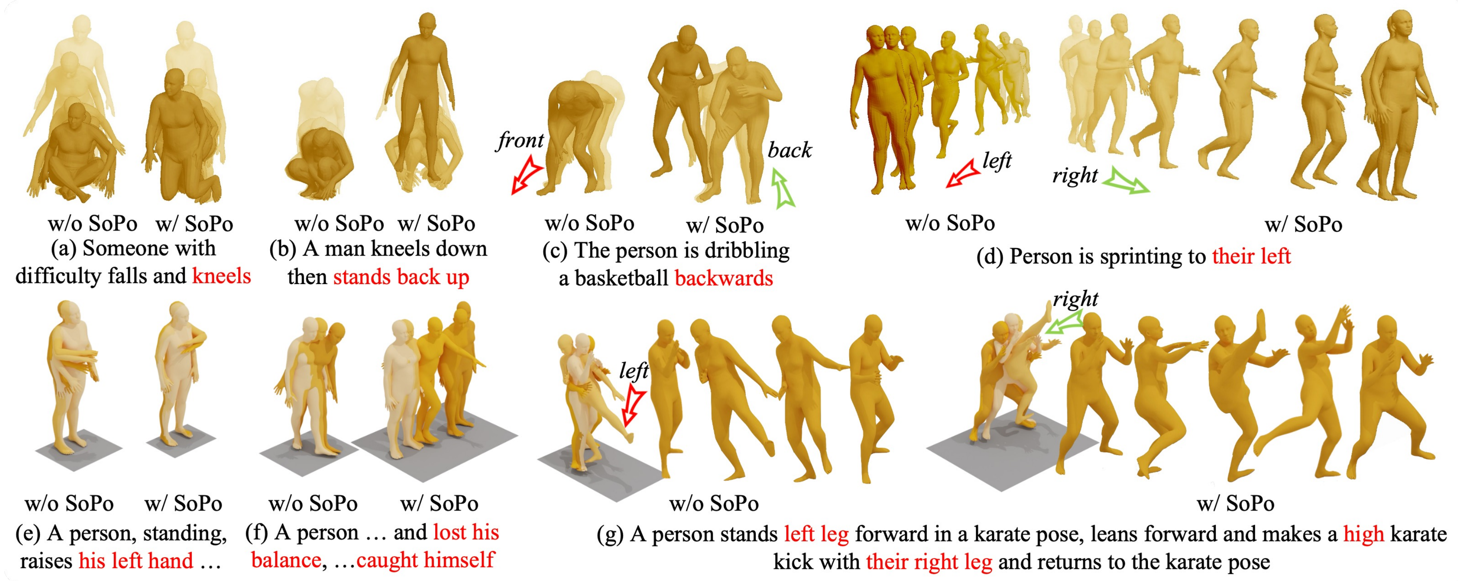
SoPo: Text-to-Motion Generation Using Semi-Online Preference Optimization

Xiaofeng Tan[1,2], Hongsong Wang[1,2*], Xin Geng[1,2], Pan Zhou[3]

[1] Southeast University, [2] Key Laboratory of New Generation Artificial Intelligence Technology, [3] Singapore Management University

Introduction

Issue: Existing text-to-motion methods struggle to generate semantically consistent motions.



Key Observation: We observe that tasks of motion understanding and discrimination are generally less complex and demonstrate superior performance compared to motion generation.

Question: How effectively can *discriminative* models improve motion *generation* quality *without any additional inference cost*?

Contributions: We propose SoPo, a semi-online preference optimization method, combining the strengths of online and offline direct preference optimization to overcome their individual shortcomings, delivering enhanced motion generation quality and preference alignment.

Motivation: Rethink Off-/Online DPO

Offline DPO: overfitting due to limited unpreferred motions.

