A Perceptron Reveals the Face of Sex

**Michael S. Gray, David T. Lawrence,
Beatrice A. Golomb, and Terrence J. Sejnowski**

April 23, 1993

Technical Report INC-9303

Institute for Neural Computation
University of California, San Diego
9500 Gilman Drive
La Jolla, California 92093-0523

Michael S. Gray is with the Computational Neurobiology Laboratory at The Salk Institute, David T. Lawrence is with the Howard Hughes Medical Institute, Beatrice A. Golomb is with The Salk Institute, and Terrence J. Sejnowski is with The Salk Institute, Howard Hughes Medical Institute, and the Institute for Neural Computation at the University of California, San Diego.

Correspondence concerning this article may be addressed to: Michael S. Gray, Computational Neurobiology Laboratory, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037 (electronic mailing address: michael@helmholtz.sdsc.edu).

Copyright © 1993 by Michael S. Gray, David T. Lawrence, Beatrice A. Golumb, and Terrence J. Sejnowski

A Perceptron Reveals the Face of Sex

Michael S. Gray, David T. Lawrence, Beatrice A. Golomb, and Terrence J. Sejnowski

Howard Hughes Medical Institute
The Salk Institute
Computational Neurobiology Laboratory
10010 North Torrey Pines Road
La Jolla, CA 92037

Phone: (619) 453-4100 ext. 534
Fax: (619) 587-0417
e-mail: michael@helmholtz.sdsc.edu

April 23, 1993

1 Introduction and Methods

Recognizing the sex of conspecifics is important. Humans rely primarily on visual pattern recognition for this task. In this note we show that a perceptron using normalized grey levels as input can classify sex from facial images with 79% accuracy. The weights in this network reveal the contribution of each region of the face to sex recognition.[1]

Previous efforts to recognize sex from human faces have involved the development of complex systems utilizing first a compression network and then a classification network (Fleming & Cottrell, 1990; Golomb, Lawrence, & Sejnowski, 1991). A database of 90 faces was used (O'Toole, Millward, & Anderson, 1988) to train a simpler sex recognition network. The images were scaled to normalize the distance between the two eyes and the perpendicular distance between the eyes and mouth. All images were adjusted to the same average brightness. The original faces were randomly subsampled at 5 different resolutions (10 pixels x 10 pixels, 15 x 15, 22 x 22, 30 x 30, and 60 x 60) to produce 5 separate databases. A sample face is shown in Figure 1a. Because not all photos were exactly head-on, each database was doubled in size (to 180 faces) by flipping each face image horizontally and including this new image in the database. This procedure removed any systematic lateral differences which could have been exploited by the network.

The ability to learn to recognize sex from faces was investigated with network models using two different architectures: (1) a simple perceptron, and (2) a standard multi-layer perceptron (MLP). Both architectures were trained on each of the 5 databases of uncompressed facial images. In the simple perceptron model, the inputs (the face image) were directly connected to a single output unit. The MLP model included a layer of 10

---

[1] The task is referred to by other authors as "gender" recognition, but "gender" is properly a grammatical term: specifically, words have gender, people don't. In English, the term "gender" has attained increasing currency as a euphemistic substitute for (male vs female) "sex". However, "To talk of persons or creatures of the masculine or feminine g., meaning of the male or female sex, is either a jocularity (permissible or not according to context) or a blunder." (Fowler's Modern English Usage, Oxford University Press, 2nd ed. 1985, revised 1983).

hidden units between the input and output units. A jackknife training procedure was used. For each architecture at each resolution, 9 separate networks were trained (with conjugate gradient) on a different subset of the faces, with the remaining faces used to test generalization performance.

2 Results and Discussion

The simple perceptron and the MLP performed remarkably similarly across all databases (see Figure 1b). An ANOVA comparing the performance of the two architectures showed that they were not significantly different at any resolution ($p > .05$). Examination of the weights of one of the simple perceptron networks at the end of training (see Figure 1c) shows that the network's ability to discriminate the length and width of the nose, the shape of the cheeks, and the size of the mouth all contributed to correct performance.

Recent efforts to model human performance on sex recognition have been very successful. Golomb *et al* (1991) achieved peak model performance of 91.9% correct, compared to 88.4% for humans. In the present work, no compression was performed on the input images. The results show that a simple perceptron architecture was found to perform as well as a MLP. The high degree of similarity between the results of the two architectures suggests that a substantial part of the problem is linearly separable. This simple perceptron, with less than 2% of the number of parameters in the model by Golomb *et al* (1991), reached a peak performance level of 78.9% correct. The sex recognition task in humans may in fact be simpler than previously believed.

Acknowledgments

This study was supported by grants from the National Science Foundation and the Office of Naval Research.

References

Fleming, M., & Cottrell, G. W. (1990). Categorization of faces using unsupervised feature extraction. *Proceedings of IJCNN-90, 2,* 65-70.

Golomb, B. A., Lawrence, D. T., & Sejnowski, T. J. (1991). SexNet: A neural network identifies sex from human faces. In D. S. Touretzky & R. Lippman (Eds.), Advances in Neural Information Processing Systems 3. San Mateo, CA: Morgan Kaufmann.

O'Toole, A. J., Millward, R. B., & Anderson, J. A. (1988). A physical system approach to recognition memory for spatially transformed faces. *Neural Networks, 1,* 179-199.

