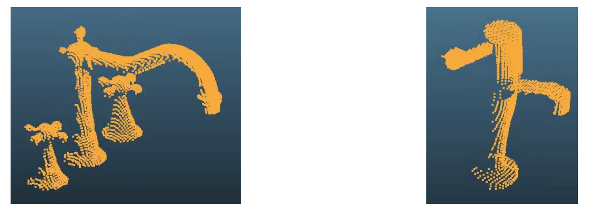
PointPatchRL: Masked Reconstruction Improves RL on Point Clouds

Balázs Gyenes[1] Nikolai Franke[1] Philipp Becker[1] Gerhard Neumann[1]

[1]Autonomous Learning Robots, Karlsruhe Institute of Technology, Karlsruhe

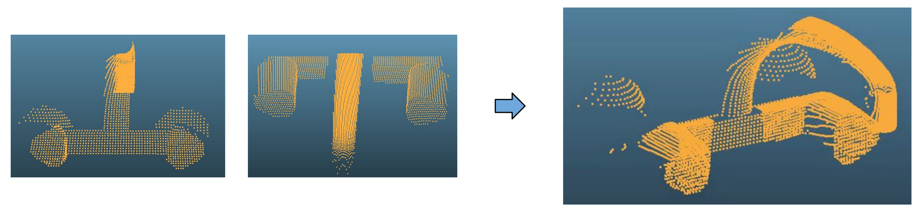
Point Clouds

- capture **task-relevant geometry** of deformable or varying objects



- disentangle **occlusions** in 3D space

- combine **multiple camera** views



Method

(see also main figure)

PointPatchRL builds on point patching paradigm of PointMAE and PointGPT

1. Split point clouds into (overlapping) **patches**

$$C = \text{fps}(X), \quad C \in \mathbb{R}^{n \times 3}$$

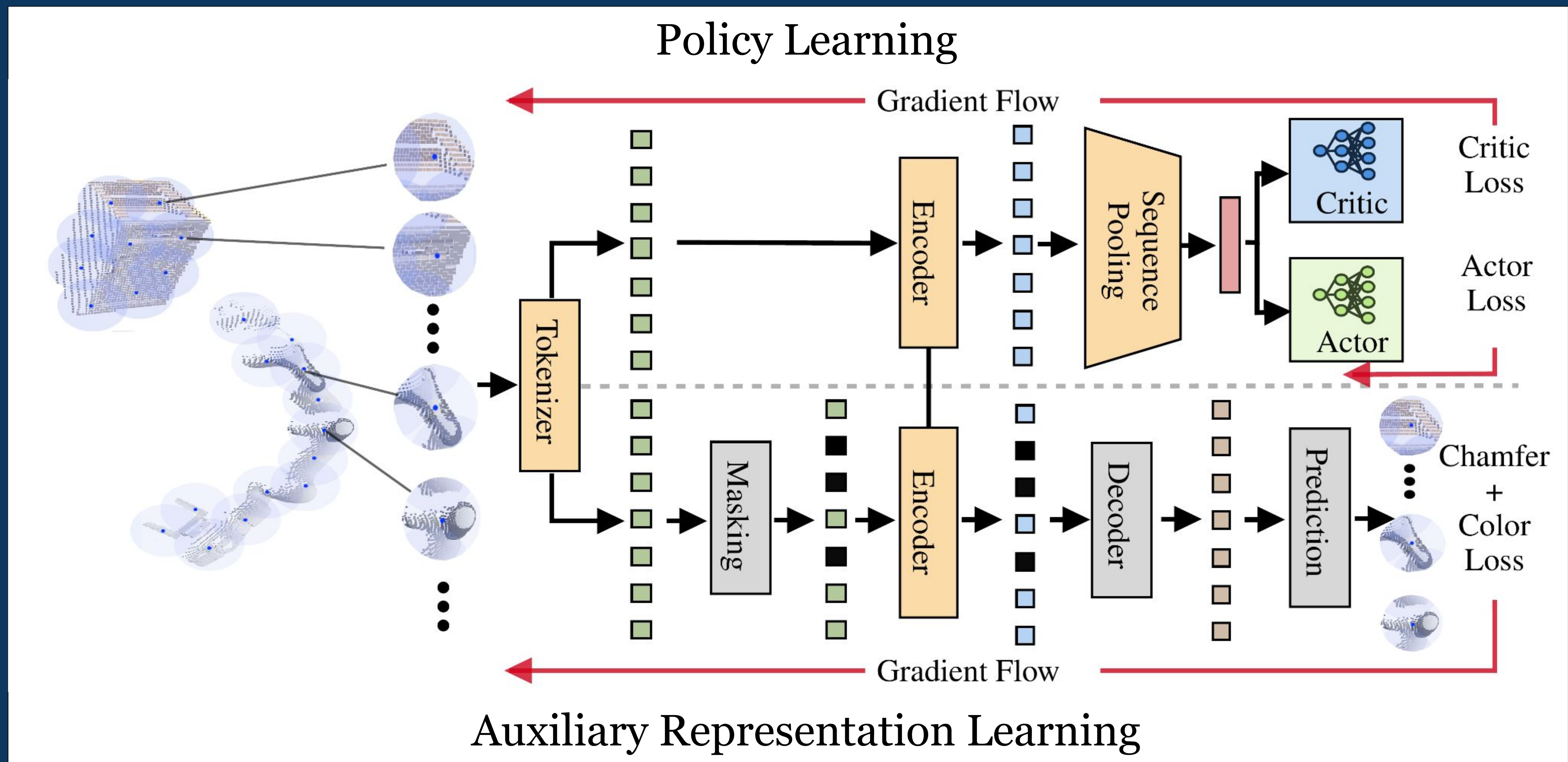
$$P = \text{knn}(X, C), \quad P \in \mathbb{R}^{n \times k \times 3}$$

