Text2Tex: Text-driven Texture Synthesis via Diffusion Models

Team4
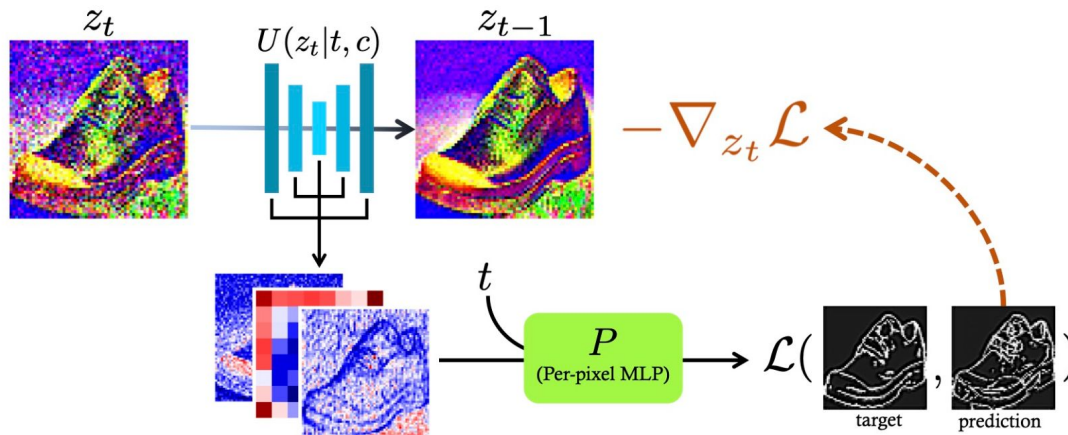Chanryeol Lee, Chanhyuk Lee

Review of Previous Presentation

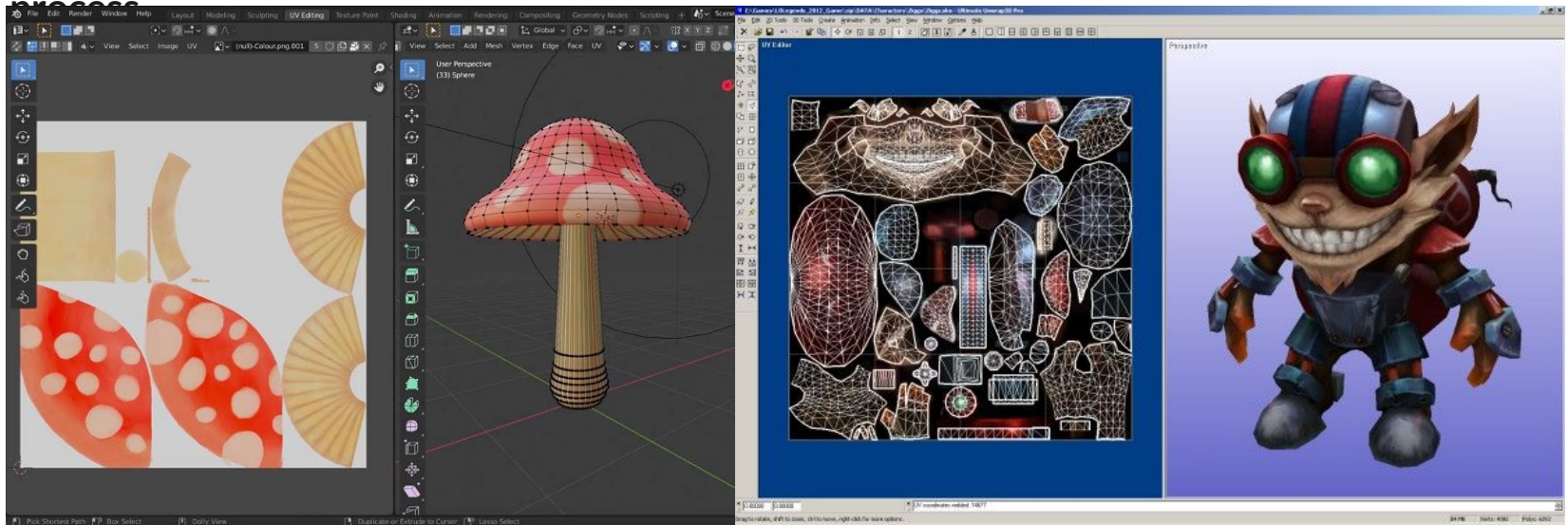**Sketch-Guided Text-to-Image DIffusion Models**

