AMAGO-2: Breaking the Multi-Task Barrier in Meta-Reinforcement Learning with Transformers

Jake Grigsby, Justin Sasek, Samyak Parajuli, Daniel Adebi, Amy Zhang, Yuke Zhu

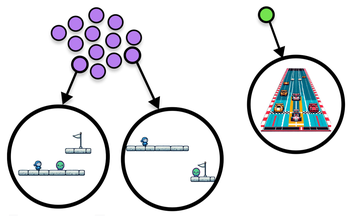
The University of Texas at Austin

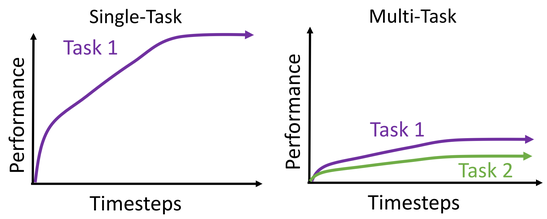
UT Robot Perception & Learning Lab

Meta-RL and the Multi-Task Barrier

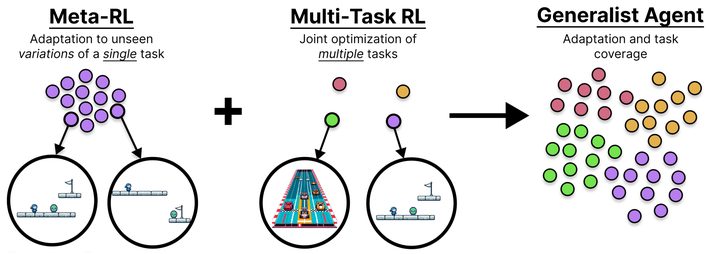
Black-Box Meta-RL adapts to unfamiliar situations using memory of previous attempts ("*in-context learning*") . It focuses on the engineering problem of training sequence models with RL.

We want to train on as many tasks as possible to generalize at test-time, but **efforts to add new reward functions introduce multi-task RL problems** that prevent us from learning the tasks we already have.





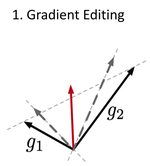
Our work studies a simple and **scalable** solution to this problem that breaks the "multi-task barrier" and lets us train sequence models with RL in more diverse multi-task environments.

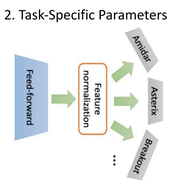


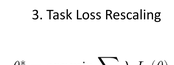
Breaking the Multi-Task Barrier

Multi-Task RL has solutions, but they scale with task count and require task labels →

There's an easy issue we can fix: **multi-task RL losses depend on current value estimates.** The scale of each task's loss evolves unevenly during training. Tricks like **turning Q-learning into a classification** can help patch this.

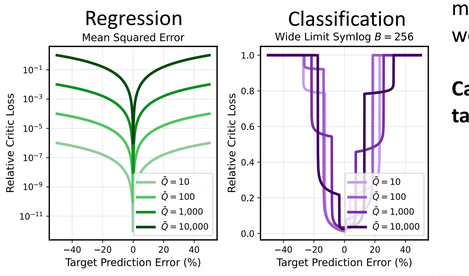




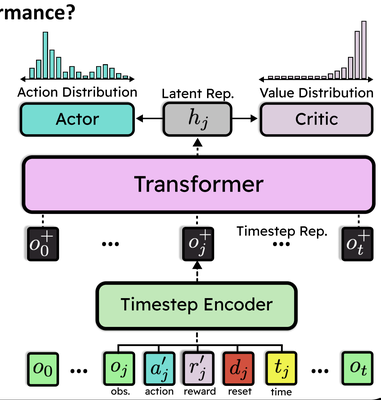


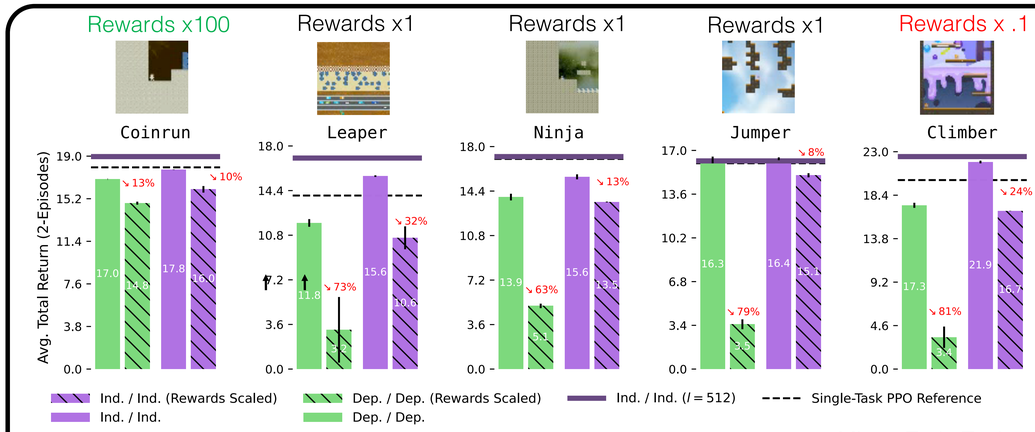
We train Transformers on multi-task training sets of meta-learning and memory problems. We share many low-level details with sequence modeling, but we're still doing end-to-end off-policy actor-critic RL.

