[ECCV 2024]

LGM: Large Multi-View Gaussian Model for High-Resolution 3D Content Creation

Jiaxiang Tang[1*], Zhaoxi Chen[2], Xiaokang Chen[1], Tengfei Wang[3], Gang Zeng[1], and Ziwei Liu[2]

[1] National Key Lab of General AI, Peking University
[2] S-Lab, Nanyang Technological University
[3] Shanghai AI Lab

Presenter: Gyeongsu Cho

Fri Feb 21, 2025

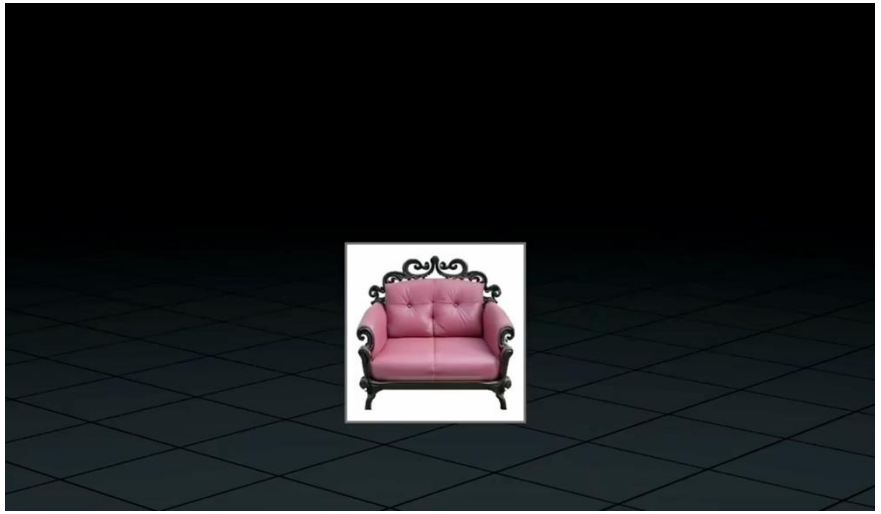
Contents

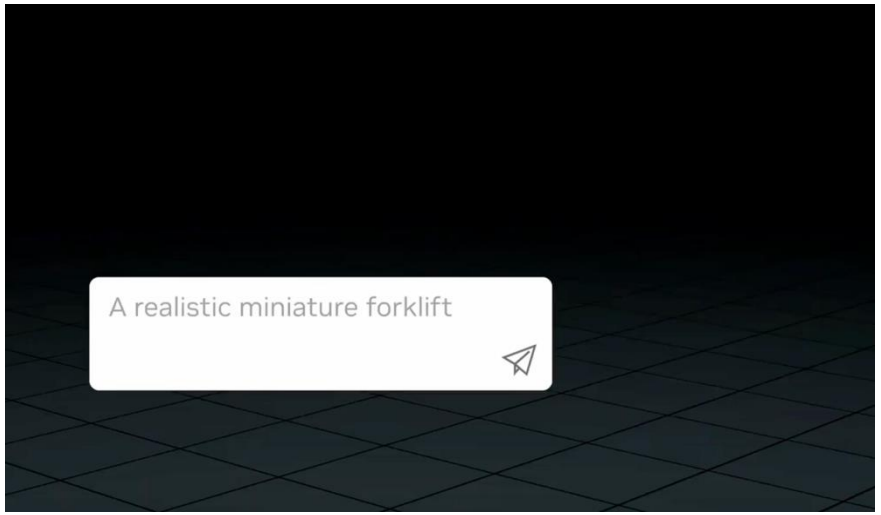
Introduction

LGM

LGM generates high-resolution 3D Gaussians in 5 seconds from single-view images or texts

