Interpretable reinforcement learning

About me

About me

Overview

- Motivation
- Problem statement
- Approaches
- Final method
- Results

Motivation

Robots would be great

Robots don’t understand the world

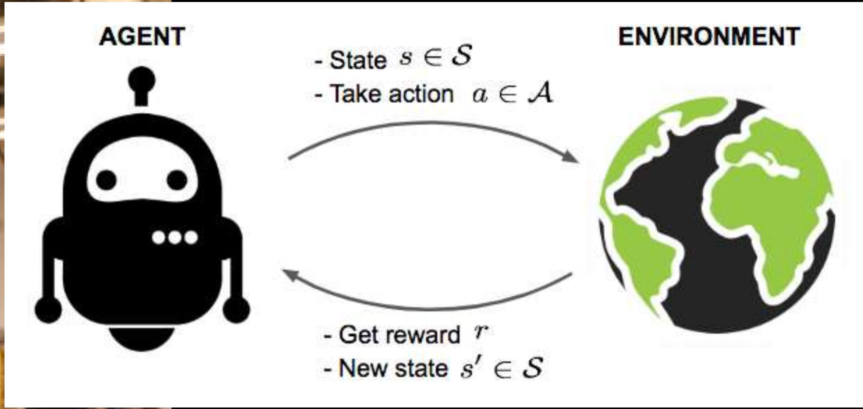
Supervised data is expensive

Built-in workflow pricing for labeling with Amazon Mechanical Turk

If you use a vendor, the cost per label is set by the vendor. You can see each vendor's pricing details in AWS Marketplace. If you use Amazon Mechanical Turk for labeling, you are charged per object per labeler. We recommend that you use multiple labelers per object to improve label accuracy.

Workflow	Suggested price per labeler
Image classification	$0.012
Text classification	$0.012
Named Entity Recognition (NER)	$0.024
Bounding box	$0.036
Semantic Segmentation	$0.84

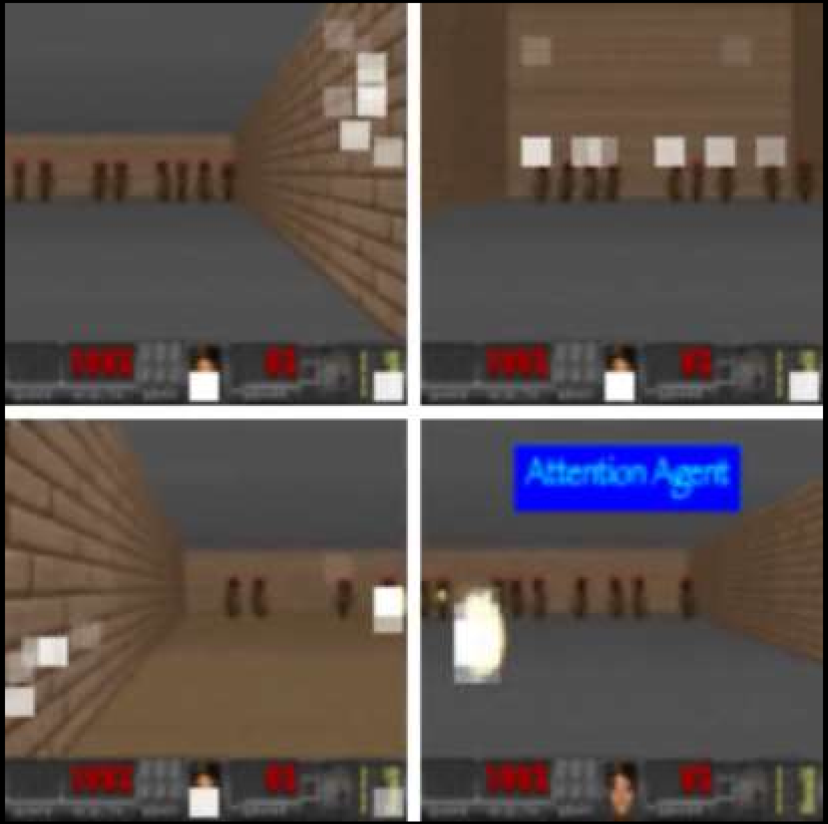
Robots should explore by themselves



How can the robot learn human concepts?

?

How can the robot learn human concepts?



Problem statement

Interpretable reinforcement learning

Procgen

Object-based reinforcement learning

Investigating Human Priors for Playing Video Games

Rachit Dubey Pulkit Agrawal Deepak Pathak Tom Griffiths Alexei A. Efros

University of California, Berkeley

ICML 2018

[Download Paper] [Github Code]

