Deep Reinforcement Learning for Robotic Grasping from Octrees

Learning Manipulation from Compact 3D Observations

Master's Thesis

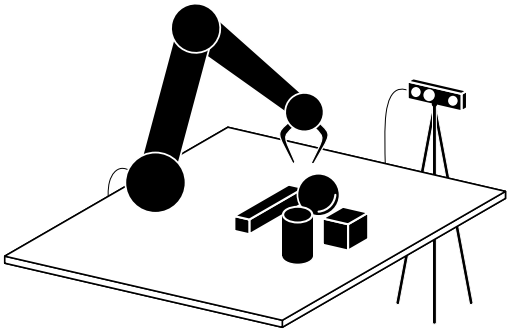
June 25, 2021

Andrej Orsula

Aalborg University
Denmark

AALBORG UNIVERSITY

Vision-Based Robotic Grasping of Diverse Objects



Vision-Based Robotic Grasping of Diverse Objects

Approach

Approaches

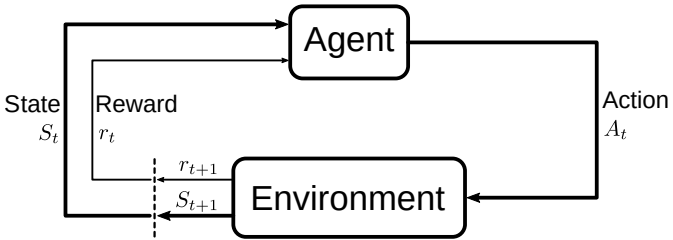
- Analytical
- Empirical
  - Supervised learning
  - Imitation learning
  - Reinforcement learning

Vision-Based Robotic Grasping of Diverse Objects

Approach

Approaches

- Analytical
- Empirical
  - Supervised learning
  - Imitation learning
  - **Reinforcement learning**



Task Definition

Agent

- High-level controller
  - Gripper pose
  - Gripper action

Environment

- Objects
- Robot
  - Low-level controllers
- Physics and visuals

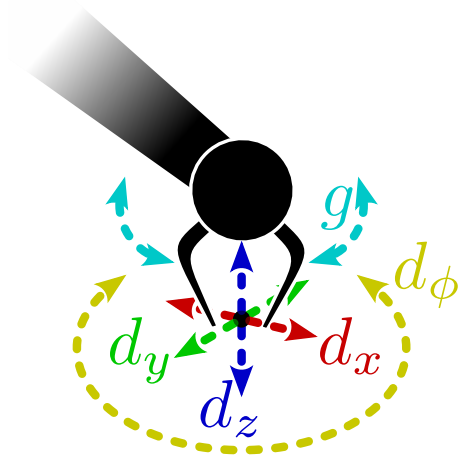
Episodic Task

- Success
  - Lifting an object
- Failure
  - Pushing all objects away
- Max 100 time steps
  - ~40 s (simulation)

Task Definition
Action Space

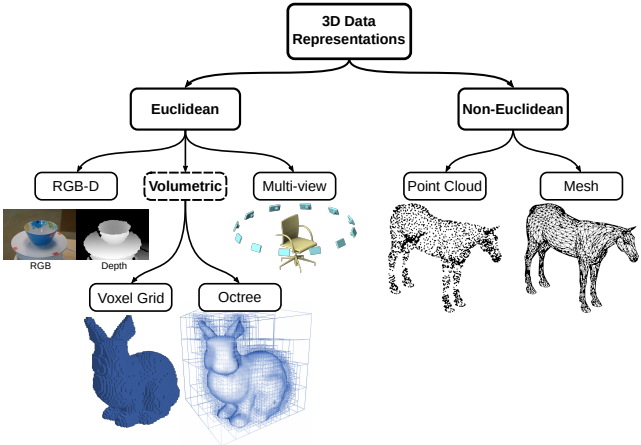
Actions in Cartesian Space

- Translational displacement
  - d_x
  - d_y
  - d_z
- Gripper rotation
  - d_ϕ
- Gripper actions (open/close)
  - g

