

# RGBD<sup>2</sup>: Generative Scene Synthesis via Incremental View Inpainting using RGBD Diffusion Models

CVPR 2023

Jiabao Lei<sup>1</sup>, Jiapeng Tang<sup>2</sup>, Kui Jia<sup>1,3</sup>

<sup>1</sup>South China University of Technology

<sup>2</sup>Technical University of Munich, <sup>3</sup>Peng Cheng Laboratory



# Task

## Scene Reconstruction

Input

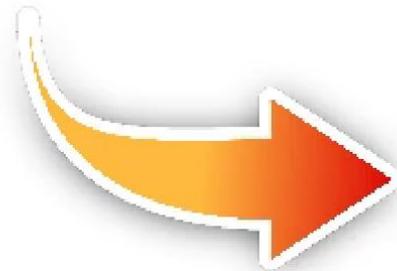
A Sparse Set of  
Posed **RGBD**  
Images



Output

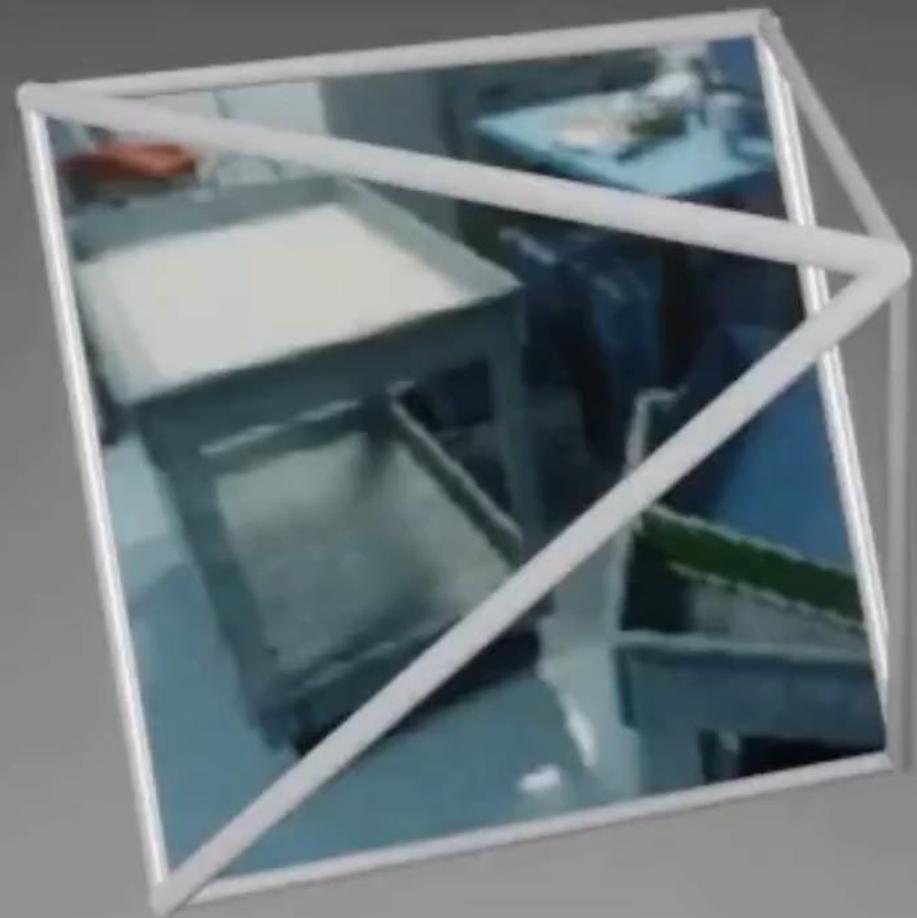
Colored Mesh

sparse views



newly generated  
(inpainted)



