Microsoft

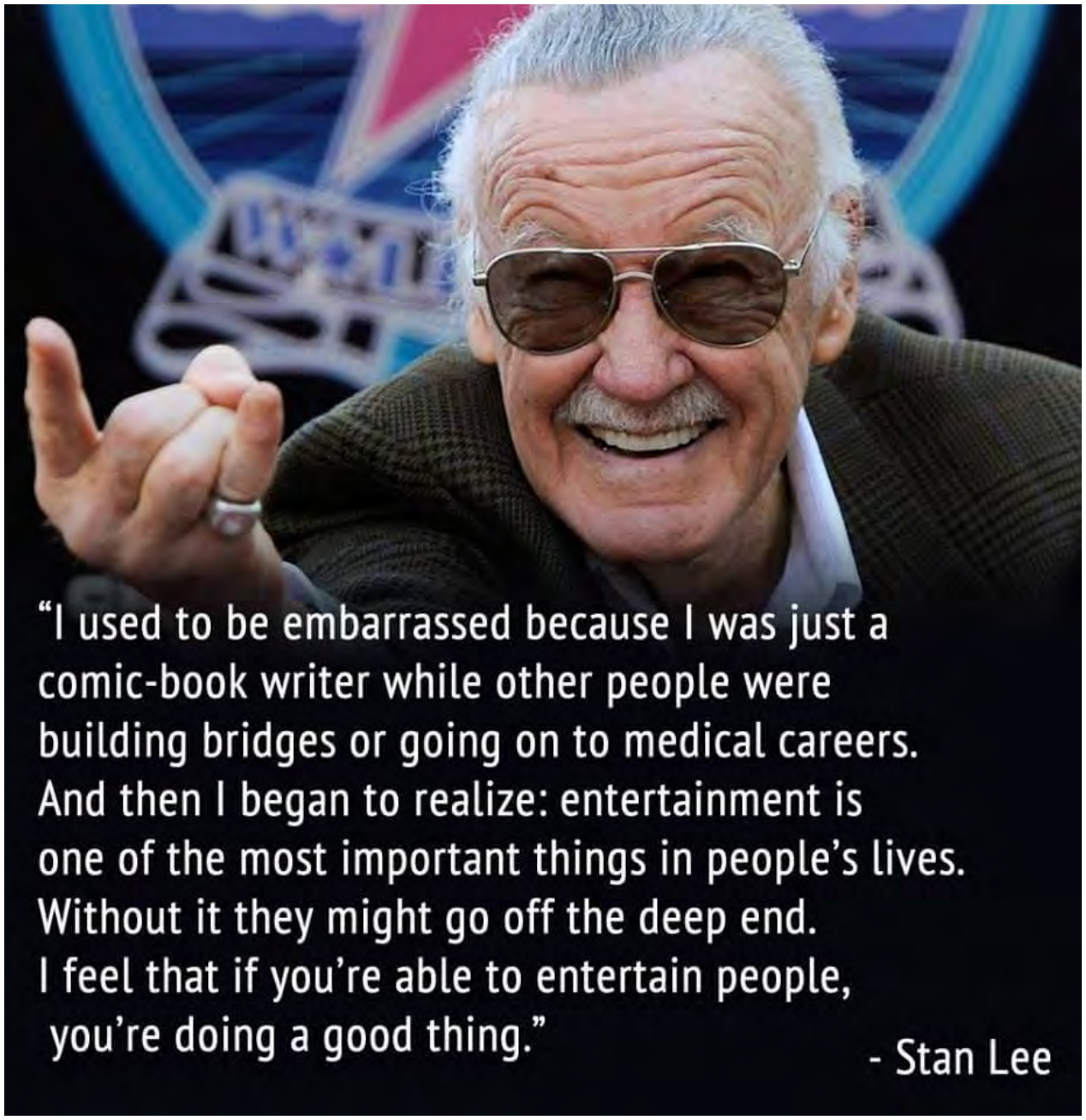
Generalization in RL with Selective Noise Injection

Sam Devlin

Game Intelligence
Microsoft Research, Cambridge, UK
aka.ms/gameintelligence

@smdvln


E3 2019 BRIEFING
Jump in.
THE MICROSOFT THEATER
IN XBOX PLAZA
Sam Devlin
Microsoft Research
MY PRONOUNS:
HE


“I used to be embarrassed because I was just a comic-book writer while other people were building bridges or going on to medical careers. And then I began to realize: entertainment is one of the most important things in people’s lives. Without it they might go off the deep end. I feel that if you’re able to entertain people, you’re doing a good thing.”
- Stan Lee

