NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

New Generative Models for 3D Content Generation

Xingang Pan

Assistant Professor

College of Computing and Data Science

Nanyang Technological University

24 Mar 2025

3D Content Creation

3D content lays the foundation for broad applications

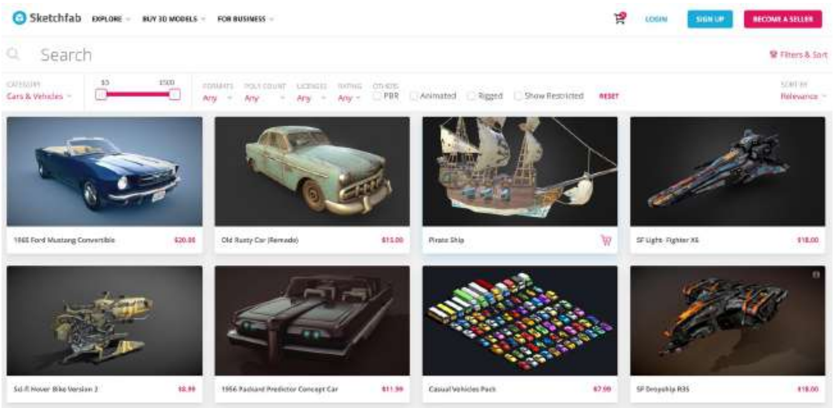
Movies / VFX

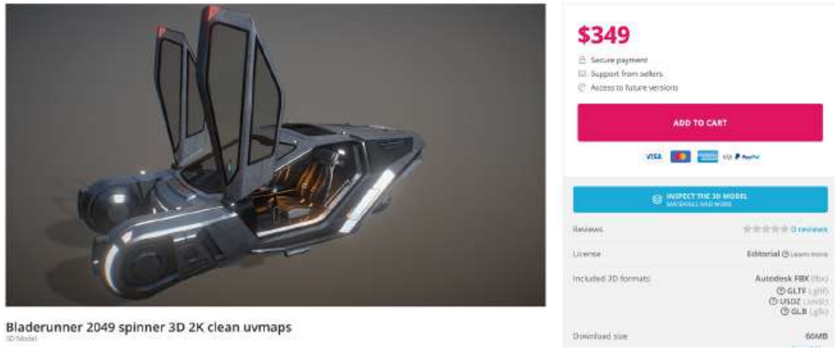
Gaming

AR / VR / Metaverse

Manufacturing

3D assets: limited and expensive





3D Content Creation via Generative Models

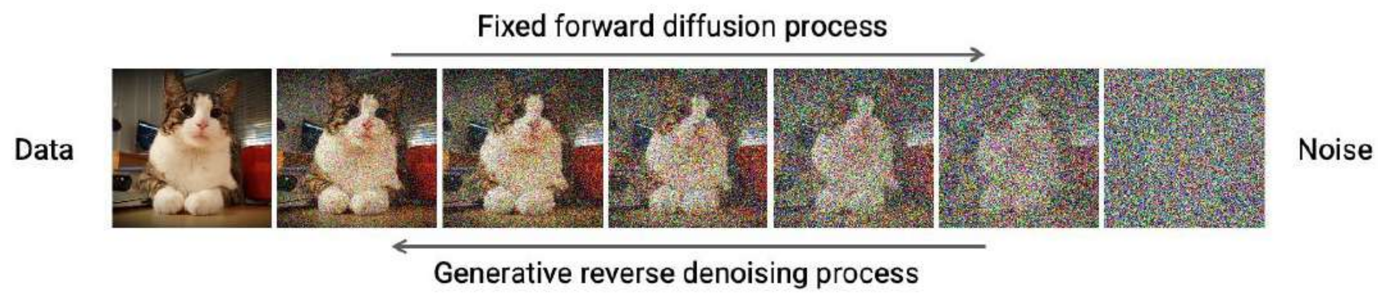
$z \sim \mathcal{N}(0, \mathrm{I})$

Generative Model

Objverse, https://objaverse.allenai.org/

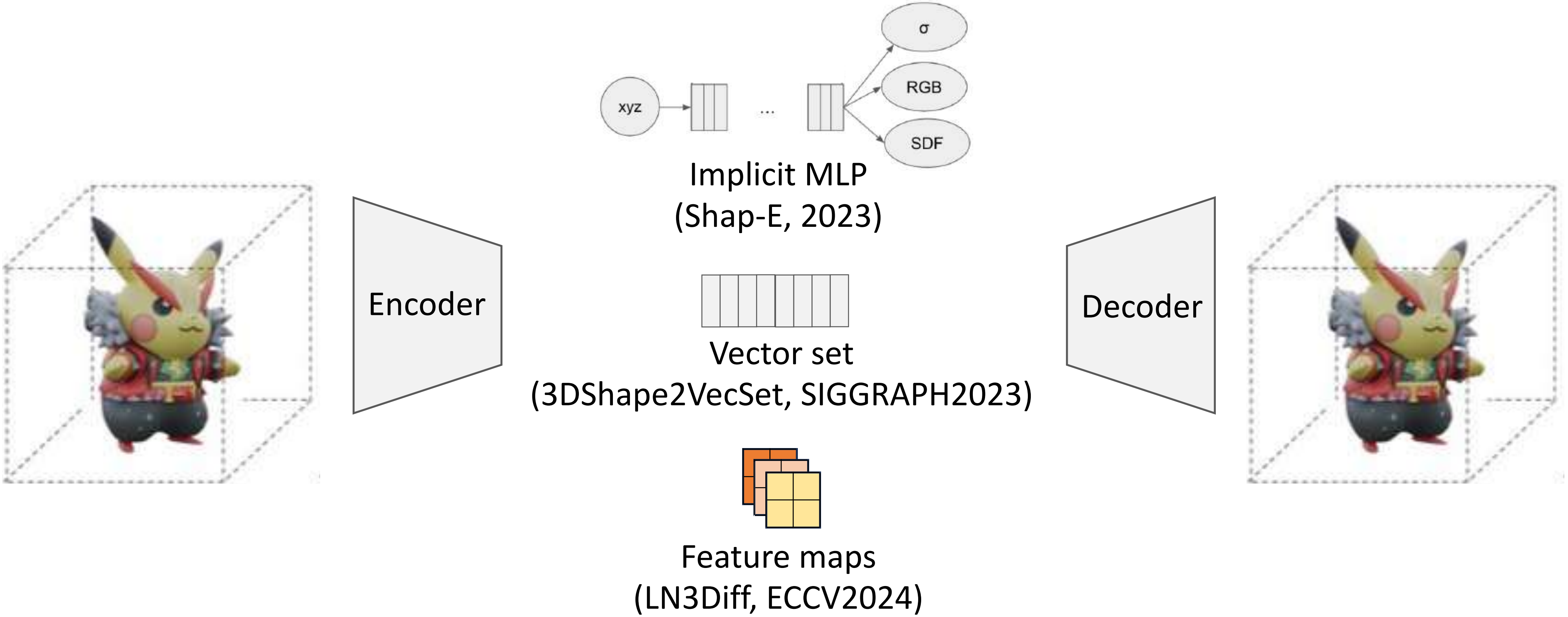
Generative Models

Diffusion Model



Autoregressive Model

Background on Native 3D Diffusion Models



- Lack of explicit 3D-aware latent space for interactive editing.
- Lack of high-quality texture and efficient 3D VAE encoding from 2D inputs.

