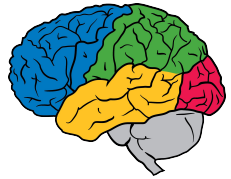
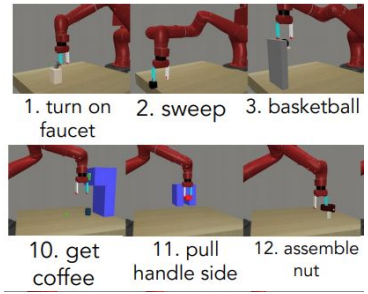
Google Research



Generalization Mysteries in Reinforcement Learning

xingyousong@, Yiding Jiang, neyshabur@, stephentu@,
rishabhagarwal@, alexirpan@, yingjiemiao@, and many others

Example Benchmarks around RL “Generalization”



MetaWorld

MineRL

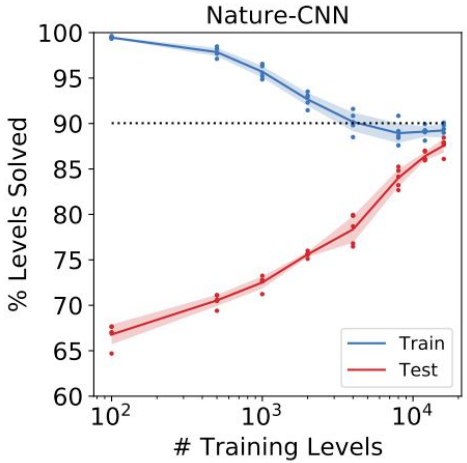
Gym Retro

ProcGen

Paper Bibliography in Appendix.

Our “Generalization” Definition

- **Zero-Shot:** Finite training set of MDPs, evaluate on test set of MDPs.
- **Distributional:** All MDP’s sampled from distribution
- **Overfitting:** Reward gap b/w train + test



K. Cobbe et al. *Quantifying Generalization in Reinforcement Learning* (ICML, 2019)

What Causes Overfitting in RL?

Sonic the Hedgehog - Gym Retro

- Sonic the HedgeHog (Gym Retro): Saliency (Red) suggests overfitting to background

Sonic the Hedgehog - Gym Retro

- Agent can train even if it only saw the timer!

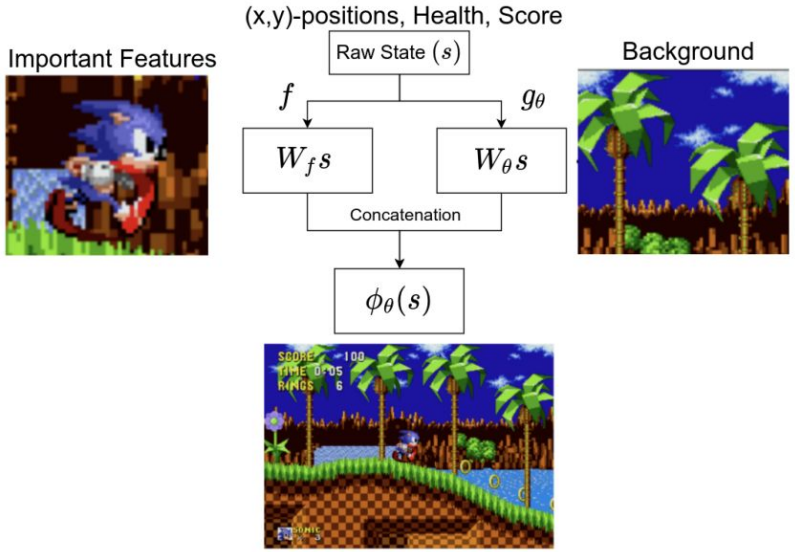
Settings	IMPALA	NatureCNN
NoScoreBoard	1250	1141
ScoreBoard	1130	1052

Test Rewards

Sonic in Action - Example Video

Observational Overfitting

- Any single MDP -> distribution of MDP's via constructing "observation functions"
- f-function stays the same
- g-function changes per level



Simplest Possible Benchmark: LQR

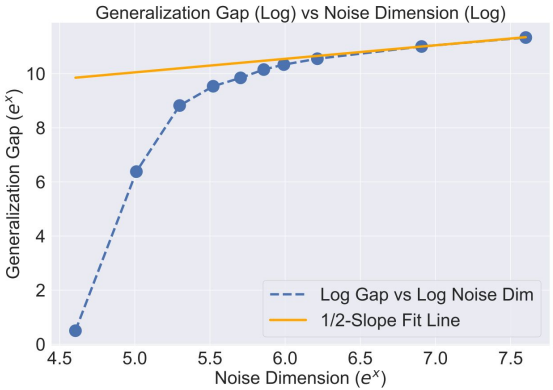
- Take any standard LQR

$$\begin{aligned} \text{minimize} \quad & E_{s_0 \sim \mathcal{D}} \left[\tfrac{1}{2} \sum_{t=0}^{\infty} s_t^T Q s_t + a_t^T R a_t \right], \\ \text{subject to} \quad & s_{t+1} = A s_t + B a_t, a_t = K o_t \end{aligned}$$

$$o_t = \begin{bmatrix} W_f \\ W_\theta \end{bmatrix} s_t$$

High-Dimensional Distractors

W_{theta} varies across each domain d causing overfitting.

