CONFORM: A Project to Create Crowd-Sourced Open Neuroscience fMRI Foundation Models

Anonymous Author(s)
Affiliation
Address
email

Abstract

We propose CONFORM (Crowd-Sourced Open Neuroscience fMRI Foundation Model), a project that will bring together recent advances in neural data processing and analysis with a novel, crowd-sourced infrastructure. This transformative approach will overcome several current challenges in creating a foundational human fMRI model for vision: collecting massive amounts of data from a handful of participants is neither scalable nor sustainable; the number of participants is small for such datasets; stimulus diversity is limited; and generalizability to different populations is poor. CONFORM will overcome these limitations by combining a powerful generative denoising method (PSN), a scalable framework for aggregating existing fMRI datasets (MOSAIC), and a meta-learning model that enables generalization with much smaller data from new participants (BraInCoRL). Our collaborative effort will produce models built on unprecedented scale and diversity—ultimately with hundreds of participants and hundreds of thousands of naturalistic image and movie stimuli—and provide the tools for continuous expansion of the underlying dataset. This "crowd-sourced" approach will allow many more researchers to leverage state-of-the-art NeuroAI methods using the scale of data they typically collect, democratizing access to powerful models and accelerating scientific discovery for a wide range of neuroscientific domains and populations.

1 Background and Introduction

Creating a foundational human fMRI model is a critical next step for extending modern neuroAI [1]. To achieve this, the model must generalize across both individuals and tasks, which requires a large volume of data with many participants, observations, and diverse stimuli. Historically, a significant impediment has been that most fMRI studies have small sample sizes and a low number of observations per session; the latter also leading to poor stimulus diversity. As a result, typical fMRI experiments sample only a tiny fraction of the human population and the vast space of real-world visual, auditory, or linguistic inputs. These limitations impeded efforts to draw robust conclusions from fMRI data and to integrate insights from modern AI systems into our understanding of the human brain—a challenge that is exacerbated by the inherently noisy BOLD signal.

In visual neuroscience, a first step in meeting this challenge has already been taken through the collection of large-scale fMRI datasets, which typically include brain responses from a small number of participants each scanned over many repeated sessions (15-40 hours-long sessions), who view a large number of stimuli (5000-10,000 stimuli per participant) [2, 3, 4, 5]. This approach of "deeply sampling" a small number of participants increases the statistical power of experiments [6, 7], and enables powerful parameter-rich, within-subject models. While this approach of collecting large datasets from small groups of participants has led to hundreds of publications and impactful discoveries, even this strategy is neither sustainable nor scalable for both scientific and practical reasons:

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

1. Successful data collection at this scale depends on *heroic* efforts by both experimenters and participants. The time commitment and scheduling complexities are onerous: participants, experimenters, and scanners must remain consistently healthy and available (e.g., in both the BOLD5000 and NSD datasets at least one participant failed to complete the study [2, 3]; in the THINGS dataset one participant was canceled due to "technical issues" [4]).
2. Even with this extraordinary amount of effort, data was collected from only 3-8 participants – a small number that does not support the hoped-for population diversity expected of human neural foundation models.
3. Stimulus diversity is necessarily limited by small participant pools and the need for stimulus repeats and/or overlap across participants [8]. Even within a single recurring participant, only a limited number of observations are possible.
   Moreover, controlled tasks and stimulus selection methods have further reduced diversity in the visual images included in each dataset: NSD uses only COCO images (only 80 object categories [9], which leave gaps in many regions of natural image space [10]), BOLD5000 uses COCO as well as SUN [11] and ImageNet [12] images, and THINGS uses a larger number of "concepts", but depicted as single cropped objects that show little context [4].
4. Creating the infrastructure for data management and distribution is a considerable technical challenge. Short-term it requires a robust and replicable data processing pipeline and a reliable platform for data distribution. Long-term it requires stability—years later the distribution website should remain readily accessible.
5. The monetary costs of collecting data can present a challenge to any single lab (e.g., five participants across 25 x one hour scans could easily cost on the order of $100,000) and risks over-representing the interests of the small number of labs with the necessary resources.