Motivation

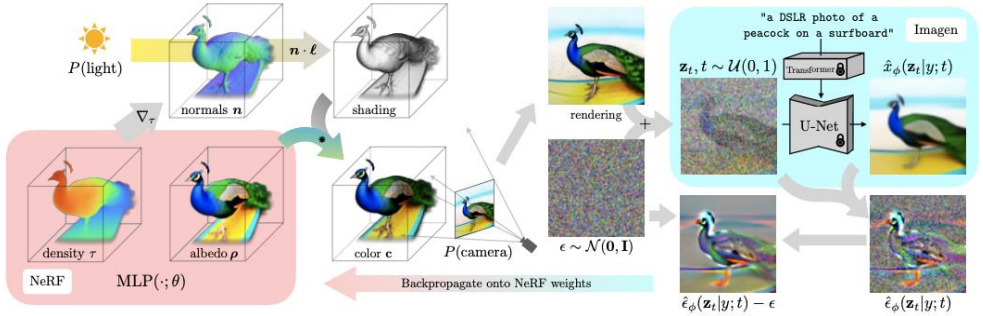




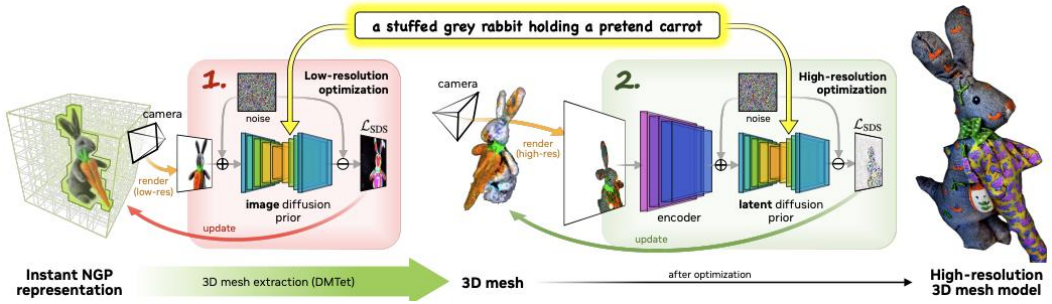
Manual 3D creation takes a lot of time

Problem: Slow, low resolution

Optimization based

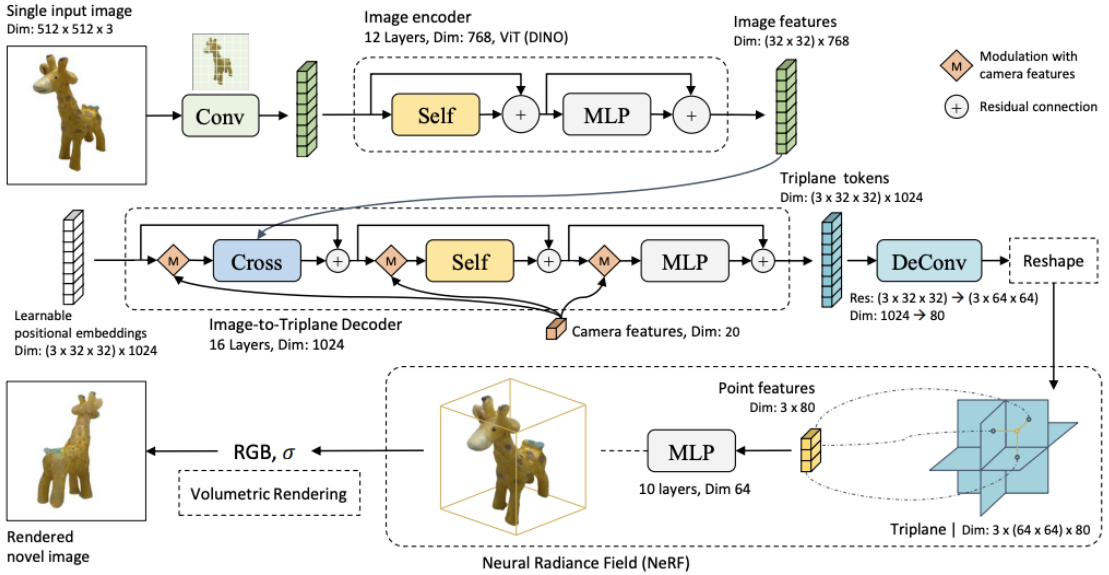


[ICLR 2023] DreamFusion



[CVPR 2023] Magic3d

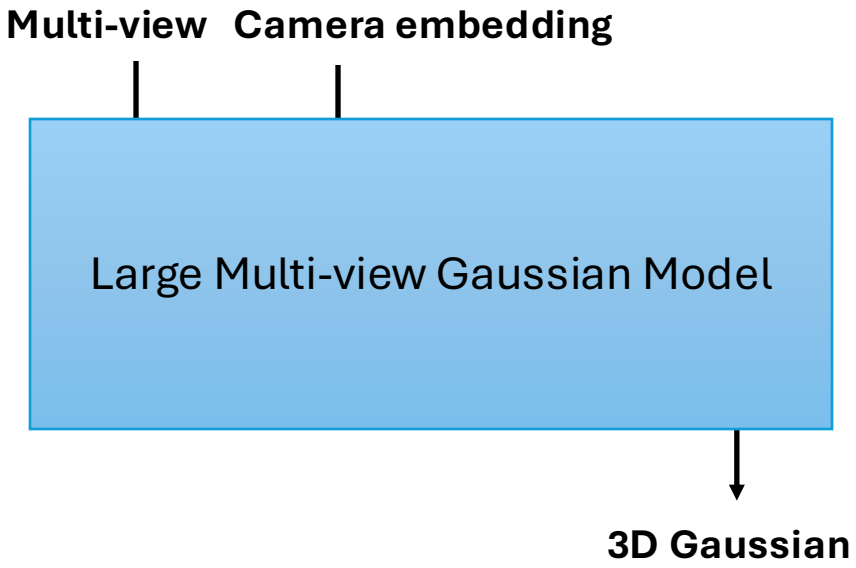
Feed-forward based

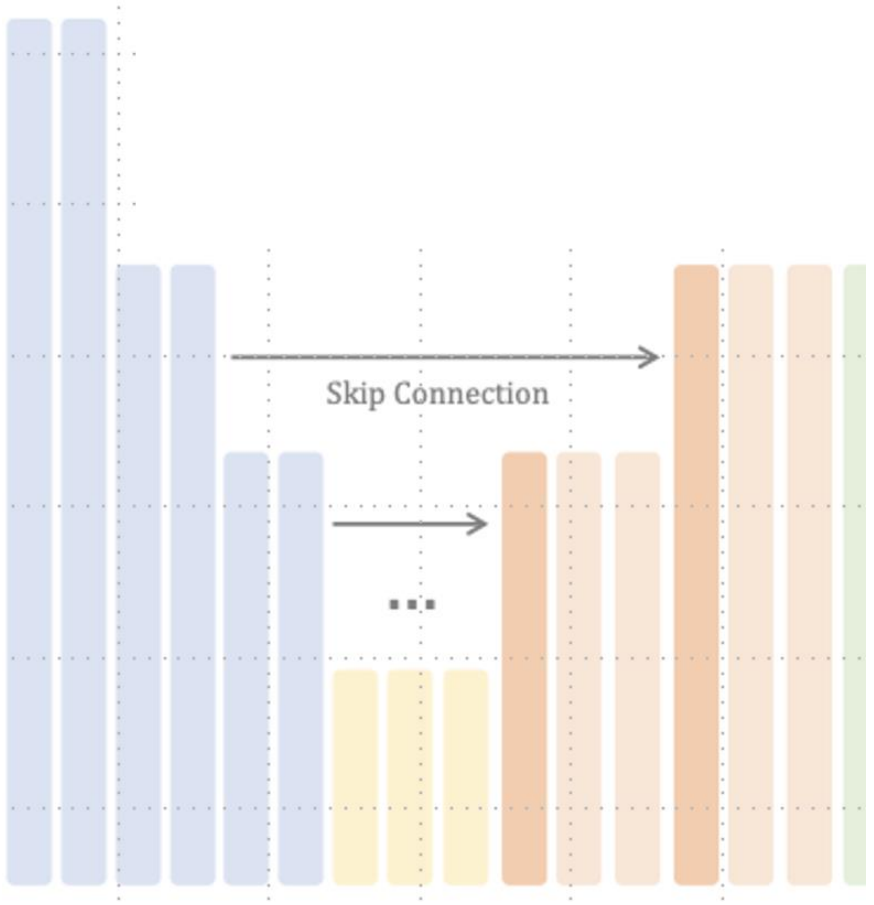


[ICLR 2024] LRM

Existing methods are slow and predict low resolution 3D output

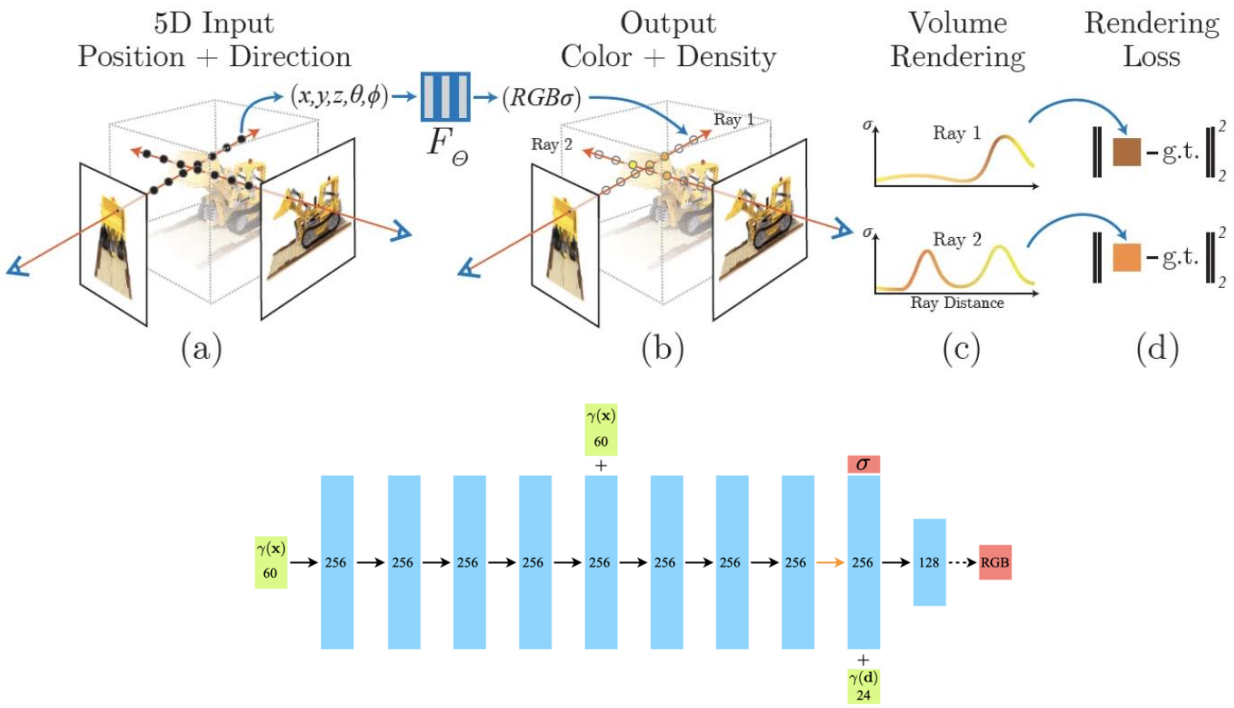
Key Idea



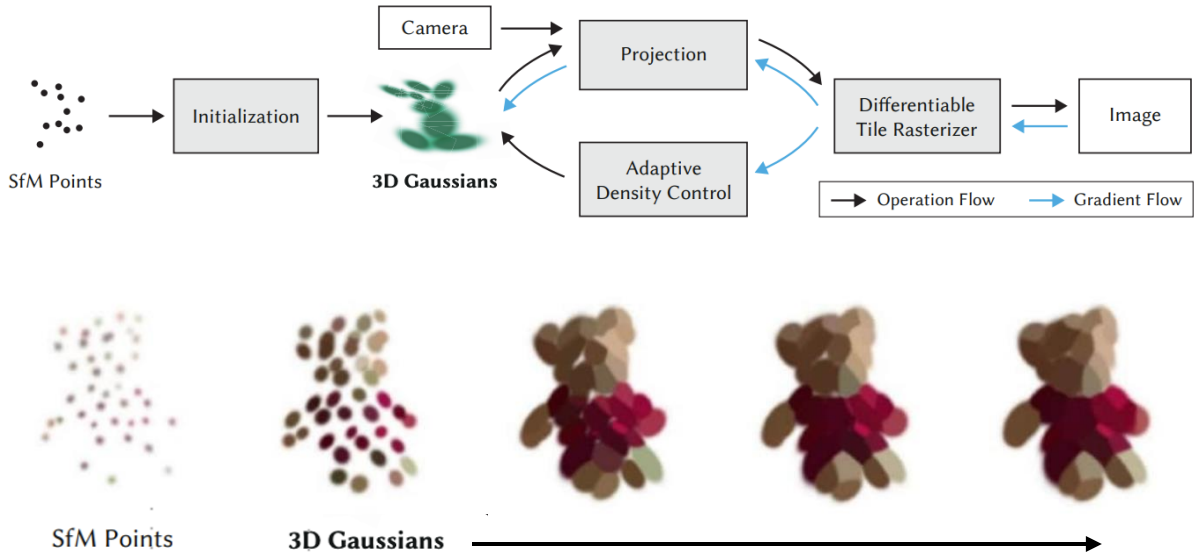


1. Novel use of large multi-view Gaussian features to represent 3D scenes.
2. An asymmetric U-Net architecture that fuses multi-view information efficiently.

