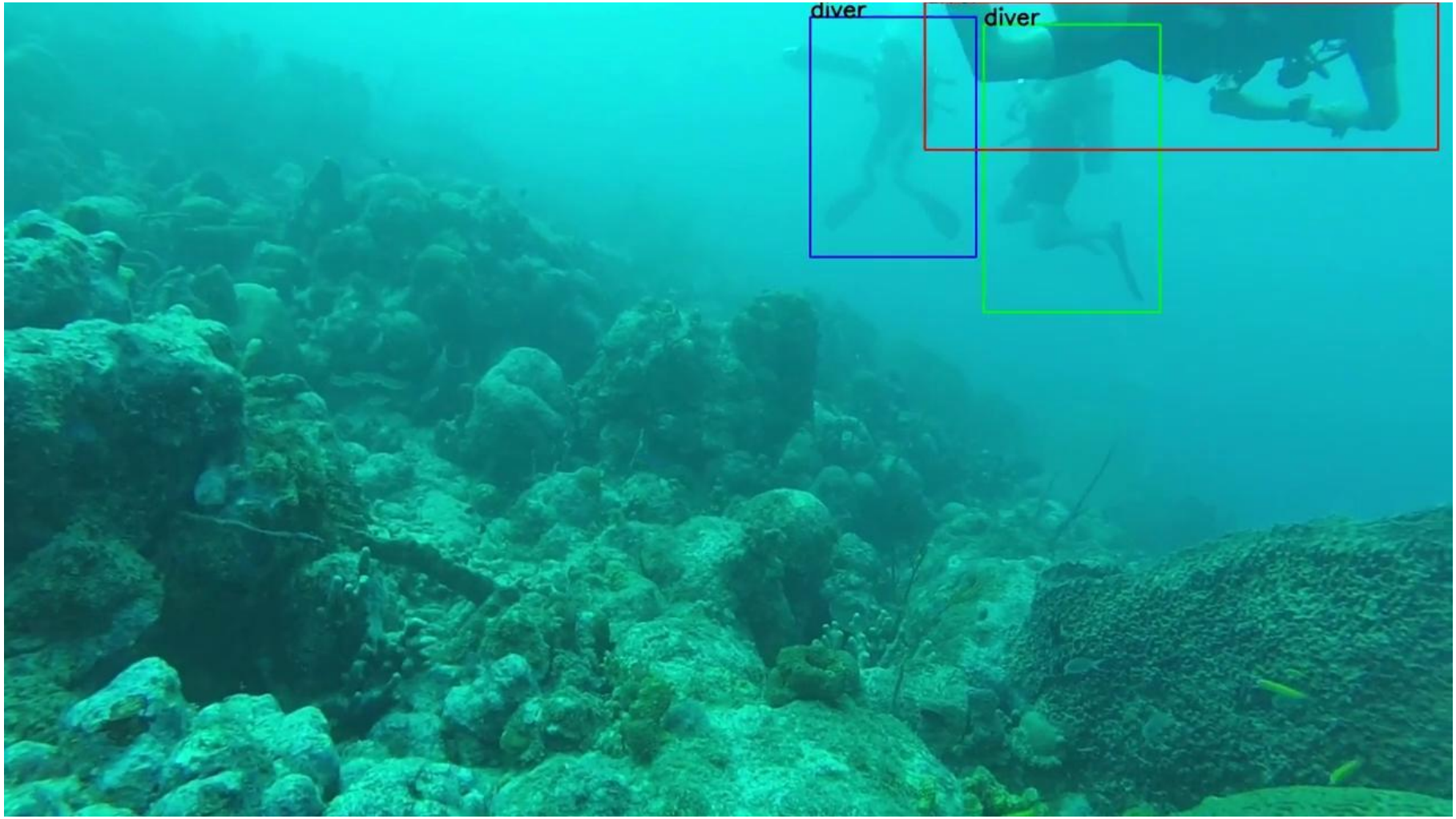
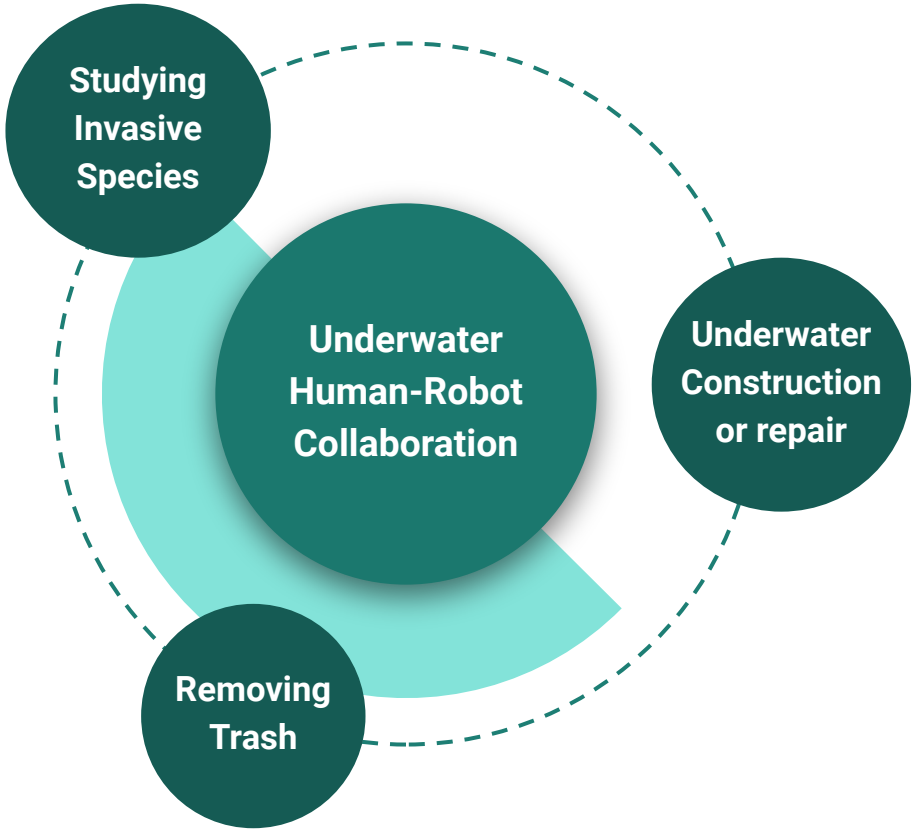
Predicting the Future Motion of Divers for Enhanced Underwater Human-Robot Collaboration

