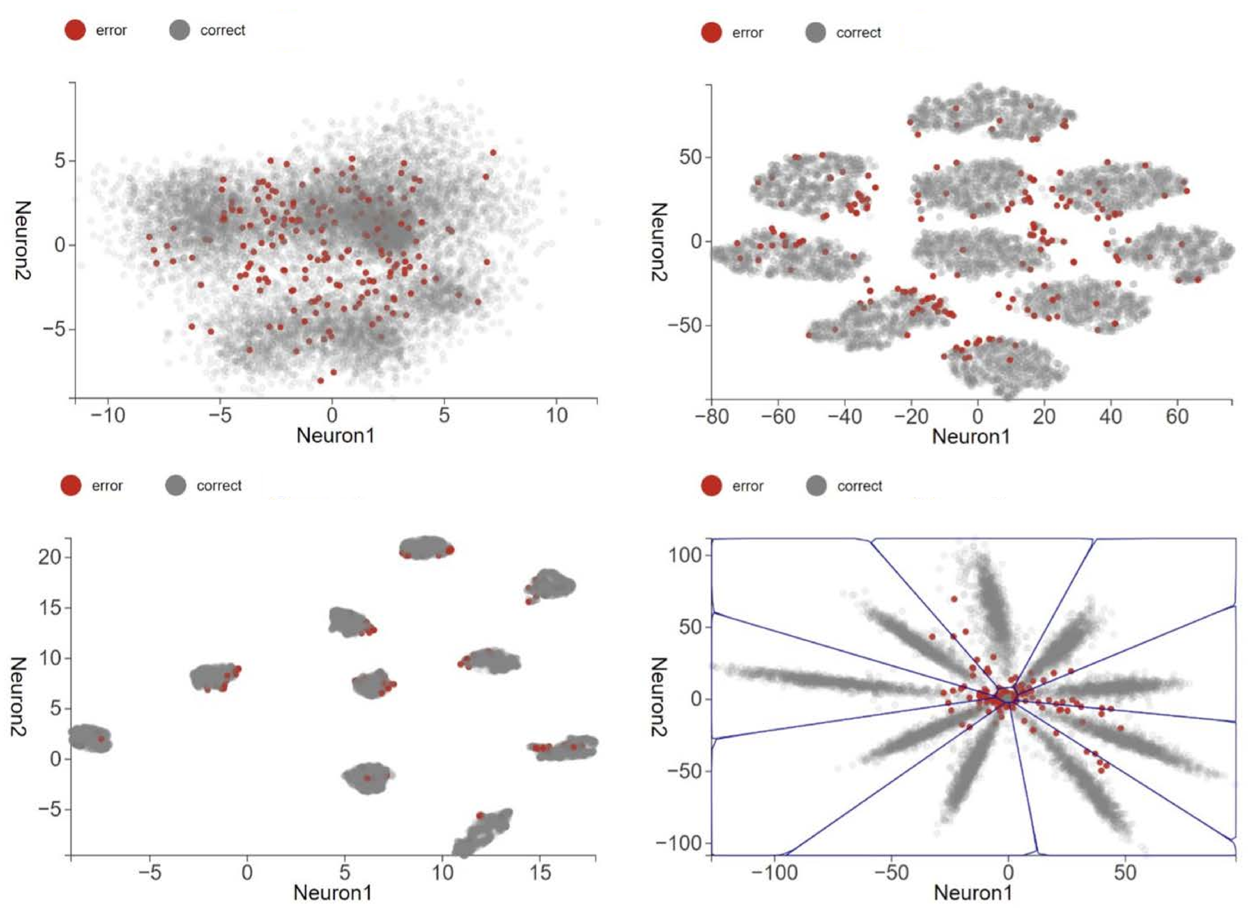
A Geometric Visual Comparative Analysis of Neural Network Model

Zhimin Li (U of U), Shusen Liu (LLNL), Xin Yu (U of U), Kaikhura Bhavya (LLNL), Jie Cao (U of U), James Daniel Diffenderfer (LLNL), Peer-Timo Bremer (LLNL), Valerio Pascucci (U of U)

