Extended Abstract Track

MAPS: A Dataset for Controlled Probing of Representational Topology in Vision Models

Editors: List of editors' names

Abstract

Neural activity exhibits low-dimensional organization across brain areas, behaviors, and species. While prior work has shown that behaviors shape the geometry and topology of neural manifolds, the structure of sensory representations remains less understood. In this work, we use artificial neural networks to investigate the topology of neural representations through continuous changes in visual features. We introduce MAPS (Manifolds of Artificial Parametric Scenes), a dataset of objects rendered in 3D with systematic parameter sweeps across hue, camera angle, lighting, and size. Each parameter defines a specific topology (e.g., a ring or an interval), with combined parameters yielding product manifolds. We show that, despite being trained on images without continuous transformations, pretrained vision models capture the topology of our controlled input manifolds. We envision to expand MAPS with additional objects and transformations, and to move beyond topology toward analyzing the geometry of neural representations.

Keywords: Neural representations, topology, pretrained vision models, datasets

1. Introduction

Neural population activity exhibits a fundamental organizing principle: it unfolds within low-dimensional manifolds whose layout reflects task goals, input statistics, and inductive biases (Perich et al., 2025). Striking examples include grid cells encoding spatial location on a torus (Gardner et al., 2022), motor cortex activity forming behavioral manifolds (Brennan and Proekt, 2019), and visual cortex exhibiting a 2-sphere topology for orientation processing (Singh et al., 2008). While these studies demonstrate that neural systems can preserve and exploit the topological structure of their input domain, whether low-dimensional manifolds systematically organize sensory representations remains less understood.

Artificial neural networks (ANNs) offer a controlled testbed to explore the topological structure of stimulus encoding, owing to their lower overall complexity in comparison to brain circuits, the accessibility of neural activity and the much lower measurement noise. Yet, systematic investigations of how models handle topological manipulations in stimulus features remain limited, in part because existing parameterized datasets rely on primitive shapes with limited semantic richness (Johnson et al., 2017; Higgins et al., 2017; Kim and Mnih, 2018), or are restricted to low-resolution renderings (Gondal et al., 2019).

Here, we present MAPS (Manifolds of Artificial Parametric Scenes), a high-resolution dataset of semantically meaningful ImageNet objects with systematic single-feature manipulations. By varying individual properties, such as color, size, orientation, or lighting, while holding all other features constant, we trace out how such manipulations map to trajectories in the representational space of ANNs, investigating whether these feature variations are encoded into smooth or rather fragmented structures.

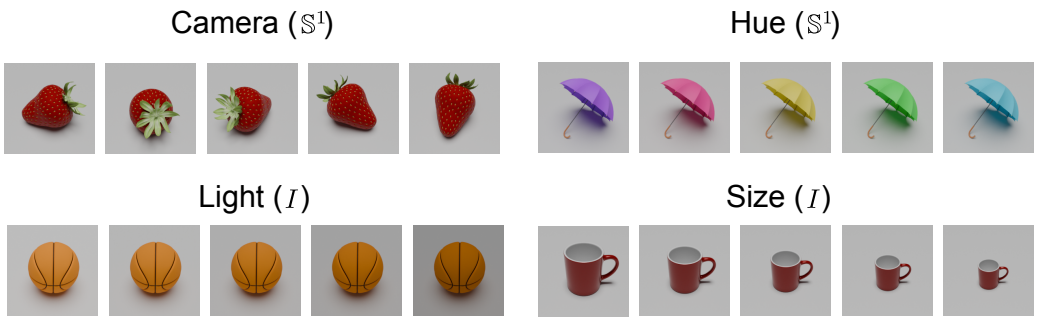

Figure 1: Overview of the MAPS dataset showcasing four feature manipulations: camera angle (azimuthal angle), hue, light, size. Cyclic features (hue, camera angle) have circular topology ($\mathbb{S}^1$), while bounded features (light, size) have interval topology (I).

Our contributions: (i) we provide MAPS, a controlled dataset for probing representational topology in vision models; (ii) using topological data analysis on pretrained vision models, we show that the activity patterns of early layers capture the input transformation topology.

