Uncovering High-Dimensional Visual Neural Structures: A Topological Approach to Representational Similarity Analysis

Coco Wang; Reza D. Farivar
Cognitive Science Program, McGill University Department of Ophthalmology and Visual Science, McGill University

Abstract

Understanding brain function across regions requires robust analytical techniques that can bridge brain-activity measurement, behavioral data, and computational modeling. Traditional approaches, such as Representational Similarity Analysis (RSA), have proven effective in investigating visual neural activity by constructing Representational Dissimilarity Matrices (RDMs). However, RSA's reliance on linear similarity metrics can obscure deeper, high-dimensional structures in neural data. Representational Topology Analysis (RTA), as a more powerful alternative, leverages Topological Data Analysis (TDA) and Persistent Diagrams (PDs) with machine learning and statistical inference to capture topological features such as clusters, loops, and voids within neural representations. By utilizing the TDApplied R package developed by Shael Brown and Dr. Reza Farivar, this research extends RTA on the Natural Scenes Dataset (NSD), a large-scale fMRI dataset, to explore important topological features underlying visual neural activities. This research aims to demonstrate the efficacy of RTA in revealing novel insights into neural coding and provide a replicable testbed for further exploration of topological structures in brain activity patterns.

INTRODUCTION

Understanding brain function across regions requires sophisticated analytical techniques. It has been a fundamental challenge in system neuroscience to quantitatively bridge three key branches of research: brain-activity measurement, behavioral measurement, and computational modeling (1). There were successful cases of constructing computational model for a single neuron across different stimuli (2, 3, 4), but the model became undiscernible when the analysis advances to a large scale of organization, in which things can be further complicated if a one-to-one correspondence between a fMRI voxel and a single model unit cannot be assumed. To avoid this spatial correspondency problem, Kriegeskorte et al. (1) introduced Representational Similarity Analysis (RSA), a framework that abstracts neural activity patterns into representational dissimilarity matrices (RDMs). Instead of comparing directly the activity patterns, they compared these patterns associated with each pair of conditions (5) to produce dissimilarity matrices—an approach known as second-order isomorphism (6). In the study done by Drucker and Aguirre in 2009 (7) that investigated how neural representations of shape similarity vary across subregions of the lateral occipital complex (LOC), for instance, relied on explicit similarity analyses of neuronal activity patterns in human fMRI. In the famous research of Allen et al in 2022 published in Nature Neuroscience (8) which involves subjects viewing 750 natural scenes, RSA was utilized as one of its major ways of probing brain representations. However, despite its widespread adoption, RSA is found to be insufficient to capture similarities in certain cases (9, 10), which could obscure deeper, non-linear structures in high-dimensional neural data.

For instance, two computational systems deemed highly similar by RSA may, in reality, rely on fundamentally different mechanisms (11). Singh et al had argued that the topological structure of activity patterns in V1 when the cortex is spontaneously active is consistent with the topology of a two sphere (12). Similar property was discovered in Curto's research that discussed topology and its underlying neural code, in which the hexagonal domain of the spiking activity of rats' grid cells is "precisely a torus" (13). Since RSA evaluates only linear similarities, it risks false equivalences between representational spaces (e.g. grid cells and V1 simple cells) that differ topologically (Fig. 1). To overcome these limitations, Brown and Farivar proposed Representational Topology Analysis (RTA), incorporating Persistent Diagrams (PDs) with machine learning and statistical inference (9). Given any input distance matrix of a dataset, persistence diagram is not only capable of identifying inherent structures of the neural patterns like clusters (H_0), loops (H_1), and voids (H_2) but also providing valuable information about which features are significantly influential. As the number of input data points grows, the topological features will eventually converge to the underlying features (14) regardless of possible noise in the dataset (15). Instead of relying on second-order isomorphism of RDMs, RTA leverages second-order isomorphisms of persistence diagrams, creating a topological analytic space better suited for detecting complex, higher-order representational structures in neural data.