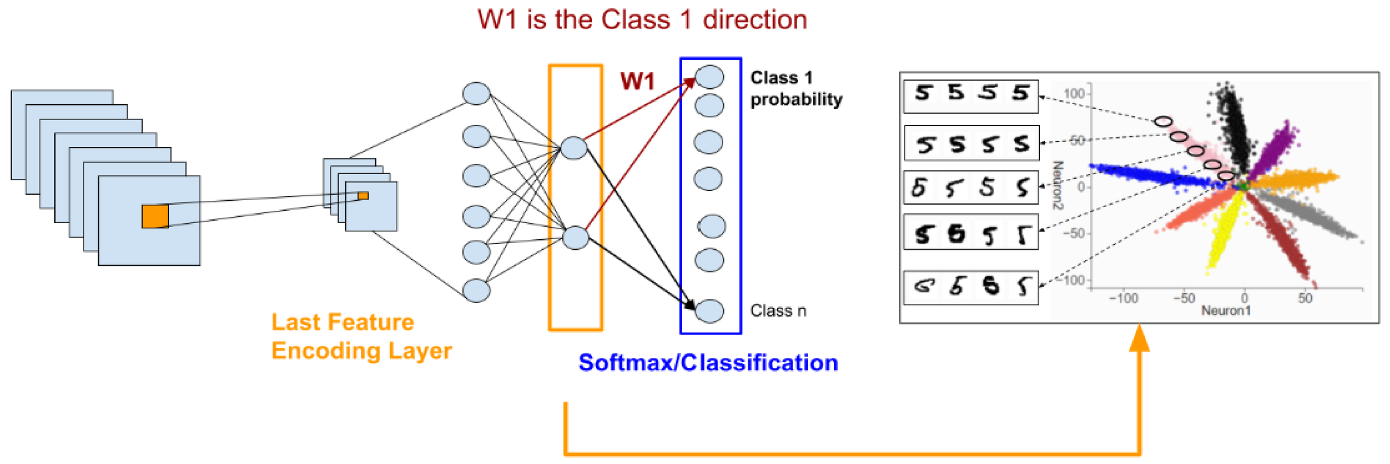
Understand the High Dimensional Latent Feature Space of Neural Network Model is Challenging

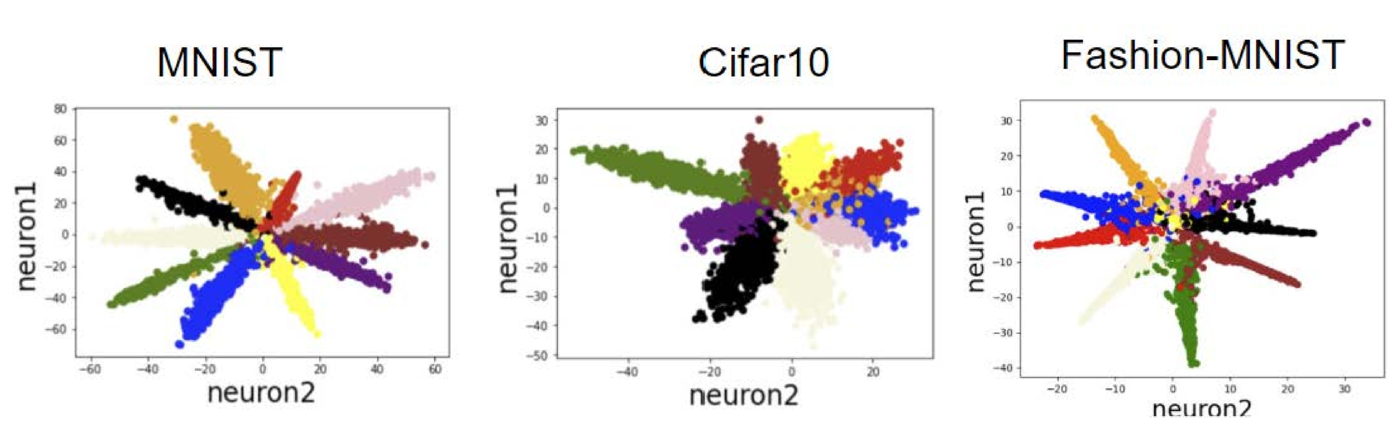


The visualization of PCA, t-SNE, UMap and model probe feature embedding on neural network latent feature space (MNIST dataset)

- High dimensional space is naturally difficult for human to understand
- Dimension reduction which projects data into low dimensional space distorts the distance between samples and leads to highly uncertain results.

