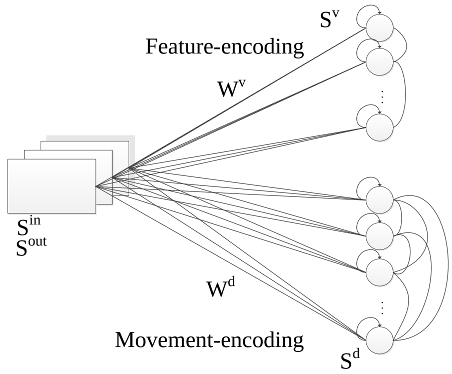
Learning Features and Transformations with a Predictive Horizontal Product Model

J. Zhong, C. Weber, S. Wermter
Department of Computer Science, University of Hamburg, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany
{zhong,weber,wermter}@informatik.uni-hamburg.de

According to the theory of two parallel visual pathways, the 'dorsal pathway' encodes spatial information, invariant of stimulus-specific properties, while the 'ventral pathway' encodes object feature identity, invariant of positions and sizes. Commonly, they are referred to as the 'where' and 'what' pathways. Such a distinction between spatial transformation- and identity encoding cells is already evident in the diverse response properties of V1 complex cells [2]: some complex cells are direction- and speed selective, independent of spatial frequency, hence resembling neurons in Medial Temporal Lobe(MT) of the dorsal pathway (they predict 'where'). Other complex cells are selective to spatial frequencies independent of speed, hence coding feature identity ('what'). Interestingly, to compensate for upstream and downstream neural transmission delays, the 'where' pathway should maintain a *future* position of an object. This can be accounted for by the representation of movement direction and velocity in the dorsal pathway. It is important, for instance, when a person detects a temporal pattern change in visual target stimulus.

We propose a new architecture which learns object position/motion and feature identity in an unsupervised fashion based on a predictive model. The network output S^{out} at time t predicts the input S^{in} at future time step $t+1$, which is expressed as a cost $(S^{out}(t) - S^{in}(t+1))^2$. As shown in the figure, we separate the encoding of 'object movement' and 'object feature' in the dorsal-like layer and the ventral-like layer respectively. While both pathways are structured symmetrically, the encoding of an object's position and movement is encouraged in the dorsal-like layer by fast reacting cells. Position- and motion-invariant encoding of identity is encouraged in the ventral-like layer by a slow activity trace. Due to the existence of direction-selective cells, the information predicts the input's future position taking into account its moving direction. To combine the 'what' and 'where' representations at the level of the output, the horizontal product is used: the output is generated by multiplying outputs of sub-models by $S^{out} = S^d W^d \odot S^v W^v$, where $\odot$ indicates element-wise multiplication, S^d and S^v represent activations in dorsal-like and ventral-like layers, W^d and W^v represent corresponding weighting matrices. A similar division of 'what" and 'where' information using horizontal product has been done by Köster et al. [1] together with Independent Component Analysis, however without specific movement representation. The horizontal product model keeps computational effort limited. For example, if there are I input units, considering T transformations and F features, a full bilinear model has $I \times T \times F$ connections, while the horizontal product model uses only $2I \times (T + F)$ connections. Using the horizontal product in reconstruction, we can obtain movement and feature predictions; in this way, the reconstruction is able to predict the movement assuming the same object identity appears in the visual field.



The experimental results based on artificially generated data show that information of object feature and position have been successfully separated in an unsupervised manner: the activations of feature-encoding units remain stable when a given object moves over the input layer, while different patterns appear in the movement-encoding units indicating the movement directions. In particular, the fast responding dorsal-like cells have become direction selective while the slow responding ventral-like cells encode 'object identity'. These experimental results are analogous to the recording of different populations of complex cells in V1 [2]. The predictive function of our model, which is motivated by neurophysiological findings of predictive receptive field shifts [3] and behavioural findings of visual responsibility in movement prediction [4], could be further implemented in robot vision applications.

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

In: Proc. 16th International Conference on Cognitive and Neural Systems (ICCNS), 2012