GaussianAnything: Interactive Point Cloud Latent Diffusion for 3D Generation

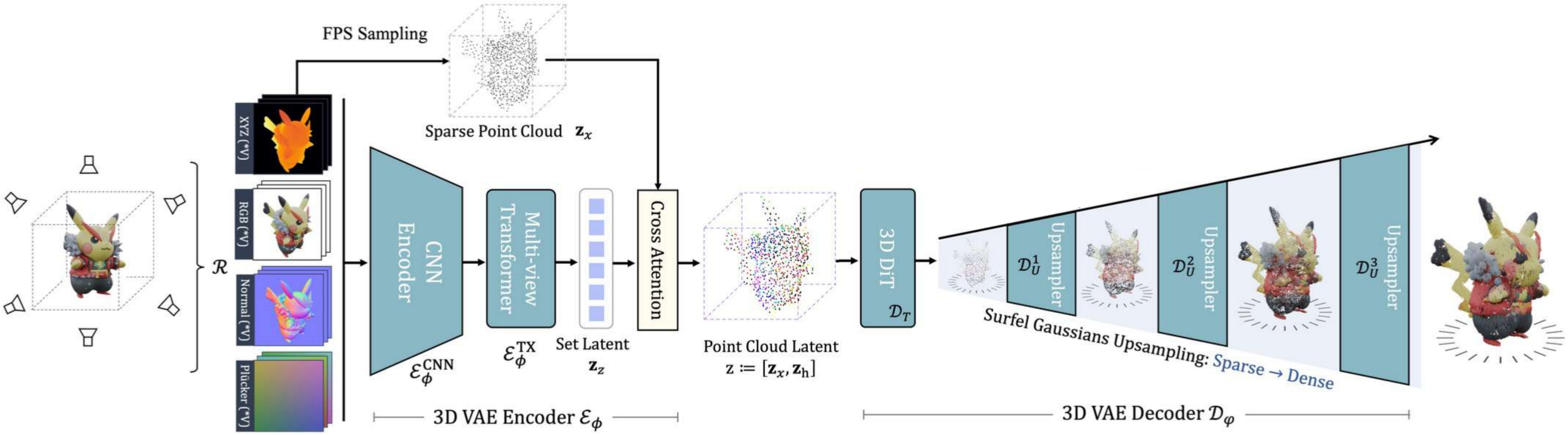
Yushi Lan[1] Shangchen Zhou[1] Zhaoyang Lyu[2] Fangzhou Hong[1]
Shuai Yang[3] Bo Dai[2] Xingang Pan[1] Chen Change Loy[1]
[1]S-Lab, NTU Singapore [2]Shanghai AI Lab [3]Peking University

https://nirvanalan.github.io/projects/GA/

Lan et al, GaussianAnything: Interactive Point Cloud Latent Diffusion for 3D Generation, ICLR2025

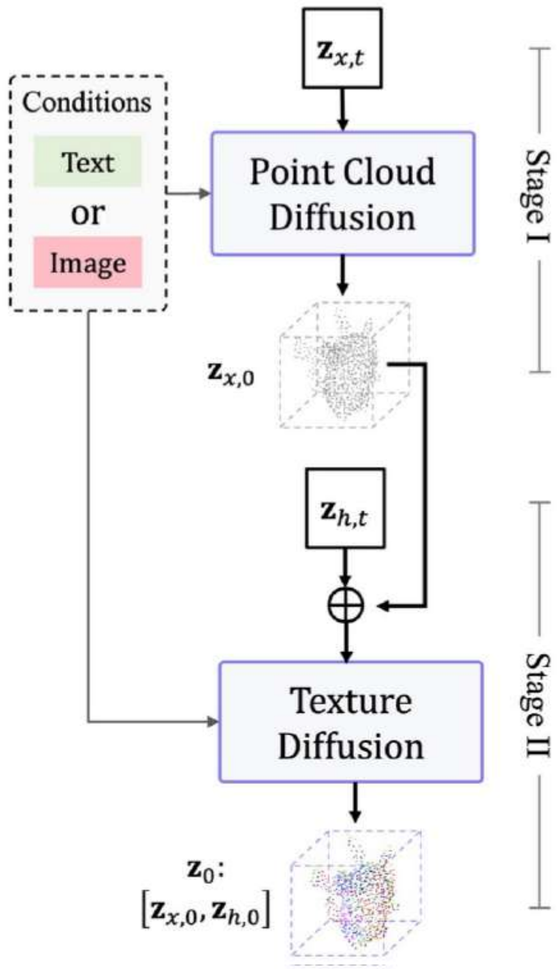
3D VAE with Structured Latent

Input	**Latent**	**Output**
Point cloud + multi-view RGB, Depth, Normal	Point cloud + Embedding for each point	High-quality surfel Gaussians

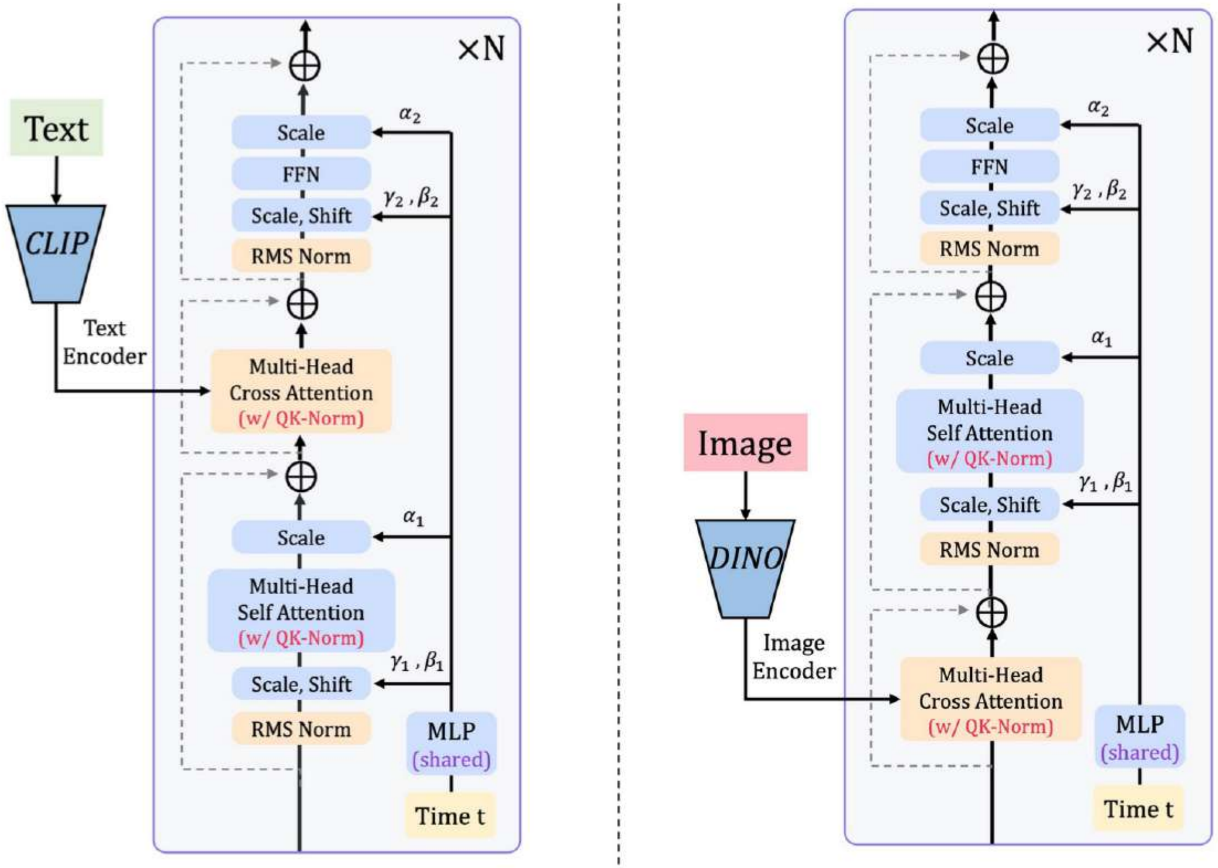


Pipeline of the 3D VAE of GaussianAnything.