AI for Players

North star: Enable new types of game experiences

Example: AI agents that adapt to user generated content

AI for Game Devs

North star: Make RL accessible to all game developers

Examples: Robust RL algorithms, interpretable and controllable behaviour

Microsoft

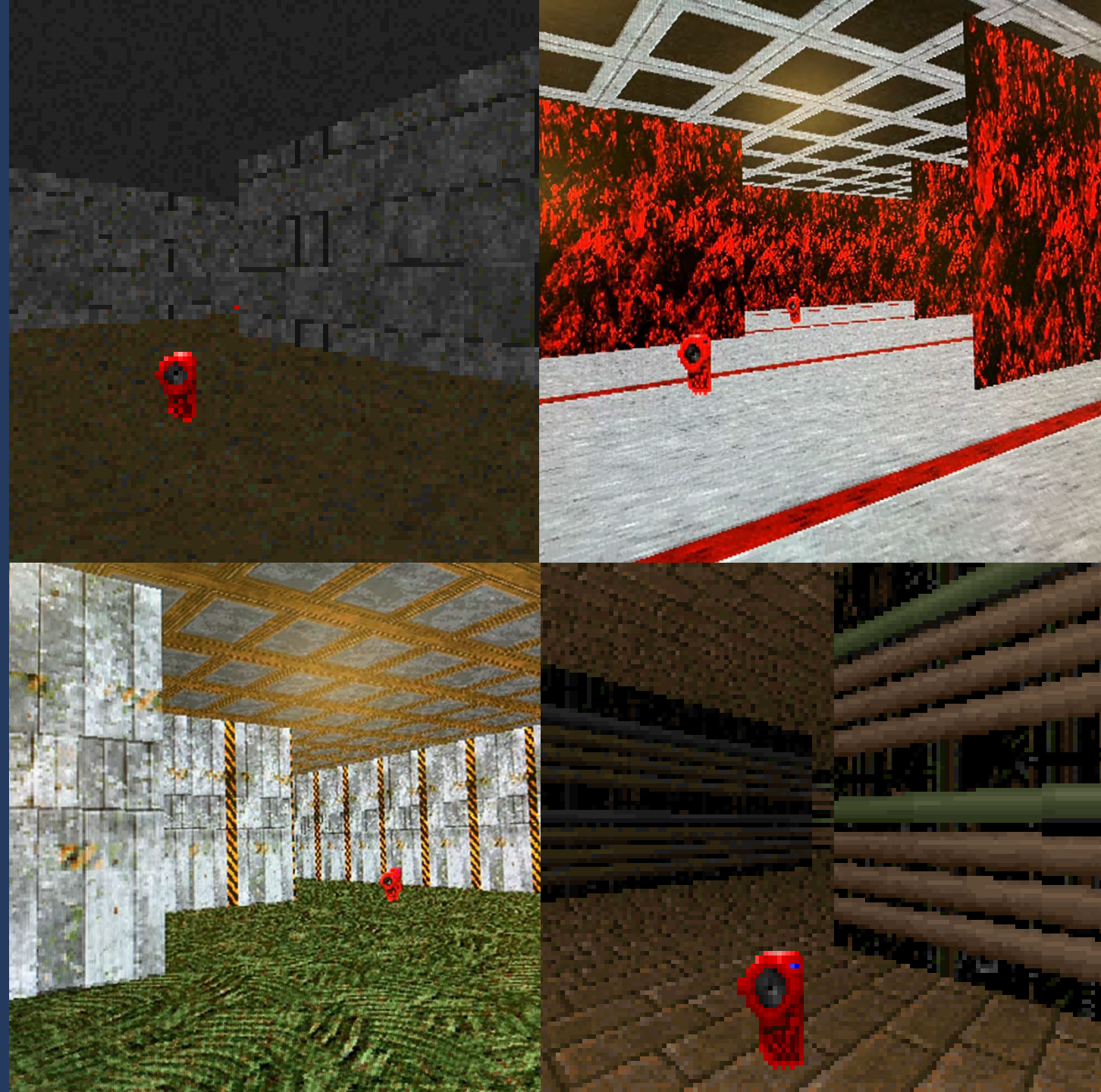
MazeExplorer

A Customisable 3D Benchmark for Assessing Generalisation in Reinforcement Learning