Theorem 1. *Given a preference motion dataset $\mathcal{D}$, a reference model π_{ref}, and ground-truth preference distribution p_{gt}, the gradient of $\nabla_\theta \mathcal{L}_{\text{off}}$ can be written as:*

$$\nabla_\theta \mathcal{L}_{\text{off}}(\theta) = \mathbb{E}_{c\sim\mathcal{D},x^{1:K}} \nabla_\theta D_{KL}(p_{\text{gt}}||p_\theta). \tag{4}$$

Here $p_\theta(x^{1:K}|c) = \prod_{k=1}^{K} p_\theta(x^k|c)$ represents the likelihood that policy model generates motions $x^{1:K}$ matching their rankings, where $p_\theta(x^k|c) = \frac{(\exp h_\theta(x^k,c))^\beta}{\sum_{j=k}^{K}(\exp h_\theta(x^j,c))^\beta}$.

Online DPO: biased sampling, resulting in even high-preference samples being incorrectly categorized as low-preference motions.

Theorem 2. *Given a reward model r and a reference model π_{ref}, for the online DPO loss $\mathcal{L}_{\text{on}}$, its gradient is:*

$$\nabla_\theta \mathcal{L}_{\text{on}}(\theta) = \mathbb{E}_{c\sim\mathcal{D},x^{1:K}} \nabla_\theta\, p_{\bar\pi_\theta}(x^{1:K}|c) D_{KL}(p_r||p_\theta), \tag{6}$$

where $p_{\bar\pi_\theta}(x^{1:K}|c) = \prod_{k=1}^{K} p_{\bar\pi_\theta}(x^k|c)$ with $p_{\bar\pi_\theta}(x^k|c)$ being the generative probability of policy model to generate x^k conditioned on c, and $p_\theta(x^k) = \frac{(\exp h_\theta(x_k,c))^\beta}{\sum_{j=k}^{K}(\exp h_\theta(x_j,c))^\beta)^\beta}$ denotes the likehood that policy model generates motion x_k with the k-th largest probability.