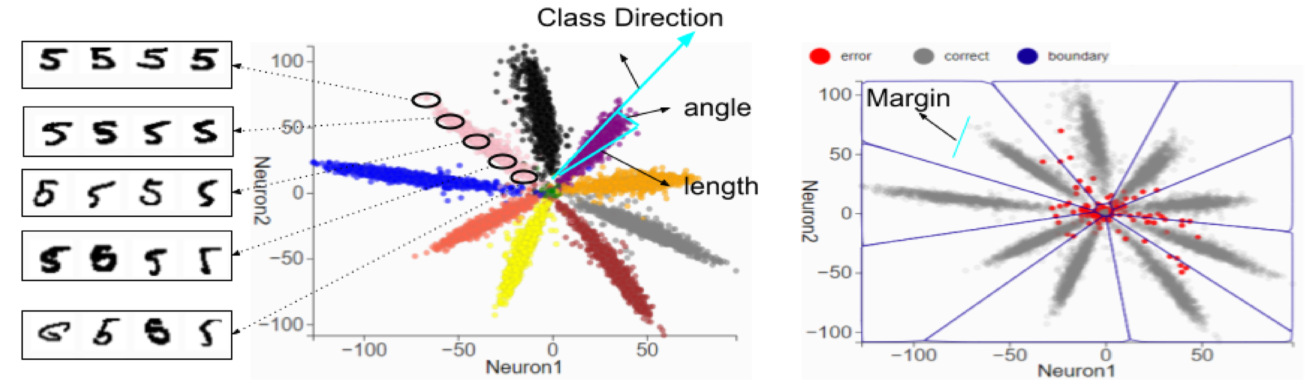
The Geometric Shape of Feature Representation in 2D Space





- An intuition about the shape of feature representation in high dimensional space is helpful.

Describe the Geometry of Feature Representation in High Dimensional Space



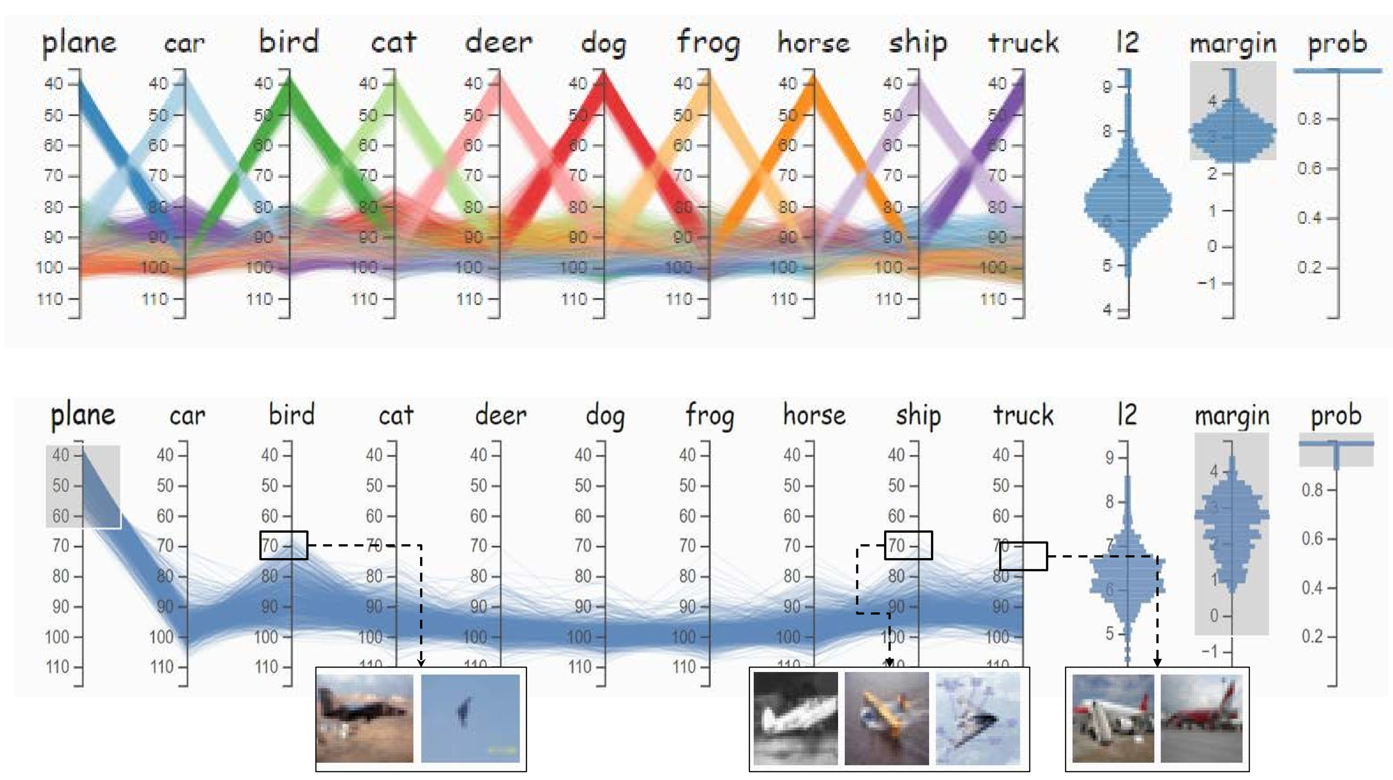
Class Direction: represents a class direction in high dimensional space (e.g. dog, cat).

