Deep Reinforcement Learning and Complex Environments

Raia Hadsell

DeepMind

End-to-end Deep Learning for robots?

slide from V. Vanhoucke

End-to-end Deep Learning for robots?

2010: Speech Recognition

Audio → **Deep Net** → Text

slide from V. Vanhoucke

End-to-end Deep Learning for robots?

2010: Speech Recognition

Audio → **Deep Net** → Text

2012: Computer Vision

Pixels → **Deep Net** → Labels

slide from V. Vanhoucke

End-to-end Deep Learning for robots?

2010: Speech Recognition

Audio → **Deep Net** → Text

2012: Computer Vision

Pixels → **Deep Net** → Labels

2014: Machine Translation

Text → **Deep Net** → Text

slide from V. Vanhoucke

End-to-end Deep Learning for robots?

2010: Speech Recognition

Audio → Deep Net Text

2012: Computer Vision

Pixels → Deep Net → Labels

2014: Machine Translation

Text → Deep Net → Text

2017: Robotics?

Sensors → Perception → World Model → Planning → Control → Action

slide from V. Vanhoucke

Robotics is different

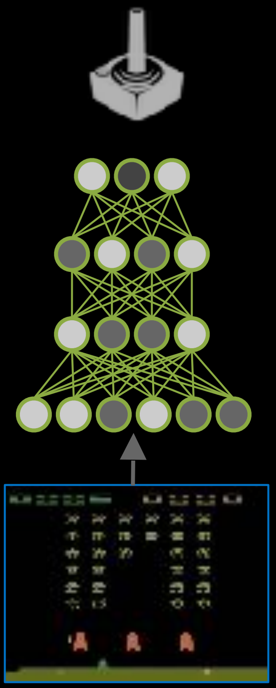
LABELS

Robotics is different

SENSORS $\rightleftarrows$ ACTIONS

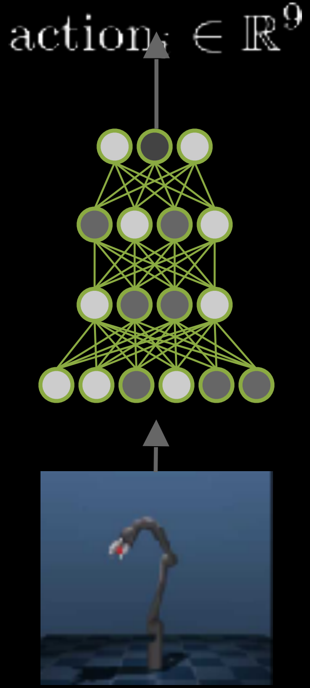
Deep Reinforcement Learning

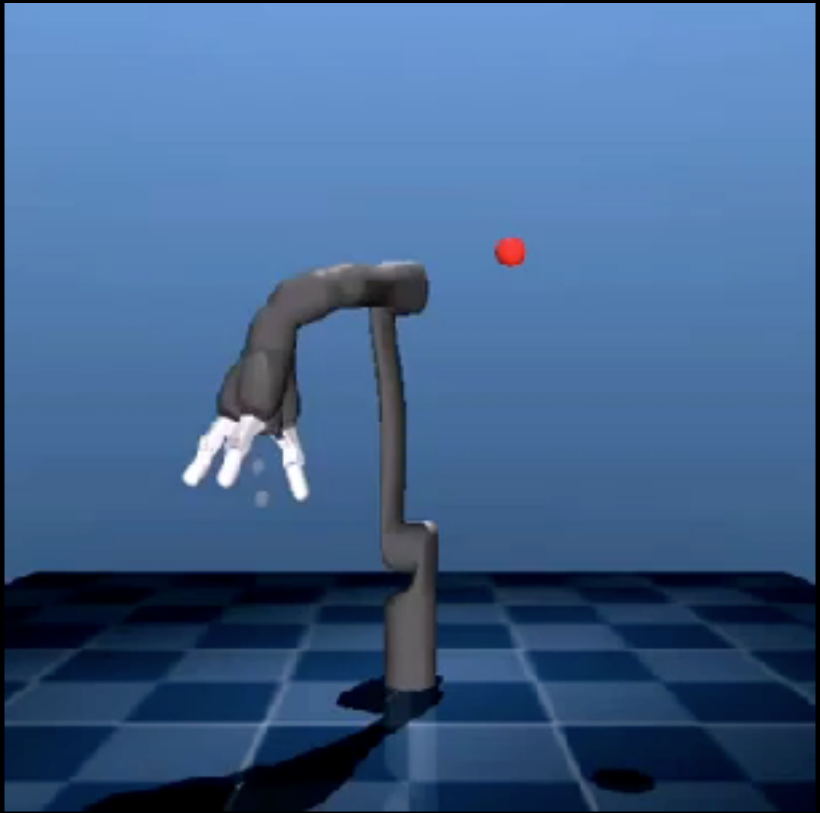
General Atari Player



[Mnih et al, *Playing Atari with Deep Reinforcement Learning*, 2014]

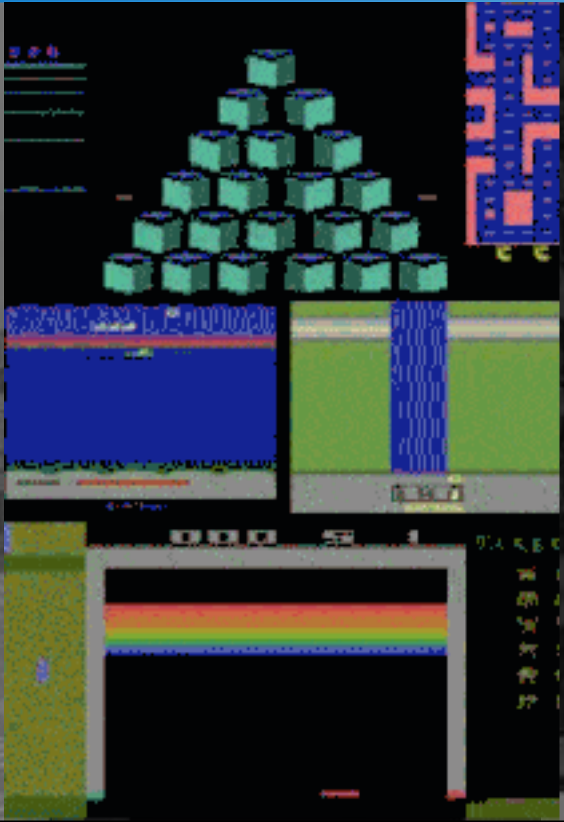
9DOF Random reacher





- Can deep RL agents learn multiple tasks?
- Can deep RL agents learn efficiently?
- Can deep RL agents learn from real data?
- Can deep RL agents learn continuous control?

Lab Mazes



Multiple Tasks
&
Lifelong learning

StreetLearn

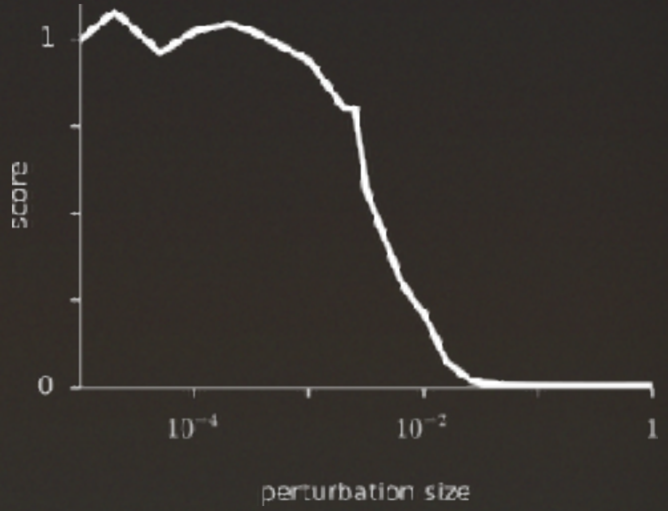
Parkour

Lifelong Learning - 3 challenges

1. Catastrophic forgetting
2. Positive transfer
3. Specialization and generalization

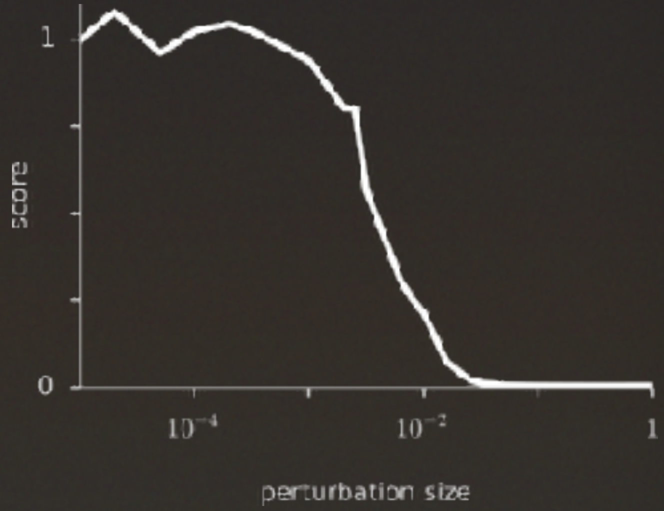
Catastrophic forgetting

- Well-known phenomenon
- Especially severe in Deep RL

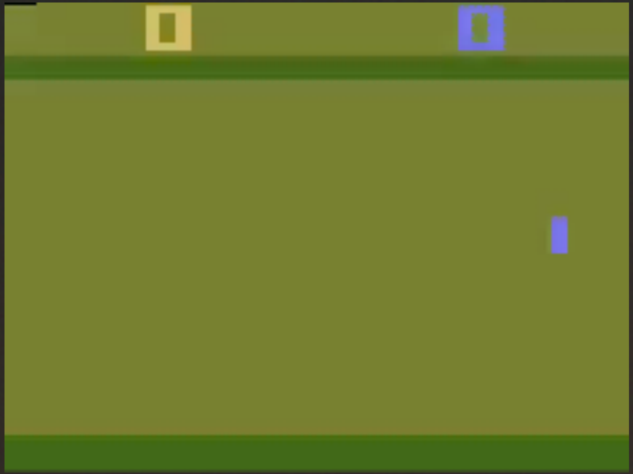


Catastrophic forgetting

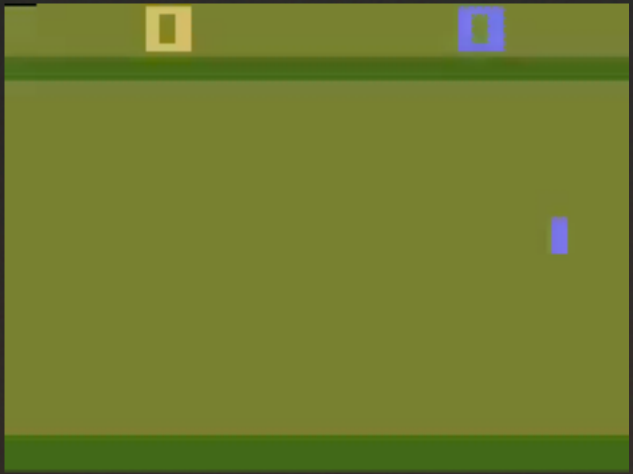
- Well-known phenomenon
- Especially severe in Deep RL