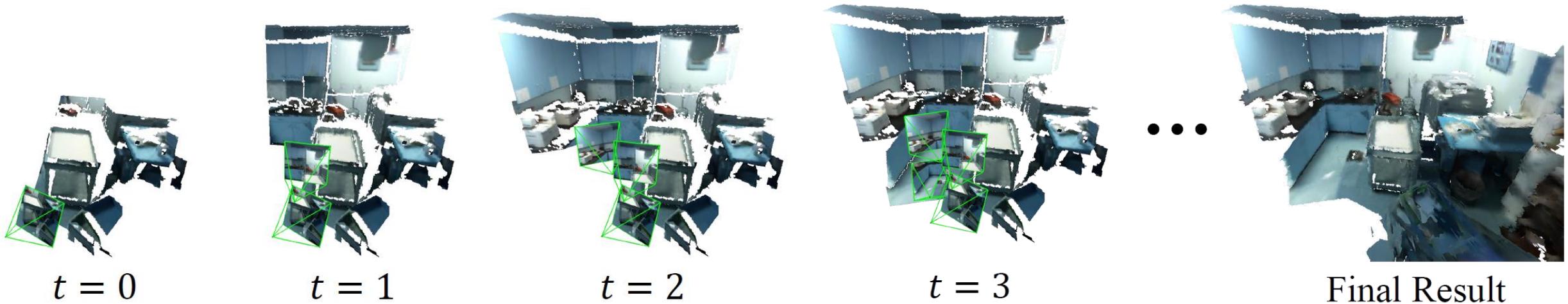


# Implementation Details



# Basic Idea

Generate what the camera sees when it moves in the scene

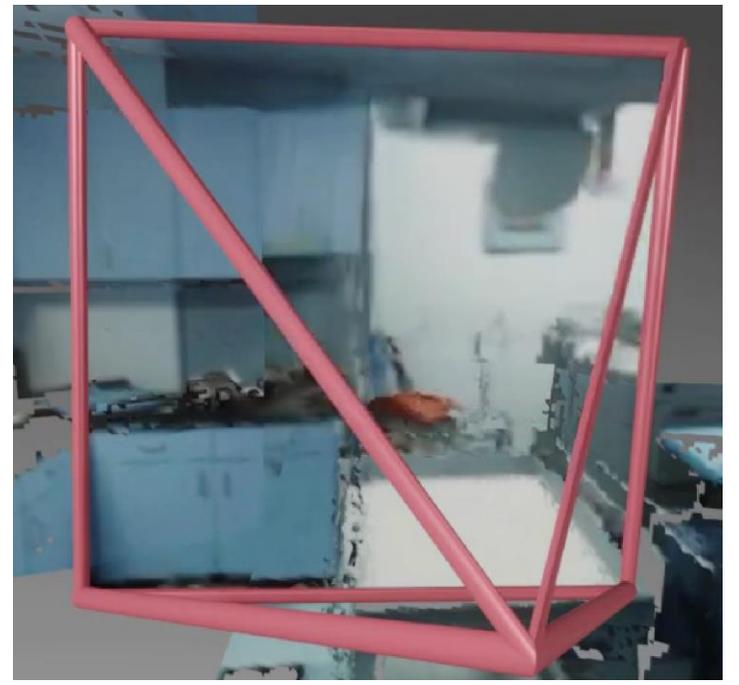
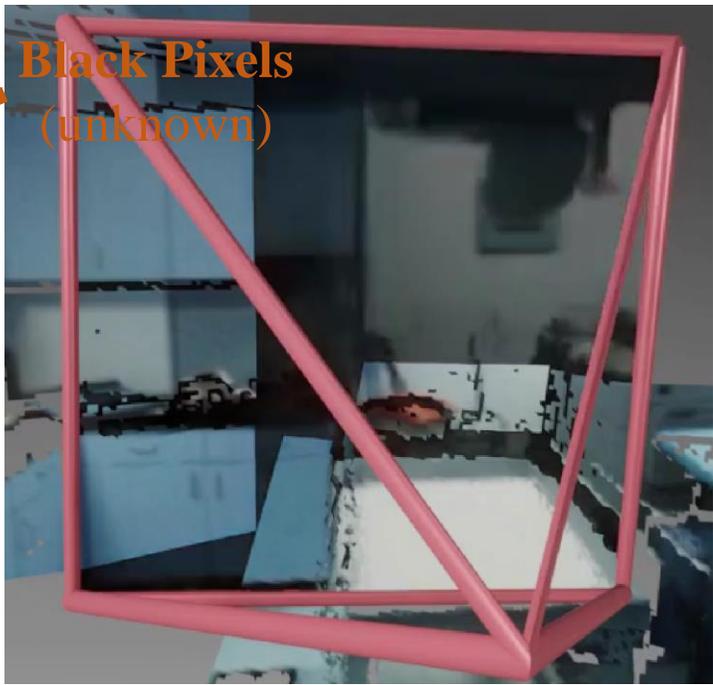
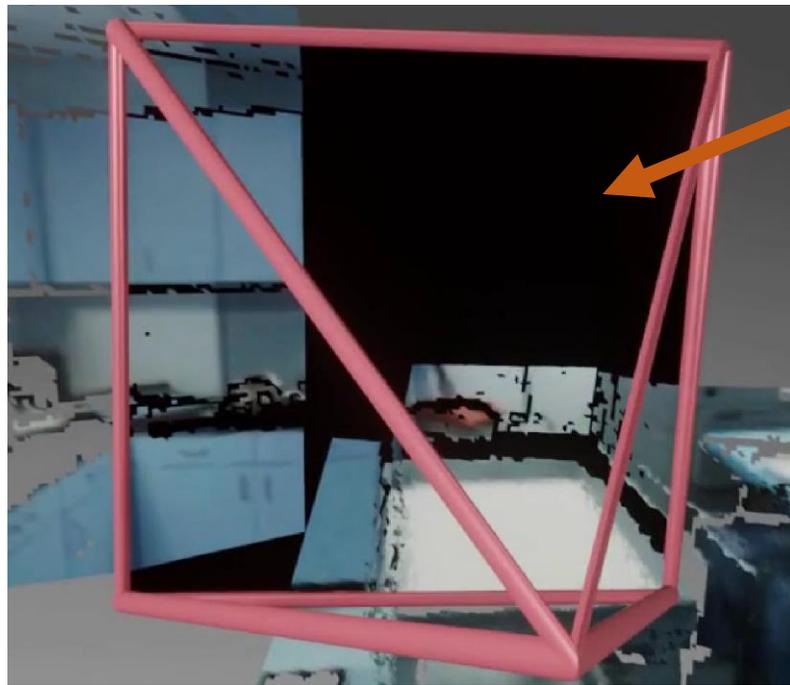


as camera **moves**...



# Implementation Details – Progressive RGBD Inpainting

Use a versatile **RGBD Diffusion** (**RGBD<sup>2</sup>**) Model



reverse diffusion sampling...

