Seeing Beyond Seeing with Enhanced Deep Tracking

Zhiyang He, BS
Stanford University
CS 229 Final Project

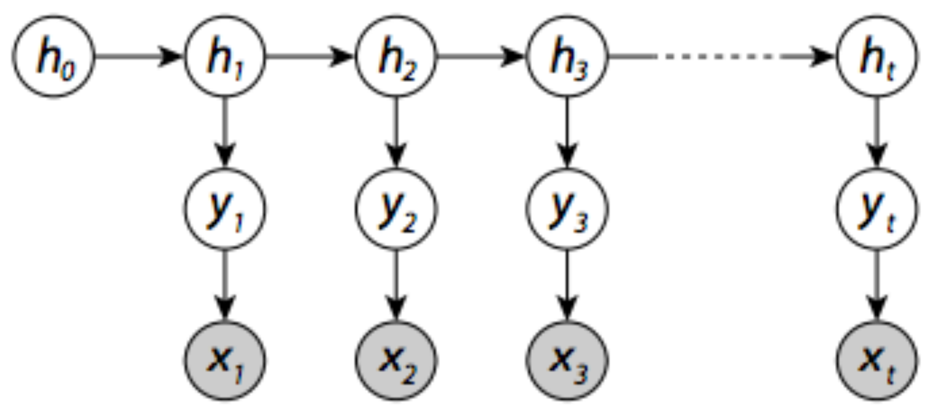
Abstract

In this work I present an enhanced end-to-end framework for tracking partially observable dynamic environment, using raw sensor input. This study is a further investigation of Deep Tracking, a recently proposed theoretical framework for unsupervised prediction of space occupancy.

Compared to traditional visual tracking methods which heavily relies on feature engineering, the predictive framework in this study demonstrates very good potential without requiring the same level of pre-existing knowledge.

Introduction

We want to be able to see objects that are occluded in raw sensor data. If we model the events as Markov process, this means we want to uncover the truth state y, behind our observation x.



We assume the generative model above. *h_i* captures complete of real world states. Given *x_1:t*, our goal is to find *P(y_t|x_1:t).*

$$P(y_{1:N}, x_{1:N}, h_{1:N}) = \prod_{t=1}^{N} P(x_t|y_t)P(y_t|h_t)P(h_t|h_{t-1}),$$

The unsupervised learning challenge is we do not have knowledge of y_t as direct feedback to our prediction. Thus we rephrase the question to predict *P(y_t+n|x_1:t).* Standard node dropout theory ([6]) shows this is viable. We use Baysian Filtering for reconstruction of belief state. The belief weight matrix and iteration matrix can be modeled by neural network as learning parameters.

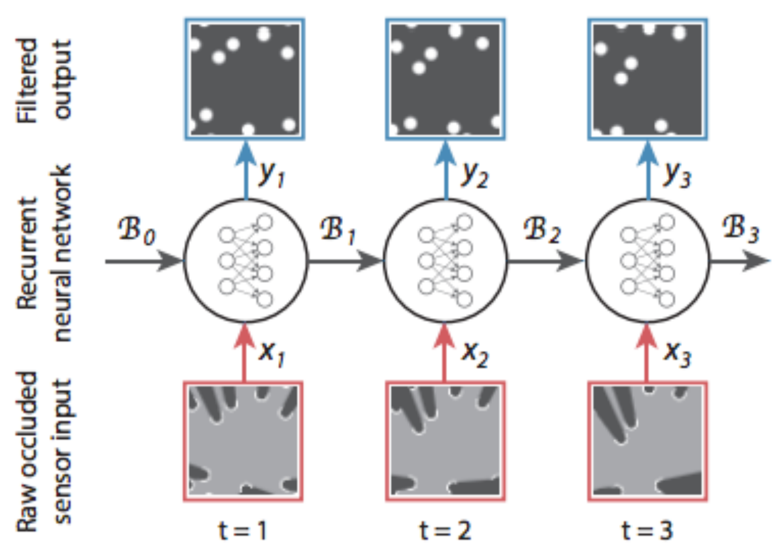


Figure 3 Baysian filtering process

Deep Learning Model

The original research proposes the four layer recurrent neural network. It uses convolutional operations followed by a sigmoid nonlinearity as processing step at each layer.

