Modelling Complex Cells of Early Visual Cortex using Predictive Coding

Angelo Franciosini, Victor Boutin and Laurent Perrinet

Aix Marseille Univ, CNRS, INT, Inst Neurosci Timone, Marseille, France

Introduction. Predictive Coding (PC) is an influential framework introduced by Rao & Ballard [1] to model neural processes in the primary visual cortex of mammals (V1). PC exploits the hierarchical structure of sensory information into a bi-directional update scheme: Higher-level cortical layers predict at best the activity of the lower-level ones and send the prediction through feedback connections. This prediction is compared to the activity of the lower-level layers to generate a prediction error that is sent to the upper layer through feed-forward connections [2]. Interestingly, PC gives a possible explanation to extra-classical receptive fields effects in V1 [1], this is also in line with the abundance of feedback connectivity in the brain [3]. Additionally, this model has provided an elegant way to model task-driven learning in the brain by approximating error back-propagation, commonly used in deep neural networks, only by means of Hebbian plasticity and local computations [4, 5]. When implemented in a recurrent neural network, with the addition of sparsity constraints, PC can explain the emergence of edge sensitive cells in low-level visual areas as well as more specific descriptors in higher cortical areas [6]. We show that such a model, called Sparse Deep Predictive Coding network (SDPC), can also account for the topological organization of the primary visual cortex when imposing a max-pooling operator operator across small groups of neurons. Moreover, we show that the resulting model encodes for edges of specific orientation independently of their phase, a behaviour analogous to the one observed in neural recordings of complex cells [7, 8].
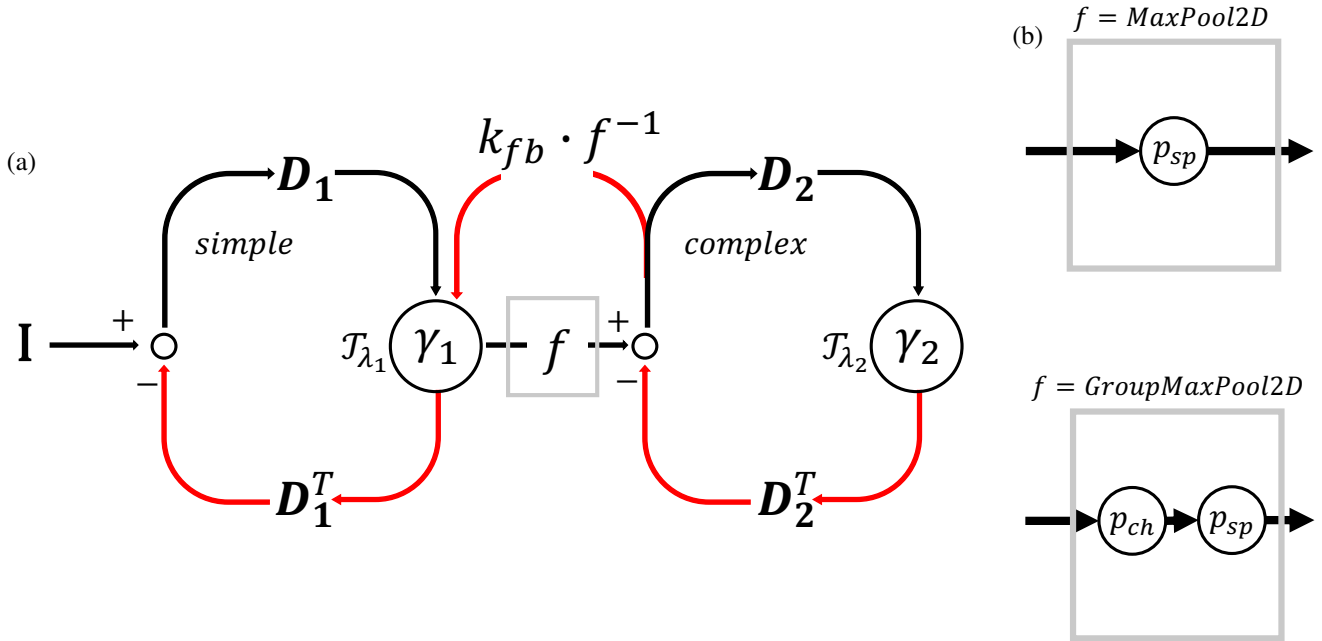

