

Dynamic, context-dependant spike response of visual neurons reflects inference and learning in natural movies

Sophie Deneve, Timm Lochmann, Udo Ernst

Group for Neural Theory, Ecole Normale Supérieure, Collège de France, CNRS.

Experimental studies have demonstrated that stimuli outside the classical receptive field (cRF) of a neuron can nevertheless strongly modulate the cell's response to stimuli inside the cRF. These non-classical receptive field (ncRF) properties can drastically change the neuron's behaviour, e.g. its preferred orientation. They also act differently depending on stimulus contrast, enhancing the firing rate when the contrast inside the cRF is low, and suppressing the activity when the contrast is high. The extent and influence of the cRFs and ncRFs varies over time: RF sharpen and preferred spatial frequency increases at longer delays. These observations hint at a coarse to fine coding continuum as more information becomes available: At low contrast or for short presentations, neurons integrate information from a wider range of channels, decreasing detection threshold at the cost of a lower precision. At high contrast and longer presentations, the summation field sharpens as neurons achieve higher precision. Several mechanistic models have been proposed to account for the observed phenomena. However, few studies have investigated the computational role of ncRFs. Here we show that ncRF properties and their contextual dependencies emerge from natural visual statistics, more precisely, from detecting the presence of dynamic, independent features in natural movies. Therefore, by drawing a parallel between the spatio-temporal statistics of the visual input and the plasticity and dynamics of a spiking neural networks, this computational model proposes a new implementation for ncRFs modulatory effects.

The assumption that neural systems learn the underlying cause of their sensory input have successful accounts for linear visual response properties, e.g. for the shape of cRF of V1 simple cells. However, sensory systems have to estimate the state of constantly changing variables (i.e. the presence of oriented edges in a movie) from noisy streams of stochastic data (photons, spikes), rather than static images. Natural visual input consists of objects that all appear and disappear randomly and independently, and superpose each other. When learning such a generative model, local spatial and temporal correlations in natural movies lead to center surround, and simple cell-like RFs. Moreover, due to the temporal dynamics embedded in the network, the predicted cRFs and ncRFs vary as a function of time and contrast, and account for a very large family of contextual effects in LGN and V1. Interestingly, such a model can be implemented in a spiking neural network, composed of adapting leaky integrate and fire neurons with feedforward connections and lateral divisive inhibition. The resulting model also accounts for detailed properties of spike response functions, such as the Poisson-like statistics, the transient nature of visual responses, and the dependency of the spike-triggered average on stimulus contrast. This model is not based on rate coding units, but on the assumption that spikes signal an increase in the probability of presence of the neuron's preferred object, and thus, extract and transmit the maximal amount of information about the objects composing a scene.

This approach provides several new insights about the nature of visual processing. First it suggest that visual neurons are better described by their "causal field" (i.e. the predicted consequence of the appearance of their preferred object in the scene, which is invariant) rather than their RF. Second, by taking time into account, it supports the idea that neural processing is adjusted to the spatial and temporal statistics of its input and involved in probabilistic inference and decision making at the level of single spikes. Finally, divisive inhibition (as opposed to subtractive inhibition) emerges as an essential component of visual processing to implement explaining away, i.e. solve ambiguities between different possible interpretations of a visual scene, the strongest inhibition occurring when the center and surround stimuli are identical, rather than being a non-selective form of gain modulation.