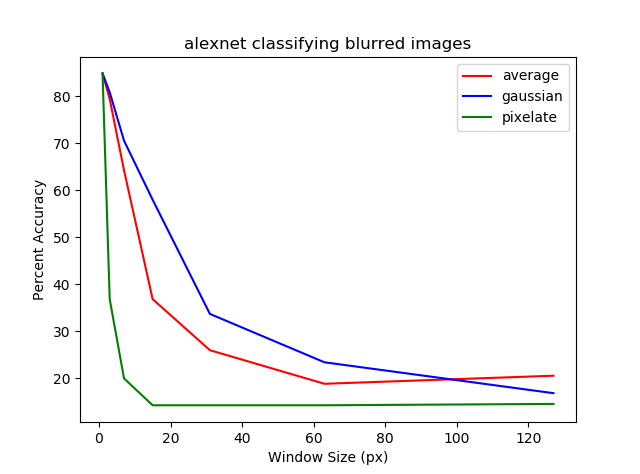
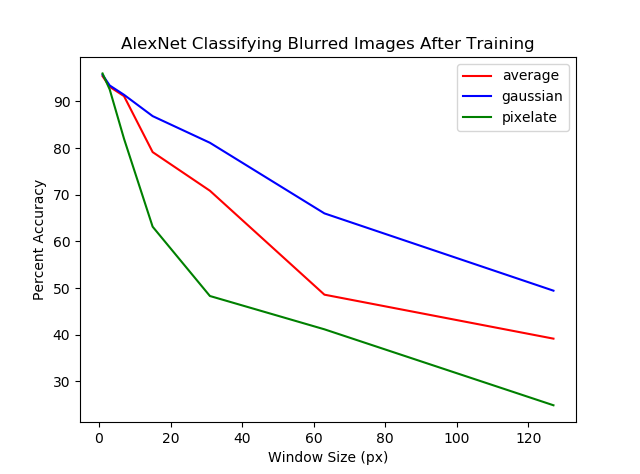
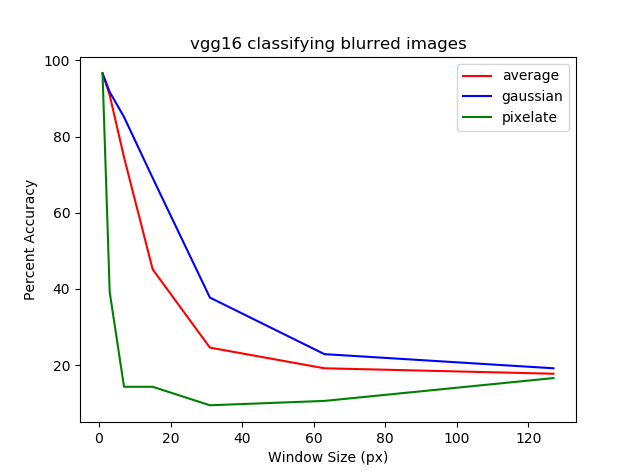
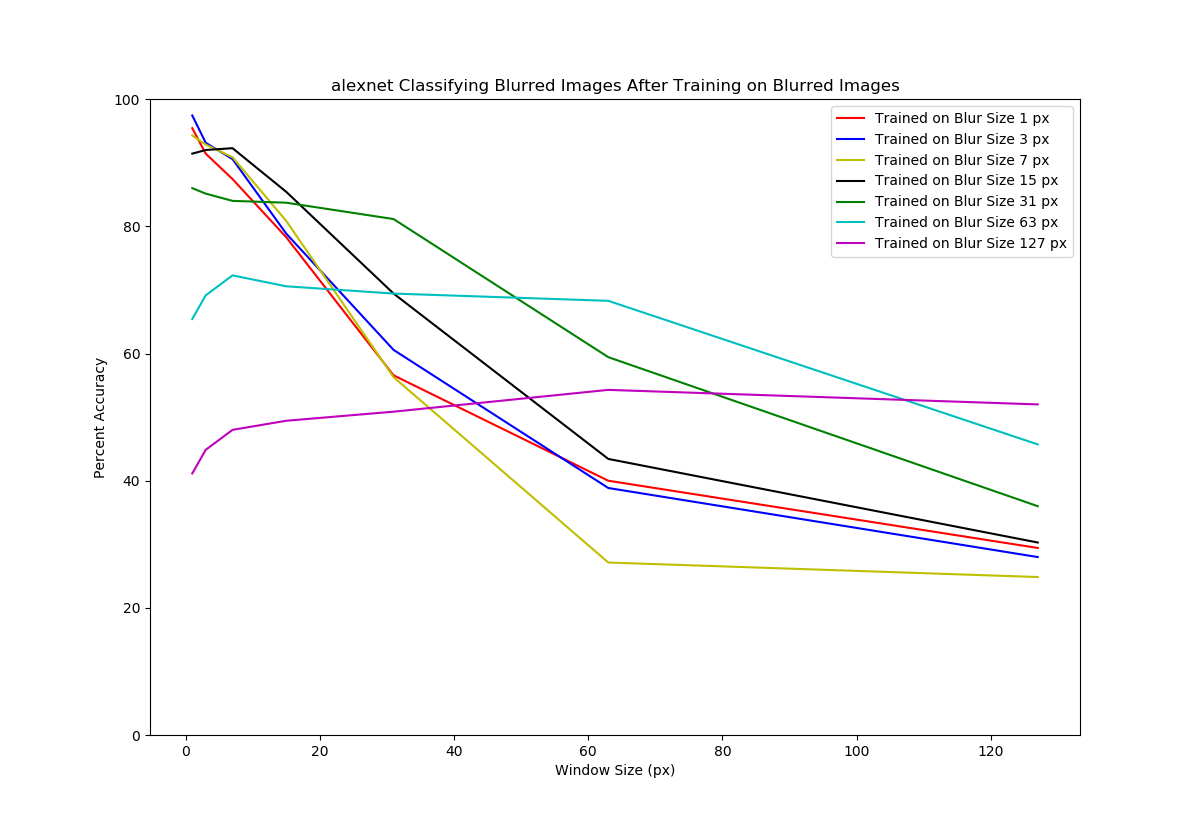
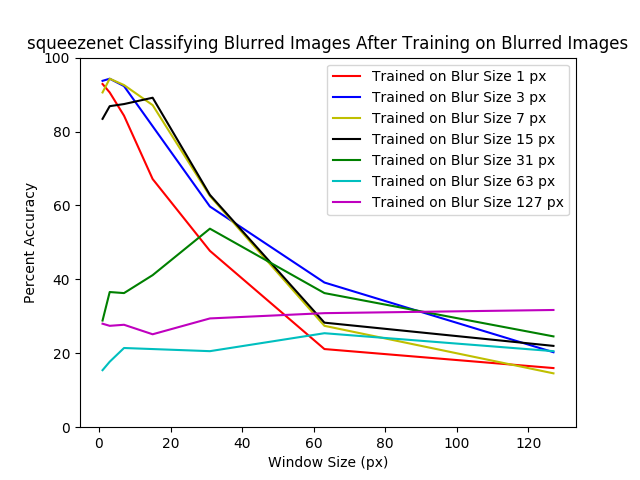
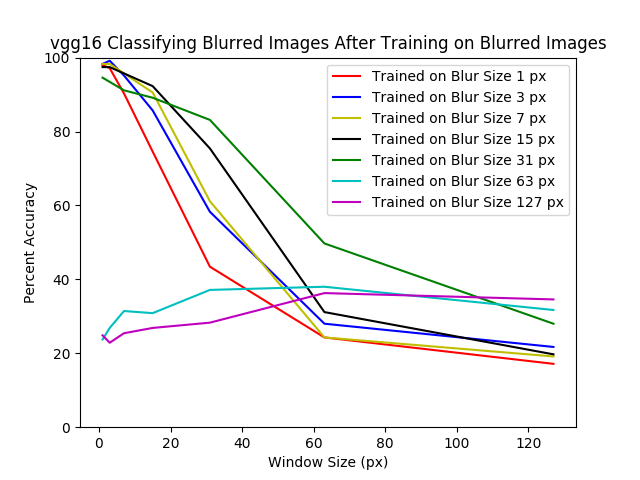
Note: Blur window of 1 = no blur

