Function determines structure in complex neural networks

Tatyana O. Sharpee[1]

Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037

In the course of our daily lives, we rely on our ability to quickly recognize visual objects despite variations in the viewing angle or lighting conditions (1). Most of the time, this process occurs so naturally that it is easy to underestimate the computational challenge it poses. In fact, as much as one third of our cortex is involved in visual object recognition (2). Thus far, we have not been able to create artificial object recognition systems that can match the performance of the human visual system, especially when one takes into account the energetic constraints. However, understanding the principles of visual object recognition has major implications not only because of potential wide-ranging practical applications but also because these principles are likely to hold clues to how sensory systems work in general. After all, there is a growing body of experimental and computational evidence that similar principles (3) might be at work during visual, auditory, or olfactory object recognition tasks. In PNAS, Yamins et al. present an important advance in this research direction by demonstrating how to optimize artificial networks to match the discrimination ability of primates when it comes to categorizing objects (4). Importantly, optimized networks share many features with the corresponding networks in the brain. Thus, although the structure of neural networks in the brain could have been largely determined by the idiosyncrasies of the previous evolutionary trajectory, they seem to reflect instead a unique optimal solution. This in turn offers more support for theoretical and computational pursuits to find optimal circuit organization in the brain (5).

Primate vision is supported by a number of interconnected areas (6) that are thought to form a hierarchy of processing stages. The early stages, such as the primary visual cortex (V1), contain neurons that are selective for relatively simple image features such as edges (7) and respond only when these features appear at specific positions in the visual field. Neurons at later stages can be selective for complex combinations of visual components.

For example a contour's curvature is an important characteristic for neurons in the intermediate area V4 (8). In the inferotemporal (IT) cortex, neurons are often selective for faces and their components (9), as well as other objects of large biological significance (10). The neurons from later stages of visual processing are also often more tolerant to changes in the position of their preferred visual features (11). The best computer vision models have borrowed heavily from the known properties of the primate visual system. However, beyond edge detection and convolution (12), it has been difficult to determine unequivocally what transformations of images the subsequent areas perform.

Theoretical and computational approaches provide a way around this roadblock (4, 13–15). The idea is to study artificial networks that have been optimized to perform object recognition tasks (Fig. 1). By following this research direction, Yamins et al. achieve an important milestone in our understanding of the human visual system (4). The authors created artificial networks capable of reaching human performance levels in a variety of object recognition tasks. The optimized networks provide substantial improvements in object recognition tasks compared with previous studies (13). Two features of these networks were crucial for their success. First, the networks were based on the so-called "convolutional" network modules (12), where the number of free parameters that specify connections within a module is reduced by taking advantage of symmetries, such as translation invariance, inherent to vision. Second, modules within the larger networks were organized hierarchically and optimized independently, taking advantage of recent breakthroughs in machine learning for training complex networks (16).

The often stated argument against using optimization principles to understand biological function is that biological solutions are likely to reflect more the previous development during evolution than the engineering constraints imposed by energy

train the network to characterize natural scenes

IT-like
IT
adjust to match recorded neurons
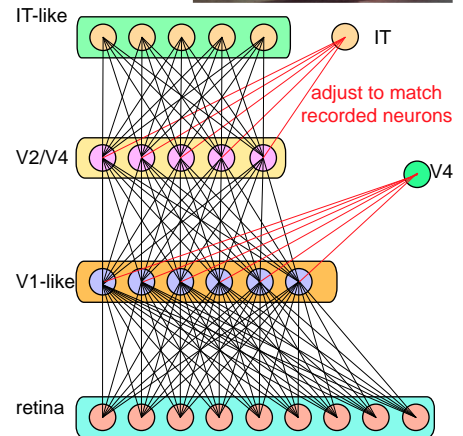V2/V4
V4
V1-like
retina

Fig. 1. Sensory processing in the brain relies on a series of nonlinear transformations coupled with pooling of signals from previous stages (13, 15). The connections (black lines) within such networks were optimized to characterize natural scenes and objects present in them (4). The optimized networks matched human performance over a wide range of object recognition tasks (4). These networks also proved useful in accounting for neural responses at different stages of processing in the primate visual system. A weighted pooling of signals (red lines represent connections optimized for a particular recording) from second or third tier of optimized networks accounted for the largest fraction of response variance among neurons in the intermediate area V4 within the ventral visual stream where object recognition is thought to take place (4). A weighted pooling of signals from the top layer of the model accounted for a large percentage of variance in the responses of neurons in the inferotemporal (IT) cortex, the final stage of ventral visual stream in primates.

and requirements for conveying signals from the environment. Further, even if biological systems are primarily shaped by their constraints, there are still likely to be many equally good solutions to a given problem. In this case, finding any given near-optimal solution within optimized networks using theoretical and computational investigations may yield little insight into the organization of networks in the brain.

