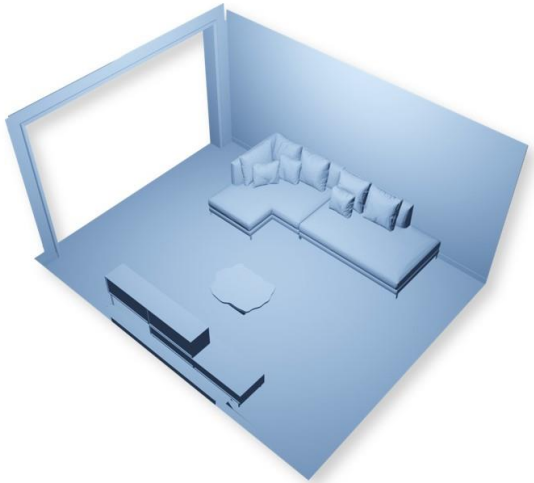
High-quality Texture Synthesis for 3D Scenes

Zhenyu Chen – 陈振宇
Visual Computing & AI Group
Technical University of Munich

How do we design appearance for a scene?

Scene Geometry



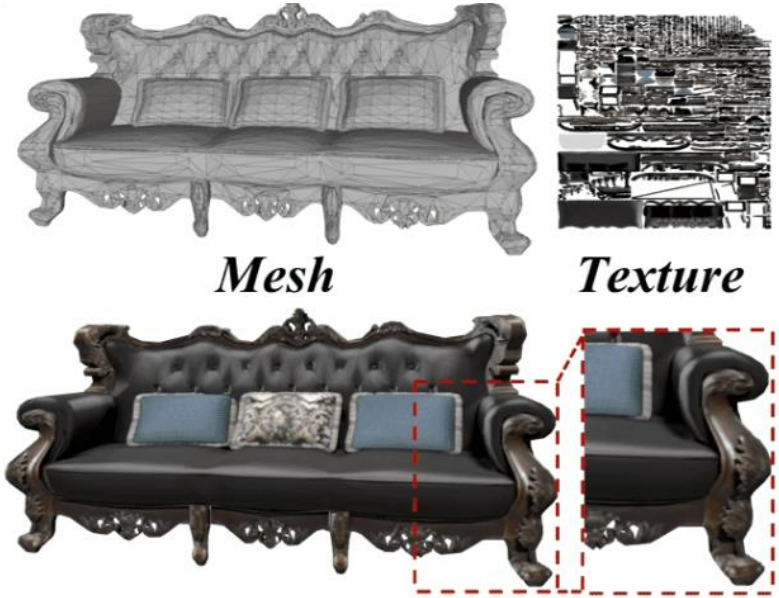
?

Prompt: "a Scandinavian style living room"

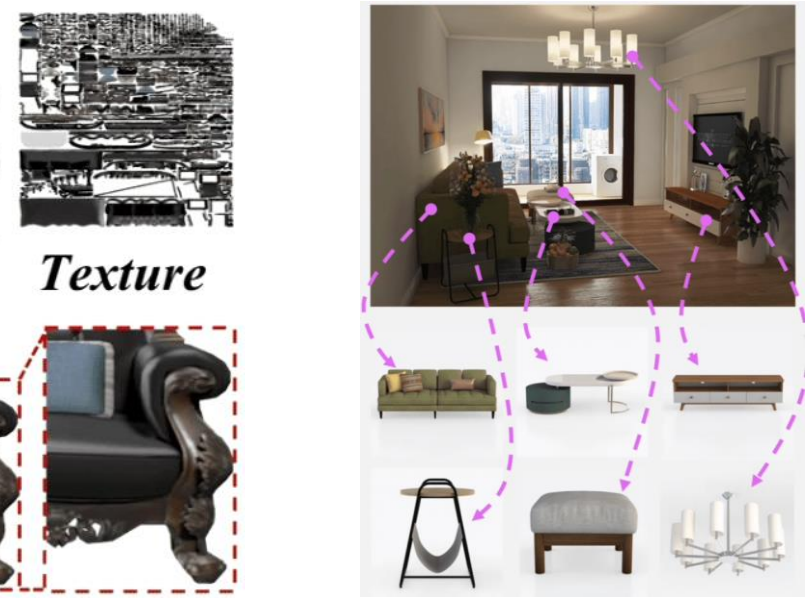
How do we design appearance for a scene?

