ReinDiffuse: Crafting Physically Plausible Motions with Reinforced Diffusion Model

Gaoge Han*
Northwest A&F University
hangaoge@nwafu.edu.cn

Mingjiang Liang*
University of Technology Sydney
breeliang@gmail.com

Jinglei Tang†
Northwest A&F University
tangjinglei@nwsuaf.edu.cn

Yongkang Cheng
Tencent AI Lab
2571642004@qq.com

Wei Liu
University of Technology Sydney
wei.liu@uts.edu.au

Shaoli Huang†
Tencent AI Lab
shaol.huang@gmail.com

Method	FID↓	Penetrate↓	Float↓	Skate↓
MDM [3]	0.544	11.291	18.876	1.406
PhysDiff [4]	0.433	0.998	2.601	0.512
ReinDiffuse(Ours)	**0.385**	**0.000**	**1.126**	**0.363**

Table 1. Text-to-motion results using PhysDiff's metrics on HumanML3D [1]. ↓ means closer to real is better. **Bold** indicate the best results.

Quantitative Comparison using PhysDiff's Metrics. In Table 1, we present the quantitative comparison results using PhysDiff's metrics on the HumanML3D. We demonstrate superior performance in these indicators compared to PhysDiff. PhysDiff's metrics mainly focus on three physical issues including Penetrate, Float and Skate. For ground penetration (Penetrate), the distance between the ground and the lowest body mesh vertex below the ground is calculated. For floating (Float), the distance between the ground and the lowest body mesh vertex above the ground is computed. To account for geometry approximation, both Penetrate and Float have a 5 mm tolerance. For foot sliding (Skate), foot joints that make contact with the ground in two consecutive frames are identified, and their average horizontal displacement within these frames is calculated.

Method	FID↓	R-Precision↑	Skate ratio→	Float (m)→	Penetrate (m)↓	Clip (m)↓
Real	0.002	0.797	0.057	0.704	0.000	0.000
MDM [3]	0.544	0.611	0.102	1.757	0.048	0.014
Ours(w/o RL)	0.423	0.608	0.078	1.261	0.031	0.009
Ours(w/o IS)	0.428	0.613	0.069	0.911	0.016	0.005
Ours	**0.385**	**0.622**	**0.058**	**0.711**	**0.000**	**0.000**

Table 2. Ablation studies on HumanML3D [1]. "IS" denotes Importance Sampling. "RL" denotes Reinforcement Learning.

Ablation Studies on The Effect of Importance Sampling. In reinforcement learning training, we use importance sampling techniques with the initialization of a pretrained model. The pretrained policy is handling exploration, while the RL policy utilizes the exploration rewards for updates. To validate the effectiveness of this strategy, we conduct experiments on vanilla reinforcement learning, as shown in Table 2. Specifically, vanilla reinforcement learning employs a single RL policy responsible for both exploration and exploitation. From the experimental results, it is evident that our reinforcement learning strategy achieves better physical fidelity and motion quality. This superior performance is due to its effective separation of exploration and exploitation in reinforcement learning scenarios, leading to better adaptability and efficiency.

Ablation Studies on The Effect of using Reinforcement Learning. In Table 2, we also investigate supervised-based fine-tuning without using reinforcement learning. As previously mentioned, our reward function is not directly differentiable, and therefore, we employ the masked loss for supervised fine-tuning. The experimental results demonstrated that using reinforcement learning is superior to supervised fine-tuning. Our physical rewards are calculated without GT measurements for motion plausibility, making gradient-based learning difficult. The success of RLHF [2] also demonstrate the advantage of enhancing the model's alignment capabilities on probability outputs using Reinforcement Learning.

*Gaoge Han and Mingjiang Liang are co-first authors.
†Corresponding author: Jinglei Tang, Shaoli Huang.

References

[1] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5152–5161, June 2022. 1

[2] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances*

in Neural Information Processing Systems, 35:27730–27744, 2022. 1

[3] Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Amit H Bermano, and Daniel Cohen-Or. Human motion diffusion model. *arXiv preprint arXiv:2209.14916*, 2022. 1

[4] Ye Yuan, Jiaming Song, Umar Iqbal, Arash Vahdat, and Jan Kautz. Physdiff: Physics-guided human motion diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16010–16021, 2023. 1