Vision Research 119 (2016) 16–28

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

A computational exploration of complementary learning mechanisms in the primate ventral visual pathway

Courtney J. Spoerer *, Akihiro Eguchi *, Simon M. Stringer

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

A R T I C L E I N F O

Article history:
Received 11 March 2015
Received in revised form 21 September 2015
Accepted 8 December 2015
Available online 19 January 2016

Keywords:
Visual object recognition
Continuous transformation
Trace learning
Inferior temporal cortex

A B S T R A C T

In order to develop transformation invariant representations of objects, the visual system must make use of constraints placed upon object transformation by the environment. For example, objects transform continuously from one point to another in both space and time. These two constraints have been exploited separately in order to develop translation and view invariance in a hierarchical multilayer model of the primate ventral visual pathway in the form of continuous transformation learning and temporal trace learning. We show for the first time that these two learning rules can work cooperatively in the model. Using these two learning rules together can support the development of invariance in cells and help maintain object selectivity when stimuli are presented over a large number of locations or when trained separately over a large number of viewing angles.

1. Introduction

In vision, it is important to correctly identify an object in the environment as being the same despite changes in the retinal image. Over successive stages in the visual system, neurons develop response properties that are invariant to the size, position, and view of an object (Rolls, 1992; Rolls, 2000; Rolls & Deco, 2002; Desimone, 1991; Tanaka, Saito, Fukada, & Moriya, 1991). Cells in inferior temporal cortex (IT) that show invariance to the translation (Op de Beeck & Vogels, 2000; Kobotake & Tanaka, 1994; Ito, Tamura, Fujita, & Tanaka, 1995; Tovee, Rolls, & Azzopardi, 1994), size (Rolls & Baylis, 1986; Ito et al., 1995), contrast (Rolls & Baylis, 1986), lighting (Vogels & Biederman, 2002), spatial frequency (Rolls, Baylis, & Leonard, 1985; Rolls, Baylis, & Hasselmo, 1987), and view (Hasselemo, Rolls, Baylis, & Nalwa, 1989; Booth & Rolls, 1998) of objects have been reported.

Developing invariant recognition of objects involves associating together representations of the same object under different conditions. In the particular case of translation invariance, this would mean developing associations between the neural representations of an object in different spatial locations on the retina. In order to develop these associations, the visual system can exploit constraints placed upon object translation by the environment. For example, when an object translates from one point to another, it does so in a manner that is continuous in both space and time. These same constraints can be exploited for the development of view invariance, as different views of an object also appear in a spatially and temporally continuous manner.

One method for developing translation invariant representations utilizes the temporally continuous nature of object translation. Neurophysiological evidence suggests that the brain might use this type of information to develop translation invariant representations of objects (Li & DiCarlo, 2008). As breaking temporal continuity causes neurons to lose their selective responses to different objects. Different approaches have been developed in order to understand how the brain might exploit this temporally continuity, such as using inputs representing temporal context to guide learning (Becker, 1999), learning high probability sequences of visual input in order to infer the object being presented (George & Hawkins, 2005), and extracting slowly changing features in the visual inputs to analyze the transform invariant representations (Berkes & Wiskott, 2005; Wiskott & Sejnowski, 2002).

Temporal information can also be used to develop invariant representations of objects by incorporating a temporal trace into associative learning rules (Földiák, 1991; Rolls, 1992; Wallis & Rolls, 1997). This encourages neurons to respond to stimulus image transforms that occur close together in time. The advantage of this approach is that it can arise naturally out of biophysically realistic spiking neural networks when longer time constants for synaptic conductance are introduced (Evans & Stringer, 2012). Increasing this time constant keeps the neuron active for longer as it lengthens the time period over which current leaks into the

---

* Corresponding authors.
E-mail addresses: courtney.spoerer@mrc-cbu.cam.ac.uk (C.J. Spoerer), akhiro.eguchi@psy.ox.ac.uk (A. Eguchi).

http://dx.doi.org/10.1016/j.visres.2015.12.008
0042-6989/© 2016 The Authors. Published by Elsevier Ltd.

postsynaptic neuron, thus allowing temporal trace learning to occur. Therefore, it is feasible that this type of learning could occur in the brain without requiring a specific architecture to operate.

A second method for developing translation and view invariance, known as continuous transformation (CT) learning, depends on the spatial continuity of object transformation (Stringer, Perry, Rolls, & Proske, 2006). As an object moves smoothly from one location to another, it will also appear in several intermediate positions. Each of these intermediate positions will be highly overlapping with the adjacent locations that the object appears in as it moves across the environment. Therefore, each of these adjacent locations would be likely to activate a common postsynaptic neuron that associates each of the positions together. This would result in the cell developing translation invariant response properties.

