HUMAN DEMONSTRATIONS AND RL DATA YIELD DIFFERENT CHUNKING EFFECTIVENESS: ACTIVITY SPARSITY AS A KEY INDICATOR

Anonymous authors
Paper under double-blind review

ABSTRACT

Action chunking has emerged as a promising technique for robotic learning, yet it remains unclear when an action-chunking policy can be more effective. Our extensive empirical studies reveal that action chunking policies trained on human demonstrations perform well on standard benchmarks with moderate chunking sizes, whereas those trained on reinforcement learning (RL)-generated data often degrade as the chunk size increases. We investigate this intriguing human-RL gap in chunking effectiveness through the lens of activity sparsity. Our findings show that human demo datasets generally exhibit higher activity sparsity than RL-generated datasets across tasks. Building on this insight, we introduce sparsity regularization into the RL data-generation process and show that the resulting datasets can improve the effectiveness of action chunking policies.

1 INTRODUCTION

Action chunking has emerged as an important technique in imitation learning and robotic control (Chi et al., 2023; Zhao et al., 2023; Zitkovich et al., 2023; Team et al., 2024; Kim et al., 2024; Black et al., 2026). Rather than predicting a single action at each timestep, an action-chunking policy generates a sequence of future actions and executes some or all of that sequence before replanning. The effectiveness of this approach, however, varies substantially across settings. In some cases, policy performance is maintained or improved as the chunk size increases from short to moderate horizons; in others, performance begins to degrade at relatively short horizons. This variability raises a fundamental question: under what conditions is action chunking most effective?

We investigate this question by studying action chunking policies trained on different datasets under otherwise matched experimental settings, including the task and control interface. We find a pronounced human–RL gap in chunking effectiveness: when chunking configurations increase from small to moderate sizes, policies trained on human demonstrations generally maintain or improve their performance, whereas those trained on RL-generated data often degrade. We observe this pattern across robomimic (Mandlekar et al., 2021) Lift, Can, and Square settings under end-effector control, as well as Franka Kitchen (Gupta et al., 2019) under joint-velocity control.

This human–RL gap motivates us to investigate which differences in the underlying behavior might help explain it. We compare several structural properties of the training data and identify activity sparsity as the strongest indicator of chunking effectiveness in our comprehensive analysis. We characterize activity sparsity using two complementary measures: Mean Inactive Dimensions (MID), which measures the average number of control dimensions with near-zero activity, and Hoyer sparsity (Hoyer, 2004), which measures the relative concentration of activity across dimensions. Both measures exhibit consistently moderate-to-strong positive correlations with chunking effectiveness and distinguish datasets yielding high versus low chunking effectiveness across the evaluated tasks. The other tested properties do not receive comparable support from both analyses. We further find that human datasets generally exhibit higher activity sparsity than RL-generated datasets. These findings suggest that this structural difference may reflect behavioral structure that supports greater chunking effectiveness, helping explain the human–RL gap.

Motivated by these findings, we intervene the data-generation process by introducing sparsity regularization into the RLPD (Ball et al., 2023) actor objective and evaluate policies trained on the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

resulting trajectories. We find that this intervention can improve the chunking effectiveness of policies trained on RL-generated data, suggesting that the data-generation objective can be designed to better support learning with longer action chunks.

Our main contributions can be summarized as follows:

1. We identify a human–RL gap in chunking effectiveness across multiple manipulation tasks and control interfaces: as chunking configurations increase from small to moderate sizes, policies trained on human demonstrations generally maintain or improve their performance, whereas those trained on RL-generated data often degrade.
2. We identify activity sparsity as a key indicator of chunking effectiveness, receiving consistent support from both correlation and distinguishability analyses that the other tested properties do not.
3. We show that sparsity regularization in the RL data-generation objective can improve the chunking effectiveness of policies trained on the resulting datasets, providing a direction for generating data that better supports learning with longer action chunks.

2 Human–RL Gap in Chunking Effectiveness: An Empirical Discovery

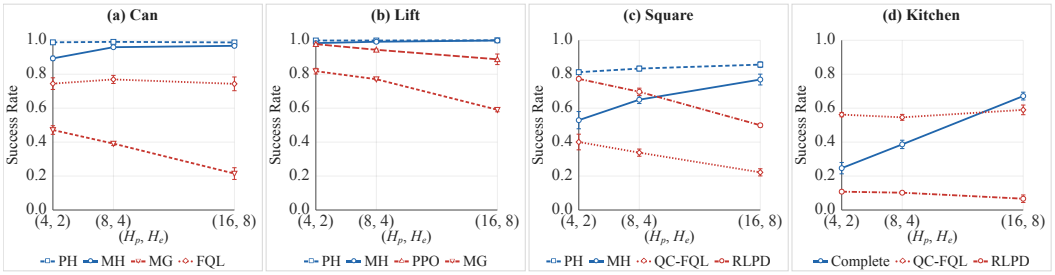


Figure 1: Human demonstration datasets generally exhibit greater chunking effectiveness than RL-generated datasets. Success rate is shown across action-chunking configurations (H_p, H_e) on (a) Can, (b) Lift, (c) Square, and (d) Franka Kitchen. Blue and red curves denote human and RL-generated datasets, respectively. Points and error bars show the mean and standard deviation across three seeds, with each seed averaged over its last 10 evaluations.

In this study, we have carried out extensive experimental studies on the effectiveness of action chunking policies across multiple benchmarks. Specifically, we first compare action chunking policies trained on three Can datasets: Proficient-Human (PH) demonstrations, Multi-Human (MH) demonstrations, and Machine-Generated (MG) dataset, which contains trajectories generated by SAC (Haarnoja et al., 2018) policies. Figure 1(a) shows that as the chunking configuration increases from $(4, 2)$ to $(16, 8)$, policies trained on PH and MH maintain or improve their performance, whereas the policy trained on MG exhibits substantial degradation. These observations motivate the hypothesis that policies trained on human demonstrations generally exhibit greater chunking effectiveness than those trained on RL-generated data.

We next investigate whether the human–RL gap in chunking effectiveness exhibits in other datasets and tasks. Interestingly, across the robomimic tasks (Figure 1(a)–(c)), we observe a consistent pattern: as (H_p, H_e) increases from $(4, 2)$ to $(16, 8)$, policies trained on human demonstrations maintain or improve their success rates, whereas those trained on RL-generated data generally exhibit substantial performance drops. The only RL dataset for which performance remains approximately stable is FQL (Park et al., 2025) on Can. We hypothesize that its objective, which encourages the RL policy to remain close to a BC policy trained on human demonstrations, may help prevent severe degradation.

