WHICH STIMULI TO SCAN? SMALL IMAGE-SET SELECTION FOR IMAGE-TO-FMRI ENCODER TRAINING

Anonymous authors
Paper under double-blind review

ABSTRACT

Functional MRI (fMRI) offers a powerful way to study human visual perception. Recently, Image-to-fMRI encoders trained on hundreds of participant-specific image–fMRI pairs were shown to accurately predict the participant's fMRI response to new images. However, collecting such extensive training data is often infeasible due to high scanning costs and the limited availability of participants. Achieving higher accuracy with smaller training sets would substantially broaden the applicability of these encoding models, facilitating both large-scale studies of population variability and studies of how visual responses change with experience or intervention. This raises a fundamental question: ***If we can measure fMRI responses to only a small number of images, which images should we choose to scan?*** We show that image-sets selected to provide high coverage of the brain-response space of one subject (with many fMRI scans) generalize well to new subjects and significantly improve their image-to-fMRI encoding in the low-data regime (tens of scanned images). We explore image-set selection both in brain-response space and image representation space, and show that both outperform random image sampling, with brain-response-based selection providing particularly strong gains. Moreover, image-sets selected using NSD neural representations transfer well to other fMRI datasets (e.g., LAION-fMRI, 3T Generic Object Decoding dataset). These results provide a step toward more data-efficient stimulus acquisition for fMRI experiments.

1 INTRODUCTION

Understanding how the human brain represents visual information, and how these representations vary across individuals, is a central goal of visual neuroscience. Characterizing this variation can shed light on perceptual experience, behavioral traits, and clinical conditions (Charest et al., 2014; Karvelis et al., 2018; Mennen et al., 2019), motivating studies that extend across larger populations (Marek et al., 2022). Adaptive experiments and neurofeedback create a complementary need to rapidly characterize neural representations as they change in response to stimuli, feedback, or intervention (Lorenz et al., 2016; Sitaram et al., 2017; Debettencourt et al., 2015).

Image-to-fMRI encoding models are a powerful tool for characterizing variation in visual representations across and within individuals, as they can be trained to predict subject-specific fMRI responses to new images with high accuracy (Kay et al., 2008; Beliy et al., 2019; Adeli et al., 2025). Recent advances in *cross-subject* encoding further enable knowledge learned from extensively scanned participants to be transferred to a new individual, reducing the required number of participant-specific training pairs to the low hundreds (Wen et al., 2018; Gu et al., 2022; Beliy et al., 2024). However, collecting this amount of training data can still be infeasible in many circumstances due to scan costs and the limited availability of participants, limiting the uptake of these methods.

We therefore ask whether further improvements in training efficiency can be achieved by optimizing *which images* are selected to be scanned for a new participant, given a fixed budget of N images. Subset selection is a long-standing problem in machine learning (Sener & Savarese, 2017; Hacohen et al., 2022; Schreiber et al., 2020), and recent work in visual fMRI has begun to study stimulus-set design for large-scale acquisition, emphasizing broad coverage of visual-semantic image space (Roth & Hebart, 2025; Zerbe et al., 2026). Our focus is the practically important small-set regime, where the goal is to choose the few images to scan on a new participant that best support learning an accurate

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

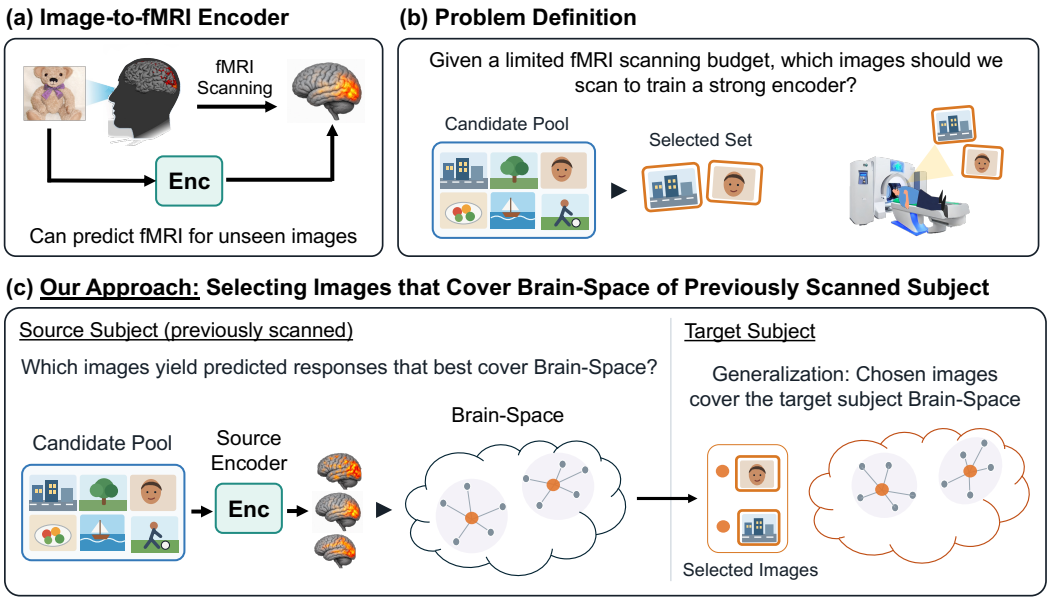


Figure 1: **Problem definition and approach.** **(a)** Image-to-fMRI encoders learn from measured image–fMRI pairs and can predict responses to unseen images. **(b)** Given a scanning budget of N images for a new subject, we ask which images from a candidate pool should be scanned to train a strong encoder. **(c)** Our approach uses an encoder trained on previously scanned source subjects to predict responses to candidate images and selects a subset whose responses broadly cover Brain-Space. This coverage transfers to a new target subject: the selected images also cover the target subject's Brain-Space, without using any target-subject fMRI during selection, providing an informative set for efficient encoder adaptation.

Image-to-fMRI encoder for that participant. Rather than defining coverage solely in image space (by selecting a set of images that broadly spans the representation space formed by many images), we instead define coverage in neural-response space. Our central hypothesis is that, because visual representations share structure across individuals, images that broadly cover this space in previously scanned participants (who were scanned on many images) can also provide an informative calibration set for a new participant, prior to any fMRI scanning on the new participant.

We build on a recent multi-subject image-to-fMRI encoder that enables data-efficient transfer to new participants (Beliy et al., 2024), and investigate whether our proposed images-to-scan selection improves image-to-fMRI prediction over random image sampling. We compare several variants of our image-set selection algorithm, exploring diversity- and coverage-based objectives in visual-feature spaces (CLIP and DINO), with our proposed method of using coverage in brain space (fMRI). Brain space is derived from previously measured participants, allowing selection without using any target-participant fMRI scans. Evaluating on all 8 subjects of the NSD dataset (Allen et al., 2022), brain-space coverage performs best, showing significant performance gains in the low-data regime.

Moreover, the usefulness of the selected images transfers beyond the participants and dataset used to define the neural space. Using an NSD-trained encoder to guide image selection also improves prediction for new participants in other fMRI datasets, including LAION-fMRI (Zerbe et al., 2026) and the 3T GOD dataset (Horikawa & Kamitani, 2017).

Our main contributions are as follows. We pose the question of which few images to scan for a new participant, treating stimulus selection itself as an optimization problem. This enables data-efficient characterization of individual visual representations where scan time is scarce: large cohorts, clinical populations, and adaptive or neurofeedback experiments. Rather than maximizing visual diversity, we select images that best cover the neural responses of previously scanned participants, and show that they also best cover those of a new participant, whose responses are unseen during selection. Shared neural structure thus guides not only model transfer, but also which data to acquire. In the low-data regime, from only a few selected scans, we predict a new participant's responses to unseen images as accurately as random sampling does with nearly twice as many images. Finally, selection

guided by one dataset generalizes to participants in other datasets, across stimulus distributions, acquisition settings, and fMRI encoders. This establishes a new use for existing large-scale fMRI datasets: leveraging already-collected data to design data-efficient acquisition for new studies.

2 Related Work

Experimental design and stimulus sampling in visual fMRI. The choice and design of visual stimuli have long shaped what can be learned from fMRI. Structured stimuli established retinotopic and receptive-field organization in visual cortex (Sereno et al., 1995; Dumoulin & Wandell, 2008), while category contrasts revealed selective regions for high-level visual categories (Malach et al., 1995; Kanwisher et al., 1997; Epstein & Kanwisher, 1998; Downing et al., 2001). In parallel, advances in experimental design and single-trial response estimation improved the information obtainable from limited scan time (Dale, 1999; Prince et al., 2022). More recently, image datasets such as GOD, NSD, and LAION-fMRI have expanded stimulus diversity to thousands of images, but remain sampled in only a few individuals (Horikawa & Kamitani, 2017; Allen et al., 2022; Zerbe et al., 2026).

The importance of small training sets in visual fMRI. Understanding individual variation in visual processing can reveal links to perception, behavioral traits, and mental health (Charest et al., 2014; Karvelis et al., 2018; Mennen et al., 2019). Reducing the measurements required per participant could enable larger cohorts, supporting population-level studies of these relationships (Marek et al., 2022). A similar need for sample efficiency arises in adaptive experiments and neurofeedback, which use real-time measurements to study or influence changing perceptual and attentional states (Lorenz et al., 2016; Sitaram et al., 2017; Debettencourt et al., 2015). In these settings, a participant's current responses and state must be inferred from short measurement windows and correspondingly few samples, because longer windows would combine observations across potentially distinct states. Also, in neurofeedback studies, reducing the number of stimulus presentations required to fit the encoding model leaves more time for the participant to engage in the learning or intervention phase.

Image-to-fMRI encoding. Image-to-fMRI encoding models predict responses to unseen visual stimuli, enabling the study of brain representations beyond the images measured in a scanner (Kay et al., 2008; Eickenberg et al., 2017; Adeli et al., 2025). Although functional organization varies across individuals (Feilong et al., 2018; Wasserman et al., 2024), visual representations also share substantial structure across participants (Haxby et al., 2011). Multi-subject encoding methods have consequently used data from multiple individuals to improve subject-specific prediction (Wen et al., 2018; Khosla et al., 2020; Beliy et al., 2024). Recent approaches show that pretrained encoders can be adapted to new participants using substantially fewer target-specific measurements (Gu et al., 2022; Beliy et al., 2024). Nevertheless, prediction becomes increasingly challenging as the available target data become very small, making further improvements in this regime particularly valuable.

Stimulus-set selection. Outside neuroscience, representative subset selection has been widely studied through core-set, clustering-based, and submodular approaches (Sener & Savarese, 2017; Hacohen et al., 2022; Schreiber et al., 2020). In visual neuroscience, Roth & Hebart (2025) focus on constructing broadly representative stimulus sets that capture visual-semantic diversity in CLIP image space and support generalizable inferences, comparing several sampling strategies. In this dataset-scale setting, they find that the diversity of the underlying stimulus pool is substantially more important for generalization than the particular sampling strategy. The LAION-fMRI dataset (Zerbe et al., 2026) follows this broad-coverage perspective, selecting natural images by optimizing effective dimensionality in CLIP space. In contrast, our focus is the small-set regime, where we select stimuli to cover neural-response space rather than image space, aiming for efficient encoder adaptation.

3 Methods

Our framework selects a small stimulus set from a candidate image pool for adapting a multi-subject image-to-fMRI encoder to a new participant. We represent the candidate images in visual-feature or predicted neural-response spaces and compare diversity- and coverage-based selection objectives. Our main approach selects the subset that best covers the neural-response space predicted for other participants, without using measured fMRI from the target participant. After selection, the corresponding target responses are used for transfer learning, adapting an encoder trained only on other participants, and performance is evaluated on unseen images.

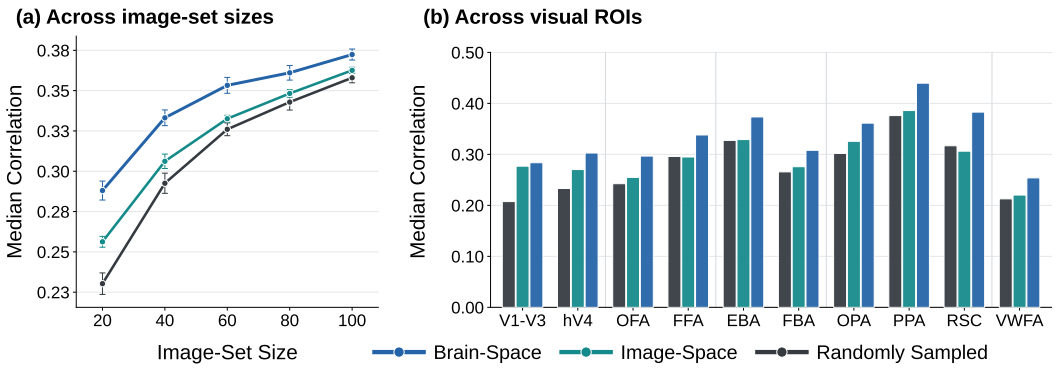


Figure 2: **Stimulus selection across image-set sizes and visual ROIs on NSD.** Image-Space uses k-means in CLIP full-token space, and Brain-Space uses linear-NCSNR-weighted coverage. (a) Median voxelwise Pearson correlation on the fixed test set, averaged equally across eight participants and nine candidate pools. Error bars show descriptive ± 2 Cousineau–Morey-corrected standard errors across participants. (b) At $N = 20$, the same Brain-Space selected image sets are evaluated within visual ROIs; bars show the participant-averaged median voxelwise correlations.

3.1 Problem Setup and Target-Participant Adaptation

Given a candidate image pool $\mathcal{C}$ and a measurement budget N, we aim to select a subset $S \subseteq \mathcal{C}$, with $|S| = N$, that supports training an accurate image-to-fMRI model for a new target subject. While the method can operate on any candidate image pool, our experiments use candidate images that already have measured fMRI responses for the target subject, allowing alternative selected sets to be evaluated without new scanning. These target responses are not used for encoder pretraining or image selection; selection uses only features from pretrained visual models or fMRI predictions from encoders trained on other subjects. Once S is chosen, we use the selected image–fMRI pairs $\{(x, \mathrm{fMRI}(x)) \mid x \in S\}$ to adapt the multi-subject encoder of Beliy et al. (2024), pretrained only on other subjects. Here, $\mathrm{fMRI}(x)$ denotes the target subject's measured fMRI response to image x. All selection methods use the same transfer-learning procedure and are evaluated on a fixed test set disjoint from the candidate pool. Encoder and training details are provided in Appendix C.1.

3.2 Image Selection Representation Spaces

For stimulus-set selection, we compare two types of representations. **(i) Visual representations.** We extract features for each candidate image using pretrained CLIP (Radford et al., 2021) and DINOv2 (Oquab et al., 2023) image encoders. To construct the image's representation, we use either CLS representations, which summarize each image with a single token, or full patch-token representations, which retain features from individual image patches. **(ii) Brain representations.** Since target-subject fMRI is unavailable during selection, we use a multi-subject image-to-fMRI encoder to predict each candidate image's fMRI responses for other individuals, termed *source subjects*. The target subject is excluded from both encoder pretraining and the source subjects. We standardize the predictions, weight them by source-voxel reliability using normalized noise-ceiling signal-to-noise ratio (NCSNR; see Appendix A.5), and concatenate them across source subjects to construct a brain representation for each image. This approach tests whether shared functional structure allows image sets selected in source-subject fMRI space to generalize to the target subject. Details of both visual-feature and predicted-fMRI representations are provided in Appendix C.2.

3.3 Stimulus-Set Selection Objectives

Let $\mathcal{C}$ denote the candidate image pool. Given an image representation, we compare several criteria for selecting a subset $S \subseteq \mathcal{C}$ of size N. These objectives capture complementary notions of a useful stimulus set: representing the full candidate population, spanning diverse directions of variation, or identifying representative clusters. We apply the same objectives in both visual and brain spaces whenever applicable, allowing us to separate the effect of the representation from that of the selection rule. Uniform random sampling serves as a representation-independent baseline.

Table 1: **Full-brain NSD stimulus-selection results.** Comparison of how the representation space and selection objective affect downstream encoding performance across image budgets. We compare Random Sampling with selection in two representation spaces: (i) Image-Space and (ii) Brain-Space. Within each space, we compare three selection objectives: Effective-Dimensionality (ED), k-means, and Coverage. We report median voxelwise Pearson correlation for different image-set sizes, averaged across nine candidate pools within each participant and across eight NSD participants.

