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

Anonymous Author(s)
Affiliation
Address
email

Abstract

Text-to-motion generation is essential for advancing the creative industry but often presents challenges in producing consistent, realistic motions. To address this, we focus on fine-tuning text-to-motion models to consistently favor high-quality, human-preferred motions—a critical yet largely unexplored problem. In this work, we theoretically investigate the DPO under both online and offline settings, and reveal their respective limitation: overfitting in offline DPO, and biased sampling in online DPO. Building on our theoretical insights, we introduce Semi-online Preference Optimization (SoPo), a DPO-based method for training text-to-motion models using "semi-online" data pair, consisting of unpreferred motion from online distribution and preferred motion in offline datasets. This method leverages both online and offline DPO, allowing each to compensate for the other's limitations. Extensive experiments demonstrate that SoPo outperforms other preference alignment methods, with an MM-Dist of 3.25% (vs e.g. 0.76% of MoDiPO) on the MLD model, 2.91% (vs e.g. 0.66% of MoDiPO) on MDM model, respectively. Additionally, the MLD model fine-tuned by our SoPo surpasses the SoTA model in terms of R-precision and MM Dist. Visualization results also show the efficacy of our SoPo in preference alignment. Code will be released publicly.

1 Introduction

Text-to-motion generation aims to synthesize realistic 3D human motions based on textual descriptions, unlocking numerous applications in gaming, filmmaking, virtual and augmented reality, and robotics [1–4]. Recent advances in generative models [5–7], particularly diffusion models [1, 2, 8–14], have significantly improved text-to-video generation. However, text-to-motion models often encounter challenges in generating consistent and realistic motions due to several key factors.

Firstly, models are often trained on diverse text-motion pairs where descriptions vary widely in style, detail, and purpose. This variance can cause inconsistencies, producing motions that do not always meet realism or accuracy standards [15, 16]. Secondly, text-to-motion models are probabilistic, allowing diverse outputs for each description. While this promotes variety, it also increases the chances of generating undesirable variations [4]. Lastly, the complexity of coordinating multiple flexible human joints results in unpredictable outcomes, increasing the difficulty of achieving smooth and realistic motion [16]. Together, these factors limit the quality and reliability of current methods of text-to-motion generation.

In this work, we focus on refining text-to-motion models to consistently generate high-quality and human-preferred motions, a largely unexplored but essential area given its wide applicability. To our knowledge, MoDiPO [9] is the only work directly addressing this. MoDiPO applies a preference alignment method, DPO [17], originally developed for language and text-to-image models, to the text-to-motion domain. This approach fine-tunes models on datasets where each description pairs

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

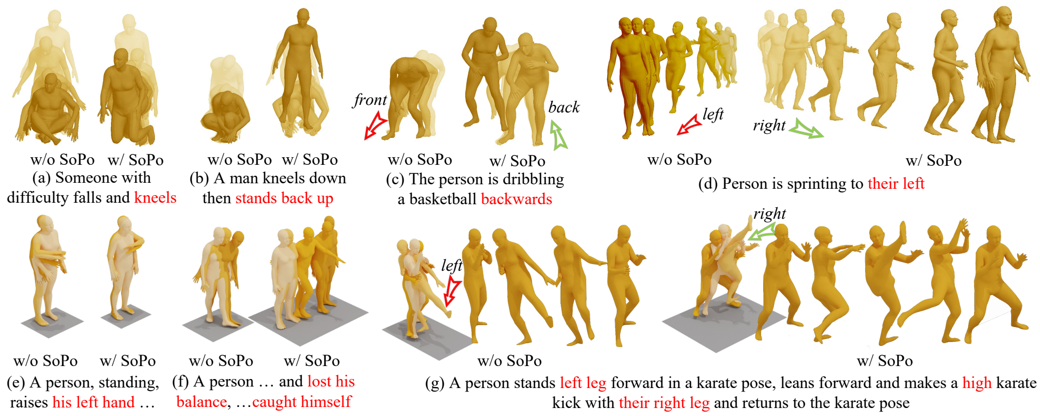

Figure 1: Visual results on HumanML3D dataset. We integrate our SoPo into MDM [13] and MLD [1], respectively. Our SoPo improves the alignment between text and motion preferences.

with both preferred and unpreferred motions, guiding the model toward more desirable outputs. Despite MoDiPO's promising results, challenges remain, as undesired motions continue to arise, as shown in Fig. 1. Unfortunately, this issue is still underexplored, with limited efforts directed at advancing preference alignment approaches to mitigate it effectively.

Contributions. Building upon MoDiPO, this work addresses the above problem, and derives some new results and alternatives for text-to-motion generation alignment. Particularly, we theoretically investigate the limitations of online and offline DPO, and then propose a Semi-Online Preference Optimization (SoPo) to solve the alignment issues in online and offline DPO for text-to-motion generation. Our contributions are highlighted below.

Our first contribution is the explicit revelation of the limitations of both online and offline DPO. Online DPO is constrained by biased sampling, resulting in high-preference scores that limit the preference gap between preferred and unpreferred motions. Meanwhile, offline DPO suffers from overfitting due to limited labeled preference data, especially for unpreferred motions, leading to poor generalization. This leads to inconsistent performance in aligning preferences for existing methods.

Inspired by our theory, we propose a novel and effective SoPo method to address these limitations. SoPo trains models on "semi-online" data pairs that incorporate high-quality preferred motions from offline datasets alongside diverse unpreferred motions generated dynamically. This blend leverages the offline dataset's human-labeled quality to counter online DPO's preference gap issues, while the dynamically generated unpreferred motions mitigate offline DPO's overfitting.

Finally, extensive experimental results like Fig. 1 show that our SoPo significantly outperforms the SoTA baselines. For example, on the HumanML3D dataset, integrating our SoPo into MLD brings 0.222 in Diversity and 3.25% in MM Dist improvement. By comparison, combining MLD with MoDiPO only bring 0.091 and -0.01% respectively. These results underscore SoPo's effectiveness in improving human-preference alignment in text-to-motion generation.

2 Related Works

Text-to-Motion Generation. Text-to-motion generation [10, 18–24] is a key research area with broad applications in computer vision. Recently, diffusion-based models have shown remarkable progress by enhancing both the quality and diversity of generated motions with stable training [2, 11–13]. MotionDiffuse [14] is a pioneering text-driven diffusion model that enables fine-grained body control and flexible, arbitrary-length motion synthesis. Tevet et al. [13] propose a transformer-based diffusion model using geometric losses for better training and performance. Chen et al. [1] improve efficiency by combining latent space and conditional diffusion. Kong et al. [8] enhance diversity with a discrete representation and adaptive noise schedule. Dai et al. [2] present a real-time controllable model using latent consistency distillation for efficient and high-quality generation. Despite these advances, generating realistic motions that align closely with text remains challenging.

Direct Preference Optimization. Preference alignment aims to model preference distributions over different outputs under the same conditions. It has shown great success in large language models (LLMs) [17, 25], text-to-3D [26], and image generation [27–31], offering a promising solution to the aforementioned issue. Existing methods are broadly categorized into offline [27, 32] and online DPO [28–31]. Offline DPO trains on fixed datasets with preference labels from humans [27] or AI feedback [9]. In contrast, online DPO generates data online using a policy [31] or a reference model [29], and forms preference pairs via human [28] or AI feedback [32]. While effective in text-to-image generation, DPO methods for text-to-motion—e.g., MoDiPO [9]—remain underexplored and face challenges such as overfitting and insufficient preference gaps.