- Conventional designing pipeline:

1. Design Object/Room Texture

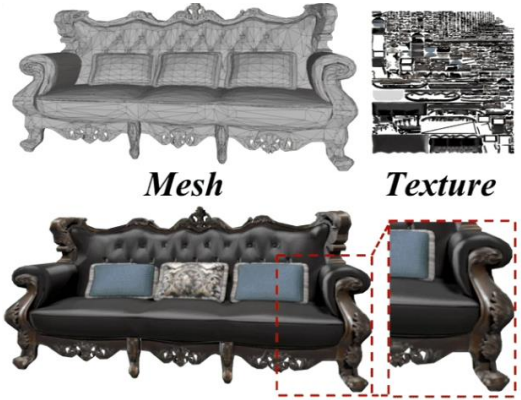


2. Design Scene Layout

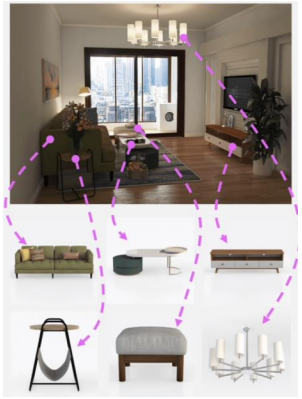


How do we design appearance for a scene?

1. Design Object/Room Texture



2. Design Scene Layout



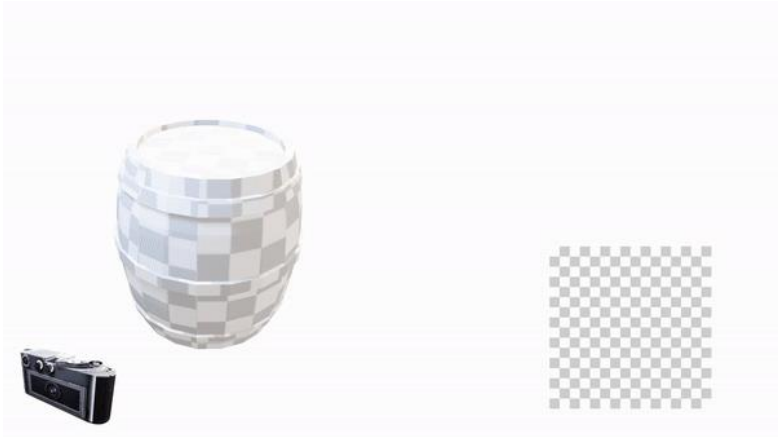
- Very expensive and time-consuming
- Style inconsistency across objects
- Hard to modify the whole scene, e.g. style, condition

⇨ *Texture synthesis – but for the whole scene!!!*

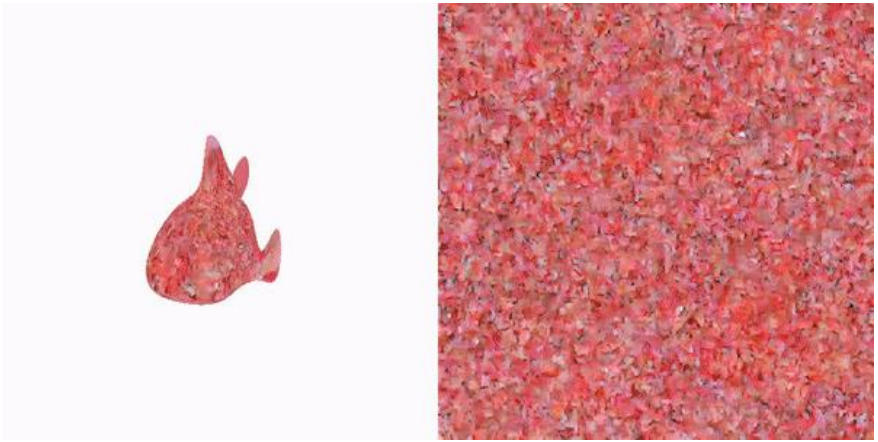
Texture synthesis works well on objects

- Two types of object-based synthesis methods:

Inpainting-based
(Text2Tex/TEXTure/TexFusion)



Optimization-based
(Latent-Paint)



- ✚ Optimization-free
- ▬ Inconsistency and artifacts

- ✚ Globally consistency
- ▬ Low quality due to latent optimizations

On scenes? Not so well…

- Text2Tex on a whole scene:

- More obvious seams and artifacts
- Style inconsistency
- Loop closure issue
- Hard to apply to the whole scene as a single mesh

On scenes? Not so well…

- Latent-Paint on a whole scene:

- Blurriness and patch-like artifacts
- Geometry and view inconsistency
- Slow convergence
- Performance drop due to latent optimization

How can we obtain high-quality textures for 3D scene meshes?

