

FlexGen: Flexible Multi-View Generation from Text and Image Inputs

Supplementary Material

A. More visualization results

We show more visualizations of the model in Figure 1. Our method can freely control the unseen area. In addition, we also show the comparison of the control ability of the model trained on ours and Cap3D dataset in Figure 2.

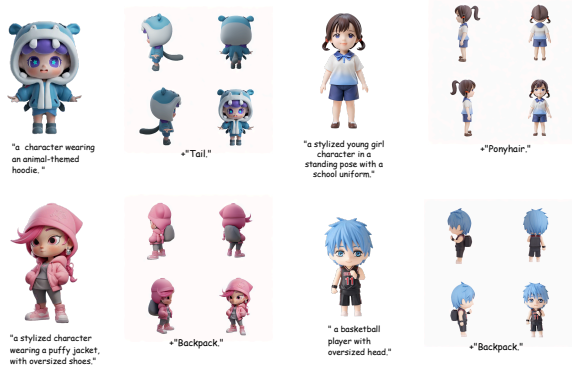


Figure 1. Illustration of more success cases

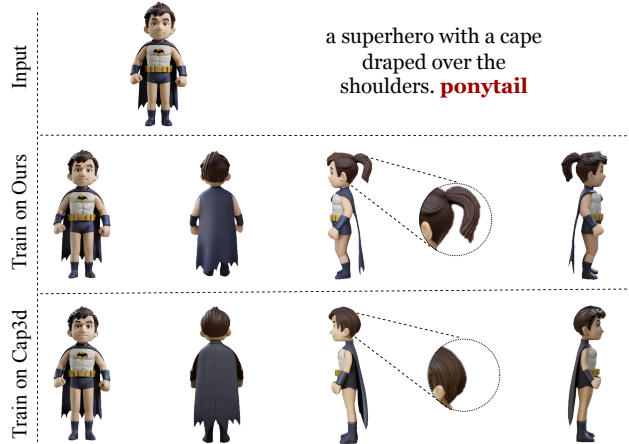


Figure 2. Controllability comparison under Our annotation and Cap3D.

B. Detail of Material Rendering

We integrate material descriptions, including attributes such as metallicity and roughness, into the text annotations to enable material-controllable generation. These descriptions are designed to correspond with the materials utilized during rendering in Blender. With Blender, we can freely adjust

the values of metallicity and roughness, allowing us to render corresponding images, as show in Figure 3. The values for metallicity and roughness range from 0 to 1. Specifically, when the values are below 0.3, the corresponding prompt is "low" and when they exceed 0.6, the corresponding prompt is "high".

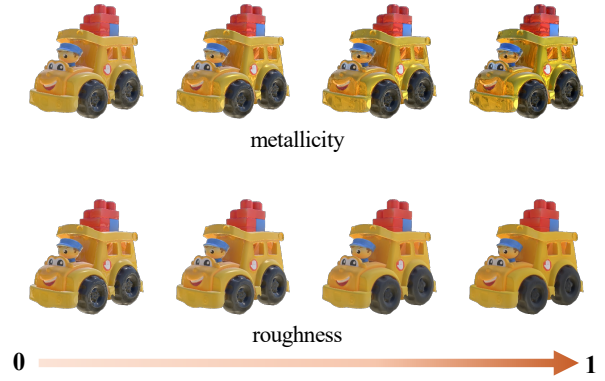


Figure 3. Visualization of an object rendered with different values of metallicity and roughness.

C. Failure Cases

Figure 4 illustrates two failure cases, where the input text is unable to serve as an editor for multiple views. This limitation arises due to the absence of relevant text descriptions in the training data.

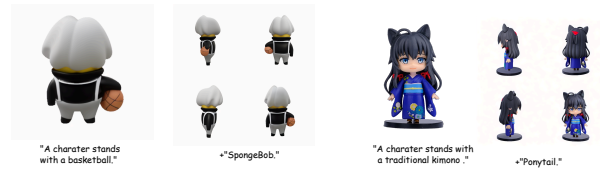


Figure 4. Illustration of some failure cases