

# Distinguishing Cubes from Spheres

March 7, 2017

In our report, we show movies “movies” that are generated in a simulated setting. They represent the pattern of light that returns from a diffuse reflection off of a cube or sphere, in the traditional three-bounce setting. At a glance, it is obvious that these movies are rich with meaning. As a proof of concept, however, we thought it worthwhile to put the frames from these movies into a machine-learning architecture, to see if frames from different sorts of movies could be distinguished from each other.

As part of this proof of concept, we generated five movies of cubes in different positions and with different temporal resolutions, and five analogous movies of spheres (represented by a finely-discretized mesh of triangles). We then shuffled all the frames together, marking the frames according to whether they were from a movie of a sphere or a movie of a cube. The classification task was to predict which of these two sorts of movies the frame belonged to.

We ran the experiment on a set of 872 movie frames drawn from 10 movies—436 from movies of cubes, and 436 from movies of spheres. We split these frames into a training set (used to find the right model) of 436 frames, a validation set (used to tune the model’s hyperparameters, such as the learning rate) of 218 frames, and a test set (used to evaluate the model’s performance) of 218 frames.

We used a neural network containing three layers: an input layer of the 1024 pixels in the  $32 \times 32$  simulated detector array, a hidden layer containing 30 pixels, and an output layer of 2 neurons (the cube-response and the sphere-response of the neural network). We ran the neural network for 200 epochs, and it reached a test accuracy of 93.6% (204/218 frames correctly classified).

The following table shows the test accuracy of the neural network after each number of epochs:

Number of Epochs	Test Accuracy
1	78.0% (170/218)
50	86.2% (188/218)
100	93.1% (203/218)
150	93.1% (203/218)
200	93.6% (204/218)

This work is preliminary for several reasons. First, it understates the awesome power that neural networks can bring to bear on problems of this kind. This is fundamentally a machine vision problem, although not of the traditional kind: given an image, we must deduce what object generated it. Many specialized techniques have been recently developed for training neural networks for machine vision problems, including convolutional layers, random dropout, and more hidden layers in the neural network, none of which were in evidence in this experiment.

In fact, our neural network lacked hard-coded geometric information of any kind. We expect that using the recent advances in general machine vision techniques will lead to improved performance. We also expect that this problem is ripe for innovation that pertains specifically to the problem of seeing around corners. For examples, neurons that are particularly well-suited to representing the functions that show up again and again in these sorts of problems (like division, or cosines) could potentially lead to a lot of improvement.

However, we should also note that this work is preliminary because it understates the difficulty of the problem. This is a simulated environment with a high signal-to-noise ratio and the presumption that our model is completely correct; although we included realistic limits on temporal and spatial resolution, we are likely to find that in an experimental setting, the problem will become much harder. And, of course, in the real world we will be presented with scenes that are much more complicated than simple cubes or spheres.

Nonetheless, we think that this proof of concept is good evidence that the intensity response of cubes will be distinguishable from that of spheres. We are very optimistic about this direction of our research.