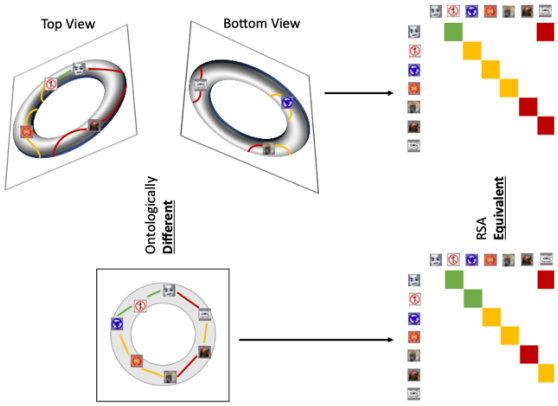

Fig. 1 | Sampled representations from a torus, projected onto an annulus, and the resulting two RDMs. The two RDMs would be considered equivalent by RSA, despite distinct topological spaces (11)

The rich information embedded in NSD further provides a robust and replicable testbed for advancing this approach. By extending RTA methodologies, this research aims to uncover novel insights into stimulus-driven fMRI neural representations in higher dimensions.

Building on this framework, I performed RTA to the neural data of the first subject involved in the natural scene dataset (NSD) experiment. Seven regions of interests (ROI) are being examined closely: Primary Visual Cortex (V1), Second Visual Area (V2), Third Visual Area (V3), Fourth Visual Area (V4), Eighth Visual Area (V8), Area Lateral Occipital 2 (L02), and Posterior-Infero-Temporal Complex (PIT). For a more detailed explanation of each ROI, please refer to HCP-MMP1 atlas online document (16).

METHODS

Subjects

This study utilized the Natural Scenes Dataset (NSD), a publicly available dataset that includes high-resolution (1.8-mm) whole-brain 7T functional magnetic resonance imaging (fMRI) data from eight carefully screened human participants. Each subject viewed approximately 9,000–10,000 unique natural scene images over the course of 30–40 scan sessions distributed across one year. Across all participants, NSD encompasses responses to 70,566 distinct natural scene images, making it an order of magnitude larger than similar fMRI datasets.

Each participant viewed 750 natural scene images per session while completing a cognitive task. Specifically, participants were asked to judge whether each image had been presented at any point in the past while maintaining central fixation. The dataset also includes extensive metadata, such as trial-wise behavioral responses and fixation stability, enabling a comprehensive analysis of neural responses to naturalistic visual stimuli. More details on the dataset can be found at https://naturalscenesdataset.org/.

Experimental Design

Trial Design. During each trial, participants maintained central fixation while being presented with a sequence of color natural scenes in a 3-s ON/1-s OFF trial structure. Their task was to determine whether each image had been previously shown at any point in the study. The stimuli were sourced from Microsoft's COCO14 dataset, which includes detailed object annotations.

Run and Session Structure. Each experimental run lasted approximately 5 minutes and comprised 62 or 63 stimulus trials, interspersed with occasional blank trials to provide baseline measurements. A single scan session consisted of 12 such runs, totaling 750 stimulus trials per session. The experimental timeline included an initial screening session (prffloc), 30–40 NSD core sessions, and two concluding sessions dedicated to synthetic image and mental imagery processing (nsdsynthetic and nsdimagery). The first NSD core session was designated as day 0 for each participant.

Behavioral Performance. To quantify participants' performance, their responses were categorized into three trial types, with the percentage of trials in which an 'old' response was given recorded and analyzed for accuracy.

Stimulus. The stimuli used in this study consisted of natural scene images sourced from the Natural Scenes Dataset. Each image was taken from the Microsoft COCO14 dataset, which includes extensive object annotations, allowing for detailed analyses of visual features. The selection of stimuli ensured a wide variety of scene compositions, promoting robust neural responses across experimental conditions.