Contact

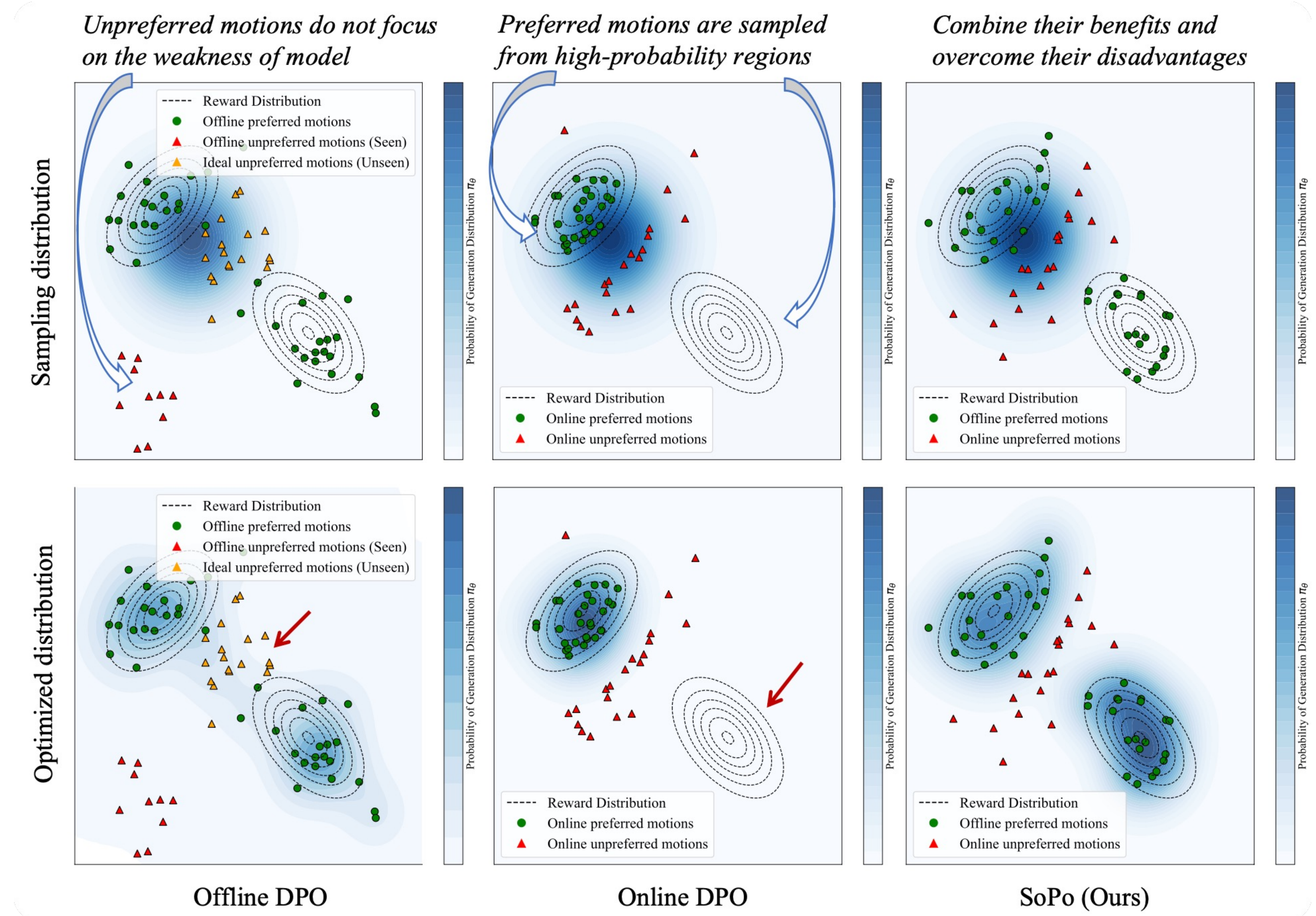
Email: txf0620@gmail.com

WeChat: txf_06_20

X: XiaofengTan85815

First Author | Project Page | Code | Paper

Toy Example: Off-/Online DPO, and SoPo



Semi-Online Preference Optimization

Insight for Unpreferred Motion Sampling

Case 1: The group $\{x^k_{\bar\pi_\theta}\}_{k=1}^K$ contains a low-preference unpreferred motion $x^l_{\bar\pi_\theta}$. Then we select these unpreferred motions iteratively which ensure diversity due to randomness of online generations and address the diversity lacking issue in offline DPO.

Case 2: The group contains no low-preference unpreferred motion $x^l_{\bar\pi_\theta}$, meaning all sampled motions are of high preference and should not be treated as unpreferred. This suggests the model performs well under condition c, so training should focus on high-quality preferred motions from offline data to further enhance generation quality.

1. Distribution Separation

$$p_{\bar\pi_\theta}(x^{1:K}_{\bar\pi_\theta}|c) = \underbrace{p_{\bar\pi_\theta}(x^{1:K}_{\bar\pi_\theta}|c)p_\tau(r(x^l_{\bar\pi_\theta},c)\geq\tau)}_{\text{relatively high-preference unpreferred motions } \bar\pi_\theta^{hu}} + \underbrace{p_{\bar\pi_\theta}(x^{1:K}_{\bar\pi_\theta}|c)p_\tau(r(x^l_{\bar\pi_\theta},c)<\tau)}_{\text{valuable unpreferred motions } \bar\pi_\theta^{vu}},$$

$$\mathcal{L}_{\text{vu}} = -\mathbb{E}_{(x^w,c)\sim\mathcal{D}} Z_{vu}(c)\mathbb{E}_{x^{1:K}_{\bar\pi_\theta}\sim\bar\pi_\theta^{vu*}(\cdot|c)} \log\sigma\big(\beta\mathcal{H}_\theta(x^w, x^l_{\bar\pi_\theta}, c)\big),$$

$$\mathcal{L}_{\text{hu}} = -\mathbb{E}_{(x^w,c)\sim\mathcal{D}} Z_{hu}(c)\mathbb{E}_{x^{1:K}_{\bar\pi_\theta}\sim\bar\pi_\theta^{hu*}(\cdot|c)} \log\sigma\big(\beta\mathcal{H}_\theta(x^w, x^l_{\bar\pi_\theta}, c)\big),$$

