Learning Features and Transformation Encoding based on a Generative Horizontal Product Model

J. Zhong, C. Weber, S. Wermter

University of Hamburg, Department of Informatics, Knowledge Technology, Vogt-Kölln-Straße 30, D - 22527 Hamburg, Germany

http://www.informatik.uni-hamburg.de/WTM/

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Biological Evidence

- According to the theory of two parallel visual pathways, the 'dorsal pathway' encodes transformations, invariant of stimulus-specific properties, while the 'ventral pathway' encodes object identity, invariant of positions and sizes, etc.
- Because of the delays from upstream and downstream neural transmission, the 'where' pathway should maintain a future position of an object, which can be accounted for by the representation of movement direction and velocity in the dorsal pathway. This is also evidenced by recordings in some complex cells, which are direction- and speed selective independent of spatial frequency, hence resembling neurons in MT of the dorsal pathway [2] (they predict 'where').

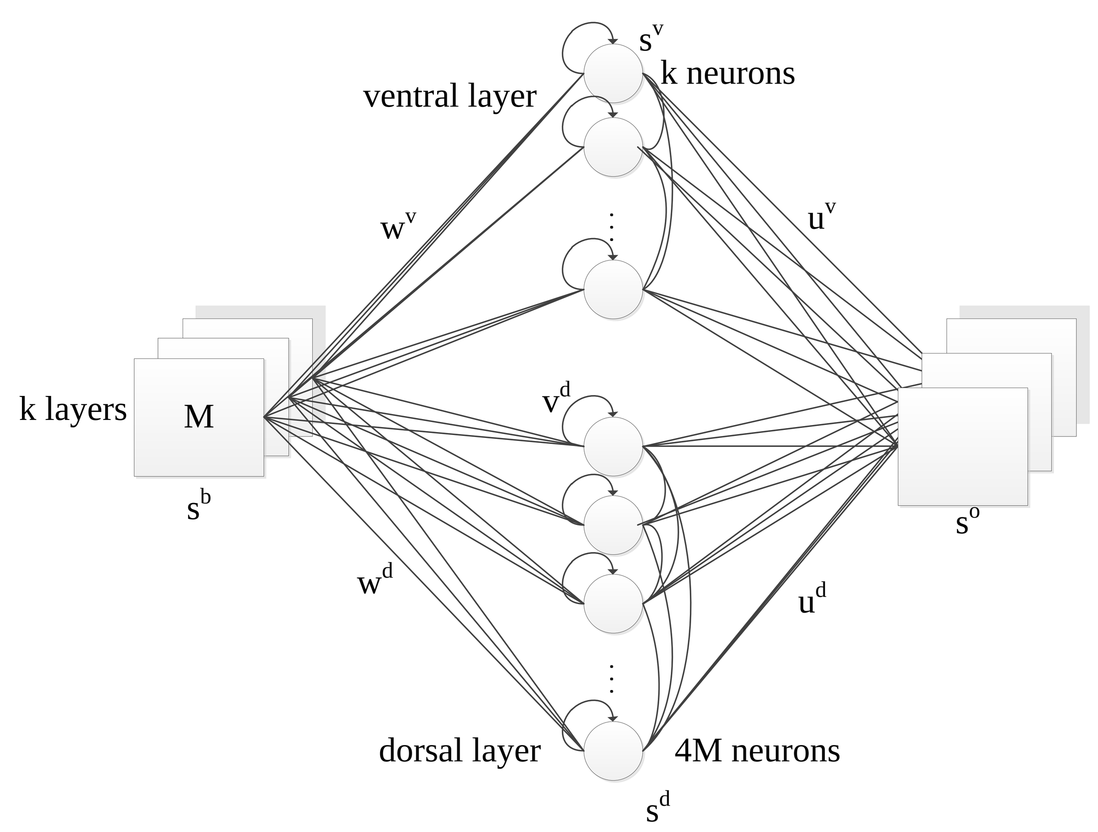
Motivation

- The horizontal product model together with Independent Component Analysis (ICA) [1] was applied to separate the location of image features from their identities.
- From this ICA model, we propose a method that can extract two or more components of information into separate pathways from input data. Unlike [1], our model encodes motion to predict its *future* input: both pathways incorporate recurrent connections so as to capture the observed response properties of complex cells.
- The output is then generated by multiplying outputs from sub-models via a horizontal product. The horizontal product model reduces computational effort: assuming that there are I input units, considering T transformations and F features, a full bilinear model has $I \times T \times F$ connections, but only $2I \times (T + F)$ connections are needed in this network.

