Learning to Navigate
. . . at City Scale

Raia Hadsell
Senior Research Scientist
DeepMind

[BBH Brazil for Renault / Art: Pedro Utzeri]

Navigation

Where am I?

Where am I going?

Where did I start?
How distant is A from B?
What is the shortest path from A to B?
Have I been here before?
How long until we get there?

Real world

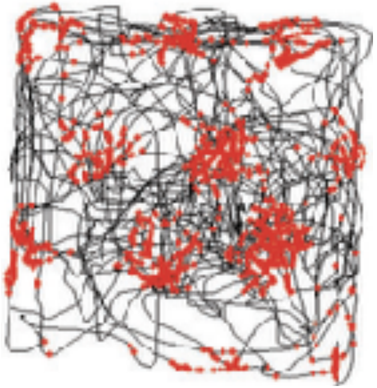
Modularity and transfer learning

Exploration

Multi-task prediction of sensory data

Memory

One-shot navigation in unseen environment



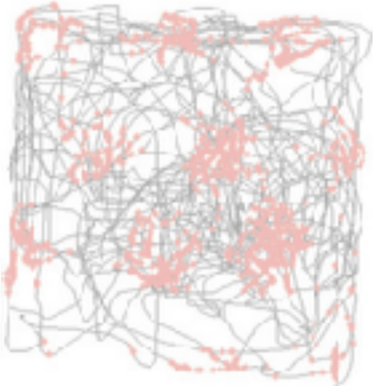
Representation

Grounding in neuroscience

Real world
Modularity and
transfer learning

Exploration
**Multi-task prediction
of sensory data**

Memory
One-shot navigation
in unseen environment



Representation
Grounding in
neuroscience

Can we teach agents to explore partially observed environments?

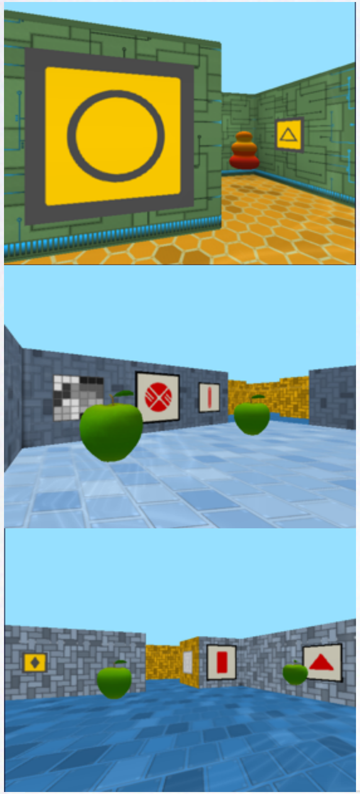
Learning to Navigate in Complex Environments

Piotr Mirowski*, Razvan Pascanu*, Fabio Viola, Hubert Soyer, Andy Ballard, Andrea Banino, Misha Denil, Ross Goroshin, Laurent Sifre, Koray Kavukcuoglu, Dharsh Kumaran and Raia Hadsell

[MIT News / Photo: Mark Ostow]

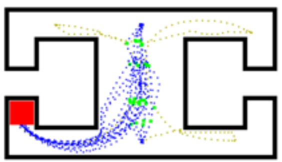
Navigation mazes

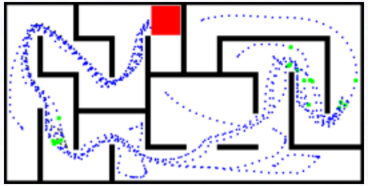
[Beattie et al (2016)
“DeepMind Lab”,
github.com/deepmind/lab]

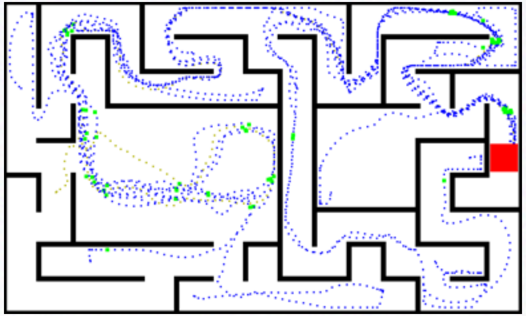


+10

+1

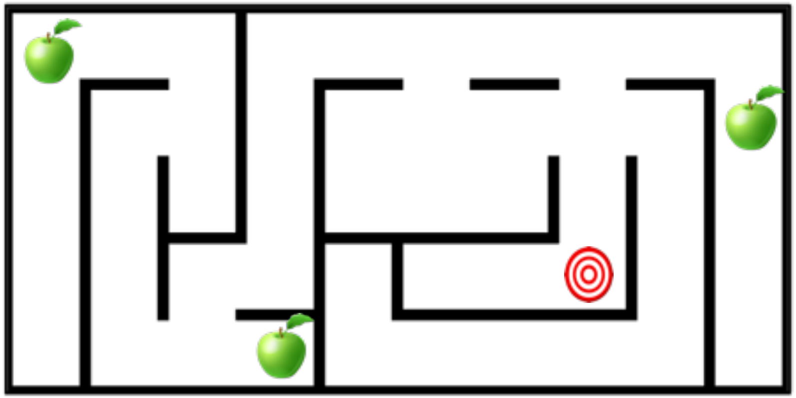


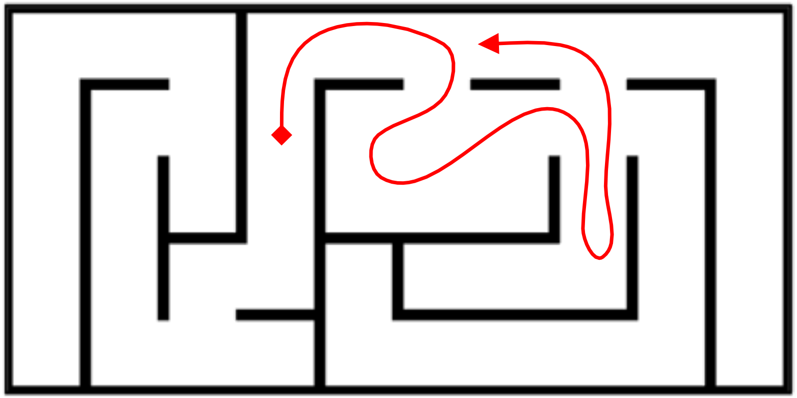




Within episode:
Fixed **goal** (static or randomly changing b/w episodes)
Random **respawns**

