Sujet de stage Master 2 :

Modèle de diffusion de l'action humaine 4D

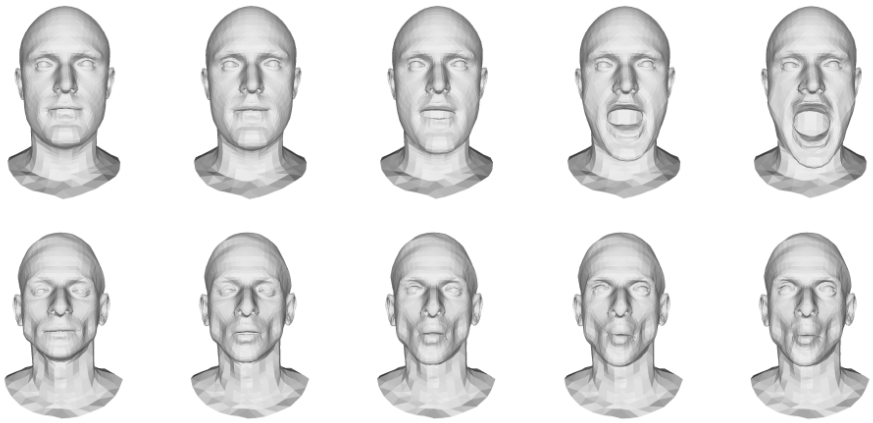

Figure 1: Séquences d'expressions faciales générées par notre modèle. Le processus inverse du modèle de diffusion a été guidé par le label d'expression : « mouth extreme » (en haut) et « cheeks in » (en bas).

Organisme d'accueil

— Laboratoire ICube: Le laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie (The Engineering science, computer science and imaging laboratory), http://icube.unistra.fr/

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

Lieu de travail et salaire

Place de l'hôpital, Strasbourg (67), France. Le stage se déroulera dans l'équipe de recherche MLMS (Machine Learning, Modélisation & Simulation) située sur le site hospitalier du laboratoire, à 10 min à pied du cœur du centre-ville de Strasbourg, classé au patrimoine mondial de l'UNESCO.

Salaire : 600€/mois environ pour une durée de 6 mois.

Encadrants

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

Date de début

Février – avril 2023.

Contexte

La vision robotique pour la cognition humaine ne fonctionne souvent pas bien dans la situation réelle, malgré les résultats perturbateurs obtenus en vision par ordinateur et en intelligence artificielle. Alors que la plupart des données d'entraînement ont été collectées dans des arrière-plans bien conditionnés et faciles à isoler, les vidéos sauvages du monde réel peuvent contenir diverses conditions environnementales telles que l'éclairage, les motifs d'arrière-plan et, plus notoirement, les occlusions. Cette dernière devient la source de problèmes

récurrents de la cognition humaine par les robots-soins en situation d'interne. De grandes variations dans la forme du corps, les mouvements, les vêtements et les interactions fréquentes avec des objets contribuent également à la difficulté. Une façon prometteuse d'améliorer les performances cognitives consiste à augmenter les données d'entraînement, ce qui coûte cher, malheureusement. Notre objectif est de développer un modèle génératif 4D, qui sera utilisé pour générer un ensemble de données synthétiques pour de telles tâches de cognition humaine basées sur la vision.

Objectifs

Les modèles génératifs basés sur le modèle probabiliste de diffusion de débruitage (DDPM) [1] ont montré des résultats remarquables, comme l'ont montré certains des travaux récents sur la synthèse d'images [2], la génération de nuages de points [3] et la génération de mouvement humain [4, 5]. Récemment, nous avons réussi à adapter DDPM à la tâche de génération d'expression de visage 3D (i.e. visage 4D). Alors que le modèle a été entrainé de manière inconditionnelle, son processus inverse peut être conditionné par divers signaux de condition, tels que des étiquettes d'expression, du texte, des séquences partielles ou simplement une géométrie faciale. Cela nous permet de développer efficacement plusieurs tâches en aval impliquant diverses générations conditionnelles, dont beaucoup se sont avérées plus performantes que les méthodes de pointe.

Dans ce stage, nous étendrons notre modèle de diffusion d'expression faciale 4D à un modèle de diffusion d'action humaine. Plusieurs tâches *downstream* seront définies, pour chacune desquelles une génération conditionnelle sera développée.

Compétence attendue

— Formation en informatique, apprentissage automatique, apprentissage profond ou traitement du signal

— Expérience de la programmation en Python et Pytorch/Tensorflow

— Bonne notion de la modélisation cinématique est un plus

— Bonnes compétences en communication

Application

Envoyez votre CV et vos relevés de notes (licence et master) à seo@unistra.fr.

Bibliographie

[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.