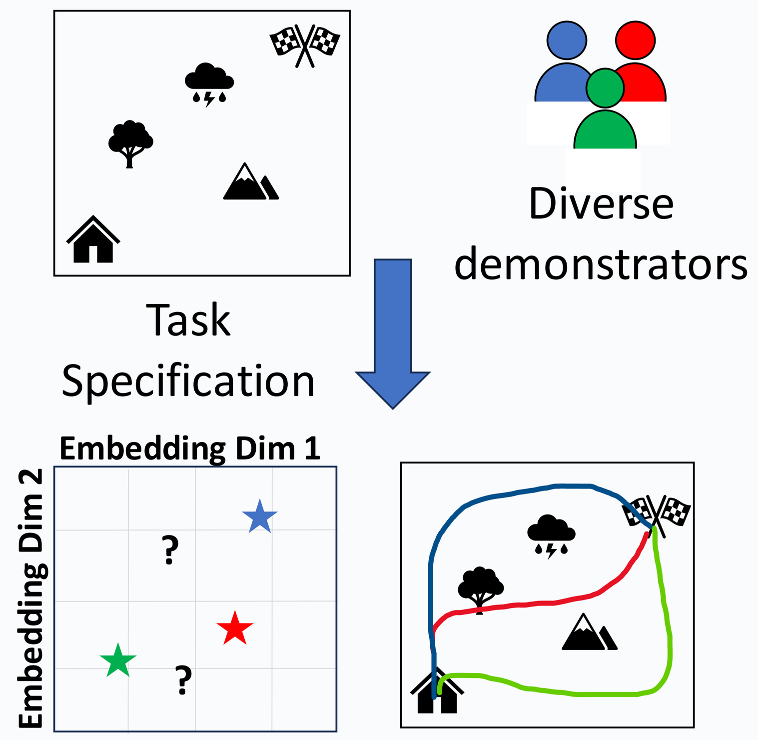
Generalized Behavior Learning from Diverse Demonstrations

Varshith Sreeramdass, Rohan Paleja, Letian Chen, Sanne van Waveren, Matthew Gombolay
{vsreeramdass, rpaleja3, letian.chen, sanne}@gatech.edu, matthew.gombolay@cc.gatech.edu