Luke Harries*, Sebastian Lee*,
Jaroslaw Rzepecki, Katja Hofmann, Sam Devlin

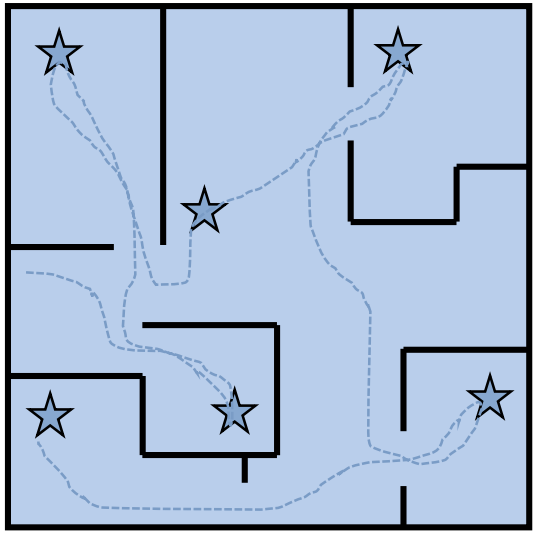
IEEE Conference on Games 2019

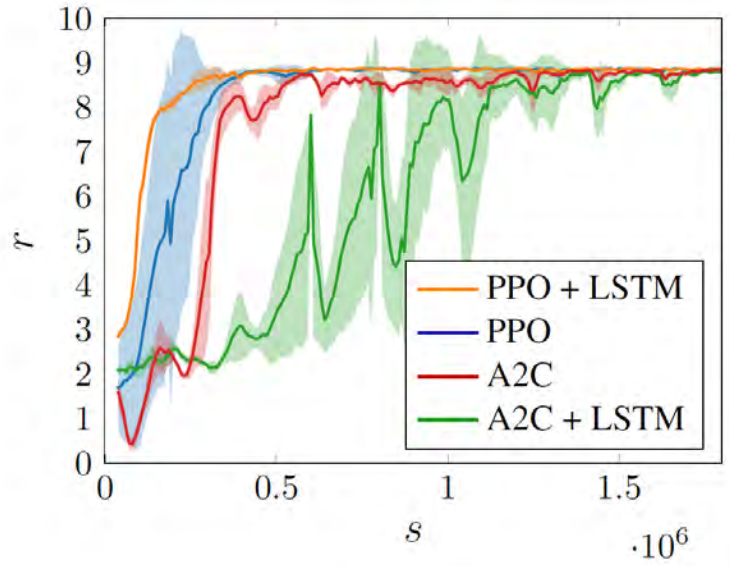
https://github.com/microsoft/MazeExplorer



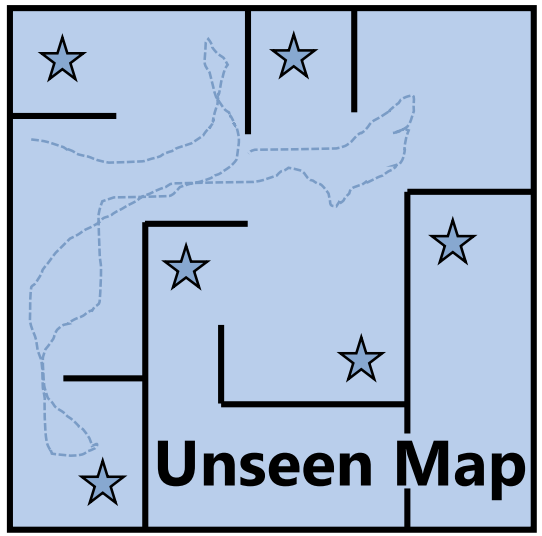
Generalization in Reinforcement Learning