We further examine whether this pattern extends to Franka Kitchen, which uses joint-velocity control rather than the end-effector control used in robomimic. As shown in Figure 1(d), the policy trained on the human Complete dataset improves substantially as (H_p, H_e) increases from $(4, 2)$ to $(16, 8)$,

whereas the policy trained on QC-FQL (Li et al., 2025) data remains approximately stable and the policy trained on RLPD data declines slightly. These results suggest that the human–RL gap in chunking effectiveness extends to a different control interface.

3 Activity Sparsity as a Key Indicator of Chunking Effectiveness

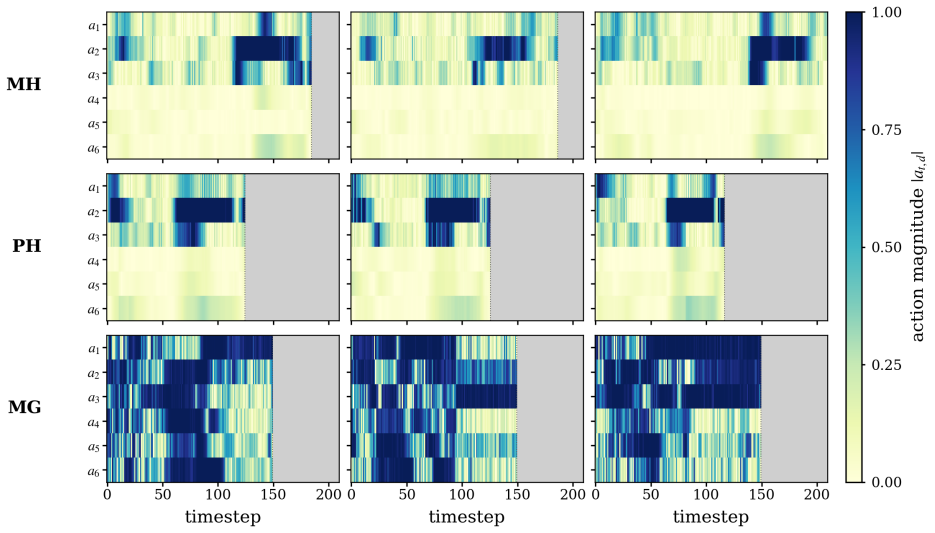


Figure 2: Human demonstrations exhibit visibly sparser action activity than machine-generated trajectories on Can. We plot $|a_{t,d}|$ for the six normalized end-effector delta-pose dimensions of action trajectories, for MH, PH and MG(SAC) data samples. Clearly, MH and PH demos exhibit activity sparsity, whereas MG data exhibits more dense simultaneous activity across multiple dimensions. (Lighter colors indicate near-zero actions; gray denotes unused timesteps after trajectory termination.)

The human–RL gap identified in the previous section raises a natural question: which differences in the underlying behavior might explain this disparity? To investigate, we evaluate five structural properties of the training datasets as potential indicators of chunking effectiveness: activity sparsity, trajectory incoherence, trajectory diversity, temporal sparsity, and chunk-space complexity. Our consideration of activity sparsity is motivated by the sparser action patterns observed in Can human demonstrations(shown in figure 2), which also yield greater chunking effectiveness. Prior work suggests that action chunking can better model pauses and behavioral variability in human demonstrations and improve temporal consistency (Zhao et al., 2023; Chi et al., 2023; Liu et al., 2024). These findings motivate us to investigate whether temporal sparsity (the prevalence of pauses), trajectory diversity (variation across demonstrations), and trajectory incoherence (temporal action variation) also indicate differences in chunking effectiveness. We additionally consider chunk-space complexity based on the intuition that greater effective dimensionality of action-chunk variation may make longer action sequences harder to learn.

Across the evaluated tasks, our comprehensive analyses consistently identify activity sparsity as the strongest indicator, with greater activity sparsity generally associated with more effective chunking. None of the other properties receives comparable support across the analyses. We further observe that human demonstration datasets generally exhibit greater activity sparsity than RL-generated datasets, suggesting that this structural distinction may help explain the observed human–RL gap.

The remainder of this section is organized as follows. We first introduce the structural properties and their quantification in subsection 3.1 and 3.2, followed by the correlation and distinguishability analyses and robustness results in subsection 3.3. Finally, we investigate in depth the relation between activity sparsity between human and RL-generated datasets, and connect these findings to the human–RL gap in subsection 3.4.

In what follows, all structural metrics are computed under the following action-space assumptions. Each trajectory $\tau \in \mathcal{D}$ contains T_τ action vectors $\mathbf{a}_t^\tau \in [-1, 1]^D$, where D is the number of action dimensions used in the analysis. Zero denotes no commanded motion along the corresponding dimension, while -1 and 1 denote the maximum commands in the negative and positive directions, respectively.

3.1 Notions of Activity Sparsity

To get a more concrete sense of activity sparsity, we depict in Figure 2 the action patterns in the Can setting. Notably, the MG dataset, which exhibits low chunking effectiveness, contains substantial activity across many control dimensions simultaneously. In contrast, the PH and MH datasets, which exhibit high chunking effectiveness, show more near-zero action coordinates and activity concentrated in fewer dimensions. These observations motivate us to investigate activity sparsity as a potential indicator of chunking effectiveness.

In what follows, we characterize activity sparsity using two widely used methods: 1) the prevalence of near-zero activity dimensions, measured by Mean Inactive Dimensions (MID), and 2) the relative concentration of activity across dimensions, measured by Hoyer sparsity.

Mean Inactive Dimensions (MID). MID measures the average number of action dimensions whose magnitudes fall below an inactivity threshold:

$$\mathrm{MID}(\mathcal{D}) = \frac{1}{N_\mathcal{D}} \sum_{\tau \in \mathcal{D}} \sum_{t=1}^{T_\tau} \sum_{d=1}^{D} \mathbb{I}\left(\left|a_{t,d}^\tau\right| < \epsilon\right), \qquad N_\mathcal{D} = \sum_{\tau \in \mathcal{D}} T_\tau. \tag{1}$$

Here, $\mathbb{I}(\cdot)$ is the indicator function and ϵ is the inactivity threshold. MID ranges from 0 to D, with larger values indicating more inactive dimensions per timestep.

We use $\epsilon = 0.05$ for all datasets. This choice is guided by a Gaussian-mixture analysis (Dempster et al., 1977) of absolute action values pooled in a dataset-balanced manner across the robomimic datasets. Using the six continuous end-effector motion dimensions, the analysis identifies a low-magnitude component with a fitted mean of approximately 0.05. Details of the mixture fitting are provided in Appendix A.