Human gameplay on game version without any object priors

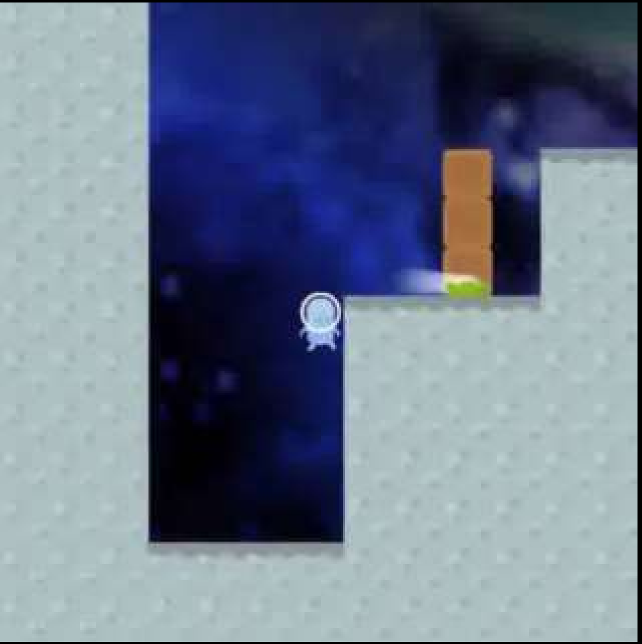
Human gameplay on original game version

Goal: add an object detector

Image → Object detector → Objects → RL

Approaches

1. Use a pretrained vision model (Detectron)

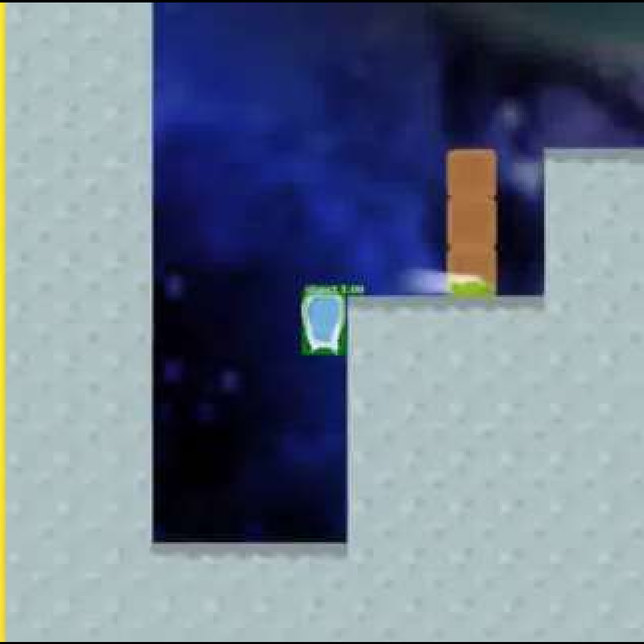


Original

1. Use a pretrained vision model (Detectron)

Original

Optical flow



Object detector

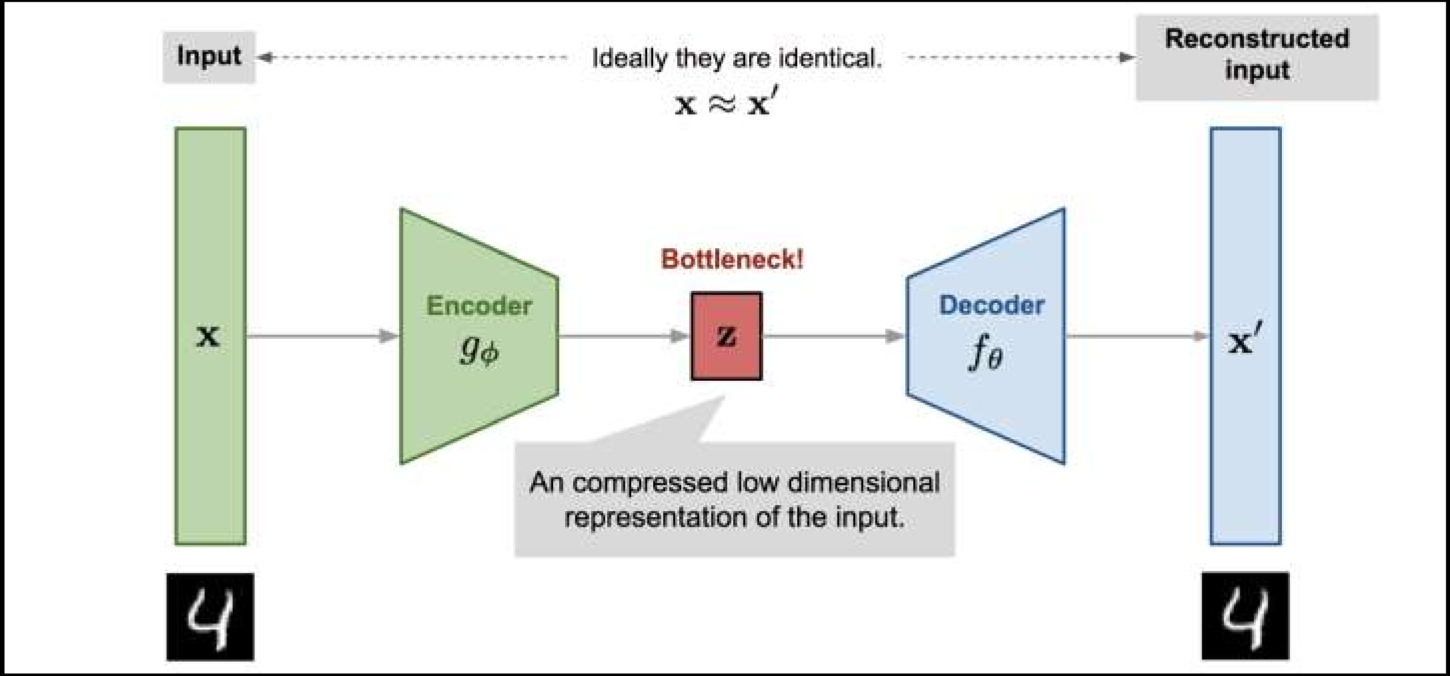
1. Use a pretrained vision model (Detectron)

Original

Optical flow

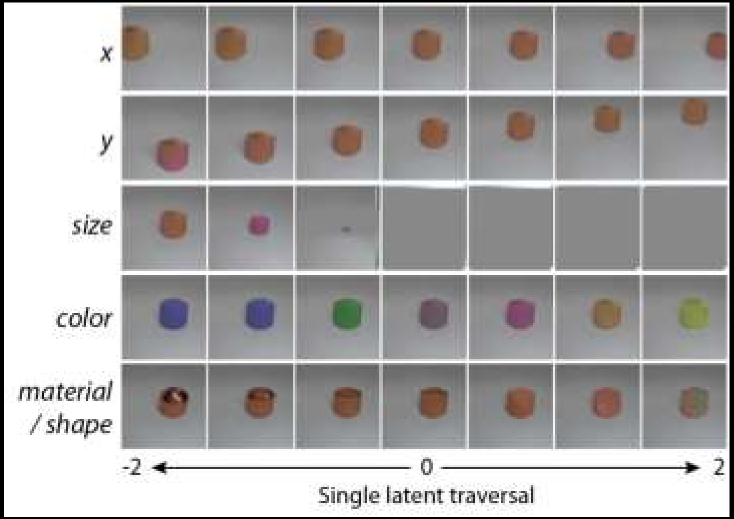
No objects
detected

2. Self-supervised learning



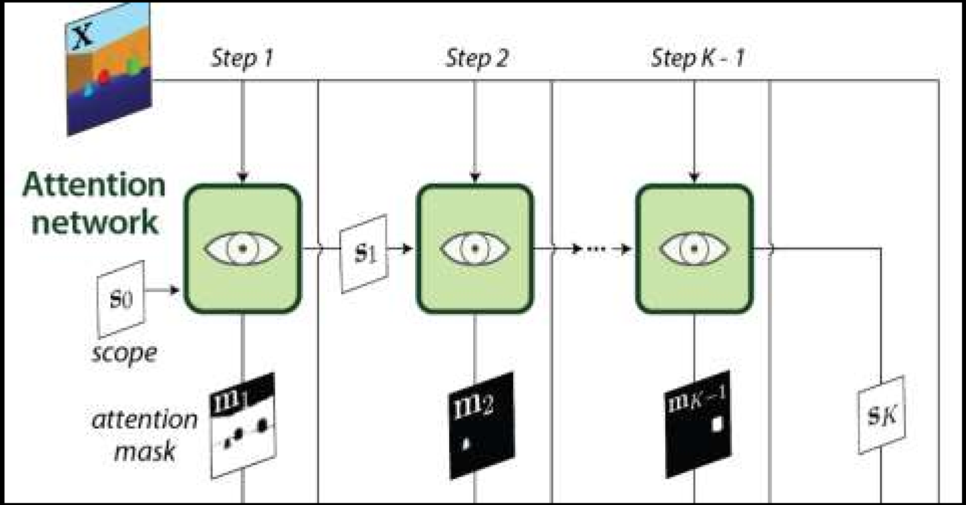
2. Self-supervised learning

a) Represent image as an array of object vectors



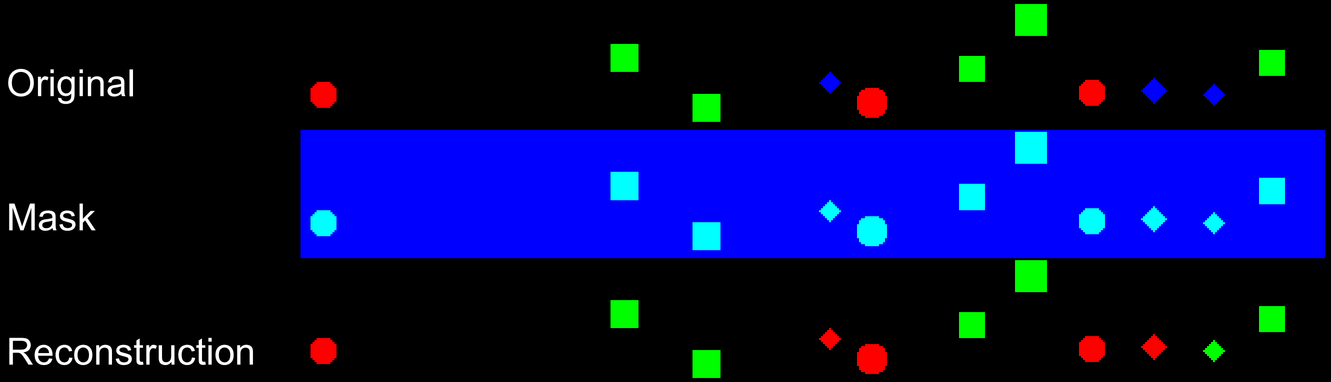
2. Self-supervised learning

b) Represent image as a list of (object category, object mask)



