"Tailored" Neural Networks to Improve Image Classification

Lerner, B., Guterman, H., Dinstein, I. and Romem, Y.*
Department of Electrical and Computer Engineering
Ben-Gurion University of the Negev
Beer-Sheva, Israel 84105

* The Institute of Medical Genetics, Soroka Medical Center
Beer-Sheva, Israel 84105

Abstract

The concept of "tailored" neural network is inspired by the concept of grouping in the visual cortex of the mammalian brain. This biological animated concept was implemented to develop "tailored" neural networks for image classification improvement. Each "tailored" network was specialized to classify a different class of vectors. This was done by employing separate training and using specific features in each class. Image classification improvement was tested by the chromosome classification application. For chromosome classification, the probability of correct classification using the "tailored" networks was 2.5% higher than the probability achieved by a conventional neural network (97.6% versus 95.1%). This improvement was found to be higher when lower quality features were employed. It is expected that the improvement will increase whenever the image classification task will become more and more complicated.

1. Introduction

Image classification using multilayer perceptron (MLP) neural networks has become widespread in the computer vision and neural networks communities. The neural network classifier has the advantage of being fast (highly parallel), easily trainable and capable of creating arbitrary partitions of feature space. However, image classification using an MLP depends on a series of various procedures generally held according to practical considerations. In most vision applications these stages precede the classification itself and are motivated by a mathematical analysis and/or engineering concepts. Even the MLP classifier itself, when applied to a complicated classification task fails very often to correctly classify the input data. The mammalian visual cortex seems not to suffer from this kind of problems. It simply does not function as our classical image classifiers do. Image projected from the retina onto the visual cortex parallelly spread among a series of cell clusters, each of which performs its own special analysis and synthesis. Each retinal area is analyzed over and over again, column after column, and again in neighboring cortical regions, with respect to a number of different variables such as position, orientation and color [2]. From a large series of experiments, it became apparent that in area 17 simple and complex neurons with similar receptive field axis orientation are neatly stacked on top of each other in discrete columns. *Separate* columns exist for each axis orientation. Other functional variables are also grouped in columnar aggregates of cells. In cortical areas of the monkey beyond area 17 of the visual cortex, there exist columns of cells with well-defined color sensitivity and other columns in which the direction of movement of the visual stimulus is important. Cortical structure and functional organization go hand in hand [2].

This work was supported in part by the Paul Ivanier Center for Robotics and Production Management, Ben-Gurion University, Beer-Sheva, Israel.

This study suggests imitating several aspects of the mammalian visual cortex structure and functions in order to improve image classification. We introduce the "tailored" neural network that gains class separability by both extracting specific class features and by tuning the training to a specific class only. The discrete cortex columns are implemented through the discrete "tailored" neural networks.

To implement the suggested concept we have used the chromosome classification task as a model for image classification. The chromosome classification task is well known, yet without satisfying solutions [1], [10], [13]. Several studies on chromosome classification using neural networks were held in the last two years, most of them relate to feature extraction and selection [3]-[6] and other relate to the performance and the optimization of the MLP NN as chromosome classifier [7]-[9]. Therefore, this study can be also regarded as a continuation study, in the way to establish a neural network classifier for human chromosome.

2. The "tailored" network

The "tailored" network tries to mimic the outlined biological concept and to optimize its performance. By training an MLP NN to classify *only* vectors of *one* class we enable (and actually force) it to employ all the relevant information of this class feature data in order to perfectly distinguish between the class and the rest of the classes. The network does not need to distinguish vectors of several classes but only "the" class vectors compare to "other" classes vectors. By training several networks, each of them specialized on vectors of a different class, we can get a special network for each class. When a test vector will be introduced to *all* the "tailored" networks, the network with the highest output value will indicate the correct classification of this vector (a kind of "winner-take-all" mechanism). Decomposition of the task to several simpler tasks makes the solution easier and yields better performance, in *vivo,* as well as, while solving difficult engineering problems. Furthermore, it is known that a large network may perform perfectly on the training set but fails to interpolate as well as a smaller network.