Hoyer sparsity. Hoyer sparsity measures the relative concentration of action magnitude across dimensions. Unlike MID, it does not require an inactivity threshold and is invariant to uniform rescaling of a nonzero action vector. For $\mathbf{a} \neq \mathbf{0}$, it is defined as

$$H(\mathbf{a}) = \frac{\sqrt{D} - \|\mathbf{a}\|_1 / \|\mathbf{a}\|_2}{\sqrt{D} - 1}. \tag{2}$$

The measure ranges from 0, when all coordinates have equal absolute magnitude, to 1, when only one coordinate is nonzero. We compute dataset-level Hoyer sparsity by averaging over action vectors whose ℓ_2 norms exceed a numerical tolerance:

$$\bar{H}(\mathcal{D}) = \frac{1}{|\mathcal{V}_\mathcal{D}|} \sum_{(\tau,t) \in \mathcal{V}_\mathcal{D}} H(\mathbf{a}_t^\tau), \quad \mathcal{V}_\mathcal{D} = \left\{(\tau, t) : \tau \in \mathcal{D},\ 1 \le t \le T_\tau,\ \|\mathbf{a}_t^\tau\|_2 > 10^{-8}\right\}. \tag{3}$$

Vectors with ℓ_2 norm at or below 10^{-8} are excluded because the norm ratio is undefined at zero and numerically sensitive near zero.

3.2 A Close Look at Other Structural Properties

For comparison, we also investigate four other potentially relevant properties of action structure: trajectory incoherence, trajectory diversity, temporal sparsity, and chunk-space complexity. These comparisons are used to determine whether activity sparsity offers stronger evidence as an indicator of chunking effectiveness than these alternative descriptions of the action data.

Trajectory incoherence. Trajectory incoherence describes how much actions change over a fixed temporal interval. We measure it using the normalized k-step action difference:

$$D(k) = \mathbb{E}_{i,t}\left[\left\|\mathbf{a}_{t+k}^{(i)} - \mathbf{a}_{t}^{(i)}\right\|_2\right], \qquad \bar{D} = \frac{D(k)}{\mathbb{E}_{i,t}\left[\left\|\mathbf{a}_{t}^{(i)}\right\|_2\right]}. \tag{4}$$

The numerator averages over all valid timestep pairs within trajectories, while the denominator averages over all timesteps in the dataset. This normalization expresses temporal differences relative to the average action magnitude. We use $k = 8$, with lower values of $\bar{D}$ indicating lower trajectory incoherence.

Trajectory diversity. Trajectory diversity describes how action sequences differ across demonstrations. We linearly resample each trajectory along each action dimension to a common length $T_0 = 100$, then compute the average pairwise distance:

$$d(i,j) = \frac{1}{T_0}\sum_{t=1}^{T_0}\left\|\tilde{\mathbf{a}}_t^{(i)} - \tilde{\mathbf{a}}_t^{(j)}\right\|_2, \qquad Dist. = \frac{2}{N(N-1)}\sum_{1\le i<j\le N} d(i,j), \tag{5}$$

where $\tilde{\mathbf{a}}_t^{(i)}$ denotes the resampled action. Lower Dist. values indicate lower trajectory diversity.

Temporal sparsity. Temporal sparsity describes how frequently overall action activity is near zero. We measure it using the pause ratio, defining a timestep as idle when the root-mean-square action magnitude falls below $\epsilon_{\text{pause}} = 0.05$:

$$\text{Pause} = \frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\mathbb{I}\left(\frac{\left\|\mathbf{a}_t^{(i)}\right\|_2}{\sqrt{D}} < \epsilon_{\text{pause}}\right). \tag{6}$$

The ratio is computed over all timesteps in the dataset, with higher values indicating greater temporal sparsity.

Chunk-space complexity. Chunk-space complexity describes the effective dimensionality of variation among action chunks. We form sliding windows of length $k = 8$ within each trajectory and flatten each window into a vector $\mathbf{x}_t^{(i)} = [(\mathbf{a}_t^{(i)})^\top, \ldots, (\mathbf{a}_{t+k-1}^{(i)})^\top]^\top \in \mathbb{R}^{kD}$. We pool and center these vectors, compute the eigenvalues λ_j of their sample covariance matrix, and calculate its effective rank (Roy & Vetterli, 2007):

$$\text{Rank} = \exp\left(-\sum_{j:\lambda_j>0} p_j \log p_j\right), \qquad p_j = \frac{\lambda_j}{\sum_\ell \lambda_\ell}. \tag{7}$$

Lower effective rank indicates that variation among action chunks is concentrated in fewer directions, corresponding to lower chunk-space complexity.

3.3 From Structural Properties to Chunking effectiveness

We examine the relation between action chunking effectiveness and these structural properties, using two complementary analyses. Spearman correlation assesses how closely each property's metric ranks datasets according to the chunking effectiveness of policies trained on them. The distinguishability test assesses whether a separating value can distinguish datasets whose trained policies exhibit high versus low chunking effectiveness. Both analyses are conducted within each task, and we compare their results across tasks to identify which properties consistently indicate chunking effectiveness.

We first quantify chunking effectiveness as the success-rate gain obtained by moderately increasing the action-chunking configuration from a baseline with the smallest prediction horizon supported by the policy:

$$\Delta_{\text{chunk}}(\mathcal{D}) = \text{SR}_{\mathcal{D}}(H_{nk}, H_{nk/2}) - \text{SR}_{\mathcal{D}}(H_k, H_{k/2}). \tag{8}$$

where k is the smallest supported prediction horizon and n controls the magnitude of the configuration increase. For the diffusion policy implementation used in our experiments, $k = 4$, giving the

baseline configuration $(H_p, H_e) = (4, 2)$. We operationalize a moderate increase using $n \in \{2, 4\}$, corresponding to configurations $(8, 4)$ and $(16, 8)$. The main text uses $n = 4$, measuring chunking effectiveness by $\Delta_{\text{chunk16}}(\mathcal{D}) = \text{SR}_{\mathcal{D}}(H_{16}, H_8) - \text{SR}_{\mathcal{D}}(H_4, H_2)$, while Appendix B reports results for $n = 2$, denoted by Δ_{chunk8}.

We classify a dataset as having high chunking effectiveness if $\Delta_{\text{chunk}} \geq -0.5$ percentage points and low chunking effectiveness otherwise. This allows near-zero negative changes are treated as approximately unchanged performance rather than assigned to the low-effectiveness category.

