Master internship subject

3D Human Motion Diffusion Model

Hosting institute

ICube Laboratory (The Engineering science, computer science and imaging laboratory) at the University of Strasbourg is a leading research center in Computer Science, with more than 300 permanent researchers, with the recently opened AI graduate school supported by the French government.

Work place and salary

The thesis work will take place in the MLMS (Machine Learning, Modeling & Simulation) research team of the ICube laboratory (The Engineering science, computer science and imaging laboratory) of the University of Strasbourg, a leading research center with more than 300 permanent researchers. The workplace is located on the hospital site of the laboratory, a 10-minute walk from the heart of downtown Strasbourg, listed as a UNESCO World Heritage Site.

650 euros net monthly

Supervisors

– director: Hyewon Seo (ICube, Univ. Strasbourg)

– co-supervisors: Cédric Bobenrieth (ICAM, Strasbourg)

Staring date

February – April 2025.

Work description

Human motion generation is a key task in computer graphics, crucial for applications involving virtual characters, such as film production or virtual reality experiences. Recent deep learning methods, particularly generative models, started to make significant contributions in this domain. While early neural methods focused on the unconditional generation of vivid and realistic human motion sequences, more recent methods guide the motion generation using various conditioning signals, including action class, text, and audio. Among them, the diffusion-based model has shown significant success, dominating research frontiers [TRG*23, KKC23, ZCP*24, DMGT23].

Motivated by these recent successes, we will develop action-conditioned human motion generator based on a diffusion model. In particular, we will aim at the generation of daily actions in residential settings, in the view of augmenting training data for the action recognition models. To achieve this goal, we will deploy a diffusion-based motion generation, based on our previous works [ZFY*24, XS24]. To condition the generation using an action class or a text description, we will adopt CLIP [RKH*21] as a text encoder to embed the text prompt and use a trainable tensor as embeddings for different action classes.

The work can be structured as the following tasks:

1. **Data rearrangements:** We will rearrange/select the datasets that are at our disposal in such a way that they can be seamlessly used as training data.
2. **Unconditional motion generation:** The first step is to train a diffusion model
3. **Action-conditioned motion generation**: The aforementioned model will be extended towards conditional generation task. To achieve highly precise conditioned sampling without the need for training auxiliary models, we will take a classifier-free guidance approach.
4. **Experiments**: The developed model will be parameter-tuned, tested, and compared with the state-of-the-art models. Additionally, a number of ablation studies will be conducted to assess the impact of various components on performance.

Candidate profile

− Solid programming skills in Python

− Working skills in Blender for 3D modeling and animation

− Experience in Deep Learning (Diffusion model)

− Good communication skills

Application

Send your CV and academic records (Bachelor and Master) to cedric.bobenrieth@icam.fr and seo@unistra.fr, for (a) possible interview(s).

Bibliography

[DMGT23] DABRAL R., MUGHAL M. H., GOLYANIK V., THEOBALT C.: Mofusion: A framework for denoising-diffusion-based motion synthesis. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (2023), pp. 9760–9770.

[KKC23] KIM J., KIM J., CHOI S.: Flame: Free-form language-based motion synthesis & editing. In Proceedings of the AAAI Conference on Artificial Intelligence (2023), vol. 37, pp. 8255–8263.

[RKH∗21] G., AGARWAL S., SASTRY G., ASKELL A., MISHKIN P., CLARK J., et al.: Learning transferable visual models from natural language supervision. In International conference on machine learning (2021), PMLR, pp. 8748–8763.

[TRG∗23] TEVET G., RAAB S., GORDON B., SHAFIR Y., COHEN-OR D., BERMANO A. H.: Human motion diffusion model. In The Eleventh International Conference on Learning Representations (2023).

[XS24] Xue K. and Seo H., Shape Conditioned Human Motion Generation with Diffusion Model, arXiv preprint https://arxiv.org/abs/2405.06778.

[ZCP∗24] ZHANG M., CAI Z., PAN L., HONG F., GUO X., YANG L., LIU Z.: Motiondiffuse: Text-driven human motion generation with diffusion model. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024).

[ZFY*24] Zou K., Faisan S., Yu B., Valette S., Seo H., 4D Facial Expression Diffusion Model, ACM Transactions on Multimedia Computing, Communications and Applications, https://dl.acm.org/doi/10.1145/3653455.