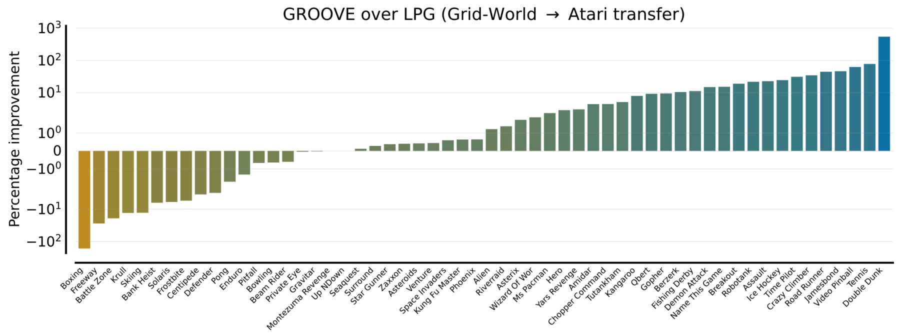
Discovering General RL Algorithms with Adversarial Environment Design [JJH+23]

Artur Ganzha 10019651

Introduction

- **Background:**
  - Meta-learning update rules have shown promise in discovering robust RL algorithms
- **Challenges:**
  - Generalization gap when applying meta-learned algorithms to unseen environments
- **Objective:**
  - Examine the impact of meta-training distributions on the generalization performance of RL algorithms
  - Introduce a novel approach for automatically generating curricula to maximize the regret of a meta-learned optimizer
  - Performance of GROOVE over LPG on Atari
  - Outperforms after learned of grid world levels
  - Figure shows a implicit baseline of LPG performance



Approach and Results

- **GROOVE Methodology:**
  - Build on Unsupervised Environment Design principles (**= dynamically generating environments during training, based on performance**)
  - Trains on Grid-World environments and is evaluated on Atari
- **Algorithmic Regret:**
  - Approximation of regret (Meta-learner performance - baseline performance) used for selecting meta-training tasks
  - Co-trains an antagonist (Baseline) agent using a known RL algorithm to estimate regret
- **Results**
  - Achieves superior generalization compared to LPG
  - Significant improvement in out-of-distribution on Atari even though it was trained exclusively on Grid-World levels