

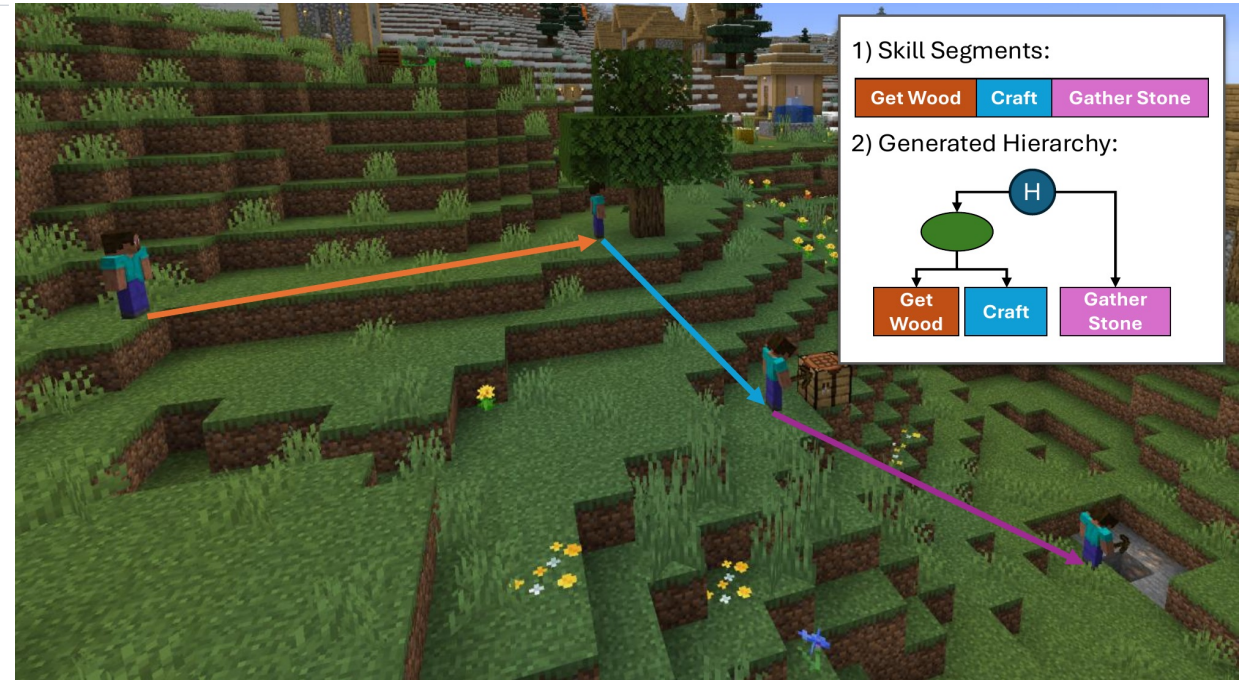
Unsupervised Hierarchical Skill Discovery

We **segment skills** and recover their **compositional hierarchy** from **observation-only trajectories**. No actions, rewards, or annotations are required.

Damion Harvey¹, Geraud Nangue Tasse^{1,2}, Benjamin Rosman^{1,2}, Branden Ingram^{1,2}, Steven James^{1,2}

¹ University of the Witwatersrand

² MIND Institute, Johannesburg



Long-horizon tasks crush flat RL agents

*Humans plan in **skills**. Flat RL agents drown at the **action level**.*

1

humans plan in skills

Walk · pickup · sit — not primitive muscle commands. Flat RL has no skill vocabulary, so long horizons drown it.

2

hierarchies give us three wins

Reuse · Abstraction · Smarter exploration. The right chunks make hard problems tractable.

3

from observations alone

No actions. No rewards. No annotations.
Action-labelled video is rare —
observation-only video is abundant.

The Supervision Gap

Existing methods assume signals that raw observations don't carry.