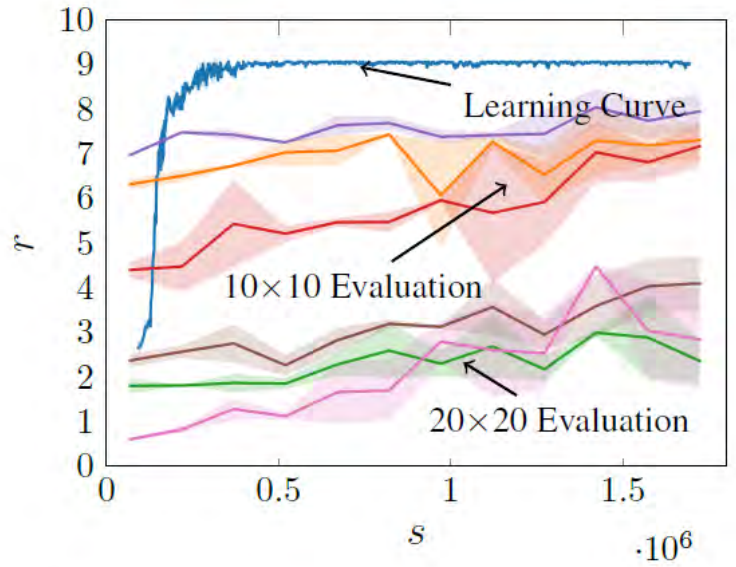
Train





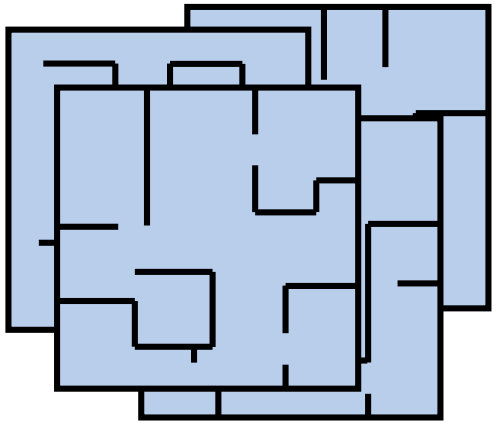
Test



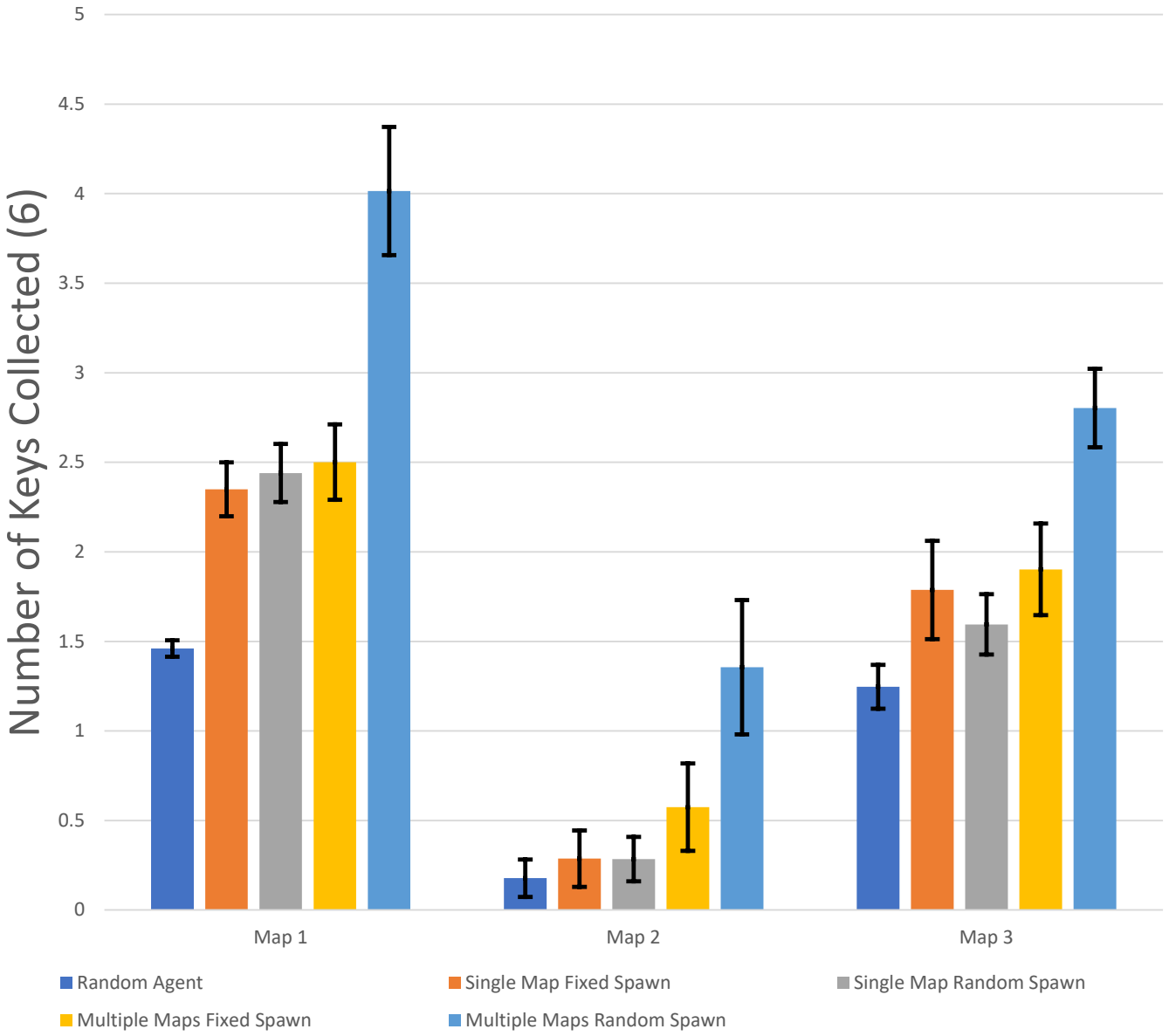


Domain Randomisation Improves Generalisation

Multiple Maps



Random Spawn Location



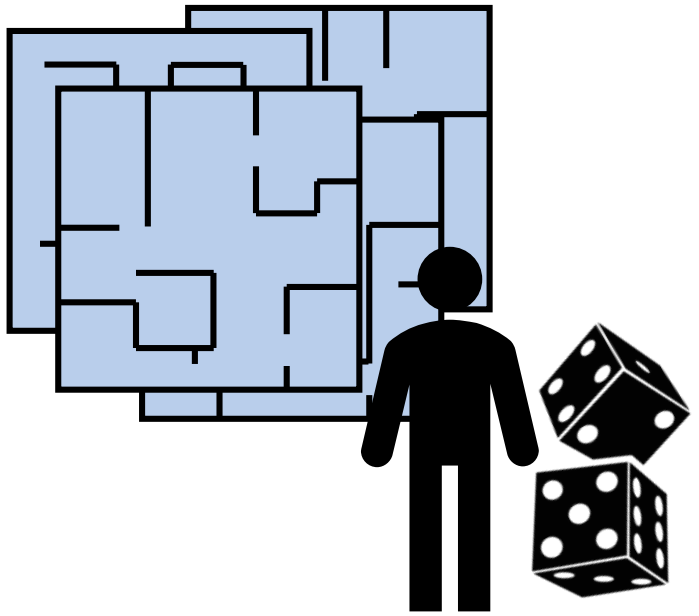
Generalization in Reinforcement Learning

Domain Randomization
e.g. MazeExplorer [CoG 2019]

