Exploring the role of image domain in self-supervised DNN models of rodent brains

Anonymous Author(s)
Affiliation
Address
email

Abstract

Biological visual systems have evolved around the efficient coding of natural image statistics in order to support recognition of complex visual patterns. Recent work shown that deep neural networks are able to learn similar representations to those measured in visual areas in animals, suggesting they may serve as models for the brain. Varying network architectures and loss functions has been shown to modulate the biological similarity learned representations, however the extent to which this results from exposure to natural image statistics during training has not been fully characterized. Here, we use self-supervised learning to train neural network models across a range of data domains with different image statistics and evaluate the similarity of the learned representations to neural activity of the mouse visual cortex. We find that networks trained on different domains also exhibit different responses when shown held-out natural images. Furthermore, we find that the degree of biological similarity of the representations generally increases as a function of the naturalness of the data domain used for training. Our results provide evidence for the idea that that the training data domain is an important component when modeling the visual system using deep neural networks.

1 Introduction

In recent years, it has been shown that task-driven deep neural networks (DNNs) are the highly accurate normative models for predicting the neural responses in both the primate [22, 21] and mouse ventral visual stream [2, 15, 10, 4]. Despite their success as models of the brain, it has not been fully elucidated how each component of the DNN contributes to their capacity to learn biologically similar representations. Namely, three aspects of DNNs may be essential: the **network architecture**, the **loss function**, and the **training data**.

Recent studies argue that the **network architecture** can have a significant impact on biological similarity of learned representations [1, 5, 4, 16]. [10] used a shallow, multi-stream network to achieve state-of-the art results in neural predictivity for the mouse visual cortex, showing that similarity in topology impacts similarity in representations. Early work relied on task-driven **loss functions** (e.g., object recognition), but it has recently been demonstrated that models trained using self-supervised loss functions can reach the same or greater level of neural predictivity [24]. Very few studies [5, 10], however, have investigated the the role of **training data** domain on the neural predictivity of learned representations. While it is thought that the visual stream has specifically evolved around the efficient coding of natural image statistics [18, 11], it is still not clear to what extent this may also apply to DNN models of the brain.

Here, we investigate the role natural image statistics play in the formation of biologically plausible representations within DNN models of the mouse visual cortex. To do this, we first train separate

Submitted to 4th Workshop on Shared Visual Representations in Human and Machine Visual Intelligence (SVRHM) at NeurIPS 2022. Do not distribute.

models using self-supervised learning on image datasets from visually distinct domains with varying degrees of "naturalness". We then compare the representations of these networks to each other to confirm whether they are indeed learning distinct feature representations. Finally, we probe the extent to which models trained on domains of varying "naturalness" are predictive of biological neural responses. We find that the naturalness of the training image distribution affects the degree to which the network is able to learn biologically-realistic representations, and find that this is reflected in emergence of canonical visual feature detector motifs such as Gabor filters and textured patches. We see this effect most profoundly in the earlier layers of the artificial network indicating that they are more sensitive to low-level image features. We will make all of our training and analysis code and data available upon publication. .

2 Methods

2.1 Datasets

Figure 1: Sample images from each dataset. Networks were trained on ImageNet, DomainNet, and CREMI, then evaluated on AIVC and Sensorium.

ImageNet [7] is a multiclass object categorization dataset with $C = 1000$ classes containing a training set with around 1.3 million images and a validation set of around 50 thousand images. In order to control for dataset size and since the dataset is already downsized to 64×64 (shown to be important for modeling mouse brains [10], we used the TinyImageNet [9] dataset which contains about 100k training images and about 20k validation images with only 200 classes.

DomainNet [13] is a large-scale multi-source domain adaptation dataset containing 6 domains each having 345 categories: Real (510 images per category/175,327 total), Infograph (150 images per category/53,201 total), Clipart (150 images per category/48,833 total), Painting (220 images per category/75,759 total), Sketch (220 images per category/70,386 total), and Quickdraw (500 images per category/17,2500 total).

