On the Origins of Hierarchy in Visual Processing

Angelo Franciosini
Institut de Neurosciences de la Timone
angelo.franciosini@univ-amu.fr

Laurent U. Perrinet
Institut de Neurosciences de la Timone
laurent.perrinet@univ-amu.fr

It is widely assumed that visual processing follows a forward sequence of processing steps along a hierarchy of laminar sub-populations of the neural system. Taking the example of the early visual system of mammals, most models are consequently organized in layers from the retina to visual cortical areas, until a decision is taken using the representation that is formed in the highest layer. Typically, features of higher complexity (position, orientation, size, curvature, ...) are successively extracted in distinct layers [1]. This is prevalent in most deep learning algorithms and stems from a long history of feed-forward architectures. Though this proved to be highly successful, the origin of such architectures is not known [2]. Using a generic unsupervised learning algorithm, we first trained a simple one-layer convolutional neural network on a seta of natural images with a growing number of neurons. By doing this, we could quantitatively manipulate the complexity of the representation that emerges from such learning and analyze if sub-populations within the layer could be grouped by their similarity, hence justifying the emergence of a hierarchical processing. As shown in previous studies [3], such an algorithm converges to a weight matrix that has strong analogies with the receptive fields of simple cells located in the Primary Visual Cortex of mammals (V1).

This result extends naturally to a cortical representation of the input image that encodes second-order features (edges) as neural responses arranged in a three-dimensional space, where the third dimension can be seen as a model of hyper-columns of the Primary Visual Cortex. From this bio-inspired encoding, we were able to define contours in images as simple smooth trajectories in a cortical representation space. This simple model shows that hierarchical processing may originate from the neural encoding of different visual transformations within natural images: respectively translation, rotations and zooms, which correspond to rigid translation in the cortical space. The model can be further extended to reproduce the effect of complex cells in V1 (max pooling) and feedback signals from higher cortical areas. We predict that invariance to more complex transformation like shearing (perspective) and viewpoint changes (looming) will emerge as these additional steps in sensory processing are taken into account.

Indeed, a higher level of complexity can be introduced as the cortical representation is extended from smooth trajectories (space domain) to smooth surfaces (space-time domain). As such, this justifies the extension of a simple sparse network formalism to translation invariant neural networks (such as the convolutional neural networks used in deep learning) that is able to generalize geometrical transformations, such as translation, rotations, and zooms, in an invariant bio-inspired representation [4]. This should provide some key insights into higher-order features such as co-occurrences, but also to novel categorization architectures. Indeed, such features were recently found to be sufficient to allow the categorization of images containing an animal [5]. Crucially, as the geometrical transformations develop in time, we expect that the detection of these features is made robust by dynamical processes.

References

[1] Matteo Carandini and David J Dj Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(November):1–12, jan 2012.

[2] Thomas Serre, Aude Oliva, and Tomaso Poggio. A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, 104(15):6424–6429, 2007.

[3] Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607, 1996.

[4] Laurent U. Perrinet. Sparse models for computer vision. In Gabriel Cristóbal, Laurent Perrinet, and Matthias S. Keil, editors, *Biologically Inspired Computer Vision*, chapter 13. Wiley-VCH Verlag, nov 2015.

[5] Laurent U. Perrinet and James A. Bednar. Edge co-occurrences can account for rapid categorization of natural versus animal images. *Scientific Reports*, 5:11400, jun 2015.