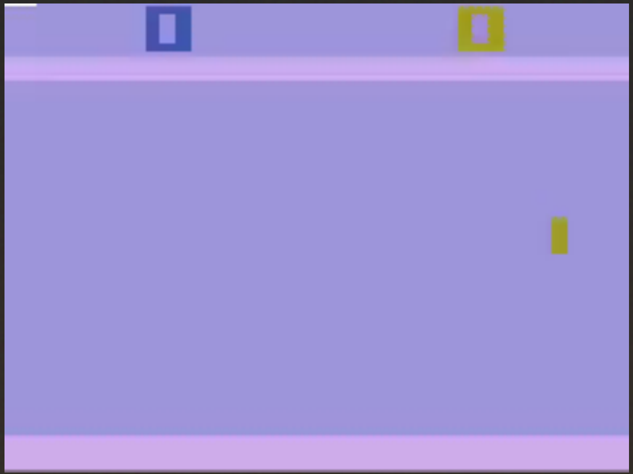


Catastrophic forgetting



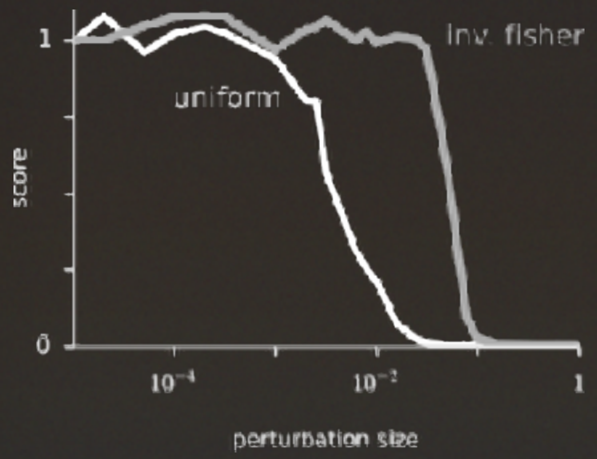
Catastrophic forgetting

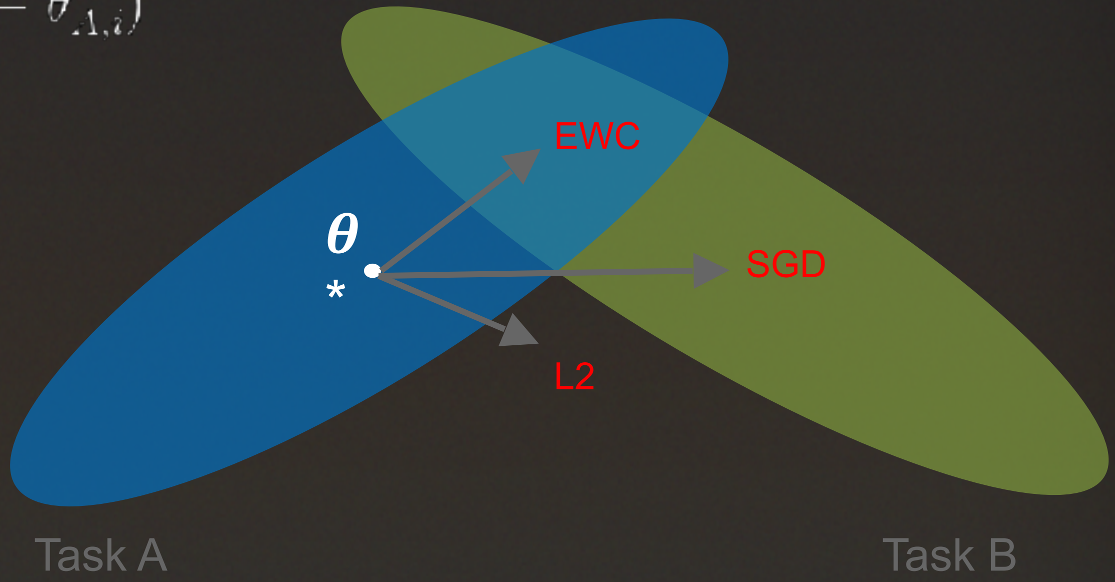




Elastic Weight Consolidation

$$\mathcal{L}(\theta) = \mathcal{L}_B(\theta) + \sum_i \frac{\lambda}{2} F_i (\theta_i - \theta^*_{A,i})^2$$





James Kirkpatrick et al (2017), “Overcoming Catastrophic Forgetting in NNs”

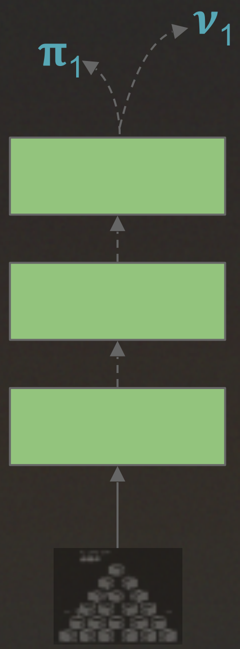
What if my tasks really don’t get along?

What if my tasks really don't get along?

Progressive Nets

- add columns for new tasks
- freeze params of learnt columns
- layer-wise neural connections

→ capacity for task-specific features
→ enables deep compositionality
→ precludes forgetting

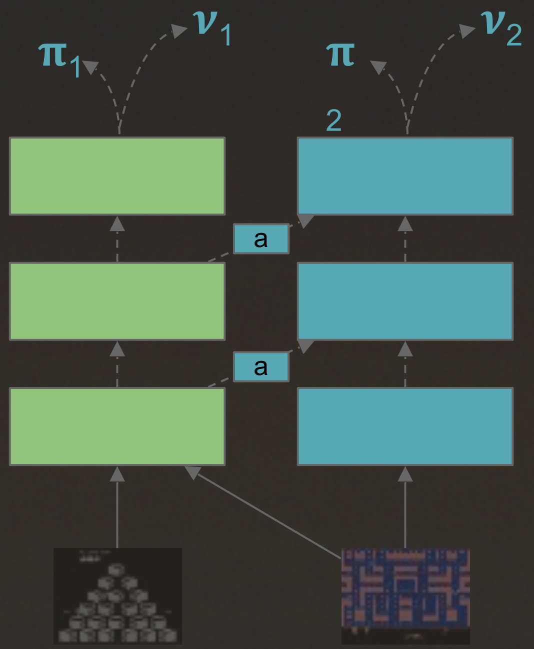


What if my tasks really don't get along?

Progressive Nets

- add columns for new tasks
- freeze params of learnt columns
- layer-wise neural connections

→ capacity for task-specific features
→ enables deep compositionality
→ precludes forgetting



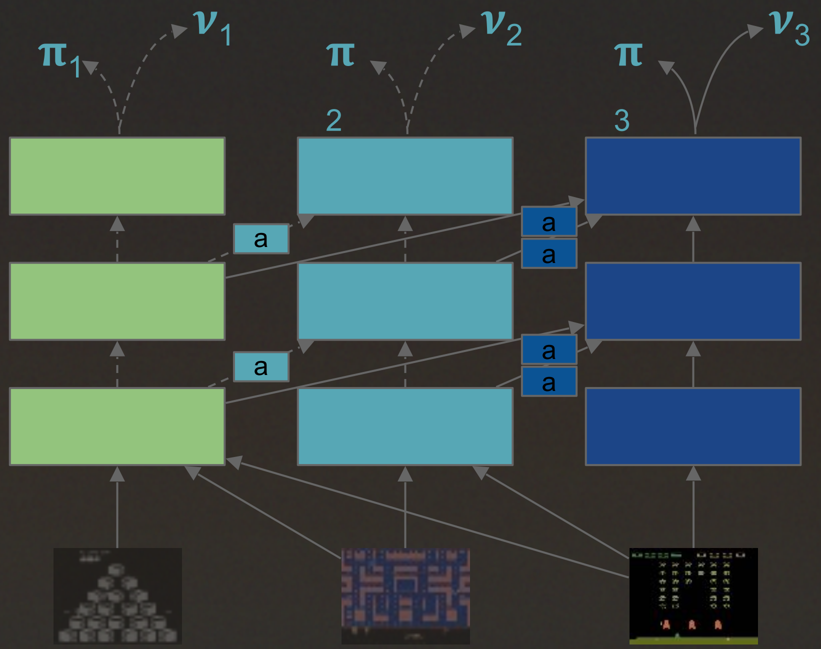
Andrei Rusu et al (2016), "Progressive Neural Networks"

What if my tasks really don’t get along?