3 Motivation: Rethink Offline & Online DPO

Preliminaries. Here we analyze DPO in MoDiPO to explain its inferior alignment performance for text-to-motion generation. To this end, we first briefly introduce DPO [17]. Let $\mathcal{D}$ be a preference dataset which comprises numerous triples, each containing a text condition c and a motion pair $x^w \succ x^l$ where x^w and x^l respectively denote the preferred motion and unpreferred one. With this dataset, Reinforcement Learning from Human Feedback (RLHF) [33] first trains a reward model $r(x, c)$ to access the quality of x under the condition c. Then RLHF maximizes cumulative rewards while maintaining a KL constraint between the policy model π_θ and a reference model π_{ref}:

$$\max_{\pi_\theta} \mathop{\mathbb{E}}_{c\sim\mathcal{D}, x\sim\pi_\theta(\cdot|c)} \left[r(x,c) - \beta D_{\text{KL}}\left(\pi_\theta(x|c) \,\|\, \pi_{\text{ref}}(x|c)\right)\right]. \tag{1}$$

Here one often uses the frozen pretrained model as the reference model π_{ref} and current trainable text-to-motion model as the policy model π_θ.

Building upon RLHF, DPO [17] analyzes the close solution of problem in Eq. (1) to simplify its loss:

$$\mathcal{L}_{\text{DPO}}(\theta) = \mathbb{E}_{(x^w, x^l, c)\sim\mathcal{D}}\Big[-\log\sigma\left(\beta\mathcal{H}_\theta(x^w, x^l, c)\right)\Big], \tag{2}$$

where $\mathcal{H}_\theta(x^w, x^l, c) = h_\theta(x^w, c) - h_\theta(x^l, c)$, $h_\theta(x, c) = \log\frac{\pi_\theta(x|c)}{\pi_{\text{ref}}(x|c)}$, and σ is the logistic function. When there are multiple preferred motions (responses) under a condition c, i.e., $x^1 \succ x^2 \succ \cdots \succ x^K$ $(K \geq 2)$, by using Plackett-Luce model [34], DPO can be extended as:

$$\mathcal{L}_{\text{off}}(\theta) = -\mathbb{E}_{(x^{1:K}, c)\sim\mathcal{D}}\Big[\log\prod_{k=1}^{K}\frac{\exp(\beta h_\theta(x^k, c))}{\sum_{j=k}^{K}\exp(\beta h_\theta(x^j, c))}\Big]. \tag{3}$$

When $K = 2$, $\mathcal{L}_{\text{off}}$ degenerates to $\mathcal{L}_{\text{DPO}}$. Since MoDiPO uses multiple preferred motions for alignment, we will focus on analyze the general formulation in Eq. (3).

3.1 Offline DPO

Analysis. In Eq. (3), its training data are sampled from an offline dataset $\mathcal{D}$. So DPO in Eq. (3) is also called "offline DPO". Here we analyze its preference optimization with its proof in App. B.1

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:*

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

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

Theorem 1 shows that the gradient of offline DPO aligns with the gradient of the forward KL divergence, $D_{KL}(p_{\text{gt}}||p_\theta)$. This suggests that the policy model p_θ (i.e., the trainable text-to-motion model) is optimized to match its distribution with the ground-truth motion preference distribution p_{gt}.

Discussion. However, since training data is drawn from a fixed dataset $\mathcal{D}$, the model risks overfitting, particularly on unpreferred samples. Due to limited annotations, text-to-motion datasets typically contain only one preferred motion group $x_c^{1:K}$ per condition c, making $p_{\text{gt}}(\cdot|c)$ resemble a one-point distribution, i.e., $p_{\text{gt}}(x_c^{1:K}|c) = 1$. In this case, minimizing $D_{\text{KL}}(p_{\text{gt}}\|p_\theta)$ reduces to maximizing likelihood: $\min D_{\text{KL}}(p_{\text{gt}}\|p_\theta) \Leftrightarrow \min -\log p_\theta(x_c^{1:K}|c)$. As a result, offline DPO progressively increases $p_\theta(x_c^{1:K}|c)$, widening the preference gap between preferred and unpreferred motions. As illustrated in Fig. 2, the model primarily learns from the fixed motion group $x_c^{1:K}$ for each c, causing the internal gap within $x_c^{1:K}$ to expand. This overfitting effect, also noted in [35], suggests that with limited unpreferred data, the model learns to avoid only specific patterns (e.g., red regions in Fig. 2) while ignoring many common unpreferred motions. Despite this limitation, the offline dataset is manually labeled and provides valuable preference information, where the gap between preferred and unpreferred motions is large, benefiting learning preferred motions.

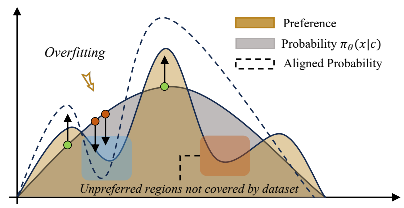

Figure 2: Overfitting in offline DPO: green/red points are preferred/unpreferred motions; blue shows bias from fixed unpreferred data, red indicates uncovered unpreferred regions.

3.2 Online DPO

Analysis. In each online DPO training iteration, the current policy model π_θ generates K samples for a given text c. A pretrained reward model r ranks them by preference as $x^1_{\bar{\pi}\theta} \succ x\bar{\pi}\theta^2 \succ \cdots \succ x\bar{\pi}\theta^K$, where $x\bar{\pi}\theta^i$ is sampled from $\pi\theta$ without gradient backpropagation. Using the Plackett-Luce model [34], the probability of $x^k_{\bar{\pi}_\theta}$ being ranked k-th is given by:

$$p_r(x^k_{\bar{\pi}_\theta}|c) = \frac{\exp r(x^k_{\bar{\pi}_\theta}, c)}{\sum_{i=k}^K \exp r(x^i_{\bar{\pi}_\theta}, c)}. \tag{5}$$

Then we can analyze online DPO below.

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:*

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

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.

See the proof in App. B.2. Theorem 2 indicates that online DPO minimizes the forward KL divergence $D_{KL}(p_r||p_\theta)$. Thus, online DPO trains the policy model π_θ, i.e., the text-to-motion model, to align its text-to-motion distribution with the online preference distribution $p_r(x|c)$.

Discussion. We discuss the training bias and limitations of online DPO. Specifically, motions with high generative probability $p_{\bar{\pi}_\theta}(x_{\bar{\pi}_\theta}|c)$ are frequently synthesized and thus dominate the training of π_θ. In contrast, motions with low generative probability—despite potentially high human preference—are rarely generated and scarcely contribute to training. Notably, when $p_{\bar{\pi}_\theta}(x_{\bar{\pi}_\theta}|c) \to 0$ but the reward $r(x_{\bar{\pi}_\theta}, c) \to 1$, the gradient still vanishes: $\lim_{p_{\pi_\theta}(x_{\bar{\pi}_\theta}|c)\to 0, r(x_{\bar{\pi}_\theta},c)\to 1} \nabla_\theta \mathcal{L}_{\text{on}} = \mathbf{0}$ (see derivation in App. B.2). This highlights a key limitation: online DPO tends to ignore valuable but infrequent preferred motions, focusing instead on commonly generated ones regardless of their actual preference.

Additionally, online DPO aligns the generative probability $p_{\bar{\pi}_\theta}(x_{\bar{\pi}_\theta}|c)$ with the preference distribution $p_r(x_{\bar{\pi}_\theta}|c)$, leading to a positive correlation. Thus, motions with higher generative probabilities often exhibit higher preferences. However, since preference rankings are determined by a reward model, roughly half of these high-preference motions—those with lower rankings k despite high scores $r(x^k_{\bar{\pi}_\theta}, c)$—are still treated as unpreferred. As a result, many unpreferred training motions retain considerable preference, reducing the preference gap compared to manually labeled offline datasets.

On the other hand, online DPO dynamically generates diverse motions, particularly unpreferred motions, in each iteration. This dynamic process enriches preference information and mitigates the overfitting observed in offline DPO, enabling the model to avoid the undesired patterns.

3.3 DPO-based methods for Text-to-Motion

Analysis. DPO in MoDiPO [9] uses an offline dataset $\mathcal{D}$ that is indeed generated by a pre-trained model π_p, denoted as:

$$\begin{cases} x^w_{\pi_p} = \operatorname{argmax}_{x^{1:K}_{\pi_p} \in \bar{\pi}_p} \exp r(x^k_{\pi_p}, c), \\ x^l_{\pi_p} = \operatorname{argmin}_{x^{1:K}_{\pi_p} \in \bar{\pi}_p} \exp r(x^k_{\pi_p}, c), \end{cases} \quad \mathcal{D} = \{(x^w_{\pi_p}, x^l_{\pi_p}, c) | c \in \text{offline textural sets}\}. \tag{7}$$

For discussion, we formulate its sampled distribution as:

$$p^{Mo}_{\text{gt}}(x_w, x_l|c) = \mathbb{I}((x_w, x_l, c) \in \mathcal{D}), \tag{8}$$

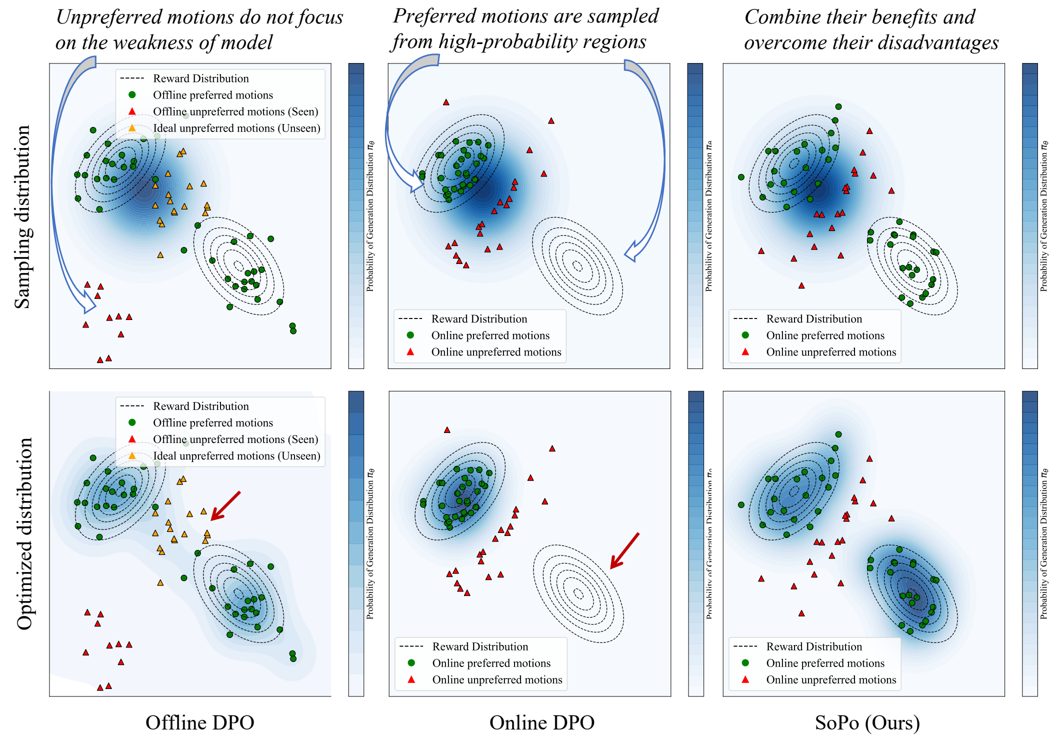

Figure 3: Comparison of offline, online DPO, and our SoPo on synthetic data. Offline DPO suffers from mining unpreferred motions with high probability, and online DPO is limited by biased sampling. Our SoPo utilizes the dynamic unpreferred motions and preferred motions from unbiased offline dataset, overcoming their advantage. Here, the blue region is the distribution of generative model.

where the indication function $\mathbb{I}(\mathcal{E}) = 1$ if event $\mathcal{E}$ happens; otherwise, $\mathbb{I}(\mathcal{E}) = 0$.

From Eq. (7), we observe that, like online DPO, MoDiPO samples preference motions from the distribution $p_{\pi_p}(x|c)$ induced by the pre-trained model π_p. This leads to two main issues like online DPO. 1) Samples with low generative probability $p_{\pi_p}(x|c)$ but high preferences $r(x, c)$ are rarely generated by π_p and thus seldom contribute to training, even though they are highly desirable motions. 2) As discussed in Sec. 3.2, the motions x_{π_p} generated by π_p typically exhibit both high generative probability and preference scores, which causes half of the preferred samples to be selected as unpreferred, skewing the model's learning process. See the detailed discussion in Sec. 3.2.

Additionally, from Eq. (8), we see that for a given condition c, MoDiPO trains on fixed preference data, similar to offline DPO. Consequently, MoDiPO is limited to avoiding only the unpreferred motions valued by the pre-trained model π_p, rather than those relevant to the policy model π_θ. Thus, it inherits the limitations of both online and offline DPO, constraining the alignment performance.

4 Semi-Online Preference Optimization

4.1 Overview of SoPo

We introduce our Semi-Online Preference Optimization (SoPo) to address the limitations in both online and offline DPO for text-to-motion generation. Its core idea is to train the text-to-motion model on semi-online data pairs, where high-preference motions are from offline datasets, while low-preference and high-diversity unpreferred motions are generated online.

As discussed in Sec. 3, offline DPO provides high-preference motions with a clear preference gap from unpreferred ones but tends to overfit due to reliance on fixed, single-source unpreferred motions. In contrast, online DPO benefits from diverse, dynamically generated data but often lacks a sufficient preference gap and overlooks low-probability preferred motions. To leverage the strengths of both, SoPo samples diverse unpreferred motions $x^l_{\pi_\theta}$ from online generation and high-preference motions

$x^w_{\mathcal{D}}$ from offline datasets, ensuring a broad gap between them. Thus, SoPo mitigates the overfitting of offline DPO and the insufficient preference gaps in online DPO. Accordingly, we arrive at our SoPo:

$$\mathcal{L}_{\text{DSoPo}}(\theta) = -\,\mathbb{E}_{(x^w,c)\sim\mathcal{D}}\mathbb{E}_{x^l\sim\bar{\pi}_\theta(x|c)}\log\sigma\Big(\beta\mathcal{H}_\theta(x^w,x^l,c)\Big), \tag{9}$$

where $\mathcal{H}_\theta(x^w,x^l,c)$ is defined below Eq. (2), x^w is preferred motion from the offline dataset, and x^l is unpreferred motion sampled from online DPO. To demonstrate the advantages of SoPo, we conduct experiments on synthetic data, as shown in Fig. 3 (Detailed experimental settings in App. C.1).

However, direct online generation of unpreferred motions from the policy model presents challenges, given the positive correlation between the generative distribution $p_{\bar{\pi}_\theta}$ and preference distribution p_r. Additionally, a large gap between preferred and unpreferred motions remains essential for effective SoPo. In Sec. 4.2 and 4.3, we receptively elaborate on SoPo's designs to address these challenges.

4.2 Online Generation for Unpreferred Motions

Here we introduce our generation pipeline for diverse unpreferred motions. Specifically, given a condition c, we first generate K motions $\{x^k_{\bar{\pi}_\theta}\}_{k=1}^K$ from the policy model π_θ, and select the one with the lowest preference value:

$$x^l_{\bar{\pi}_\theta} = \text{argmin}_{\{x^k_{\bar{\pi}_\theta}\}_{k=1}^K\sim\pi_\theta} r(x^k_{\pi_\theta},c). \tag{10}$$

However, $x^l_{\bar{\pi}_\theta}$ could still exhibit a relatively high preference $r(x^l_{\bar{\pi}_\theta},c)$ due to the positive correlation between the generative probability $p_{\bar{\pi}_\theta}$ and preference distribution p_r (see Sec. 3.2 or 3.3). To identify genuinely unpreferred motions, we apply a threshold τ to the set $\{x^k_{\bar{\pi}_\theta}\}_{k=1}^K$ and check if any preference score is below it. This leads to two training strategies based on the result.