- Guide image generation with sketch inputs
- Use latent edge predictor to predict edge of sketch
- Edges are used for conditional generation, to preserve the scene geometry

Motivation

Texture maps are human generated to match the actual 3d geometry, in **conventional rendering process**
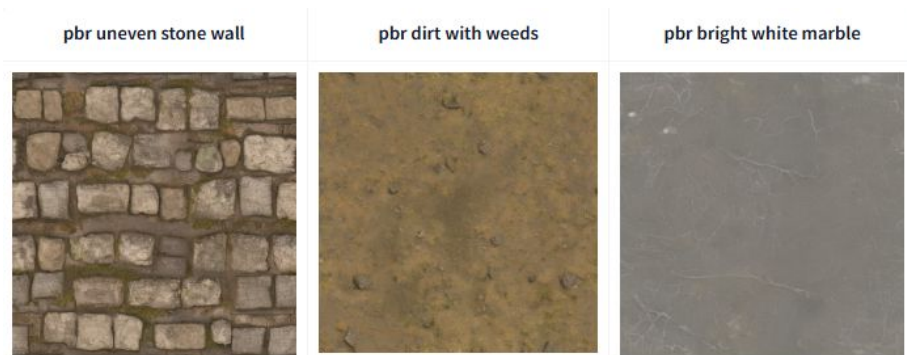

Big obstacle in fully-automated 3d scene generation, harder to automate than geometry generation

Motivation

Text2Img Generative Models (Diffusion, GAN, ..) could well express **textural descriptions in 2D**
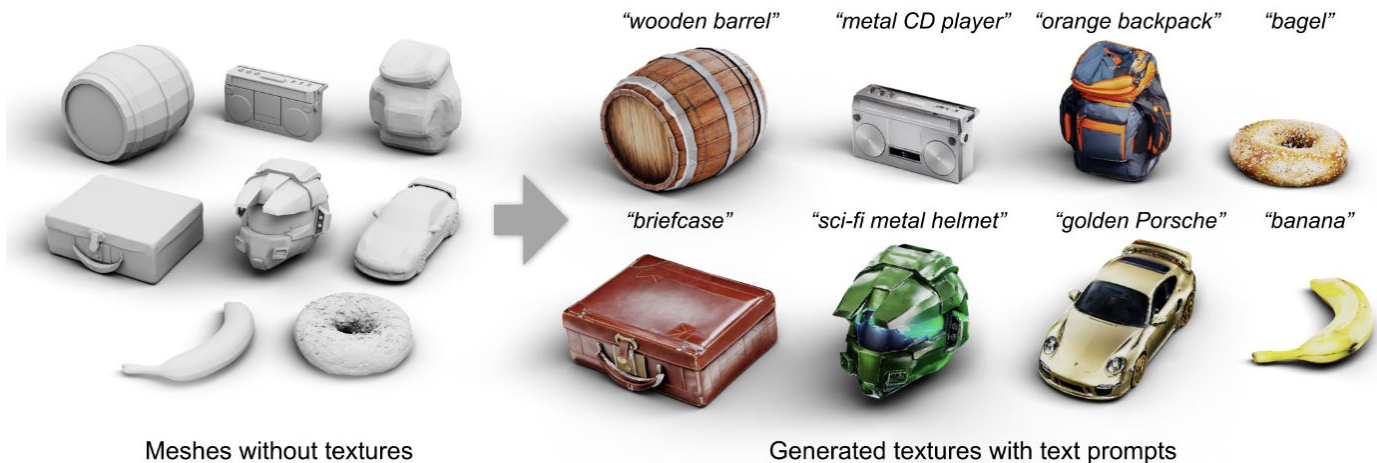

PSGAN(CVPR 2020)

pbr uneven stone wall   pbr dirt with weeds   pbr bright white marble

Stable Diffusion + Prompt

However, it's hard to generate textures with **aware of actual 3d geometry of the object!**

Chen, Dave Zhenyu, et al. "PSGAN: Pose and Expression Robust Spatial-Aware GAN for Customizable Makeup Transfer" CVPR 2020

Text2Tex

Generate high-quality textures **based on given 3D meshes** from the given text prompts!

