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

**Santiago Galella**[1,2] GALELLA@FIAS.UNI-FRANKFURT.DE
**Pamela Osuna-Vargas**[1,2] OSUNA@FIAS.UNI-FRANKFURT.DE
**Maren Wehrheim**[3,4] MAREN.WEHRHEIM@MILA.QUEBEC
**Matthias Kaschube**[1,2] KASCHUBE@FIAS.UNI-FRANKFURT.DE

1. *Frankfurt Institute for Advanced Studies, Frankfurt, Germany*
2. *Goethe University, Frankfurt, Germany*
3. *Mila - Quebec AI Institute, Montréal, Canada*
4. *York University, Toronto, Canada*

**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). 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 the topological structure of their input domain, the role such manifolds play in the organization of sensory representations remains less understood.

Artificial neural networks (ANNs) provide a controlled testbed to probe the topological structure of stimulus encoding, as they share key similarities with visual processing in the brain (Kietzmann et al., 2019), while offering reduced complexity, direct access to network activity and a lack of 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

© 2025 S. Galella, P. Osuna-Vargas, M. Wehrheim & M. Kaschube.

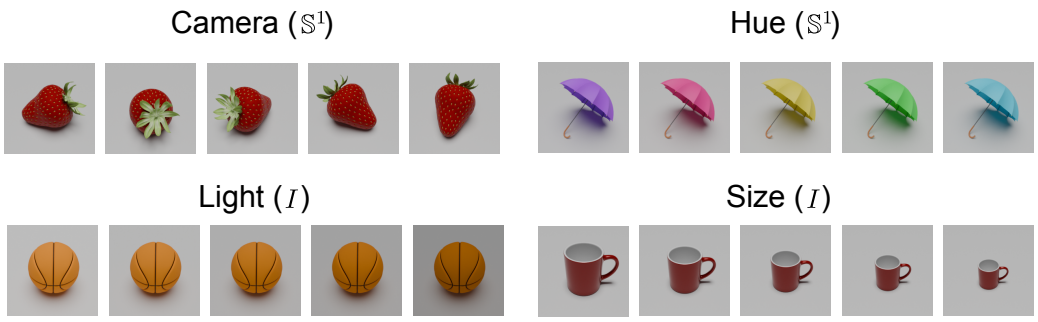


Figure 1: Overview of the MAPS dataset showcasing four feature manipulations: camera angle (azimuthal angle), hue, light, size. Cyclic features (hue, camera angle) have circular topology ($\mathbb{S}^1$), while bounded features (light, size) have interval topology ($I$).

(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 or size, while holding all other features constant, we trace out how such manipulations map to manifolds in the representational space of ANNs, investigating whether input topology is reflected in the network activations and how this topology evolves across the network hierarchy.

Our contributions: (i) we introduce MAPS, a high-resolution dataset of controlled stimuli for probing representational topology in vision models; (ii) using topological data analysis in ResNet-50, we show that early-layer activity patterns may preserve topological structure present in the input space.

## 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, see Appendix A.1 for details).

Each transformation corresponds to a specific underlying topology. Hue shifts and camera rotations form circles ($\mathbb{S}^1$), while scaling and light intensity changes form intervals ($I = [a, b]$ for $a, b \in \mathbb{R}^+$ and $b > a$). The parameter space of composed transformations has a topology given by the product of the individual manifolds. For instance, combining hue shifts with camera rotations yields a parameter space with topology $\mathbb{S}^1 \times \mathbb{S}^1 = \mathbb{T}^2$, i.e., a 2-torus.