Progressive Nets

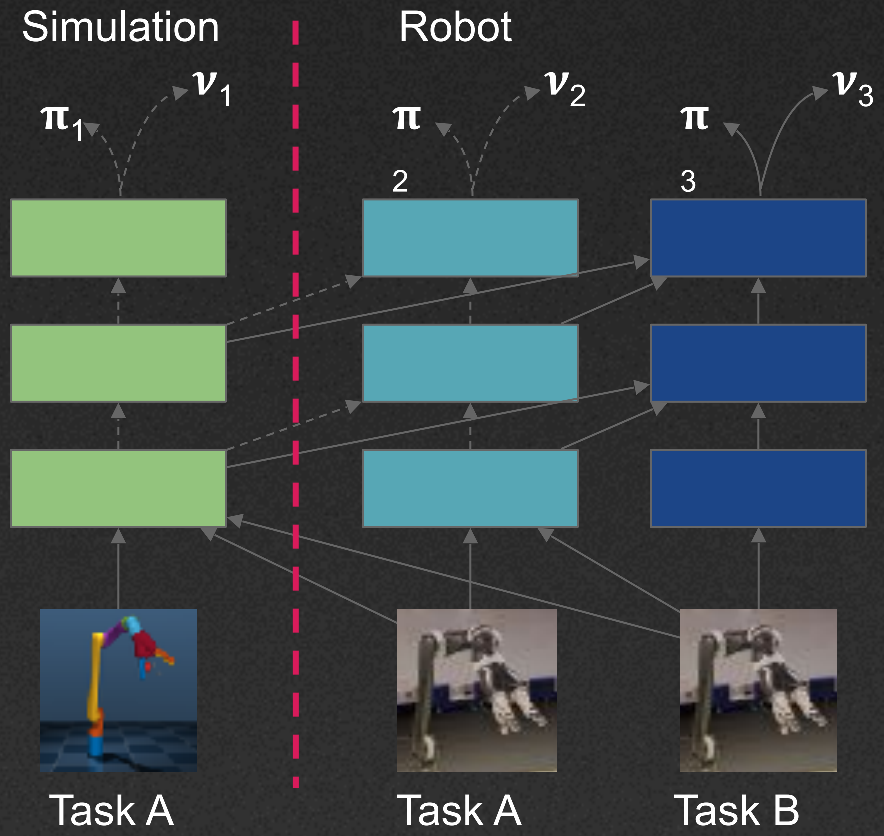
- add columns for new tasks
- freeze params of learnt columns
- layer-wise neural connections

→ capacity for task-specific features
→ enables deep compositionality
→ precludes forgetting



Andrei Rusu et al (2016), “Progressive Neural Networks”

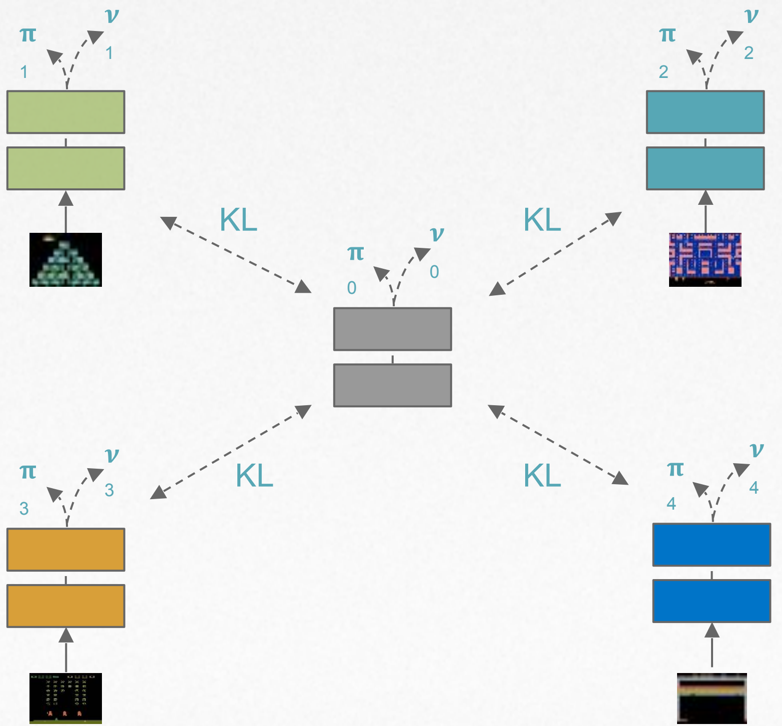
Sim-to-Real



What if my tasks really don't get along?

Distral (Distill and Transfer Learning)

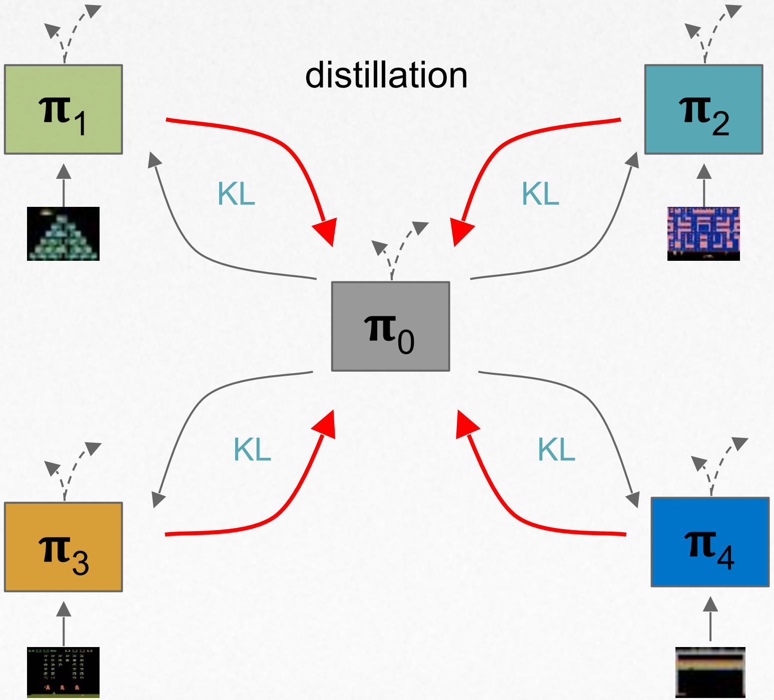
- Task-specific networks plus shared network
- KL Divergence constraint
- Regularisation in policy space rather than parameter space
- Shared policy as a communication channel between tasks



Distral (Distill and Transfer Learning)

- Task-specific networks plus shared network
- Regularisation in policy space rather than parameter space
- Shared policy as a communication channel between tasks

→ *Distillation* of knowledge into shared model enables *transfer* to tasks

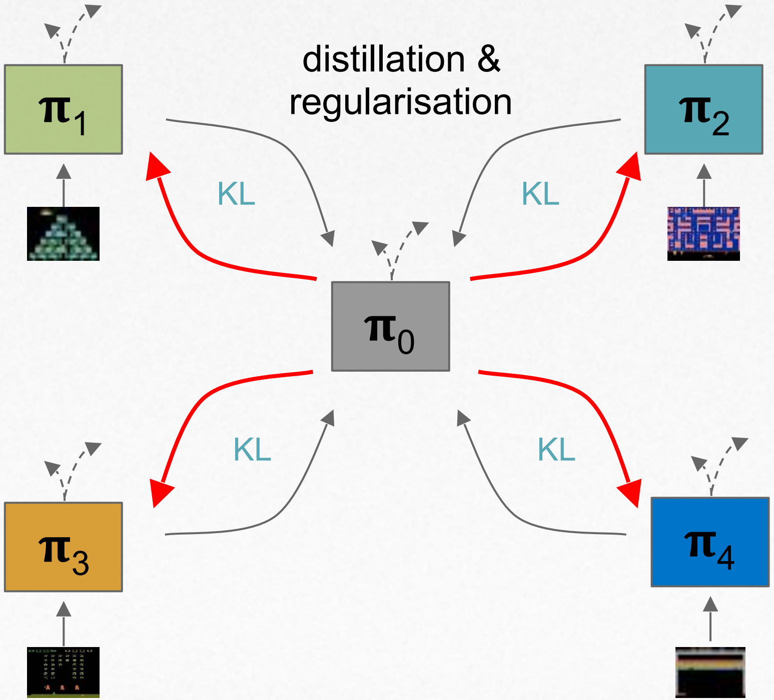


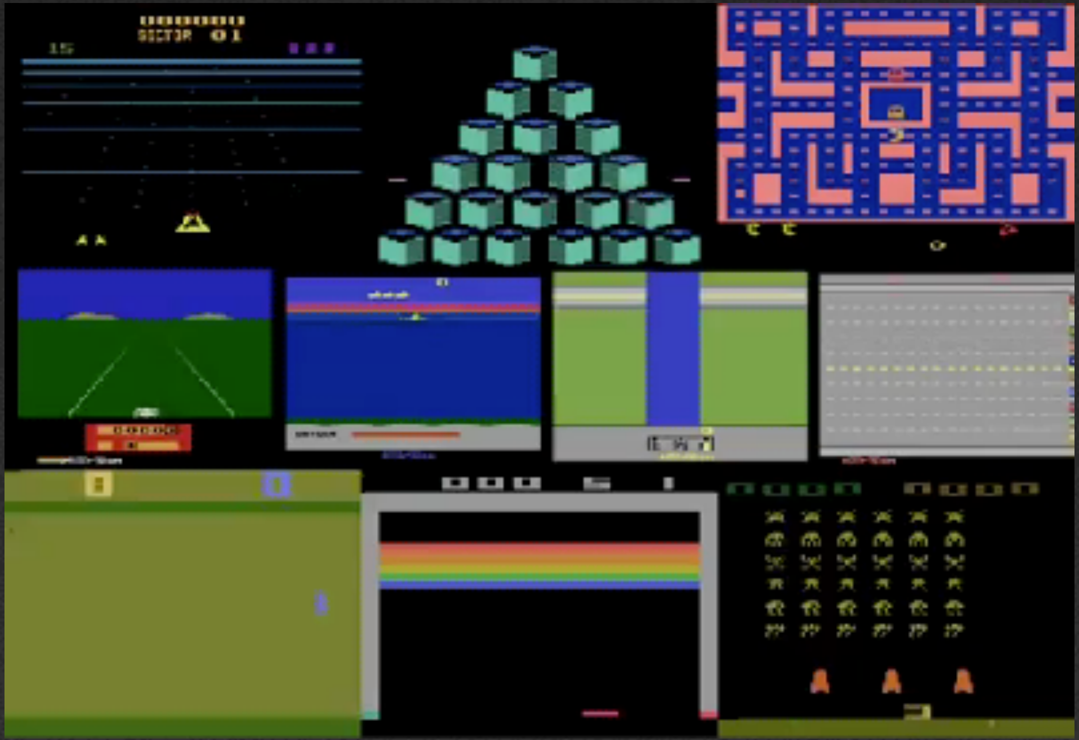
Distral (Distill and Transfer Learning)