Table 1: Within-task Spearman ρ of each metric with Δ_{chunk16}. n is the number of datasets. Average is the unweighted mean across tasks. Bold marks the largest absolute correlation in each row, including ties.

Task	n	MID	Hoyer	$\bar{D}$	Dist.	Pause	Rank
Lift	4	$+0.80$	$+0.60$	$+0.40$	-0.60	$\mathbf{+0.95}$	-0.60
Can	4	$+0.80$	$\mathbf{+1.00}$	-0.20	-0.20	$+0.80$	-0.20
Square	4	$\mathbf{+1.00}$	$+0.80$	-0.20	$\mathbf{-1.00}$	$+0.80$	-0.20
Kitchen	3	$\mathbf{+1.00}$	$\mathbf{+1.00}$	-0.50	-0.50	$\mathbf{+1.00}$	$\mathbf{-1.00}$
Average		$\mathbf{+0.90}$	$+0.85$	-0.12	-0.57	$+0.89$	-0.50

Spearman Correlation Test Table 1 shows that MID achieves the highest average Spearman correlation with Δ_{chunk16} across tasks (0.90), followed closely by the pause ratio (0.89) and Hoyer sparsity (0.85). Both activity sparsity measures and the pause ratio exhibit positive correlations in every evaluated task, suggesting that activity sparsity and temporal sparsity are informative indicators of chunking effectiveness.

However, strong rank correlation does not necessarily imply that a metric can distinguish datasets whose trained policies exhibit high versus low chunking effectiveness. For example, a dataset yielding low chunking effectiveness may have a higher sparsity value than one yielding high chunking effectiveness, even when the overall correlation is positive. To complement the ranking analysis, we therefore conduct a distinguishability test that checks whether a separating value can distinguish these two groups.

Distinguishability Test We consider both possible directions: higher metric values may indicate greater chunking effectiveness, or lower values may do so. For the higher-is-better direction, we select the smallest metric value x among datasets yielding high chunking effectiveness and check whether all datasets yielding low chunking effectiveness have metric values strictly below x. For the lower-is-better direction, we select the largest metric value x among datasets yielding high chunking effectiveness and check whether all datasets yielding low chunking effectiveness have values strictly above x. A metric passes the distinguishability test if a separating value exists within every task under the same direction.

Table 2 shows that MID and Hoyer sparsity pass the test in the higher-is-better direction, while trajectory diversity passes in the lower-is-better direction. The remaining metrics fail to consistently separate the two groups across tasks. In particular, the pause ratio fails on Square and Kitchen despite its strong correlations with chunking effectiveness.

Summary and Robustness. The two analyses identify activity sparsity as a key indicator of chunking effectiveness among the tested structural properties. Both sparsity measures exhibit strong positive correlations and consistently distinguish the two groups across the evaluated tasks. In contrast, trajectory diversity consistently separates the groups but exhibits significantly weaker correlations overall, whereas temporal sparsity shows comparable correlation strength but does not consistently distinguish them. Repeating both analyses using $\Delta_{\text{chunk8}} = S(8, 4) - S(4, 2)$ supports the same overall conclusion, suggesting that our findings are not specific to the $(16, 8)$ configuration (Appendix B).

Table 2: Per-datasetΔ_{16} ($\equiv \Delta_{\text{chunk16}}$, pp) and structural metrics. MID and Hoyer measure activity sparsity. $\bar{D}$, Dist., Pause, and Rank measure trajectory incoherence, trajectory diversity, temporal sparsity, and chunk-space complexity, respectively. Arrows indicate successful separation directions; bold values mark the corresponding separation value.

(a) Lift

Dataset	Δ_{16}	MID$^{\uparrow}$	Hoyer$^{\uparrow}$	$\bar{D}$	Dist.$^{\downarrow}$	Pause$^{\uparrow}$	Rank$^{\downarrow}$
MH	+1.5	3.60	**0.53**	0.99	**0.35**	0.19	**8.04**
PH	+0.1	**2.78**	0.57	0.92	0.32	**0.03**	4.04
PPO	−8.9	0.10	0.09	1.22	2.12	0.00	31.41
MG	−22.7	0.17	0.15	0.73	1.80	0.00	15.72

(b) Can

Dataset	Δ_{16}	MID$^{\uparrow}$	Hoyer$^{\uparrow}$	$\bar{D}^{\downarrow}$	Dist.$^{\downarrow}$	Pause$^{\uparrow}$	Rank$^{\downarrow}$
MH	+7.4	2.94	0.52	**0.77**	**0.54**	0.11	**4.39**
PH	−0.1	**1.91**	0.49	0.73	0.54	**0.03**	3.68
FQL	−0.1	2.46	**0.43**	0.53	0.37	0.07	3.03
MG	−25.7	0.22	0.18	0.78	1.69	0.00	11.08

(c) Square

Dataset	Δ_{16}	MID$^{\uparrow}$	Hoyer$^{\uparrow}$	$\bar{D}$	Dist.$^{\downarrow}$	Pause	Rank
MH	+23.9	3.16	**0.49**	0.95	0.50	0.12	5.95
PH	+4.5	**2.49**	0.50	0.88	**0.58**	0.01	4.42
QC-FQL	−17.9	2.09	0.43	0.72	0.60	0.02	3.95
RLPD	−27.3	0.24	0.18	1.25	1.93	0.00	25.68

(d) Kitchen

Dataset	Δ_{16}	MID$^{\uparrow}$	Hoyer$^{\uparrow}$	$\bar{D}^{\downarrow}$	Dist.$^{\downarrow}$	Pause	Rank$^{\downarrow}$
Complete	+42.5	2.50	0.37	**1.03**	**0.53**	0.00	6.79
QC-FQL	+2.8	**2.21**	**0.30**	0.89	0.39	0.00	**12.71**
RLPD	−4.1	0.44	0.19	1.27	1.74	0.00	53.55

3.4 Implications for the Human–RL Gap

Having identified activity sparsity as a consistent indicator of chunking effectiveness, we return to the human–RL gap. Table 2 shows that human datasets generally exhibit higher activity sparsity than RL-generated datasets under both MID and Hoyer sparsity. On Lift, Square, and Kitchen, human datasets have higher values under both measures than all evaluated RL-generated datasets. The pattern also largely holds on Can, although FQL has a higher MID than PH.

