

Human Visual Object Similarity Judgments are Viewpoint-Invariant and Part-Based as Revealed via Metric Learning

Joseph German

University of Rochester, Rochester, New York, United States

Robert Jacobs

University of Rochester, Rochester, New York, United States

Abstract

We describe and analyze the performance of metric learning systems, including deep neural networks (DNNs), on a new dataset of human similarity judgments of Fribbles, naturalistic, part-based objects. Metrics trained using pixel-based or DNN-based representations fail to explain our experimental data, but a metric trained with a viewpoint-invariant, part-based representation produces a good fit. We also find that although neural networks can learn to extract the part-based representation—and therefore should be capable of learning to model our data—networks trained with a triplet loss function based on similarity judgments do not perform well. We analyze this failure, providing a mathematical description of the relationship between the metric learning objective function and the triplet loss function. The comparatively poor performance of neural networks appears to be due to the nonconvexity of the optimization problem in network weight space. We discuss the implications for neural network research as a whole.