上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

NEURAL INFORMATION PROCESSING SYSTEMS

SceneCraft: Layout-Guided 3D Scene Generation
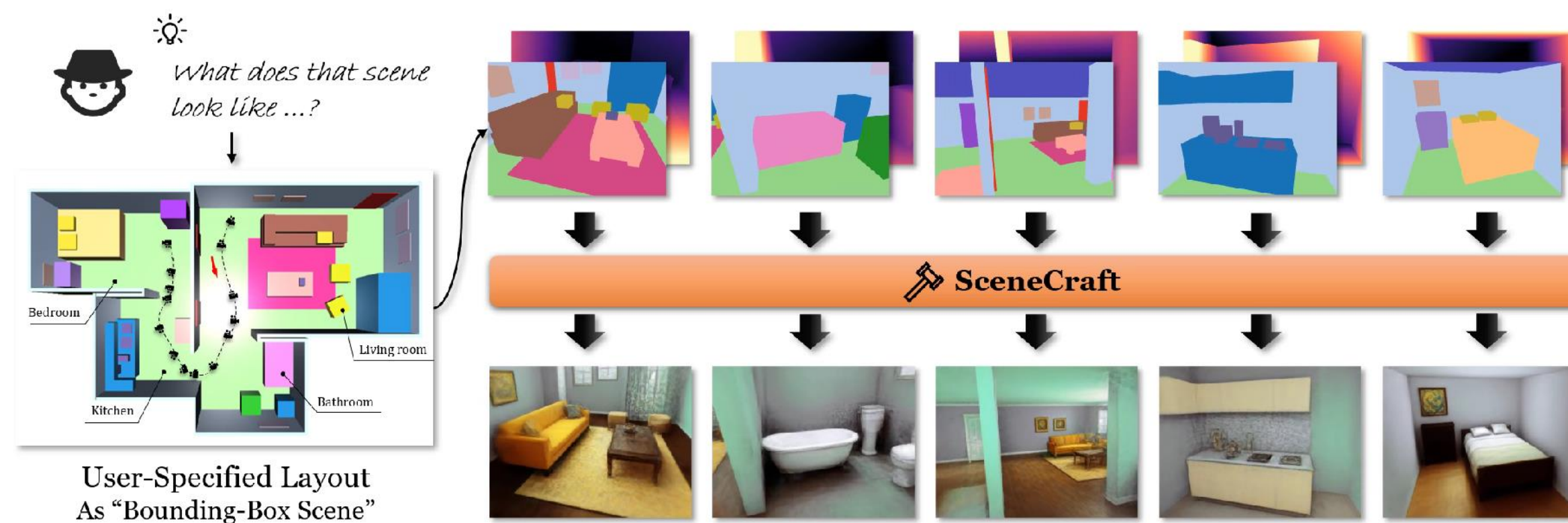
Xiuyu Yang*, Yunze Man*, Jun-Kun Chen, Yu-Xiong Wang

SceneCraft

We generate complex 3D indoor scenes conditioned on free-form layout and viewpoints

- Build Scenes like in Minecraft!
- Operate on free multi-view trajectories!
- Customize the style via texts!

Task: 3D Scene Generation

Input: User-specified 3D semantic layouts and prompts.