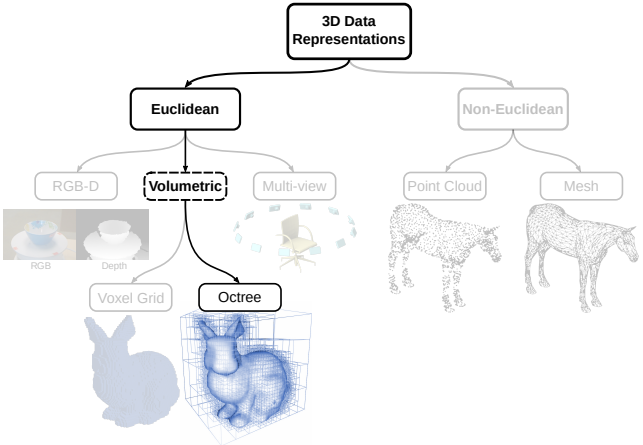


Task Definition

Observation Space

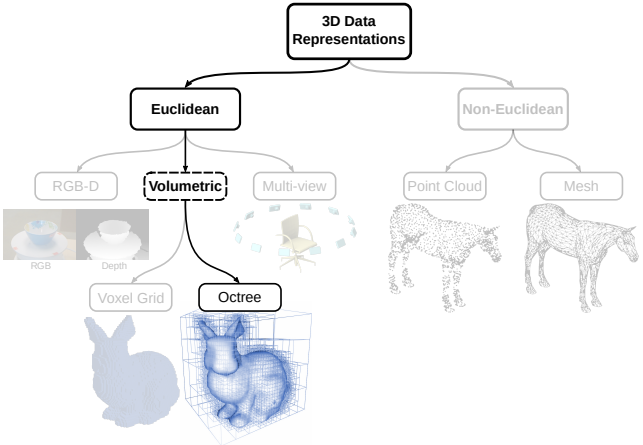


Task Definition
Observation Space



Wang et al. 2017. O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis. ACM Trans. Graph. (SIGGRAPH) 36, 4, Article 72.

Task Definition
Observation Space



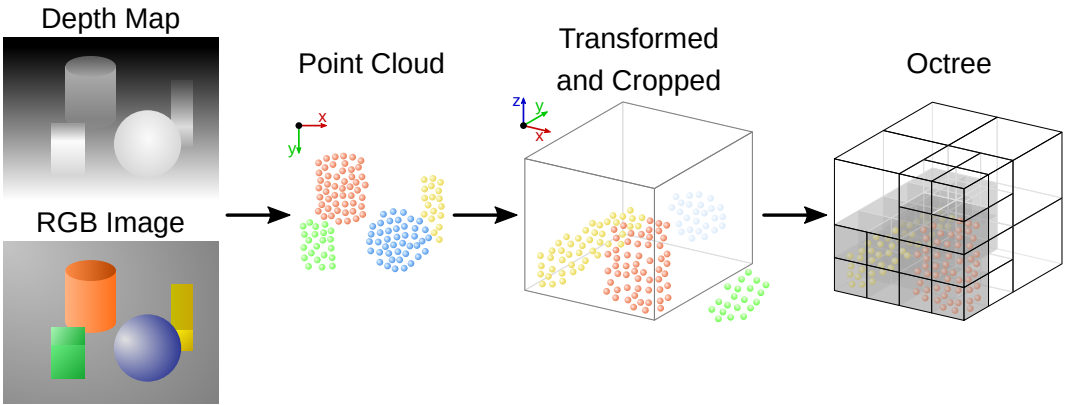
Proprioceptive Observations

- Gripper position
- Gripper rotation
- Gripper state

Wang et al. 2017. O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis. ACM Trans. Graph. (SIGGRAPH) 36, 4, Article 72.

Task Definition

Observation Space - Construction of Octree



Task Definition

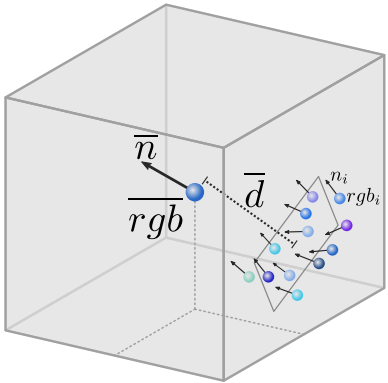
Observation Space - Features and Stacks

Features

- Spatial
  - Average normal vector $\overline{n}$
  - Average distance to points $\overline{d}$
- Colour
  - Average intensity of RGB channels $\overline{rgb}$

Observation Stacking

- Three consecutive observations



Task Definition
Reward Function

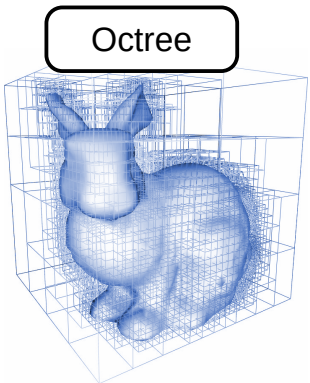
Composite Reward

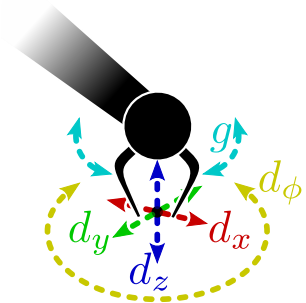
- Reach
  - $+1$ (7^0)
- Touch
  - $+7$ (7^1)
- Grasp
  - $+49$ (7^2)
- Lift
  - $+343$ (7^3)

Recurring Reward

- Collision with ground/table
  - -1
- Incentive to act quickly
  - -0.005

Task Definition
Summary





Reward Function

- Composite
  - Reach
  - Touch
  - Grasp
  - Lift
- Collision with ground/table
- Incentive to act quickly

