Editing of a 3D object represented by Gaussian Splatting

In recent years, there has been strong growth in user-generated content software, especially in the visual domain. After the release of GPT-4o from OpenAI, users created more than 700 million images in just one week. Users who had no previous exposure to computer graphics were able to easily exaggerate or change a certain image to their liking. This underscores the huge demand for generative tools that allow intuitive and direct manipulation of visual content.

While generating and editing 2D images is rapidly becoming common practice, editing in higher dimensions such as video, 3D and 4D content is much more complex. These areas pose challenges in terms of consistency, geometry, temporal continuity and user control that current systems are not yet able to fully address.

Gaussian Splatting (GS) is a major advance in computer graphics, representing 3D scenes as sets of Gaussian components with color and transparency. Its training and rendering are very efficient and generate realistic images. GS has also been adapted for dynamic 4D scenes, 2D images and video.

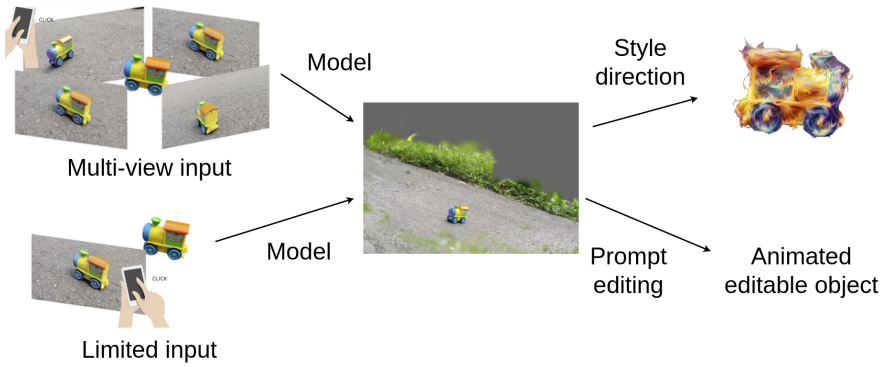

Figure 1: Example of process pipeline in a 3D scene: using the available images, an object is reconstructed with Gaussian-splatting-based models. The user can then modify the properties of the selected object.

Representing a scene/object with primitives allows for fast rendering, giving great potential in applications in 3D games or movies. However, GS models to reconstruct an object/scene need a lot of images in the input to faithfully represent the consistency of the reconstruction. It is potentially possible to generate with large generative models additional views based on a single image or text. This allows for additional data augmentation. However, inconsistencies are possible. This problem can potentially be solved by generating sequences rather than individual views, using the models to generate video, this would allow more accurate reconstruction of the scene geometry.

Editing objects represented by 3D Gaussian primitives falls into four categories: texturing, position modeling, local editing and style transfer. Editing the shape and position of a single primitive is not a difficult task, the difficulty is in editing in a consistent way a certain set of Gaussians of which the object is composed. There are methods to control the Gaussians, but unfortunately they do not support an object mesh or structure that allows precise control of the movement of 3D objects. Styling and editing such scenes is even less trivial when, through limited model input, we only have access to a limited number of images that allow for fully correct reconstruction. Potentially, object styling affecting primitive features in the case of inconsistent reconstruction could improve the visual aspect of the object.

The project would involve investigating the feasibility of reconstructing a 3D object from a single reference image or textual description. **As the final result of the research will be the creation of effective and accessible solutions to create and edit realistic 3D models with limited input data.**