CREMI(**C**ircuit **R**econstruction from **E**lectron **M**icroscopy **I**mages) is an electron microscopy dataset made up of three datasets, each consisting of two 5 μm^3 volumes (training and testing, each $1250 \times 1250 \times 125$ px of serial section EM of the adult fly brain)[1]. We treated each slice in the z dimension as a separate image and then generated 100 125×125 tiles per image resulting in a dataset with 27,750 images which we then trained on.

Sensorium is a large-scale dataset from mouse primary visual cortex containing the responses of more than 28,000 neurons across seven mice stimulated with thousands of natural images, together with simultaneous behavioral measurements that include running speed, pupil dilation, and eye movements [20].

Allen Institute Visual Coding Dataset (**AIVC**) [6, 17] is a mouse neural recording dataset containing electrophysiological and calcium fluorescence recordings of around 60 thousand neurons from the VISp, VISl, VISal, VISam, VISpm, and VISrl regions of 221 different mice. We limited our study to the neural responses of animals in response to the Natural Scenes stimulus set which consisted of 118 grayscale natural images. After following the filtering process outlined by [4] and [10] we were left with around 17,000 neurons across 98 different specimen(calcium) and 1422 neurons across 42 different specimen(neuropixels) across the six visual areas.

2.2 Network architecture and training

Because we were interested in only how the input stimuli affected the representations formed, we decided to keep the network architecture fixed and train a network on each domain. Following [10],

[1] MICCAI Challenge on Circuit Reconstruction from Electron Microscopy Images (CREMI): https://cremi.org

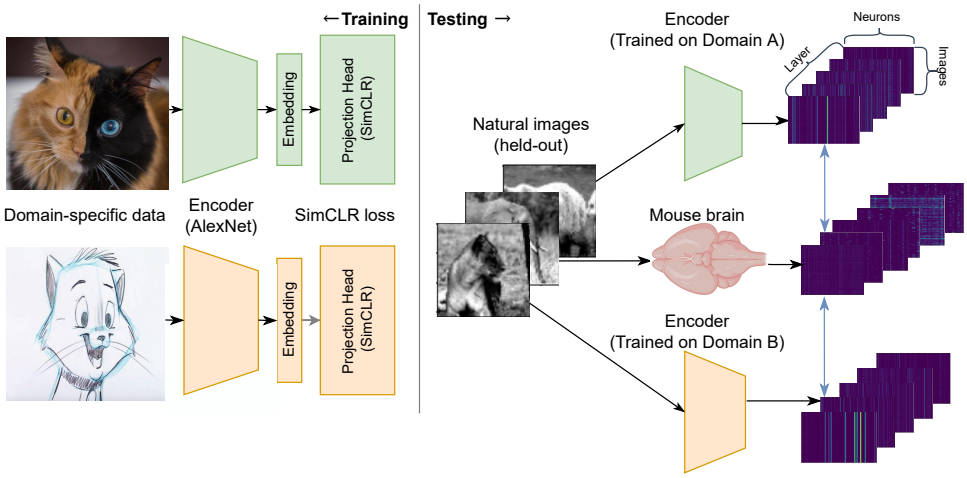

Figure 2: Experiment Pipeline. From left to right: First, we train separate SimCLR-based contrastive AlexNet models on each domain. We then extract activations from each network in response to the AIVC and Sensorium image dataset and evaluate the representational similarity learned by each network to each other as well as to biological neural recordings.

we chose to use a shallow network trained on a contrastive objective function as this has been shown to be the best DNN-based model of the mouse visual cortex. Specifically, we used Pytorch Lightning and the Lightly Self-supervised library to train an AlexNet [8] with a SimCLR loss [3] by replacing the fully-connected layers with a 2-layer MLP projection head that produces a 128-dimensional feature embedding. We trained each network for 600 epochs using the NXent-Loss [3] with temperature scaling $\tau = 0.1$. We used a LARS [23] optimizer with a batch size of 4096, learning rate of 4.8, momentum of 0.9 and weight decay of 10^{-6} and we decayed the learning rate by a factor of 0.5 if the loss did not decrease by 0.1 every 10 epochs. To ensure that the distribution of image statistics were not altered due to augmentations, we restricted them to: random-resized crop, random horizontal and vertical flips, and random gray-scaling. Every image dataset was reduced to a resolution of 64×64 before applying augmentations in order to better model the mouse's visual acuity [14, 10].

