In Touretzky, D. S., Lippman, R., (eds.) *Advances in Neural Information Processing Systems 3*, San Mateo, CA; Morgan Kaufmann, 1991.

SEXNET: A NEURAL NETWORK IDENTIFIES SEX FROM HUMAN FACES

B.A. Golomb, D.T. Lawrence, and T.J. Sejnowski
The Salk Institute
10010 N. Torrey Pines Rd.
La Jolla, CA 92037

Abstract

Sex identification in animals has biological importance. Humans are good at making this determination visually, but machines have not matched this ability. A neural network was trained to discriminate sex in human faces, and performed as well as humans on a set of 90 exemplars. Images sampled at 30x30 were compressed using a 900x40x900 fully-connected back-propagation network; activities of hidden units served as input to a back-propagation "SexNet" trained to produce values of 1 for male and 0 for female faces. The network's average error rate of 8.1% compared favorably to humans, who averaged 11.6%. Some SexNet errors mimicked those of humans.

1 INTRODUCTION

People can capably tell if a human face is male or female. Recognizing the sex of conspecifics is important. While some animals use pheromones to recognize sex, in humans this task is primarily visual. How is sex recognized from faces? By and large we are unable to say. Although certain features are nearly pathognomonic for one sex or the other (facial hair for men, makeup or certain hairstyles for women), even in the absence of these cues the determination is made; and even in their presence, other cues may override.

Sex-recognition in faces is thus a prototypical pattern recognition task of the sort at which humans excel, but which has vexed traditional AI. It appears to follow no simple algorithm, and indeed is modifiable according to fashion (makeup, hair etc). While ambiguous cases exist, for which we must appeal to other cues such as physical build (if visible), voice patterns (if audible), and mannerisms, humans are

fairly good in most cases at discriminating sex merely from photos of faces, without resorting to such adscititious cues. Can neural networks do the same?

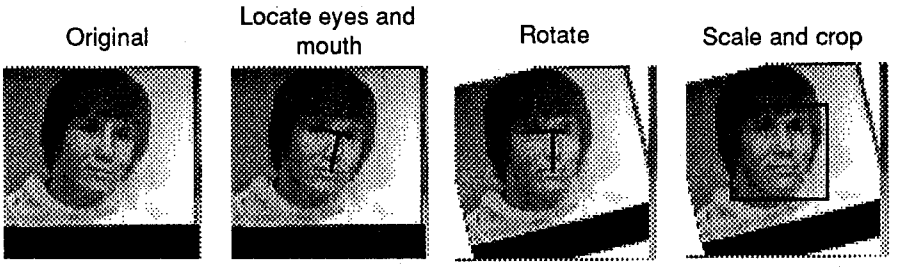

Figure 1: Preprocessing for all faces. After locating the eyes by hand, the image was rotated such that the line joining the eyes was horizontal. The distance between the eyes and the perpendicular distance to the mouth were scaled and the resulting image was cropped. Blocks of pixels were averaged to produce a final 30x30 subsampled image which served as input to the network.

2 METHOD

90 photos of young adult faces (45 male, 45 female), were used (O'Toole, Millward, & Anderson, 1988). Faces had no facial hair, no jewelry, and apparently no makeup. A white cloth was draped about each neck to eliminate possible clothing cues. Most photos were head on, but the exact angle varied.

Faces were rotated until eyes were level; scaled and translated to position eyes and mouth similarly in each image; and clipped to present a similar extent of image around eyes and mouth. Final faces were 30x30 pixels with 12 pixels between the eyes, and 8 pixels from eyes to mouth. The 256 gray-level images were adjusted to the same average brightness. (No attempt was made to equalize higher order statistics.)

Network processing entailed two stages: image compression and sex discrimination. Both networks were fully-connected three layer networks with two biases, trained with simple unadorned back-propagation (Werbos, 1974; Parker, 1986; Rumelhart, Hinton, & Williams, 1986), with a sigmoidal squashing function and a learning rate of 0.2, using Bottou and LeCun's SN2 simulator. Image compression followed the scheme of Cottrell and Fleming (1989, personal communication), who previously used compressed faces as an input to a face identity network. The compression network served to force the 30x30 images (900 inputs units) through a 40 hidden unit bottleneck, and reconstruct the image at the 900 unit output level. Thus, the input equalled the desired output. The function of this compression was twofold. First, use of compressed representations decreases the number of inputs and hence connections to the sex discrimination portion of the SexNet, allowing for faster learning and relearning of sex with different subsets of faces. Second, while simple gray-levels may adequately represent changes in face images for part of a single face in fixed lighting (Yuhas, Goldstein, Sejnowski & Jenkins, 1990), the representation of multiple faces benefits from preprocessing which extracts essential properties. In

an encoder network (Cottrell, Munro & Zipser, 1987), the compression performs a principle components analysis if the hidden units are linear. The 50 leading components reproduce reasonable likenesses of faces (Kirby & Sirovich, 1990). For nonlinear hidden units, such as those used here, the compression is more efficient and fewer are needed. The compression network trained for 2000 runs on each of 90 faces, yielding output faces that were subjectively distinct and discriminable, although not identical to the inputs. This procedure served to forge a representation of each face in the activities of only 40 units, thus providing a more tractable input (40 units rather than 900) to the sex discrimination network.