Cascaded Native 3D Diffusion



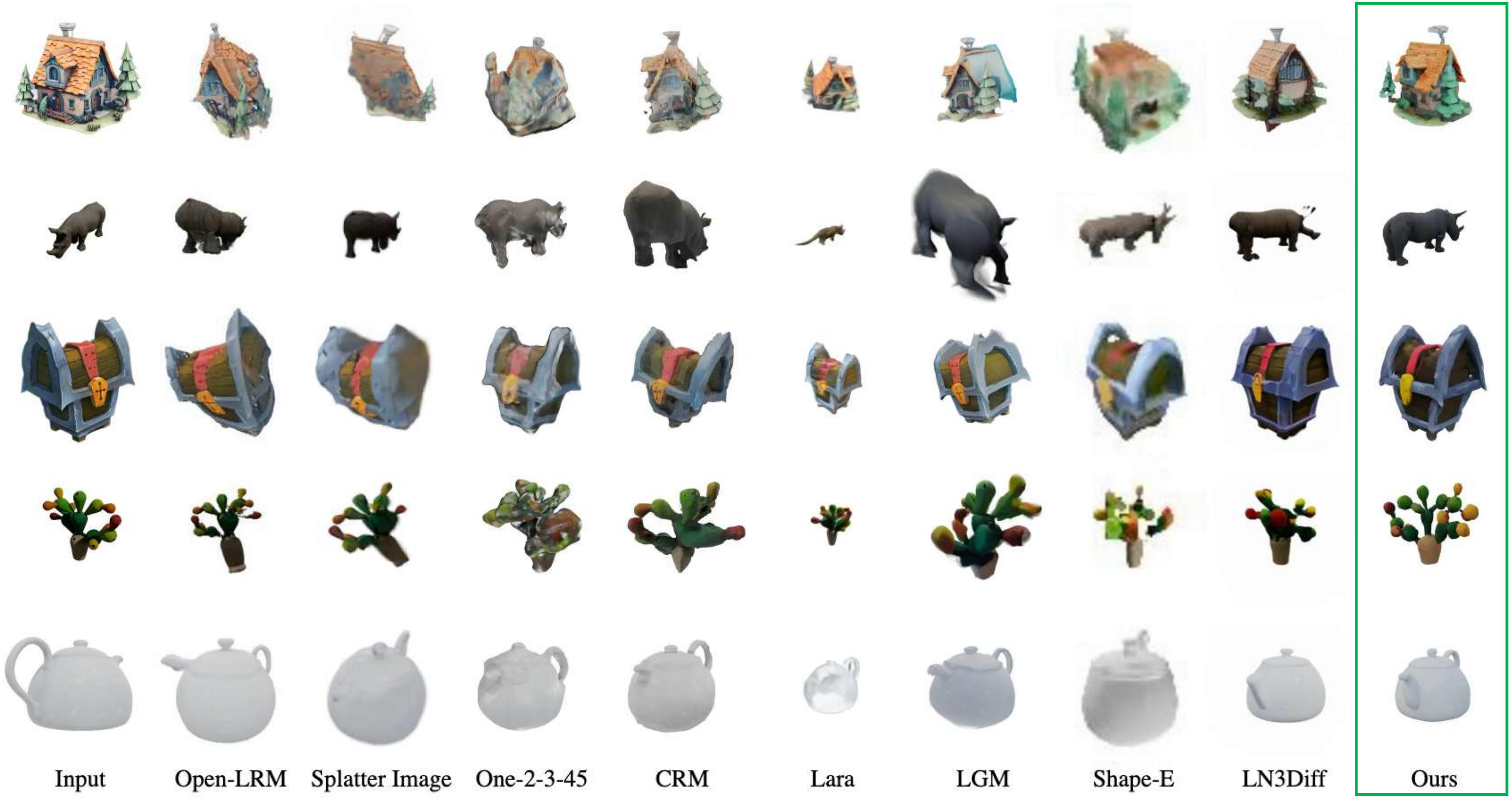
Two-Stage Diffusion



(a) DiT Block (Text condition)

(b) DiT Block (Image condition)

Qualitative Results (Image-to-3D)

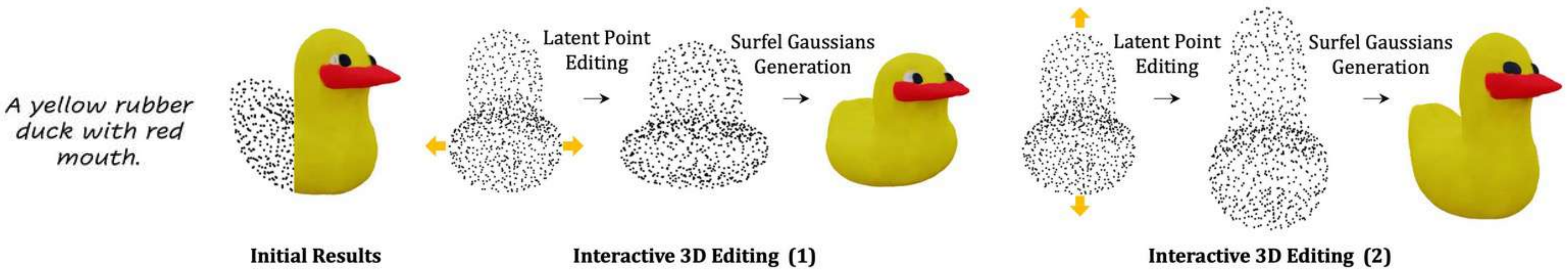


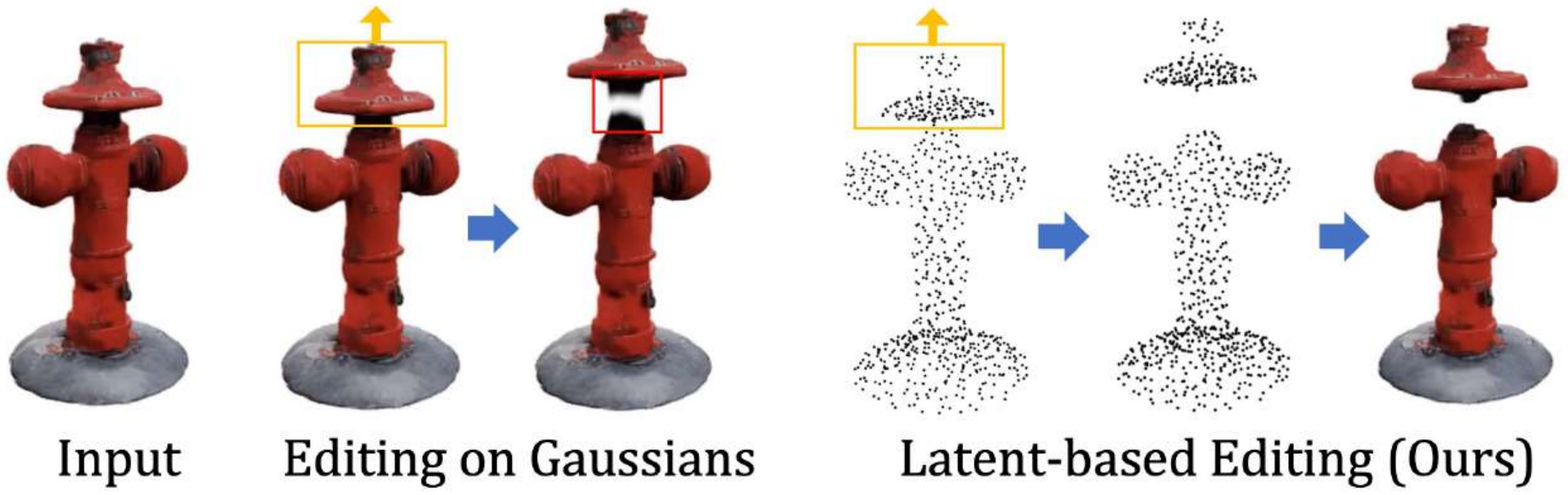
Quantitative performance (Image-to-3D)