These differences connect the preceding analyses to the human–RL gap: higher activity sparsity indicates greater chunking effectiveness, and human demonstrations generally exhibit higher activity sparsity. These findings suggest that the higher activity sparsity in human demonstrations may reflect underlying behavioral structure that supports greater chunking effectiveness, helping explain the human–RL gap.

4 Sparsity Regularization Can Narrow the Human–RL Gap in Chunking Effectiveness

The preceding sections show that policies trained on human demonstrations generally benefit more from longer action chunks than those trained on RL-generated data, and that this difference is associated with activity sparsity. These findings motivate an intervention: can explicitly encouraging sparsity during RL data generation improve downstream chunking effectiveness and thereby narrow the human–RL gap? We investigate this question by adding sparsity regularization to the RLPD actor objective and evaluating policies trained on the resulting datasets.

4.1 Sparsity-Regularized RL Objectives

We consider two regularizers that target complementary aspects of activity sparsity. The first penalizes action magnitude through an L_1 term (Tibshirani, 1996), encouraging action coordinates toward zero. The second uses a Hoyer-inspired norm ratio to encourage action magnitude to concentrate in fewer dimensions. Both regularizers are applied to the sampled action restricted to these dimensions, denoted by $a_{\mathcal{R}}$.

Original RLPD objective function. The RLPD actor minimizes the SAC-style objective

$$\mathcal{L}_{\pi}^{\text{RLPD}}(\theta) = \mathbb{E}_{s\sim\mathcal{B},\, a\sim\pi_\theta(\cdot|s)}\left[\alpha \log \pi_\theta(a \mid s) - \bar{Q}(s,a)\right], \tag{9}$$

where $\mathcal{B}$ denotes the state-sampling distribution induced by mixing the offline prior dataset and the online replay buffer, α is the entropy temperature, and $\bar{Q}(s,a) = \frac{1}{N_Q}\sum_{i=1}^{N_Q} Q_i(s,a)$ is the mean prediction of the N_Q critics.

RLPD objective function with L_1 regularization. We augment the actor objective with the mean absolute magnitude of the sampled motion command:

$$\mathcal{L}_\pi^{L_1}(\theta) = \mathcal{L}_\pi^{\mathrm{RLPD}}(\theta) + \lambda_{L_1}\mathbb{E}_{s\sim\mathcal{B},\, a\sim\pi_\theta(\cdot|s)}\left[\frac{\|a_\mathcal{R}\|_1}{D}\right]. \tag{10}$$

This penalty encourages smaller action coordinates and can thereby increase the number of dimensions below the MID inactivity threshold. However, it also penalizes overall action magnitude and does not distinguish between necessary and unnecessary motion.

RLPD objective function with Hoyer-inspired regularization. To encourage concentration across action dimensions without directly penalizing overall magnitude, we use a stabilized normalized L_1/L_2 ratio:

$$\mathcal{L}_\pi^{\mathrm{shape}}(\theta) = \mathcal{L}_\pi^{\mathrm{RLPD}}(\theta) + \lambda_{\mathrm{shape}}\mathbb{E}_{s\sim\mathcal{B},\, a\sim\pi_\theta(\cdot|s)}\left[\frac{\|a_\mathcal{R}\|_1 + \sqrt{D}\,\delta}{\sqrt{D}\,(\|a_\mathcal{R}\|_2 + \delta)}\right], \qquad \delta = 10^{-3}. \tag{11}$$

Without stabilization, minimizing the normalized norm ratio is equivalent to maximizing Hoyer sparsity for a fixed action dimensionality. The stabilized penalty approximates this scale-invariant ratio when $\|a_\mathcal{R}\|_2 \gg \delta$, while remaining defined at zero. At the zero vector, the penalty equals 1, so complete inactivity does not trivially minimize the objective.

4.2 Experimental Results

We evaluate the interventions on Can and Square using three matched seeds. Dataset sparsity is measured using MID with $\epsilon = 0.05$ and Hoyer sparsity over the six motion-action dimensions. We use CE as shorthand for $\Delta_{\mathrm{chunk16}}$, the difference in success rate between $(H_p, H_e) = (16, 8)$ and $(4, 2)$. Success rates are reported as percentages and CE as percentage points.

Table 3: Sparsity regularization during RLPD data generation on Can and Square. Success rates are for policies trained on the resulting datasets. Values are mean ± standard deviation across three matched seeds. SR is reported in percent; $\mathrm{CE} = \mathrm{SR}(16,8) - \mathrm{SR}(4,2)$ is reported in percentage points.

(a) Can: L_1 regularization.

λ_{L_1}	MID ↑	Hoyer ↑	SR(4, 2)	SR(16, 8)	CE ↑
0	0.226 ± 0.020	0.167 ± 0.012	74.7 ± 7.0	11.3 ± 2.2	-63.5 ± 4.8
3	$\mathbf{1.990 \pm 0.129}$	$\mathbf{0.432 \pm 0.017}$	68.5 ± 2.8	31.1 ± 11.0	$\mathbf{-37.5 \pm 11.1}$

(b) Square: Hoyer-inspired regularization.

λ_{shape}	MID ↑	Hoyer ↑	SR(4, 2)	SR(16, 8)	CE ↑
0	0.222 ± 0.035	0.162 ± 0.014	81.1 ± 1.9	42.3 ± 2.2	-38.9 ± 1.3
3	$\mathbf{3.042 \pm 0.163}$	$\mathbf{0.664 \pm 0.021}$	71.9 ± 5.4	39.3 ± 11.6	$\mathbf{-32.7 \pm 8.8}$

L_1 regularization on Can. As shown in Table 3a, using $\lambda_{L_1} = 3$ increases MID from 0.226 to 1.990 and Hoyer sparsity from 0.167 to 0.432. The mean chunking gain improves from -63.5 to -37.5 percentage points, reducing the performance drop by 26.0 percentage points.

Hoyer-inspired regularization on Square. With $\lambda_{\mathrm{shape}} = 3$, MID increases from 0.222 to 3.042 and Hoyer sparsity from 0.162 to 0.664 (Table 3b). The mean chunking gain improves from -38.9 to -32.7 percentage points, a reduction in degradation of 6.2 percentage points.

L_1 regularization on Square. The effect of L_1 regularization differs across tasks. On Square, both tested nonzero coefficients increase MID and Hoyer sparsity, but neither improves the mean chunking gain relative to the unregularized baseline (Table 4). CE decreases from -38.9 percentage points to -54.3 at $\lambda_{L_1} = 1$ and to -41.3 at $\lambda_{L_1} = 3$. Increasing the measured activity sparsity is therefore not sufficient to improve chunking effectiveness in these interventions.