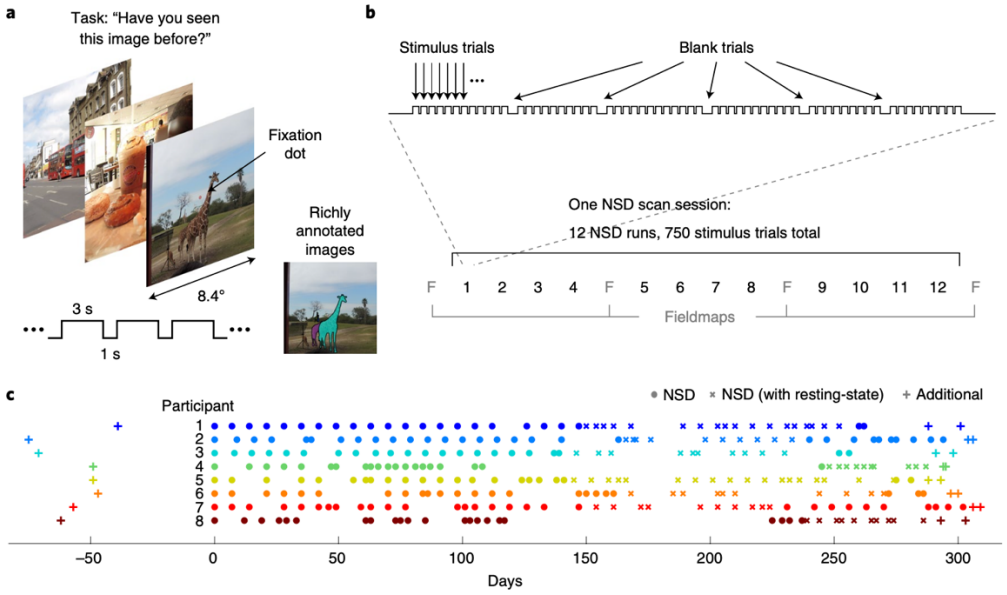

Fig. 2 | Design of the NSD experiment. a, Trial design **b,** Run and session design. **c,** Timeline of 7T fMRI scan sessions. (1)

Equipment and Material

All data processing and analysis were conducted on a MacBook Pro equipped with an Apple M3 Pro processor, running macOS 15.3.1 (24D70). The following software tools were utilized:

- Python (nibabel, numpy, matplotlib, pandas): Automated Representational Dissimilarity Matrix (RDM) construction and data processing.
- Azure Data Studio: Filtering and inspection of large-scale neural data.
- R (TDApplied): Topological data analysis (17) to extract complex structural patterns.
- AFNI & Freeview: Visualization of regions of interest (ROIs) within the brain.
- Amazon Web Services (AWS)'s Simple Storage Service (S3)

Task

In this study, we aimed to perform a Representational Similarity Analysis (RSA) and subsequently apply Topological Data Analysis (17) to explore the structure of representational dissimilarity matrices (RDMs) derived from fMRI data. The workflow consisted of several key steps, as outlined below.

Representational Similarity Analysis (RSA). First, we obtained surface-based fMRI beta values from the first session of the participant 1 in NSD. The dataset was structured in the fsaverage cortical surface space, with data stored in .mgh format. Each vertex (node) represented an activation pattern across different conditions (image stimuli). To assign each vertex to a specific region of interest (ROI), we utilized the HCP_MMP1.mgz atlas and its corresponding label file, HCP_MMP1.mgz.txt. This allowed us to map each of the 163,842 cortical nodes to one of 180 predefined brain regions. The ROI assignments were integrated into our data matrix, enabling us to analyze region-specific activation patterns.

Once the node-to-ROI mapping was established, we extracted data corresponding to selected ROIs of interest. We selected V1, V2, V3, V4 and V8 as our primary focus, Lo2 and PIT as the control group. For each ROI, we constructed a representational dissimilarity matrix (RDM) by computing the Spearman correlation distance (1 - Spearman correlation) across all conditions. The resulting RDMs captured the similarity structure of neural responses within each ROI (fig. 3).

To compare RDMs across different brain regions, we first implemented a standard RSA approach, calculating Spearman correlations between the upper triangular values of each RDM. These results were visualized in a heatmap, illustrating the degree of representational similarity between brain regions (fig. 3). The activity patterns seem to be highly similar between visual areas of V1, V2, V3 and V4, but not so much in higher visual areas. This aligns with progressive transformation of visual information in the brain. Early areas share more structured, image-driven encoding, while higher areas branch into more specialized, semantically enriched spaces (18). However, to gain deeper insights into the global structure of the RDMs, we shifted to a Topological Data Analysis (17) framework.