Representation Space	Selection Objective	Image-Set Size				
		20	40	60	80	100
–	Random	0.230	0.293	0.326	0.343	0.358
Image-Space (CLIP Full)	ED	0.215	0.271	0.307	0.326	0.343
	k-means	0.256	0.306	0.333	0.348	0.363
	Coverage	0.247	0.307	0.333	0.345	0.361
Brain-Space (fMRI)	ED	0.177	0.217	0.250	0.277	0.297
	k-means	0.278	0.323	0.346	0.353	0.369
	Coverage	**0.288**	**0.333**	**0.353**	**0.361**	**0.372**

Coverage. Our primary objective is to select a small set that represents the candidate population (Schreiber et al., 2020). Intuitively, each candidate image is considered well covered if it is similar to at least one selected image, and the objective favors sets that provide high coverage across the pool. Let $s(x_i, x_j)$ denote the nonnegative cosine similarity between the representations of two candidate images. For a selected set S, we define the facility-location score $F(S) = \sum_{x_i \in \mathcal{C}} \max_{x_j \in S} s(x_i, x_j)$. This measures how well every candidate image is represented by its most similar selected image. We maximize this facility-location objective greedily, adding at each step the image that yields the largest increase in $F(S)$, until N images are selected. This favors broad population coverage while avoiding redundant additions in already well-covered regions of the representation space.

Alternative selection objectives. Following recent work on stimulus-set design in visual fMRI (Roth & Hebart, 2025), we additionally evaluate k-means, k-center/Core-Set, and effective-dimensionality (ED) maximization as complementary selection criteria. K-means selects representative images from clusters in the representation space, while k-center emphasizes diversity by iteratively expanding coverage toward poorly represented regions. ED instead favors sets whose variation is distributed across a larger number of representation dimensions. Together, these provide clustering-, diversity-, and dimensionality-based alternatives to our population-coverage objective. Exact implementation details are provided in Appendix C.3.

4 Experiments

4.1 Experimental setup

Datasets. We use three image-fMRI datasets: NSD (8 subjects) (Allen et al., 2022), LAION-fMRI (5 subjects) (Zerbe et al., 2026), and the 3T GOD dataset (5 subjects) (Horikawa & Kamitani, 2017). Dataset and preprocessing details are provided in Appendix C.4.

Experimental protocol and compared methods. In the within-NSD setting, each subject serves in turn as the target subject for image selection and encoder adaptation. The encoder is pretrained on the other seven subjects, and Brain-Space representations are constructed from predicted responses for non-target NSD source subjects. In the cross-dataset setting, LAION-fMRI and GOD subjects serve as targets, while encoder pretraining and source-subject information come from NSD. In both settings, measured target-subject responses are withheld during image selection and used only afterward to adapt the encoder using the selected image–fMRI pairs. Our main method, *Brain-Space*, applies linear-NCSNR-weighted coverage in predicted fMRI space. We compare it with Random Sampling and k-means selection in CLIP full patch-token space (*Image-Space*), alongside the alternative representations and objectives described in Section 3.3. All methods use the same adaptation procedure (Appendix C.1).

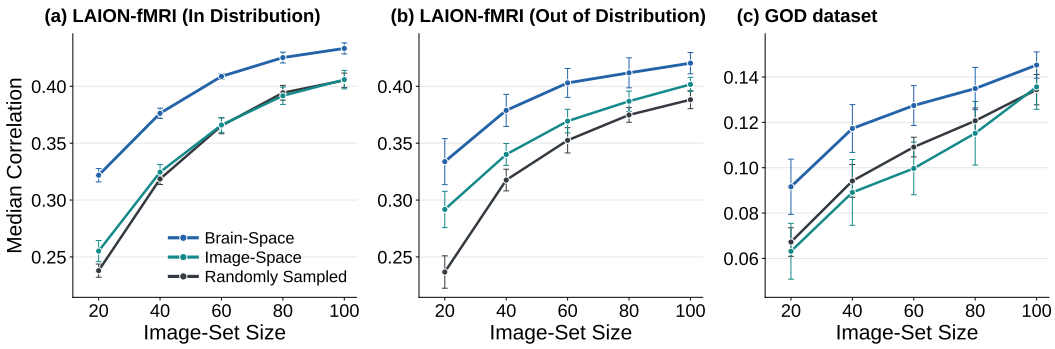


Figure 3: **NSD-derived Brain-Space selection generalizes across datasets.** Median voxelwise Pearson correlation across image-set sizes for transfer to LAION-fMRI, evaluated on both the (a) ID and (b) OOD test sets, and to the (c) GOD dataset. In all experiments, Brain-Space stimulus sets are selected using responses predicted by an encoder trained on NSD, testing whether NSD-derived predicted neural representations can guide stimulus selection in new datasets. Results are averaged over nine candidate pools within each participant and then equally across participants; error bars show descriptive ± 2 Cousineau–Morey-corrected standard errors.

Image budgets, candidate pools, and evaluation. We use budgets of $N \in \{20, 40, 60, 80, 100\}$ distinct training images, with fMRI responses averaged over available presentations of each image (Appendix C.4). To assess robustness to the candidate pool while retaining a large set of images to choose from, we evaluate nine pools per subject, each containing 80% of the training images and shared across methods. Performance is the median voxelwise Pearson correlation on a fixed test set, averaged across pools within each subject and then equally across subjects. For LAION-fMRI, we report the official in-distribution (ID) and out-of-distribution (OOD) test sets separately. Supplementary evaluations include MSE (Appendix D.1).

4.2 Selecting which stimuli to scan improves encoding performance

Optimized image sets outperform random sampling. We first examine whether selecting the training images improves encoding for a new NSD subject under a fixed image budget. Figure 2.a shows that both Image-Space and Brain-Space selection achieve higher average performance than Random sampling at every evaluated budget. At $N = 20$, participant-averaged median correlation increases from 0.230 with Random sampling to 0.256 with Image-Space selection and 0.288 with Brain-Space selection, corresponding to relative improvements of approximately 11% and 25%, respectively. The gains are largest at the smallest budgets, demonstrating the value of choosing informative training images when target-subject measurements are scarce.

Brain-Space selection outperforms Image-Space selection. Brain-Space selection achieves higher average performance than the Image-Space baseline throughout the evaluated budget range (Table 1). For example, Brain-Space selection with $N = 40$ reaches a correlation of 0.333, approximately matching Image-Space performance with $N = 60$, despite using one-third fewer training images. The advantage also holds when the selection objective is controlled: both k-means and coverage achieve higher average performance in Brain-Space than in the corresponding CLIP representations. The primary Brain-Space method improves over both baselines for all eight subjects at every budget (Appendix D.2), with paired subject-level tests remaining significant after multiple-comparison correction ($p_{\max T} = 0.008$ for each comparison and budget; Table T17). A supplementary MSE analysis also shows lower prediction error than both baselines across the evaluated budgets, indicating that the improvement is not specific to Pearson correlation (Appendix D.1).

The selection objective matters. A neural-response representation alone does not ensure effective stimulus selection. Within Brain-Space, coverage achieves the highest average performance among the evaluated objectives at every budget (Table 1). At $N = 20$, average correlation is 0.288 for coverage, compared with 0.278 for k-means and 0.177 for effective-dimensionality (ED) selection (see Appendix A.1 for an extended comparison).

Encoding improvements extend across visual regions. Figure 2.b evaluates the same selected image sets within visual ROIs at $N = 20$. Brain-Space selection improves performance across all

Table 2: **Generalization across selection encoders.** Brain-Space image sets are constructed from responses predicted by either the Universal Encoder or the trained encoder of Adeli et al. (2025) (Brain-Space*). The selection objective and downstream Universal Encoder adaptation and evaluation remain fixed.

Selection method	20	40	60	80	100
Random	.232	.294	.327	.346	.361
Image-Space	.259	.310	.335	.351	.364
Brain-Space	.288	.333	.353	.361	.372
Brain-Space*	.282	.326	.349	.357	.369

Table 3: **Generalization to an independent downstream encoder.** We use the Random, Image-Space, and Brain-Space image sets from the main NSD experiment, with Brain-Space selection based on the Universal Encoder. The selected sets are then used to train another encoder of Adeli et al. (2025), which does not have an efficient cross-subject transfer procedure.

Selection method	20	40	60	80	100
Random	.130	.154	.164	.175	.184
Image-Space	.142	.166	.172	.183	.187
Brain-Space	**.177**	**.192**	**.196**	**.201**	**.207**

displayed ROIs, but its advantage over Image-Space varies across regions. In early visual cortex (V1–V3), Image-Space and Brain-Space selection perform similarly, both substantially outperforming Random sampling. In contrast, Image-Space selection provides only modest or negligible gains over Random in face-selective (OFA and FFA) and body-selective (EBA and FBA) regions, whereas Brain-Space yields substantially larger improvements. Brain-Space also benefits scene-selective regions (OPA, PPA, and RSC), with particularly pronounced gains over Image-Space in PPA and RSC. Additional ROI-level results at $N = 60$ and $N = 100$ are provided in Appendix D.3 (Figure S4).

4.3 Generalization across datasets and encoders

NSD-derived selection transfers to LAION-fMRI. We test whether neural representations learned from NSD can guide stimulus selection for new subjects in LAION-fMRI. Candidate images are drawn from those presented in the LAION-fMRI training split and represented using responses predicted for NSD source subjects, without using measured LAION-fMRI responses during selection. The selected target-subject image–fMRI pairs are then used for encoder transfer learning. Brain-Space selection achieves higher average performance than Random sampling and Image-Space selection on both the in-distribution (ID) and out-of-distribution (OOD) test sets at every evaluated budget (Figure 3.a,b). At $N = 20$, the relative improvements over Random are approximately 35% on ID and 41% on OOD. Improvements over Random hold for every subject (Appendix D.4).

The benefit extends to GOD under different acquisition conditions. We further evaluate NSD-derived selection on the 3T GOD dataset, testing transfer across acquisition settings and to object categories absent from the target training set. Brain-Space selection achieves the highest average performance at every evaluated budget (Figure 3.c). At $N = 20$, average performance improves by approximately 37% over Random sampling and 46% over Image-Space selection. In contrast, Image-Space selection does not consistently outperform Random across budgets. The average-performance improvement of Brain-Space over Random holds for all five subjects (Appendix D.5).

Brain-Space selection generalizes across selection encoders. To test whether the benefit depends on the encoder used to construct Brain-Space, we replace the Universal Encoder with the transformer-based encoders of Adeli et al. (2025), which differ in architecture and are trained independently for each subject, keeping the selection objective and downstream adaptation fixed. The resulting selection closely tracks that of the Universal Encoder (0.282 vs. 0.288 at $N = 20$) and outperforms Random sampling and Image-Space selection at every budget (Table 2; Appendix D.6). Thus, the benefit of Brain-Space selection is not specific to a particular selection encoder.

Selected images also benefit a different downstream encoder. We next test whether the advantage also persists when changing the encoder trained on the selected images. We use the same NSD image sets to train the subject-specific fMRI encoder of Adeli et al. (2025). Brain-Space selection remains based on the Universal Encoder, while the downstream encoder is trained from scratch on each selected set, as it does not have an efficient cross-subject transfer procedure. Table 3 shows that Brain-Space selection retains its advantage over Random sampling and Image-Space selection at every evaluated budget.

Table 4: **Brain-Space selection remains effective with a single presentation per image.** We repeat LAION-fMRI adaptation using one measured response per selected training image. Brain-Space selection retains its advantage over Random and Image-Space selection across image-set sizes on both test sets. Results report median Pearson correlation averaged across 3 candidate pools and five participants.

Test	Method	Image-Set Size				
		20	40	60	80	100
ID	Random	.152	.217	.258	.285	.300
	Image-Space	.152	.210	.257	.285	.305
	Brain-Space	**.229**	**.282**	**.322**	**.340**	**.350**
OOD	Random	.143	.220	.251	.270	.288
	Image-Space	.192	.221	.263	.297	.310
	Brain-Space	**.259**	**.291**	**.326**	**.331**	**.340**

Table 5: **More distinct images versus repeated measurements.** We compare two ways of allocating 80 LAION-fMRI training presentations: 20 images presented four times and averaged (20×4), or 80 images presented once (80×1). Across methods, using more distinct images provides similar or higher encoding performance.

Test	Method	80 presentations	
		20×4	80×1
ID	Random	.247	.285
	Image-Space	.243	.285
	Brain-Space	**.319**	**.340**
OOD	Random	.235	.270
	Image-Space	.285	.297
	Brain-Space	**.330**	**.331**

4.4 Stimulus selection and repeated presentations

Our main LAION-fMRI experiments average four presentations for each training image, so an image budget of N corresponds to $4N$ measured training presentations. This raises two practical questions: whether the benefit of stimulus selection persists when adaptation uses only a single response per image, and, for a fixed presentation budget, whether it is preferable to measure more distinct images or obtain more reliable responses by repeating fewer images.

The selection advantage persists with single presentations. We repeat LAION-fMRI adaptation using a single presentation per selected image instead of the average of four, keeping the candidate pools, selected image sets, and test responses unchanged. Brain-Space selection outperforms Random and Image-Space selection at every evaluated budget on both test splits (Table 4), with a larger advantage than under four-presentation training, consistent with selection being most valuable when the training signal is limited.

More images versus repeated measurements at a fixed budget. We next compare two ways of allocating an 80-presentation training budget: 20 images presented four times and averaged, or 80 images presented once. Within each selection method, using 80 distinct images yields higher average performance on both test splits (Table 5). For Brain-Space, ID performance increases from 0.3189 to 0.3395, while OOD performance is nearly unchanged (0.3295 vs. 0.3311). At this budget, measuring more distinct images thus matches or exceeds the benefit of repetition.

5 Ablations & Analysis

We next briefly highlight the main additional analyses, with a broader set of ablations and robustness analyses provided in Appendix A. **Representation and selection-objective controls** (Appendix A.1) further support the importance of both the representation and the selection objective. **Scaling to larger image budgets** (Appendix A.2) extends the main analysis beyond $N = 100$, up to $N = 1000$, showing that the advantage of Brain-Space selection is strongest in the low-budget regime and gradually decreases as more target-subject data become available. **Visual-representation controls** (Appendices A.3 and A.4) compare CLIP and DINOv2, as well as visual representations extracted from the fMRI encoder. CLIP and DINOv2 perform similarly overall, while features adapted for response prediction do not consistently improve stimulus selection over standard pretrained visual representations and remain less effective than Brain-Space selection. **Source-voxel reliability weighting** (Appendix A.5). **Analysis of cross-subject transfer** (Appendix A.6). **Sensitivity to the source-subject set** (Appendix A.7) shows that constructing Brain-Space from a single source subject already achieves performance close to using multiple sources. **Robustness to candidate-pool choice** (Appendix A.8) is assessed using worst-observed performance across nine candidate pools. Brain-Space selection shows larger gains over Random and Image-Space selection under this summary statistic than in average performance across the evaluated NSD budgets.

Brain-Space selections are consistent across candidate pools. In Appendix A.9, we examine how Brain-Space selections change across candidate pools. In NSD, the first four images selected from two pools are identical, and many subsequent selections differ in identity but share content, such as outdoor scenes with people, bathroom interiors, and tennis players (Figure S1); similar patterns occur in LAION-fMRI and GOD. Quantitatively, exact-image recurrence across nine pools is higher for Brain-Space than for Image-Space and Random selection at every dataset and budget. At $N = 20$, 31.7–44.4% of Brain-Space selections belong to images chosen in at least five of nine pools, compared with 7.8–13.3% for Image-Space.

Analysis of selected sets across visual and brain spaces. In Appendices A.10 and A.11, we investigate how well the same selected sets cover the candidate population in three spaces: Image-Space, Brain-Space, and Target Brain-Space, and complement this with t-SNE visualizations showing where the selected images lie in these representations. Here, Brain-Space is constructed from predicted responses of the previously scanned source participants, while Target Brain-Space is defined by the measured responses of the target participant. The visualizations suggest that Brain-Space and Image-Space often select different images from similar regions of Image-Space, while Brain-Space selection covers regions in Brain-Space that are sparsely covered by Image-Space selection. Quantitatively, stronger Image-Space coverage does not necessarily translate into stronger coverage of brain responses. Brain-Space selection best covers the Target Brain-Space among the methods. This suggests that selecting images to represent the source participants' Brain-Space also selects images that better represent the range of responses observed in the new participant.