Figure 1: (a) Simplified state update scheme of the SDPC: black and red arrows indicate respectively excitatory and inhibitory connections. I is the input image, γ_1 and γ_2 are neural activity maps, D_1 and D_2 the convolutional kernels and $\mathcal{T}_\lambda$ the non-negative soft-thresholding operator as in [9]. The function f is a spatial transformation between layers and k_{fb} is a coefficient that regulates the strength of feedback information flow. (b) Scheme of the pooling functions $MaxPool2D$ and $GroupMaxPool2D$ where p_{sp} indicates a spatial max-pooling layer (with kernel size 2×2 and a stride of 2) and p_{ch} indicates a max-pooling across organized in 100 overlapping groups of 4 channels. The results shown in Fig. 2 (a) and (b) were obtained respectively by setting $f = MaxPool2D = p_{sp}$ and $f = GroupMaxPool2D = p_{ch} \circ p_{sp}$.

Methods. We implemented a model of cortical area V1 as a neural network composed of two layers, implementing simple and complex cells respectively. Our algorithm combines the architecture of a convolutional neural network with the predictive coding model proposed by Rao & Ballard [1] (see Fig. 1a) into a Sparse Deep Predictive Coding network (SDPC) architecture [6]. In this model, the state variables are sparse neural activity maps (γ_1 and γ_2) encoding the input and the feedback signals. Sparsity is imposed through a ℓ_1-norm penalty on the neural activity. The (learned) parameters of the system are convolutional kernels optimized through Hebbian learning. Note that these kernels describing the convolutional channels, D_1 and D_2, can be interpreted as synaptic weights shared across spatial locations. In order to replicate the behavior of complex cells, we extended the original algorithm [6] by introducing a more generic max-pooling function between the layers, as defined by f in Fig. 1-a,b [8]. The first one is a simple spatial pooling ($f = MaxPool2D$ in Fig. 1b), where $f(\gamma_1)$ is computed by selecting the maximum response of a neuron across different neighboring spatial locations of the activity map γ_1. Here, we added a second pooling function where in addition to the spatial pooling, the maximum activity is selected across small groups of neighboring neurons, or equivalently a two-phases pooling ($f = GroupMaxPool2D$ in Fig. 1b). The competition mechanism introduced by these operators enforces neighbouring neurons to encode for similar features. As the resulting SDPC network minimizes the prediction error ($f(\gamma_1) - D^T\gamma_2$), this should introduce some degree of tolerance with respect to small variations in the input image, making the complex-cells model account for non-linear relationships with the presented stimulus.