"wooden barrel"   "metal CD player"   "orange backpack"   "bagel"

"briefcase"   "sci-fi metal helmet"   "golden Porsche"   "banana"

Meshes without textures

Generated textures with text prompts

"Text2tex: Text-driven texture synthesis via diffusion models." CVPR 2023

Related Works: Text-to-3D

NeRF + Diffusion Based Methods:

- Optimization Based-> Long Convergence Time

Poole et al. "Dreamfusion: Text-to-3D using 2D Diffusion." ICLR 2023"
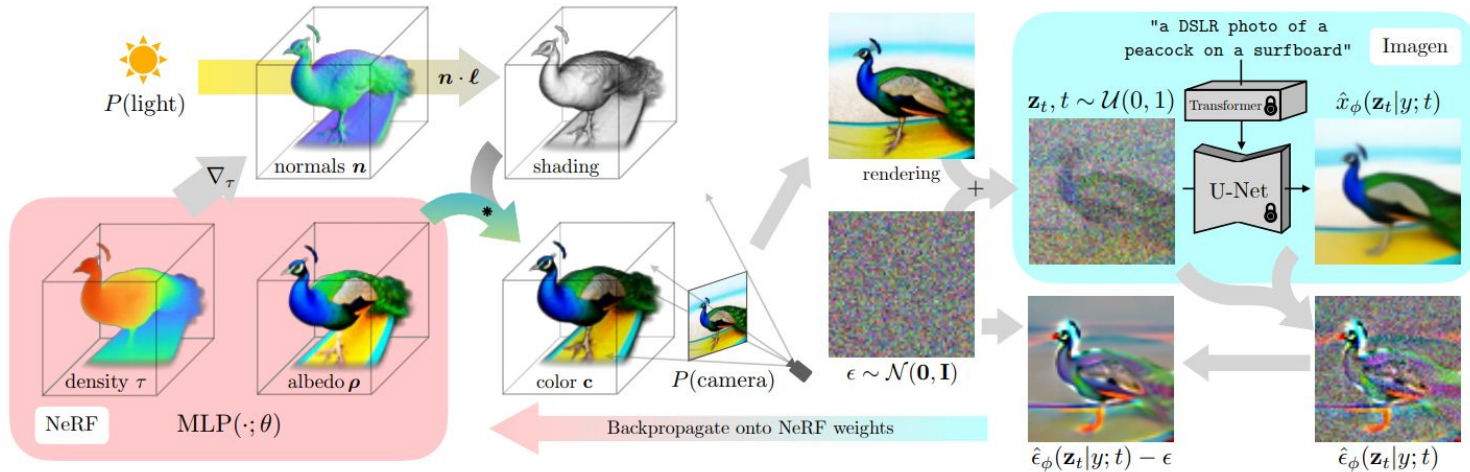
Method

Goal: Non-Optimization Approach & Minimize Human Effort in Text-to 3D

1) Use **pretrained** text-to-image diffusion model, without **additional training**
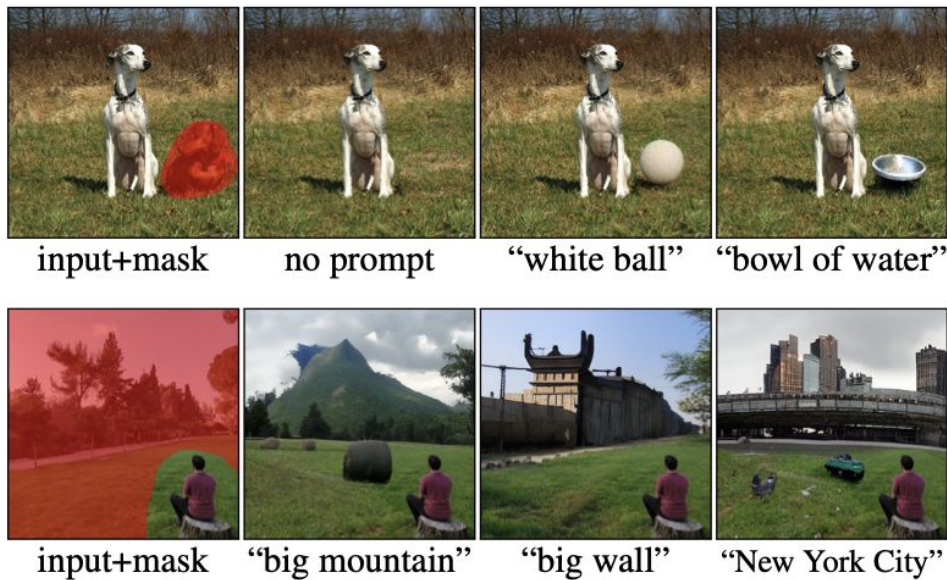
2) Dynamic View Partitioning: **Generate-then-Refine**

