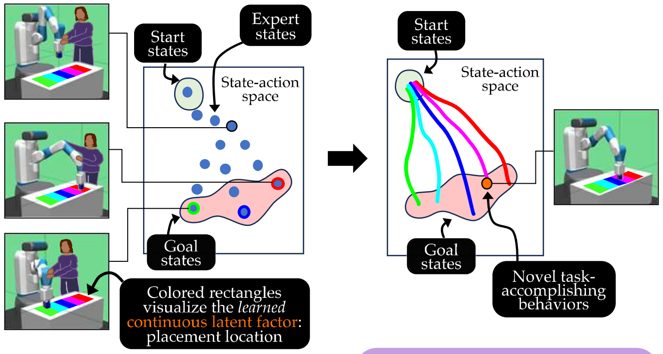
Generalized Behavior Learning from Diverse Demonstrations

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

1. Introduction



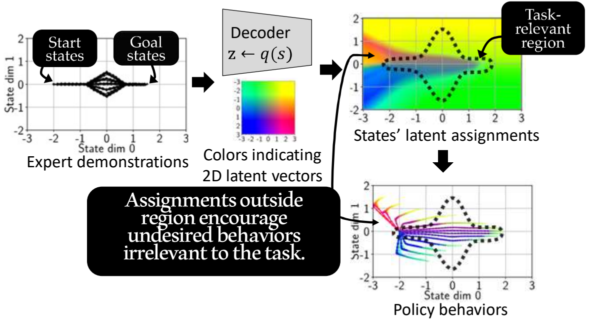
Colored rectangles visualize the *learned* continuous latent factor: placement location

- Humans exhibit natural diversity in their demonstrations.
- Learning diverse policies help adapt to env. changes or other agents.

MOTIVATION
How can we utilize diverse demos to generate novel task-accomplishing policies?

2. Prior work and Limitations

InfoGAIL [1] (IG) optimizes mutual information using a decoder to encourage coverage of diverse demos.



Expert demonstrations

Colors indicating 2D latent vectors

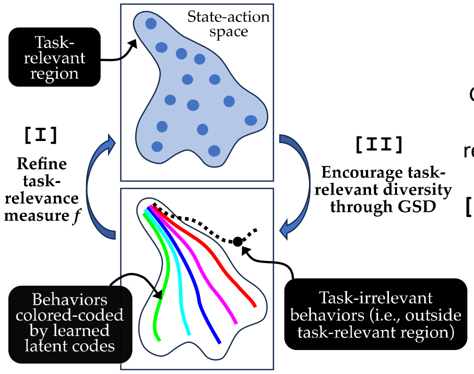
States' latent assignments

Policy behaviors

Assignments outside region encourage undesired behaviors irrelevant to the task.

KEY INSIGHT
Restricting latent assignments to relevant regions can produce diverse task-accomplishing behaviors.

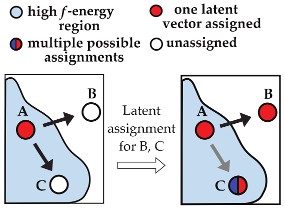
3. Our Approach: Guided Strategy Discovery (GSD)


Task-relevant region
State-action space
[I]
Refine task-relevance measure f
[II]
Encourage task-relevant diversity through GSD

Behaviors colored-coded by learned latent codes

Task-irrelevant behaviors (i.e., outside task-relevant region)

Overview:
GSD performs in parallel:
[I] extraction of a task-relevance function $f(s, a)$ based on prior work [2],
[II] optimization of **task-relevant diversity**.


high f-energy region
one latent vector assigned
multiple possible assignments
unassigned
A
B
C
Latent assignment for B, C
A
B
C

CONTRIBUTION: TASK-RELEVANT DIVERSITY
Lipschitz-based local regularization discourages coverage outside relevant regions.

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

Constraint is enforced only for task-relevant predecessors (A) by scaling on both sides

Distinct assignment for task-irrelevant state-actions (B) is prevented by pulling it towards predecessor's (A) assignment

$\oplus$ denotes concatenation.

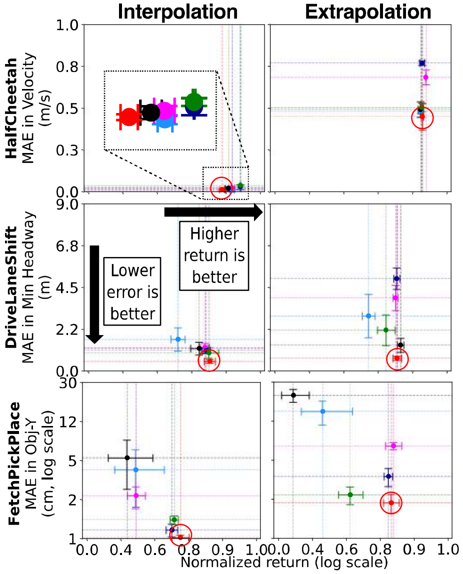
4. Quantitative Evaluation

Sample z, rollout π_z and measure true factor value. → Pick z with factor value closest to desired test value.

Interpolation
🟩🟥🟩🟥🟩

Extrapolation
🟥🟩 🟩🟥

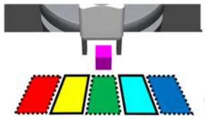
We consider known, measurable factors for the sake of evaluation.
For 1D factors, green indicates train and red indicates test intervals.



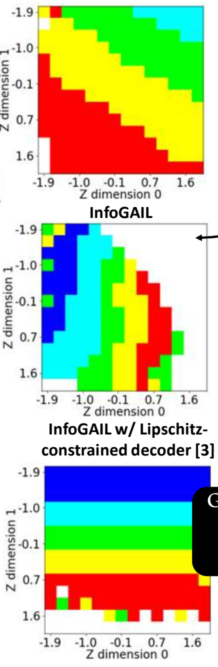
Performance Metrics:
X-Axis: Task - Episode rewards
Y-Axis: Diversity – Average error in desired factor value

Our approach outperforms baselines by ~21% **in recovery of novel behavior factors** while **matching task performance**.

5. Latent Spaces Visualization



FetchPickPlace: Object place locations shown with colors. Locations in train set are indicated by solid borders.



InfoGAIL

InfoGAIL w/ Lipschitz-constrained decoder [3]

GSD (Ours)

Baselines produce failed (empty cells) or missing behaviors (no blue cells).

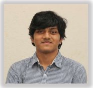
Our approach produces **well-structured latent spaces** equally representing various behaviors.

GSD represents all place locations equally.

6. Future Work

- Scale to high dimensional and non-markovian factors.
- Evaluate with real robot setups and subjective human metrics.

Authors


Varshith Sreeramdass

Rohan Paleja

Letian Chen

Sanne van Waveren

Matthew Gombolay

Acknowledgements & References

This work was supported by MIT Lincoln Laboratory grant FA8702-15-D-0001, NIH grant 1R01HL157457-01A1, NSF grant CPS-2219755, ONR grant N00014-22-1-2834, 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] Chen, L., et. al. (2020). Joint goal and strategy inference across heterogeneous demonstrators via reward network distillation. HRI.
[3] Park, S., et. al. (2021). Lipschitz-constrained unsupervised skill discovery. In *ICRL*.