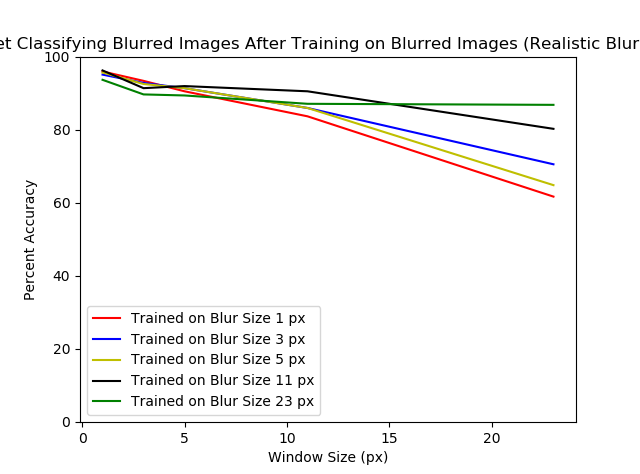
These were done on Imagenette (7 classes, trained on 1300 images each, tested on 50 images each; all photos had different aspect ratios with the smallest dimension being 320px, but when I processed the images, I made them square and 224x224). The dataset was blurred with 3 different types of blur with windows of 1, 3, 7, 15, 31, 63, and 127. [I may have lost some them, but the VGG and SqueezeNet graphs looked similar.]

