

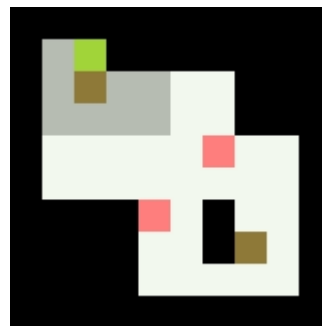
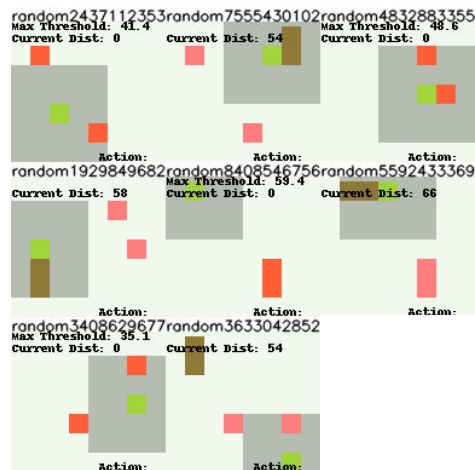
Research Experience

Charles Zhang

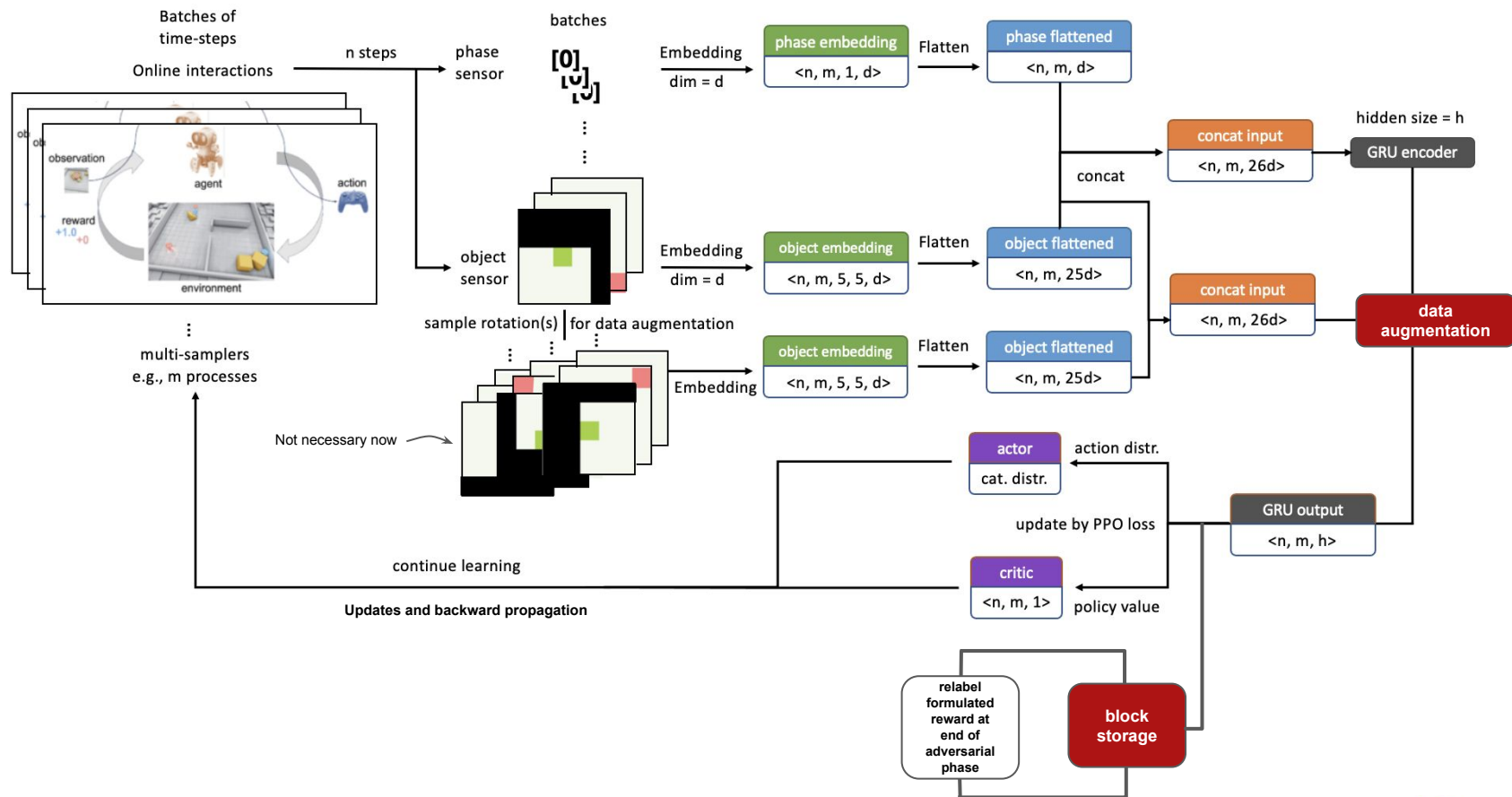


Reset Free RL with Sink States via Adversarial Gameplay

- Lead the research project mentored by [Luca Weihs](#) at [Allen Institute for AI \(AI2\)](#).
- Meanwhile, contributed codes for baseline experiments and pretrained models of various of tasks from the MuJoCo group under [OpenAI gym](#) environments to the [project](#) of [AllenAct](#), and provide a [Tutorial](#).
- Used **Dynamism, Embeddings, Recurrent Models (LSTM, GRU), Data Augmentation by MDP Homomorphism under Rotation Invariance**, adaptive (and probably conservative in the future) irreversible penalties with (weighted) binary cross entropy (BCE) losses (policies) for irreversible costs, separate policies for repeated two phases, and the **basic experience rollout block storage along with an adversarial reward reply storage training pipeline**.