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.

To operationalize this, we apply: 1) distribution separation and 2) training loss amendment.

(1) Distribution separation: With a threshold τ, we separate the distribution $p_{\bar{\pi}_\theta}(x^{1:K}_{\bar{\pi}_\theta}|c)$ into two sub-distributions:

$$p_{\bar{\pi}_\theta}(x^{1:K}_{\bar{\pi}_\theta}|c) = \underbrace{p_{\bar{\pi}_\theta}(x^{1:K}_{\bar{\pi}_\theta}|c)p_\tau(r(x^l_{\bar{\pi}_\theta},c)\geq\tau)}_{\text{relatively high-preference unpreferred motions } \bar{\pi}^{hu}_\theta} + \underbrace{p_{\bar{\pi}_\theta}(x^{1:K}_{\bar{\pi}_\theta}|c)p_\tau(r(x^l_{\bar{\pi}_\theta},c)<\tau)}_{\text{valuable unpreferred motions } \bar{\pi}^{vu}_\theta}, \tag{11}$$

where $p_{\bar{\pi}_\theta}(x^{1:K}|c) = \prod_{k=1}^K p_{\bar{\pi}_\theta}(x^k|c)$, $p_{\bar{\pi}_\theta}(x^k|c)$ is the generative probability of policy model π_θ to generate x^k conditioned on c, $p_\tau(r(x^l_{\bar{\pi}_\theta},c)\geq\tau)$ is the probability of the event $x^l_{\bar{\pi}_\theta}\geq\tau$, and $p_\tau(r(x^l_{\bar{\pi}_\theta},c)\leq\tau)$ has similar meaning.

Eq. (11) indicates that the online generative distribution $\bar{\pi}_\theta(x^{1:K}_{\bar{\pi}_\theta}|c)$ can be separated according to whether the sampled motion $x^{1:K}_{\bar{\pi}_\theta}$ group contains valuable unpreferred motions. Accordingly, our objective loss in Eq. (9) can also be divided into two ones: $\mathcal{L}_{\text{DSoPo}}(\theta) = \mathcal{L}_{\text{vu}}(\theta) + \mathcal{L}_{\text{hu}}(\theta)$, where $\mathcal{L}_{\text{vu}}(\theta)$ targets valuable unpreferred motions and $\mathcal{L}_{\text{hu}}(\theta)$ targets high-preference unpreferred motions:

$$\begin{aligned}\mathcal{L}_{\text{vu}} &= -\,\mathbb{E}_{(x^w,c)\sim\mathcal{D}}Z_{vu}(c)\mathbb{E}_{x^{1:K}_{\bar{\pi}_\theta}\sim\bar{\pi}^{vu*}_\theta(\cdot|c)}\log\sigma\big(\beta\mathcal{H}_\theta(x^w,x^l_{\bar{\pi}_\theta},c)\big),\\ \mathcal{L}_{\text{hu}} &= -\,\mathbb{E}_{(x^w,c)\sim\mathcal{D}}Z_{hu}(c)\mathbb{E}_{x^{1:K}_{\bar{\pi}_\theta}\sim\bar{\pi}^{hu*}_\theta(\cdot|c)}\log\sigma\big(\beta\mathcal{H}_\theta(x^w,x^l_{\bar{\pi}_\theta},c)\big),\end{aligned} \tag{12}$$

where $\mathcal{H}_\theta(x^w,x^l_{\bar{\pi}_\theta},c)$ is defined in Eq. (2), $p_{\bar{\pi}^{vu*}_\theta}(\cdot) = \frac{p^{vu}_{\bar{\pi}_\theta}(\cdot)}{Z_{vu}(c)}$ and $p^{hu*}_{\bar{\pi}_\theta}(\cdot) = \frac{p^{hu}_{\bar{\pi}_\theta}(\cdot)}{Z_{hu}(c)}$ respectively denote the distributions of valuable unpreferred and high-preference unpreferred motions. Here $Z_{vu}(c) = \int p_{\bar{\pi}^{vu}_\theta}(x)\mathrm{dx}$ and $Z_{hu}(c) = \int p_{\bar{\pi}^{hu}_\theta}(x)\mathrm{dx}$ are the partition functions, and are unnecessary to be computed in our implementation (Nore discussion are provided in App. B.3).

(2) Training loss amendment: As discussed above, unpreferred motions in case 2 have relatively high-preference (score $\geq r$), and thus should not be classified into unpreferred motions for training.

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

$$\mathcal{L}_{\mathrm{USoPo-hu}}(\theta) = -\mathbb{E}_{(x^w,c)\sim\mathcal{D}} Z_{hu}(c) \log\sigma\Big(\beta h_\theta(x^w,c)\Big), \quad \mathcal{L}_{\mathrm{USoPo}}(\theta) = \mathcal{L}_{\mathrm{USoPo-hu}}(\theta) + \mathcal{L}_{\mathrm{vu}}(\theta). \tag{13}$$

See more discussion on $\mathcal{L}_{\mathrm{USoPo}}/\mathcal{L}_{\mathrm{DSoPo}}$ in App. B.4.

4.3 Offline Sampling for Preferred Motions

As discussed, online DPO suffers from a limited preference gap between preferred and unpreferred motions. While high-quality motions from offline datasets can help mitigate this issue, they may not always differ significantly from generated motions—especially when the model is well-aligned with the dataset. Thus, motions with larger preference gaps (Sec. 4.2) are crucial and should be prioritized.

To utilize the generated unpreferred motion set $\mathcal{D}_c$ conditioned on c from Sec. 4.2, we calculate its proximity with the unpreferred motions in $\mathcal{D}_c$ using cosine similarity:

$$S(x^w) = \min_{x^k_{\bar{\pi}_\theta}\sim\mathcal{D}_c} \cos(x^w, x^k_{\bar{\pi}_\theta}).$$

Then we reweight the loss using $\beta_w(x_w) = \beta(C - S(x^w))$ with a constant $C \geq 1$:

$$\begin{aligned}\mathcal{L}_{\mathrm{SoPo}}(\theta) = &- \mathbb{E}_{(x^w,c)\sim\mathcal{D}, x^{1:K}_{\bar{\pi}_\theta}\sim\bar{\pi}^{vu*}_\theta(\cdot|c)} Z_{vu}(c)\Big[\log\sigma\Big(\beta_w(x^w)h_\theta(x^w,c) - \beta h_\theta(x^l,c)\Big)\Big] \\ &- \mathbb{E}_{(x^w,c)\sim\mathcal{D}} Z_{hu}(c) \log\sigma\Big(\beta_w(x^w)h_\theta(x^w,c)\Big).\end{aligned} \tag{14}$$

As similar samples have similar preferences, this reweighting strategy guides the model to prioritize preferred motions with a significant preference gap from unpreferred ones. Accordingly, this reweighting strategy relieves and even addresses the small preference gap issue in online DPO.

