Available online at www.sciencedirect.com

ScienceDirect

Neurocomputing 70 (2007) 1808–1812

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Modelling the statistical processing of visual information

Mia Šetić, Domagoj Švegar, Dražen Domijan*

Department of Psychology, Faculty of Philosophy, University of Rijeka, I. Klobučarića 1, HR-51000 Rijeka, Croatia

Available online 6 November 2006

Abstract

Recent psychophysical investigations showed that humans have the ability to compute the mean size of a set of visual objects. The investigations suggest that the visual system is able to form an overall, statistical representation of a set of objects, while the information about individual members of the set is lost. We proposed a neural model that computes the mean size of a set of similar objects. The model is a feedforward, two-dimensional neural network with three layers. Computer simulations showed that the presented model of statistical processing is able to form abstract numerical representation and to compute the mean size independently from the visual appearance of objects. This is achieved in a fast, parallel manner without serial scanning of the visual field. The mean size is computed indirectly by comparing the total activity in the input layer and in the third layer. Therefore, the information about the size of individual elements is lost. An extended model is able to hold statistical information in the working memory and to handle the computation of the mean size for surfaces with empty interiors.

Keywords: Dendritic computation; Number; Presynaptic inhibition; Space; Statistical processing

1. Introduction

The visual system is confronted with a multitude of objects in everyday environment. How does the visual system cope with this complexity when it is probed to compute the size of objects? Recent psychophysical investigations suggest that humans are able to compute the mean size of a set of similar objects presented visually [1,4,5]. Ariely [1] used a set of dots with different radius in a member identification task and in a mean discrimination task. In the member identification task, the participants were asked to judge whether a test dot was presented in a set of dots. In the mean discrimination task, the participants were asked to judge whether the test dot represents an average dot size in the set. A surprising finding was that the participants were able to accurately solve the mean discrimination task but had difficulty with the member identification task. It suggests that the visual system is able to form an overall, statistical representation of a set of objects, while the information about individual members of the set is lost. In a detailed analysis of this finding, Chong and Treisman [4,5] showed that the statistical analysis of the size is performed in a fast, parallel manner without serial scanning. They also showed that the visual system actually computes the mean size and not the median.

How is statistical processing implemented in the brain? In this report, it is argued that a model of number detection and discrimination proposed by Domijan [9] can account for observed properties of the visual statistical processing. The model was intended to explain how abstract representation of numbers arises in a neural network. Such representation is not sensitive to the position of objects, their size or density, or any other visual properties. The model was able to simulate properties of number sensitive neurons discovered in the prefrontal cortex of monkeys [16]. Also, the model simulates the results from psychophysical investigations of the number discrimination [7].

2. Model description

The model is a feedforward, two-dimensional neural network with three layers. It is assumed that all required computations are feedforward in order to achieve the rapid

*Corresponding author. Tel.: +385 51 315233; fax: +385 51 315228.
E-mail addresses: mia-setic@ffri.hr (M. Šetić), dsvegar@ffri.hr (D. Švegar), ddomijan@ffri.hr (D. Domijan).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2006.10.069

formation of numerical representation and to avoid dynamic complexities associated with recurrent networks. Therefore, the model is described in an algebraic form but it can be easily transformed into a real-time form.

The first layer is an input layer defined as $I_{ij} = 1$ if an object occupies location $\{i,j\}$ or $I_{ij} = 0$ if that location is empty. It roughly corresponds to the striate cortex where the surface representation is formed. The dimensions of the network are denoted with $i = 1, \ldots, M$ and $j = 1, \ldots, N$. The second layer employs two mechanisms: gradient synaptic weights and dendritic multiplication. Signals from the first layer are differentially weighted depending on their spatial position in the network, meaning that every location in the network receives a unique amount of tonic activation. Furthermore, the input activity is convolved with Gaussian kernels which models properties of neurons in the parietal cortex. The total node activity is multiplied by a corresponding node in the input layer. Dendritic multiplication prevents interference from the nodes that do not represent objects but are active due to the activation in the neighbourhood. The multiplication simply silences them because they do not receive direct support from the input layer. Both mechanisms are necessary for obtaining a proper transformation from the visual input to the number representation as shown by computer simulations in [9].