One step further in the trail to mimic the visual cortex functioning could be training and testing the network using the best selected features of each class. In a similar way to the visual cortex (e.g., area 17) [2], we can extract features, optimize their selection and train, for a specific task, a network based on these features. This work examined the use of the "knock-out" algorithm [12] to select the best features to represent each class. Each class was represented by a *different* set of *optimal* (in the sense of the "minimum variance" [12]) features. Every vector, in each specific class, either a training or a test vector, was represented by this set of specific optimal features of *its* class. Training and testing the "tailored" networks was done based on these "class optimized" sets of features.

3. The methodology

Our data set included 481 chromosome images of types "2", "4", "13", "19" and "x". The features that were used for the classification were based on the density profile of the chromosomes. The density profile extracts the typical banded structure of the chromosomes and yields a suitable representation for the classification [1], [5], [6], [10]. The density profile was normalized both in length (64-dimensional) and in value to the [-0.5,0.5] range [6].

Each "tailored" network is based on a two-layer feedforward neural network trained by the backpropagation (bp) learning algorithm [11]. The number of input units was set by the feature space (64-dimensional) while the number of output units was determined by the number of classes to be classified (5-dimensional). The number of hidden units of the network was set according to the Principal Component Analysis (PCA), applied to the feature vectors. The number was set to be the number of the largest eigenvalues, the sum of which accounts for more than a pre-specified percentage of the sum of all the eigenvalues. In all the simulations, this number was set according to a pre-specified percentage of 90%.

Optimization of the neural network parameters regarding the chromosome data was made elsewhere [7]-[8]. Learning rate (μ) was set to be 0.026, momentum constant (α) to be 0.97 and the training cycle was set to be 4000 epochs.

Figure 1 sketches the procedure we have developed to examine "tailored" network performance. First, we used the "knock-out" algorithm to select the optimal features (in the sense of "minimum variance") in *each* class. We chose the best 40 features, although even a smaller number of features is adequate to represent the chromosomes without performance degradation [5]. Second, we partitioned the chromosome feature data set into training and test sets, where 80% of the vectors were chosen randomly to be in the training set. Then, we trained separately the 5 different networks based on the specific class features. Each training vector had 5 different representations each of them was used in the specific network. The desired output of training vectors belonged to the correct specific class was set to "1" while this output for *all* the rest of the vectors was set to "0". After training, vectors from the test set were tested by *all* the networks at once (test vectors had 5 different representations, as well). The largest output among these networks indicates the right class of the tested vector (winner-take-all). All the simulations were repeated 3 times, with the same network parameters but with different sets of randomly chosen training vectors, and the results were averaged. This procedure yields specialized networks we have, therefore, called "tailored" networks.

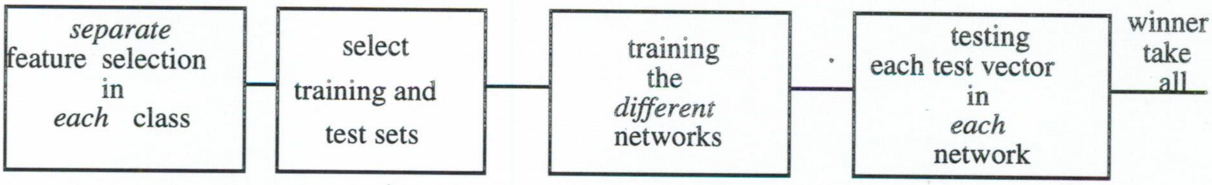

Figure 1. Implementation of a "tailored" network.

4. Simulation results

Figure 2 and Figure 3 show, respectively, the probability of correct training set classification and the probability of correct test set classification of the "tailored" networks. Training, as expected, is perfect in (almost) all the classes. Non-perfect training exists in classes with small training sets and/or in classes with considerable variations of image representation. In Figure 3, each network tests vectors according to the optimal features of the class it represents. The only significant columns in the Figure are the highest ones (winner-take-all), where there is an agreement between network type and the tested vector representation.