2. Self-supervised learning

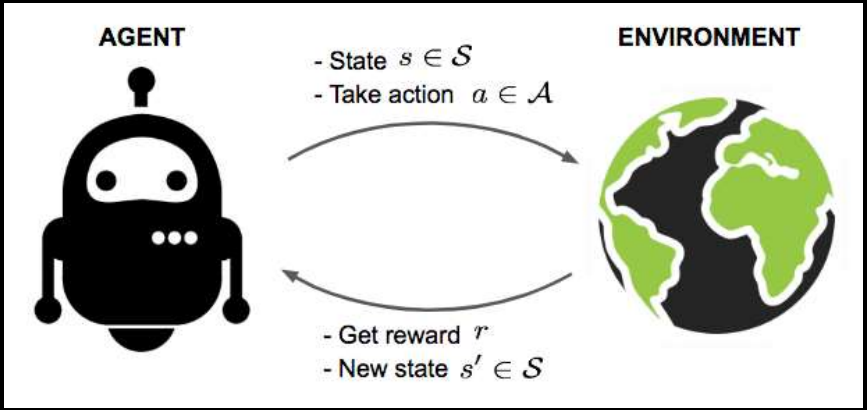
- Unreliable
- Encoder misses small objects
- Ignores task context



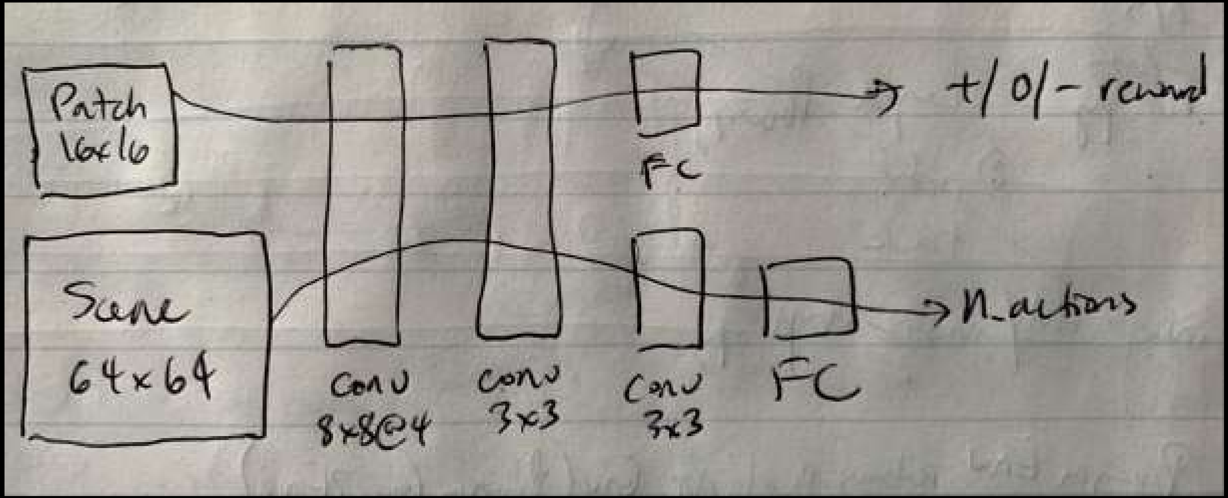
Approaches
(part 2)

Maybe we can add indirect biases

3. Model agent-object interactions

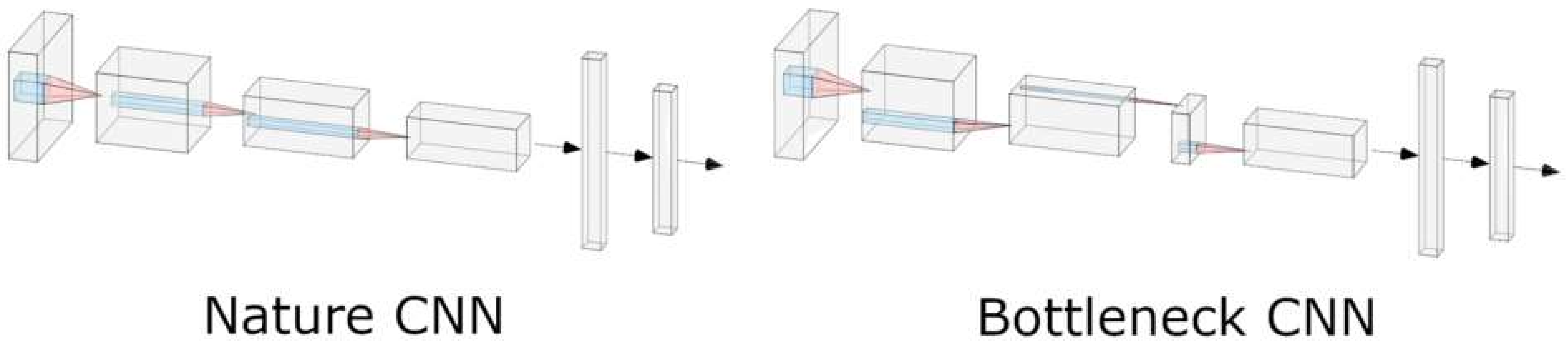


3. Model agent-object interactions



4. Add a channel bottleneck

- Prior: only a few object types
- Prior: true dimensionality is much lower



Experiment

- Train with 40 object textures
- Test on 6 new textures
- Same dynamics: agent needs to collect good objects and avoid bad objects