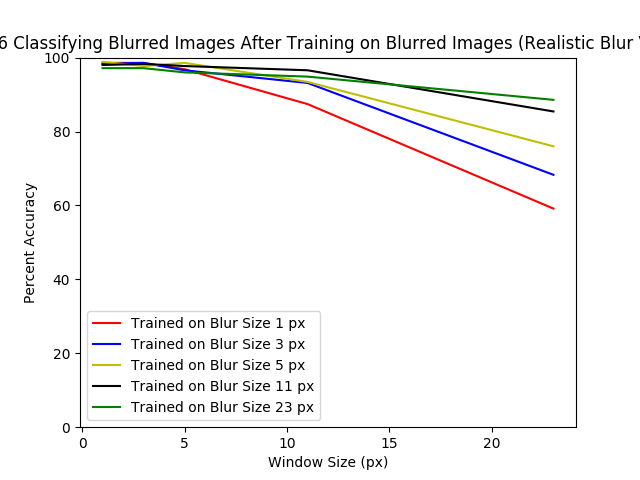






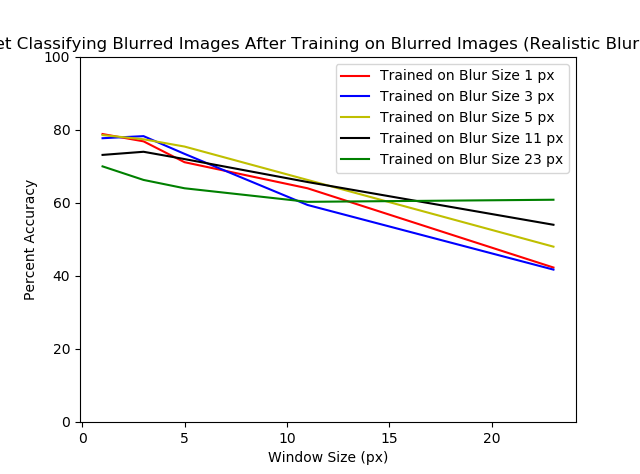
Each CNN was trained on images blurred with a window of 1, 3, 7, 15, 63, or 127, then tested on all the same window sizes. Still using Imagenette. Each level of training beats all others when tested on the same level, but unlike the PNAS paper, each level of training doesn’t always peak at the same level of testing.





