KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

SHAP-Explainable Image-to-Topology Regression

Charles Fanning
Supervisor: Dr. Bin Luo

GRP-021

INTRODUCTION

Intel Scene Classification Original Images

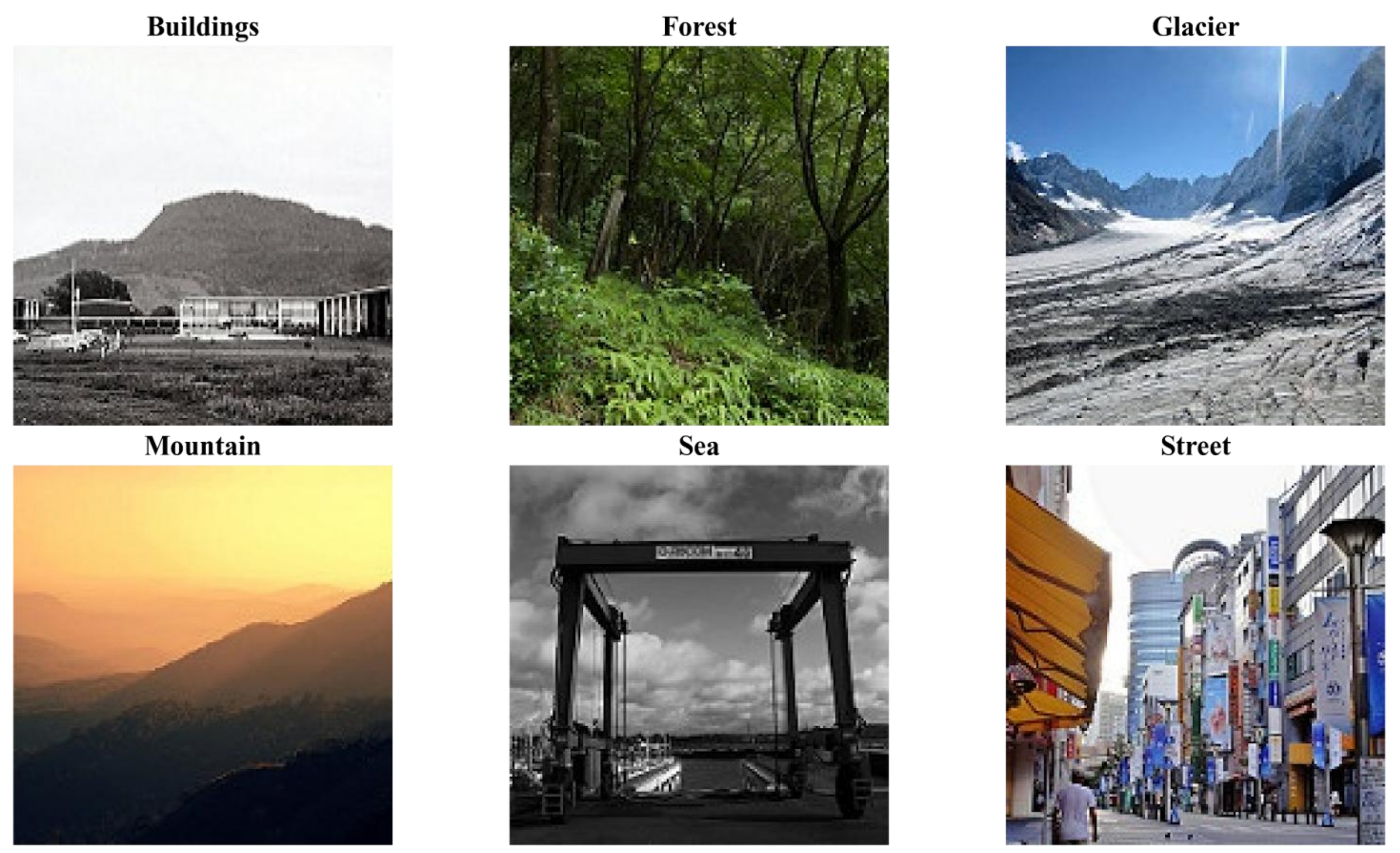

Figure 1. *Samples from the Intel Scene Classification dataset, each from one of the six scene categories: Buildings, Forest, Glacier, Mountain, Sea, and Street.*