Given sparse rewards… **… explore and learn spatial knowledge**





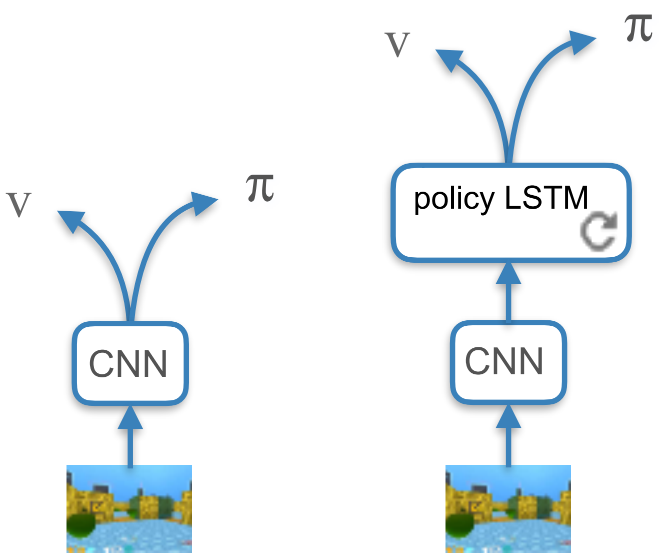
Accelerate reinforcement learning through **auxiliary losses**

Derive **spatial knowledge** from **auxiliary tasks**:

- **Depth** prediction
- Local **loop closure** prediction

Assess **navigation skills** through **position decoding**

Agent training



Advantage actor critic reinforcement learning

[Mnih, Badia et al (2015)

"Asynchronous Methods for Deep Reinforcement Learning"]

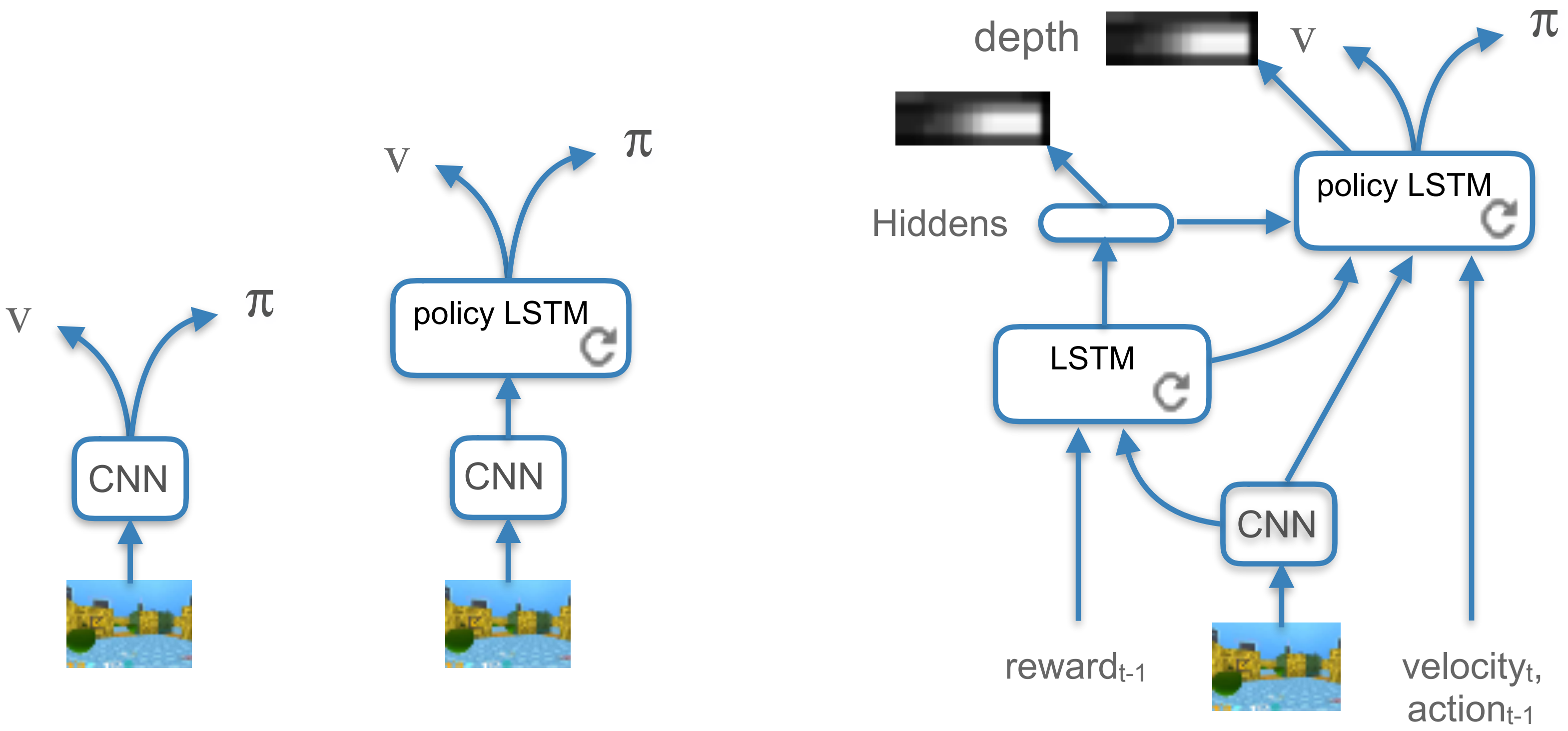
Agent observes state s_t and takes action a_t

