**Comparative Analysis – 3D NeRF Asset Generation Models**

**Abstract:**

Generative models have good capabilities on text and image generations. The next frontier for Generative AI is to generate 3D images/models which can be used for Games development and Computer vision applications.

There are many new models which are showing promising results in the 3D generation from few/single images. These models develop NeRF assets by leveraging transformer and diffusion architecture.

**Introduction:**

NeRF 3D asset model is a technique that uses neural radiance fields (NeRF) to represent and render 3D scenes from 2D images. [NeRF is a neural network that maps 3D coordinates and viewing directions to densities and colours](https://arxiv.org/abs/2209.14988). By ray marching through the scene, NeRF can synthesize novel views of the scene from any camera pose.

**Open-Source 3D NeRF Asset Generation Models:**

There is several open-source software that implement NeRF or its variants for different tasks, such as text-to-3D, image-to-3D, or mesh generation. Here are some examples:

1. [Shap-E](https://github.com/openai/shap-e) is a conditional generative model for 3D assets that can produce implicit functions from text prompts or images. Shap-E uses a diffusion process to generate a 3D image from a 2D text-to-image diffusion model, and then optimizes a randomly-initialized NeRF to match the 2D renderings.
2. [DreamFusion](https://github.com/ashawkey/stable-dreamfusion) is a text-to-3D model that uses a pretrained 2D text-to-image diffusion model to perform text-to-3D synthesis. DreamFusion introduces a loss based on probability density distillation that enables the use of a 2D diffusion model as a prior for optimization of a parametric image generator. Using this loss in a DeepDream-like procedure, DreamFusion optimizes a NeRF via gradient descent such that its 2D renderings from random angles achieve a low loss.
3. [Instant-NGP](https://github.com/ashawkey/stable-dreamfusion) is a system that implements four neural graphics primitives, being NeRF, signed distance functions (SDFs), neural images, and neural volumes. Instant-NGP uses a multiresolution hash encoding that permits the use of a smaller network without sacrificing quality, thus significantly reducing the number of floating point and memory access operations. Instant-NGP achieves a speedup of several orders of magnitude, enabling training of high-quality neural graphics primitives in a matter of seconds, and rendering in tens of milliseconds at a resolution of 1920x1080.

**Comparation Criteria:**

A possible comparison of these software could be based on the following criteria:

* **Input modality:** text or image
* **Output representation:** implicit function or NeRF
* **Training data:** required or not
* **Training time:** seconds or minutes
* **Rendering quality:** resolution and fidelity
* **Rendering speed:** milliseconds or seconds
* **Additional features:** lighting, relighting, compositing, etc.

**3d Asset Generation:**

The following are the results obtained after generating the 3d asset models from the above-mentioned open-source generative models:

[DreamFusion](https://github.com/ashawkey/stable-dreamfusion):

We regret to inform you that DreamFusion, the text-to-3D synthesis tool, is currently under maintenance. DreamFusion is an open-source project that relies on shadcn, a collection of React components, as a dependency. However, shadcn is also undergoing maintenance and has caused some compatibility issues with DreamFusion. As a result, we are unable to run and render the NeRF 3D model, which is a neural network that represents 3D scenes as a radiance field, using DreamFusion at this time.

[Shap-E](https://github.com/openai/shap-e):

