Transfer Learning by Modeling A Distribution over Policies

Disha Shrivastava*[,1,2], Eeshan Gunesh Dhekane*[,1,2], Riashat Islam[1,3]

Mila[1], Université de Montréal[2], McGill University[3]

Introduction

- **VFunc: Deep Bayesian Learning** framework [1]:
  - Model joint distribution $p(f,z)$ over **function space** using latent variable z
  - **Efficient function sampling** by marginalizing z
  - Training **entropy-regularized objective** with **variational lower bound**
- **Transfer Learning**: ***Modeling a distribution over policies*** can lead to improved exploration in the target environment

Methodology

- **Maximization Objective***

$$\mathcal{F} = \mathbb{E}_{f\sim p(f)}[R(D|f) + \log \bar{p}(f)] + H(f)$$

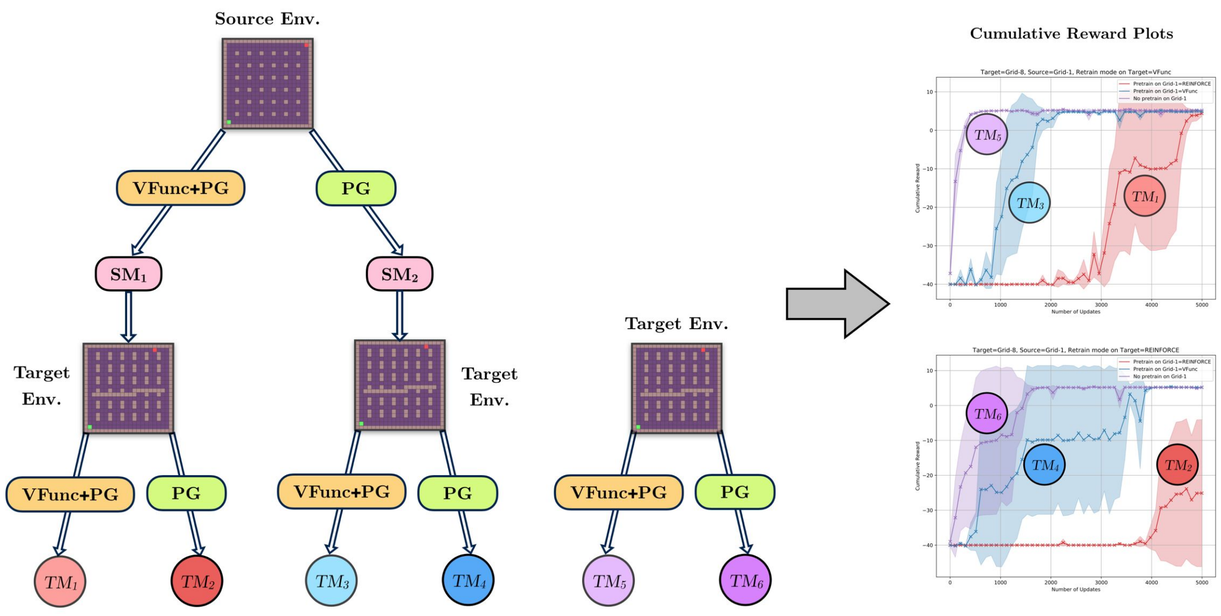
Entropy Term; Reward Term; Policy Distribution Prior

- **Variational Lower Bound on $H(f)$***

$$H(f) \geq H(z) + \mathop{\mathbb{E}}_{(f,z)\sim p(f,z)}[\log q(z|f)] + H(f|z)$$