Formally, activity of the node in the second layer, x_{ij}, at spatial position $\{i,j\}$ are given by

$$x_{ij} = I_{ij} \times \left[\sum_m \sum_n G_{mnij} I_{mn} + W_{ij} J \right]. \quad (1)$$

The nodes in the second layer compute Gaussian weighted sum of their input. Gaussian kernel, G_{mnij}, is given by

$$G_{mnij} = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(m-i)^2 + (n-j)^2}{2\pi\sigma^2} \right]. \quad (2)$$

The size of the receptive field or the spatial spread is defined by a standard deviation of Gaussian kernel, σ. The nodes in the second layer also receive a tonic input or bias, J, which is multiplied by synaptic weights, W_{ij}, defined as

$$W_{ij} = 0.5[i + (j-1)N]/NM. \quad (3)$$

The tonic input generates a gradient of activity values where every spatial location receives a unique amount of excitation. Synaptic weights, W_{ij}, are normalized by NM in order to prevent excessive amount of activation in networks with large dimensions. The tonic input endows the nodes with small differences in the activity in the case where the neighbouring nodes receive the same amount of excitation from their receptive fields. This can occur with large objects that can cover the whole receptive field of several nodes. It could be considered as a spatial code which provides every location with a unique activity level. Dendritic multiplication between the total excitation (receptive field input plus tonic input) and the input signal, I_{ij}, at the corresponding location is denoted with an $\times$ sign.

The third layer nodes, y_{ij}, described as

$$y_{ij} = f\left[x_{ij}\right] - \sum_p \sum_q f\left[x_{pq} - x_{ij}\right] \quad (4)$$

compute the winner-takes-all (WTA) function, restricted to the four nearest neighbour locations, p and q, defined as $\{(i-1,j), (i+1,j), (i,j-1), (i,j+1)\}$. The rectified output of the third layer, Y_{ij}, is given by

$$Y_{ij} = g\left[y_{ij}\right]. \quad (5)$$

The function $f(\,)$ is defined as $f(a) = 1$ if $a > 0$ and $f(a) = 0$ if $a \leqslant 0$ and function $g(\,)$ is a linear threshold defined as $g(a) = a$ if $a > 0$ and $g(a) = 0$ if $a \leqslant 0$. The third layer computes the activity differences between the neighbouring nodes and allows only the locally most active nodes to survive the competition and to represent objects. The particular node will remain active only if it receives stronger input than its immediate neighbours. One neighbour node with larger activity will be sufficient to inhibit the node below its firing threshold. In the third layer, every object is represented by a single node, which enables computation of the input numerosity. The activity difference is computed using the presynaptic inhibition from the target node to the inhibitory axons from the neighbouring nodes. The physiological substrate for this feedforward presynaptic inhibition is a glutamate transmitter spill-over on the presynaptic GABA receptors on the nearby inhibitory axons [2].

3. Results

The network behaviour is tested using computer simulations shown in Fig. 1. The network dimensions were $M = 50$, $N = 50$, and the standard deviation of Gaussian kernel was set to $\sigma = 4$. Instead of disks, we used squares as the input objects. Columns in Fig. 1 depict different network layers and the rows show network responses to different input configurations. The first column denotes the input. The second column shows the second layer activation after the convolution of the input with Gaussian kernel. The third column indicates the second layer activation after the dendritic multiplication takes place, which removes some of the noise in the network activation. It is an important step for an adequate object representation. Otherwise, more than one node may be active for a single object in the third layer, whose activation is illustrated in the fourth column. The top row in Fig. 1. shows the network ability to extract invariant number representation from the input with many objects of different sizes. The middle row in Fig. 1. shows that the network is robust with respect to the change in the size of the receptive fields in the second layer and with respect to changes in the input density. Here, $\sigma = 6$ and objects are rearranged so as to be maximally close to each other.