למה output 5-מימדי ולא 2-מימדי (או חד-מימדי)

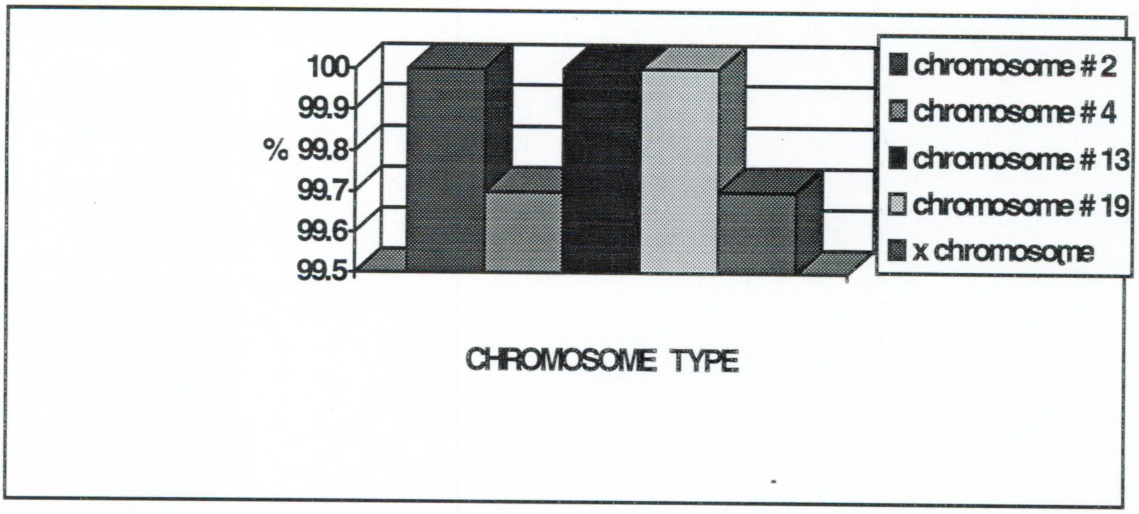

Figure 2. The probability of correct training set classification using the 5 "tailored" networks.

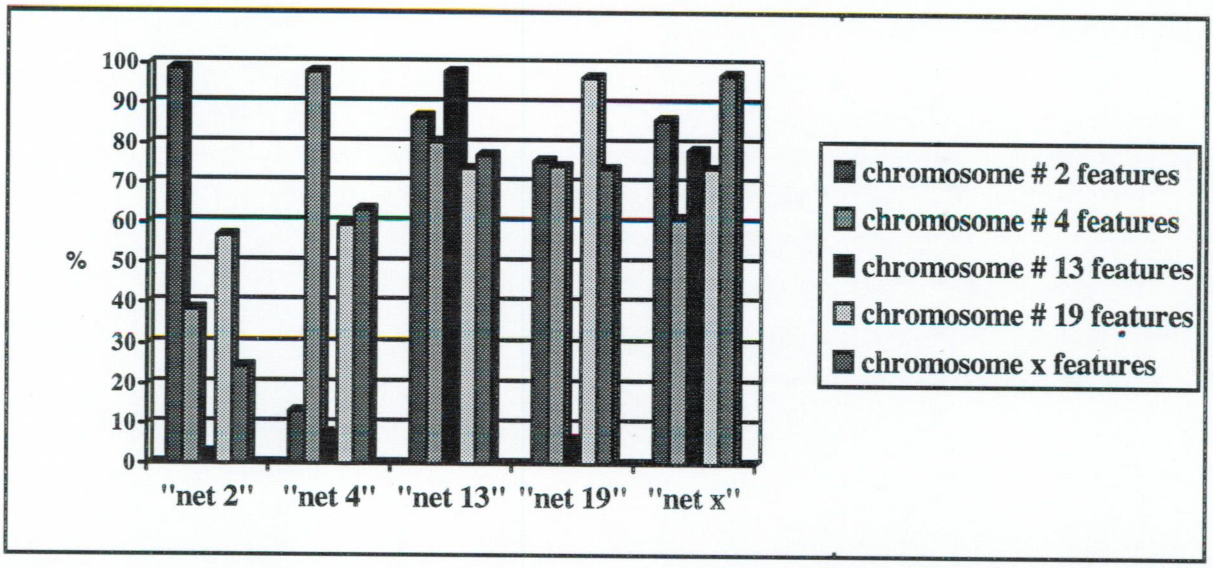

Figure 3. The probability of correct test set classification using the 5 "tailored" networks.

