Neural network model develops border ownership representation through visually guided learning

Akihiro Eguchi *, Simon M. Stringer

Oxford Centre for Theoretical Neuroscience and Artificial Intelligence, Department of Experimental Psychology, University of Oxford, Oxford, UK

ABSTRACT

As Rubin's famous vase demonstrates, our visual perception tends to assign luminance contrast borders to one or other of the adjacent image regions. Experimental evidence for the neuronal coding of such border-ownership in the primate visual system has been reported in neurophysiology. We have investigated exactly how such neural circuits may develop through visually-guided learning. More specifically, we have investigated through computer simulation how top-down connections may play a fundamental role in the development of border ownership representations in the early cortical visual layers V1/V2. Our model consists of a hierarchy of competitive neuronal layers, with both bottom-up and top-down synaptic connections between successive layers, and the synaptic connections are self-organised by a biologically plausible, temporal trace learning rule during training on differently shaped visual objects. The simulations reported in this paper have demonstrated that top-down connections may help to guide competitive learning in lower layers, thus driving the formation of lower level (border ownership) visual representations in V1/V2 that are modulated by higher level (object boundary element) representations in V4. Lastly we investigate the limitations of our model in the more general situation where multiple objects are presented to the network simultaneously.

1. Introduction

As Rubin's famous vase (Fig. 1) demonstrates, our visual perception tends to assign luminance contrast borders to one or other of the adjacent image regions, as if they serve as occluding contours (von der Heydt, Zhou, & Friedman, 2003). This is an example of *feature binding* in vision, in this case binding a luminance contrast border to a particular object. Representing such binding relationships between visual features is essential to the ability of the visual system to interpret and *make sense* of complex visual scenes. Experimental evidence for the neuronal coding of such border-ownership in the primate visual system has arisen in a neurophysiology study carried out by Zhou, Friedman, and von der Heydt (2000).

Zhou et al. (2000) have shown that the responses of simple cells in earlier cortical stages of visual processing such as V1 and V2, which respond preferentially to oriented edges, are also modulated by which side of an object or figure the edge occurs on. This is the case even when the figure/background cues lie well outside the classical receptive field of the neuron, which in area V1 is approx-

imately 1 degree in size. Such neurons are referred to as *border ownership cells*. Sugihara, Qiu, and von der Heydt (2011) later reported that the border ownership signal emerges with a latency of 61 ms, which is about 13 ms later than the onset of orientation selectivity. This suggests that the global image context specifying border ownership modulates the activity of these neurons. In other words, there must be a mechanism that enables the contextual information to be conveyed to these early stage visual neurons in V1 and V2. It has been proposed that these kinds of border ownership responses in area V1 represent a form of feature binding, and so may be important for understanding how primate vision may solve the problem of feature binding more generally.
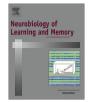
Some theoreticians have suggested that the context integration required for border ownership representations in V1 and V2 can be achieved via lateral propagation of signals within a layer via horizontal fibres (Baek & Sajda, 2005; Nishimura & Sakai, 2004; Zhaoping, 2005). However, Sugihara et al. (2011) have argued that the conduction velocity of horizontal fibres is too slow (most of them being between 0.1 and 0.4 m/s (Angelucci & Bullier, 2003)) to produce the border ownership signals within the short latency observed in neurophysiology studies. Furthermore, Sugihara et al. (2011) showed that varying the distance between the target border and the visual features that carry contextual information about the 'owner' of the border does not in fact influence the latency before

* Corresponding author.
E-mail address: akihiro.eguchi@psy.ox.ac.uk (A. Eguchi).

Fig. 1. Rubin's Vase (Rubin, 1915).

the border ownership signals arise. Therefore, they concluded that context influence by horizontal signal propagation alone is highly unlikely.

On the other hand, the feedforward (bottom-up) and feedback (top-down) connections between successive visual stages have fast-conducting axons, with conduction velocities of between 2 and 6 m/s, which is about ten times faster than cortical horizontal fibres (Angelucci & Bullier, 2003). Accordingly, both Craft, Schtze, Niebur, and von der Heydt (2007) and Jehee, Lamme, and Roelfsema (2007) have proposed models that involve hypothetical 'grouping circuits' within a higher cortical layer that capture the contextual information about local boundary elements, and these contextual signals are then relayed down through feedback connections to modulate responses in an earlier layer. They proposed that the larger receptive fields in the higher layer allow the network to employ 'grouping circuits' without having to rely on slow lateral propagation of signals. Nevertheless, it still remains a challenge to understand exactly how such neural circuits may be learned. The objective of the current study is to investigate the learning mechanisms that underpin the development of border ownership cells in the primate visual brain, in terms of synaptic modification guided by visual experience and consequent neural adaptation throughout a hierarchy of cortical stages. Moreover, given the proposed role of border ownership cells in feature binding, which is essential for integrating the visual features within a scene, the simulations described below provide a step towards understanding how the brain learns to make sense of the visual world.