3 Experiments and Results

3.1 Training data domain has an impact on neural predictivity

We first asked whether training contrastive models on different image domains would result in significantly different distributions of activations in response to a (potentially out-of-domain) stimulus set of natural images (AIVC and Sensorium). In other words, we wanted to verify the extent to which DNNs learn the same set of features independent of the data domain. For each network, we extracted the activations from every convolutional layer by forward-propagating each image from the AIVC stimulus set until the activation layer succeeding that convolutional layer to get an $(h \times w \times c)$ tensor.

Given the actions, we calculated the l_2-norm of each image activation to measure the distribution of how much each network responded to common images. For each image, we calculated the pairwise Euclidean distance between the activations from each network in order to get an estimate for how far away the representations of the different networks were. Our results (Figure 3) show that the distribution of activations is indeed different across networks indicating that the networks are learning, unrelated features as a function of data domain.

To verify this result, we hypothesized that if the activations of each network are truly distinct, then it should be possible to find a hyperplane that separates each domain from each other within the activation space. Using the same activations, we trained a separate linear support vector machine (SVM) to classify which domain each set of activations came from. We first flattened the tensors into an $(n \times h \cdot w \cdot c)$ dataset, then performed a 5-fold cross-validation and measured the accuracy, precision, and recall scores for each. Indeed, the SVM was able to classify which domain a given

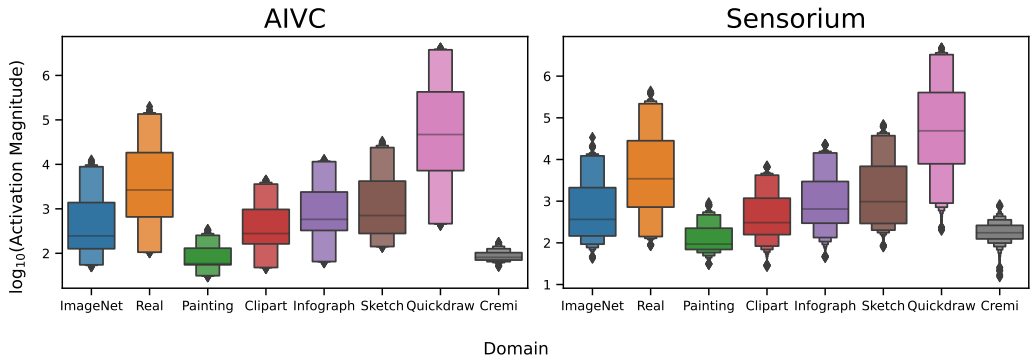

Figure 3: Distributions of activation magnitudes in response to natural images differ across networks trained on different domains. Network activations were extracted per convolutional layer in response to each image in both the AIVC dataset (*left*) and the Sensorium Dataset (*right*)