Value $\mathrm{V}(\mathrm{s}_t; \theta_V)$ and policy $\pi(a_t|s_t;\theta)$ are updated with estimate of policy gradient given by the k-step advantage function A

Policy term: $\nabla_\theta \log \pi(a_t|s_t;\theta) A(s_t, a_t; \theta_V)$

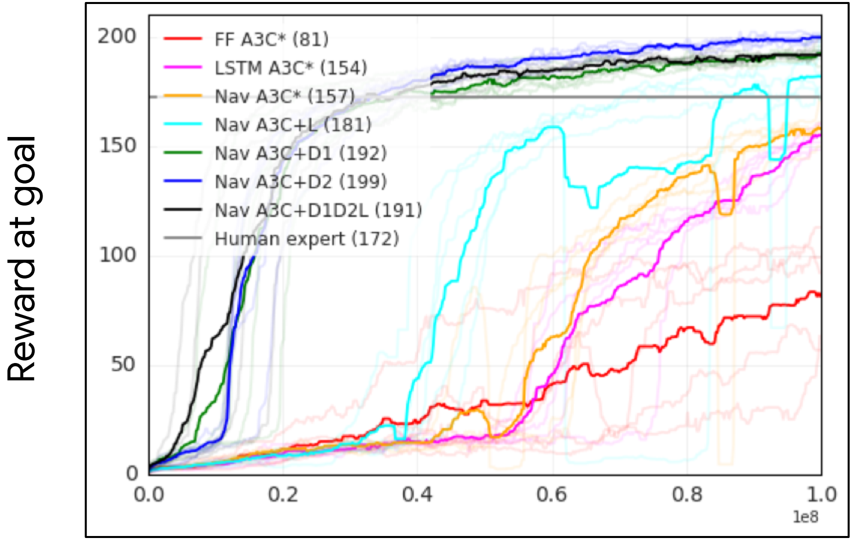
$$A(s_t, a_t; \theta_V) = \sum_{i=0}^{k-1} \gamma^i r_{t+i} + \gamma^k V(s_{t+k}; \theta_V) - V(s_t; \theta_V)$$

Navigation agent architectures



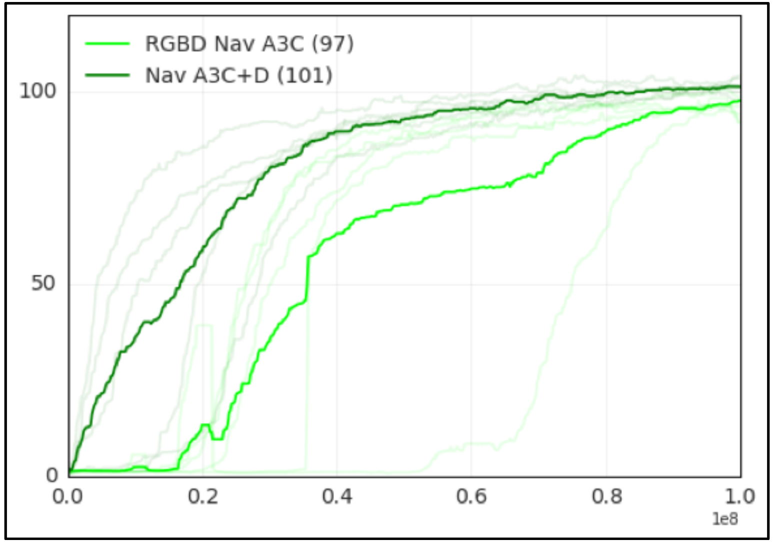
Results on large static mazes

Importance of auxiliary tasks

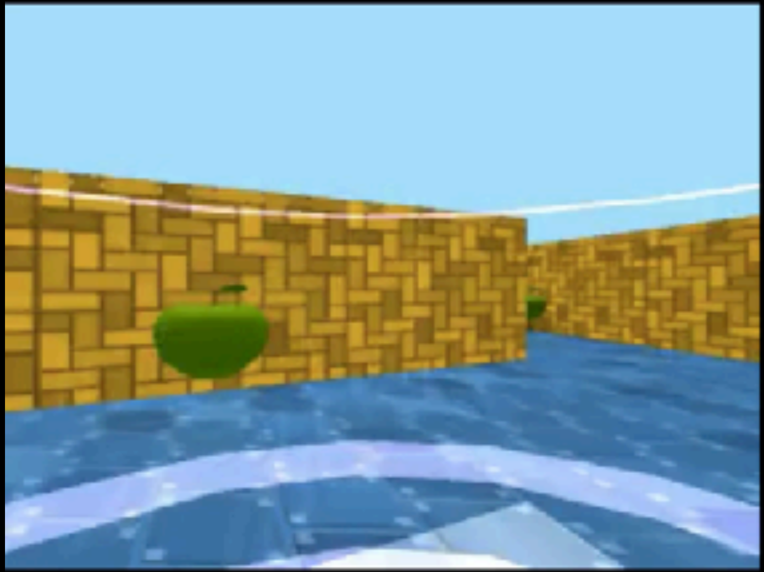
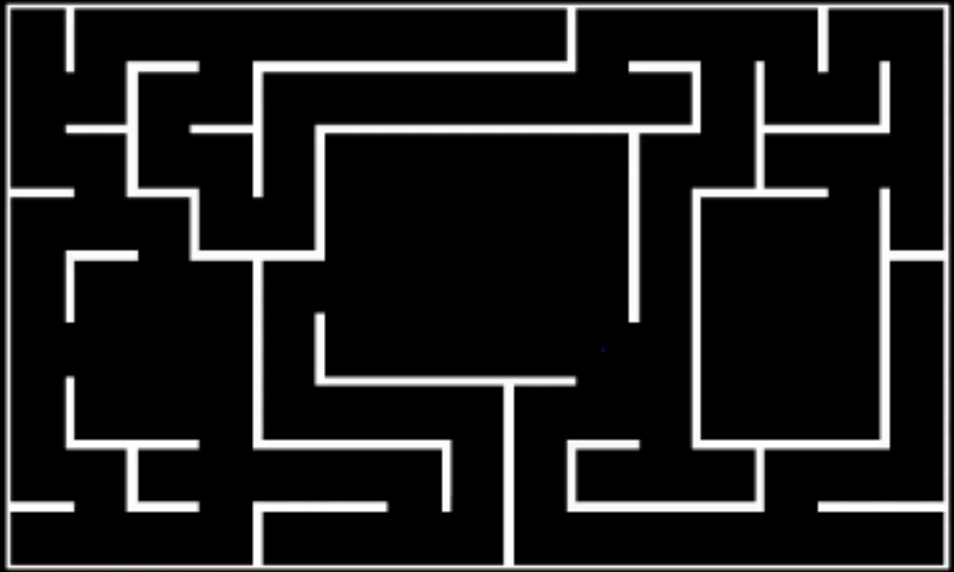


