Understanding sensory systems by training models in virtual worlds

Chengxu Zhuang[1*], Damian Mrowca[2], Daniel Yamins[1]

[1]Department of Psychology, University of Stanford
[2]Computer Science Department, University of Stanford
Jordan Hall Rm 427, Stanford, CA 94305, USA

*To whom correspondence should be addressed; E-mail: chengxuz@stanford.edu.

In this paper, we trained models in virtual worlds for specific goals and then showed that those models could predict the brain areas realted to the same goals well. We frst built two virtual worlds for models to learn two behaviors, recognizing objects by detecting their shapes using simulated-whiskers and recovering 3D shapes and therefore categories from 2D images collected in simulation world. After training those models well, we then showed that those models could predict the related neural responses for those behaviors as well. Both these models shared the similarity that they were trained to give normals of the object surfaces as intermediate outputs. And by explicitly constraining the models to do so, we were expecting that this would actually help the models for better explaining the responses of both intermediate layers and final layers of brain areas in interest. We also hope that this work would inspire future researches to utilize the considerable literature about functions of intermediate layers to improve the performances of models both on finishing

tasks and modelling brains.

Introduction

Brains have done remarkable work by actively analyzing environment information and making decisions upon that. Among all systems in brains, sensory systems are usually concentrated on analyzing some particular environment information, for example, light for visual systems, sound for auditory systems, and inputs from whiskers for mouse somatosensory systems. The goal of those sensory systems is to extract the useful semantic information from the complex raw input data, which could be described as untangling the behavior-related dimensions (such as category) from other irrelevant dimensions (such as translation and rotation of the objects)[22]. And in this work, we are especially interested in visual systems and mouse somatosensory systems.

A lot of work has been done to explore both two systems. While those systems differ from their input data, number of overall neurons, and specific structures as well as organizations, they are believed to have the similarity of consisting of several consecutive regions that are distinguishable on both structures and functions [6]. For visual systems, starting from work of Hubel and Wiesel[9], there have been a large number of hierarchical models developed to explain the response patterns of them[17, 19, 7, 2, 15]. Similarly in somatosensory systems, researchers also find evidence showing hierarchical processing for somatosensory input in both human and primates[16, 10, 11].

Hierarchical models are also used widely in artificial intelligence to help design better systems for various tasks. The recent work using deep neural networks (DNNs) has achieved significant improvements on object recognition, speech recognition, and numerous of other artificial intelligence tasks[12, 8, 13]. Those deep neural networks are all composed of multiple simple neural network layers in series, where the computation in single layer is usually simple but non-linear and stacks of those simple non-linear computations finally make up of some highly complicated non-linear computations. Additionally, those models are also believed to be biologically plausible and therefore could be good candidates for models of related brain systems.

Furthermore, researchers have also found that the hierarchical models optimized for performances on object recognition tasks could also serve as a good model for IT areas in primates, which are believed to be the responsible areas for object recognition in brains [20, 21, 4]. Inspired by this, we are building performance optimized hierarchical models for specific tasks that we think V4 (an intermediate layer in visual systems) in human and primates and mouse somatosensory systems are performing. And after having the models, we could use them to explain the responses of those brain areas.

Both two areas of interest are poor understood. V4 is believed to encode intermediate level of object features and show strong attentional modulation[18], while IT areas are believed to encode high level of object features as object category. And mouse could use their whiskers to detect object shape, position, and texture of object surface[3, 5, 1, 14]. Thus the task that we are interested in for V4 neurons is predicting normals of object surfaces using 2D images as input. Once we have a good model for V4, we could further add some extra layers on top of it to predict the object category from normals. Those extra layers then could be treated as models for IT areas. For mouse somatosensory cortex, we are using the similar task but with input now collected through simulated-whiskers to simulate the input to mouse barrel cortex. Via doing this, we are trying to model S1 (primary somatosensory cortex) as normal predictors and then S2 (secondary somatosensory cortex) as object category detector. With explicitly modeling the functions of intermediate layers, we hope that we could have a better model for the whole systems.

However, for optimizing those deep hierarchical models, one would need a large nubmer of examples with corresponding labels. And in our work, it is either too difficult to collect the corresponding labels (for example, the normals of object surfaces given the 2D images) or the desired example itself (for example, the input from simulated whiskers). Thus we need to create virtual worlds for our tasks and train our models there.

Methods

Results

References

[1] Ehsan Arabzadeh, Erik Zorzin, and Mathew E. Diamond. Neuronal encoding of texture in the whisker sensory pathway. *PLoS Biology*, 3(1), 2005.

[2] Yoshua Bengio. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[3] Yves Boubenec, Daniel E Shulz, and Georges Debrégeas. Whisker encoding of mechanical events during active tactile exploration. *Front Behav Neurosci*, 6(November):74, 2012.

[4] Charles F. Cadieu, Ha Hong, Daniel L K Yamins, Nicolas Pinto, Diego Ardila, Ethan A. Solomon, Najib J. Majaj, and James J. DiCarlo. Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLoS Computational Biology*, 10(12):1–35, 2014.

[5] Mathew E Diamond, Moritz von Heimendahl, Per Magne Knutsen, David Kleinfeld, and Ehud Ahissar. 'Where' and 'what' in the whisker sensorimotor system. *Nat Rev Neurosci*, 9(8):601–612, 2008.

[6] Daniel J Felleman and David C Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex*, 1(1):1–47, 1991.

[7] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

[8] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[9] D H Hubel and T N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148:574–591, 1959.

[10] Koji Inui, Xiaohong Wang, Yohei Tamura, Yoshiki Kaneoke, and Ryusuke Kakigi. Serial processing in the human somatosensory system. *Cerebral Cortex*, 14(8):851–857, 2004.

[11] Yoshiaki Iwamura. Hierarchical somatosensory processing. *Current Opinion in Neurobiology*, 8(4):522–528, 1998.

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1–9, 2012.

[13] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[14] Daniel H. O'Connor, Simon P. Peron, Daniel Huber, and Karel Svoboda. Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron*, 67(6):1048–1061, 2010.

[15] Nicolas Pinto, David Doukhan, James J DiCarlo, and David D Cox. A high-throughput screening approach to discovering good forms of biologically inspired visual representation. *PLoS Comput Biol*, 5(11):e1000579, 2009.

[16] T P Pons, P E Garraghty, David P Friedman, and Mortimer Mishkin. Physiological evidence for serial processing in somatosensory cortex. *Science (New York, N.Y.)*, 237(4813):417–420, 1987.

[17] Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–1025, 1999.

[18] Anna W. Roe, Leonardo Chelazzi, Charles E. Connor, Bevil R. Conway, Ichiro Fujita, Jack L. Gallant, Haidong Lu, and Wim Vanduffel. Toward a Unified Theory of Visual Area V4. *Neuron*, 74(1):12–29, 2012.

[19] 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.

[20] D L Yamins, H Hong, and C Cadieu. Hierarchical Modular Optimization of Convolutional Networks Achieves Representations Similar to Macaque IT and Human Ventral Stream. *Advances in neural information processing systems*, pages 1–9, 2013.

[21] Daniel L K Yamins, Ha Hong, Charles F Cadieu, Ethan a Solomon, Darren Seibert, and James J DiCarlo. Performance-optimized hierarchical models predict neural responses

in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23):8619–24, jun 2014.

[22] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3):356–365, 2016.