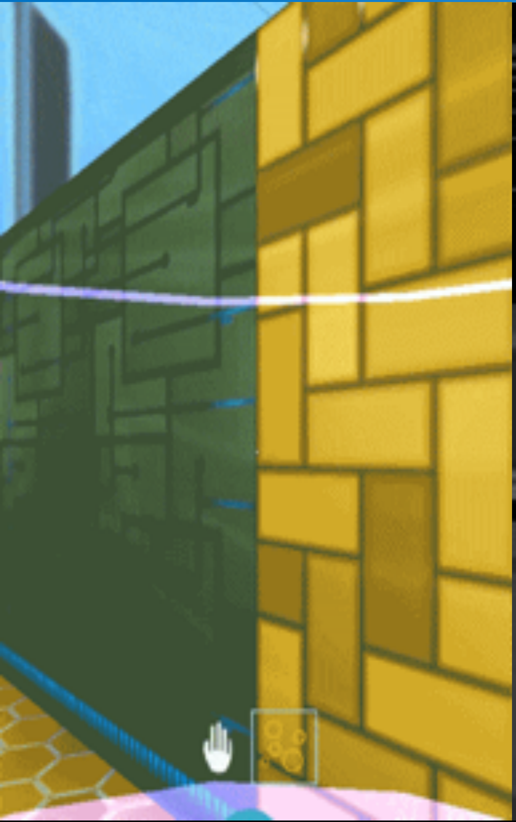
- Task-specific networks plus shared network
- Regularisation in policy space rather than parameter space
- Shared policy as a communication channel between tasks

→ *Distillation* of knowledge into shared model enables *transfer* to tasks

→ *Regularisation* of shared model gives stability and robustness







Lab Mazes
&
Auxiliary Learning

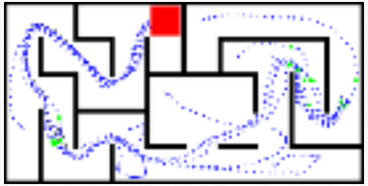
Multiple Tasks
&
Lifelong learning

StreetLearn

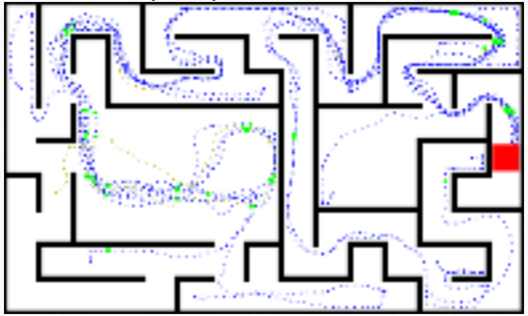
Parkour

Navigation mazes

3600 steps/episode



10800 steps/episode



Game **episode:**

1. Random start
2. Find the goal (+10)
3. Teleport randomly
4. Re-find the goal (+10)
5. Repeat (limited time)

Variants:

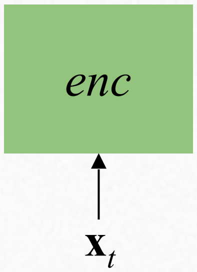
Static maze, static goal

Static maze, random goal

Random maze

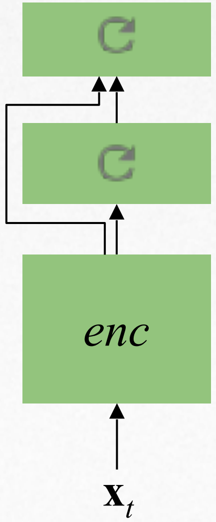
Nav agent architecture

1. Convolutional encoder and RGB inputs



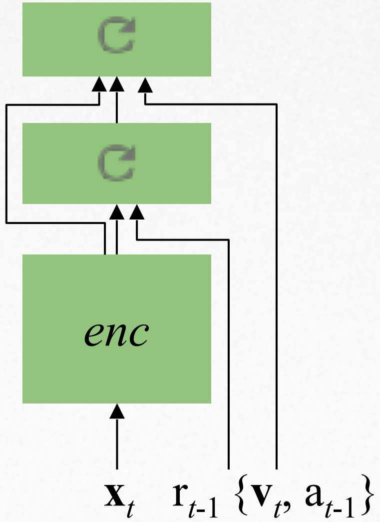
Nav agent architecture

1. Convolutional encoder and RGB inputs
2. Single or stacked LSTM with skip connection



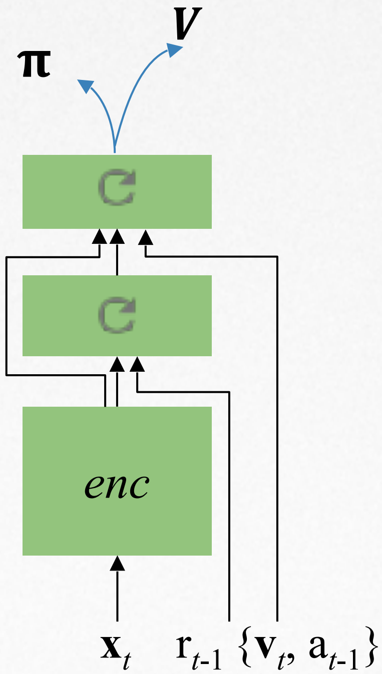
Nav agent architecture

1. Convolutional encoder and RGB inputs
2. Stacked LSTM
3. Additional inputs (reward, action, and velocity)



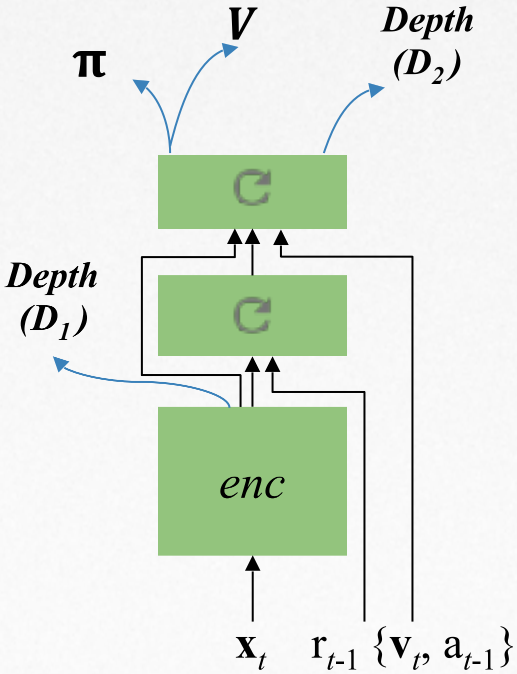
Nav agent architecture

1. Convolutional encoder and RGB inputs
2. Stacked LSTM
3. Additional inputs (reward, action, and velocity)
4. RL: Asynchronous advantage actor critic (A3C)



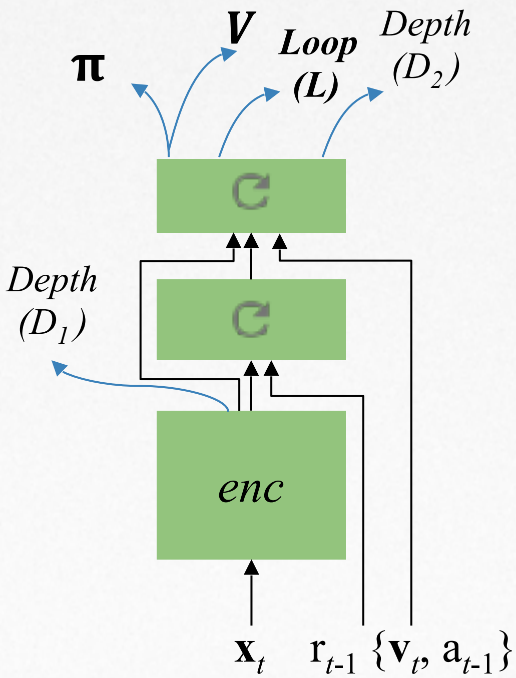
Nav agent architecture

1. Convolutional encoder and RGB inputs
2. Stacked LSTM
3. Additional inputs (reward, action, and velocity)
4. RL: Asynchronous advantage actor critic (A3C)
5. Aux task 1: Depth predictors