Angle: is the angle between class directions and feature representation vectors.

Margin: is the minimum distance to the decision boundary.

Length (L2 Norm): is the L2 norm of feature vectors.

Visualize the Geometry of Feature Representation in High Dimensional Space



Why We Care about the Geometry of Feature Representation in High Dimensional Space?

CNN Architecture	rc-angle	rc-l2	rc-margin.
VGG16	-0.636	0.516	0.685
resnet18	-0.715	0.192	0.696
resnet50	-0.7172	0.053	0.6677
resnet152	-0.7158	-0.035	0.7228
densenet121	-0.726	-0.015	0.6732

TABLE 2

The result of the **Imagenet-C** validation dataset. This table shows the Pearson correlation coefficient between different geometric features and models' robustness (*supplementary equation (1)*). Angle and margin show a significant correlation with robustness. The correlation between l2 and robustness is moderate or subtle.

- A sample's geometric location is correlated with its decision robustness

CNN Architecture	rc-angle	rc-l2	rc-margin.
VGG16	-0.6739	0.3924	0.6534
VGG19	-0.6667	0.3984	0.6514
resnet18	-0.6795	0.1737	0.7282
resnet50	-0.6933	0.083	0.7058
Densenet121	-0.6816	0.0252	0.7154

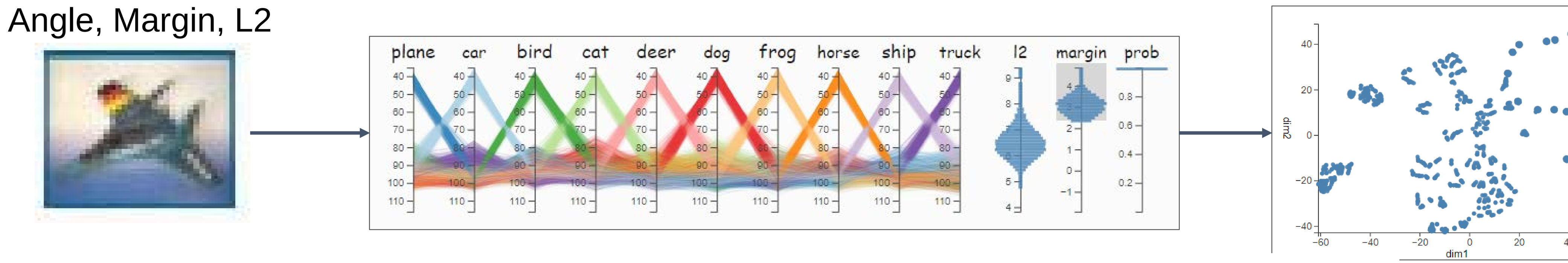
TABLE 3

The result of the **Imagenet** validation dataset. This table shows the Pearson correlation coefficient between different geometric features and model magnitude pruning (*supplementary equation (2)*). Angle and margin show a significant correlation with pruning vulnerability. The correlation between l2 and pruning vulnerability is subtle.