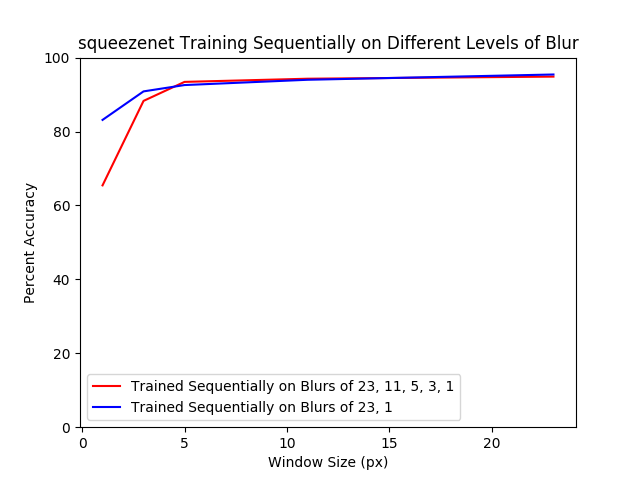
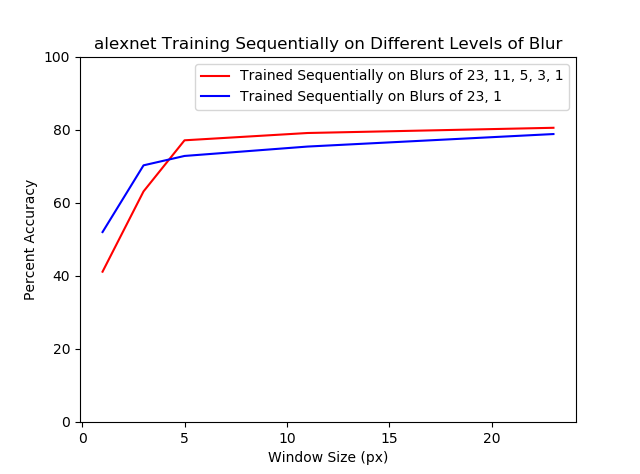
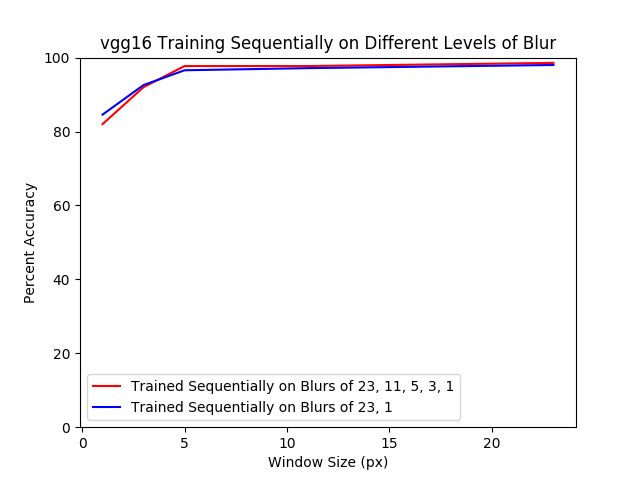
Top left = AlexNet, Top Right = SqueezeNet, Bottom = VGG16

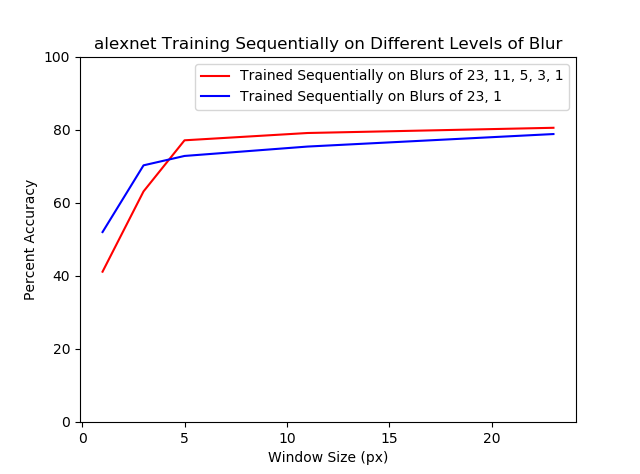
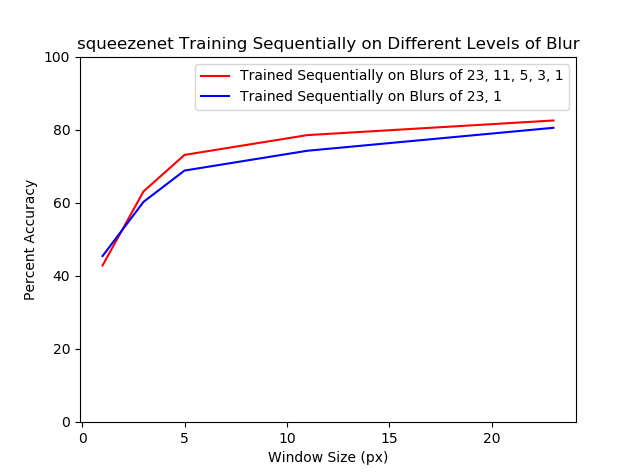
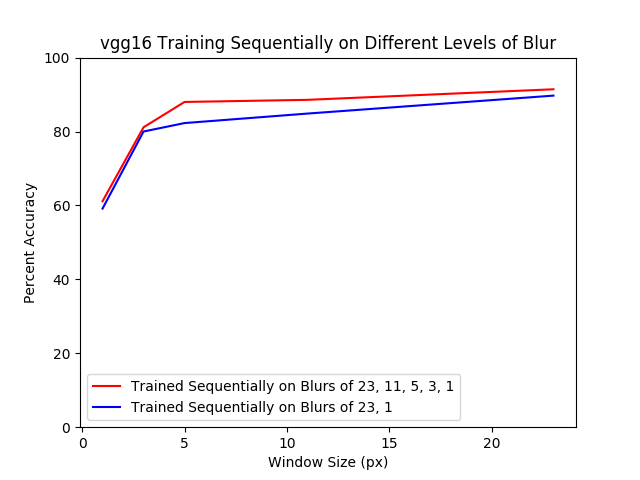
Same as previous experiment, but using blur windows of 1, 3, 5, 11, and 23, which we said we more biologically realistic and a good general progression from infant vision to normal vision. At least for AlexNet and VGG, I would say training on a higher amount of blur gives the best overall performance; it’s slightly worse at no and little blur amounts, but significantly better at higher levels of blur.