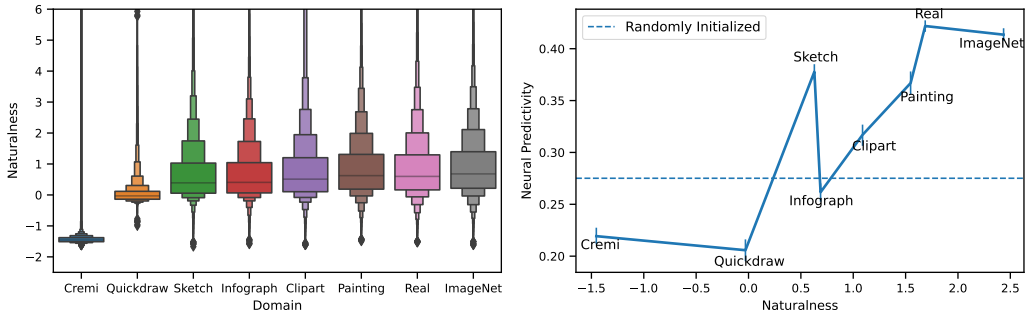

Figure 4: *Left*: Naturalness of each domain approximated by $-\log|p-2|$ after fitting the power spectrum of a sample of images from each domain to the form $\frac{1}{f^p}$. *Right*: The neural predictivity of the first convolutional layer of ANNs trained various domains plotted against its naturalness. The neural predictivity was calculated against the calcium recordings from the AIVC dataset.

convolutional layer's activation was from with near certainty (classification precision of 0.9876 and 0.9998 for AIVC and Sensorium respectively).

3.2 Models trained on different data domains also have different degrees of biological similarity

Given that models trained on different domains exhibit different responses to natural images, we next asked whether these differences in representations were associated with changes in neural predictivity. We calculated the neural predictivity of each layer of each model to each visual region of the AIVC dataset as follows. We first mapped the model activations to their biological counter part on one half of the image set using a partial least squares regression model. We then calculated the inter-animal consistency metric described in [10] using the held out set. However, in our case we decided to keep the measured predictivity per layer rather than for the overall model so we did not take the maximum over all the convolutional layers. We ran 100 trials for each domain using different train-test splits each time.

We find that models trained on different data domains exhibit significant differences in neural predictivity ($p = 1.48 \times 10^{-45}$, $p = 3.46 \times 10^{-58}$, and $p = 2.44 \times 10^{-79}$ using a one-sided Mann-Whitney U test for AIVC-calcium, AIVC-neuropixels, and Sensorium respectively).

3.3 Training with more natural image domains increases neural predictivity

We hypothesized that this association may be explained by the "naturalness" of the image domain from which representations were derived. To measure this quantity, we first conducted a spectral

Table 1: Neural Predictivity Results.

Domain	Naturalness	PLS-AIVC(calcium)	PLS-AIVC(ephys)	PLS-Sensorium
CREMI	-1.453	0.219	0.614	0.129
Quickdraw	-0.029	0.331	0.786	0.147
Sketch	0.629	0.305	0.750	0.129
Infograph	0.688	0.278	0.755	0.143
Clipart	1.089	0.303	0.744	0.143
Painting	1.548	0.260	0.667	0.154
Real	1.687	0.342	0.808	0.156
ImageNet	**2.436**	**0.352**	**0.810**	**0.170**

analysis on the statistics of a subsample of 1000 images in each dataset to evaluate their adherence to the $\frac{1}{f^p}$ power law that has been previously associated with natural images owing, in part, to their texture bias [19, 18]. We used the distance of the average p value to the idealized $p = 2$ as an approximation for the "naturalness" of the domain.

We find that more natural domains like ImageNet ($p = 2.06 \pm 0.76$) and DomainNet-Real ($p = 1.97 \pm 0.79$) are closer to the previously reported $p = 2$ [19] than less natural domains like DomainNet-Quickdraw ($p = 1.04 \pm 0.31$) and CREMI ($p = 6.19 \pm 0.54$) (Figure 4).

Next, using this parameter, we compared the naturalness of models trained on different domains to the neural predictivity of the learned representations. We found that models trained on more natural domains generally appear to exhibit higher neural predictivity (Figure 4, Table 1). This effect is especially clear in early convolutional layers (conv1), with mixed trends for deeper layers deeper (**See Appendix**).

3.4 Biological visual system motifs emerge when training with more natural data

Previous work has shown that Gabor-like filter banks emerge in the earlier layers of the visual stream due to the tuning of these neurons to edges and other natural statistics [18], and in computational models of vision [12]. We wondered whether there was a similar functional reason for the gap in neural similarity between naturalistic and non-naturalistic domains. Although qualitative, visualizing the first convolutional kernel of our models suggests that the Gabor-like filters are learned more frequently in models trained on more natural data (**See Appendix**).