Pretrained Model: Depth2Image

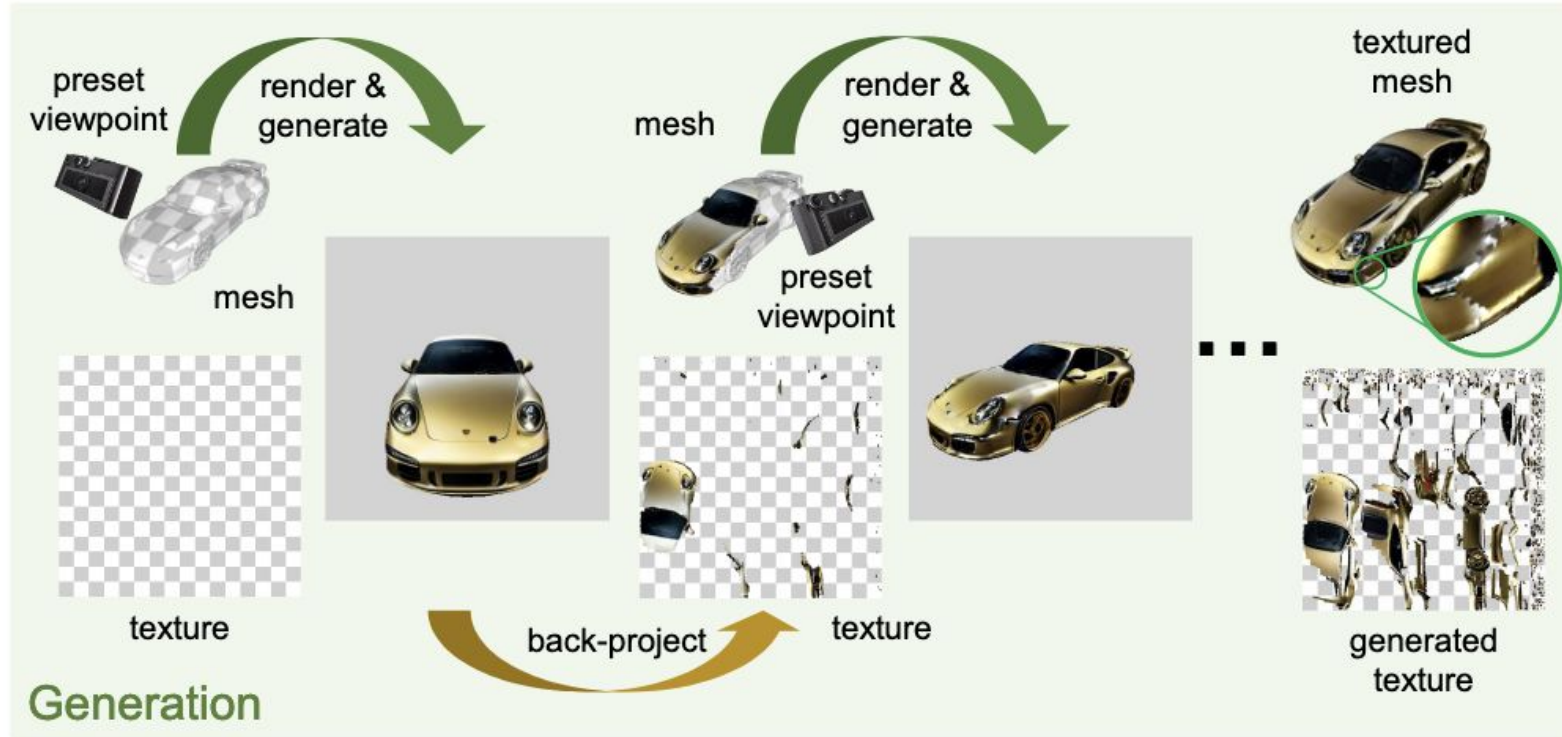From Depth information of pixels, generate appropriate image aware of 3d geometry

Depth-to-Image, https://huggingface.co/docs/diffusers/api/pipelines/stable_diffusion/depth2img

Depth-Aware Image Inpainting

Depth2Image generate entire image => use inpainting mask to guide sampling process