**Prompt**: “A Lion”

**Rendered Model:**

**Batch size = 4**

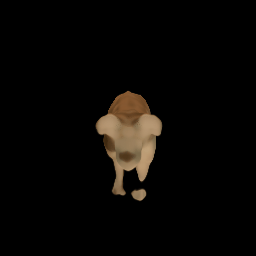
**Size = 64**

**Rendering speed:** 110sec



**Batch size = 1 , size = 256 Batch size = 1, Size = 32**

**Rendering speed:** 2min 28 sec. **Rendering speed:** 32sec.

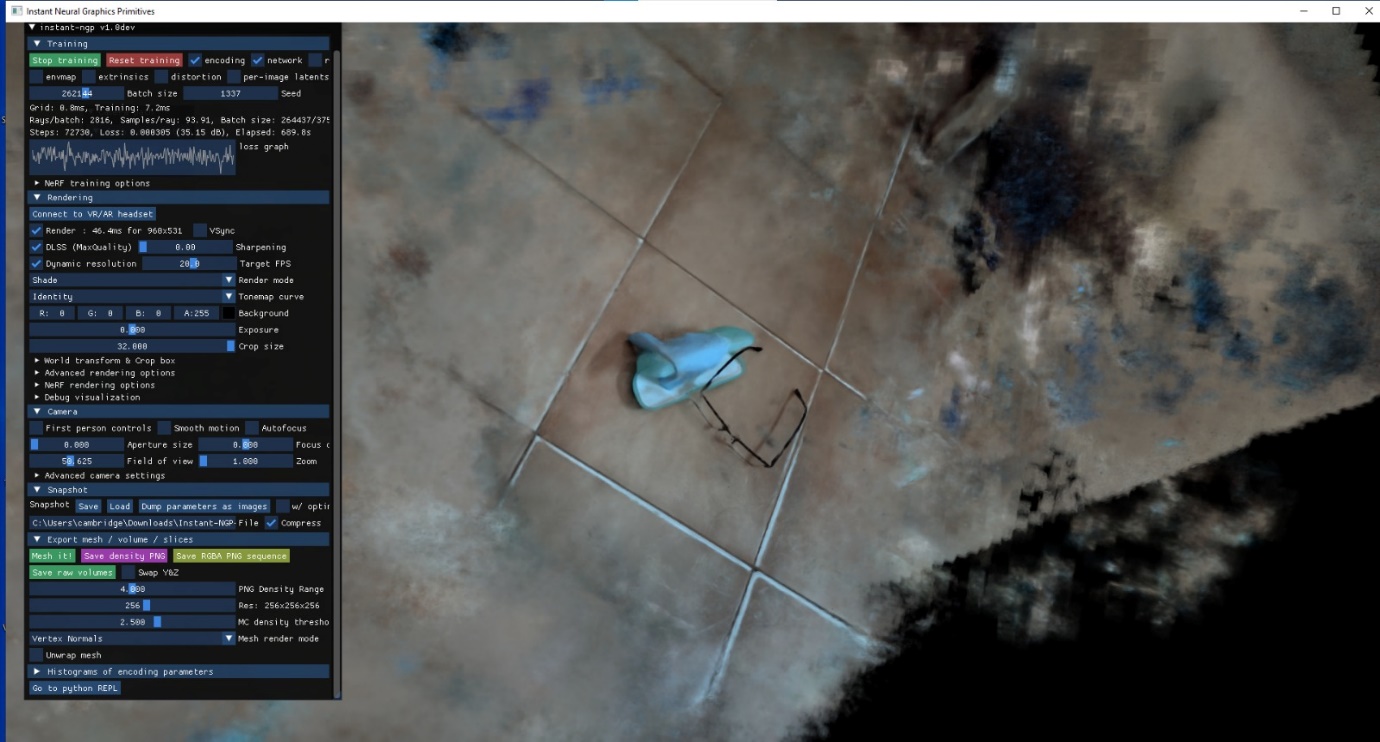


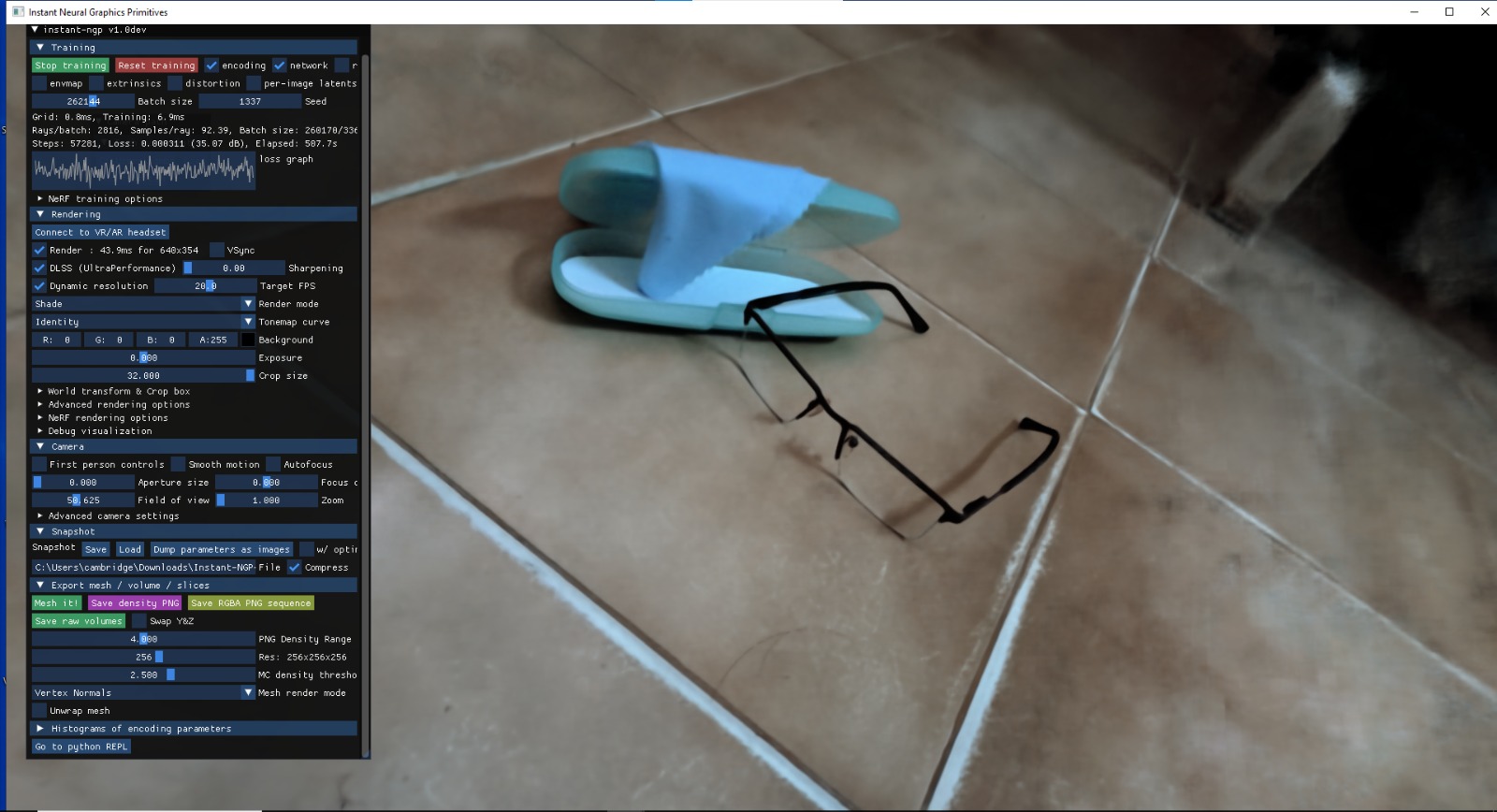
[Instant-NGP](https://github.com/ashawkey/stable-dreamfusion):

