ShapeY: Measuring Shape Recognition Capacity Using Nearest Neighbor Matching

1. The Core Capability of a Recognizing System: Knowing What Looks Like What

The main idea behind the ShapeY benchmarking system:
For a competent recognizing system, a reference view should match most closely to another view of the same object, regardless of the number of distractor object views.

2. The ShapeY Dataset

• 20 different object categories (Airplane, Chair, Lamp, etc)
• 10 instances in each category
• 341 3-D views of each object
• Total of 68,200 images

3. Generating the 3-D Views of Each Object

• Each object was rendered in 31 “series” of 11 3-D views (341 views total)
• Series involved shifts (X, Y) and/or rotations (Pitch, Roll, yaw)

4. Benchmarking Task

5. Results (ResNet50, pre-trained on ImageNet)

6. Examples of Matching Errors for ResNet50

7. Adding Additional Challenge: Contrast Reversal

8. Conclusion

• ShapeY measures shape recognition capacity using simple nearest-neighbor matching in the embedding space
• ResNet50’s embedding space is badly tangled for purposes of shape recognition, leading to poor matching performance even with mild 3-D viewpoint transformations
• Other training schemes such as the SimCLRv1 showed even worse performance
• ShapeY currently contains ~1% of the number of shape categories that humans easily master (including children); we expect performance of systems like ResNet to deteriorate to near zero as the number of object categories is increased to human levels.

Find ShapeY here: https://github.com/njw0709/ShapeY