The second, sex-discrimination portion, or SexNet had 40 inputs (the activities of the 40 hidden units of the compression net), 2, 5, 10, 20 and 40 hidden units, and one output unit. Training consisted of encouraging, by gradient descent (Rumelhart, et al., 1986) the network to produce a "1" for men, and a "0" for women. Values greater than 0.5 were accounted "male", and those less than 0.5 female. In a control experiment we trained a 900x40x1 backpropagation network directly on the raw images. This network performed well on the training set but was unable to generalize.

Since the proper measure of performance of the network is human performance on the same faces, a pseudorandomized face order was established, by which even vs odd sequential digits of pi coded male vs female for 45 faces, and, to equalize males and females, the order was repeated with reverse parity for the second 45 faces. No visual reference to the faces influenced the order. 5 humans were tested on these 90 faces, with two binary decisions for each face: sex and certainty (sure vs unsure). Subjects had unlimited time, and could scrutinize faces in any manner. For comparison, 8 tests of the SexNet were undertaken, each training on a different 80 faces, leaving a distinct set of 10 untrained faces for testing.

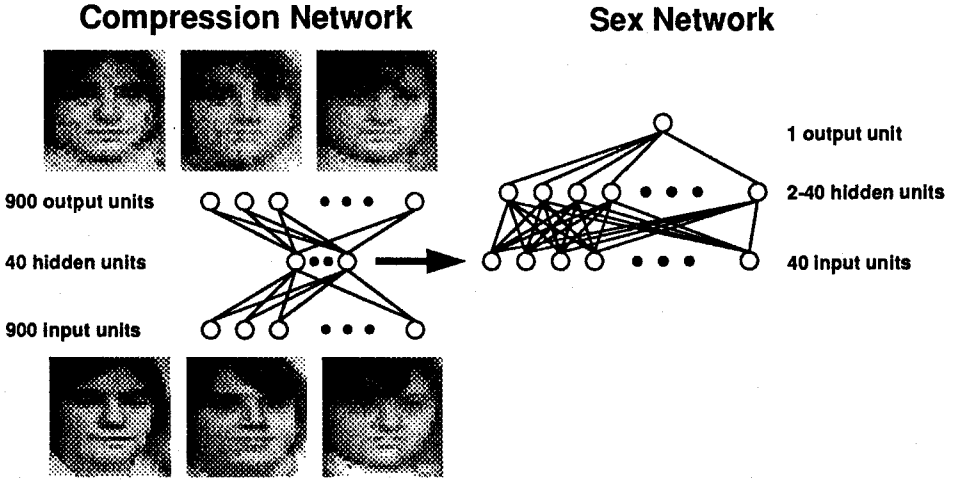

Figure 2: Two-stage network for discrimating sex from faces. The compression network encodes the normalized faces into 40 hidden units, which are then used as inputs to a sex network. The 30x30 input image has 256 gray levels per pixel. The output of the sex network is 1 for male and 0 for female.

3 RESULTS

Psychophysical studies of 5 humans on the 90 faces revealed errors of 8, 10, 12, 8 and 14, corresponding to 8.9, 11.1, 13.3, 8.9 and 15.5%, with an average error of 11.6%. The SexNet with 10 hidden units gave errors on test faces of 15, 0, 20, 0, 20, 10, 0 and 0%, for an average of 8.1%.

Similar errors seemed to affect the net and humans. One male face gave particular trouble to the SexNet, being mis-sexed when a test face, and taking long to train when a training face. This same face was (erroneously) judged "female", "sure" by all 5 human observers.

On one preliminary trial the SexNet correctly assigned all ten test faces, but misjudged two of the 80 training faces: the problematic male hitherto noted, to which it assigned the androgynous value of 0.495, and another male on which it performed wretchedly, with a value between 0.2 and 0.3, despite copious training. The SexNet proved right: The face was a clear female whose sex value had been mistranscribed.

4 DISCUSSION

Gender can be recognized by humans even when lesions of cerebral cortex in humans cause prosopagnosia, a selective impairment in the ability to recognize individual faces (Tranel, Damasio & Damasio, 1988; Damasio, Damasio & Van Hoesen, 1982). Thus, gender recognition in humans, as in our network, does not depend on the ability to identify individuals. Single neurons in the superior temporal sulcus of visual cortex, as well as the amygdala, respond selectively to faces and such neurons may participate in facial discrimination tasks similar to those of the SexNet (Rolls, 1984;Baylis, Rolls & Leonard, 1985).

We have shown that the complex visual pattern recognition task of recognizing the sex of human faces can be adequately performed by a neural network without prior feature selection and with minimal preprocessing. Human performance was matched by a using a 900x40x900 Cottrell-style back-propagation image compression network, the activities of whose hidden units served as inputs to a back-propagation SexNet; no efforts to optimize the network were needed to match human performance.

