

Learning to Learn and Structure Learning in Model Spaces for Small Sample Visual Recognition

Yu-Xiong Wang

April 3, 2017

The Robotics Institute
Carnegie Mellon University

Thesis Committee:

Martial Hebert, Chair

Deva Ramanan

Ruslan Salakhutdinov

Andrew Zisserman, University of Oxford

Yann LeCun, Facebook AI Research & New York University

*Thesis proposal submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in Robotics*

©Y.X. Wang, 2017

Abstract

Understanding how to recognize novel categories from few examples for both humans and machines remains a fundamental challenge. Humans are remarkably able to grasp a new category and make meaningful generalization to novel instances from just few examples. By contrast, state-of-the-art machine learning techniques and visual recognition systems typically require thousands of training examples and often break down if the training sample set is too small.

This thesis focuses on endowing visual recognition systems with such low-shot learning ability. Our key insight is that the visual world is well structured and highly predictable in not only feature spaces but also model spaces. In this spirit, we address key technical challenges and explore different and complementary perspectives. (1) Inspired by developmental learning, we progressively grow a convolutional neural network with increased model capacity when transferring on target tasks through fine-tuning. (2) We leverage prior work in learning to learn by encoding a generic, category agnostic transformation via a deep model regression network from models learned from few samples to models learned from large enough sample sets. (3) We capture a more generic, richer description of the visual world by encouraging the top layer units of convolutional neural networks to learn diverse sets of low-density separators during an additional phase of unsupervised meta-learning.

In proposed work, we aim to extend these perspectives to (1) progressively self-grow a network with compositional or interleaved modules when learning from continuously evolving data streams and tasks; (2) cast small-sample recognition itself as a learning problem by leveraging the entire learning process and model dynamics; and (3) hallucinate additional examples by leveraging the joint regularity from generative adversarial learning and large-scale unsupervised data. Finally, combining these approaches and perspectives, we further propose learning predictive model structures through exploration and exploitation.