Dream Fusion: Text-to-3D using 2D Diffusion

Review by Andrew Milko, 01.02.2023

Ben Poole

Google Research

Ajay Jain

UC Berkeley

Jon Barron

Google Research

Ben Mildenhall

Google Research

Quick Summary

- using a pretrained 2D text-to-image diffusion model (Google's **Imagen**) to perform 3D synthesis
- no need in 3D training data
- no need to make modifications in **Imagen**

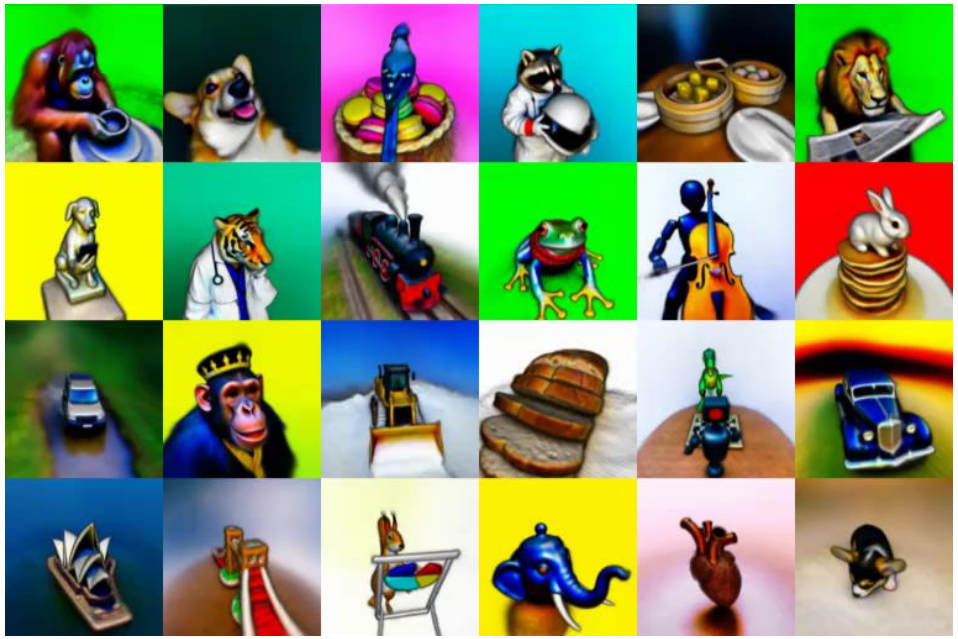
Before & after projects

time

	text-to-image	text-to-3D	text-to-video
January 2021	DALL-E *(Open AI)*		
April 2022	DALL-E 2 *(Open AI)*		
May 2022	Imagen *(Google)*		
August 2022	Stable Diffusion *(Stability AI)*		
September 2022		Dream Fusion *(Google)* / Get3D *(Nvidia)*	Make-A-Video *(Meta AI)*
November 2022	Stable Diffusion Version 2 *(Stability AI)*		
December 2022		3D Avatar Diffusion *(Microsoft)*	