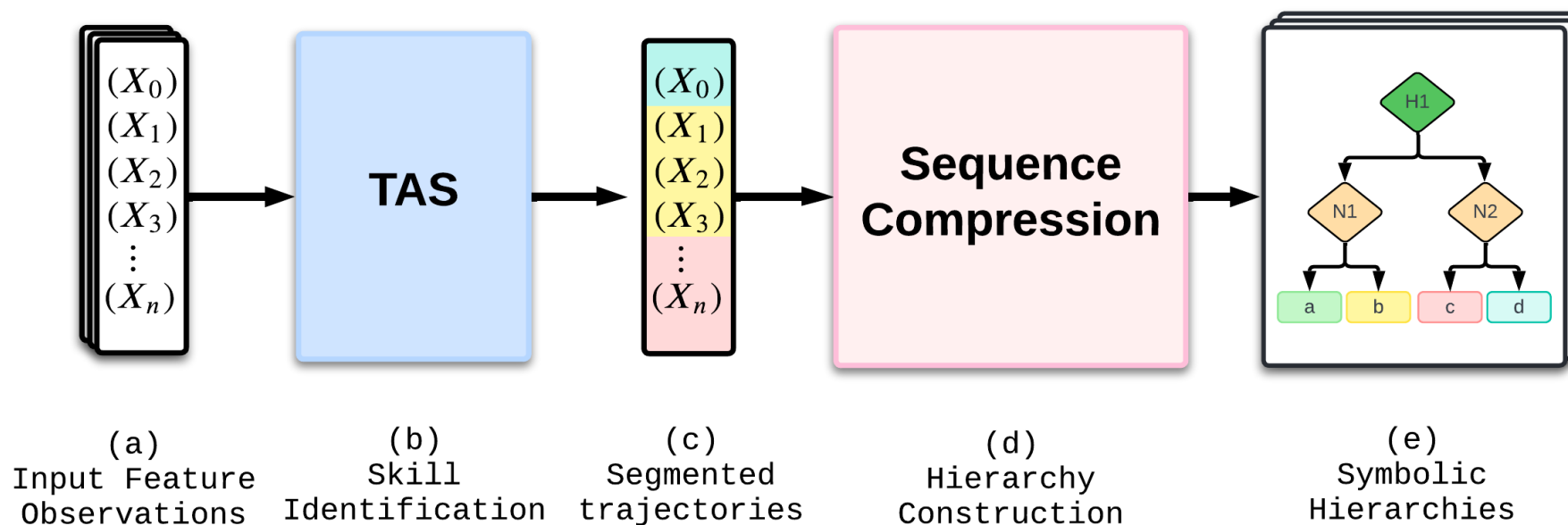
Method	Action Labels	Sub-task Order	K (nb. of skills)	Other
HiSD (Ours)	–	–	✓	–
CompLE	✓	✓	✓	segments per trajectory
OMPN	✓	✓	✓	hierarchy depth

- Action labels are scarce in real-world demonstrations.
- Sub-task orderings inject the very structure we're trying to discover.
- HiSD only assumes a loose upper bound K on the number of skills.

[1] CompLE — Kipf et al., ICML 2019 · [2] OMPN — Lu et al., ICLR 2021

HiSD: segment, then compress

Two unsupervised stages: TAS & Sequence Compression



1

Temporal Action Segmentation

Drop in any action segmenter from CV. We use ASOT (current SOTA)

2

Modified Sequitur

Grammar compression over skill symbols, computed over all episodes to find a composition.