Method	FID↓	KID(%)↓	MUSIQ↑	P-FID↓	P-KID(%)↓	COV(%)↑	MMD(‰)↓
OpenLRM	38.41	1.87	45.46	35.74	12.60	39.33	29.08
Splatter-Image	48.80	3.65	30.33	19.72	7.03	37.66	30.69
One-2-3-45 (V=12)	88.39	6.34	59.02	72.40	30.83	33.33	35.09
CRM (V=6)	45.53	1.93	64.10	35.21	13.19	38.83	28.91
Lara (V=4)	43.74	1.95	39.37	32.37	12.44	39.33	28.84
LGM (V=4)	19.93	0.55	54.78	40.17	19.45	50.83	22.06
Shape-E	138.53	11.95	31.51	20.98	7.41	61.33	19.17
LN3Diff	29.08	0.89	50.39	27.17	10.02	55.17	19.94
Ours	24.21	0.76	65.17	8.72	3.22	59.50	15.48

Text-to-3D performance

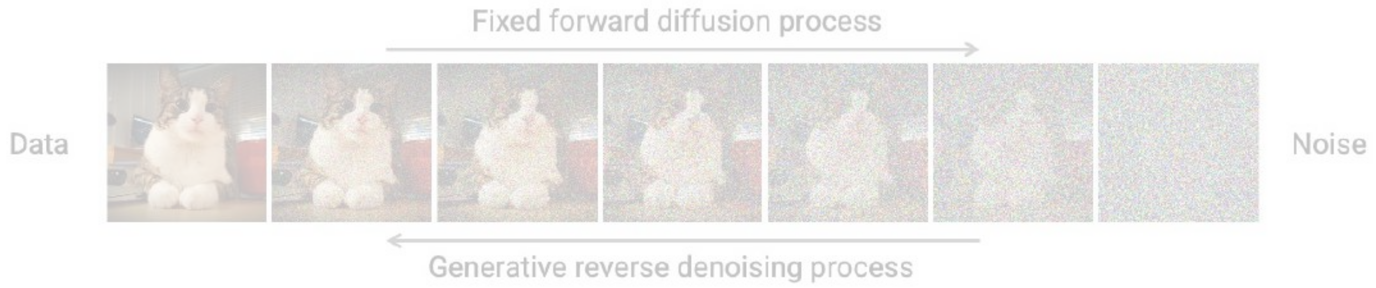
Interactive 3D Editing





Generative Models

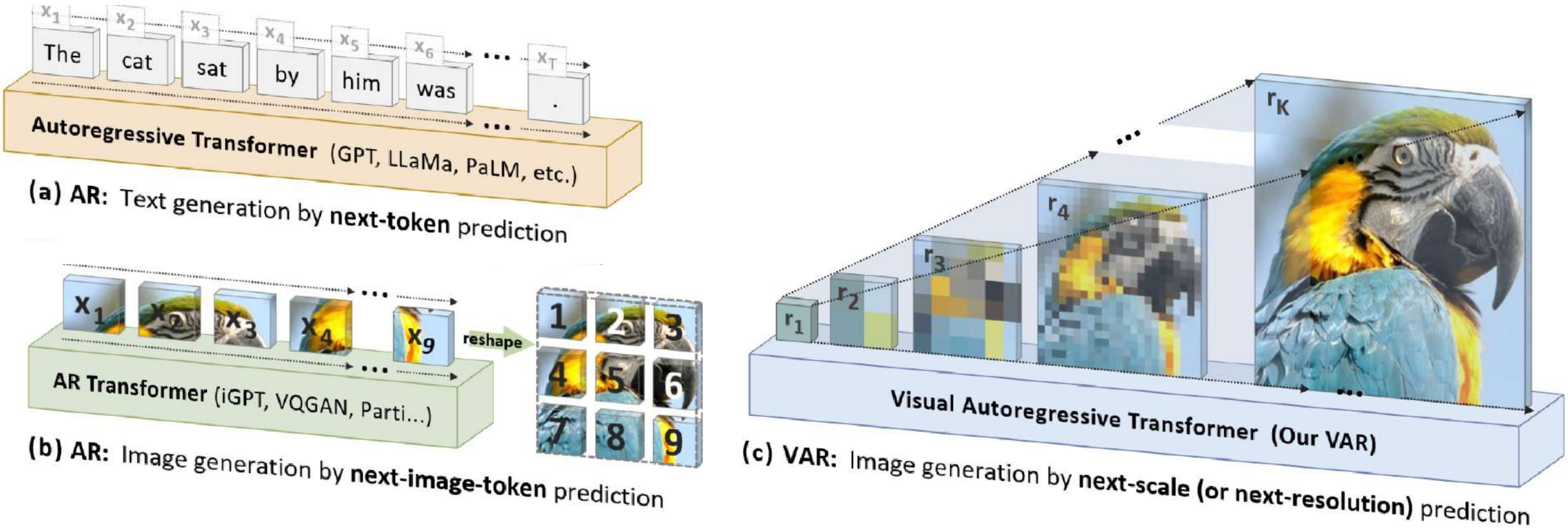
Diffusion Model



- A native 3D Diffusion Model, GaussianAnything:
  Structured latent space, better design, better performance

Autoregressive Model

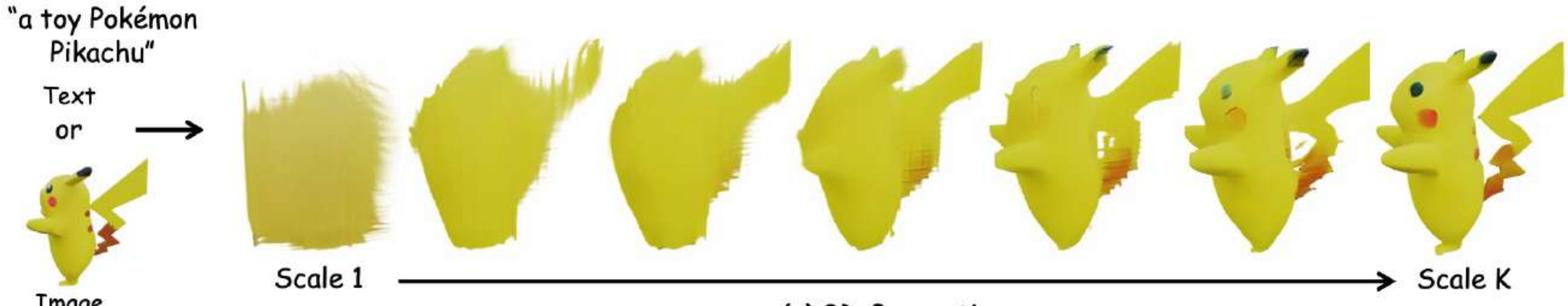
Autoregressive Generation



Tian et al, Visual Autoregressive Modeling: Scalable Image Generation via Next-Scale Prediction, NeurIPS2024

SAR3D: Autoregressive 3D Object Generation and Understanding via Multi-scale 3D VQVAE

