Workshop on Deep Learning and the Brain

Andrew Michael Saxe (asaxe@stanford.edu)
Center for Mind, Brain, and Computation,
Department of Electrical Engineering,
Stanford University
316 Jordan Hall, Stanford, CA 94305 USA

Keywords: Deep learning; Neural networks; Sensory processing

Introduction

Deep learning methods rely on many layers of processing to perform sensory processing tasks like visual object recognition, speech recognition, and natural language processing (Bengio & LeCun, 2007). By learning simpler features in lower layers, and composing these into more complex features in higher layers, deep learning systems take advantage of the compositional nature of many real world tasks. To recognize cars, for instance, a deep system might first build wheel detectors, window detectors, etc, in lower layers, before combining these into a car detector at a higher layer. Deep learning has emerged as a central tool in the engineering disciplines due to its impressive performance in a range of applications, from visual object classification (Krizhevsky, Sutskever, & Hinton, 2012; Ciresan, Meier, & Schmidhuber, 2012) to speech recognition (Mohamed, Dahl, & Hinton, 2012) and natural language processing (Collobert & Weston, 2008).

Parts of the brain (and in particular the visual system) appear to share some of these features. Anatomically, they consist of a series of processing layers than can be arranged hierarchically (Felleman & Van Essen, 1991). And functionally, neural responses show a progression of complexity from lower to higher levels (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005), and these representations change with experience.

In light of these similarities, this workshop will explore the implications of deep learning for our understanding of the brain and mind. To what degree can the brain be considered “deep”? How central is depth to its function? What insights from machine learning can inform work in cognitive science, and visa versa? How does depth impact both the dynamics of learning in a neural network, and the content of what is learned? How might deep learning models illuminate phenomena of interest to cognitive scientists such as perceptual learning, language acquisition, cognitive development, and category formation?

The participants in this workshop have been chosen to present a broad range of perspectives on deep learning in the cognitive sciences. They span computational and empirical approaches, and allow for critical contact with other theoretical perspectives. The recent rapid progress on deep learning within the machine learning community, and the growing number of deep learning-based models in the cognitive sciences, makes this workshop both timely and important for the cognitive science community.

Goals and scope

The goal of the workshop is to explore the relevance of recent deep learning advances to cognition, to bring together cognitive science-oriented deep learning researchers, and to facilitate exchanges between the machine learning and cognitive science communities.

While deep learning has been a persistent thread of research in the cognitive sciences from the very beginning, a goal of the workshop is to provide a focal point for this community and a forum for important discussions and collaborations that can span methodological approaches. Because of the domain general nature of deep learning methods, these approaches can serve to unite a diverse set of researchers focusing on a variety of phenomena.

In addition, the workshop will demonstrate the ability of deep learning models to address phenomena at a variety of different scales and levels of detail, with talks covering material from receptive field models in retina and early visual cortices, through mid-level vision and object recognition, to semantic cognition.

Workshop organization

The main feature of the workshop will be a series of invited talks meant to span a broad range of perspectives on deep learning and the brain, and concentrated mostly on visual processing. Visual object recognition is the area most studied in prior deep learning work both in machine learning and cognitive science, and hence makes a natural first focus for a workshop. Although the talks will address recent research, by their diverse perspectives they will also constitute a good introduction to the field for those who have not engaged with deep learning before. The workshop is planned as a full day workshop, and each speaker will have approximately 30 minutes, to leave time for questions and discussion following the talks. Depending on time considerations, the workshop will close with a panel discussion to allow the audience further interaction with the speakers, and to permit speakers from different backgrounds to engage each other on themes that have emerged during the day.

The workshop will also accept submissions of abstracts for posters to be presented during lunch and coffee breaks. Accepted poster submissions will be made available from the workshop website. The aim of the poster sessions is to showcase the much broader range of issues relevant to cognitive science that are being tackled using deep learning methods which, due to time limitations, cannot be included in the main program.

Workshop organizer

Andrew Saxe is a PhD student with Jay McClelland and Andrew Ng at Stanford University. He has contributed to deep learning research on machine learning (A. M. Saxe, McClelland, & Ganguli, 2014), sensory neuroscience (A. Saxe, Bhand, Mudur, Suresh, & Ng, 2011), and semantic cognition (A. Saxe, McClelland, & Ganguli, 2013).

Audience

This workshop aims to bring together a diverse set of researchers that span the disciplinary spectrum of cognitive science, from artificial intelligence, machine learning, and neural networks, to cognitive psychology and neuroscience. Because of their domain-general nature, deep learning methods have been applied or may be applicable to modeling natural language processing, auditory perception, visual number sense, semantic cognition, and beyond. In light of this, we expect the workshop to be of broad interest to the cognitive science community and estimate around 75 participants.

Presenters

The following speakers have confirmed their ability to participate.

Merav Ahissar
Edmond and Lily Safra Center for Brain Sciences
Department of Psychology
Hebrew University of Jerusalem

Yoshua Bengio
Department of Computer Science and Operations Research
University of Montreal

Gary Cottrell
Department of Computer Science and Engineering
University of California, San Diego

David Cox
Departments of Molecular and Cellular Biology and Computer Science
Harvard University

Chris Eliasmith
Canada Research Chair in Theoretical Neuroscience
Departments of Philosophy and Systems Design Engineering
University of Waterloo

Surya Ganguli
Department of Applied Physics
Stanford University

Jay McClelland
Department of Psychology
Stanford University

Tomaso Poggio
Department of Brain and Cognitive Sciences
McGovern Institute for Brain Research
Computer Science & Artificial Intelligence Lab
Massachusetts Institute of Technology

Josh Tenenbaum
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

References

Bengio, Y., & LeCun, Y. (2007). Scaling learning algorithms towards AI. In L. Bottou, O. Chapelle, D. DeCoste, & J. Weston (Eds.), *Large-scale kernel machines* (pp. 1–41). MIT Press.

Ciresan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column Deep Neural Networks for Image Classification. In *Ieee conf. on computer vision and pattern recognition* (pp. 3642–3649).

Collobert, R., & Weston, J. (2008). A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In *Proceedings of the 25th international conference on machine learning.*

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, N.Y. : 1991)*, *1*(1), 1–47.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in neural information processing systems 25.*

Mohamed, A., Dahl, G., & Hinton, G. (2012, January). Acoustic Modeling Using Deep Belief Networks. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(1), 14–22.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(June), 1102–1107.

Saxe, A., Bhand, M., Mudur, R., Suresh, B., & Ng, A. Y. (2011). Unsupervised learning models of primary cortical receptive fields and receptive field plasticity. In *Advances in neural information processing systems 25.*

Saxe, A., McClelland, J. L., & Ganguli, S. (2013). Learning hierarchical category structure in deep neural networks. In M. Knauff, M. Paulen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the cognitive science society* (pp. 1271–1276). Austin, TX: Cognitive Science Society.

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2014). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. In Y. Bengio & Y. LeCun (Eds.), *International conference on learning representations.* Banff, Canada.