Each of the methods discussed so far consider how spatial and temporal constraints could each individually contribute to the development of invariant representations. However, in the real world, information provided by each of these constraints is available to the visual system simultaneously. Psychophysical evidence suggests that object-selective view-invariant recognition is improved when stimuli transform in a temporally and spatially continuous manner, compared to spatially continuous transformation alone (Perry, Rolls, & Stringer, 2006). It is important to understand how an observer might simultaneously utilize the benefits of spatial and temporal continuity in object transformation when developing invariant representations. This effect could be explained by the visual system using CT learning and temporal trace learning in tandem.

In this paper, we will explore how CT learning and temporal trace learning can operate together to help develop view and translation invariance using a hierarchical model of the ventral visual pathway, VisNet (Wallis & Rolls, 1997; Rolls & Milward, 2000), illustrated in Fig. 1. Both trace and CT learning have been tested extensively in the rate-coded VisNet model (Wallis & Rolls, 1997; Stringer et al., 2006), and so we shall use VisNet to study how these two learning mechanisms may be combined in the same rate-coded model.

2. Methods

2.1. The VisNet model

2.1.1. Hierarchical neural network architecture of the model

The architecture of the model used in this paper, VisNet (Wallis & Rolls, 1997), is developed according to the following principles:

(i) A series of hierarchical competitive networks with local graded inhibition and excitation.
(ii) Convergent connections to each neuron from a topologically corresponding region of the preceding layer.
(iii) Synaptic plasticity based on a biologically-plausible local learning rule, such as the Hebb rule or trace rule.

As mentioned above, the forward connections to individual cells in VisNet are derived from a topologically corresponding location in the preceding layer. The probability of each connection forming follows a Gaussian distribution. These distributions are defined by a radius containing approximately 67% of the connections from the preceding layer. The values employed in the current study are given in Table 1. The gradual increase in the receptive field of cells in successive layers reflects the known physiology of the primate ventral visual pathway (Freeman & Simoncelli, 2011; Pasupathy, 2006; Pettet & Gilbert, 1992).

2.1.2. Pre-processing of the visual input by Gabor filters

Before images are presented to layer 1 of VisNet, they are preprocessed by a set of Gabor filters that correspond to the known response profiles of V1 simple cells (Jones & Palmer, 1987; Cumming & Parker, 1999). Filtering the images produces a unique set of inputs that are then presented to layer 1 of the model. The input filters used are computed by the following equations:

$$g(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right) \tag{1}$$

with the following definitions:

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \end{aligned} \tag{2}$$

where x and y specify the position of a light impulse in the visual field (Petkov & Kruizinga, 1997), σ controls the number of such periods inside the Gaussian window, θ defines the orientation of the feature, ψ defines the phase, and γ sets the aspect ratio that determines the shape of the receptive field. In each experiment, an array of Gabor filters is generated at each of 256 × 256 retinal locations with the parameters given in Table 2.

The outputs of the Gabor filters are passed to the neurons in layer 1 of VisNet according to the synaptic connectivity given in Table 1. Each layer 1 neuron received connections from 400 randomly chosen Gabor filters within a topologically corresponding region of the retina.

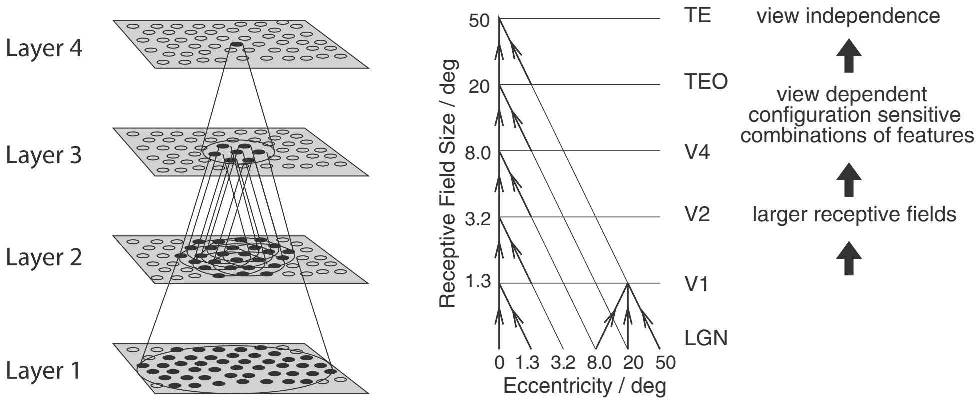

Fig. 1. (Left) A schematic representation of VisNet. The model consists of a hierarchy of competitive networks with feed-forward connections between them. Convergence in the model is designed so that cells in the final layer of the model have a receptive field that covers the whole of the input retina. (Right) Convergence in the visual system. V1, visual cortex area V1; TEO, posterior inferior temporal cortex; TE, inferior temporal cortex (IT).