Preliminaries: 3D Gaussian Splatting



[ECCV 2020] NeRF

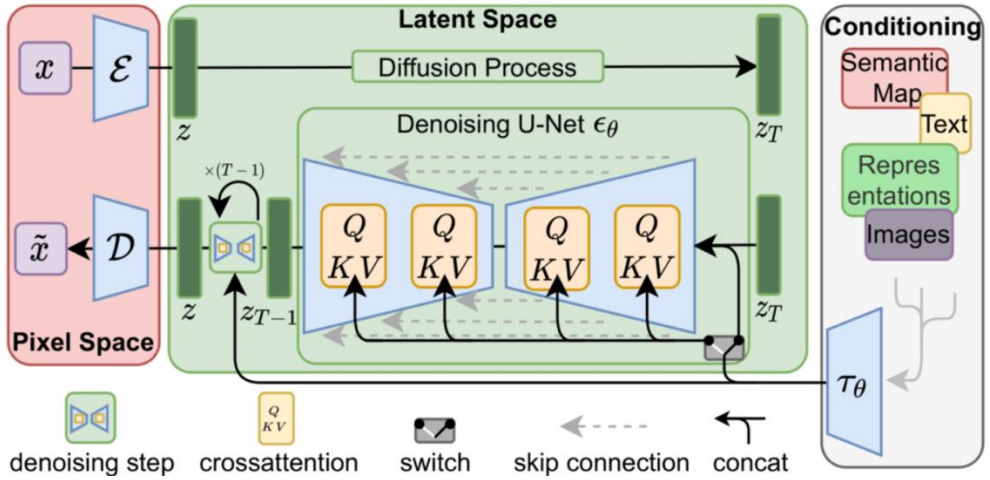


$$\Theta_i = \{\mathbf{x}_i, \mathbf{s}_i, \mathbf{q}_i, \alpha_i, \mathbf{c}_i\}$$

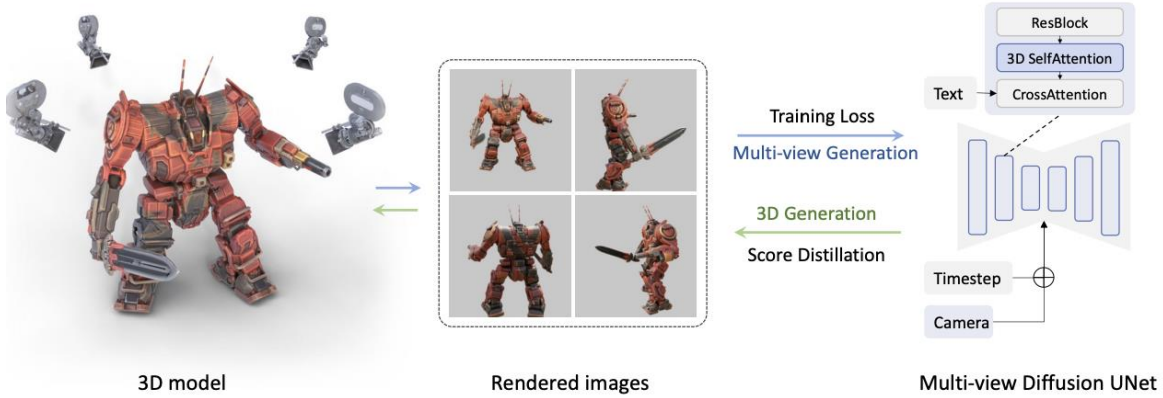
[SIGGRAPH 2023] 3D Gaussian Splatting

3D Gaussian splatting employs a collection of 3D Gaussians to represent 3D data

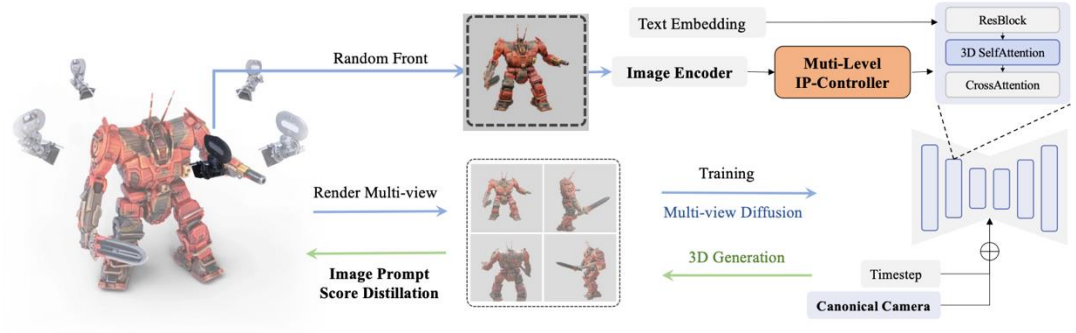
Preliminaries: Multi-view Diffusion Model



[CVPR 2022] Latent Diffusion



[ICLR 2024] MVDream

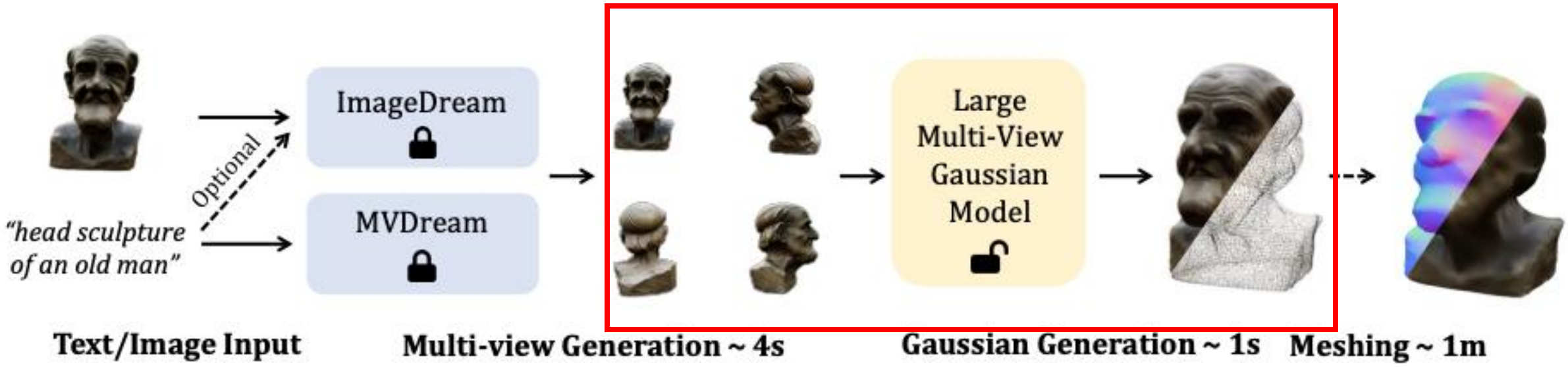


[arXiv 2024] ImageDream

Multi-view diffusion models can generate multi view images by training on 3D datasets.