4.4 SoPo for Diffusion-Based Text-to-Motion

Recently, diffusion text-to-motion models have achieved remarkable success [2, 6, 11, 12], enabling the generation of diverse and realistic motion sequences. Inspired by [27], we derive the objective function of SoPo for diffusion-based text-to-image generation (See proof in App. B.5):

$$\mathcal{L}^{\mathrm{diff}}_{\mathrm{SoPo}} = \mathcal{L}^{\mathrm{diff}}_{\mathrm{SoPo-vu}} + \mathcal{L}^{\mathrm{diff}}_{\mathrm{SoPo-hu}}, \tag{15}$$

$$\begin{aligned}\mathcal{L}^{\mathrm{diff}}_{\mathrm{SoPo-vu}} &= -\mathbb{E}_{t\sim\mathcal{U}(0,T),(x^w,c)\sim\mathcal{D},x^{1:K}_{\bar{\pi}_\theta}\sim\bar{\pi}^{vu*}_\theta(\cdot|c)} Z_{vu}(c)\Big[\log\sigma\Big(-T\omega_t\big(\beta_w(x_w)(\mathcal{L}(\theta,\mathrm{ref},x^w_t) - \beta\mathcal{L}(\theta,\mathrm{ref},x^l_t))\Big)\Big] \\ \mathcal{L}^{\mathrm{diff}}_{\mathrm{SoPo-hu}} &= -\mathbb{E}_{t\sim\mathcal{U}(0,T),(x^w,c)\sim\mathcal{D}} Z_{hu}(c)\Big[\log\sigma\Big(-T\omega_t\beta_w(x_w)\mathcal{L}(\theta,\mathrm{ref},x^w_t)\Big)\Big],\end{aligned} \tag{16}$$

where $\mathcal{L}(\theta,\mathrm{ref},x_t) = \mathcal{L}(\theta,x_t) - \mathcal{L}(\mathrm{ref},x_t)$, and $\mathcal{L}(\theta/\mathrm{ref},x_t) = \|\epsilon_{\theta/\mathrm{ref}}(x_t,t) - \epsilon\|_2^2$ denotes the loss of the policy or reference model. Equivalently, we optimize the following form

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

where $x^l = \mathrm{argmin}_{\{x^k_{\bar{\pi}_\theta}\}^K_{k=1}\sim\pi_\theta} r(x^k_{\pi_\theta}, c)$. Proof and more details are provided in App. A.

5 Experiment

Datasets & Evaluation Metrics. We evaluate SoPo on two widely used datasets, HumanML3D [3] and KIT-ML [36], focusing on two key aspects: alignment and generation quality. Alignment is assessed using R-Precision and MM Dist, while generation quality is measured by Diversity and FID.

Implementation Details. Due to limited preference-labeled motion data, we use existing datasets (e.g., HumanML3D, KIT-ML) as offline preferred motions. For online generation of unpreferred motions, we use TMR, a text-to-motion retrieval model [37], as the reward model. Hyperparameters K and τ are tuned through preliminary experiments to balance performance and efficiency, with $\tau = 0.45$, $C = 2$, and $\beta = 1$ in Eq. (14). We set $K = 4$ for MDM [38] and $K = 2$ for MLD [1]. All models are trained in 100 minutes on a single NVIDIA GeForce RTX 4090D GPU. Since MLD* [2] is tailored for HumanML3D, we use MLD [1] for KIT-ML. Further details are in App. C.2.

Table 1: **Quantitative results of preference alignment methods for text-to-motion generation on the HumanML3D test set.** Results are borrowed from those reported in [9]. The subscripts in each cell denotes the relative performance change. Superscript "†" marks the largest improvement across all models; gray background highlights the largest improvement for each model. "Time*" denotes estimated online/offline motion generation time, with "1X" as the time for MLD [1] to generate all HumanML3D motions and "K" (unspecified in [9], typically 2∼6) as the number of motion pairs.

Methods	Time*	R-Precision ↑			MM Dist ↓	Diversity →	FID ↓
		Top 1	Top 2	Top 3			
Real	-	$.511^{\pm.003}$	$.703^{\pm.003}$	$.797^{\pm.002}$	$2.974^{\pm.008}$	$9.503^{\pm.065}$	$.002^{\pm.000}$
MLD [1]	+0 X	-	-	$.755^{\pm.003}$	$3.292^{\pm.010}$	$9.793^{\pm.072}$	$.459^{\pm.011}$
+ MoDiPO-T [9]	+121K X	-	-	$.758^{\pm.002}{}_{+0.40\%}$	$3.267^{\pm.010}{}_{+0.76\%}$	$9.747^{\pm.073}{}_{+0.046}$	$.303^{\pm.031}{}_{+33.9\%}$
+ MoDiPO-G [9]	+121K X	-	-	$.753^{\pm.003}{}_{-0.26\%}$	$3.294^{\pm.010}{}_{-0.01\%}$	$9.702^{\pm.075}{}_{+0.091}$	$.281^{\pm.031}{}_{+38.8\%}$
+ MoDiPO-O [9]	-	-	-	$.677^{\pm.003}{}_{-10.3\%}$	$3.701^{\pm.013}{}_{-12.4\%}$	$9.241^{\pm.079}{}_{-0.018}$	$.276^{\pm.007}{}_{+39.9\%}{}^{\dagger}$
+ SoPo (Ours)	+20 X	-	-	$.763^{\pm.003}{}_{+1.06\%}$	$3.185^{\pm.012}{}_{+3.25\%}{}^{\dagger}$	$9.525^{\pm.065}{}_{+0.268}{}^{\dagger}$	$.374^{\pm.007}{}_{+18.5\%}$
MDM [13]	+0 X	$.418^{\pm.005}$	$.604^{\pm.005}$	$.703^{\pm.005}$	$3.658^{\pm.025}$	$9.546^{\pm.066}$	$.501^{\pm.037}$
+ MoDiPO-T [9]	+121K X	-	-	$.706^{\pm.004}{}_{+0.42\%}$	$3.634^{\pm.026}{}_{+0.66\%}$	$9.531^{\pm.073}{}_{+0.015}$	$.451^{\pm.031}{}_{+9.98\%}$
+ MoDiPO-G [9]	+121K X	-	-	$.704^{\pm.001}{}_{+0.14\%}$	$3.641^{\pm.025}{}_{+0.46\%}$	$9.495^{\pm.071}{}_{+0.035}$	$.486^{\pm.031}{}_{+2.99\%}$
MDM (fast) [13]	+0 X	$.455^{\pm.006}$	$.645^{\pm.007}$	$.749^{\pm.004}$	$3.304^{\pm.023}$	$9.948^{\pm.084}$	$.534^{\pm.052}$
+ SoPo (Ours)	+60 X	$.479^{\pm.006}{}_{+5.27\%}{}^{\dagger}$	$.674^{\pm.005}{}_{+4.50\%}{}^{\dagger}$	$.770^{\pm.006}{}_{+2.80\%}{}^{\dagger}$	$3.208^{\pm.025}{}_{+2.91\%}$	$9.906^{\pm.083}{}_{+0.042}$	$.480^{\pm.046}{}_{+10.1\%}$

5.1 Main Results

Settings. We evaluate SoPo for preference alignment and motion generation, comparing it with state-of-the-art preference alignment [9] and text-to-motion methods [1, 7]. For fairness, we fine-tune MLD [1] and MDM [13] with SoPo, using a fast diffusion variant [13] with 50 sampling steps. We also fine-tune MLD* [2] as a stronger baseline. Since MLD* is not adapted to KIT-ML, we use MLD [1] and MoMask [39] for diffusion-based and autoregressive methods, respectively.

Comparison with Preference Alignment Methods. Table 1 compares preference alignment methods. MoDiPO, a DPO-based method for motion generation, faces overfitting and biased sampling issues [17]. Conversely, our SoPo method uses diverse high-probability unpreferred and high-quality preferred motions, improving generation quality and reducing unpreferred motions. SoPo excels in most metrics except FID, with R-Precision gains of 5.27%, 4.50%, and 2.80% (vs. baseline 0.42%) and a 3.25% MM Dist. improvement (vs. MoDiPO's -12.4% to $+0.76\%$). SoPo boosts Diversity by 0.268 (vs. MoDiPO's -0.018 to 0.091). Despite MoDiPO's slight FID edge, SoPo's results are comparable, owing to conservative training on low-probability, high-preference samples. SoPo also eliminates pairwise labels and cuts preference data generation time to ∼1/10 of that MoDiPO.