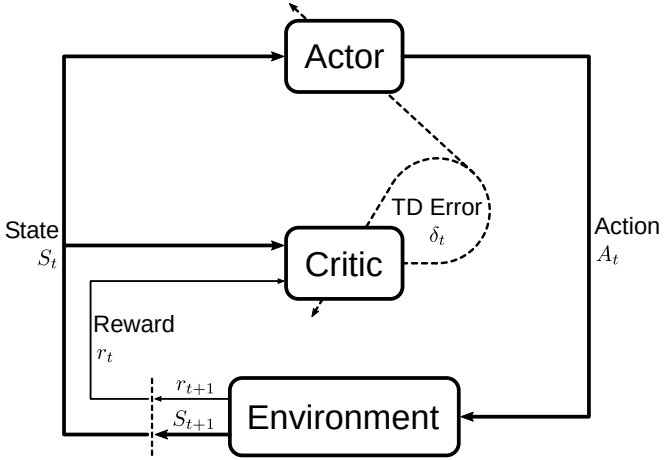
Deep Reinforcement Learning
Algorithms

Actor-Critic Algorithms

- TD3
- SAC
- TQC

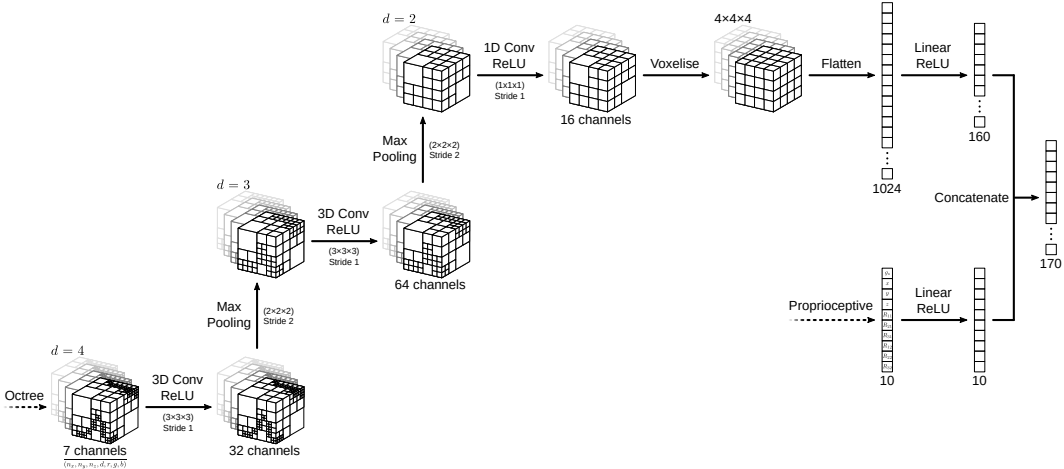
Implementation

- Stable Baselines3



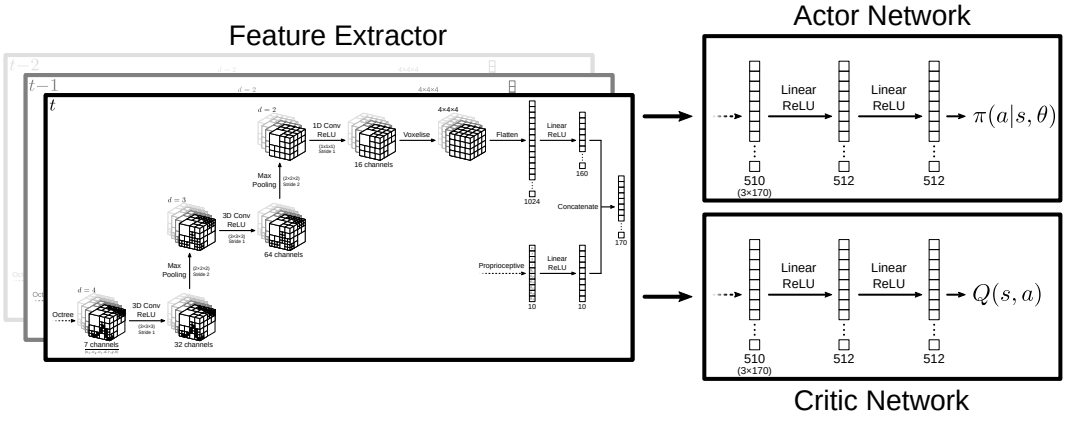
Deep Reinforcement Learning

Octree-Based Feature Extractor



Deep Reinforcement Learning

Full Actor-Critic Network Architecture



Simulation Environment
Selection

Simulators

- MuJoCo
- PyBullet
- Gazebo Classic
- Ignition Gazebo
- Isaac
- Webots
- Unreal Engine
- Unity
- Unigine
- RaiSim
- ...

Simulation Environment
Selection

Simulators

- MuJoCo
- PyBullet
- Gazebo Classic
- **Ignition Gazebo**
- Isaac
- Webots
- Unreal Engine
- Unity
- Unigine
- RaiSim
- ...

