The neural network for view-invariant object recognition and classification

Natalia Efremova, Nobuhiko Asakura and Toshio Inui

Department of Intelligent Science and Technology, Graduate School of Informatics, Kyoto University
Faculty of Technology Bldg.#2, Yoshida-honmachi, Sakyo-ku, 606-8501 Kyoto, Japan
E-mail:(efremova,asakura,inui)@cog.ist.i.kyoto-u.ac.jp

Abstract— We propose a neural network of ventral visual pathway for recognition and classification of 3D real-world objects. The model consists of hierarchy of modules, representing the visual areas V1-V4 and inferotemporal cortex (IT). The architecture of the model represents the self-organized map (SOM) of functional radial-basis function (RBF) modules. We show that the proposed architecture is capable of recognizing and categorizing 3D real-world objects in a way consistent with recent fMRI findings.

Keywords— Neural Network, Object Recognition, Object Classification, Mental Rotation, Inferotemporal cortex

1. Introduction

It has been reported that the primate visual system is able to perform classification and categorization of objects based on their shape, and that humans rely heavily on shape similarity among objects for object categorization and identification [1].

The visual information is processed in two major parallel pathways: ventral and dorsal. The ventral pathway is involved in the recognition of color, shape and texture of the objects. It includes visual areas V1, V2, V4, and IT. These areas are organized in a retinotopic manner, but with different degrees of resolution [2].

In this study, we propose a cortical architecture for hierarchical visual perceptual processing, resembling the ventral visual stream in primate visual cortex. We show that a neural network model based on this architecture can perform recognition and classification of objects. We further show that the model can create a similarity map of the objects. Finally we compare the model’s performance with experimental data from humans and non-human primates.

2. Model architecture

The first level of the model consists of local orientation detectors that model simple cells in the primary visual cortex (V1). These detectors are Gabor-like filters, which are reported to be similar to the receptive fields (RFs) in V1. The orientation detectors have four preferred orientations. The RF sizes of these detectors correspond to those in monkey V1. The next level contains position-invariant bar detectors, corresponding to complex-like cells in area V1, or to neurons in areas V2 and V4. The combinations of the features extracted at these earlier stages are then processed with the neurons of IT cortex.

The architecture of IT cortex module is based on the notion of modular-network SOM [3]. The IT cortex is represented in the model by the SOM of RBF-units. The output of the network is presented by hundred SOM-units, organized in a square lattice. The hidden units of the RBF network are Gaussian functions. The learning algorithm for SOM of RBFs can be described in terms of four processes. First, the outputs of all functional modules are evaluated for each input-output data vector pair. After evaluating all outputs for all inputs, the errors of all modules for each data class are evaluated. The module which minimizes the error is determined as the best matching module, i.e. the winner module. Further, the learning rate of each module is adjusted according to a neighborhood function: the further away from the winner, the smaller the learning rate. Finally all functional modules are updated by the back-propagation learning algorithm. The back-propagation learning is repeated until all modules are sufficiently updated. These processes are repeated until the network gets to a steady state. The detailed description of the model can be found in [4].

3. Simulation experiment

We conducted a simulation experiment to investigate this network’s ability to classify objects and to create a similarity map of the objects. In the experiment, we used three computer-generated novel 3D objects as stimuli (Fig.1). The inputs of the network were individual views of each object rotated in depth in ten-degree steps.

Fig.2 shows the result of the simulation. We found that the majority of SOM-units exhibit tuning to a particular object. The best view of the tuned object (i.e. the one that causes the best response in the RBF module) is depicted on the corresponding lattice square. We can see that different objects are represented in different regions of the output map. This demonstrates that the network can classify the input objects and form a similarity map based on their shape.

In the resulting map each of the input objects is recognized by nearly equal number of SOM-units. Within the region of SOM-units responsible for each input object, the best views differ, but similar views appear to be located nearby. Thus, the network is likely to classify objects according to their prototypical views. Taken together, these results suggest that the proposed network has an ability to categorize and represent objects in two dimensions: shape and view similarity.

Figure 1. Stimulus set used in the experiments.
The behavior of the proposed network is consistent with the preferable view. The activation graph usually has one peak for a specific view of the object, and the activation level declines smoothly during the rotation of the object in-depth. In some cases the activation graph has two peaks: the second peak shows the activation of the hidden unit for the presentation of the mirror-view counterpart of the object, and the activation level declines smoothly. The activation graph usually has one peak for a specific view of the object, and the activation level declines smoothly during the rotation of the object in-depth. In some cases the activation graph has two peaks: the second peak shows the activation of the hidden unit for the presentation of the mirror-view counterpart of the preferable view.

4. Discussion.

The behavior of the proposed network is consistent with the corresponding properties of monkey IT cortex [6]. In addition, our previous study has shown that the method of storage of information in the SOM of RBFs is similar to the organizational map of the IT region: the inner representations of the input objects in the RBF centers resemble the columnar organization of the IT cortex [4]. Furthermore, the organization of the similarity map generated by the current model can be compared with the pattern of horizontal activation of the IT area, as was described by Tanaka [7].

Op de Beeck et al. obtained clear evidence of the existence of a large-scale shape map in monkey IT cortex from their recent fMRI study [8]. Their experiments revealed topography of selectivity of IT neurons for novel 3D object classes. In the current study, using the same type of objects as stimuli, we obtained the similarity map of those objects that is qualitatively similar to the shape map in the IT cortex.

Additionally, we would like to point out that the similarity map can qualitatively reproduce the pattern of activation of the IT region revealed by optical imaging experiments [9]. The results of these experiments showed that the presentation of a single feature activated multiple spots in the IT region. They further revealed a partial overlapping between the spots evoked by similar features. We found that these properties are consistent with the activation of SOM-units around the borders between different regions in the similarity map, each of which is tuned to a particular object.

Riesenhuber and Poggio proposed a basic scheme for hierarchical visual processing in the ventral pathway [10]. Visual areas in the ventral pathway (V1-V4, IT, PFC) were considered as hierarchically organized modules. This constitutes a hierarchy of increasingly sophisticated representations, extending the model of simple to complex cells of Hubel and Wiesel [11]. The current study further extends this architecture by adding the representations of shape and view similarity to the IT cortex module.

A natural extension of the current study is to construct the global neural network for object recognition that includes the dorsal pathway for mental rotation. Inui and Ashizawa have proposed a basic architecture of such a network [12]. This model consists of two subsystems, one for mental rotation, which is a model of parietal network, and the other for object recognition, which is a model of inferior temporal cortex. The introduction of a mental-rotation type mechanism may improve the current model’s robustness in recognition of objects from non-canonical viewpoints.

References