We have compared the classification results using these nets to the results when only *one* MLP network is trained to classify the chromosomes and the features are *common* to all classes. The 5 "tailored" networks yielded *average* probability of correct training set classification of 99.88% compare to 99.3% of the "one" network [6] and 97.6% with the test set compares to 95.1% of the "one" network [6]. When the simulation was repeated on slightly better features, the improvement due to the use of "tailored" networks, was slightly lower. Therefore, the advantage of using "tailored" networks is greater whenever "low quality" features are considered. The superiority of the "tailored" networks comparing to the "one" MLP NN is assumed to become even higher when the classification task will be complicated. Consequently, when all the 24 chromosome types will be considered, the "tailored" networks are believed to perform even better. However, not to be forgotten, that time and memory requirements are multiplied by the number of classes to be classified.

5. Discussion

The concept of "tailored" neural networks to improve image classification performance is introduced. Each "tailored" neural network is specialized to classify vectors of a different class. This was done by employing separate training and using specific features in each class. For chromosome classification, the probability of correct classification using the "tailored" networks was 2.5% higher than the probability achieved by a conventional neural network that uses common features in all the classes (97.6% versus 95.1%). This improvement was found to be higher when lower quality features were employed. It is expected that the improvement will increase whenever the image classification task will become more and more complicated.

6. References

1. Groen, F.C.A., ten Kate, T.K., Smeulders, A.W.M. & Young, I.T. (1989). Human chromosome classification based on local band descriptors. *Pattern Recognition Letters*, **9**, 211-222.
2. Kuffler, S.W, Nicholls, J.G & Martin, A.R. (1984). *From neuron to brain.* Sinauer Associates Inc.
3. Lerner, B., Guterman, H. & Dinstein, I. (1992). On classification of human chromosomes. *Neural Networks for Learning, Recognition and Control*, a research conference at Boston University, May 14-16.
4. Lerner, B., Guterman, H., Dinstein, I. & Romem, Y. (1993). Classification of human chromosomes by two-dimensional Fourier transform components. *WCNN'93*, Portland, July 11-15, 793-796.
5. Lerner, B., Levinstein, M., Rosenberg, B., Guterman, H., Dinstein, I. & Romem, Y. (1993). Feature selection and chromosome classification using an MLP neural network. (Submitted to *ICNN'94*).
6. Lerner, B., Guterman, H., Dinstein, I. & Romem, Y. (1993). Medial Axis Transform based features and neural network classifier for human chromosome classification. (Submitted for publication).
7. Lerner, B., Guterman, H., Dinstein, I. & Romem, Y. (1993). Human chromosome classification using multilayer perceptron neural network. (Submitted for publication).
8. Lerner, B., Guterman, H., Dinstein, I. & Romem, Y. (1993). Learning curves and optimization of multilayer perceptron neural network for chromosome classification. (Submitted to *WCNN'94*).
9. Lerner, B., Guterman, H., Dinstein, I. & Romem, Y. (1993). A Comparison of Multilayer Perceptron Neural Network and Bayes Piecewise Classifier for Chromosome Classification. (Submitted to *ICNN'94*).
10. Piper, J., Granum, E., Rutovitz, D. & Ruttledge, H. (1980). Automation of chromosome analysis. *Signal Processing*, **2**, 203-221.
11. Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning internal representations by error propagation. In Rumelhart, D.E., McClelland, J.L. and the PDP research group, *Parallel Distributed Processing*, vol. 1, chap. 8, Cambridge: MIT Press.
12. Sambur, M.R. (1975). Selection of acoustic features for speaker identification. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **ASSP-23**, 176-182.
13. Wu, Q., Suetens, P. & Oosterlinck, A. (1987). Toward an expert system for chromosome analysis. *Knowledge-Based Systems*, **1**, 43-52.