Good objects

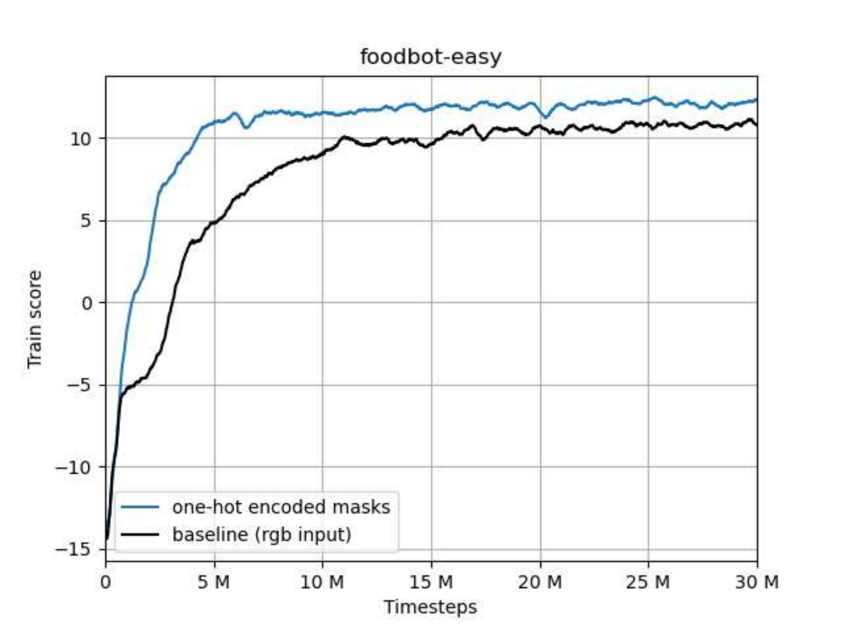
Bad objects

Final method

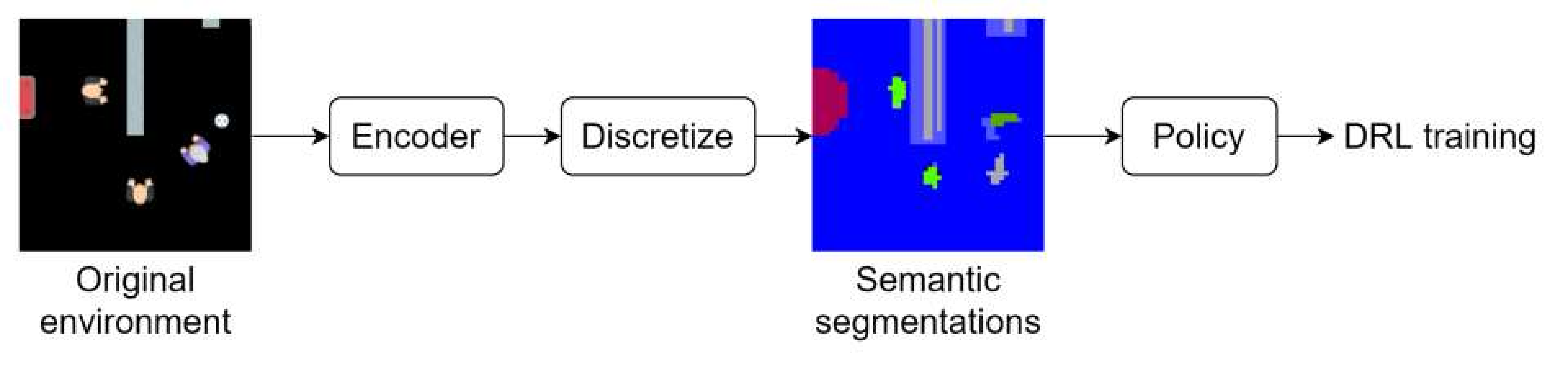
Idea: categorize objects

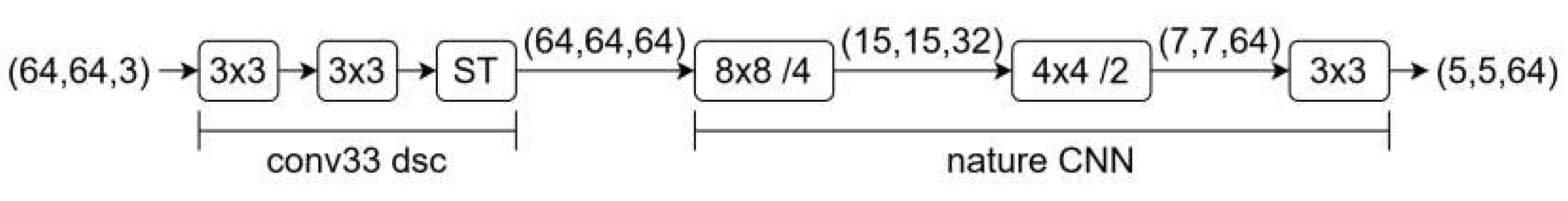
Sanity check

- Ground truth categories improve agent performance

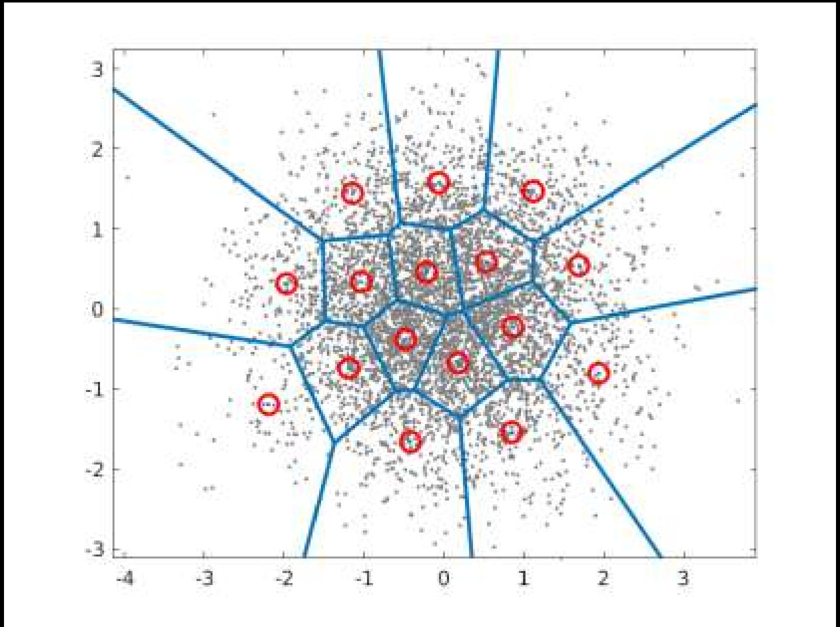


Architecture





Discretization



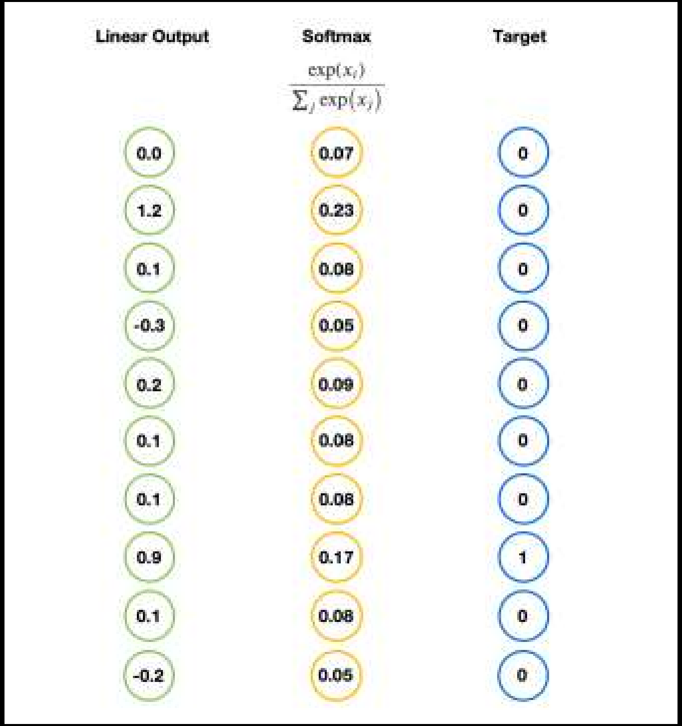
Vector quantization

Discretization

$$q_i = \frac{exp(z_i/T)}{\sum_j exp(z_j/T)}$$

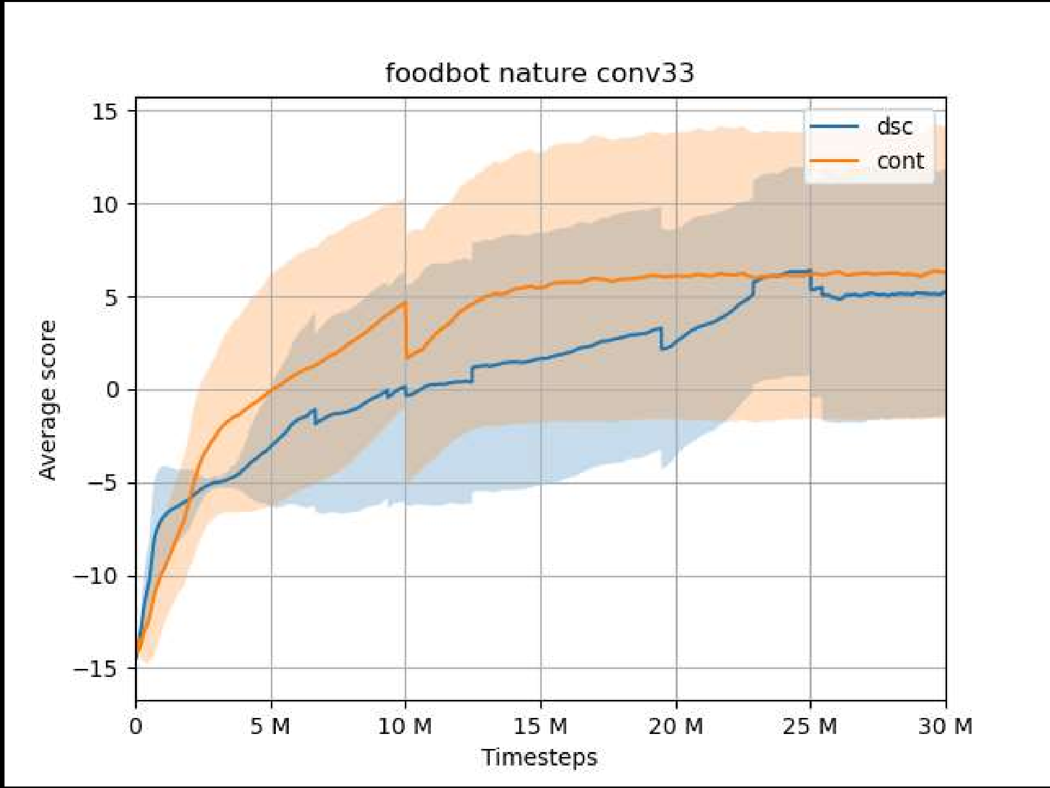
Softmax annealing

Discretization



Pass-through gradients