2. Training loss amendment

Accordingly, we rewrite the loss $\mathcal{L}_{\text{hu}}(\theta)$ into $\mathcal{L}_{\text{USoPo}-\text{hu}}(\theta)$ for filtering them:

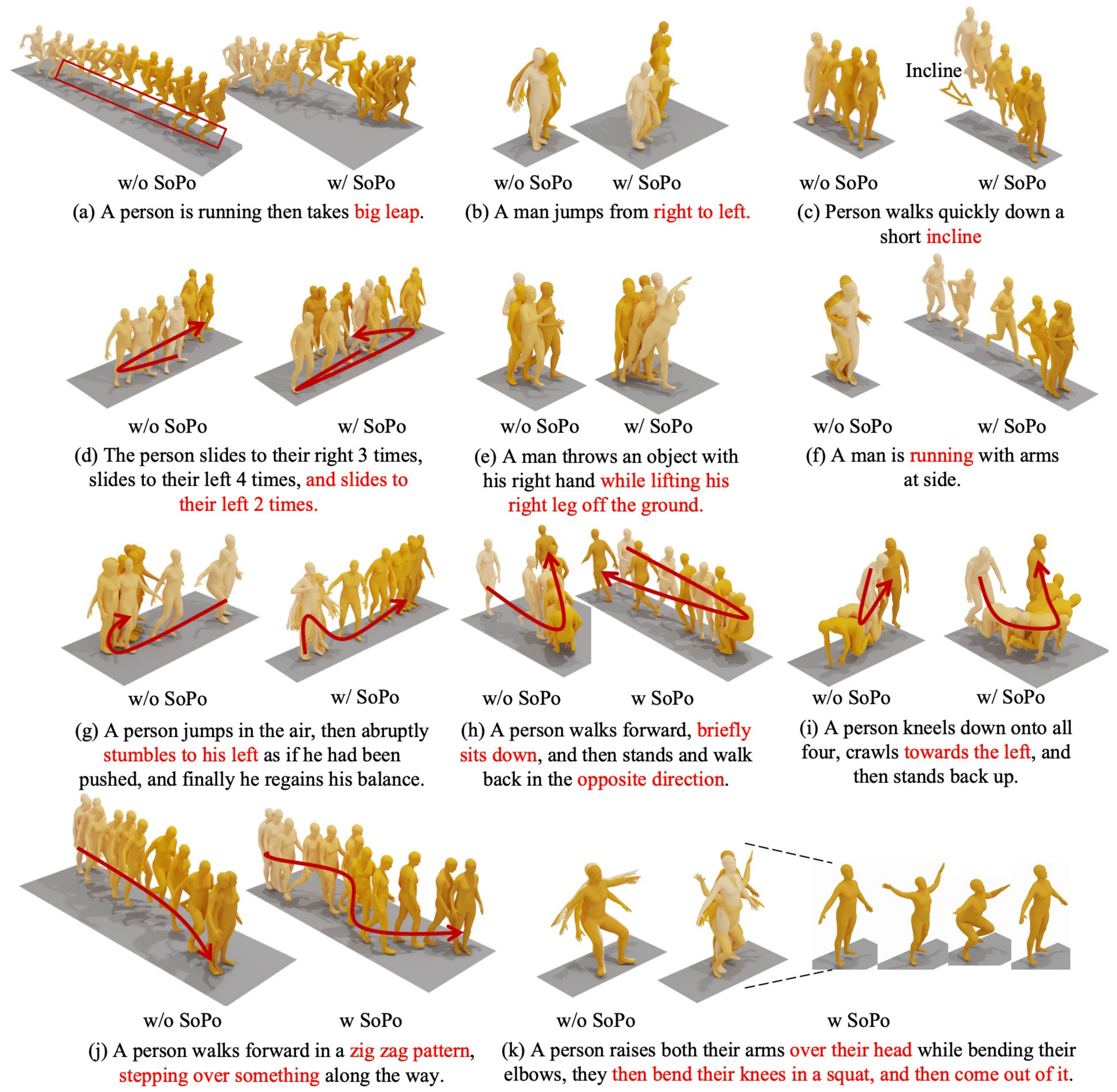
$$\mathcal{L}_{\text{USoPo}-\text{hu}}(\theta) = -\mathbb{E}_{(x^w,c)\sim\mathcal{D}} Z_{hu}(c)\log\sigma\Big(\beta h_\theta(x^w,c)\Big),\quad \mathcal{L}_{\text{USoPo}}(\theta) = \mathcal{L}_{\text{USoPo}-\text{hu}}(\theta) + \mathcal{L}_{\text{vu}}(\theta)$$