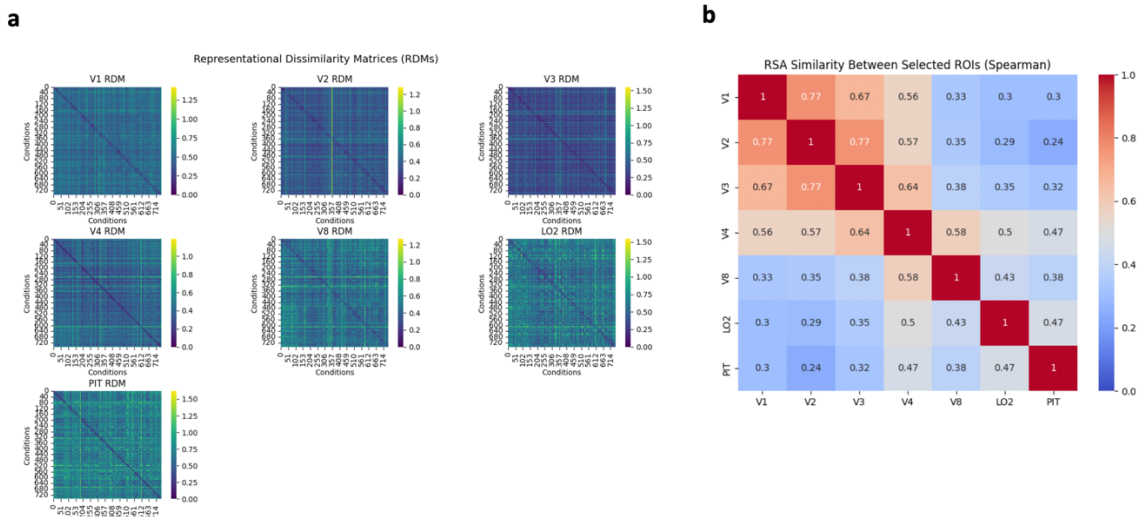

Fig. 3 | Representational Similarity Analysis (RSA). a, representational dissimilarity matrices (RDM) of seven selected region of interests **b,** second-order isomorphism of RDMs calculated using Spearman correlation distance

Representational Topology Analysis (RTA)

To uncover higher-order structural patterns in fMRI data, I applied Topological Data Analysis (17) using the TDApplied and RTA packages in R. The goal was to analyze surface-based fMRI Representational Dissimilarity Matrices (RDMs) from early and higher visual areas (e.g., V1, V2, V3, V4, LO2, PIT), in order to detect topological features such as loops (1-dimensional homology, h1) that might reflect meaningful stimulus relationships in the brain's representational space.

Bootstrapping for statistical significance

Each region's RDM was subjected to bootstrapping via the *bootstrap_persistence_thresholds function* (17) to identify statistically significant persistent features. This method involves repeated random sampling of the RDM to generate a null distribution of persistence diagrams, allowing for empirical p-value estimation. In this study, we are only considering two dimensions (maxdim = 1) and keeping features in the top 2.5% of the null distribution (thresh = 2).

Due to the random nature of bootstrapping, results varied slightly between runs. I applied 10 runs to each ROIs on average. However, consistent trends emerged:

- **V1** consistently yielded 4 to 6 significant loop features. One example output is shown in Fig. 4.
- **V2** and **V3** often exhibited 1 to 2 features.
- **V4**, **V8**, **LO2**, and **PIT** generally lacked significant features across runs, suggesting a lower level of topological structure in their RDMs.

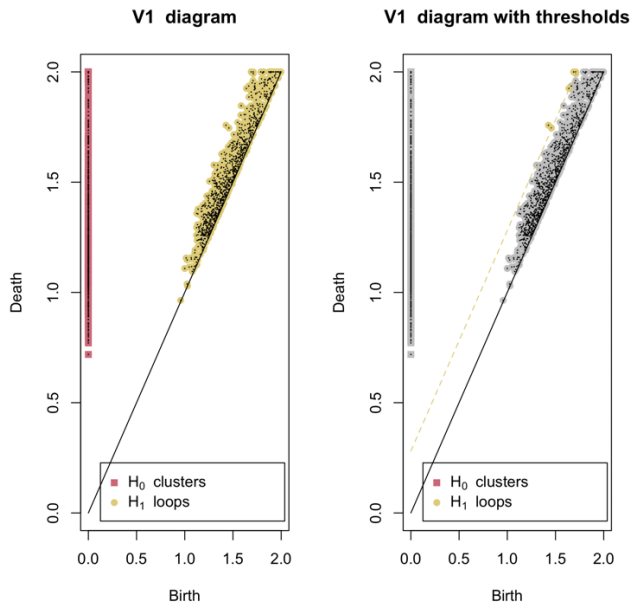

Fig. 4 | Persistent Diagram (PD) of RDM of V1. This plot shows that six significant H1 features are found.