Environment steps

Depth prediction as auxiliary task outperforms using depth as inputs



Environment steps

- 3D, first person environment
- partially observed
- procedural variations

… *but it's not real*

Real world

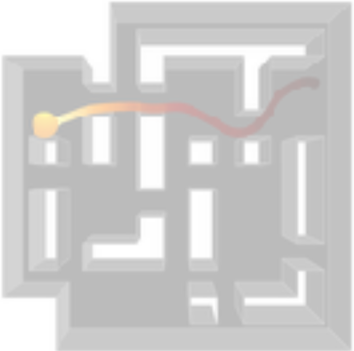
Modularity and transfer learning

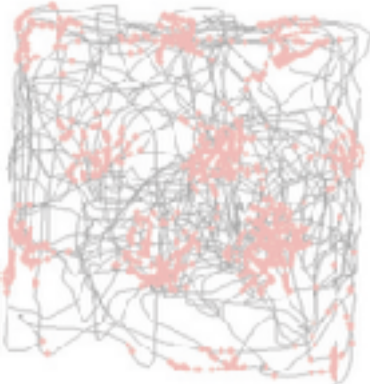
Exploration

Multi-task prediction of sensory data

Memory

One-shot navigation in unseen environment





Representation

Grounding in neuroscience

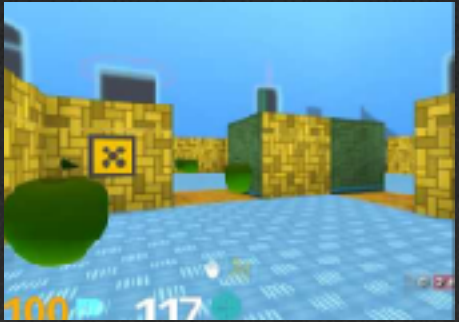
Can we solve navigation tasks in the real world?

Learning to Navigate in Cities Without a Map

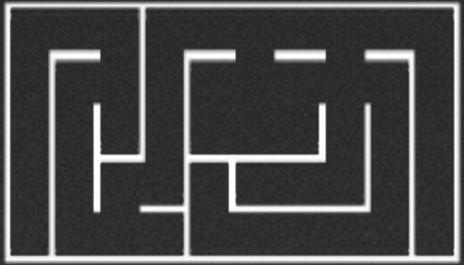
Piotr Mirowski*, Matthew Koichi Grimes, Mateusz Malinowski, Karl Moritz Hermann, Keith Anderson, Denis Teplyashin, Karen Simonyan, Koray Kavukcuoglu, Andrew Zisserman and Raia Hadsell

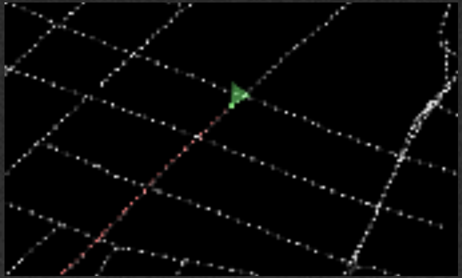
arxiv.org/abs/1804.00168

Can we solve navigation tasks in the real world?



Google
Street View





Street View as an RL environment: StreetLearn

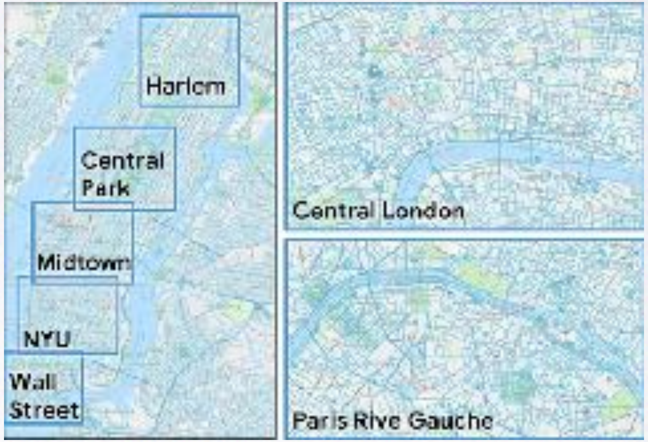
Street View image

Google Maps graph

RGB panoramic image
(we crop it and render at 84x84)