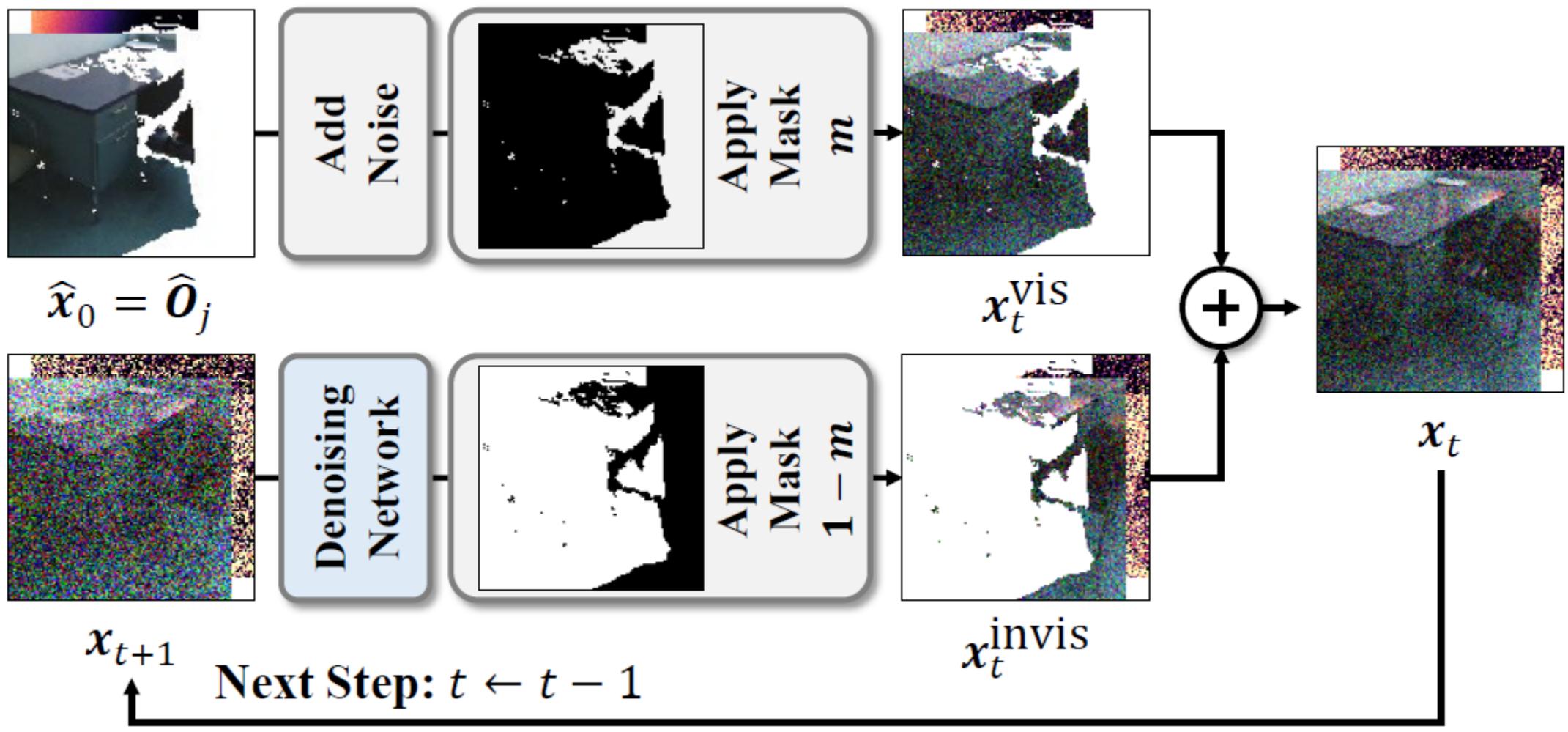
$s = 0$

$s = 50$



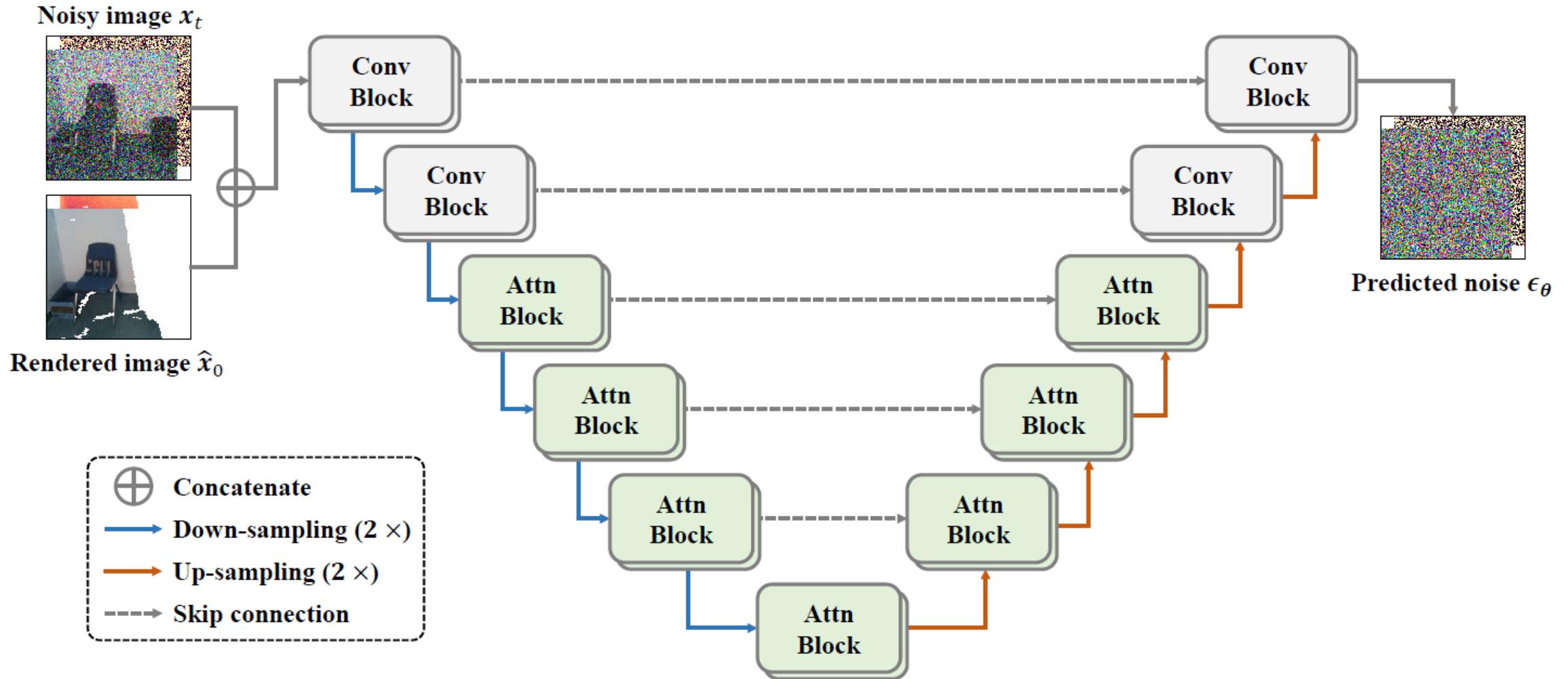
# Implementation Details – Masked Inpainting