Tanmay Agarwal, Michael Fulton, Junaed Sattar

International Conference on Intelligent Robots and Systems (IROS) 2021

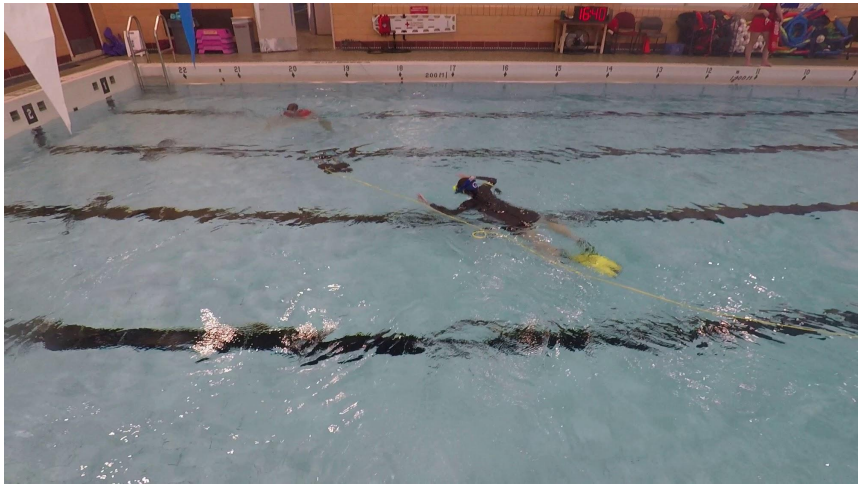

diver
diver

Why is this important?



Why do we need predictions for diver following or leading?

1. Challenges in current methods:



Source: Interactive Robotics and Vision Lab, University of Minnesota

Source: McGill Mobile Robotics Lab

2. Enabling robot leading behaviors

What methods might work best?