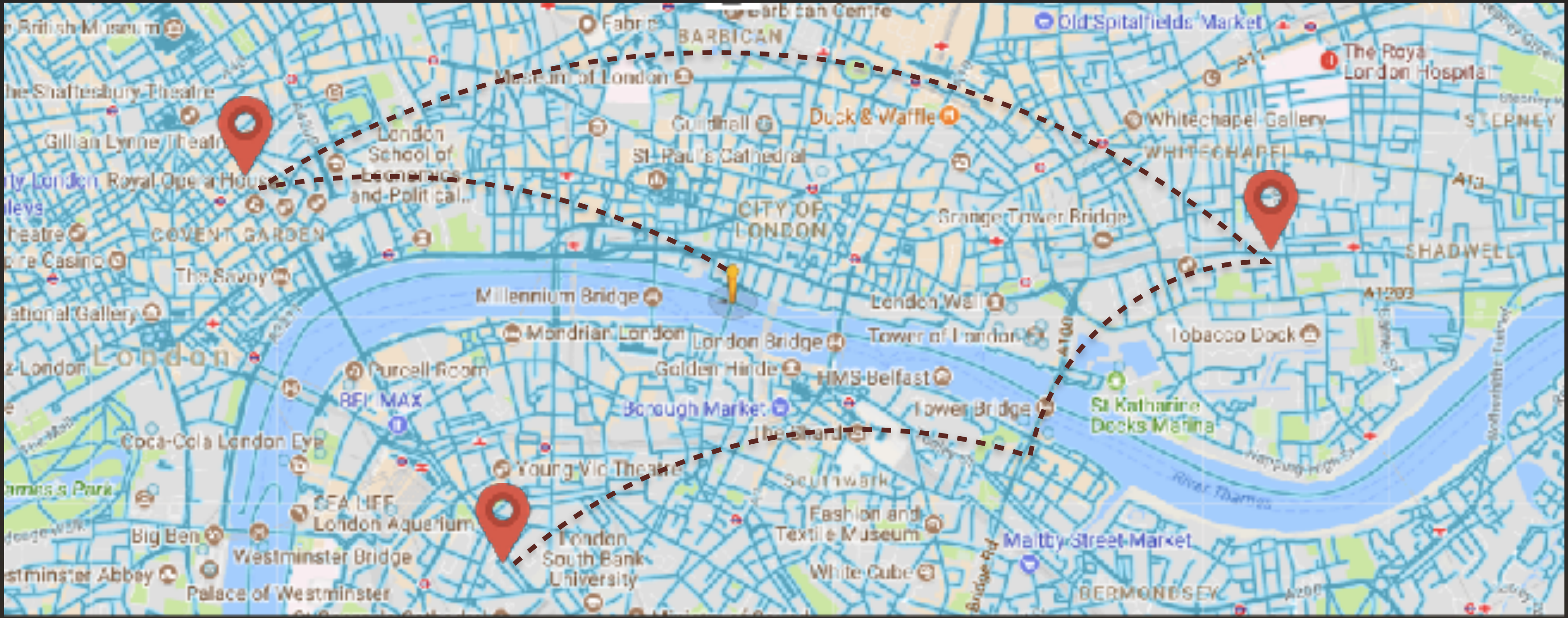
Actions:
move to the next node,
turn left/right

New York, London, Paris



- 14,000 to 60,000 nodes (panoramas) per “city”, covering range of 3.5-5km
- Discrete action space allows rotating in place and stepping to next node
- **Multi-city dataset and RL environment will be released later this year**

The Courier Task



DeepMind

The Knowledge

- Test to get a black cab license in London
- Candidates study for 3-4 years
- Memorize 25,000 roads and 20,000 named locations
- By the time they've passed the exam,

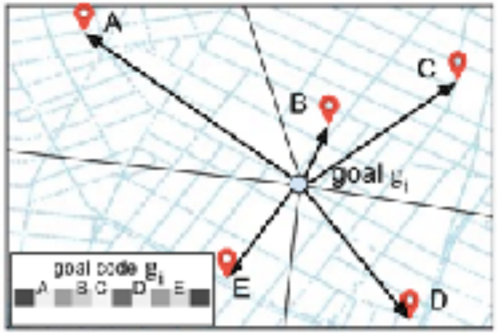
their hippocampuses are 'significantly enlarged'.

Woollett & Maguire. 2011. Acquiring "the Knowledge" of London's Layout Drives Structural Brain Changes. Current Biology

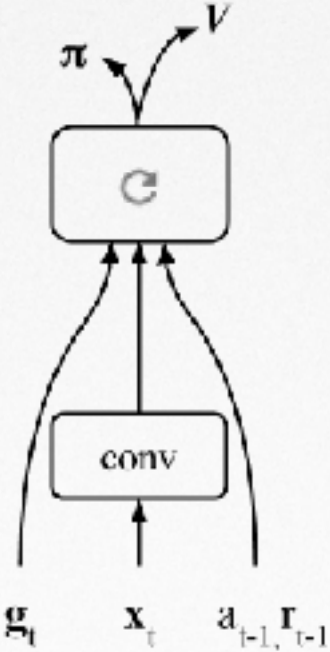
SCOTCH HOUSE
Hyde Park Corner
HYDE PARK CORNER
(Closed Midnight - 5am) New Ride
Prince of Wales Gate
SOUTH CARRIAGE
CONSTITUTION
(Closed to traffic on Sen
Mem. Gates
Buckingham Palace
GORE
A315
ROAD
KNIGHTSBRIDGE
Barracks
GROSVENOR
GROSVENOR CR.
A302
BUCKINGHAM PALACE
GARDENS
The Royal Mews
Knightsbridge
EXHIBITION ROAD
A4 ROAD
SLOANE ST.
A3216
Harrods
KNIGHTSBRIDGE
Coll.
PRINCES GATE MEWS
Brompton Oratory
Victoria & Albert Museum
Science Museum
BROMPTON
BROMPTON ROAD
PONT ST.
B319
BEAUFORT GS.
HANS
THURLOE PL.
ROAD
30
Sth. Kensington
PELHAM
LA Fitness
B304
ONSLOW
SLOANE SQUARE
Sloane Square
BELGRAVIA
BUCKINGHAM
Coach Sta.
PALACE
VICTOR