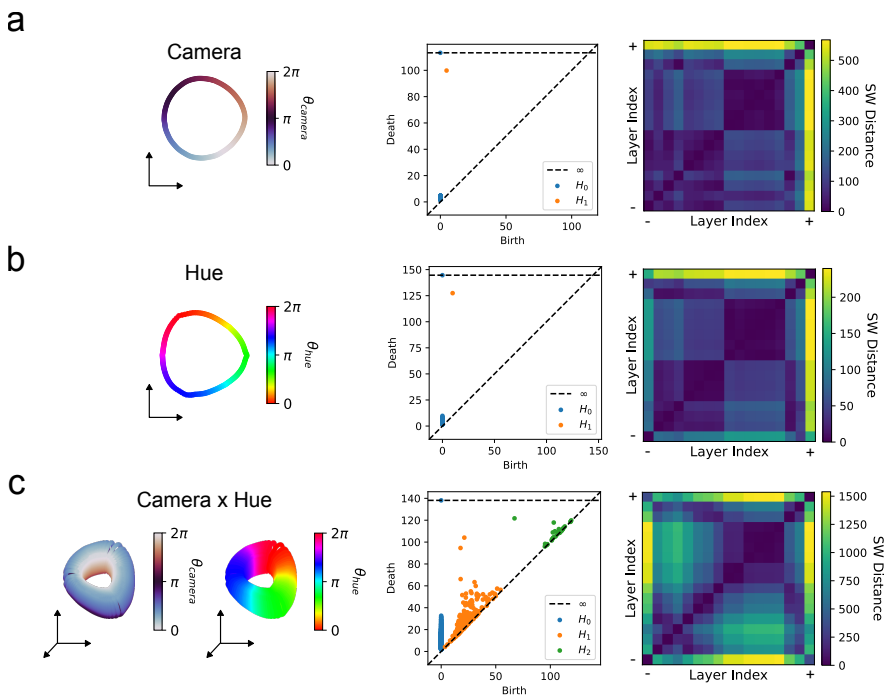


Figure 2: Topological analysis of individual (**a**, **b**) and joint (**c**) camera angle and hue manipulations in umbrella renderings. The model's first ReLU activations are projected using Isomap (first column). From a subsample of such point cloud, a persistence diagram is computed (second column). 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 **c**, 10,000 images were rendered, and 1000 subsampled.

### 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 in Appendix A.1.

Following Gardner et al. (2022), we projected neural activations onto 10 principal components using PCA. We then 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 (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), computing a weighted average of the $H_0$ and $H_1$ features for single transformations, and additionally including $H_2$ for product manifolds. For visualization, we used Isomap (Tenenbaum et al., 2000) to the PCA-projected activations (20 neighbors). Figure 4 in the Appendix A.2 shows an overview of the process detailed in this section.

## 3. Results & Analysis

Figure 2 shows the representational topology for three manifolds of the umbrella object from MAPS: (a) camera rotation, (b) hue shift, and (c) their joint transformation. For each, we display low-dimensional projections (first column), persistence diagrams from the first ReLU layer (second column), and representational dissimilarity matrices (RDMs) measuring Sliced Wasserstein distances between persistence diagrams across layers (third column).

Right after the first ReLU activation, ResNet-50 captures the parameter space topology for both single-parameter and product manifolds. The persistence diagrams reveal correct topological signatures: single loops ($H_1$) for hue and camera transformations, and torus topology with two loops ($H_1$) and one void ($H_2$) for their combination (Fig. 2, second column). However, the manifolds behave differently across the network hierarchy. Both camera and hue maintain low SW distance across layers, though hue shows smaller SW distance between the first and last ReLU layers (Fig. 2, third column). These differences arise primarily from shifts in the birth–death scales of the loops rather than structural changes. For the camera-hue product manifold, torus topology emerges in early layers but disappears in deeper layers, leaving only residual topology in the final representations.

Additional parameters show similar topology preservation in early layers (Appendix A.2).Classification performance across parameter variations are in 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 varying individual features of 3D objects. Unlike existing datasets, which focus on disentanglement through sparse parameter sampling or provide multi-view images at low resolution, MAPS densely samples the parameter spaces at high resolution to capture continuous manifold structure. We demonstrated the usefulness of MAPS for studying the topology of representations in ANNs. Our exploratory results suggest that ResNet-50, despite being trained solely on isolated images, develops internal representations that mirror key topological features of the input space.

MAPS has some limitations: it relies on synthetic scenes with reduced ecological complexity and object diversity, and it varies only a small set of factors (e.g., hue, rotation, lighting) that are typically coupled in natural vision. These simplifications may not fully capture natural image statistics and can yield cleaner or simpler topological structures than those found in real environments. Nevertheless, our preliminary results show that controlled parametric datasets, combined with topological analysis, could reveal fundamental principles of representational structure in both artificial and biological neural systems. 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.

## Acknowledgments

We thank Jens-Bastian Eppler and Lorenzo Butti for helpful discussions. This research was supported by the German Research Foundation (DFG) through the research units FOR 5368 ARENA, SPP 2041 Computational Connectomics and GRK 2566 iMOL, the Speyer'sche Hochschulstiftung, and the Connected Minds Postdoctoral Fellowship (supported by CFREF).

## 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.

Tim C. Kietzmann, Patrick McClure, and Nikolaus Kriegeskorte. Deep neural networks in computational neuroscience, 01 2019. URL https://oxfordre.com/neuroscience/view/10.1093/acrefore/9780190264086.001.0001/acrefore-9780190264086-e-46.

