

Deep learning and computational neuroscience

Erik De Schutter

Computational Neuroscience Unit, Okinawa Institute of Science and Technology Graduate University, Japan

Especially young colleagues are fascinated by the potential of deep learning for neuroscience. This was obvious at the recent Society for Neuroscience meeting in Washington DC, where the few posters that had the magical words in their title attracted large crowds of attendees who seemed almost exclusively in their twenties. The success of deep learning of data representation has led to impressive applications in image, video and speech processing ¹. Compared to these, recent advances in applying reinforcement learning to playing games are outright mind blowing, with AlphaGo Zero achieving superhuman performance in just three days of training on a single machine with specialized hardware ². It is, therefore, easy to predict that the interest in deep learning among young computational neuroscientists will only increase, but the reality may be more complex than they surmise. In this Editorial, I will focus on the question of correspondence between deep learning and how the brain works³. I will not consider the many opportunities of applying deep learning as a supporting technology.

The original breakthrough leading to the success of deep learning tested the method on an image recognition task, classifying handwritten digits ⁴. Correspondingly, most of the applications of deep learning to computational neuroscience are about understanding the visual system (including the posters at the recent Society for Neuroscience meeting). As pointed out in a recent

¹ Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, “Deep Learning” *Nature* 521 (2015): 436–44, doi:10.1007/s10994-013-5335-x.

² David Silver et al., “Mastering the Game of Go Without Human Knowledge” *Nature* 550 (2017): 354–59, doi:10.1038/nature24270.

³ A more extensive analysis emphasizing the brain to deep learning connection can be found in Demis Hassabis et al., “Neuroscience-Inspired Artificial Intelligence” *Neuron* 95 (2017): 245–58, doi:10.1016/j.neuron.2017.06.011.

⁴ Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh, “A Fast Learning Algorithm for Deep Belief Nets.” *Neural Computation* 18 (2006): 1527–54, doi:10.1162/neco.2006.18.7.1527.

review ⁵ one category of deep learning models, goal-driven hierarchical convolutional neural networks, has been very successful at predicting neural responses in several layers of primate visual cortex, including V1, V2, V4 and inferior temporal cortex (IT). But the authors also point out that this success is probably due to convolutional neural networks closely mimicking the overall architecture of cortex ⁵, in particular implementing features similar to receptive fields of increasing size across the hierarchy. This leads to a warning that any correspondences between deep learning methods and the brain may not generalize to all deep learning. In fact, though the field of machine learning has clearly been inspired by neuroscience ³, it has never seen this as a limitation on the methods it can use. For example, the breakthrough referred to earlier ⁴ was a method to teach layers in a multilayer network one at a time, something that is hard to imagine occurring in a real brain. Deep learning networks have typically also many more layers than corresponding brain systems and one of the current hypes are “very deep” models with tens of layers ⁶. A recent breakthrough, also used in AlphaGo Zero, are residual networks where shortcut connections are used that connect units in lower layers directly with units in higher layers ⁶. Residual networks are an example of deep learning methods that do not reflect real neural systems, this would be like V1 densely projecting directly to V4 or IT. Conversely, there are well known brain circuits that have clearly quite different architectures than visual cortex, like for example the olfactory system.

Another difference between deep learning and human brains is the number of training examples required, with millions of labeled images needed to learn simple categorization tasks ⁵. In fact, deep learning would not exist if the digital revolution hadn't made big data available. Fortunately, the human brain is better at generalizing from smaller sets of experiences, but recent machine learning approaches try to mimic this ³. Conversely, we may not learn to recognize some features because we do not routinely train ourselves on labeled data. An example is the recent controversial study claiming that a deep network learned to recognize sexual preferences of people by analyzing pictures on a dating website ⁷. In newspapers, this was reported as a demonstration of how artificial intelligence can now beat the human mind. In real life, however, people often don't know who is gay or not (we lack the label) and I doubt any rational person would consider training themselves by going through all the ads on a dating site.

⁵ Daniel L K Yamins and James J DiCarlo, “Using Goal-Driven Deep Learning Models to Understand Sensory Cortex” *Nature Neuroscience* 19 (2016): 356–65, doi:10.1017/S0140525X0001863X.

⁶ Kaiming He et al., “Deep Residual Learning for Image Recognition” *Conference on Computer Vision and Pattern Recognition* 2016.