Intel Scene Classification Dataset

- ~25,000 color images of natural and urban scenes
- Six categories: Buildings, Forest, Glacier, Mountain, Sea, Street
- Most images are 150 × 150 pixels (minor outliers excluded)
- Originally released for an Analytics Vidhya challenge by Intel
- Commonly used for scene classification benchmarks
- Pre-split into:
  - ~14,000 training
  - ~7,000 validation
  - ~3,000 test

Average RGB Color Channel Distributions per Class

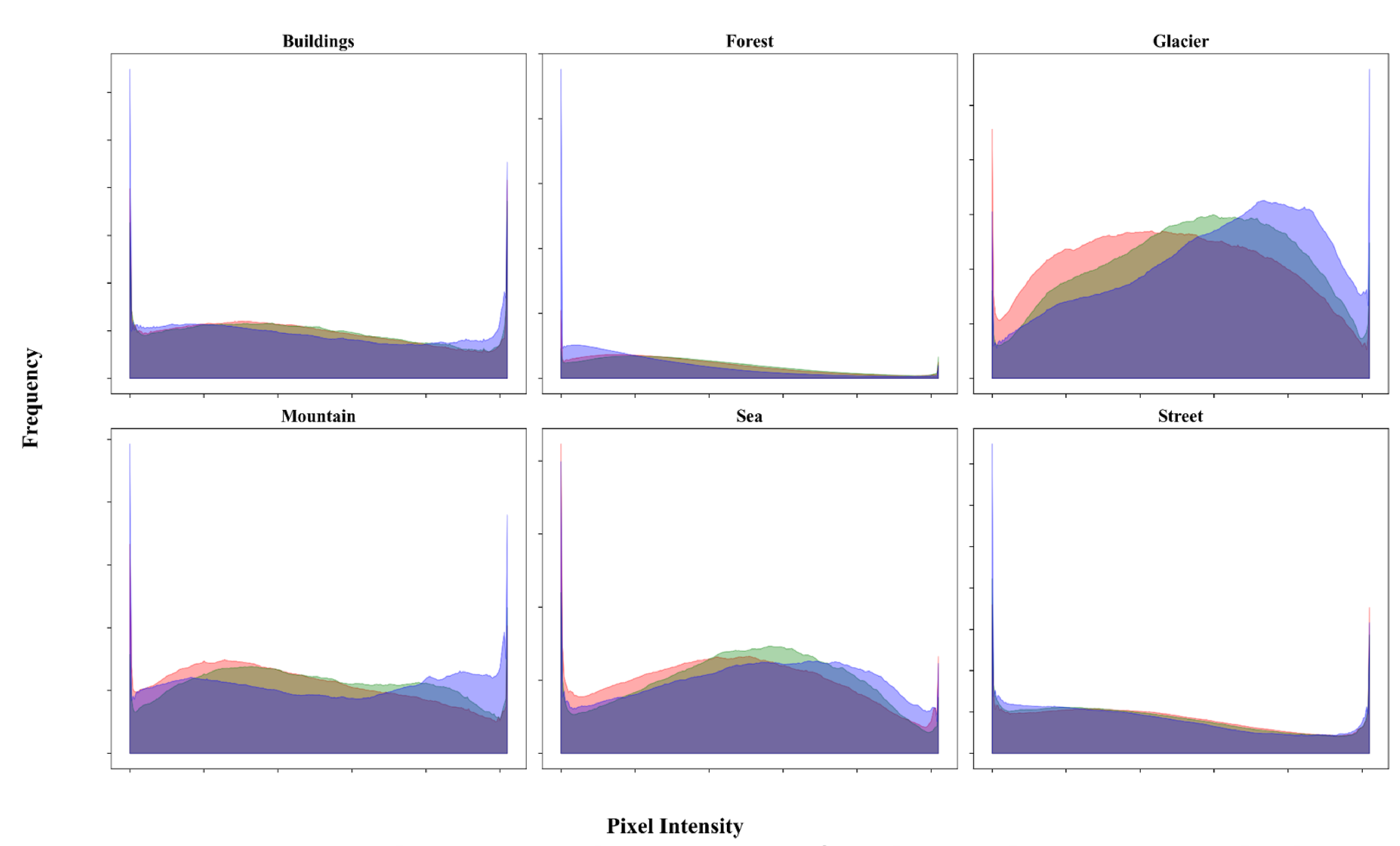

Figure 2. *Average pixel-intensity histograms for the Red, Green, and Blue channels across the entire dataset.*

Motivation

Previous work has used deep learning models to **predict persistent homology**–based representations of data. However, we still are not sure whether these models *actually learn* topological structure or are just overfit to the output vectorization. We evaluate this by analyzing which input regions most influence the predicted persistence landscape using **SHAP**.

We train a **DenseNet-121** model to predict 300-dimensional **persistence landscape** vectors derived from **cubical filtrations** on grayscale versions of scene images. We then compare the SHAP attribution maps to the hole locations extracted from persistent homology alone.

