Neuro-Bayesian Methods for Realtime Scene Perception

by

Austin J. Garrett

Department of Electrical Engineering and Computer Science

Proposal for Thesis Research in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

December 2019

Author .....
Department of Electrical Engineering and Computer Science
December 11, 2019

Certified by .....
Vikash Mansinghka
Research Scientist
Thesis Supervisor

Accepted by .....
Placeholder
Chairman, Department Committee on Graduate Theses

Neuro-Bayesian Methods for Realtime Scene Perception

by

Austin J. Garrett

Submitted to the Department of Electrical Engineering and Computer Science
on December 11, 2019, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

TODO: replace me

Thesis Supervisor: Vikash Mansinghka
Title: Research Scientist

Chapter 1

Thesis Proposal

1.1 Introduction

Hook: The past decade has seen a flourishing of highly powerful computer vision techniques.

Gap: Yet the shortcomings of neural techniques have only grown increasingly urgent in proportion to their dominance. [IDENTIFY COMMON FAILURE MODES OF NEURAL TECHNIQUES].

Motivation: Meanwhile Monte Carlo and sampling-based techniques have seen rapid growth, with the creation of multiple probabilistic programming systems that can leverage the power of generative modeling to produce robust and intuitive models that are capable of expressing the problems above. [POTENTIALLY LIST HOW BAYESIAN REASONING CAN HANDLE THE FAILURE MODES ABOVE]

In practical applications, Bayesian techniques have often been neglected due to the computational demands of performing inference on high-dimensional latent parameters. The lack of tools

We propose the development of probabilistic facilities for Bayesian real-time scene perception. In particular, we propose the development of three main components:

- Toolkit corresponding to ROS pipeline, visualization, and models all implemented in the Gen ecosystem. [This point requires a bit of clarification of what

Task	Expected Completion
Development of realtime ROS-based particle filter framework with visualization	Jan 15th
Two-level neuro-predictive generative model and associated inference procedure	Jan 30th
Some project involving replacing the neural component with more distributed detectors that can be modeled probabilistically (hierarchical generalization of the research from January). This needs to be broken up into more digestible tasks, and we need consideration of exactly what form this takes.	April 30th

exactly we're providing, as opposed to just throwing our development environment at the reader and saying "we made some code".]

- A series of problems/a dataset exploring a domain of common sense reasoning tasks that traditional neural techniques fail to handle.

- A methodology for composing generative models with bottom-up neural detectors in the domain of visual scene perception for fast and robust scene understanding.[1]

