Identifiable Token Correspondence for World Models

Youngin Kim* Ray Sun* Inho Kim Bumsoo Park Hyun Oh Song

Seoul National University · KRAFTON

** Equal contribution*

github.com/snu-mllab/Identifiable-Token-Correspondence

Problem

- Token-based world models **hallucinate** over long horizons
- Objects **duplicate, disappear,** and **transmute**
- Hallucinated rollouts corrupt the imagined trajectories used to train the policy

(a) True rollout

(b) Dedieu et al. (2025)

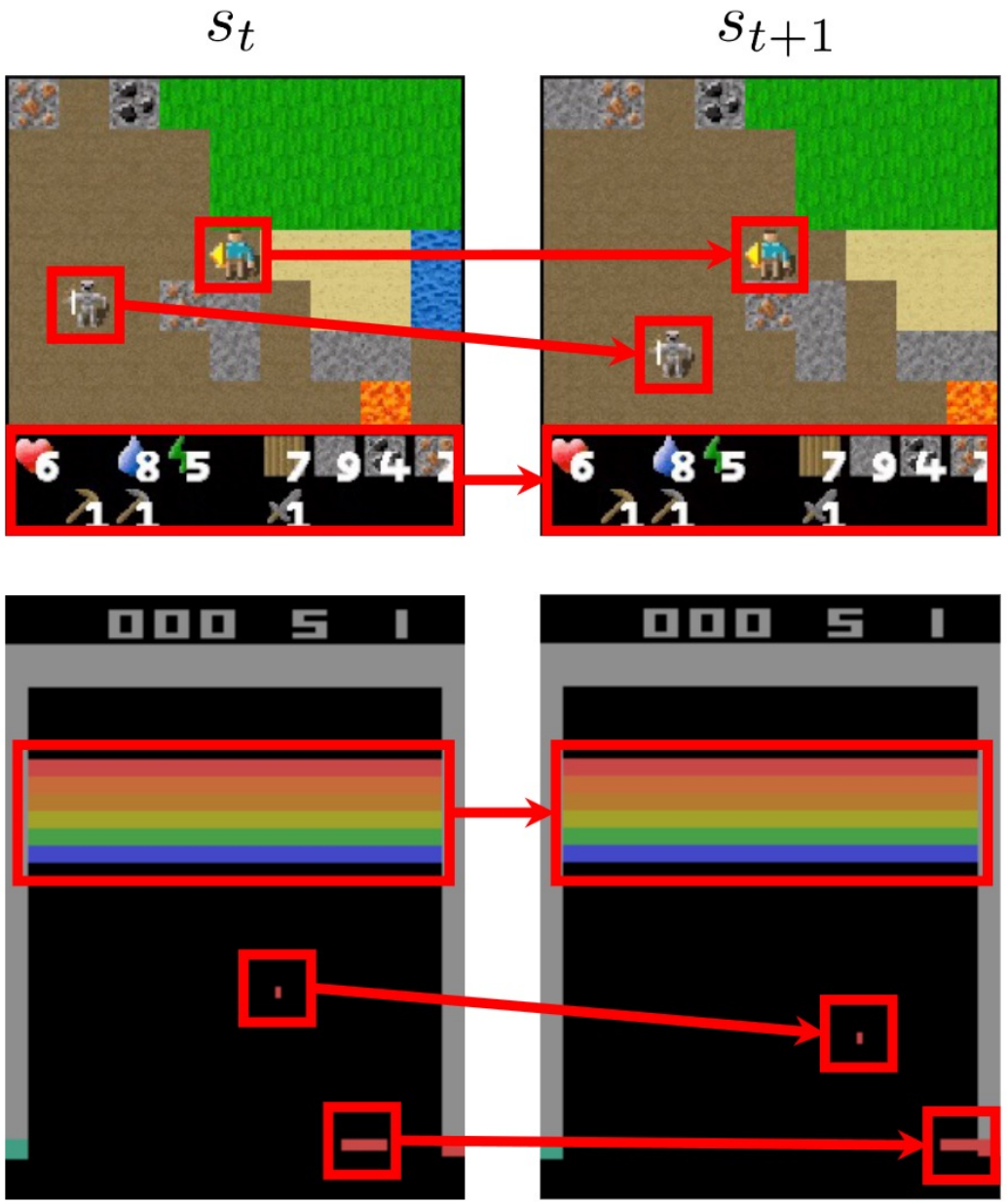