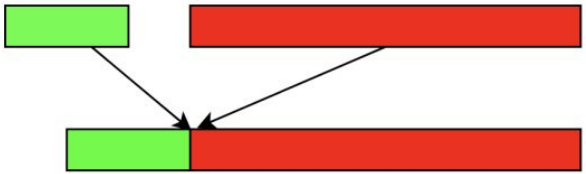


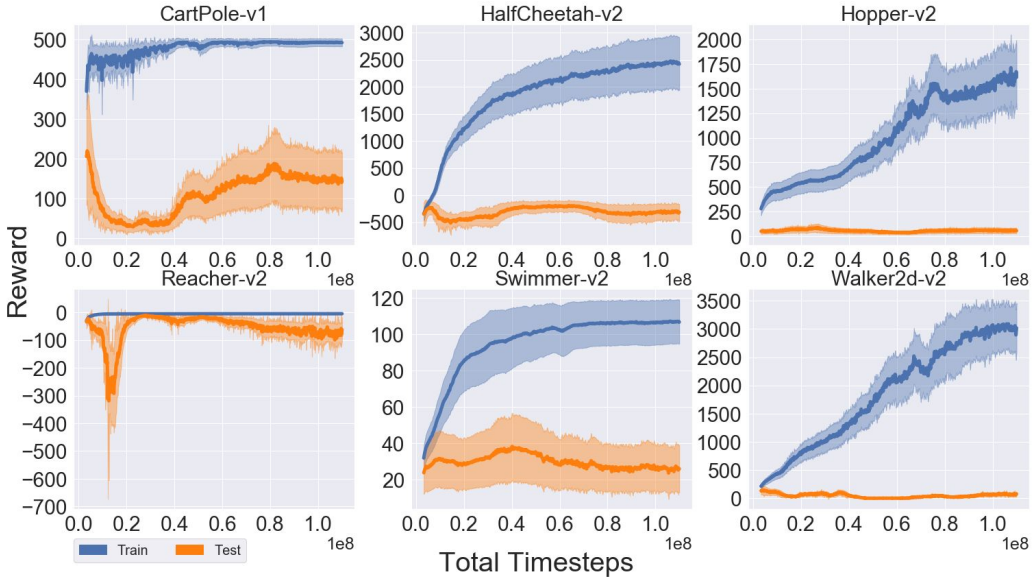
Figures adapted from Rishabh Agarwal.

Another Simple Benchmark: 1D State Mujoco

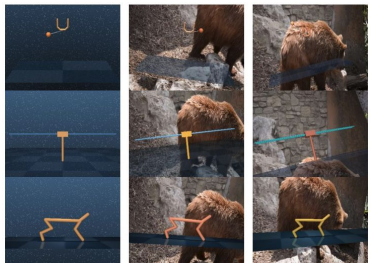
- Don’t need to drop 2D image backgrounds in DM-Control

$$o_t = \begin{bmatrix} W_f \\ W_\theta \end{bmatrix} s_t$$

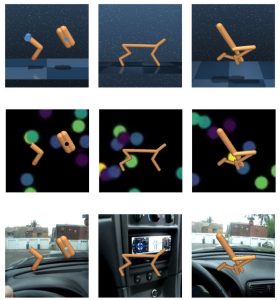




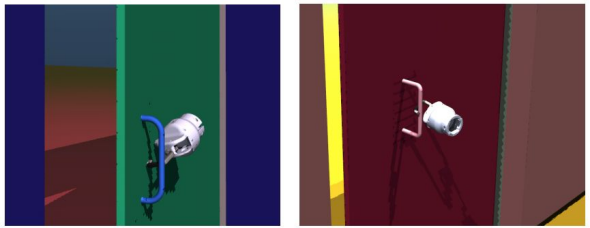
Explosion of Observational Overfitting Benchmarks



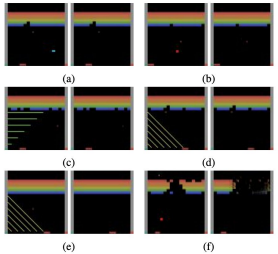
[Stone'21]



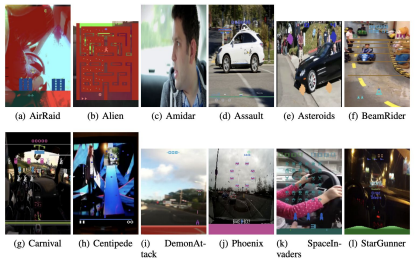
[Zhang'21]



[Sonar'20]



[Gamrian'18]



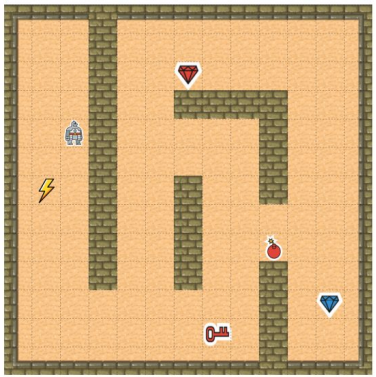
[Zhang'18]

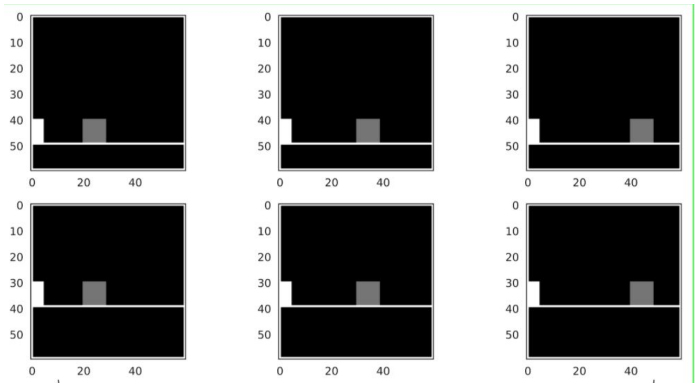
Paper Bibliography in Appendix.