The Courier Task

- **Random start** and **target**
- Navigation **without a map**
- Reward shaped when close to goal (<200m)
- Actions: rotate left, right, or step forward
- Inputs for the agent at every time point *t*:
    - 84x84 RGB **image observations**
    - landmark-based **goal description**



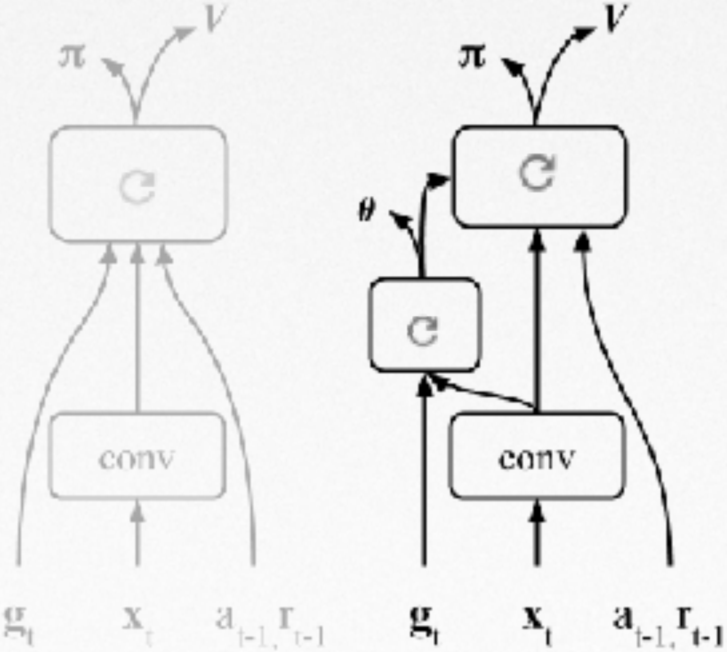
Architecture



a. *GoalNav* agent

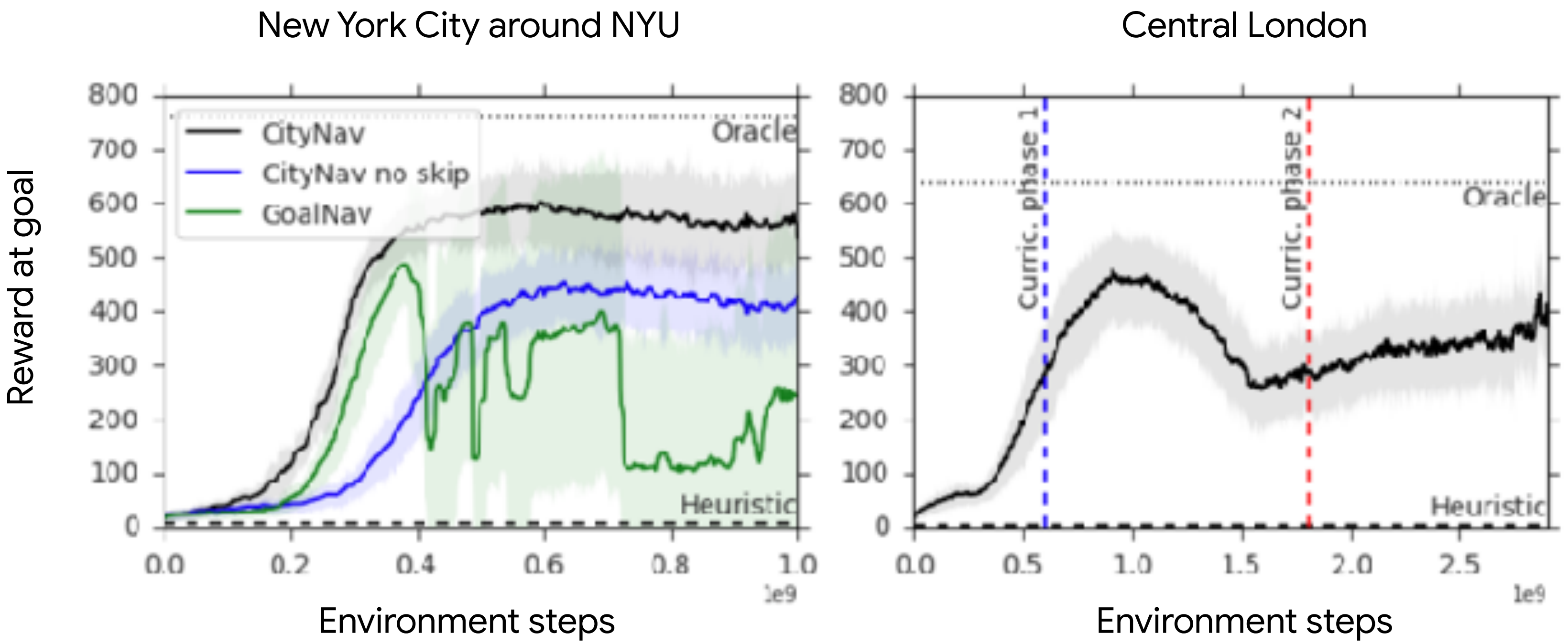
[Mnih, Badia et al (2015)
"Asynchronous Methods for Deep Reinforcement Learning"]