2. Dataset & Methods

2.1. MAPS dataset

To create MAPS, we imported 3D object models into Blender 4.5.1 (Blender Online Community, 2025) and rendered controlled image sets with systematic variations in hue, camera angle, object size, and light intensity (Fig. 1). These transformations were scripted using the Blender Python API and designed to isolate individual feature dimensions, avoiding the confounds inherent in natural datasets (see Appendix A.1 for details).

Each transformation is associated with an underlying topology: hue shifts and camera rotations correspond to circles ($\mathbb{S}^1$), while scaling and light intensity changes correspond to intervals ($I = [a, b]$ for $a, b \in \mathbb{R}^+$ and $b > a$). Composing transformations produces a new manifold whose topology is given by the product of the individual manifolds. For example, combining hue shifts and camera rotations yields a parameter space with topology $\mathbb{S}^1 \times \mathbb{S}^1 = \mathbb{T}^2$, i.e., a 2-torus.

2.2. Topological analysis in ANNs

We rendered images of the umbrella object under systematic feature transformations and analysed activations after the first ReLU non-linearity from a ResNet-50 model pretrained on ImageNet (He et al., 2016). Value ranges and sampling densities are shown in Appendix A.1.

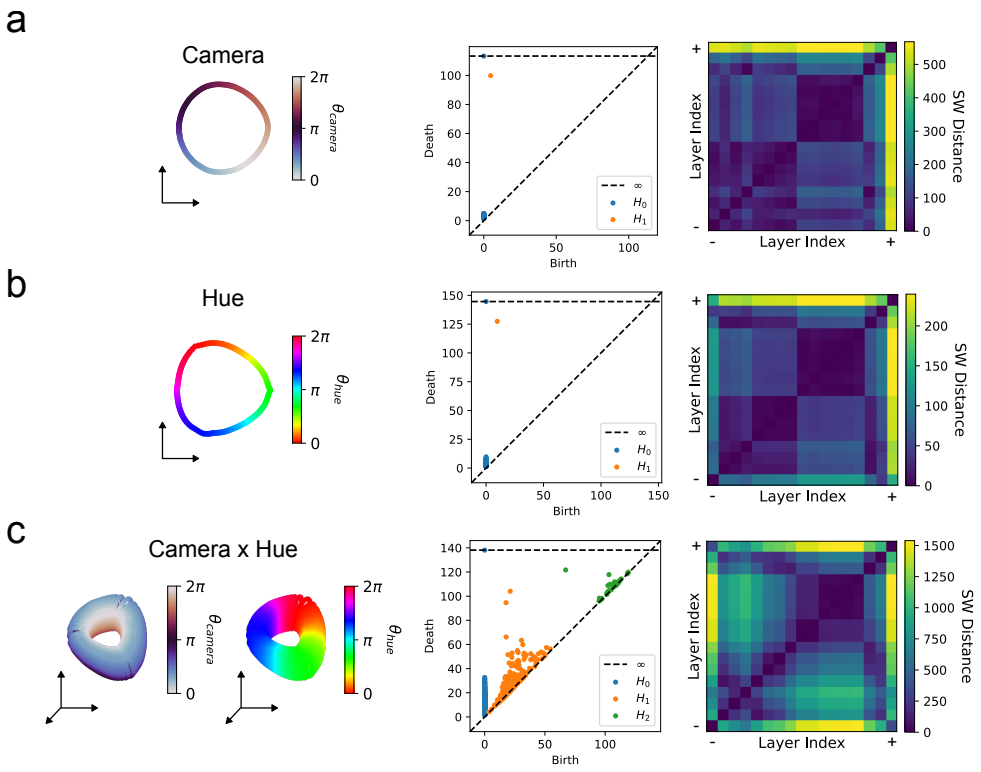

Figure 2: Topological analysis of individual (**a**, **b**) and combined (**c**) camera angle and hue manipulations in umbrella renderings. The model's first ReLU activations are projected using Isomap (first column). From a subsample of such point cloud, a persistence diagram is computed (second column). Each layer is then characterized by their persistence diagram. Layer-wise topological similarity is computed using Sliced Wasserstein distance, and a representation dissimilarity matrix is built (third colum). For the product of camera and hue, 10,000 images were rendered, from which 1000 were subsampled to efficiently compute the persistence diagram.

