Reinforcement Learning with Unity 3D: Autonomous Garbage Collector

Sangram Gupta  Damian Bogunowicz  HyunJun Jung

Chair for Computer Aided Medical Procedures & Augmented Reality
07.02.2019
OKTOBERFEST!!!
1 billion euros

value of the Oktoberfest to the Munich economy
6 million visitors celebrate Oktoberfest in Munich every year
1000 tons

total amount of waste produced at Oktoberfest
- Massive events
- Large scale operation
- Functional 24/7
- Autonomous, intelligent
Damian
Passionate about making machines autonomous and intelligent.

HyunJun
Biomedical Computing student, loves Computer Vision and Deep learning.

Sangram
Exploring new technologies in Computer Vision and also into getting decent grades.
Environment

Collect!

Avoid!
Perception, Cognition, Action
Software

- unity

- TensorFlow

- GitHub

Algorithm

- Semantic Segmentation
  (SegNet, Badrinarayan et al., 2015

- Behavioral Cloning
  (Bain and Sommut, 1999

- Proximal Policy Optimization
  (Schulman et al., 2017

- Our own heuristic

https://github.com/GeorgeSeif/Semantic-Segmentation-Suite
**Actions**

- → ← ✗ Left/Right/Empty
- ↑ ↓ ✗ Forward/Backward/Empty
- 🟢 ✗ ✗ Grabber On/Grabber Off

**Rewards**

- Collect non-collectible item
- Slam against the wall
- Slam against the obstacle
- Punishment per step
- Punishment per grabber activation
- Reward for forward movement
- Collect garbage
Behavioral Cloning

- Short training time
- Only as clever as human player
- Good for naive agents
Behavioral Cloning

Student Brain

Teacher Brain
PPO: Single-Agent

- Long training time
- Increase punishments slowly
- About 40h of training
- Great learning experience!
PPO: Heuristic

- PPO for navigation and heuristic for collection
- Feasible for simple action
- Medium training time
Heuristics: API Perspective

External Communicator (Unity) → Observations → RL network → Heuristic → modify

Observations → Place holders → RL network → model.py

RL network → New Action → policy.py

Heuristic → Initial Action → model.py

ml-agents
Heuristics: Algorithm

One hot segmentation (Merged for visualization)

Channel 2 (Garbage)

Depth image (inversed)

max

if

Greater than Threshold?

Interrupt (Collect)
PPO: with SegNet
Two Approaches:

1. Train PPO with SegNet
   - Easiest way to implement
   - It takes about 5s to generate an observation

2. Train PPO network separately
   - Combine two only in test time
   - Tricky to implement
   - No effect on performance during training
PPO + SegNet: API perspective

External Communicator (Unity)

Observations

Actions

Global Variables

SegNet weights

PPO weights

SegNet (w. scope)

PPO Network (w.o. scope)

model.py

policy.py

trainer_controller.py

ml-agents
SegNet In Action:

- Computationally expensive
- Reflects real world implementation (RealSense camera)
- Easy modification of its objective
A simple modification:

I need to collect channel 2
I should not collect channel 5
I have to avoid channel 1,4
Plot twist: The Furniture Collector

I collect furniture now!
The Furniture Collector In Action:
Room for improvement

- Install the actual mechanism for garbage collection
- Deploy the algorithm on machine can handle real-time semantic segmentation
- Transfer the knowledge from simulation to a real robot with RealSense camera
- Make the world a better and cleaner place!
Outlook for the future: fleet of autonomous robots
Thank You For Attention!