Fantastic DNimals and where to find them

Scholte, H.S.

DOI
10.1016/j.neuroimage.2017.12.077

Publication date
2018

Document Version
Final published version

Published in
NeuroImage

License
CC BY

Link to publication

Citation for published version (APA):
Scholte, H. S. (2018). Fantastic DNimals and where to find them. *NeuroImage*, *180*(A), 112-113. https://doi.org/10.1016/j.neuroimage.2017.12.077

NeuroImage 180 (2018) 112–113

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Fantastic DNimals and where to find them

H. Steven Scholte

Department of Brain & Cognition, University of Amsterdam, The Netherlands

The introduction of deep neural networks (DNNs, Krizhevsky et al., 2012; Lecun et al., 1998) has altered the fields of computer vision and machine learning and these networks are starting to have an impact on the field of biological vision. Kendrick Kay's paper in this issue is therefore timely in addressing two important questions in this area, namely: i) how DNNs can be used to study vision, ii) how they relate to other modeling approaches.

A central question Kay (2017) asks is how DNNs relate to the tri-level framework of Marr (1982). Within this framework a differentiation is made between the level of computational theory (what is the goal and logic of the process being carried out?), the representation and algorithm level (what is the representation of the input and output and how is this processed?), and the implementation level (how is this process implemented in hardware?). Kay asks to what degree it can be reasonably argued that DNNs are a good functional or mechanistic model of biological visual system. He finds them, beyond the most trivial inferences, lacking in this regard (Kay, 2017, section 8).

I fully agree that it is problematic to relate DNNs to Marr's modeling levels, but at the same time I do not see this as a fundamental problem regarding the use of DNNs for understanding vision. Instead, I propose we consider DNNs in a similar way to how we consider animal models. An animal model for cognition typically emerges when an animal shows behavior that can be studied in a systematic fashion; such a model increases in importance when the underlying anatomy and/or physiology of the animal can be linked to the behavior of interest. Animal models are relevant for understanding the human brain because they provide the opportunity to systematically measure and disrupt neural activity which is not possible with human subjects. Insights from such models can subsequently be applied in the development of different types of Marrian models.

Implementations of neural network models are currently being studied because they have an impressive performance in identifying objects and show a wide range of 'behavior' similar to that of humans (Kheradpisheh et al., 2016; Kubilius et al., 2016). Methods for analysing these networks are being developed at a rapid rate (Kietzmann et al., 2017; Zintgraf et al., 2016) and are yielding insight into how these networks operate and how their behavior is related to architectural elements and computations. **These results are used as measurements from a model organism, and do not require, nor suggest, that DNNs are implementation models of the animal visual system.**

Treating the rapidly expanding wide world of DNNs as a zoo of animal models opens the doors to the study of fantastic beasts like the model by Yang et al. (2015) that can recognize and generate action commands necessary for cooking using two parallel DNNs, one for classifying the hand grasp and the other for object recognition which are integrated by a third command unit. Studying these types of models could yield insight in when it is sensible to separate abstract rule generation from processing visual information and how this can be done. Another creature of interest might be the Deep Driving model in which a car is controlled using a direct perception based approach to estimate the affordances of driving (Chen et al., 2015) allowing the comparison of models based on affordance with models based on mid-level visual representations.

Beyond studying these beasts, and having the possibility to inspect all units, we can also alter the methods and material with which DNNs are trained and even alter the architecture of these animals (all without any animal welfare-concerns). For instance, Cichy et al. (Cichy et al., 2017) compared the impact of training the same DNN architecture with an object- or scene-based image database. We have recently shown that training DNNs on unrelated tasks creates segregated representations while training on related tasks does not (Scholte et al., 2017). This shows how architectural structure develops within DNNs for solving multiple tasks. This approach opens the way to training DNNs with multiple pathways solving multiple tasks, and relating these to human neuroimaging data.

The suggestion that DNNs can be considered as similar to animal models is limited by the fact that a DNN is not a biological organism with an evolutionary history that is shared with primates.

However, this problem is mitigated by the fact that both DNNs and primates 'grow up' (are "trained") in a world with natural stimuli.

Furthermore, a shared evolutionary history can induce similarities because of vestigiality or because the same mechanisms are used to solve the same problems (Gould, 1980). Given that these similarities in tuning develop because of an overlap in 'growing up/training', in combination with a similar basic architecture, (Yamins and DiCarlo, 2016) this further strengthens the case that DNNs can be used to study the fundamentals of perception.

By further training and developing these DNN animals - let's call them DNimals - we have the possibility to make much stronger inferences about what type of training and computational mechanisms explain

E-mail address: h.s.scholte@uva.nl.

https://doi.org/10.1016/j.neuroimage.2017.12.077
Received 28 September 2017; Received in revised form 12 December 2017; Accepted 22 December 2017
Available online 27 December 2017

behavior and patterns of activity in neuroscience data. Insights from studying DNimals can subsequently be used to make formal models at the algorithm and implementation level.

References

Chen, C., Seff, A., Kornhauser, A., Xiao, J., 2015. Deepdriving: learning affordance for direct perception in autonomous driving. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2722–2730.

Cichy, R.M., Khosla, A., Pantazis, D., Oliva, A., 2017. Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. NeuroImage 153, 346–358.

Gould, Stephen Jay, 1980. Senseless signs of history. In: The Panda's Thumb: More Reflections in Natural History. W.W. Norton & Company, New York, pp. 27–34.

Kay, K., 2017. Principles for models of neural information processing. bioRxiv 129114.

Kheradpisheh, S.R., Ghodrati, M., Ganjtabesh, M., Masquelier, T., 2016, April 21. Humans and Deep Networks Largely Agree on Which Kinds of Variation Make Object Recognition Harder. arXiv [cs.CV]. Retrieved from. http://arxiv.org/abs/1604.06486.

Kietzmann, T.C., McClure, P., Kriegeskorte, N., 2017. Deep Neural Networks in Computational Neuroscience. bioRxiv. Retrieved from. http://biorxiv.org/content/early/2017/05/04/133504.abstract.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems, vol. 25. Curran Associates, Inc, pp. 1097–1105.

Kubilius, J., Bracci, S., Op de Beeck, H.P., 2016. Deep neural networks as a computational model for human shape sensitivity. PLoS Comput. Biol. 12 (4), e1004896.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.

Marr, D., 1982. Vision: a Computational Investigation into the Human Representation and Processing of Visual Information, vol. 2. Inc., New York, NY, 4–2.

Scholte, H.S., Losch, M.M., Ramakrishnan, K., de Haan, E.H.F., Bohte, S.M., 2017, June 6. Visual Pathways from the Perspective of Cost Functions and Multi-task Deep Neural Networks. arXiv [q-bio.NC]. Retrieved from. http://arxiv.org/abs/1706.01757.

Yamins, D.L.K., DiCarlo, J.J., 2016. Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. 19 (3), 356–365.

Yang, Y., Li, Y., Fermüller, C., Aloimonos, Y., 2015. Robot learning manipulation action plans by" watching" unconstrained videos from the world wide web. In: AAAI, pp. 3686–3693.

Zintgraf, L.M., Cohen, T.S., Welling, M., 2016, March 8. A New Method to Visualize Deep Neural Networks. arXiv [cs.CV]. Retrieved from. http://arxiv.org/abs/1603.02518.