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 their 3D models from Sketchfab (see Table 1). Because ImageNet includes both very specific and general object categories, we focused on 10 accessible, widely recognizable objects. 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 $224 \times 224$ pixel resolution to match the input size of pretrained vision models (Figure 3). Transformations were scripted using the Blender Python API.

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 |

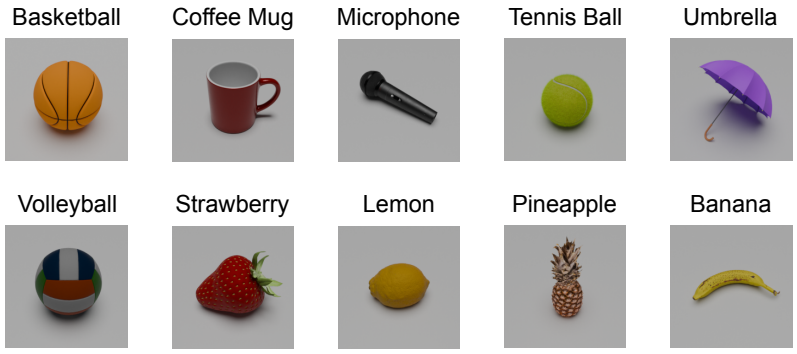


Figure 3: 3D renderings of all objects in MAPS.

For each transformation parameter, we defined value ranges and linearly sampled points within these ranges. For camera angle, we swept $\theta_{\text{camera}} \in [0, 2\pi)$; for hue, $\theta_{\text{hue}} \in [0, 2\pi)$;

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 5) and for hue and size (Figure 6), for individual and joint feature manipulation. Figure 4 shows the step of the pipeline to extract the topology from neural activations.

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 5, 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 5, third column).

We perform the same analysis for a joint manipulation of hue and size, and show the results in Figure 6. The low-dimensional projection of the activations seem to form a cylindrical surface (Figure 6, 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 6, second column) suggesting a more complex underlying structure. This may be due to sampling artifacts, as discretely 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.

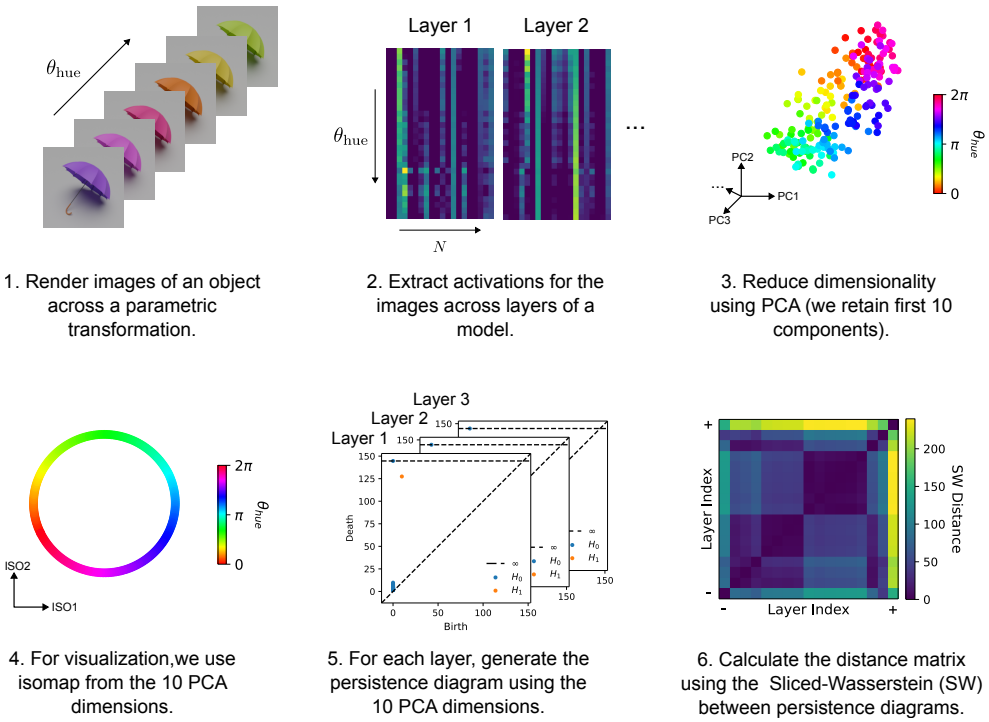


Figure 4: Pipeline to extract the topology of neural representations in ANNs.