Agent

Actions

Observations
Rewards



Environment

Microsoft

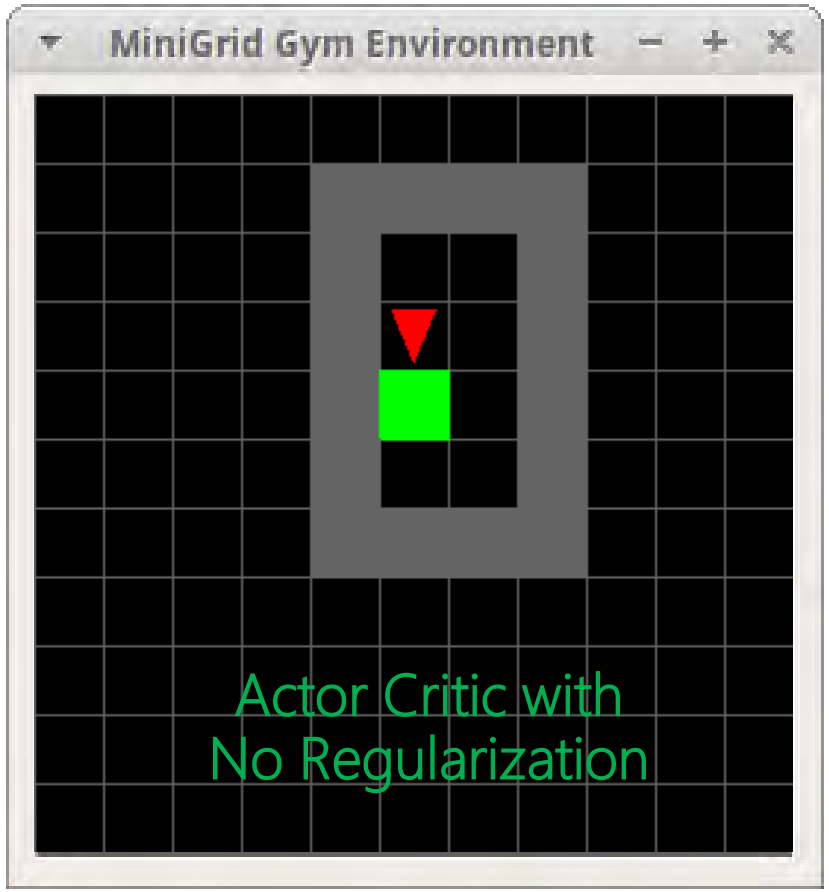
Generalization in RL with Selective Noise Injection and Information Bottleneck

Maximilian Igl (University of Oxford)
Kamil Ciosek , Yingzhen Li, Sebastian Tschiatschek,
Cheng Zhang, Sam Devlin, Katja Hofmann.