METHODS

Persistence Diagram of Intel Scene Classification Images

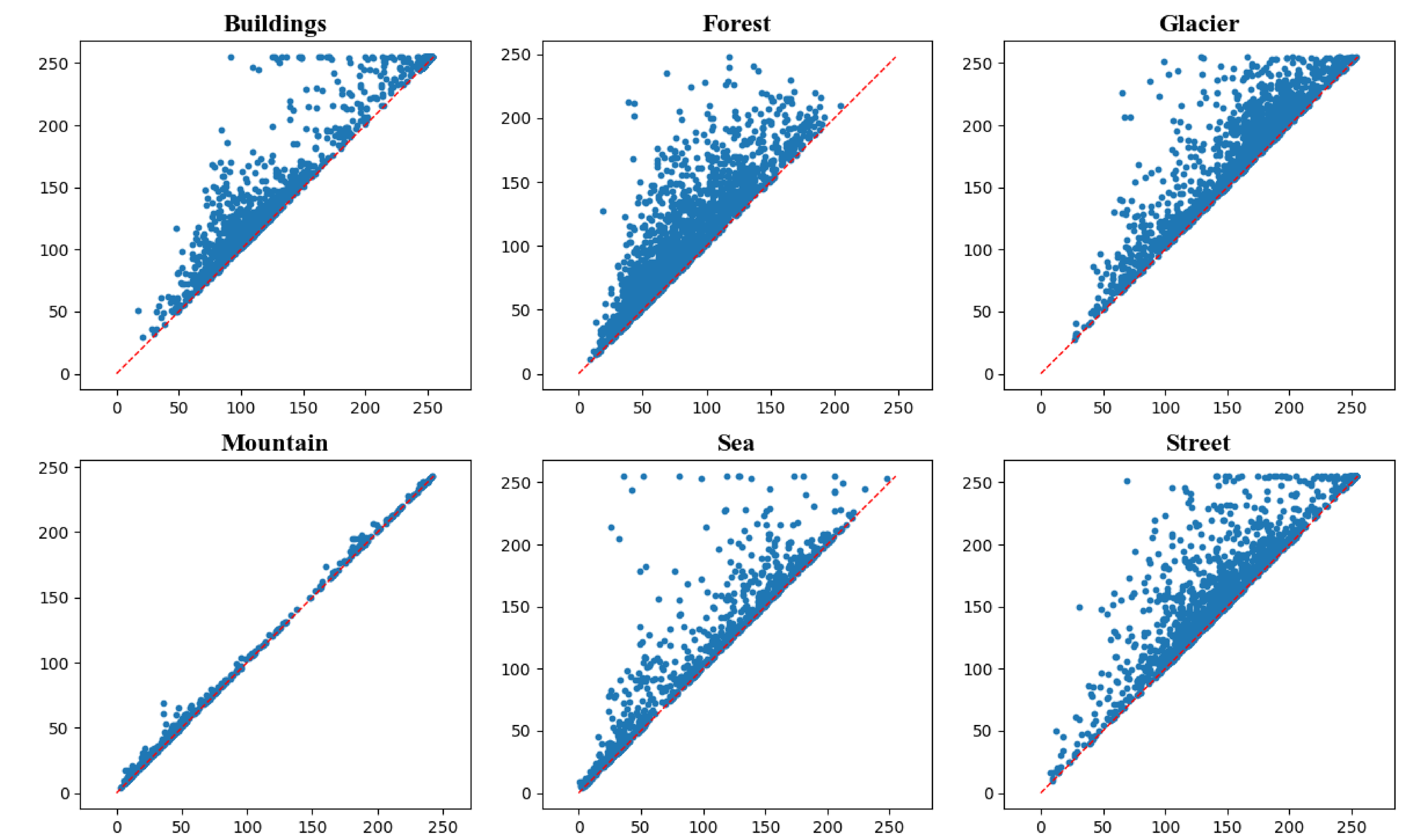

Figure 3. *Persistence diagrams for one representative image from each class in the Intel Scene Classification dataset. Each point represents the birth and death times of a topological feature, plotted on the horizontal and vertical axes, respectively.*