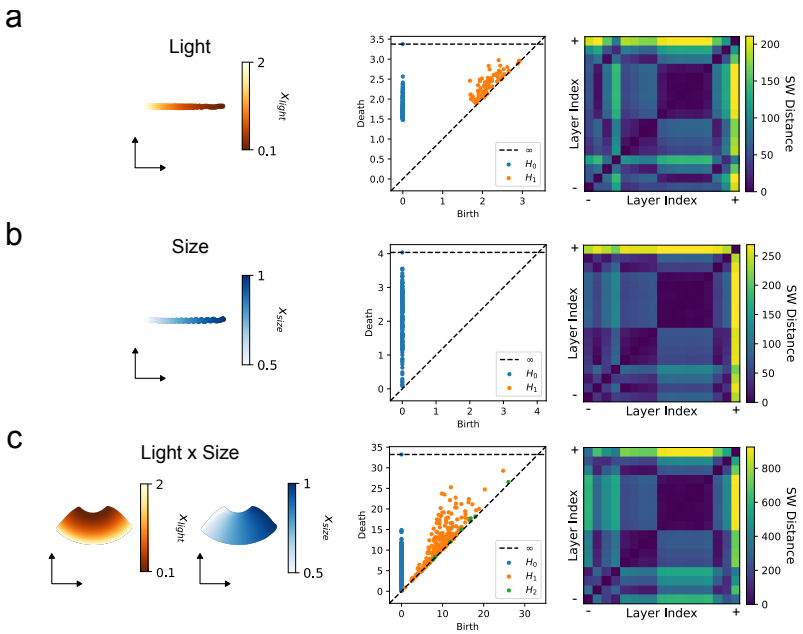


Figure 5: Topological analysis of individual (**a**, **b**) and combined (**c**) light and size manipulation in the umbrella renderings. The model's first ReLU activations are projected using Isomap (first column). From a subsample of such point cloud, a persistence diagram is computed (second column). 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.

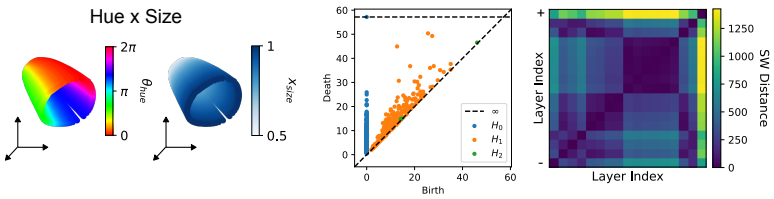


Figure 6: Topological analysis of joint hue-size manipulations in umbrella renderings. Columns are as for Figure 5. 10,000 images were rendered for the low-dimensional visualization of the network activations, and 1000 were subsampled when computing the persistence diagram.

### 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 7 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.

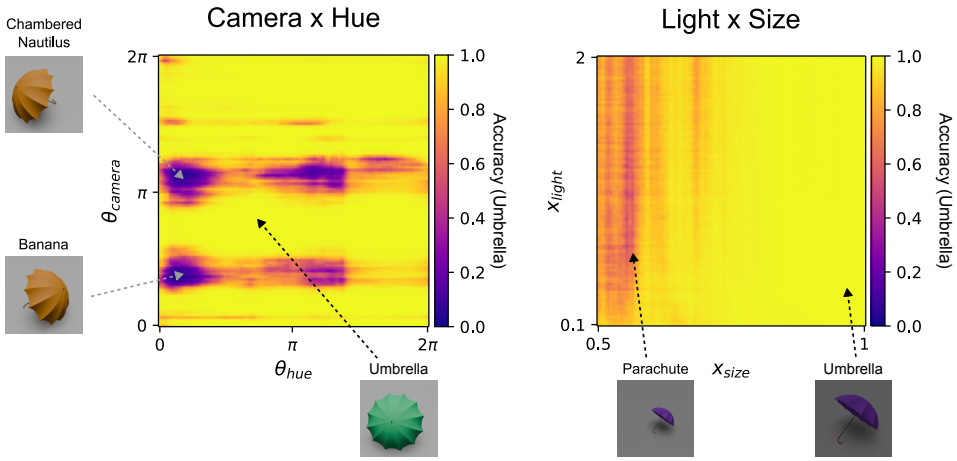


Figure 7: Classification accuracy heatmaps for umbrella across joint feature manipulation. **Left:** camera angle vs. hue. **Right:** light intensity vs. object size.