In my enhanced learning model, I added an additional network layer. This layer is set up for the purpose of tracking extra information. The hope is that it can capture different sizes, trajectories, as well as rotations. In order to maintain computation time, I changed sigmoid non-linear operators to ReLU.

Figure 1. Application in Autonomous driving

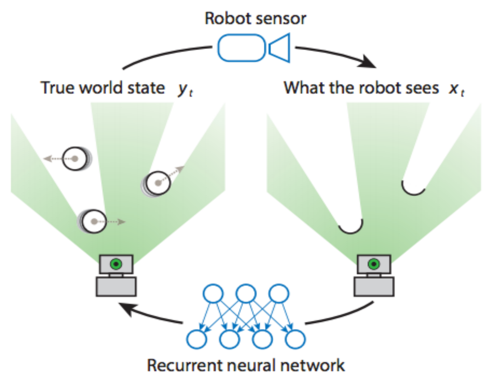


Figure 2. Robotic Representation Model

Dataset

The data sets used in this study are entirely synthesized. This is done for two reasons. First, the limited time and resource makes it difficult to collaborate with robotics lab and obtain real sensor data. Second, since the Deep Tracking framework is still under theoretical phase, its limited power means simple synthetic sets are most suitable for study purpose.

The data generation procedures for original study is not provided. Thus I implemented my own lua script to generate combinations of geometric shapes in space, and model these shapes to move in different patterns. The data sets are stored as native Torch 7 files, the size of each individual set (with 1million frames) is 7.31G before compression, 2.9G each after.

	Variations	Predictable?
Shape	Triangle, Square, Mixed	Yes
Speed	Non-uniform	Yes
Rotation	Non-stationary	Less satisfactory
Trajectory	Non-straight	Less satisfactory

Table 1. Synthetic Datasets and partial results

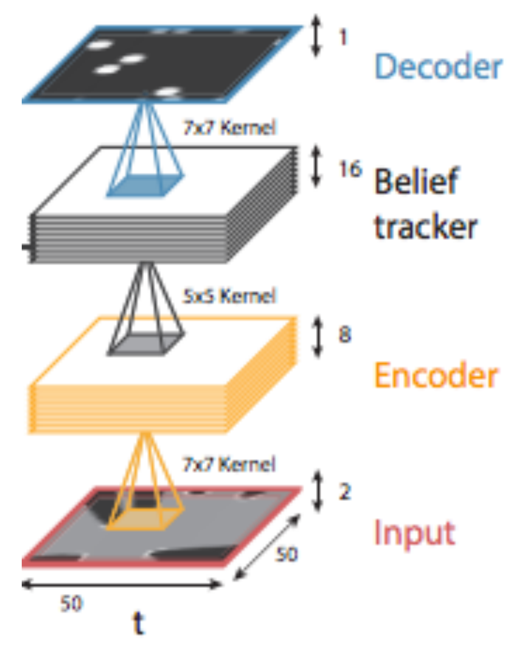


Figure 4 Original Learning Model Structure

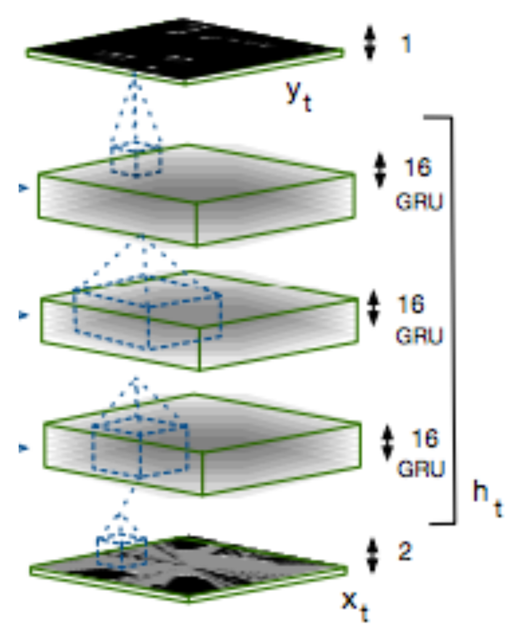


Figure 5 Enhanced Learning Model Structure

Results

By training round and triangular data on our improved neural network model, we achieved satisfactory results. We are not only able to reconstruct the back side of the light-bearing object, but also "see" the occluded objects hidden in the shadow.