3. SoPo for Diffusion

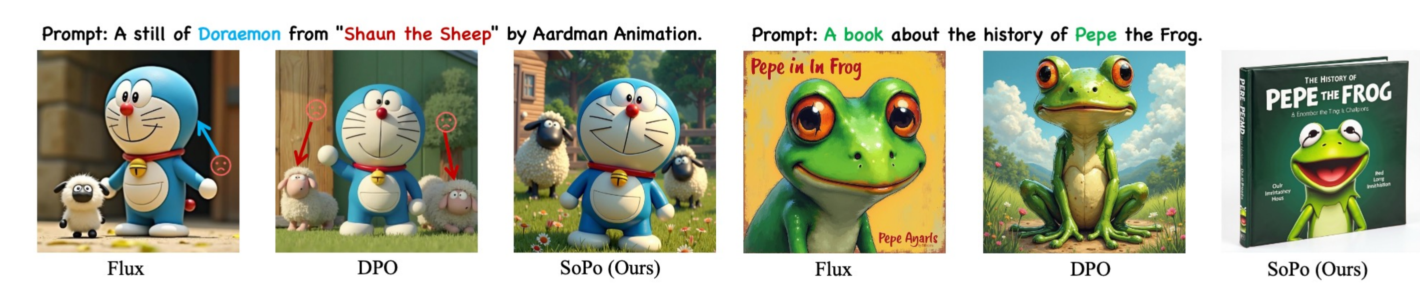
$$\mathcal{L}^{\text{diff}}_{\text{SoPo}}(\theta) = -\mathbb{E}_{t\sim\mathcal{U}(0,T),(x^w,c)\sim\mathcal{D},x^{1:K}_{\bar\pi_\theta}\sim\bar\pi_\theta(\cdot|c)} \begin{cases} \log\sigma\Big(-T\omega_t\big(\beta_w(x_w)\mathcal{L}(\theta,\text{ref},x^w_t) - \beta\mathcal{L}(\theta,\text{ref},x^l_t)\big)\Big), & \text{if } r(x^l,c)<\tau, \\ \log\sigma\Big(-T\omega_t\beta_w(x_w)\mathcal{L}(\theta,\text{ref},x^w_t)\Big), & \text{otherwise.} \end{cases} \tag{17}$$