**Aim**: Image to 3d model rendering

**Rendering speed:** Around 2 minutes

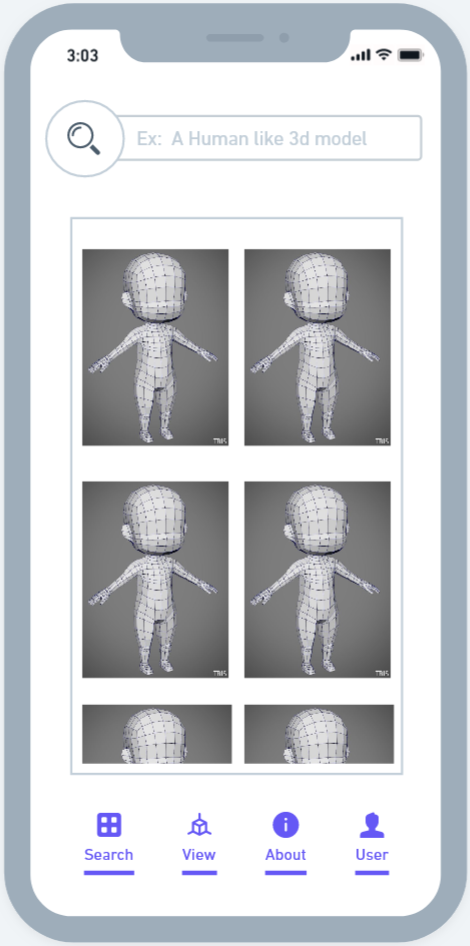
GPU used: NVIDIA RTX A6000

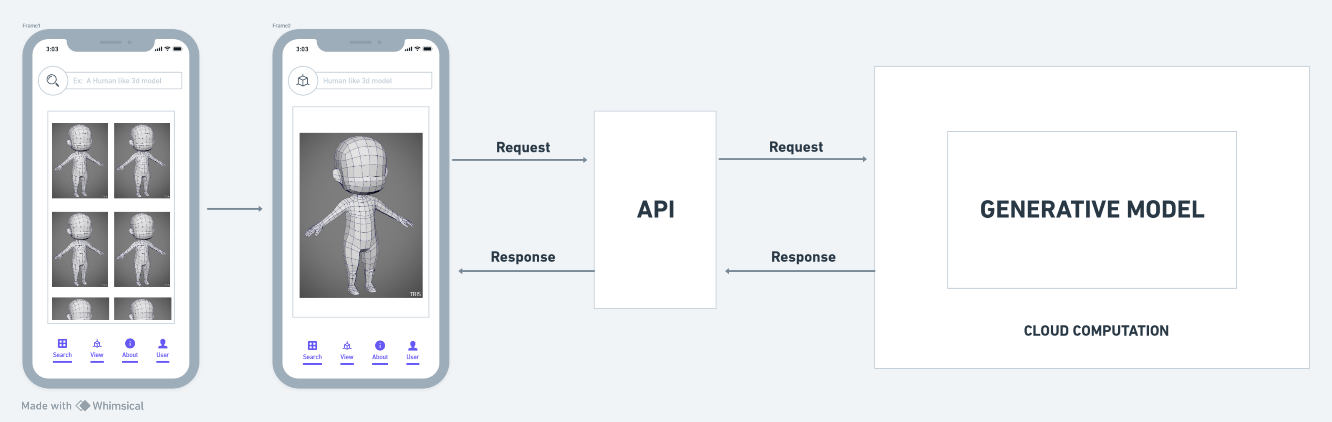
**Rendered Model:**



**Application:**

Link: [Blueprint](https://whimsical.com/nerf-mob-app-8A4u9cM72Haqtn6myB8kCZ)

****



**Conclusion:**

NeRF generative models shows promising results in the 3D model generation from few/single images or with prompt instead. Based on the analysis of the generated renders, it is clear that there is a scope for development in the field.

Neural Radiance Fields (NeRF) are a novel technique for synthesizing photorealistic 3D scenes from 2D images. However, the performance of NeRF models depends on the type of input they receive. When the input is textual, such as a natural language description of a scene, the NeRF models struggle to produce high-quality and consistent outputs. This is because text is an ambiguous and sparse representation of visual information, and it requires a lot of semantic understanding and inference to generate realistic images. Therefore, the text-based NeRF models are still in the experimental phase and need more time and research to improve their capabilities and robustness.

On the other hand, when the input is image-based, such as a collection of photos of a scene taken from different viewpoints, the NeRF models can achieve impressive results. This is because images are a rich and direct representation of visual information, and they provide enough cues and constraints for the NeRF models to reconstruct the 3D geometry and appearance of the scene. The image-based NeRF models can generate realistic and appealing outputs that can be viewed from any angle and under any lighting condition. These outputs can have various applications in industrial domains, such as virtual reality, augmented reality, 3D printing, and digital art. However, the image-based NeRF models also have some limitations and challenges, such as requiring a large number of input images, taking a long time to render, and being sensitive to noise and occlusion. The image-based NeRF models generate realistic and appealing outputs that can be applied in industrial settings, given the appropriate conditions.

In controlled environment (with proper inputs of images parameters and increased cycles of renders etc), one can expect highly rendered 3d models which can be used in the industry which helps in efficient management of time.