- The methodology is first tested in the reset free **egocentric** Sokoban environment, with extra “done” and “setting goal” actions to test the feasibility (and fine-tuning for the real Sokoban task => plan to train from script about this in the future).



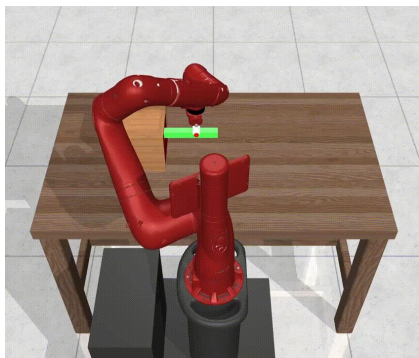
Reset Free RL with Sink States via Adversarial Gameplay



Reset Free RL with Sink States via Adversarial Gameplay

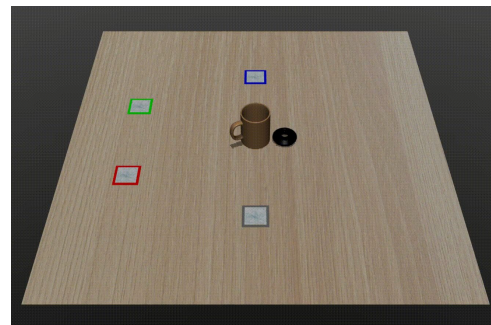
Sawyer Peg

- Community benchmark success rate: 0% (recently 40% with demonstrations, reset every $1e5$ steps, even with fine-tuning!)
- Tried with varied types of distributions, parameterized standard deviation, and implemented Random Network Distillation (**RND**) PPO in AllenAct. Resolved few bugs & long-term instabilities in MuJoCo. New (simpler) reward shaping speeds up 10x.
- Main challenges: exploration & out of task-relevant states (almost irreversible) in the long term.



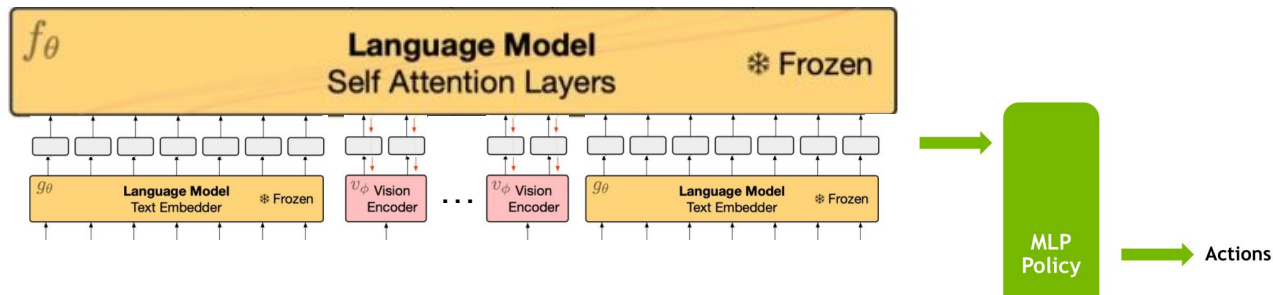
Tabletop Manipulation

- **Multi-tasks reset free** version (just a sanity check in continuous space and everything).
- Vanilla model converges in 500k steps, which is already the benchmark, with new (simpler) reward shaping.



Incoming work: solve challenges, (sparse) + intrinsic, more complicated tasks & env, irreversibility (e.g. narrow the table in Sawyer Peg)...

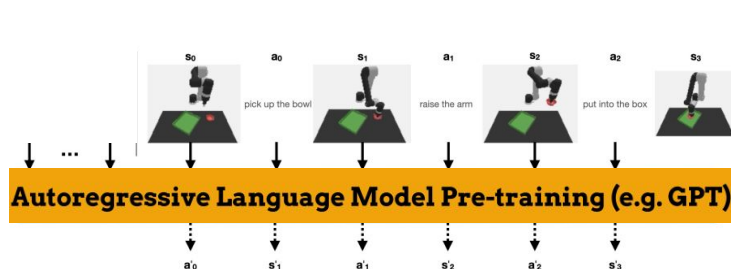
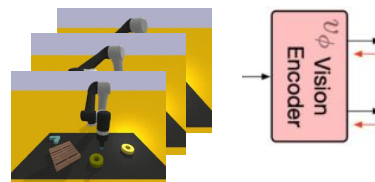
Scalable Robot Learning with Transformers (still under confidential, sorry...)



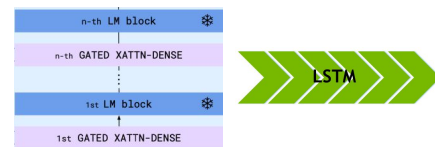
Some novel things....

Joined the research with people from SVL and Nvidia, supervised by Jim Fan (Nvidia) and Yuke Zhu (UT-Austin and Nvidia).