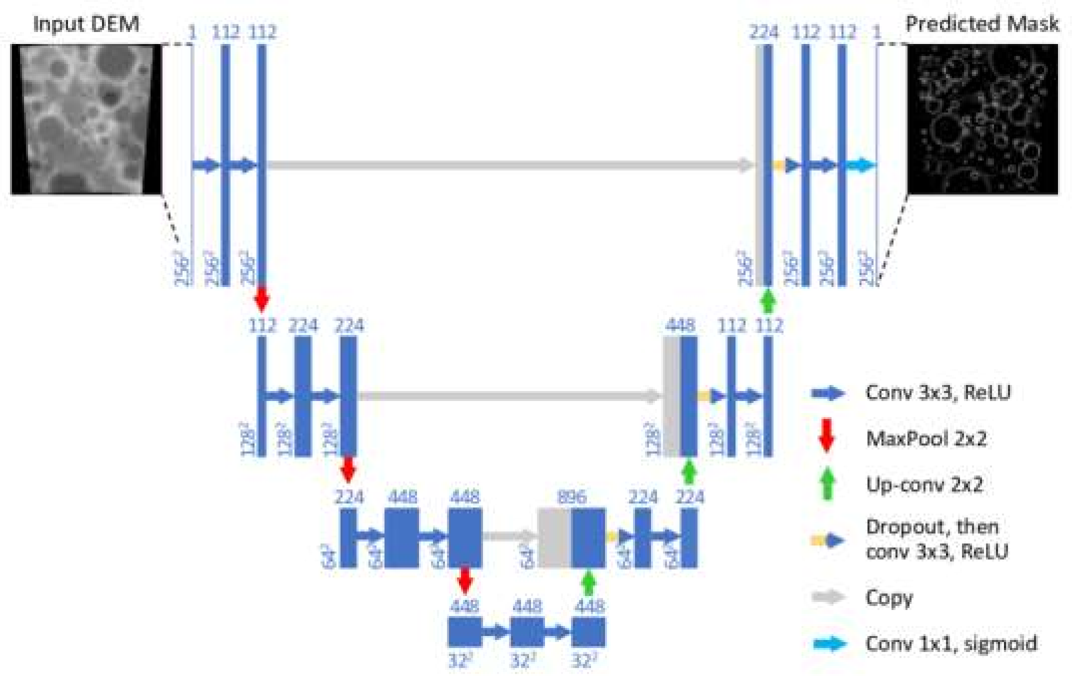
Encoder



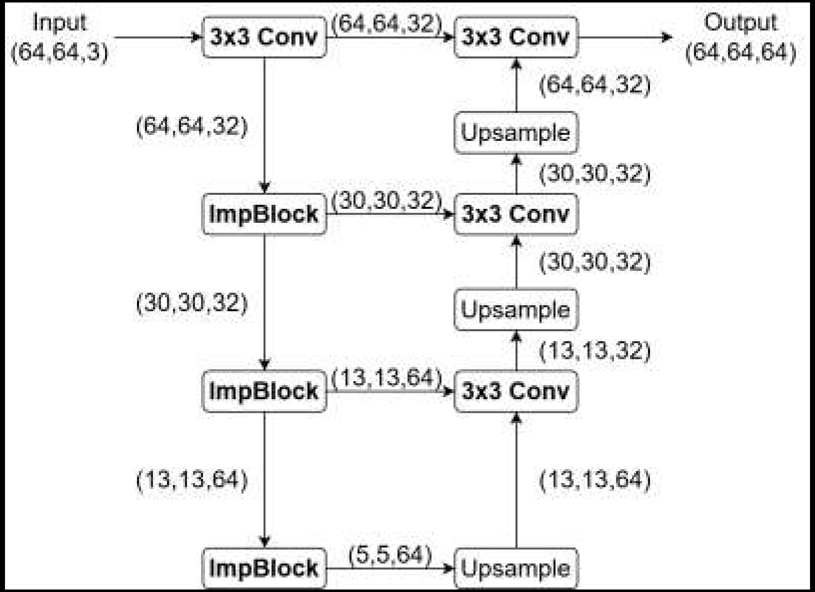
foodbot nature conv33
dsc
cont
15
10
5
0
−5
−10
−15
Average score
0
5 M
10 M
15 M
20 M
25 M
30 M
Timesteps

Encoder

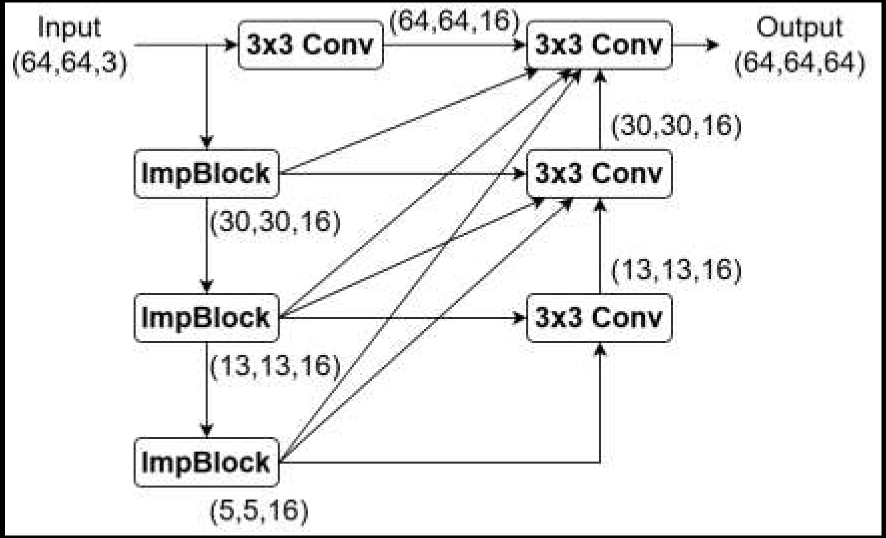
- U-Net paper



Encoder



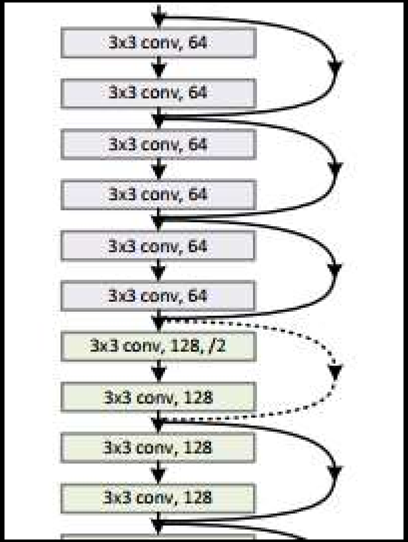
Input
(64,64,3)
3x3 Conv
(64,64,32)
3x3 Conv
Output
(64,64,64)
(64,64,32)
(64,64,32)
Upsample
(30,30,32)
ImpBlock
(30,30,32)
3x3 Conv
(30,30,32)
(30,30,32)
Upsample
(13,13,32)
ImpBlock
(13,13,64)
3x3 Conv
(13,13,64)
(13,13,64)
ImpBlock
(5,5,64)
Upsample