6 Discussion

Rapidly characterizing an individual's visual representations from only a few image–fMRI pairs is practically important. Image-to-fMRI encoders extend what can be learned from fMRI measurements by predicting participant-specific responses to unscanned images, while cross-subject transfer further reduces the amount of data required to fine-tune a pre-trained model on a new participant. This makes the choice of the few images that are actually scanned an important part of the problem. Our work treats this as an experimental-design problem, optimizing the small image set specifically to support efficient encoder transfer.

The improvement over Random sampling establishes the value of optimized stimulus selection, but the additional advantage over Image-Space is also significant. We hypothesize that this gap reflects an imperfect correspondence between visual and neural representations. Although deep networks capture aspects of visual brain representations, as assessed by benchmarks such as Brain-Score (Schrimpf et al., 2018; 2020), this correspondence remains incomplete (Xu & Vaziri-Pashkam, 2021). Similarities in image-feature space may therefore differ from those in Brain-Space, so a set covering image space well does not necessarily provide good coverage of brain responses. Indeed, selecting images to maximize Image-Space coverage yields the lowest coverage of the target participant's measured responses among the methods compared in this analysis (Appendix A.10). The ROI results provide a complementary pattern: at $N = 20$, Image-Space and Brain-Space selection perform similarly in early visual cortex (V1–V3), while Brain-Space provides larger gains in higher-level visual areas, including face- and body-selective regions and especially the scene-selective PPA and RSC (Figure 2.b). One possible interpretation is that mapping images into Brain-Space emphasizes distinctions underweighted by generic visual similarities that may be particularly relevant to higher-level visual responses, making the selected sets more representative of neural responses.

The largest gains appear in the low-data regime, where the choice of each measured image has the greatest influence on the information available for adaptation. One possible explanation is that, as the budget grows, even random sampling increasingly captures the diversity of the candidate pool, leaving less room for optimized selection to improve performance. Extended discussion and limitations are provided in Appendix B. Overall, our results support the hypothesis that images covering the neural-response space of previously scanned subjects also form informative training sets for a new individual, allowing selection before collecting any target-subject fMRI. This principle holds consistently across individuals and generalizes across datasets and encoders, showing that shared neural structure can guide not only multi-subject encoder transfer, but also the choice of which images to scan.

7 AI USE DISCLOSURE

We used generative AI tools to assist with code implementation and to improve the clarity, fluency, and grammar of the manuscript. The authors reviewed all AI-assisted code and text and take full responsibility for the final manuscript.

REFERENCES

Hossein Adeli, Minni Sun, and Nikolaus Kriegeskorte. Transformer brain encoders explain human high-level visual responses. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Emily J Allen, Ghislain St-Yves, Yihan Wu, Jesse L Breedlove, Jacob S Prince, Logan T Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, et al. A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, 25(1):116–126, 2022.

Roman Beliy, Guy Gaziv, Assaf Hoogi, Francesca Strappini, Tal Golan, and Michal Irani. From voxels to pixels and back: Self-supervision in natural-image reconstruction from fmri. *Advances in Neural Information Processing Systems*, 32, 2019.

Roman Beliy, Navve Wasserman, Amit Zalcher, and Michal Irani. The wisdom of a crowd of brains: A universal brain encoder. *arXiv preprint arXiv:2406.12179*, 2024.

Ian Charest, Rogier A Kievit, Taylor W Schmitz, Diana Deca, and Nikolaus Kriegeskorte. Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences*, 111(40):14565–14570, 2014.

Anders M Dale. Optimal experimental design for event-related fmri. *Human brain mapping*, 8(2-3): 109–114, 1999.

Megan T Debettencourt, Jonathan D Cohen, Ray F Lee, Kenneth A Norman, and Nicholas B Turk-Browne. Closed-loop training of attention with real-time brain imaging. *Nature neuroscience*, 18 (3):470–475, 2015.

Paul E Downing, Yuhong Jiang, Miles Shuman, and Nancy Kanwisher. A cortical area selective for visual processing of the human body. *Science*, 293(5539):2470–2473, 2001.

Serge O Dumoulin and Brian A Wandell. Population receptive field estimates in human visual cortex. *Neuroimage*, 39(2):647–660, 2008.

Michael Eickenberg, Alexandre Gramfort, Gaël Varoquaux, and Bertrand Thirion. Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, 152: 184–194, 2017.

Russell Epstein and Nancy Kanwisher. A cortical representation of the local visual environment. *Nature*, 392(6676):598–601, 1998.

Ma Feilong, Samuel A Nastase, J Swaroop Guntupalli, and James V Haxby. Reliable individual differences in fine-grained cortical functional architecture. *NeuroImage*, 183:375–386, 2018.

Zijin Gu, Keith Jamison, Mert Sabuncu, and Amy Kuceyeski. Personalized visual encoding model construction with small data. *Communications Biology*, 5(1):1382, 2022.

Guy Hacohen, Avihu Dekel, and Daphna Weinshall. Active learning on a budget: Opposite strategies suit high and low budgets. *arXiv preprint arXiv:2202.02794*, 2022.

James V Haxby, J Swaroop Guntupalli, Andrew C Connolly, Yaroslav O Halchenko, Bryan R Conroy, M Ida Gobbini, Michael Hanke, and Peter J Ramadge. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2):404–416, 2011.

Tomoyasu Horikawa and Yukiyasu Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications*, 8:15037, 2017. doi: 10.1038/ncomms15037.

Nancy Kanwisher, Josh McDermott, and Marvin M Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of neuroscience*, 17(11): 4302–4311, 1997.

Povilas Karvelis, Aaron R Seitz, Stephen M Lawrie, and Peggy Seriès. Autistic traits, but not schizotypy, predict increased weighting of sensory information in bayesian visual integration. *ELife*, 7:e34115, 2018.

Kendrick N. Kay, Thomas Naselaris, Ryan J. Prenger, and Jack L. Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–355, 2008. doi: 10.1038/nature06713.

Meenakshi Khosla, Gia H Ngo, Keith Jamison, Amy Kuceyeski, and Mert R Sabuncu. A shared neural encoding model for the prediction of subject-specific fmri response. In *International Conference on Medical image Computing and Computer-Assisted Intervention*, pp. 539–548. Springer, 2020.

Romy Lorenz, Ricardo Pio Monti, Inês R Violante, Christoforos Anagnostopoulos, Aldo A Faisal, Giovanni Montana, and Robert Leech. The automatic neuroscientist: A framework for optimizing experimental design with closed-loop real-time fmri. *NeuroImage*, 129:320–334, 2016.

Rafael Malach, JB Reppas, RR Benson, KK Kwong, H Jiang, WA Kennedy, PJ Ledden, TJ Brady, BR Rosen, and RB Tootell. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, 92(18): 8135–8139, 1995.

Scott Marek, Brenden Tervo-Clemmens, Finnegan J Calabro, David F Montez, Benjamin P Kay, Alexander S Hatoum, Meghan Rose Donohue, William Foran, Ryland L Miller, Timothy J Hendrickson, et al. Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603(7902):654–660, 2022.

Anne C Mennen, Kenneth A Norman, and Nicholas B Turk-Browne. Attentional bias in depression: understanding mechanisms to improve training and treatment. *Current opinion in psychology*, 29: 266–273, 2019.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Jacob S. Prince, Ian Charest, Jan W. Kurzawski, John A. Pyles, Michael J. Tarr, and Kendrick N. Kay. Improving the accuracy of single-trial fmri response estimates using glmsingle. *eLife*, 11:e77599, 2022. doi: 10.7554/eLife.77599.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Johannes Roth and Martin N Hebart. How to sample the world for understanding the visual system. In *8th Annual Conference on Cognitive Computational Neuroscience*, 2025.

Jacob Schreiber, Jeffrey Bilmes, and William Stafford Noble. apricot: Submodular selection for data summarization in python. *Journal of Machine Learning Research*, 21(161):1–6, 2020.

Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Kailyn Schmidt, Daniel L. K. Yamins, and James J. DiCarlo. Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv*, pp. 407007, 2018. doi: 10.1101/407007.

Martin Schrimpf, Jonas Kubilius, Michael J. Lee, N. Apurva Ratan Murty, Robert Ajemian, and James J. DiCarlo. Integrative benchmarking to advance neurally mechanistic models of human intelligence. *Neuron*, 108(3):413–423, 2020. doi: 10.1016/j.neuron.2020.07.040.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.

Martin I Sereno, Anders M Dale, John B Reppas, Kenneth K Kwong, John W Belliveau, Thomas J Brady, Bruce R Rosen, and Roger BH Tootell. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212):889–893, 1995.

Ranganatha Sitaram, Tomas Ros, Luke Stoeckel, Sven Haller, Frank Scharnowski, Jarrod Lewis-Peacock, Nikolaus Weiskopf, Maria Laura Blefari, Mohit Rana, Ethan Oblak, et al. Closed-loop brain training: the science of neurofeedback. *Nature reviews neuroscience*, 18(2):86–100, 2017.

Navve Wasserman, Roman Beliy, Roy Urbach, and Michal Irani. Functional brain-to-brain transformation with no shared data. *arXiv preprint arXiv:2404.11143*, 2024.

Haiguang Wen, Junxing Shi, Wei Chen, and Zhongming Liu. Transferring and generalizing deep-learning-based neural encoding models across subjects. *NeuroImage*, 176:152–163, 2018.

Yaoda Xu and Maryam Vaziri-Pashkam. Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nature Communications*, 12(1):2065, 2021. doi: 10.1038/s41467-021-22244-7.

Josefine Zerbe, Johannes Roth, Maggie Mae Mell, Peer Herholz, Tomas Knapen, and Martin N. Hebart. Laion-fmri: A densely sampled 7t-fmri dataset providing broad coverage of natural image diversity, 2026. URL `https://www.visionsciences.org/talk-sessions?id=642`. Talk 25.11, Scene Perception Talk Session, Vision Sciences Society Annual Meeting, May 16, 2026.

Appendix

A Ablations and Robustness Analyses

A.1 Representation and selection-objective contributions

To disentangle the contribution of the representation from that of the selection objective, we evaluate the same selection rules across visual and brain spaces. Table T1 reports this matched comparison for all eight NSD subjects, using three candidate pools and $N \in \{20, 40, 60, 80, 100\}$. The results show a clear interaction between representation and objective. Within image space, k-means is generally the strongest selection rule, particularly with CLIP full-token features. In brain space, both k-means and coverage substantially outperform ED and k-center, with coverage achieving the highest performance at every evaluated budget. Thus, the advantage of our main method does not follow simply from performing selection in brain space: the choice of selection objective is also important, with population coverage providing the strongest combination.

Table T1: **Representation-by-objective comparison on NSD.** Median voxelwise Pearson correlation on the fixed test set for all eight NSD subjects (higher is better). We apply the same selection objectives across CLIP CLS, CLIP full-token, and Brain-Space representations, with k-center included as an additional diversity baseline. Each cell first averages the same three matched candidate pools within each subject and then averages subjects equally. This matched comparison isolates the contributions of the representation and selection objective to downstream encoding performance.

Representation	**Selection**	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
-	Randomly	.232	.294	.327	.346	.361
Image-Space (CLIP CLS)	ED	.231	.290	.325	.341	.354
	k-means	.247	.303	.334	.347	.360
	k-center	.226	.282	.317	.333	.349
	Coverage	.234	.296	.332	.346	.361
Image-Space (CLIP Full)	ED	.215	.270	.309	.326	.343
	k-means	.259	.310	.335	.351	.364
	k-center	.224	.271	.300	.315	.330
	Coverage	.252	.306	.333	.345	.361
Brain-Space (fMRI)	ED	.179	.218	.252	.279	.301
	k-means	.281	.322	.347	.353	.368
	k-center	.198	.245	.279	.304	.325
	Coverage	.288	.333	.353	.361	.372

A.2 SCALING TO LARGER IMAGE BUDGETS

Our main experiments focus on the low-data regime $N \leq 100$, where stimulus selection has the largest effect. As expected, the advantage of Brain-Space selection decreases as more target-subject data become available. We nevertheless find that the benefit persists beyond the smallest image-set sizes. As shown in Table T2, at $N = 100$, Brain-Space selection still improves average performance over Random by 0.014 on NSD, 0.028 on LAION-fMRI ID, and 0.032 on LAION-fMRI OOD. On LAION-fMRI, smaller but consistent gains also remain at $N = 200$–300.

The worst-observed results show a similar, and in some cases larger, advantage. This is particularly relevant for prospective stimulus selection, where the quality of the available candidate pool is not known in advance. At $N = 100$, Brain-Space selection improves worst-observed performance by 6.1% on NSD, 9.6% on LAION-fMRI ID, and 17.6% on LAION-fMRI OOD relative to Random sampling. At $N = 200$, the corresponding relative improvements are 2.4%, 3.2%, and 6.5%, respectively, and at $N = 300$ they are 1.6%, 2.2%, and 3.1%. Thus, the benefit is strongest in the low-data regime and gradually diminishes as the training budget increases.

Table T2: **Scaling to larger image budgets.** Median voxelwise Pearson correlation on the NSD and LAION-fMRI ID and OOD fixed test sets (higher is better). Each cell reports average / worst-observed performance across candidate pools, first summarized within each participant and then averaged across participants equally. Brain-Space uses linear-NCSNR-weighted coverage in responses predicted by the Universal Encoder.

Dataset	**Method**	$N = 20$	$N = 100$	$N = 200$	$N = 300$	$N = 500$	$N = 1000$
NSD	Random	.230 / .193	.358 / .344	.378 / .370	.388 / .380	.402 / .395	.423 / .419
	Brain-Space	**.288 / .273**	**.372 / .365**	**.384 / .379**	**.391 / .386**	**.404 / .399**	**.423 / .420**
LAION-fMRI (ID)	Random	.238 / .188	.405 / .386	.449 / .439	.462 / .453	.478 / .471	**.495** / .490
	Brain-Space	**.322 / .292**	**.433 / .423**	**.460 / .453**	**.469 / .463**	**.481 / .476**	.494 / **.490**
LAION-fMRI (OOD)	Random	.237 / .160	.388 / .341	.425 / .397	.439 / .420	.455 / .442	.476 / .464
	Brain-Space	**.334 / .300**	**.420 / .401**	**.437 / .423**	**.449 / .433**	**.463 / .452**	**.478 / .468**

A.3 VISUAL REPRESENTATION AND SELECTION OBJECTIVE

We first examined how the choice of representation and selection objective affects downstream encoding performance. Table T4 compares CLIP and DINOv2 representations, using both CLS and full-token features, together with coverage, k-means, and effective-dimensionality (ED) selection. This development analysis uses NSD Subjects 1–2, three matched candidate pools, and $N \in \{20, 60, 100\}$. For each pool, selection is evaluated on the unused 20% complement of the training images; the corresponding fixed-test results are shown only as a descriptive check.

CLIP and DINOv2 perform similarly overall, while full-token representations are generally more effective than CLS features in the smallest-data regime. Brain-Space coverage also performs strongly in this initial comparison. Within CLIP Full, however, coverage and k-means are very close on Subjects 1–2 and their relative performance varies across budgets. We therefore expanded the held-out comparison used to select the final Image-Space objective to all eight NSD subjects.

As shown in Table T3, k-means outperforms coverage at every evaluated budget $N \in \{20, 40, 60, 80, 100\}$ on the unused-complement validation sets. Averaging budgets, candidate pools, and subjects equally gives a selection score of 0.3312 for k-means and 0.3243 for coverage. We therefore use CLIP full-token k-means as the primary Image-Space baseline in the main experiments. The expanded method-selection analysis also considered the other CLIP representation/objective variants; the table highlights the two closest full-token alternatives that determined the final choice.