What about other types of overfitting?

Why do Gridworlds/Non-Vision overfit?

- Maybe something temporal?
  - Agent is “expecting” something to occur in time?





C. Zhang, O. Vinyals, R. Munos, S. Bengio. *A Study on Overfitting in Deep Reinforcement Learning (2018).*
R. Tachet des Combes, P. Bachman, H. Seijen. *Learning Invariances for Policy Generalization* (ICLR Workshop, 2018)

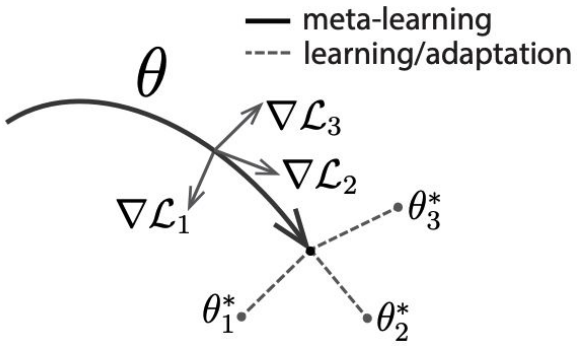
Opinion: We don’t know (no clear conceptual framework)

- Observational Overfitting isn’t specific to RL
- (Opinionated) Metrics of understanding
  - Edit specific parts of MDP to increase/decrease gen. gap
  - Clear ways to make benchmarks (empirical + theoretical)

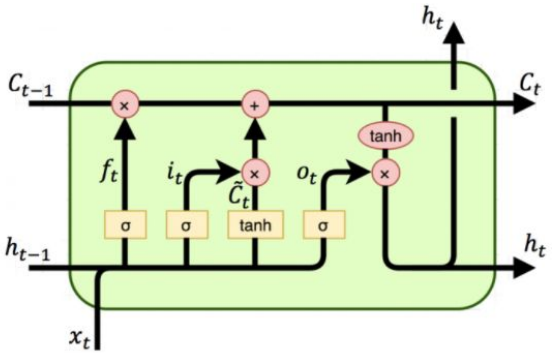
K. Xiao, L. Engstrom, A. Ilyas. *Noise or Signal: The Role of Image Backgrounds in Object Recognition* (2020).
M. Arjovsky, L. Bottou, I. Gulrajani, D. Lopez-Paz. *Invariant Risk Minimization* (2020)

Example: What is “Recurrent Overfitting”?

Is it a maze?



Is it MAML?



Is it an RNN?

What Affects Generalization?

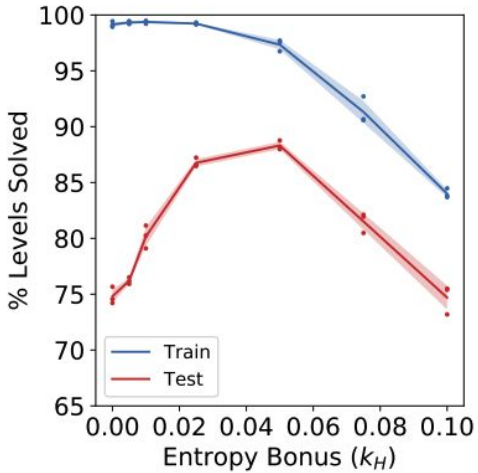
Explicit Regularization

- Invariant Representations
  - Invariant Representations for Reinforcement Learning without Reconstruction (2021)
  - Contrastive Behavioral Similarity Embeddings for Generalization in Reinforcement Learning (2021)
- Domain Randomization + Data Augmentation
  - Reinforcement Learning with Augmented Data (2020)
  - Automatic Data Augmentation for Generalization in Deep Reinforcement Learning (2020)
- Losses (L2 reg., dropout, etc.)
  - Quantifying Generalization in Reinforcement Learning (2019)
  - Generalization and Regularization in DQN (2018)

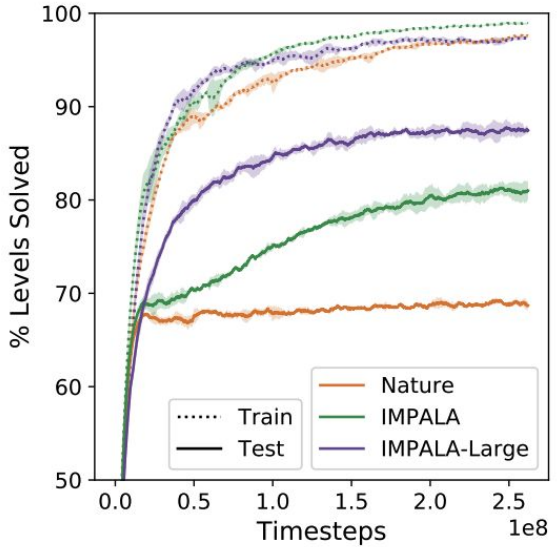
Paper Bibliography in Appendix.