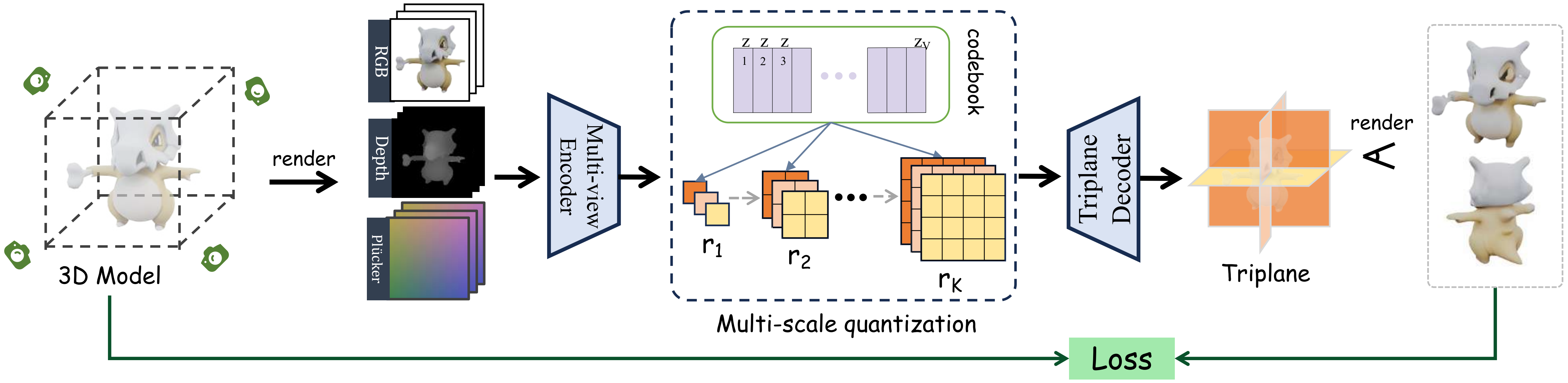
Yongwei Chen[1], Yushi Lan[1], Shangchen Zhou[1], Tengfei Wang[2], Xingang Pan[1]
[1]S-Lab, Nanyang Technological University
[2]Shanghai Artificial Intelligence Laboratory



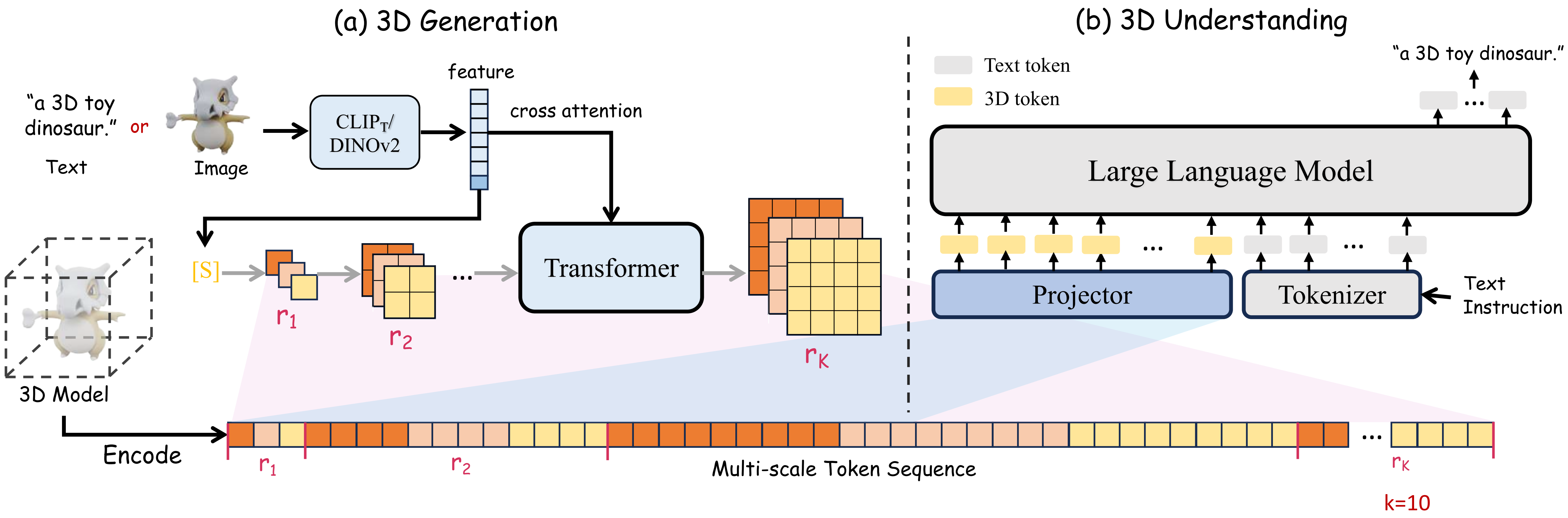
(a) 3D Generation

Chen et al, SAR3D: Autoregressive 3D Object Generation and Understanding via Multi-scale 3D VQVAE, CVPR2025

SAR3D – VAE with multi-scale quantization

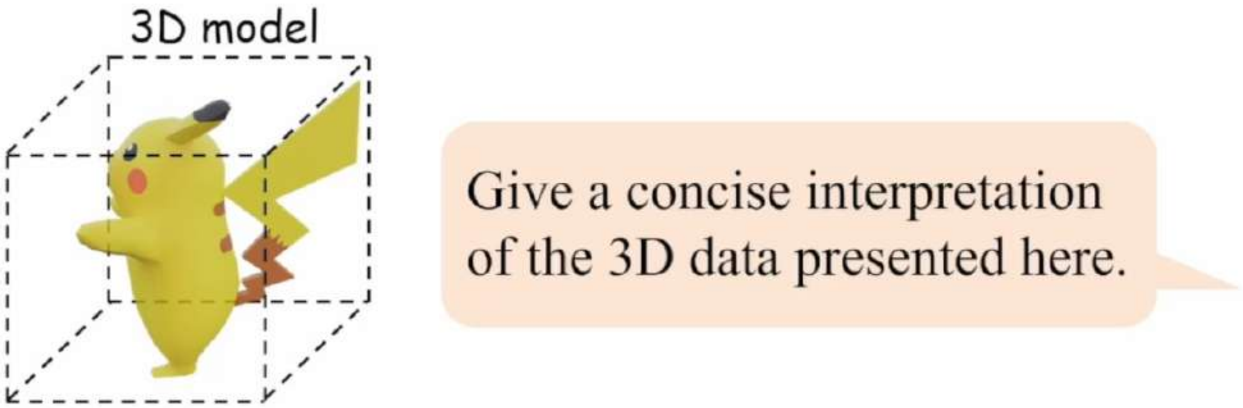


SAR3D -- Method



SAR3D -- Results

Fast 3D generation (<1s)



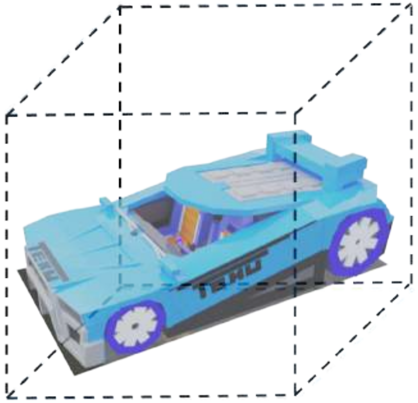
Detailed 3D understanding

SAR3D -- Experiments

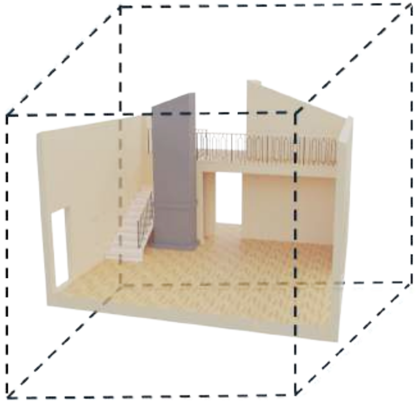
(a) Single Image to 3D

(b) Text to 3D

SAR3D -- Experiments

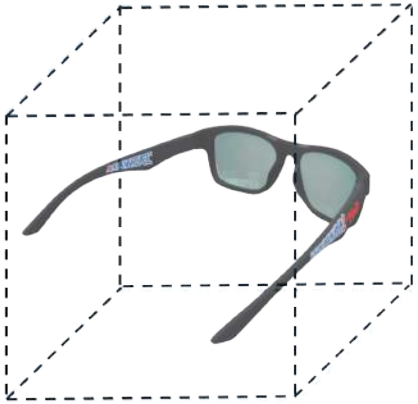


A sleek and aerodynamic blue and white racing car with a futuristic design, featuring racing stripes, a spoiler on the back, and a low profile.



