Generalization and Regularization in DQN

Jesse Farebrother **Marlos C. Machado** **Michael Bowling**

amii — University of Alberta

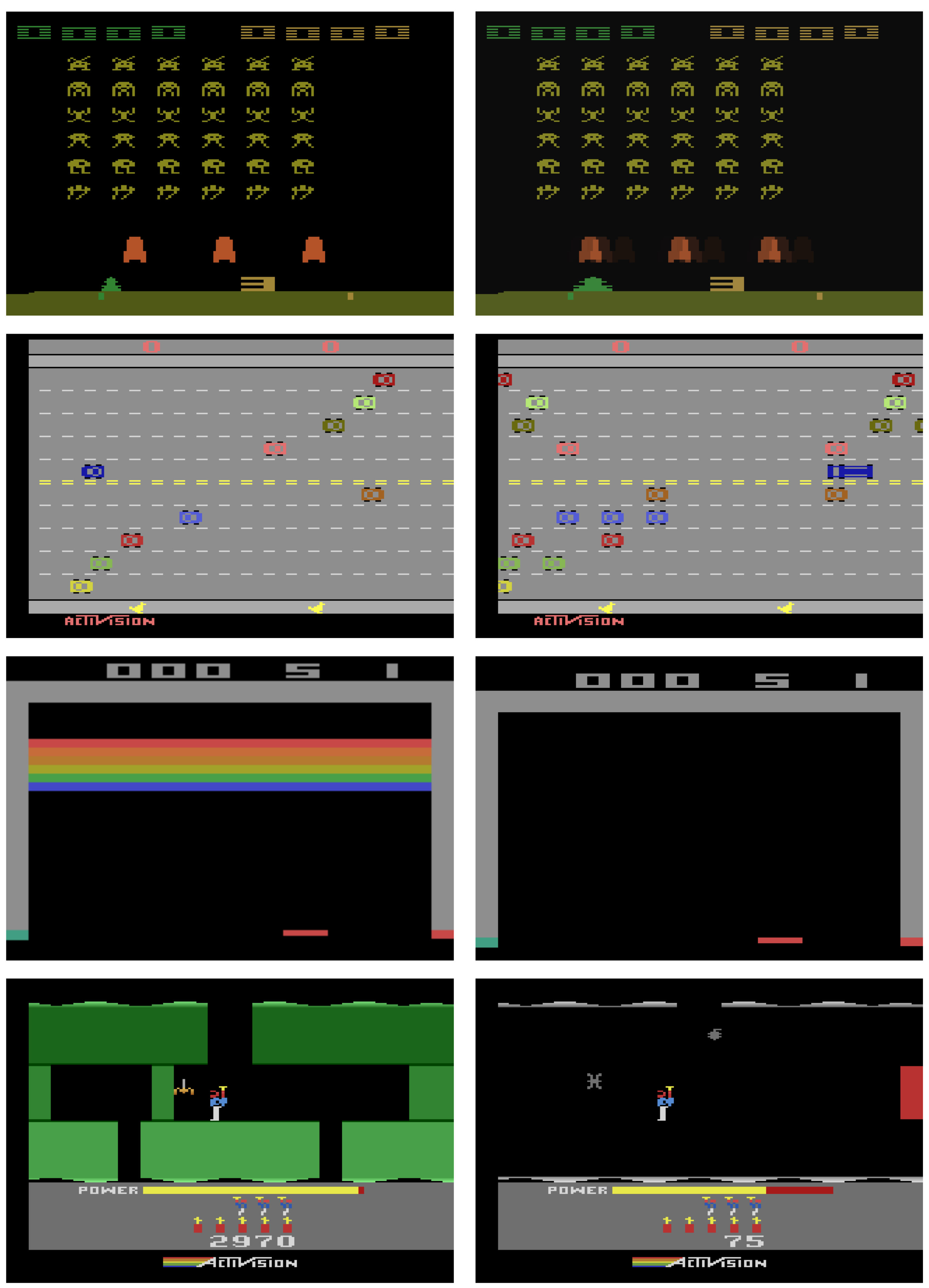
Motivation

Do deep neural networks allow RL agents to generalize to small variations of high-dimensional environments, e.g, game flavours in the ALE?