Following Gardner et al. (2022), we projected neural activations onto the principal components using PCA, here retaining 10 components. Next, we applied persistent homology to characterize the topology of activation patterns (Chazal and Michel, 2021). For each layer, we constructed a Vietoris–Rips complex and computed its homology to obtain the topological features of the representations (Tralie et al., 2018; Bauer, 2021). To compare persistent diagrams across layers, we used Sliced Wasserstein (SW) distance (Carriere et al., 2017; Purvine et al., 2023). For single features, we calculated the SW distance between layers using a weighted average of the H_0 and H_1 features, while for their product manifold, we used additionally H_2. For visualization, we used Isomap (Tenenbaum et al., 2000) on the PCA-projected activations (20 neighbors).

3. Results & Analysis

In the following analyses, we focus on the umbrella object from the MAPS dataset. Figure 2 shows the representational topology for each manifold: (a) camera rotation (b) hue shift and (c) the joint camera–hue transformation. For each case, we display low-dimensional projections (first column) and the persistence diagrams from the first ReLU layer (second column), and representational dissimilarity matrices (RDMs) measuring the Sliced Wasserstein distance between the persistence diagram of each layer (third column).

We observe that after the first ReLU activation, ResNet-50 captures the topology of the parameter space, both for single-parameter manifolds and for their product. The persistence diagrams reveal the correct topological signatures: single loops (H_1) for individual hue and camera angle transformations, and torus topology with two loops (H_1) and one void (H_2) for their combination (Fig. 2, second column). However, the manifolds exhibit different behaviors across the network hierarchy. For single feature manipulation, both camera angle and hue maintain ring topology across all layers, although for hue the distance between the first and last ReLU layers is smaller (Fig. 2, third column). The differences in SW distances mainly arise from shifts in the birth–death scales of the loops rather than from structural changes. For the product manifold of camera angle and hue, the torus topology emerges in early layers but disappears in deeper layers, with only a residual topology preserved in the final representations.

We performed the same analysis for additional parameters and also found a preserved topology of the parameter space imprinted in the network activations on early layers (see Appendix A.2).

Additionally, we evaluated the classification performance when varying hue, camera angle, and their combination (see Appendix A.3).

4. Conclusion & Future Work

In this preliminary work, we introduced MAPS, a dataset that enables the investigation of neural manifolds by systematically exploring object features of 3D objects in neural networks. We showed how the dataset can be used to study the topology of image transformations. Our exploratory results indicate that, despite ResNet-50 being trained only on individual image instances, it captures continuous input transformations quite accurately.

Future work will include extending and comparing topological analyses across different objects and models, as well as broadening the analysis to geometry and dimensionality of representations. More broadly, this work demonstrates how controlled datasets combined with topological analysis tools can provide insights on the fundamental principles of representational structures in both artificial and biological neural systems. Moreover, this work opens several promising avenues: the observed topological signatures in ANNs could inspire new hypotheses about visual representations in the brain; the controlled stimuli of MAPS could be leveraged to study models with explicit invariance and equivariance constraints; and the dataset could be expanded with additional ImageNet classes and new transformations to explore how semantic diversity and invariance properties shape representational topology.

References

Ulrich Bauer. Ripser: efficient computation of Vietoris-Rips persistence barcodes. *J. Appl. Comput. Topol.*, 5(3):391–423, 2021. ISSN 2367-1726. doi: 10.1007/s41468-021-00071-5. URL https://doi.org/10.1007/s41468-021-00071-5.

Blender Online Community. *Blender – a 3D modelling and rendering package*. Blender Foundation, 2025. URL https://www.blender.org.

Connor Brennan and Alexander Proekt. A quantitative model of conserved macroscopic dynamics predicts future motor commands. *Elife*, 8:e46814, 2019.

Mathieu Carriere, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. In *International conference on machine learning*, pages 664–673. PMLR, 2017.

Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4: 667963, 2021.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

Richard J Gardner, Erik Hermansen, Marius Pachitariu, Yoram Burak, Nils A Baas, Benjamin A Dunn, May-Britt Moser, and Edvard I Moser. Toroidal topology of population activity in grid cells. *Nature*, 602(7895):123–128, 2022.

Muhammad Waleed Gondal, Manuel Wuthrich, Djordje Miladinovic, Francesco Locatello, Martin Breidt, Valentin Volchkov, Joel Akpo, Olivier Bachem, Bernhard Schölkopf, and Stefan Bauer. On the transfer of inductive bias from simulation to the real world: a new disentanglement dataset. *Advances in Neural Information Processing Systems*, 32, 2019.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*, 2017.

Justin Johnson, Bharath Hariharan, Laurens Van Der Maaten, Li Fei-Fei, C Lawrence Zitnick, and Ross Girshick. Clevr: A diagnostic dataset for compositional language and elementary visual reasoning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2901–2910, 2017.

Hyunjik Kim and Andriy Mnih. Disentangling by factorising. In *International conference on machine learning*, pages 2649–2658. PMLR, 2018.

Matthew G Perich, Devika Narain, and Juan A Gallego. A neural manifold view of the brain. *Nature Neuroscience*, pages 1–16, 2025.

Emilie Purvine, Davis Brown, Brett Jefferson, Cliff Joslyn, Brenda Praggastis, Archit Rathore, Madelyn Shapiro, Bei Wang, and Youjia Zhou. Experimental observations of the topology of convolutional neural network activations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 9470–9479, 2023.

Gurjeet Singh, Facundo Memoli, Tigran Ishkhanov, Guillermo Sapiro, Gunnar Carlsson, and Dario L Ringach. Topological analysis of population activity in visual cortex. *Journal of vision*, 8(8):11–11, 2008.

Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

Christopher Tralie, Nathaniel Saul, and Rann Bar-On. Ripser.py: A lean persistent homology library for python. *The Journal of Open Source Software*, 3(29):925, Sep 2018. doi: 10.21105/joss.00925. URL https://doi.org/10.21105/joss.00925.

Appendix A.

A.1. MAPS dataset

We selected 10 object classes from ImageNet (Deng et al., 2009) and obtained the corresponding 3D models from Sketchfab (see Table 1 for object breakdown and author attributions). Each object was placed in a simple scene in Blender 4.5.1 (Blender Online Community, 2025) with a neutral background, camera, and light source. Images were rendered at 224 × 224 pixel resolution to match the input size of pretrained vision models (Figure 3).

Table 1: ImageNet objects used in MAPS, their Sketchfab author attribution, and the license type provided by the author.

ImageNet Index	Class	Sketchfab Author	License
430	Basketball	Blender Guru	CC Attribution
504	Coffee Mug	hungry_beagle	CC Attribution
650	Microphone	Aike	CC Attribution
852	Tennis Ball	Arman.Abgaryan	CC Attribution
879	Umbrella	CGV	CC Attribution - Non Commercial
890	Volleyball	Jeremy E. Grayson	CC Attribution
949	Strawberry	gelmi.com.br	CC Attribution
951	Lemon	svnfbgr	CC Attribution
953	Pineapple	Nevena Knežević	CC Attribution
954	Banana	matousekfoto	CC Attribution

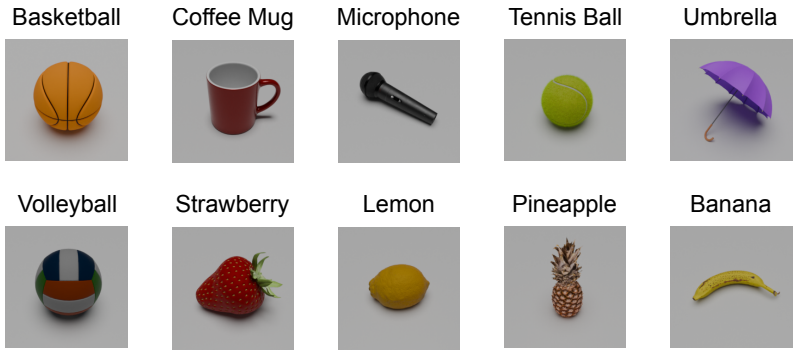

Figure 3: 3D renderings of all objects in MAPS.