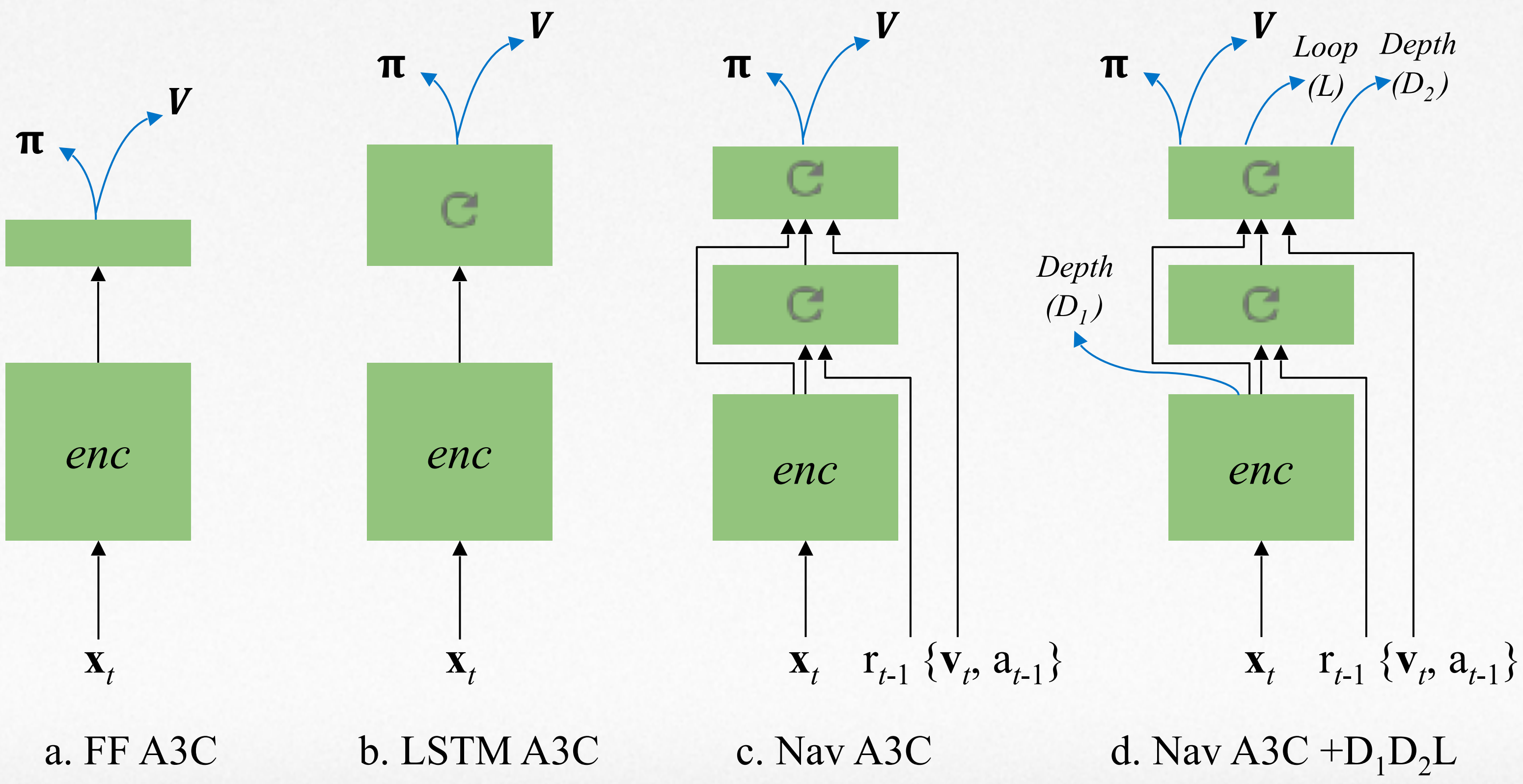


Nav agent architecture

1. Convolutional encoder and RGB inputs
2. Stacked LSTM
3. Additional inputs (reward, action, and velocity)
4. RL: Asynchronous advantage actor critic (A3C)
5. Aux task 1: Depth predictor
6. Aux task 2: Loop closure predictor



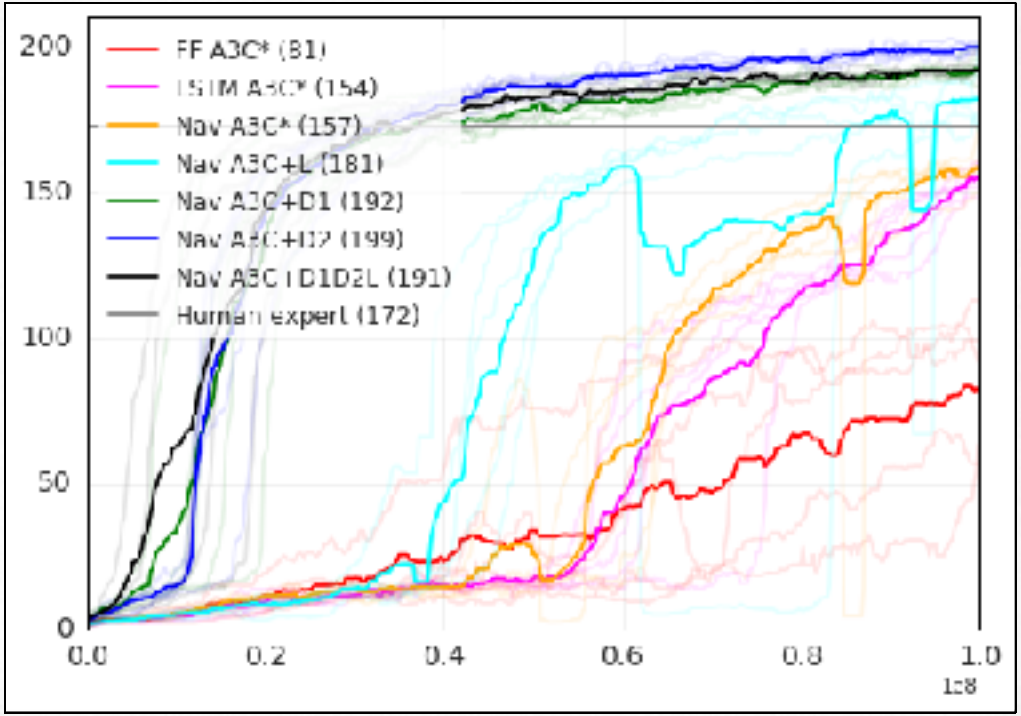
Variations in architecture

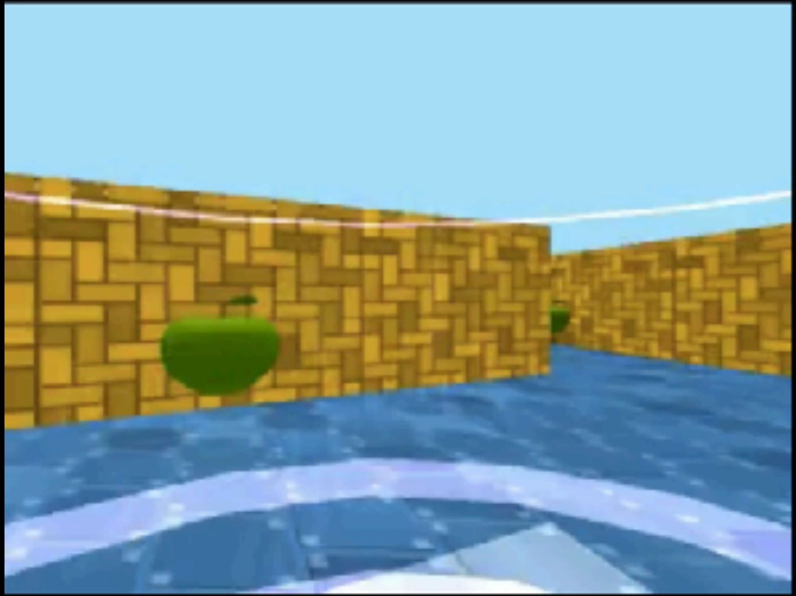


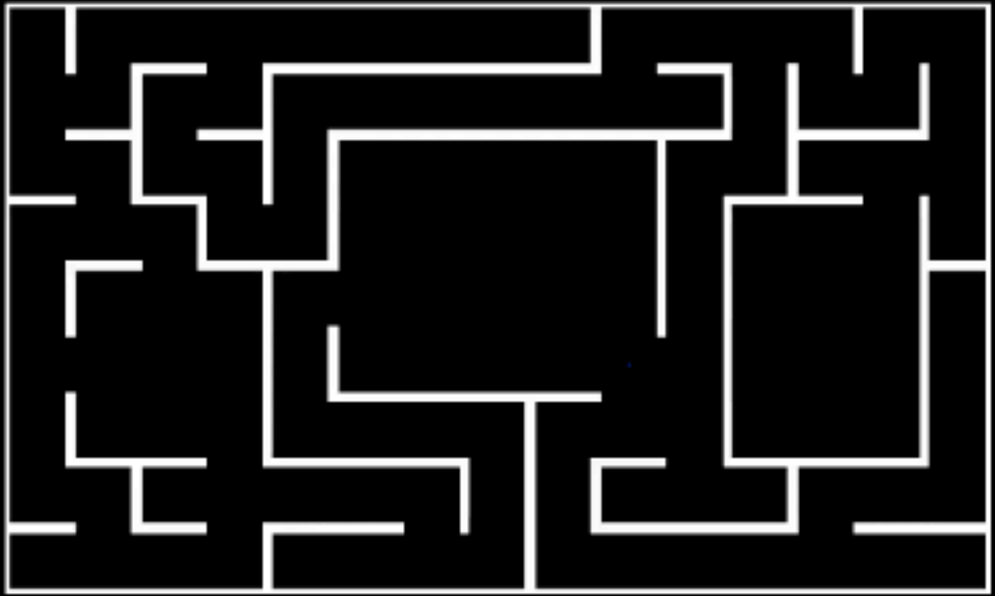
Results on large maze with static goal

+10

+1







Lab Mazes
&
Auxiliary Learning

Multiple Tasks
&
Lifelong learning

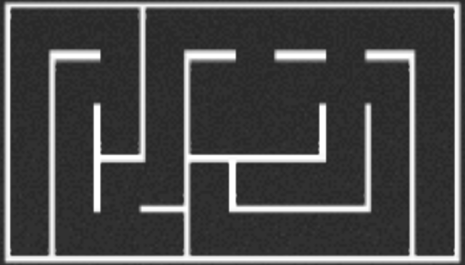
StreetLearn
&
Real woRld RL

Parkour

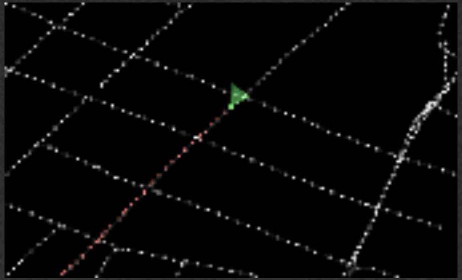
Navigation mazes in the real world?