Persistence Landscapes of Intel Scene Classification Images

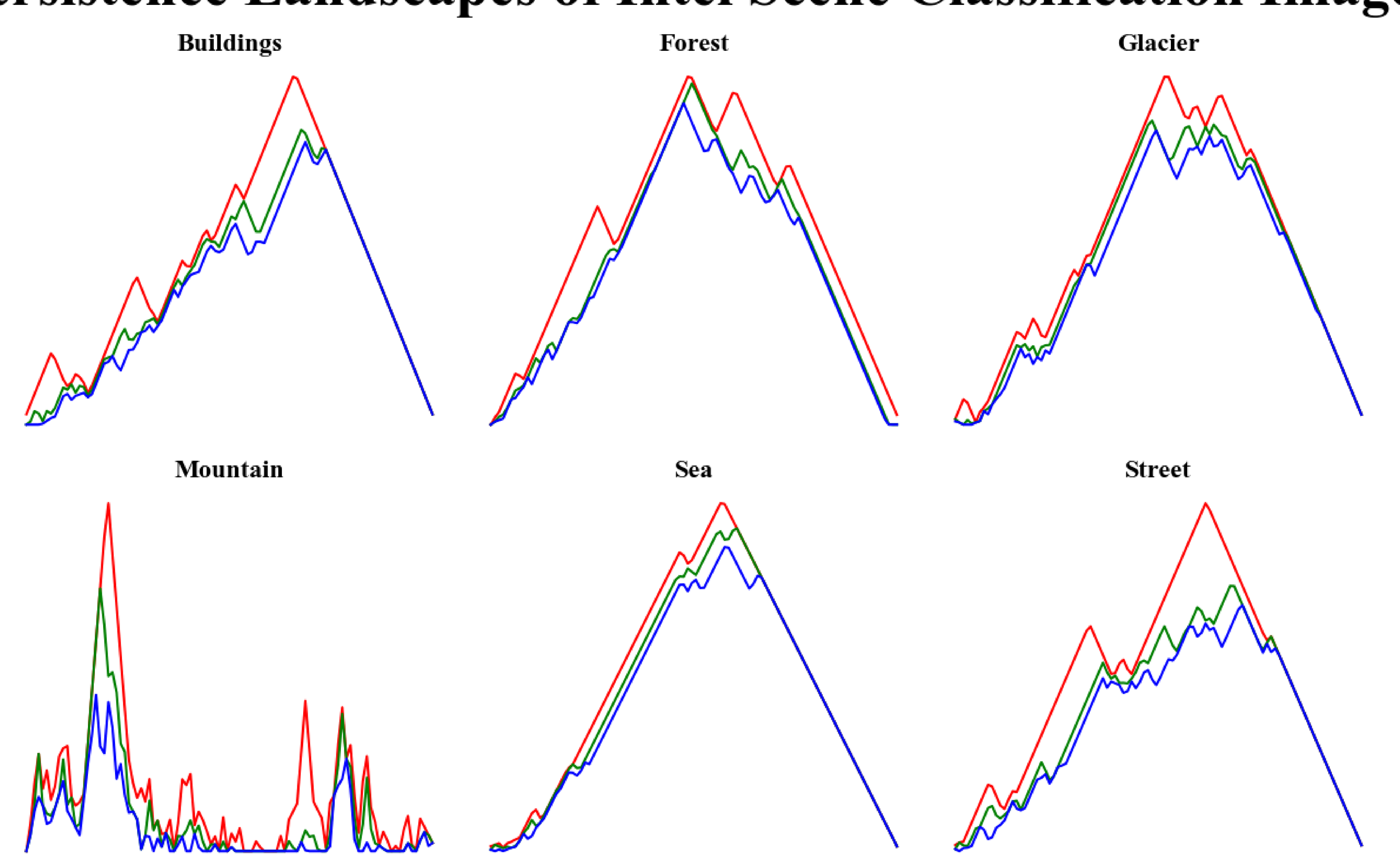

Figure 4. *Persistence landscapes for representative images from each class in the Intel Scene Classification dataset. Each plot shows the first three landscape functions.*

