Skeleton2Stage: Reward-Guided Fine-Tuning for Physically Plausible Dance Generation Supplementary Materials

Jidong Jia, Youjian Zhang, Huan Fu, and Dacheng Tao, *Fellow, IEEE,*

- Section I: Introduction of the video demo.
- Section II: Ablation study on the training strategy of the imitation policy.
- Section III: The relationship between the length of dance sequences and diversity.
- Section IV: Details about the reproduction of Morph.
- Section V: Applying our method to Bailando++.

I. Video Demo

We further provide some video examples in the supplementary materials:
(1) “ours_vs_EDGE”;
(2)“freezing_motions_by_FACT”;
(3)“Ours_vs_EDGE_w_PhysDiff”;
(4)“freezing_moitions_wo_anti-freezing_reward”.

a) “ours_vs_EDGE”: consists of videos that compare EDGE with our proposed method to further prove the effectiveness of our method. The motions named “ours” are generated by our method, while the motions named “EDGE” are generated by EDGE. We can see that our proposed method can reduce the penetration significantly. Referring to Fig. 2 and Table 1 in the main paper for analysis.

b) “freezing_motions_by_FACT”: consists of the freezing motions generated by FACT to prove that FACT tends to generate freezing motions. We can observe that the visual quality of the motions by FACT is diminished because of the freezing motions. Referring to Table 4 in the main paper for analysis.

c) “Ours_vs_EDGE_w_PhysDiff”: includes the comparison between “EDGE w/ PhysDiff*” and Ours to prove the effectiveness of RLFT. We can observe that the imitation policy struggled to imitate the complex generated dances, resulting in jittering and even falling. Referring to Fig. 3 and Table 2 in the main paper for analysis.

d) “freezing_moitions_wo_anti-freezing_reward”: consists of the comparison between our proposed method and Ours w/o AF to prove the effectiveness of anti-freezing reward. We can find that with the anti-freezing reward, the generated motions are more dynamic, while the motions without the anti-freezing reward have a relatively smaller magnitude and slower speed, which tend to freeze. Referring to Fig. 3 and Table 4 in the main paper for analysis.

J. Jia is with School of Computer Science, Shanghai Jiao Tong University, Shanghai 200240, China (jjd1123@sjtu.edu.cn).
Y. Zhang is with Bosch, Shanghai 201206, China (Youjian.ZHANG@cn.bosch.com).
H. Fu is with Youku, Alibaba, Beijing 100124, China (hufu6371@uni.sydney.edu.au).
D. Tao is with Nanyang Technological University, Singapore (dacheng.tao@ntu.edu.sg).

II. Ablation Study on Training Imitation Policy

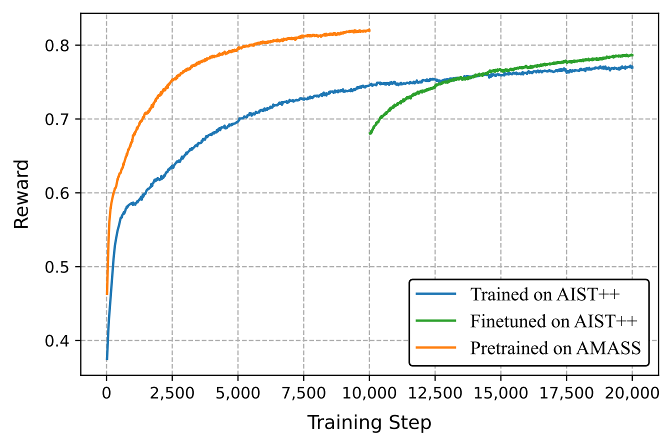

Fig. 1. Training Curves of the imitation policy trained on different datasets.

To enhance the imitation policy’s ability to mimic dance, we first train the imitation policy on the AMASS dataset [1] and fine-tune it on the AIST++ dataset [2].

Here, we conducted ablation studies on the imitation policy training strategy. Specifically, we compared training the imitation policy solely on AIST++ [2] with the strategy that pre-trained the imitation policy on AMASS [1] and then fine-tuned it on AIST++ [2]. From Fig. 1, we can observe that the second training strategy can produce an imitation policy with a stronger imitation ability. From the results in Table II, we can observe that using the imitation policy trained both on AMASS [1] and AIST++ [2] can derive a better “Penetration Rate” and “PFC”. However, for the “Penetration Rate”, we can find that the improvement is subtle. This is because the imitation policy is primarily designed to faithfully replicate the given motions and does not explicitly account for physical plausibility. Instead, the physical plausibility mainly comes from the physical constraints imposed by the physical simulator, which is instilled into the diffusion model through RLFT. For the “PFC”, the imitation policy with lower imitation ability

TABLE I
FINETUNING BAILANDO++ WITH OUR REWARDS

Method	FID_k /FID_g	Pen. Rate ↓	PFC ↓	FGD ↓	BAS ↑	Div_k /Div_g →
Bailando++ w/o RL	28.87/9.45	136.62	1.5474	46.4480	0.2204	7.49/6.78
Bailando++ w/ BA	21.71/9.66	130.10	1.9124	49.7011	0.2383	6.63/7.01
Bailando++ w/ Ours	23.98/12.35	**112.49**	**1.2722**	**34.2014**	**0.2468**	7.56/7.08

may fail to imitate certain motions, causing the controlled character to fall. This falling case will affect the evaluation of the reward and will cause abnormal foot-ground contact, ultimately degrading "PFC".