Introduction

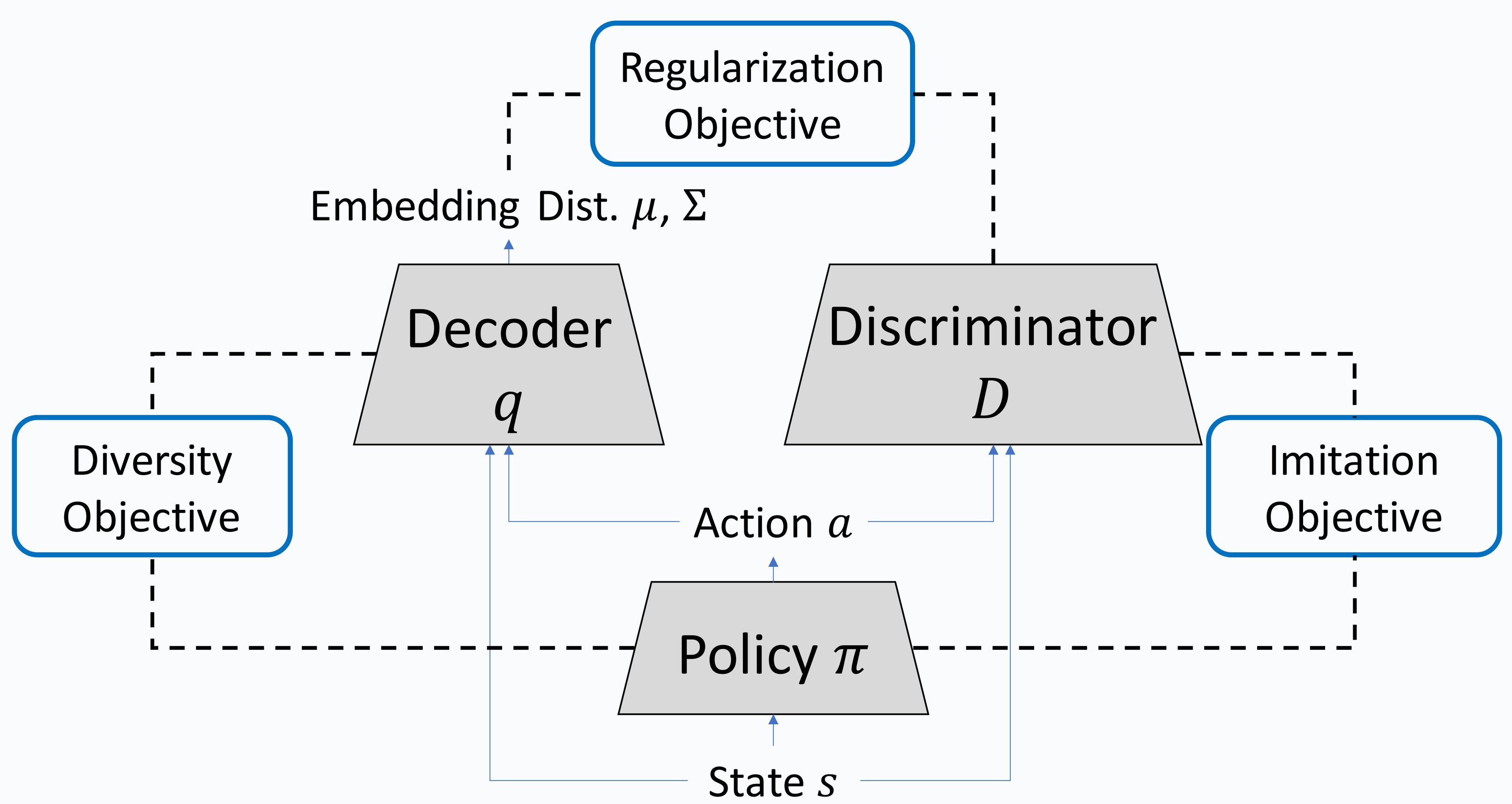
- The goal of Imitation Learning (IL) is to learn robot control policies from demonstrations.
- Demonstrations from humans can be diverse, even for well-specified tasks, due to hidden factors that are often continuous.
- **Embedding** and **generating** diverse behaviors can help personalize and improve human-robot collaboration.



Our goal is to develop a **generalizable latent space** for generating task-accomplishing behaviors that generalize to unseen factors.

Approach

- Prior work [1] balances imitation and diversity objectives. Diversity is formulated as Mutual Information (MI), where the decoder (posterior) q, influences the nature of behavior diversity.
- Naively using MI can lead to insufficient or arbitrary diversity [2].



Overview of a general multimodal IL framework.

- Spectral Normalization (SN) [2] enforces Lipschitz constraints to promote uniform variation but can be misaligned with the task.