Table T3: **Expanded held-out validation for the CLIP selection objective.** Because CLIP full-token coverage and k-means were close in the initial Subjects 1–2 development comparison, we extended their validation to all eight NSD subjects. Values are median voxelwise Pearson correlations on each pool's unused 20% complement, averaged equally across subjects and three matched candidate pools. k-means is higher at every evaluated budget; averaging equally across budgets gives selection scores of 0.3312 for k-means and 0.3243 for coverage.

Selection	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
Coverage	.2553	.3099	.3367	.3514	.3684
k-means	**.2636**	**.3167**	**.3433**	**.3593**	**.3731**

Table T4: **Matched representation and objective development ablation.** Median voxelwise Pearson correlation for NSD Subjects 1–2, averaged over three matched candidate pools (higher is better). For each pool, the unused 20% complement of the training images forms the validation set; fixed-test results are shown only as a descriptive check. We compare CLIP and DINOv2 CLS and full-token representations across coverage, k-means, and effective-dimensionality (ED) selection, together with Brain-Space representations predicted by the Universal Encoder.

		Unused-complement validation			**Fixed test**		
Objective	**Representation**	$N = 20$	$N = 60$	$N = 100$	$N = 20$	$N = 60$	$N = 100$
Random	Randomly Sampled	.2790	.4102	.4585	.2813	.4085	.4545
Coverage	CLIP CLS	.2867	.4228	.4602	.2917	.4217	.4591
	DINOv2 CLS	.2802	.4199	.4589	.2816	.4198	.4549
	CLIP Full	.3123	.4311	.4608	.3168	.4331	.4601
	DINOv2 Full	.3199	.4306	.4609	.3201	.4298	.4588
	Brain-Space	.3452	.4403	.4658	.3472	.4368	.4607
k-means	CLIP CLS	.3101	.4167	.4591	.3125	.4149	.4564
	DINOv2 CLS	.2786	.4141	.4520	.2845	.3923	.4248
	CLIP Full	.3095	.4296	.4602	.3139	.4284	.4574
	DINOv2 Full	.3098	.4214	.4640	.3136	.3992	.4319
	Brain-Space	.3312	.4324	.4653	.3284	.4292	.4575
ED	CLIP CLS	.2694	.4094	.4555	.2672	.4048	.4475
	DINOv2 CLS	.2708	.4077	.4569	.2803	.3869	.4263
	CLIP Full	.2609	.3885	.4413	.2646	.3838	.4345
	DINOv2 Full	.2522	.3875	.4357	.2496	.3803	.4281
	Brain-Space	.2056	.3066	.3889	.2026	.3030	.3846

A.4 STIMULUS SELECTION USING THE FMRI ENCODER'S VISUAL REPRESENTATIONS

The Universal Encoder uses a DINOv2-based visual backbone, raising the question of whether its learned visual representations provide better features for stimulus selection than ordinary pretrained visual features. We therefore compare frozen CLIP and DINOv2 representations with visual features extracted from different stages of the Universal Encoder, while keeping downstream adaptation fixed.

Representations and evaluation. We compare CLIP full patch-token features with four DINOv2 representations: (i) frozen final patch tokens; (ii) final patch tokens incorporating the encoder's learned low-rank adaptation (LoRA) residuals; (iii) multilevel projected features immediately before voxel attention, with LoRA enabled; and (iv) the same projected features with LoRA disabled. The projected representations combine features from layers 1, 6, 12, 18, and 24 after learned layer-specific normalization and projection. Importantly, disabling LoRA removes the backbone adaptation but retains the fMRI-trained normalization and projection layers, so this condition is distinct from ordinary frozen DINOv2.

Each representation is evaluated with Coverage and k-means using NSD Subjects 1 and 2, three matched candidate pools, and $N \in \{20, 60, 100\}$. All selected sets use the same Universal-Encoder adaptation procedure, training only the target voxel embeddings. Performance is evaluated on the fixed test set, averaging pools within each subject and then weighting subjects equally.

Frozen visual features remain competitive. Table T5 shows that encoder-derived visual features do not consistently improve stimulus selection over the frozen representations. With Coverage at $N = 20$, frozen DINOv2 achieves 0.3201, close to the strongest encoder-derived condition at 0.3209. Frozen CLIP achieves the highest performance at $N = 60$ and nearly matches the best result at $N = 100$. The effect of adaptation nevertheless depends on the selection objective: several encoder-derived representations improve over frozen DINOv2 with k-means at the larger budgets. Thus, adaptation can help particular representation–objective combinations, but does not yield a consistent advantage over the competitive frozen-feature conditions. These findings suggest that visual features learned for response prediction do not automatically provide better similarities for choosing training images, supporting the use of standard pretrained features as Image-Space controls.

Table T5: **Image-Space selection using pretrained and encoder-derived visual representations.** Median voxelwise Pearson correlation on the fixed NSD test set for Subjects 1–2, averaged over three matched candidate pools within each subject and then equally across subjects. Each representation is evaluated using Coverage and k-means. CLIP Full and DINOv2 Full denote the standard pretrained representations. DINOv2 Full (adapted) includes the visual-backbone adaptation learned within the Universal Encoder. Both DINOv2 Projected variants use the encoder's learned multilevel normalization and projection layers; they differ only in whether the adapted or base DINOv2 backbone is used. Bold marks the highest value at each image-set size.

Visual Representation	**Selection Objective**	**Image-Set Size**		
		20	60	100
CLIP Full	Coverage	0.3168	0.4331	0.4601
	k-means	0.3139	0.4284	0.4574
DINOv2 Full	Coverage	0.3201	0.4298	0.4588
	k-means	0.3136	0.3992	0.4319
DINOv2 Full (adapted)	Coverage	0.2954	0.4216	0.4547
	k-means	0.3059	0.4213	0.4572
DINOv2 Projected (adapted backbone)	Coverage	0.3089	0.4236	0.4593
	k-means	0.2768	0.4160	0.4531
DINOv2 Projected (base backbone)	Coverage	0.3209	0.4259	0.4603
	k-means	0.3117	0.4175	0.4522

A.5 SOURCE-VOXEL RELIABILITY WEIGHTING

We next test whether accounting for source-voxel reliability improves Brain-Space selection. We compare coverage using the standardized predicted responses directly with coverage after weighting each source voxel by its normalized NCSNR. This development experiment uses NSD Subjects 1–2, three matched candidate pools, and $N \in \{20, 60, 100\}$. The unused 20% complement of each candidate pool is used for validation, while fixed-test performance is reported only as a descriptive check. As shown in Table T6, linear NCSNR weighting improves validation performance at $N = 20$ and $N = 60$, while the two variants are nearly identical at $N = 100$. Under the predeclared selection criterion, which averages performance equally across validation budgets, linear NCSNR achieves 0.4171 compared with 0.4126 without weighting. We therefore use linear NCSNR weighting for Brain-Space selection in the main experiments.

Table T6: **Source-voxel reliability weighting.** Median voxelwise Pearson correlation for NSD Subjects 1–2 comparing unweighted and linear-NCSNR-weighted Brain-Space coverage (higher is better). Values are averaged over three matched candidate pools. For each pool, the unused 20% complement forms the validation set used to select the weighting scheme; fixed-test results are shown only as a descriptive check.

	Unused-complement validation			**Fixed test**		
Method	$N = 20$	$N = 60$	$N = 100$	$N = 20$	$N = 60$	$N = 100$
Random	.2790	.4102	.4585	.2813	.4085	.4545
Brain-Space Coverage (unweighted)	.3352	.4359	**.4668**	.3327	.4320	**.4634**
Brain-Space Coverage (linear NCSNR)	**.3452**	**.4403**	.4658	**.3472**	**.4368**	.4607

A.6 IMPORTANCE OF CROSS-SUBJECT MODEL TRANSFER

Our main framework combines cross-subject model transfer with optimized stimulus selection. To separate these contributions, we compare the standard transfer setting with target-only training. The target-only model retains the same pretrained visual backbone, but all fMRI-specific parameters are reinitialized and learned only from the selected target-subject data. For this comparison, both target-only and cross-subject models are first trained on matched randomly selected image sets; we then additionally report Brain-Space selection with cross-subject transfer. As shown in Table T7, cross-subject transfer provides a large improvement throughout the low-data regime. At $N = 20$, median voxelwise correlation increases from 0.121 with target-only training to 0.232 with cross-subject transfer. Brain-Space selection further increases performance to 0.288. The same pattern persists across budgets, showing that optimized stimulus selection provides an additional benefit on top of the substantial gain from cross-subject transfer.

Table T7: **Importance of cross-subject model transfer.** Median voxelwise Pearson correlation on the NSD fixed test set for all eight subjects (higher is better). Values are averaged over three matched candidate pools within each subject and then across subjects equally. Target-only training retains the pretrained visual backbone but reinitializes all fMRI-specific parameters, which are learned only from the selected target data. The Random image sets are matched between target-only and cross-subject training. Brain-Space Coverage shows the additional gain obtained by optimizing the selected images.

Training and selection	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
Target-only training + Random	.121	.153	.169	.183	.192
Cross-subject transfer + Random	.232	.294	.327	.346	.361
Cross-subject transfer + fMRI Coverage	**.288**	**.333**	**.353**	**.361**	**.372**

A.7 SENSITIVITY TO THE SOURCE-SUBJECT SET

Our main Brain-Space representation concatenates predicted responses from multiple source subjects. We therefore examine how sensitive stimulus selection is to the particular source-subject set by constructing the brain representation from each source individually, or from all four sources together. We use NSD targets S3, S4, S6, and S8 and source subjects S1, S2, S5, and S7, so that every evaluated source is distinct from every target. The comparison uses three matched candidate pools and $N \in \{20, 40, 60, 80, 100\}$. As shown in Table T8, representations constructed from a single source subject already achieve performance close to the four-source representation across budgets. Combining all four sources provides only a modest additional benefit overall. This suggests that the selection signal is not strongly dependent on pooling many source subjects.

Table T8: **Sensitivity to the source-subject set.** Median voxelwise Pearson correlation for NSD targets S3, S4, S6, and S8 using Brain-Space coverage constructed from each source subject separately or from all four sources together. Values first average three matched candidate pools within each target and then weight targets equally. The unused 20% complements provide the primary source-composition comparison; fixed-test results are shown as a descriptive check.

	Unused-complement validation					**Fixed test**				
Source set	**20**	**40**	**60**	**80**	**100**	**20**	**40**	**60**	**80**	**100**
Source S1	.2435	.2877	.3082	.3133	.3265	.2280	.2668	.2875	.2924	.3046
Source S2	.2541	.2910	.3101	.3114	.3237	.2385	.2718	.2899	.2917	.3032
Source S5	.2346	.2794	.2984	.3050	.3185	.2207	.2616	.2782	.2848	.2985
Source S7	.2403	.2828	.3054	.3108	.3232	.2235	.2647	.2850	.2905	.3030
Sources S1+S2+S5+S7	.2541	.2916	.3094	.3141	.3251	.2387	.2748	.2900	.2958	.3056

A.8 ROBUSTNESS TO CANDIDATE-POOL CHOICE

Encoding performance varies across candidate pools, and the best-performing pool is not known in advance. We therefore examine whether selection remains effective under less favorable pool choices. We report worst-observed performance by taking the minimum score across nine matched candidate pools within each subject, then averaging these minima equally across subjects. This summarizes the evaluated pools rather than providing a performance guarantee for unseen pools.

On NSD, Brain-Space selection provides larger gains over Random sampling and Image-Space selection in worst-observed performance than in average performance at every evaluated budget. Participant-level results are provided in Table T16. Corresponding results for LAION-fMRI and GOD are provided in Appendices D.4 and D.5.

A.9 CONSISTENCY OF SELECTED IMAGES ACROSS CANDIDATE POOLS

We first visualize how Brain-Space stimulus selection changes when the available candidate pool changes. Each candidate pool contains 80% of the available training images, so pools overlap substantially while differing in which images are available for selection. Figure S1 shows independently selected $N = 20$ image sets from two such candidate pools for NSD, LAION-fMRI, and GOD. Despite these differences, many exact images recur across pools, and several non-identical selections also exhibit similar semantic content or visual composition.

Visual correspondences extend beyond exact image matches. Brain-Space selection preserves both exact images and broader semantic or compositional structure across candidate pools. In NSD, for example, the first four selected images are identical across the two pools. The next several selections are not identical, but remain closely related semantically: an outdoor scene with people, a face with an object held near the face, an outdoor road-like scene, a bathroom image, and an indoor scene containing multiple people. Elsewhere in the two sets, although the ordering is no longer aligned, similar content appears again, including bus images, towers, sky-dominated scenes, animal-focused images, and tennis players. Similar patterns of both exact recurrence and semantically related selections can also be seen in the LAION-fMRI and GOD datasets.

Occurrence-weighted cross-pool consistency. We quantify exact-image recurrence across all nine matched candidate pools. For a given budget N, let $c_i(N)$ denote the number of pools in which image i is selected. We call an image consistent if it is selected in at least five of the nine pools and define

$$\text{Consistency}(N) = \frac{\sum_i c_i(N)\,\mathbf{1}[c_i(N) \geq 5]}{9N} \times 100. \tag{1}$$

The numerator counts every selection occurrence belonging to a consistent image: an image selected in seven pools contributes seven, not one. The denominator counts all $9N$ selection occurrences rather than the number of distinct selected images.

Brain-Space selection is more consistent across pools. Table T9 shows higher consistency for Brain-Space than for Image-Space or Random selection at every evaluated dataset and budget. At $N = 20$, Brain-Space consistency is 35.6% on NSD, 44.4% on LAION-fMRI, and 31.7% on GOD, compared with 13.3%, 7.8%, and 11.1% for Image-Space, respectively. The advantage persists at larger budgets, whereas Random selection has near-zero consistency under this majority-of-pools criterion. Thus, Brain-Space repeatedly identifies a recurring subset of stimuli despite changes in candidate-pool composition, complementing the robustness observed in downstream encoding performance.

Table T9: **Cross-pool consistency of selected images.** Percentage of selection occurrences belonging to images selected in at least five of nine candidate pools (Equation 1). Every occurrence of a qualifying image contributes to the numerator; the denominator is $9N$. The measure uses exact image identities, not semantic similarity. Bold marks the highest consistency within each dataset and image budget.

Dataset	Method	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
NSD	Random	0.0	0.0	0.0	0.0	0.0
	Image-Space	13.3	18.3	15.7	9.7	12.1
	Brain-Space	**35.6**	**35.3**	**33.7**	**35.7**	**38.7**
LAION-fMRI	Random	0.0	0.0	0.0	0.0	0.0
	Image-Space	7.8	3.9	3.5	1.0	1.9
	Brain-Space	**44.4**	**40.6**	**45.6**	**50.4**	**49.4**
GOD	Random	0.0	0.0	0.0	0.0	0.6
	Image-Space	11.1	3.9	2.2	3.3	3.2
	Brain-Space	**31.7**	**42.2**	**42.4**	**44.4**	**44.9**



Figure S1: **Brain-Space selections preserve recurring images and visual content across candidate pools.** Each dataset block shows $N = 20$ images selected from Candidate Pool 1 (left) and Candidate Pool 2 (right), arranged in 4×5 grids. Results are shown for NSD Subject 1, LAION-fMRI Subject 1, and the GOD selection shared across subjects. In addition to exact repetitions, distinct selected images often share content or composition, as illustrated by the NSD clock towers and tennis scenes. We first visualize how Brain-Space stimulus selection changes when the available candidate pool changes. Each candidate pool contains 80% of the available training images, so pools overlap substantially while differing in which images are available for selection.

A.10 CROSS-SPACE EVALUATION OF SELECTED IMAGE SETS

We examine what makes the selected images representative, and whether coverage optimized in predicted source responses carries over to measured target responses. We compare Random sampling, Image-Space k-means, Image-Space Coverage, and Brain-Space Coverage at $N = 20$ and $N = 60$. Each selected set is held fixed and evaluated in three representations: CLIP full-token space, predicted source Brain-Space, and measured target Brain-Space. Results are averaged across three matched candidate pools within each of the eight NSD subjects, followed by equal-subject averaging. Measured target responses are used only for this post hoc analysis, never for selection.