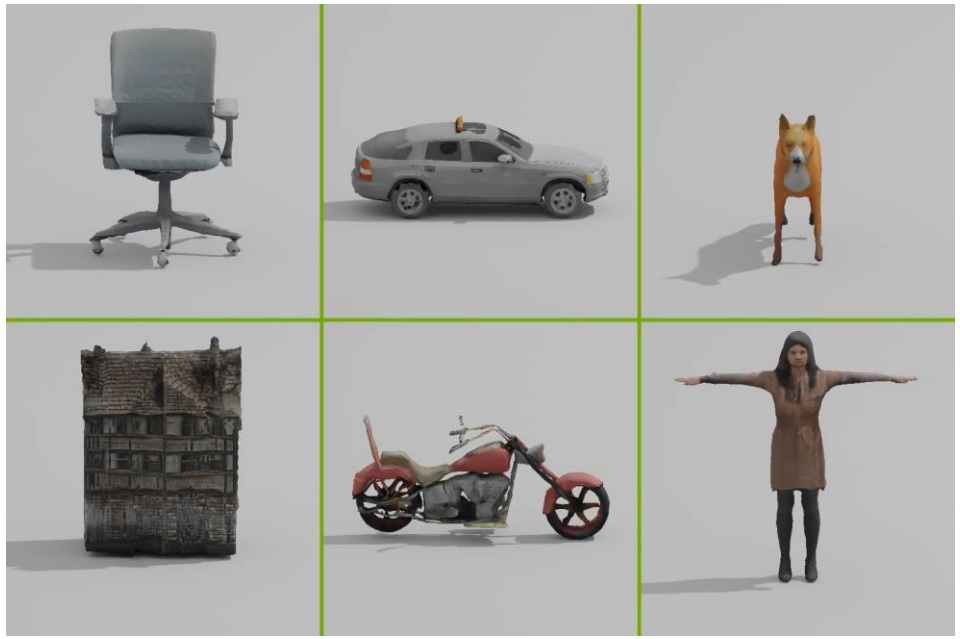
Competitors

Dream Fusion (Google)



- 3D training data: **none**
- quality: **bad**
- areas of application: **few** (for now) / **many** (in future)
- construction composition of objects: **good**
- open source: **none**
- free-to-use: **none**

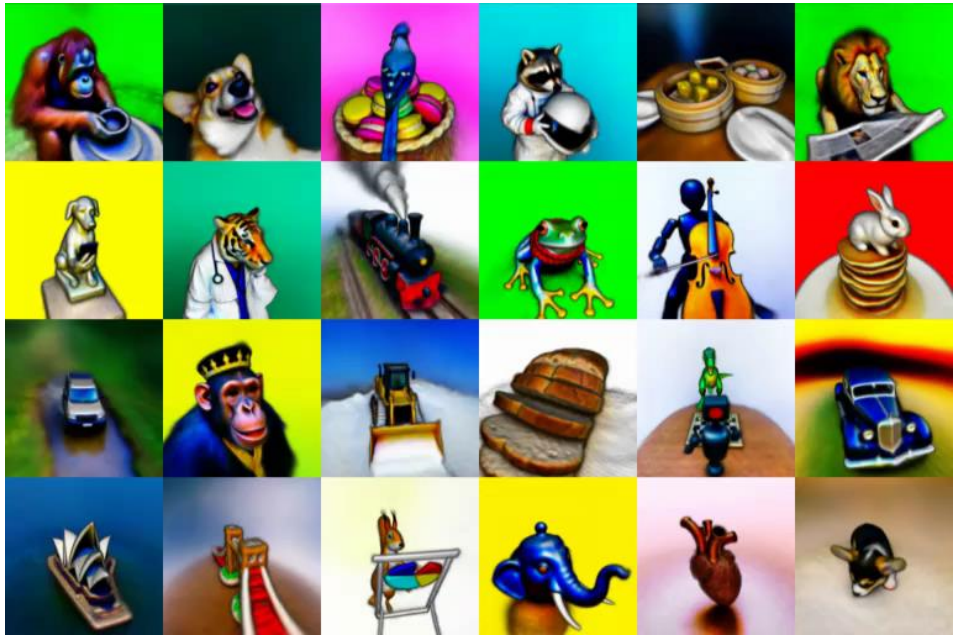
GET3D (Nvidia)



- 3D training data: **ShapeNet dataset**
- quality: **good**
- areas of application: **many**
- construction composition of objects: **none**
- open source: **yes**
- free-to-use: **yes**

Competitors

Dream Fusion (Google)



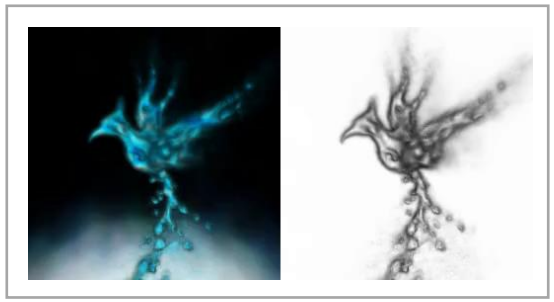
- 3D training data: **none**
- quality: **bad**
- areas of application: **few** (for now) / **many** (in future)
- construction composition of objects: **good**
- open source: **none**
- free-to-use: **none**