LSTMs

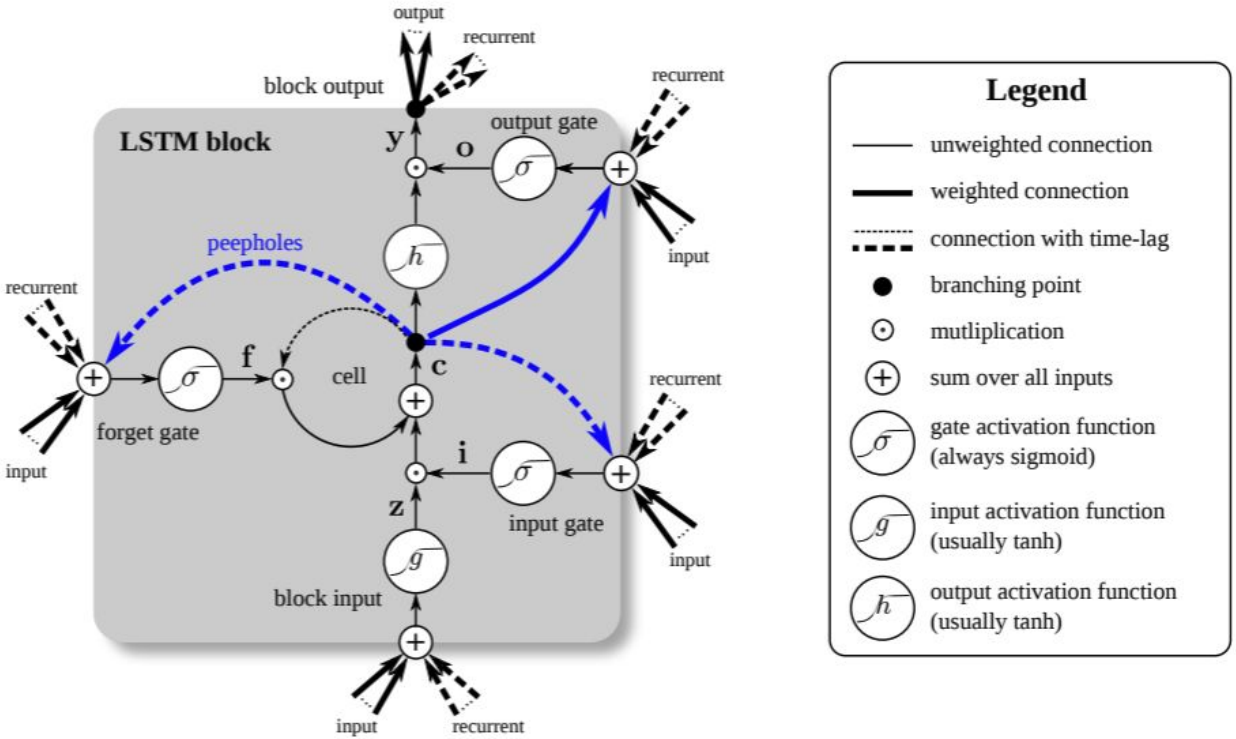
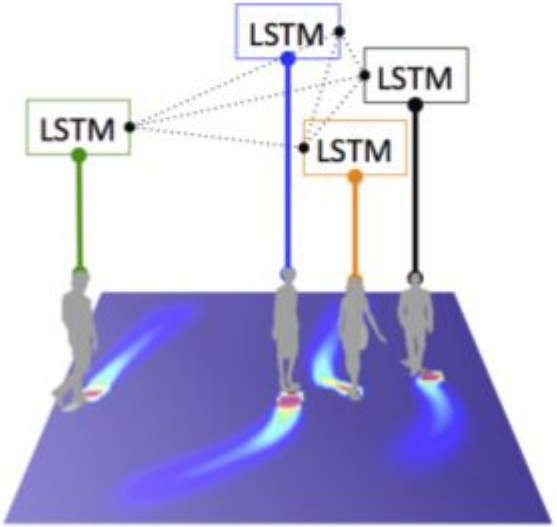


Image source:
K. Greff et. al. "LSTM: A Search Space Odyssey" in IEEE Transactions on Neural Networks and Learning Systems, 2016.