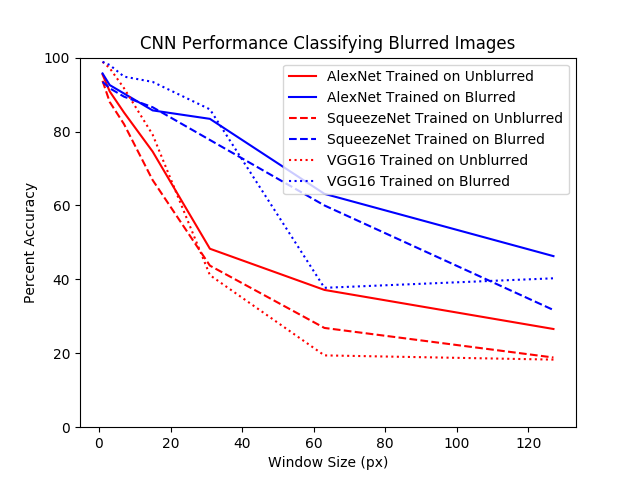


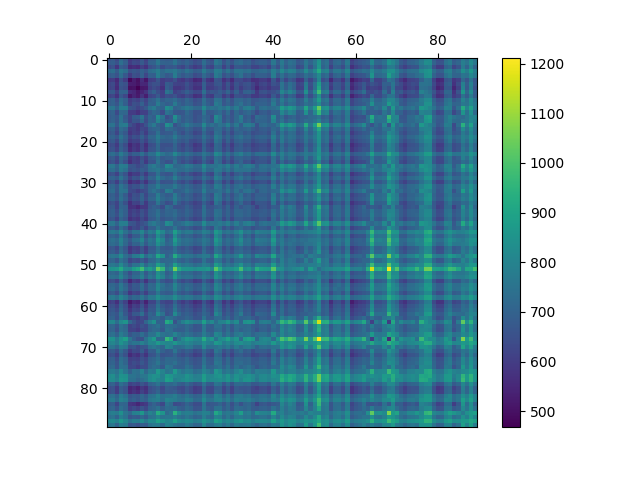
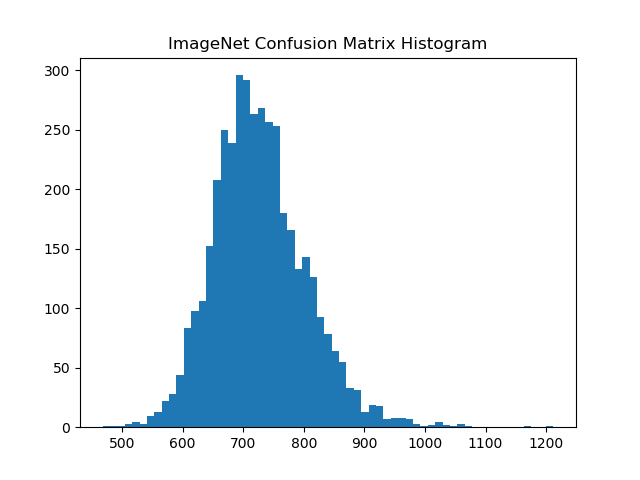
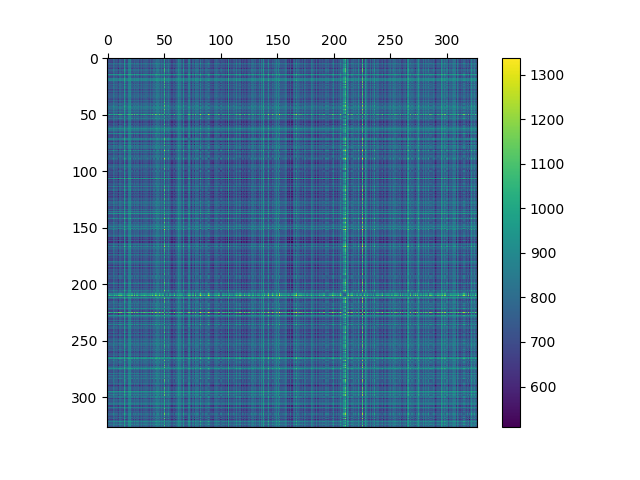
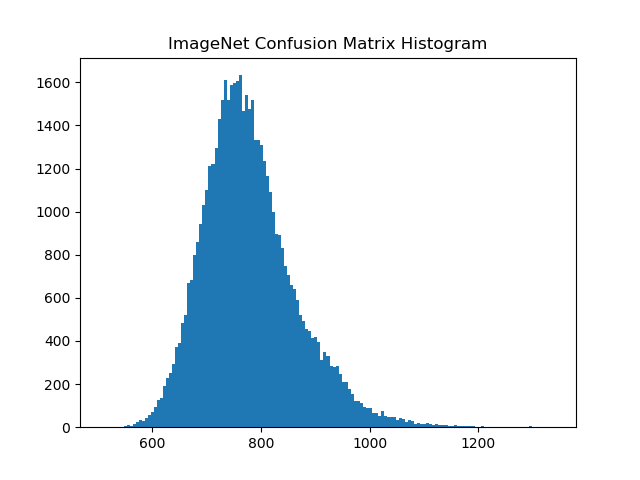
Top left = AlexNet, Top Right = SqueezeNet, Bottom = VGG16