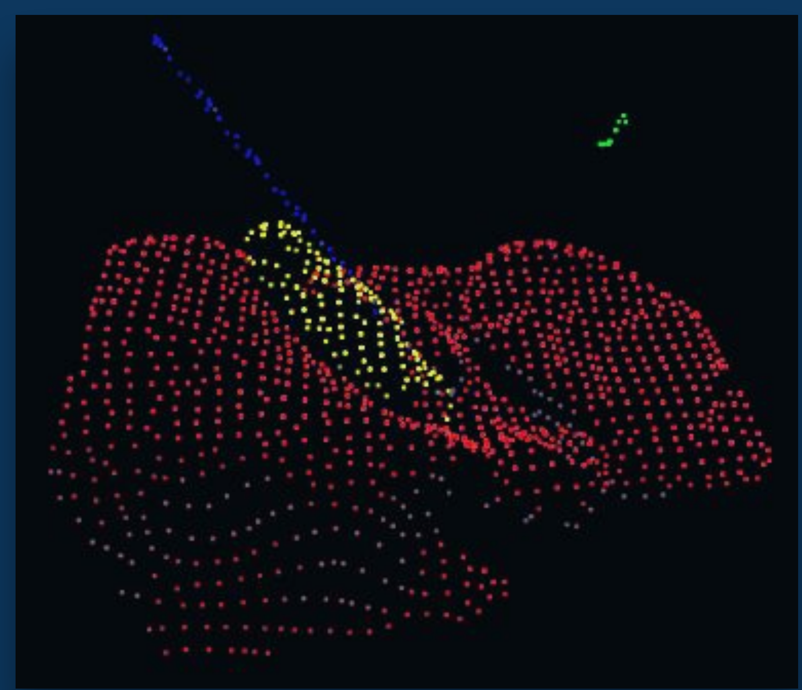
2. Convert each patch into a **token** and add **positional encoding**
3. Copy tokens and **mask** some of them
4. **Encode** both sequences with Transformer layers
5. Compute feature vector using **sequence pooling**, use MLPs to predict value and policy, and compute **SAC critic and actor** losses
6. **Decode** masked sequence
7. Reconstruct point cloud. Loss is chamfer distance for positions and MSE for colors

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} d_{\text{CD}}(P_i^{\text{pd}}, P_i^{\text{gt}}) + \mathcal{L}^{\text{rgb}}(P_i^{pd}, P_i^{\text{gt}})$$