Results. To test this prediction, we trained two SDPC networks using $f = MaxPool2D$, and a $f = GroupMaxPool2D$, respectively. Both networks are composed of two layers characterized by 121 channels of convolutional neurons each. Each network was trained for 50 epochs on the STL-10 data-set of colored natural images [10]. Fig. 2 shows the 1st layer of kernels after learning. In presence of a classical spatial pooling ($MaxPool2D$) the network is able to extract localized edge detectors (Fig. 2a) analogous to the receptive fields of simple cells in V1 [11]. In this case, the disposition of the filters is invariant to permutations of the channels. While this model is able to account for invariance to small shifts of the input stimulus [12], it is not able to explain the disposition of edge sensitive simple cells in V1. On the other hand, when the two stages pooling is applied ($GroupMaxPool2d$), a topological structure emerges along with formation of the kernels (Fig. 2b). Interestingly, the topological structure is solely a consequence of the feedback connection coming from the 2nd layer: Enforcing the pooling across neighbouring channels constrains neighbouring kernels to encode for similar features. In particular, edge-like filters with similar orientation and phase tend to be grouped in neighboring channels. This organization shows qualitatively strong similarities with the formation of macro-columns structures as found in V1, for which edges of similar orientation, frequency, and color are shown to be encoded by groups of neighbouring neurons around a pinwheel [13].
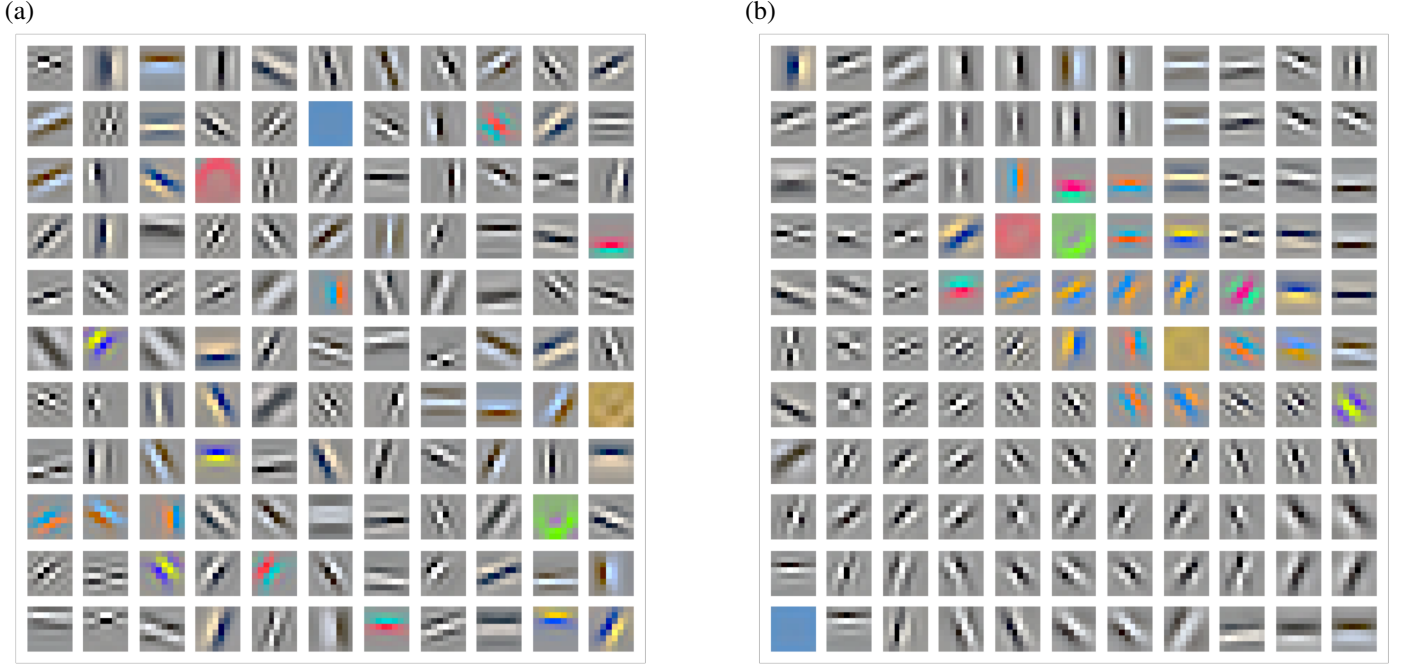
(a) (b)

Figure 2: (a) Emergence of Gabor-like kernels $\boldsymbol{D}_1$ in the 1st layer of the network (121 kernels, 8×8 pixels) learned from the STL10 databse [10] for $k_{fb} = 4$ and $f = MaxPool2D$. Their relative position in the grid is random. (b) Kernels learned when introducing group sparsity with $k_{fb} = 4$ and $f = GroupMaxPool2D$. The red square indicates one of the pooling groups used by p_{ch}.

