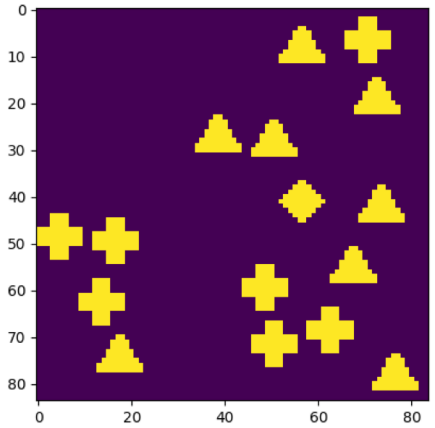
Symbolic Transfer in Games in Deep Reinforcement Learning

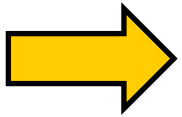
Team Members:
Taylor Qin, Wei Zhou, Junming Zhao
Tutor: Ekaterina

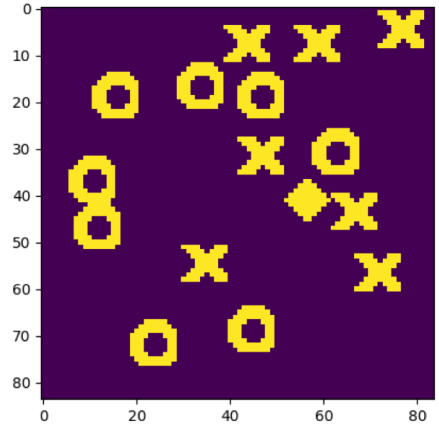
What is our research about?

To transfer between games with similar rules but different symbols.

Positive object: ✚
Negative object: ▲







Positive object: **X**
Negative object: **O**

Game rule:

Let the Agent (♦) hit as many positive objects as possible; while avoiding negative objects.

Why is our research important?

Why transfer?

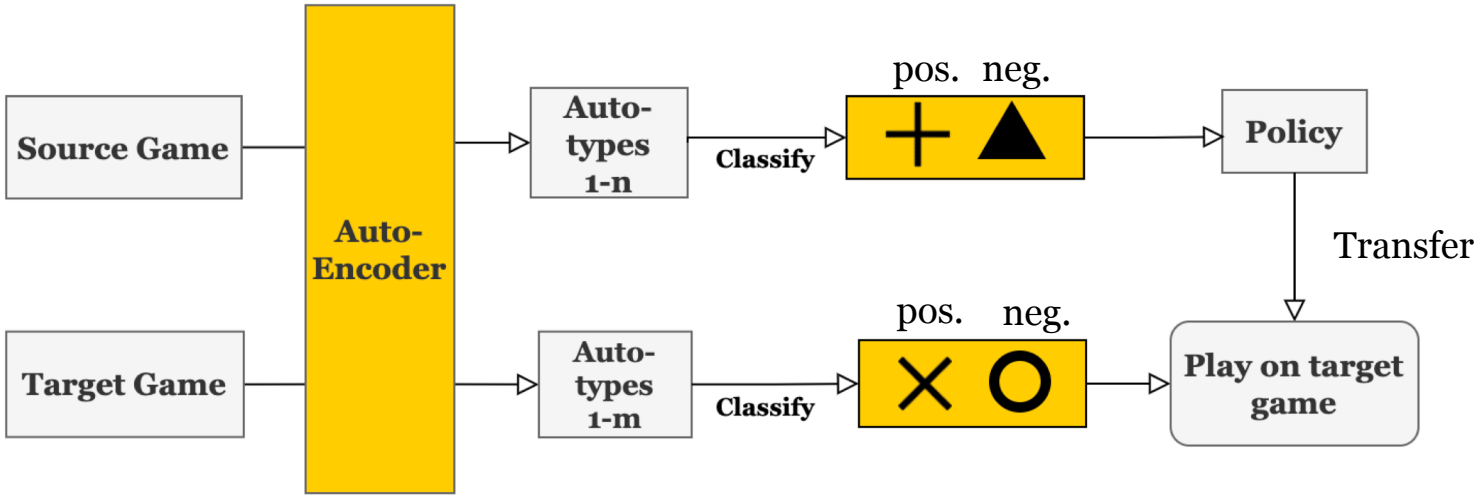
- Reuse prior knowledge
- Save training time
- Less data needed
- Better performance for modelling

Why extract symbols?

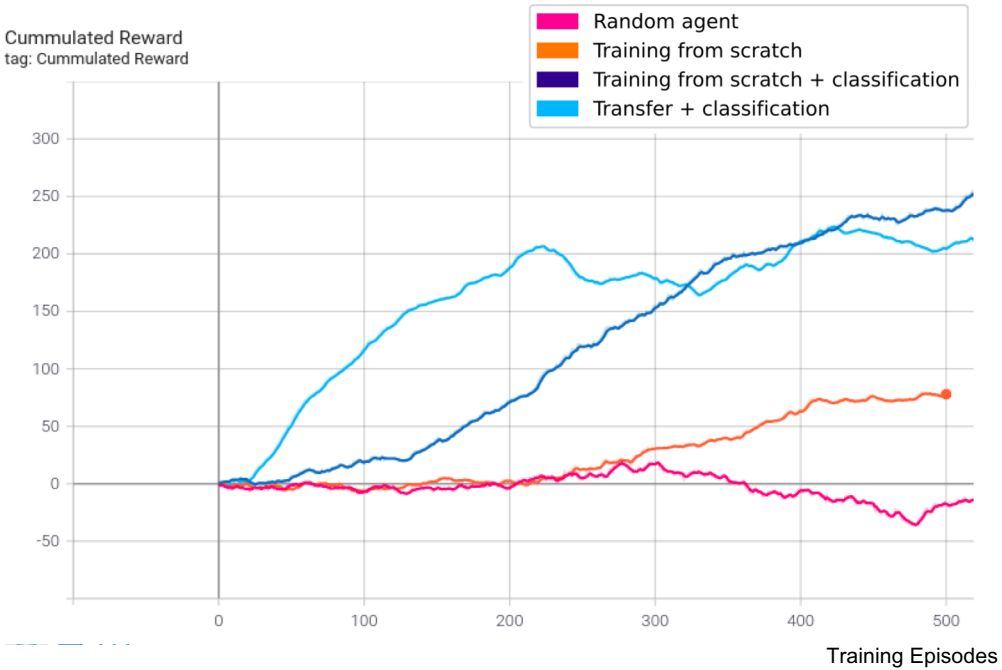
- Less spatial complexity
- Complex symbols adaptable
- Human-readable
- Denoising

What methods do we use?

- Unsupervised Symbol Extraction: **Auto-Encoder** (pixels -> auto-types)
- Classification: **Probe Policy** (auto-types -> pos./neg.)



Experiment Result



- **Classification** significantly improves training efficiency.
- Our **transfer** method accelerates the learning.

- Performance not stable enough

Thanks!

Please come to our tutorial presentation!

Wednesday May 26, 3pm – 5pm