Comparison with Motion Generation Methods. We evaluate SoPo on HumanML3D [3], with results in Table 3. Using preference alignment, SoPo surpasses state-of-the-art methods in R-Precision, MM Dist, and FID, achieving the **best performance**. Although MotionGPT [41] has slightly higher Diversity (9.584 vs. 9.528), SoPo improves R-Precision by 6.46%, FID by 33.5%, and MM Dist by 5.34%. Compared to Motion Mamba and CrossDiff, SoPo increases Diversity by 0.287 and reduces MM Dist by 12.5%. It also enhances MLD*'s FID by 61.3%. On KIT-ML (Table 2), SoPo with MoMask [39] achieves the **best results** across all metrics: Top-k R-Precision (0.446, 0.673, 0.797), MM Dist (2.783), and FID (0.176). MLD with SoPo consistently outperforms its original version, confirming SoPo's effectiveness across various model architectures.

Table 2: **Comparison of text-to-motion generation performance on the KIT-ML dataset.**

Method	R Precision ↑			FID ↓	MM Dist ↓	Diversity →
	Top 1	Top 2	Top 3			
Real	0.424	0.649	0.779	0.031	2.788	11.08
TEMOS [38]	0.370	0.569	0.693	2.770	3.401	10.91
T2M [3]	0.361	0.559	0.681	3.022	2.052	10.72
MLD [1]	0.390	0.609	0.734	0.404	3.204	10.80
T2M-GPT [40]	0.416	0.627	0.745	0.514	3.007	10.86
MotionGPT [41]	0.366	0.558	0.680	0.510	3.527	10.35
MotionDiffuse[14]	0.417	0.621	0.739	1.954	2.958	11.10
Mo.Mamba [7]	0.419	0.645	0.765	0.307	3.021	**11.02**
MoMask [39]	0.433	0.656	0.781	0.204	**2.779**	10.71
MLD [1]$_{+\text{SoPo}}$	0.412	0.646	0.759	0.384	3.107	10.93
MoMask [39]$_{+\text{SoPo}}$	**0.446**	**0.673**	**0.797**	**0.176**	2.783	10.96

5.2 Ablation Studies

Impact of Sample Size K. Due to computational and memory constraints, we recommend keeping $K < 8$. As shown in Table 4, increasing K significantly improves generation quality. A larger sample pool allows the reward model to better evaluate and filter unpreferred motions, leading to more accurate guidance and higher-quality results.

Table 3: **Quantitative comparison of state-of-the-art text-to-motion generation on the HumanML3D test set.** 'MLD*" refers to the enhanced reproduction of MLD [1] from [2]. For a fair comparison, we selected the "LMM-T" [42] with a similar size to ours.

Methods	Year	R-Precision ↑				MM Dist ↓	Diversity →	FID ↓
		Top 1	Top 2	Top 3	Avg.			
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 [38]	2022	0.424$^{\pm0.002}$	0.612$^{\pm0.002}$	0.722$^{\pm0.002}$	0.586	3.703$^{\pm0.008}$	8.973$^{\pm0.071}$	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$^{\pm0.002}$	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$^{\pm0.066}$	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$^{\pm0.082}$	0.473$^{\pm0.013}$
Fg-T2M [5]	2023	0.418$^{\pm0.005}$	0.626$^{\pm0.004}$	0.745$^{\pm0.004}$	0.596	3.114$^{\pm0.015}$	10.930$^{\pm0.083}$	0.571$^{\pm0.047}$
M2DM [8]	2023	0.416$^{\pm0.004}$	0.628$^{\pm0.004}$	0.743$^{\pm0.004}$	0.596	3.015$^{\pm0.017}$	11.417$^{\pm0.082}$	0.515$^{\pm0.029}$
MotionGPT [41]	2023	0.492$^{\pm0.003}$	0.681$^{\pm0.003}$	0.778$^{\pm0.002}$	0.650	3.096$^{\pm0.008}$	**9.528**$^{\pm0.071}$	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$^{\pm0.049}$	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$^{\pm0.159}$	0.352$^{\pm0.109}$
MoDiPO-T [9]	2024	-	-	0.758$^{\pm0.002}$	-	3.267$^{\pm0.010}$	9.747$^{\pm0.073}$	0.303$^{\pm0.031}$
PriorMDM [12]	2024	0.481	-	-	-	5.610	9.620	0.600
LMM-T[1] [42]	2024	0.496$^{\pm0.002}$	0.685$^{\pm0.002}$	0.785$^{\pm0.002}$	0.655	3.087$^{\pm0.012}$	9.176$^{\pm0.074}$	0.415$^{\pm0.002}$
CrossDiff[3] [11]	2024	-	-	0.730	-	3.358	9.577	0.281
Motion Mamba [7]	2024	0.502$^{\pm0.003}$	0.693$^{\pm0.002}$	0.792$^{\pm0.002}$	0.662	3.060$^{\pm0.009}$	9.871$^{\pm0.084}$	0.281$^{\pm0.011}$
MLD* [1, 2]	2023	0.504$^{\pm0.002}$	0.698$^{\pm0.003}$	0.796$^{\pm0.002}$	0.666	3.052$^{\pm0.009}$	9.634$^{\pm0.064}$	0.450$^{\pm0.011}$
MLD* [2] + SoPo	-	**0.528** $_{+4.76\%}$	**0.722** $_{+3.44\%}$	**0.827** $_{+3.89\%}$	**0.692** $_{+3.90\%}$	**2.939** $_{+3.70\%}$	9.584 $_{+38.1\%}$	**0.174** $_{+61.3\%}$

Table 4: **Ablation study on alignment methods, thresholds τ, and sampled number K.**

Methods	R-Precision ↑			MM Dist ↓	Diversity →	FID ↓
	Top 1	Top 2	Top 3			
MDM (fast) [13]	.455	.645	.749	3.304	9.948	.534
+DSoPo	.460 $_{+1.08\%}$	.655 $_{+1.55\%}$	.756 $_{+0.93\%}$	3.297 $_{+0.02\%}$	9.925 $_{+0.033}$	.495 $_{+7.30\%}$
+SoPo w/o VU	.460 $_{+1.08\%}$	.656 $_{+1.71\%}$	.756 $_{+0.93\%}$	3.295 $_{+0.02\%}$	9.915 $_{+0.033}$	.486 $_{+8.98\%}$
+USoPo	.473 $_{+3.96\%}$	.668 $_{+3.57\%}$	.767 $_{+2.40\%}$	3.226 $_{+2.36\%}$	**9.901** $_{+0.047}$	.556 $_{-4.12\%}$
+SoPo	**.479** $_{+5.27\%}$	**.674** $_{+4.50\%}$	**.770** $_{+2.80\%}$	**3.208** $_{+2.91\%}$	9.906 $_{+0.042}$	**.480** $_{+10.1\%}$
+SoPo ($\tau = 0.40$)	.475 $_{+4.40\%}$	.661 $_{+2.48\%}$	.768 $_{+2.53\%}$	3.272 $_{+0.97\%}$	10.04 $_{-0.088}$	.600 $_{-12.4\%}$
+SoPo ($\tau = 0.45$)	**.479** $_{+5.27\%}$	**.674** $_{+4.50\%}$	**.770** $_{+2.80\%}$	**3.208** $_{+2.91\%}$	9.906 $_{+0.042}$	**.480** $_{+10.1\%}$
+SoPo ($\tau = 0.50$)	.468 $_{+2.86\%}$	.663 $_{+2.79\%}$	.764 $_{+2.01\%}$	3.256 $_{+1.45\%}$	9.900 $_{+0.048}$	.491 $_{+8.05\%}$
+SoPo ($\tau = 0.55$)	.466 $_{+2.41\%}$	.660 $_{+1.86\%}$	.763 $_{+1.87\%}$	3.263 $_{+1.24\%}$	9.896 $_{+0.041}$	.430 $_{+19.5\%}$
+SoPo ($\tau = 0.60$)	.461 $_{+1.31\%}$	.656 $_{+1.71\%}$	.758 $_{+1.20\%}$	3.288 $_{+0.48\%}$	**9.803** $_{+0.145}$	**.399** $_{+25.3\%}$
+SoPo ($K = 2$)	**.480** $_{+5.50\%}$	.671 $_{+4.03\%}$	**.771** $_{+2.94\%}$	3.212 $_{+2.78\%}$	9.907 $_{+0.041}$	.502 $_{+5.99\%}$
+SoPo ($K = 4$)	.479 $_{+5.27\%}$	**.674** $_{+4.50\%}$	.770 $_{+2.80\%}$	**3.208** $_{+2.91\%}$	**9.906** $_{+0.042}$	**.480** $_{+10.1\%}$