Predicted vs True Persistence Landscapes

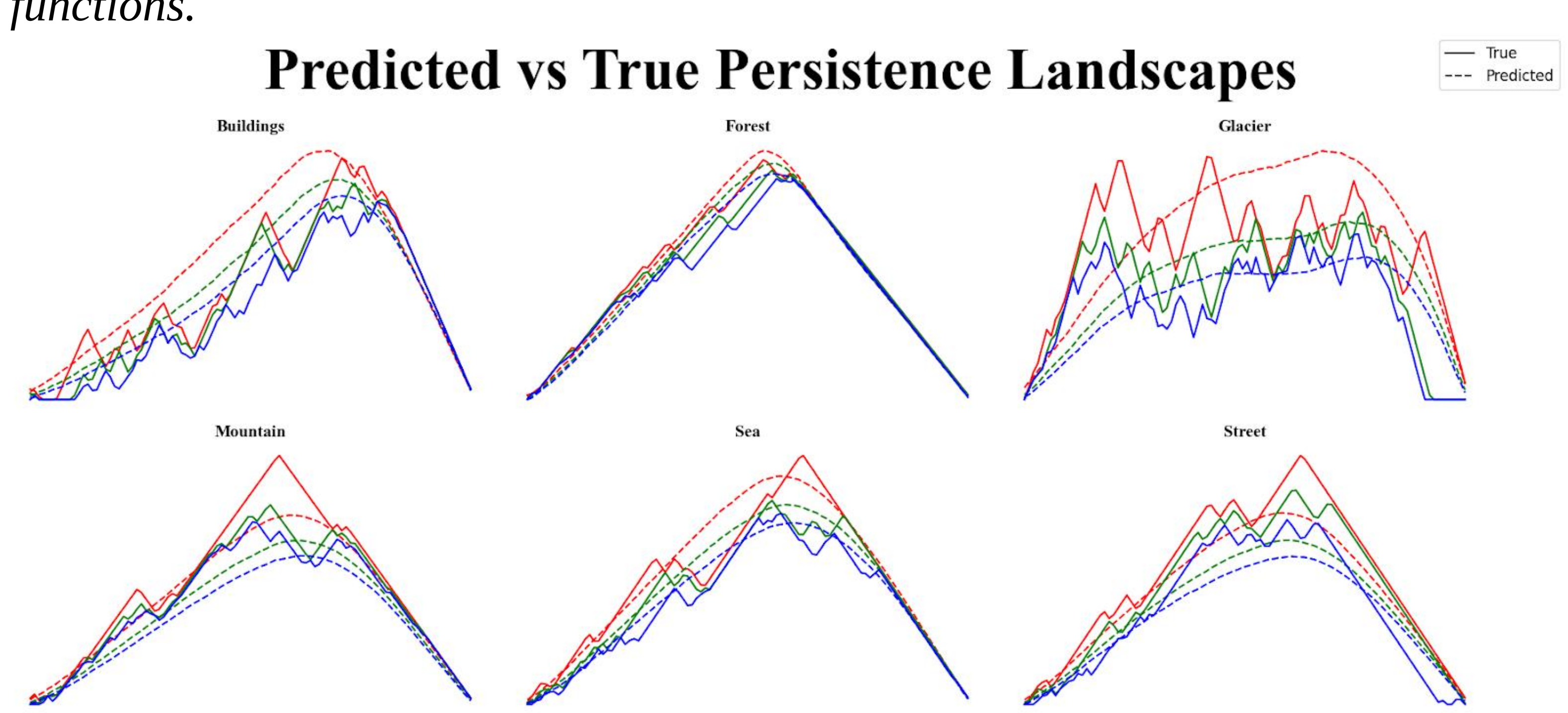

Figure 5. *Predicted vs. true persistence landscapes for representative images from each class in the Intel Scene Classification dataset. Each plot shows the first three landscape functions (red, green, and blue, respectively), with the model's predicted curves (dotted) compared to the ground truth (solid).*