Tables T10 and T11 report four complementary geometric summaries. Coverage measures the average nonnegative cosine similarity between each candidate image and its most similar selected image. Induced-cluster distortion summarizes the compactness of the candidate groups represented by the selected images, while worst-case distance measures how far the least well-represented candidate lies from its nearest selected image. Effective dimensionality (ED) describes how broadly the selected-set variation is distributed across dimensions. Population-based scores use the complete candidate pool, whereas ED characterizes the selected set itself. Comparisons are made within each evaluation space, rather than between the absolute scores of different representations.

Visual-space coverage does not ensure neural-response coverage. Image-Space Coverage achieves the highest CLIP coverage at both budgets, slightly exceeding Image-Space k-means. However, this ordering does not carry over to neural-response space. At $N = 20$, Image-Space Coverage reaches a CLIP coverage score of 0.469, compared with 0.465 for k-means, but its measured target-response coverage is lower: 0.292 versus 0.303. It also falls below Random sampling in both predicted and measured response coverage at both budgets. Thus, improving how well a set represents the candidate population in CLIP space does not necessarily improve how well it represents the corresponding neural responses. The comparison between Image-Space and Brain-Space Coverage holds the selection objective fixed, highlighting the importance of the representation in which coverage is optimized.

The coverage advantage transfers to measured target responses. Brain-Space Coverage achieves the highest coverage in predicted source space, as expected from its objective. More importantly, its advantage extends to the target participant's measured responses. At $N = 20$, measured-response coverage reaches 0.335, compared with 0.303 for both Image-Space k-means and Random sampling. At $N = 60$, the corresponding scores are 0.381, 0.364, and 0.359. Brain-Space Coverage also yields the lowest measured-space distortion and worst-case distance at both budgets. These results support the central premise of cross-subject selection: images chosen to represent source-subject response patterns also form a more representative set in the target participant's measured response space, without using those measurements during selection.

Better neural coverage does not require higher dimensionality. Brain-Space Coverage has the lowest ED among the compared methods in both predicted and measured neural spaces at both budgets, despite achieving the highest coverage. In CLIP space, however, its ED remains close to that of Random sampling. The selected images therefore need not maximize neural-response dimensionality to represent the candidate population effectively. This distinction suggests that the benefit is associated with representative response patterns rather than simply spreading selected-set variation across more dimensions. The metrics also capture different aspects of representativeness: at $N = 60$, Random sampling has a slightly lower worst-case distance in predicted Brain-Space, even though Brain-Space Coverage provides higher average coverage. The latter therefore improves population representation without necessarily optimizing every geometric criterion.

Table T10: **Cross-space evaluation of selected image sets at $N = 20$.** The same selected NSD image sets are evaluated in CLIP full-token space, predicted source Brain-Space, and measured target Brain-Space. Population-based scores use the complete candidate pool; effective dimensionality (ED) describes the selected set. Results average three matched pools within each subject and then weight the eight subjects equally. Target responses are used only for post hoc evaluation. Arrows indicate the optimization direction of each metric, and bold marks the best value within each evaluation space. Scores should be compared between methods within a space, not across spaces.

Evaluation space	Selection method	Coverage ↑	Distortion ↓	Worst ↓	ED ↑
CLIP Full	Random	.430	.608	1.346	**18.84**
	Image-Space k-means	.465	.606	1.302	18.57
	Image-Space Coverage	**.469**	**.603**	**1.299**	18.48
	Brain-Space Coverage	.434	.607	1.345	18.79
Predicted Brain-Space	Random	.532	.605	1.799	7.02
	Image-Space k-means	.523	.607	1.820	7.21
	Image-Space Coverage	.456	.628	2.201	**7.46**
	Brain-Space Coverage	**.619**	**.559**	**1.714**	5.23
Measured Brain-Space	Random	.303	.806	1.887	**12.42**
	Image-Space k-means	.303	.803	1.910	12.21
	Image-Space Coverage	.292	.810	1.989	12.01
	Brain-Space Coverage	**.335**	**.795**	**1.872**	10.36

Table T11: **Cross-space evaluation of selected image sets at $N = 60$.** Evaluation follows the protocol of Table T10, using the complete candidate population for coverage, induced-cluster distortion, and worst-case distance, and the selected images for ED. Results average three matched pools within each subject and then weight the eight subjects equally. Measured target responses are used only for post hoc evaluation. Brain-Space Coverage retains the highest predicted- and measured-response coverage despite having lower neural-response ED. Bold marks the best value for each metric within an evaluation space.

Evaluation space	Selection method	Coverage ↑	Distortion ↓	Worst ↓	ED ↑
CLIP Full	Random	.443	.600	1.337	**57.86**
	Image-Space k-means	.470	.599	1.301	57.25
	Image-Space Coverage	**.474**	**.594**	**1.294**	56.06
	Brain-Space Coverage	.452	.601	1.324	57.55
Predicted Brain-Space	Random	.612	.521	**1.576**	**9.42**
	Image-Space k-means	.613	.522	1.579	9.25
	Image-Space Coverage	.580	.536	1.822	9.16
	Brain-Space Coverage	**.675**	**.485**	1.581	6.66
Measured Brain-Space	Random	.359	.760	1.797	22.01
	Image-Space k-means	.364	.760	1.803	20.73
	Image-Space Coverage	.350	.764	1.849	**22.45**
	Brain-Space Coverage	**.381**	**.753**	**1.790**	18.04

A.11 VISUALIZING SELECTION ACROSS VISUAL AND NEURAL REPRESENTATIONS

We visualize where the different methods select images from the same candidate population. Figures S2 and S3 show NSD Subject 1, Pool 1, at $N = 20$ and $N = 60$, respectively. Columns show t-SNE projections of CLIP CLS, CLIP full-token, and predicted Brain-Space representations. Gray points denote all candidate images, and highlighted points denote selected images. Each column uses the same coordinates across methods and budgets. Both Image-Space methods select using CLIP full-token features; the CLS column provides an additional visual view.

Brain-Space selects different images from many of the same visual neighborhoods. In the CLIP projections, Image-Space k-means and Brain-Space Coverage both select images across many distinct regions of the candidate population. Their selections often occupy similar neighborhoods, particularly along the right-hand side of the CLIP maps. However, the two methods share no exact selected images in the displayed subject and pool at either budget. The apparent overlap therefore reflects different images from nearby visual neighborhoods, not identical selections. Other locations contain Brain-Space selections without a nearby k-means representative. Thus, Brain-Space selection retains broad visual variety while changing both which neighborhoods are represented and which images represent them.

The neural map reveals what visual-space selection leaves out. The rightmost column makes the distinction clearer. At $N = 20$, several central and lower-left portions of the neural map contain Brain-Space selections but no Image-Space selections. Meanwhile, some images chosen by the Image-Space methods lie close together in the neural projection, leaving other portions sparsely represented. Brain-Space Coverage distributes its selections more broadly through these locations. The important difference is therefore not simply whether the images appear varied in CLIP space: images representing different visual neighborhoods need not provide the same breadth of neural-response patterns. This interpretation is supported by the quantitative analysis, where Brain-Space Coverage achieves higher coverage in both predicted source responses and measured target responses (Appendix A.10).

The selection objective also changes the visible pattern. Even within CLIP space, k-means and coverage produce noticeably different selections. At $N = 20$, Image-Space Coverage places several images close together in the projections, particularly in the bottom cluster, whereas k-means distributes its representatives across more visibly separated locations. These projected groupings do not imply worse CLIP coverage: Image-Space Coverage actually achieves the highest coverage in the original CLIP representation. This example illustrates why the maps are useful for seeing differences between selected sets, but cannot replace the high-dimensional quantitative comparisons.

Larger budgets make the differences less pronounced. At $N = 60$, more parts of each map contain selected images under every method, including Random sampling. Brain-Space still selects a different set of representatives, but the contrast is less pronounced than at $N = 20$. This agrees with the quantitative analysis, where its neural-coverage advantage over Random and Image-Space k-means becomes smaller. Together, the analyses suggest that Brain-Space selection is especially useful when a small set must represent the candidate population: it preserves many of the visual neighborhoods represented by Image-Space selection while prioritizing a different, more representative combination of neural-response patterns.

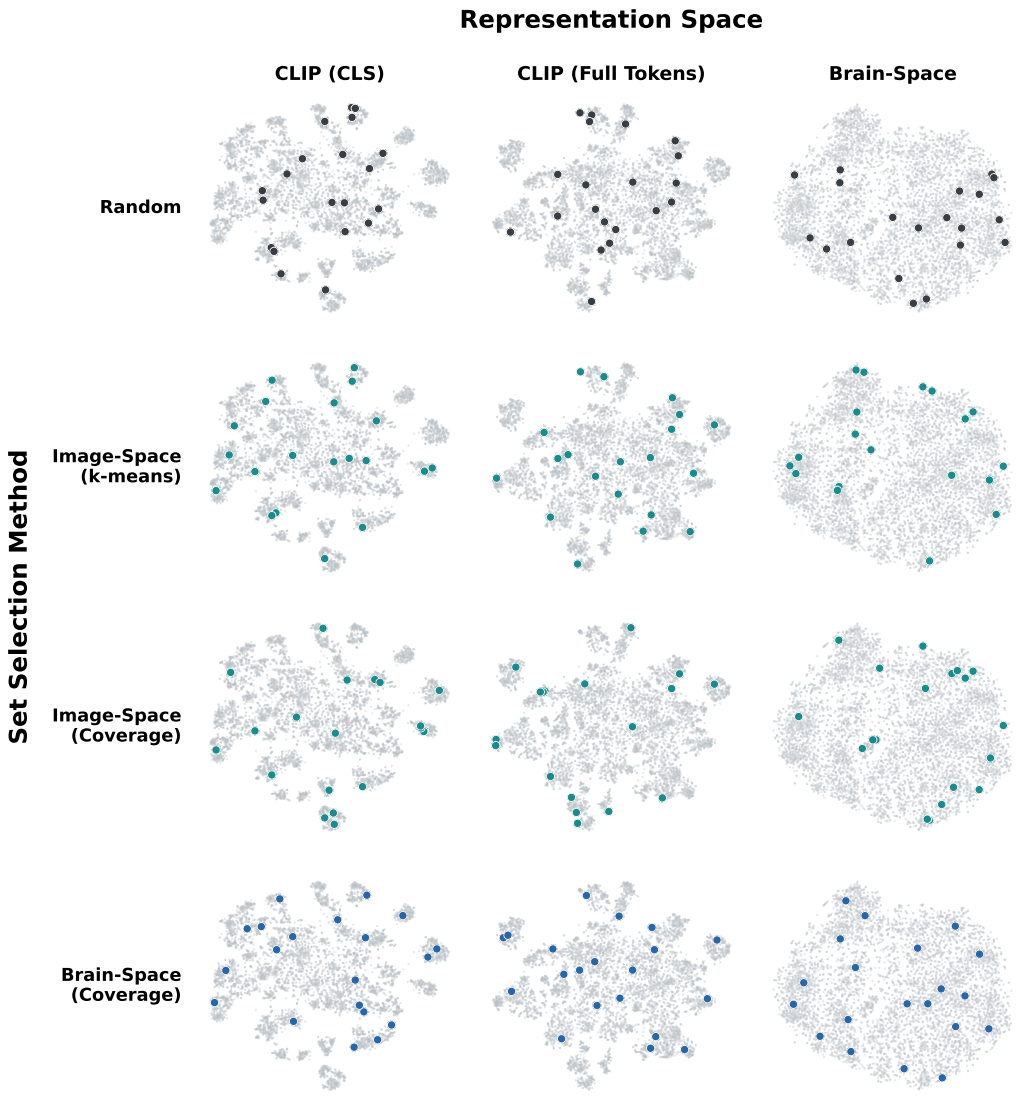


Figure S2: **Selected images occupy different neighborhoods across visual and neural representations at $N = 20$.** Rows show Random sampling, Image-Space k-means, Image-Space Coverage, and Brain-Space Coverage for NSD Subject 1, Pool 1. Columns show independently computed PCA-50/t-SNE projections of CLIP CLS, CLIP full-token, and predicted Brain-Space representations. Gray points denote all candidate images; highlighted points denote the selected set. Coordinates are fixed across methods within each column. Image-Space selection uses full-token features; CLS is an additional visualization only. The maps illustrate representation-dependent selection patterns, not quantitative distances or coverage in the original spaces.

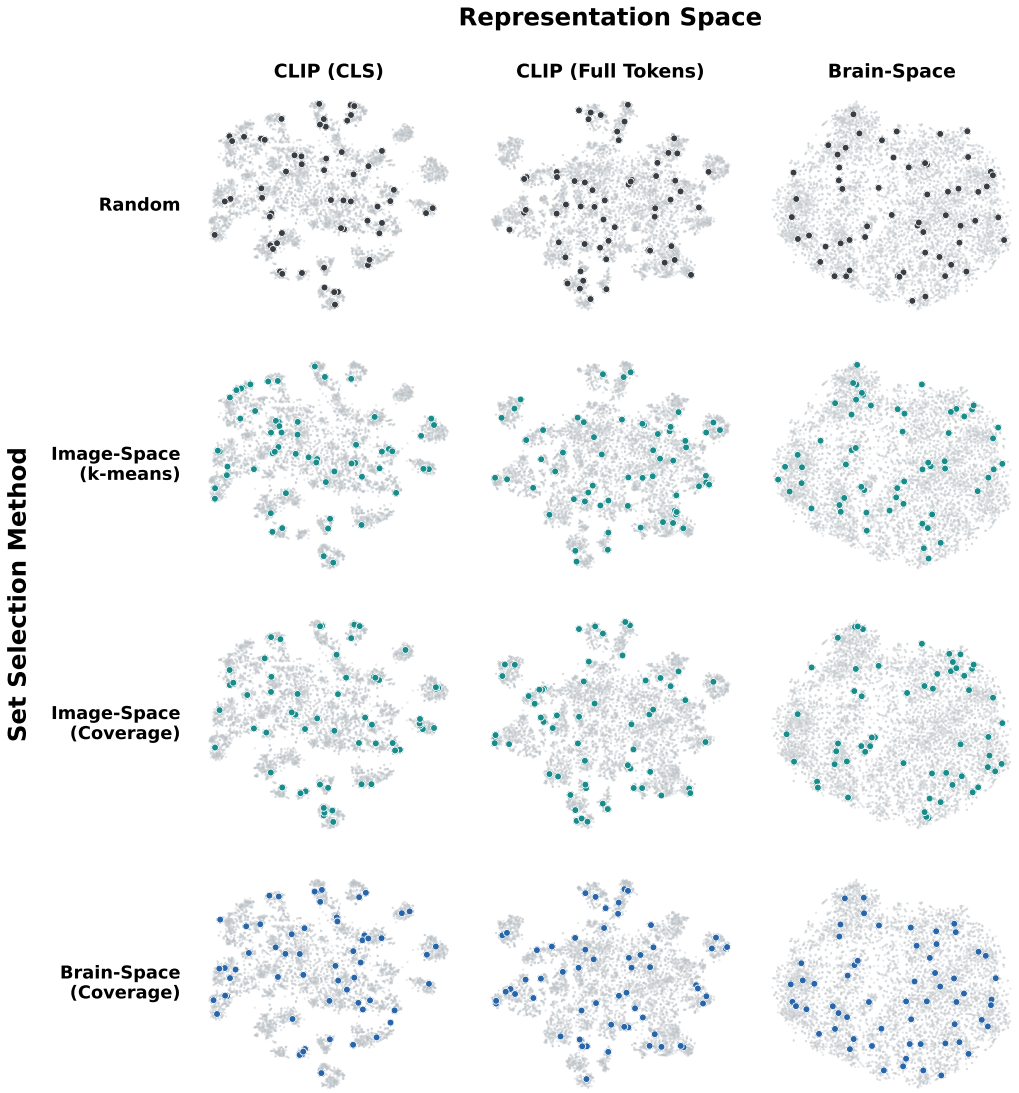


Figure S3: **Selection geometry at the larger budget of $N = 60$.** The same candidate images and representation maps as in Figure S2 are shown, now highlighting the sets selected at $N = 60$. More neighborhoods contain selected images under all methods, although their locations remain representation-dependent. Gray points denote the candidate population and highlighted points denote selected images. Both Image-Space methods select in CLIP full-token space, while the CLS column is an additional view. Quantitative comparisons use the original representations, not the t-SNE coordinates.