Model

- A three layer architecture which learns object identity and position in an unsupervised fashion based on a predictive model.
- The hidden layer contains two independent sets of neurons which represent 'object transformation' and 'object identity'.
- The recurrent connections in the hidden layers predict movement in the dorsal-like layer and maintain a persistent representation of an object in the ventral-like layer.
- The network output s^o is obtained via the horizontal product

$$s^o = s^v u^v \odot x^d u^d$$



Algorithms

- The hidden units' activities of both pathways are defined as

$$y_j^{v,d}(t) = \sum_i s_i^b(t) w_{ji}^{v,d} + \sum_i s_i^b(t-1)\bar{w}_{ji}^{v,d} + \sum_{j'} s_j^{v,d}(t-1) v_{jj'}^{v,d}$$

- The transfer functions in both hidden layers employ a logistic function and a soft-max function (omitting the superscripts of v, d):

$$z_j = \frac{1}{1 + exp(-a_j y_j + b_j)}$$
$$s_j = \frac{exp(z_j)}{\sum_{j'} exp(z_{j'})}$$

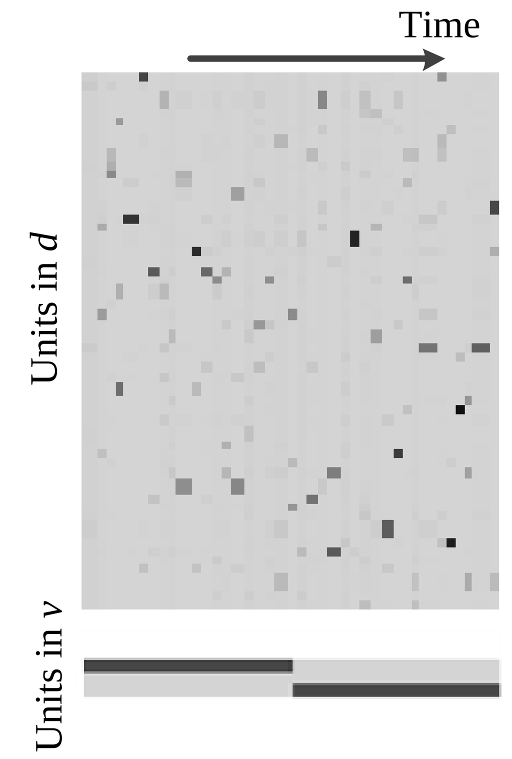
- The logistic function has two local modifiable parameters a and b, leading to regular firing on the hidden layer, which is inspired by the intrinsic plasticity of neurons [4].
- The training progress is determined by a cost function:

$$C = \frac{1}{2} \sum_t^T \sum_k^N (s_k^b(t+1) - s_k^o(t))^2$$

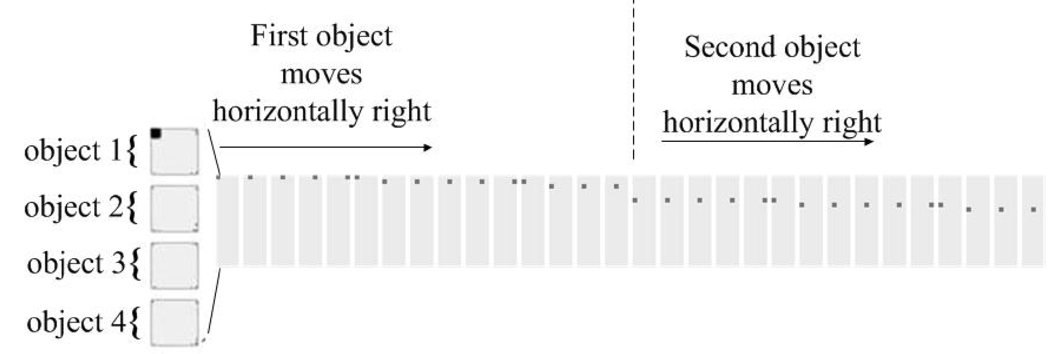
where $s_i^b(t+1)$ is the one-step ahead input, as well as the desired output, $s_k^o(t)$ is the current output, T is the total number of available time-step samples and N is the number of output nodes, which equals the number of input nodes.

Experiment

The artificially generated input data mimics moving objects, i.e. their positions change quickly but their identity changes rarely. In this dataset, only one object appears at one unique position in any time-step. This minimalistic set up sketches a hyper-column in V1 that processes oriented lines of 4 different orientations at 5×5 possible positions.



(a) Network activations in hidden layers.



(b) Partial training samples while one object moving horizontally rightwards.

(c) Network output given the above input.

Discussion

- The object identity and position have been successfully separated in an unsupervised manner; especially the activation in the dorsal-like units are analogous to the recording of 'direction-selective' complex cells in V1; ventral-like cells are transformation invariant.
- Recurrent connections are highlighted in this model, which store previous movement information, and serve a predictive function. Prediction in the visual system is apparent by neurophysiological findings of predictive receptive field shifts [3] and behavioral findings of visual responsibility in movement prediction [5]. We believe that any cortical area should compensate its processing delays via prediction.

References

[1] U. Köster, J. Lindgren, M. Gutmann, and A. Hyvärinen. Learning natural image structure with a horizontal product model. *Independent Component Analysis and Signal Separation*, pages 507–514, 2009.

[2] N. Priebe, S. Lisberger, and J. Movshon. Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *J Neurosci*, 26(11):2941–2950, 2006.

[3] M. Sommer and R. Wurtz. Influence of the thalamus on spatial visual processing in frontal cortex. *Nature*, 444(7117):374–377, 2006.

[4] C. Weber and J. Triesch. A sparse generative model of V1 simple cells with intrinsic plasticity. *Neural comput*, 20(5):1261–1284, 2008.

[5] M. Wexler and F. Klam. Movement prediction and movement production. *J Exp Psychol*, 27(1):48–64, 2001.

Acknowledgment

This research has been partially supported by the EU projects RobotDoc under 235065 from the FP7, Marie Curie Action ITN and KSERA by the European Commission under n°2010-248085.