Implicit Regularization - “Accidental” Factors

- Hyperparameters
  - Entropy matters alot
  - Gamma matters in other works



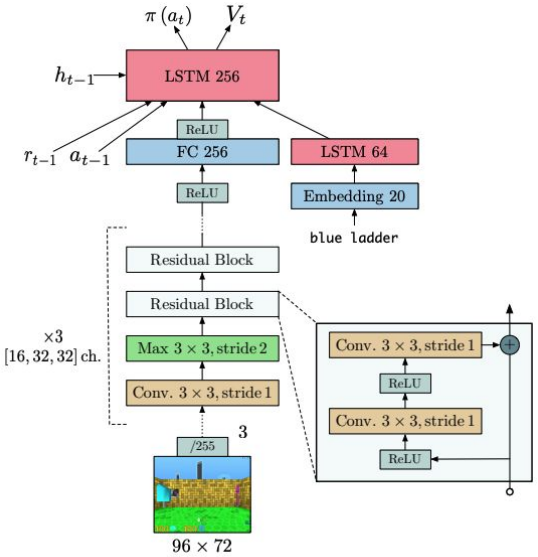
- Architectures
  - IMPALA-Large > IMPALA > NatureCNN



K. Cobbe et al. *Quantifying Generalization in Reinforcement Learning* (ICML, 2019)

Implicit Regularization - Architectures

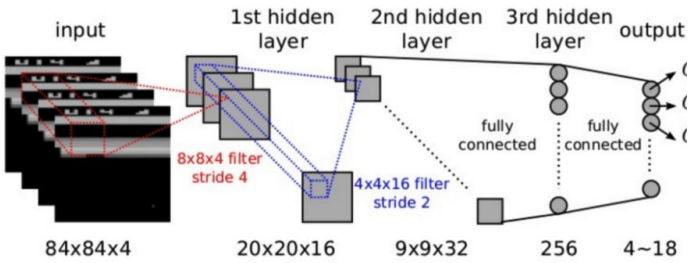
Why is:



IMPALA-CNN

>

?

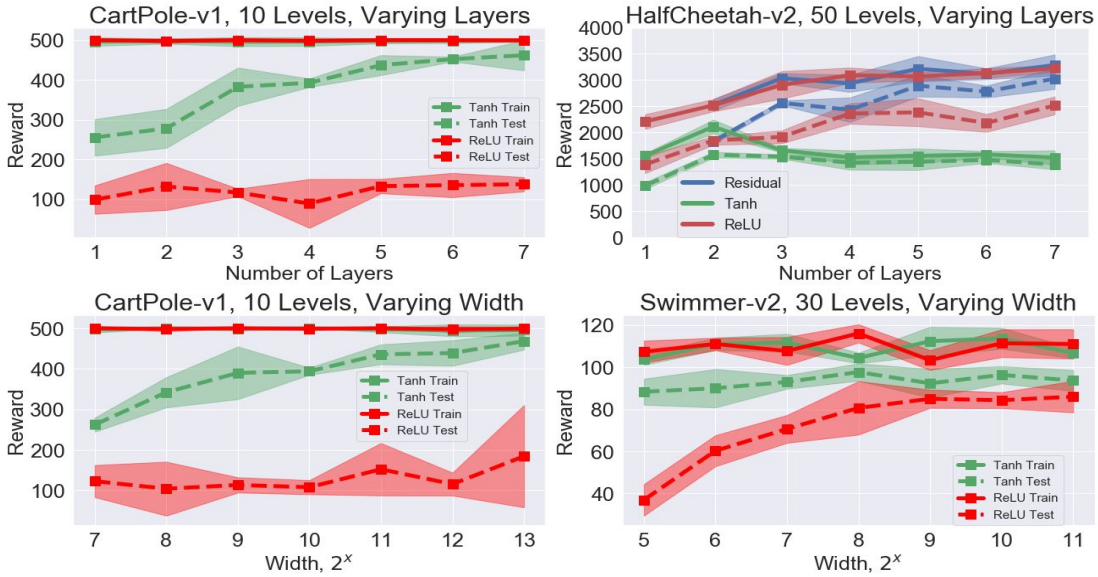


NatureCNN

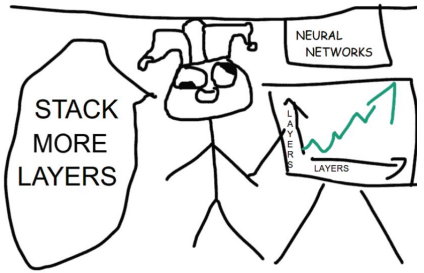
L. Espeholt et al. *IMPALA: Scalable Distributed Deep-RL with Importance Weighted Actor-Learner Architectures* (2018)
V. Minh et al. *Playing Atari with Deep Reinforcement Learning* (2013)