Output: 3D-consistent and realistic scenes aligned with spatial layouts.
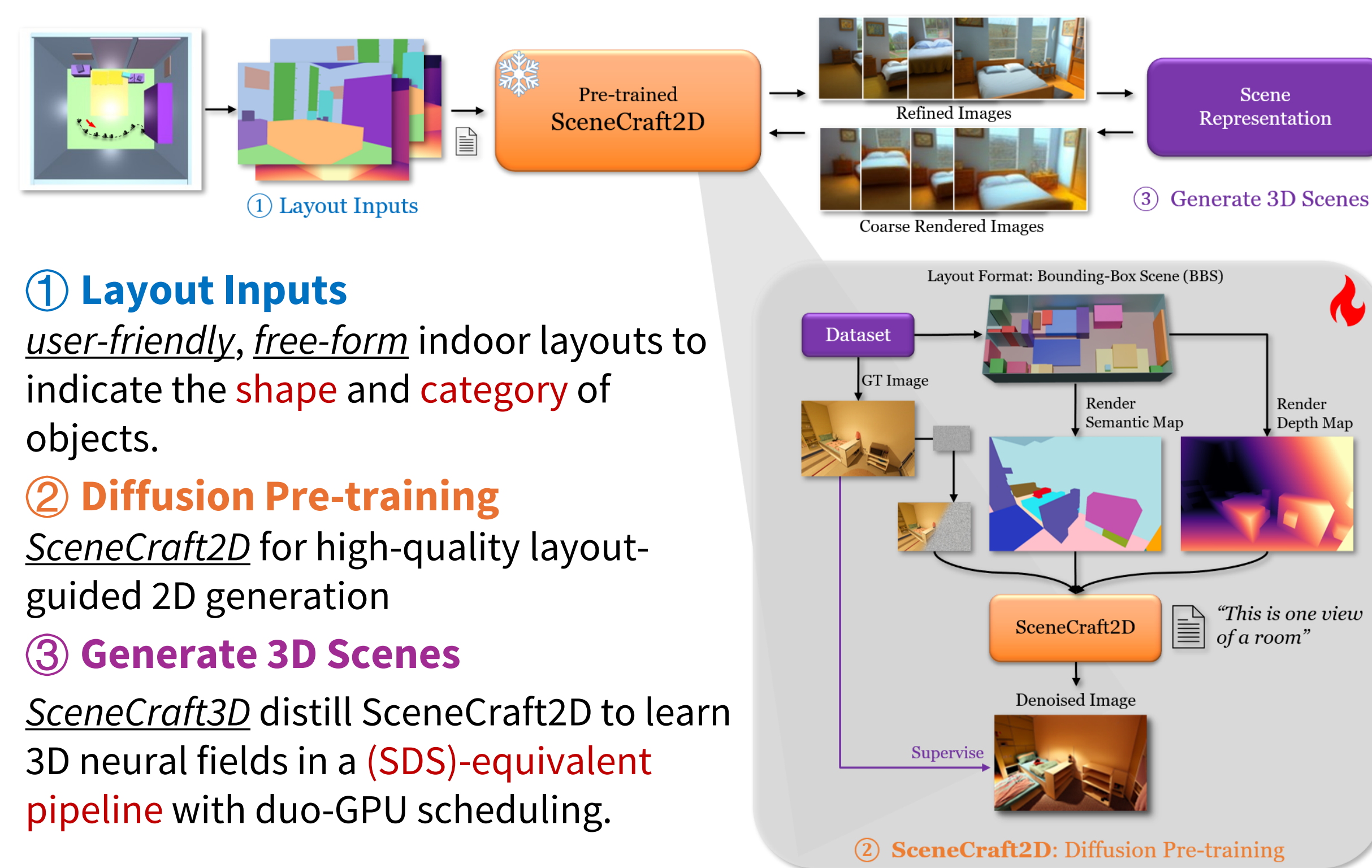
Previous Limitations:
1. Lack of layout control
2. Limited to simple single-room

What does that scene look like ...?

User-Specified Layout As "Bounding-Box Scene"

Framework: Distillation-Guided Scene Generation

① Layout Inputs
② Pre-trained SceneCraft2D → Refined Images → Scene Representation
Coarse Rendered Images
③ Generate 3D Scenes

① Layout Inputs
user-friendly, *free-form* indoor layouts to indicate the shape and category of objects.

② Diffusion Pre-training
SceneCraft2D for high-quality layout-guided 2D generation

③ Generate 3D Scenes
SceneCraft3D distill SceneCraft2D to learn 3D neural fields in a (SDS)-equivalent pipeline with duo-GPU scheduling.