Impact of Objective Functions. We fine-tune MDM [13] using four objectives: DSoPo (Eq. (12)), USoPo (Eq. (13)), SoPo without value-unpreferred (VU), and full SoPo (Eq. (14)). As shown in Table 4, DSoPo alleviates limitations of offline/online DPO (Sec. 4.1) and improves FID by 7.30%. Removing VU further boosts FID to 8.98% by emphasizing preferred motions that differ from unpreferred ones. USoPo, using a threshold τ to filter unpreferred motions, enhances R-Precision (+3.96%), MM Dist (+2.36%), and Diversity (+0.047), though FID slightly drops (–4.12%). Combining all advantages, SoPo achieves the best results: +5.27% R-Precision and +10.1% FID.

Impact of Cut-Off Thresholds τ. Table 4 reports results with τ ranging from 0.40 to 0.60. A lower τ leads to stricter filtering, yielding more reliable unpreferred motions. As τ decreases, R-Precision and MM Dist improve, indicating better alignment. In contrast, higher τ values improve FID and Diversity, suggesting enhanced generative quality due to exposure to more diverse samples.

Visualization. We visualize results of our SoPo and existing methods, provided in App. C.3.

6 Conclusion

In this study, we introduce a semi-online preference optimization method: a DPO-based fine-tune method for the text-to-motion model to directly align preference on "Semi-online data" consisting of high-quality preferred and diverse unpreferred motions. Our SoPo leverages the advantages both of online DPO and offline DPO, to overcome their own limitations. Furthermore, to ensure the validity of SoPo, we present a simple yet effective online generation method along with an offline reweighing strategy. Extensive experimental results show the effectiveness of our SoPo.

Limitation discussion. SoPo relies on a reward model to motion quality evaluation and identify usable unpreferred samples. However, research on reward models in the motion domain remains scarce, and current models, trained on specific datasets, exhibit limited generalization. Consequently, SoPo inherits these limitations, facing challenges in seamlessly fine-tuning diffusion models with reward models across diverse, open-domain scenarios.

References

[1] Xin Chen, Biao Jiang, Wen Liu, Zilong Huang, Bin Fu, Tao Chen, and Gang Yu. Executing your commands via motion diffusion in latent space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18000–18010, 2023. 1, 2, 7, 8, 9

[2] Wenxun Dai, Ling-Hao Chen, Jingbo Wang, Jinpeng Liu, Bo Dai, and Yansong Tang. Motionlcm: Real-time controllable motion generation via latent consistency model. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, editors, *European Conference on Computer Vision*, pages 390–408, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-72640-8. 1, 2, 7, 8, 9

[3] Chuan Guo, Shihao Zou, Xinxin Zuo, Sen Wang, Wei Ji, Xingyu Li, and Li Cheng. Generating diverse and natural 3d human motions from text. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5142–5151, 2022. 7, 8, 9

[4] Biao Jiang, Xin Chen, Wen Liu, Jingyi Yu, Gang Yu, and Tao Chen. Motiongpt: Human motion as a foreign language. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 20067–20079. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/3fbf0c1ea0716c03dea93bb6be78dd6f-Paper-Conference.pdf. 1

[5] Yin Wang, Zhiying Leng, Frederick W. B. Li, Shun-Cheng Wu, and Xiaohui Liang. Fg-t2m: Fine-grained text-driven human motion generation via diffusion model. In *IEEE/CVF International Conference on Computer Vision*, pages 21978–21987, 2023. 1, 9

[6] Zan Wang, Yixin Chen, Baoxiong Jia, Puhao Li, Jinlu Zhang, Jingze Zhang, Tengyu Liu, Yixin Zhu, Wei Liang, and Siyuan Huang. Move as you say, interact as you can: Language-guided human motion generation with scene affordance. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 433–444, 2024. 7, 9

[7] Zeyu Zhang, Akide Liu, Ian Reid, Richard Hartley, Bohan Zhuang, and Hao Tang. Motion mamba: Efficient and long sequence motion generation. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, editors, *European Conference on Computer Vision*, pages 265–282. Springer Nature Switzerland, 2024. 1, 8, 9

[8] Hanyang Kong, Kehong Gong, Dongze Lian, Michael Bi Mi, and Xinchao Wang. Priority-Centric Human Motion Generation in Discrete Latent Space . In *IEEE/CVF International Conference on Computer Vision*, pages 14760–14770, Los Alamitos, CA, USA, October 2023. IEEE. 1, 2, 9

[9] Massimiliano Pappa, Luca Collorone, Giovanni Ficarra, Indro Spinelli, and Fabio Galasso. Modipo: text-to-motion alignment via ai-feedback-driven direct preference optimization, 2024. URL https://arxiv.org/abs/2405.03803. 1, 3, 4, 8, 9

[10] Ekkasit Pinyoanuntapong, Pu Wang, Minwoo Lee, and Chen Chen. MMM: Generative Masked Motion Model . In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1546–1555. IEEE, 2024. 2

[11] Zeping Ren, Shaoli Huang, and Xiu Li. Realistic human motion generation with cross-diffusion models. *European Conference on Computer Vision*, 2024. 2, 7, 9

[12] Yoni Shafir, Guy Tevet, Roy Kapon, and Amit Haim Bermano. Human motion diffusion as a generative prior. In *The Twelfth International Conference on Learning Representations*, 2024. 7, 9

[13] Guy Tevet, Sigal Raab, Brian Gordon, Yoni Shafir, Daniel Cohen-or, and Amit Haim Bermano. Human motion diffusion model. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=SJ1kSyO2jwu. 2, 8, 9

[14] Mingyuan Zhang, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang, and Ziwei Liu. Motiondiffuse: Text-driven human motion generation with diffusion model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(6):4115–4128, 2024. 1, 2, 8, 9

[15] Qiaosong Qi, Le Zhuo, Aixi Zhang, Yue Liao, Fei Fang, Si Liu, and Shuicheng Yan. Diffdance: Cascaded human motion diffusion model for dance generation. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, page 1374–1382, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701085. doi: 10.1145/3581783.3612307. URL https://doi.org/10.1145/3581783.3612307. 1

[16] Wentao Zhu, Xiaoxuan Ma, Dongwoo Ro, Hai Ci, Jinlu Zhang, Jiaxin Shi, Feng Gao, Qi Tian, and Yizhou Wang. Human motion generation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 1

[17] Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, Red Hook, NY, USA, 2024. Curran Associates Inc. 1, 3, 8

[18] Junfan Lin, Jianlong Chang, Lingbo Liu, Guanbin Li, Liang Lin, Qi Tian, and Chang-wen Chen. Being comes from not-being: Open-vocabulary text-to-motion generation with wordless training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23222–23231, 2023. 2

[19] Jianrong Zhang, Yangsong Zhang, Xiaodong Cun, Yong Zhang, Hongwei Zhao, Hongtao Lu, Xi Shen, and Ying Shan. Generating human motion from textual descriptions with discrete representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14730–14740, 2023.