Simulation Environment
Ignition Gazebo

Physics

Rendering

Simulation Environment
Ignition Gazebo

Physics

Rendering

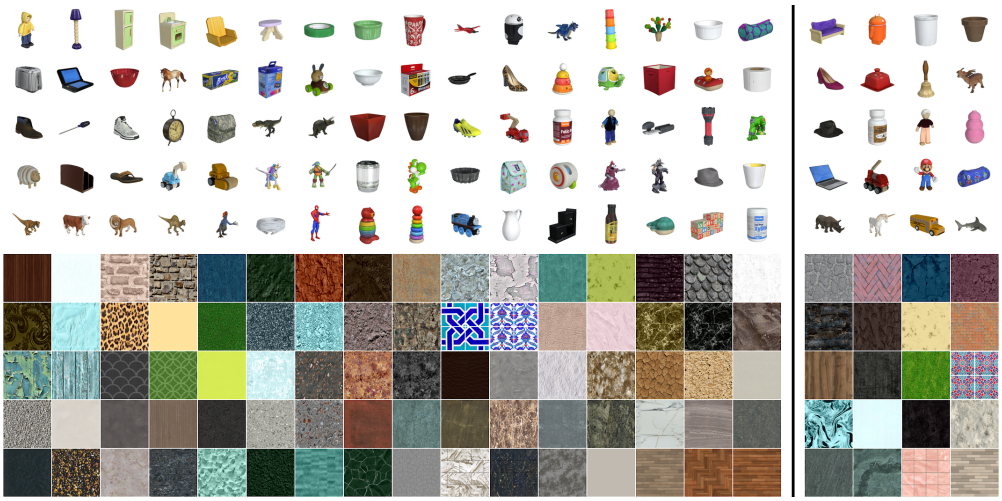
Gym-Ignition

- Interface for Ignition Gazebo
- Tooling for creation of OpenAI Gym environments
  - Compatibility with RL frameworks

Ferigo et al. 2020. Gym-Ignition: Reproducible Robotic Simulations for Reinforcement Learning. In 2020 IEEE/SICE International Symposium on System Integration (SII). 885–890.

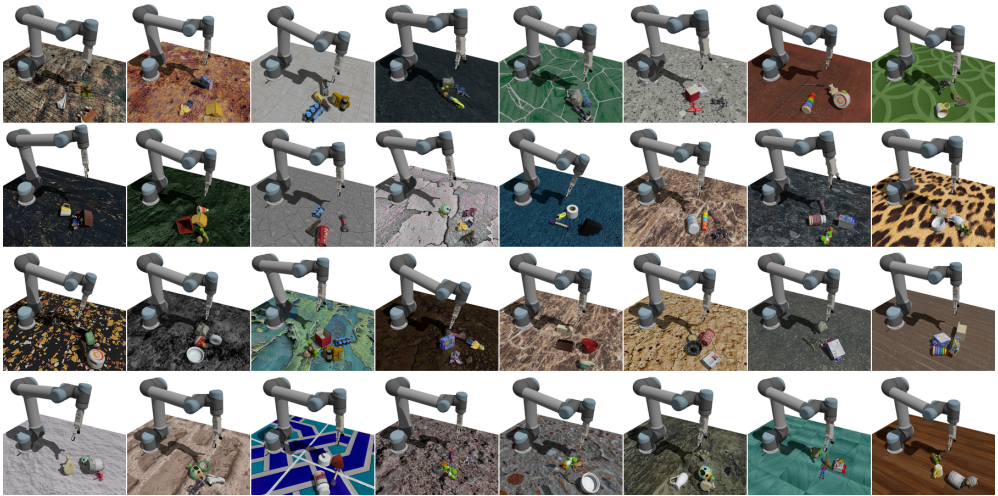
Simulation Environment

Datasets



Simulation Environment

Domain Randomisation

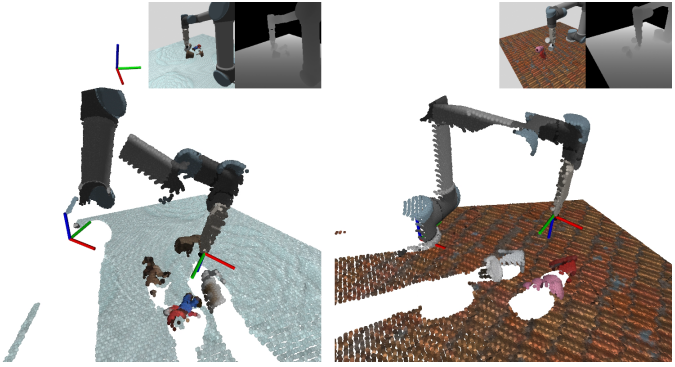


Simulation Environment

Domain Randomisation

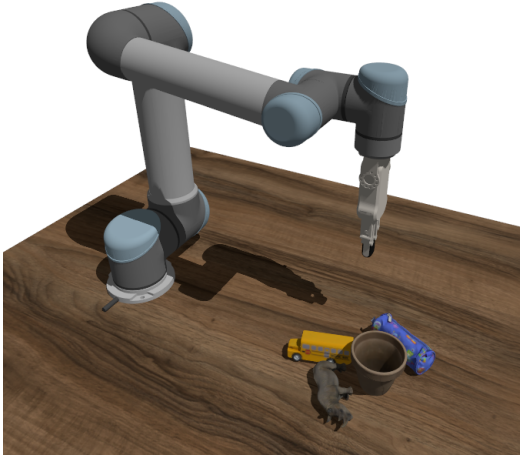
Random

- Object
  - Model
  - Scale
  - Mass
  - Friction
  - Pose
- Ground plane texture
- Initial robot configuration
- Camera
  - Pose
  - Sensory noise



