

Diffusion-Based Conditional Mesh Synthesis for Rapid 3D Visual Design

Khandelwal, Anant (School: Thomas Jefferson High School for Science and Technology)

Motati, Sritan (School: Thomas Jefferson High School for Science and Technology)

Sood, Siddhant (School: Thomas Jefferson High School for Science and Technology)

As digital design has become ubiquitous over the last few decades, 3D creations have become much more advanced through Computer-Aided Design (CAD). However, this approach can be tedious, time-consuming, and inaccessible, especially when creating the kinds of smooth and curved designs necessary in art. To accelerate 3D artwork, this research presents CloudGen, a revolutionary text-to-3D printing pipeline to create 3D models from natural language prompts, allowing artists to shape objects using text. To this end, a Latent Diffusion Model (LDM) was implemented to generate print-ready 3D structures based on text inputs. First, an encoder-decoder mechanism transforms point clouds into functions which parametrize 3D shapes. Then, using pre-trained CLIP embeddings to convert text prompts into numerical representations, a vision transformer creates such 3D shapes, repeatedly iterating upon the object embedding towards the shape designated by the prompt. Once the initial synthesis is complete, the shape generated is smoothed using the bilateral filter algorithm, thereby removing any inconsistencies and instabilities. Further, to enable users to modify their generations, a novel editing framework was developed to retrace diffusion steps and incorporate feedback prompts. The end-to-end generation architecture has been implemented in a user-friendly web interface paired with a virtual reality-powered design viewer. Both quantitative and qualitative experiments validate the efficacy of the proposed text-to-3D printing pipeline. With this research, CloudGen establishes the foundation for human-in-the-loop creation of 3D art pieces.