4 Discussion

Overall, our results show that the input data distribution can have significant impacts on the biological similarity of the representations learned by self-supervised models of the ventral visual stream. Importantly, we demonstrate that the "naturalness" of the data domain has a correlation with the neural predictivity of the learned representations. This effect is most pronounced in the earlier layers of the models, which we hypothesize might be a result of the the earlier layers serving as low level feature detectors (e.g., edge detectors) which would change accordingly with respect to the low level image statistics of the data that it is exposed to. Given that the brain is optimized to natural image statistics, its representations should differ from one optimized for images from other domains. One major limitation of our study, however, is that while our more natural image domains are more "natural" in the spectral sense, they are not natural in the ethological sense. In other words, animals such as mice are less likely to have evolved to learn from images of cars and houses, than they are of shrubbery and predators. Going forward, we believe that it is important not just to take into consideration the architecture and loss function as hyperparameters when modeling the brain with deep neural networks, but also the domain of the training data.

5 Acknowledgements

We would like to thank Aran Nayebi and Nathan Kong for their support in replicating their work.

References

[1] Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like? | bioRxiv. URL https://www.biorxiv.org/content/10.1101/407007v2.abstract.

[2] Santiago A. Cadena, Fabian H. Sinz, Taliah Muhammad, Emmanouil Froudarakis, Erick Cobos, Edgar Y. Walker, Jake Reimer, Matthias Bethge, Andreas Tolias, and Alexander S. Ecker. How well do deep neural networks trained on object recognition characterize the mouse visual system? July 2022. URL https://openreview.net/forum?id=rkxcXmtUUS.

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. Technical Report arXiv:2002.05709, arXiv, June 2020. URL http://arxiv.org/abs/2002.05709. arXiv:2002.05709 [cs, stat] type: article.

[4] Colin Conwell, David Mayo, Andrei Barbu, Michael Buice, George Alvarez, and Boris Katz. Neural Regression, Representational Similarity, Model Zoology & Neural Taskonomy at Scale in Rodent Visual Cortex. In *Advances in Neural Information Processing Systems*, volume 34, pp. 5590–5607. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/2c29d89cc56cdb191c60db2f0bae796b-Abstract.html.

[5] Colin Conwell, Jacob S. Prince, George A. Alvarez, and Talia Konkle. What can 5.17 billion regression fits tell us about artificial models of the human visual system? December 2021. URL https://openreview.net/forum?id=i_xiyGq6FNT.

[6] Saskia E. J. de Vries, Jerome A. Lecoq, Michael A. Buice, Peter A. Groblewski, Gabriel K. Ocker, Michael Oliver, David Feng, Nicholas Cain, Peter Ledochowitsch, Daniel Millman, Kate Roll, Marina Garrett, Tom Keenan, Leonard Kuan, Stefan Mihalas, Shawn Olsen, Carol Thompson, Wayne Wakeman, Jack Waters, Derric Williams, Chris Barber, Nathan Berbesque, Brandon Blanchard, Nicholas Bowles, Shiella D. Caldejon, Linzy Casal, Andrew Cho, Sissy Cross, Chinh Dang, Tim Dolbeare, Melise Edwards, John Galbraith, Nathalie Gaudreault, Terri L. Gilbert, Fiona Griffin, Perry Hargrave, Robert Howard, Lawrence Huang, Sean Jewell, Nika Keller, Ulf Knoblich, Josh D. Larkin, Rachael Larsen, Chris Lau, Eric Lee, Felix Lee, Arielle Leon, Lu Li, Fuhui Long, Jennifer Luviano, Kyla Mace, Thuyanh Nguyen, Jed Perkins, Miranda Robertson, Sam Seid, Eric Shea-Brown, Jianghong Shi, Nathan Sjoquist, Cliff Slaughterbeck, David Sullivan, Ryan Valenza, Casey White, Ali Williford, Daniela M. Witten, Jun Zhuang, Hongkui Zeng, Colin Farrell, Lydia Ng, Amy Bernard, John W. Phillips, R. Clay Reid, and Christof Koch. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature Neuroscience*, 23(1):138–151, January 2020. ISSN 1546-1726. doi: 10.1038/s41593-019-0550-9. URL https://www.nature.com/articles/s41593-019-0550-9. Number: 1 Publisher: Nature Publishing Group.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, June 2009. doi: 10.1109/CVPR.2009.5206848. ISSN: 1063-6919.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

[9] Ya Le and Xuan Yang. Tiny ImageNet Visual Recognition Challenge. pp. 6.