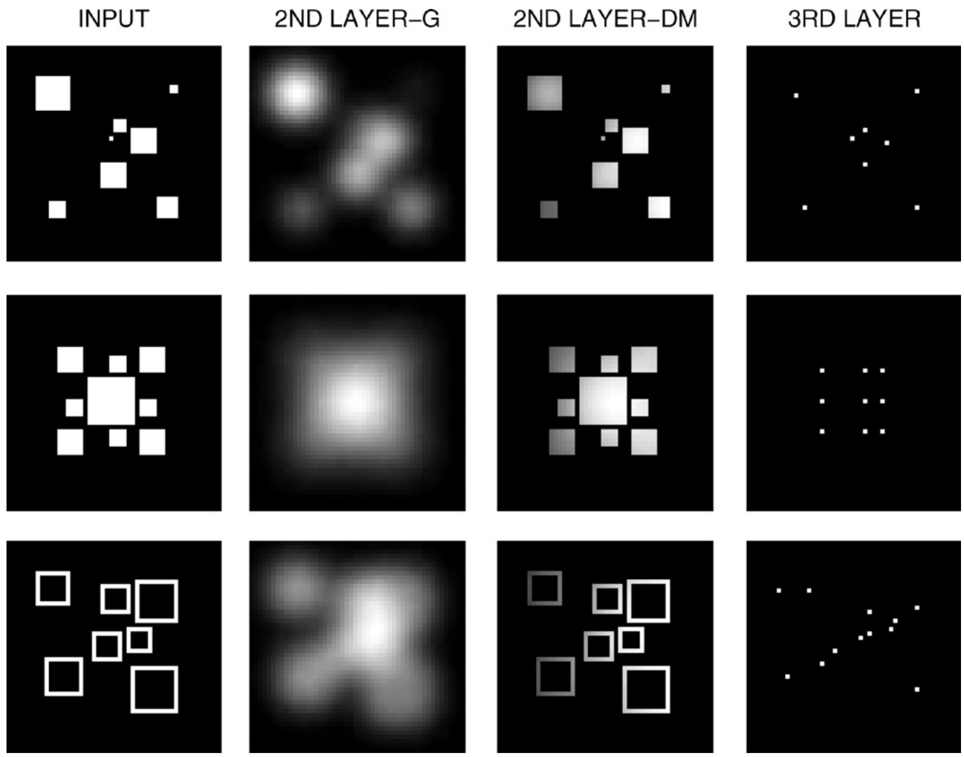

Fig. 1. Computer simulations of the proposed network architecture for the statistical processing of the visual input. The columns denote different network layers and computations. The rows denote the network response to different input configurations. G—convolution with Gaussian kernel. DM—dendritic multiplication.

The bottom row in Fig. 1. illustrates the network limitation when it is confronted with geometrically more complex objects which have empty interiors. In this case, the network is not able to achieve the proper representation of input numerosity.

With several additions, the model can be applied to the statistical processing of size. If we assume that there exists a node that computes the sum of activation in the third layer and a separate node that computes the sum of activation in the input layer, then, the ratio between activities of these two nodes represents a mean size of a set of objects. These nodes simply count the number of active nodes in the corresponding layers because in both layers only two values are possible (0 for empty location and 1 for location occupied by the object). Formally, they are described as

$$I = \sum_i \sum_j I_{ij}; \ Y = \sum_i \sum_j Y_{ij}. \tag{6}$$

Furthermore, we may assume that there is a third node which computes the ratio of output of two summation nodes in order to provide information about the mean size. Ratio computation is possible in a simple neuron model with divisive or shunting inhibition [3,12]. For example, consider a node, z, defined by the following equation:

$$\frac{dz}{dt} = -Az + I - zY. \tag{7}$$

The node z receives excitation from the I node, and inhibition from the Y node. Inhibition multiplicatively interacts with the activity of the node z. The parameter A describes a passive decay of activity. Due to the fact that this is a feedforward model, the node z will quickly reach an equilibrium $dz/dt = 0$ and the steady-state solution is given by

$$z = \frac{I}{A + Y} \tag{8}$$

which is proportional to the ratio I/Y [12]. With $A \ll Y$, the node z computes the exact mean size. We hypothesize that the read-out of the mean size information could occur in the parietal or the prefrontal cortex where the number sensitive neurons are discovered. The same neurons could respond to statistical information also. The computation of the mean size is achieved indirectly without explicit computation of the size of individual elements in the set. Consequently, the model is not able to solve the member identification task which is consistent with experimental findings. The mean computation is fast because it is based on the feedforward network and it does not require serial scanning of the visual input.

Although the model is conceptualized as a feedforward network, it is possible to include self-excitatory connections in order to sustain activity in the third layer after the input vanishes. Due to the fact that single nodes are used to represent object locations, it is sufficient to include only the self-excitation (without the recurrent collaterals to other nodes) which will drive the node to the saturation level and keep it active until external reset is issued. The model should be able to sustain activity because in the mean identification task, the presentation of a set of objects and a test object is made sequentially. Moreover, Chong and Treisman [4] observed no difference in performance with an increased temporal gap between the presentation of the set of objects and the test object indicating that statistical representation was held in the working memory.

It should be noted that Chong and Treisman [4,5] used open circles instead of filled circles in their experiments. Open circles pose a challenge to the present model because it is designed to operate on filled surfaces. This is illustrated in Fig. 1 (bottom row) where it is shown that the third layer greatly overestimates the number of objects when the surface interiors are empty. However, the model could be augmented with a pre-processing stage, which segment the interior of the circles as separate surfaces than background. This is possible in a model of visual segmentation proposed by Domijan [10]. The segmentation network assigns different activity level to the nodes that represent different surfaces or objects while the nodes that code the same surfaces obtain the same activity level due to the activity spreading through the local excitatory connections. How this is achieved is illustrated in Fig. 2. The input to the segmentation network (Fig. 2a) is same as in the Fig. 1 (bottom row). The segmentation network is initialized with the gradient of activity values similar to the weight gradient used in Eq. (3) (Fig. 2b). After the convergence, every surface is labelled with a unique activity label. Activity values are uniformly spread along the whole surfaces (Fig. 2c). Background locations receive the largest possible activity value and they are easily distinguished from other surfaces. An inverse transformation of activity values in the segmentation network enables formation of representation where the background is assigned the value of zero and all squares along with their interiors are represented with the value of one (Fig. 2d). Such a representation is appropriate as an input for the network for statistical processing.

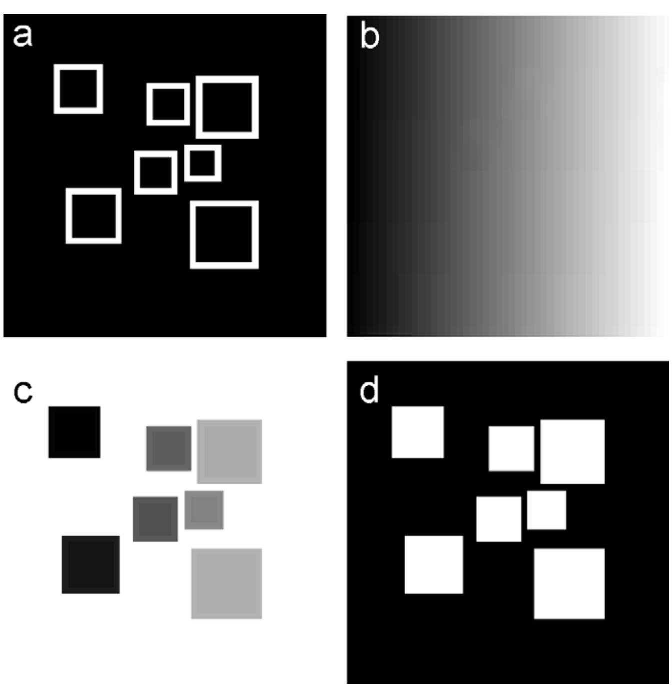

Fig. 2. Computer simulation of the segmentation network: (a) input; (b) initial state of the network; (c) network activity after the convergence; (d) inverse transformation which removes the background and retains the object representation.

