MODELLING THE VISUAL CORTEX USING ARTIFICIAL NEURAL NETWORKS FOR VISUAL IMAGE RECONSTRUCTION

G. Qiu

University of Derby, UK

Abstract: In this paper, we present an artificial neural network model with some special neurons that are designed to function as the *feature detectors* found in the visual cortex and apply it to the reconstruction of visual images. The model is a multilayer feedforward neural network. The neurons in the first hidden layer of the network are feature detectors of various scales and orientations. The connection strengths between the input and the first hidden layer are pre-set (and fixed) such that the outputs of this layer are some visually important features of various scales. The rest of the connection strengths in the network are decided through learning via the backpropagation algorithm in a self-supervising manner. Computer simulations were conducted, and the results seem to show that the artificial neural network model proposed in this paper is consistent with its biological counterpart (the visual cortex) in terms of the visual features detected by its feature detectors and its ability to reconstruct near perfect input images from the output of these feature detectors. It is also shown that the system can be used for effective image data compression. Simulation results are presented which show promising potential of the new system for image coding applications.

INTRODUCTION

There are 100 million neurons in the visual cortex. Only the smallest fraction of these have been thoroughly studied in attempts to discover their response characteristics. Based on current technology and our limited understanding of the visual cortex, it would be impossible to simulate the *complete* function of the visual cortex using computer programs. However, it might be possible to simulate *some* of the known characteristics of the visual cortex using modern computational technologies. In this work, an attempt has been made to use artificial neural network technology to perform a very small part of this task.

It has been demonstrated by researchers that there are several places in the cortex where there is a rather direct topological mapping of the external visual world, with specific points in the environment corresponding to specific points in the cortex. This may mean that much of the analysis of features of visual input occurs within individual cortical neurons and patches of adjacent cortex. Much of the pioneering work in this area was done by David Hubel and Torstein Wiesel. In the 1960s, they discovered that there are certain types of *feature detectors* in the visual cortex of cats [1,2]. These feature detectors were labelled as simple cells and complex cells. The simple cells never seem to respond to diffuse illumination covering the whole screen, but respond strongly to stimuli such as bars or edges at particular orientations. The complex cells have larger receptive fields than do simple cells. Like simple cell, complex cells respond maximally to stimuli when they are in a particular orientation and they also seem to generalise their response over a wider area of the visual field.

In this paper, we present an artificial neural network model with some special neurons that are designed to function as the feature detectors found in the visual cortex and apply it to the reconstruction of visual images. The model is a multilayer feedforward neural network. The neurons in the first hidden layer of the network are feature detectors of various scales and orientations. The connection strengths between the input and the first hidden layer are pre-set (and fixed) such that the outputs of this layer are some visually important features of various scales. The rest of the connection strengths in the network are decided through learning via the backpropagation algorithm [3] in a self-organising manner. The purpose of building such a model is twofold. Firstly we would like to explore the possibility of using artificial neural network technology to simulate the psychophysical discoveries with the expressed goal of building biologically plausible computational models. Secondly, we would like to explore such biologically inspired computational

'Artificial Neural Networks', 26–28 June 1995, Conference Publication No. 409, © IEE, 1995

models for engineering applications. Computer simulations were conducted, and the results seem to show that the artificial neural network model is consistent with its biological counterpart (the visual cortex) in terms of the visual features detected by its feature detectors and its ability to reconstruct near perfect input images from the output of these feature detectors. We also demonstrate that such a system can be used for effective image data compression. Simulation results are presented which show promising potential of the new system for image coding applications.

A NEURAL NETWORK MODEL WITH HIERARCHICAL VISUAL FEATURE DETECTORS

The schematic diagram of a neural network model with hierarchical visual feature detectors is shown in Fig. 1. The first hidden layer of the network is the feature detector layer. The neurons in this layer responsed to image features of different scales. It has been shown that the feature detectors in the visual cortex are sensitive to edge patterns of various orientations. Therefore, a reasonable choice of the image features detected in this layer would be the first order directional derivatives of different scale. The first order derivatives not only detect the changes in grey scale intensities (edges), also the combination of the directional derivatives in the higher layer can determine the orientations of the edge patterns. In other words, the neurons in the first hidden layer are directional derivative calculators of different scale. Therefore, the connection strengths between the input and the first hidden layer can be pre-set (and fixed) to achieve this goal. The neurons in the rest of the layers are ordinary neurons, i.e. they calculate the weighted sum of the input signals and pass it through a transfer function to produce their outputs. The connection strengths of these layers are determined through learning via the backpropagation algorithm. The number of neurons in the output layer is set to be the same as that of the input layer. When an image is fed to the network, it is also used as the desired output of the network for training purpose, i.e. the network learns to reconstruct the input image.

