:[tivive@uow.edu.au], [a.bouzerdoum@ieee.org]

Abstract

Demographic features, such as gender, are very important for human recognition and can be used to enhance social and biometric applications. In this paper, we propose to use a class of convolutional neural networks for gender classification. These networks are built upon the concepts of local receptive field processing and weight sharing, which makes them more tolerant to distortions and variations in two dimensional shapes. Tested on two separate data sets, the proposed networks achieve better classification accuracy than the conventional feedforward multilayer perceptron networks. On the Feret benchmark dataset, the proposed convolutional neural networks achieve a classification rate of 97.1%.

1. Introduction

The gender classification task was initially considered in psychophysical studies only, in the context of understanding the human visual processing and identifying the key facial features that distinguish male from female individuals. More recently, however, it has attracted a great deal of attention from the computer vision community; the disparity between facial femininity and masculinity is an important piece of information that can be used to improve the performance of other applications, such as face recognition and intelligent human-computer interfaces.

To the best of our knowledge, the most successful gender classification technique reported to date was proposed by Moghaddam and Yang [4] using support vector machines (SVMs). The authors evaluated their SVM-classifier on 1755 face images of size 21×12 from the FERET database. The images were first preprocessed for contrast and geometric shape variations, and then the face region was masked to remove the hair information. Based on five-fold cross validation and using 20% of the training set as support vectors, they obtained a classification rate of 96.6%. Instead of using

a single classifier, Wu *et al.* [9] employed Adaboost learning algorithm to select a series of look-up table (LUT) weak classifiers for gender classification. The training procedure was performed on a large training set of more than 11,000 face images, and their Adaboost-classifier achieved 88% accuracy with 200 selected LUTs. Golomb *et al.* [2] used a two layer neural network, known as SEXNET, to discriminate between male and female classes and obtained a classification accuracy of 91.9% based on a database containing 90 images. To alleviate the effect of the "curse of dimensionality", some researchers have combined pre-processing techniques with neural network (NN) classifiers. In [7], Sun *et al.* used principle component analysis (PCA) to compress the input space into low dimensional features (i.e, eigenvectors), and applied a genetic algorithm (GA) for selecting a set of eigen-vectors to represent the two classes. Tested on frontal face images, their hybrid technique achieved a classification rate of 88.7% on a test set of 400 images. The methods presented so far have proven that NNs can be used for the gender classification task. However, most of these methods include several pre-processing stages, in order to reduce the dimensionality of the input space and the influence of some external factors, such as the orientation of the face pattern and variations in lighting conditions.

In this article, we apply shunting inhibitory convolutional neural networks (SICoNNets) [8] for gender classification. In general, convolutional neural networks (CoNNs) have the ability to perform feature extraction and classification within the network structure through learning and achieve a certain degree of invariance while preserving the spatial topology of the input data. Moreover, a certain amount of prior knowledge about the task is incorporated into the network. This makes SICoNNets more robust to geometric distortions and other 2-D shape variations. The remainder of this paper is organized as follows. Section 2 presents the SICoNNet architectures used for gender classification. The gender database and the evaluation procedure are explained in Section 3. Experimental results are presented in Section 4, followed by concluding remarks in

Section 5.

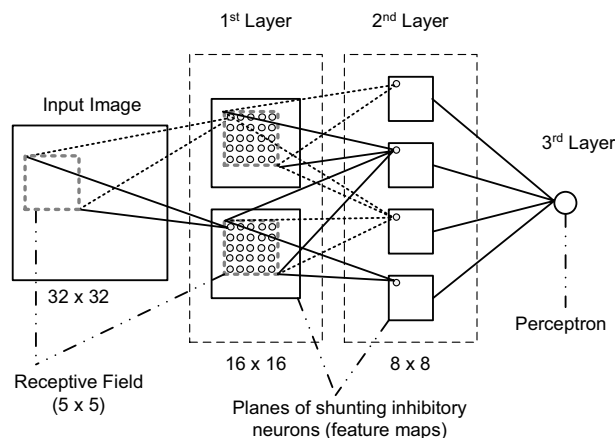## 2. Convolutional Neural Network Models

Recently, we developed three generic convolutional neural network architectures that use a special type of processing element known as *shunting inhibitory neuron* [8]. This kind of neuron has been applied in other NN models for supervised pattern classification and regression and has been shown to be more powerful than the sigmoid neuron or perceptron [1]. Here, two CoNN structures are developed for gender classification, namely the binary- and toeplitz-connected architectures. Each network consists of three processing layers: two hidden layers and one output layer. The input layer is a two-dimensional (2-D) image of size 32×32, acting as the network retina receiving inputs from the environment. The hidden neurons are organized into planes called *feature maps*. The first hidden layer can have any number of feature maps, but the second layer has twice as many feature maps as the first one. The feature maps are sub-sampled by a factor of two; that is, the feature maps in the first hidden layer are of size 16×16 and those in the second layer are of size 8×8. Each hidden neuron is connected locally to a small input region of size 5×5, called the *receptive field*. All neurons from the same feature map receive inputs from different input regions so as to cover the entire input plane, but share the same set of synaptic weights (weight sharing); however, different feature maps in the same hidden layer use different set of weights. The weight sharing strategy has two advantages: (i) it eases the curse of dimensionality as each neuron only has to concern itself with the processing of a limited amount of information; (ii) each neuron behaves as a feature detector which extracts the same elementary visual feature from different positions in the input plane.