Social-LSTM



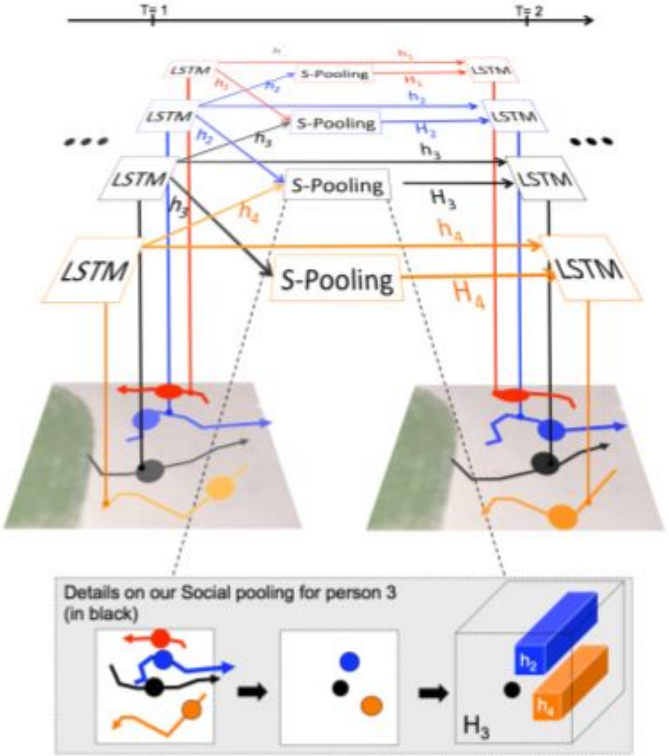


Image source (both):
A. Alahi et. al. “Social-LSTM: Human Trajectory Prediction in Crowded Spaces” in IEEE Computer Vision and Pattern Recognition (CVPR) 2016

Challenges

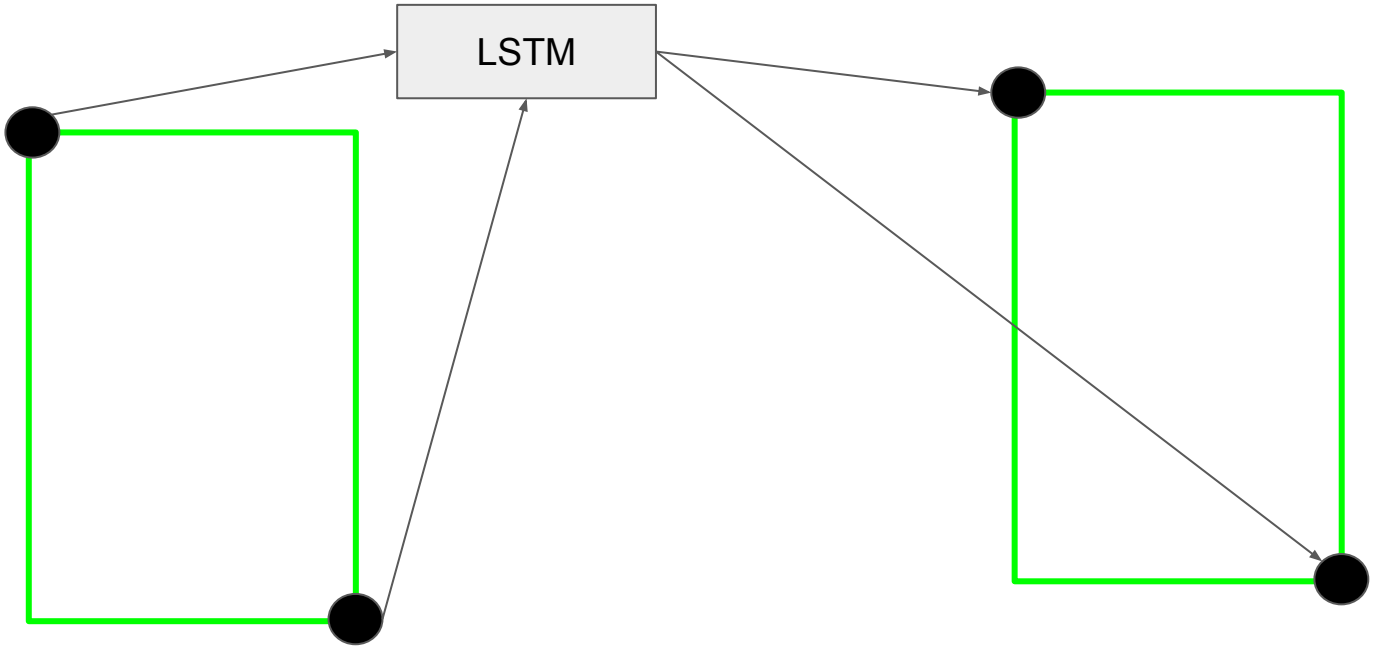
1. Pedestrian Algorithms - 2D Representation
2. Camera Ego Motion

