Impact of Stimulus Statistics on Activity Patterns in a Model of Mouse Visual Cortex

Parsa Torabian (p2torabi@uwaterloo.ca)
Department of Systems Design Engineering, University of Waterloo
200 University Avenue West, Waterloo, ON Canada

Yinghan Chen (y977chen@uwaterloo.ca)
Department of Systems Design Engineering, University of Waterloo
200 University Avenue West, Waterloo, ON Canada

Shahab Bakhtiari (shahab.bakhtiari@umontreal.ca)
Department of Psychology, University of Montreal
Montreal, QC, Canada

Bryan Tripp (bptripp@uwaterloo.ca)
Department of Systems Design Engineering, University of Waterloo
200 University Avenue West, Waterloo, ON Canada

Abstract

A number of studies have found that deep networks can be significantly predictive of brain activity in mice, monkeys, and humans. In this work, we used an architecture that closely resembles the mouse visual cortex (MouseNet) and tested the impact of training data statistics on representational similarity with data recorded from the mouse visual system. We used the Unity engine to create eight video datasets with stimulus properties that were either realistic or artificial in three dimensions: environment, motion statistics, and optics of the modelled eye. We used each of these datasets to train the MouseNet model using a self-supervised objective. We found significant, area-dependent variations in similarity scores across different training data conditions. Notably, models trained with realistic environments consistently yielded the highest increases in similarity scores, particularly in higher areas of MouseNet. In contrast, the realistic motion conditions caused area-selective improvements or degradations in similarity scores. Furthermore, across conditions and network instances, we found that self-supervised loss and top-1 accuracy were poorly correlated with similarity to cortical representations. These results are an important step in developing models that more fully account for stimulus-driven mouse brain activity.

Keywords: Dataset; Visual Statistics; CNN; MouseNet; Mouse; Visual Cortex; Representations

Introduction

Deep artificial neural network (ANN) responses are often found to predict responses of brain networks to the same stimuli (Yamins et al., 2014; Storrs, Kietzmann, Walther, Mehrer, & Kriegeskorte, 2021). In general, ANN visual representations depend on statistics of training data, but it is unclear how these statistics affect similarity with the brain. We hypothesized that training data with properties that were more realistic and/or ethologically relevant to mice would result in more mouse brain-like responses. To test this idea, we used a video-game engine to create datasets that differed in several dimensions of realism and used them to train MouseNet (Shi, Tripp, Shea-Brown, Mihalas, & Buice, 2022), a convolutional network that is modelled on the mouse visual cortex. Consistent with our hypothesis, we found that datasets based on an ethologically relevant meadow environment resulted in more brain-like representations than datasets based on an artificial spaceship environment. Other dimensions of realism had inconsistent effects.

Realistic Natural Environment

We used two environments to generate training stimuli: A realistic and ethologically relevant ”meadow” environment and an artificial ”spaceship” environment (Figure 1). The meadow environment was more visually complex and contained natural features such as swaying grass, rocks, dirt paths, trees, apples, cliffs, and clouds. The spaceship environment had grey metallic walls, roofs, and ceilings punctuated by bright neon light strips across the walls and floor.

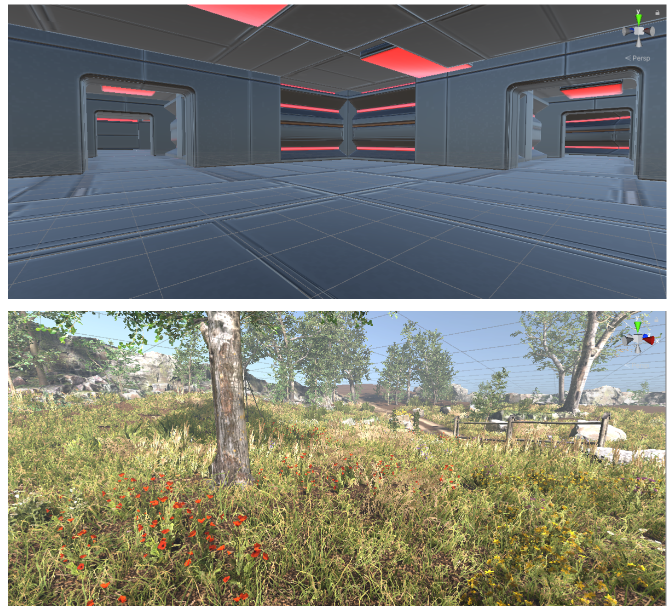

Figure 1: Screenshots of the Non-Realistic vs Realistic Environment environment from within the Unity Video Game Engine

Realistic Motion Statistics

We created stimulus videos by moving binocular cameras through these environments using two distinct kinds of motion. First, we developed a VARMA model of mouse-head motion from motion-tracking data (courtesy of Adrien Peyrache). Second, we used predominantly straight-line motion punctuated by sharp turns.