To connect the feature maps of one layer to the next layer, two partial-connection schemes have been adopted: the binary- and toeplitz-connection schemes. In the binary scheme, each feature map is connected to two feature maps in the succeeding layer, similar to a binary tree. In the toeplitz scheme, each feature map may have one-to-one or one-to-many links with feature maps of the preceding layer; however, the connection pattern between two successive layers has the form of a toeplitz matrix. For example, the toeplitz-connections between the two feature maps of the first hidden layer (L1) and the four feature maps of the second hidden layer (L2) are shown in Table 1. Suppose the four feature maps of L2 are labeled 1 to 4 and the two feature maps of L1 are labeled A and B. Feature maps 1 and 4 form one-to-one connections with feature maps A and B, respectively, whereas feature maps 2 and 3 are both connected to feature maps A and B. In other words, the connections appear along the diagonal of the connection matrix in the form

Table 1. Connections between feature maps in L1 and L2 of the toeplitz-connected CoNN.

L2 Feature Map	Connections from L1 to L2		
1	A		
2	B	A	
3		B	A
4			B

of a Toeplitz matrix. Figure 1 presents a schematic diagram of the toeplitz-connected network with one output neuron.

Figure 1. The network structure of a three-layer toeplitz-connected CoNN.

The novelty of this CoNN in comparison to earlier convolutional models is that the feature map is made up of shunting inhibitory neurons. The neural activity of a shunting inhibitory neuron is given by

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

where

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

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

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

The parameters $C_{L,k}$ and $D_{L,k}$ are the set of weights, $b_{L,k}$ 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, F_L is the size of the feature map at the Lth layer, and S_{L-1} is the number of feature maps in the $(L-1)^{th}$ layer to which the neuron is connected. All the neurons in a feature map share the same bias parameters

and passive decay term. This implies that a feature map consists of one shunting inhibitory neuron which replicates itself into a 2-D array.

At the output layer, there is one perceptron or sigmoid neuron, whose output is used to classify the 2-D input pattern into male or female face. The output neuron receives inputs from all feature maps of the last hidden layer, i.e., the output neuron is fully connected to all the neurons in layer L2. The response of the output neuron is given by

$$y = h\Big(\sum_{i=1}^{S_N} w_i z_i + b\Big), \qquad (2)$$

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

3. Gender Databases and Evaluation Procedure

The classification performance of the trained CoNNs and MLPs were evaluated on two face databases based on the five-fold cross validation. The first database (DB-1) is the Feret benchmark database [5] with 1152 male and 610 female of up-right frontal face images. The second database (DB-2) is the face and skin detection database created by Phung *et al.* [6], where 8000 face images (4000 males and 4000 females) were used for training and testing; these face images are slightly rotated and shifted in position, and surrounded by other background information, as shown in Fig. 2. Before the evaluation of the classification performances of the different networks, some preliminary experiments had been conducted, which showed that using face images folded along the Y-axis (mirrored face images) in the training and test sets improved the classification performance. Therefore, all face images are folded along the Y-axis to double the number of patterns in the training set; during the testing phase, each face image is evaluated by the trained network including its mirrored version, and the average of both network responses is taken as the final classification score. All the face patterns in the training and test sets were histogram equalized and the image pixels were scaled to the range $[-1, 1]$. The desired outputs corresponding to the male and female images were labeled 1 and -1, respectively.

4. Experimental Results

To determine an appropriate size of the network for this classification task, several CoNNs with different number of feature maps in the hidden layers were trained and tested,

(a)

(b)
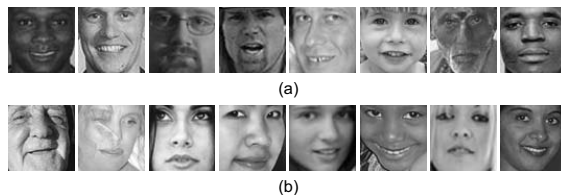
Figure 2. Face patterns from the face database [6]: (a) male patterns and (b) female patterns.

ranging from 2 to 4 feature maps in L1 and 4 to 8 feature maps in L2. In the first hidden layer, the activation functions g_L and f_L are the hyperbolic tangent and exponential functions, respectively. In L2, g_L is a logarithmic sigmoid function, whereas a hyperbolic tangent is used at the output. During the training and initialization of the CoNNs, the denominator of (1) is constrained to be greater than zero so as to prevent division by zero. Hence, the passive decay rate, $a_{L,k}$, is constrained as follows:

$$a_{L,k}(i,j) + Y_{L,k}(i,j) \geq \varepsilon > 0, \qquad (3)$$

where ε is a small positive constant.

