

Deep Reinforcement Learning for Robotic Grasping from Octrees

Learning Manipulation from Compact 3D Observations

Master's Thesis

June 25, 2021

Andrej Orsula

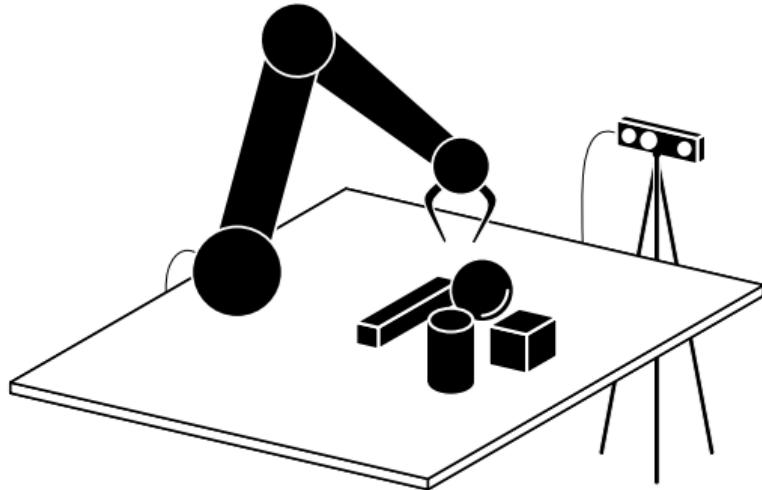
Aalborg University
Denmark



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

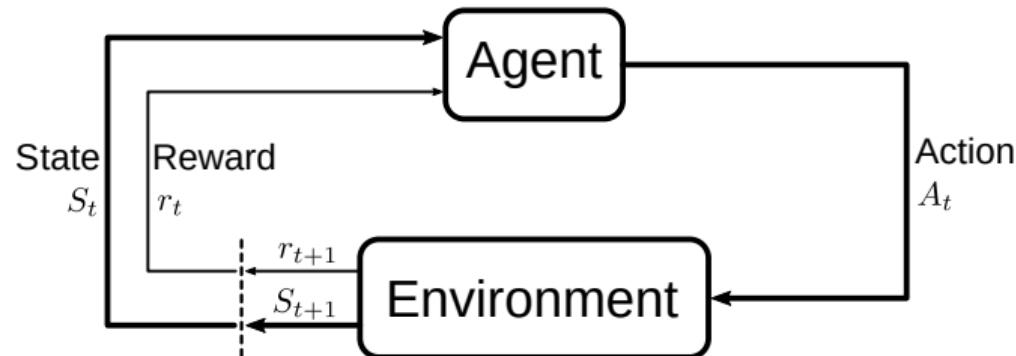


Vision-Based Robotic Grasping of Diverse Objects

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)

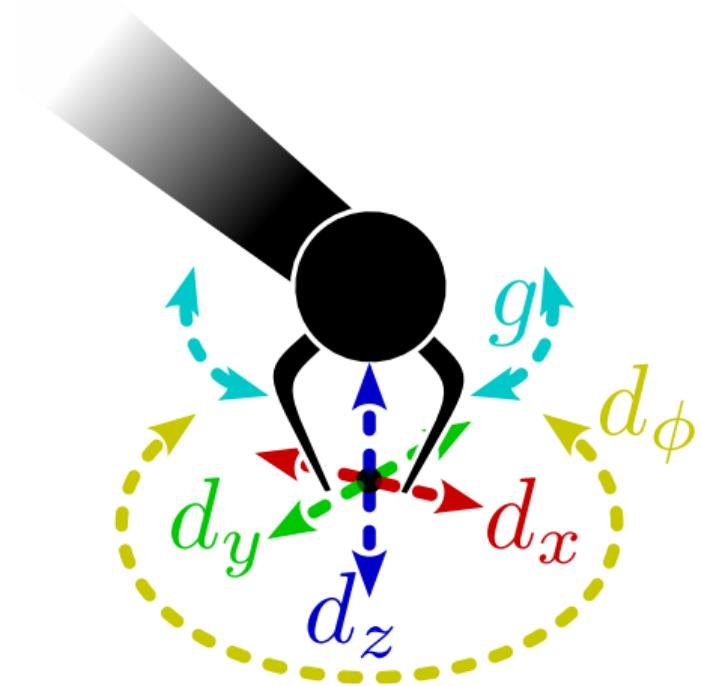


Task Definition

Action Space

Actions in Cartesian Space

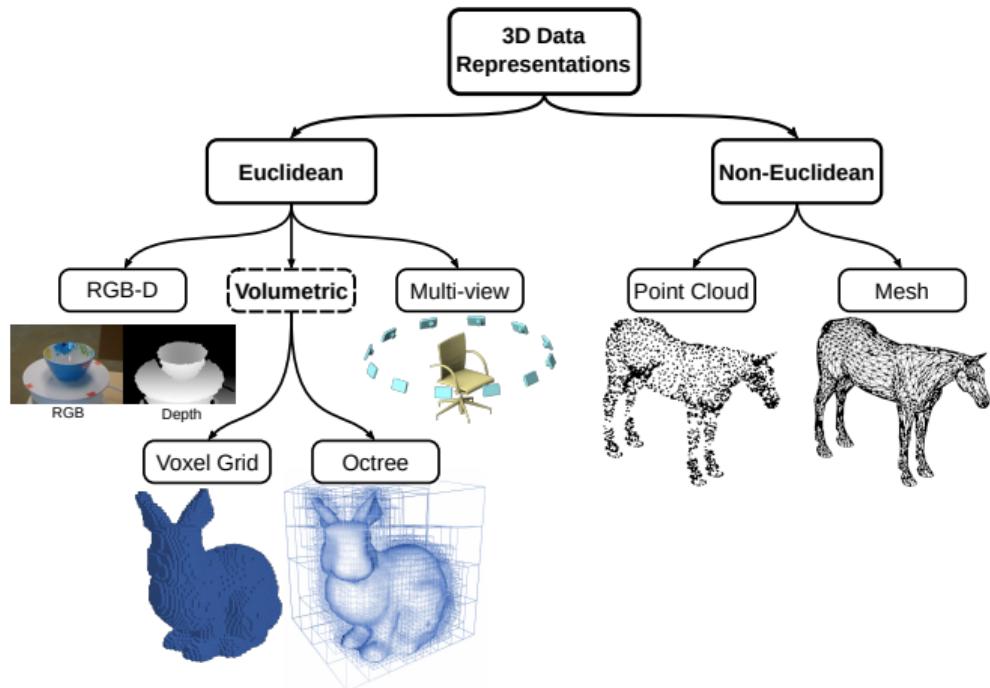
- ▶ Translational displacement
 - ▶ d_x
 - ▶ d_y
 - ▶ d_z
- ▶ Gripper rotation
 - ▶ d_ϕ
- ▶ Gripper actions (open/close)
 - ▶ g





Task Definition

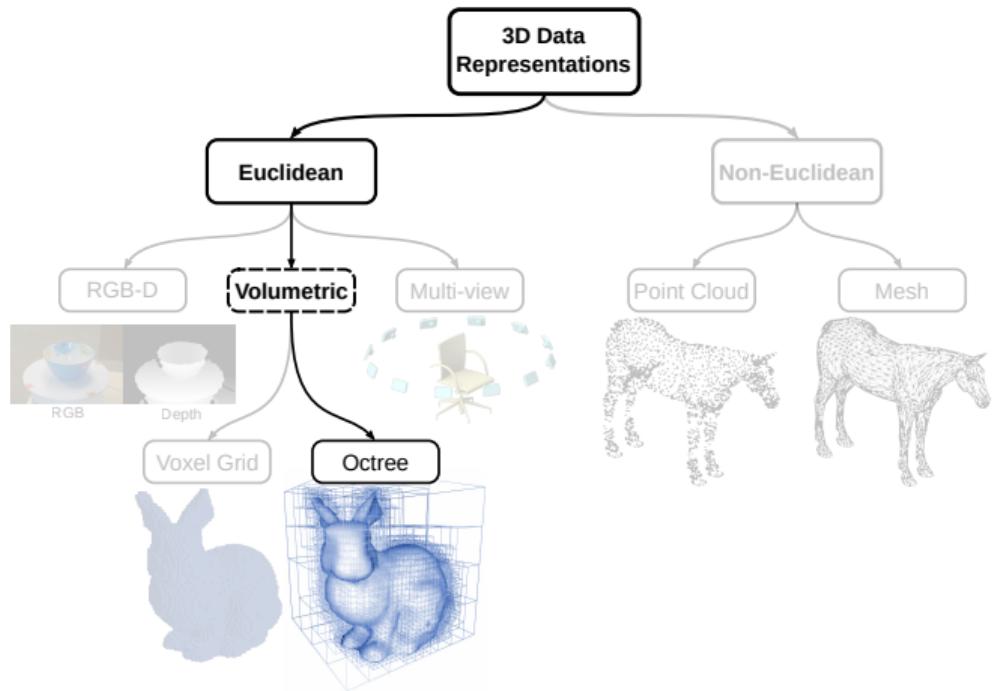
Observation Space





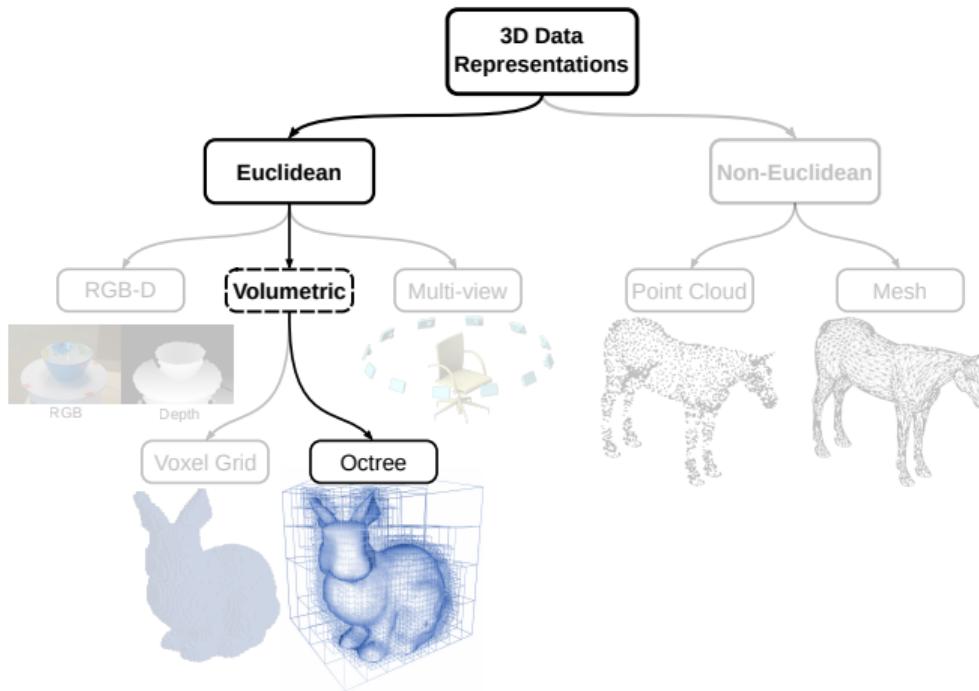
Task Definition

Observation Space



Task Definition

Observation Space



Proprioceptive Observations

- ▶ Gripper position
- ▶ Gripper rotation
- ▶ Gripper state



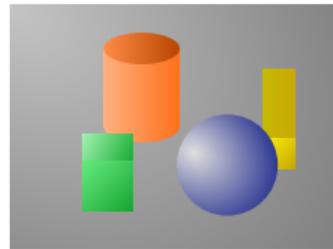
Task Definition

Observation Space - Construction of Octree

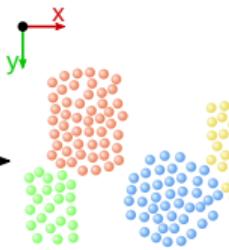
Depth Map



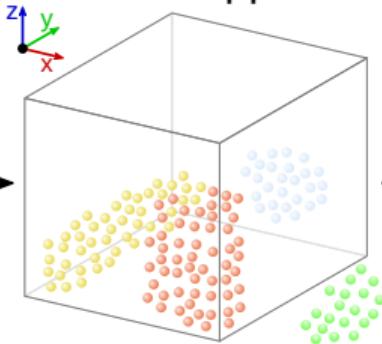
RGB Image



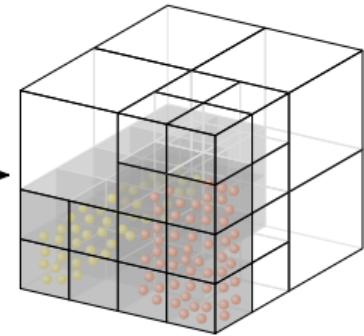
Point Cloud



Transformed
and Cropped



Octree



Task Definition

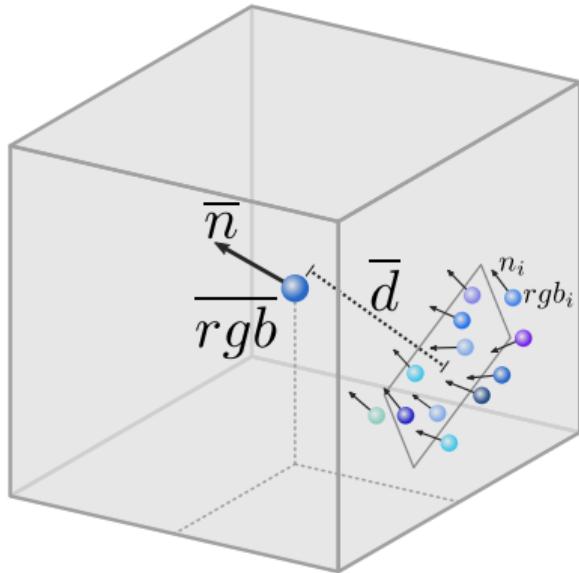
Observation Space - Features and Stacks

Features

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

Observation Stacking

- ▶ Three consecutive observations





Task Definition

Reward Function

Composite Reward

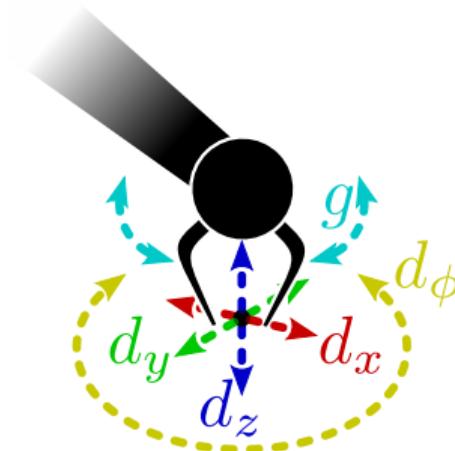
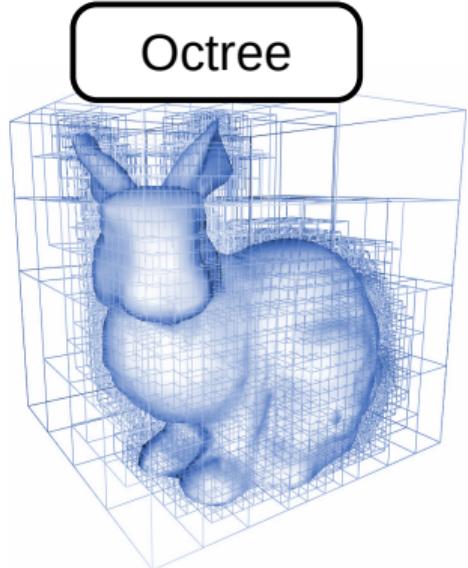
- ▶ Reach
 - ▶ $+1 (7^0)$
- ▶ Touch
 - ▶ $+7 (7^1)$
- ▶ Grasp
 - ▶ $+49 (7^2)$
- ▶ Lift
 - ▶ $+343 (7^3)$

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



Reinforcement Learning

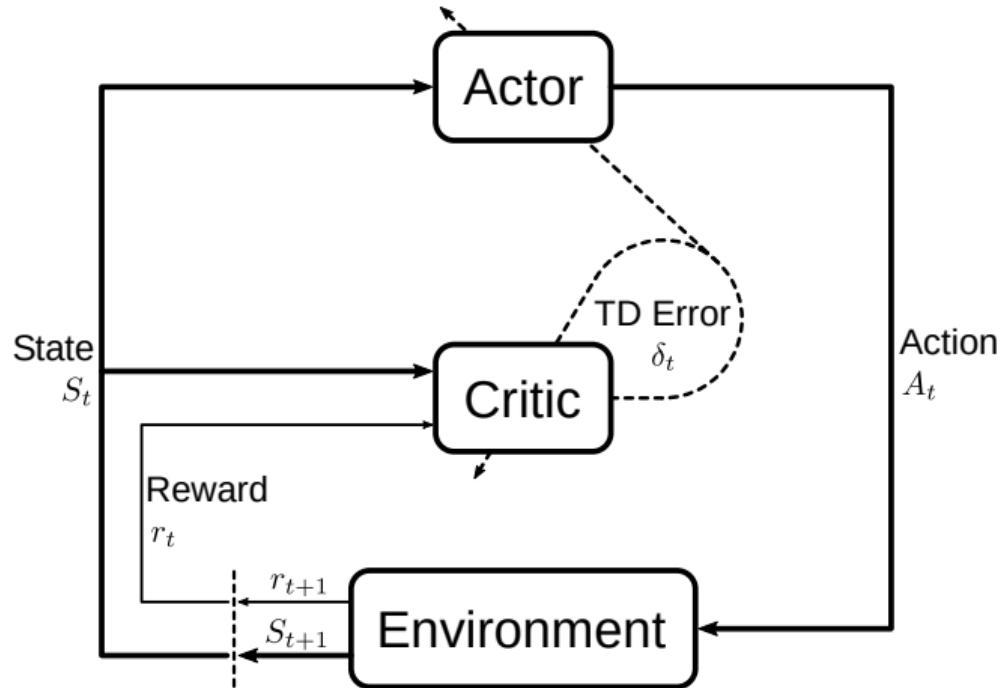
Algorithms

Actor-Critic Algorithms

- ▶ TD3
- ▶ SAC
- ▶ TQC

Implementation

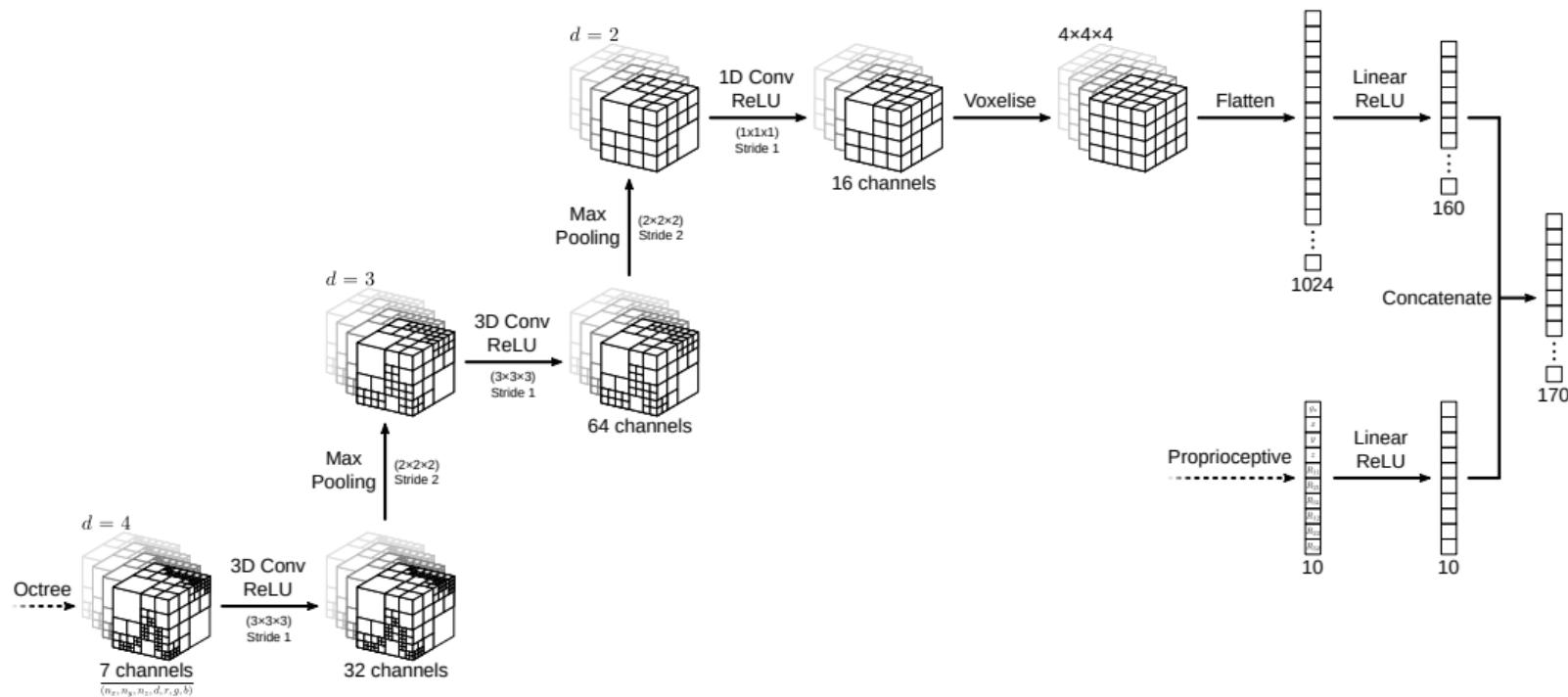
- ▶ Stable Baselines3





Deep Reinforcement Learning

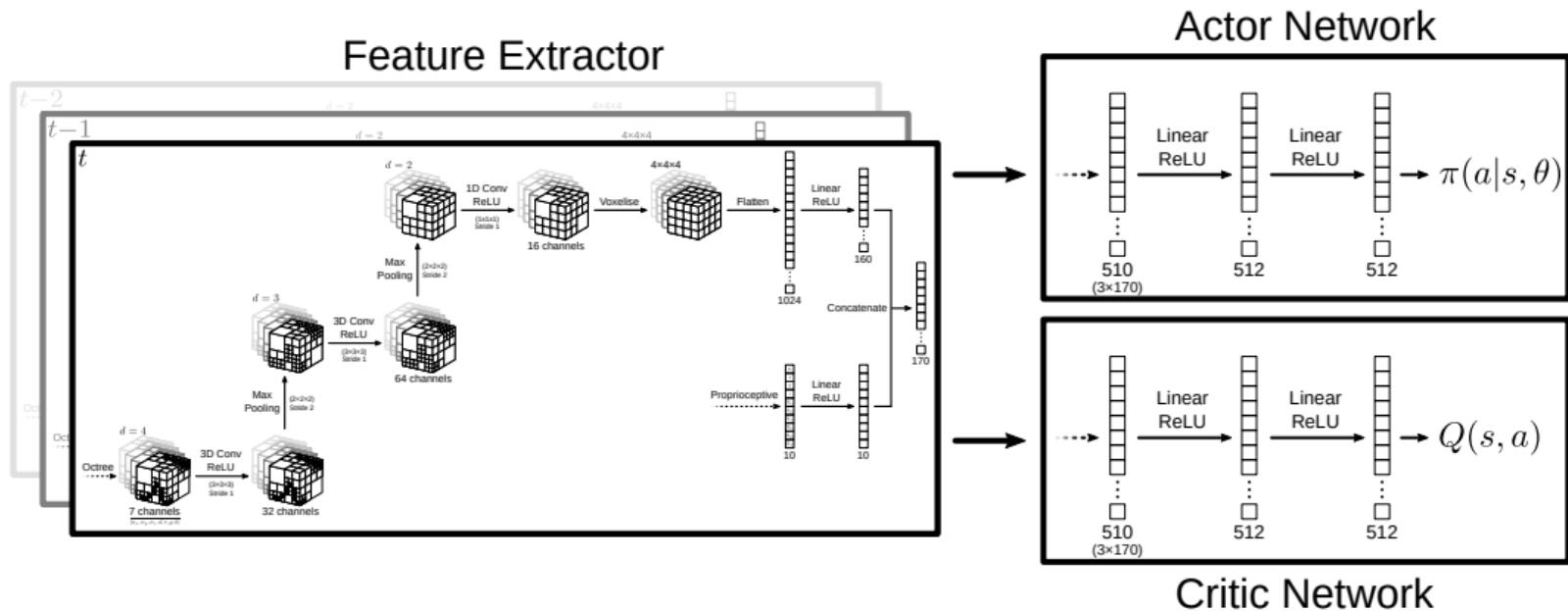
Octree-Based Feature Extractor





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



Gym-Ignition

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

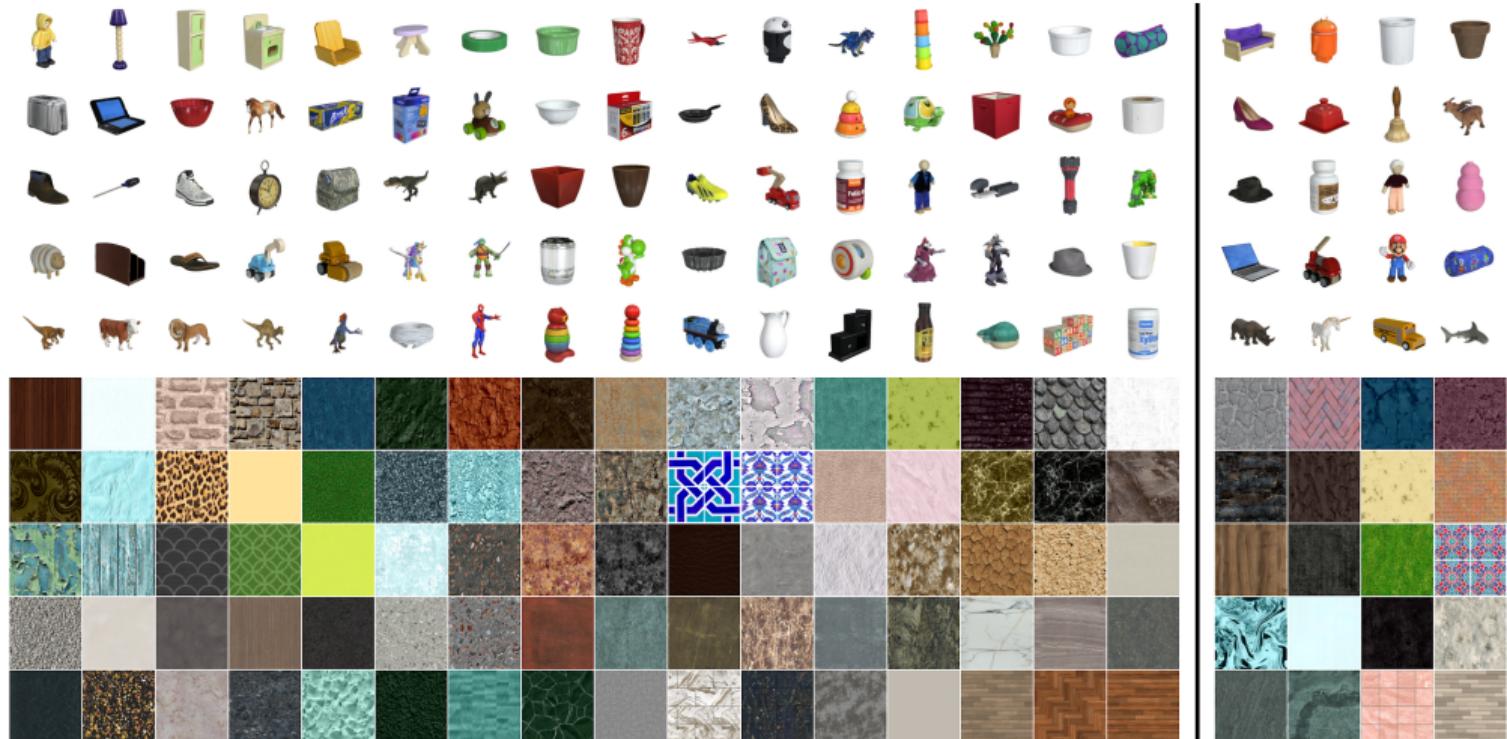
Rendering





Simulation Environment

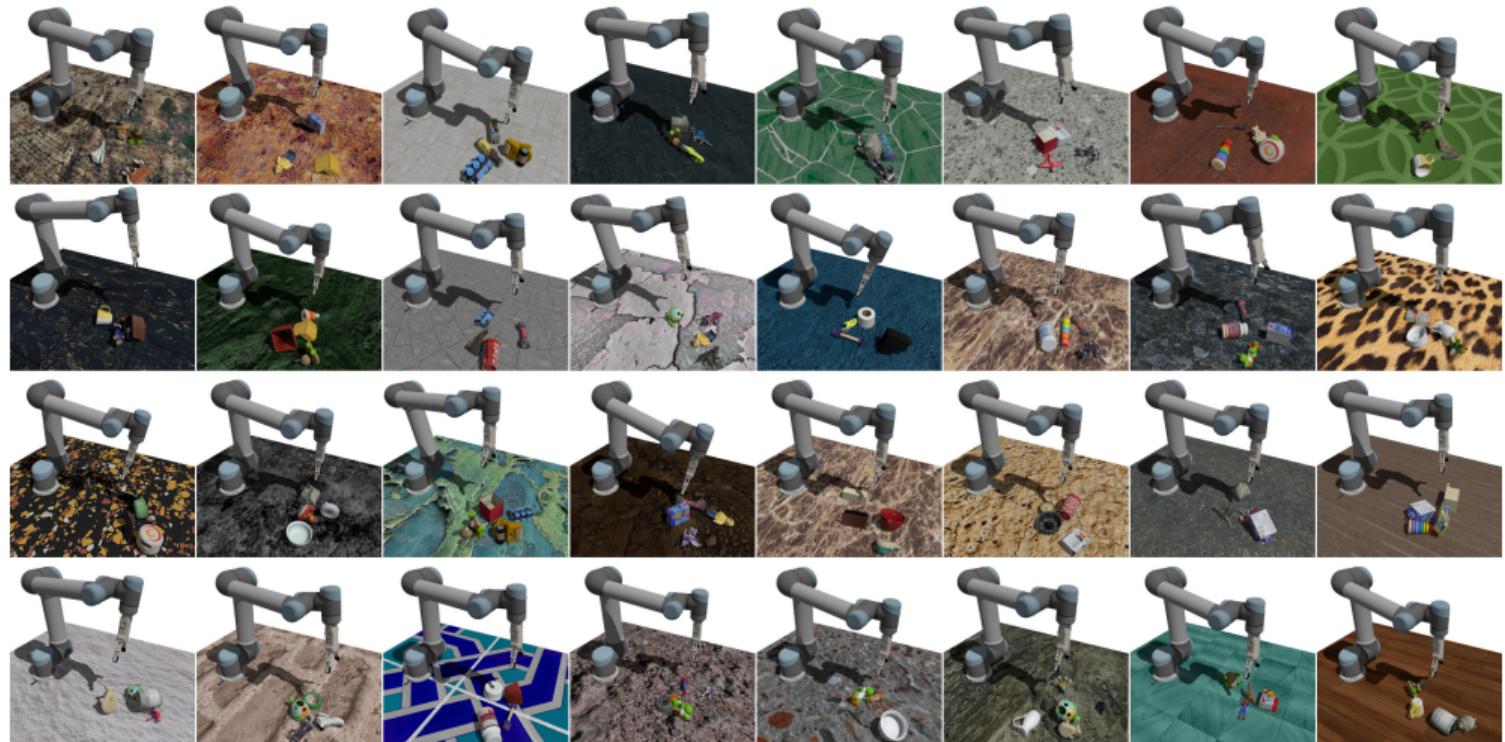
Domain Randomisation





Simulation Environment

Domain Randomisation

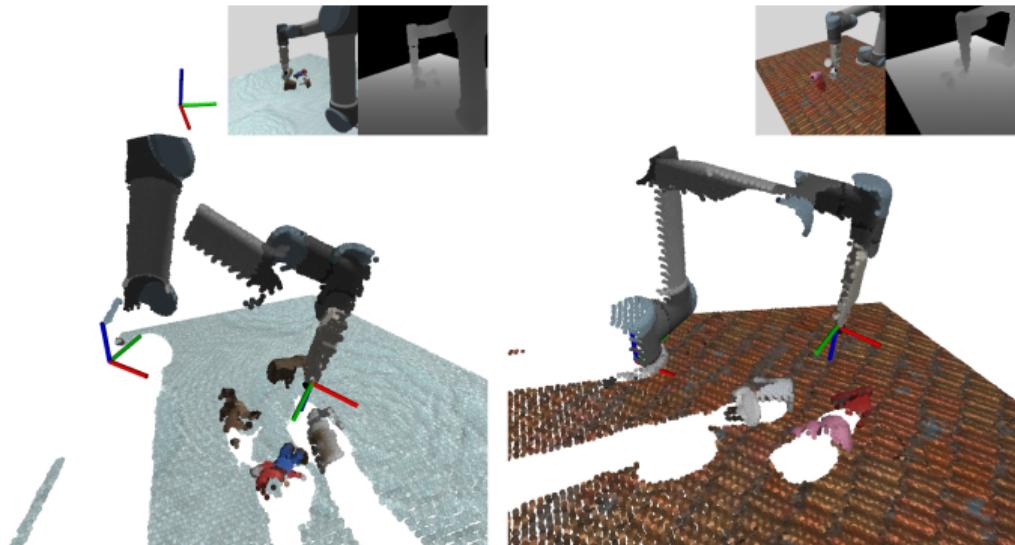


Simulation Environment

Domain Randomisation

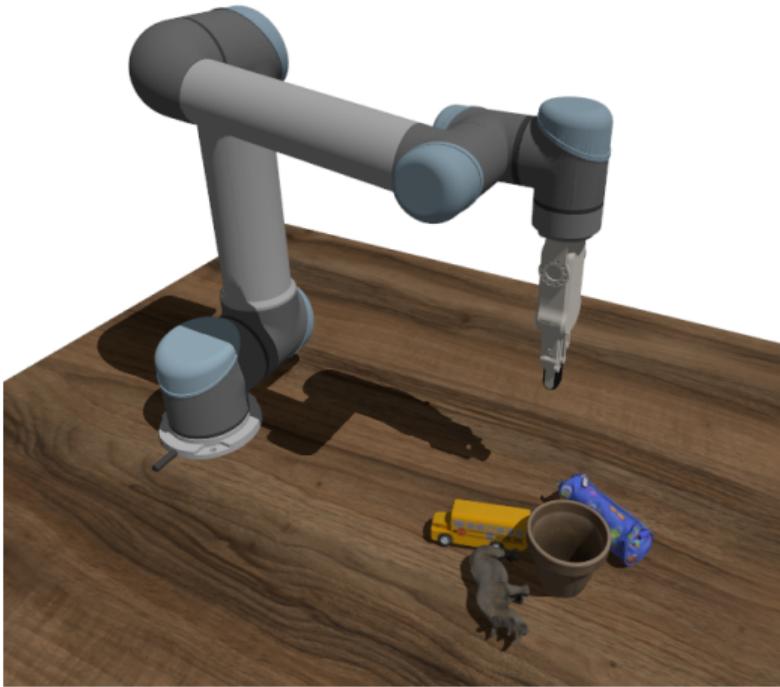
Random

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



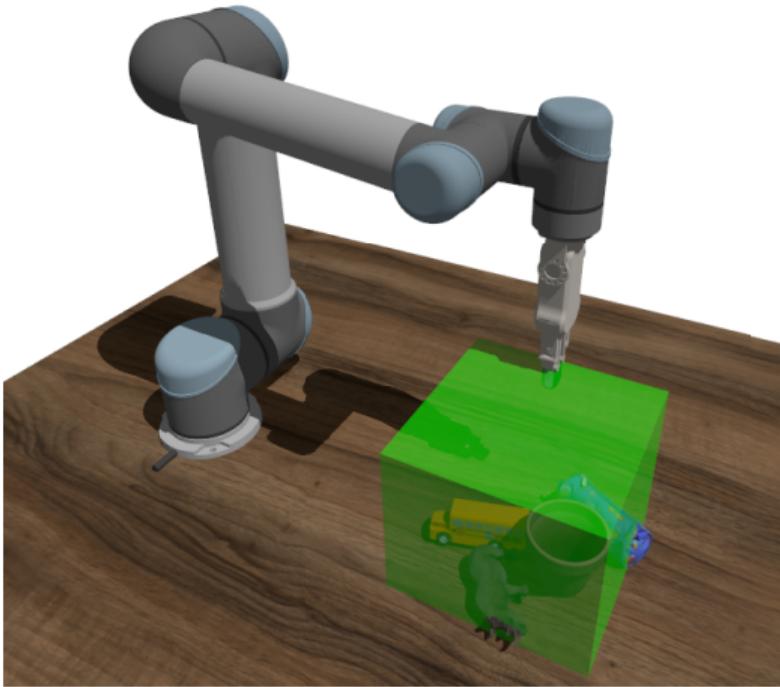