4. Discussion

The proposed model is based on several biophysically realistic mechanisms. The assumption that synaptic weights could be a function of spatial locations has been previously used in models of sensorimotor transformations and in the target selection in the parietal cortex. For instance, Groh [11] used the simple linear function between weights and locations in a model of transformation from the sensory place code to the motor rate code. In a model of target selection, Hahnloser et al. [14] introduced pointer neurons which redirect activity in the recurrent network according to the attentional demands. The pointer neurons made excitatory connections with the other neurons. Synaptic weights of excitatory connections are assumed to depend on the spatial location in order to achieve smooth transitions between targets.

A multiplicative interaction is assumed to occur between different dendritic branches. In a recent computational study of synaptic integration in a dendritic tree, Poirazi et al. [17] discovered that the sum of independent sigmoid functions is a good approximation of the input–output relations in a pyramidal neuron. However, they ignored interactions between distal and proximal dendritic branches, which are better described with multiplication. Therefore, Häusser and Mel [15] argued for a more elaborate three-layer model of dendritic integration with the summation in the first layer and the multiplication in the second layer before the output is sent to the soma of the neuron.

The presynaptic inhibition has been previously used in the computational study of the WTA function where it is implemented as a recurrent network. Yuille and Grzywacz [18] found that their network exhibited sensitivity to initial conditions, that is, the network was not able to reset itself when new input had been presented. Furthermore, the WTA is obtained only by biophysically unrealistic assumptions about the network connectivity. On the other hand, the proposed model implements presynaptic inhibition as a feedforward mechanism and therefore avoids problems associated with recurrent networks such as dynamic instabilities and insensitivity to the input fluctuations. Also, synaptic weights of the unit strength are sufficient to obtain the WTA behaviour without losing precision. The presynaptic inhibition mediated by the glutamate spill-over from the excitatory node to the axons of the inhibitory interneurons has been discovered recently in the ventrobasal thalamus of rats [2]. Axonal terminals of the inhibitory interneurons in the thalamus are endowed with kainate glutamergic receptors, whose activation depresses the GABA release.

The presented model of computing the statistical description of the visual input is a simple extension of the model of the number discrimination proposed by Domijan [9]. The model has several advantages over other models of numerical processing. It does not depend on the serial scanning of the input as the model by Grossberg and Repin [13]. It uses a single spatial scale which does not confuse single large object with many small ones, as it is possible in a multi-scale model by Changeux and Dehaene [6]. Here we showed that the same model could be applied in the modelling of the mean size discrimination task. Therefore, the model makes a testable prediction that during the statistical processing of the size, the same parietal circuit is employed as for computing of the approximate numerical magnitude. In other words, the mean-size computation will activate an intraparietal sulcus which is known to be involved in a neural representation of the mental number line. This activation is distinguished from the activation of the parietal cortex during serial scanning of the input guided by attention which involves the posterior superior parietal cortex [8]. A potential concern with the model is that it utilized different computational mechanisms at every network stage. It is an open research question whether it is possible to obtain the same results in a more uniform architecture.

In conclusion, the model of numerical processing is able to compute the mean size of a set of similar objects. This is achieved in a fast, parallel manner without serial scanning of the visual field. The mean size is computed indirectly by comparing the total activity in the input layer and in the third layer. Therefore, the information about the size of individual elements is not represented while the mean size of a set is computed. The extended model is able to hold statistical information in the working memory and to handle computation of the mean size for open circles as well. The model employs several biophysically realistic mechanisms such as gradient synaptic weights, dendritic multiplication, divisive inhibition, and presynaptic inhibition by the glutamate spill-over.

