3D Assets Generator Project Proposal

Group Members:

Team Member1: Yizhou Zhang (Student ID: 21253387, Email: yzhangww@connect.ust.hk)

Team Member2: Yifu Wang

Team Member3: Zhenyuan Hu (Student ID: 21270579, Email: zhubl@connect.ust.hk)

Team Member4: Sen Yang (Student ID: 21255696, Email: syangdt@connect.ust.hk)

Project Background:

The rapid evolution of computer vision and generative AI has shifted focus from 2D image synthesis to 3D asset creation, which is pivotal for emerging applications such as augmented reality, gaming, and the metaverse. Traditionally, producing high-quality 3D models demands extensive manual effort, specialized skills, and time-consuming processes like sculpting or scanning. Recent advancements in generative models, including diffusion-based techniques and neural radiance fields, are revolutionizing this domain by enabling the creation of detailed 3D objects from simple inputs like text prompts or 2D images. These methods democratize 3D content generation, allowing for scalable production of assets that align closely with user descriptions or visual references.

This project addresses key challenges in 3D generation, such as ensuring geometric accuracy, visual realism, and semantic fidelity. By leveraging state-of-the-art models, we aim to build a pipeline that transforms textual or image inputs into usable 3D model files with textures. This aligns with the course's emphasis on innovative computer vision applications, particularly in handling multi-modal inputs and evaluating 3D outputs quantitatively and qualitatively. Our motivation stems from the potential impact on industries like virtual reality, where efficient 3D asset generation can accelerate content creation workflows. We select this project to explore cutting-edge techniques like Diffusion Models and 3D Gaussians, building on foundational papers in NeRFs and recent open-source advancements.

Preliminary Methodology

We propose a generative AI pipeline to create 3D assets using Diffusion Models, enhanced with 3D Gaussian Splatting for efficient rendering. The system will process text prompts or 2D images to produce textured 3D models in formats such as .obj or .glb, meeting the project's requirements.

The methodology encompasses input preprocessing where detailed prompts describing materials, styles, and proportions like a wooden table with a matte finish are crafted, and for image inputs, edge detection and background simplification are performed using OpenCV to isolate key features. Model implementation relies on adopting Hunyuan3D 2.0 as the base model for its advanced text-to-3D synthesis capabilities, integrating 3D Gaussian techniques for optimization, which includes encoding inputs with CLIP, generating geometry via diffusion sampling, and applying textures through neural rendering with iterative refinement. Post-processing involves refining generated assets in Blender to apply textures and optimize meshes, ensuring compatibility with standard 3D formats. Evaluation assesses performance using Hausdorff Distance for geometric quality, PSNR for texture fidelity, and qualitative reviews for semantic consistency, with testing covering diverse inputs such as everyday objects and complex scenes to validate the system's robustness.

The pipeline will be built using PyTorch, with potential integration of neural radiance fields for enhanced view synthesis if needed. Computational efficiency will be ensured through GPU acceleration and resolution adjustments during development to manage resource demands effectively.

Project Work Timeline

Weeks 1-2 (October 2025): Conduct literature review, coordinate team efforts, and set up dataset preparation. At the end of this phase, we will finalize model selection and acquire synthetic benchmarks such as ShapeNet.

Weeks 3-5 (Late October - Early November 2025): Develop and train the base generation pipeline. At the end of this phase, we will produce initial 3D assets and conduct preliminary qualitative assessments.

Weeks 6-8 (Mid-November 2025): Integrate post-processing steps and establish evaluation metrics. At the end of this phase, we will complete the end-to-end pipeline with quantitative analysis on test cases.

Weeks 9-10 (Late November - Early December 2025): Refine the system to handle complex geometries, address edge cases, and prepare the technical report and presentation slides. At the end of this phase, we will ensure submission readiness with a polished prototype.

This timeline promotes balanced contributions from all members, supported by weekly progress meetings. We aim to have a functional prototype by Week 6, allowing time for iterations to achieve high geometric quality and semantic consistency.