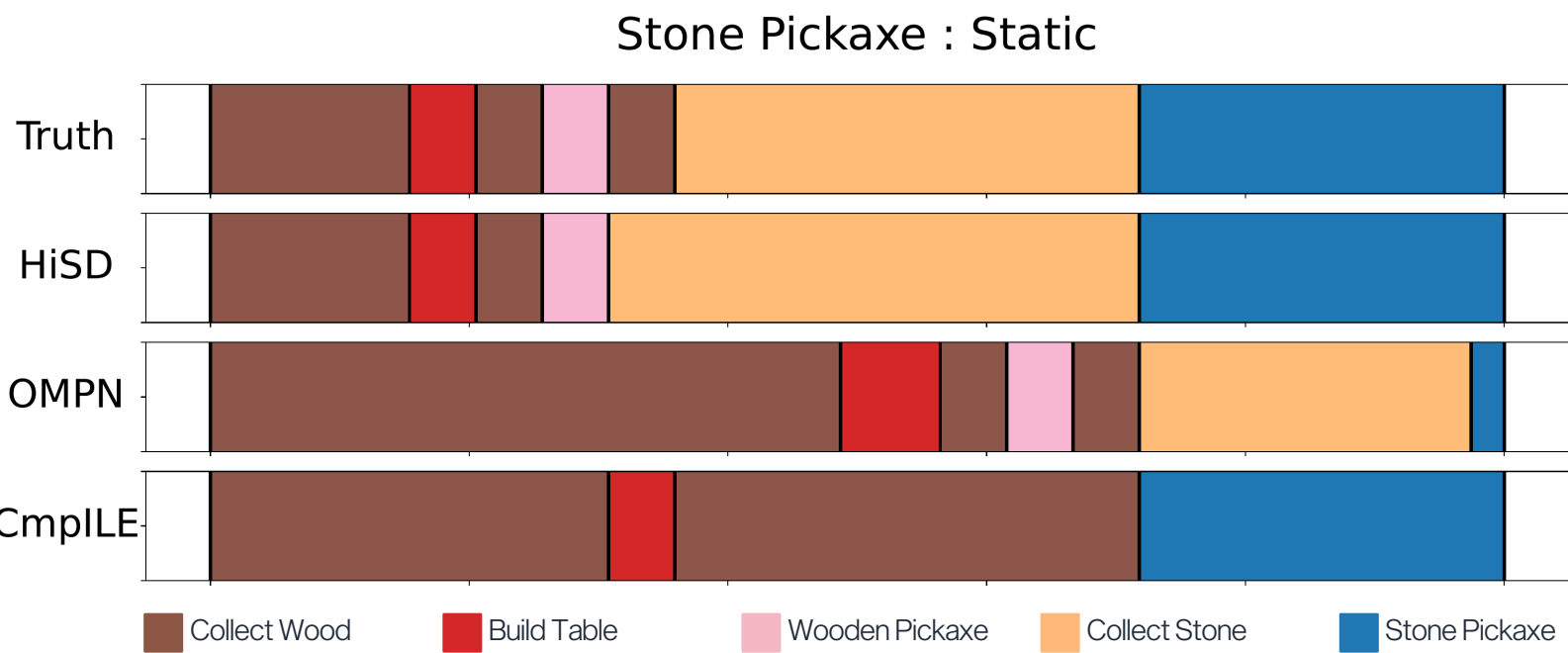
3

Output

Reusable terminal skills + composite sub-routines, for each unique sequence of skills.

HiSD wins on skill re-use

HiSD wins on skill segmentation and re-use accuracy.



Full-mIoU • HiSD vs baselines

Craftax • Stone-Pickaxe-Static

HiSD	65%
OMP	26%
CompILE	40%

Minecraft • Mapped

HiSD	33%
OMP	8%
CompILE	5%

HiSD collapses visually-different, semantically-equivalent interactions into one skill.

[1] CompILE — Kipf et al., ICML 2019 · [2] OMPN — Lu et al., ICLR 2021 · [3] Craftax — Matthews et al., ICML 2024 · [4] MineRL — Guss et al., IJCAI 2019

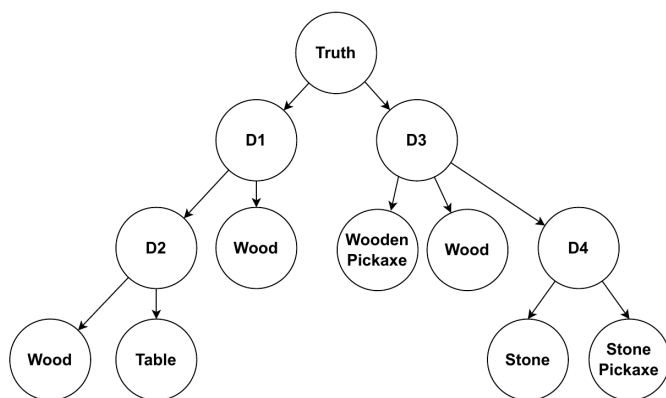
Hierarchies that actually have meaning

Good hierarchies generalise across the dataset.