- A sample's geometric location is correlated with network pruning

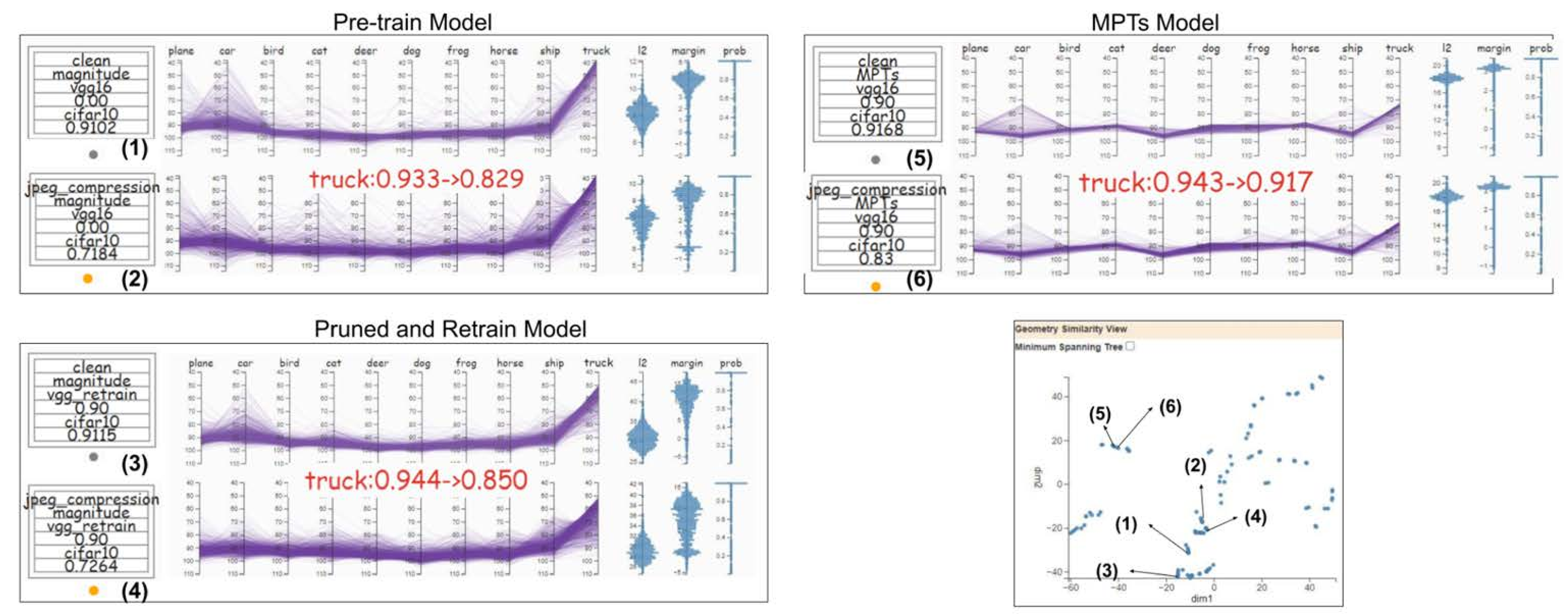
A Scalable Approach to Compare the Geometry of Neural Network Feature Representation