Insight

Existing approaches

- Regenerate every next-frame token from scratch

Our approach

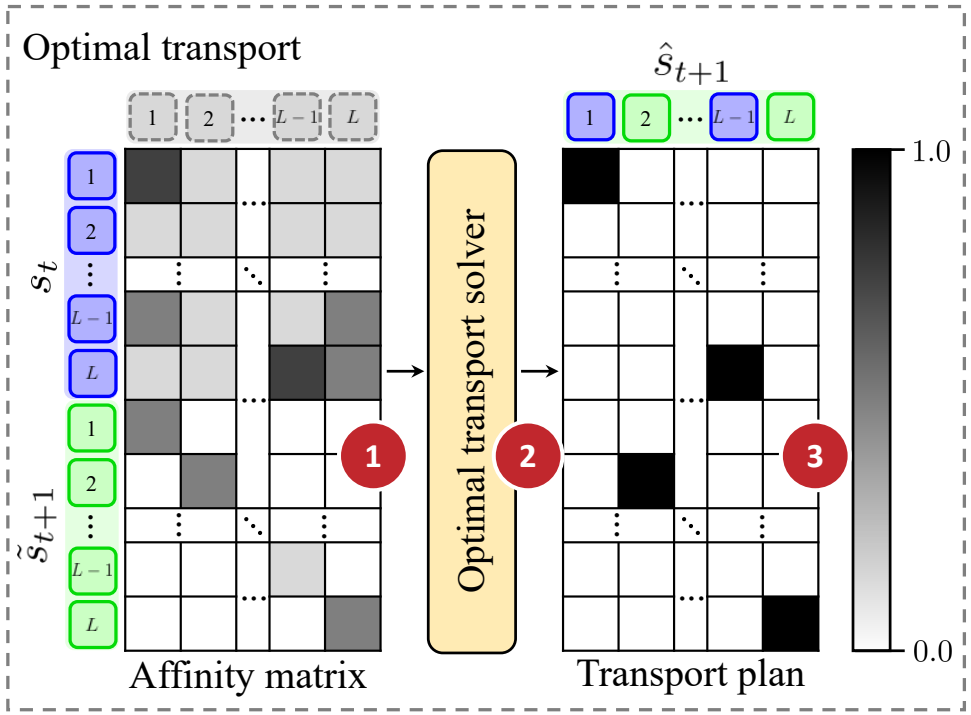
- Same tokens, different positions
- So we **find correspondences** and **copy** them

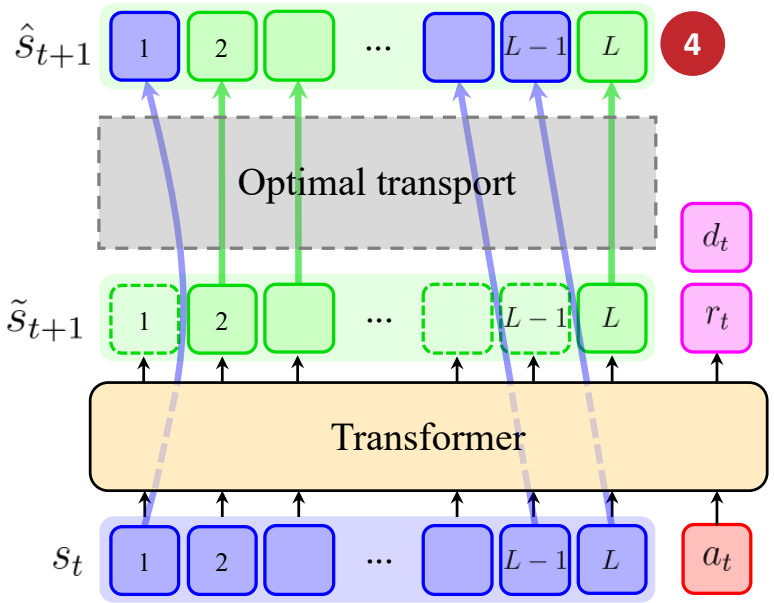


Method

- ITC is a step added at **decoding time**
- **Optimal-transport solver** between transformer predictions and next-state tokens
- Each output is **copied** or **newly generated** — architecture and loss unchanged