3D Avatar Diffusion (Microsoft)

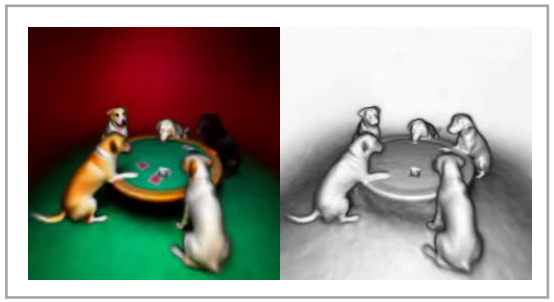
- 3D training data: **Face Synthetics dataset**
- quality: **good**
- areas of application: **few** (only human avatars)
- construction composition of objects: **good**
- open source: **none**
- free-to-use: **none**

Strength of Dream Fusion

1. generated 3D objects correspond to the text input well
2. capable of creating not just a singular 3D objects but entire 3D scenes
3. realistic interactions and intersections of objects



difficult materials



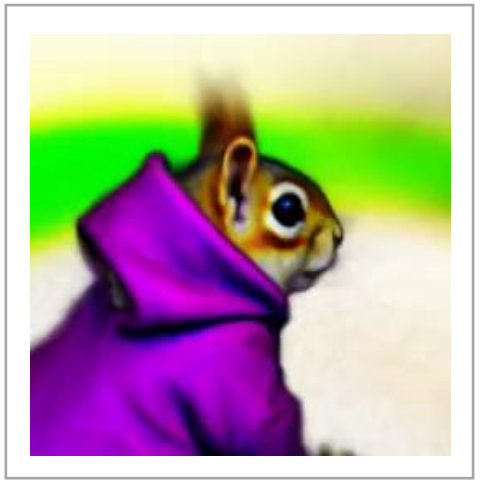
objects composition / intersection

volumetric effects based on environment

Weaknesses of Dream Fusion

1. **quality issues**: blurry, dark, unnatural high-contrast colors

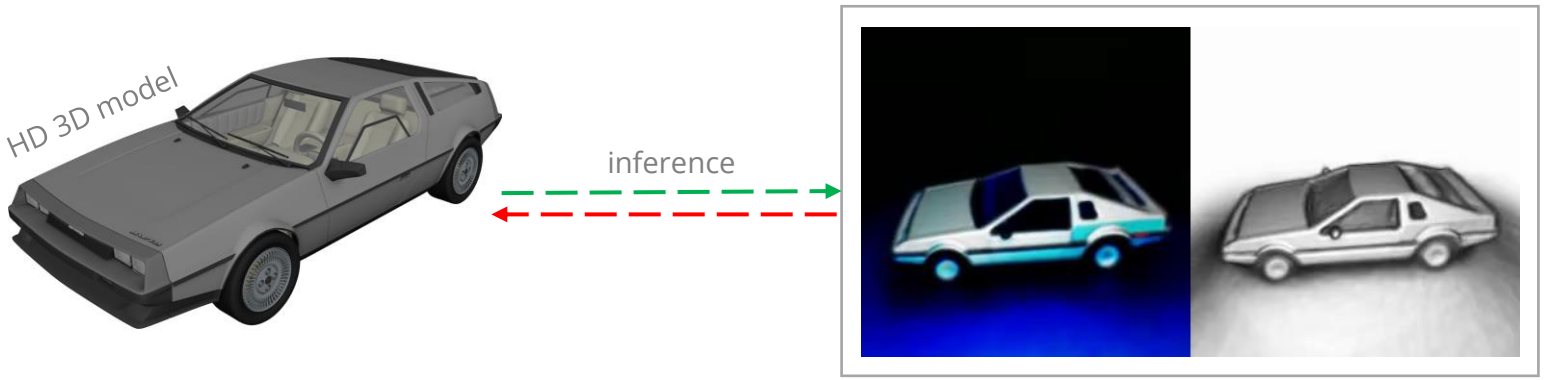
2. **multi-facing problem** (a.k.a. Janus problem)