Figure Captions

Figure 1. (a.) The face of a woman in the database. This image was sampled at 60 pixels by 60 pixels. (b.) Generalization performance for the simple perceptron and MLP networks. Each bar in the graph represents the average correct performance of 9 networks in the classification of faces in the test set. The black filled bars represent the simple perceptron networks, and the white unfilled bars represent the MLP networks. (c.) Connection weights for a simple perceptron (30 pixels x 30 pixels) at the completion of training. White squares represent positive weights (males) and black squares represent negative weights (females). The size of the square within each cell is proportional to the magnitude of the weight. The three black circles indicate the location of the center of the right and left eye and the mouth.

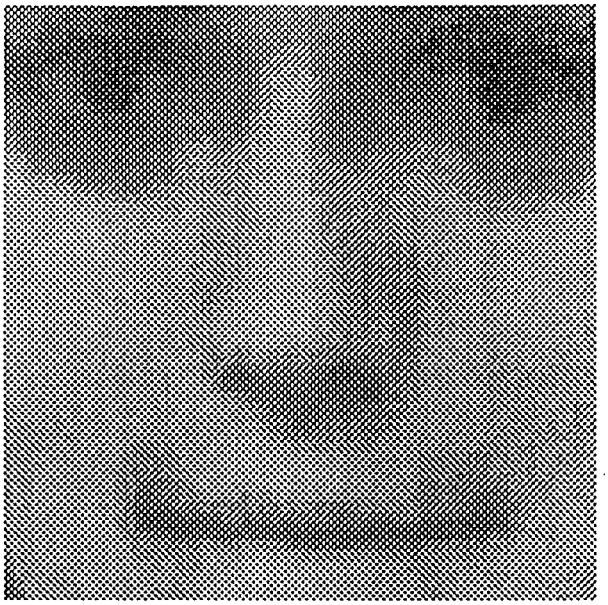

Figure 1. (a.)

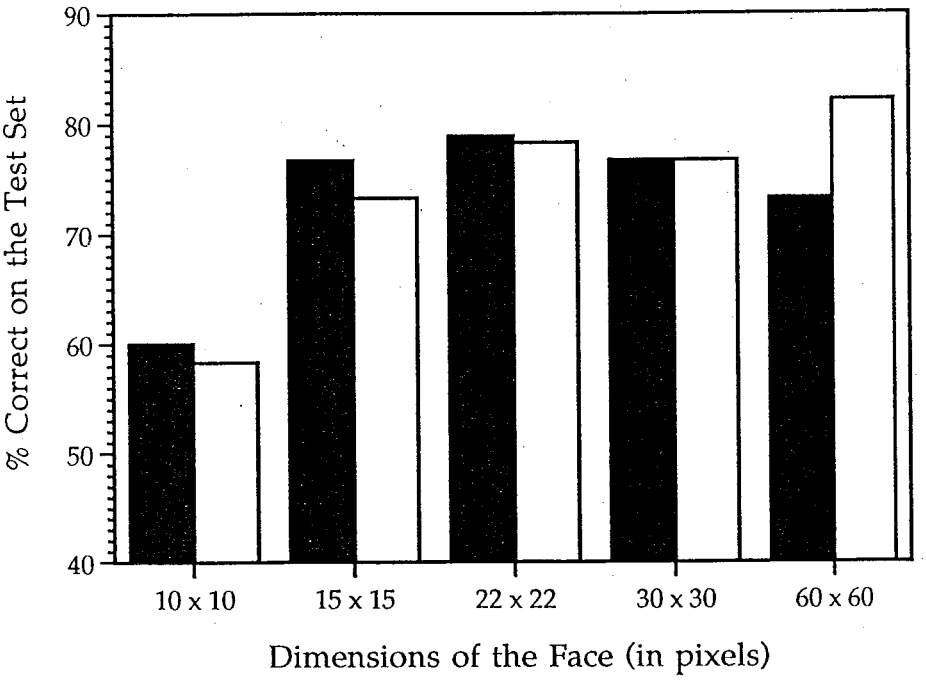

Figure 1. (b.)

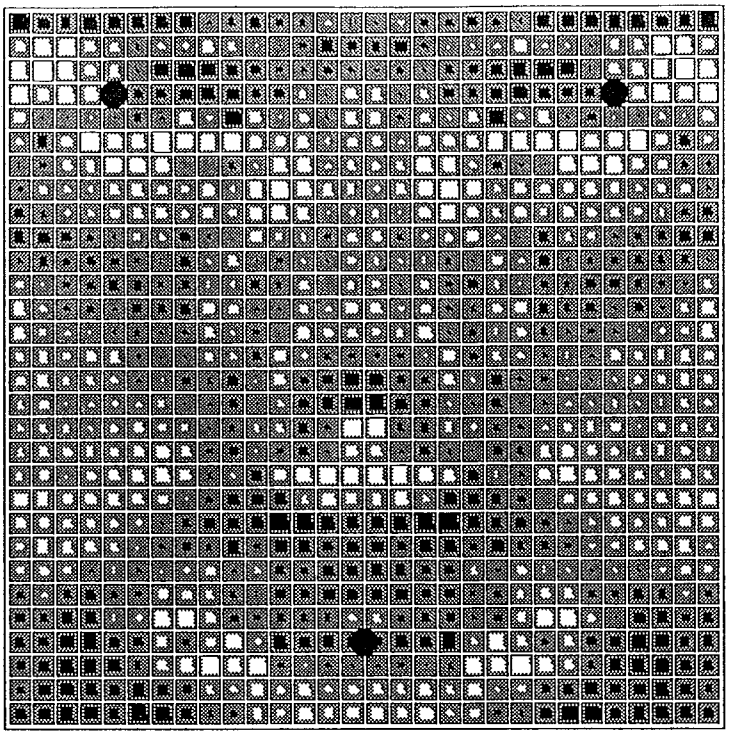

Figure 1. (c.)