7

Ground Truth

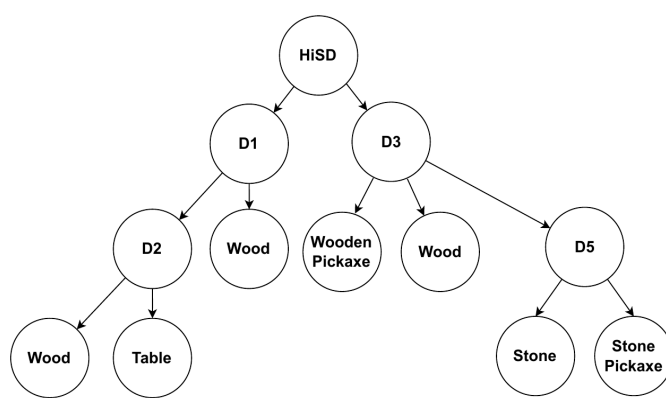
unique trees in the dataset



47

HiSD (ours)

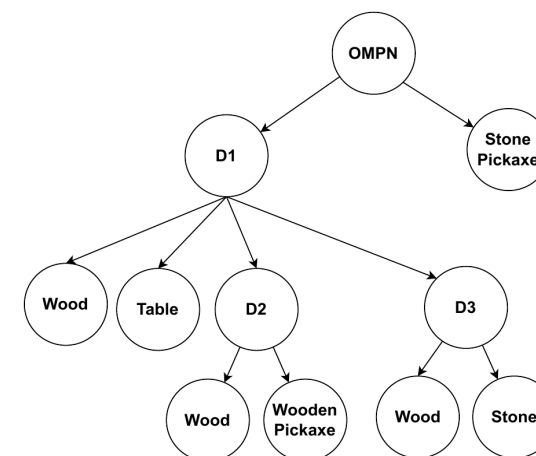
best tree matches the truth structure



500

OMP

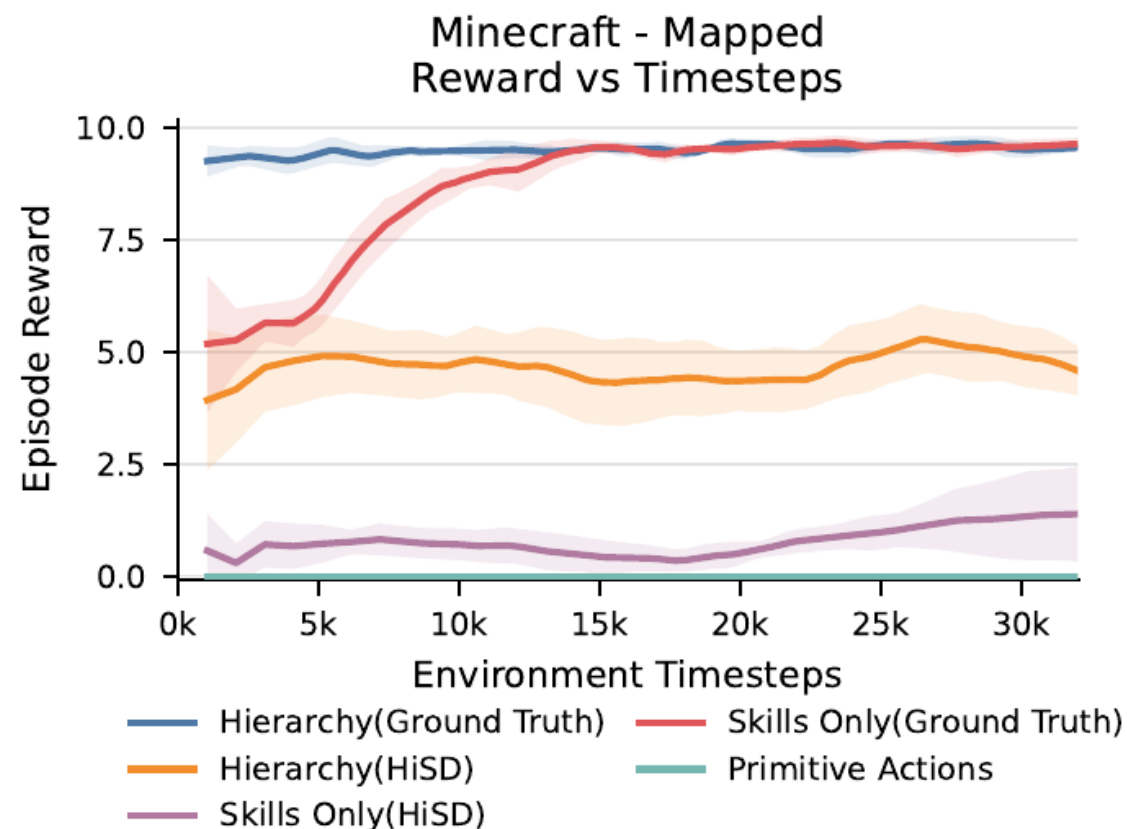
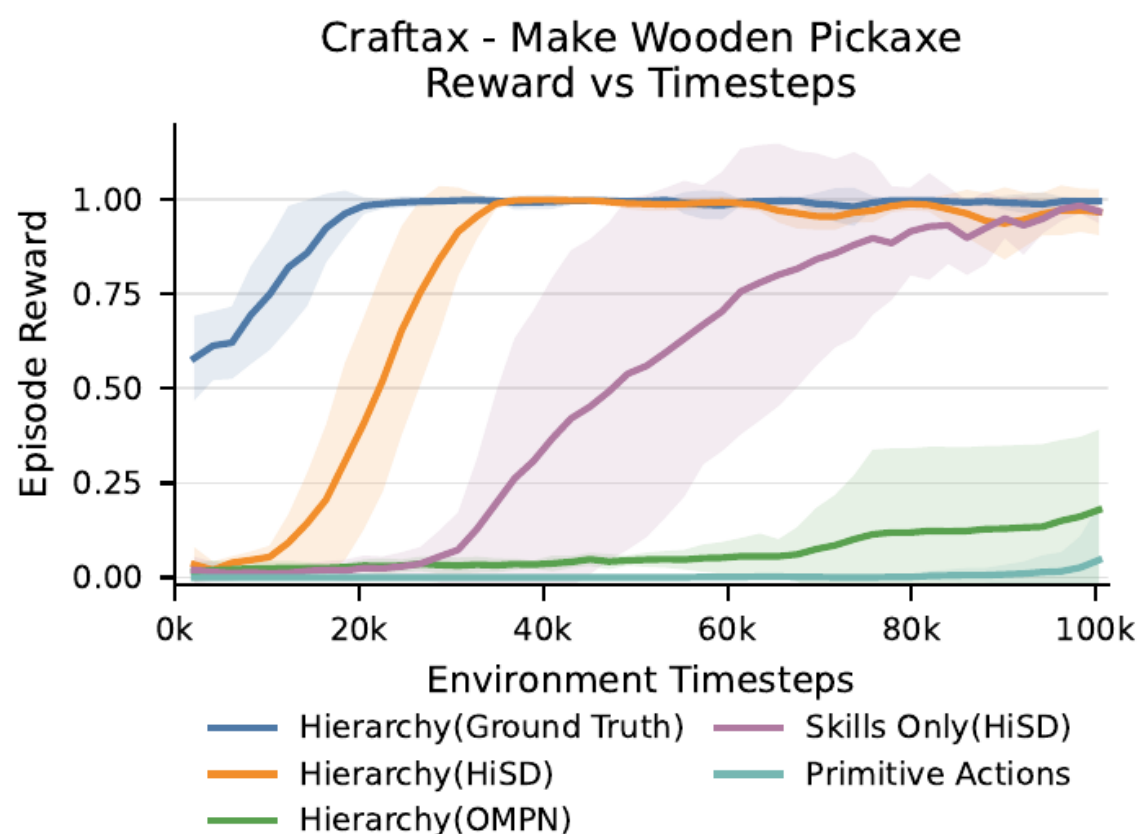
one ad-hoc tree per episode



HiSD's tree is structurally identical to ground truth. OMP collapses every episode under a different degenerate subtree.

Discovered Structure Unlocks RL

Hierarchy carries the gain. HiSD succeeds while primitive PPO fails.



Unsupervised hierarchies, learnt from demos.

Across Craftax and Minecraft, *HiSD recovers reusable, semantically meaningful skill hierarchies* that match ground truth and accelerate downstream RL.

K

the only supervision

No actions, no rewards, no orderings, no segments. Just a loose prior on the number of skills.

47 vs 500

unique trees on Stone-Pickaxe-Random

HiSD's tree recovers the ground-truth structure. OMPN finds a different tree for every episode.

HiSD $\approx$ GT

downstream RL with HiSD options

HiSD-Hierarchy reaches near-ground-truth reward on Craftax Wooden-Pickaxe and Minecraft Collect-Log.