In conclusion, the physical plausibility primarily depends on the physical constraints of the physical simulator, which is then instilled into the diffusion model by RLFT. However, an imitation policy with a relatively strong ability is also essential to ensure that most physically plausible motions can be replicated. Thus we can get a more accurate imitation reward during RLFT.

TABLE II
THE ABLATION STUDY ON THE TRAINING DATASET OF IMITATION POLICY

Method	Dataset	Penetration Rate ↓	PFC↓
EDGE	-	176.44	0.8715
Ours	①	94.65	0.9712
	①&②	90.38	0.7928

① represents AIST++ [2] and ② represents AMASS [1]. The column "Dataset" means the dataset used to train the imitation policy.

III. DIVERSITY OF THE DANCE RESULTS

In this section, we further evaluate the diversity of generated dances on 5-second dance clips. By comparing the results calculated on 5-second and 20-second dance clips in Table III, we can observe that the diversity of the generated dance increases when tested on 5-second clips. This also supports the statement in the main paper.

TABLE III
DIVERSITY OF GENERATED DANCES ON 5-SECOND AND 20-SECOND CLIPS

Method	Div_k /Div_g →
EDGE(20s)	2.03/3.79
Ours(20s)	1.95/4.16
EDGE(5s)	3.55/5.03
Ours(5s)	3.55/4.79
Ground Truth	8.72/7.77

IV. THE REPRODUCTION OF MORPH

Since the authors only open-sourced the imitation policy they use, which is based on ASE [3], and they didn't provide any guidance on its integration with the generation model (e.g., EDGE). So we attempted to train their imitation policy on dance motions generated by EDGE. We then utilized this trained policy to imitate the generated dances. However, despite the reward curve in Fig. 2 indicating convergence, we observed a success rate of only 12.5%.

Furthermore, an analysis of the released code reveals that the imitation policy does not incorporate body collision checks. Consequently, their method fails to bridge the gap between skeleton motion generation and body mesh visualization (e.g., body interpenetration). Moreover, as a method necessitating an imitation policy during inference, it suffers from limitations similar to those of PhysDiff, as detailed in Section V-C in the main paper.

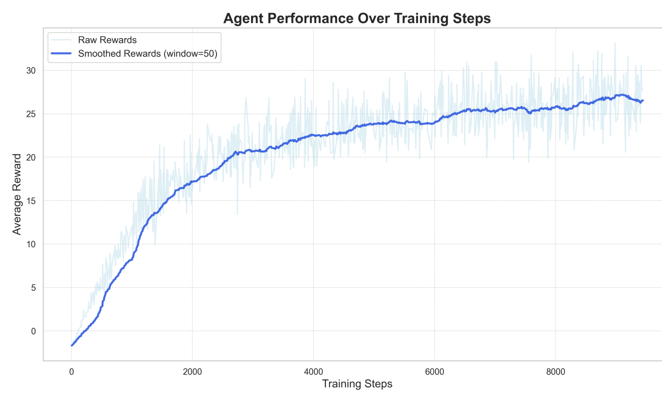

Fig. 2. Morph Reward Curve.

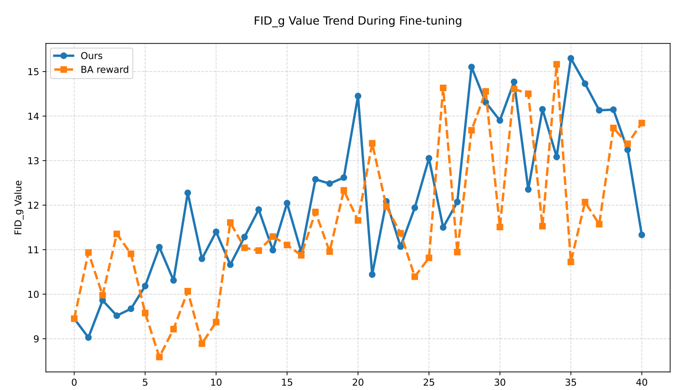

Fig. 3. The Performance of FID_g during the fine-tuning process of Bailando++.

V. FINETUNING BAILANDO++ WITH OUR REWARDS

We also try finetuning the Bailando++ using the imitation and anti-freezing reward. As Bailando++ is not a diffusion model, we finetune it using the same method as [4]. The results are show in Table I, where "Bailando++ w/ BA" means finetuning Bailando++ with Beat Alignment Reward as suggested in [4]. We find that our method also exhibits superior improvement among all metrics. And $\mathrm{FID_g}$ becomes a bit worse because of the distribution shift during RL-finetuning as shown in Fig. 3.

REFERENCES

[1] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "AMASS: Archive of motion capture as surface shapes," in *International Conference on Computer Vision*, Oct. 2019, pp. 5442–5451.

[2] R. Li, S. Yang, D. A. Ross, and A. Kanazawa, "Learn to dance with aist++: Music conditioned 3d dance generation," 2021.

[3] X. B. Peng, Y. Guo, L. Halper, S. Levine, and S. Fidler, "Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters," *ACM Trans. Graph.*, vol. 41, no. 4, Jul. 2022.

[4] L. Siyao, W. Yu, T. Gu, C. Lin, Q. Wang, C. Qian, C. C. Loy, and Z. Liu, "Bailando++: 3d dance gpt with choreographic memory," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 12, pp. 14 192–14 207, 2023.