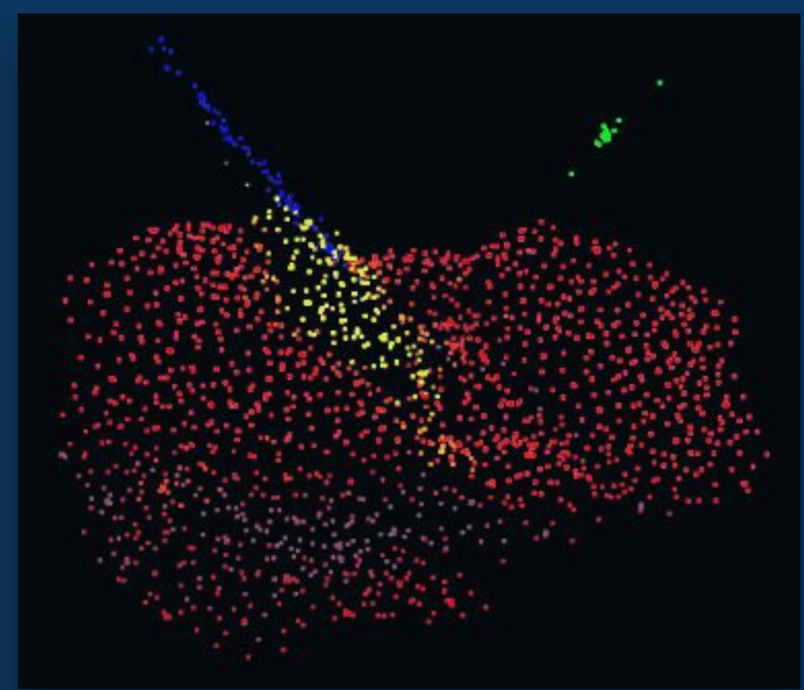
Karlsruhe Institute of Technology

Autonomous Learning Robots Lab





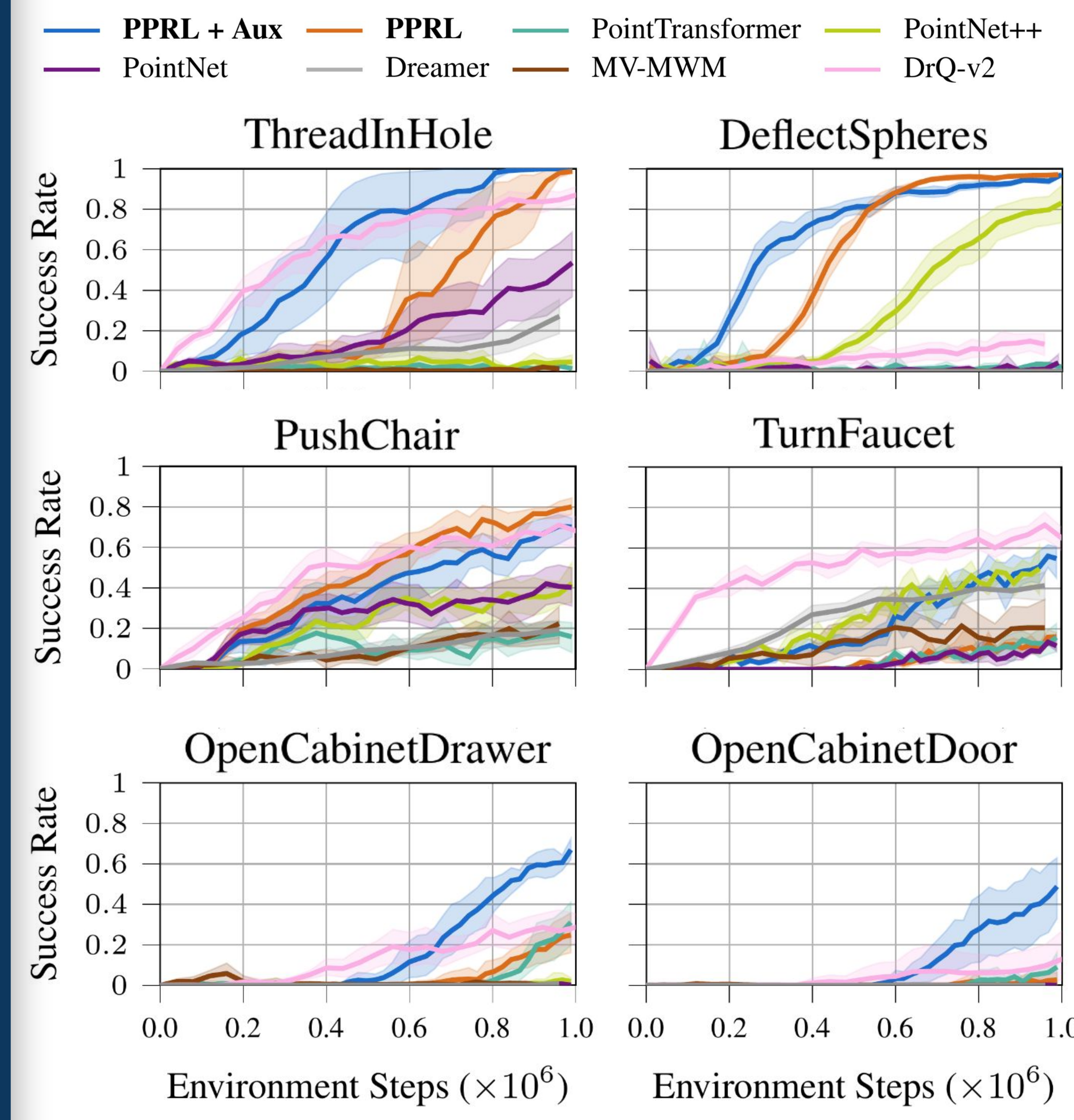
Ground Truth



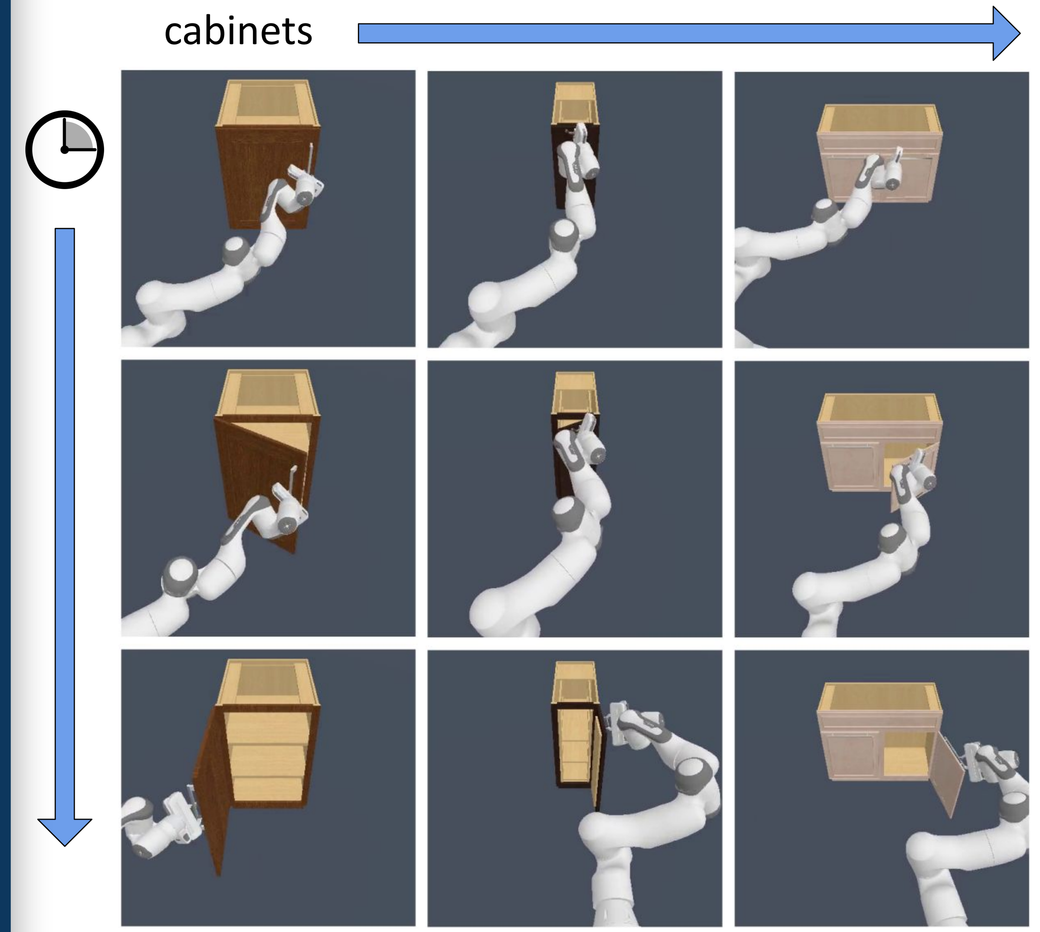
Reconstruction

Experiments

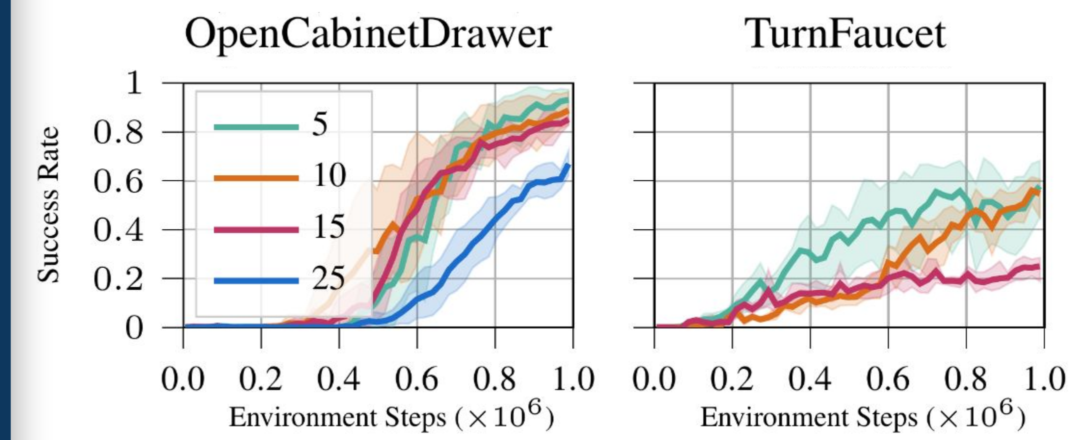
Favourable performance on challenging tasks compared to baselines



Diverse strategies conditioned on object geometry (same policy for each)



Encoder has sufficient capacity to handle more models - up to a critical point



- Train and test agents on increasing numbers of object models [5,10,15,25]