$$||\mu_{q(\cdot\mid s,a)} - \mu_{q(\cdot\mid s',a')}|| \leq ||s - s'||$$

Constraints with Spectral Normalization

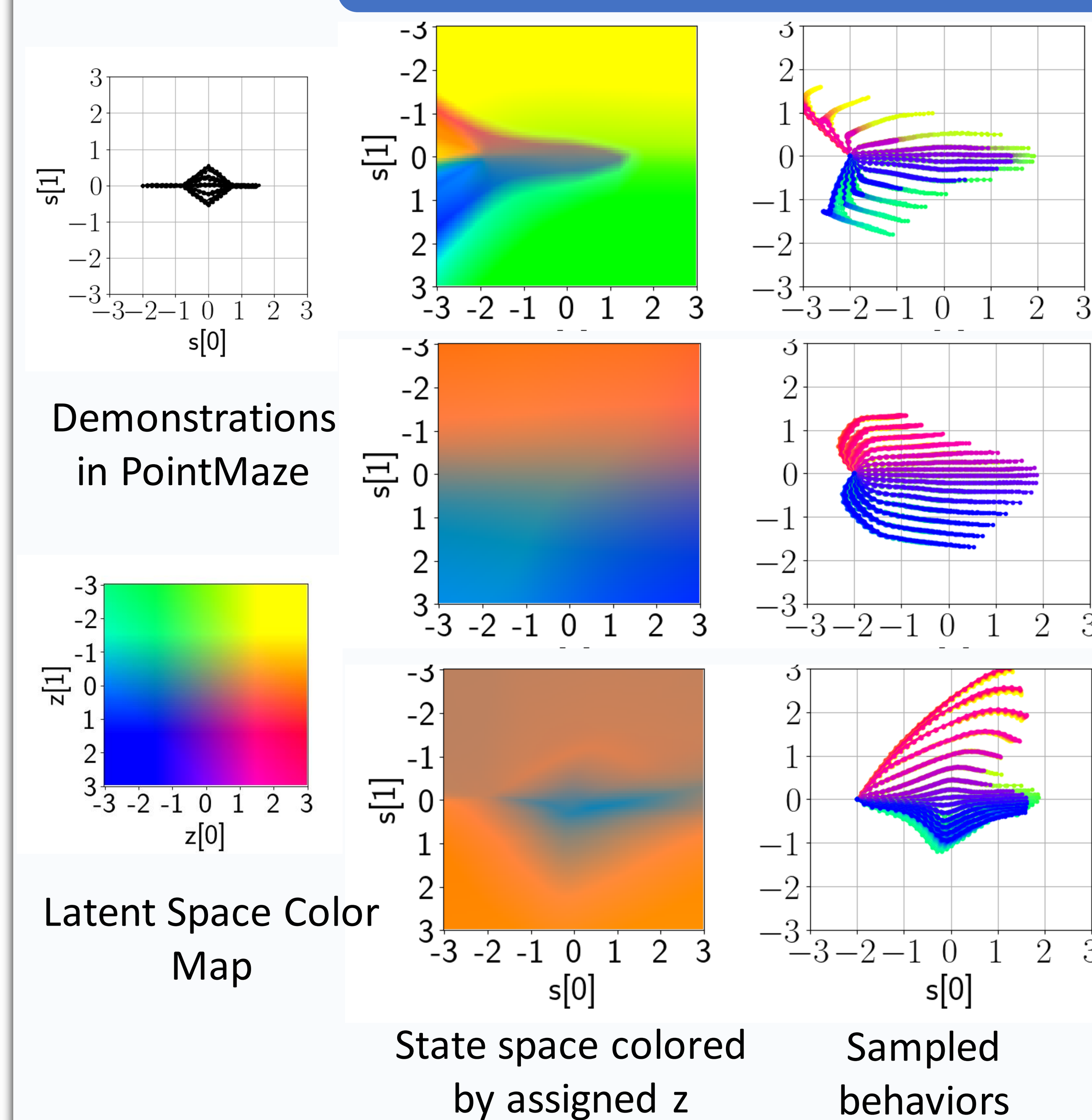
Our method **Guided Strategy Discovery (GSD)** uses a novel regularization scheme where constraints are locally modulated, driving the diversity to be task-relevant.

$$D(s,a) = \sigma(\lambda_S \cdot f(s,a) + b)$$

$$||\mu_{q(\cdot\mid s,a)} - \mu_{q(\cdot\mid s',a')}|| \leq f(s,a) \cdot ||s - s'||$$