Despite their increased scale relative to standard fMRI studies, these datasets still present significant challenges in the construction of NeuroAI models. The number of observations and participants is still small for purposes of model training, and data quality is dependent on preprocessing methods. More importantly, prediction accuracy and decoding performance are typically high only when trained and tested within the same participant—due to inherent structural and functional differences between individual brains and, at present, weak methods for generalizing across them. Consequently, when models are applied across participants, even within the same study, their performance and decoding capabilities decrease dramatically.

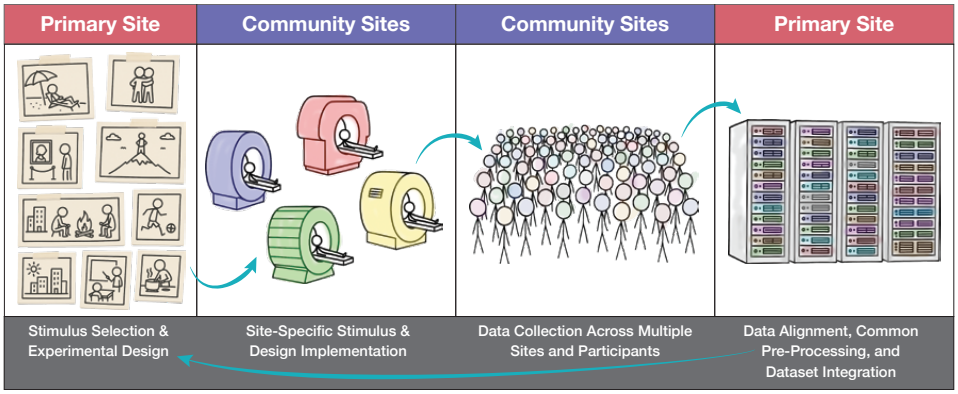

Figure 1: **CONFORM workflow.** A single, optimized experimental design is distributed to multiple sites for data collection. The collected data is then centralized for preprocessing, alignment, and integration into a foundational dataset. This process creates a continuous feedback loop, allowing the dataset to grow in size and diversity, which informs future experimental design and provides the basis for a strong foundation model.

2 Towards a dynamic foundation model for vision fMRI

We propose CONFORM (Crowd-Sourced Open Neuroscience fMRI Foundation Model)–a strategy for building foundational human vision fMRI models through community-contributed datasets and models. Following previous efforts in systems neuroscience [13], we propose to leverage multi-site crowd-sourcing to enable collection of larger and more diverse datasets, along with new computational advances to facilitate coherent analysis. As detailed below, the building blocks of CONFORM are already in place, spanning four key domains:

1. *A larger-scale and highly diverse dataset* that aggregates close to 100 participants and 100,000s of natural scenes depicting 1000's of object categories/concepts in context. "MOSAIC" [14] is a scalable framework for combining extant fMRI datasets [2, 3, 4, 5, 15, 16, 17, 18], using common preprocessing and registration, into a single, extremely large-scale and extensible vision dataset.

2. *Higher data quality* through an enhanced preprocessing pipeline to improve the signal-to-noise ratio of measured BOLD responses. Building on *GLMSingle* [8] and Generative Modeling of Signal and Noise (*GSN [19]*), we will develop *PSN* (Partitioning of Signal and Noise)—a powerful, generative low-rank denoising method that optimally separates signal from noise in neural data, outperforming trial-averaging and PCA, especially when noise is structured or complex (as in fMRI).

3. *Enhanced generalization* to new participants from outside-of-dataset studies using "BraInCoRL"—a meta-learned in-context foundation model that enables generalization using only a small amount of additional data [20].

4. *Crowd-sourcing infrastructure* to support the continuous integration of data from new studies across unique participants and data collection sites.