- a DSLR photo of a squirrel + wearing a purple hoodie
- a DSLR photo of a squirrel + wearing a purple hoodie + eating a hamburger

Additional experiments ideas

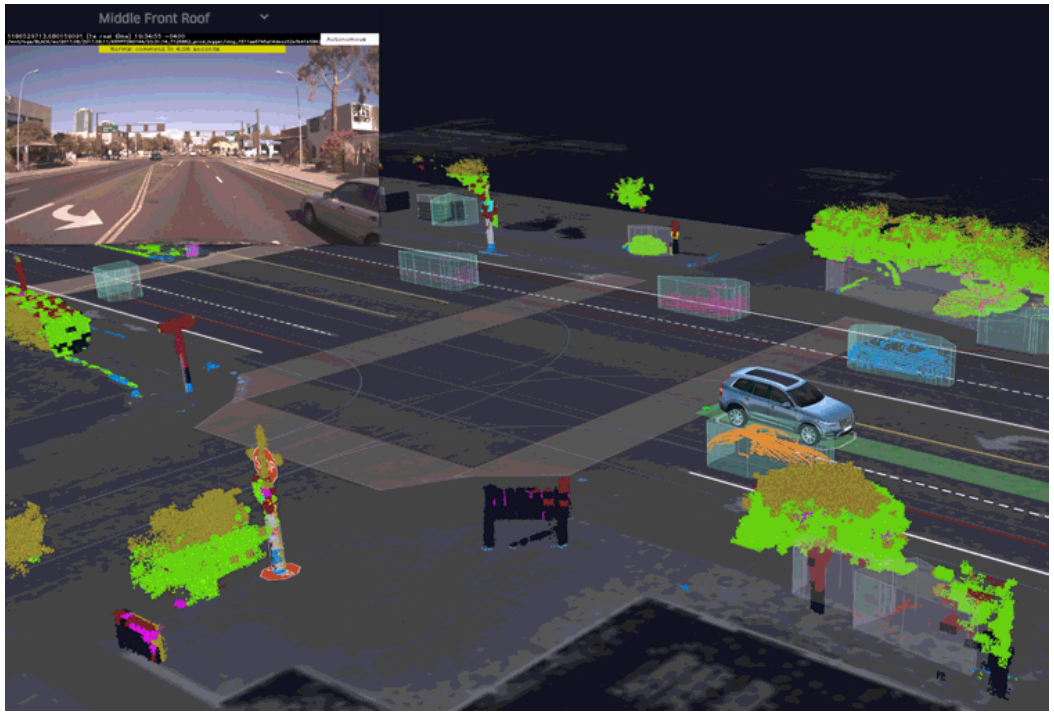
1. need a more lightweight version of **Dream Fusion** that can ran quickly on your local machine with low RAM
2. generate multiple outputs by a single input: how can we generate multiple 3D scenes with the same text?
3. adding into the training pipeline ground truth objects in the form of manually created high quality 3D models



Application Idea (right now)

Generated by **Dream Fusion** 3D models can be used in
AI training applications / simulators

- **fact:** for training self-driving cars developers use 3D simulators, which should be filled with different types of 3D assets: *trees, roads, pedestrians, other cars, motorcycles, bikes, houses and etc.*
- **task:** object recognition by *images* + *lidar scans*
- **problem:** simulators have limited and identical packs of 3D assets
- **consequence:** overfitting of AI system (bad accuracy)
- **solution:** generate many different 3D assets to increase diversity
- **result**: removed overfitting



Sources

1. [2209.14988] DreamFusion: Text-to-3D using 2D Diffusion (arxiv.org)
2. DreamFusion: Text-to-3D using 2D Diffusion (dreamfusion3d.github.io)
3. DreamFusion: Text-to-3D using 2D Diffusion! - YouTube
4. Google's DreamFusion AI: Text to 3D - YouTube
5. DreamFusion: Text-to-3D using 2D Diffusion | OpenReview