A 7T fMRI dataset of synthetic images for out-of-distribution modeling of vision

Alessandro T. Gifford (alessandro.gifford@gmail.com)
Department of Education and Psychology, Freie Universität Berlin
Kaiserswerther Str. 16-18, 14195 Berlin, Germany

Radoslaw M. Cichy (rmcichy@zedat.fu-berlin.de)
Department of Education and Psychology, Freie Universität Berlin
Kaiserswerther Str. 16-18, 14195 Berlin, Germany

Thomas Naselaris (tnaselar@gmail.com)
Department of Neuroscience, University of Minnesota
321 Church Street SE, Minneapolis, MN 55455, United States

Kendrick Kay (kay@umn.edu)
Department of Radiology, University of Minnesota
2021 6th St SE, Minneapolis, MN 55455, United States

Abstract

Large-scale visual neural datasets such as the Natural Scenes Dataset (NSD) are boosting NeuroAI research by enabling computational models of the brain with performances beyond what was possible just a decade ago. However, these datasets lack out-of-distribution (OOD) components, which are crucial for the development of more robust models. Here, we address this limitation by releasing NSD-synthetic, a dataset consisting of 7T fMRI responses from the eight NSD subjects for 284 carefully controlled synthetic images. We show that NSD-synthetic's fMRI responses are OOD with respect to NSD, that brain encoding models exhibit reduced performance when tested OOD on NSD-synthetic compared to when tested in-distribution (ID) on NSD, and that OOD tests on NSD-synthetic reveal differences between encoding models not detected by ID tests—specifically, self-supervised deep neural networks better explain neural responses than their task-supervised counterparts. These results showcase how NSD-synthetic enables OOD generalization tests that facilitate the development of more robust models of visual processing, and the formulation of more accurate theories of human vision.

Keywords: visual neuroscience; NSD; fMRI; data release; out-of-distribution generalization; encoding models

Introduction

Large-scale visual neural datasets such as NSD (Allen et al., 2022) are enabling new progress in computational models of the brain, leading to new theoretical insights (Doerig et al., 2022; Khosla, Murty, & Kanwisher, 2022). However, despite the unprecedented size of these datasets, their stimuli typically live within the same visual distribution. Therefore, it remains difficult to use these datasets for OOD generalization tests essential for the development of more robust computational models of the brain. To enable these crucial OOD generalization tests, here we release a companion dataset to NSD called *NSD-synthetic* (in the following we refer to NSD as *NSD-core*, to distinguish it from NSD-synthetic). NSD-synthetic consists of fMRI responses for an additional scan session from the same 8 subjects of NSD-core. During this session, fMRI responses were measured to 284 carefully controlled synthetic (non-naturalistic) images (**Figure 1A**).

Methods and Results

NSD-synthetic's fMRI Responses Are Out-of-distribution With Respect To NSD-core

To be well suited for OOD generalization tests, NSD-synthetic's fMRI responses should exhibit a different distribution compared to NSD-core's responses. To ascertain this, we reduced NSD-synthetic and NSD-core's fMRI responses to two dimensions using multidimensional scaling and compared the resulting two-dimensional MDS embeddings. We found that the fMRI responses for NSD-synthetic form a distinct cluster compared to the fMRI responses from NSD-core (**Figure 1B**). NSD-synthetic's fMRI responses cluster based on the image class to which they belong, and responses for scene images are closer in the MDS embedding space to NSD-core, which is reasonable given that NSD-core consists of fMRI responses to scenes. In summary, NSD-synthetic's fMRI responses are differently distributed (i.e., OOD) with respect to NSD-core.

Brain Encoding Models Exhibit Reduced Performance When Tested Out-of-distribution On NSD-synthetic

Next, we asked whether NSD-synthetic is useful for OOD generalization tests of computational models of the brain, using encoding models (Naselaris, Kay, Nishimoto, & Gallant, 2011). The encoding models consisted of linear regressions that mapped image features from AlexNet onto the fMRI responses of each vertex. We trained the linear regression weights using image features and fMRI responses from NSD-core, and tested their generalization performance both ID (using NSD-core data not used for model training) and OOD (using NSD-synthetic). We found that the encoding models trained on NSD-core predict visual fMRI responses better ID than OOD (**Figure 1C**). The fact that OOD generalization is lower is crucial, as it indicates that further improvement is needed to capture stimulus-to-brain-response relationships, thus validating NSD-synthetic's usefulness for model and theory development.