## Masked Inpainting on RGBD Images



# Implementation Details – Model

## Network Architecture (U-Net)

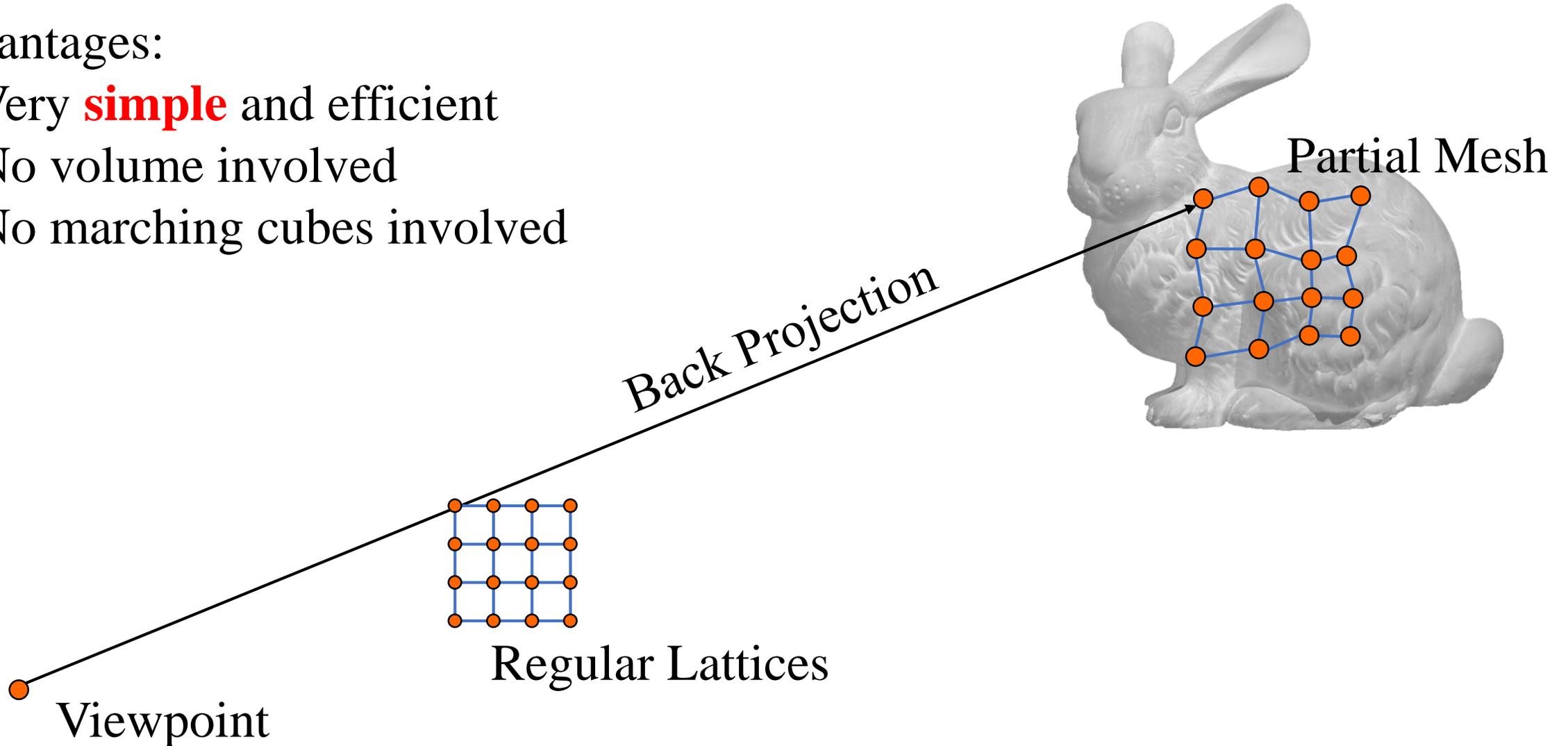


# Implementation Details – Back Projection

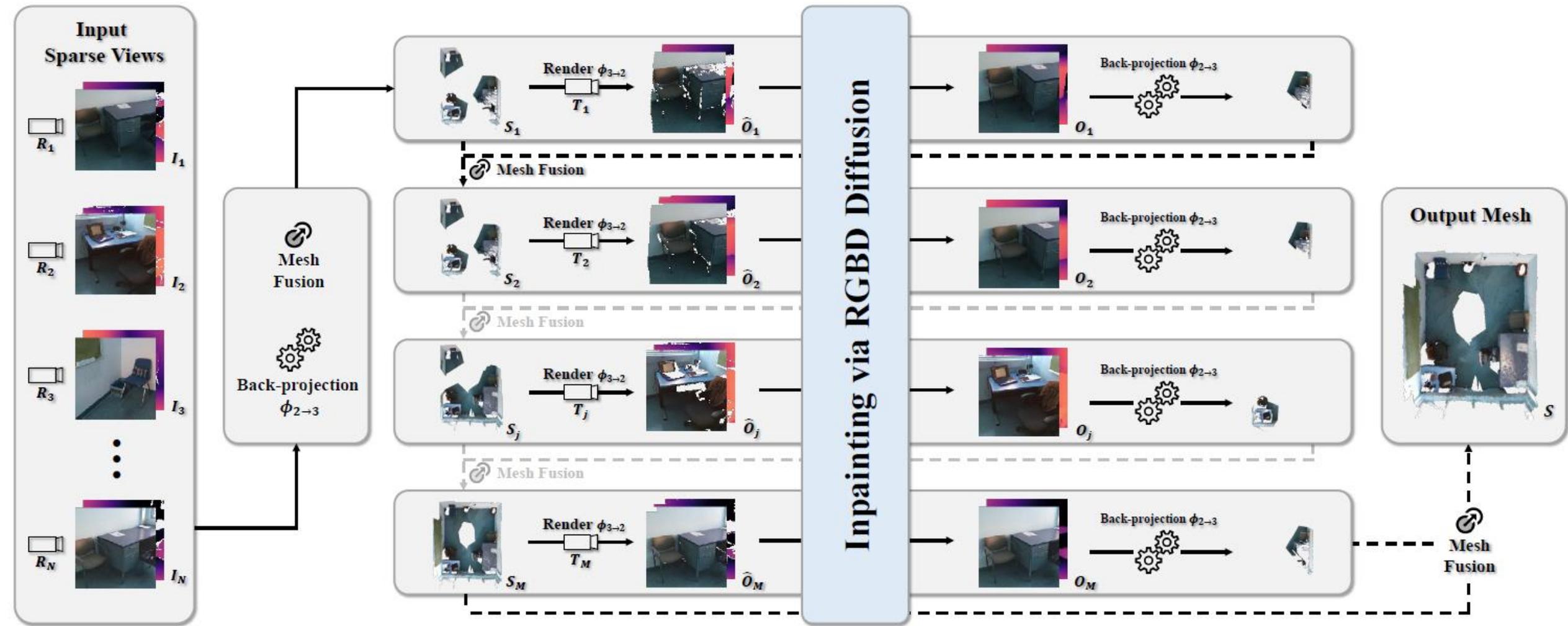
Directly convert an RGBD image into a partial surface mesh