Implicit Regularization - Architecture

- Residual Layers, Overparameterization, Nonlinearities



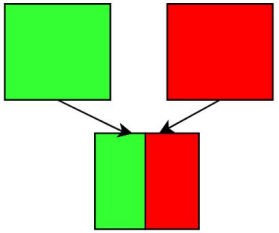
1D State Mujoco Task

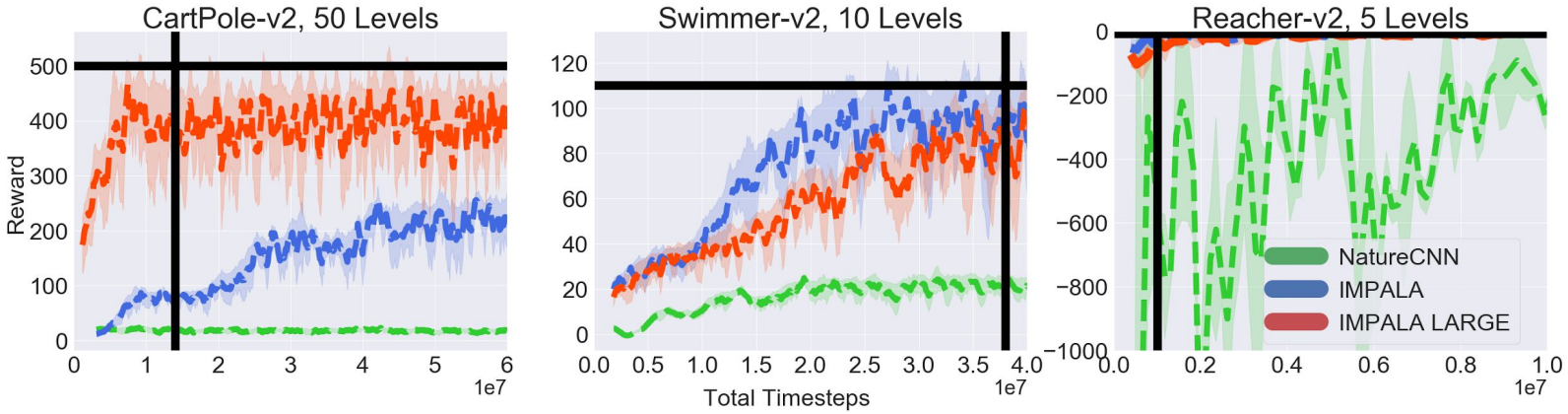


He was right all along.

Implicit Regularization - Architecture

Ranking also occurs if I make a **2D State Mujoco Task.**

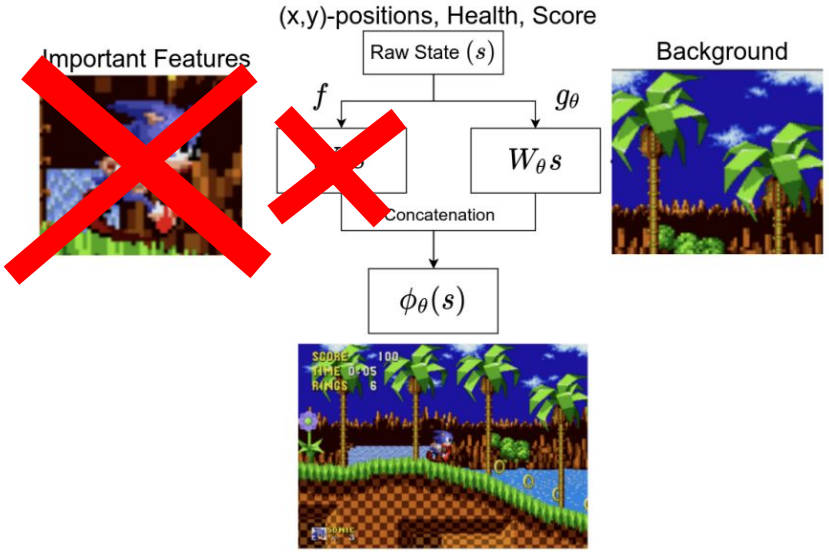




Implicit Regularization - Architecture Memorization

Which memorizes the most?

- NatureCNN (600K Params)
- IMPALA (622K Params)
- IMPALA-LARGE (823K Params)



More parameters = More memorization?

Wrong!

Implicit Regularization - How Strong is it?