Discriminator form and constraints with GSD

Visualization



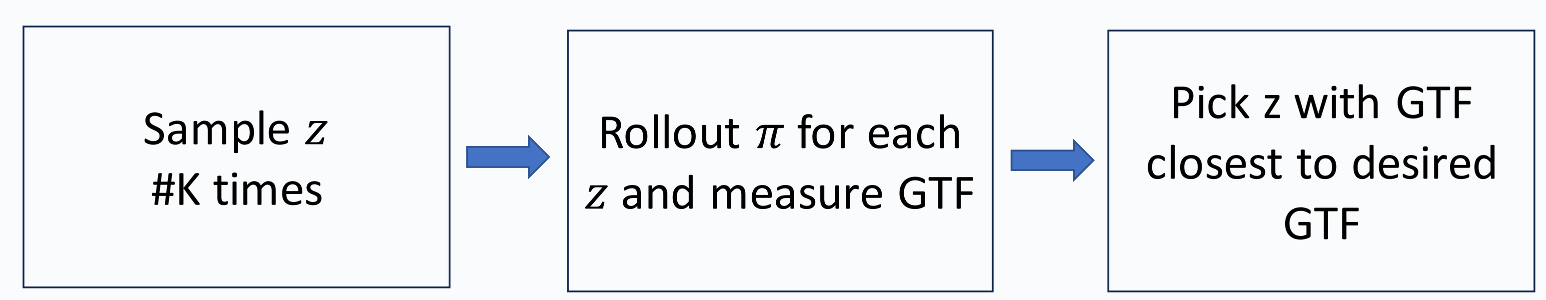
No regularization leads to arbitrary or insufficient diversity.

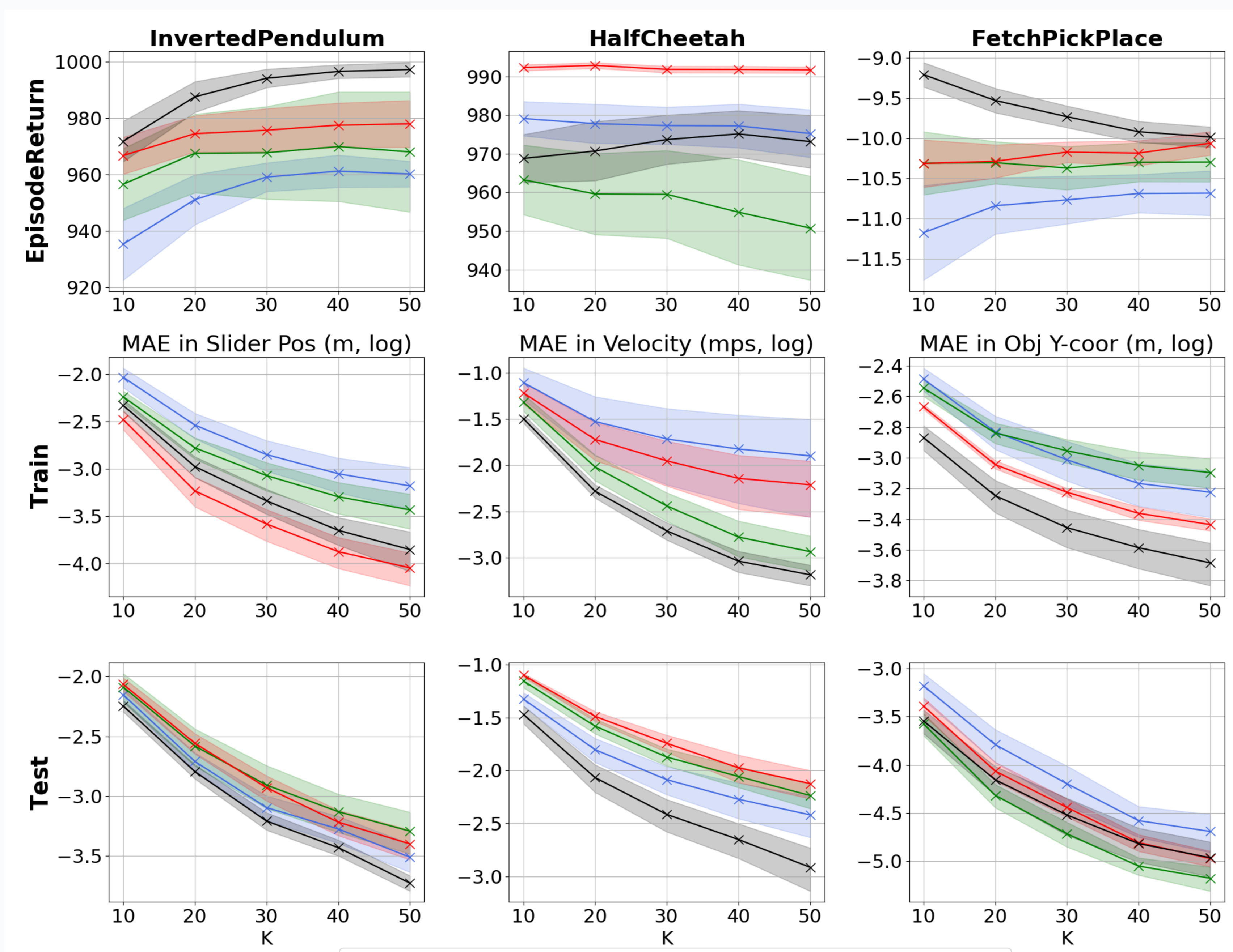
Spectral Normalization leads to uniform variation, but behaviors diverge from the demonstrated goal.