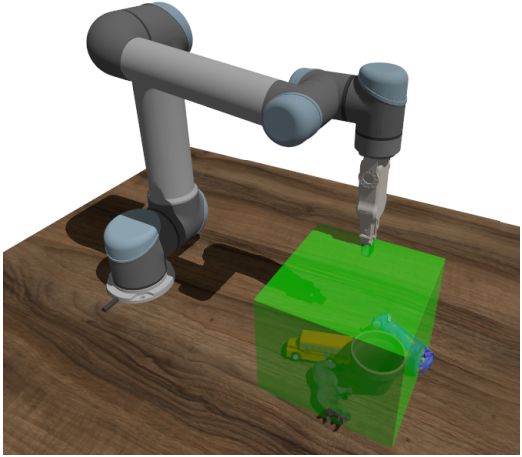
Simulation Environment

Environment for Training



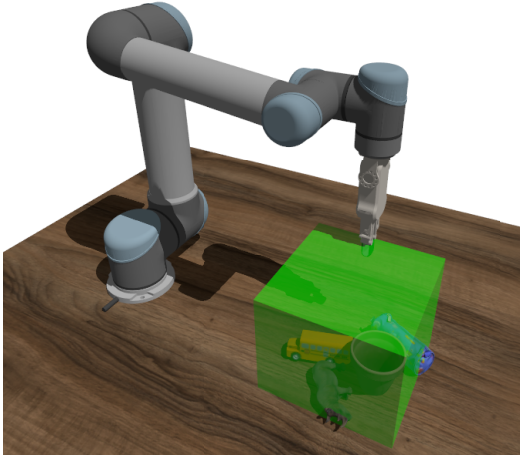
Simulation Environment

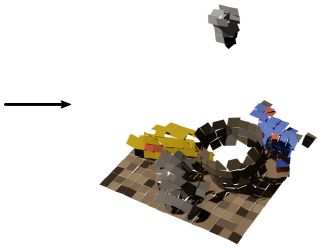
Environment for Training



Simulation Environment

Environment for Training





Training
Hyperparameters

Optimisation

- Automatic (Optuna)
- Manual

Hyperparameter	**TD3**	**SAC**	**TQC**
Optimisation Algorithm	Adam		
Learning Rate Schedule	Linear, $1.5 \cdot 10^{-4} \rightarrow 0$		
Mini-batch Size	32		
Update Frequency	After Every Episode		
Gradient Steps per Update	100		
Replay Buffer Size	40000		
Discount Factor γ	0.999		
Target Update Rate τ	$5 \cdot 10^{-5}$		
Number of Critics	2		
Activation Function	ReLU		
Exploratory Action Noise	$\mathcal{N}(0, 0.025)$		
Target Policy Noise	$\mathcal{N}(0, 0.25)$	—	—
Initial Entropy Coefficient	—	0.1	
Entropy Target	—	$-dim(\mathcal{A}) = -5$	
Number of Atoms	—	—	25
Number of Truncated Atoms	—	—	3

Training
Demonstrations and Curriculum

Demonstrations

- Automatic collection of samples
  - Simple scripted policy
    - 19% success rate
- 5k Collected transitions
  - Replaced after 40k steps (buffer size)

Curriculum

- Scaling of environment difficulty
  - Number of objects
    - $1 \rightarrow 4$
  - Spawn area
    - $2.4 \text{ cm} \times 2.4 \text{ cm} \rightarrow 24 \text{ cm} \times 24 \text{ cm}$
- Full problem at 60% success rate

Results
Overview

Experiments

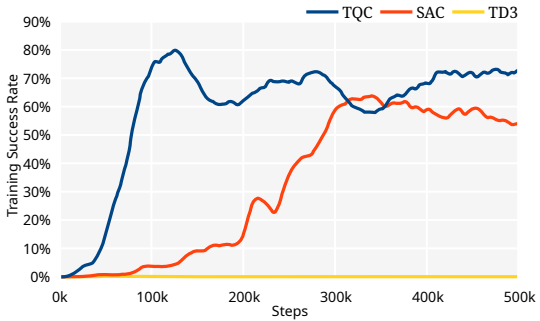
- Comparison of actor-critic algorithms
- Comparison of 2D/2.5D/3D observations
- Invariance to robot
- Sim-to-real transfer

Ablation Studies

- No demonstrations
- No curriculum
- No colour features
- No proprioceptive
- Shared/separate feature extractor
  - Among actor and critics
  - Among observation stacks