B Extended Discussion and Limitations

Reusing existing neural data for new participants. Large, densely sampled fMRI datasets such as NSD already provide rich measurements from a small number of extensively scanned participants. Our framework asks how these existing resources can be reused to reduce the measurements required for additional individuals. A source-trained encoder provides the neural-response representation used for stimulus selection, and our cross-dataset results show that this source resource need not come from the target dataset: an NSD-trained encoder also guides useful selection for LAION-fMRI and GOD. We further find that constructing Brain-Space from the predicted responses of a single source subject performs close to using four sources (Appendix A.7), suggesting that the selection signal itself does not require pooling predictions from many source subjects. This analysis keeps encoder pretraining fixed, and therefore addresses the source information needed for selection rather than the data required to train the initial encoder.

Retrospective evaluation and target-participant independence. Restricting candidate pools to images with existing fMRI measurements allows different selection methods to be compared using real responses under matched conditions. Although the measurements were collected previously, the selection procedure does not access the target participant's responses. The target is excluded from encoder pretraining and from the source subjects defining Brain-Space; only after selection are the chosen target image–fMRI pairs used for adaptation. The evaluation therefore preserves the central information constraint of selecting stimuli before scanning a new participant. Importantly, selection does not require measured source-subject responses to the candidate images, since these responses are predicted by the source-trained encoder. The cross-dataset experiments further test this procedure on new participants and stimulus collections, including the different acquisition setting of the 3T GOD dataset. These experiments provide a controlled evaluation of the selection rule, while remaining distinct from acquiring new scans after selection (Appendix C.5).

Visual-feature controls and predicted neural responses. The DINOv2 comparisons are particularly relevant because the Universal Encoder itself uses a DINOv2-based visual backbone. Our visual-feature controls include both CLIP and DINOv2 representations, with comparisons across selection objectives (Appendix A.3). They therefore address whether the advantage of Brain-Space reflects only the choice of a different generic visual representation. The results support the usefulness of defining selection through predicted neural responses, beyond simply replacing CLIP features with DINOv2 features.

Different selection behavior on GOD. On GOD, CLIP full-token k-means does not consistently improve over Random sampling, whereas Brain-Space selection improves average encoding performance at every evaluated budget (Appendix D.5). GOD has a smaller, category-structured candidate collection, and its test categories are disjoint from those used for training (Appendix C.4). One possible explanation is that images representing clusters in the training pool's visual-feature space are not necessarily the most informative examples for predicting neural responses to unseen categories. The smaller candidate pool may also influence the benefit of selection.

C Additional Technical Details

C.1 Encoder and transfer-learning details

Multi-subject encoder. We use the universal image-to-fMRI encoder of Beliy et al. (2024), which combines a DINOv2-based visual backbone with shared prediction parameters and learned voxel-specific embeddings. This architecture supports joint learning across participants while retaining participant-specific fMRI predictions. For each target participant in NSD, we use an encoder pretrained on the other seven participants. For transfer to LAION-fMRI and GOD, we initialize from an NSD-pretrained encoder.

Adaptation to a new participant. After stimulus selection, we initialize voxel embeddings for the new participant and optimize them using only the selected image–fMRI pairs. All pretrained model parameters remain frozen; only the new participant's voxel embeddings are learned. All selected pairs are used for adaptation.

Training configuration. We train for 100 epochs using Adam with AMSGrad, a learning rate of 10^{-3}, and a batch size of 32. We use 224-pixel image inputs, float32 computation, and the existing small-shift image augmentation. These adaptation settings are held fixed across selection methods in the main comparison.

C.2 Implementation details for visual and brain representations

Visual-feature representations. We extract features using pretrained CLIP ViT-L/14 (Radford et al., 2021) and DINOv2 ViT-L/14-reg (Oquab et al., 2023), keeping both models frozen without fine-tuning on fMRI data. For each model, we consider a global CLS representation and a full patch-token representation. For CLIP, we use tokens from the final visual-transformer block after the model's final layer normalization and learned visual projection. The resulting 768-dimensional CLS token is L2-normalized. For the full-token representation, we exclude the CLS token, flatten the 256 projected 768-dimensional patch tokens into a 196,608-dimensional vector, and L2-normalize the complete flattened vector. For DINOv2, we use the final normalized CLS and patch-token outputs. Register tokens are excluded. The 1,024-dimensional CLS token is L2-normalized, while the 256 normalized 1,024-dimensional patch tokens are flattened into a 262,144-dimensional vector and L2-normalized.

Predicted-fMRI representations. We predict each candidate image's fMRI responses for the source participants using the multi-subject encoder of Beliy et al. (2024). In the NSD experiments, the encoder is pretrained on the seven non-target participants. The representation uses predicted responses for source participants 1, 2, 5, and 7, excluding the target participant when applicable. In the cross-dataset experiments, NSD-trained encoders predict responses to candidate images from LAION-fMRI and GOD. For each source participant, predictions are standardized separately for every voxel using the mean and standard deviation across the candidate pool. For voxels with near-zero variance, we set the standardization denominator to one. Raw NCSNR values are made finite and nonnegative, normalized to have mean one within the included source-voxel block, clipped to $[0.1, 5]$, and normalized again to mean one. Each standardized predicted-response coordinate is multiplied by its resulting normalized NCSNR value. The weighted response vectors are concatenated across source participants to form one representation per image. Each participant therefore contributes a separate feature block, and no correspondence between participants' voxels is required. Coverage selection uses cosine similarity clipped below at zero. All standardization statistics are derived only from candidate-image predictions, and all reliability weights come from source-participant data; no measured target-participant fMRI is used during image selection.

C.3 Stimulus-set selection details

We apply the selection objectives to the visual and brain representations described in Appendix C.2. Let $\mathbf{z}_i$ denote the L2-normalized representation of candidate image x_i, and let $K_{ij} = \mathbf{z}_i^\top \mathbf{z}_j$ be their cosine Gram matrix. K-means, k-center, and effective-dimensionality selection retain negative cosine similarities; only coverage clips similarities to $[0, 1]$.

Candidate pools. In the nine-pool NSD comparison, each candidate pool contains 80% of the available training images for the target subject. Specifically, for T available images and pool seed $p \in \{0, \ldots, 8\}$, we sample $M = \lfloor 0.8T \rfloor$ positions without replacement using NumPy's `default_rng(p).choice(T, M, replace=False)`. The sampled positions are sorted to define the pool's image ordering. Pools may overlap and are held fixed across selection methods and representations. The fixed test images are excluded from pool construction, and measured target-subject fMRI is unavailable during stimulus selection.

Random selection. For each target subject and candidate pool, the Random baseline is a uniformly sampled ordering without replacement, shared across representation and objective comparisons; its prefixes define the different budgets and also initialize ED and k-center selection.

Facility-location coverage. The reference population and selectable population are both the candidate pool $\mathcal{C}$. We define $s_{ij} = \mathrm{clip}(K_{ij}, 0, 1)$ and maximize

$$F(S) = \sum_{x_i \in \mathcal{C}} \max_{x_j \in S} s_{ij}, \qquad |S| = N.$$

Selection starts from the empty set with zero coverage. At each step, we compute the marginal gain $F(S \cup \{x\}) - F(S)$ for every remaining candidate x, evaluating each gain over all reference images. There is no candidate or reference subsampling. Before selecting the maximum, we add a fixed candidate-specific jitter drawn uniformly from $[0, 10^{-9})$ using seed 0, making the selection deterministic in the presence of tied or nearly tied gains.

Effective-dimensionality selection. For a selected set S, we extract its cosine Gram matrix K_S and center it within the selected set:

$$K_c = HK_SH, \qquad H = I - \frac{1}{|S|}\mathbf{1}\mathbf{1}^\top.$$

Effective dimensionality is measured by the participation ratio,

$$D_{\text{eff}}(S) = \frac{\mathrm{tr}(K_c)^2}{\|K_c\|_F^2} = \frac{\left(\sum_j \lambda_j\right)^2}{\sum_j \lambda_j^2},$$

where λ_j are the eigenvalues of K_c. A covariance normalization factor is unnecessary because it cancels in the ratio. Selection begins with the first two distinct images from the matched Random ordering. At each subsequent step, we add the remaining candidate that maximizes the effective dimensionality of the enlarged set. Every remaining candidate is evaluated using algebraically vectorized updates of the centered-Gram trace and Frobenius norm, without candidate subsampling or approximate search. Scores are considered numerically tied using relative tolerance 10^{-12} and absolute tolerance 10^{-14}. Ties are resolved using a deterministic candidate priority, with a seed derived from the target subject, candidate pool, and representation.

K-means selection. For a budget of N images, we partition the candidate pool into N clusters using deterministic initialization. Clustering is performed using squared Euclidean distance between L2-normalized representations, which is equivalent to cosine geometry. After convergence, we select the candidate image nearest to each cluster centroid, yielding one representative image per cluster. We use a deterministic initialization for reproducibility and run Lloyd updates for at most 100 iterations. Empty clusters are reassigned deterministically using points from non-singleton clusters. K-means is fit independently for each budget N, so the selected sets are not nested across budgets.

K-center selection. We use deterministic farthest-first traversal, initialized with the first image from the matched Random ordering. At each subsequent step, we select the candidate whose minimum squared Euclidean distance to the selected set is largest. For unit-normalized representations,

$$d^2(x_i, x_j) = 2 - 2K_{ij}.$$

Cosine similarities are not clipped at zero. Only the resulting squared distances are lower-bounded at zero to correct numerical rounding errors. Distance ties are identified using relative tolerance 10^{-7} and absolute tolerance 10^{-8}, and resolved using a fixed seeded candidate priority. The seed is derived from the target subject, candidate pool, representation, and selection objective.

Selection across budgets. Random, coverage, effective-dimensionality, and k-center sets are nested across budgets: each smaller set is the corresponding prefix of a single ordering. For the three greedy methods, this ordering is constructed through the largest evaluated budget. K-means is instead fitted independently for each N, with budget-specific initialization, so its selected sets are not nested.

C.4 Datasets and response preparation

Table T12 summarizes the retained training and evaluation images. Training counts refer to the available images before constructing the 80% candidate pools. Throughout the paper, "full-brain" evaluation refers to the complete retained visual-cortex output for each dataset, rather than every anatomical brain voxel.

Table T12: Dataset splits used in our experiments. Training counts precede candidate-pool construction. Test/ID denotes the fixed test set for NSD and GOD and the in-distribution test set for LAION-fMRI. Counts refer to retained images after dataset-specific validity and repetition requirements.

Dataset	Subjects	Training	Test/ID	OOD
NSD	S1, S2, S5, S7	9,000	1,000	–
	S3, S6	8,481	930	–
	S4, S8	8,302	907	–
LAION-fMRI	sub-01	4,662	1,166	371
	sub-03	4,662	1,167	370
	sub-05	4,661	1,167	371
	sub-06	4,663	1,166	371
	sub-07	4,662	1,165	371
GOD	S1–S5	1,200	50	–

NSD. We use the prepared NSD responses from the Universal Encoder data pipeline (Beliy et al., 2024). The preparation reads `betas_fithrf_GLMdenoise_RR` responses in fsaverage space and applies the Algonauts all-vertices visual-cortex mask in each hemisphere. Responses are standardized voxelwise within acquisition sessions and then averaged across available presentations of each image. Subjects S1, S2, S5, and S7 have three presentations per training and valid test image; S3, S4, S6, and S8 have one, two, or three. The packaged outputs contain 39,548 retained vertices for S1–S5 and S7, 39,198 for S6, and 39,511 for S8. We apply no additional target-voxel reliability threshold.

LAION-fMRI. We use the GLMsingle-tedana response estimates for subjects sub-01, sub-03, sub-05, sub-06, and sub-07. Responses are z-scored voxelwise within acquisition sessions, averaged by image, and projected to fsaverage7. Our fixed-four condition averages exactly four presentations per image for training and both evaluation splits; images with fewer than four presentations are excluded. Invalid or constant session voxels are set to zero, and nonfinite projected vertices are excluded. The bilateral visual-ROI-union masks retain 41,575, 37,744, 36,798, 35,842, and 41,892 vertices for these subjects, respectively. Candidate images come from the official training split. We evaluate the same fitted model on the official `random_0` ID test and the official `ood` test. The latter tests a stimulus-category shift: its 371 stimuli include illusions, Gabor patches, abstract shapes, unusual arrangements, and isolated cutouts, whereas training uses regular images. The repetition requirement leaves 370 or 371 OOD images per subject, as shown in Table T12.

GOD. We use all five subjects and the provided visual-cortex masks. The training set contains 1,200 images from 150 categories, with eight images per category and one presentation per image. The test set contains 50 images from 50 categories disjoint from the training categories, each presented 35 times (Horikawa & Kamitani, 2017). We use the existing preprocessed response arrays without further response normalization, and verify the stored test averages against the 35 individual presentations. The visual-cortex masks contain 4,466, 4,404, 4,643, 4,133, and 4,370 voxels for S1–S5, respectively.

C.5 CANDIDATE POOLS AND MATCHED COMPARISONS

For the primary experiments, we construct multiple candidate pools to assess sensitivity to the available image population. NSD and LAION-fMRI candidate pools contain 80% of the available training images, sampled without replacement. GOD uses category-stratified candidate pools of 960 images. Within each experiment, the same candidate pools are used across selection methods, so comparisons are matched with respect to the available images. GOD uses the same candidate pools and selected image sets for all five subjects because they share the same training stimuli. The main NSD, LAION-fMRI, and GOD comparisons use nine candidate pools. Some development and supporting analyses use three matched pools, as specified in their corresponding appendix sections and table captions. All reported adaptation runs use the same fixed training seed within each matched comparison.

C.6 EVALUATION METRICS AND STATISTICAL ANALYSIS

Prediction metrics. For each fitted encoder, Pearson correlation is computed between predicted and measured responses across evaluation images, separately for each retained target voxel. The reported fit-level score is the unweighted median of these voxelwise correlations. Zero-variance cases use a stabilized denominator and contribute zero correlation. Reported per-voxel arrays are required to be finite. For the supplementary NSD MSE analysis, we average squared prediction errors over all fixed-test image–voxel pairs in the units of the preprocessed, voxelwise-standardized responses.

Aggregation and worst-observed performance. For each method and image-set size, we first average the fit-level scores across candidate pools within each subject, then average subjects with equal weight. Subject-specific standard deviations summarize variation across candidate pools. For worst-observed performance, we take the minimum pool score separately within each subject and then average these minima across subjects. This statistic describes the least favorable outcomes among the evaluated pools; it is neither a population lower bound nor a guarantee for an unseen pool.

Paired inference. For each NSD comparison, we subtract the baseline score from the method score after averaging pools within each subject. The mean of these eight paired differences is the reported effect size. Positive differences favor the method for Pearson correlation; negative differences favor it for MSE. We compute two-sided paired sign-flip tests using the absolute studentized mean difference as the test statistic, enumerating all 2^8 subject-level sign assignments. Confidence intervals are 95% percentile intervals obtained from 100,000 bootstrap resamples of subjects with replacement. Method pairing is retained within each resampled subject. Candidate pools and voxels are not treated as independent replicates.

Multiple-comparison correction. We apply exact Westfall–Young step-down max-T correction within each declared experiment family. The primary NSD Pearson family contains the three pairwise comparisons among Random, Image-Space, and Brain-Space at five image-set sizes, for 15 endpoints, using the nine-pool subject summaries. The supplementary MSE analysis uses a separate 15-endpoint family based on its three-pool summaries. Each alternative-encoder experiment retains its original 30-endpoint family: six pairwise comparisons among Random, Image-Space, Universal-Encoder Brain-Space selection, and Transformer Brain-Space selection at five image-set sizes. The full family is retained even when a displayed table presents only a subset of these comparisons. Bootstrap intervals summarize individual effect estimates; multiplicity adjustment is applied to the reported p-values.

Figure error bars. The main cross-dataset figure uses the same nine-pool subject means as the corresponding tables. At each image-set size, we compute Cousineau–Morey within-subject standard errors, with the Morey correction for three repeated methods. Error bars show ± 2 such standard errors. These bars describe within-subject variation across methods; they are not pairwise confidence intervals or significance tests.