Input
(64,64,3)
3x3 Conv
(64,64,16)
3x3 Conv
Output
(64,64,64)
(30,30,16)
ImpBlock
3x3 Conv
(30,30,16)
(13,13,16)
ImpBlock
3x3 Conv
(13,13,16)
ImpBlock
(5,5,16)

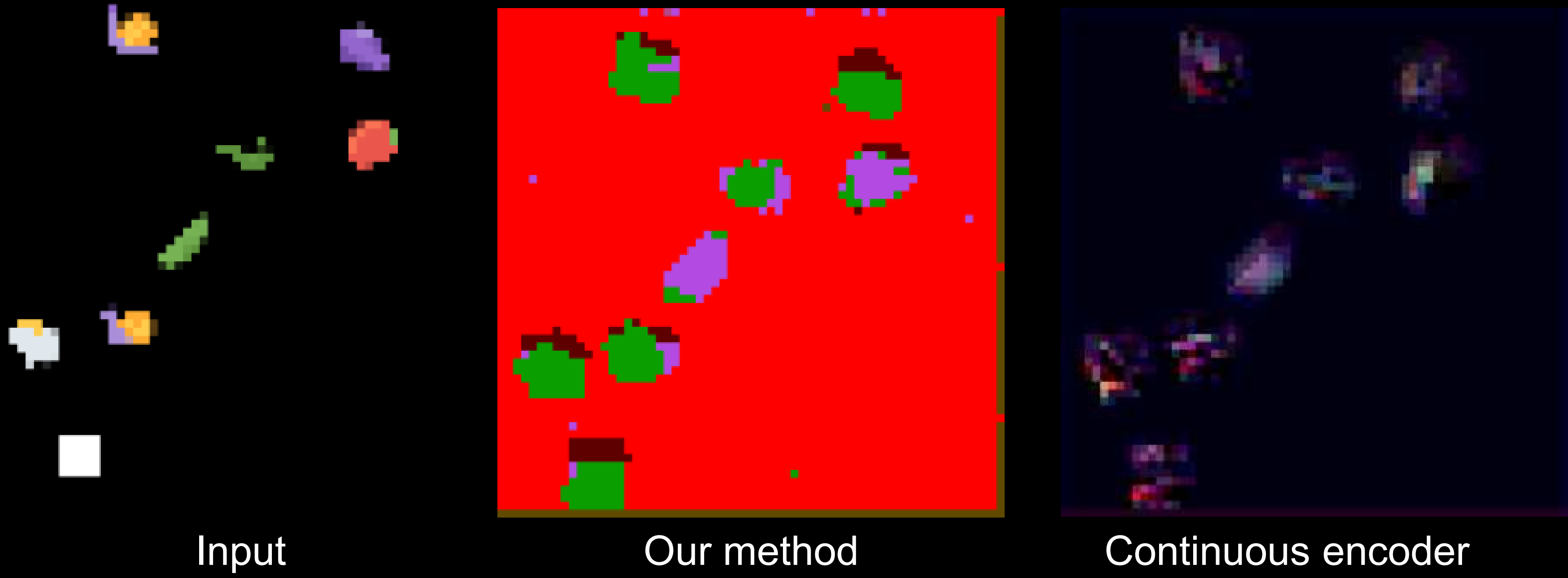
Encoder

- Resnet performs just as well

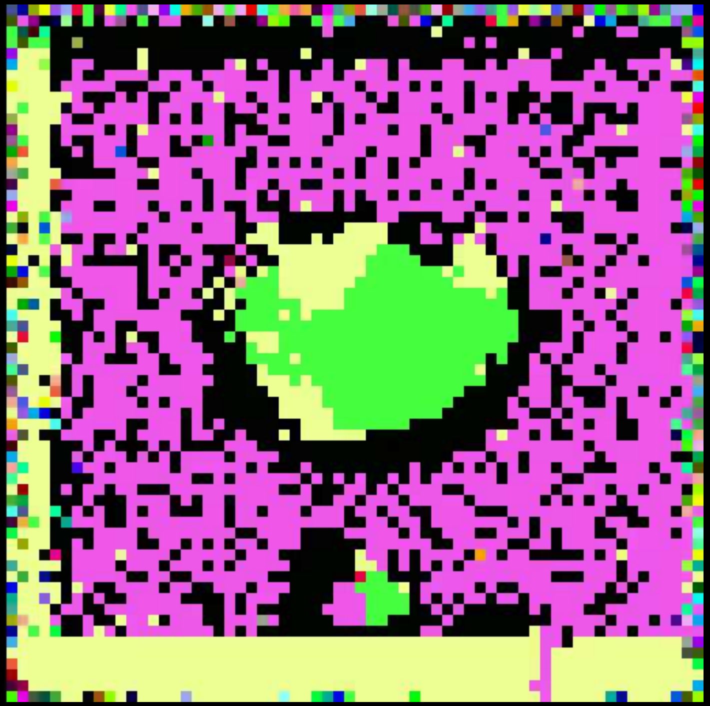


Results