Simulation Environment

Environment for Training



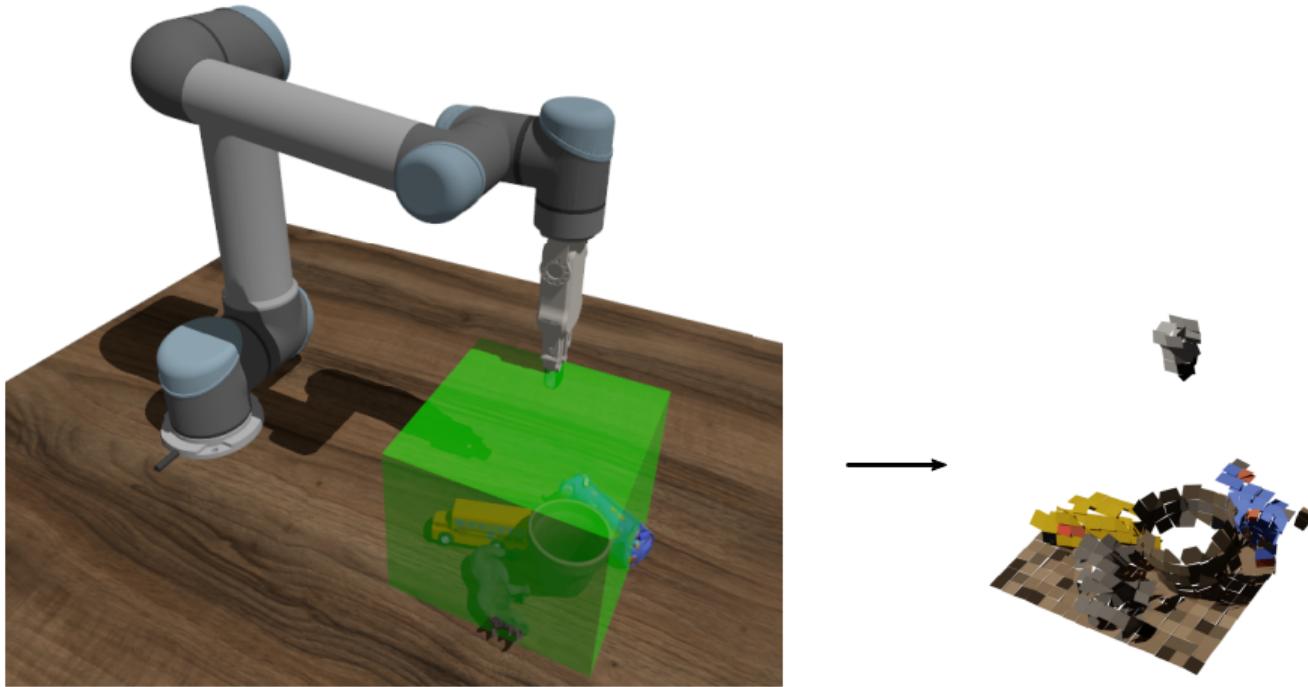
Simulation Environment

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

Thank you for your time



AALBORG UNIVERSITY