Mapping Features to Stimuli

For each detected loop, we extracted the underlying trials from **subsetted_representatives** output of bootstrapping method. In the NSD experiments, image stimuli are presented and repeated in a specific pattern. In the Natural Scenes Dataset (NSD), each of the 10,000 images was shown three times across 30,000 trials, with a fixed trial ordering shared across participants (8). The first 1,000 image slots were reserved for a shared set ("shared1000"), while the remaining 9,000 were unique per participant. To control repetition timing and ensure balanced memory difficulty, image presentations were scheduled using a circular sampling procedure. For each image, a mixture of a von Mises distribution (60%) and a uniform distribution (40%) was used to determine three repetition times, with the von Mises peak randomly placed on a conceptual circle. This circular trial layout was then unwrapped into a linear sequence, split evenly into 40 scanning sessions (750 trials/session), ensuring a uniform distribution of repetitions and minimizing clustering effects across the experiment.

According to the ordering, we matched each trial to the image stimulus via a trial-to-nsdID mapping. This allowed visual inspection of the actual stimuli contributing to each topological feature (Fig. 5).

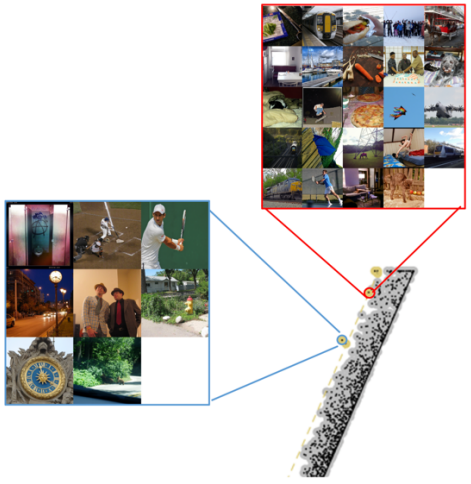

Fig. 5 | Trial-to-nsdID mapping. This plot zooms in two significant H1 features (out of six) of V1. Their underlying trials and corresponding NSD images are displayed.

Vietoris-Rips (VR) Graph Construction and Visualization

To visualize the structure of each feature, I used the Vietoris–Rips graph method (*vr_graphs()*), which constructs a VR graph from the filtered RDM of interest (Method 1). VR graphs provide an intuitive, geometric view of the similarity structure encoded in the RDM, highlighting clusters, transitions, and loops, which correspond to topological features revealed through persistent homology. By filtering the original RDM to include only the trials forming a significant loop, I generated a reduced distance matrix that preserves only the relevant stimulus relationships. Then, using *plot_vr_graph()*, I rendered the graph with node colors corresponding to different features and overlaid each node with its corresponding image stimulus using *rasterImage()*.

I explored an alternative approach (Method 2) where the full VR graph was constructed first, and then subgraphs corresponding to specific features were extracted. However, this method proved more limited due to constraints in plot_vr_graph(), which expects an object directly from vr_graphs(). Therefore, Method 1 was preferred for flexibility and interpretability.

Observations and Interpretation

The resulting VR graphs offered an interpretable, geometric visualization of stimulus clusters forming topological loops (fig. 6). However, there are several things to note: not all features detected using bootstrapping are valid, and some of the features vary across different funs of bootstrapping. The first phenomenon happens when the persistence diagram showed a loop, but the representative cycle could not be reliably reconstructed. At the same time, some cycles might be borderline significant and may show up differently across runs — they could have meaningful persistence values but ambiguous or unstable representatives.