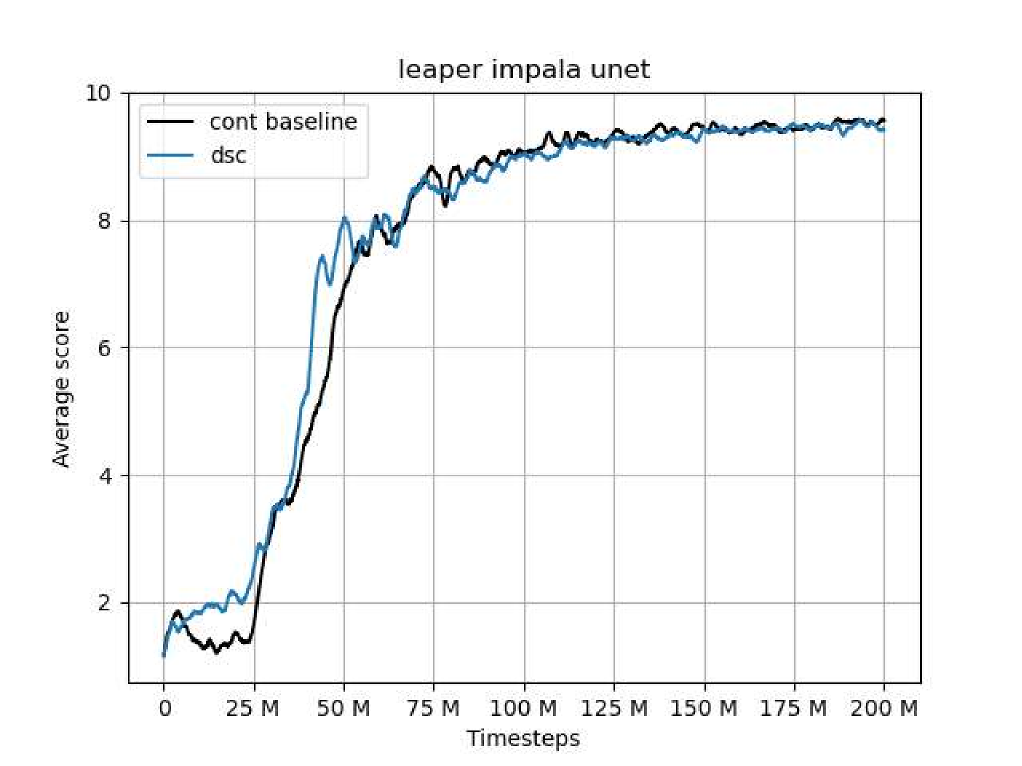
Latents



Latents



No significant impact on performance

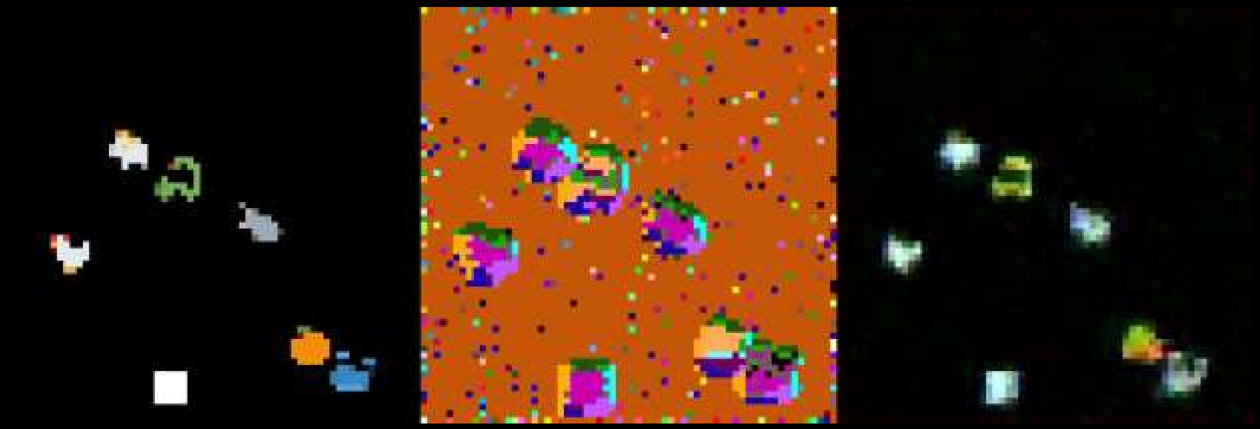


Ablation

- What if we use a reconstruction loss?

Ablation

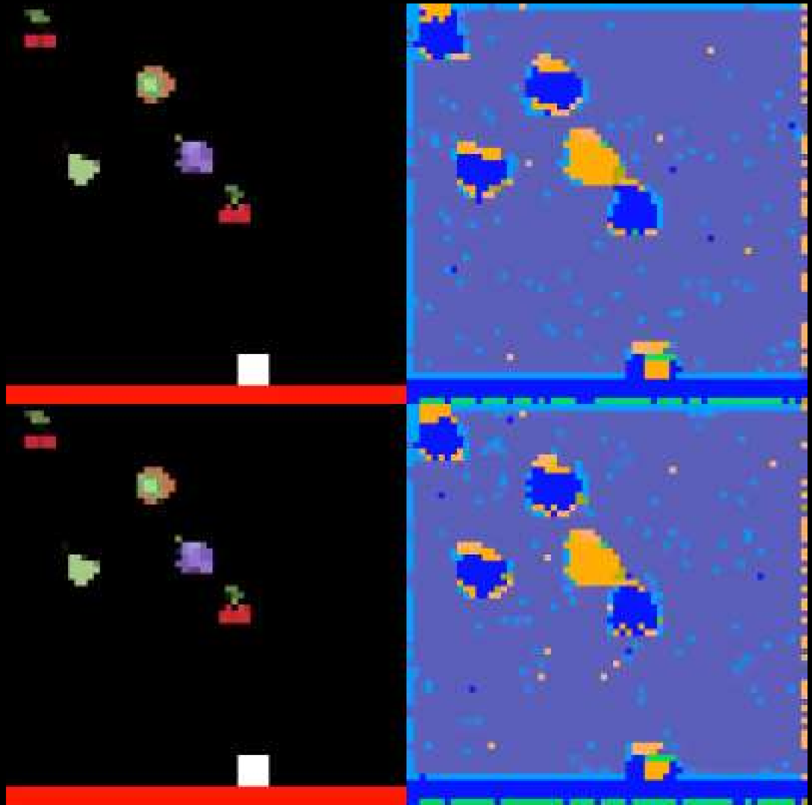
- What if we use a reconstruction loss?