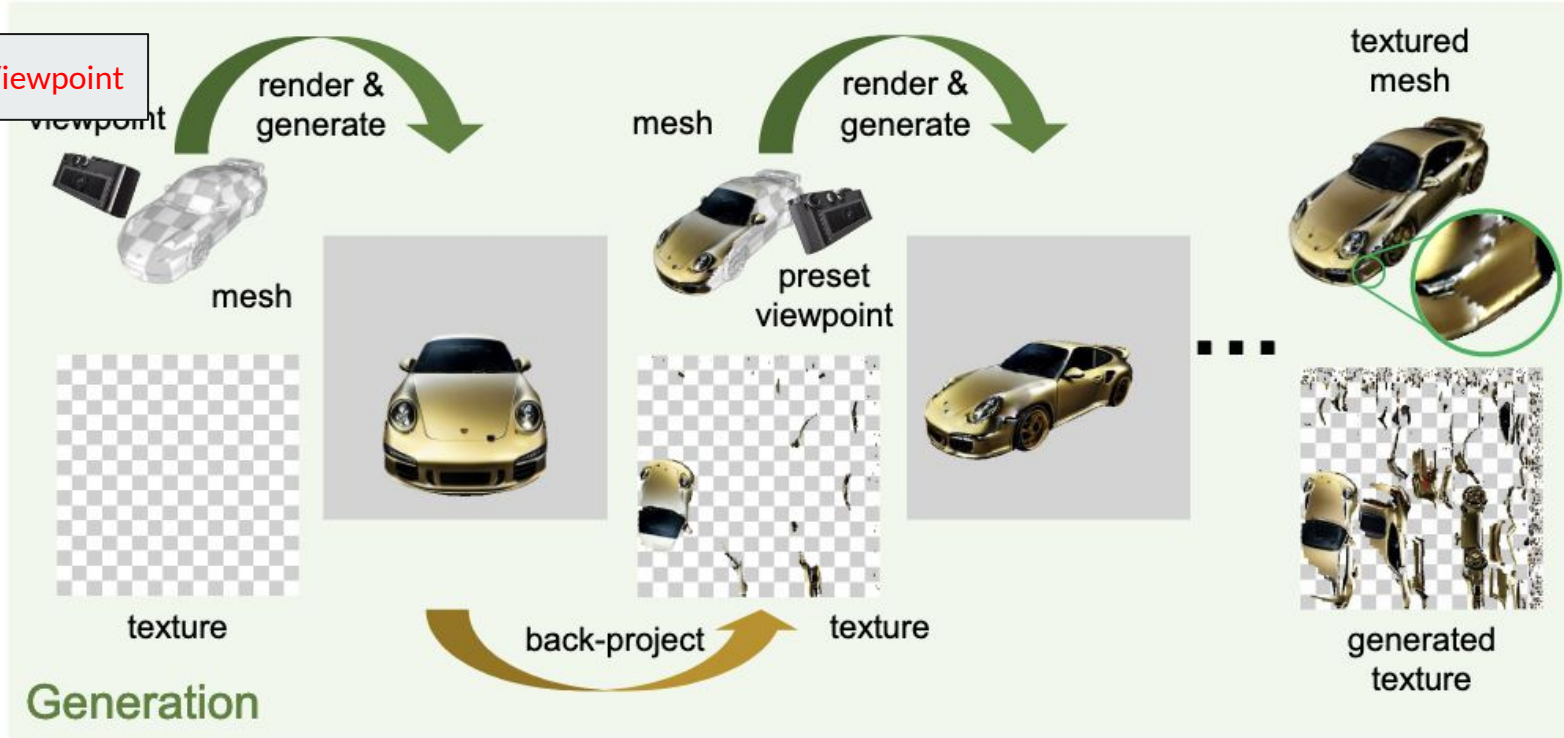input+mask   no prompt   "white ball"   "bowl of water"