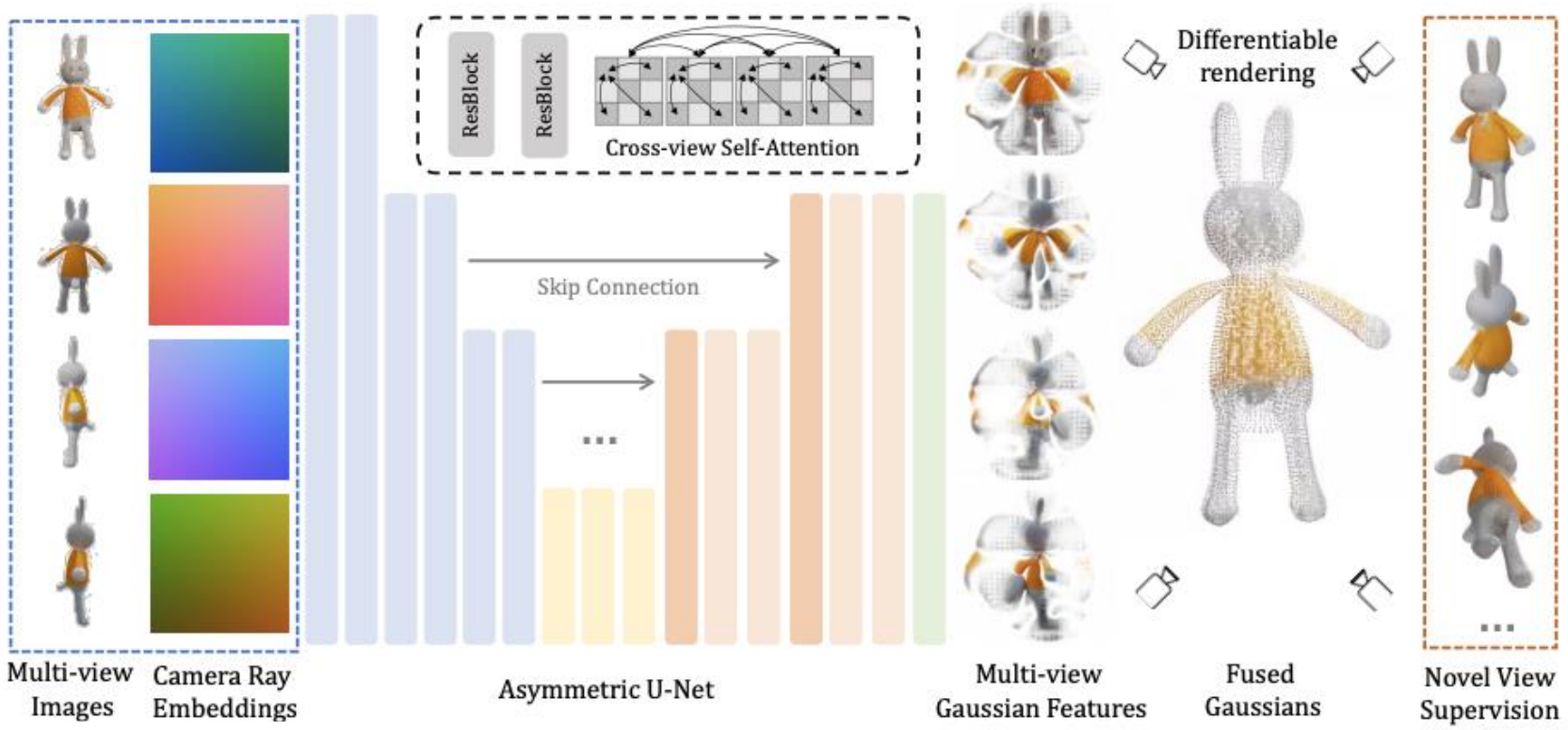
Method

Pipeline of LGM



1. Generate multi-view images using diffusion models (e.g., MVDream, ImageDream)
2. Fuse these images with an asymmetric U-Net to predict and combine 3D Gaussians

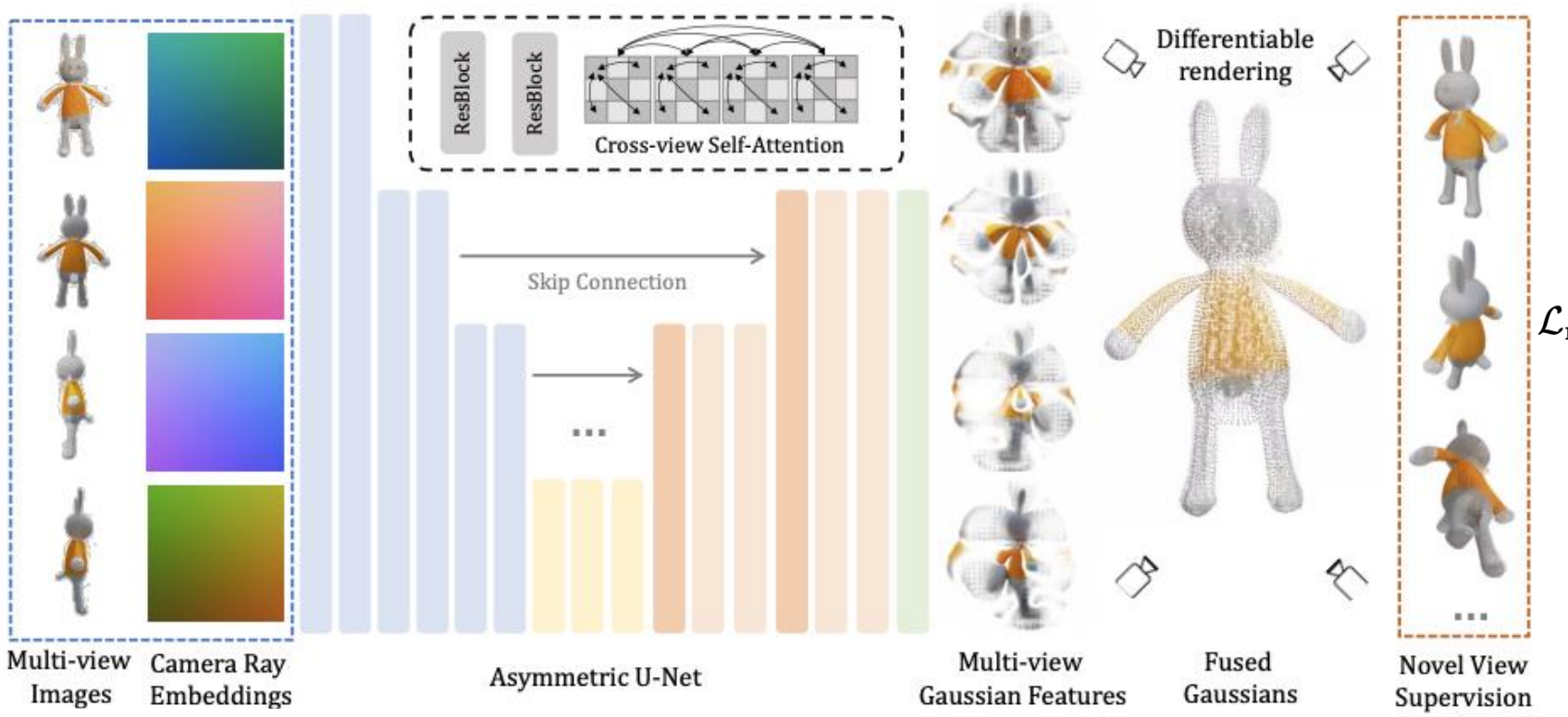
Architecture of LGM



Asymmetric U-Net: multiple down-sampling and up-sampling blocks, but with a twist

the output resolution is intentionally lower than the input

Architecture of LGM



$$\mathbf{f}_i = \{\mathbf{c}_i, \mathbf{o}_i \times \mathbf{d}_i, \mathbf{d}_i\}$$