Building on these methodological advances and the lessons learned from distributed large-scale fMRI datasets [2, 3, 4, 5, 15], CONFORM will be a unique collaborative modeling strategy that will enable the creation of large-scale vision foundation fMRI models on datasets with improved signal quality, more participants, greater stimulus diversity, and which, critically, generalizes to new participants and studies in low data regimes. Longer term— across labs, participants, and MRI systems, we further propose a "crowd-sourced" community-driven effort to collect and integrate new data, thereby continuously improving the models. Given the challenges of collecting ever-larger and more diverse datasets at a single site, we suggest that crowd-sourcing is the only tenable solution for building appropriate-scale, truly foundational neural datasets. However, developing a viable crowd-sourcing infrastructure at this scale remains an unsolved challenge with a very high risk/reward tradeoff.

We are taking on this challenge by integrating and further developing recent advances in fMRI preprocessing, data aggregation, and generalization. CONFORM will also include the infrastructure for continuously expanding the dataset's size and the diversity of its stimuli [21]. Our project will use a two-pronged approach for data contributions: locally directed and globally directed.

The *locally directed* model is straightforward: the CONFORM distribution website will also accept contributions. In contrast to other neural data repositories [22], we will provide detailed specifications for the acceptable designs, stimuli, tasks, and data formats to ensure submissions can be seamlessly integrated into CONFORM with high data quality. One attractive aspect of a locally directed model is that CONFORM may be able to re-purpose extant data that was already collected for a different purpose, thus giving new life to data that may have been otherwise dormant for years. At the same time, processing all available public data is not feasible. As an alternative, we will facilitate researchers re-analyzing their datasets with our pipeline. Our goal with the locally directed model is to be as inclusive as possible with stimuli and tasks, even with necessary limitations.

The *globally directed* model is more ambitious and forward-looking, and offers a greater potential payoff. We will provide a complete, turn-key study design to participating research sites, streamlining the data collection process (Fig. 1). We will optimize the selection of stimulus images to achieve the best possible distribution of images within natural image space across many participants [10]. We will also optimize for repeated stimuli and partial stimulus overlap across the population. Similarly, we will optimize the study design with respect to scanning parameters and trial structure. Collaborators will be able to specify both the length of scan sessions and the total number of participants they contribute. They will then be provided with complete scan protocols, experimental control files, and stimulus images. An interface on the same website used for distribution will allow them to download these files and upload their collected data for incorporation into the dataset.

CONFORM's framework towards a scalable foundation fMRI model will enable powerful insights into human vision. Datasets within CONFORM will continue to grow in size and stimulus diversity as the community contributes data. Critically, the resultant models will achieve improved generalization to new participants across diverse subpopulations, requiring only a relatively small amount of data per individual. As such, CONFORM will dramatically broaden the accessibility of NeuroAI methods, empowering researchers in a much wider range of scientific domains to make new discoveries.

2.1 Improving data quality—PSN

The recently introduced *GLMSingle* preprocessing pipeline dramatically improves the signal-to-noise ratio of measured BOLD responses acquired using standard fMRI methods [8]. In parallel, the Generative Modeling of Signal and Noise technique (*GSN*, [19]) has established a new paradigm for accurately estimating the parameters of the signal and noise distributions that give rise to the observed measurements. We will build upon the GLMSingle and GSN approaches in developing "PSN" (Partitioning of Signal and Noise)—a generative low-rank denoising method that optimally separates signal from noise in neural data, improving the performance and interpretability of downstream computational models.

PSN will address a core challenge in building a truly foundational fMRI dataset by maximizing the amount of stimulus-driven information (signal) that can be recovered from each participant's measurements, while partitioning out the influence of other sources of variability (noise). Conventional denoising strategies such as trial averaging are straightforward and widely used, but they rely on the assumptions that noise is independent across trials and uncorrelated between voxels. In actuality, these assumptions are often violated in fMRI data, where noise can be structured, spatially correlated, and non-stationary. Similarly, PCA-based low-rank denoising identifies directions of highest variance but does not explicitly distinguish between signal and noise, leading to bias when noise variance is large or when signal and noise share overlapping subspaces [19, 23], .

PSN addresses these limitations by extending the GSN framework [19] to produce denoised trial-averaged data optimized for downstream modeling. GSN estimates separate covariance structures for the signal and noise directly from repeated-trial measurements. These estimates define a signal-aware basis for low-rank reconstruction, allowing us to selectively preserve dimensions most likely to reflect stimulus-driven activity while discarding those dominated by noise.