A small, wooden house with a rectangular shape, staircase leading up to the entrance, and a patio area in front.

A wooden desk and chair set with a rectangular shape, featuring a simple and minimalistic design. The desk has a wooden top with a metal base, while the chair has a wooden seat and backrest.



A unique pair of black and green sunglasses with a slim and curved frame, featuring green lenses and a distinctive design.

3D Captioning. Given a 3D model, our method can generate captions that contain both category and details.

SAR3D -- Comparison

Method	FID↓	KID(%)↓	MUSIQ↑	COV(%)↑	MMD(‰)↓	Latency-V100 (s) ↓
Splatter-Image	48.80	3.65	30.33	37.66	30.69	0.83
OpenLRM	38.41	1.87	45.46	39.33	29.08	7.21
One-2-3-45 (V=12)	88.39	6.34	59.02	33.33	35.09	59.23
Lara (V=4)	43.74	1.95	39.37	39.33	28.84	11.93
CRM (V=6)	45.53	1.93	64.10	38.83	28.91	22.10
LGM (V=4)	19.93	0.55	54.78	50.83	22.06	3.87
Shap-E	138.53	11.95	31.51	61.33	19.17	9.54
LN3Diff	29.08	0.89	50.39	55.17	19.94	7.51
GaussianAnything	24.21	0.76	65.17	59.50	15.48	15.02
SAR3D-NeRF	22.55	0.42	65.77	74.17	13.63	1.64
SAR3D-Flexicubes	27.30	0.63	67.24	71.50	15.25	2.92

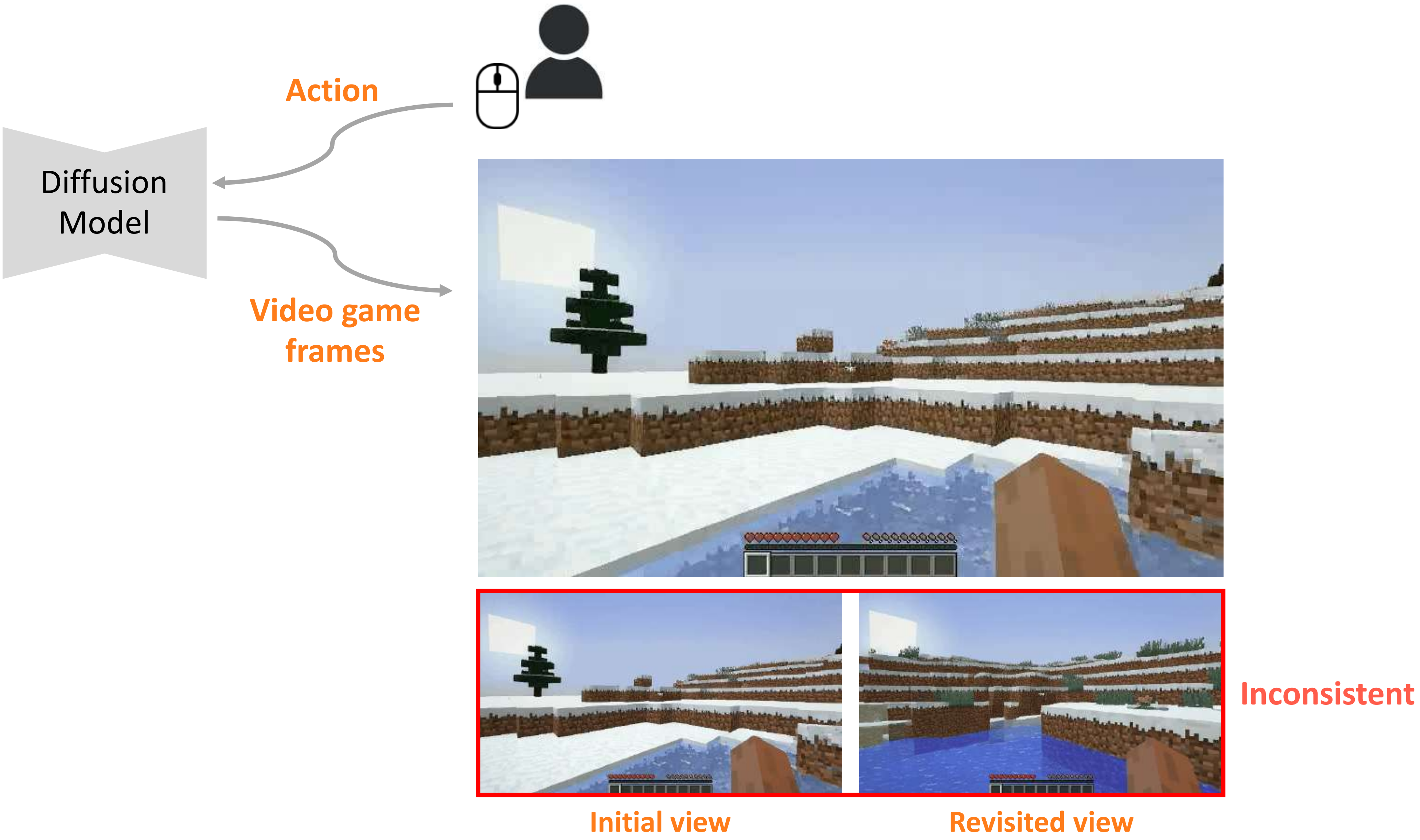
- Autoregressive Model for 3D generation can perform as well as diffusion models while being more efficient

Generative AI as 3D Game Engine?