However the prediction accuracy drops when more complexities are introduced: object rotation, curved trajectories, etc. Due to the current computation limitation, we have yet to know whether this could be improved with more model iteration, or if we need more complex neural network models.

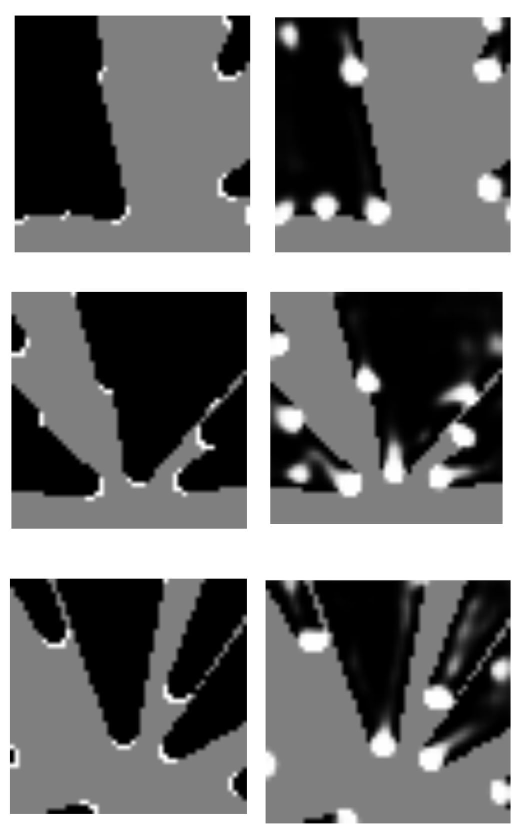


Figure 6 Round Data with 3000 Iterations (10 hours)

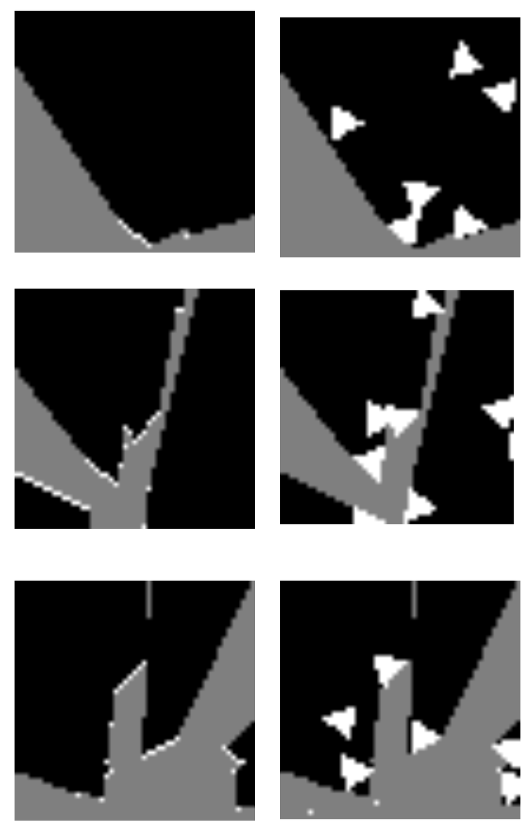


Figure 7 Triangular Data with 45000 iterations (1 week)

Further Work

My study shows the limitation of the current DeepTracking model. Next steps of work include more quantitative error analysis, or improvement of neural network architecture. The latter is more challenging because of time limit.

Contact

Zhiyang He
Stanford University
Email: hzyjerry@stanford.edu
Website: https://github.com/hzyjerry

References

1. P. Ondruska and I. Posner, "Deep tracking: Seeing beyond seeing using recurrent neural networks," in *The Thirtieth AAAI Conference on Artificial Intelligence (AAAI)*, Phoenix, Arizona USA, February 2016.
2. P.Ondruska, J. Dequaire, D. Z. Wang, and I. Posner, "End-to-end tracking and semantic segmentation using recurrent neural networks," *arXiv preprint arXiv:1604.05091, 2016.*
3. Graves A. "Generating sequences with recurrent neural networks," *arXiv preprint arXiv:1308.0850*
4. Dequaire J., Rao D., Ondurska P., Wang Z. D. and Posner I., "Deep tracking on the move: learning to track the world from a moving vehicle using recurrent neural networks," *arXiv*:1609.09365v1.
5. Srivastava N., Hinton G., Krizhevsky A., Sutskever I. and Salakhutdinov R. "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research* 15(1):1929-1958