Master 2 internship subject:
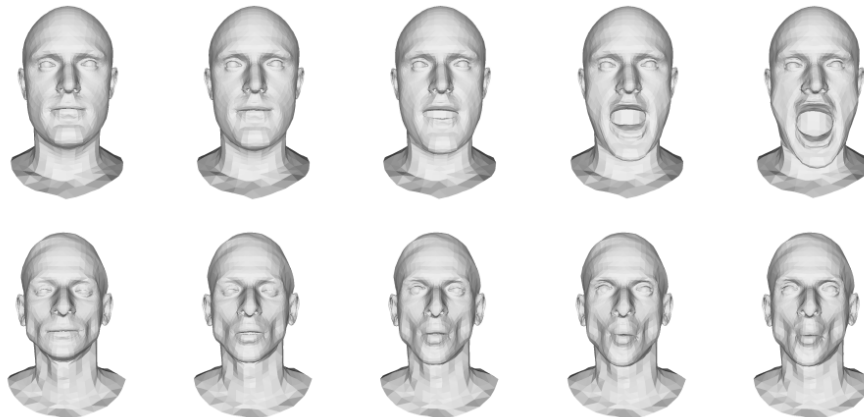# 4D Human Action Diffusion Model

Figure 1: Facial expression sequences generated by our model. The reverse process of the diffusion model has been guided by the expression label: "mouth extreme" (top), and "cheeks in" (bottom).

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.

— CNRS, Alsace delegation, France, https://www.alsace.cnrs.fr/

Work place and salary

The internship 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.

Salary : 600€/month approximately for the duration of 6 months.

Supervisors

— Hyewon Seo, Kaifeng Zou, Sylvain Faisan {seo, kaifeng.zou, faisan}@unistra.fr

Staring date

February – April 2023.

Context

Robot vision for human cognition often fails to work well in the real-world situation, despite the disruptive results achieved in Computer Vision and Artificial Intelligence. While most training data have been

collected in well-conditioned, easy-to-isolate backgrounds, wild videos from the real-world may contain various environmental conditions such as lighting, background patterns, and, most notoriously, occlusions. The latter becomes the source of recurrent problems of human cognition by care-robots in the in-house situation. Large variations in body shapes, motions, clothes, and frequent interactions with objects also contribute to the difficulty. One promising way to improve the cognition performance is to augment the training data, which is expensive, unfortunately. Our objective is to develop a 4D generative model, which will be used to generate synthetic dataset for such vision-based human cognition tasks.

Work description

Generative models based on Denoising Diffusion Probabilistic Model (DDPM) [1] have shown remarkable results, as have shown by some of the recent works on image synthesis [2], point cloud generation [3], and human motion generation [4, 5]. Recently, we have successfully shown to adapt DDPM to the 3D face expression (i.e. 4D face) generation task. While the model has been trained unconditionally, its reverse process can be conditioned by various condition signals, such as expression labels, text, partial sequences, or simply a facial geometry. This allows us to efficiently develop several downstream tasks involving various conditional generation, many of which have shown to outperform state-of-the-art methods.

In this internship, we will extend our 4D Facial expression Diffusion Model to a human action diffusion model. Several downstream tasks will be defined, for each of which a conditional generation will be developed.

Candidate profile

— Background in computer science, machine learning, deep learning or signal processing

— Programming experience in Python and Pytorch/Tensorflow

— Good notion of kinematic modeling is a plus

— Good communication skillsApplication

Application

Send your CV and your academic transcriptions (Bachelor and Master) to seo@unistra.fr.

Bibliography

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33:6840–6851, 2020.

[2] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes, et al. Photorealistic text-to-image diffusion models with deep language understanding.

[3] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2837–2845, 2021.

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

[5] Mingyuan Zhang, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang, and Ziwei Liu. Motiondiffuse: Text-driven human motion generation with diffusion model. arXiv preprint arXiv:2208.15001, 2022.