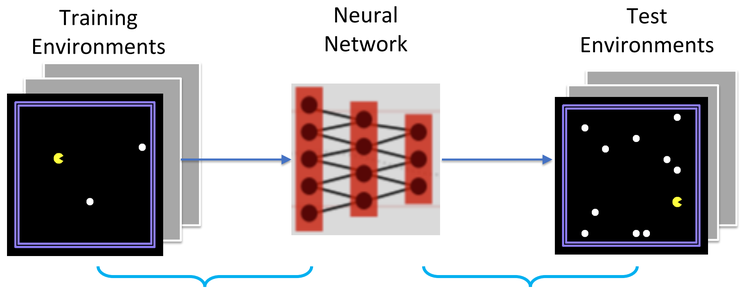
Refactoring Policy for Compositional Generalizability using Self-Supervised Object Proposals

Tongzhou Mu[1*], Jiayuan Gu[1*].
Zhiwei Jia[1], Hao Tang[2], Hao Su[1]
[1]UC San Diego, [2]SJTU
[*]Equal contribution

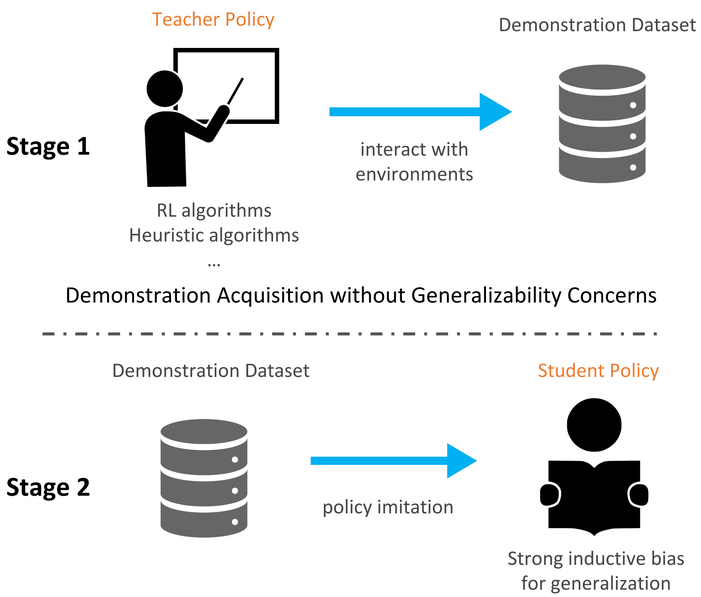
Abstract

We study how to learn a policy with compositional generalizability. We propose a two-stage framework, which *refactorizes* a high-reward teacher policy into a generalizable student policy with strong inductive bias. Particularly, we implement an object-centric GNN-based student policy, whose input objects are learned from images through self-supervised learning. Empirically, we evaluate our approach on four difficult tasks that require compositional generalizability, and achieve superior performance compared to baselines.