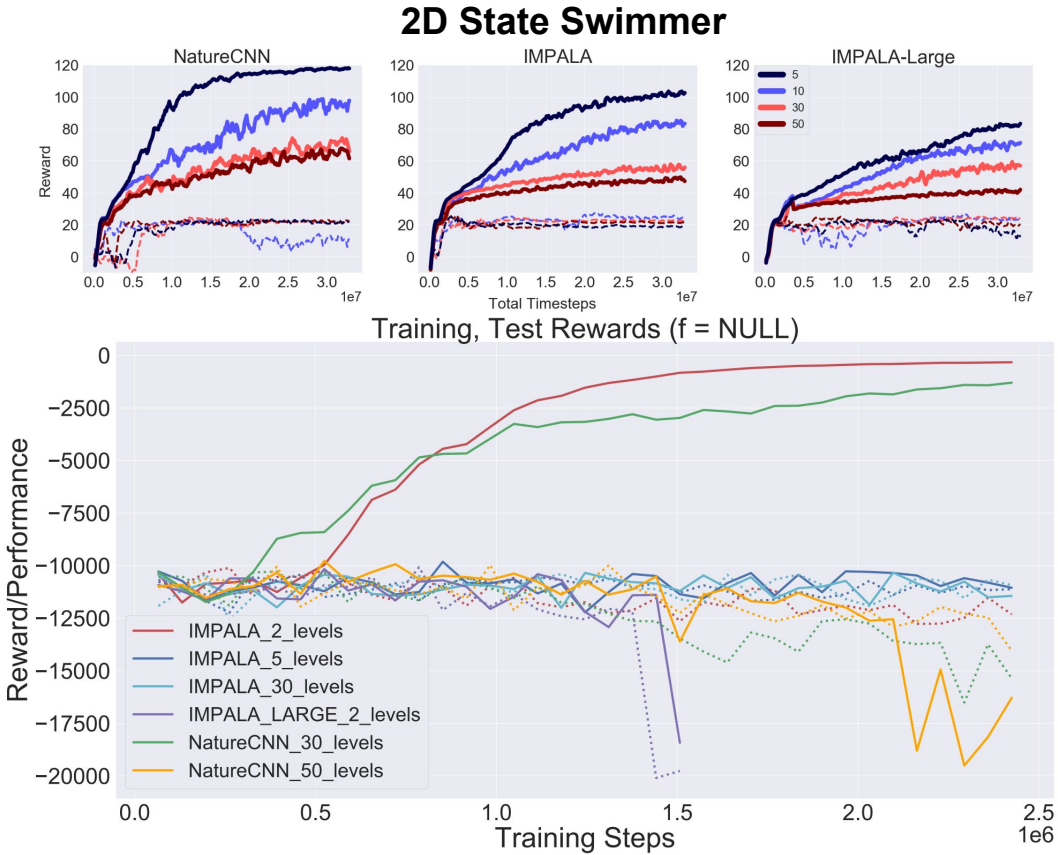
IMPALA-LARGE memorizes the **least.**

Implicit Regularization is **VERY strong.**

Memorization Capacities:

NatureCNN: 30-50
IMPALA: 2-5
IMPALA-LARGE: <2

2d State LQR



How to Predict Generalization?

How do you know beforehand that you've overfitted?

AI Camera Ruins Soccer Game For Fans After Mistaking Referee's Bald Head For Ball

69.7K SHARES | Share on Facebook | Share on Twitter

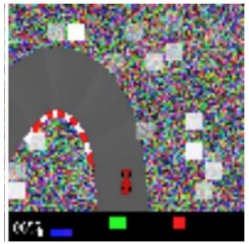
Clever Hans

Horse

Clever Hans was a horse that was claimed to have performed arithmetic and other intellectual tasks. After a formal investigation in 1907, psychologist Oskar Pfungst demonstrated that the horse was not actually performing these mental tasks, but was watching the reactions of his trainer. Wikipedia

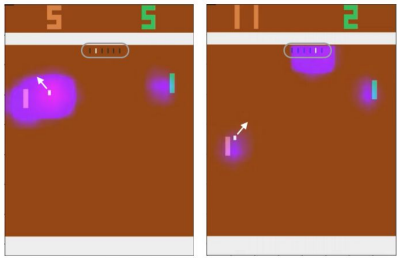
Human-in-the-loop methods

Saliency



Tang’20

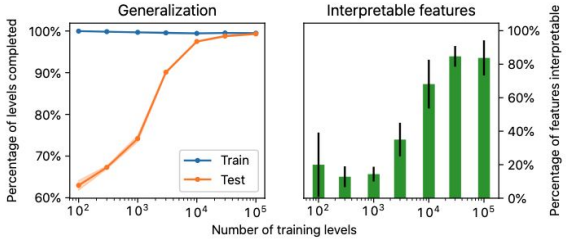
Song’20



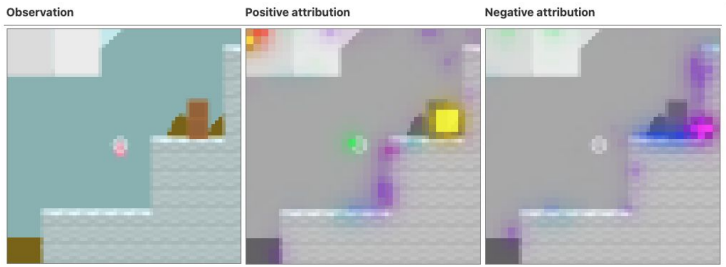
(a) Pong: control (b) Pong: overfit

Greydanus’17

Interpretable Features



Hilton’20



Hilton’20

Paper Bibliography in Appendix.

Systematic ways

Inspiration: Use knowledge from **supervised learning.**

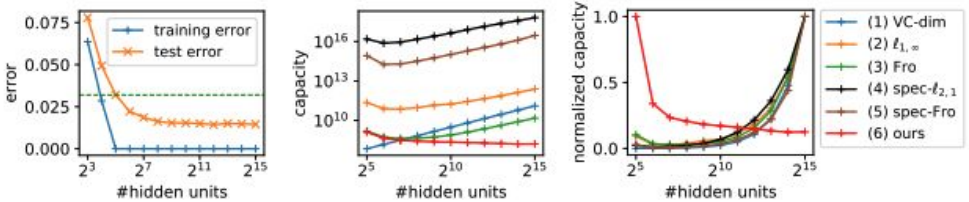
Generalization Bounds: $L_0(f) \leq \hat{L}_\gamma(f) + 2\frac{\mathcal{R}_m(\mathcal{F})}{\gamma} + \sqrt{\frac{8\ln(2/\delta)}{m}}$

Rademacher/Lipschitz/Network Weights: $\mathcal{R}_m(\mathcal{F}) \leq \sqrt{\frac{4^d \ln(n_{\text{in}}) \prod_{i=1}^d \|W_i\|_{1,\infty}^2 \max_{\mathbf{x}\in\mathcal{S}} \|\mathbf{x}\|_\infty}{m}}$