[20] Zixiang Zhou and Baoyuan Wang. Ude: A unified driving engine for human motion generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5632–5641, 2023.

[21] Ke Fan, Junshu Tang, Weijian Cao, Ran Yi, Moran Li, Jingyu Gong, Jiangning Zhang, Yabiao Wang, Chengjie Wang, and Lizhuang Ma. Freemotion: A unified framework for number-free text-to-motion synthesis. In *European Conference on Computer Vision*, pages 93–109. Springer, 2024.

[22] Mingyuan Zhang, Xinying Guo, Liang Pan, Zhongang Cai, Fangzhou Hong, Huirong Li, Lei Yang, and Ziwei Liu. Remodiffuse: Retrieval-augmented motion diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 364–373, 2023.

[23] Shuai Li, Sisi Zhuang, Wenfeng Song, Xinyu Zhang, Hejia Chen, and Aimin Hao. Sequential texts driven cohesive motions synthesis with natural transitions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9498–9508, 2023.

[24] Matthias Plappert, Christian Mandery, and Tamim Asfour. Learning a bidirectional mapping between human whole-body motion and natural language using deep recurrent neural networks. *Robotics and Autonomous Systems*, 109:13–26, 2018. ISSN 0921-8890. doi: https://doi.org/10.1016/j.robot.2018.07.006. URL `https://www.sciencedirect.com/science/article/pii/S0921889017306280`. 2

[25] Shangmin Guo, Biao Zhang, Tianlin Liu, Tianqi Liu, Misha Khalman, Felipe Llinares, Alexandre Rame, Thomas Mesnard, Yao Zhao, Bilal Piot, et al. Direct language model alignment from online ai feedback. *arXiv preprint arXiv:2402.04792*, 2024. 3

[26] Junliang Ye, Fangfu Liu, Qixiu Li, Zhengyi Wang, Yikai Wang, Xinzhou Wang, Yueqi Duan, and Jun Zhu. Dreamreward: Text-to-3d generation with human preference. *arXiv preprint arXiv:2403.14613*, 2024. 3

[27] Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion Model Alignment Using Direct Preference Optimization . In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8228–8238, Los Alamitos, CA, USA, June 2024. IEEE Computer Society. doi: 10.1109/CVPR52733.2024.00786. URL `https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.00786`. 3, 7

[28] Kai Yang, Jian Tao, Jiafei Lyu, Chunjiang Ge, Jiaxin Chen, Weihan Shen, Xiaolong Zhu, and Xiu Li. Using Human Feedback to Fine-tune Diffusion Models without Any Reward Model . In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8941–8951, Los Alamitos, CA, USA, June 2024. IEEE Computer Society. doi: 10.1109/CVPR52733.2024.00854. URL `https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.00854`. 3

[29] Daoan Zhang, Guangchen Lan, Dong-Jun Han, Wenlin Yao, Xiaoman Pan, Hongming Zhang, Mingxiao Li, Pengcheng Chen, Yu Dong, Christopher Brinton, and Jiebo Luo. Seppo: Semi-policy preference optimization for diffusion alignment, 2024. URL `https://arxiv.org/abs/2410.05255`. 3

[30] Zichen Miao, Zhengyuan Yang, Kevin Lin, Ze Wang, Zicheng Liu, Lijuan Wang, and Qiang Qiu. Tuning timestep-distilled diffusion model using pairwise sample optimization, 2024. URL `https://arxiv.org/abs/2410.03190`.

[31] Zhanhao Liang, Yuhui Yuan, Shuyang Gu, Bohan Chen, Tiankai Hang, Ji Li, and Liang Zheng. Step-aware preference optimization: Aligning preference with denoising performance at each step. *arXiv preprint arXiv:2406.04314*, 2024. 3

[32] Sanghyeon Na, Yonggyu Kim, and Hyunjoon Lee. Boost your own human image generation model via direct preference optimization with ai feedback. *ArXiv*, abs/2405.20216, 2024. URL https://api.semanticscholar.org/CorpusID:270123365. 3

[33] Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. NIPS'17, page 4302–4310. Advances in Neural Information Processing Systems. 3

[34] R. L. Plackett. The analysis of permutations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 24(2):193–202, 1975. ISSN 00359254, 14679876. URL http://www.jstor.org/stable/2346567. 3, 4

[35] Banghua Zhu, Michael Jordan, and Jiantao Jiao. Iterative data smoothing: Mitigating reward overfitting and overoptimization in RLHF. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 62405–62428. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/zhu24e.html. 4

[36] Matthias Plappert, Christian Mandery, and Tamim Asfour. The kit motion-language dataset. *Big Data*, 4 (4):236–252, 2016. doi: 10.1089/big.2016.0028. URL https://doi.org/10.1089/big.2016.0028. PMID: 27992262. 7

[37] Mathis Petrovich, Michael J. Black, and Gül Varol. Tmr: Text-to-motion retrieval using contrastive 3d human motion synthesis. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9454–9463, 2023. doi: 10.1109/ICCV51070.2023.00870. 7

[38] Mathis Petrovich, Michael J. Black, and Gül Varol. TEMOS: Generating diverse human motions from textual descriptions. In *European Conference on Computer Vision*, 2022. 7, 8, 9

[39] Chuan Guo, Yuxuan Mu, Muhammad Gohar Javed, Sen Wang, and Li Cheng. Momask: Generative masked modeling of 3d human motions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1900–1910, 2024. 8

[40] Jianrong Zhang, Yangsong Zhang, Xiaodong Cun, Yong Zhang, Hongwei Zhao, Hongtao Lu, Xi Shen, and Ying Shan. Generating human motion from textual descriptions with discrete representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14730–14740, June 2023. 8

[41] Biao Jiang, Xin Chen, Wen Liu, Jingyi Yu, Gang Yu, and Tao Chen. Motiongpt: Human motion as a foreign language. *Advances in Neural Information Processing Systems*, 36:20067–20079, 2023. 8, 9

[42] Mingyuan Zhang, Daisheng Jin, Chenyang Gu, Fangzhou Hong, Zhongang Cai, Jingfang Huang, Chongzhi Zhang, Xinying Guo, Lei Yang, Ying He, and Ziwei Liu. Large motion model for unified multi-modal motion generation. In *European Conference on Computer Vision*, page 397–421. Springer, 2024. 9

[43] Han Liang, Jiacheng Bao, Ruichi Zhang, Sihan Ren, Yuecheng Xu, Sibei Yang, Xin Chen, Jingyi Yu, and Lan Xu. Omg: Towards open-vocabulary motion generation via mixture of controllers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 482–493, 2024. 9

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims in the abstract accurately reflect our contributions.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of this work in Sec. 6.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: The proof are provided in App. B.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide complete experimental details in Sec. **??**.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: We plan to release the code and detailed documentation after the acceptance of the paper.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We describe the complete experimental details and hyperparameter choices in Sec 5.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The confidence intervals based on 20 independent repetitions are reported in Table 1.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We report the computational resource requirements of our proposed method in Sec 5.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: Our research aligns with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: Our paper focuses on preference alignment for motion generation without addressing broader societal implications.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We plan to release the code and datasets after the acceptance of the paper.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We plan to release the code and detailed documentation after the acceptance of the paper.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This work does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This work does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: This paper does not use LLMs.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.