Realistic Optics of the Modelled Eye

To toggle realism in the vision of the agent, we used two sets of optical settings for the cameras that captured video stimuli. Non-realistic vision employed Unity's default pinhole camera. Realistic vision was based on mouse-eye focal depth and field of view, and averaging the red and green color channels (Prusky, Alam, & Douglas, 2006; Geng et al., 2011; Ali & Klyne, 1985).

Naming Convention

The datasets are coded with E, V, M to denote Realistic Environment, Realistic Vision, and Realistic Motion. A slash (/) is used to represent the non-realistic setting. Thus, dataset EVM is the dataset recorded using naturalistic motion within the meadow environment and mouse-approximated optical properties. Similarly, $///$ is the dataset recorded with straight line motion in the spaceship environment using the default Unity camera.

Methods

We used Dense Predictive Coding (DPC) (Han, Xie, & Zisserman, 2019), a variant of Contrastive Predictive Coding (CPC) (van den Oord, Li, & Vinyals, 2018), in which negative samples are drawn from both spatial and temporal indices of training video dataset. For each dataset condition, we trained five MouseNet models each with a different random seeds. We also randomly initialized 30 MouseNet models to compare as an "Untrained" or baseline condition. To compare the MouseNet representations with mouse brain activation, we used the Allen Brain Observatory open 2-photon calcium imaging dataset (de Vries et al., 2019) and the recordings corresponding to 30-seconds of natural movie presentation. Those same natural movies were presented to the MouseNet models and we used Representational Similarity Analysis (RSA) (Kriegeskorte, Mur, & Bandettini, 2008) as our similarity score. All changes in similarity scores are reported as the coefficients of a linear model compared to randomly initialized untrained models and only reported if significant (p-value $<$ 0.05).

Results

Realistic Environment Boosts Similarity

In the average across all Mouse-CNN area pairs (cells of the RSA matrix) and random seeds, the only models which outperformed the randomly initialized MouseNet models were those with the realistic naturalistic environment (Figure 2). The $E//$ dataset yielded the highest average increase in similarity score by 4.4%, while training with the $EV/$ and EVM datasets boosted similarity by 3.9% and 3.4% respectively.

Area Specific Effects

While the datasets with realistic environment conditions resulted in higher similarities overall, they reduced similarity with mouse areas VISp and VISam. Furthermore, training with dataset E/M reduced VISp similarity by -7.7% consistently across seeds.

The only dataset that significantly increased similarity to mouse area VISam was condition $//M$. This is notable given that this area is particularly associated with visual motion processing (Bakhtiari, Mineault, Lillicrap, Pack, & Richards, 2021; Sit & Goard, 2020). The $//M$ dataset increased similarity to VISam by 1.2% whereas other datasets yielded decreases of -1.3% to -4.9%.

Contrastive Predictive Loss vs Brain Similarity

Contrastive predictive training is effective at training predictive models of brain activation (Nayebi et al., 2023),

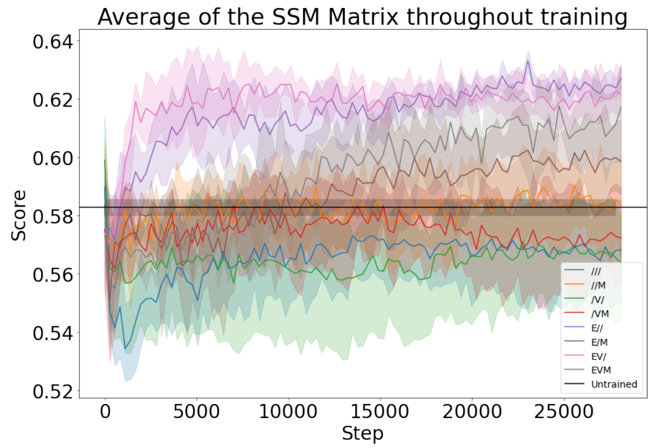

Figure 2: Similarity scores throughout training showing which dataset conditions perform better than a randomly initialized model in the aggregate.

but we found that contrastive loss and accuracy were not closely related to similarity between different input conditions. For example, $///$ and $/V/$ led to the lowest loss and highest training accuracy, but these models performed poorly in terms of average similarity (Figure 3). Conversely, $E//$ and $EV/$ show the highest similarity scores despite larger contrastive losses, relative to the other dataset conditions.

Higher loss may arise from stimuli that are more complex or less predictable, which may force the model to learn more sophisticated representations. These representations may in turn be more biologically realistic.