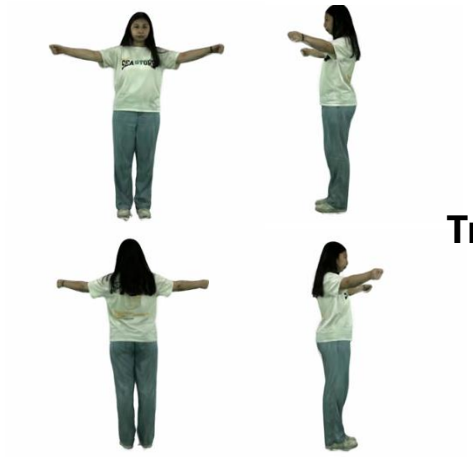
$$\mathcal{L}_{\text{rgb}} = \mathcal{L}_{\text{MSE}}(I_{\text{rgb}}, I_{\text{rgb}}^{\text{GT}}) + \lambda \mathcal{L}_{\text{LPIPS}}(I_{\text{rgb}}, I_{\text{rgb}}^{\text{GT}})$$

$$\mathcal{L}_{\alpha} = \mathcal{L}_{\text{MSE}}(I_{\alpha}, I_{\alpha}^{\text{GT}})$$

Asymmetric U-Net: multiple down-sampling and up-sampling blocks, but with a twist

the output resolution is intentionally lower than the input

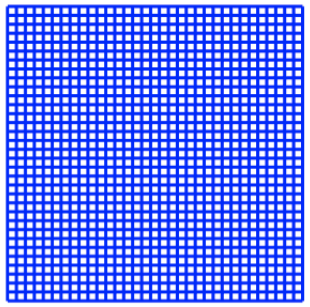
Training Strategies & Data Augmentation

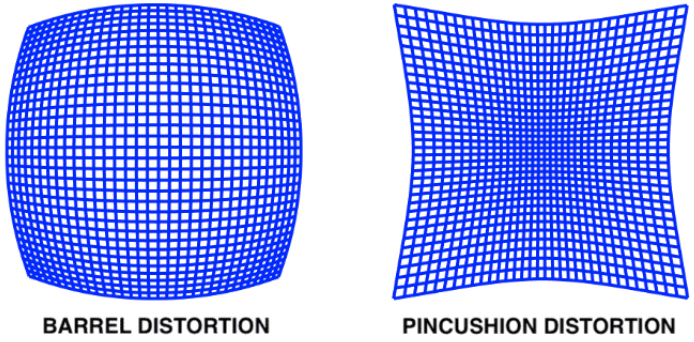


Training MV

Inference MV

Domain gap between training and inference





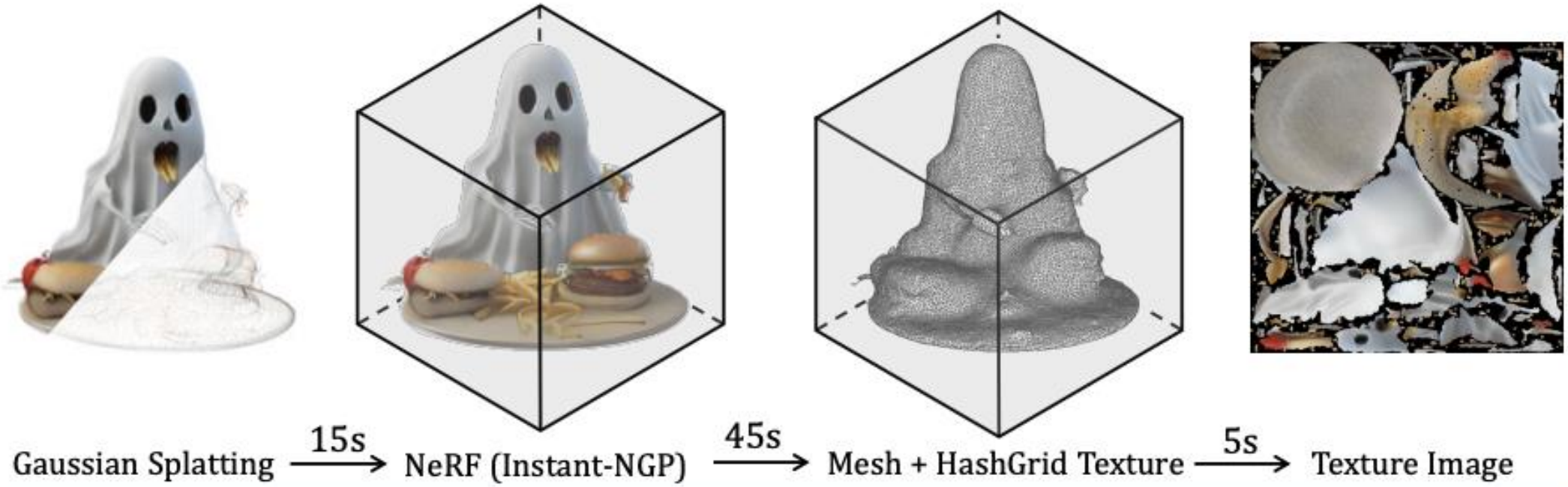
Grid distortion

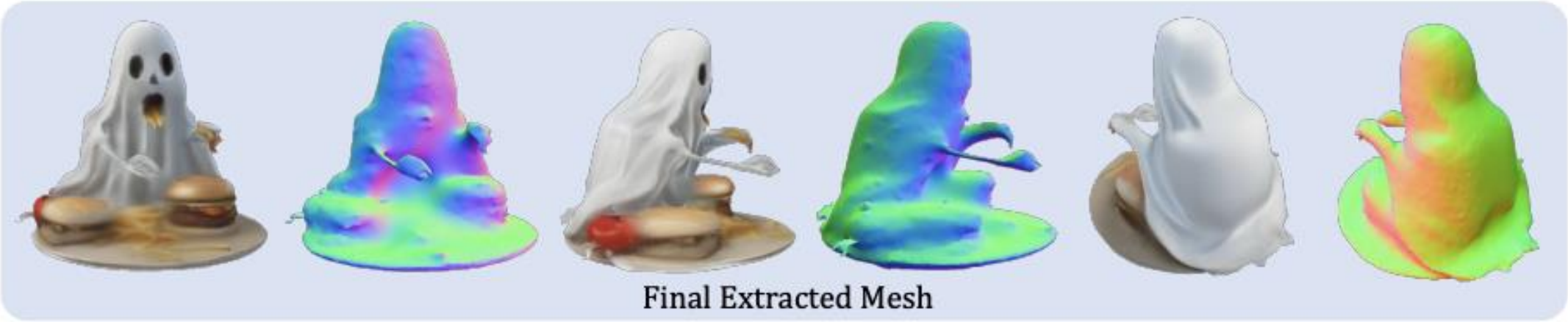
Orbital camera jitter

Grid distortion: Randomly distort views (except the reference)