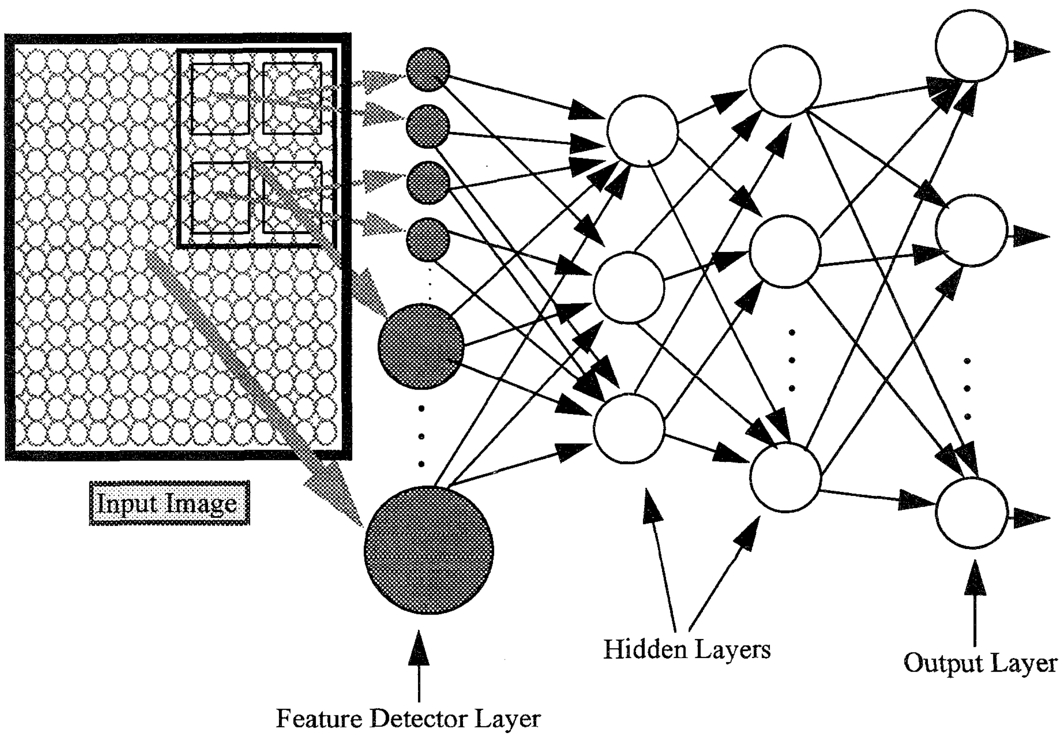

Fig. 1 A neural network with hierarchical visual feature detectors

SIMULATION

In the simulation, we divided the input images into 8 x 8 blocks, i.e. there were 64 input neurons (and therefore 64 output neurons) in the network. In this simulation, two-level hierarchy feature detector neurons were used. The blocks were fed to the network one at a time. The feature detector layer has 10 neurons, whose outputs are, horizontal and vertical directional derivatives of the four non-overlapping 4 x 4 blocks within the 8 x 8 block and two directional derivatives of the 8 x 8 block. Let $I(k, l)$, $k = 0, 1, \cdots, 7$, $l = 0, 1, \cdots, 7$, be the pixel values in an 8 x 8 image block, the feature layer generates the following 10 outputs.