Advantages:

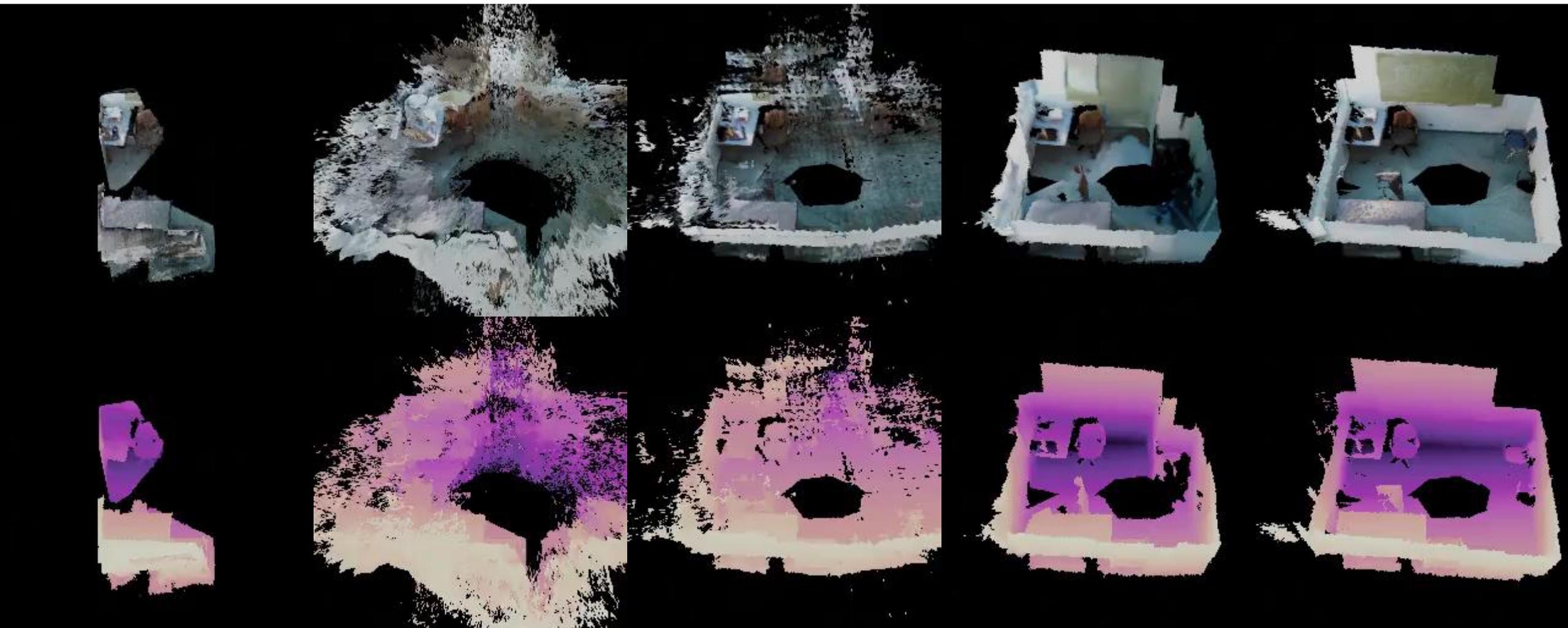
- ✓ Very **simple** and efficient
- ✓ No volume involved
- ✓ No marching cubes involved