For comparison purposes, several MLPs were also developed for gender classification, where the number of hidden neurons in each layer matches the number of feature maps used in the CoNNs. The input size to the MLPs is the same as that of the CoNNs, but arranged in a form of column vector of length 1024. The output layer has one neuron, and the activation functions used across all layers are the hyperbolic tangent function. The trainable weights of the CoNNs and MLPs are adapted by the Levenberg-Marquardt algorithm, in which the Hessian matrix and the gradient vector are computed, in terms of the Jacobian matrix, using the modified backpropagation algorithm proposed by Hagan [3].

Table 2 and 3 present the classification rates of the binary- and toeplitz-connected networks, respectively, and Table 4 lists the performances of the MLPs. The number of weights for each trained network is given in the second column of the tables, followed by the number of feature maps (or perceptrons) used in each hidden layer. The correct classification rates for males (M) and females (F) and the total classification rate (T. R.) based on the two face databases are listed in the remaining columns. On the Feret database, the CoNNs achieve classification rates between 95.9% and 97.1%, whereas the MLPs achieve classification rates between 95.1% and 95.6%. The highest classification accuracy of 97.1%, achieved by the binary-connected CoNN BC-03, is greater than that achieved by the state-of-the-art SVM-based system of Moghaddam and Yang [4], which achieves a classification rate of 96.6%. On the second database, our CoNNs achieve classification rates in the range

[87.8%, 88.8%], whereas MLPs achieve classification rates in the range [82.7%, 83.8%]. These results show that not only do the CoNNs have the capability to discriminate face image between a male and a female, but also they are robust against slight variations and distortions in the input image. Furthermore, MLPs have significantly more weights than CoNNs. For example, a CoNN with 6 feature maps has 575 weights, whereas an equivalent MLP, with 2 hidden units in the first layer and 4 in the second layer, has 2067 weights, which is 3.6 times more weights than the CoNN. With an additional 3 hidden units, the MLP (MLP-02) has a significant increase of 1039 weights. This is because each neuron in the MLP is fully connected with all the neurons in the previous layer where each connection is a trainable weight. In contrast, the CoNN encompasses a weight sharing mechanism, which constrains the same set of weights to be shared across all neurons within the same feature map.

Table 2. Gender classification performance of the binary-connected networks (BCs) based on Feret and Phung databases.

Network Index	No. weights	No. Feat. Maps		Classification rate (%)					
		L1	L2	DB-1 (Feret)			DB-2 (Phung)		
				M	F	T. A.	M	F	T. R.
BC-01	575	2	4	97.1	94.3	96.1	86.1	90.2	88.1
BC-02	862	3	6	97.1	93.6	95.9	88.4	87.8	88.1
BC-03	1149	4	8	98.6	94.3	97.1	87.0	90.3	88.7

Table 3. Gender classification performance of the toeplitz-connected networks (TCs) based on Feret and Phung databases.

Network Index	No. weights	No. Feat. Maps		Classification rate (%)					
		L1	L2	DB-1 (Feret)			DB-2 (Phung)		
				M	F	T. A.	M	F	T. R.
TC-01	575	2	4	96.9	94.3	96.0	85.9	89.7	87.8
TC-02	862	3	6	97.1	95.4	96.5	88.7	88.9	88.8
TC-02	1149	4	8	97.6	94.3	96.4	87.0	89.9	88.4

Table 4. Gender classification performance of the MLPs based on Feret and Phung databases.

Network Index	No. weights	No. F. maps		Classification rate (%)					
		L1	L2	DB-1 (Feret)			DB-2 (Phung)		
				M	F	T. A.	M	F	T. R.
MLP-01	2067	2	4	96.6	92.1	95.1	80.5	86.0	82.7
MLP-02	3106	3	6	96.7	92.5	95.2	81.5	86.0	83.8
MLP-03	4149	4	8	96.8	93.4	95.6	81.8	86.0	83.6

5. Conclusion

In this paper, shunting inhibitory convolutional neural networks are used to solve the gender classification problem. Two different architectures were developed for gender classification: the binary- and toeplitz-connected architectures. Experiments were conducted on two datasets: the Feret and Phung *et al.* databases. On the Feret database, the CoNNs perform as well as the state-of-the-art gender classification systems, such as SVM-based systems, with much less pre-processing. In fact, the proposed CoNNs achieve the highest classification rate of 97.1%. In comparison to MLPs, the CoNNs have better performance and fewer trainable parameters. When tested on images from the Phung database, the CoNNs achieve classification rates 5% higher than MLPs. The superior performance of the proposed CoNNs over the MLPs demonstrate that CoNNs are more robust to geometric and photometric distortions.

References

[1] G. Arulampalam and A. Bouzerdoum. Application of shunting inhibitory artificial neural networks to medical diagnosis. In *Proc. of the Seventh Australian and New Zealand Intelligent Information Systems Conference*, pages 89–94, Perth, 2001.

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

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

[4] B. Moghaddam and M.-H. Yang. Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):707–711, 2002.

[5] 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*, 22(10):1090–1104, 2000.

[6] 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*, 27(1):148 – 154, 2005.

[7] 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*, volume 3, pages 2433 – 2438, 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*, 16(3):541–556, 2005.

[9] 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*, pages 104–110, 2003.

COMPUTER SOCIETY