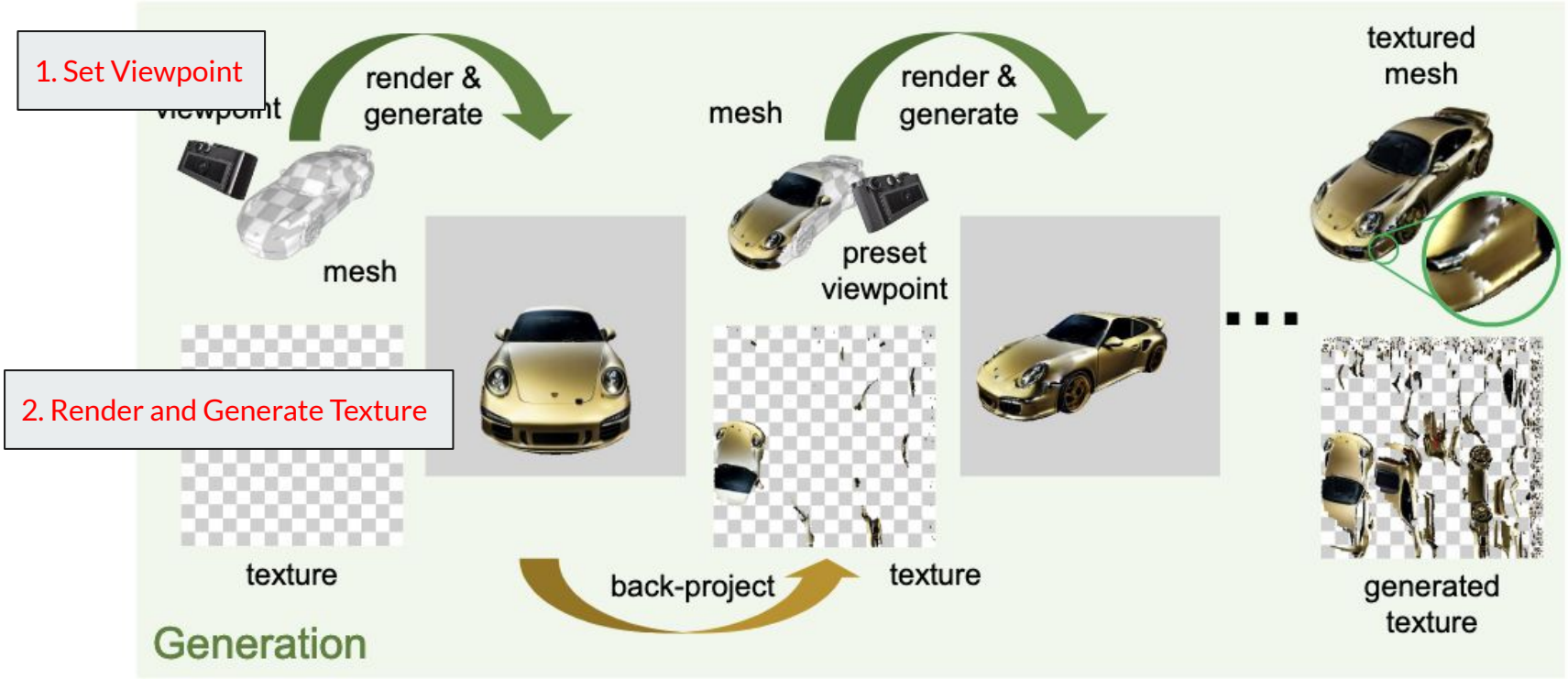
input+mask   "big mountain"   "big wall"   "New York City"