One possible explanation is that L_1 regularization changes both the distribution of activity across dimensions and the overall action magnitude. By encouraging low action magnitude, it may suppress

Table 4: L_1 regularization during RLPD data generation on Square. Values are mean $\pm$ standard deviation across three matched seeds. Success rates of policies trained on the resulting datasets are reported in percent; CE is reported in percentage points.

λ_{L_1}	MID ↑	Hoyer ↑	SR$(4,2)$ ↑	SR$(16,8)$ ↑	CE ↑
0	0.222 ± 0.035	0.162 ± 0.014	$\mathbf{81.1 \pm 1.9}$	$\mathbf{42.3 \pm 2.2}$	$\mathbf{-38.9 \pm 1.3}$
1	1.030 ± 0.073	0.383 ± 0.010	66.8 ± 4.7	12.5 ± 2.3	-54.3 ± 6.4
3	$\mathbf{2.500 \pm 0.146}$	$\mathbf{0.485 \pm 0.006}$	53.4 ± 4.5	12.1 ± 4.3	-41.3 ± 6.6

unnecessary movement as well as task-relevant corrections. We hypothesize that this side-effect is particularly consequential for the precise alignment and insertion required by Square. The Hoyer-inspired penalty instead encourages concentration across dimensions without directly penalizing overall magnitude, which may help explain its more favorable mean CE on this task.

Implications. These interventions support the relationship between activity sparsity and chunking effectiveness: L_1 regularization on Can and Hoyer-inspired regularization on Square increase sparsity and improve mean CE. More importantly, these improvements partially narrow the human–RL gap, suggesting a direction for generating data that better supports learning with longer action chunks.

5 Related Work

Action Chunking in Sequential Decision Making. Action chunking is widely used in imitation learning and robotic control (Fu et al., 2024; Zhao et al., 2024). Methods such as ACT and Diffusion Policy generate sequences of future actions rather than predicting each action independently (Zhao et al., 2023; Chi et al., 2023). Generalist robot policies and vision-language-action models, including Octo and π_0, also adopt action-chunking architectures (Team et al., 2024; Black et al., 2026; Liu et al., 2025; Kim et al., 2025). Subsequent work improves how action chunks are generated and executed. Bidirectional Decoding uses backward coherence and forward contrast to balance temporal consistency with responsiveness to new observations (Liu et al., 2024), while real-time and asynchronous execution methods address inference latency (Black et al., 2025; Shukor et al., 2025). Action chunking has also been extended to reinforcement learning: Q-chunking operates in a chunked action space to support temporally coherent exploration and multi-step value learning (Li et al., 2025).

Understanding the Effectiveness of Action Chunking. Recent work investigates the mechanisms and limitations of action chunking. Zhang et al. (Zhang et al., 2026) provide a control-theoretic account of how action chunking and exploratory data collection can mitigate compounding errors under appropriate stability conditions. Lazzati et al. (Lazzati et al., 2026) investigate mechanisms including non-Markovian expressivity, delayed observations, and implicit ensembling, and examine whether demonstrator non-Markovianity predicts the gains of action chunking over single-step policies. Other analyses examine multimodal modeling failures in behavioral cloning (Mazza et al., 2026) and the role of data consistency in action-chunking Q-learning (Li et al., 2026). Our work complements these perspectives by examining how training datasets differ in their support for increasing chunking configurations from small to moderate sizes, focusing on activity sparsity and interventions in RL data generation.

6 Conclusion

We study when action-chunking policies are more effective. Across multiple tasks and control interfaces, we identify a human–RL gap: policies trained on human demonstrations generally maintain or improve their performance, whereas those trained on RL-generated data often degrade. We identify activity sparsity as a key indicator of chunking effectiveness: two complementary measures consistently distinguish datasets with high and low chunking effectiveness and exhibit moderate-to-strong positive correlations with it within the evaluated tasks. Furthermore, sparsity regularization in the RL data-generation objective can improve downstream chunking effectiveness. Together, these findings highlight the importance of activity sparsity in understanding chunking effectiveness and suggest a direction for generating data that better supports action-chunking policies.

REPRODUCIBILITY STATEMENT

The appendix specifies the notation, bridge derivation, learning schedules, simulation and real-world hyperparameters, the robot action interface, task definitions, and evaluation procedures. It reports individual simulation outcomes, real-world success counts, and the aggregation rules and totals for physical execution efficiency.

AI USE STATEMENT

In this work, we did not use generative AI tools for helping develop theoretical models or conceptual frameworks, constructing mathematical claims, proposing hypotheses, designing research methodology or experiments. We used generative AI tools to assist with literature search and summarization, modifying figures, editing code, and organizing and polishing the paper. We have carefully reviewed all AI-assisted work, including checking summarized related work for potential hallucinations and verifying AI-generated code for correctness. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Philip J. Ball, Laura Smith, Ilya Kostrikov, and Sergey Levine. Efficient online reinforcement learning with offline data, 2023. URL https://arxiv.org/abs/2302.02948.

Kevin Black, Manuel Y Galliker, and Sergey Levine. Real-time execution of action chunking flow policies. *arXiv preprint arXiv:2506.07339*, 2025.

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Lucy Xiaoyang Shi, James Tanner, Quan Vuong, Anna Walling, Haohuan Wang, and Ury Zhilinsky. π_0: A vision-language-action flow model for general robot control, 2026. URL https://arxiv.org/abs/2410.24164.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. In *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1): 1–22, 1977.

Zipeng Fu, Tony Z Zhao, and Chelsea Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. *arXiv preprint arXiv:2401.02117*, 2024.

Abhishek Gupta, Vikash Kumar, Corey Lynch, Sergey Levine, and Karol Hausman. Relay policy learning: Solving long-horizon tasks via imitation and reinforcement learning. *arXiv preprint arXiv:1910.11956*, 2019.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pp. 1861–1870. Pmlr, 2018.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *CoRR*, abs/2006.11239, 2020. URL https://arxiv.org/abs/2006.11239.

Patrik O Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov):1457–1469, 2004.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

Moo Jin Kim, Chelsea Finn, and Percy Liang. Fine-tuning vision-language-action models: Optimizing speed and success. *arXiv preprint arXiv:2502.19645*, 2025.

Filippo Lazzati, Kyle Stachowicz, William Chen, Alberto Maria Metelli, Andrew Wagenmaker, and Sergey Levine. Why does action chunking improve behavioral cloning performance in robotic control? *arXiv preprint arXiv:2608.02547*, 2026.