observation

observation

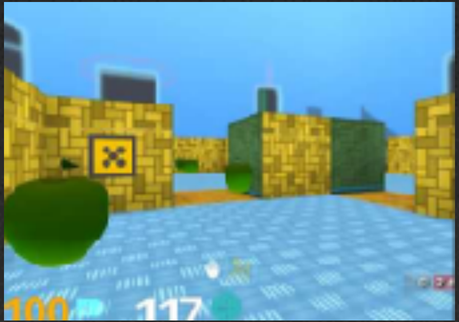


structure



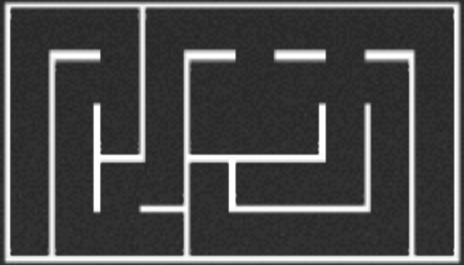
structure

StreetView as an RL environment: StreetLearn

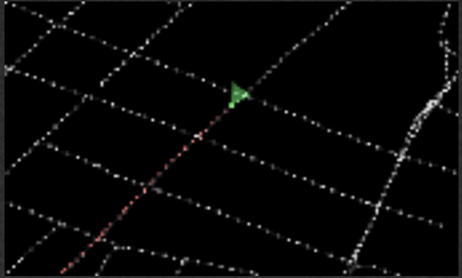


observation

observation



structure

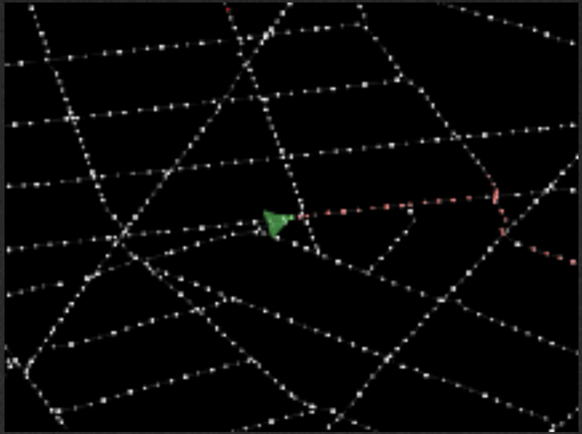


structure

- RGB image cropped from panorama (84x84)
- Goal location

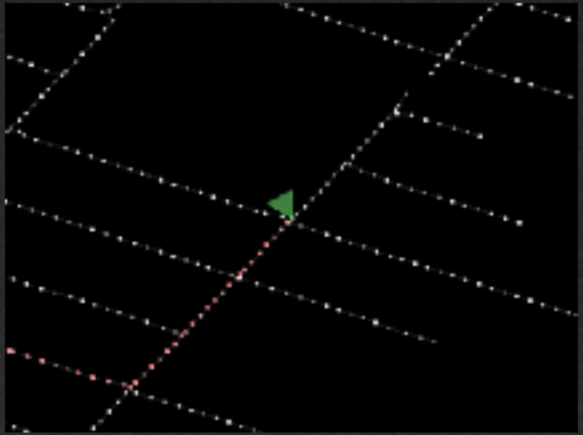
Actions: move to next node, rotate view 20° or 60°

StreetView as an RL environment: StreetLearn



left or right?

StreetView as an RL environment: StreetLearn

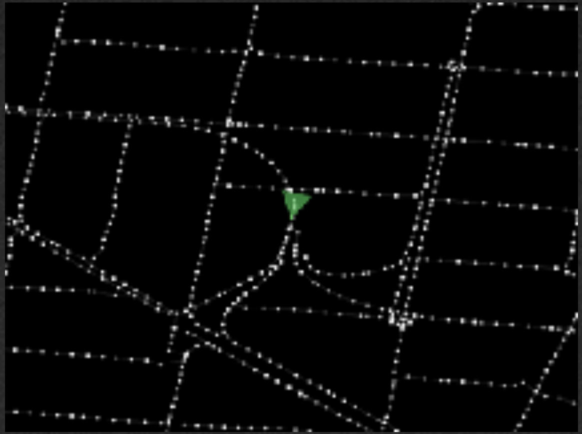


Looks like a road, but it's a park entrance

StreetView as an RL environment: StreetLearn

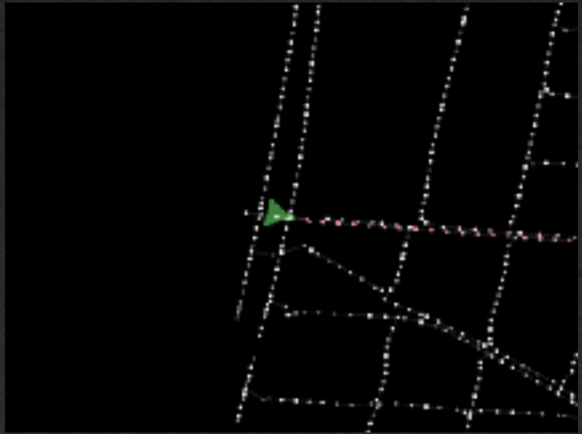
west side highway

StreetView as an RL environment: StreetLearn



curved roads and tunnels

StreetView as an RL environment: StreetLearn



really, tunnels!

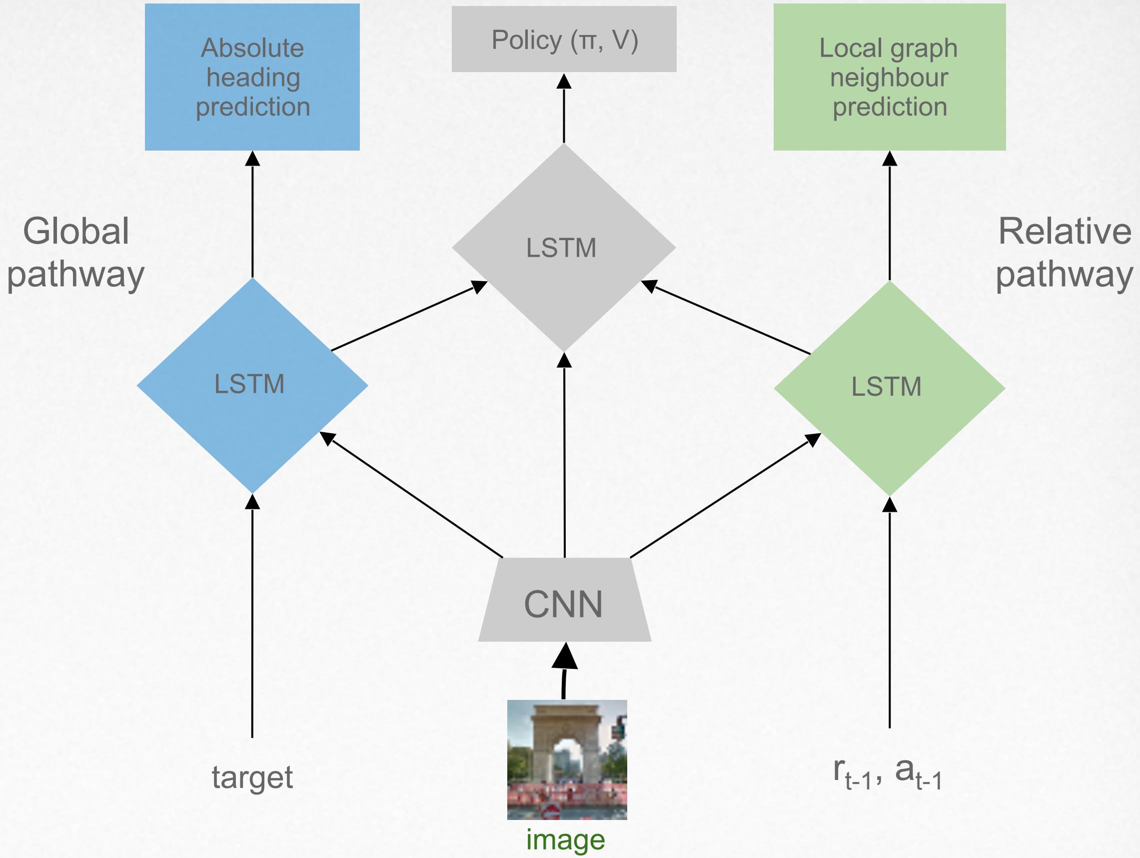
StreetLearn: The Courier Task

1. Spawn randomly and navigate to a random target location.

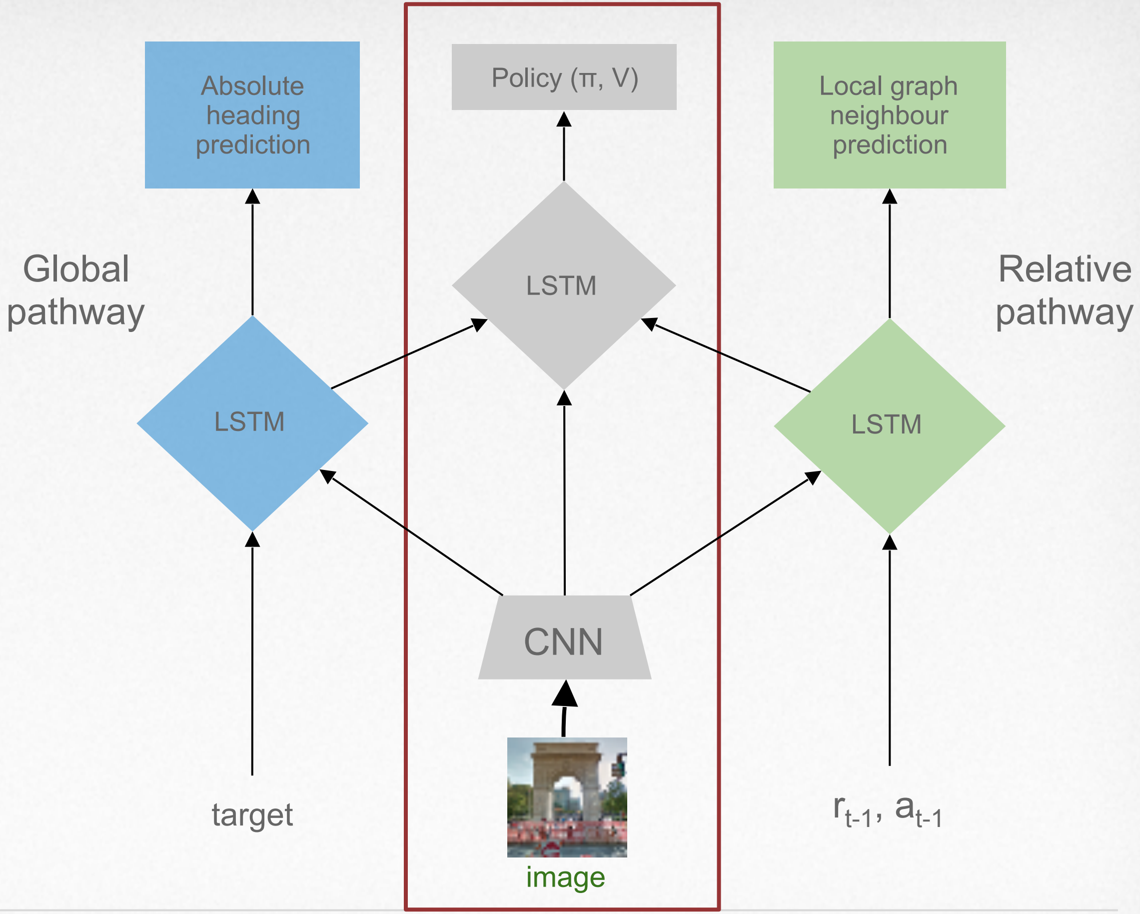
2. Start receiving reward when close to target (within 400m).