Layout Format: Bounding-Box Scene (BBS)

Dataset → GT Image → Render Semantic Map → Render Depth Map

SceneCraft2D → "This is one view of a room"

Denoised Image

Supervise

② **SceneCraft2D**: Diffusion Pre-training

Complex Scene Generations (Scene A~D)

Indoor scenes of *large and irregular scale* and *complex room layouts*. E.g., Apartments composed of multiple interconnected rooms.
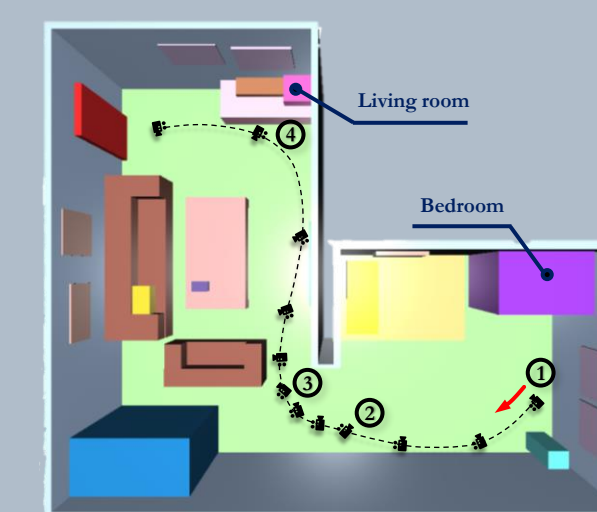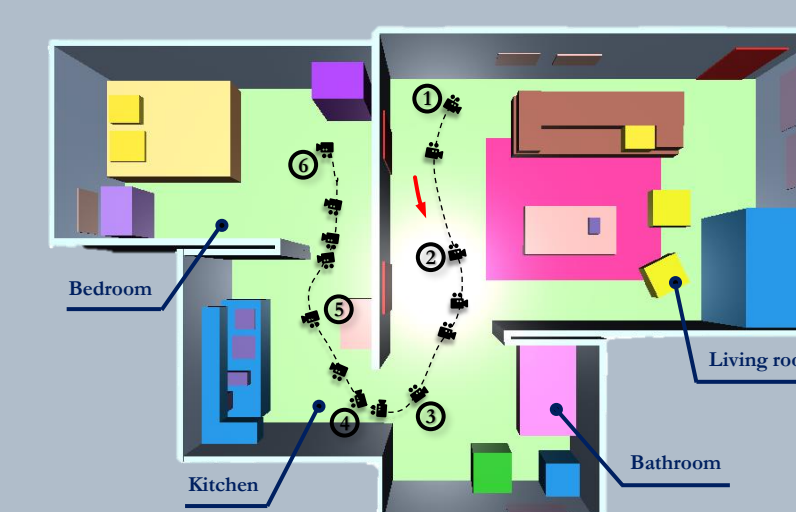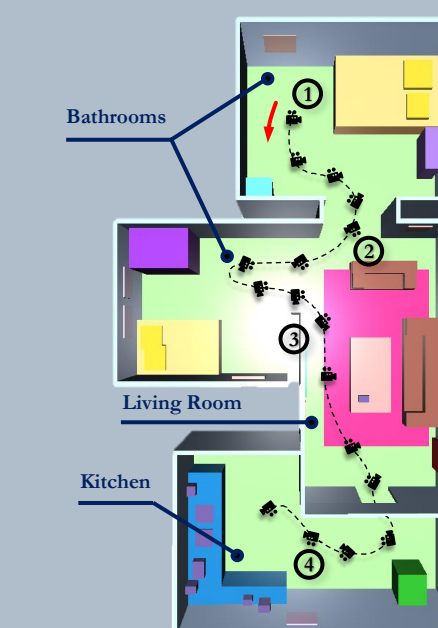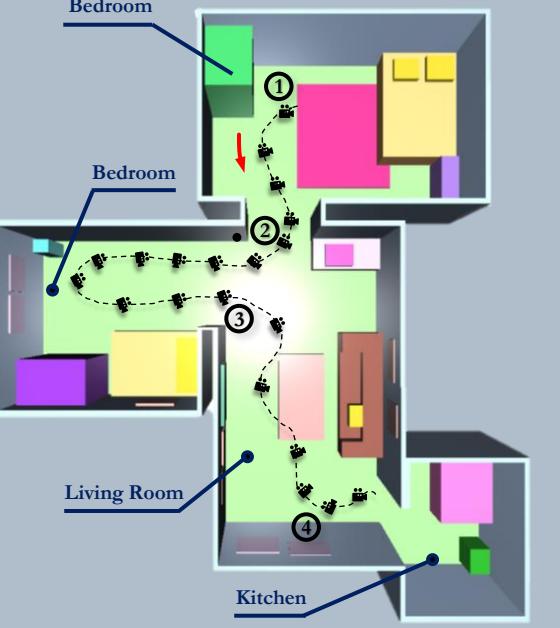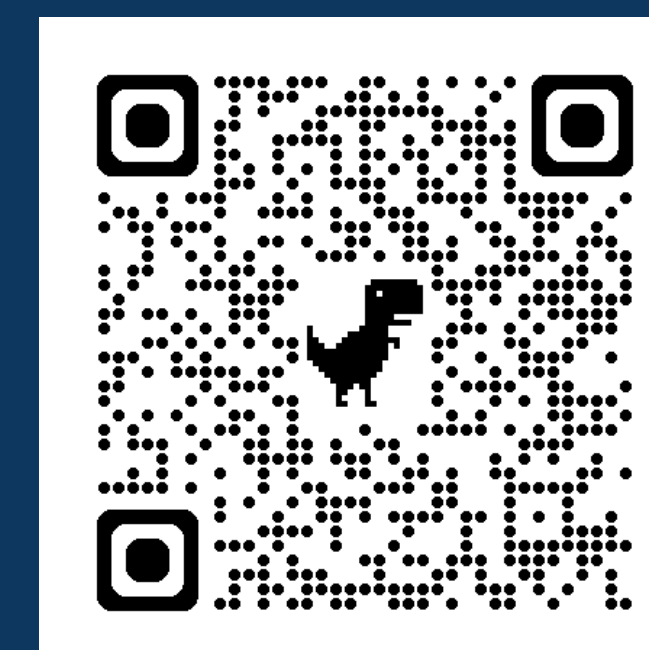
Performance of SceneCraft2D