Can regularization methods in supervised learning, e.g., dropout, weight decay, be leveraged to allow deep RL agents to generalize to these variations?

Do these regularization methods allow for more adaptable representations, i.e., can we fine-tune the representation to improve sample efficiency?

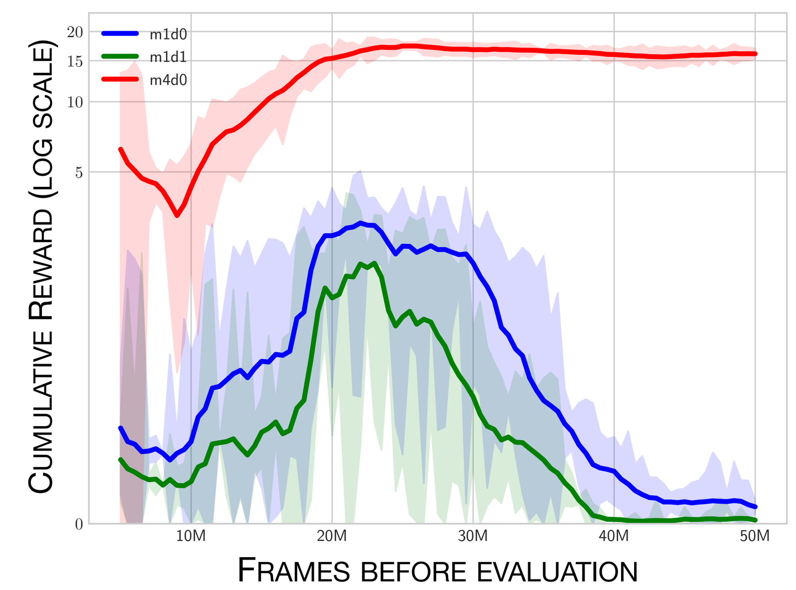
Game flavours in the ALE



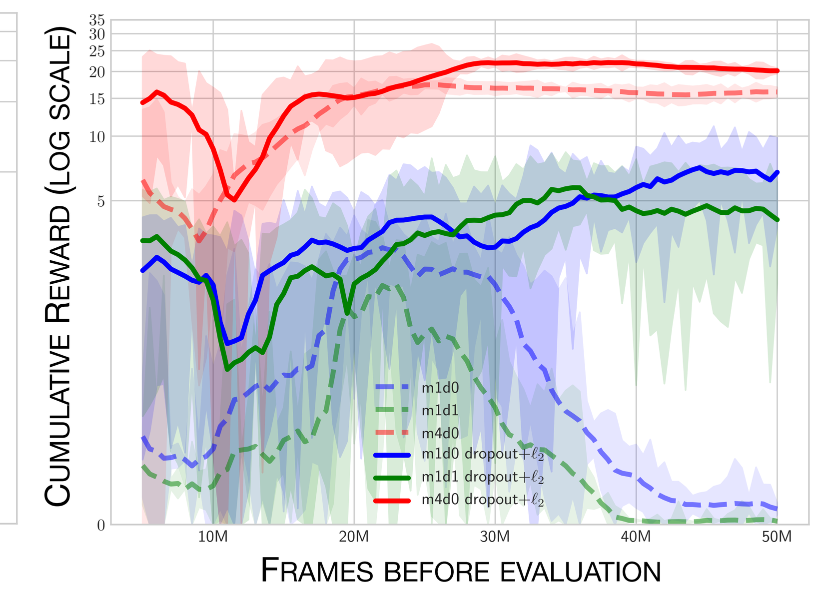
Generalization And Overfitting

- Evaluate the learned policy from the default flavour to every other game flavour
- We observe the agent overfitting to the default flavour in some games when evaluation during training

Freeway Policy Evaluation



Freeway Policy Evaluation w/ Regularization



Regularization in deep RL

- Employ dropout and weight decay during training to study the effect on generalization
- Dropout and weight decay work in tandem and improves evaluation performance on some flavours

DQN objective with weight decay:

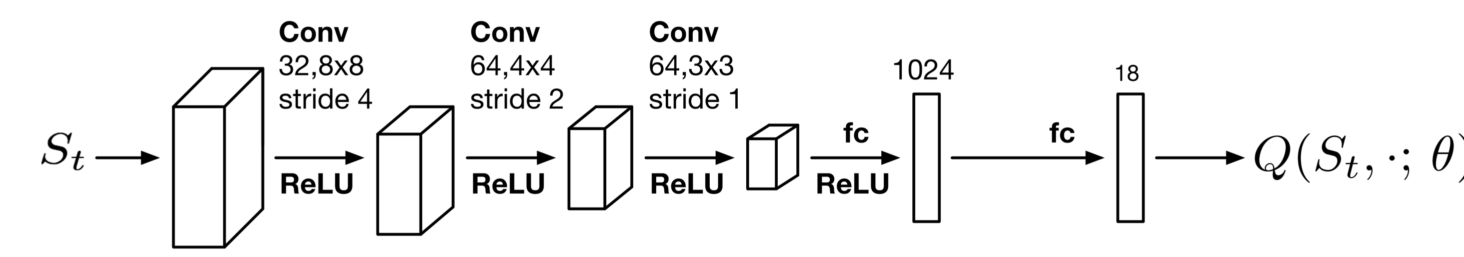
$$\min_{\theta} \frac{\lambda}{2} \parallel \theta \parallel_2^2 + \mathbb{E}_{S_t, A_t, R_{t+1}, S_{t+1} \sim U(\cdot)}\left[\left(R_{t+1} + \max_{a' \in A} Q(S_{t+1}, a'; \theta^-) - Q(S_t, A_t; \theta)\right)^2\right]$$

Drop neural unit according to: $d_j^{(l)} \sim$ **Bernoulli**(p)

Game	Variant	Eval. With Regularization		Eval. Without Regularization		Learn from Scratch	
Freeway	M1D0	5.8	(3.5)	0.2	(0.2)	4.8	(9.3)
	M1D1	4.4	(2.3)	0.1	(0.1)	0.0	0.0
	M4D0	20.6	(0.7)	15.8	(1.0)	29.9	(0.7)
Hero	M1D0	116.8	(76.0)	82.1	(89.3)	1425.2	(1755.1)
	M2D0	30.0	(36.7)	33.9	(38.7)	326.1	(130.4)
Breakout	M12D0	31.0	(8.6)	43.4	(11.1)	67.6	(32.4)
Space Invaders	M1D0	456.0	(221.4)	258.9	(88.3)	753.6	(31.6)
	M1D1	146.0	(84.5)	140.4	(61.4)	698.5	(31.3)
	M9D0	290.0	(257.8)	179.0	(75.1)	518.0	(16.7)

Value function fine-tuning

- Re-use the regularized representation post-training from one flavour to fine-tune on a different flavour
- Fine-tuning from a regularized representation improves sample efficiency and outperforms training from scratch in most flavours
- Re-learning co-adaptations between regularized layers and randomly initialized layers doesn't provide any immediate benefit



Game	Variant	Fine-Tune 50M		Regularized Fine-Tune 50M		Scratch 100M	
Freeway	M1D0	22.5	(7.5)	25.4	(0.2)	7.5	(11.5)
	M1D1	17.4	(11.4)	25.4	(0.4)	2.5	(7.3)
	M4D0	31.4	(0.5)	32.2	(0.5)	32.8	(0.2)
Hero	M1D0	496.7	(362.8)	4104.6	(2192.8)	5026.8	(2174.6)
	M2D0	92.5	(26.2)	211.0	(100.6)	323.5	(76.4)
Breakout	M12D0	69.1	(14.9)	96.1	(11.2)	55.2	(37.2)
Space Invaders	M1D0	926.1	(56.6)	1033.5	(89.7)	979.7	(39.8)
	M1D1	799.4	(52.5)	920.0	(83.5)	906.9	(56.5)
	M9D0	574.1	(37.0)	583.0	(17.5)	567.7	(40.1)

Takeaways

- DQN struggles to generalize to even slight variations of the underlying MDP
- Regularization methods designed to prevent deep neural networks from overfitting improve generalization and adaptability of DQN