ETH Dataset

Hotel Dataset

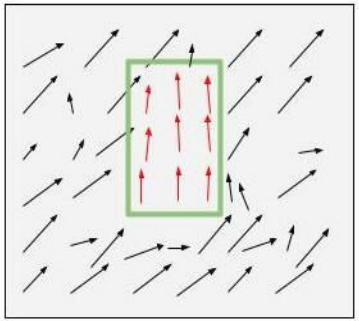
Solutions

1. Propagate 2 points

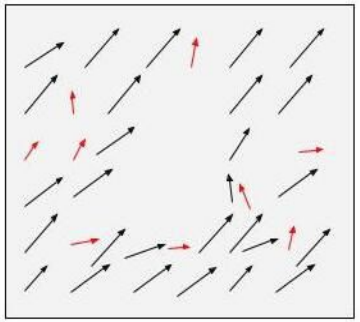


Solutions

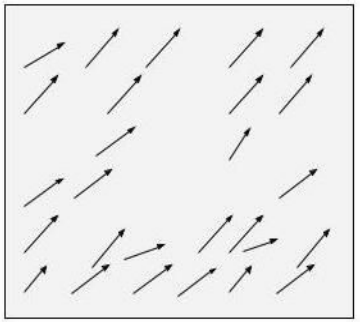
2. Stabilize Bounding Boxes



Initial Dense Optical Flow

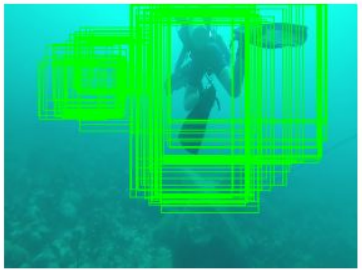


Bounding Box Removed

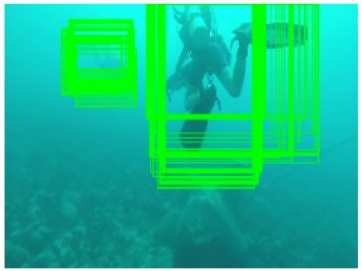


Thresholded

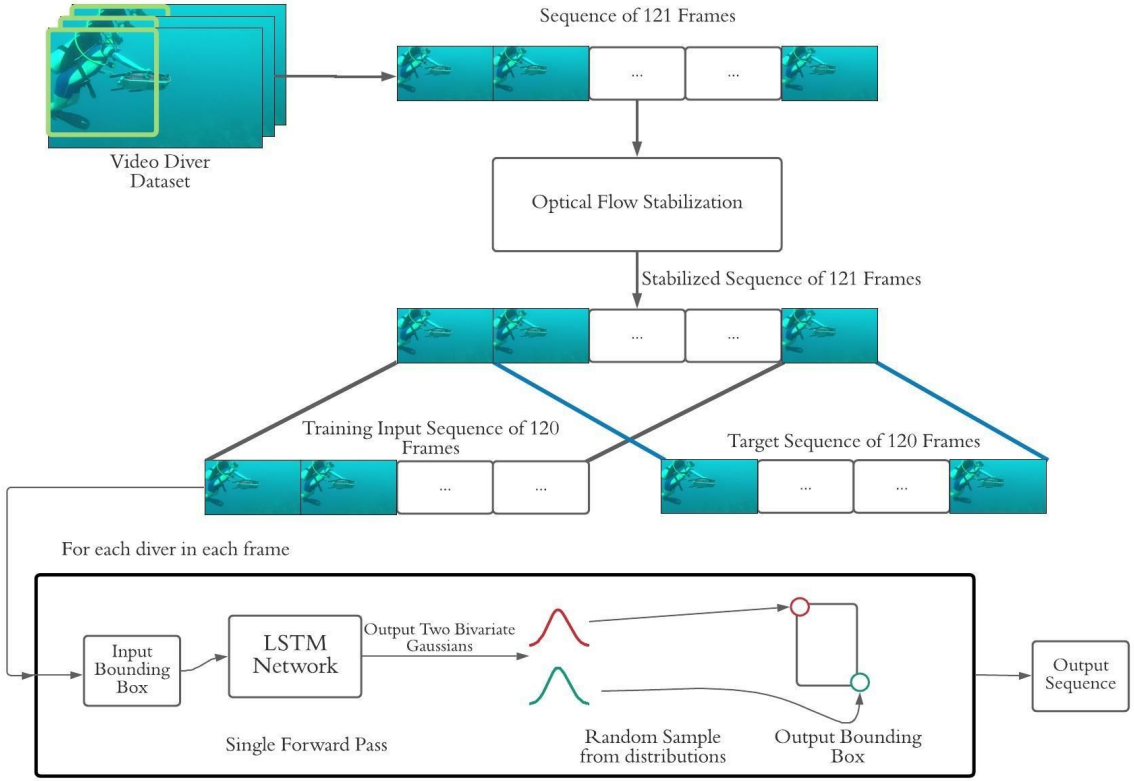