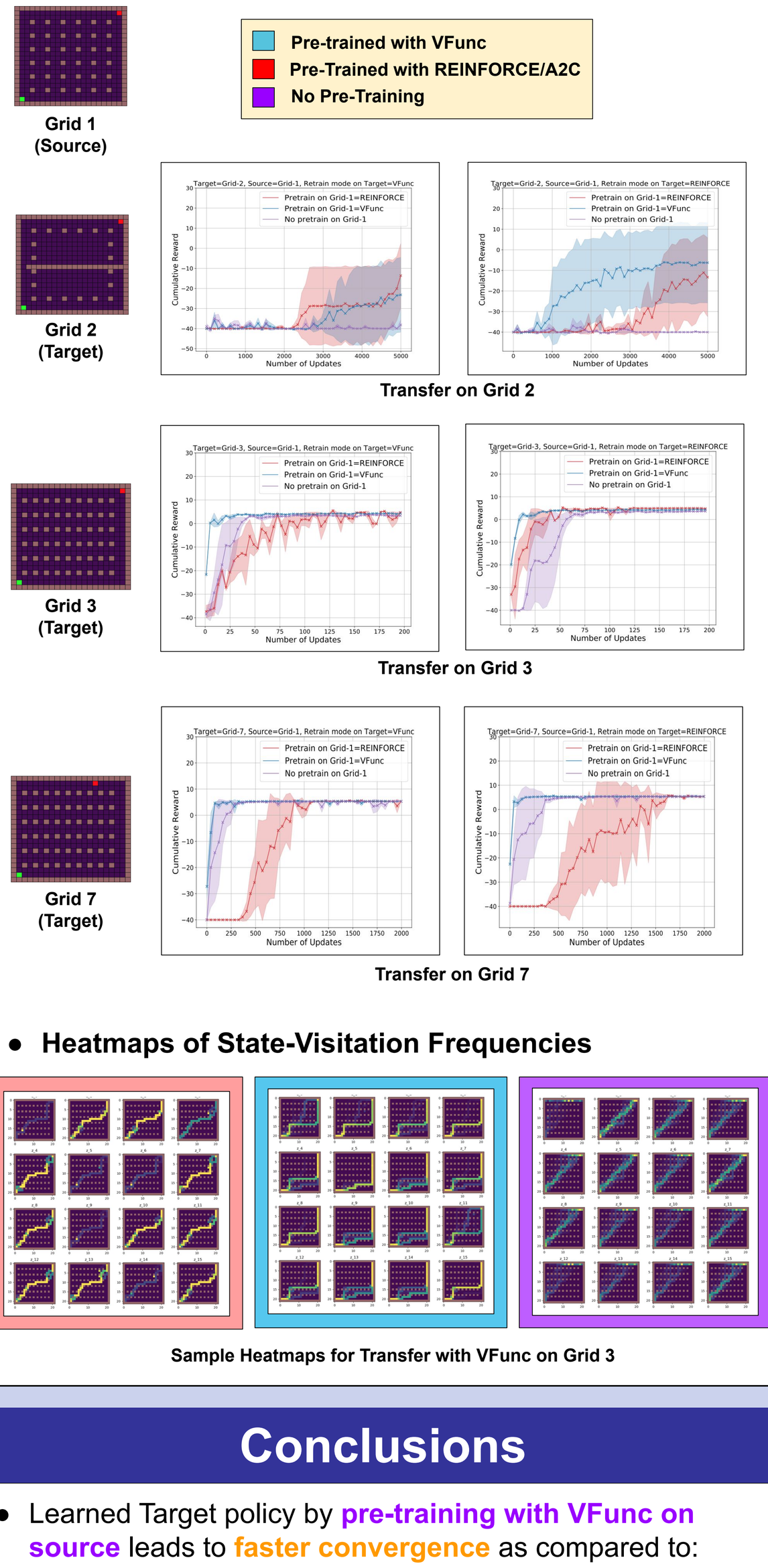
* Taken from [1]

