

Abstract: The performance of recognition systems has grown by leaps and bounds these last 5 years. However, modern recognition systems still require thousands of examples per class to train. Furthermore, expanding the capabilities of the system by introducing new visual concepts again requires collecting thousands of examples for the new concept. In contrast, humans are known to quickly learn new visual concepts from as few as 1 example, and indeed require very little labeled data to build their powerful visual systems from scratch. The requirement for large training sets also makes it infeasible to use current machine vision systems for rare or hard-to-annotate visual concepts or new imaging modalities.

I will talk about some of our work on reducing this need for large labeled training sets. I will describe novel loss functions for training convolutional network-based feature representations as well as new ways of learning data augmentation strategies so that new concepts can be learned from a few examples. I will also discuss how this question of supervision plays out for other “richer” computer vision tasks beyond image classification, and how thinking carefully about how we teach machines to do tasks can lead to rich dividends.