Orbital camera jitter: Introduce small perturbations in camera poses

Mesh Extraction Pipeline of LGM





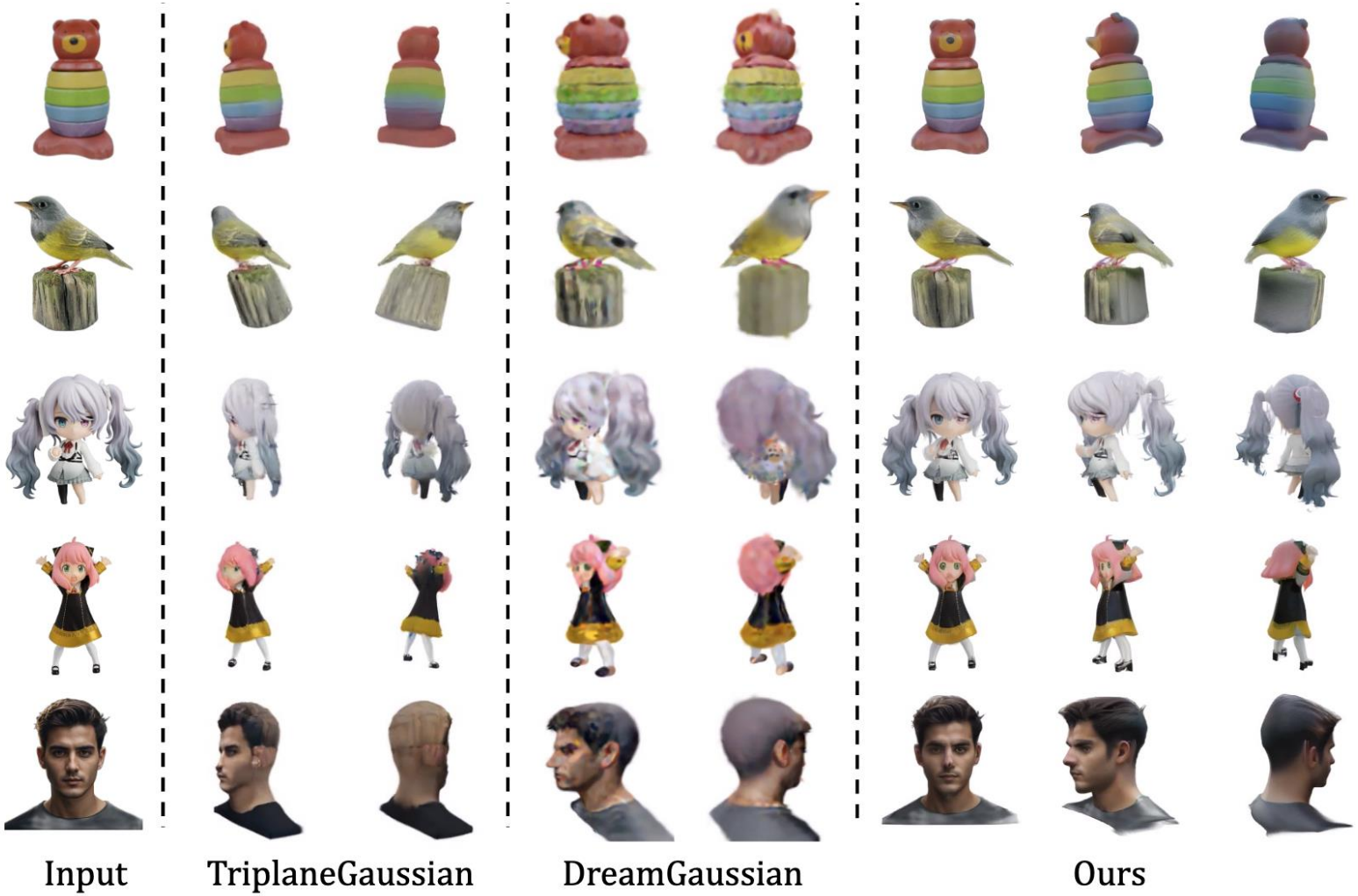
To ready-to-use, we should extract mesh from 3D Gaussian.

Experiments

Experiments

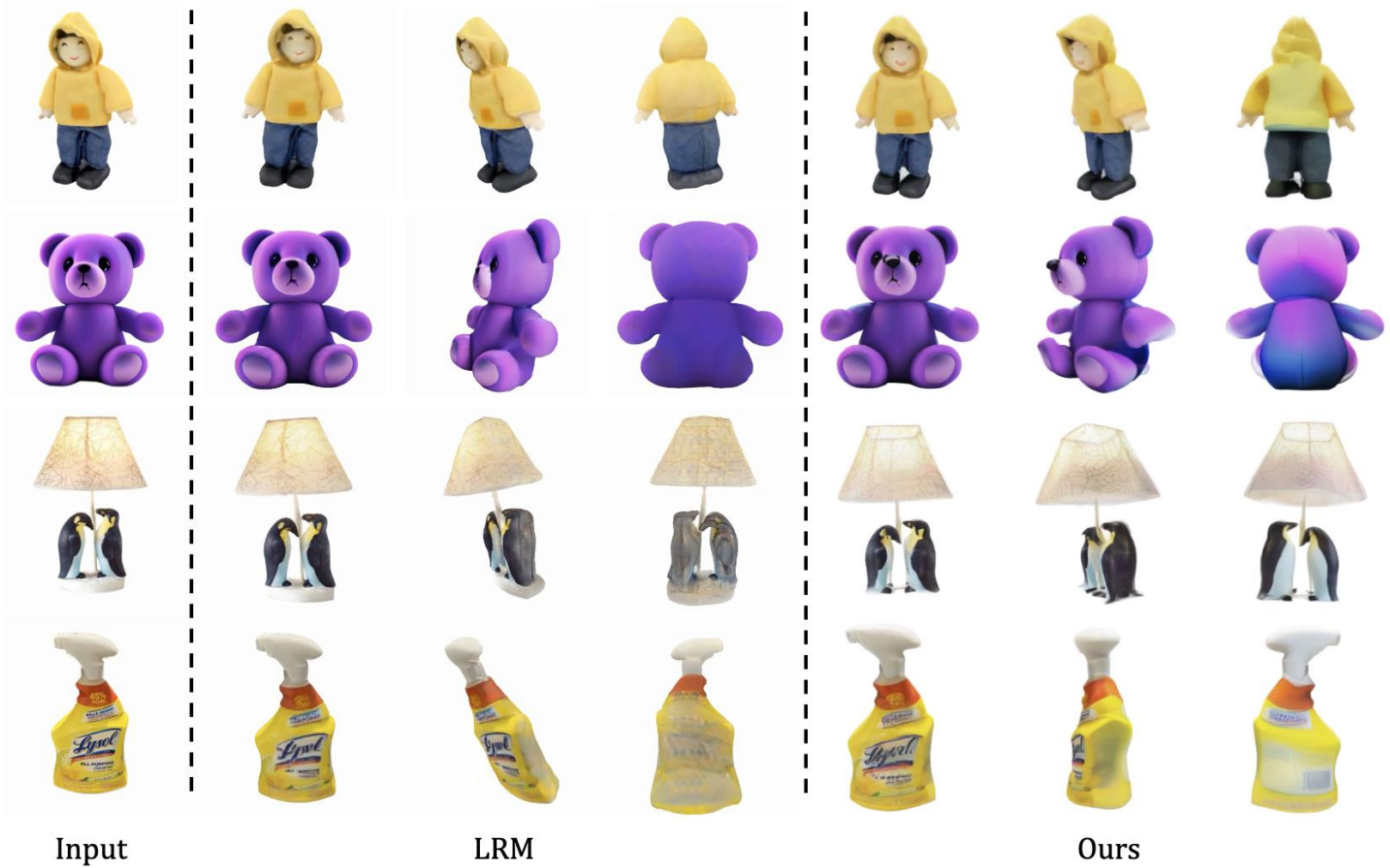
- Dataset
  - A filtered subset of the Objaverse dataset (80K)
- Baseline
  - Optimization-based: DreamGaussian
  - Feed-forward-based: TriplainGaussian, LRM
- Evaluation metrics
  - Human preference evaluation