Original Boxes



Stabilized Boxes



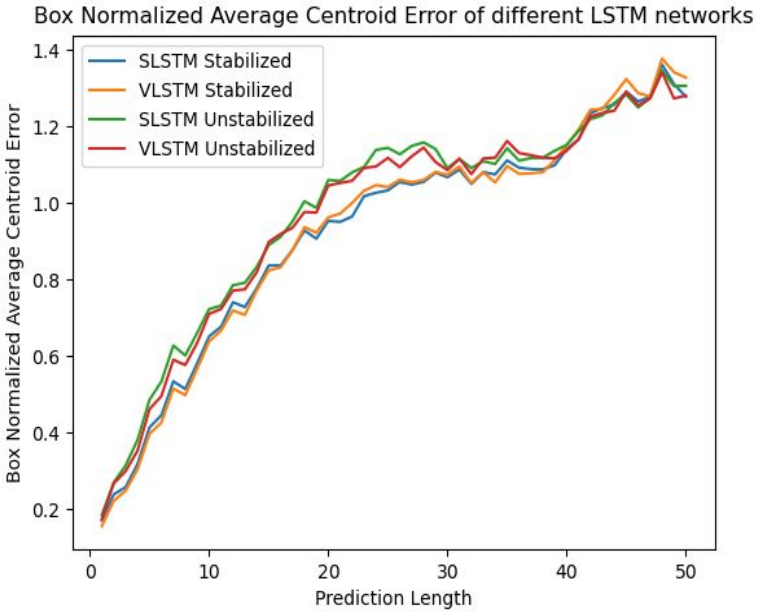
Training Methodology

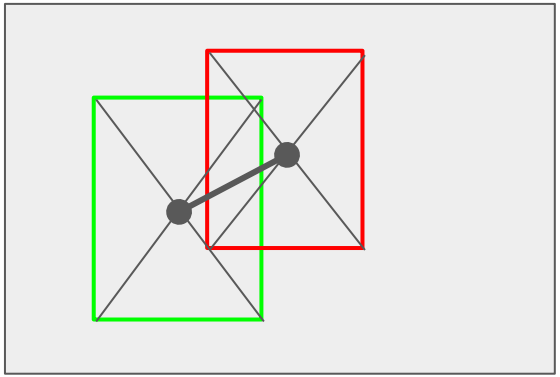


Video Diver Dataset (VDD):
https://conservancy.umn.edu/handle/11299/219283

Results

Box Normalized Average Centroid Error





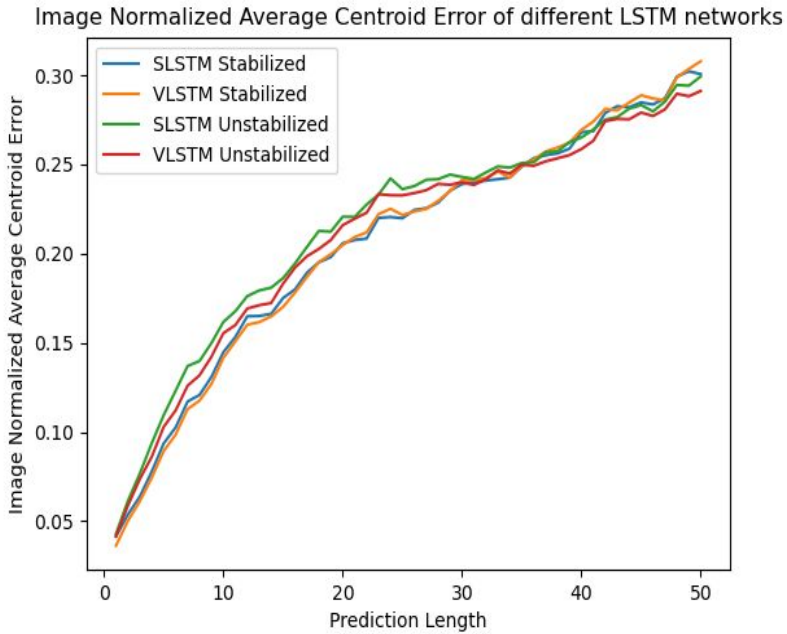
Ground Truth Prediction

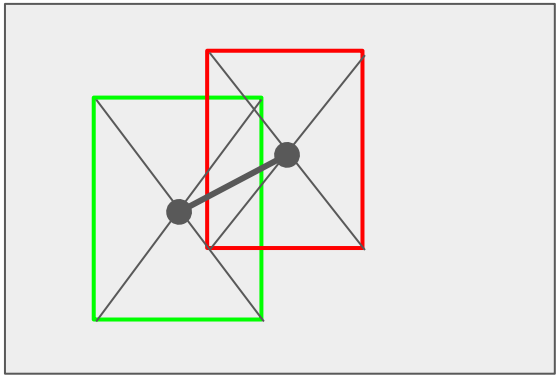
Normalization:

$$\frac{x_i}{box_{width}} \, , \, \frac{y_i}{box_{height}}$$

Results

Image Normalized Average Centroid Error





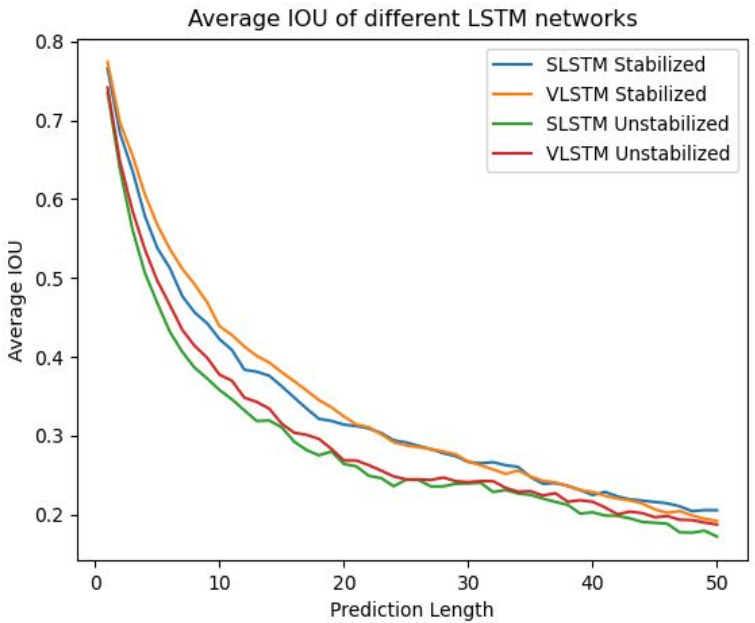
Ground Truth Prediction

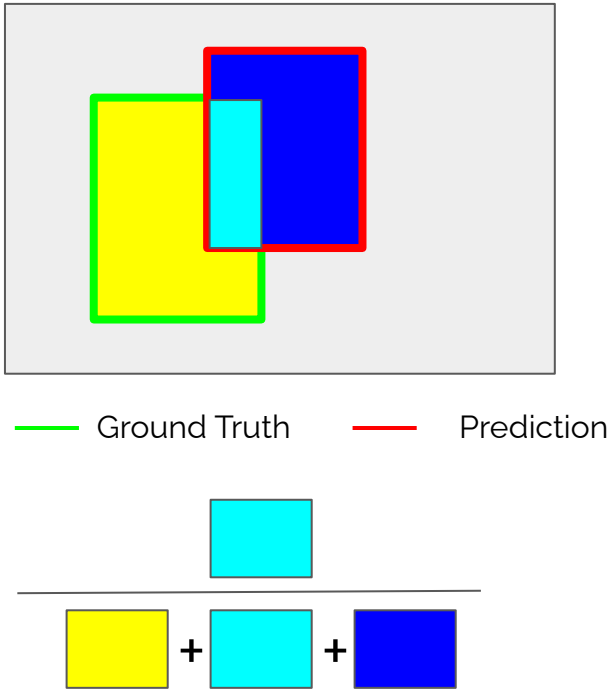
Normalization:

$$\frac{x_i}{img_{width}}, \frac{y_i}{img_{height}}$$

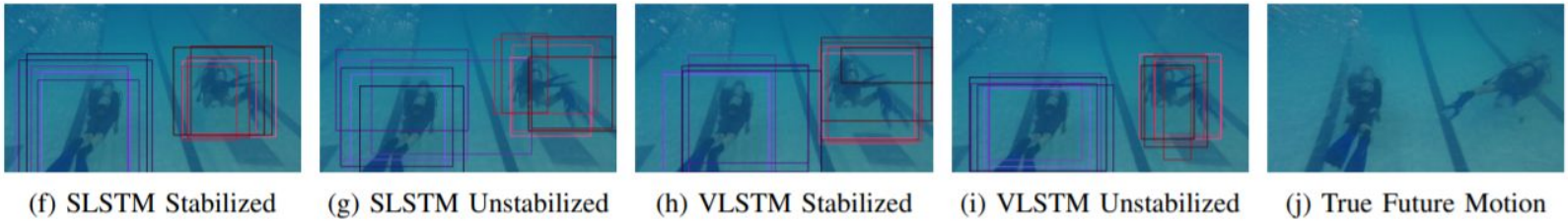
Results

Average Intersection Over Union





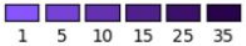
Results



(f) SLSTM Stabilized (g) SLSTM Unstabilized (h) VLSTM Stabilized (i) VLSTM Unstabilized (j) True Future Motion

(k) SLSTM Stabilized (l) SLSTM Unstabilized (m) VLSTM Stabilized (n) VLSTM Unstabilized (o) True Future Motion

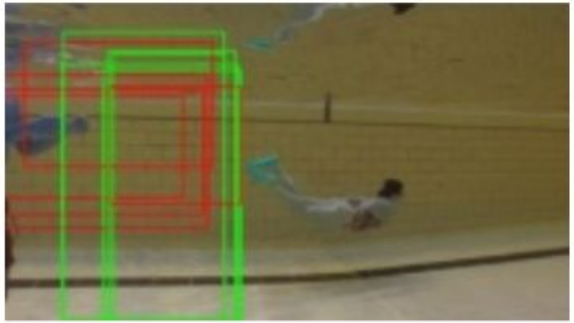
Predicted Frame Legend for Diver 1



Predicted Frame Legend for Diver 2

Failure Cases

(a) Last Observed Frame



(b) Predicted Boxes 50 Frames in the future

(c) Last Observed Frame

(d) Predicted Boxes 100 Frames in the future

Inference Time

Model Type	Vanilla LSTM	Social LSTM
Stabilized	558 ms	772 ms
Unstabilized	527 ms	737 ms

Table 4.1: Inference Time on a Jetson TX2

Inference Time: 0.5 s

Prediction Length: 1.5 s

Conclusion

- **First** method for prediction of diver motion
- Adapted Social and Vanilla LSTM
- Introduced an optical flow based stabilization method
- Reliable Predictions **1.5 seconds** into the future
- Inference rate of **2 Hz**

Future Work

- New methods for trajectory estimation
- Datasets recorded from a robot's perspective
- Encode diver features (such as pose, orientation) into the LSTM state