NeurIPS 2019

Regularization in Reinforcement Learning

Issue #1: RL Agents Act In A Low-Data Regime Early On In Training



Information Bottleneck Actor Critic

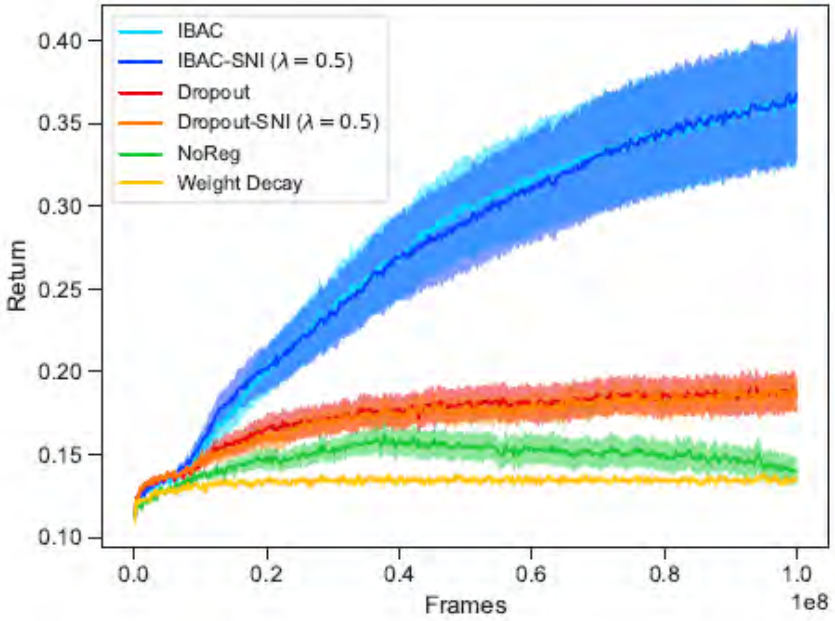
States → Z → Actions

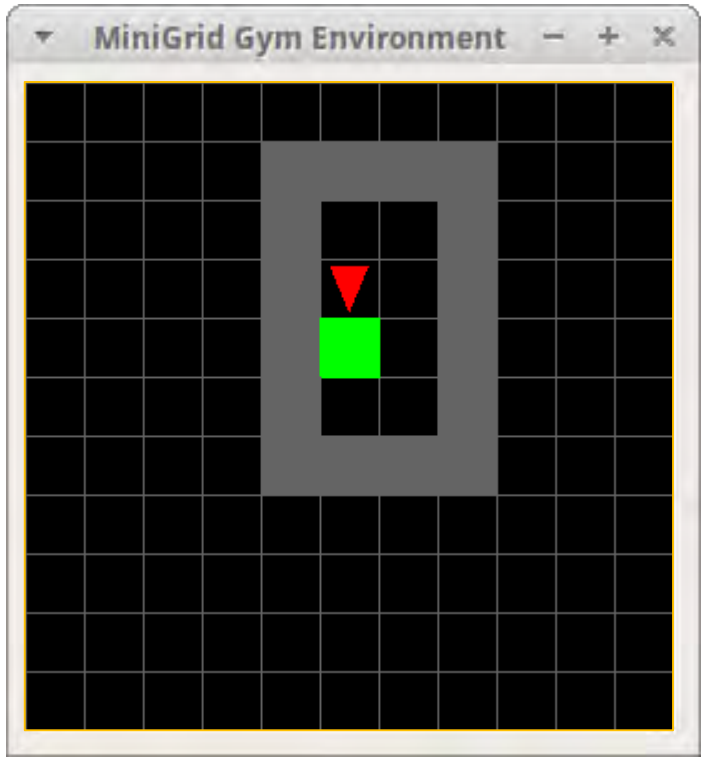
ENCODER

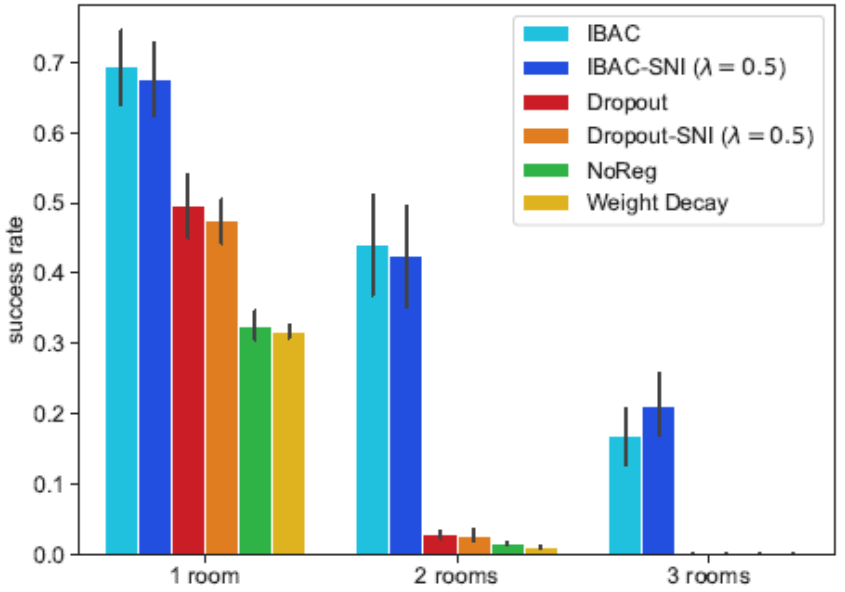
Minimize Mutual Information between States Input and Z

DECODER

Maximize Mutual Information between Actions Output and Z



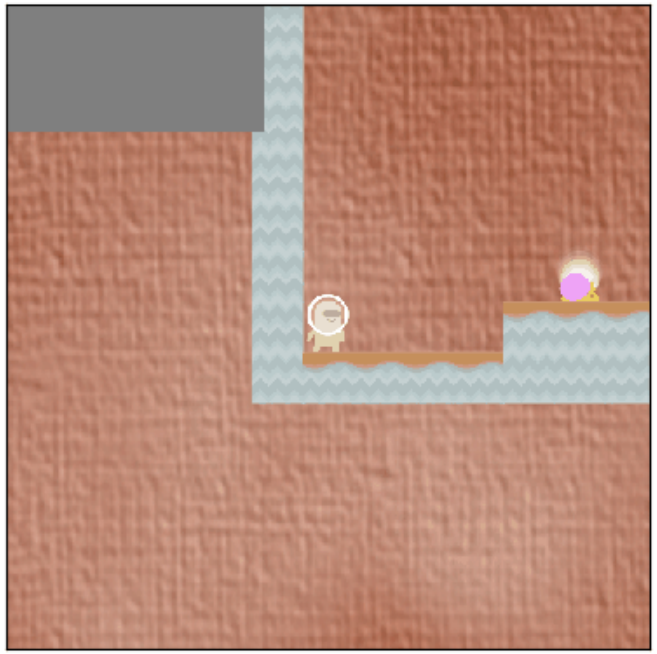




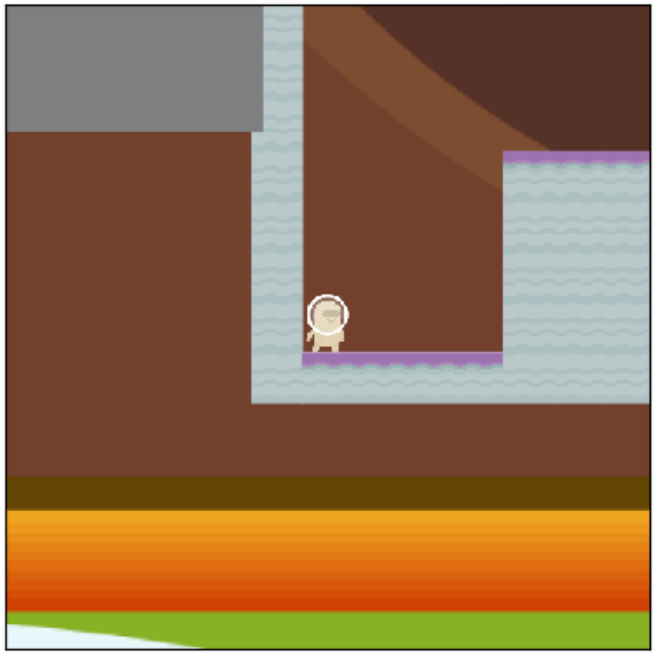
Regularization in Reinforcement Learning