[10] Aran Nayebi, Chengxu Zhuang, Anthony M. Norcia, Nathan C. L. Kong, Justin L. Gardner, and Daniel L. K. Yamins. Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation. Technical report, bioRxiv, July 2022. URL https://www.biorxiv.org/content/10.1101/2021.06.16.448730v3. Section: New Results Type: article.

[11] B A Olshausen and D J Field. Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7(2):333–339, January 1996. ISSN 0954-898X. doi: 10.1088/0954-898X_7_2_014. URL https://doi.org/10.1088/0954-898X_7_2_014. Publisher: Taylor & Francis _eprint: https://doi.org/10.1088/0954-898X_7_2_014.

[12] Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, June 1996. ISSN 1476-4687. doi: 10.1038/381607a0. URL https://www.nature.com/articles/381607a0. Number: 6583 Publisher: Nature Publishing Group.

[13] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment Matching for Multi-Source Domain Adaptation. Technical Report arXiv:1812.01754, arXiv, August 2019. URL http://arxiv.org/abs/1812.01754. arXiv:1812.01754 [cs] type: article.

[14] G. T. Prusky, P. W. West, and R. M. Douglas. Behavioral assessment of visual acuity in mice and rats. *Vision Research*, 40(16):2201–2209, 2000. ISSN 0042-6989. doi: 10.1016/s0042-6989(00)00081-x.

[15] Jianghong Shi, Eric Shea-Brown, and Michael A. Buice. Comparison Against Task Driven Artificial Neural Networks Reveals Functional Organization of Mouse Visual Cortex. Technical Report arXiv:1911.07986, arXiv, November 2019. URL http://arxiv.org/abs/1911.07986. arXiv:1911.07986 [q-bio] type: article.

[16] Jianghong Shi, Michael A. Buice, Eric Shea-Brown, Stefan Mihalas, and Bryan Tripp. A Convolutional Network Architecture Driven by Mouse Neuroanatomical Data. Technical report, bioRxiv, October 2020. URL https://www.biorxiv.org/content/10.1101/2020.10.23.353151v1. Section: New Results Type: article.

[17] Joshua H. Siegle, Xiaoxuan Jia, Séverine Durand, Sam Gale, Corbett Bennett, Nile Graddis, Greggory Heller, Tamina K. Ramirez, Hannah Choi, Jennifer A. Luviano, Peter A. Groblewski, Ruweida Ahmed, Anton Arkhipov, Amy Bernard, Yazan N. Billeh, Dillan Brown, Michael A. Buice, Nicolas Cain, Shiella Caldejon, Linzy Casal, Andrew Cho, Maggie Chvilicek, Timothy C. Cox, Kael Dai, Daniel J. Denman, Saskia E. J. de Vries, Roald Dietzman, Luke Esposito, Colin Farrell, David Feng, John Galbraith, Marina Garrett, Emily C. Gelfand, Nicole Hancock, Julie A. Harris, Robert Howard, Brian Hu, Ross Hytnen, Ramakrishnan Iyer, Erika Jessett, Katelyn Johnson, India Kato, Justin Kiggins, Sophie Lambert, Jerome Lecoq, Peter Ledochowitsch, Jung Hoon Lee, Arielle Leon, Yang Li, Elizabeth Liang, Fuhui Long, Kyla Mace, Jose Melchior, Daniel Millman, Tyler Mollenkopf, Chelsea Nayan, Lydia Ng, Kiet Ngo, Thuyahn Nguyen, Philip R. Nicovich, Kat North, Gabriel Koch Ocker, Doug Ollerenshaw, Michael Oliver, Marius Pachitariu, Jed Perkins, Melissa Reding, David Reid, Miranda Robertson, Kara Ronellenfitch, Sam Seid, Cliff Slaughterbeck, Michelle Stoecklin, David Sullivan, Ben Sutton, Jackie Swapp, Carol Thompson, Kristen Turner, Wayne Wakeman, Jennifer D. Whitesell, Derric Williams, Ali Williford, Rob Young, Hongkui Zeng, Sarah Naylor, John W. Phillips, R. Clay Reid, Stefan Mihalas, Shawn R. Olsen, and Christof Koch. A survey of spiking activity reveals a functional hierarchy of mouse corticothalamic visual areas. preprint, Neuroscience, October 2019. URL http://biorxiv.org/lookup/doi/10.1101/805010.