Additionally, we observed qualitatively that when $f = GroupMaxPool2D$, the activity of the model in the second layer shows an invariance to the phase of an oriented edge as is exhibited by biological V1 complex cells. To specifically test this property, we fed both networks with a set of oriented edge-like images (Gabors) [14] with the same center but different phases ($\phi \in [0, \pi)$) and different orientations ($\theta \in [\frac{3}{8}\pi, \frac{1}{8}\pi)$), see Fig. 3-a. We then evaluated the difference in activity with respect to a reference stimulus ($\phi_{REF} = 0, \theta_{REF} = \frac{3}{8}\pi$) when $f = MaxPool2D$ and $f = GroupMaxPool2D$. The results of this experiment are shown in Fig. 3-b and c. In Fig. 3-b, a network showing a topological organization appears to be strongly phase invariant, with changes in network activity being only dependent on variations in orientation θ: Changes in phase marginally modify the complex cells activity. In this case, the network shows also a relative tolerance for variations in the stimulus orientation. On the other hand, Fig.3-c shows the network without a topological structure and which does not generalize over the range of inputs: Changes in phase induce variations in network activity. The invariant behaviour of a network trained with $f = GroupMaxPool2D$ is explained by the induced topological organization of its 1st layer. Indeed, neighboring kernels tend to encode similarly oriented edges with different phases, and pooling across these neurons leaves the resulting response unchanged. As a results, this makes the model complex cells encode for one oriented stimulus independently of its local properties. The same response will be evoked for different local contrast (phase) and small variations in stimulus orientation. This behaviour has already been observed in complex cells in V1 and likely represents a key computational step to build an invariant representation of the input [7, 8].

Conclusions. We showed that a two-layered SDPC model of V1 can predict highly non-linear cell behaviors observed in the mammals' visual cortex [7]. We suggest that invariance to the input stimuli in complex cells could emerge directly from imposing a topological organization to V1, rather than from a simple pooling across spatial locations. We predict that we should find such differences between species which lack a topological organization (mice for instance) and other which have such an organization (such as primates). Moreover, such a behaviour would be of key importance for implementing object recognition in a neural substrate. Indeed, natural images are known to be efficiently described by co-linear and co-circular sets of edges described by Gabor filters of similar orientation disposed along smooth trajectories [15]. Complex cells are likely implementing a hierarchical model [13] by exploiting such regularities together with other functional structures of the visual cortex.
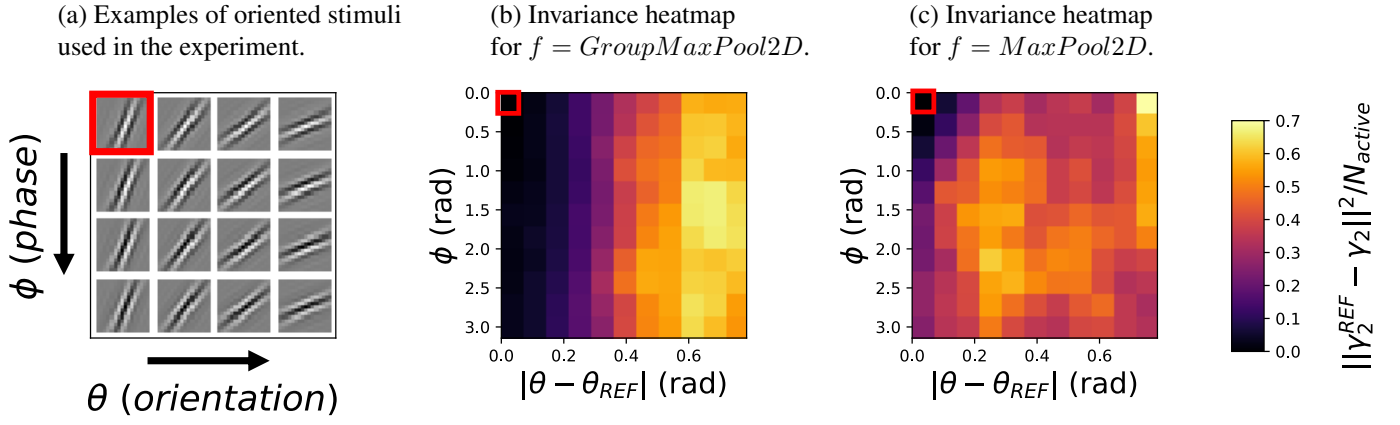
(a) Examples of oriented stimuli used in the experiment.

