

# Enhancing Dance-to-Music Generation via Negative Conditioning Latent Diffusion Model

Changchang Sun<sup>1</sup>   Gaowen Liu<sup>2</sup>   Charles Fleming<sup>2</sup>   Yan Yan<sup>1\*</sup>

<sup>1</sup>University of Illinois Chicago   <sup>2</sup>Cisco Research

{csun47, yyan55}@uic.edu   {gaoliu, chfleming}@cisco.com

## 1. Additional Results

To show the input-output correspondence in terms of dance-music beats and the quality of generated music of PN-Diffusion and baseline methods (CDCD [1] and LORIS [2]), we provide complementary qualitative demos of these three methods in the Supplementary Material, where the dance segments corresponding to different times are similar. By comparing these dance videos, we can easily draw the conclusion that the performance of PN-Diffusion in generating highly-synchronized and high-quality music for the given dance video is better than baseline methods. In details, for both CDCD and LORIS, they mainly produce music that is merely a collection of rhythmic sounds and is significantly different from the actual music needed. In contrast, the music generated by our method(PN-Diffusion) is more in line with real-world requirements.

## References

- [1] Ye Zhu, Yu Wu, Kyle Olszewski, Jian Ren, Sergey Tulyakov, and Yan Yan. Discrete contrastive diffusion for cross-modal music and image generation. In *ICLR*, 2023. 1
- [2] Jiashuo Yu, Yaohui Wang, Xinyuan Chen, Xiao Sun, and Yu Qiao. Long-term rhythmic video soundtracker. In *ICML*, pages 40339–40353, 2023. 1

---

\*corresponding author