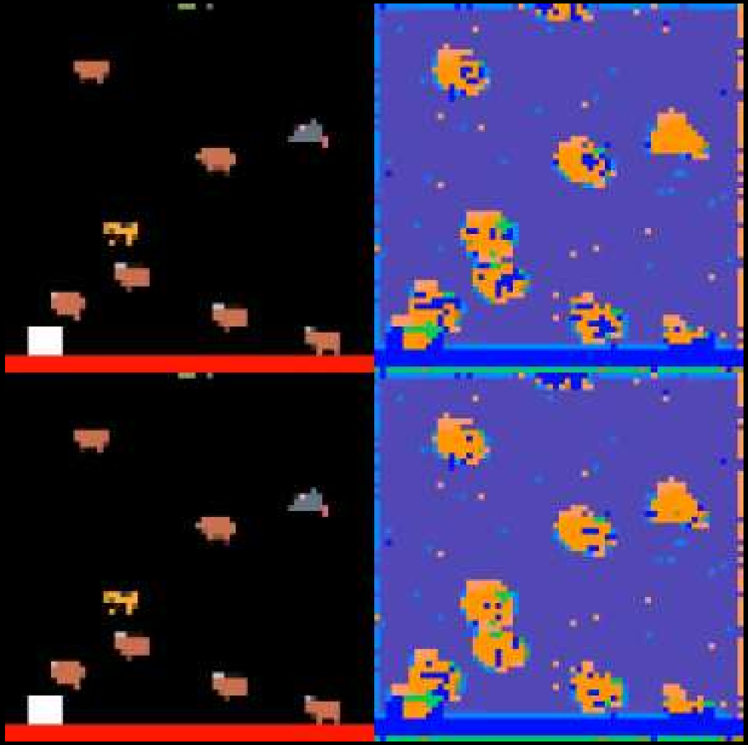
Correcting mistakes

Original

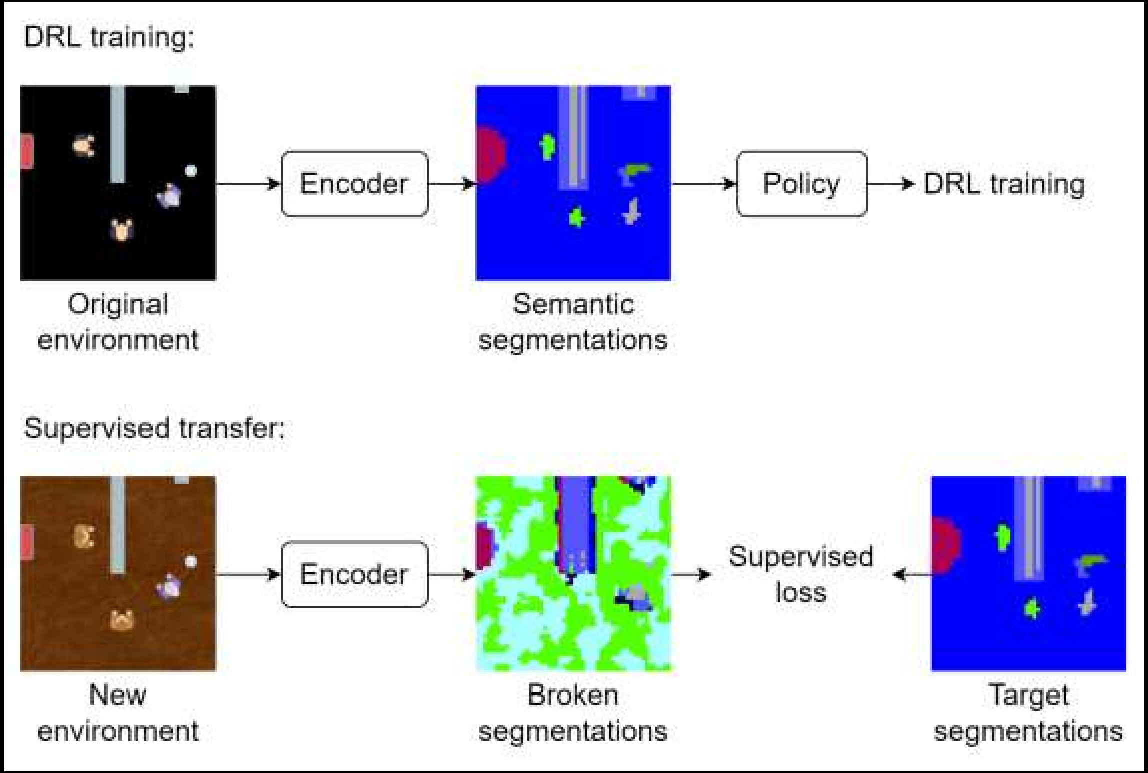
Grapes = good



Teaching new objects



Transferring textures



Transfer

Theme and background transfer

	PAD		BC		RIA					
n_steps	**131k**	**1M**	**8k**	**131k**	**0**	**16**	**32**	**64**	**256**	**1024**
leaper	-127	-127	74	164	-51	19	54	66	76	88
fruitbot	9	9	16	90	9	55	66	77	70	83
miner	-17	-15	42	37	45	50	49	50	79	90
dodgeball	-8	-8	5	22	-7	29	35	32	50	40
starpilot	-6	-6	4	70	6	21	28	33	43	55

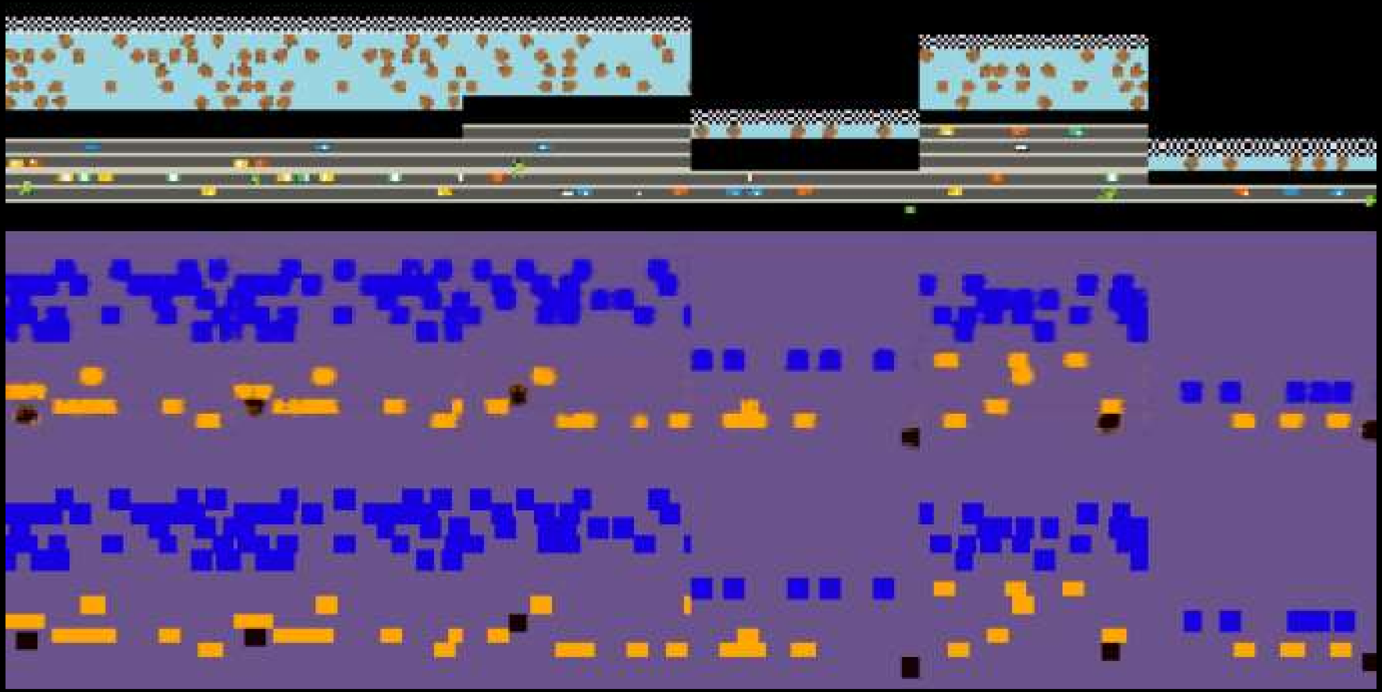
Table 1: Average normalized scores in test environments with different backgrounds and object textures.

Transferring games

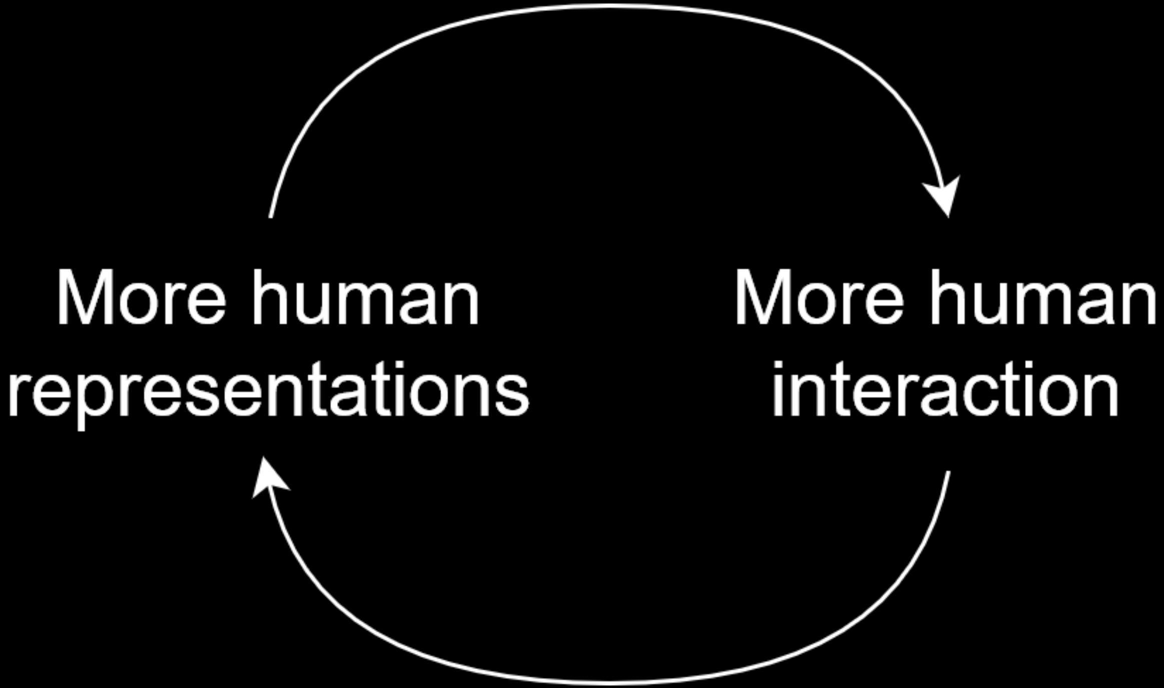
Input

Output

Target



Conclusion



Code: github.com/allenz/interpretable-rl