*Stacked obs,
Segmentations,
Embeddings, e.t.c.*

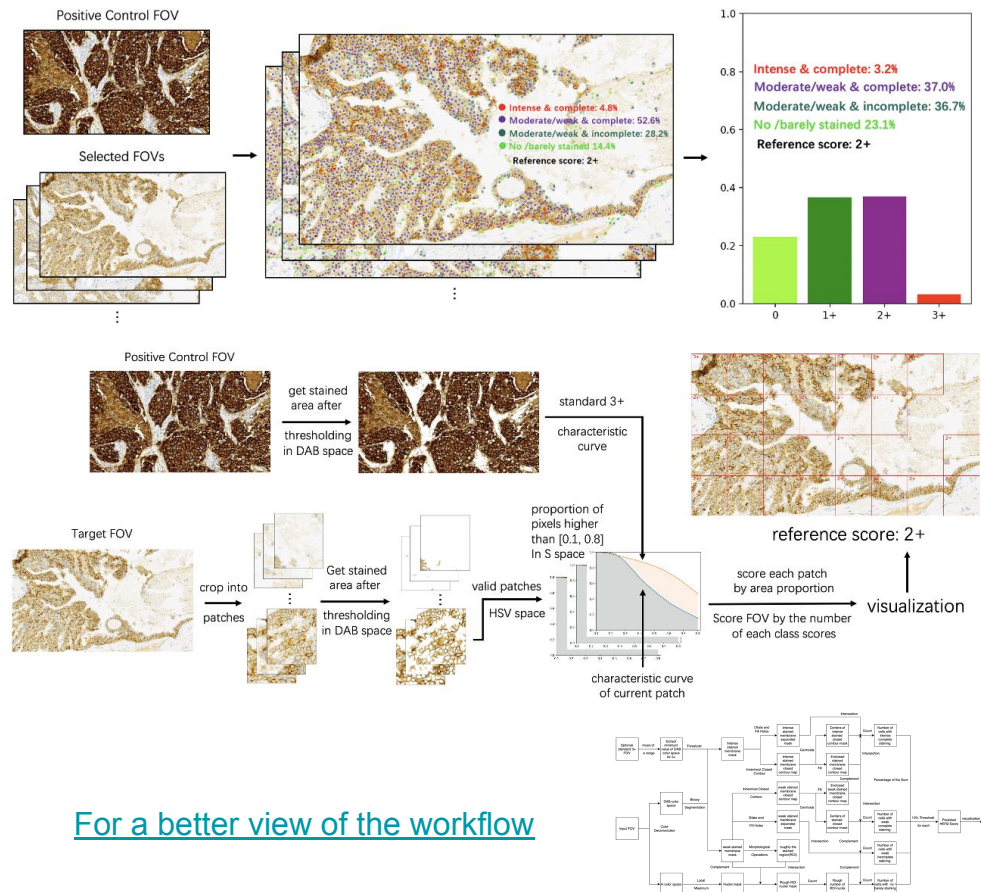


Other option/availability
Xattn/lstm encoder/policy



Automated Scoring System of HER2 in Pathological Images under the Microscope

- Led the research project mentored by CTO, co-CEO [Shuhao Wang](#) at [Thorough Images](#).
- Applied the **National Innovation Patent** (China) and would be **deployed** in People's Liberation Army General Hospital.
- [arXiv](#), 18th European Congress on Digital Pathology (ECDP), 2022
- We propose a HER2 automated scoring system that strictly follows the HER2 scoring guidelines simulating the real workflow of HER2 scores diagnosis by pathologists. Unlike the previous work, our method takes the positive control of HER2 into account to make sure the assay performance for each slide, eliminating work for repeated comparison and checking for the current field of view (FOV) and positive control FOV, especially for the borderline cases
- Our rigorous workflow along with the flexible interactive adjustment in demand substantially assists pathologists to finish the HER2 diagnosis faster and improves the robustness and accuracy



Other 2 Research - summer research at Macalester College

- RL for Area Coverage Path Planning (CPP)
 - Main contributor, the first time I realized the gap between the simulation and real deployment was considerable
 - [My technical report](#)
 - My first opportunity to dig into the RL research at Macalester College mentored by visiting professor Esra Kadioglu Urtis (now at Wake Forest University)
 - DRL were implemented for single and dual UAV CPP. Moreover, I came up with a graph-based RL method with surrogate tasks where the weights of the graph were rewards such that operations in adjacency matrices only allowed the available directions. As the area coverage could be considered as visiting as many way- points as possible, the optimal policy was equivalently corresponding to the minimum values derived by finding the shortest path from the start to the end way-points. This strategy significantly empowered the agent to successfully complete the task in irregular and much larger environments.
- Characterization of Rectifiable Measures that are Carried by Lipschitz Graphs
 - Main contributor for this geometric theory research
 - [Post Link](#), the paper is under revision and will be released and submit soon. The abstract is accepted by [Joint Mathematics Meeting \(JMM\) 2022](#), AMS Contributed Paper Session on Functions of Complex Variables, Measure and Integration Theory (1 of 6 presenters), and American Mathematical Society-Pi Mu Epsilon (AMS-PME) Poster Session