



Figure 1: CRTR makes representations reflect the temporal structure of combinatorial tasks. t-SNE visualization of representations learned by CRTR (right) and CRL (left) for Sokoban. Colors correspond to trajectories; three frames from two trajectories are shown in the center and linked to their representations. CRL embeddings cluster tightly within trajectories, failing to capture global structure and limiting their usefulness for planning. In contrast, CRTR organizes representations meaningfully across trajectories and time (vertical axis).

contrastive representations overfit to instance-specific context, instead of reflecting environment dynamics. Consequently, models fail to adequately capture the temporal structure that is vital for effective decision-making. This failure mode — such as when the model overfits to wall layouts in Sokoban (Section 4.1) — manifests as a collapse of trajectory representations into small, disconnected clusters, as illustrated in Figure 1 (left).

To solve this, we introduce Contrastive Representations for Temporal Reasoning (CRTR), a simple, theoretically grounded CL method that uses in-trajectory negatives. By design, CRTR forces the model to distinguish between temporally distant states within the *same* episode. This mechanism prevents it from exploiting irrelevant context — such as visual or layout cues — and instead encourages learning temporally meaningful embeddings that reflect the problem’s relevant dynamics. This echoes recent findings in neuroscience, where hippocampal representations of overlapping routes diverge during learning despite visual similarity [10]. Our approach similarly prioritizes temporally meaningful structure over reliance on visual cues.

We evaluate CRTR across challenging combinatorial domains: Sokoban, Rubik’s Cube, N-Puzzle, Lights Out, and Digit Jumper. Due to their large, discrete state spaces, sparse rewards, and high instance variability, they serve as challenging testbeds for evaluating whether learned representations can support efficient, long-horizon combinatorial reasoning. In each case, CRTR significantly improves planning efficiency over standard contrastive learning, and approaches or surpasses the performance of strong supervised baselines.

Our main contributions are the following:

1. We identify and analyze a critical failure mode in standard contrastive learning, showing its inability to capture relevant temporal or causal structure in problems with a complex temporal structure.
2. We propose Contrastive Representations for Temporal Reasoning (CRTR), a novel and theoretically grounded contrastive learning algorithm that utilizes in-trajectory negative sampling to learn high-quality representations for complex temporal reasoning.
3. We demonstrate that CRTR outperforms existing methods on 4 out of 5 combinatorial reasoning tasks, and that its representations, even without explicit search, enable solving the Rubik’s Cube from arbitrary initial states using fewer search steps than BestFS—albeit with longer solutions.

2 Related Work

We build upon recent advances in self-supervised RL and contrastive representation learning.

Contrastive learning. Contrastive learning is widely used for discovering rich representations from unlabeled data [12, 11, 32, 53] to improve learning downstream tasks [63]. Contrastive