[18] Eero P Simoncelli and Bruno A Olshausen. Natural Image Statistics and Neural Representation. *Annual Review of Neuroscience*, 24(1):1193–1216, March 2001. ISSN 0147-006X, 1545-4126. doi: 10.1146/annurev.neuro.24.1.1193. URL https://www.annualreviews.org/doi/10.1146/annurev.neuro.24.1.1193.

[19] Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: Computation in Neural Systems*, 14(3):391–412, January 2003. ISSN 0954-898X. doi: 10.1088/0954-898X_14_3_302. URL https://doi.org/10.1088/0954-898X_14_3_302. Publisher: Taylor & Francis _eprint: https://doi.org/10.1088/0954-898X_14_3_302.

[20] Konstantin F. Willeke, Paul G. Fahey, Mohammad Bashiri, Laura Pede, Max F. Burg, Christoph Blessing, Santiago A. Cadena, Zhiwei Ding, Konstantin-Klemens Lurz, Kayla Ponder, Taliah Muhammad, Saumil S. Patel, Alexander S. Ecker, Andreas S. Tolias, and Fabian H. Sinz. The sensorium competition on predicting large-scale mouse primary visual cortex activity, 2022. URL https://arxiv.org/abs/2206.08666.

[21] Daniel L. K. Yamins and James J. DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3):356–365, March 2016. ISSN 1546-1726. doi: 10.1038/nn.4244. URL https://www.nature.com/articles/nn.4244. Number: 3 Publisher: Nature Publishing Group.

[22] 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*, 111(23):8619–8624, June 2014. doi: 10.1073/pnas.1403112111. URL `https://www.pnas.org/doi/10.1073/pnas.1403112111`. Publisher: Proceedings of the National Academy of Sciences.

[23] Yang You, Igor Gitman, and Boris Ginsburg. Large Batch Training of Convolutional Networks. Technical Report arXiv:1708.03888, arXiv, September 2017. URL `http://arxiv.org/abs/1708.03888`. arXiv:1708.03888 [cs] type: article.

[24] Chengxu Zhuang, Siming Yan, Aran Nayebi, Martin Schrimpf, Michael C. Frank, James J. DiCarlo, and Daniel L. K. Yamins. Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences*, 118(3):e2014196118, January 2021. doi: 10.1073/pnas.2014196118. URL `https://www.pnas.org/doi/10.1073/pnas.2014196118`. Publisher: Proceedings of the National Academy of Sciences.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See 4
   (c) Did you discuss any potential negative societal impacts of your work? [No] As far as we are aware there are no potential negative societal impacts to our work
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [No] Our work is not theoretical
   (b) Did you include complete proofs of all theoretical results? [No] Our work is not theoretical
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] We will provide all code and data upon request or upon publication to maintain anonymity. See 1
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] see 2
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] See figures in 3
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] **See Appendix**
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [No] As far as we are aware there are no licenses to mention
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No] All data used in this project was open sourced
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No] There is no personally identifiable information or offensive content
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [No] We did not crowd source or use human subjects

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [No] We did not crowd source or use human subjects

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [No] We did not crowd source or use human subjects