Architecture



a. *GoalNav* agent b. *CityNav* agent

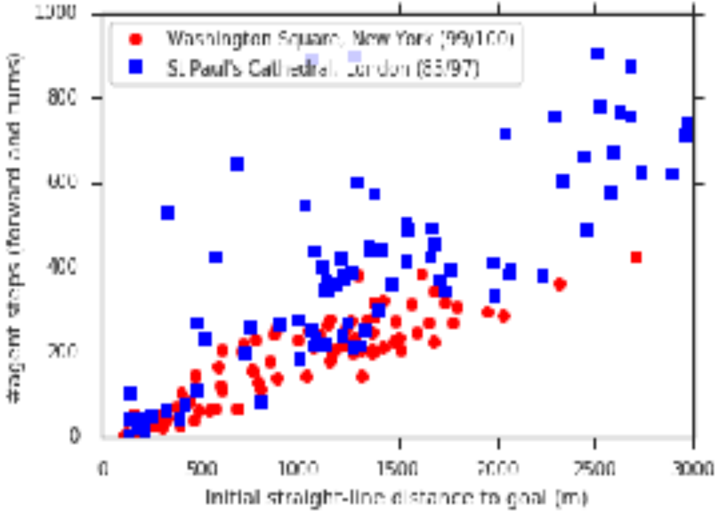
Successful learning on all 3 cities



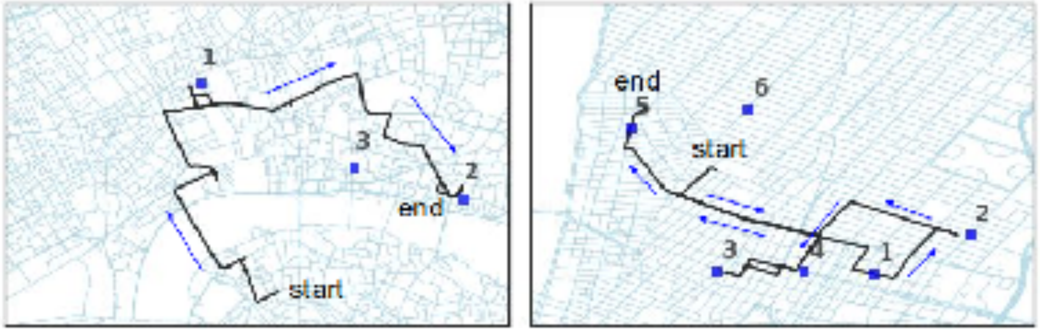

Matel821

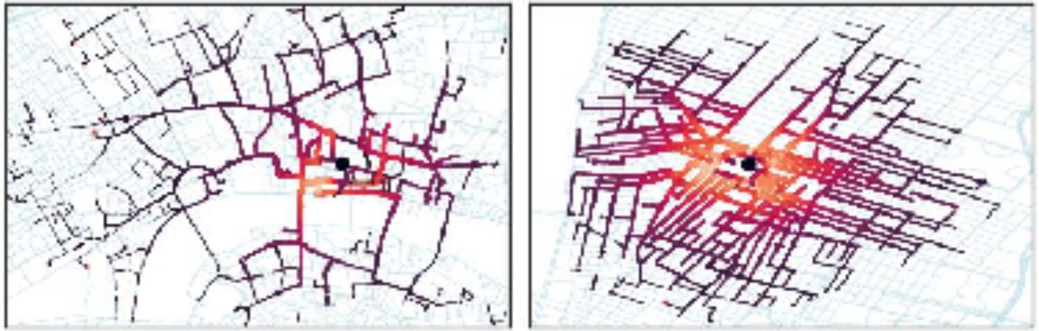

Chang's Noodle

Analysis of goal acquisition



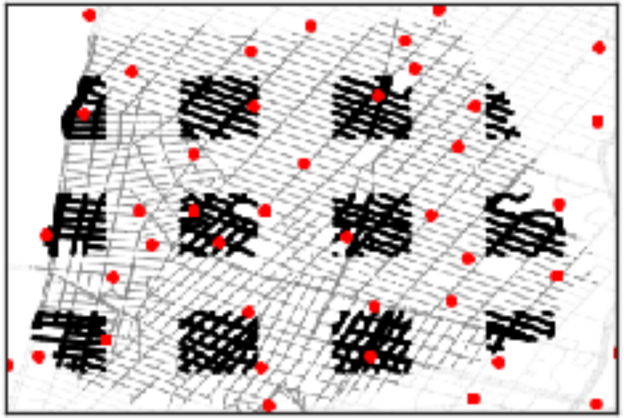
Examples of 1000-step episodes





Examples of value function for the same target

Generalization on new goal areas

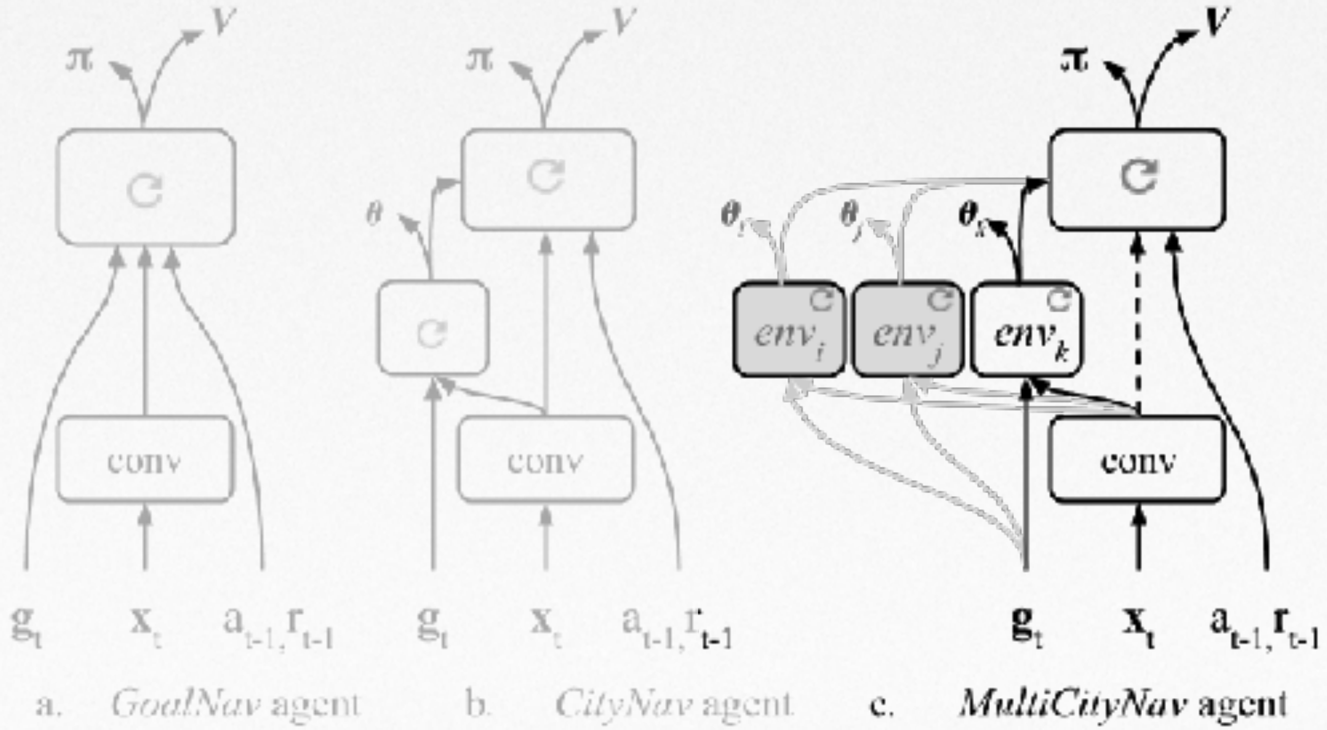


Goal locations held-out during training and **landmark locations**

Grid Size	Train Rewards	Test Rewards	Test Fail	Test $T_{\frac{1}{2}}$
FINE	655	567	11%	229
MEDIUM	637	293	20%	184
COARSE	623	164	38%	243

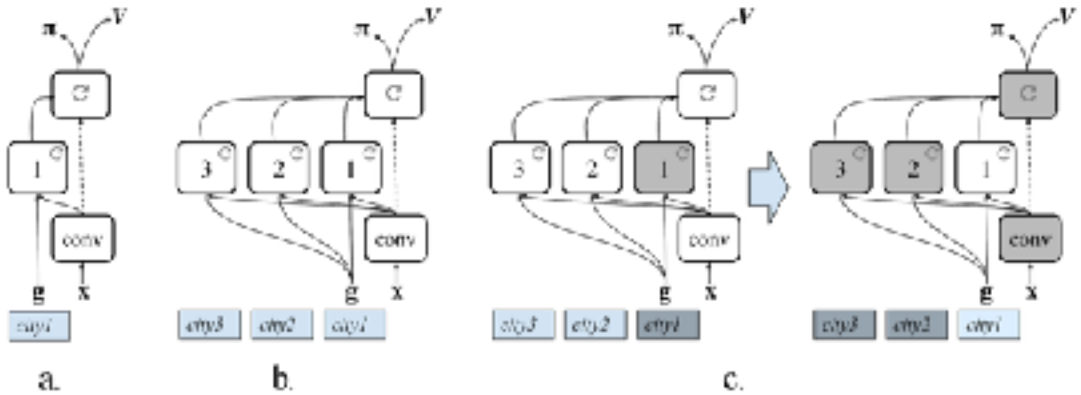
Table 1. CityNav agent same-city generalization performance (goal acquisition reward) when separating a training and a held-out set of destination locations shows that the agent performs worse as the size of the held-out area increases. In addition to the reward metric and a fail metric, we also compute the *half-trip time* ($T_{\frac{1}{2}}$, or the number of steps necessary to reach halfway to the goal) to understand the lower performance.