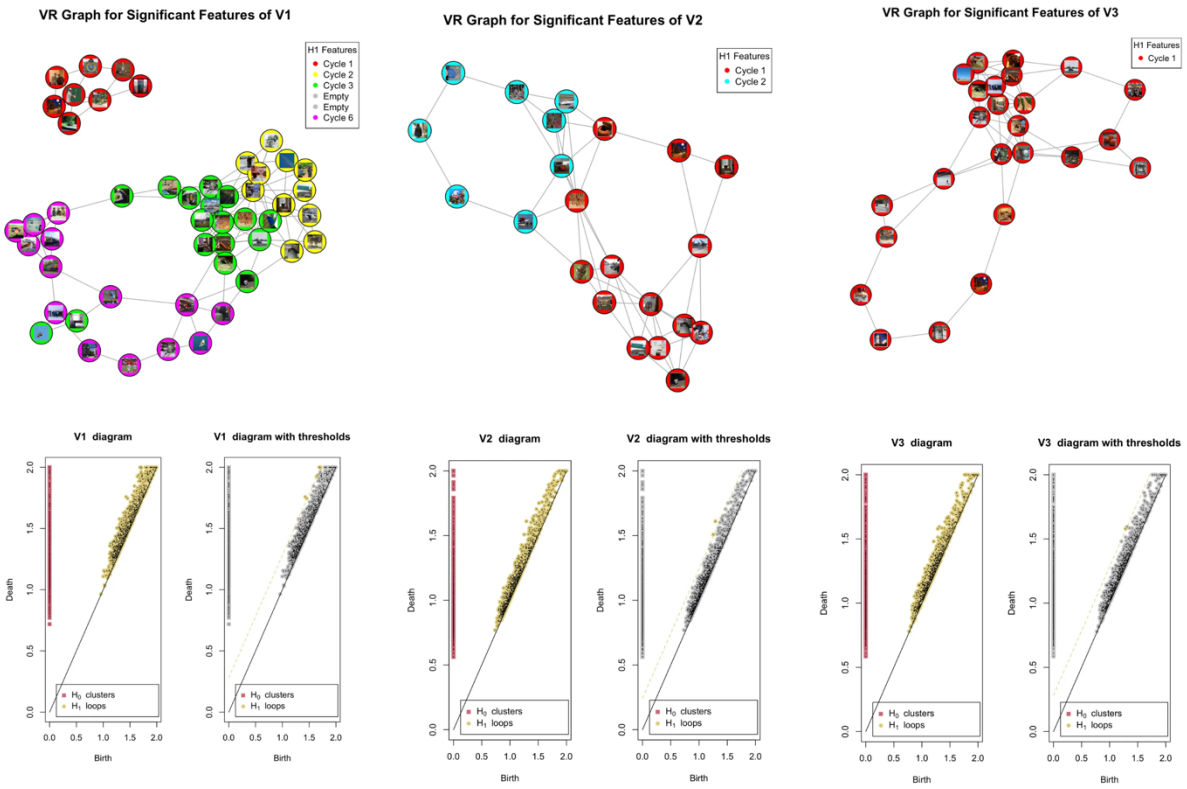

Fig. 6 | Vietoris-Rips (VR) Graph for significant features in V1 to V3 and corresponding PDs. Some cycles have empty nodes and are not real topological feature with valid representative cycles.

In the VR graph of the primary visual cortex (V1), one of the topological features—Feature 1—was distinctly separated from the rest. While these images in feature 1 do not form an obvious semantic or categorical group, they exhibit prominent vertical edges, elongated high-contrast structures, or distinct spatial frequency patterns—such as the edges of the door or the sharp contrast lines of sports equipment and poles. Given that V1 neurons are highly sensitive to line orientation, especially vertical and diagonal edges, and respond selectively to spatial frequency and edge direction, the grouping of these visually distinct images into a single feature might suggest a shared activation pattern for vertical lines across V1's early visual detectors.

In contrast, the VR graph of the third visual area (V3) revealed a single, prominent topological feature shaped like a figure "8," composed of two interconnected loops. Notably, one of these loops clustered animal-related stimuli more tightly than observed in earlier visual areas like V1 or V2. This emerging structure suggests a potential transition from low-level visual representations (e.g., edges, textures, symmetry) toward mid-level or even proto-semantic groupings. Although V3 is still considered part of early visual cortex, it is associated with the processing of global shapes and motion and hence may begin to reflect more integrated representations of visual scenes.