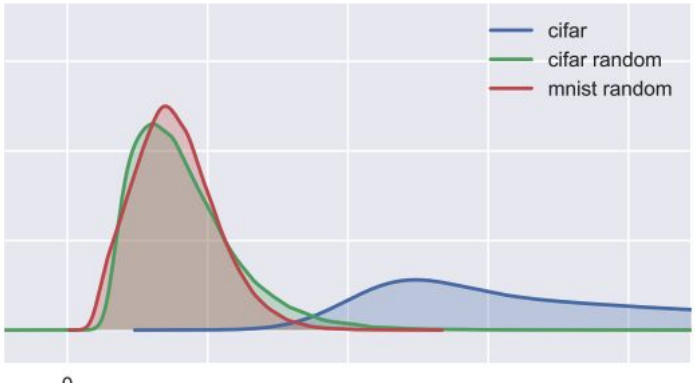
Margin Distributions: $(x, y) \mapsto \frac{F_{\mathcal{A}}(x)_y - \max_{i\neq y} F_{\mathcal{A}}(x)_i}{R_{\mathcal{A}}\|X\|_2/n},$

Systematic ways?

Generalization Methods have great success in SL:



Generalization Gap Bounds



Margin Distributions

What about RL?

B. Neyshabur et al. *Towards Understanding the Role of Over-Parametrization in Generalization of Neural Networks* (ICLR, 2019)
P. Bartlett et al. *Spectrally-normalized margin bounds for neural networks* (NeurIPS, 2017)

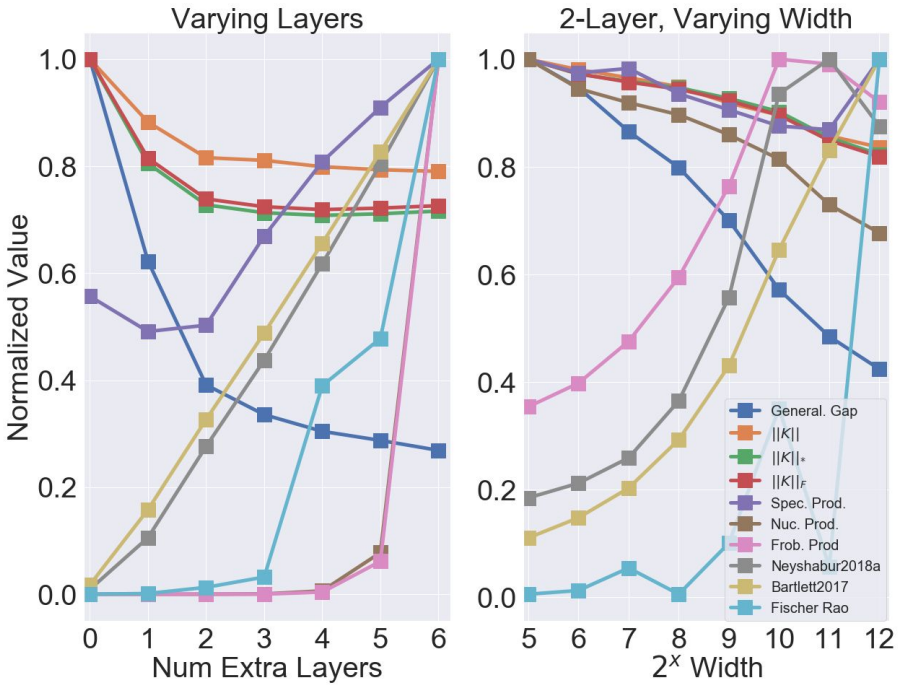
Systematic ways...?

Simple Case: 1D Projected LQR

$$K = K_1 K_2 ... K_j$$

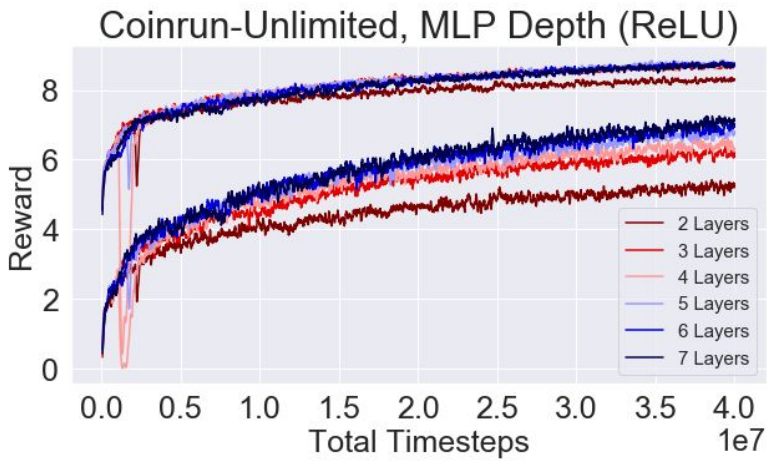
As a function of overparameterization:

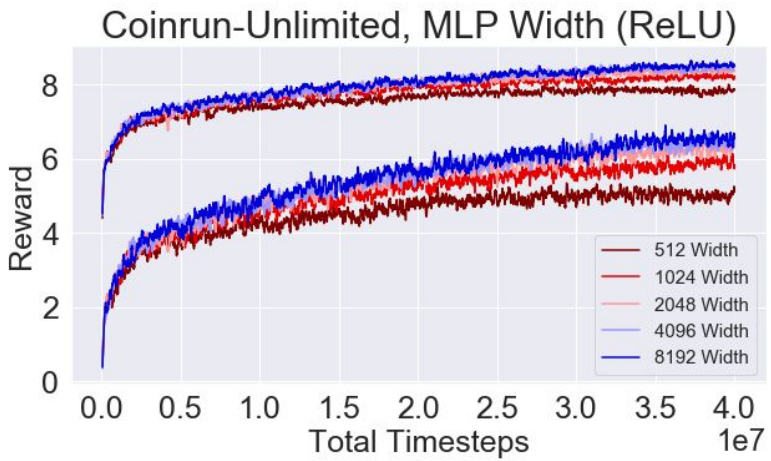
- **Generalization Gap** Decreases
- **E2E Policy Norm** Decreases
- **Successful SL Bounds**...Increase??