# Pipeline Overview



# Experiment – Sparsity 5% (only **3 input views**)



DSNGP

DDP

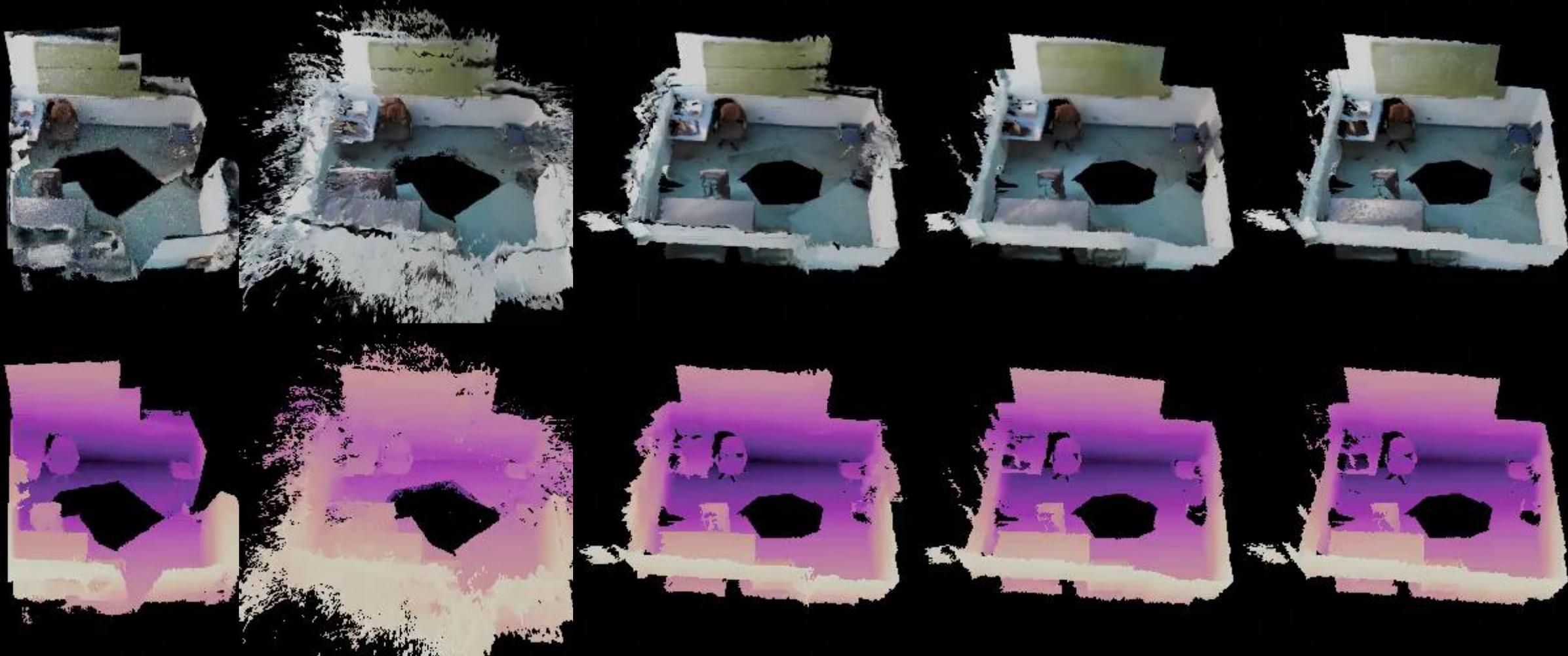
NRGBD

**OUR**

GT



# Experiment – Sparsity 20% (only **11 input views**)



DSNGP

DDP

NRGBD

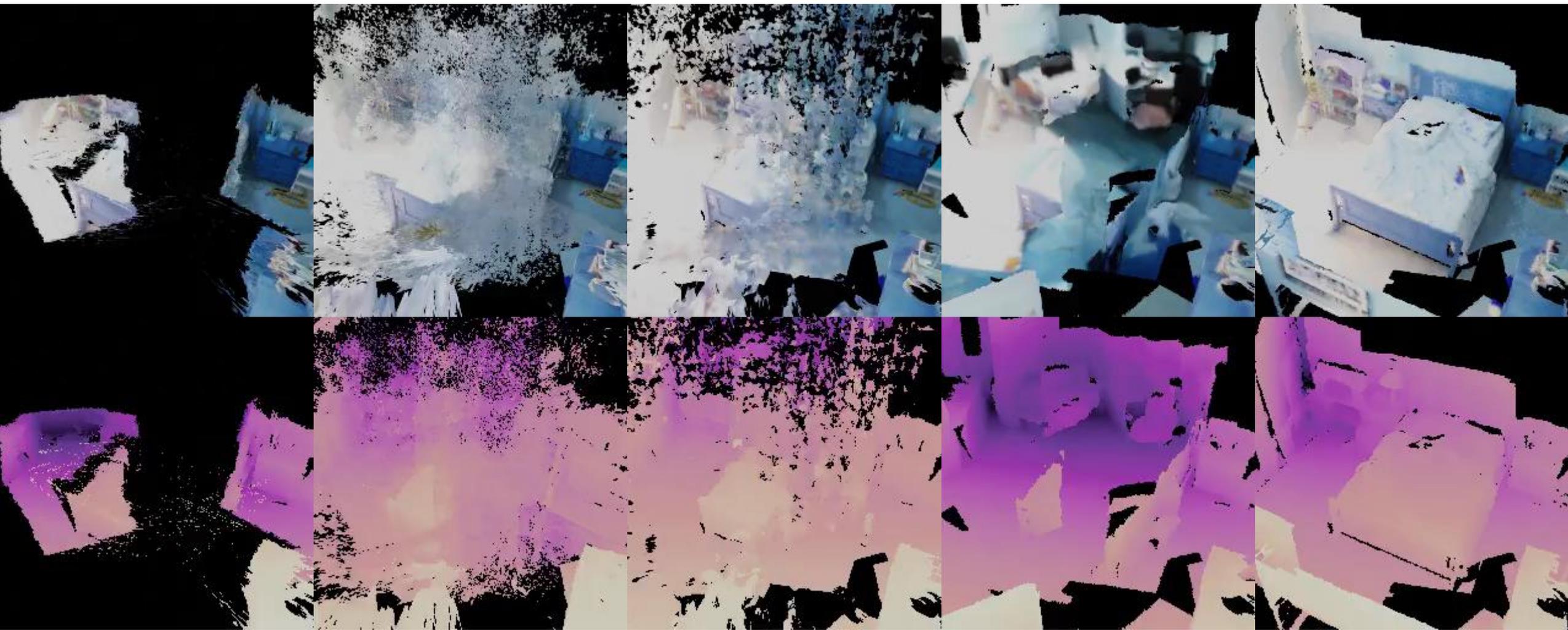
**OUR**

GT



\* Presented videos are rendered from meshes with Blender-3.4

# Experiment – Sparsity 5% (only 4 input views)



DSNGP

DDP

NRGBD

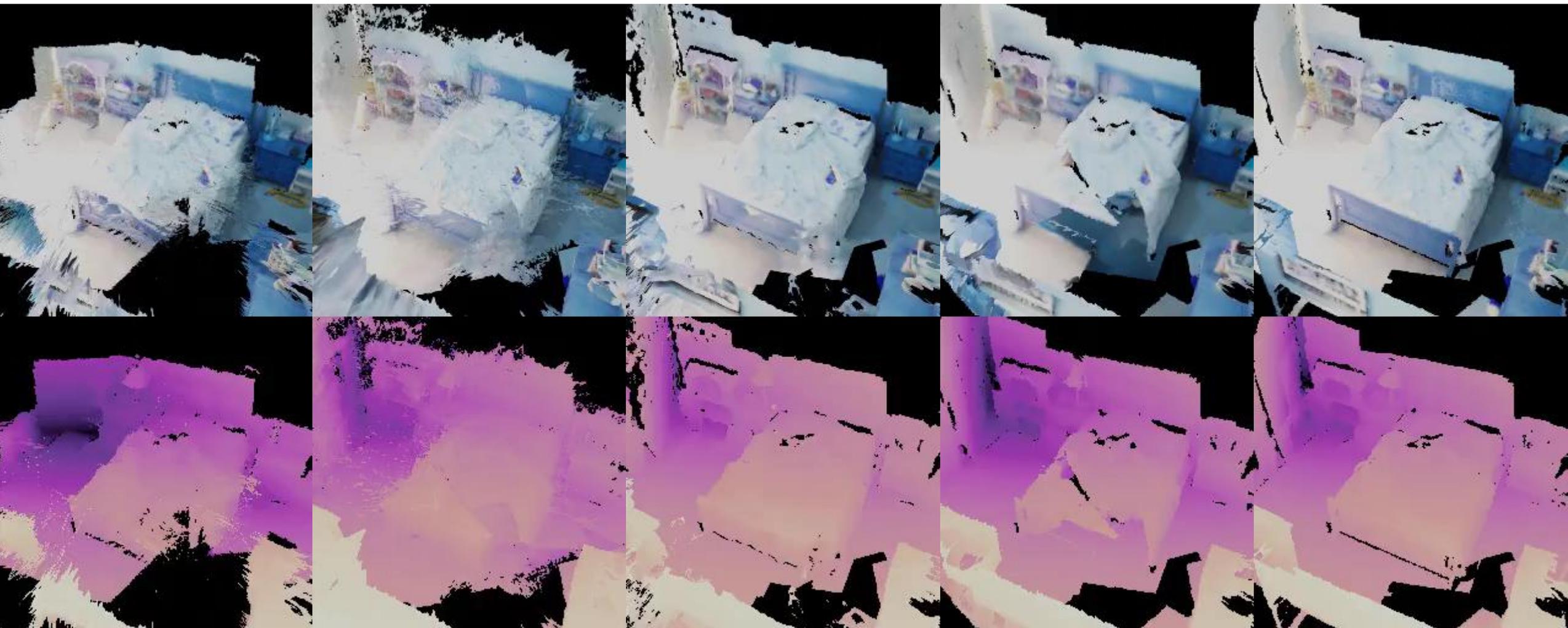
**OUR**

GT



\* Presented videos are rendered from meshes with Blender-3.4

