nature portfolio

Corresponding author(s): Yu Zhang

Last updated by author(s): Dec 29, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
□	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
□	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
□	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	□	A description of all covariates tested
□	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
□	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
□	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	□	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	□	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
□	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Psychophysics Toolbox 3.0.14, MATLAB R2018a, Meadows web-based platform (http://meadows-research.com).
Data analysis	Standard preprocessing steps were performed using fmriprep pipeline v24.0.0 (Esteban et al., 2019), including slice timing correction, head motion correction, co-registration of the functional images to the participant's T1w image, and nuisance regression of motion parameters, white matter (WM), and cerebrospinal fluid (CSF) signals. The functional images were subsequently normalized to the ICBM152 template by combining the linear functional-to-structural transformation with the nonlinear warping from individual structural space to the MNI space. We estimated single-trial brain responses to each scene image using the GLMsingle model (Prince et al., 2022). The resulting voxel-wise beta maps were projected onto the fsaverage cortical surface template (Fischl et al., 2004), resulting in a vertex-wise brain response map (163,842 vertices per hemisphere) for each scene image. These vertex-wise brain maps served as the target variable for all subsequent encoding model analyses.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The Natural Scenes Dataset(NSD) and COCO datasets are public and accessible to all researchers. The preprocessed 7T fMRI dataset from NSD can be accessed via https://naturalscenesdataset.org/. The natural scenes and the corresponding object categories and text captions from COCO can be downloaded from https://cocodataset.org/dataset/home.htm. All unimodal and multimodal encoding models and analysis code will be made avaliable upon request to ensure reproducibility.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender

The NSD dataset consists of eight healthy adult participants (6 females and 2 males; Age: 19-34). The four participants used in this study were reported as biological females.

Reporting on race, ethnicity, or other socially relevant groupings

The primary documentation for the Natural Scenes Dataset (NSD) does not explicitly report the race or ethnicity of individual participants.

Population characteristics

The NSD dataset consists of eight healthy adult participants (6 females and 2 males; Age: 19-34), each viewing 10,000 natural scenes over a one-year period.

Recruitment

Participants were recruited from the University of Minnesota and the surrounding Twin Cities community through public advertisements. The recruitment process was designed to identify individuals capable of participating in a demanding, long-term longitudinal neuroimaging study.

Ethics oversight

The data used in this study were obtained from the Natural Scenes Dataset (NSD). The original experimental protocol and all data collection procedures were reviewed and approved by the University of Minnesota Institutional Review Board (IRB).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The NSD dataset consists of eight healthy adult participants （6 females and 2 males; Age: 19-34), each viewing 10,000 natural scenes over a one-year period.

Data exclusions

For the present analysis, we focused exclusively on a subset of the NSD dataset, consisting of four participants (subj01, subj02, subj05, and subj07) who completed the full protocol of all 40 sessions, providing a total of 10,000 unique image trials per subject. These four subjects were most commonly used in previous studies, allowing for direct comparison with previous state-of-the-art results.

Replication

We conducted a systematic replication of both unimodal and multimodal encoding models for the four participants (subj01, subj02, subj05, and subj07). Each participant viewed 9,000 unique scene images that were used to train individualized, vertex-wise encoding models. Besides, a fixed set of 1,000 Shared scene images, viewed by all four participants across multiple repetitions, served as the independent test set for evaluating model performance and ensuring cross-subject consistency.

Randomization

For each subject, 9,000 unique images were randomly selected from the Microsoft COCO database for model training, while a fixed set of 1,000 images was reserved for cross-subject testing. Despite the difference in specific image instances, the massive scale of sampling (10,000 trials per subject) ensures that the high-dimensional space of natural scene features is comprehensively and consistently represented across all participants. This allows for the training of robust, subject-specific encoding models that are statistically comparable.

Blinding

The 1,000 shared test images and their corresponding neural responses were fully sequestered during the model training phase. All

Blinding: hyperparameter tuning and vertex-wise weight estimations were performed exclusively on the 9,000 unique training trials. A strict data masking protocol was maintained for all encoding models and all participants.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☒	☐	Clinical data
☒	☐	Dual use research of concern
☒	☐	Plants

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☐	☒	MRI-based neuroimaging

Plants

Seed stocks: *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes: *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication: *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

Magnetic resonance imaging

Experimental design

Design type: The core NSD experiment is task-based and has an event-related design. The prf experiment is task-based and has a continuous design. The floc experiment is task-based and has an event-related design.

