Modelling Human Perception of High Gloss Materials using Neural Networks

Konrad E. Prokott, Hideki Tamura, Roland W. Fleming

For several years machine learning models have been able to match or even exceed human performance on object recognition tasks. There have been many reports of similarities between deep neural networks (DNNs) and human observers and there are many examples where the differences between DNNs and human perception become strikingly apparent. We take a look at machine learning in the context of modelling human material perception, specifically human perception of high gloss materials. We ask how good a model of human observers a neural network can be, how we can best find a good model within the vast space of possible models, and what we could learn from such a model.

First, we created a dataset of 149 000 images of objects rendered in a high gloss or near-matte textured material. We tested human observers in a binary classification task (high gloss – low gloss) and trained several classifiers to do the same task. For randomly chosen images models and humans perform well and therefore correlate highly with one another, so it is hard to distinguish between the models. We therefore painstakingly assembled a 'diagnostic set' of images in which human responses are systematically decorrelated from ground truth. A good model of human perception should predict the specific response patterns of humans, including failures. We conducted a Bayesian hyperparameter search, training several thousand networks to look for those architectures that most resemble human responses on the diagnostic image set. Our results show that a wide range of architectures can show similar response patterns to humans.

Second, we sought to model observers' ability to recognize specular highlights on surfaces. We created a dataset of 154 000 computer-generated images that show a perturbed, glossy surface, covering the entire image. To make the task more challenging, we added texture patterns to the surfaces. To test human observers, we generated predictions based on the ground truth specular image as well as a very simple model using only an intensity threshold. We compared human responses to these two predictions across a range of spatial frequencies of surface geometries and different texture patterns. We then searched for neural networks that imitate the pattern of responses.