- **Schematics of Transfer Learning with VFunc**



Results: GridWorld

- **Transfer Results on GridWorld [3] [REINFORCE]**

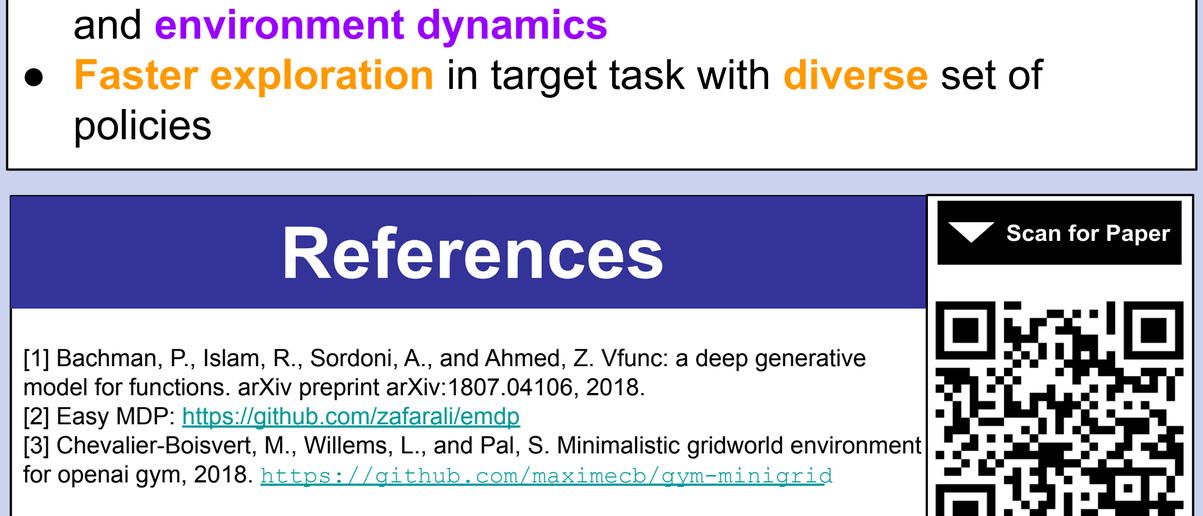


Transfer on Grid 2

Transfer on Grid 3

Transfer on Grid 7

- **Heatmaps of State-Visitation Frequencies**



Sample Heatmaps for Transfer with VFunc on Grid 3

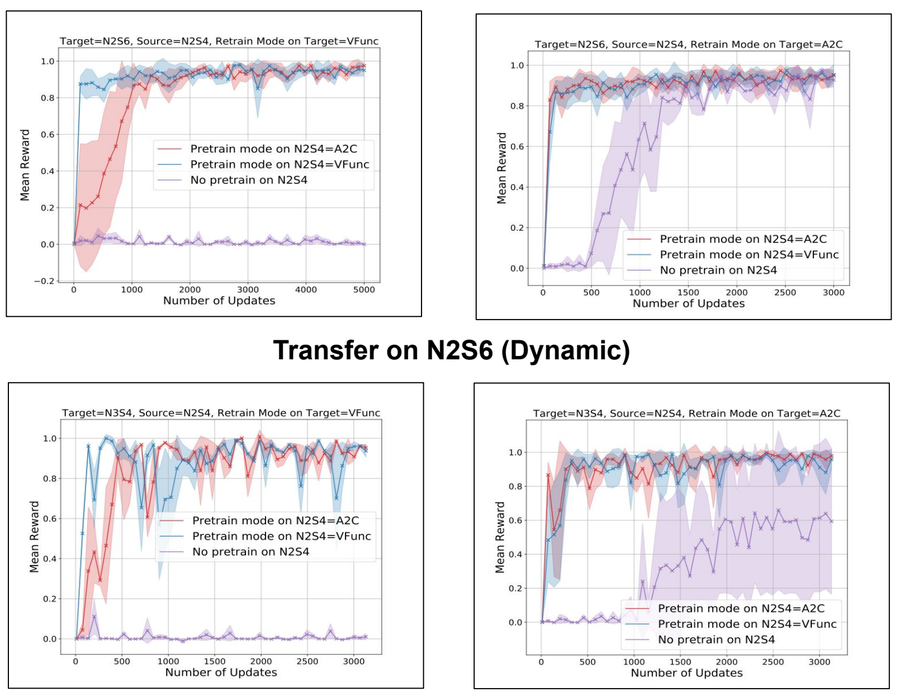
Results: MiniGrid

- **Multi-Room Envs [2]:**

Static
Dynamic

Name	# Rooms	Room Size	Type
N2S4	2	4	*Source*
N2S6	2	6	*Target*
N3S4	3	4	*Target*

- **Transfer Results on MiniGrid [A2C]:**



Transfer on N2S6 (Dynamic)

Transfer on N3S4 (Static)

Conclusions

- Learned Target policy by **pre-training with VFunc on source** leads to **faster convergence** as compared to:
  - **Pre-training with other PG algorithms on source**
  - **Training from scratch on target**
- Difference more apparent with **increasing difficulty level** and **environment dynamics**
- **Faster exploration** in target task with **diverse** set of policies

References

[1] Bachman, P., Islam, R., Sordoni, A., and Ahmed, Z. Vfunc: a deep generative model for functions. arXiv preprint arXiv:1807.04106, 2018.
[2] Easy MDP: https://github.com/zafarali/emdp
[3] Chevalier-Boisvert, M., Willems, L., and Pal, S. Minimalistic gridworld environment for openai gym, 2018. https://github.com/maximecb/gym-minigrid

Scan for Paper