Experiments

Text-to-Motion Qualitative Results



Text-to-Image Qualitative Results



Text-to-Motion Quantitative Results

Methods	Time*	R-Precision ↑ Top 1	Top 2	Top 3	MM Dist ↓	Diversity →	FID ↓
Real	-	$0.511^{\pm0.003}$	$0.703^{\pm0.003}$	$0.797^{\pm.002}$	$2.974^{\pm0.008}$	$9.503^{\pm0.065}$	$0.002^{\pm0.000}$
MLD [1]	+0 X	$0.453^{\pm0.003}$	$0.679^{\pm0.003}$	$0.755^{\pm0.003}$	$3.292^{\pm0.010}$	$9.793^{\pm0.072}$	$0.459^{\pm0.011}$
+ MoDiPO-T [9]	+121K X	$0.455^{\pm0.002}$	$0.682^{\pm0.003}$	$0.758^{\pm0.002}_{+0.40\%}$	$3.267^{\pm.010}_{+0.76\%}$	$9.747^{\pm0.073}_{+0.046}$	$0.303^{\pm0.031}_{+33.9\%}$
+ MoDiPO-G [9]	+121K X	$0.452^{\pm0.003}$	$0.678^{\pm0.003}$	$0.753^{\pm0.003}_{-0.26\%}$	$3.294^{\pm0.010}_{-0.01\%}$	$9.702^{\pm.075}_{+0.091}$	$0.281^{\pm0.031}_{+38.8\%}$
+ MoDiPO-O [9]	-	$0.406^{\pm0.003}$	$0.609^{\pm0.003}$	$0.677^{\pm0.003}_{-10.3\%}$	$3.701^{\pm0.013}_{-12.4\%}$	$9.241^{\pm.079}_{-0.018}$	$0.276^{\pm0.007}_{+39.9\%}$†
+ SoPo (Ours)	+20 X	$0.463^{\pm0.003}_{+2.21\%}$	$0.682^{\pm0.003}_{+2.23\%}$	$0.763^{\pm0.003}_{+1.06\%}$	$3.185^{\pm0.012}_{+3.25\%}$†	$9.525^{\pm0.065}_{+0.268}$†	$0.374^{\pm0.007}_{+18.5\%}$
MDM [13]	+0 X	$0.418^{\pm0.005}$	$0.604^{\pm0.005}$	$0.703^{\pm0.005}$	$3.658^{\pm0.025}$	$9.546^{\pm0.066}$	$0.501^{\pm0.037}$
+ MoDiPO-T [9]	+121K X	$0.421^{\pm0.006}$	$0.635^{\pm0.005}$	$0.706^{\pm0.004}_{+0.42\%}$	$3.634^{\pm.026}_{+0.66\%}$	$9.531^{\pm0.073}_{+0.015}$	$0.451^{\pm0.031}_{+9.98\%}$
+ MoDiPO-G [9]	+121K X	$0.420^{\pm0.006}$	$0.632^{\pm0.005}$	$0.704^{\pm0.001}_{+0.14\%}$	$3.641^{\pm0.025}_{+0.46\%}$	$9.495^{\pm0.071}_{+0.035}$	$0.486^{\pm0.031}_{+2.99\%}$
MDM (fast) [13]	+0 X	$0.455^{\pm0.006}$	$0.645^{\pm0.007}$	$0.749^{\pm0.004}$	$3.304^{\pm0.023}$	$9.948^{\pm0.084}$	$0.534^{\pm0.052}$
+ SoPo (Ours)	+60 X	$0.479^{\pm0.006}_{+5.27\%}$†	$0.674^{\pm.005}_{+4.50\%}$†	$0.770^{\pm0.006}_{+2.80\%}$†	$3.208^{\pm0.025}_{+2.91\%}$	$9.906^{\pm.083}_{+0.042}$	$0.480^{\pm0.046}_{+10.1\%}$