One higher visual area that might provide appropriate top-down modulatory signals is V4, which contains neurons that represent the localised boundary contour elements of objects (Layton, Mingolla, & Yazdanbakhsh, 2012). The responses of these neurons are sensitive to both the shape of the boundary element and where the element is with respect to the centre of mass of the object (Pasupathy & Connor, 2001; Pasupathy & Connor, 2002). Hence each of the neurons encodes that a specific border element belongs to a particular object - i.e. a kind of border ownership representation. A subpopulation of these neurons will provide a distributed representation of the entire boundary of the object. Furthermore, the neurons are able to respond invariantly as the object is shifted across different locations on the retina over a modest range.

The visually-guided development of such V4 cells has been previously investigated in a computational modelling study with an established neural network model, VisNet, of the primate ventral visual pathway (Eguchi, Mender, Evans, Humphreys, & Stringer, 2015). The network architecture consisted of a hierarchy of cortical visual layers, with each layer modelled as a competitive neural network (Wallis & Rolls, 1997). Whenever an image was presented to the network, visual signals propagated through feedforward plastic synaptic connections between successive layers. Within each competitive layer, the excitatory cells competed with each other to respond to the current visual stimulus. In the brain, competition between excitatory cells is implemented via inhibitory interneurons. Although to save computational expense in VisNet, competition between excitatory neurons is modelled more directly using local filters. During an initial period of training with visual objects, the feedforward synaptic connections between successive layers of the network are continually modified using local, biologically plausible, associative learning rules. The competition within each layer then forces individual neurons to learn to respond selectively to a particular stimulus class, with different neurons responding to different kinds of stimulus. Competitive learning is a very simple unsupervised learning paradigm that allows neurons to discover important features of the stimulus input patterns (Rumelhart & Zipser, 1985). Eguchi et al. (2015) showed that the gradual increase in the receptive field size of neurons through successive layers of the visual system (Gross, Bender, & Rocha-Miranda, 1969; Pettet & Gilbert, 1992) allows V4 neurons access to local image information specifying how localised luminance contrast contours belong to adjacent object regions. As a result, cells in the higher layer of their hierarchical competitive neural network model developed neuronal response properties similar to those reported by Pasupathy and Connor (2001, 2002) when the model was trained on a number of real world objects.

In this paper, we extend the previous purely feedforward model of Eguchi et al. (2015) by incorporating both feedforward (bottom-up) and feedback (top-down) connections. This extended model architecture is used to investigate how the edge-detecting simple cells in the earliest layer of the network, which corresponds to visual areas V1/V2 in the primate brain, may develop border ownership representations via top-down modulation from neurons in the output layer, which corresponds to visual area V4. The necessary feedforward and feedback synaptic connectivity within the network is set up by visually-guided learning using a biologically plausible, local, trace learning rule (Foldiak, 1991) as the network is trained on a collection of differently shaped visual object stimuli. We go on to show how these border ownership signals in the earliest layer evolve dynamically during the 300 ms time course of a stimulus presentation, as reported by Sugihara et al. (2011) and Jehee et al. (2007). We then investigate the limitations of the model in the more general situation where multiple objects are presented to the network simultaneously.

1.1. Hypothesis

Eguchi et al. (2015) have shown that when an established hierarchical neural network model of the primate ventral visual pathway, VisNet (Wallis & Rolls, 1997), is trained on 177 images of real world objects, which rotated in plane through 360° and shifted across a 3 × 3 grid of nine different retinal locations, the neurons in the higher layers of the model learn to represent local boundary contour elements. Individual neurons are tuned to boundary elements with a specific curvature at a particular location with respect to the centre of mass of the object. Moreover, the neurons respond invariantly as an object is translated across different retinal locations. These are the same neuronal response properties as observed in area V4 of the primate visual system by Pasupathy and Connor (2002). Although they have reported that the translation invariant responses of V4 neurons are only over a modest