In higher-order regions—including V4, V8, LO2, and PIT—no statistically significant topological features were consistently detected. But it is important to note that the current analysis is solely based on a single session from the first participant in the NSD experiment, which includes over 40 sessions per participant and 8 participants in total. Due to greater response variability and abstract coding in higher-order areas like LO2 and PIT, it is likely that more robust, stable topological patterns would emerge with a larger dataset spanning multiple sessions and participants.

Discussion

For future studies, it will be essential to extend the current analysis beyond a single session and participant. Including data from multiple sessions and individuals can offer a more robust and generalizable picture of topological structures across brain regions.

Equally important is the development of systematic metrics to interpret and evaluate the image content within each topological feature. While persistent homology reveals topological signatures in neural similarity structure, the semantic or visual basis of these features often remains opaque. One promising approach would be to apply dimensionality reduction techniques such as PCA or t-SNE to extract dominant visual components from the images in each feature, offering insights into shared low-level or mid-level properties. Furthermore, machine learning methods—such as clustering or supervised classification—could be employed to quantify and predict the relationships between visual features and neural topologies. Finally, a kind of Representational Similarity Analysis (RSA) applied directly to image embeddings (e.g., from CNNs or CLIP models) could be used to compare image-based similarity to brain-derived topologies, helping to bridge the gap between image-level and neural-level representations.

REFERENCES

1. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. Front. systems neuroscience 2, 4.
2. David, S. V., and Gallant, J. L. (2005). Predicting neuronal responses during natural vision. Network 16, 239–260
3. Koch, C. (1999). Biophysics of Computation: Information Processing in Single Neurons. New York, Oxford University Press.
4. Rieke, F., Warland, D., De Ruyter van Steveninck, R., and Bialek, W. (1999). Spikes: Exploring the Neural Code. Cambridge, MA, MIT Press.
5. Edelman, S. (1998). Representation is representation of similarities. Behav. Brain Sci. 21, 449–498.
6. Roger N Shepard, Susan Chipman, Second-order isomorphism of internal representations: Shapes of states, Cognitive Psychology, Volume 1, Issue 1, 1970, Pages 1-17, ISSN 0010-0285, https://doi.org/10.1016/0010-0285(70)90002-2.
7. Drucker, D. M., & Aguirre, G. K. (2009). Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cerebral cortex (New York, N.Y. : 1991)*, *19*(10), 2269–2280. https://doi.org/10.1093/cercor/bhn244
8. Allen, E.J., St-Yves, G., Wu, Y. *et al.* (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat Neurosci* 25, 116–126. https://doi.org/10.1038/s41593-021-00962-x
9. Dujmovi´ c, M., Bowers, JS., Adolfi, F., Malhotra, G. (2022). The pitfalls of measuring representational similarity using representational similarity analysis. bioRxiv 1.
10. Chen, X., Martin, R., & Fischer-Baum, S. (2021). *Challenges for using Representational Similarity Analysis to Infer Cognitive Processes: A Demonstration from Interactive Activation Models of Word Reading*. https://escholarship.org/uc/item/opn3664s
11. Brown, S., & Farivar, R. (2024). The Topology of Representational Geometry. *bioRxiv*, 2024-02.
12. Singh, G., et al. (2008). Topological analysis of population activity in visual cortex. J. vision 8, 11.1–18.
13. Curto, C. (2016, May 6). *What can topology tell us about the neural code?* arXiv.org. https://arxiv.org/abs/1605.01905
14. Chazal, F., Glisse, M., Labruère, C., & Michel, B. (2014, January 27). *Convergence rates for persistence diagram estimation in Topological Data Analysis*. PMLR. https://proceedings.mlr.press/v32/chazal14.html
15. Edelsbrunner, N., Letscher, N., & Zomorodian, N. (2002). Topological persistence and simplification. *Discrete & Computational Geometry*, *28*(4), 511–533. https://doi.org/10.1007/s00454-002-2885-2
16. Atlases — neuroimaging core 0.1.1 documentation. (n.d.). Retrieved from https://neuroimaging-core-docs.readthedocs.io/en/latest/pages/atlases.html#id4
17. Brown, S., & Farivar, R. (2025, January 20). Machine Learning and Inference for Topological Data Analysis. Retrieved from https://cran.r-project.org/web/packages/TDApplied/vignettes/ML_and_Inference.html
18. Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365. https://doi.org/10.1038/nn.4244