Refactoring Policy for Compositional Generalizability using Self-Supervised Object Proposals

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

Stage 1
RL algorithms
Heuristic algorithms

Training Environments
Neural Network
Test Environments

Optimization Challenge
Generalization Challenge

Teacher Policy

Demonstration Acquisition without Generalizability Concerns

Demonstration Dataset
Student Policy

Stage 2
policy imitation

Refactor Demonstration into Compositional Generalizable Policy

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.

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.

Experiments
- Multi-MNIST
Task is 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.

- 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.

- BigFish
A game from ProcGen benchmark. 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.