Avrahami, Omri, Dani Lischinski, and Ohad Fried. "Blended diffusion for text-driven editing of natural images." CVPR 2022

Progressive Texture Generation

Progressive Texture Generation

1. Set Viewpoint

Progressive Texture Generation

1. Set Viewpoint

2. Render and Generate Texture

Progressive Texture Generation

1. Set Viewpoint

2. Render and Generate Texture

3. Back-Project into Texture Map
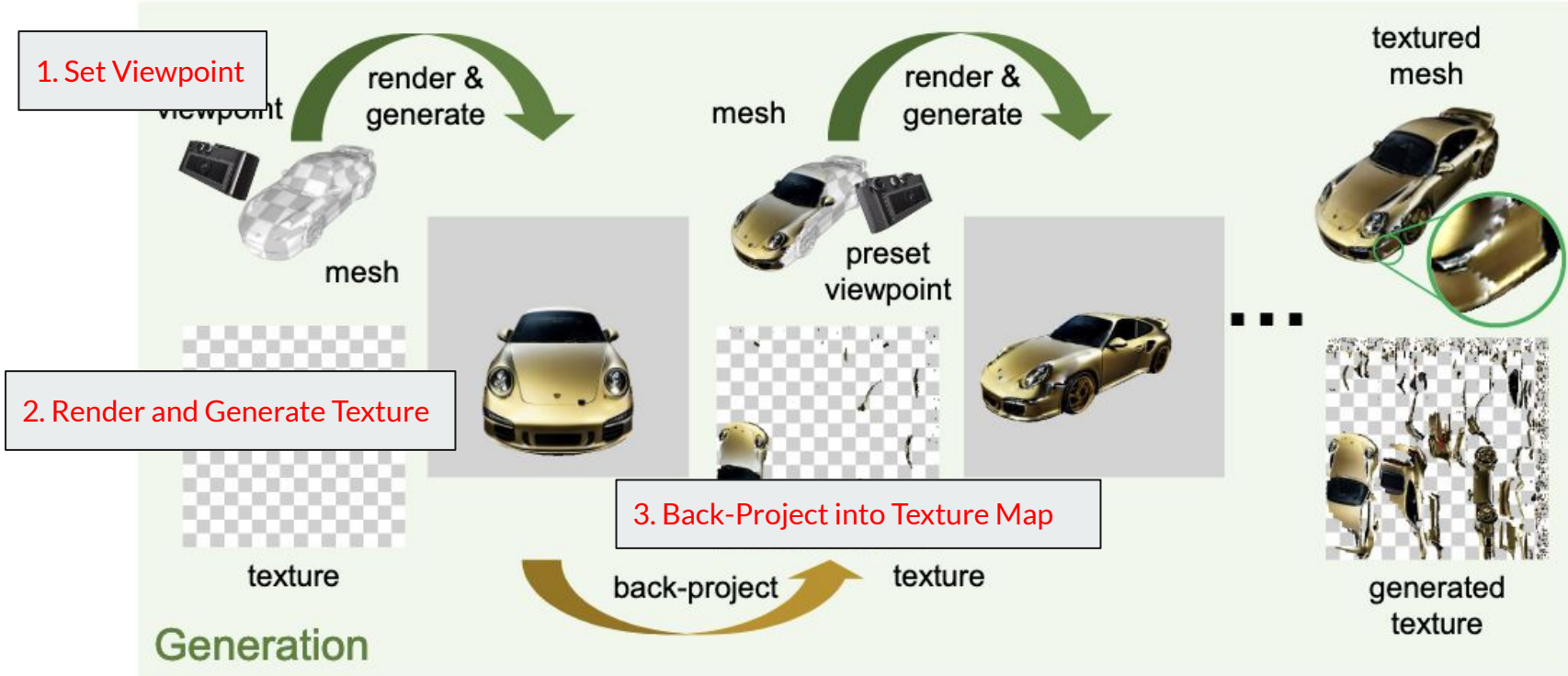
Limitations

Inconsistency between textures generated from different viewpoints, especially on **curved surface**
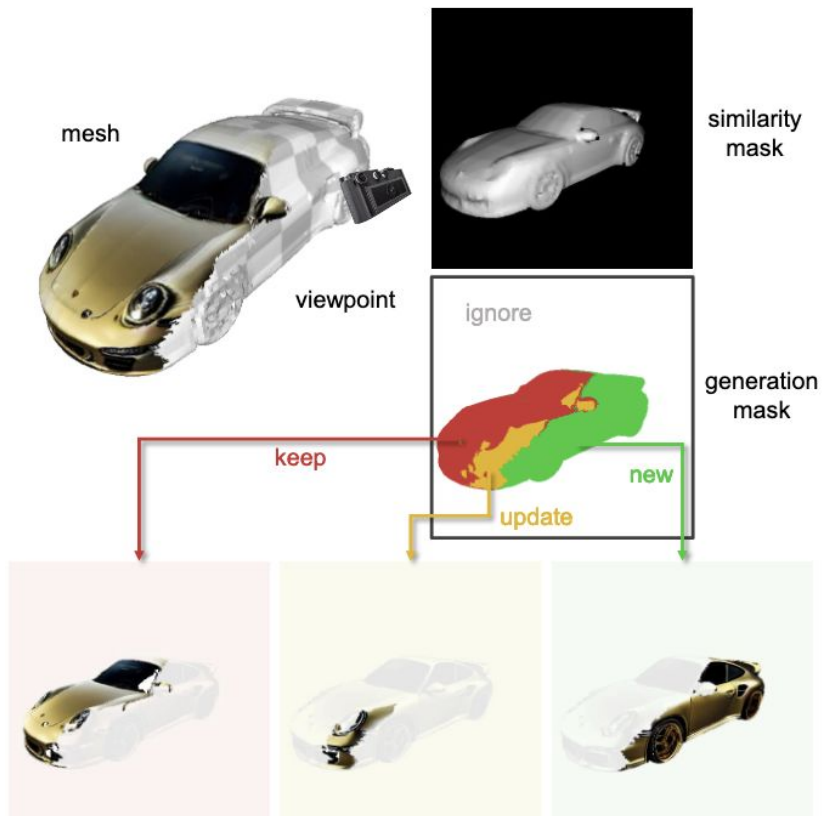
"orange backpack"

