AI Model for Text/image-to-3D Image Rendering

Juve Barajas, Ken Chika, Gwangwoo Lee

05/08/2024

Table of Contents

Abstract

- Our project showcases an innovative method to generate 3D object files from text or image inputs using a pre-trained HuggingFace model called Shap_E. Leveraging Shap_E's capabilities, the project produces 3D representations that are then processed for display in a scene. This technology provides a more accessible method for creating 3D object files.

Introduction

- Our project focuses on automating the conversion of images or text, using an advanced AI model.
- We used a GPU server for computational resources because it was too much to run on our local machines.
- Our software is capable of rendering 3D object files at an accuracy of the model’s capabilities.

Project Design

- Text/Image-to-3D conversion
    - We utilize shap_e's inference to generate 3D models given an image/text prompt.
- Command line arguments
    - Command line arguments are used to help the P3 project understand where the generated file was located on disk.
- P3
    - We modified our P3 project to be able to handle the new object files and allowed it to be called by a python script with command line arguments.

Implementation

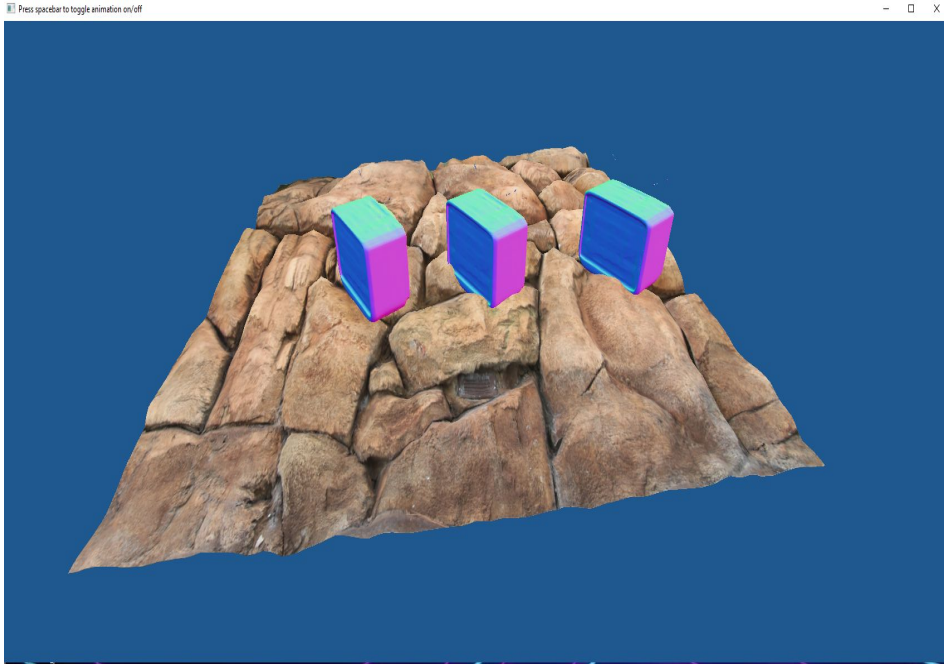
- Text/Image-to-3D conversion
  - Model loading uses text300M/image300M for interpreting text/image inputs.
  - Diffusion process is used to guide the model to generate 3D latent representations.
  - Latent sampling is used to generate the 3D models.
- Command line arguments
  - ‘text’: Text prompt for generating 3D objects.
  - ‘image’: Image file path for generating 3D objects.
  - ‘output_object_path’: Path to object file output directory.
  - ‘output_object_filename’: Filename of the output object file.
- Compiled P3
  - It was necessary to compile P3 project before being able to invoke it in the python script.
  - Computed vector normals to color by normal.

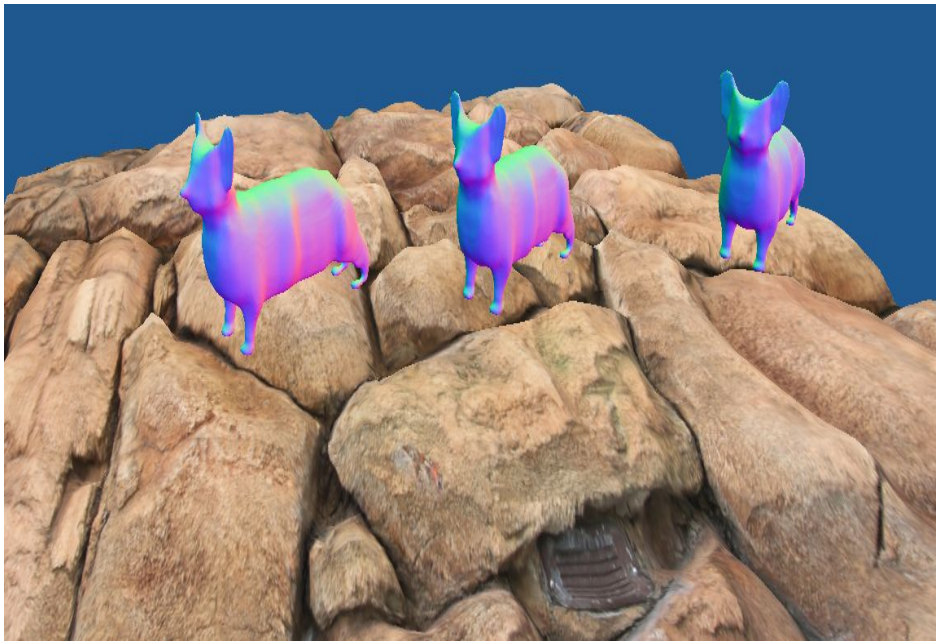
Demo

- https://drive.google.com/file/d/1gaVxQErgIY7PbFQsG9wXI8SIwQK4nBQg/view?usp=sharing

Results

- 1950’s tv (Text prompt)
- Corgi (Image prompt)





Code and resource organization

- Github
  - We organized all the python code within a github repository, but worked on the P3 modifications together without uploading it due to github's limited capacity for resource management.

Challenges and improvements

- Vertex normals
    - A big challenge was due to the inability of shap_e to generate vertex normals in the object files it generated. To overcome this, we implemented a new method in the Shape class to be able to calculate vertex normals on the fly if there were not present in the object file.
- Server-side GUI
    - A huge setback was trying to display the P3 program on a GPU server. The problem is that the GUI portion of our application was running on the server side rather than client side. One simple method of relaying the GUI information to the client side is X11 forwarding. The only problem is that we do not have admin permissions on the server, so we could not enable it.

Learning outcomes

- Technical knowledge
  - Gained advanced understanding and experience with AI 3D modeling, particularly in handling complex model outputs and integrating them into existing rendering pipelines.
- Problem solving
  - Developed new strategies for overcoming technical challenges, such as calculating vertex normals and implementing client-side solutions for server-based GUI problem.
- Collaboration and resource management
  - Learned skills in managing computational resources and collaborating on code development under constraints, reflecting a realistic software development environment.

Socially responsible

- Generating accurate and detailed 3D models from image or text data is limited by specialized knowledge, making the problem inaccessible. This project aims to make 3D content more accessible using AI.

Conclusion

- From conceptualization to realization, our project has advanced our technical skills and deepened our understanding of AI and its transformative potential in the computer graphics industry, uncovering new avenues for innovation and application.

Thank you!