For each transformation parameter, we defined value ranges and linearly sampled points within these ranges. For camera angle, we swept $\theta_{\text{camera}} \in [0, \pi)$; for hue, $\theta_{\text{hue}} \in [0, \pi)$; for light intensity, $x_{\text{light}} \in [0.1, 2]$; and for object size, $x_{\text{size}} \in [0.5, 1]$. For single-parameter

sweeps, we rendered 200 images using Blender's Cycles engine, a path-tracing renderer that simulates light for photorealistic images. For two-parameter sweeps (10,000 images), we used Blender's Eevee engine, a rasterization-based renderer optimized for real-time graphics, which allowed us to scale to large sweeps. Table 2 details the parameter ranges, the sampling densities, and the rendering engine.

Table 2: Sampling range and density for the manipulated parameters.

Parameter	**Value range**	**# Sampled images**	**Rendering engine**
Camera angle	$[0, 2\pi)$	200	Cycles
Hue	$[0, 2\pi)$	200	Cycles
Light intensity	$[0.1, 2]$	200	Cycles
Object size	$[0.5, 1]$	200	Cycles
Combined (two-param.)	-	10 000	Eevee

A.2. Additional results on topological analysis of ANNs

We repeated the topological analysis steps from Section 2.2 to umbrella renderings for light and size (Figure 4) and for hue and size (Figure 5), for individual and joint feature manipulation.

For light and size, we observe a 1-D interval replicating the topology of their parameter space, and a plane for their joint manipulation (Figure 4, first and second column). The representational diagrams in these three cases show a block-like structure, possibly corresponding to the different stages of ResNet-50 processing hierarchy (early convolution, then four bottleneck stages) (Figure 4, third column).

We perform the same analysis for a joint manipulation of hue and size, and show the results in Figure 5. The low-dimensional projection of the activations seem to form a cylindrical surface (Figure 5, first column), preserving the product topology of the parameter spaces from hue and object size. The expected persistent diagram would consist of one persistence H_1 feature, and one H_0 component persisting to infinity. However, we find multiple persistent H_1 points (Figure 5, second column) suggesting a more complex underlying structure. This may be due to sampling artifacts, as discretly sampling from the cylindrical surface could create transient cycles before revealing the circular topology. Finally, in the RDMs, we observe a pattern of distances similar to those in camera and hue and also in light and size, which reflects the different stages of the model.