RESULTS & DISCUSSION

Combined SHAP (Blue) & Topological (Red) Dot Overlays

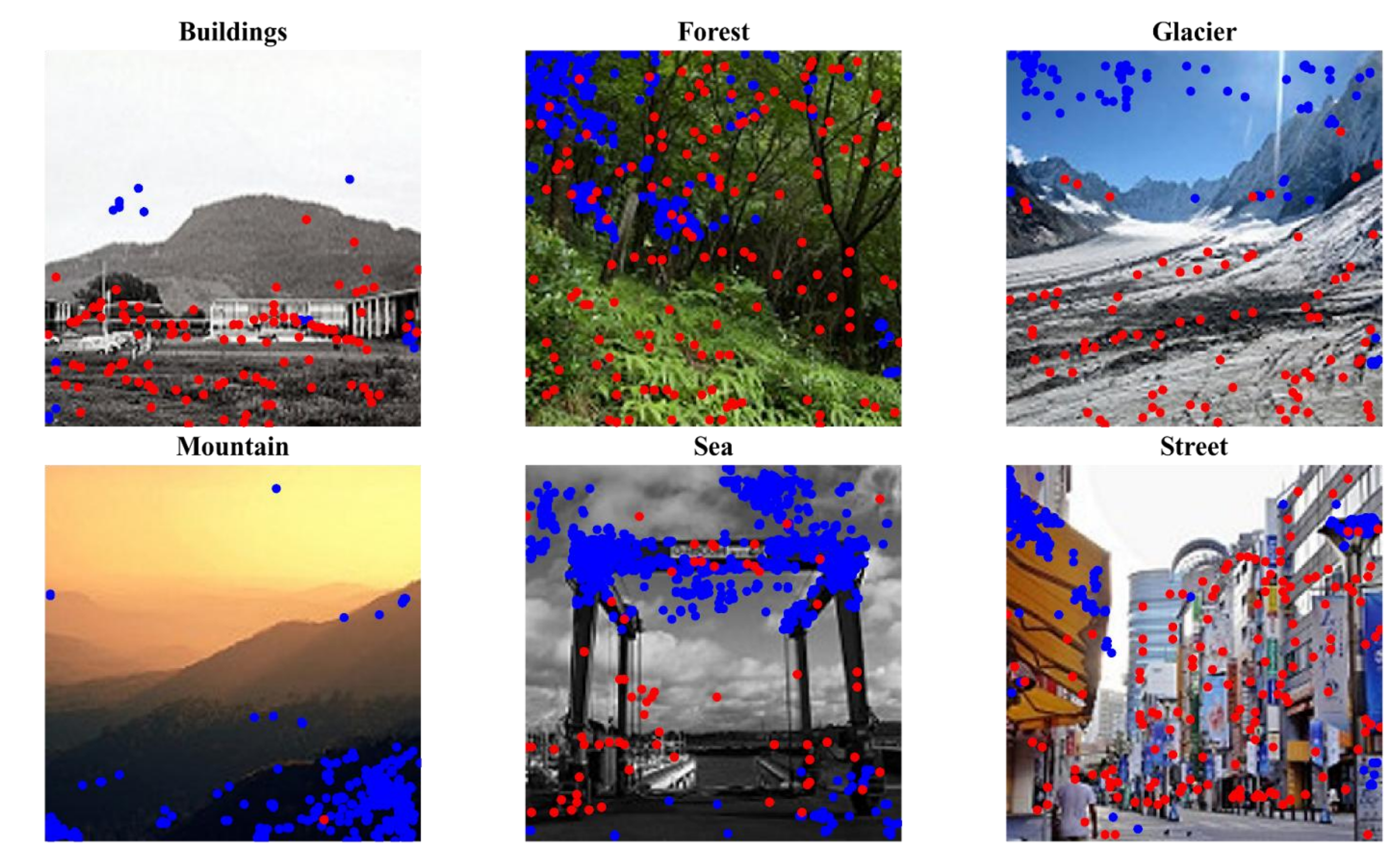

Figure 6. Combined overlay of SHAP (blue) and topological (red) dot positions for each scene category. Pixels with SHAP values above a threshold are marked in blue. Pixels with 1-dimensional topological features with persistence values above a certain threshold are marked in red.

Class	Mean Overlap Ratio	Mean Baseline Ratio	Difference
buildings	0.0068	0.0091	-0.0023
sea	0.0059	0.028	-0.022
street	0.0054	0.0089	-0.0035
mountain	0.0033	0.0072	-0.0039
glacier	0.0026	0.011	-0.008
forest	0.0022	0.0036	-0.0014

Figure 7. Mean overlap ratio between SHAP explained heatmaps and 1-dimensional topological features across each class. A baseline is given between thresholded SHAP-identified pixels against the heatmap baseline.

Discussion

Across all six classes, the overlap between **SHAP** attribution heatmaps and 1-dimensional topological features was consistently lower than the baseline (Figure 7). The worst performing observation was in the *sea* category, where the **SHAP**-topology overlap was 2.2 percent lower than baseline. The smallest difference was in the forest class, with a difference of 0.14%.

Overall, SHAP explanations consistently aligned with 1-dimensional topological features less than random sampling would suggest.

The predicted **persistence landscapes** from the **DenseNet-121** model are significantly different from those produced by **persistent homology**. The **persistence landscapes** generated through persistent homology are *rigid* and exhibit many *local minima* and *maxima*. The predicted ones tend to be smoother and either unimodal or bimodal.

There are also significant differences in threshold behaviors between the two approaches. The number of **persistent features** — features that contribute significantly to persistent homology — vary independently of each other. That is, the number of persistent features detected via **SHAP** is not determined or even dependent on the number of persistent features determined by persistent homology.

Sources

[1] Analytics Vidhya. (n.d.). *Practice problem: Intel Scene Classification Challenge*. Retrieved from https://www.analyticsvidhya.com/datahack/contest/practice-problem-intel-scene-classification-challe/

[2] Zomorodian, A., & Carlsson, G. E. (2004). Computing persistent homology. *Discrete & Computational Geometry*, 33(2), 249–274. https://doi.org/10.1007/s00454-004-1146-2

[3] Bubenik, P. (2012). Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16, 77-102. Retrieved from http://www.jmlr.org/papers/volume16/12-167/12-167.pdf

[4] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. Retrieved from https://arxiv.org/abs/1705.07874