Animaj
PICTURE THIS!

Internship:
Construction of a 3D mesh auto-encoder

Animaj presentation

Animaj (https://www.animaj.com/) is a next-generation, international media company that creates, manages and delivers engaging and premium brands to kids and families everywhere. We acquire and transform high-quality intellectual properties (IPs) into multi-platform franchises, using a digital-first approach. Examples of our IPs are Pocoyo, Kidibli and Hey Kids.

Harnessing real-time children’s content trends detection, cutting-edge audience engagement tools, and AI-driven animation production, we are propelling new technologies in kids' entertainment. Our content, produced independently, aims to inspire children to dream, explore, and create.

Subject

In the Engineering department, our team is harnessing the power of Deep Learning methodologies to enhance the efficiency of 3D animation production workflows. Our focus encompasses cutting-edge topics such as retargeting (transferring poses seamlessly from one character or motion capture to another), motion retrieval, motion style transfer, and text-based motion generation. For example, you can find a blog post describing our work on retargeting here:
https://www.animaj.com/post/animaj-motion-to-motion-transfer
Each of these areas involves the utilization of a 3D stylized character as both input and/or output for our models.

Various representations of 3D characters exist, each serving specific purposes. Some methods rely on the skeleton definition for character representation, making them suitable for retargeting [1] and motion generation [4, 5]. However, these approaches struggle to generalize to custom characters used in production due to differing skeleton architectures.

An alternative character representation involves a 3D triangular mesh, incorporating face definitions and vertex positions [2, 3]. This method offers increased flexibility in skeleton construction and provides an accurate portrayal of a character's 3D shape. Nevertheless, this flexibility comes at a cost. Achieving a detailed character surface requires a triangular mesh with a high face and vertex count, often in the order of ten thousand faces and approximately half as many vertices.

A significant drawback emerges from the substantial memory and computational resources needed for such detailed representations, limiting these models to single-frame and single-character operations. As a result, extending these models along the temporal axis for tasks like motion recognition or generation, as well as modeling interactions between characters, becomes impractical. For example, this limitation results in human motion generation methods typically being constrained to the generation of human motion within a fixed representation [4, 5].

Moreover, the unstructured nature of the representation derived from faces and vertices poses challenges for integration into Transformer-like architectures. The quadratic complexity inherent in Transformers makes the application of this representation in such models difficult. Transformer models, renowned for their effectiveness across diverse modalities, including text, image, video, and audio, are particularly well-suited for multimodal tasks. Leveraging Transformers would allow for the development of versatile multimodal models, such as text/image/video/audio to pose/motion models.

The internship's primary goal is to conduct a thorough review of current literature within the field of mesh representation. Subsequently, the aim is to design and implement a 3D mesh auto-encoder with the ability to accurately capture the shapes of objects and efficiently compress them into a latent representation. Leveraging latent space has proven highly effective in generative models, as seen in text-to-image models [6, 7]. The potential applications of this auto-encoder are diverse. Primarily, it has the potential to substantially decrease the computational burden on models by facilitating direct operation within the latent space. This not only unlocks the possibility of constructing motion models that work seamlessly on sequences of meshes but also enables the extension of models to handle multiple characters, incorporating interactions between them. Secondly, the latent space can be employed in generative models, facilitating the translation from text/image/video to pose/motion.

References

1. Skeleton-Aware Networks for Deep Motion Retargeting
2. Skeleton-free Pose Transfer for Stylized 3D Characters
3. HMC: Hierarchical Mesh Coarsening for Skeleton-free Motion Retargeting
4. GMD: Guided Motion Diffusion for Controllable Human Motion Synthesis
5. PhysDiff: Physics-Guided Human Motion Diffusion Model
6. High-Resolution Image Synthesis with Latent Diffusion Models
7. Latent Consistency Models

Application

If you are interested in this project, please apply directly to the following position:
https://www.welcometothejungle.com/fr/companies/animaj/jobs/deep-learning-scientist-intern_paris