References

[1] D. Ariely, Seeing sets: representation by statistical properties, Psychol. Sci. 12 (2001) 157–162.

[2] K.E. Binns, J.P. Turner, T.E. Salt, Kainate receptor (GluR5)-mediated disinhibition of responses in rat ventrobasal thalamus allows a novel sensory processing mechanism, J. Physiol. 551 (2003) 525–537.

[3] M. Carandini, D.J. Heeger, Summation and division by neurons in primate visual cortex, Science 264 (1994) 1333–1336.

[4] S.C. Chong, A. Treisman, Representation of statistical properties, Vis. Res. 43 (2003) 393–404.

[5] S.C. Chong, A. Treisman, Statistical processing: computing the average size in perceptual groups, Vis. Res. 45 (2005) 891–900.

[6] S. Dehaene, J.P. Changeux, Development of elementary numerical abilities: a neural model, J. Cogn. Neurosci. 5 (1993) 390–407.

[7] S. Dehaene, G. Dehaene-Lambertz, L. Cohen, Abstract representations of numbers in the animal and human brain, Trends Neurosci. 21 (1998) 355–361.

[8] S. Dehaene, M. Piazza, P. Pinel, L. Cohen, Three parietal circuits for number processing, Cogn. Neuropsychol. 20 (2003) 487–506.

[9] D. Domijan, A neural model of quantity discrimination, NeuroReport 15 (2004) 2077–2081.

[10] D. Domijan, Recurrent network with large representational capacity, Neural Comput. 16 (2004) 1917–1942.

[11] J.M. Groh, Converting neural signals from place codes to rate codes, Biol. Cybern. 85 (2001) 159–165.

[12] S. Grossberg, Nonlinear neural networks: principles mechanisms, and architectures, Neural Network 1 (1988) 17–61.

[13] S. Grossberg, D.V. Repin, A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes, Neural Network 16 (2003) 1107–1140.

[14] R. Hahnloser, R.J. Douglas, M. Mahowald, K. Hepp, Feedback interactions between neuronal pointers and maps for attentional processing, Nat. Neurosci. 2 (1999) 746–752.

[15] M. Häusser, B.W. Mel, Dendrites: bug or feature?, Curr. Opin. Neurobiol. 13 (2003) 372–383.

[16] A. Nieder, D.J. Freedman, E.K. Miller, Representation of the quantity of visual items in the primate prefrontal cortex, Science 297 (2002) 1708–1711.

[17] P. Poirazi, T. Brannon, B.W. Mel, Pyramidal neuron as a 2-layer neural network, Neuron 37 (2003) 989–999.

[18] A.L. Yuille, N.M. Grzywacz, A winner-take-all mechanism based on presynaptic inhibition feedback, Neural Comput. 1 (1989) 334–347.

Mia Šetić is a Ph.D. student at the University of Zagreb, Croatia. She received B.A. in Psychology from the University of Rijeka, Croatia. Her research interests include behavioural investigations of interaction between perception and abstract conceptual processing. Also she is interested in computational modelling of perception, attention, and numerical cognition. She is a member of Association for Psychological Science.

Domagoj Švegar is a Ph.D. student at the University of Ljubljana, Slovenia. He received B.A. in Psychology from the University of Rijeka, Croatia. His research interests include behavioural testing and computational modelling of visual working memory.

Dražen Domijan is an Assistant Professor of Psychology at the University of Rijeka, Croatia. He received B.A. in Psychology from the University of Rijeka, Croatia, and Ph.D. in Psychology from the University of Zagreb, Croatia. His research interests include computational modelling of visual perception, attention and cognition. In particular, he studied how models of neural networks with dendrites may help to explain various cognitive phenomena and how artificial dendrites may increase computational power of neural networks. He is a member of International Neural Network Society and Association for Psychological Science.