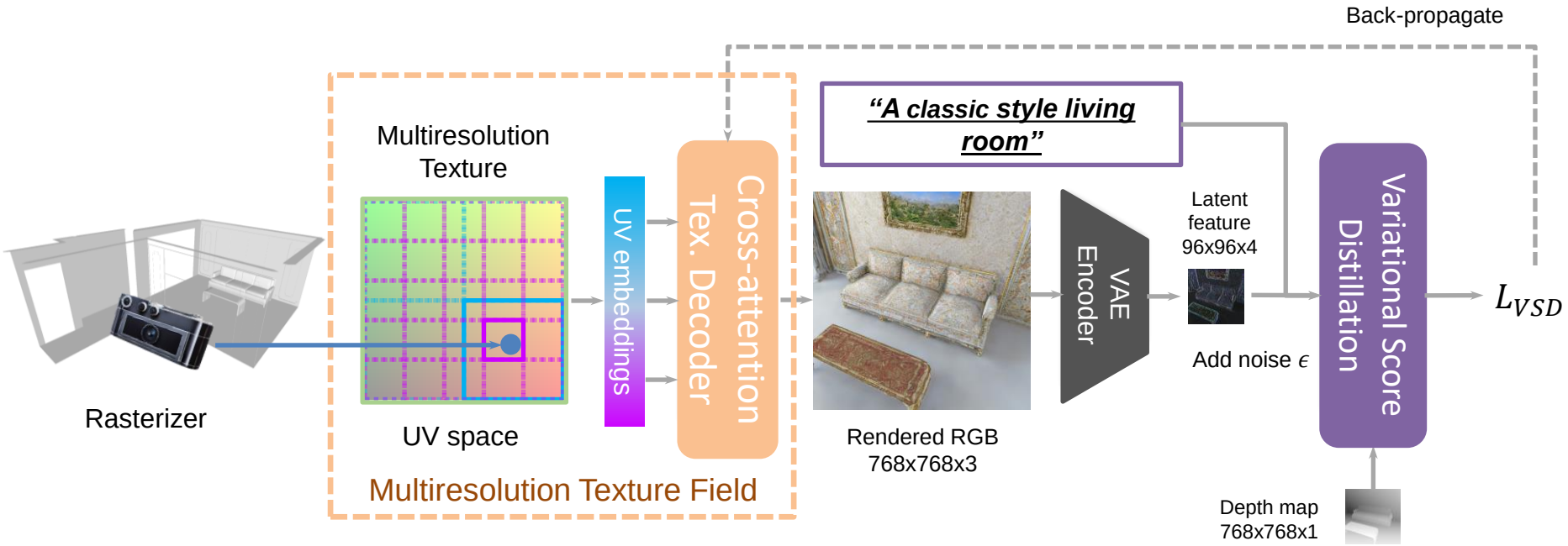
- What we aim to achieve:
  - High-quality appearance in visible views
  - Global view- and geometry consistency
  - No seams and stretching artifacts
  - Flexible for objects in different sizes

- To synthesize 3D scene textures, we need:
  - A global optimization method
  - A flexible representation

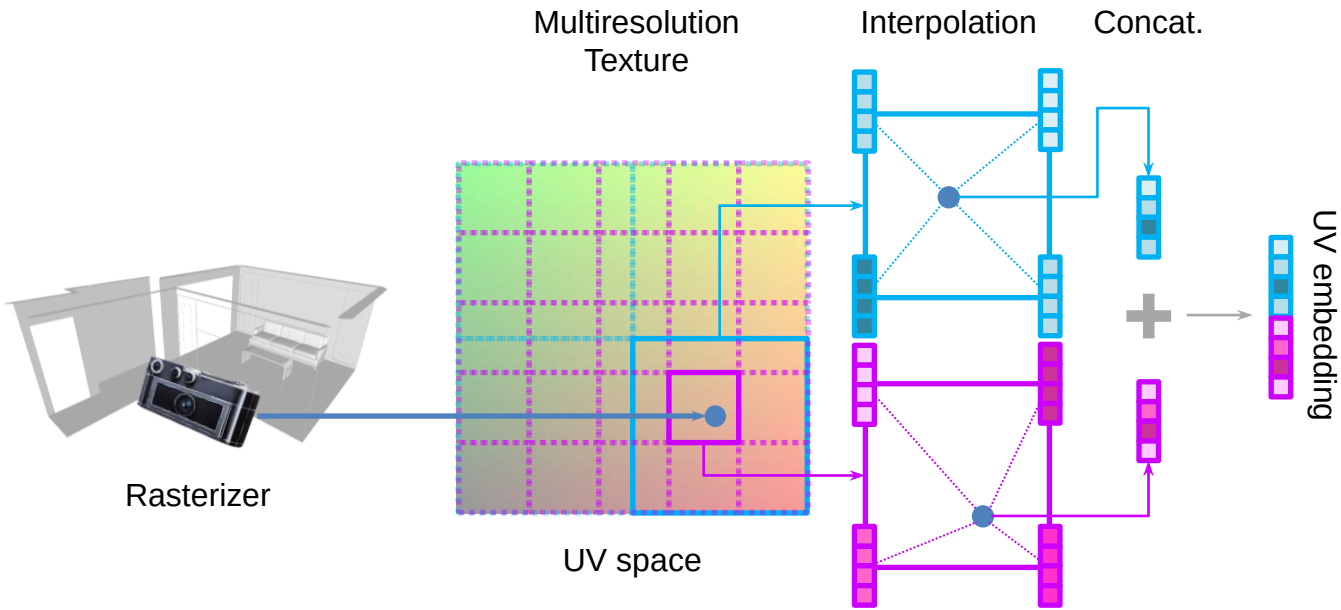
Generating Textures for 3D Scenes

- [CVPR 24' Highlight] SceneTex: High-quality Texture Synthesis for Indoor Scenes via Diffusion Priors