Qiyang Li, Zhiyuan Zhou, and Sergey Levine. Reinforcement learning with action chunking. *arXiv preprint arXiv:2507.07969*, 2025.

Qiyang Li, Seohong Park, and Sergey Levine. Decoupled q-chunking. In *International Conference on Learning Representations*, volume 2026, pp. 145315–145389, 2026.

Songming Liu, Lingxuan Wu, Bangguo Li, Hengkai Tan, Huayu Chen, Zhengyi Wang, Ke Xu, Hang Su, and Jun Zhu. Rdt-1b: a diffusion foundation model for bimanual manipulation. In *International Conference on Learning Representations*, volume 2025, pp. 29982–30009, 2025.

Yuejiang Liu, Jubayer Ibn Hamid, Annie Xie, Yoonho Lee, Maximilian Du, and Chelsea Finn. Bidirectional decoding: Improving action chunking via guided test-time sampling. *arXiv preprint arXiv:2408.17355*, 2024.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Ajay Mandlekar, Danfei Xu, Josiah Wong, Soroush Nasiriany, Chen Wang, Rohun Kulkarni, Li Fei-Fei, Silvio Savarese, Yuke Zhu, and Roberto Martín-Martín. What matters in learning from offline human demonstrations for robot manipulation. In *5th Annual Conference on Robot Learning*, 2021.

Lorenzo Mazza, Massimiliano Datres, Ariel Rodriguez, Sebastian Bodenstedt, Gitta Kutyniok, and Stefanie Speidel. Understanding multimodal failure in action-chunking behavioral cloning. *arXiv preprint arXiv:2605.22493*, 2026.

Seohong Park, Qiyang Li, and Sergey Levine. Flow q-learning. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=KVf2SFL1pi.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021. URL http://jmlr.org/papers/v22/20-1364.html.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Cham, 2015. Springer International Publishing.

Olivier Roy and Martin Vetterli. The effective rank: A measure of effective dimensionality. In *2007 15th European signal processing conference*, pp. 606–610. IEEE, 2007.

Mustafa Shukor, Dana Aubakirova, Francesco Capuano, Pepijn Kooijmans, Steven Palma, Adil Zouitine, Michel Aractingi, Caroline Pascal, Martino Russi, Andres Marafioti, et al. Smolvla: A vision-language-action model for affordable and efficient robotics. *arXiv preprint arXiv:2506.01844*, 2025.

Octo Model Team, Dibya Ghosh, Homer Walke, Karl Pertsch, Kevin Black, Oier Mees, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, et al. Octo: An open-source generalist robot policy. *arXiv preprint arXiv:2405.12213*, 2024.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

Thomas TCK Zhang, Daniel Pfrommer, Chaoyi Pan, Nikolai Matni, and Max Simchowitz. Action chunking and data augmentation yield exponential improvements in behavior cloning for continuous spaces. In *The Fourteenth International Conference on Learning Representations*, 2026.

Tony Z. Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. In *Proceedings of Robotics: Science and Systems*, Daegu, Republic of Korea, July 2023. doi: 10.15607/RSS.2023.XIX.016.

Tony Z Zhao, Jonathan Tompson, Danny Driess, Pete Florence, Kamyar Ghasemipour, Chelsea Finn, and Ayzaan Wahid. Aloha unleashed: A simple recipe for robot dexterity. *arXiv preprint arXiv:2410.13126*, 2024.

Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In *Conference on Robot Learning*, pp. 2165–2183. PMLR, 2023.

A GAUSSIAN-MIXTURE ANALYSIS OF ACTION MAGNITUDES.

We analyzed the first six continuous end-effector motion dimensions of the RoboMimic action vectors and excluded the sign-based gripper dimension. For each selected action component, we considered its magnitude

$$x_{t,j} = |a_{t,j}|, \qquad j \in \{1, \dots, 6\}.$$

To prevent datasets with more timesteps from dominating the fit, we constructed a dataset-balanced pool by sampling the same number of timesteps from each dataset. All six selected dimensions were then assigned equal weight.

We fitted a two-component univariate Gaussian mixture model to the pooled action magnitudes:

$$p(x) = \pi_0 \mathcal{N}(x; \mu_0, \sigma_0^2) + \pi_1 \mathcal{N}(x; \mu_1, \sigma_1^2), \qquad \pi_0 + \pi_1 = 1,$$

where the component with the smaller mean was designated as the low-magnitude component. The fitted parameters were

$$\pi_0 = 0.523, \qquad \mu_0 = 0.0515, \qquad \sigma_0 = 0.0451,$$

for the low-magnitude component, and

$$\pi_1 = 0.477, \qquad \mu_1 = 0.5886, \qquad \sigma_1 = 0.3212,$$

for the broader active component. Thus, the center of the fitted low-magnitude component was approximately 0.05. Because a Gaussian distribution is symmetric, its model-implied median is equal to its mean; nevertheless, 0.0515 is reported here as the fitted component mean rather than as an empirical median.

This analysis was used only to motivate 0.05 as a candidate low-command threshold. Threshold sensitivity is evaluated independently over a prespecified range.

B SENSITIVITY TO Δ_{chunk8}

As discussed in 3.3, we additionally evaluate the same measure of chunking effectiveness under a smaller configuration increase, from $(4, 2)$ to $(8, 4)$, to assess sensitivity of our claim that activity sparsity is a key indicator to chunking effectiveness.

Table 5: Within-task Spearman ρ of each metric with $\Delta_{\text{chunk8}} = S(8,4) - S(4,2)$. n is the number of datasets. Average is the unweighted mean across tasks. Bold marks the largest absolute correlation in each row, including ties.

Task	n	MID	Hoyer	$\bar{D}$	Dist.	Pause	Rank
Lift	4	$+0.80$	$+0.60$	$+0.40$	-0.60	$\mathbf{+0.95}$	-0.60
Can	4	$\mathbf{+1.00}$	$+0.80$	-0.40	-0.40	$\mathbf{+1.00}$	-0.40
Square	4	$\mathbf{+1.00}$	$+0.80$	-0.20	$\mathbf{-1.00}$	$+0.80$	-0.20
Kitchen	3	$\mathbf{+0.50}$	$\mathbf{+0.50}$	$\mathbf{+0.50}$	$\mathbf{+0.50}$	$\mathbf{+0.50}$	$\mathbf{-0.50}$
Average		$\mathbf{+0.82}$	$+0.67$	$+0.08$	-0.38	$+0.81$	-0.42

Spearman Correlation Test Table 5 shows that MID, Pause,and Hoyer has moderate to strong correlation strength, with value ranging from 0.5 to 0.1. This suggests that both activity sparsity and temporal sparsity can indicate chunking effectiveness under Δ_{chunk8}.