What about real RL?

- Overparameterization helps in CoinRun.

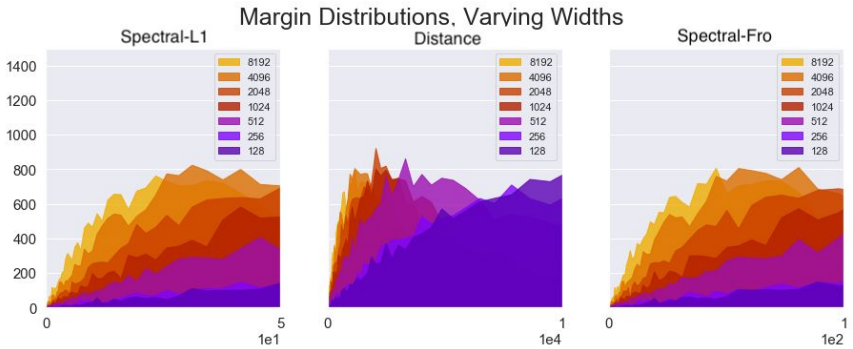


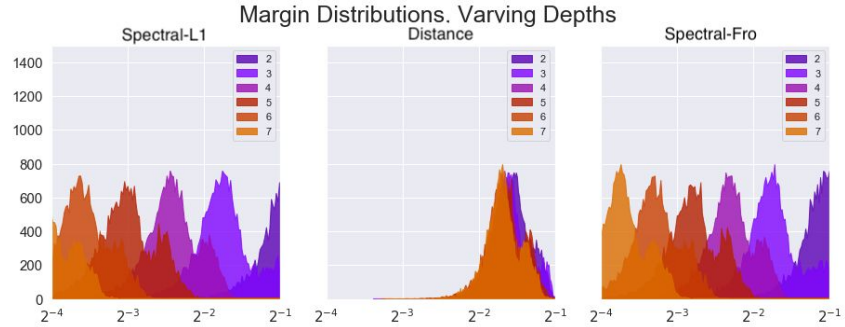


Margin Distributions don’t say anything

$$(x, y) \mapsto \frac{F_{\mathcal{A}}(x)_y - \max_{i \neq y} F_{\mathcal{A}}(x)_i}{R_{\mathcal{A}} \|X\|_2 / n},$$

- Use (state, action) from replay buffer as (x,y)
- **Expect:** Increasing parameterization, distribution shifts right
- **Actual:** Increasing parameterization, distribution shifts left
- Denominator (weight norms) too strong **:(**





Key Questions

- What causes overfitting in RL, besides observational overfitting?
  - What is a good framework to study this?
- How do you explain effects of implicit regularization?
  - Neural Tangent Kernels in RL?
- How do you predict generalization without explicitly testing on eval env?
  - **Practical** Generalization Theories in RL?

Thank you!

Feel free to reach out!

Appendix

S. Gamrian et al. *Transfer Learning for Related Reinforcement Learning Tasks via Image-to-Image Translation* (ICML, 2019).
A. Zhang et al. *Natural Environment Benchmarks for Reinforcement Learning* (2018)
K. Cobbe et al. *Quantifying Generalization in Reinforcement Learning* (ICML, 2019)
A. Nichol et al. *Gotta Learn Fast: A New Benchmark for Generalization in RL* (2018)
P. Bartlett et al. *Spectrally-normalized margin bounds for neural networks* (NeurIPS, 2017)
B. Neyshabur et al. *Towards Understanding the Role of Over-Parametrization in Generalization of Neural Networks* (ICLR, 2019)
Y. Tang et al. *Neuroevolution of Self-Interpretable Agents* (GECCO, 2020)
X. Song et al. *Observational Overfitting in Reinforcement Learning* (ICLR 2020)
T. Yu et al. *Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement Learning* (CoRL 2019)
W. Guss et al. *The MineRL 2020 Competition on Sample Efficient Reinforcement Learning using Human Priors* (NeurIPS 2020 Competition)
K. Cobbe et al. *Leveraging Procedural Generation to Benchmark Reinforcement Learning* (ICML 2020)
K. Cobbe et al. *Quantifying Generalization in Reinforcement Learning* (ICML 2019)
A. Stone et al. *The Distracting Control Suite -- A Challenging Benchmark for Reinforcement Learning from Pixels* (2021)
A. Zhang et al. *Invariant Representations for Reinforcement Learning without Reconstruction* (ICLR 2021)
J. Hilton et al. *Understanding RL Vision* (Distill.pub, 2020)
A. Sonar et al. *Invariant Policy Optimization: Towards Stronger Generalization in Reinforcement Learning* (2020)
S. Greydanus et al. *Visualizing and Understanding Atari Agents* (ICML 2018)
C. Zhang et al. *A Study on Overfitting in Deep Reinforcement Learning* (2018).
R. Tachet des Combes, et al. *Learning Invariances for Policy Generalization* (ICLR Workshop, 2018)