Dynamic View Partitioning

Generate Similarity Mask based on **Surface Normal**

Based on similarity mask,
Partition generation mask into 4 types

1) **New**: not yet textured
2) **Update**: high similarity score (perpendicular to view angle)
3) **Keep**: low similarity score
4) **Ignore**: background

Texture Refinement with Automatic Viewpoint Selection

Refine the generated textures from **automatically chosen** additional viewpoints
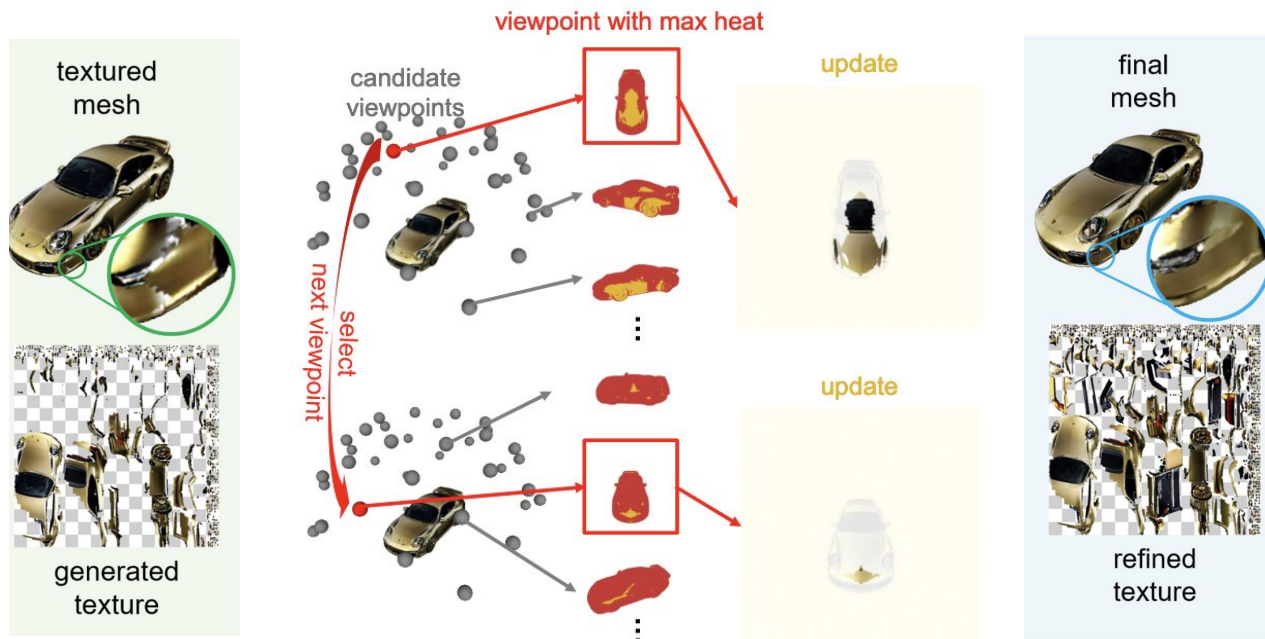
Results

Column headers: Geometry, CLIPMesh, Text2Mesh, Latent-Paint, Ours, Objaverse GT

Row labels: "a compass", "an ambulance", "a backpack", "a dumpster"

Results

Geometry   Texturify   Ours

https://daveredrum.github.io/Text2Tex/

Quantitative Results

Method	FID ↓		KID (x10^{-3}) ↓	
Text2Mesh [34]	45.38	(+9.7)	10.40	(+2.7)
CLIPMesh [37]	43.25	(+7.6)	12.52	(+4.8)
Latent-Paint [33]	43.87	(+8.1)	11.43	(+3.7)
Text2Tex (Ours)	**35.68**		**7.74**	

Method	FID ↓		KID (x10^{-3}) ↓	
Texture Fields [42]	177.15	(+130.2)	17.14	(+12.8)
SPSG [16]	110.65	(+63.7)	9.59	(+5.2)
LTG [65]	70.76	(+23.8)	5.72	(+1.4)
Texturify [56]	59.55	(+12.6)	4.97	(+0.6)
Text2Tex (Ours)	**46.91**		**4.35**	

Quiz