Out-of-distribution Generalization Tests Reveal Differences Between Brain Models Not Detected By In-distribution Tests

Finally, we asked whether out-of-distribution generalization tests reveal differences, between brain encoding models based on highly similar deep neural networks, that are not detected by ID tests. Specifically, we compared encoding models based on ResNet-50, a convolutional neural network pretrained on task-supervised image classification, with encoding models based on MoCo, consisting of the same ResNet-50 architecture but trained with self-supervised contrastive learning (Zbontar, Jing, Misra, LeCun, & Deny, 2021). When tested ID, we found that ResNet-50 outperformed MoCo in higher-level visual areas (blue areas in **Figure 1D**, left panel), and that MoCo outperformed ResNet-50 in lower-level visual areas (red areas in **Figure 1D**, left panel). When tested OOD, MoCo instead outperformed ResNet-50 across both lower- and higher-level visual areas (red areas in **Figure 1D**, right panel). The ID difference between the two models was smaller (peaking at 5% of absolute explained variance difference, **Figure 1D**, left panel) compared to their OOD difference (peaking at 10% of absolute explained variance difference, **Figure 1D**, right panel), indicating that OOD tests allow to better disentangle brain models compared to ID tests. Thus, OOD tests reveal differences not detected by ID tests, showing that NSD-synthetic provides unique insight essential for developing more robust models of the brain.