1.2 Related Work

TODO: replace me

1.2.1 Placeholder

TODO: replace me

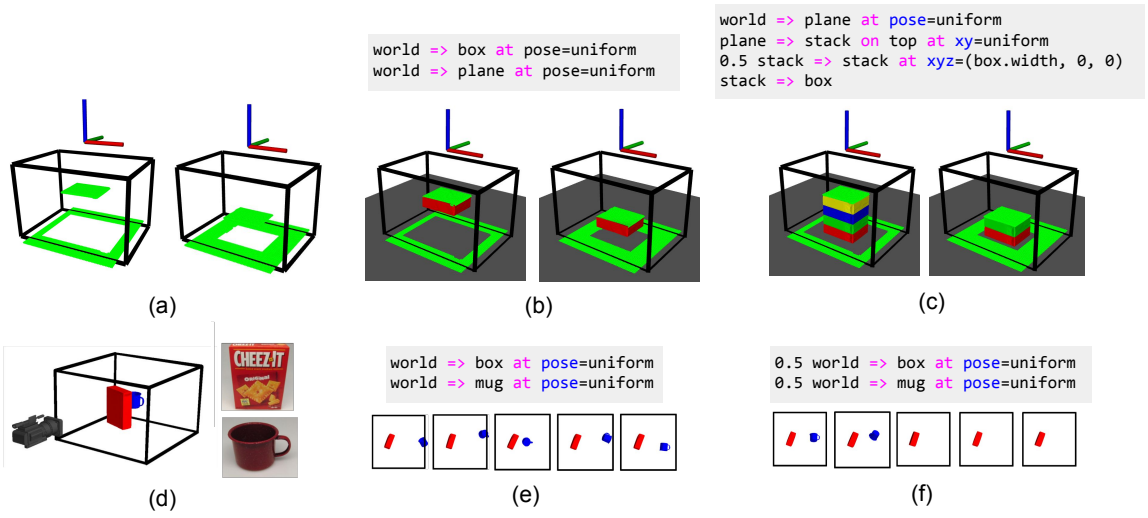

Figure 1-1: Two scenarios with prior knowledge specified as programs in our probabilistic scene description language. (a) shows two depth measurements made by a depth camera. (b) shows a program in which boxes have random poses in 3D space, and the resulting inferred pose of the box. (c) shows a program that assumes boxes are in stacks that rest on the floor, and the inferred number of boxes that explain the data for each observation. (d) shows another scenario, where a depth camera observes a box that is occluding a mug. (e) shows a program that asserts that both objects exist (but at unknown poses). The resulting inferences show the mug must exist somewhere behind the box to be consistent with this knowledge and the observation. (f) shows a program that allows for either object to not exist, and the resulting joint inferences about the mug's existence and pose.

1.3 Proposed Work

1.3.1 Toolkit

TODO: insert work from Nov-Jan in here

1.3.2 Dataset

TODO: Insert current work from neurips and lafi. The dataset serves as a simultaneous exploration of the research gap left by purely neural techniques, and a series of preliminary solutions (perhaps both purely symbolic and neuro-symbolic) that use Bayesian methods to address this gap.

This part is a highly relevent research goal, due to the undervaluing of semantic structuring of information by the neural network community. In some sense we are

trying to provide a solution to a problem that is not fully visible to the community, or at the very least is thought to be "too hard" for current techniques. Hard grounding metrics, or at least a fuller range of explicitly specified common-sense tasks will go a long way in changing the over-skeptical perspective toward Bayesian techniques trying to answer this question. Thus, an important part of the thesis work will be identifying the gap that exists between types of problems that we'd like to solve (common sense reasoning), and how neural networks currently fail to address these problems, as well as providing as much of a solution as we can in the time available, which we believe will be compelling enough to demonstrate the utility of Bayesian methods in modern CV pipelines.

1.3.3 Neuro-Symbolic Predictive Coding OR Generative Modeling for Neuro-Predictive Error Correction

In some ways the current state-of-the-art in both neural and probabilistic techniques have orthogonal and complementary strengths and weaknesses. While neural techniques exhibit strong performance in parsing out useful statistics from massively high-dimensional data with relatively little compute, they struggle to enforce semantically meaningful constraints on those statistics. On the other hand, probabilistic techniques can naturally capture symbolic systems and thus are much more expressible when it comes to representing and reasoning about highly structured semantic content. However, the dominant paradigms for inference treats it mostly as a top-down problem in which latent scene parameters are estimated via mostly-blind random walks. This "guess-and-check" methodology creates a bottleneck in propagating information from low-level features to the statistics of interest.

We believe there is a natural road to combining the strength of these two approaches. We propose a compositional technique where each system is used to leverage its own abilities. We propose

This corresponds to a meta-modeling predictive coding problem, in which we represent and abstract the behavior of the black-box neural model with a generative

model that leverages the intuition we have about the failure modes of these neural systems. Our generative procedure is not attempting to predict the images themselves, but rather construct a robust model of the detector itself in order to enrich the unstructured and noisy outputs of the detector with structured information.

There is also a possible path to a more fully distributed set of models, via a reformulation of the inference procedure as

PROPOSED FIGURE: A spurious DOPE detection is pictured. Below, the abstract scene graph is visualized, along with a percentage of occlusion. The more abstract model captures this via increasing the uncertainty of the pose estimation.

1.4 Conclusion

TODO: replace me

1.4.1 Placeholder

TODO: replace me

Bibliography

[1] Jonathan Tremblay, Thang To, Balakumar Sundaralingam, Yu Xiang, Dieter Fox, and Stan Birchfield. Deep object pose estimation for semantic robotic grasping of household objects. *CoRR*, abs/1809.10790, 2018.