Minecraft powered purely by Generative AI

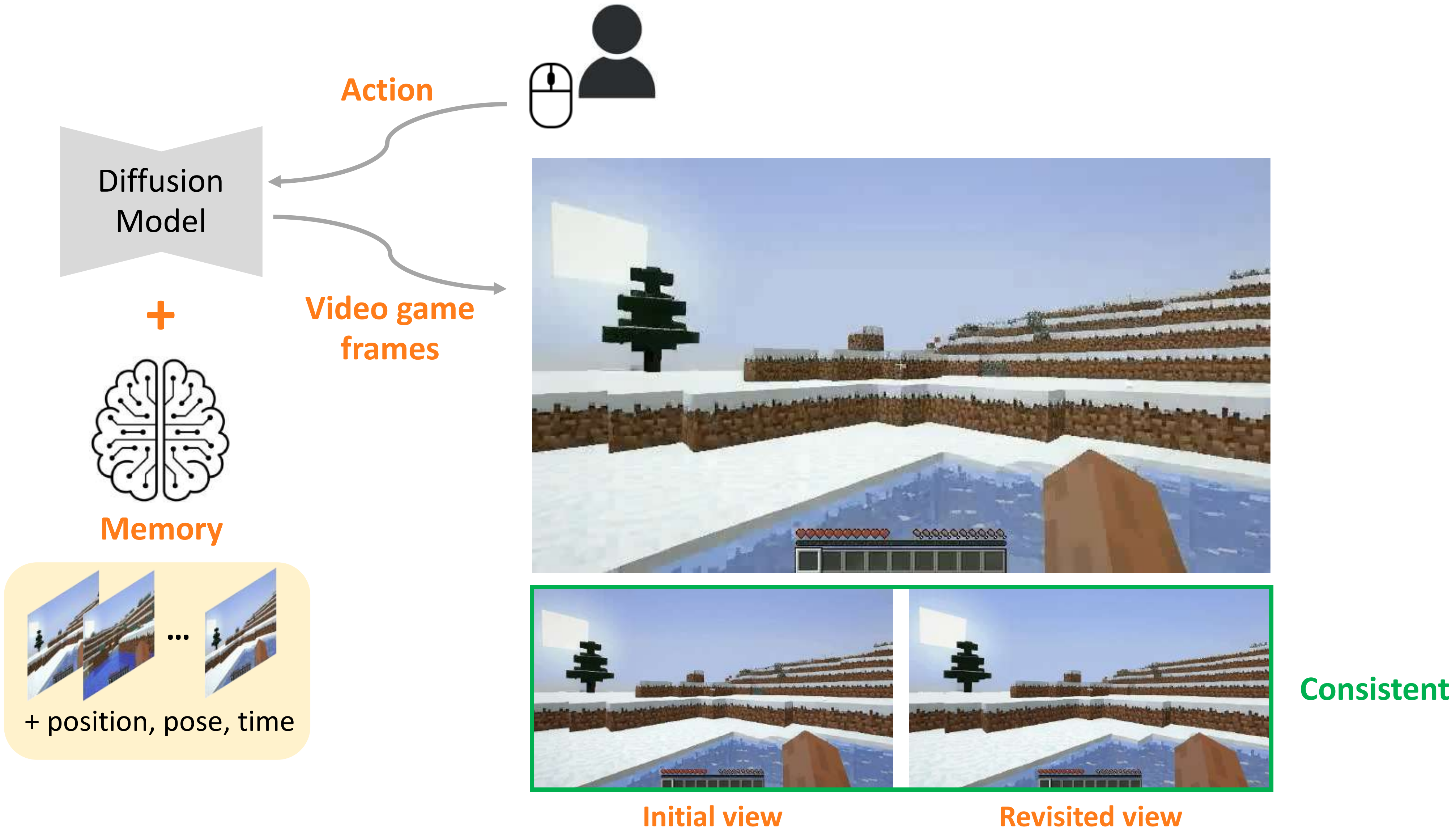
Xiao et al, To appear on arXiv, 2025

Generative AI as 3D Game Engine



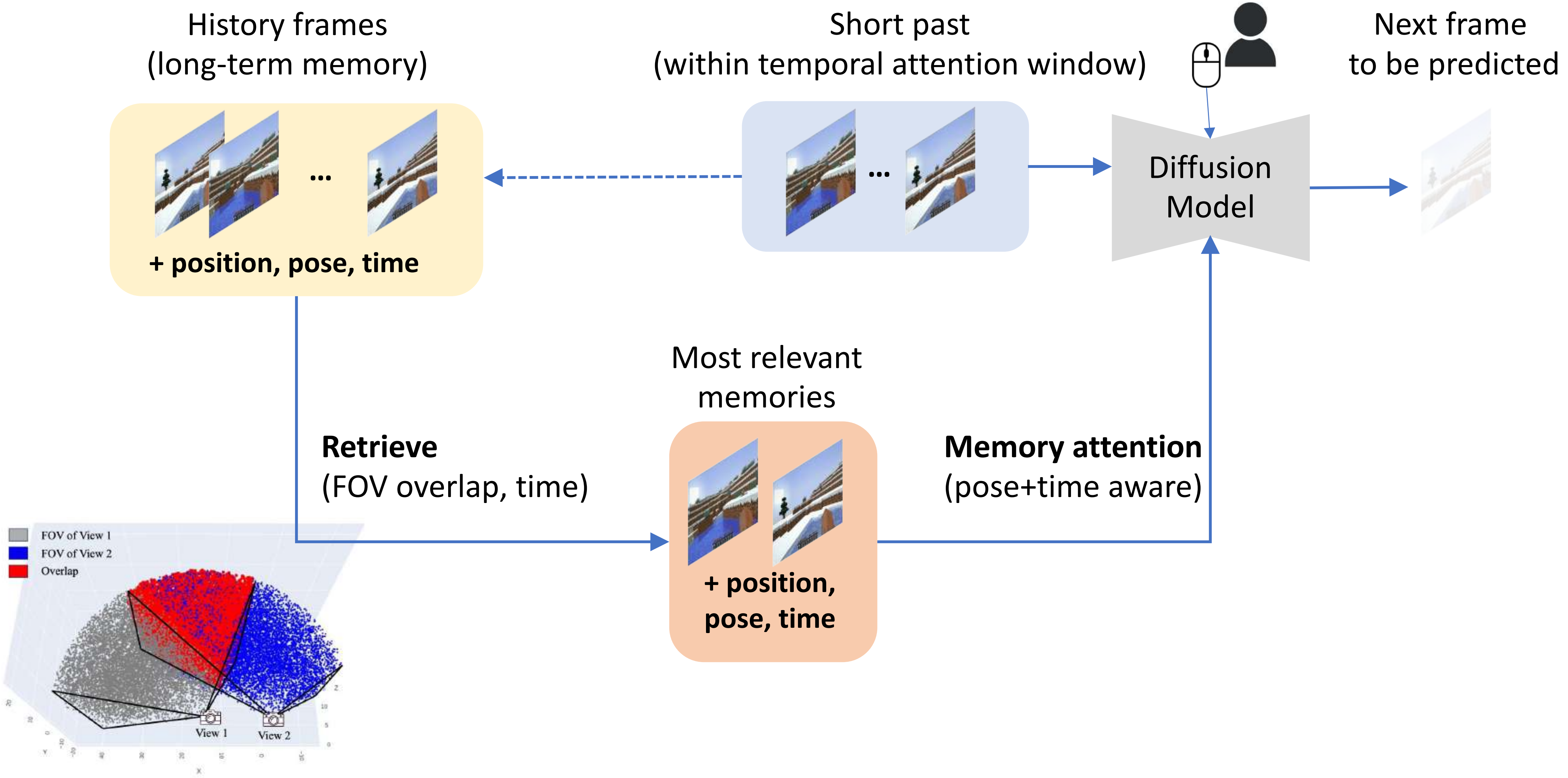
Oasis: A Universe in a Transformer https://oasis-model.github.io/