Design specifications: In the NSD experiment, images were presented for 3 seconds, and were followed by a minimum of 1 second of gap before the next trial. Many thousands of distinct images were presented over the course of many distinct scan sessions, with a maximum number of presentations per distinct image of 3.

Behavioral performance measures: Button presses and associated reaction times for each trial in the NSD experiment were recorded.

Acquisition

Imaging type(s): Functional, structural, diffusion

Field strength: 7T

Sequence & imaging parameters: The primary fMRI sequence involved gradient-echo EPI, FOV 216 mm x 216 mm, matrix size 120 x 120, slice thickness 1.8 mm, orientation axial, TR 1.6 s, TE 22.0 ms, and flip angle 62°.

Area of acquisition: Whole-brain scans

Diffusion MRI ☒ Used ☐ Not used

Parameters: 99-100 directions; b-values of 0, 1,500, and 3,000; no cardiac gating

Preprocessing

Preprocessing software: Custom Python codes, FreeSurfer, FSL, and ANTs. The resulting data were projected onto the fsaverage surface, providing the vertex-wise representations used for our subject-specific encoding models.

Normalization

he NSD data were prepared in a variety of spaces including subject-native space and atlas spaces (MNI, fsaverage).

Normalization template

MNI152 and fsaverage templates

Noise and artifact removal

The data-driven analysis method GLMdenoise and the statistical technique of ridge regression were used. These methods can account for a variety of sources of noise (e.g., physiological, motion, scanner artifacts, effects of collinearity). A version of the GLM results that omit these noise removal methods is also provided.

Volume censoring

No censoring

Statistical modeling & inference

Model type and settings

Single-trial beta-values, representing the fMRI response to each scene image, were estimated using the GLMsingle model (Prince et al., 2022). The resulting voxel-wise beta maps were projected onto the fsaverage cortical surface template (Fischl et al., 2004), resulting in a vertex-wise brain response map (163,842 vertices per hemisphere) for each scene image. These vertex-wise brain maps served as the target variable for all subsequent encoding model analyses.

Effect(s) tested

Pearson's correlation

Specify type of analysis: □ Whole brain □ ROI-based ☒ Both

Anatomical location(s)

Subject-specific encoding models were trained separately for each participant and each vertex, using their unique 9,000 training images. We quantified the each encoding model by computing Pearson's correlation coefficients (r) between the predicted and observed fMRI responses across 1,000 test scene images.
We localized cortical visual areas with the Kastner atlas (Wang et al., 2015b), spanning regions primary regions (V1-V3), extrastriate cortex (V3A/B, V4), ventral occipital areas (VO1/2), parahippocampal areas (PHC1/2), lateral occipital areas(LO1/2), temporal occipital areas (TO1/2, encompassing hMT+), intraparietal sulcus areas (IPS0-5), frontal eye field (FEF) and supplementary eye fields (SEF), registered to the fsaverage surface and thresholded at 25%.

Statistic type for inference

(See Eklund et al. 2016)

For statistical inference, we utilized Pearson correlation (r) between predicted and observed neural responses on the held-out test set. All p-values were corrected for multiple comparisons across vertices using False Discovery Rate (FDR).

Correction

To correct for multiple comparisons across the large number of cortical vertices (163,842 vertices in the fsaverage surface), we applied a False Discovery Rate (FDR) correction to the prediction accuracy maps.

Models & analysis

n/a	Involved in the study
☒	□ Functional and/or effective connectivity
☒	□ Graph analysis
□	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis

For each cortical vertex, we quantified the prediction accuracy of encoding models using the Pearson correlation coefficient (r) between predicted and observed brain responses across the held-out test set of 1,000 shared scene images. To correct for multiple comparisons across the large number of cortical vertices, we applied a False Discovery Rate (FDR) correction to the prediction accuracy maps. Only vertices with FDR corrected, p<0.01 were retained as significant predictions in the encoding models.
Our vertex-wise encoding framework employed three distinct families of deep learning architectures, including Vision, Language, and Multimodal fusion models. For vision-based models, we utilized two prominent architectures, ResNet and Vision Transformer (ViT), to extract latent features of natural scenes across different scales and levels of abstraction. The language-based models, Multi-hot Encoding (MultiHot) and Bidirectional Encoder Representations from Transformers (BERT), captured semantic and linguistic representations of the scene images. Finally, a dedicated multimodal fusion approach was implemented via the Vision-and-Language Transformer (ViLT) architecture.