1 Build affinity matrix **2** Solve optimal transport **3** Binarize **4** Decode (copy or sample)





Results

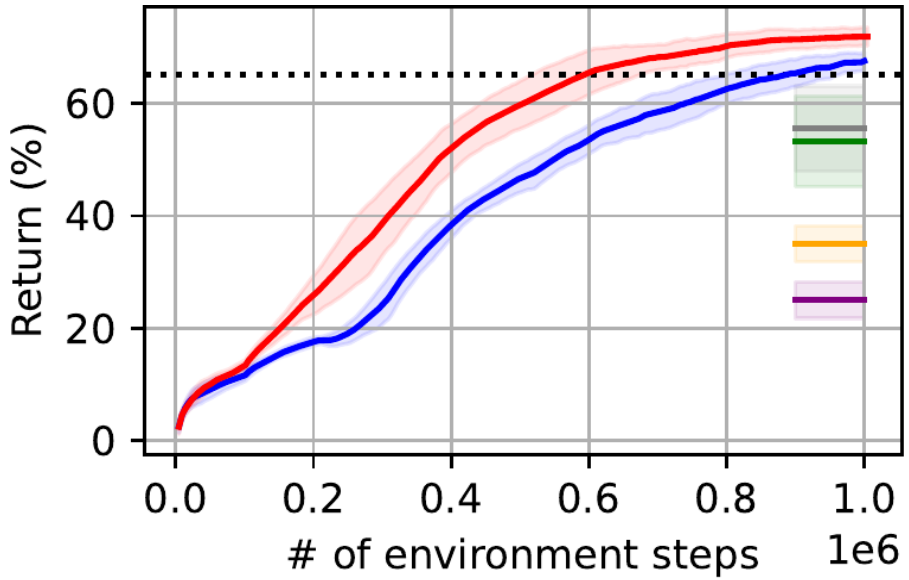
Craftax-classic

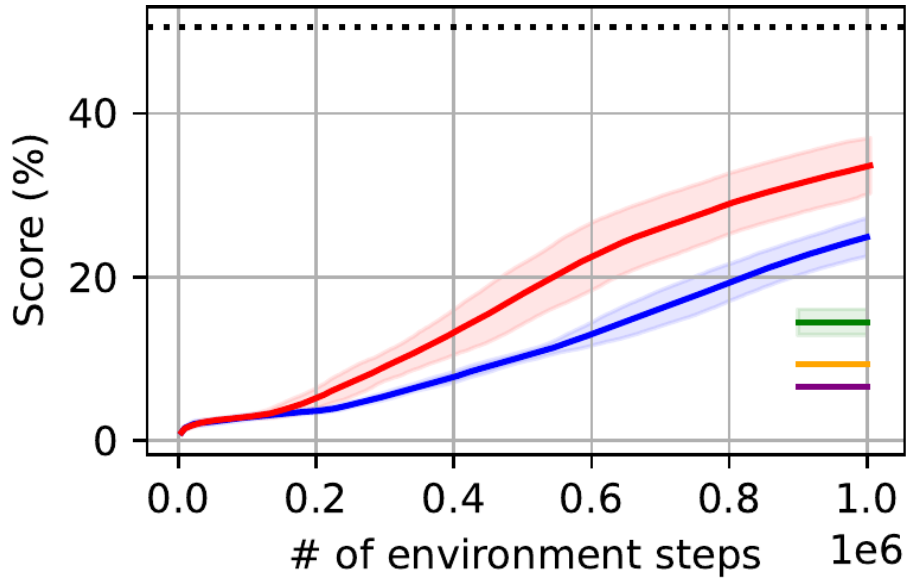
Long horizon, open-world survival game

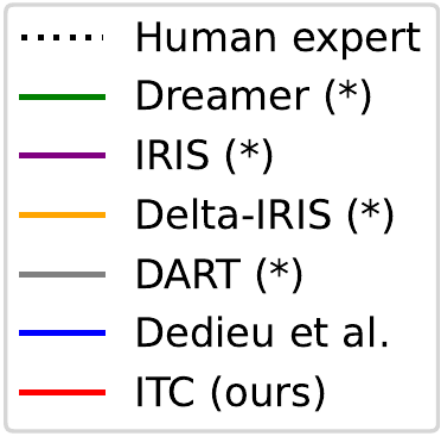
vs Dedieu et al. (2025)

69.6 → **72.4** return

27.9 → **35.6** score







Results

Craftax

harder, larger environment

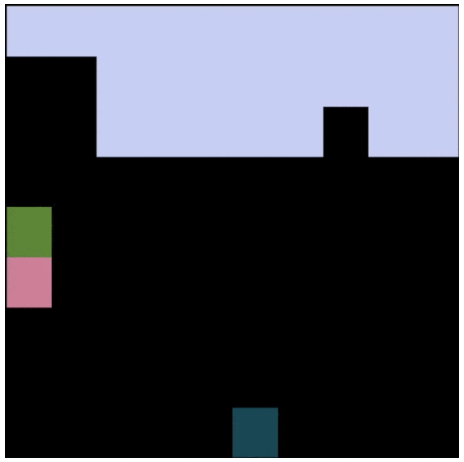
vs Dedieu et al. (2025)

5.44 → **7.09** return

1.53 → **2.40** score

MinAtar

4 games · generality of dynamics



vs Dedieu et al. (2025)

44.8 → **50.0** Asterix

93.9 → **99.5** Breakout

71.1 → **71.3** Freeway

186.2 → **188.9** SpaceInv.