Results

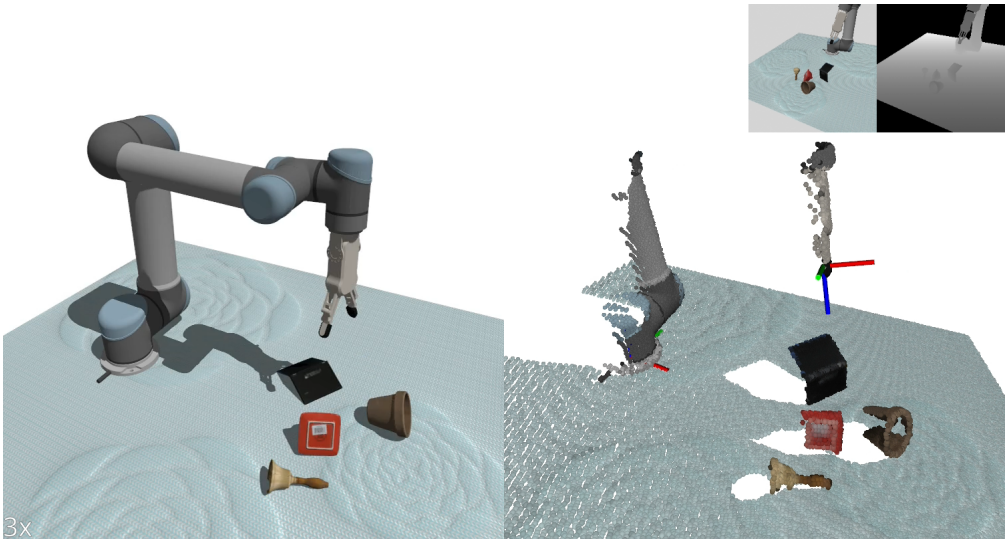
Comparison of Actor-Critic Algorithms



	TQC	SAC	TD3
Success Rate	77%	64%	0%
Episode Length	14.0	29.8	—

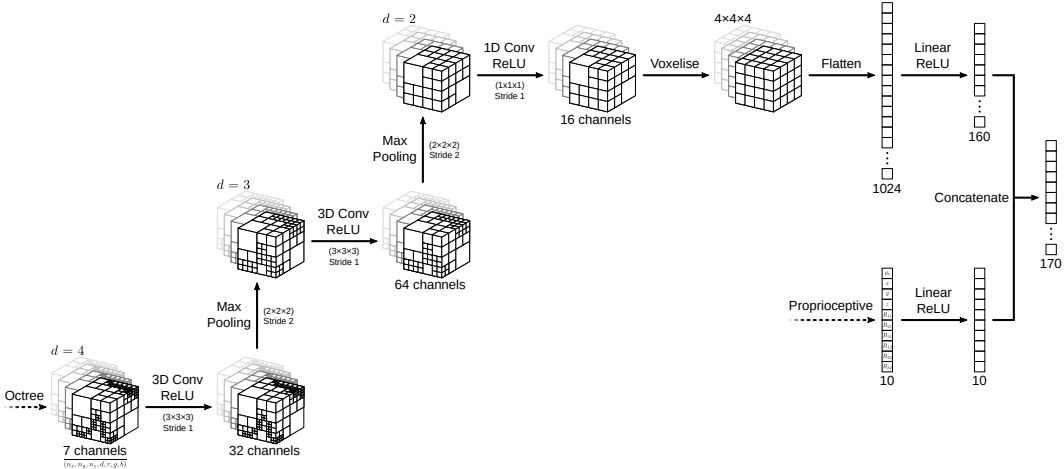
Results

Agent Trained with TQC (Video Example)



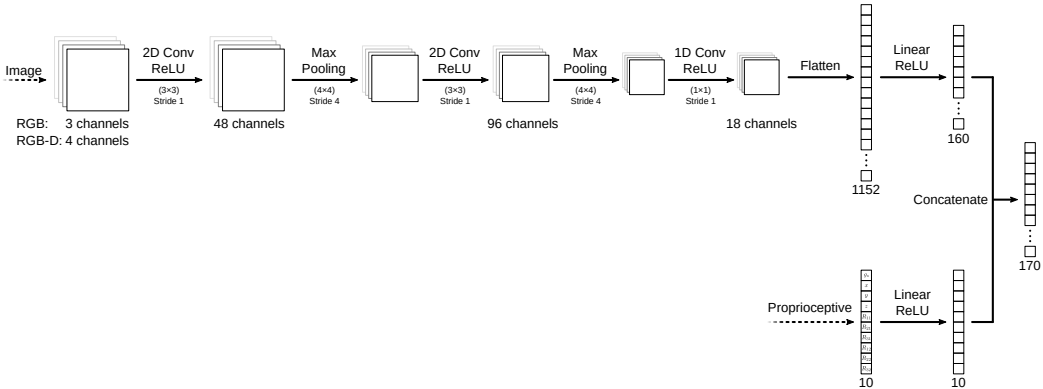
Results

Comparison of 2D/2.5D/3D Observations - Feature Extractor



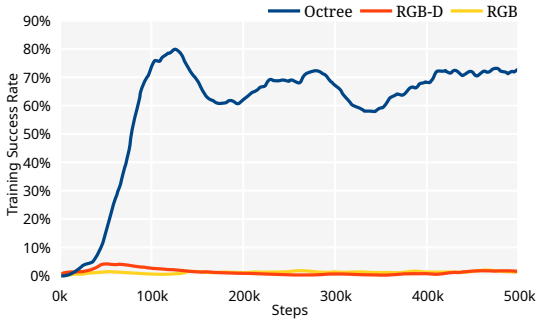
Results

Comparison of 2D/2.5D/3D Observations - Feature Extractor



Results

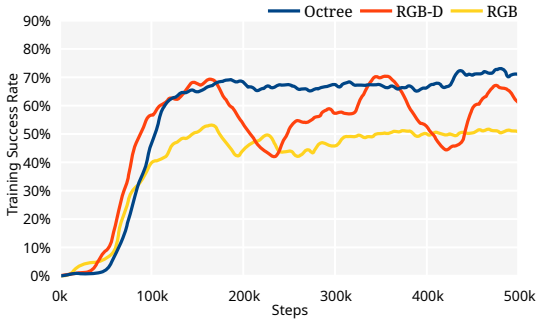
Comparison of 2D/2.5D/3D Observations - Random Camera Pose