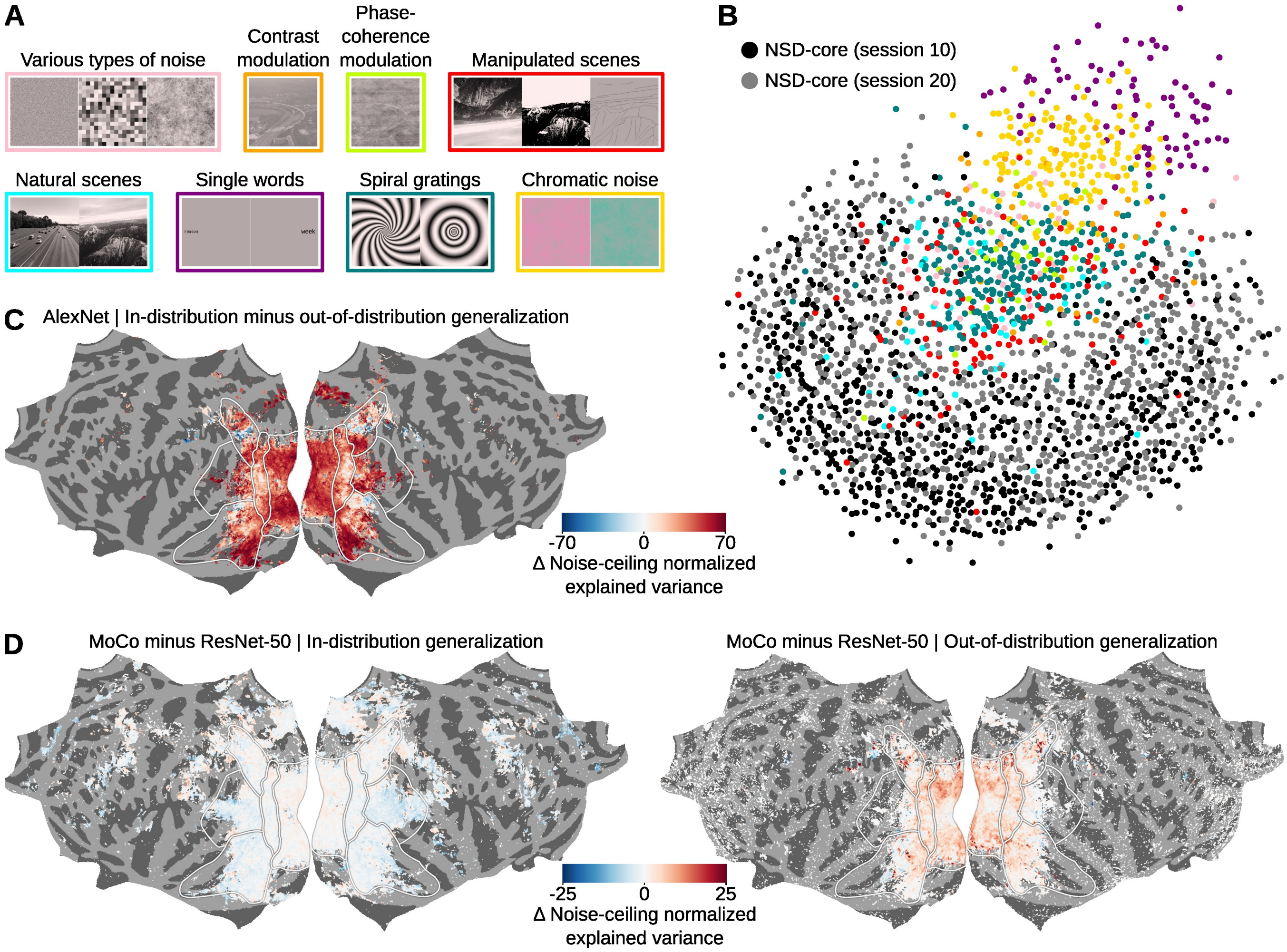

Figure 1: NSD-synthetic enables OOD modeling of vision. A. Example stimulus images in NSD-synthetic, divided in eight classes. **B.** Multidimensional scaling embedding of subject-aggregated fMRI responses from NSD-synthetic (colored dots) and NSD-core (black/gray dots). **C.** Subject-average difference between encoding models ID and OOD encoding accuracy, plotted on flattened cortical surfaces. White contours indicate early, intermediate, ventral, lateral, and dorsal visual stream. **D.** Difference between MoCo and ResNet-50's encoding accuracy.

Discussion

These findings establish NSD-synthetic as a useful out-of-distribution companion to NSD-core. Beyond indicating that further model improvement is required to accurately describe visual representations in the brain, NSD-synthetic also suggests directions for closing the gap between ID and OOD generalization performance.

One direction of model improvement comes from consideration of the properties of NSD-synthetic's stimuli. Since the stimuli consist of a multitude of diverse and carefully parameterized images, they facilitate discovery of the specific OOD visual features to which computational models fail to generalize, which in turn provide explicit objectives for the engineering of more robust models (Madan et al., 2024).

Another direction of model improvement comes from benchmarking the performance of different computational models on NSD-synthetic and isolating model properties leading to best OOD generalization (Ren & Bashivan, 2024). Because each model embeds a different hypothesis of visual processing, these OOD generalization tests can help adjudicate between competing hypotheses, therefore facilitating theory formation. As an example of this, we found that encoding models based on self-supervised deep neural networks better generalize OOD than encoding models based on their task-supervised counterparts, in line with recent work proposing self-supervision as a more plausible account of coding in visual cortex (Prince, Alvarez, & Konkle, 2024).

In conclusion, we invite researchers to take advantage of this new dataset in building more robust and accurate computational models of the brain, leading to better theories of vision. All data are publicly available, including data in ready-to-use preprocessed formats[1]. The code to reproduce the analyses of the NSD data shown in this paper is available on GitHub[2].

[1] http://naturalscenesdataset.org
[2] https://github.com/gifale95/NSD-synthetic

Acknowledgments

Collection of the NSD dataset was supported by NSF IIS-1822683 (K.K.) and NSF IIS-1822929 (T.N.). This work was supported by NIH grant R01EY034118 (K.K.), German Research Council (DFG) grants (CI 241/1-3, CI 241/1-7, INST 272/297-1) (R.M.C.), and European Research Council (ERC) Consolidator grant (ERC-CoG-2024101123101) (R.M.C.). We thank J. Bosten, B. Broderick, and A. White for consultation regarding stimulus design. We also thank E. Allen and Y. Wu for collecting the neuroimaging data, and the HPC Service of FUB-IT, Freie Universität Berlin for computing time (DOI: http://dx.doi.org/10.17169/refubium-26754).

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., . . . Kay, K. (2022). A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25(1)*, 116–126.

Doerig, A., Kietzmann, T., Allen, E., Wu, Y., Naselaris, T., Kay, K., & Charest, I. (2022). Semantic scene descriptions as an objective of human vision. *arXiv preprint arXiv:2209.11737*.

Khosla, M., Murty, N., & Kanwisher, N. (2022). A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition. *Current Biology*, *32(19)*, 4159-4171.

Madan, S., Xiao, W., Cao, M., Pfister, H., Livingstone, M., & Kreiman, G. (2024). Benchmarking out-of-distribution generalization capabilities of dnn-based encoding models for the ventral visual cortex. *arXiv preprint arXiv:2406.16935*.

Naselaris, T., Kay, K., Nishimoto, S., & Gallant, J. (2011). Encoding and decoding in fmri. *Neuroimage*, *56(2)*, 400-410.

Prince, J., Alvarez, G., & Konkle, T. (2024). Contrastive learning explains the emergence and function of visual category-selective regions. *Science Advances*, *10(39)*, eadl1776.

Ren, Y., & Bashivan, P. (2024). How well do models of visual cortex generalize to out of distribution samples? *PLOS Computational Biology*, *20(5)*, e1011145.

Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S. (2021). Barlow twins: Self-supervised learning via redundancy reduction. *arXiv preprint arXiv:2103.03230*.