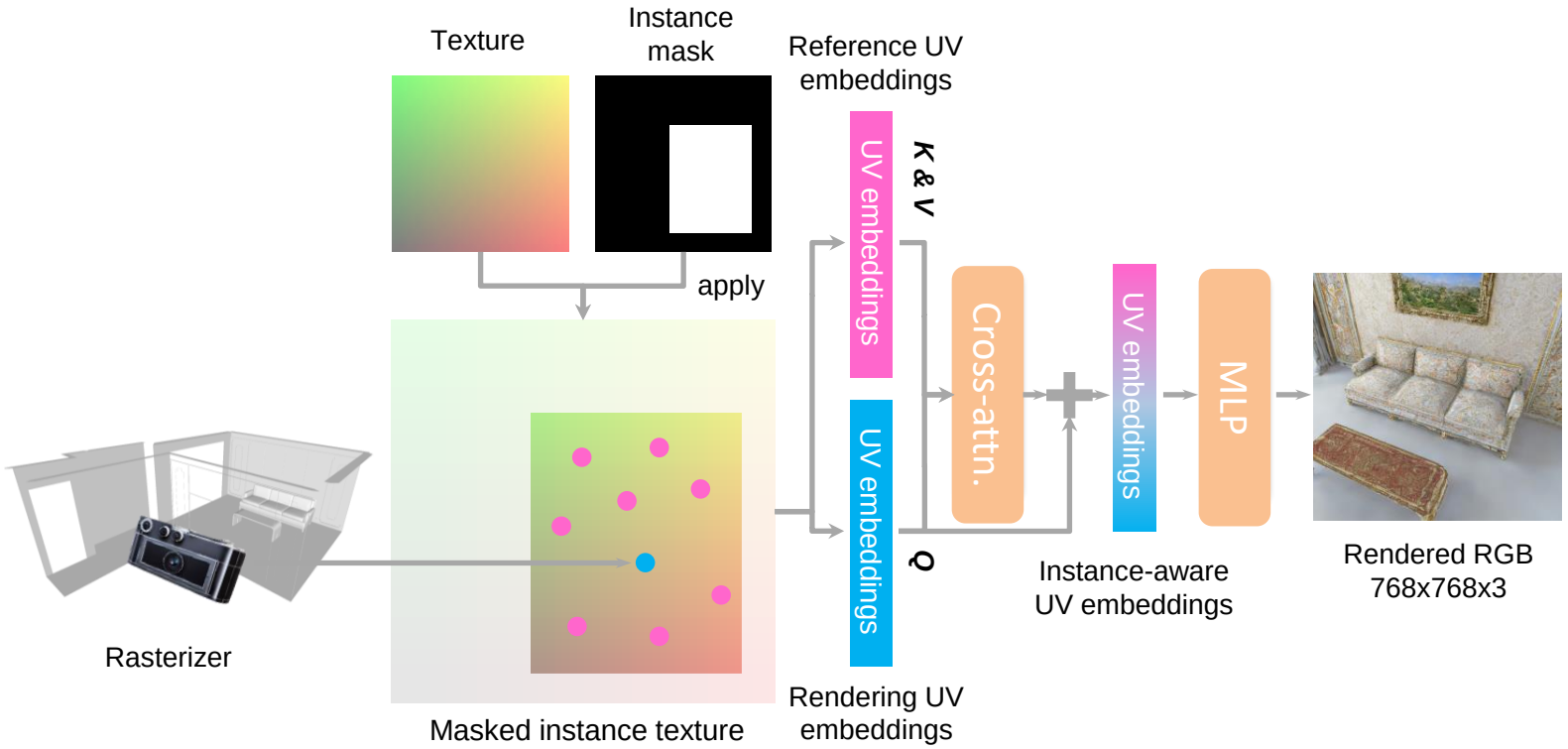
Architecture Overview



Multiresolution Texture



Cross-Attention Texture Decoder



Quantitative Results

Similarity between rendered views and input prompt

Realism of the rendered views

Method	CLIP	Inception Score
Latent-Paint	18.37	1.96
MVDiffusion	18.47	2.83
Text2Tex	20.83	2.87
Ours	**22.18**	**3.33**

We evaluate 10 scenes with 2 different prompts for each scene. We render 200 views in each scene.

User Study

We present 10 random views for the target scene

7. On a scale of 1 to 5, how would you rate the scene texture overall? Please ingore the white background. You might consider aesthetics, realism, smoothness, or even your personal taste. *

(very bad) 1 2 3 4 5 (very good)

8. On a scale of 1 to 5, how well do you think the presented appearance match the style in following description? *

a midcentury style bedroom

(very bad) 1 2 3 4 5 (very good)

75 human users rate the visual quality and prompt fidelity on a scale of 1-5.