$$f_{4dx}(i,j) = \sum_{k=0}^{3}\sum_{l=0}^{3} I(4i+k, 4j+l)G_{4x}(k,l) \tag{1}$$

$$f_{4dy}(i,j) = \sum_{k=0}^{3}\sum_{l=0}^{3} I(4i+k, 4j+l)G_{4y}(k,l) \tag{2}$$

for all $i = 0, 1$ and $j = 0,1$, and

$$f_{8dy} = \sum_{k=0}^{7}\sum_{l=0}^{7} I(k,l)G_{8y}(k,l) \tag{3}$$

$$f_{8dx} = \sum_{k=0}^{7}\sum_{l=0}^{7} I(k,l)G_{8x}(k,l) \tag{4}$$

where the convolution kennels are defined as follows:

$$G_{4x}(i,j) = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix}$$

$$G_{4y}(i,j) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{bmatrix}$$

$$G_{8x}(k,l) = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{-1} \end{bmatrix}$$

$$G_{8y}(k,l) = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{-1} & \mathbf{0} \end{bmatrix}$$

where the bolded 0 represents a 4 x 4 matrix with all of its elements equal to 0, the bolded 1 represents a 4 x 4 matrix with all of its elements equal to 1, and the bolded -1 represents a 4 x 4 matrix with all of its elements equal to -1.

The desired output of the network, D is defined as

$$D((4k+i),\ (4l+j)) = I((4k+i),\ (4l+j)) - mean(2k+l) \tag{5}$$

for all $k = 0,1$; $l = 0, 1$; $i = 0, 1, 2, 3$; and $j = 0, 1, 2, 3$, and where

$$mean(2k+l) = \frac{1}{16}\sum_{i=0}^{3}\sum_{j=0}^{3} I((4k+i),\ (4l+j)) \text{ for } k = 0, 1;\ l = 0, 1$$

A single hidden layer with 30 neurons was used in the network. When the training process has completed, the network was used to reconstruct the input image. In the reconstruction process, the block means are added back to the neural network outputs to obtain the reconstructed image pixel values, therefore the reconstructed block, I_r is defined by

$$I_r((4k+i),\ (4l+j)) = Y((4k+i),\ (4l+j)) + mean(2k+l) \qquad (6)$$

for all $k = 0,1$; $l = 0, 1$; $i = 0, 1, 2, 3$; and $j = 0, 1, 2, 3$, where Y is the output of the neural network.

RESULTS: Simulations have been performed on a large number of grey scale monochrome images. All the results indicated that the network can reconstruct visually near perfect images. Here we show one of the images. Fig. 2 (a) shows the original 8 bit per pixel Girl image. Fig. 2 (b) shows the network reconstructed Girl image. It is seen that the network has achieved near perfect reconstruction of the original image.

(a) (b)

Fig. 2 (a) Original Girl image, (b) Reconstructed image of (a)

APPLICATION TO IMAGE DATA COMPRESSION

The above system can be easily extended for image coding applications. In the example of the above section, if we quantized all the outputs of the feature detector layer and the 4 x 4 block means using an 8-bit scalar quantizer, these quantized parameters can be used as the compressed features for the 8 x 8 image block, achieving a data compression ratio of 5.3 : 1. There are other more efficient techniques available to code the feature detector layer outputs, which allow us to achieve much higher compression ratios. Furthermore, the nature of the biologically inspired system allows us to employ the psychovisual redundancy of the image easily.

To take full advantage of the network described in the above sections for image coding application, we make the following observations. Firstly, in most natural images, edges only constitute a small fraction of the image contents, thus we would expect for most blocks in the image, the directional derivatives will be very small. Secondly, human vision is very much more sensitive to intermediate spatial frequencies than to very low or very high spatial frequencies [4]. Based on these observations, we present in the following a possible approach to apply the described network to image coding.

Quantizing Directional Derivatives

We first design a quantizer to code the directional derivatives of the 4 x 4 blocks. To do this, we collect the horizontal directional derivative of all the 4 x 4 blocks, f_{4dx}, from the training images. A competitive neural network [5] is used to cluster these data samples into several classes. In the results presented below, three classes are used corresponding to the horizontal derivative being negative large (XNL), small (XS) and positive large (XPL) respectively. The same clustering process is then applied to the vertical directional derivative of all the 4 x 4 blocks, f_{4dy}, to obtain three classes of the vertical derivative corresponding to it being negative large (YNG), small (YS) and positive large (YPL) respectively. Notice it is important to cluster the directional derivatives separately, so that the quantizer codes the edge structures (edge orientations) more accurately. A block is then classified into one of the following nine classes:

(XNL, YNL)	(XNL, YS)	(XNL, YPL)
(XS, YNL)	(XS, YS)	(XS, YPL)
(XPL, YNL)	(XPL, YS)	(XPL, YPL)

The cluster centre co-ordinates for each direction are stored and used during the coding stage to classify an image block into one of the nine classes. The quantizer output is the cluster centre co-ordinates of the appropriate class.