GSD (ours) leads to diverse behaviors that consistently progress towards the demonstrated goal.

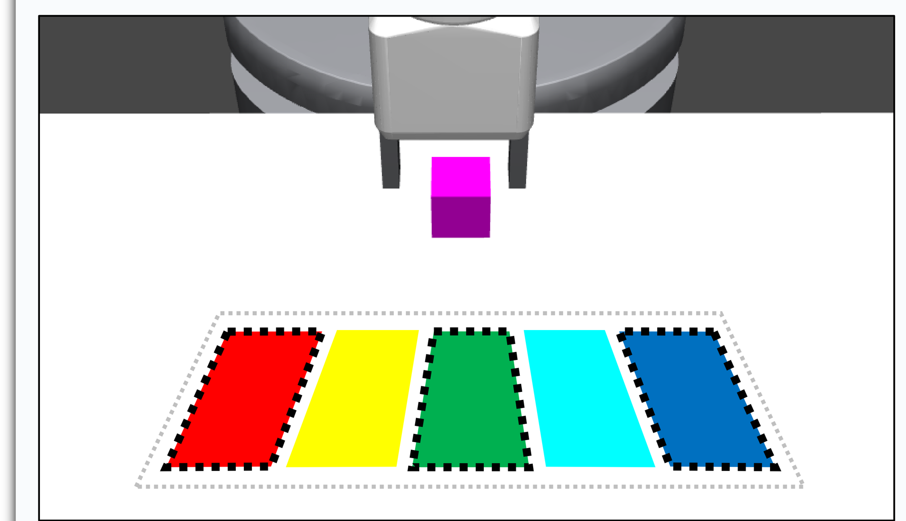
Evaluation

We consider domains and demonstrations with known and measurable ground truth factors (GTFs). Generalization is measured with error between desired values and those of generated behaviors.



