Leveraging large scale text-to-image models for realistic human motion synthesis

Keywords: Deep generative models, diffusion, human motion, hybrid models

Abstract: This internship focuses on the challenging task of generating realistic 3D bodies in motion. The BEDLAM dataset [1] has emerged as a recent solution, offering RGB videos with realistic 3D human body models. Notably, this research highlights that neural networks trained solely on computer-simulated data can achieve very accurate results for 3D human body pose and shape estimation from real RGB images. Yet, this dataset is mostly generated using traditional computer graphics techniques that rely on 3D engines, which can limit the diversity and realism of the generations.

Aside from that, there has been a recent development in text-to-image models that can generate highly realistic images based on textual descriptions [2,3]. This progress is mainly due to the advancement of Deep Diffusion Probabilistic Models (DDPM) that can be trained at scale. However, relying solely on text to generate images is not sufficient for accurately representing body poses and generating realistic synthetic data.

In this project, we explore the possibility of combining a 3D renderer with recent diffusion models to create hybrid models. The goal is to leverage the flexibility of traditional 3D rendering techniques and the realism of text-to-image models to improve the generation of synthetic data. In particular, we plan to base our approach on recent image editing methods such as [4]. If successful, we plan to evaluate the impact of this increased realism on downstream tasks when training neural networks on our synthetic data.

Supervisors: Stéphane Lathuilière (Telecom Paris), Pierre Perrault (Idemia)

Profile: Candidates must have completed or be in the final stages of defending their MSc degree. They should possess a strong foundation in computer vision, which encompasses machine learning, specifically deep learning, along with proficient coding skills in PyTorch.

Host institution: Multimedia team - Telecom Paris, with support of Idemia

Possibility of PhD extension after the internship.

Applying: To apply, candidates are required to send an email to stephane.lathuiliere@telecom-paris.fr and pierre.perrault@idemia.com including the following:

- A cover letter demonstrating their interest and suitability for the topic.
- Their CV.

- A transcript of their MSc grades.
- Some references or recommendation letters.

Applications will be reviewed on a rolling basis.

References:

[1] Black, M. J., Patel, P., Tesch, J., & Yang, J. (2023). BEDLAM: A Synthetic Dataset of Bodies Exhibiting Detailed Lifelike Animated Motion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[2] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems

[3] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition

[4] Brooks, T., Holynski, A., & Efros, A. A. (2023). Instructpix2pix: Learning to follow image editing instructions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition