Ryan Filgas
Computer Vision Week 7 Summary

The google researchers writing "Distilling the Knowledge in a Neural Network" develop on previous work of Caruana et al to distill the benefits of ensemble learning to singular models using an approach that is significantly improved . The researchers also elaborate on distilling the knowledge of large models into smaller ones that retain the knowledge of the larger networks. Their method in general relies on using the softmax of the larger model as a parameter in training the distilled model.

Researchers tried various methods on the mnist dataset such as removing all of the threes or just leaving 7s and 8s, and by adjusting starting weights were able to reduce error with an 800 layer network achieving close to the performance of a 1200 hidden layer network. This research was taken further to work on a voice recognition dataset using an android voice recognition model as a baseline. They trained an ensemble model and distilled it into a smaller model which out-performed the baseline model. Last researchers use a massive internal google dataset to distill a model that was trained for 6 months. They use a general models confusion matrix to delegate work to specialists.