Training of the System for Image Coding

To train the network for image coding applications, the same training procedure described above is applied. The directional derivatives of the 4 x 4 blocks are replaced by the cluster centre co-ordinates of their corresponding class from the directional derivative quantizer described above. The directional derivatives of the 8 x 8 block are obtained from the quantized block means of the 4 x 4 blocks (see definitions of the convolution kernel G_{8x} and G_{8y}). The quantized block means are also used to obtain the desired output of the network.

Image Coding Results

To code an 8 x 8 block, the block means of the four non-overlapping 4 x 4 blocks are coded using a scalar quantizer. The directional derivatives of the four 4 x 4 blocks are coded using the quantizer described above. The four quantized block means of the 4 x 4 block and the four codeword indices of the directional derivatives quantizer constitute the compressed data of an 8 x 8 block. Reconstruction of the block is achieved according to equation (6).

In the experiments, it was found that a large majority of the 4 x 4 blocks have their directional derivatives fell in the class (XS, YS). Hence a variable bit rate coding strategy was used when coding the directional derivatives of the 4 x 4 blocks. One bit was used to indicate whether a block belongs to the class (XS, YS) or one of the other eight classes. Thus, if a block belongs to the class (XS, YS), one bit is used for the class index, and if it belongs to one of the other eight classes, four bits are used to code its directional derivatives.

Computer simulations have been performed on a large number of images and the results are very encouraging. We show here one of the examples. Fig. 3 (a) shows the compressed Girl image when the directional derivatives and block means were coded using an 8-bit scalar quantizer. In this case the bit rate is 1.5 bits per pixel. The peak signal to noise ratio (PSNR) is 37 dB. Fig. 3 (b) shows the compressed Girl image when the directional derivatives were coded using the quantizer described in this paper and the block means coded using a 6-bit scalar quantizer. The bit rate achieved in this case is 0.61 bit per pixel. The PSNR value is 32 dB. It is seen that at 0.61 bit per pixel (Fig. 3 (b)), the visual quality obtained by the technique is quite satisfactory.

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) \ \mathbf{dB}$$

where

$$MSE = \frac{1}{N}\sum_{i=0}^{N-1}\left(I(i) - I_r(i)\right)^2$$

and N is the number of pixels in the image, $I(i)$ represents the original pixel values and $I_r(i)$ represents the reconstructed pixel values.

(a) (b)

Fig. 3, (a) Compressed Girl image at 1.5 bits per pixel, PSNR = 37 dB
(b) Compressed Girl image at 0.61 bit per pixel, PSNR = 32 dB

CONCLUDING REMARKS

In this paper, a neural network model, inspired by the visual cortex has been presented. Simulation results show that it can reconstruct near perfect images from a few visually important multi-scale features. It is also shown the system can be used for effective image coding applications. Initial experimental results have demonstrated promising performance in image coding applications.

REFERENCES

1. D. H. Hubel and T. N. Wiesel, 1979, “Brain mechanisms of vision”, Scientific American, Vol. 241, pp. 150-162

2. D. H. Hubel and T. N. Wiesel, 1962, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”, Journal of Physiology, Vol. 160, pp. 106-154

3. D. E. Rumelhart, G. E.. Hinton and R. J. Williams, 1986, “Learning internal representations by error propagation”, in Parallel Distributed Processing, Vol. 1, Chapter 8, D E. Rumelhart and J. L. McClelland (Eds.), Cambridge, MA, MIT Press,

4. F. W. Campbell and J. G. Robson, 1968, “Application of Fourier analysis to the visibility of gratings” Journal of Physiology, Vol. 197, pp. 551-566

5. T. Kohonen, 1988, *Self-organization and associative memory*, Second Edition, Springer-Verlag, Berlin