User Study

How good does the appearance look?

How well does the texture match the prompt?

Method	Visual Quality	Prompt Fidelity
Latent-Paint	1.57	2.11
MVDiffusion	3.09	3.12
Text2Tex	2.62	3.04
Ours	**4.40**	**4.29**

Qualitative Results


Latent-Paint


MVDiffusion

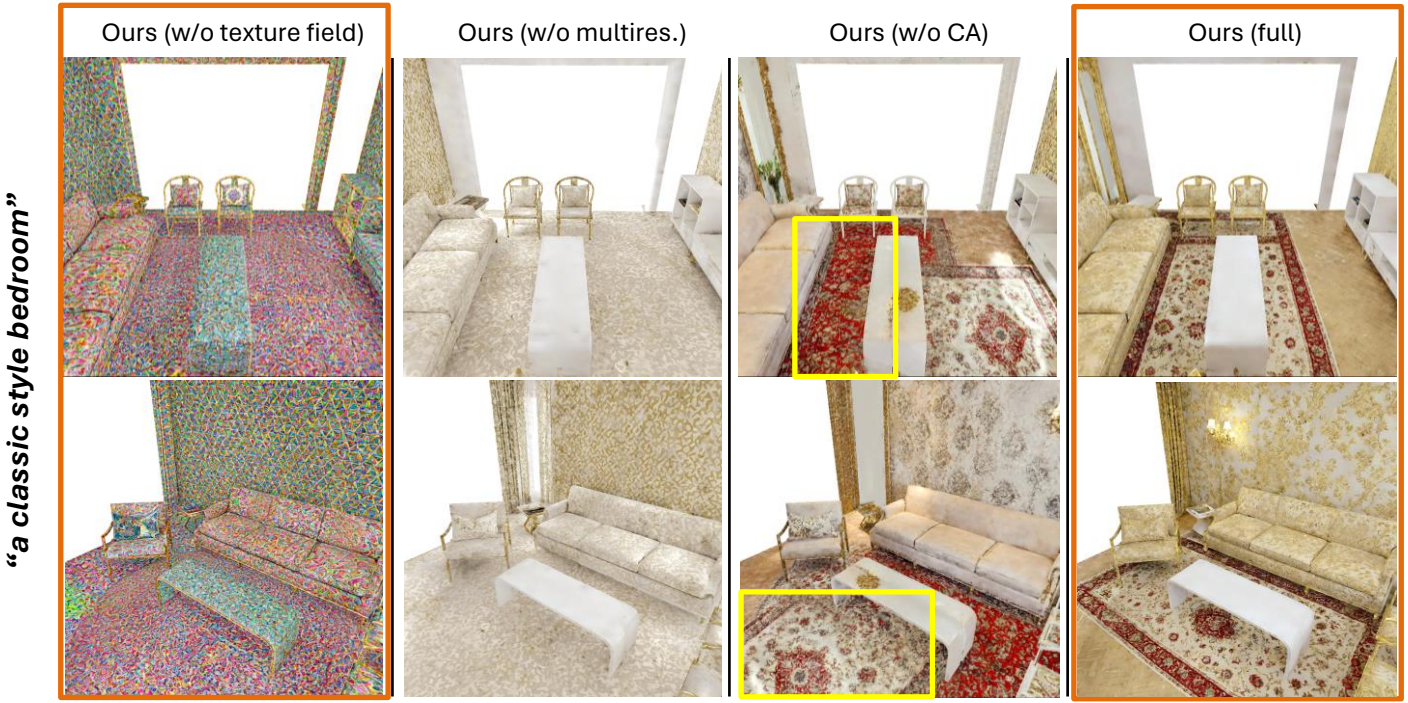

Text2Tex


Ours


Latent-Paint
MVDiffusion
Text2Tex
Ours

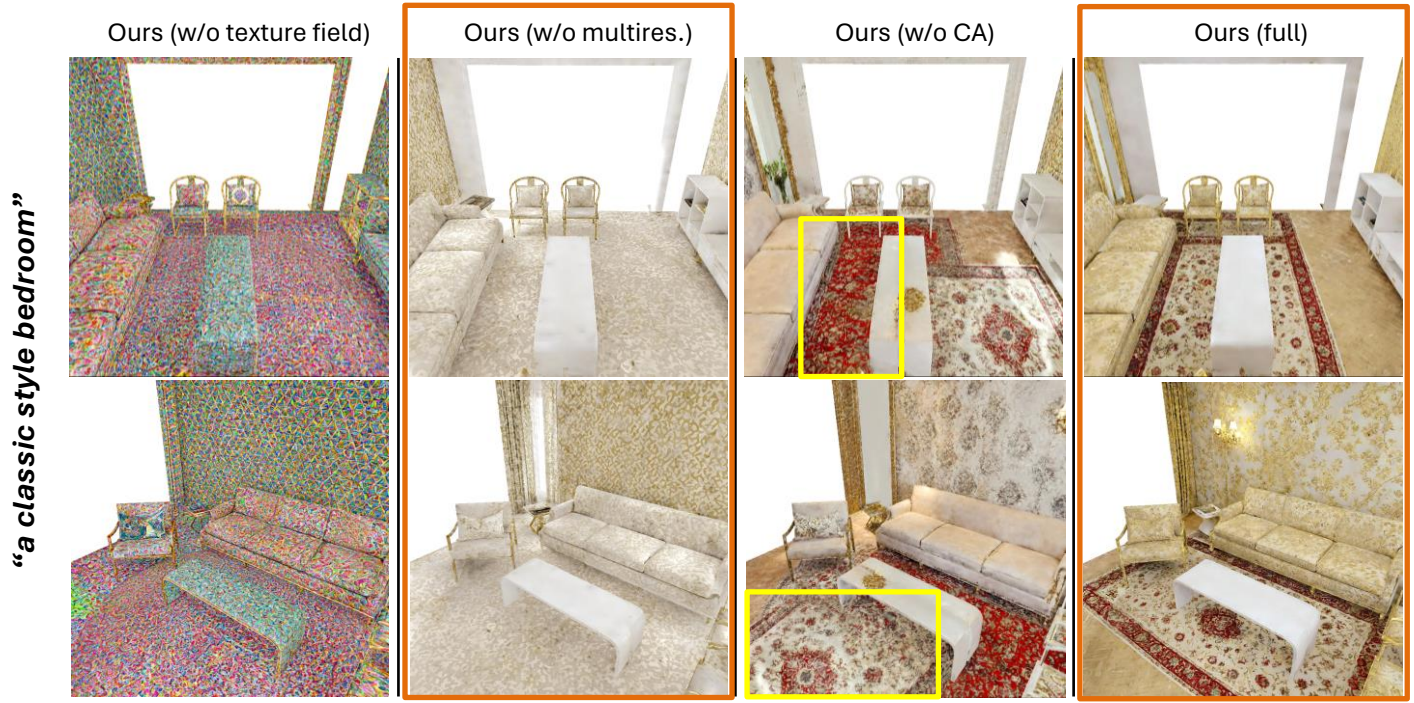
Ablation Study

- Is the texture field better than the RGB texture?



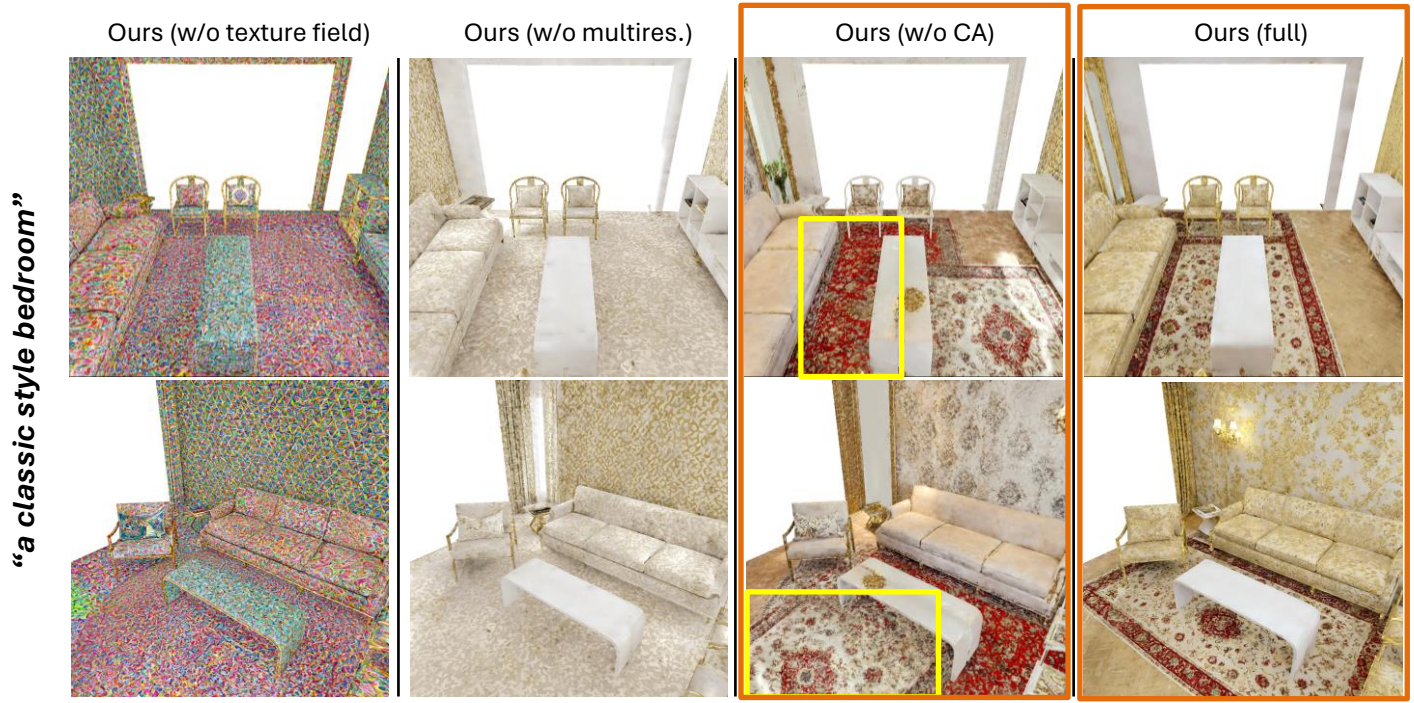
Ablation Study

- Does multiresolution texture help?



