COGNITIVE NEUROSCIENCE

A deeper look at vision and memory

Allen et al. introduce the Natural Scenes Dataset — high-resolution fMRI data from eight individuals scanned as they collectively viewed more than 70,000 natural images and performed a continuous recognition task. This resource promises to yield insights into visual perception and memory and to help bridge cognitive neuroscience and artificial intelligence.

Thomas L. Botch, Caroline E. Robertson and Emily S. Finn

Human neuroscience is now solidly in the age of 'big data'. Over the past decade, researchers have recognized that understanding the brain will require sampling it at extraordinary scales, and have gone to great lengths to collect and share large datasets. In doing so, they must make critical decisions about which dimension(s) of a dataset should be 'big'.

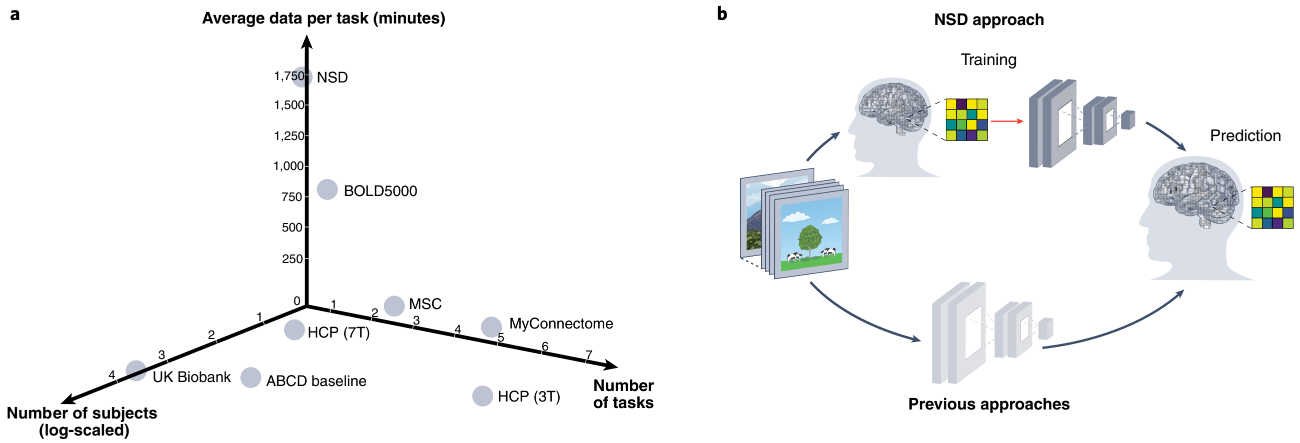

Fig. 1 | The Natural Scenes Dataset. a, The Natural Scenes Dataset (NSD) uses high-resolution fMRI of a single cognitive task to deeply sample neural responses to an unprecedented breadth of stimuli (more than 70,000 natural scene images). Previous big fMRI datasets broadly, but shallowly, sampled across a larger number of cognitive tasks and/or participants (HCP, Human Connectome Project; ABCD, Adolescent Brain Cognitive Development; MSC, Midnight Scan Club). **b**, The massive size of the NSD enabled Allen et al. to create an end-to-end model of neural responses. The authors trained a deep convolutional neural network (CNN) exclusively with fMRI-based neural activity to predict brain responses to unseen stimuli. By contrast, previous studies have relied on pre-trained CNNs to predict brain responses.

Some endeavors have collected limited amounts of data from large numbers of subjects[1–3], whereas others have sampled deeply from a small number of subjects[4,5]. While each approach is best suited to certain scientific questions, one commonality across both types of existing dataset is that functional scan paradigms are typically limited to the resting state and/or a short battery of traditional tasks designed to broadly sample across multiple domains of cognition (for example, working memory, language, emotion).