Architecture

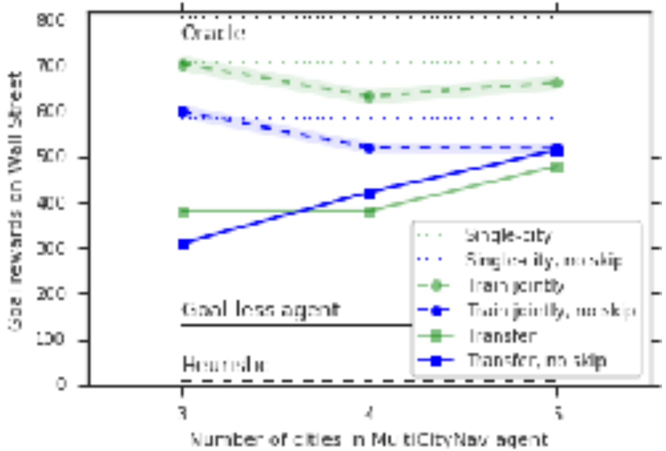


Multi-city modular transfer

Given a sequence of cities (regions of NYC), compare the following



single joint modular transfer



Successful navigation in target city,
even though the convnet and policy LSTM are frozen
and only the goal LSTM is trained.

Moreover, we note that the transfer success is correlated to number of cities seen during pre-training.

Train
in multiple environments

Many thanks to many collaborators!

- Learning to navigate in complex environments (ICLR 2017)
  Piotr Mirowski*, Razvan Pascanu*, Fabio Viola, Hubert Soyer, Andy Ballard, Andrea Banino, Misha Denil, Ross Goroshin, Laurent Sifre, Koray Kavukcuoglu, Dharsh Kumaran and Raia Hadsell

- Learning to navigate in cities without a map (NIPS 2018)
  Piotr Mirowski*, Matthew Koichi Grimes, Keith Anderson, Denis Teplyashin, Mateusz Malinowski, Karl Moritz Hermann, Karen Simonyan, Koray Kavukcuoglu, Andrew Zisserman, Raia Hadsell

www.deepmind.com www.raiahadsell.com