DLCV 2024: Assignment 3
Deadline: 30th April

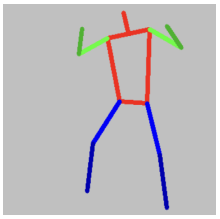
Instructions

1. Your submission should be a zip file containing the codes (.py or .ipynb), readme.txt file, and your report in pdf format.
2. Please include comments in your code. The readme.txt file should contain information on the organisation of your files, packages used along with their versions, python version, and instructions on running the code.
3. While you are encouraged to go through online repositories to learn best practices and tricks, please avoid directly copying from somewhere.
4. Please submit a report on your observations, results, plots, and analysis in pdf format. This assignment carries 25% weightage for code and 75% weightage for your analysis in the report.

[**Motion Diffusion Models**]: In this assignment, you will implement a diffusion model to learn the motion of the human dance video. The goal of this assignment is to train generative models trained on a single dance sequence and can generate new dance variations inspired by the training video sequence. The reference paper to follow is SinMDM: Single Motion Diffusion. As a simpler version, we will implement some components from this paper on 2D human skeleton. Below are the deliverables from this assignment:

1. Implement a UNet ("punet") model explained in Sec.4, that maps a sequence of 2D keypoints to another sequence of 2D keypoints.
2. Sample a set of N equal-length sequences from the original sequence of time T to create a dataset of N small sequences. (You can experiment with the amount of overlap and frequency of sampling). Try different values of N and report your observations on the model's performance.
3. Implement a Denoising Diffusion Probabilistic Models training paradigm for denoising the motion sequence. Train the DDPM model with the standard MSE diffusion loss. Report the training curves and your observations.
4. Generate multiple output sequences from the trained DDPM models and generate a dance video using the render.py file.

[Code] We provided a code snippet for rendering the skeleton video given the sequence of the 2D keypoints. Once you have trained your model, you can generate a sequence of 2D keypoints from the motion diffusion model and render a video using `'renderer.py'` script. To generate the video, you have to save the keypoints in a `.json` file in the same format as `uptown_funk.json`. Example rendering gave the input sequence of the 2D keypoints.



[Dataset details] We have provided a motion dataset in .json format. The .json file contains a set of 2D keypoint locations for the human body tracked over T timesteps. Each row in the table corresponds to the 2D keypoint locations at a particular time point. We have 13 keypoints for the skeleton: `'nose'`, `'left_shoulder'`, `'right_shoulder'`, `'left_hip'`, `'right_hip'`, `'left_elbow'`, `'right_elbow'`, `'left_hand'`, `'right_hand'`, `'left_knee'`, `'right_knee'`, `'left_leg'`, `'right_leg'`.