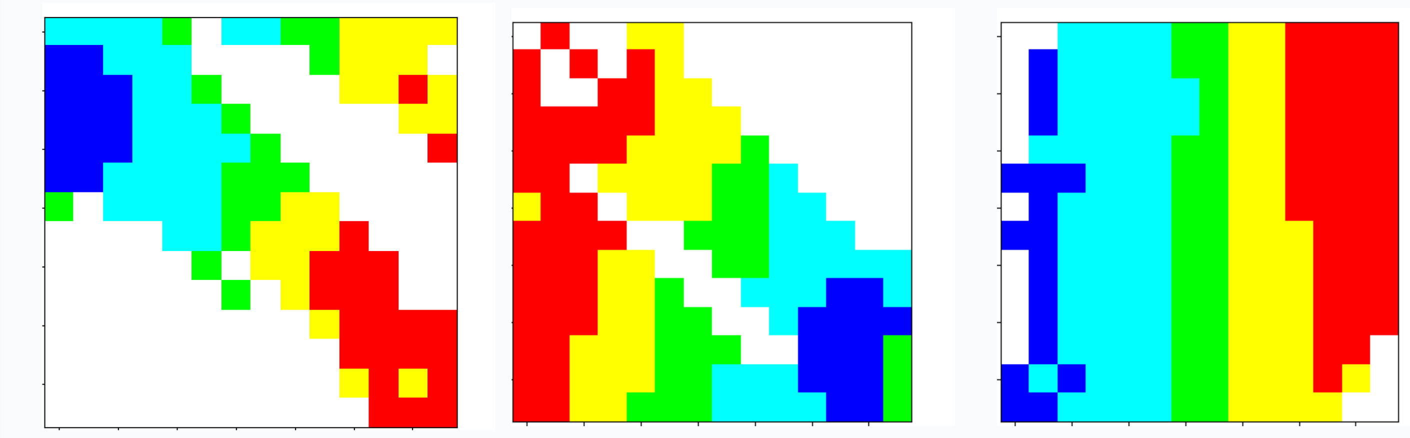


Qualitative Analysis



FetchPickPlace: Different locations at which object can be placed are visualized with colors. Dotted boundaries indicate training demonstrations.

Latent spaces visualized with locations at which objects are placed by the generated behaviors. White regions indicate failure to place in the relevant regions (below).



Behaviors with **no regularization** and **SN** either fail to accomplish the task or represent the relevant regions disproportionately.

Behaviors with **GSD (ours)** accomplish the task and represent all relevant regions well.

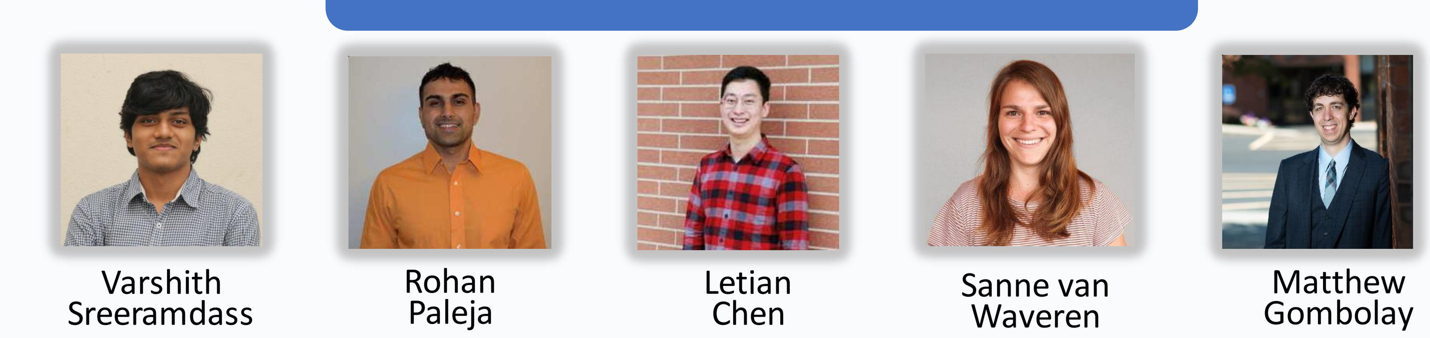
Conclusion

We investigate the use of MI in multimodal IL and propose a novel regularization scheme that improves generalization performance across three continuous control domains.

Future Work

- Explore latent conditioned discriminators to allow generalization beyond train demonstrations.
- Learn from real human demonstrations and evaluate generalization subjectively.

Authors



Acknowledgements & References

This work was supported by NSF grant CPS-2219755, ONR grant N00014-22-1-2834, MIT Lincoln Laboratory grant FA8702-15-D-0001, and a grant from Ford Motor Company.

[1] Li, Y., et. al.. (2017). Infogail: Interpretable imitation learning from visual demonstrations. *Advances in NeurIPS, 30*.
[2] Park, S., et. al. (2021, October). Lipschitz-constrained unsupervised skill discovery. In *ICRL*.