Table 6: Per-dataset Δ_8 ($\equiv \Delta_{\text{chunk8}}$, pp) and structural metrics. MID and Hoyer measure activity sparsity. $\bar{D}$, Dist., Pause, and Rank measure trajectory incoherence, trajectory diversity, temporal sparsity, and chunk-space complexity, respectively. Arrows indicate successful separation directions; bold values mark the corresponding separation value.

(a) Lift

Dataset	Δ_8	MID$^\uparrow$	Hoyer$^\uparrow$	$\bar{D}$	Dist.$^\downarrow$	Pause$^\uparrow$	Rank$^\downarrow$
MH	$+0.8$	3.60	**0.53**	0.99	**0.35**	0.19	**8.04**
PH	$+0.0$	**2.78**	0.57	0.92	0.32	**0.03**	4.04
PPO	-3.3	0.10	0.09	1.22	2.12	0.00	31.41
MG	-4.7	0.17	0.15	0.73	1.80	0.00	15.72

(b) Can

Dataset	Δ_8	MID$^\uparrow$	Hoyer$^\uparrow$	$\bar{D}^\downarrow$	Dist.$^\downarrow$	Pause$^\uparrow$	Rank$^\downarrow$
MH	$+6.6$	2.94	0.52	**0.77**	**0.54**	0.11	**4.39**
FQL	$+2.5$	2.46	**0.43**	0.53	0.37	0.07	3.03
PH	$+0.3$	**1.91**	0.49	0.73	**0.54**	**0.03**	3.68
MG	-8.0	0.22	0.18	0.78	1.69	0.00	11.08

(c) Square

Dataset	Δ_8	MID$^\uparrow$	Hoyer$^\uparrow$	$\bar{D}$	Dist.$^\downarrow$	Pause	Rank
MH	$+12.1$	3.16	**0.49**	0.95	0.50	0.12	5.95
PH	$+2.1$	**2.49**	0.50	0.88	**0.58**	0.01	4.42
QC-FQL	-6.3	2.09	0.43	0.72	0.60	0.02	3.95
RLPD	-7.6	0.24	0.18	1.25	1.93	0.00	25.68

(d) Kitchen

Dataset	Δ_8	MID$^\uparrow$	Hoyer$^\uparrow$	$\bar{D}$	Dist.	Pause	Rank$^\downarrow$
Complete	$+14.0$	**2.50**	**0.37**	1.03	0.53	0.00	**6.79**
RLPD	-0.6	0.44	0.19	1.27	1.74	0.00	53.55
QC-FQL	-1.6	2.21	0.30	0.89	0.39	0.00	12.71

Distinguishability Test Table 6 shows that MID and Hoyer sparsity pass the test in the higher-is-better direction. The remaining metrics fail to consistently separate the two groups across tasks. In particular, the pause ratio fails on Square and Kitchen despite its strong correlations with chunking effectiveness.

C Experimental Details

Codebase. We conduct our experiments on action chunking effectiveness using the Robomimic codebase[1].

Policy architecture. All chunking variants use the same UNet-based diffusion policy in Robomimic (`algo_name = diffusion_policy`), with a 1D temporal UNet backbone (Ronneberger et al., 2015). The UNet configuration is fixed across experiments: diffusion-step embedding dimension 256, channel widths $[256, 512, 1024]$, kernel size 5, and group normalization with 8 groups. We use EMA on policy parameters with power 0.75. For diffusion training/inference, we use DDPM (Ho et al., 2020) with 100 train timesteps and 100 inference timesteps, squared-cosine beta schedule (`squaredcos_cap_v2`), clipped samples, and ϵ-prediction.

Observation/action setup. We use low-dimensional state observations (no RGB/depth/scan for the main Robosuite tasks), with frame stacking of 2 and observation horizon $H_o = 2$. Action outputs are predicted autoregressively in chunks; training sequence length is set equal to prediction horizon (`seq_length` $= H_p$).

Optimizer and training schedule. We use AdamW (Loshchilov & Hutter, 2019)with cosine learning-rate scheduling. The base learning rate is 1×10^{-4}, warmup is 500 steps, and L2 regularization coefficient is 1×10^{-6} (as configured in Robomimic templates). Batch size is 100 and training runs for 2000 epochs per configuration. Other optimization hyperparameters are left at framework defaults.

Rollout evaluation during training. Rollout evaluation is enabled throughout training. For each evaluation point, we run $n = 50$ episodes, with `terminate_on_success=True`. For Lift/Can/Square experiments, rollout horizon is 400. Evaluation is logged every 50 epochs (via `rollout.rate=50`).

Seeds and reported metric. Each $(\text{task}, \text{dataset}, H_p, H_e)$ configuration is trained with 3 random seeds $(0, 1, 2)$. For paper tables/plots, we follow the same post-processing used in our scripts: for

[1] `https://github.com/ARISE-Initiative/robomimic`

each run, we average the *last 10* logged rollout-success values; then we report mean and standard deviation of these per-run values across the 3 seeds.

Compute resources. Experiments are run on NVIDIA RTX 3090 GPUs. Each run uses approximately 2 GB GPU memory. Per-run wall-clock time ranges from roughly 1 to 2.5 days, depending on chunking configuration; notably, smaller chunking horizons are slower in our implementation.

D Dataset Details

Tasks. Main results use Robomimic low-dimensional manipulation tasks: Lift and Can (and Square in supplementary trend analyses).

Successful-trajectory protocol. All datasets used in our chunking analysis are built from *successful trajectories only*. This avoids conflating chunking effects with failure-heavy behavior modes.

Human datasets (PH, MH). For PH and MH, we use the full available successful dataset for each task.

RL-derived datasets (MG, PPO, FQL). To keep dataset scale comparable to PH/MH, we use 300 successful trajectories for each RL-derived dataset:

- **MG(SAC):** the original collected pool contains > 1000 trajectories and includes failures; we first filter successful trajectories, then randomly sample 300.
- **PPO:** we train PPO with Stable-Baselines3 (Raffin et al., 2021), then collect successful rollouts from the last 5 checkpoints (60 successful rollouts per checkpoint, total 300).
- **QC-FQL/FQL/RLPD:** The training of these algorithms is based on the Q-Chunking codebase[2] and follows its experimental setup. We collect successful rollouts from the last five checkpoints, with 60 successful rollouts per checkpoint, yielding 300 trajectories in total.

[2] https://github.com/ColinQiyangLi/qc