Ablation Study

- Does cross-attention texture decoder help?



Application

Scene Stylization


Baroque


Bohemian


French


Japanese


Baroque


Bohemian

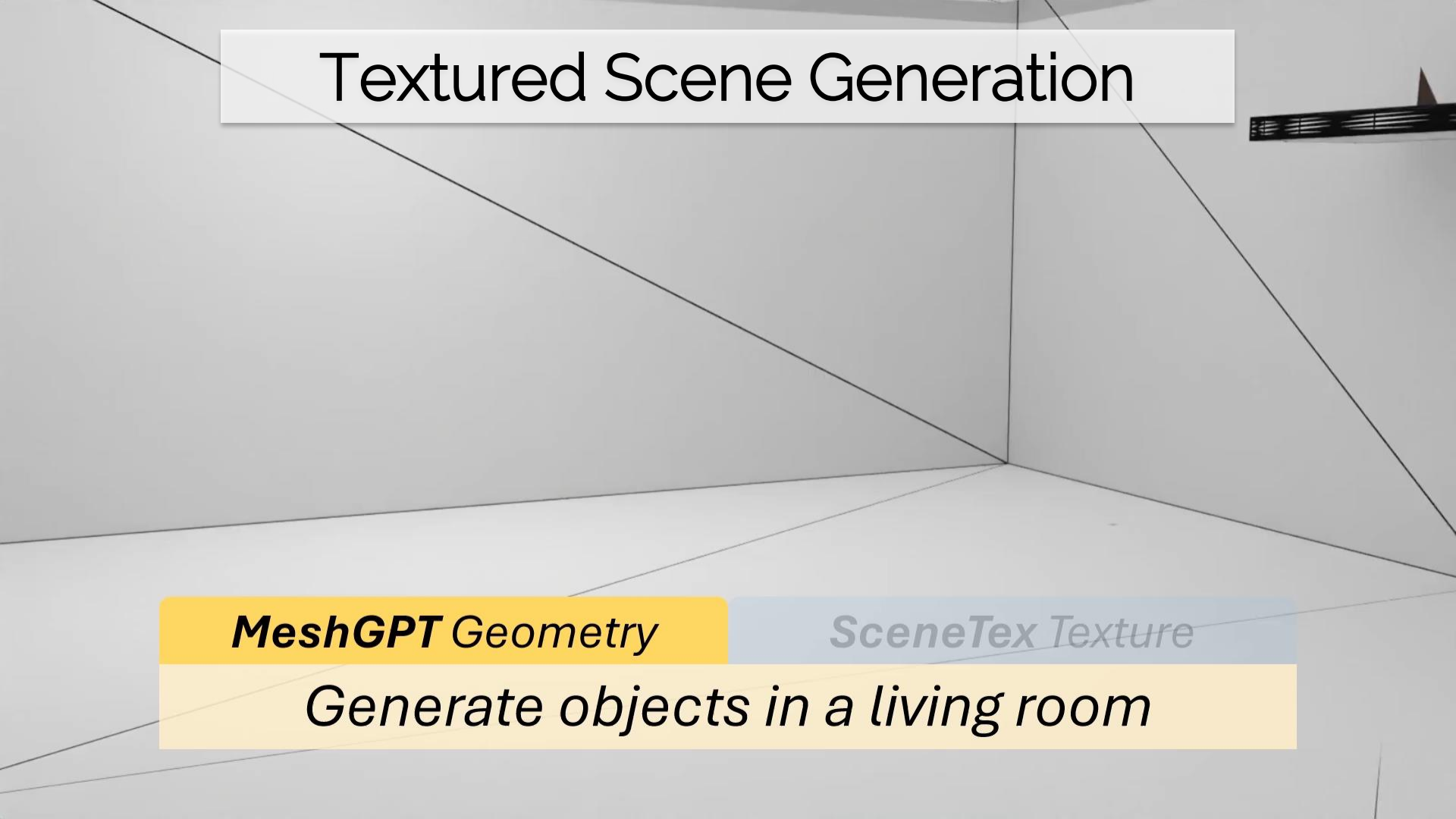

French


Japanese

Application

Textured Scene Generation

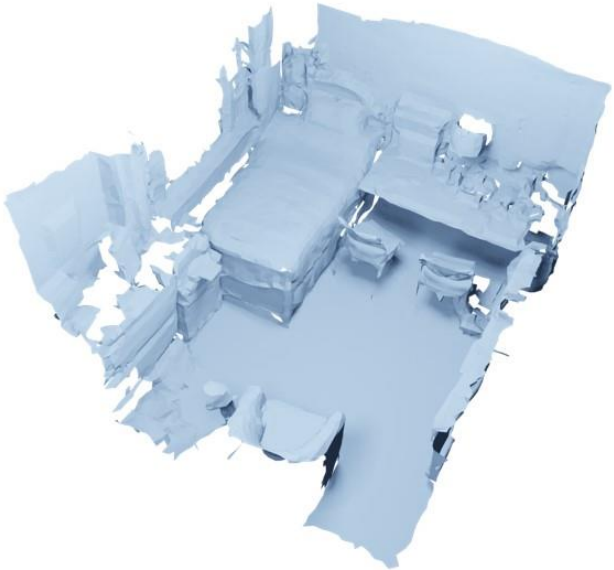
Textured Scene Generation



Application

Real-world 3D indoor scene

“A Japanese style bedroom”



ScanNet geometry

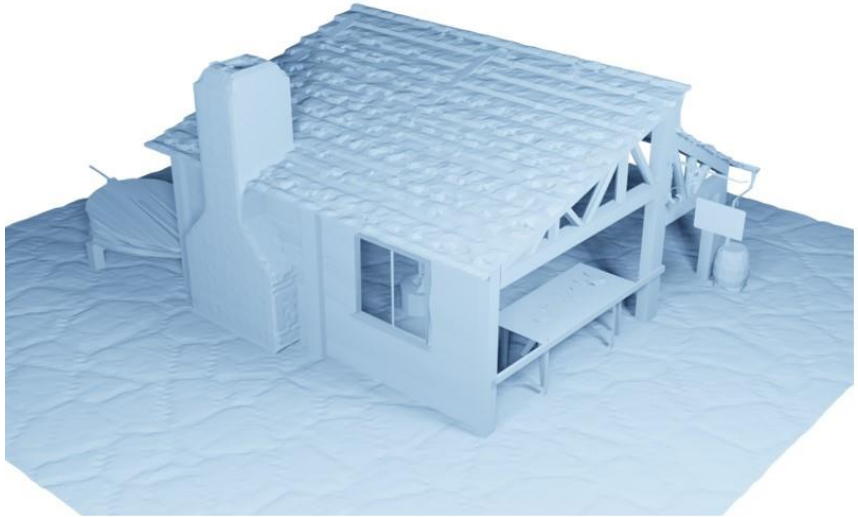
ScanNet color

w/ our texture

Application

