Towards unsupervised learning of invariance to a large class of transformations

W. Einhäuser, C. Kayser, K. P. Körding, P. König - Institute of Neuroinformatics (UNI/ETH Zürich)

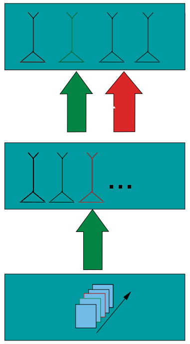
Eidgenössische Technische Hochschule Zürich

Abstract

Humans and animals can recognize objects nearly invariantly to various transformations, such as translations, scaling or changes in illumination or viewpoint. However, it is not a priori obvious, which features of an object are relevant to its recognition. This is the rationale to build systems which themselves learn to extract the relevant information from generic properties of their input. Here we present a model which exploits the fact, that different features of natural scenes change on different temporal scales. In a self-supervised scheme cells of the model integrate information over an extended period of time. By only maintaining information which is common to stimuli over this period, fast varying information is discarded, while slowly changing information is preserved, yielding invariance to fast variables. We show, that this model is capable of learning the translation invariance of complex cells from natural image sequences. Furthermore, invariance to other variables such as orientation and color can be obtained by just varying a single parameter, namely by increasing the time constant the system is operating on. Additionally, by de-correlating groups of cells with different time constants, cells can be obtained which are invariant to slower variables, while still specific to faster ones. Extending the proposed model into a hierarchical scheme might thus be a fruitful approach in pushing forward invariant object recognition.

The model

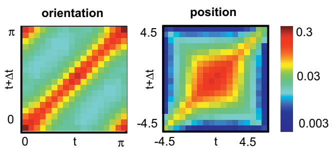
Based on the classical Hubel & Wiesel model we propose a feedforward architecture consisting of three layers, the input layer, the middle layer and the top layer. The input layer receives a continuous stream of 10x10 pixel wide patches drawn from natural videos (see below). The middle layer employs a competitive Hebbian learning rule and acquires simple cell like receptive fields. Founded on recent physiological findings, that the induction of LTP vs LTD depends on the relative timing of pre- and postsynaptic spike, a the top layer cell of highest activity increases its connection to a middle layer cell if the middle layer cell was highly active a time-step Δt before. This enables the network to learn information which is correlated over time while discarding temporally uncorrelated information. Discarding part of the information enables the network to acquire invariance to vast varying variables.



The input

In order to approximate the input the visual system is naturally exposed to, we record videos by a camera mounted to the head of a freely behaving cat. This results in a continuous input stream which preserves the temporal structure of the cat's visual input.
Since the learning rule exploits temporal correlations, we investigate the correlation of local orientation and position over time. We find that orientation changes slower than position. This generic property of natural scenes enables the learning rule to learn translation invariant orientation detectors - complex cells.

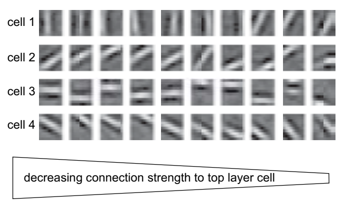
Recording setup for natural videos.



In natural scenes position changes faster than orientation.

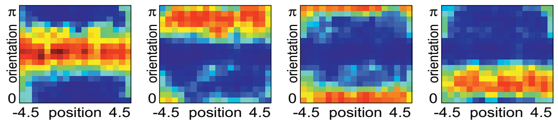
Translation Invariance of Top Layer Cells

The figure to the right depicts the receptive fields of 40 middle layer cells. All of these are tuned for orientation and position as typical for cortical simple cells.
The receptive fields are sorted according to the strength of their synaptic connection to different top layer cells. Each top layer cell connects strongest to simple cells which share the same orientation preference irrespective of their position tuning. Thus, the top layer neurons exhibit translation invariance like cortical complex cells.



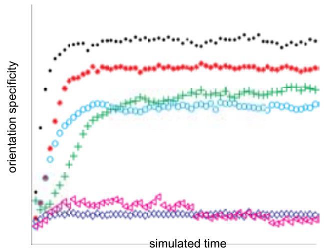
Top layer cell learn complex cell properties

In order to quantify the responses of top layer cells, we probe a converged network with bar stimuli. The responses of the four top layer cells are depicted colorcoded below. Whereas all show a precise orientation tuning, none of them exhibits any tuning towards position of the stimulus. Thus we can conclude that these cells detect stimulus orientation independently from the position of the stimulus and thus resemble complex cells.



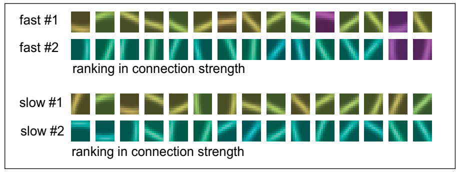
Stability & controls

A critical issue for network models is the stability of the acquired network properties over time. The red graph on the right shows that top layer cells learn complex cell properties which remain stable over time. The properties are independent of the total number of middle layer cells (black, cyan and green plots). However, complex cell properties are not learnt if the temporal structure of the input is impaired (magenta and blue plots). This shows that only the combination of a **temporal learning rule** and the natural **temporal structure of the input** is sufficient to learn the properties of complex cells.



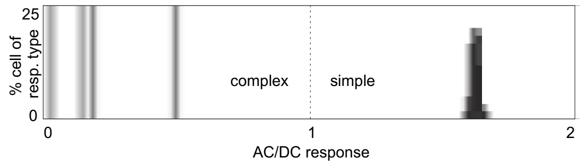
Extension to multiple variables

The model can straightforward be extended to learn invariance to other variables. By increasing the time constant Δt the model obtains invariance to transformations other than translation. In the example below all top layer cells are specific to color, while only fast cells (small Δt) are specific to orientation, whereas slow cells are invariant to orientation. As before all of those cells are invariant to translation. This indicates that higher level invariances can also be learnt in the proposed framework.



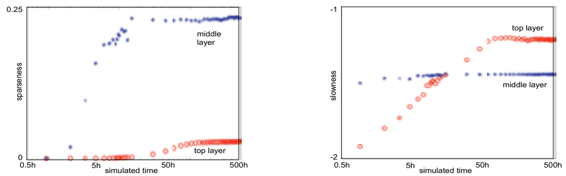
Relation to physiological results

A measure commonly used to classify simple from complex cells, is the AC/DC ratio of a cell in response to a drifting sinewave grating. A cell is classified simple if this ratio is larger than 1 and complex otherwise. We find that also according to this criterion the model's top layer neurons are classified as complex cells and the middle layer cells as simple.



Relation to Goalfunction approaches

A number of recent studies addresses the possibility that simple as well as complex cells are optimized with respect to certain analytically given Goalfunctions. The graphs below show that two of the most common objectives - sparseness and slowness - increases for middle layer cells during learning, whereas only slowness shows a significant increase for top layer cells. Therefore the present study is in compliance with studies suggesting that simple and complex cells optimize sparseness, whereas only complex cells optimize slowness.



This work was supported by Honda R&D Europe - Germany (WE), the center of neuroscience Zurich - ZNZ (CK), the Collegium Helveticum (KPK) and the Swiss National Fond - grant no. 31-65415.01 (PK)