Critically, PSN relies on cross-validation to determine the optimal number of signal dimensions to retain, with thresholds chosen either at the multi-voxel or single-voxel level, depending on the data's heterogeneity in feature tuning and signal-to-noise ratio. This cross-validated tailoring of denoising parameters will be particularly important given CONFORM's aim of integrating large, multi-site datasets, where measurement quality can vary widely across participants, scanners, and brain regions.

In simulations with known ground truth, PSN consistently recovers more accurate signal estimates than trial averaging or PCA-based methods, achieving lower variance without introducing substantial bias. Applied to real datasets, including primate electrophysiology and human fMRI, PSN yields substantial gains in cross-validated encoding model performance and improves the interpretability of model-derived feature visualization (manuscript in preparation). In the context of CONFORM, applying PSN to every contributed dataset ensures that all data entering the foundation model is maximally informative, optimized for data quality and reliability, and robust to the structured noise sources inherent in large-scale, crowd-sourced fMRI. Finally, because non-stimulus-driven sources of neural variability may themselves be of scientific interest, PSN also enables these components to be cleanly separated for downstream analyses that focus on modeling noise rather than signal.

2.2 Integration of fMRI data across studies—MOSAIC

Individual fMRI experiments face practical constraints that create trade-offs between the number of participants, the number of experimental trials, and stimulus diversity. Any resulting conclusions are thus limited in scope. However, the aggregation of existing fMRI datasets, here called MOSAIC (Meta-Organized Stimuli And fMRI Imaging data for Computational modeling), achieves a vastly larger scale useful for measuring cross-dataset and cross-subject generalization and training of high-parameter artificial neural networks.

MOSAIC [14] currently preprocesses eight event-related fMRI vision datasets (Natural Scenes Dataset [3], Natural Object Dataset [5], BOLD Moments Dataset [15], BOLD5000 [2], Human Actions Dataset, Deeprecon [17], Generic Object Decoding [18], and THINGS [4]) with a shared pipeline and registers all data to the same cortical surface space. Single-trial beta values in MOSAIC are estimated using GLMsingle and a high integrity test-train split is curated across datasets.

At present, MOSAIC contains 430,007 fMRI-stimulus pairs from 93 participants across 162,839 unique image stimuli. The stimuli are further divided into 144,360 training stimuli, 18,145 test stimuli, and 334 synthetic stimuli for rigorous model training and evaluation. Their shared preprocessing pipeline uses open source frameworks and is thus compatible with methods advancements

such as PSN and expansion to other registration spaces such as subject native. Crucial to CONFORM, datasets can be added to MOSAIC *post-hoc* regardless of experimental design, acquisition, and size.

MOSAIC is a critical first step to enable researchers to overcome individual dataset limitations and tackle complex research questions at an unprecedented scale. The MOSAIC dataset and preprocessing code will be available soon for download. In tandem with the MOSAIC team, the larger CONFORM community will work to leverage MOSAIC's extensible design to allow the seamless integration of new datasets, creating an evolving foundation for collaborative human vision research.

2.3 Generalizing across participants and studies in a low data regime—BraInCoRL

Different datasets may utilize different stimuli, employ different scanning parameters, and collect data from diverse populations. This makes it challenging to build generalizable models that predict neural activity across diverse participants. Traditional approaches require large, participant-specific fMRI datasets, limiting their scalability for clinical and research applications. This variability in cortical organization – driven by anatomical and functional differences, developmental experiences, and learning –necessitates a framework that can adapt to new individuals with minimal data while capturing shared functional principles of visual processing.