Angle, Margin, L2



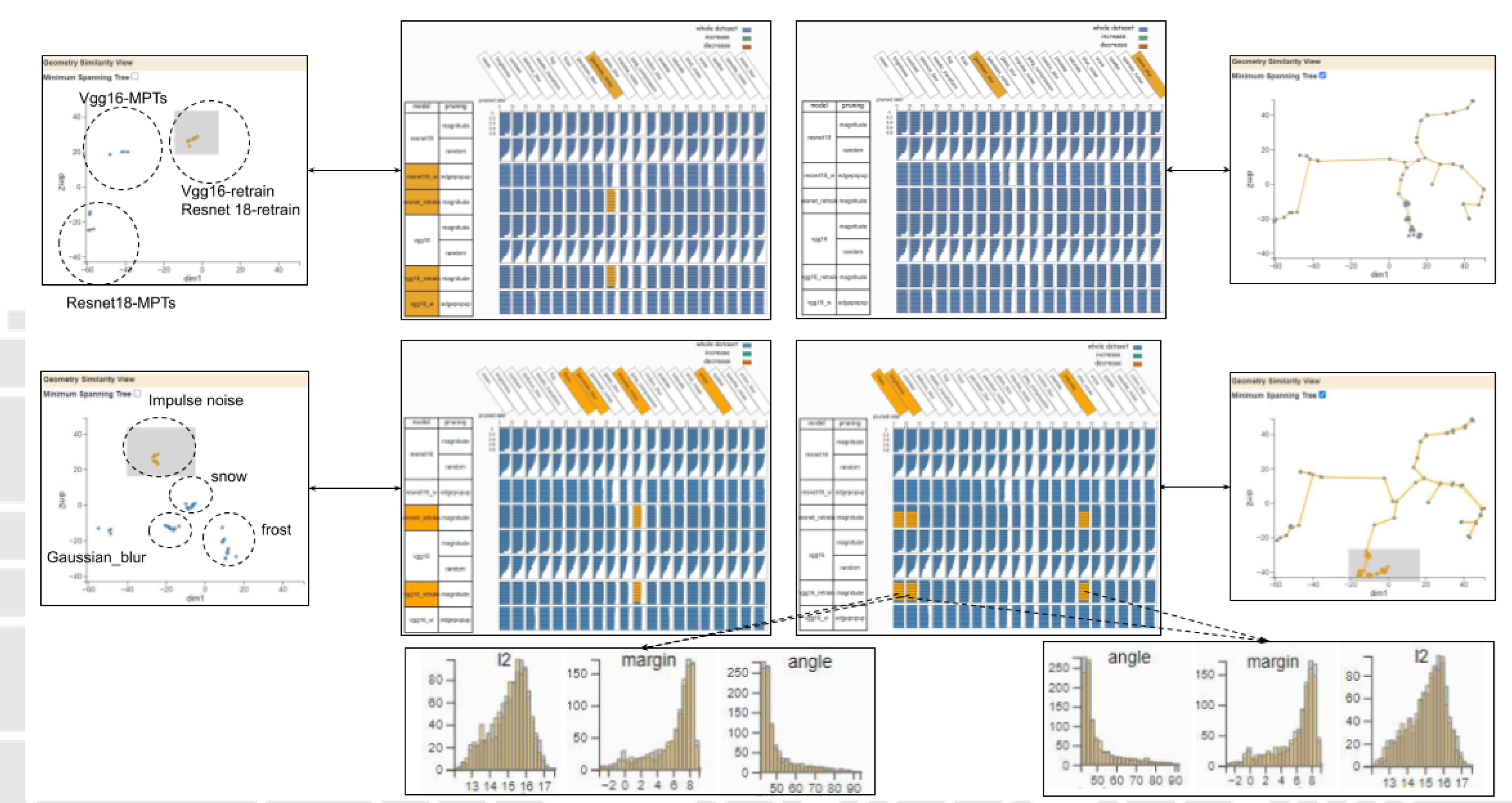
- Using the geometric property of samples to construct a model geometric embedding for a scalable model comparison.

What Is the Geometric Difference Between the State-of-Art Model and a Regular Model?



- The MPTs model which is the state-of-art robustness model has a different geometric feature distribution than the regular trained model

Comparing the Similarity Between Evaluation Benchmarks/Network Pruning



- We can utilize the geometric properties of feature representation to compare the similarity between evaluation benchmark and pruning methods.

THE UNIVERSITY OF UTAH

www.sci.utah.edu