	Octree	RGB-D	RGB
Success Rate	77%	5%	3%
Episode Length	14.0	36.5	51.0

Results

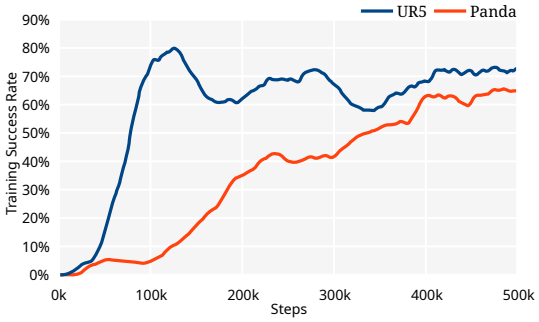
Comparison of 2D/2.5D/3D Observations - Static Camera Pose



	Octree	RGB-D	RGB
Success Rate	81.5%	59%	35%
Episode Length	24.6	9.4	9.3

Results

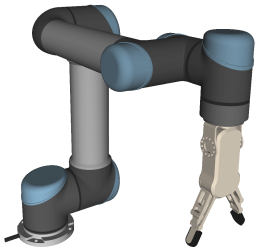
Invariance to Robot - Training and Evaluation



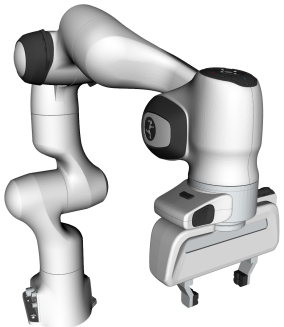
	UR5	Panda
Success Rate	77%	61.5%
Episode Length	14.0	27.1

Results

Invariance to Robot - Transfer of Policy

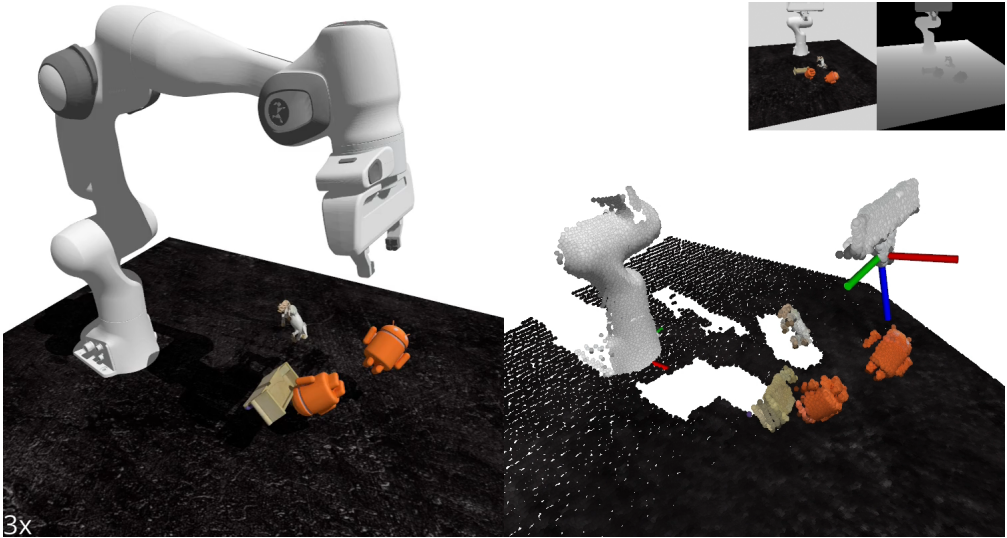


		Evaluation	
		UR5	**Panda**
Training	**UR5**	77%	27.5%
	Panda	75%	61.5%



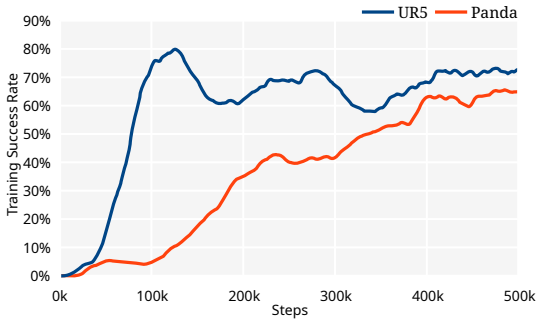
Results

Policy Trained on Panda (Video Example)



