Generalizing to new What and Where: essential for vision

What does it mean to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking.

David Marr

- Networks are brittle: don't work well beyond training distribution.
- In a multi-task setting: Can't see all combinations of all tasks. We need to generalize!
- Object Detection, VQA, Image Captioning all need to jointly reason about what and where simultaneously.
- Can networks generalize to unseen category-pose combinations?
- Ex: Can a model trained on Ford shown from front + Mitsubishi from side, predict category and pose for the third image below?

Category-Pose Datasets

- MNIST variations: (i) MNIST-Position (ii) MNIST-Scale
- Lab dataset: Natural image dataset of physical toy objects on a turntable viewed from different viewpoints + intraclass variation.
- (Introducing our new!) Bledsed-Cars dataset:
  - Varying architectural design
  - Adding quantitatively controlled bias
  - Varying architectural design

Varying architectural design

Adding quantitatively controlled bias

Research Questions

Q1. When do neural networks generalize to unseen what-where combinations?
- What is the impact Data Diversity?
- What is the impact of Architectural Choices?

Q2. What are the neural mechanisms driving this behavior?
- Do selective and invariant representations have a role to play?
- What behaviors emerges at the individual neuron level?
- How do findings change in a multi task setting?

Results I: When do networks generalize?

Finding I. Data diversity helps generalize even to unseen combinations.
Finding II. Separate "what" and "where" gives better generalization.

Results I: how do networks generalize?

Finding III. Do neurons get specialized? Yes!

Results II: How and why do networks generalize?

- Data diversity can help networks extrapolate.
- Separate architectures extrapolate better than shared ones.
- Specialization serves as a quantitative framework to understand generalization in multi-task settings.
- Separate architectures facilitate specialization, which leads to better performance.

Future Work

- Extend to other tasks like depth, and segmentation among others.
- Mixed selectivity of neurons.
- Can electricity and invariance be defined for a group of neurons?
- Extending beyond 2 tasks.