3D outdoor scene

“A blacksmith shop covered in snow”

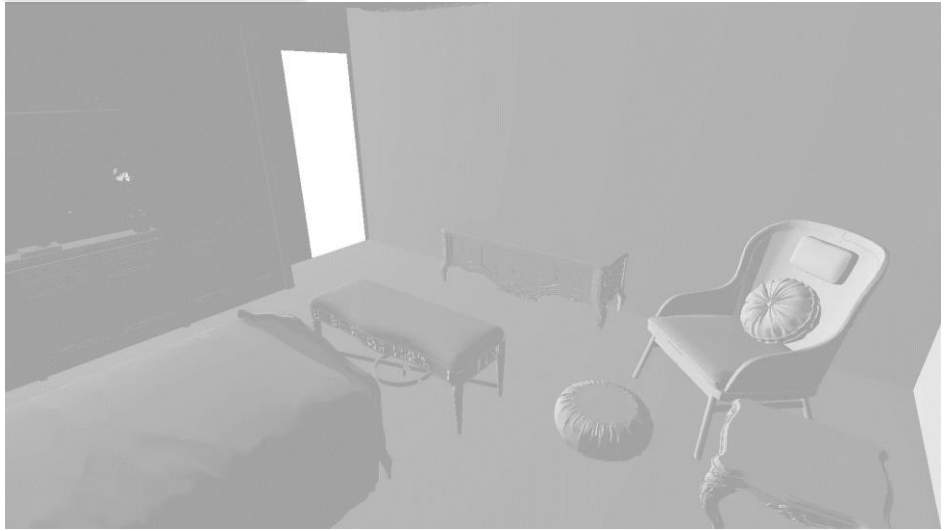


Scene geometry

Ours

Limitations

Geometry



Textured Scene

- Baked-in illumination -> PBR textures
- Object hallucinations -> stronger conditional signals, e.g. normal maps, canny edge

High-quality Texture Synthesis for 3D Scenes

Zhenyu Chen – 陈振宇
Visual Computing & AI Group
Technical University of Munich

青稞社区

扫描二维码，关注我的视频号

关注视频号
预约更多直播

关注“青稞AI”
获取更多技术干货！

扫一扫上面的二维码图案，加我为朋友。

添加小助手
申请进群交流