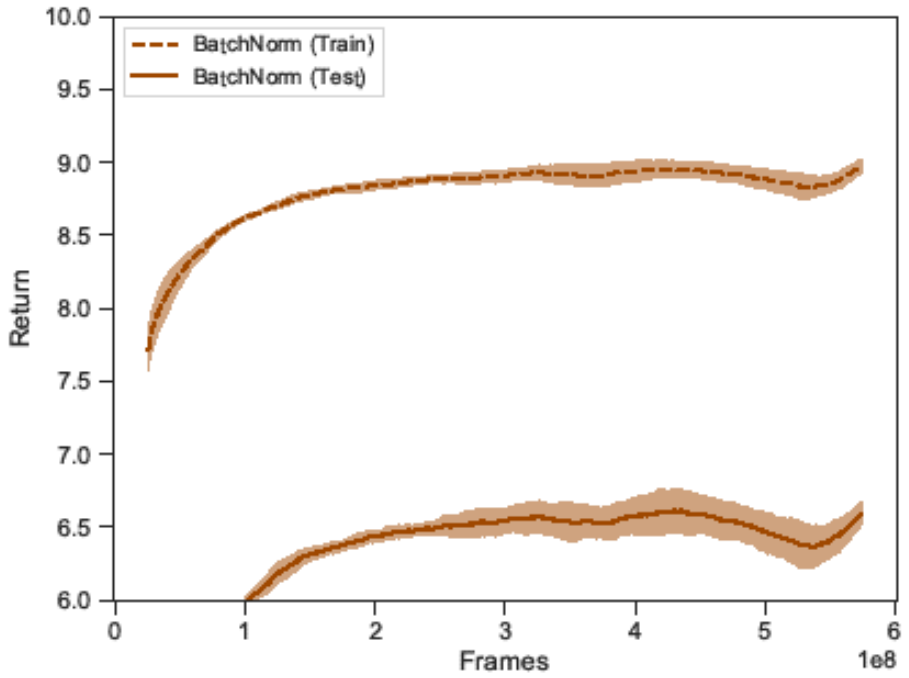
Issue #2: RL Agents Generate Their Own Dataset

BatchNorm (Train)



BatchNorm (Test)

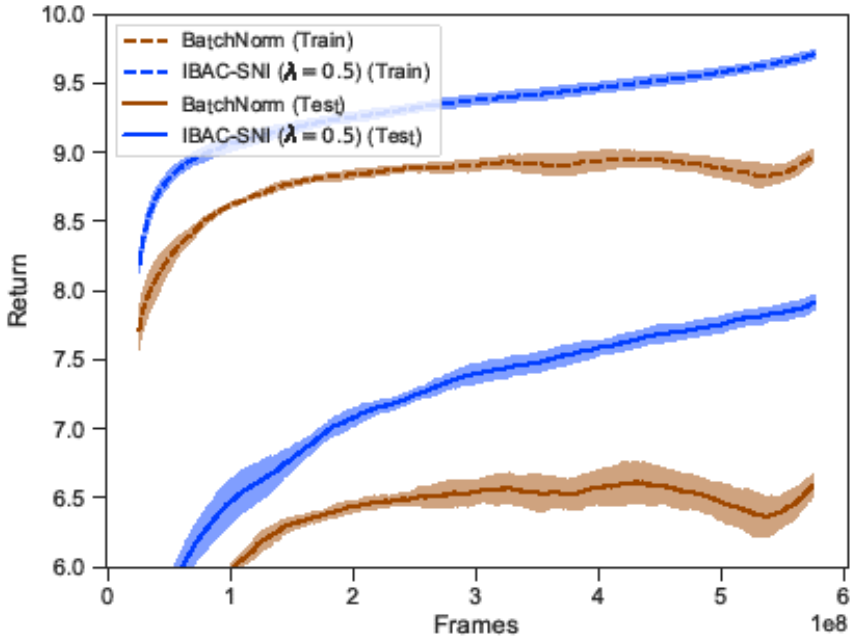




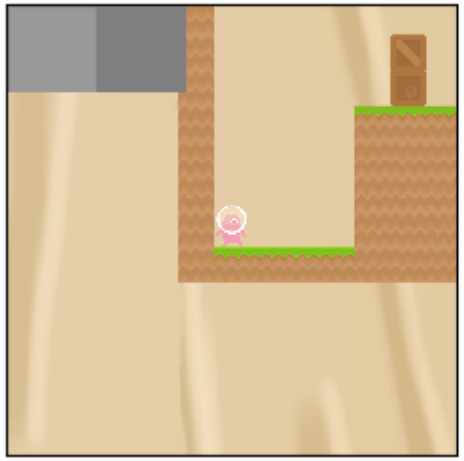
Selective Noise Injection (SNI) Improves Generalisation

Issue #2: RL Agents Generate Their Own Dataset

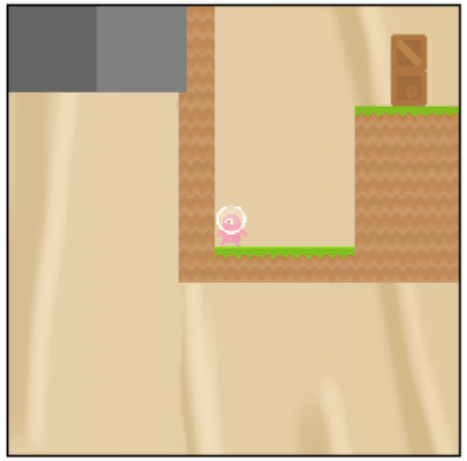
Do Not Add Stochastic Regularization To Rollout Policy or Critic