# Experiment – Sparsity 20% (only **13 input views**)



DSNGP

DDP

NRGBD

**OUR**

GT



# Experiment – Ablation Studies

## Seed? Trajectory?



(a) original

(b) seed

(c) trajectory

## Efficiency?

Time	Backend	Repre.	Training	Optimization	Rendering
(DS-)NGP [13, 43]	CUDA C++	NeRF	—	~ 3 min.	~ 0.05 sec.
N-RGBD [3]	PyTorch	NeRF	—	~ 2 hours	~ 2 sec.
DDP [56]	PyTorch	NeRF	~ 1 day	~ 1 hour	~ 1 sec.
Ours	PyTorch	Mesh	~ 3 days	—	~ 3 sec.

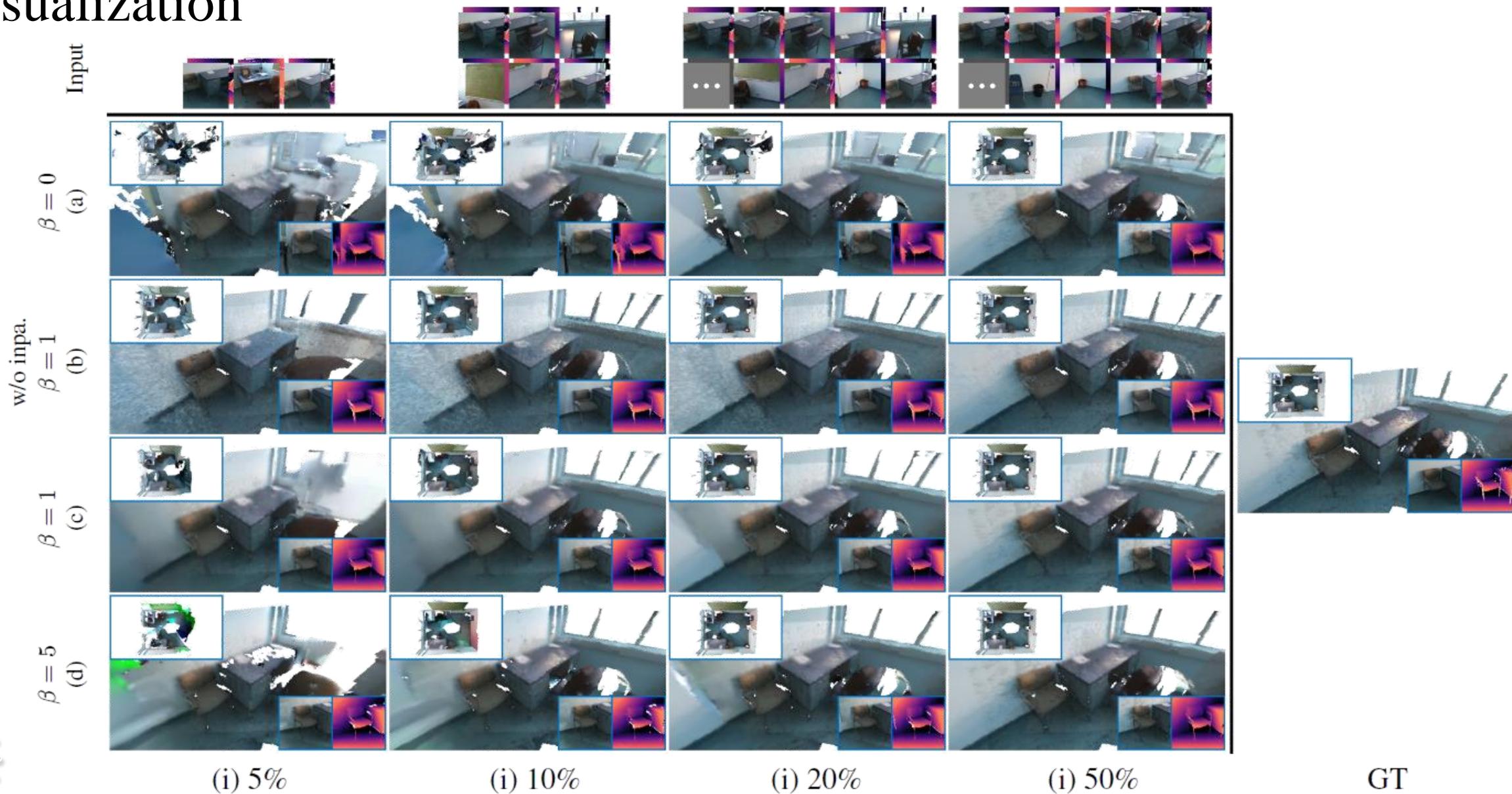
## Classifier-free Conditioning? Inpainting?