The SexNet performance was similar to humans' not just by percent errors. Not only did it correctly sex previously unseen faces as can we, but it had difficulties on faces which also posed difficulties for humans. Indeed the SexNet correctly sexed one female face despite being labeled male during training. It had evidently done a fine job of abstracting what distinguishes the sexes.

Failure of humans and the network on the same face suggests a means by which to handle the net's difficulty, in analogy with human strategies. When a face persists in being wrongly judged (say female) long after others seem stably correct, one shouldn't emend male-female categories too drastically to accommodate it; the face could be a fluke, and one may encounter another nearly identical face which is in fact female. The human strategy confronted with a "training face" (one for which sex is known by other criteria) would consist in making a special category for the individual; and having that provide input to overrule the facial information. This

would permit outliers to be correctly identified without adverse consequences to generalization.

Although the SexNet task has limited utility of itself – after all, humans sex human faces fine – extensions of this work have application. For instance, it is not known whether faces differ for male and female rhesus monkeys. By training a neural network to discriminate the sex of a monkey, then comparing the network's performance on untrained faces, better than chance performance would imply that there exist facial sex differences in rhesus monkey faces – answering a question of some ethological significance.

Another important area of application is to the recognition of facial expressions. Some emotional states, such as anger, surprise, and happiness are associated with facial expressions that are stable across cultures (Ekman, 1989). Our approach to recognizing sex can also be used to recognize human emotion from facial expression. Indeed, we have devised a preliminary ExpressioNet, which capably distinguishes among (both training and test examples of) 8 different facial expressions, a precursor to network automation of Ekman and Friesen's facial action coding system (Ekman & Friesen, 1975).

A variety of congenital medical disorders (such as Down syndrome) are accompanied by craniofacial anomalies (Dyken, & Miller, 1980), resulting in distinctive "facies", or facial appearances. Some are subtle or rare, and not often recognized by physicians. It may be possible to to screen normal from affected infants or children using special purpose neural networks. We hope to extend our work to include neural nets for diagnosing William's syndrome, or infantile hypercalcemia, in which children's' faces are "elfin-like" (Bellugi, Bihrle, Trauner, Jernigan, & Doherty, 1990; Trauner, Bellugi, & Chase, 1989). Williams' faces compare to normals in a manner which recalls the male/female distinction in that no isolated well described features occur in all of one but none of the other. Early diagnosis is important because these children often have associated cardiac defects requiring surgical correction.

On a final, more frivolous note, the same strategy, using personality indices rather than sex for the second phase of the net, could, at last, scientifically test the tenets of anthroposcopy (physiognomy), according to which personality traits can be divined from features of the face and head.

Acknowledgements

We are indebted to Dr. A. O'Toole for providing the images used in this study, and to Shona Chattarji for helping with graphics. This research was supported, in part, by the Drown Foundation.

References

Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1985). Selectivity between faces in the responses of as population of neurons in the cortex in the superior temporal sulcus of the monkey. Brain Research, 342, 91-102.

Bellugi, U., Bihrle, A., Trauner, D., Jernigan, T., & Doherty, S. (1990). Neuropsychological, neurological, and neuroanatomical profile of Williams syndrome children.

American Journal of Medical Genetics, In Press,

Cottrell, G. W., Munro, P., &, Zipser, D. (1987). Image compression by back propagation: An example of extensional programming. UCSD Institute for Cognitive Science Technical Report ICS-8702.

Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: anatomic basis and neurobehavioral mechanisms. Neurology, 32, 331-341.

Dyken, P. R., & Miller, M. D. (1980). Facial Features of Neurologic Syndromes . St. Louis, Missouri: C.V. Mosby Company.

Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. In H. W. a. J. Manstead (Ed.), Handbook of psychophysiology: Emotion and social behavior (pp. 143-164). London: John Wiley and Sons.

Ekman, P., & Friesen, W. V. (1975). Unmasking the face: A guide to recognizing emotions from facial clues. New Jersey: Prentice Hall.

Kirby, M., & Sirovich, L., (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(1), 103-108

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.

Parker, D. B. (1986). A comparison of algorithms for neuron-like cells. In J. S. Denker (Ed.), Neural networks for computing New York: American Institute of Physics.

Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. Human Neurobiology, 3, 209-222.

Rumelhart, D. E., Hinton, G., & Williams, R. J. (1986). Learning internal representation by error propagation. In D. E. R. a. J. L. McClelland (Ed.), Parallel Distributed Processing, Explorations in the microstructure of cognition (pp. 318-362). Cambridge, Mass.: MIT Press.

Tranel, D., Damasio, A. R., & Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. Neurology, 38, 690-696.

Trauner, D., Bellugi, U., & Chase, C. (1989). Neurologic features of Williams and Down Syndromes. Pediatric Neurology, 5(3), 166-168.

Werbos, P. (1974). Beyond Regression: New tools for prediction and analysis in the behavioral sciences. Harvard University,

Yuhas, B. P., Goldstein, M. H. Jr., Sejnowski, T. J., & Jenkins, R. E. (1990). Neural network models of sensory integration for improved vowel recognition. Proceedings of the IEEE, 78(10), 1658-1668