3. If target is reached (100m), navigate to a new random target.

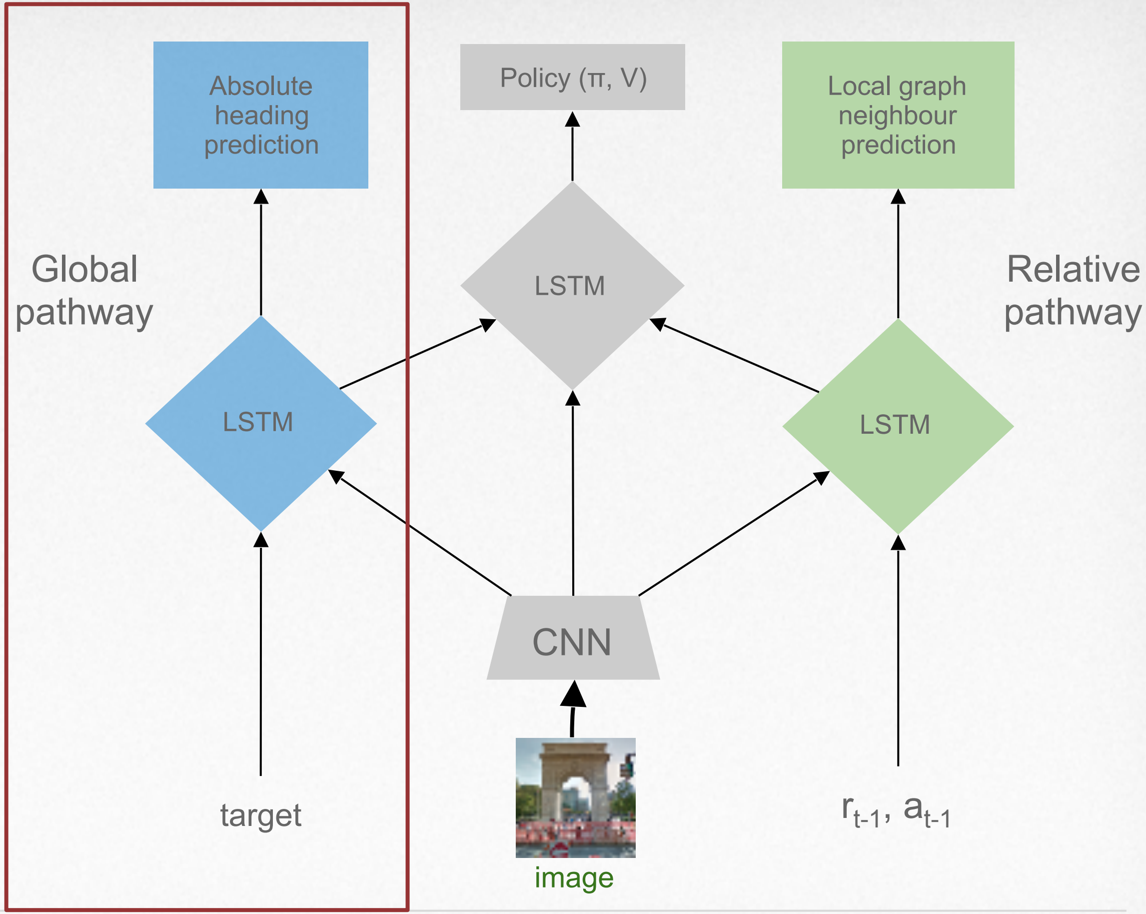
Agent architecture



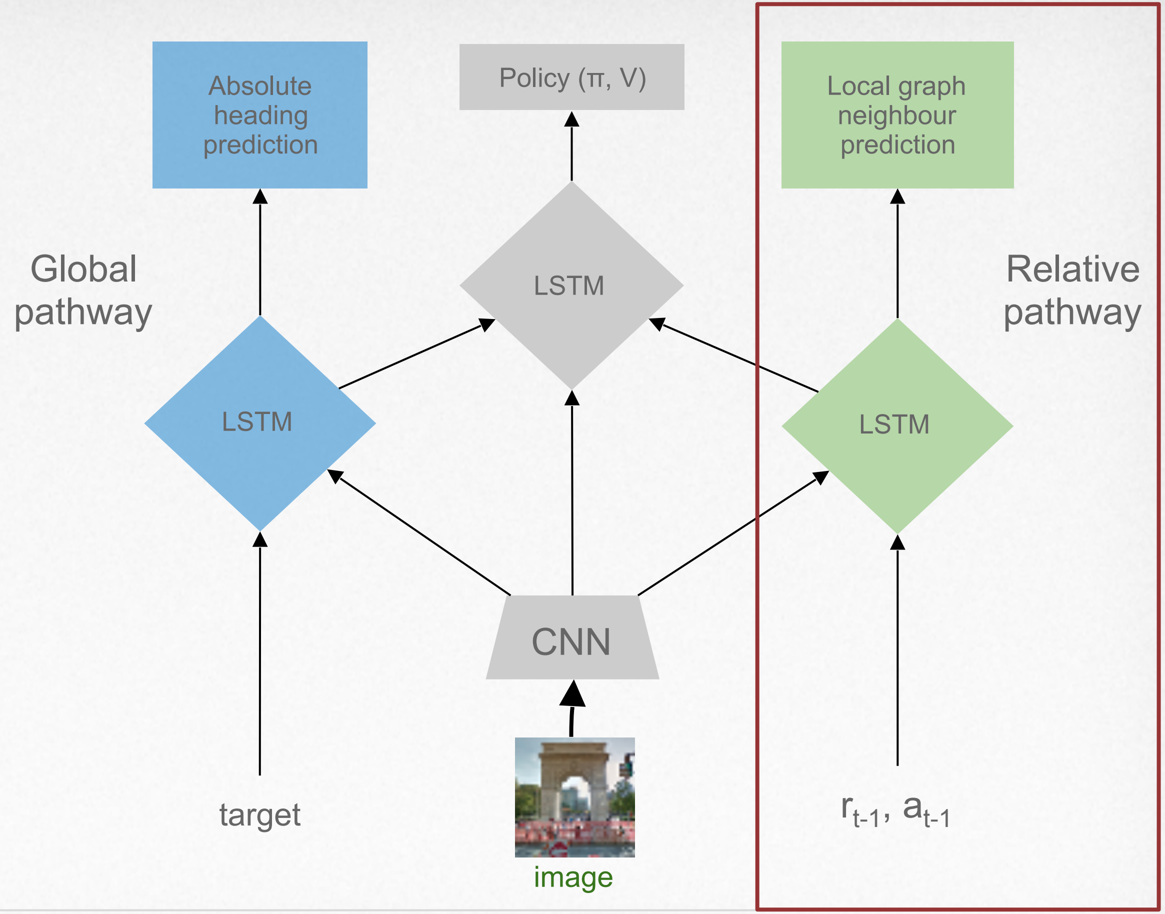
Agent architecture



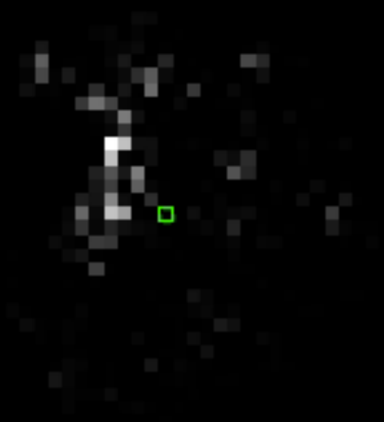
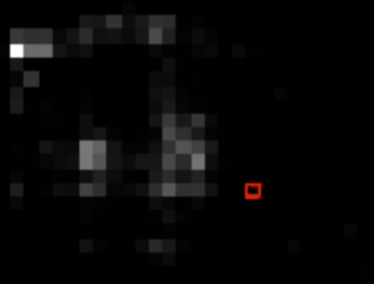
Agent architecture



Agent architecture





Lab Mazes
&
Auxiliary Learning

Multiple Tasks
&
Lifelong learning

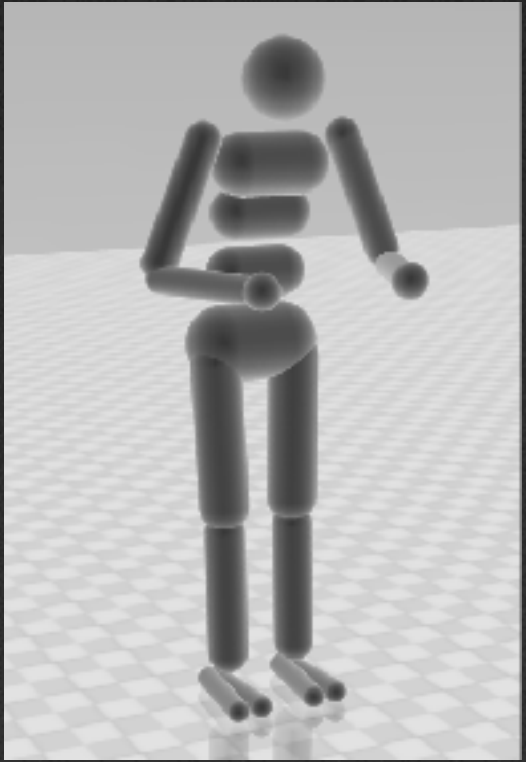
StreetLearn
&
Real woRld RL

Parkour
&
Continuous control

Proprioceptive and exteroceptive observations

Proprioceptive --
"near the body":