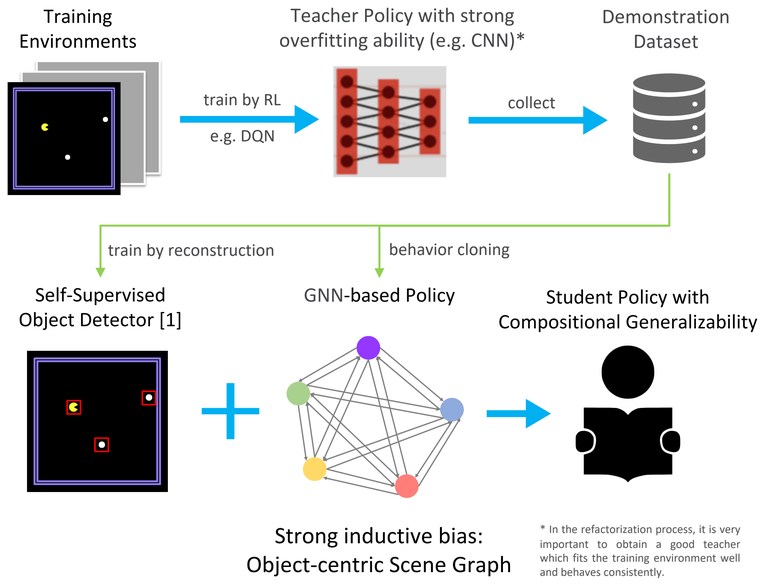
Compositional Generalization in RL



Policy Refactorization

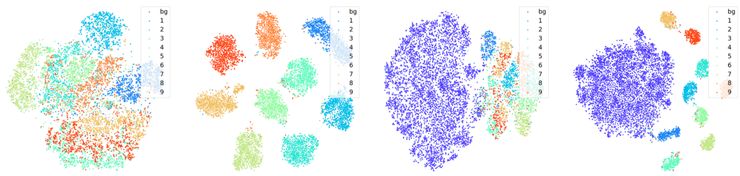


Refactor Into an Object-centric Policy



Task-relevant Knowledge Discovery

Our framework supports interpretable model diagnosis, and the object attributes emerge by clustering the learned object features.



(a) CIFAR-Recon (b) CIFAR-Task (c) ImageNet-Recon (d) ImageNet-Task

This figure corresponds to the Multi-MNIST task, which is elaborated in the right column. This is the t-SNE visualization of the learned object features by the self-supervised object detector (CIFAR/ImageNet-Recon) and our policy GNN (CIFAR/ImageNet-Task). It is observed that task-driven object features are more distinguishable compared to reconstruction-driven ones.

Conclusion

- Refactorization decouples policy optimization on training environments and finding a generalizable policy for testing environments. The decoupling introduces two simpler problems to be solved independently compared with the classical way of solving them together.
- In difficult environments that require sophisticated reasoning, long-range interaction, or unfamiliar background, GNN-based student policy shows stronger performance and robustness.

Reference

[1] Lin, Zhixuan, et al (2019). "SPACE: Unsupervised Object-Oriented Scene Representation via Spatial Attention and Decomposition." In: 7th International Conference on Learning Representations.

Experiments

- Multi-MNIST

Task is to to calculate the summation of the digits in the image with complicated backgrounds (from CIFAR or ImageNet). We show that object-centric graph can be a strong inductive bias for compositional generalizability.

Training Set Test Set

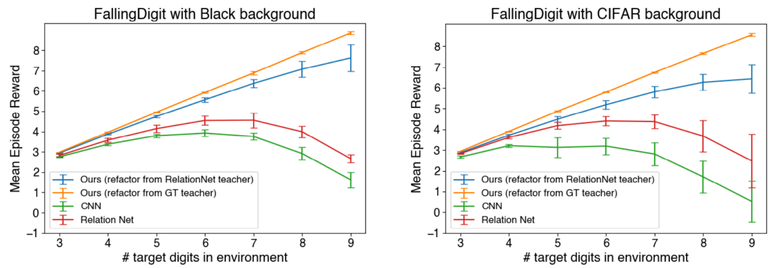
Method	Train Acc	Test Acc
CNN	90.5(2.9)	12.0(2.1)
Relation Net	96.4(0.8)	8.4(4.7)
Ours	80.2(0.2)	**51.2(1.2)**

- FallingDigit

A Tetris-like game. A digit is falling from the top and the player needs to control it to hit the digit with the closest value lying on the bottom. We show that the student network with object-centric graph inductive bias can refactorize the teacher policy into a compositional generalizable policy.

(a) FallingDigit-Black (b) FallingDigit-CIFAR

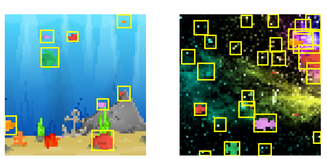
Figure 3: Examples of FallingDigit games with different backgrounds. Object proposal generated by our improved SPACE are annotated in green bounding boxes.



The mean rewards got by different methods in the FallingDigit environments with different target digits. Our refactorized GNN-based policy is trained on the environment with 3 target digits, and can generalize well to the environment with 9 target digits.

- BigFish

A game from ProcGen benchmanrk. Task is to make the green fish grow by eating smaller fishes and avoiding larger ones. We should that our object-centric GNN policy is more robust to the environment with unseen backgrounds.



Training Set Test Set

Method	Test on unseen backgrounds
CNN	4.40(1.90)
Relation Net	4.54(1.22)
Ours (refactor from CNN)	**6.05(2.44)**