Following High-level Navigation Instructions on a Simulated Quadcopter with Imitation Learning

Valts Blukis, Nataly Brukhim, Andrew Bennett, Ross A. Knepper and Yoav Artzi
Instruction Following Task

go to the left side of plane

- Instruction Understanding
- Object Recognition
- Instruction Grounding
- Continuous Corrections
- Spatial Memory
- Geometric Reasoning

Solve using:
End-to-end differentiable neural network and tricks!
Robot Mapping Systems

Image source: octomap.github.io

Image source: pirobot.org/blog/0015/
Neural Instruction Following Architectures

- Map from first-person images to actions
- Need to learn how to reason about changing observations
Our Approach: Neural Networks with Mapping

- Add explicit **Camera Projection** and **Differentiable Mapping**
- Reason about the instruction on a static map
- Automatically handle changing first-person observations
Step 1: Feature Extraction

Using Residual Neural Network

Each pixel in the feature map encodes an image neighbourhood
Step 2: Deterministic Projection

Project features from camera image plane to environment ground
Transform from first-person to third-person

Feature Map
(Image Plane in Camera Frame)

Projected Features
(Map Frame)
Step 3: Map Accumulation

Add features into the Semantic Map over time

Projected features (time $T$)
Step 4: Instruction Grounding

Semantic Map

Recognized airplane

Inferred goal location

I\times I Filter

Grounding Map

9\times 9 Filter

Goal Map

\textit{go to the left side of plane}

LSTM
Step 5: Control

- Output the velocity command, given Grounding and Goal maps
- Sent to quadcopter’s flight-controller
Imitation Learning

Modified variant of DAgger
Trade convergence guarantees for speed and memory efficiency
Imitation Learning in Random Environments

3500 Instructions + Environments
Ground-truth trajectories
63 Landmarks
252 Possible Tasks

Go to right side of mushroom

Total number of rollouts:
3500 oracle
2000 policy
Outperform standard NN with no mapping
Very close to oracle performance
Go to the left side of plane

Action

Feature Extraction -> Mapping

Image Features

Semantic Map

9x9 Filter

Grounding Map

LSTM

Instruction Embedding

1x1 Filter

Goal Map

MLP

Go to the left side of plane