Generative AI as 3D Game Engine

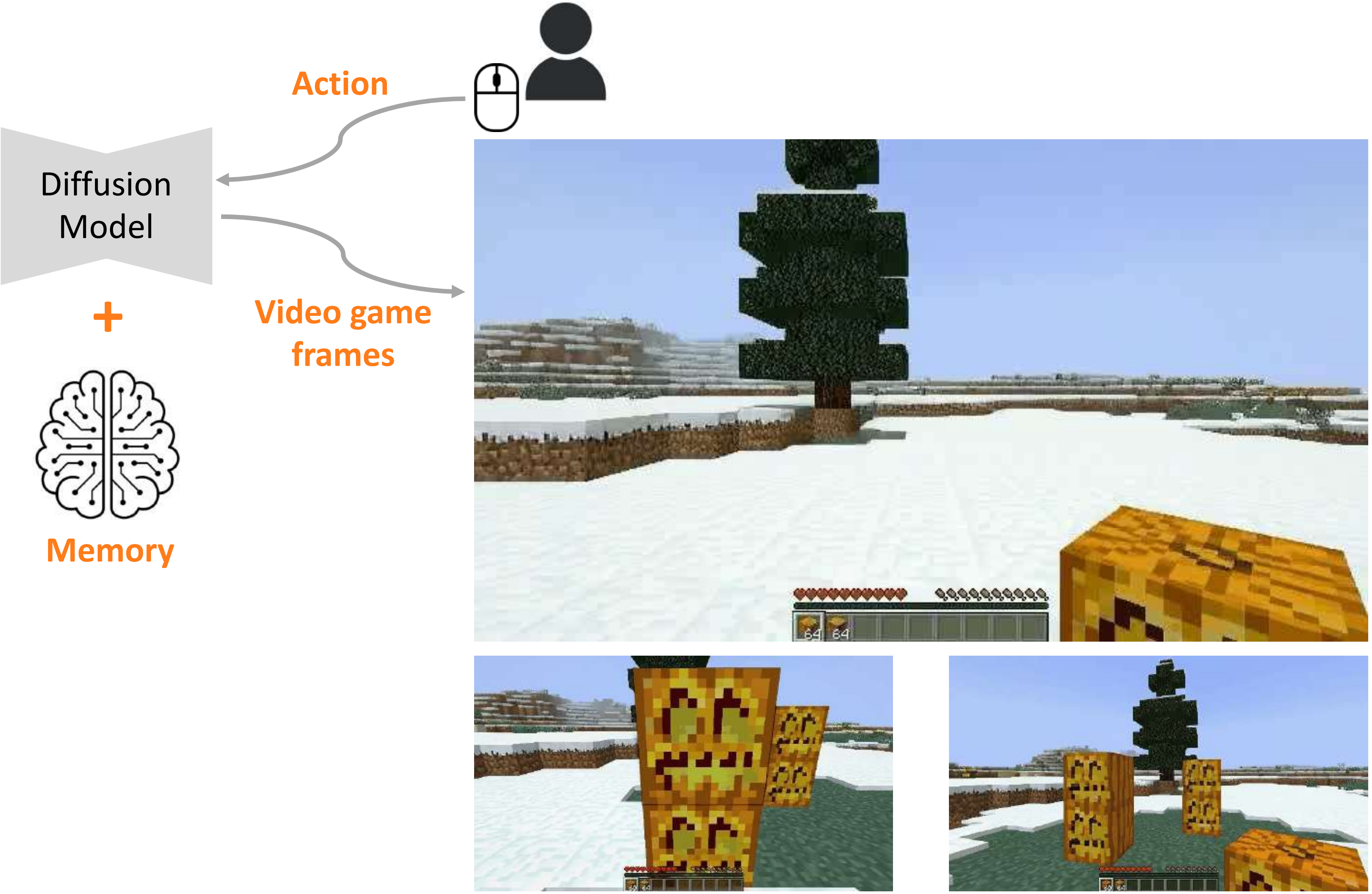


Xiao et al, To appear on arXiv, 2025

World Generation with Memory



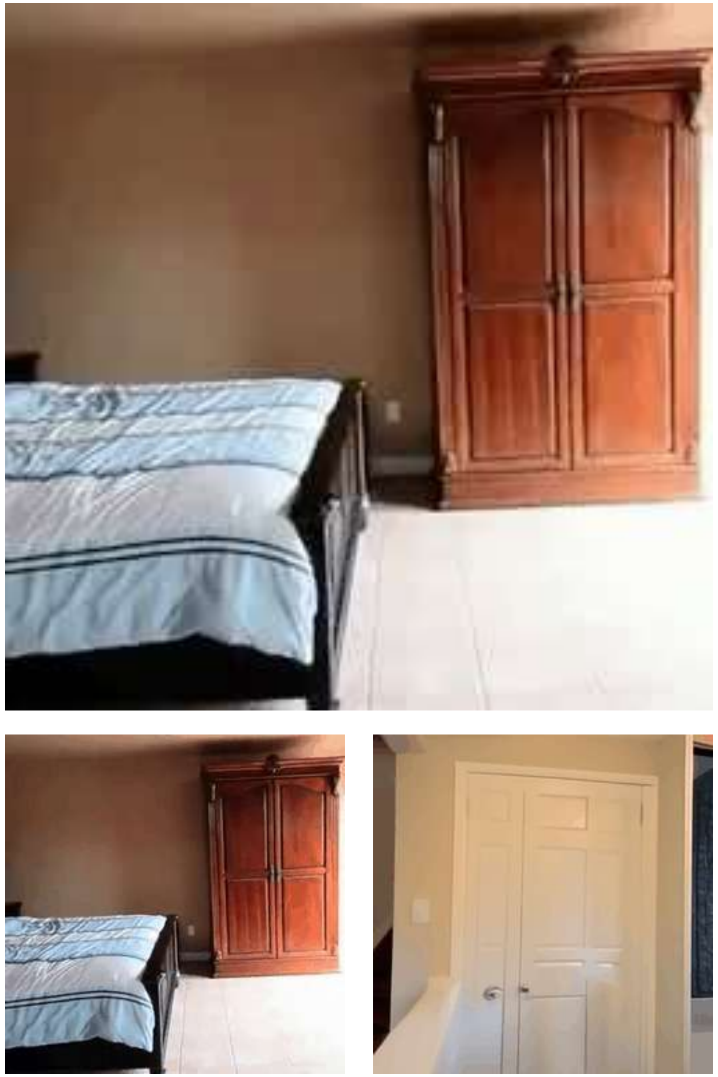
Generative AI as 3D Game Engine



Initial view

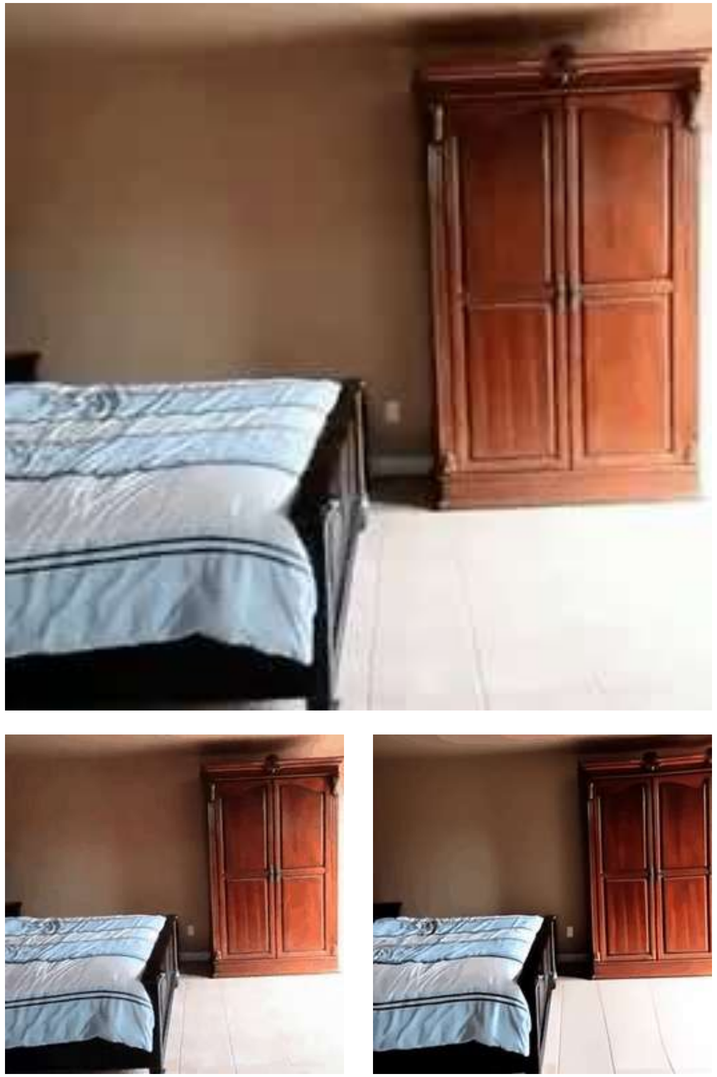
Revisited view

Real Scene Results



Initial view Revisited view

w/o Memory



Initial view Revisited view

w/ Memory

Real Scene Results

Initial view Revisited view

Initial view Revisited view

Conclusion

Fast 3D generation (<1s)

- A native 3D Diffusion Model, GaussianAnything: structured latent space, better design, better performance
- Autoregressive Model for 3D generation can perform as well as diffusion models while being more efficient
- When building 3D playable worlds via video diffusion models, **Memory** is important!

Open problems

- 3D object -> Rigging -> **Animation**
- 3D **scene** generation
- **CAD** generation, 3D object to CAD
- **Physics**-aware generation

Thank you!