To address this, BrainCoRL (Brain In-Context Representation Learning) [20] leverages meta-learning and transformer-based in-context learning to predict voxelwise neural responses from few-shot examples without fine-tuning. Inspired by how large language models adapt to new tasks through contextual examples, BrainCoRL treats each voxel's response function as a learnable mapping that can be inferred from limited data. The model is trained across multiple participants to discover shared functional principles of visual processing, enabling it to rapidly adapt to new individuals without additional fine-tuning. BrainCoRL outperforms traditional voxelwise encoding models in low-data regimes, generalizes to entirely new fMRI datasets acquired with different scanners and protocols, and provides interpretable insights into cortical selectivity through its attention mechanisms. Notably, the framework can also link neural responses to natural language descriptions, opening new possibilities for query-driven functional mapping of the visual cortex. By dramatically reducing the data requirements for accurate neural encoding models, this work paves the way for more scalable and personalized applications in both basic neuroscience and clinical settings, where understanding individual differences in brain organization is crucial for diagnosis and treatment.

3 Impact and Conclusions

Although existing large-scale fMRI datasets have been valuable, used in hundreds of studies to support a wide range of novel scientific discoveries, they are limited by their single-site, small-N approach. To move beyond this, we propose CONFORM—a unique crowd-sourcing strategy that leverages recent advances in data processing, data aggregation, analysis, and a new crowd-sourced infrastructure. This new approach directly addresses the financial and logistical challenges of collecting large datasets while enabling unprecedented stimulus diversity. However, simply crowd-sourcing data is not enough; CONFORM's success will be predicated on the specific data and modeling optimizations we introduce to handle the multifaceted noise inherent in fMRI. Moreover, by creating models that can effectively predict new data with only a small amount of information, we will dramatically broaden the accessibility of NeuroAI methods. This will empower a much wider range of researchers to leverage the power of modern AI using the typical scale of data they collect, ultimately accelerating scientific discovery.

Critically, generalizing across individuals requires addressing both biological differences and technical noise sources, such as artifacts from different scanners and motion. We directly tackle these challenges through a three-pronged approach: (1) *Data Acquisition:* Collect a limited amount of data from each participant, including repeated and partially overlapping stimuli across the population, to boost both data quality and stimulus diversity. (2) *Denoising:* Apply a two-level denoising strategy. Use GLMsingle to optimize the signal-to-noise ratio within each subject and, then, apply PSN to separate stimulus-related variance from idiosyncratic noise, improving data quality and interpretability. (3) *Alignment:* Learn a mapping from the denoised data into a shared representational space, thereby allowing us to make accurate predictions across individuals. This can be achieved through advanced methods such as BrainCoRL, which does not require overlapping stimuli, or using standard functional alignment techniques that rely on overlapping stimuli in the denoised data.

By integrating and advancing these tools to create a true foundational model, we can answer downstream questions using the dataset population to make predictions about new individuals or clinical populations. For example, recent advances in visualizing and labeling neural representations of object

categories [24, 25] could be extended to autistic individuals, thereby providing a much clearer picture of the encoding of atypically processed visual information (e.g., human faces). Thus, a wide range of research domains will have access to modern AI methods using only the scale of data they typically collect. Ultimately, this generalizability will enable the next generation of insights into brain function across a much wider range of populations.

References

[1] Alessandro T Gifford, Benjamin Lahner, Pablo Oyarzo, Aude Oliva, Gemma Roig, and Radoslaw M Cichy. What opportunities do large-scale visual neural datasets offer to the vision sciences community? *Journal of Vision*, 24(10):152–152, 2024.

[2] Nadine Chang, John A Pyles, Austin Marcus, Abhinav Gupta, Michael J Tarr, and Elissa M Aminoff. Bold5000, a public fmri dataset while viewing 5000 visual images. *Scientific Data*, 6(1):1–18, 2019.

[3] 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.

[4] Martin N Hebart, Oliver Contier, Lina Teichmann, Adam H Rockter, Charles Y Zheng, Alexis Kidder, Anna Corriveau, Maryam Vaziri-Pashkam, and Chris I Baker. Things-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *eLife*, 12:e82580, feb 2023.

[5] Zhengxin Gong, Ming Zhou, Yuxuan Dai, Yushan Wen, Youyi Liu, and Zonglei Zhen. A large-scale fMRI dataset for the visual processing of naturalistic scenes. *Scientific Data*, 10(1):559, 2023.