Results

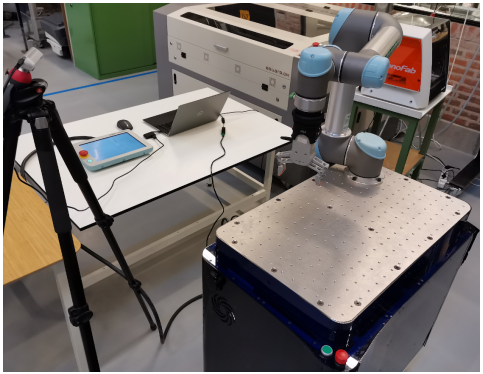
Invariance to Robot - Training and Evaluation



	UR5	Panda
Success Rate	77%	61.5%
Episode Length	14.0	27.1

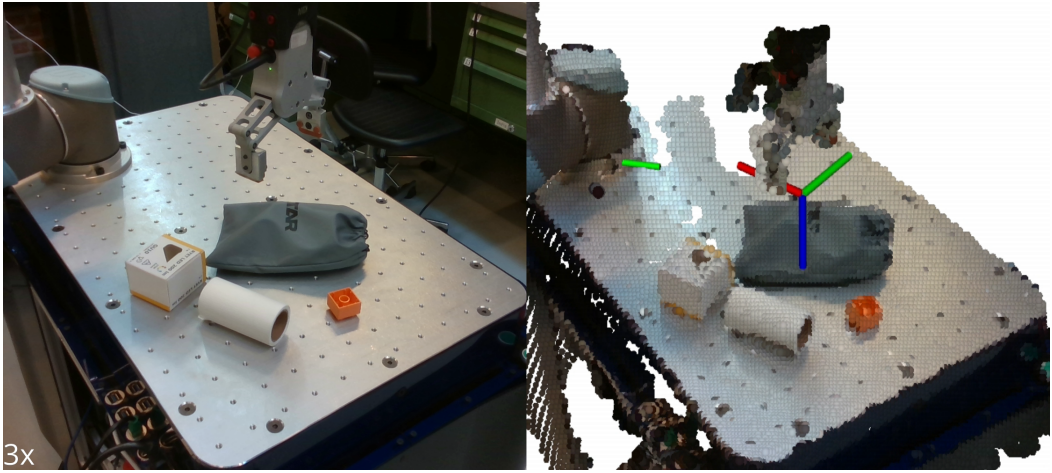
Results

Sim-to-Real Transfer - Setup

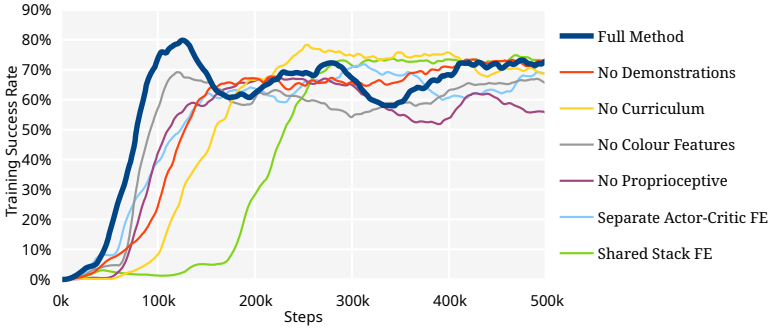


Results

Sim-to-Real Transfer (Video Example)



Results
Ablation Studies



	Full Method	No Demon-strations	No Curriculum	No Colour Features	No Proprio-ceptive	Separate Actor-Critic FE	Shared Stack FE
Success Rate	77%	84%	70.5%	66.5%	75%	68.5%	79%
Episode Length	14.0	24.5	19.9	29.4	23.0	27.5	22.8

Conclusion

Primary Contributions

- Simulation environment with domain randomisation
- Octree observations for end-to-end grasping with DRL

Performance and Scalability

- Slow data collection and training
- Parallel workers are needed, with asynchronous updates of policy

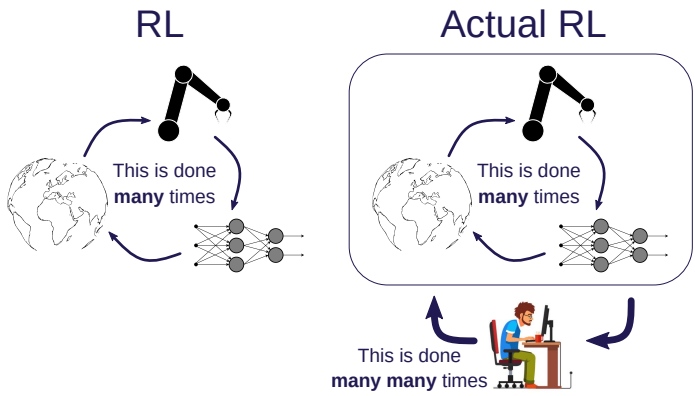
Reproducibility

- Not great for DRL
- Pre-built Docker image

```
drl_grasping/docker/run.bash andrejorsula/drl_grasping:latest ros2 run drl_grasping ex_enjoy_pretrained_agent.bash
```

Perspective on Reinforcement Learning

Great approach for learning complex robotics tasks! However, ...



Inspired by: Sergey Levine. 2020. CS 285 at UC Berkeley - L23

Thank you for your time!

AALBORG UNIVERSITY