⁷ www.economist.com/news/science-and-technology/21728614-machines-read-faces-are-coming-advances-ai-are-used-spot-signs

Returning to the AlphaGo Zero example ², the success of deep learning may soon be surpassed by reinforcement learning, which is again directly based on neuroscience concepts ^{3,8}. In this case the big data challenge was overcome by having AlphaGo play games against itself, so no prior data was required. This may be conceptually similar to many forms of learning in infancy, where trial and error clearly play a very big role. But while being successful at playing Go and chess seems a big achievement because of the close to infinite number of possible game states, the important limiting metric for reinforcement learning is the number of possible actions and that number is, clearly, limited for any board game. Nevertheless, it is probably worthwhile to carefully investigate lessons that can be drawn from AlphaGo Zero.

Finally, I want to report on an interesting recent report ⁹ that shows a fairly realistic way to solve one of the most vexing problems in mapping machine learning to the brain: the credit assignment problem. The first machine learning revolution in the 80ies was based on the discovery of the back-propagation algorithm ¹⁰ and it is well known that real brains have no back-propagation ³, i.e. transmission of weight updates from higher layers to lower layers. In more modern terms, this is called the credit assignment problem: which neurons in lower layers directly contributed to the final behavioral outcome? Interestingly, leading researchers in deep learning are quite concerned about finding solutions to this conundrum ¹¹, but till now only partial answers were proposed. Guerguiev et al. ⁹ propose to use the dendritic structure of neurons and this is a remarkably complete scheme, although there are still a few unresolved issues. Specifically, sparse feedback connections carrying higher-order feedback to the apical dendrites are used to drive changes in synaptic weights in basal dendrites that receive sensory input. The dendrites are essential because they provide for physical separation of the two inputs onto the same neuron. I will not describe the results in detail, but encourage reading of the paper.

⁸ Michael L Littman, "Reinforcement Learning Improves Behaviour From Evaluative Feedback" *Nature* 521 (2015): 445–51, doi:10.1177/0278364913495721.

⁹ Jordan Guerguiev, Timothy P Lillicrap, and Blake A Richards, "Towards Deep Learning with Segregated Dendrites.," *eLife* 6 (2017), doi:10.7554/eLife.22901.

¹⁰ D E Rumelhart, Geoffrey E Hinton, and R J Williams, "Learning Representations by Back-Propagating Errors" in *Cognitive Modeling*, ed. T A Polk and C M Seifert, (MIT Press, 1986).

¹¹ Timothy P Lillicrap et al., "Random Synaptic Feedback Weights Support Error Backpropagation for Deep Learning" *Nature Communications* 7 (2016): 13276, doi:10.1038/ncomms13276; Bengio Y, Lee D-H, Bornschein J, and Lin Z. (2015): Towards biologically plausible deep learning. arXiv. arXiv:1502.04156.

From my perspective, this is quite an ironic result. As an avid modeler of dendrites¹², I have been frustrated by the swing of the field of computational neuroscience towards point neuron network modeling. I have seen a gradual loss of interest among students of summer schools in computational neuroscience in simulating single neuron models and the shift is also noticeable in recent textbooks¹³. If what Guerguiev and colleagues (including a scientist working at DeepMind) propose⁹ is true, then understanding how neural networks learn complex tasks will require models that include both apical and basal dendrites, though the models themselves do not have to be very complex. This nicely matches the recent demonstration in in vivo experiments of the behavioral importance of dendritic spikes¹⁴ and suggests a bright future for modeling dendrites.

¹² Volker Steuber et al., “Cerebellar LTD and Pattern Recognition by Purkinje Cells” *Neuron* 54 (2007): 121–36, doi:10.1016/j.neuron.2007.03.015; H Anwar et al., “Stochastic Calcium Mechanisms Cause Dendritic Calcium Spike Variability” *The Journal of Neuroscience* 33 (2013): 15848–67, doi:10.1523/JNEUROSCI.1722–13.2013.

¹³ Wulfram Gerstner et al., *Neuronal Dynamics* (Cambridge University Press, 2009), doi:10.1017/CBO9781107447615.

¹⁴ Naoya Takahashi et al., “Active Cortical Dendrites Modulate Perception” *Science* 354 (2016): 1587–90, doi:10.1126/science.aah6066.