Atari 100K

26 games · non-grid observation

vs Simulus (Cohen et al., 2025)

0.990 → **1.092** IQM (↑)

0.412 → **0.376** Opt. Gap (↓)

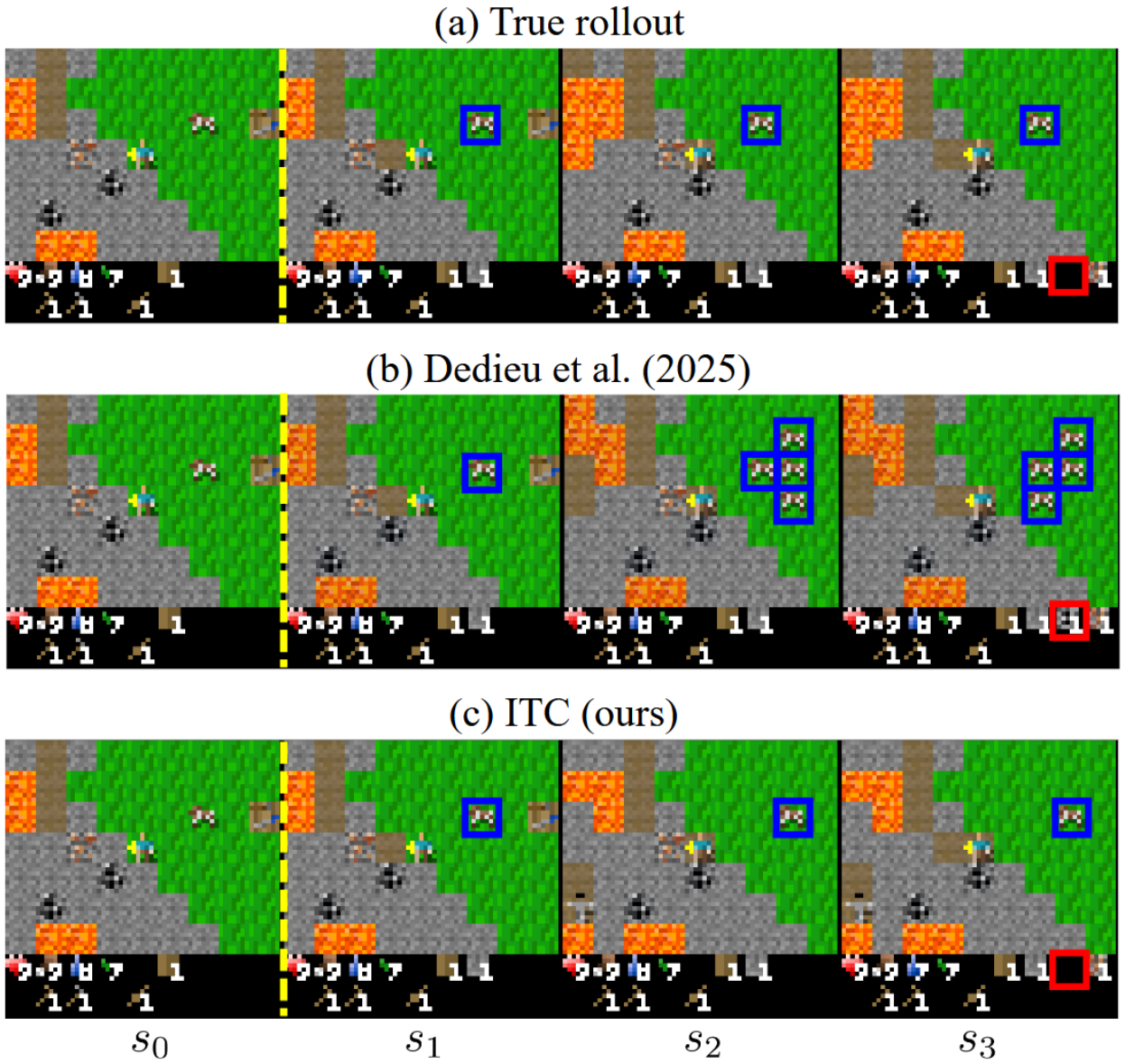
Prediction Quality

Next frame accuracy

exact next-state match · 10K held-out Craftax-classic transitions

46.9 → 51.1% overall **33.8 → 38.4%** with creatures

- Largest gains on **hard creature-containing transitions**
- Per-step gains compound across rollouts
- Better policy return and score



Impact

- **ITC** models token correspondence across frames as **optimal transport**
- Eliminates duplication, disappearance, and transmutation
- With negligible overhead

Drop-in

No architecture or training changes.
Works on existing backbones.

Cheap

Only +2.8% total training time.

General

Patch and VQ-VAE tokenizers.
New SOTA on Craftax-classic, Craftax, MinAtar, and Atari 100K.

github.com/snu-mllab/Identifiable-Token-Correspondence