Hessisches Kompetenzzentrum
für Hochleistungsrechnen

Predicting Human Similarity Judgments Using Normalizing Flows

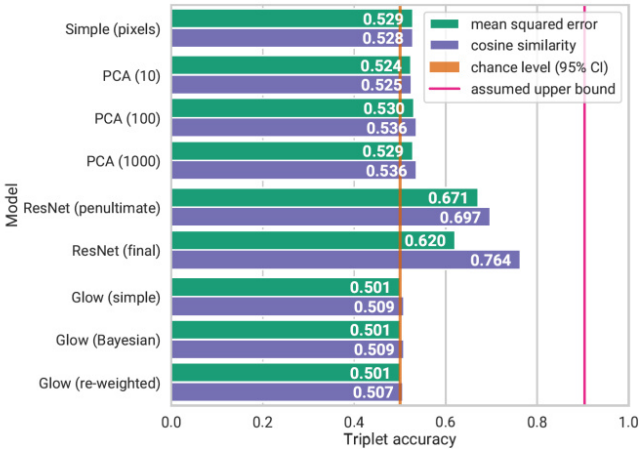

Figure 1: Comparison of predictive performance of different models. Distances between raw pixel values, PCA embedded vectors, ResNet50 last and second last layer activations, Glow latent representations, Glow latent posteriors estimates, and distances between Glow reweighted latent vectors were used for predicting similarities.

Introduction

Recent work has shown that deep neural networks are able to predict human similarity judgments with high accuracy (e.g., how similar two images appear to a human). However, these results have been accomplished primarily with supervised discriminative models, although generative models are currently growing strongly in popularity and have achieved impressive results in many other areas. Normalizing Flows are a type of generative model that learn a deterministic mapping between data and latent representations. We investigate whether Glow – a specific Normalizing Flow architecture – is able to form a latent space that is aligned with human similarity perception.

Methods

We first trained a Glow model on the Caltech-UCSD Birds-200-2011 dataset. For comparison, we also calculated a Principal Component Analysis of the dataset using 10, 100, and 1000 components and trained a ResNet50 to classify the bird images. All of these models were then used to predict human similarity judgments on bird images. To do so, we calculated the pairwise mean squared error and cosine distance between the vector embeddings of the images obtained from the different models. The Normalizing Flow had 252 million trainable parameters and the ResNet50 had 24 million trainable parameters, that is why we had to rely on the high performance computing infrastructure for training and evaluation.

Project Manager
Dominik Straub

Researchers
Lukas Maninger

Principal Investigator
Prof. Constantin Rothkopf

Project Term
2022 - 2023

Clusters
Lichtenberg II Cluster Darmstadt

Software
TensorFlow

Results

We are able to replicate the good prediction performance of deep classifiers using the ResNet50, but the accuracies achieved by our Normalizing Flow are barely above chance level and worse than even pixel-based distance metrics. Additionally, we observed that our discriminative model performs better on similarity judgments that stem from 8- rank-2 trials (given a query image, subjects have to select the most and the second most similar image out of eight possible references) than on data from 2-choose-1 trials (given a query image, subjects have to select the more similar image out of two possible references).

Discussion

We argue that current Normalizing Flows are not able to capture similarity relations in their latent space. Unfortunately, we cannot give a definite answer to why they perform so poorly. We hypothesize, that possible reasons are the lack of supervision (i.e., class information), the general problem of Normalizing Flows with learning complex datasets, and insufficient compression of information due to the complete invertibility of the model. On the other hand, our results suggest that the high accuracy of deep classifiers might be mainly explained by the fact that humans usually judge birds of the same species (i.e., belonging to the same class) to be similar.

Last Update: 2023-08-02 14:35