Experiments



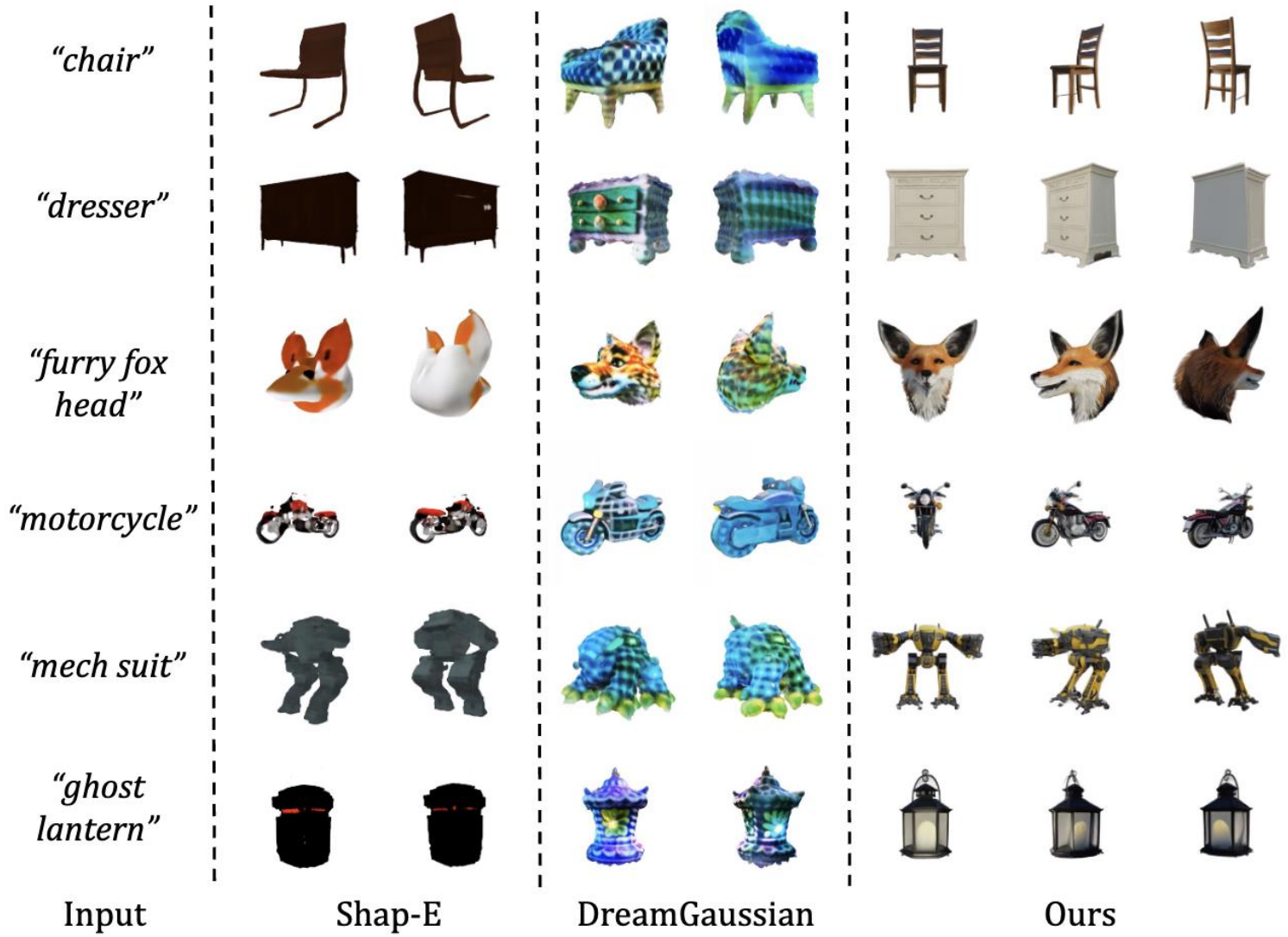
Qualitative comparisons of LGM

Experiments



Qualitative comparisons of LGM

Experiments



Qualitative comparisons of LGM for text-to-3D

Experiments

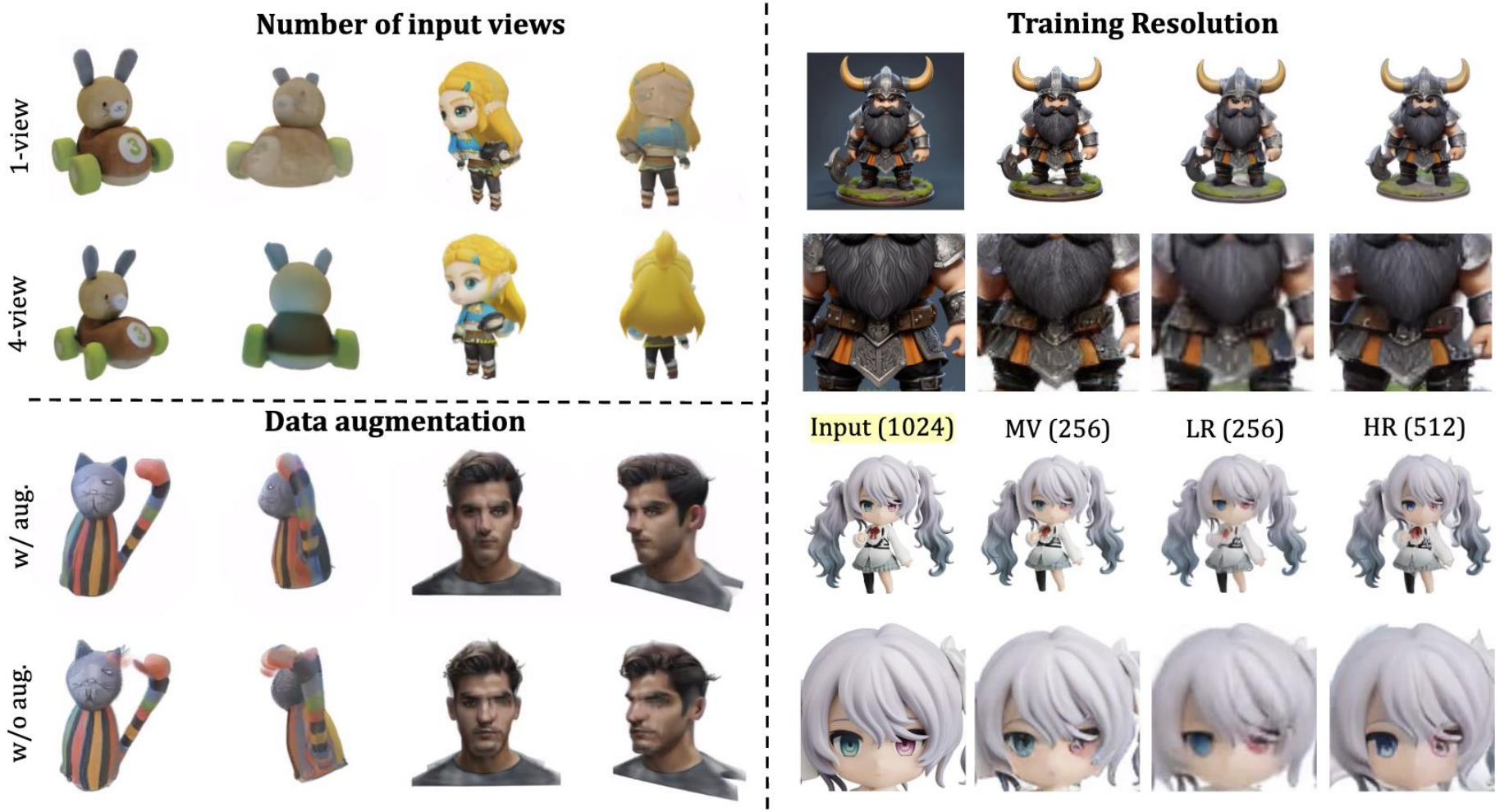
Diversity results of LGM

Experiments

	Image Consistency ↑	Overall Quality ↑
DreamGaussian [47]	2.30	1.98
TriplaneGaussian [62]	3.02	2.67
LGM (Ours)	**4.18**	**3.95**