This feature makes such previous datasets useful for investigating functional brain organization within and across individuals and cognitive states, but less useful for informing computational theories of neural processes within a single cognitive system. For this, one would need dense sampling of a single cognitive system, with an extensive stimulus set and a carefully crafted task paradigm to drive the underlying neural system in all its richness and complexity. This new approach would effectively flip the axes of breadth and depth: whereas previous datasets have gone deep on subjects (either high-*N*-sparsely-sampled or low-*N*-densely-sampled), sometimes with a breadth of paradigms to tap into different cognitive systems, a dataset in this new vein would go deep on a single system, with a breadth of stimuli and a targeted paradigm to more completely probe that particular system (Fig. 1a).

Vision presents an optimal system of choice given its long-standing role as a testbed for computational theories of cognition. In this issue of *Nature Neuroscience*, Allen et al. describe the curation and release of the Natural Scenes Dataset (NSD): a neuroimaging dataset of remarkable scale collected during the presentation of richly annotated natural scenes[6]. Over the course of a year, the authors used high-resolution 7T fMRI to sample the brain responses of eight individuals to more than 70,000 images, resulting in more than 30 h of data per individual. In addition, by designing novel analysis techniques and benchmarking data quality, they remove burdens of quality assurance from other research groups and enable them to focus on developing insights from future analyses. This uniquely rich catalog of neural responses to complex, real-world images represents a key resource for understanding the computational principles and fine-grained functional organization of the human visual cortex.

The vast and diverse set of natural scene stimuli used in the NSD affords an unprecedented opportunity to deepen our understanding of the functional organization of visually responsive cortex. Previous studies have typically used hundreds to thousands of images[7], an order of magnitude fewer than used in the NSD. Such studies have revealed numerous organizing principles of the ventral temporal cortex, including multiple granularities of category-level organization, gradients of animacy and eccentricity, and shared representational spaces across participants[8]. Yet even with stimulus sets of this size, the inherent complexity of natural scenes is such that the number of visual features is still disproportionately larger than the number of stimuli, meaning that vast portions of natural image space remain uncharacterized. The sheer expanse of the NSD stimulus set, along with the computer-vision annotations given for each image, supplies the amount of data needed to accelerate our understanding of how the visual system dissects and represents the full richness of the visual world.

Two key methodological advances accompany the dataset. First, the authors develop a novel approach to maximize the signal-to-noise ratio of fMRI data. Classic fMRI studies present stimuli in rapid succession and derive neural responses by fitting a model time-locked to the onset of each displayed stimulus. However, with such rapid stimulus presentations, neural responses can overlap, affecting the quality of the extracted signal and limiting the number of stimuli that can be presented. To address this, the authors introduce additional steps which more accurately extract and model the neural signal and demonstrate impressive performance gains over existing methods. In and of itself, this innovative method will bolster the quality of event-related fMRI analyses, including re-analysis and interpretation of existing datasets.

A second methodological advance relates to applications in machine learning. In one of the first demonstrations in the fMRI field[9], the authors trained a deep convolutional neural network (CNN) exclusively with brain responses to a set of training stimuli to predict brain responses to held-out stimuli (an end-to-end model; Fig. 1b). Previous fMRI and neurophysiology research has revealed striking correspondences between the processing stages of CNNs and the hierarchical organization of the ventral visual cortex[10]. However, owing to the amount of data required to train CNNs from scratch, prior fMRI studies have used CNNs trained to perform tasks such as object recognition and then transferred learning of the pre-trained image features to predict neural responses, rather than training the CNN directly on brain activity related to those tasks. Here, Allen et al. report that a brain-optimized CNN modestly but consistently outperforms a pre-trained, task-optimized CNN (AlexNet) in predicting early and middle visual pathway responses. This finding opens exciting possibilities for testing computational hypotheses of visual processing using models trained directly on fMRI data.

