Refactoring Policy for Compositional Generalizability using Self-Supervised Object Proposals

Tongzhou Mu*, Jiayuan Gu*, Zhiwei Jia, Hao Tang, Hao Su
Compositional Generalization

Generalize?
Challenges of Classical RL

Optimization Challenge

Training environments

Neural Network
Challenges of Classical RL

Neural Network

Test environments

Generalization Challenge

Refactoring Policy for Compositional Generalizability using Self-Supervised Object Proposals
Refactoring Policy for Compositional Generalizability using Self-Supervised Object Proposals

Refactorization

Stage 1

Teacher Policy

Demonstration Dataset

interact with environments

RL algorithms
Heuristic algorithms...

Demonstration Acquisition without Generalizability Concerns
Refactorization

Stage 2

Demonstration Dataset → Student Policy

policy imitation

Refactorize Demonstration into Compositional Generalizable Policy

Strong inductive bias for generalization
Object-centric Policy

Self-Supervised Object Detector + Graph Neural Network = Student Policy

Strong inductive bias: Object-centric Scene Graph

Refactoring Policy for Compositional Generalizability using Self-Supervised Object Proposals
Experiments

- Flexible number of objects
- Random object arrangement
- Composition of foreground/background

Multi-MNIST  FallingDigit  BigFish
Multi-MNIST

Object-centric graph can be a strong inductive bias for compositional generalizability

<table>
<thead>
<tr>
<th>Method</th>
<th>Train Acc</th>
<th>Test Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>90.5(2.9)</td>
<td>12.0(2.1)</td>
</tr>
<tr>
<td>Relation Net</td>
<td>96.4(0.8)</td>
<td>8.4(4.7)</td>
</tr>
<tr>
<td>Ours</td>
<td>80.2(0.2)</td>
<td>51.2(1.2)</td>
</tr>
</tbody>
</table>
FallingDigit

The student network with object-centric graph inductive bias can refactorize the teacher policy into a compositional generalizable policy.

Train on 3 digits

CNN-based RL policy fails to generalize to 9 digits

GNN-based refactorized policy generalizes to 9 digits
BigFish

More robust to different composition of foreground and background

Training Environments

Test Environments
Conclusion

• Refactorization through a proper student network with strong inductive bias can ease optimization and achieve compositional generalizability.

• In difficult environments with sophisticated reasoning, long-range interaction, or unfamiliar background, GNN-based student policy shows stronger performance and robustness.

• We implement an effective object-centric policy learning framework with an improved self-supervised object detector.
Thank you!