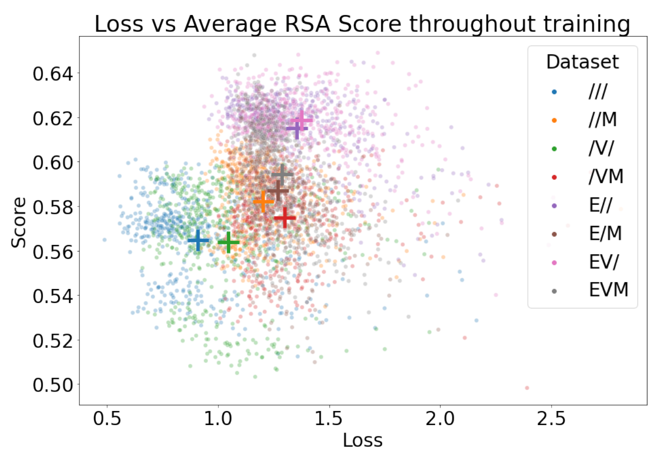

Figure 3: Representational similarity vs training loss at each epoch. '+' Markers represent the average similarity score and loss throughout training.

Acknowledgments

Supported by the Natural Sciences and Engineering Research Council of Canada, RGPIN-05855. SB is funded by NSERC Discovery grants (RGPIN-2023-03875). Mouse motion data courtesy of Adrien Peyrache.

References

Ali, M. A., & Klyne, M. A. (1985). *Vision in vertebrates*. Retrieved from https://doi.org/10.1007/978-1-4684-9129-6 doi: 10.1007/978-1-4684-9129-6

Bakhtiari, S., Mineault, P., Lillicrap, T., Pack, C., & Richards, B. (2021). The functional specialization of visual cortex emerges from training parallel pathways with self-supervised predictive learning. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, & J. W. Vaughan (Eds.), *Advances in neural information processing systems* (Vol. 34, pp. 25164–25178). Curran Associates, Inc.

de Vries, S. E. J., Lecoq, J. A., Buice, M. A., Groblewski, P. A., Ocker, G. K., Oliver, M., ... et al. (2019, Dec). *A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex.* Nature Publishing Group.

Geng, Y., Schery, L. A., Sharma, R., Dubra, A., Ahmad, K., Libby, R. T., & Williams, D. R. (2011, 2). Optical properties of the mouse eye. *Biomedical optics express*, *2*(4), 717. Retrieved from https://doi.org/10.1364/boe.2.000717 doi: 10.1364/boe.2.000717

Han, T., Xie, W., & Zisserman, A. (2019). Video representation learning by dense predictive coding. *CoRR*, *abs/1909.04656*. Retrieved from http://arxiv.org/abs/1909.04656

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008, Oct). *Representational similarity analysis - connecting the branches of systems neuroscience.* Frontiers.

Nayebi, A., Kong, N. C. L., Zhuang, C., Gardner, J. L., Norcia, A. M., & Yamins, D. L. K. (2023, Oct). *Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation.* Public Library of Science. Retrieved from https://doi.org/10.1371/journal.pcbi.1011506

Prusky, G. T., Alam, N. M., & Douglas, R. M. (2006, 11). Enhancement of vision by monocular deprivation in adult mice. *The journal of neuroscience*, *26*(45), 11554–11561. Retrieved from https://doi.org/10.1523/jneurosci.3396-06.2006 doi: 10.1523/jneurosci.3396-06.2006

Shi, J., Tripp, B., Shea-Brown, E., Mihalas, S., & Buice, M. A. (2022, 9). MouseNet: A biologically constrained convolutional neural network model for the mouse visual cortex. *PLOS computational biology/PLoS computational biology*, *18*(9), e1010427. Retrieved from https://doi.org/10.1371/journal.pcbi.1010427 doi: 10.1371/journal.pcbi.1010427

Sit, K. K., & Goard, M. J. (2020, Jul). *Distributed and retinotopically asymmetric processing of coherent motion in mouse visual cortex.* Nature Publishing Group. Retrieved from https://www.nature.com/articles/s41467-020-17283-5

Storrs, K. R., Kietzmann, T. C., Walther, A., Mehrer, J., & Kriegeskorte, N. (2021, 8). Diverse deep neural networks all predict human inferior temporal cortex Well, after training and fitting. *Journal of cognitive neuroscience*, 1–21. Retrieved from https://doi.org/10.1162/jocn_a_01755 doi: 10.1162/jocn_a_01755

van den Oord, A., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *CoRR*, *abs/1807.03748*. Retrieved from http://arxiv.org/abs/1807.03748

Yamins, D., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014, 5). 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), 8619–8624. Retrieved from https://doi.org/10.1073/pnas.1403112111 doi: 10.1073/pnas.1403112111