This valuable resource promises many future empirical insights. First and foremost, because the NSD involves dense sampling of high-resolution neural activity in a small group of carefully selected individuals, it offers a unique lens on the functional organization of visually responsive cortex. Fine-grained functional topography can be highly idiosyncratic, even when local neuroanatomical features are shared across brains[11]. As a result, the organizing principles of cortex, including boundaries between discrete functional areas or networks and gradients of distributed representations, can often be recovered only through careful single-subject analyses[5,11,12]. Recent single-subject work has begun to probe cortical organization with targeted task manipulations, uncovering separable, interdigitated networks that subserve distinct cognitive processes[13,14]. The massive amount of high-quality, within-subject data contained within the NSD raises the exciting possibility of discovering new organizing principles of vision and memory. The detailed auxiliary scans collected for each subject — population receptive field maps, functional localizers, resting state, diffusion, and several structural contrasts — enable future studies using the NSD to precisely situate each subject's neural responses in the context of their own, idiosyncratic functional organization.

The NSD is also poised to provide key insights into the neural mechanisms of visual memory. Participants performed a continuous image recognition task across the 30–40 scan sessions, representing the longest visual memory fMRI study to date and an exciting opportunity to study the transformation from novel (perceptual) to familiar (mnemonic) representations in the brain. Earlier research has reported a posterior–anterior gradient of perceptual-to-mnemonic representations in the ventral temporal cortex[15], specialized category-selective memory areas[14], and representational transformations between perceptual and remembered stimuli[16]. However, little is known about how these gradients, areas, or representational transformations emerge over time as novel stimuli become familiar, an exciting potential offering of the NSD. In future versions of NSD-related data, the authors will release an fMRI visual imagery dataset along with a behavioral dataset containing representational similarity judgments for the natural scene stimuli. These current and future resources will catalyze a detailed understanding of how perceptual–mnemonic representations change over time in the brain.

Overall, the NSD is a valuable contribution to multiple fields and subfields of cognitive neuroscience and artificial intelligence. With the curation and dissemination of this exceptional dataset, the authors continue the admirable data-sharing tradition in the human neuroimaging community and create an exciting platform for new discoveries. ❐

Thomas L. Botch, Caroline E. Robertson and Emily S. Finn ✉

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, USA.
✉*e-mail: emily.s.finn@dartmouth.edu*

Published online: 16 December 2021
https://doi.org/10.1038/s41593-021-00966-7

References

1. Van Essen, D. C. et al. *Neuroimage* **80**, 62–79 (2013).
2. Miller, K. L. et al. *Nat. Neurosci.* **19**, 1523–1536 (2016).
3. Casey, B. J. et al. *Dev. Cogn. Neurosci.* **32**, 43–54 (2018).
4. Poldrack, R. A. et al. *Nat. Commun.* **6**, 8885 (2015).
5. Gordon, E. M. et al. *Neuron* **95**, 791–807.e7 (2017).
6. Allen, E. J. et al. *Nat. Neurosci.* https://doi.org/10.1038/s41593-021-00962-x (2021).
7. Chang, N. et al. *Sci. Data* **6**, 49 (2019).
8. Grill-Spector, K. & Weiner, K. S. *Nat. Rev. Neurosci.* **15**, 536–548 (2014).
9. Seeliger, K. et al. *PLOS Comput. Biol.* https://doi.org/10.1371/journal.pcbi.1008558 (2021).
10. Yamins, D. L. K. & DiCarlo, J. J. *Nat. Neurosci.* **19**, 356–365 (2016).
11. Kanwisher, N. *Proc. Natl Acad. Sci. USA* **107**, 11163–11170 (2010).
12. Braga, R. M. & Buckner, R. L. *Neuron* **95**, 457–471.e5 (2017).
13. DiNicola, L. M., Braga, R. M. & Buckner, R. L. *J. Neurophysiol.* **123**, 1144–1179 (2020).
14. Steel, A., Billings, M. M., Silson, E. H. & Robertson, C. E. *Nat. Commun.* **12**, 2632 (2021).
15. Peer, M., Ron, Y., Monsa, R. & Arzy, S. *eLife* **8**, e47492 (2019).
16. Bainbridge, W. A., Hall, E. H. & Baker, C. I. *Cereb. Cortex* **31**, 1898–1913 (2021).

Competing interests

The authors declare no competing interests.