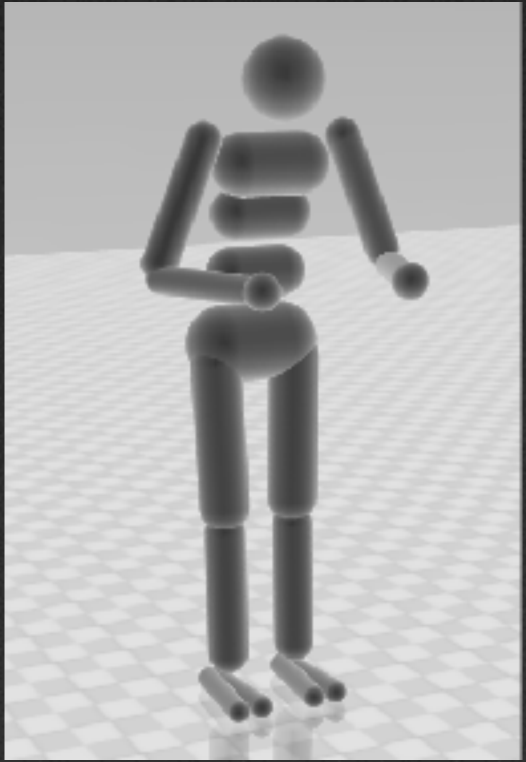
- Joint angles & velocities
- Touch sensors
- Positions and velocities of limbs in body coordinate frame

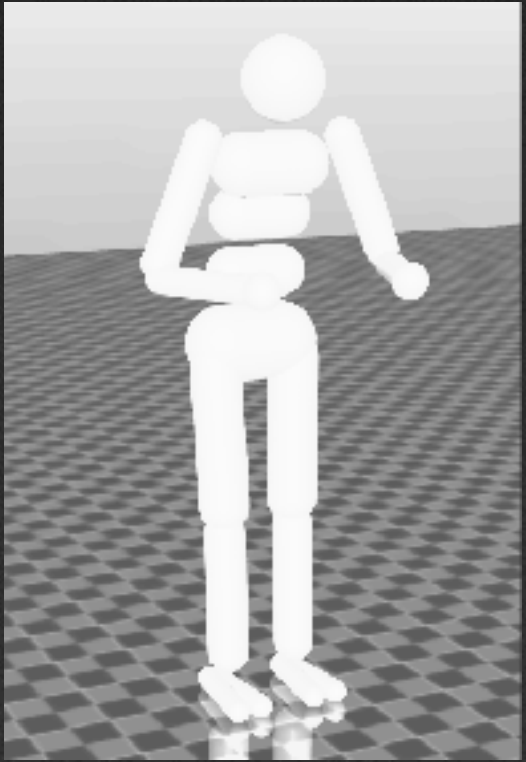


Proprioceptive and exteroceptive observations

Proprioceptive --
"near the body":

- Joint angles & velocities
- Touch sensors
- Positions and velocities of limbs in body coordinate frame

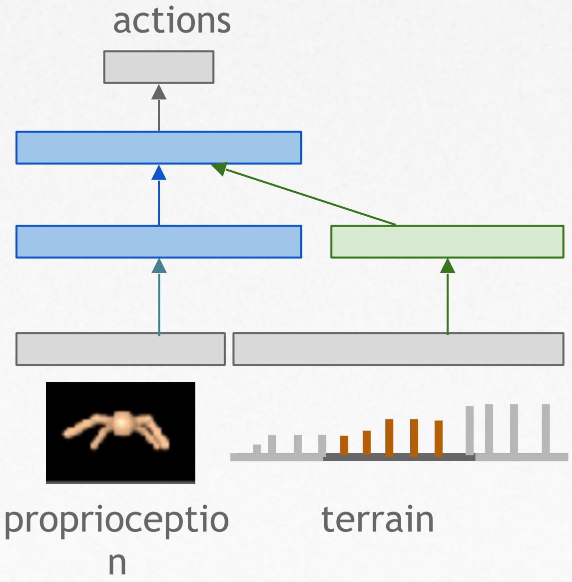




Exteroceptive --
"away from the body":

- Position / velocity in global coordinate frame
- Task-related (e.g. goal position)
- Vision

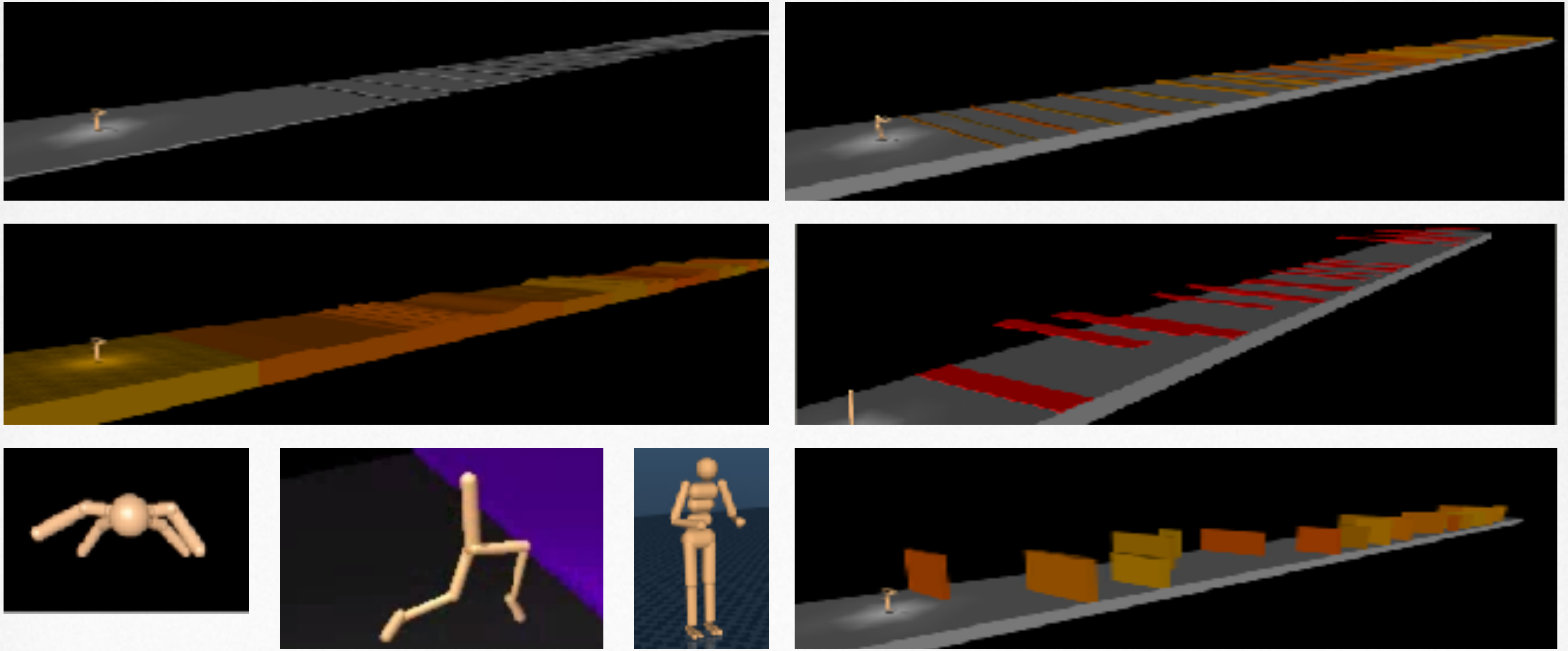
Rich environments for skill discovery: setup



Training

- Proximal policy optimization
  [Schulman et al.]
- Batched policy gradient
- Trust region
  ("gradient-based TRPO")
- High-performance
  implementation:
  - Distributed (multiple
    workers)
  - Synchronous gradient updates

Single uniform reward, based on forward progress



Humanoid: learned behaviors

- 27 DoFs
- 21 actuators

- Can deep RL agents learn multiple tasks?
- Can deep RL agents learn efficiently?
- Can deep RL agents learn from real data?
- Can deep RL agents learn continuous control?

Overcoming catastrophic forgetting in NNs, 2016
James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, Raia Hadsell

Progressive Neural Networks, 2016
Andrei A. Rusu, Neil C. Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, Raia Hadsell

Distral: Robust Multitask RL, 2017
Yee Whye Teh, Victor Bapst, Wojciech Marian Czarnecki, John Quan, James Kirkpatrick, Raia Hadsell, Nicolas Heess, Razvan Pascanu

Learning to navigate in complex environments, 2017
Piotr Mirowski*, Razvan Pascanu*, Fabio Viola, Hubert Soyer, Andrew J. Ballard, Andrea Banino, Misha Denil, Ross Goroshin, Laurent Sifre, Koray Kavukcuoglu, Dharshan Kumaran, Raia Hadsell

Learning and transfer of modulated locomotor controllers, 2016
Nicolas Heess, Greg Wayne, Yuval Tassa, Timothy Lillicrap, Martin Riedmiller, David Silver

Emergence of Locomotion Behaviours in Rich Environments, 2017
Nicolas Heess, Dhruva TB, Srinivasan Sriram, Jay Lemmon, Josh Merel, Greg Wayne, Yuval Tassa, Tom Erez, Ziyu Wang, S. M. Ali Eslami, Martin Riedmiller, David Silver

Thank you!

DeepMind