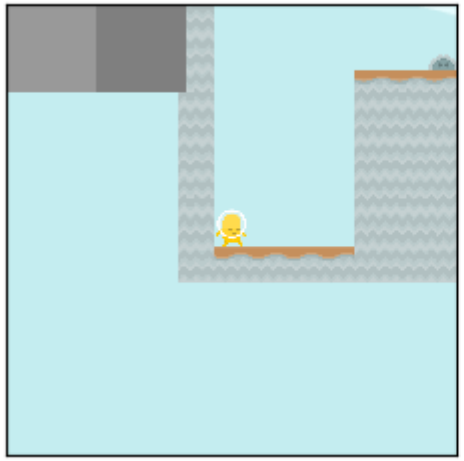
IBAC



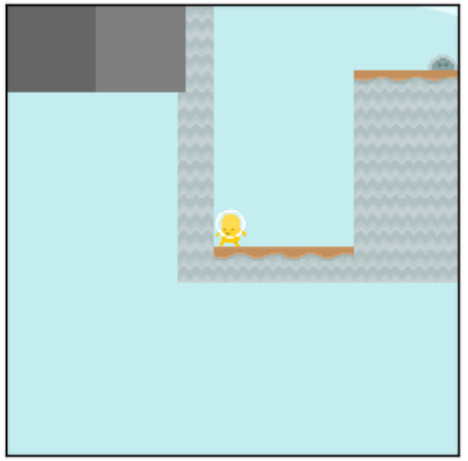
BatchNorm



IBAC



BatchNorm



Conclusion

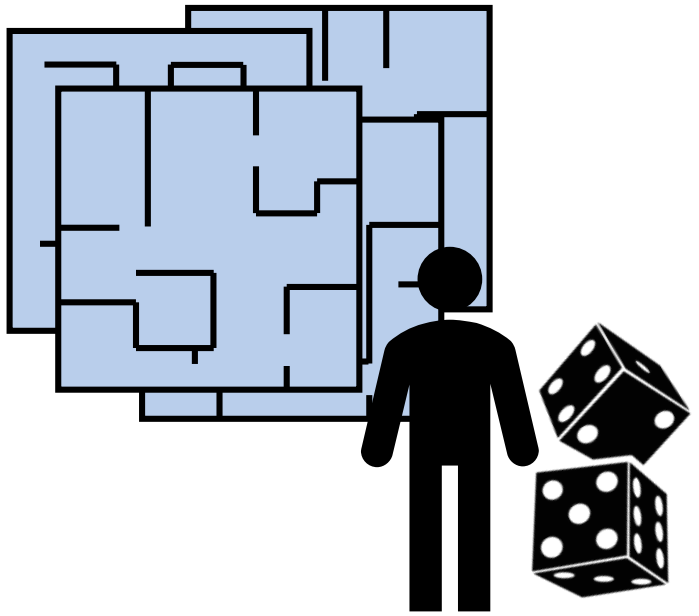
Generalization in RL with Selective Noise Injection and Information Bottleneck [NeurIPS 2019]

Agent

Actions

Observations
Rewards

Domain Randomization
e.g. MazeExplorer [CoG 2019]



Environment

Future Work

Generalization in Multiplayer Games

http://aka.ms/marlo

Agents collaborate to catch animal in a small enclosure

One agent collects and caries treasure to a goal, the other defends the team from attackers

Agents collaborate to build a structure, but the faster agent earns more rewards

Microsoft

Generalization in RL with Selective Noise Injection

Sam Devlin

Game Intelligence
Microsoft Research, Cambridge, UK
aka.ms/gameintelligence

@smdvln

Selective Noise Injection (SNI)

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}_{\pi^r_\theta(a_t|s_t)}\left[\sum_t^T \frac{\pi_\theta(a_t|s_t)}{\pi^r_\theta(a_t|s_T)} \nabla_\theta \log \pi_\theta(a_t|s_t)(r_t + \gamma V_\theta(s_{t+1}) - V_\theta(s_t))\right]$$

$$\min_\theta \mathbb{E}\left[\left(\gamma V^\perp_\theta(s_{t+1}) + r_t - V_\theta(s_t)\right)^2\right]$$

In RL adding regularization techniques can decrease quality of gradient due to:

1. Worse rollout policy leads to less observed rewards + prematurely ending episodes
2. High variance in the policy leads to high off-policy correction term
3. High variance in critic estimation

$$\nabla_\theta J(\pi_\theta) = \mathcal{G}_{\text{AC}}(\pi^r_\theta, \pi_\theta, V_\theta)$$

$$\mathcal{G}^{\text{SNI}}_{\text{AC}}(\pi^r_\theta, \pi_\theta, V_\theta) = \lambda \mathcal{G}_{\text{AC}}(\bar{\pi}^r_\theta, \bar{\pi}_\theta, \bar{V}_\theta) + (1-\lambda)\mathcal{G}_{\text{AC}}(\bar{\pi}^r_\theta, \pi_\theta, \bar{V}_\theta)$$

Selective Noise Injection Improves Generalisation