Remarkably, Yamins et al. demonstrate that neither of these concerns applies to neural circuits in the brain (4). First, they found that the optimized networks exhibited the same general principles of organization that are known to exist in the visual systems. Specifically, units from earlier processing stages were more specific for a given position in the visual field and selective for simpler contour elements than units from subsequent stages of processing. Second, the optimized networks helped to account remarkably well for the neural responses recorded from area V4 and the IT cortex. A simple pooling of signals from the top level of the optimized network (Fig. 1) accounted, on average, for 50% of variance in the responses of IT neurons to various natural stimuli. This level of achievement is unprecedented and is comparable to what could only previously be achieved in early (17, 18) and intermediate (19, 20) stages of visual processing. Notably, adding more layers to the optimized network and pooling from its top layer did not always increase the amount of explained variance in the recorded neural responses from the intermediate visual areas. For example, when accounting for responses of neurons in area V4, the best predictive power was achieved when pooling signals from the second and third layers, with substantially smaller performance when pooling signals from the highest level in the optimized network. Overall, these results demonstrate that, although networks in the brain are very complex, their organization can be understood by studying their top-level function.

Despite this important progress, many challenges remain. As the authors point

Yamins et al. achieve an important milestone in our understanding of the human visual system.

out, the principles for how images are transformed from stage to stage remain unresolved (4). The optimized networks analyzed thus far include up to four stages of nonlinear processing (4). The primate visual system involves 10 or more stages. It might be nontrivial to make current optimization techniques work for larger and deeper networks that begin to approximate the number of neurons in the brain. The ironic challenge is that as we build increasingly more accurate models of the brain they become just as difficult to analyze and interpret as the brain itself. However, as the present study demonstrates, important advances can be made by bringing together findings and ideas from machine learning and neuroscience. Let us hope that further welding between these two disciplines will help us understand the principles of object recognition soon.

1 Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381(6582):520–522.
2 Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1(1): 1–47.
3 King AJ, Nelken I (2009) Unraveling the principles of auditory cortical processing: Can we learn from the visual system? *Nat Neurosci* 12(6):698–701.
4 Yamins DLK, et al. (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci USA* 111: 8619–8624.
5 Bialek W (2013) *Biophysics: Searching for Principles* (Princeton Univ Press, Princeton).
6 Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19:109–139.
7 Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195(1):215–243.
8 Connor CE, Brincat SL, Pasupathy A (2007) Transformation of shape information in the ventral pathway. *Curr Opin Neurobiol* 17(2): 140–147.
9 Tsao DY, Livingstone MS (2008) Mechanisms of face perception. *Annu Rev Neurosci* 31:411–437.
10 Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4(8):2051–2062.

11 Gawne TJ, Martin JM (2002) Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J Neurophysiol* 88(3):1128–1135.
12 LeCun Y, et al. (1989) Backpropagation applied to handwritten zip code recognition. *Neural Comput* 1(4):541–551.
13 Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci USA* 104(15): 6424–6429.
14 Cadieu CF, Olshausen BA (2012) Learning intermediate-level representations of form and motion from natural movies. *Neural Comput* 24(4):827–866.
15 Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2(11):1019–1025.
16 Bengio Y (2009) Learning deep architectures for AI. *Foundations Trends Machine Learning* 2(1):1–127.
17 Rowekamp RJ, Sharpee TO (2011) Analyzing multicomponent receptive fields from neural responses to natural stimuli. *Network* 22(1-4):45–73.
18 David SV, Vinje WE, Gallant JL (2004) Natural stimulus statistics alter the receptive field structure of v1 neurons. *J Neurosci* 24(31): 6991–7006.
19 David SV, Hayden BY, Gallant JL (2006) Spectral receptive field properties explain shape selectivity in area V4. *J Neurophysiol* 96(6): 3492–3505.
20 Sharpee TO, Kouh M, Reynolds JH (2013) Trade-off between curvature tuning and position invariance in visual area V4. *Proc Natl Acad Sci USA* 110(28):11618–11623.