User Study of LGM

Experiments



Ablation study of LGM

Experiments

Limitations of LGM

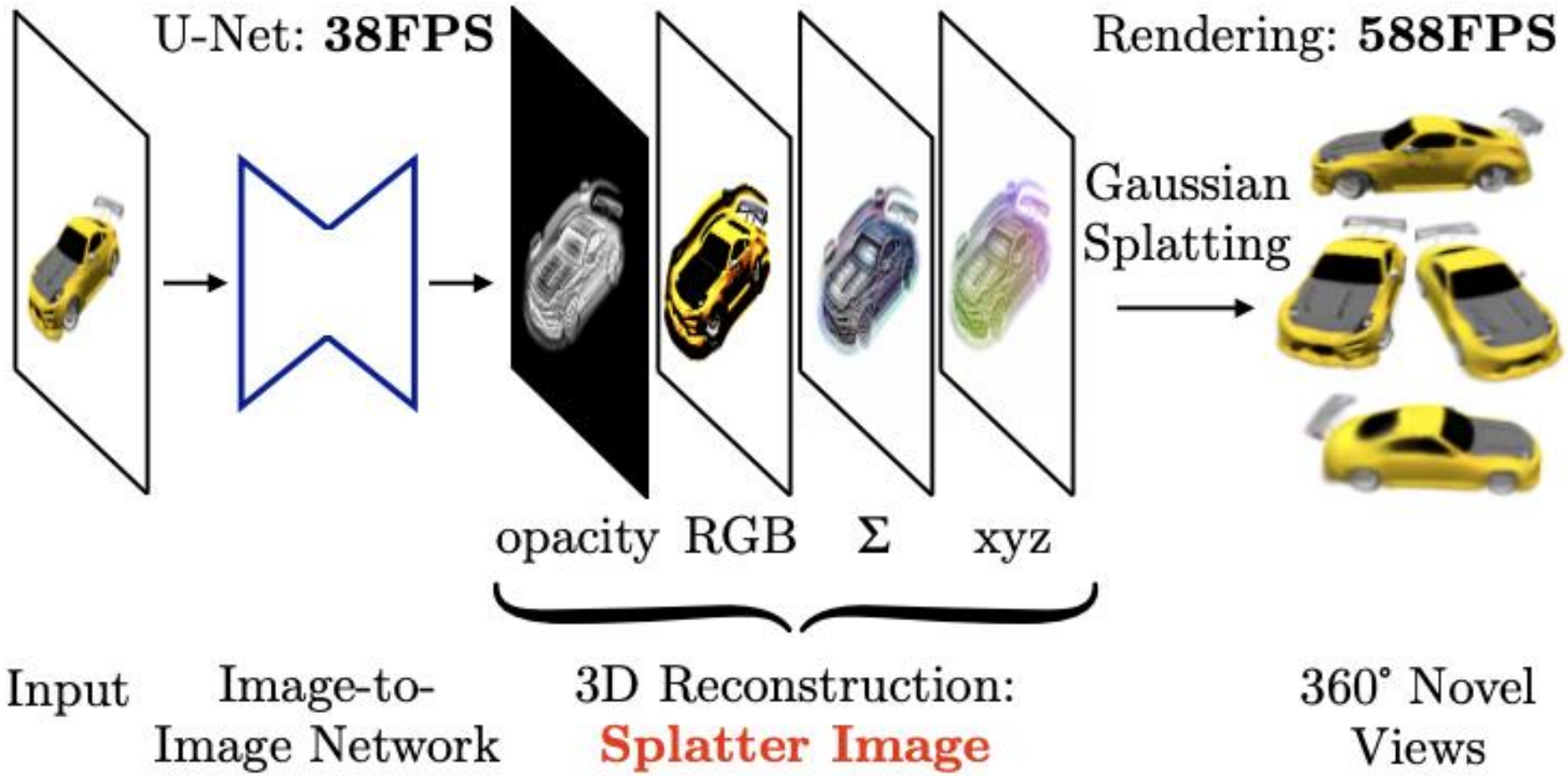
Conclusion

Conclusion

LGM generates high-resolution 3D Gaussians in 5 seconds from single-view images or texts

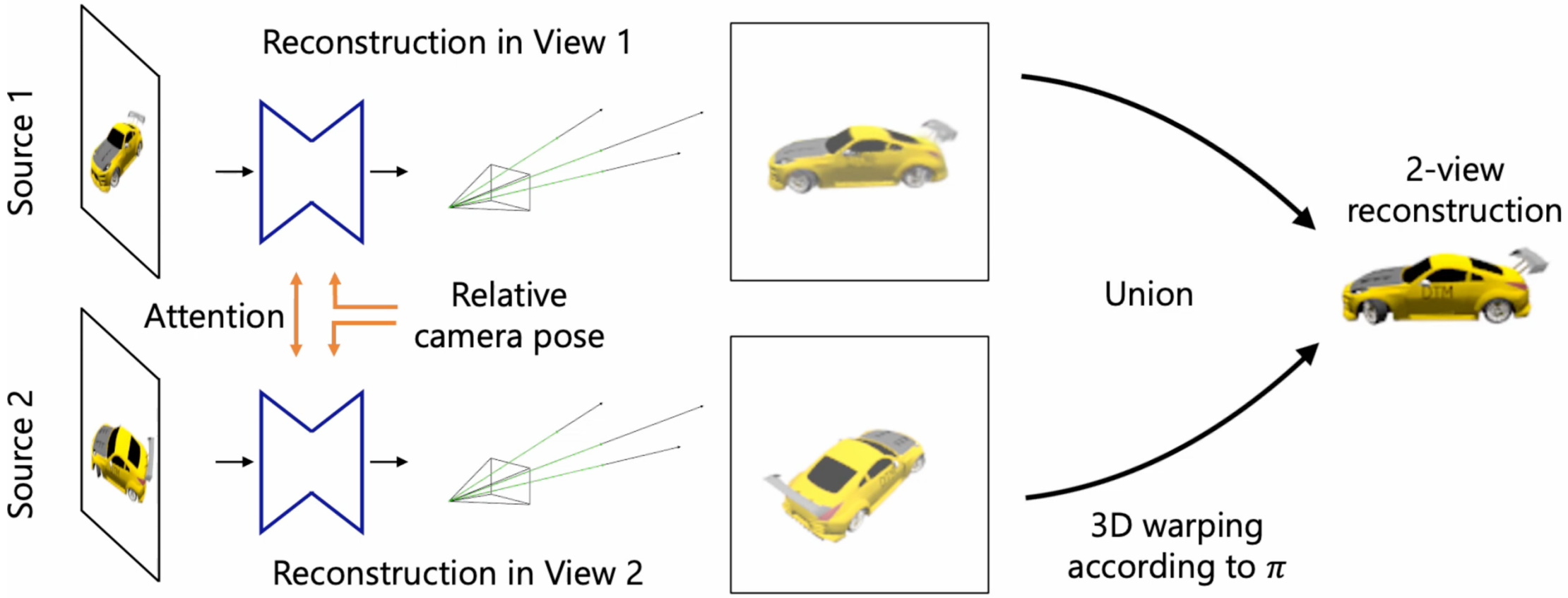
Appendix

Architecture of Splatter Image



Splatter Image predicts one 3D Gaussian per pixel from a single view

Architecture of Splatter Image



Splatter Image predicts one 3D Gaussian per pixel from a single view