D Additional Results

D.1 Robustness across evaluation metrics

Our primary evaluation uses median voxelwise Pearson correlation. To test whether the observed improvement depends on this correlation-based measure, we additionally evaluate prediction accuracy using mean squared error (MSE). As shown in Table T13, Brain-Space selection also reduces prediction error relative to Random and Image-Space selection across the evaluated budgets. At $N = 20$, MSE decreases from 0.4595 with Random sampling to 0.4434 with Brain-Space selection.

Participant-level statistical tests are reported in Table T14. Brain-Space selection significantly reduces MSE relative to Random across all evaluated budgets, while its advantage over Image-Space is strongest at the smaller budgets. Together with the Pearson-correlation results, this shows that the benefit of Brain-Space selection is not specific to the primary evaluation metric.

Table T13: **NSD performance using mean squared error.** Mean squared error on the fixed NSD test set for all eight participants (lower is better). Values are computed in the preprocessed voxelwise standardized response space, first averaged across three matched candidate pools within each participant and then across participants equally. Image-Space uses CLIP full-token k-means and Brain-Space uses linear-NCSNR-weighted coverage.

Method	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
Random	.4595	.4405	.4292	.4279	.4204
Image-Space	.4518	.4357	.4272	.4266	.4196
Brain-Space	**.4434**	**.4290**	**.4218**	**.4231**	**.4168**

Table T14: **Participant-level statistical analysis of NSD MSE.** At each image-set size, the three matched candidate pools are first averaged within each of the eight participants. $\Delta\mathrm{MSE}$ denotes the paired method-minus-baseline difference; negative values favor the method. We report participant-bootstrap 95% confidence intervals, the number of participants with a negative difference (Wins), exact two-sided paired sign-flip p_{raw} values, and Westfall–Young step-down max-T family-wise corrected p_{maxT} values across the 15 MSE comparisons.

N	**Paired comparison**	$\Delta\mathrm{MSE}$ **[95% CI]**	**Wins**	p_{raw}	p_{maxT}
20	Brain-Space − Random	-.0161 [-.0202, -.0116]	8/8	0.008	0.008
	Brain-Space − Image-Space	-.0085 [-.0108, -.0058]	8/8	0.008	0.008
	Image-Space − Random	-.0077 [-.0098, -.0054]	8/8	0.008	0.008
40	Brain-Space − Random	-.0115 [-.0147, -.0085]	8/8	0.008	0.008
	Brain-Space − Image-Space	-.0067 [-.0091, -.0045]	8/8	0.008	0.008
	Image-Space − Random	-.0048 [-.0082, -.0012]	5/8	0.062	0.172
60	Brain-Space − Random	-.0073 [-.0100, -.0051]	8/8	0.008	0.008
	Brain-Space − Image-Space	-.0054 [-.0068, -.0037]	8/8	0.008	0.008
	Image-Space − Random	-.0019 [-.0045, .0006]	6/8	0.219	0.523
80	Brain-Space − Random	-.0048 [-.0069, -.0028]	8/8	0.008	0.023
	Brain-Space − Image-Space	-.0036 [-.0059, -.0015]	7/8	0.016	0.086
	Image-Space − Random	-.0012 [-.0044, .0014]	5/8	0.570	0.766
100	Brain-Space − Random	-.0036 [-.0049, -.0024]	8/8	0.008	0.008
	Brain-Space − Image-Space	-.0028 [-.0051, -.0003]	7/8	0.070	0.227
	Image-Space − Random	-.0008 [-.0029, .0018]	6/8	0.562	0.766

D.2 PARTICIPANT-LEVEL CONSISTENCY ON NSD

The improvements in the aggregate NSD results are also consistent across individual participants. Tables T15 and T16 report average and worst-observed performance across the nine candidate pools for each participant. Brain-Space selection improves over Random sampling for every participant at every evaluated budget, with the largest gains in the smallest-data regime. Participant-level statistical tests for the primary comparisons are reported in Table T17.

Table T15: **Participant-level NSD performance across candidate pools.** Median voxelwise Pearson correlation on the fixed NSD test set for Random, Image-Space, and Brain-Space selection (higher is better). Each cell reports the mean ± sample standard deviation across nine matched candidate pools for that participant. The standard deviation reflects sensitivity to candidate-pool composition rather than biological variability. Bold indicates the highest mean for each participant and image-set size.

Participant	**Method**	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
S1	Random	.280 ± .017	.358 ± .014	.407 ± .013	.437 ± .008	.451 ± .007
	Image-Space	.303 ± .013	.378 ± .008	.410 ± .010	.436 ± .009	.449 ± .005
	Brain-Space	**.335 ± .013**	**.399 ± .008**	**.430 ± .005**	**.444 ± .007**	**.456 ± .005**
S2	Random	.301 ± .024	.378 ± .014	.425 ± .014	.447 ± .007	.461 ± .006
	Image-Space	.320 ± .010	.395 ± .012	.434 ± .011	.447 ± .010	.465 ± .005
	Brain-Space	**.358 ± .012**	**.417 ± .007**	**.444 ± .009**	**.457 ± .004**	**.469 ± .002**
S3	Random	.209 ± .033	.264 ± .025	.299 ± .017	.312 ± .011	.331 ± .009
	Image-Space	.247 ± .009	.292 ± .014	.308 ± .011	.323 ± .009	.339 ± .009
	Brain-Space	**.280 ± .010**	**.321 ± .006**	**.335 ± .004**	**.341 ± .004**	**.350 ± .003**
S4	Random	.208 ± .012	.269 ± .009	.298 ± .007	.311 ± .007	.324 ± .008
	Image-Space	.238 ± .016	.265 ± .015	.300 ± .006	.310 ± .007	.325 ± .008
	Brain-Space	**.266 ± .010**	**.309 ± .009**	**.326 ± .008**	**.330 ± .006**	**.342 ± .005**
S5	Random	.338 ± .021	.415 ± .014	.449 ± .011	.463 ± .008	.480 ± .006
	Image-Space	.375 ± .014	.425 ± .007	.456 ± .010	.466 ± .011	.482 ± .006
	Brain-Space	**.401 ± .007**	**.453 ± .005**	**.472 ± .005**	**.480 ± .003**	**.493 ± .003**
S6	Random	.195 ± .024	.249 ± .012	.276 ± .009	.292 ± .009	.304 ± .011
	Image-Space	.209 ± .014	.261 ± .011	.283 ± .009	.303 ± .008	.314 ± .009
	Brain-Space	**.246 ± .009**	**.281 ± .009**	**.300 ± .007**	**.307 ± .004**	**.317 ± .006**
S7	Random	.182 ± .021	.239 ± .029	.270 ± .025	.288 ± .017	.309 ± .014
	Image-Space	.216 ± .020	.264 ± .021	.284 ± .012	.306 ± .015	.320 ± .011
	Brain-Space	**.257 ± .008**	**.295 ± .011**	**.317 ± .014**	**.325 ± .009**	**.337 ± .008**
S8	Random	.130 ± .020	.168 ± .013	.185 ± .010	.194 ± .009	.205 ± .008
	Image-Space	.143 ± .014	.169 ± .010	.186 ± .012	.196 ± .010	.207 ± .006
	Brain-Space	**.160 ± .012**	**.190 ± .007**	**.202 ± .009**	**.204 ± .008**	**.215 ± .008**

Table T16: **Participant-level NSD worst-observed performance.** Median voxelwise Pearson correlation on the fixed NSD test set for Random, Image-Space, and Brain-Space selection (higher is better). Each cell reports the minimum performance observed across the nine matched candidate pools for that participant. This is a descriptive robustness measure over the evaluated pools, not a population lower bound. Bold indicates the highest worst-observed value for each participant and image-set size.

Participant	Method	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
	Random	.250	.334	.388	.427	.441
S1	Image-Space	.288	.365	.396	.423	.436
	Brain-Space	**.311**	**.388**	**.422**	**.436**	**.451**
	Random	.261	.354	.400	.437	.454
S2	Image-Space	.302	.367	.418	.428	.456
	Brain-Space	**.339**	**.406**	**.425**	**.450**	**.464**
	Random	.162	.214	.277	.295	.315
S3	Image-Space	.230	.268	.293	.306	.325
	Brain-Space	**.269**	**.310**	**.328**	**.335**	**.347**
	Random	.185	.255	.289	.301	.310
S4	Image-Space	.221	.246	.289	.301	.317
	Brain-Space	**.253**	**.296**	**.315**	**.322**	**.335**
	Random	.303	.396	.432	.454	.472
S5	Image-Space	.359	.414	.434	.438	.473
	Brain-Space	**.392**	**.447**	**.461**	**.474**	**.488**
	Random	.156	.226	.266	.273	.285
S6	Image-Space	.185	.242	.271	.295	.299
	Brain-Space	**.230**	**.269**	**.291**	**.302**	**.306**
	Random	.143	.167	.217	.250	.282
S7	Image-Space	.182	.220	.256	.278	.302
	Brain-Space	**.250**	**.273**	**.288**	**.314**	**.324**
	Random	.086	.140	.166	.178	.193
S8	Image-Space	.117	.154	.168	.176	.198
	Brain-Space	**.140**	**.179**	**.189**	**.189**	**.203**

Table T17: **Participant-level statistical tests for the primary NSD comparisons.** At each image-set size, the nine matched candidate pools are first averaged within each of the eight participants. Δr denotes the paired method-minus-baseline difference in median voxelwise Pearson correlation; positive values favor the method. We report participant-bootstrap 95% confidence intervals, the number of participants with a positive difference (Wins), exact two-sided paired sign-flip p_{raw} values, and Westfall–Young step-down max-T family-wise corrected p_{maxT} values across the 15 Pearson comparisons.

N	Paired comparison	Δr [95% CI]	Wins	p_{raw}	p_{maxT}
	Brain Coverage − Random	.0578 [.0480, .0661]	8/8	0.008	0.008
20	Brain Coverage − CLIP Full k-means	.0318 [.0263, .0365]	8/8	0.008	0.008
	CLIP Full k-means − Random	.0260 [.0194, .0325]	8/8	0.008	0.008
	Brain Coverage − Random	.0406 [.0331, .0482]	8/8	0.008	0.008
40	Brain Coverage − CLIP Full k-means	.0270 [.0223, .0326]	8/8	0.008	0.008
	CLIP Full k-means − Random	.0136 [.0062, .0207]	7/8	0.023	0.031
	Brain Coverage − Random	.0273 [.0215, .0343]	8/8	0.008	0.008
60	Brain Coverage − CLIP Full k-means	.0207 [.0160, .0256]	8/8	0.008	0.008
	CLIP Full k-means − Random	.0066 [.0036, .0097]	8/8	0.008	0.016
	Brain Coverage − Random	.0182 [.0119, .0255]	8/8	0.008	0.008
80	Brain Coverage − CLIP Full k-means	.0128 [.0089, .0168]	8/8	0.008	0.008
	CLIP Full k-means − Random	.0054 [.0010, .0103]	5/8	0.070	0.070
	Brain Coverage − Random	.0144 [.0100, .0193]	8/8	0.008	0.008
100	Brain Coverage − CLIP Full k-means	.0098 [.0064, .0133]	8/8	0.008	0.008
	CLIP Full k-means − Random	.0046 [.0018, .0076]	7/8	0.023	0.031

D.3 Performance across visual ROIs at larger image-set sizes

Figure S4 extends the main-text ROI analysis to image-set sizes of $N = 60$ and $N = 100$. We compare Random sampling, Image-Space selection, and Brain-Space selection within individual visual ROIs. For each method, the same full-brain-selected image sets are evaluated across all ROIs. This analysis examines how encoding performance is distributed across visual regions as the image budget increases.

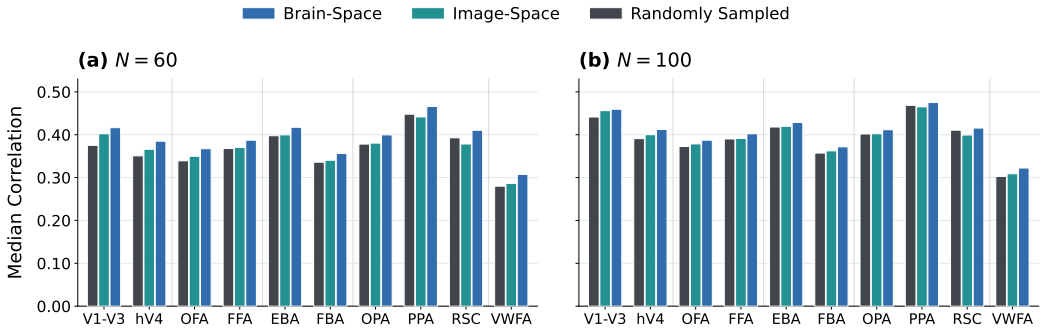


Figure S4: **Performance across visual ROIs at larger image-set sizes.** Median voxelwise Pearson correlation on the NSD test set for Random sampling, Image-Space selection, and Brain-Space selection at (a) $N = 60$ and (b) $N = 100$. For each method, the same full-brain-selected image sets are evaluated separately within each visual ROI. V1–V3 combines the ventral and dorsal subdivisions of V1, V2, and V3. Participants are weighted equally.

D.4 PARTICIPANT-LEVEL RESULTS ON LAION-FMRI

Tables T18 and T19 report participant-level average and worst-observed performance on the LAION-fMRI in-distribution (ID) test set, while Tables T20 and T21 report the corresponding out-of-distribution (OOD) results. Both evaluations use the same stimulus-selection and adaptation procedure, differing only in the test split. Brain-Space selection improves over Random sampling for every participant at every evaluated budget, in both average and worst-observed performance on both test sets. The gains are generally largest at the smallest budgets, showing that the aggregate cross-dataset improvements are consistent across individual participants.

Table T18: **Participant-level average performance on LAION-fMRI (ID).** Median voxelwise Pearson correlation on the in-distribution test set for Random sampling, Image-Space selection (CLIP ViT-L/14 full-token k-means), and Brain-Space selection (NCSNR-weighted coverage in NSD-predicted fMRI space). For each participant, values report the mean $\pm$ sample standard deviation across nine candidate pools. The Average block reports the equal-weighted mean of participant means. Responses are averaged over four repetitions.

Participant	**Method**	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
S1	Random	.213 ± .028	.289 ± .027	.338 ± .020	.365 ± .013	.373 ± .013
	Image-Space	.224 ± .029	.284 ± .021	.329 ± .009	.351 ± .009	.370 ± .010
	Brain-Space	**.286 ± .014**	**.343 ± .013**	**.378 ± .007**	**.395 ± .007**	**.402 ± .005**
S3	Random	.254 ± .043	.357 ± .023	.413 ± .013	.450 ± .012	.463 ± .007
	Image-Space	.260 ± .035	.355 ± .015	.400 ± .012	.433 ± .015	.444 ± .010
	Brain-Space	**.343 ± .017**	**.409 ± .009**	**.449 ± .007**	**.476 ± .007**	**.488 ± .007**
S5	Random	.255 ± .046	.340 ± .023	.393 ± .011	.424 ± .008	.437 ± .007
	Image-Space	.294 ± .027	.359 ± .015	.392 ± .013	.427 ± .008	.439 ± .012
	Brain-Space	**.341 ± .019**	**.395 ± .015**	**.431 ± .012**	**.447 ± .009**	**.456 ± .009**
S6	Random	.224 ± .033	.293 ± .023	.330 ± .019	.353 ± .013	.365 ± .014
	Image-Space	.242 ± .021	.304 ± .018	.345 ± .014	.363 ± .014	.378 ± .005
	Brain-Space	**.310 ± .021**	**.358 ± .007**	**.385 ± .006**	**.396 ± .004**	**.404 ± .002**
S7	Random	.245 ± .029	.313 ± .026	.354 ± .020	.379 ± .015	.388 ± .017
	Image-Space	.257 ± .022	.320 ± .011	.365 ± .012	.384 ± .017	.398 ± .006
	Brain-Space	**.330 ± .007**	**.376 ± .006**	**.400 ± .004**	**.412 ± .002**	**.417 ± .004**
Average	Random	.238	.319	.365	.394	.405
	Image-Space	.255	.325	.366	.392	.406
	Brain-Space	**.322**	**.376**	**.409**	**.425**	**.433**

Table T19: **Participant-level worst-observed performance on LAION-fMRI (ID).** Median voxelwise Pearson correlation on the in-distribution test set for Random sampling, Image-Space selection (CLIP ViT-L/14 full-token k-means), and Brain-Space selection (NCSNR-weighted coverage in NSD-predicted fMRI space). For each participant, values report the minimum performance observed across nine candidate pools. The Average block reports the equal-weighted mean of participant minima. These minima describe robustness across the evaluated pools and are not population lower bounds. Responses are averaged over four repetitions.