[6] Thomas Naselaris, Emily Allen, and Kendrick Kay. Extensive sampling for complete models of individual brains. *Current Opinion in Behavioral Sciences*, 40:45–51, 2021. Deep Imaging - Personalized Neuroscience.

[7] Daniel H. Baker, Greta Vilidaite, Freya A. Lygo, Anika K. Smith, Tessa R. Flack, Andre D. Gouws, and Timothy J. Andrews. Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, 26(3):295–314, 2021.

[8] 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, nov 2022.

[9] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In *Computer Vision - ECCV 2014*, pages 740–755. Springer International Publishing, 2014.

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

[11] Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. SUN database: Large-scale scene recognition from abbey to zoo. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pages 3485–3492, 2010.

[12] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. ImageNet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015.

[13] International Brain Laboratory. An international laboratory for systems and computational neuroscience. *Neuron*, 96(6):1213–1218, December 2017.

[14] Benjamin Lahner, N. Apurva Ratan Murty, and Aude Oliva. MOSAIC: An aggregated fMRI dataset for robust and generalizable vision research. *Journal of Vision*, 25(9):2263–2263, 2025.

[15] Benjamin Lahner, Kshitij Dwivedi, Polina Iamshchinina, Monika Graumann, Alex Lascelles, Gemma Roig, Alessandro Thomas Gifford, Bowen Pan, SouYoung Jin, N. Apurva Ratan Murty, Kendrick Kay, Aude Oliva, and Radoslaw Cichy. Modeling short visual events through the BOLD moments video fMRI dataset and metadata. *Nature Communications*, 15(1):6241, 2024.

[16] Ming Zhou, Zhengxin Gong, Yuxuan Dai, Yushan Wen, Youyi Liu, and Zonglei Zhen. A large-scale fmri dataset for human action recognition. *Scientific Data*, 10(1):415, 2023.

[17] Guohua Shen, Tomoyasu Horikawa, Kei Majima, and Yukiyasu Kamitani. Deep image reconstruction from human brain activity. *PLoS computational biology*, 15(1):e1006633, 2019.

[18] Tomoyasu Horikawa and Yukiyasu Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature communications*, 8(1):15037, 2017.

[19] Kendrick Kay, Jacob S Prince, Thomas Gebhart, Greta Tuckute, Jingyang Zhou, Thomas Naselaris, and Heiko H Schütt. Disentangling signal and noise in neural responses through generative modeling. *PLOS Computational Biology*, 21(7):e1012092, 2025.

[20] Muquan Yu, Mu Nan, Hossein Adeli, Jacob S. Prince, John A. Pyles, Leila Wehbe, Margaret M. Henderson, Michael J. Tarr, and Andrew F. Luo. Meta-learning an in-context transformer model of human higher visual cortex, 2025.

[21] Aria Yuan Wang, Kendrick Kay, Thomas Naselaris, Michael J Tarr, and Leila Wehbe. Better models of human high-level visual cortex emerge from natural language supervision with a large and diverse dataset. *Nat Mach Intell*, 5:1415–1426, 2023.

[22] Christopher J Markiewicz, Krzysztof J Gorgolewski, Franklin Feingold, Ross Blair, Yaroslav O Halchenko, Eric Miller, Nell Hardcastle, Joe Wexler, Oscar Esteban, Mathias Goncavles, Anita Jwa, and Russell Poldrack. The openneuro resource for sharing of neuroscience data. *eLife*, 10:e71774, 2021.

[23] Dean A Pospisil and Jonathan W Pillow. Revisiting the high-dimensional geometry of population responses in visual cortex. *bioRxiv*, pages 2024–02, 2024.

[24] Andrew Luo, Maggie Henderson, Leila Wehbe, and Michael Tarr. Brain diffusion for visual exploration: Cortical discovery using large scale generative models. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 75740–75781. Curran Associates, Inc., 2023.

[25] Andrew Luo, Margaret Marie Henderson, Michael J. Tarr, and Leila Wehbe. Brainscuba: Fine-grained natural language captions of visual cortex selectivity. In *The Twelfth International Conference on Learning Representations*, 2024.