Can a simple change in loss function improve multi-task performance?



	Scale Dependent	Scale Resistant
Critic Loss:	Value Regression	Value Classification
Actor Loss:	Policy Gradients	Advantage-Filtered Imitation Learning

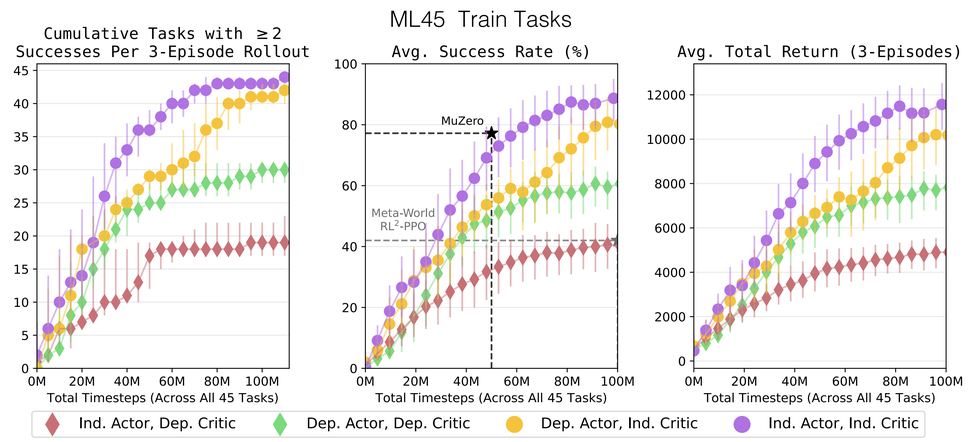


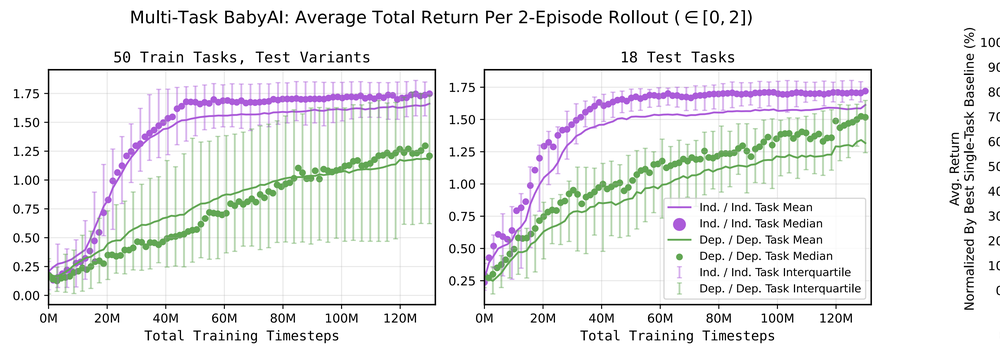


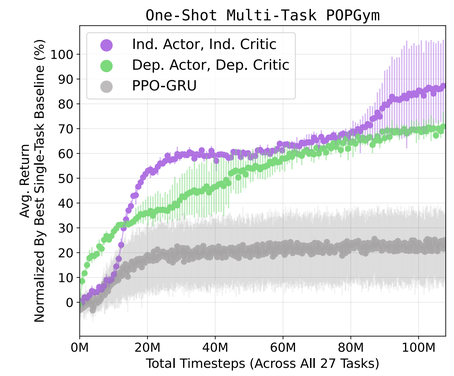
← We can use 5 Procgen games to demonstrate the impact of value scale on multi-task sequence policies by multiplying rewards by a constant. Scale-resistant updates perform better as value estimates diverge.

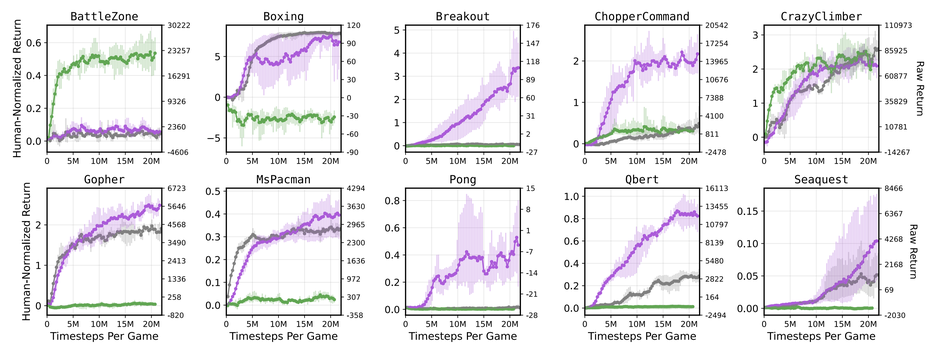
This issue arises naturally in domains that are combinations of different meta-RL tasks. A famous example is the **Meta-World ML45 training set**, which is effectively a multi-task RL problem →

We create two new large-scale examples: **Multi-Task POPGym** and **Multi-Task BabyAI**. Scale-resistant objectives like value classification let us learn from more training tasks, and create an opportunity to reach towards more generalist domains.









We emphasize that black-box meta-RL is a general case of multi-task RL, and inherits many of its challenges.

← For example, classification losses can help a Transformer **play 10 Atari games at the same time**

Scan the QR code to find our paper, code, and a more detailed summary.