Methods	Year	R-Precision ↑ Top 1	Top 2	Top 3	Avg.	MM Dist ↓	Diversity →	Multimodal ↑	FID ↓
Real	-	$0.511^{\pm0.003}$	$0.703^{\pm0.003}$	$0.797^{\pm0.002}$	0.670	$2.794^{\pm0.008}$	$9.503^{\pm0.065}$	-	$0.002^{\pm0.000}$
TEMOS [40]	2022	$0.424^{\pm0.002}$	$0.612^{\pm0.002}$	$0.722^{\pm0.002}$	0.586	$3.703^{\pm0.008}$	$8.973^{\pm0.071}$	$0.368^{\pm0.018}$	$3.734^{\pm0.028}$
T2M [3]	2022	$0.457^{\pm0.002}$	$0.639^{\pm0.003}$	$0.740^{\pm0.003}$	0.612	$3.340^{\pm0.008}$	$9.188^{\pm0.002}$	$2.090^{\pm0.083}$	$1.067^{\pm0.002}$
MDM [13]	2022	$0.418^{\pm0.005}$	$0.604^{\pm0.005}$	$0.703^{\pm0.005}$	0.575	$3.658^{\pm0.025}$	$9.546^{\pm0.066}$	$\mathbf{2.799}^{\pm0.072}$	$0.501^{\pm0.037}$
MLD [1]	2023	$0.481^{\pm0.003}$	$0.673^{\pm0.003}$	$0.772^{\pm0.002}$	0.642	$3.196^{\pm0.016}$	$9.724^{\pm0.082}$	$2.413^{\pm0.079}$	$0.473^{\pm0.013}$
MotionGPT [42]	2023	$0.492^{\pm0.003}$	$0.681^{\pm0.003}$	$0.778^{\pm0.002}$	0.650	$3.096^{\pm0.008}$	$\mathbf{9.528}^{\pm0.071}$	$2.008^{\pm0.084}$	$0.232^{\pm0.008}$
MotionDiffuse [14]	2024	$0.491^{\pm0.004}$	$0.681^{\pm0.002}$	$0.782^{\pm0.001}$	0.651	$3.113^{\pm0.018}$	$9.410^{\pm0.049}$	$1.553^{\pm0.042}$	$0.630^{\pm0.011}$
OMG [43]	2024	-	-	$0.784^{\pm0.002}$	-	-	$9.657^{\pm0.085}$	-	$0.381^{\pm0.008}$
Wang et. al. [6]	2024	$0.433^{\pm0.007}$	$0.629^{\pm0.007}$	$0.733^{\pm0.006}$	0.598	$3.430^{\pm0.061}$	$9.825^{\pm0.159}$	2.835	$0.352^{\pm0.109}$
MoDiPO-T [9]	2024	$0.455^{\pm0.003}$	$0.682^{\pm0.003}$	$0.758^{\pm0.002}$	-	$3.267^{\pm0.010}$	$9.747^{\pm0.073}$	$2.663^{\pm0.111}$	$0.303^{\pm0.031}$
PriorMDM [12]	2024	$0.481^{\pm0.002}$	-	-	-	$5.610^{\pm0.023}$	$9.620^{\pm0.074}$	-	$0.600^{\pm0.053}$
LMM-T[1] [41]	2024	$0.496^{\pm0.002}$	$0.685^{\pm0.002}$	$0.785^{\pm0.002}$	0.655	$3.087^{\pm0.012}$	$9.176^{\pm0.074}$	$1.465^{\pm0.048}$	$0.415^{\pm0.002}$
CrossDiff[3] [11]	2024	-	-	$0.730^{\pm0.003}$	-	$3.358^{\pm0.011}$	$9.577^{\pm0.082}$	-	$0.281^{\pm0.016}$
Motion Mamba [7]	2024	$0.502^{\pm0.003}$	$0.693^{\pm0.002}$	$0.792^{\pm0.002}$	0.662	$3.060^{\pm0.009}$	$9.871^{\pm0.084}$	$2.294^{\pm0.058}$	$0.281^{\pm0.011}$
MLD* [1, 2]	2024	$0.504^{\pm0.002}$	$0.698^{\pm0.003}$	$0.796^{\pm0.002}$	0.666	$3.052^{\pm0.009}$	$9.634^{\pm0.064}$	$2.267^{\pm0.082}$	$0.450^{\pm0.011}$
MLD* [2]+ SoPo	2025	$\mathbf{0.528}_{+4.76\%}$	$\mathbf{0.722}_{+3.44\%}$	$\mathbf{0.827}_{+3.89\%}$	$\mathbf{0.692}_{+3.90\%}$	$\mathbf{2.939}_{+3.70\%}$	$9.584_{+38.1\%}$	$2.301^{\pm0.076}$	$\mathbf{0.174}_{+61.3\%}$