Participant	Method	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
	Random	.168	.233	.298	.342	.352
S1	Image-Space	.177	.260	.317	.337	.352
	Brain-Space	**.265**	**.325**	**.368**	**.387**	**.392**
	Random	.200	.326	.390	.434	.451
S3	Image-Space	.213	.331	.389	.410	.429
	Brain-Space	**.306**	**.394**	**.443**	**.463**	**.473**
	Random	.190	.301	.376	.412	.430
S5	Image-Space	.260	.337	.372	.409	.421
	Brain-Space	**.296**	**.366**	**.407**	**.429**	**.437**
	Random	.173	.264	.298	.331	.344
S6	Image-Space	.202	.279	.321	.342	.372
	Brain-Space	**.273**	**.349**	**.372**	**.390**	**.400**
	Random	.207	.257	.315	.350	.355
S7	Image-Space	.228	.304	.352	.353	.391
	Brain-Space	**.320**	**.370**	**.394**	**.410**	**.411**
	Random	.188	.276	.335	.374	.386
Average	Image-Space	.216	.302	.350	.370	.393
	Brain-Space	**.292**	**.361**	**.397**	**.416**	**.423**

Table T20: **Participant-level average performance on LAION-fMRI (OOD).** Median voxelwise Pearson correlation on the out-of-distribution test set for Random sampling, Image-Space selection (CLIP ViT-L/14 full-token k-means), and Brain-Space selection (NCSNR-weighted coverage in NSD-predicted fMRI space). For each participant, values report the mean $\pm$ sample standard deviation across nine candidate pools. The Average block reports the equal-weighted mean of participant means. Responses are averaged over four repetitions. Bold indicates the highest mean within each participant or Average block at each image-set size.

Participant	Method	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
	Random	.195 $\pm$ .031	.261 $\pm$ .020	.291 $\pm$ .026	.310 $\pm$ .020	.317 $\pm$ .022
S1	Image-Space	.232 $\pm$ .036	.267 $\pm$ .020	.299 $\pm$ .010	.311 $\pm$ .013	.328 $\pm$ .009
	Brain-Space	**.259 $\pm$ .020**	**.309 $\pm$ .017**	**.336 $\pm$ .014**	**.344 $\pm$ .012**	**.349 $\pm$ .010**
	Random	.268 $\pm$ .044	.372 $\pm$ .024	.410 $\pm$ .027	.436 $\pm$ .024	.454 $\pm$ .020
S3	Image-Space	.328 $\pm$ .041	.389 $\pm$ .021	.415 $\pm$ .027	.452 $\pm$ .016	.455 $\pm$ .013
	Brain-Space	**.404 $\pm$ .013**	**.447 $\pm$ .006**	**.466 $\pm$ .012**	**.476 $\pm$ .013**	**.486 $\pm$ .009**
	Random	.281 $\pm$ .054	.374 $\pm$ .023	.416 $\pm$ .016	.427 $\pm$ .029	.445 $\pm$ .010
S5	Image-Space	**.360 $\pm$ .027**	.398 $\pm$ .025	.420 $\pm$ .013	.448 $\pm$ .015	.454 $\pm$ .011
	Brain-Space	**.360 $\pm$ .026**	**.408 $\pm$ .026**	**.440 $\pm$ .021**	**.455 $\pm$ .017**	**.463 $\pm$ .013**
	Random	.215 $\pm$ .076	.275 $\pm$ .054	.302 $\pm$ .046	.328 $\pm$ .035	.343 $\pm$ .032
S6	Image-Space	.246 $\pm$ .032	.303 $\pm$ .014	.329 $\pm$ .025	.339 $\pm$ .030	.369 $\pm$ .009
	Brain-Space	**.311 $\pm$ .027**	**.359 $\pm$ .016**	**.379 $\pm$ .013**	**.392 $\pm$ .011**	**.400 $\pm$ .010**
	Random	.224 $\pm$ .055	.306 $\pm$ .061	.345 $\pm$ .044	.373 $\pm$ .032	.381 $\pm$ .037
S7	Image-Space	.293 $\pm$ .032	.344 $\pm$ .022	.384 $\pm$ .013	.386 $\pm$ .023	.402 $\pm$ .017
	Brain-Space	**.335 $\pm$ .016**	**.371 $\pm$ .010**	**.394 $\pm$ .013**	**.393 $\pm$ .017**	**.405 $\pm$ .015**
	Random	.237	.318	.353	.375	.388
Average	Image-Space	.292	.340	.369	.387	.402
	Brain-Space	**.334**	**.379**	**.403**	**.412**	**.420**

Table T21: **Participant-level worst-observed performance on LAION-fMRI (OOD).** Median voxelwise Pearson correlation on the out-of-distribution test set for Random sampling, Image-Space selection (CLIP ViT-L/14 full-token k-means), and Brain-Space selection (NCSNR-weighted coverage in NSD-predicted fMRI space). For each participant, values report the minimum performance observed across nine candidate pools. The Average block reports the equal-weighted mean of participant minima. These minima describe robustness across the evaluated pools and are not population lower bounds. Responses are averaged over four repetitions. Bold indicates the highest value within each participant or Average block at each image-set size.

Participant	**Method**	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
S1	Random	.150	.220	.250	.284	.279
	Image-Space	.166	.230	.282	.286	.314
	Brain-Space	**.220**	**.277**	**.314**	**.329**	**.336**
S3	Random	.216	.330	.343	.379	.405
	Image-Space	.265	.349	.370	.431	.434
	Brain-Space	**.387**	**.437**	**.451**	**.461**	**.472**
S5	Random	.206	.333	.384	.360	.434
	Image-Space	.318	**.353**	.401	.420	.434
	Brain-Space	**.323**	.351	**.402**	**.426**	**.441**
S6	Random	.087	.206	.240	.274	.295
	Image-Space	.204	.277	.303	.280	.355
	Brain-Space	**.263**	**.338**	**.366**	**.375**	**.380**
S7	Random	.141	.152	.238	.298	.291
	Image-Space	.256	.319	.367	.346	**.383**
	Brain-Space	**.307**	**.359**	**.381**	**.366**	.378
Average	Random	.160	.248	.291	.319	.341
	Image-Space	.242	.306	.344	.353	.384
	Brain-Space	**.300**	**.353**	**.383**	**.391**	**.401**

D.5 PARTICIPANT-LEVEL RESULTS ON GOD

Tables T22 and T23 report participant-level average and worst-observed performance on the Generic Object Decoding (GOD) dataset. Because all five participants share the same candidate images, each selection method uses the same selected image set across participants for a given candidate pool and budget. Brain-Space selection improves average performance over Random sampling for every participant at every evaluated budget. Worst-observed performance also improves in nearly all participant–budget comparisons with Random sampling. The advantage over Image-Space selection is less uniform across participants, although Brain-Space achieves the highest participant-averaged performance at every evaluated budget under both summaries.

Table T22: **Participant-level average performance on GOD.** Median voxelwise Pearson correlation on the held-out test set for Random sampling, Image-Space selection (CLIP ViT-L/14 full-token k-means), and Brain-Space selection (NCSNR-weighted coverage in NSD-predicted fMRI space). For each participant, values report the mean ± sample standard deviation across nine candidate pools. The Average block reports the equal-weighted mean of participant means.

Participant	**Method**	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
S1	Random	.032 ± .020	.037 ± .022	.045 ± .018	.056 ± .016	.063 ± .013
	Image-Space	**.039 ± .028**	.040 ± .024	.046 ± .032	.051 ± .011	**.071 ± .020**
	Brain-Space	.037 ± .010	**.043 ± .012**	**.054 ± .014**	**.062 ± .010**	.065 ± .016
S2	Random	.068 ± .018	.086 ± .027	.091 ± .027	.105 ± .025	.117 ± .025
	Image-Space	.041 ± .031	.048 ± .034	.073 ± .041	.073 ± .027	.097 ± .032
	Brain-Space	**.088 ± .025**	**.112 ± .024**	**.119 ± .021**	**.127 ± .024**	**.127 ± .017**
S3	Random	.099 ± .020	.139 ± .028	.165 ± .029	.177 ± .026	.202 ± .021
	Image-Space	.098 ± .027	.146 ± .025	.160 ± .016	.181 ± .012	.205 ± .017
	Brain-Space	**.133 ± .014**	**.168 ± .014**	**.185 ± .018**	**.190 ± .012**	**.212 ± .012**
S4	Random	.087 ± .022	.132 ± .022	.155 ± .015	.169 ± .015	.187 ± .020
	Image-Space	.081 ± .023	.130 ± .025	.126 ± .028	.158 ± .020	.182 ± .013
	Brain-Space	**.128 ± .017**	**.162 ± .020**	**.173 ± .018**	**.183 ± .015**	**.197 ± .013**
S5	Random	.051 ± .028	.077 ± .019	.090 ± .021	.097 ± .026	.103 ± .027
	Image-Space	.057 ± .021	.082 ± .017	.094 ± .026	.112 ± .018	.122 ± .011
	Brain-Space	**.071 ± .029**	**.103 ± .016**	**.106 ± .015**	**.113 ± .015**	**.125 ± .013**
Average	Random	.067	.094	.109	.121	.134
	Image-Space	.063	.089	.100	.115	.136
	Brain-Space	**.092**	**.117**	**.127**	**.135**	**.145**

Table T23: **Participant-level worst-observed performance on GOD.** Median voxelwise Pearson correlation on the held-out test set for Random sampling, Image-Space selection (CLIP ViT-L/14 full-token k-means), and Brain-Space selection (NCSNR-weighted coverage in NSD-predicted fMRI space). For each participant, values report the minimum performance observed across nine candidate pools. The Average block reports the equal-weighted mean of participant minima. These minima describe robustness across the evaluated pools and are not population lower bounds.

Participant	**Method**	$N = 20$	$N = 40$	$N = 60$	$N = 80$	$N = 100$
S1	Random	-.010	-.009	.004	.032	**.046**
	Image-Space	-.004	.000	.010	.035	.039
	Brain-Space	**.021**	**.017**	**.022**	**.043**	.043
S2	Random	.041	.050	.059	.070	.072
	Image-Space	-.004	.004	.032	.033	.067
	Brain-Space	**.051**	**.074**	**.089**	**.104**	**.109**
S3	Random	.062	.088	.126	.142	.178
	Image-Space	.055	.094	.127	.161	.182
	Brain-Space	**.107**	**.144**	**.152**	**.171**	**.200**
S4	Random	.043	.111	.136	.145	.164
	Image-Space	.045	.098	.082	.123	.162
	Brain-Space	**.090**	**.136**	**.151**	**.164**	**.182**
S5	Random	-.007	.056	.058	.071	.073
	Image-Space	.026	.049	.051	.082	.097
	Brain-Space	**.034**	**.074**	**.086**	**.098**	**.106**
Average	Random	.026	.059	.077	.092	.106
	Image-Space	.024	.049	.060	.087	.109
	Brain-Space	**.061**	**.089**	**.100**	**.116**	**.128**

D.6 PARTICIPANT-LEVEL STATISTICAL TESTS FOR ENCODER GENERALIZATION

We additionally perform participant-level paired inference for the two encoder-generalization experiments. Table T24 reports results when the independently trained encoder of Adeli et al. (2025) is used to construct the Brain-Space representation for stimulus selection, while the Universal Encoder is retained for downstream evaluation. Brain-Space selection outperforms Random sampling at every evaluated budget after multiple-comparison correction. Improvements over Image-Space are also significant at $N = 20$ and $N = 60$, while the remaining budgets show positive but non-significant differences after correction.

Table T25 reports the complementary experiment, in which image sets are selected using the Universal Encoder and evaluated using the independently trained encoder of Adeli et al. (2025). Brain-Space selection significantly outperforms Random sampling at every evaluated budget and Image-Space selection at all budgets except $N = 40$ after multiple-comparison correction.

Table T24: **Participant-level inference for the selection-encoder experiment.** Brain-Space image sets are selected using responses predicted by the independently trained fMRI encoder of Adeli et al. (2025), while all selected sets are evaluated using the Universal Encoder. At each image-set size, repeated fits are first averaged within each of the eight NSD participants. Δr denotes the paired difference in median voxelwise Pearson correlation, with a participant-clustered bootstrap 95% confidence interval. Wins reports the number of participants with a positive paired difference. p_{raw} is the exact two-sided paired sign-flip p-value, and p_{maxT} is the Westfall–Young step-down max-T family-wise corrected p-value.

N	**Paired comparison**	Δr **[95% CI]**	**Wins**	p_{raw}	p_{maxT}
20	Adeli Brain-Space − Random	.0497 [.0362, .0616]	8/8	0.008	0.008
	Adeli Brain-Space − Image-Space	.0227 [.0149, .0307]	8/8	0.008	0.008
40	Adeli Brain-Space − Random	.0313 [.0244, .0378]	8/8	0.008	0.008
	Adeli Brain-Space − Image-Space	.0161 [.0051, .0275]	6/8	0.031	0.234
60	Adeli Brain-Space − Random	.0224 [.0180, .0268]	8/8	0.008	0.008
	Adeli Brain-Space − Image-Space	.0146 [.0094, .0201]	8/8	0.008	0.008
80	Adeli Brain-Space − Random	.0108 [.0084, .0130]	8/8	0.008	0.008
	Adeli Brain-Space − Image-Space	.0060 [.0015, .0103]	7/8	0.039	0.258
100	Adeli Brain-Space − Random	.0083 [.0059, .0103]	8/8	0.008	0.008
	Adeli Brain-Space − Image-Space	.0056 [-.0000, .0111]	5/8	0.125	0.516

Table T25: **Participant-level inference for the downstream-encoder experiment.** Image sets are selected using the Universal Encoder and evaluated using the independently trained fMRI encoder of Adeli et al. (2025). At each image-set size, repeated fits are first averaged within each of the eight NSD participants. Δr denotes the paired difference in median voxelwise Pearson correlation, with a participant-clustered bootstrap 95% confidence interval. Wins reports the number of participants with a positive paired difference. p_{raw} is the exact two-sided paired sign-flip p-value, and p_{maxT} is the Westfall–Young step-down max-T family-wise corrected p-value.

N	**Paired comparison**	Δr **[95% CI]**	**Wins**	p_{raw}	p_{maxT}
20	Universal Brain-Space − Random	.0463 [.0289, .0599]	7/8	0.016	0.023
	Universal Brain-Space − Image-Space	.0348 [.0208, .0475]	7/8	0.016	0.047
	Image-Space − Random	.0115 [.0024, .0223]	5/8	0.070	0.289
40	Universal Brain-Space − Random	.0373 [.0254, .0461]	8/8	0.008	0.016
	Universal Brain-Space − Image-Space	.0251 [.0137, .0343]	7/8	0.016	0.055
	Image-Space − Random	.0121 [.0071, .0185]	8/8	0.008	0.070
60	Universal Brain-Space − Random	.0324 [.0228, .0406]	8/8	0.008	0.016
	Universal Brain-Space − Image-Space	.0239 [.0173, .0300]	8/8	0.008	0.016
	Image-Space − Random	.0085 [.0025, .0137]	6/8	0.039	0.203
80	Universal Brain-Space − Random	.0260 [.0183, .0329]	8/8	0.008	0.016
	Universal Brain-Space − Image-Space	.0179 [.0139, .0225]	8/8	0.008	0.008
	Image-Space − Random	.0081 [.0031, .0128]	7/8	0.031	0.125
100	Universal Brain-Space − Random	.0221 [.0172, .0263]	8/8	0.008	0.008
	Universal Brain-Space − Image-Space	.0198 [.0139, .0251]	8/8	0.008	0.016
	Image-Space − Random	.0023 [-.0034, .0087]	4/8	0.531	0.781