Research on supervised learning focuses on training high-quality models by curating large datasets that capture a diverse general distribution. This approach has well known scalability challenges namely: 1) accurate general models are computationally expensive for inference 2) collecting and labeling large datasets requires extensive human effort and 3) datasets need to be repeatedly curated due to shifts in the target distribution. However, in many contexts, the goal is to build a model for a small fraction of the general distribution of images. The limited diversity in these scenarios opens up the possibility to train models specialized to these target scenarios. In this thesis, we embrace this reality and explore the following question: When and how can model specialization mitigate scalability challenges in training and deploying models?

Answering this question requires addressing the challenges that arise in building specialized models which are effective and practical. Firstly, specializing a model requires methods for identifying data specific to the target task and obtaining supervision on that data. Secondly, targeting a subset of data requires model training techniques which can cope with small and potentially imbalanced training data. Finally, at inference time one needs to know if a specialized model can be trusted to produce accurate predictions. In this thesis, we contribute methods to address these challenges in multiple contexts and demonstrate that specializing models by exploiting semantic and temporal structure in the task improves inference efficiency and enables effective machine supervision.

We leverage semantic specialization to transform image classification architectures into conditional architectures for efficient inference. We propose training methods to specialize model components for visually similar classes and dynamically choose components that need to be invoked during inference. We use temporally specialized models for efficient video segmentation. We propose an online distillation system for continuously training a fast compact model specialized to short temporal windows in a video stream. In the final part of the thesis, we propose to use semantically specialized models to reduce human effort in curating datasets for rare categories. We propose an iterative method which combines labels from specialized models and human feedback to train models for rare categories.

Thesis Committee:
- Kayvon Fatahalian
- Deva Ramanan
- Dave Andersen
- Bill Mark, Google
- Ross Girshick, Facebook

Thesis Summary:
https://drive.google.com/open?id=1Rqi7ocnGar8nFwi89_36wqeZLS5Ixj-D