"An Icy Winter Children's Bedroom."

"A Vampire's bloody bedroom."

Quantitative Evaluation

Method	2D Metrics		3D Quality	
	CS↑	IS↑	3DC↑	VQ↑
Text2Room	22.98	4.20	3.11	3.06
MVDiffusion	23.85	**4.36**	3.20	3.35
Set-the-scene	21.32	2.98	3.53	2.41
SceneCraft (Ours)	**24.34**	3.54	**3.71**	**3.56**

competitive performance in 2D metrics and 3D quality
- **3DC:** 3D consistency of scenes;
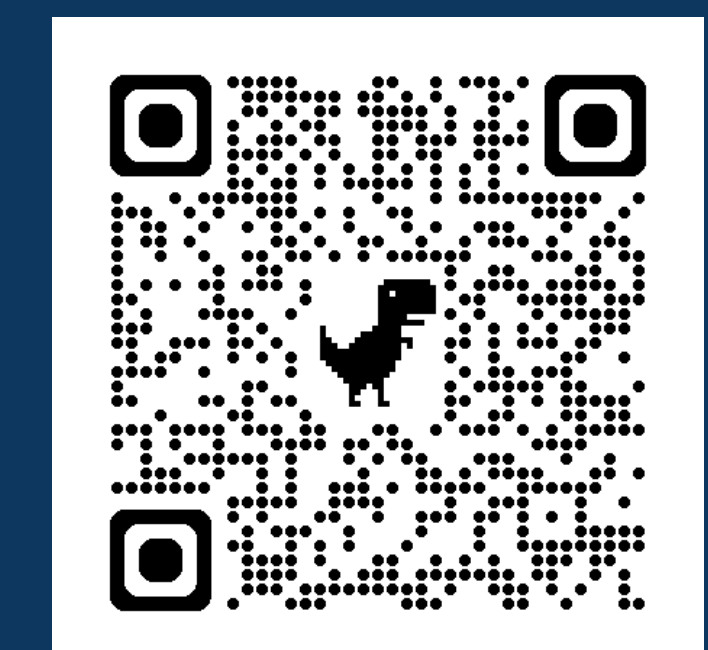- **VQ:** Overall Visual Quality;

Scene A

Scene B

Scene C

Scene D

Scene A

Scene B

Scene C

Scene D

Website / Paper

Code