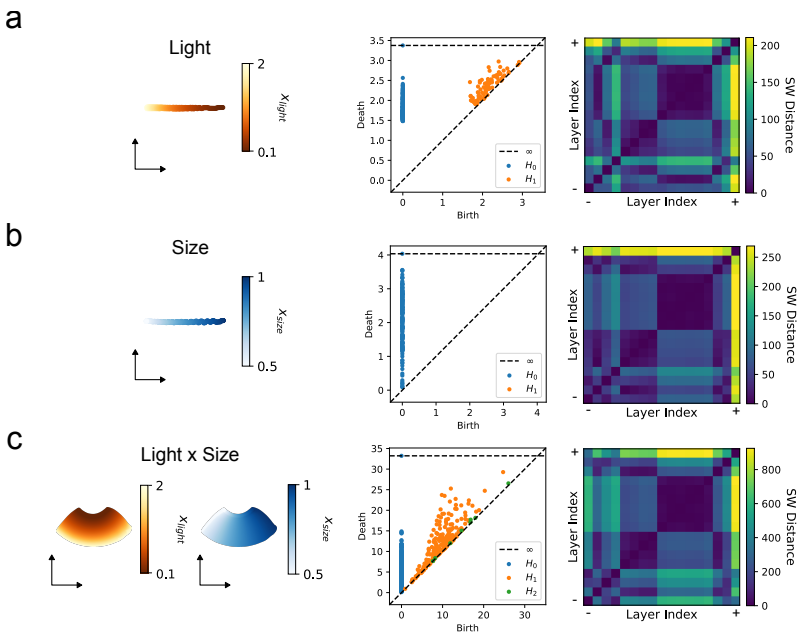

Figure 4: Topological analysis of individual (**a**, **b**) and combined (**c**) light and size manipulation in the umbrella renderings. The model's first ReLU activations are projected using Isomap (first column). From a subsample of such point cloud, a persistence diagram is computed (second column). Each layer is then characterized by their persistence diagram. Pairwise distances between layers are computed using Sliced Wasserstein distance, and a representation dissimilarity matrix is built (third colum). For the light-size product, 10,000 images were rendered, from which 1000 were subsampled for efficient computation of the persistence diagram.

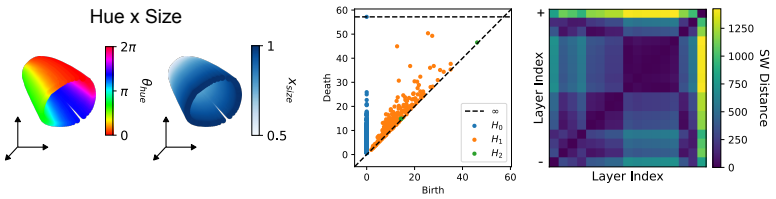

Figure 5: Topological analysis of joint hue-size manipulations in umbrella renderings. Columns are as for Figure 4. 10,000 images were rendered for the low-dimensional visualization of the network activations, and 1000 were subsampled when computing the persistence diagram.

A.3. Classification performance across the parameter space

We evaluate how single and joint feature manipulations affect classification performance by extracting the prediction probabilities across our parameter space. Figure 6 shows classification accuracy for umbrella under joint feature manipulation using ResNet-50. The model robustly maintains correct umbrella classification across most parameter combinations, indicating realistic rendering quality. For joint light and size variations (right panel), classification remains stable, except for a moderate accuracy drop at small sizes. However, joint hue and camera angle manipulations (left panel) reveal systematic vulnerabilities: accuracy drops significantly for red-to-orange hues and cyan colors, and when the object is viewed from the sides ($\theta_{\text{camera}} \approx \pi$ and $\theta_{\text{camera}} \approx \frac{\pi}{3}$). In these failure cases, the model misclassifies umbrella as chambered nautilus, bathing cap, parachute, pinwheel, or banana.

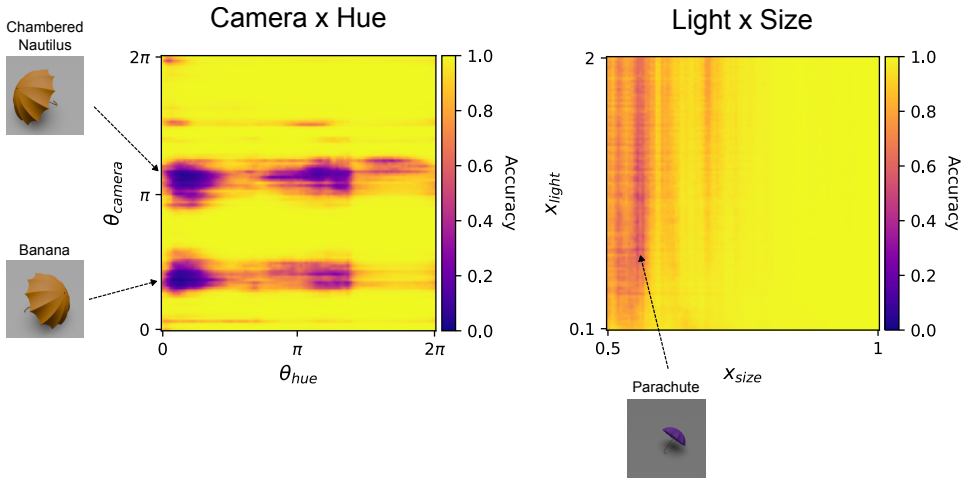

Figure 6: Classification accuracy heatmaps for umbrella across joint feature manipulation. **Left:** camera angle vs. hue. **Right:** light intensity vs. object size.