Factors		Visual												Geometric											
		PSNR <sub>color</sub>				SSIM <sub>color</sub>				LPIPS <sub>color</sub>				MSE <sub>depth</sub>				CD <sub>mesh</sub>				Comp. <sub>mesh@0.1m</sub>			
Cond.	Inpa.	5%	10%	20%	50%	5%	10%	20%	50%	5%	10%	20%	50%	5%	10%	20%	50%	5%	10%	20%	50%	5%	10%	20%	50%
		9.33	9.30	9.27	9.45	0.331	0.330	0.330	0.333	0.637	0.636	0.637	0.635	1.309	1.304	1.310	1.293	3061	2777	2463	1758	0.513	0.598	0.711	0.862
	✓	12.4	14.7	16.5	17.9	0.411	0.496	0.557	0.583	0.520	0.446	0.393	0.359	1.001	0.837	0.761	0.730	1934	850	443	149	0.600	0.781	0.881	0.931
✓		12.5	13.6	14.7	16.1	0.444	0.473	0.513	0.556	0.449	0.409	0.362	<b>0.315</b>	0.897	0.808	<b>0.662</b>	<b>0.595</b>	1163	<b>699</b>	176	<b>99.0</b>	<b>0.751</b>	0.817	0.887	0.928
✓	✓	<b>14.6</b>	<b>16.0</b>	<b>17.4</b>	<b>18.4</b>	<b>0.522</b>	<b>0.555</b>	<b>0.593</b>	<b>0.603</b>	<b>0.448</b>	<b>0.399</b>	<b>0.359</b>	0.338	<b>0.825</b>	<b>0.805</b>	0.688	0.628	<b>1058</b>	902	<b>156</b>	100	0.747	<b>0.839</b>	<b>0.909</b>	<b>0.936</b>

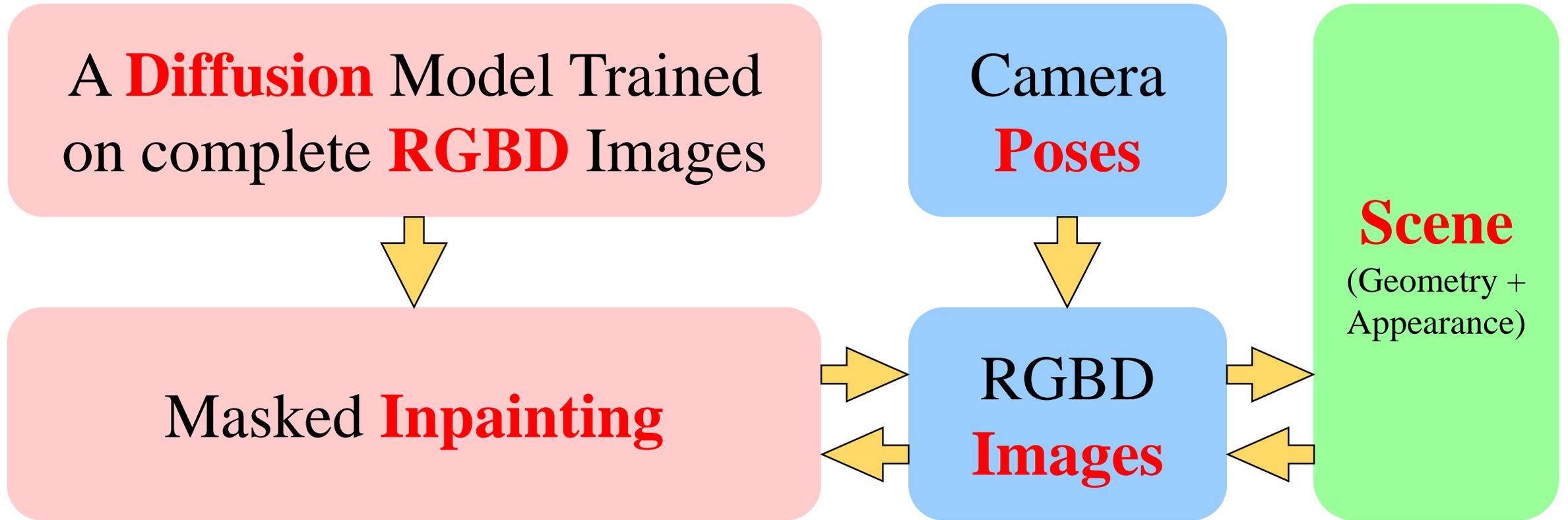


# Experiment – Ablation Studies

## Visualization



# Take-home Message



**Thank You**