(b) Invariance heatmap for $f = GroupMaxPool2D$.

(c) Invariance heatmap for $f = MaxPool2D$.

Figure 3: Variation in network activity for different phases and orientations of the input stimulus with respect to a reference edge. (a) example of different stimuli generated by varying the phase ϕ, in the range $\{0, \pi\}$, and the angle θ, in the range $\{\frac{1}{8}\pi, \frac{3}{8}\pi\}$. The stimulus used as reference for this experiment is highlighted by the red square. (b-c) The distance between the reference and the considered filter in computed as the ℓ_2 norm of the difference between the model complex cells response to the two stimuli, γ_2^{REF} and γ_2 for the current stimulus, normalized by the number of active cells for the input, N_{active}. We show the change in activity respectively for a network showing a topological structure trained with $f = GroupMaxPool2D$ (b) and for a network with no topological structure trained with $f = MaxPool2D$ (c).

References

[1]   Rajesh PN Rao and Dana H Ballard. "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects". In: *Nature neuroscience* 2.1 (1999), p. 79.

[2]   Stewart Shipp. "Neural elements for predictive coding". In: *Frontiers in psychology* 7 (2016), p. 1792.

[3]   Hillel Adesnik and Alexander Naka. "Cracking the Function of Layers in the Sensory Cortex". In: *Neuron* 100.5 (2018), pp. 1028–1043.

[4]   James CR Whittington and Rafal Bogacz. "Theories of Error Back-Propagation in the Brain". In: *Trends in Cognitive Sciences* 23 (2019), pp. 235–250.

[5]   James CR Whittington and Rafal Bogacz. "An approximation of the error backpropagation algorithm in a predictive coding network with local Hebbian synaptic plasticity". In: *Neural computation* 29.5 (2017), pp. 1229–1262.

[6]   Victor Boutin et al. "Meaningful representations emerge from Sparse Deep Predictive Coding". In: *arXiv preprint arXiv:1902.07651* (2019).

[7]   Jan Antolik and James A Bednar. "Development of maps of simple and complex cells in the primary visual cortex". In: *Frontiers in computational neuroscience* 5 (2011), p. 17.

[8]   Ko Sakai and Shigeru Tanaka. "Spatial pooling in the second-order spatial structure of cortical complex cells". In: *Vision Research* 40.7 (2000), pp. 855–871.

[9]   Amir Beck and Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". In: *SIAM journal on imaging sciences* 2.1 (2009), pp. 183–202.

[10]   Adam Coates, Andrew Ng, and Honglak Lee. "An analysis of single-layer networks in unsupervised feature learning". In: *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. 2011, pp. 215–223.

[11]   Bruno A Olshausen and David J Field. "Sparse coding with an overcomplete basis set: A strategy employed by V1?" In: *Vision research* 37.23 (1997), pp. 3311–3325.

[12]   Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), p. 436.

[13]   David H Hubel. "Brain and visual perception". In: *Perception*. Vol. 34. 2005, pp. 3–3.

[14]   Sylvain Fischer et al. "Self-Invertible 2D Log-Gabor Wavelets". In: *International Journal of Computer Vision* 75.2 (Jan. 2007), pp. 231–246. ISSN: 0920-5691. DOI: 10.1007/s11263-006-0026-8.

[15]   Laurent U Perrinet and James A Bednar. "Edge co-occurrences can account for rapid categorization of natural versus animal images". In: *Scientific reports* 5 (2015), p. 11400.