Same as previous experiment, but using Imagewoof. All the numbers are the same as Imagenette, but the classes are all different dog breeds, which means the classes are much more similar. The accuracies have all lowered compared to Imagenette, which makes sense. The trend in the Imagenette version is still here in that trained on higher levels of blur gives better results when tested on high blur and more consistent performance throughout (flatter line), but the difference is that trained on high blur does significantly worse when tested at no and little blur levels.

Each CNN was trained on Imagenette blurred with a starting window size (23) and then on the same data with lower and lower window sizes (the same as the realistic blur levels). Then that model was tested on all the realistic blur levels. This is supposed to be similar to how an infant’s vision gets better over time as it learns. The PNAS paper’s graphs look different in terms of peaks, but they still found blurred-to-high-res training to be the most effective out of the different sequences they tried (blurred-to-blurred, high-res-to-high-res, high-res-to-blurred, blurred-to-high-res). Based off their graphs, the reason our graphs dip at lower levels of blur is because they were first trained on a window of 23. But the fact that they are pretty flat the rest of the way (especially with VGG) is interesting. What would happen if we had increased/decreased training time on each level based on real life development? What if we had given a few more “years” at normal vision (trained multiple times at level 1 at the end of the sequence)?

Same as above experiment, but with Imagewoof. Same trends, but the flatness is less flat (but the more gradual sequence retains it better, and it’s clearer overall that it is better for most of the way). In VGG, the gradual sequence stays above the other line the whole time, but not for AlexNet or SqueezeNet (???).

Each CNN was trained on unblurred Imagenette and classified all blur levels (1, 3, 7, 15, 31, 63, 127). Each CNN was also trained on each blur level and classified the same blur level it was trained on. This I guess just shows that it’s easier to classify blurred images when the model is trained on blurred images (this is just the same as the first page of figures, but all together).

Confusion matrix for ImageNet on erdos with 90 classes, and a histogram of the distances between each class. Each pair of classes was compared using 10 images from each (100 comparisons per pair). Below is the same, but with ImageNet on puppet with 327 classes. The histogram looks like a nice bell curve. Weirdly, I found that some cross-class comparisons were more similar than some within-class comparisons (either this says something interesting or the 10 images I chose for each class is too small a sample size).