Synthesizing motions of arbitrary skeletal topologies from a single training example using denoising diffusion probabilistic models

Author: DXZB7

Supervisor: Dr. Yuzuko Nakamura

Current Progress

1. Literature review – reviewed the following papers pertaining to state-of-the-art techniques of motion generation, as well as ways to modify generative models to only work with a single training example in cases where data is sparse:
    a) Li, Peizhuo, Kfir Aberman, Zihan Zhang, Rana Hanocka, and Olga Sorkine-Hornung. "GANimator: Neural Motion Synthesis from a Single Sequence." ACM Transactions on Graphics 41, no. 4 (July 2022): 1–12. https://doi.org/10.1145/3528223.3530157.
    b) Nikankin, Yaniv, Niv Haim, and Michal Irani. "SinFusion: Training Diffusion Models on a Single Image or Video." arXiv, June 19, 2023. http://arxiv.org/abs/2211.11743.
    c) Tevet, Guy, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit H. Bermano. "Human Motion Diffusion Model." arXiv, October 3, 2022. http://arxiv.org/abs/2209.14916.
    d) Zhang, Mingyuan, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang, and Ziwei Liu. "MotionDiffuse: Text-Driven Human Motion Generation with Diffusion Model." arXiv, August 31, 2022. http://arxiv.org/abs/2208.15001.
2. Used Adobe Mixamo to create several animations with non-human skeleton topology to use as training examples in the proposed model. Since animations are saved as industry standard .bvh files, I have also implemented a parser in Python to convert them into the

matrix representation suggested by papers in my literature review. Also created a parser for the reverse process to convert any matrix representation generated by the proposed model into a standard .bvh file. This is important since the .bvh files can be read by software such as Blender, which will be used for qualitative analysis of generated animations.

3. Working on initial prototype of diffusion model. Main aim of the model is to work with only a single training example, so model design considerations revolve around ensuring generation diversity. Using results from papers like SinFusion, a CovNext-based architecture is being implemented.

Remaining Work

1. Finish off prototype of CovNext-based architecture to generate results from training examples. Preliminary evaluation of model outputs can be done to measure model diversity.
2. Investigate implementation of losses such as velocity loss, position loss, and foot-contact loss to improve coherence of generated motions.
3. Investigate differences between 1D and 2D convolutions on the output quality. While 2D is more suited to spatial data such as images, it is likely that the temporal nature of animations works better with 1D convolutions.