

# Single-view 3D Scene Reconstruction with High-fidelity Shape and Texture



Yixin Chen<sup>1\*</sup> Junfeng Ni<sup>1,2\*†</sup> Nan Jiang<sup>1,3†</sup> Yaowei Zhang<sup>1</sup> Yixin Zhu<sup>3</sup> Siyuan Huang<sup>1</sup>

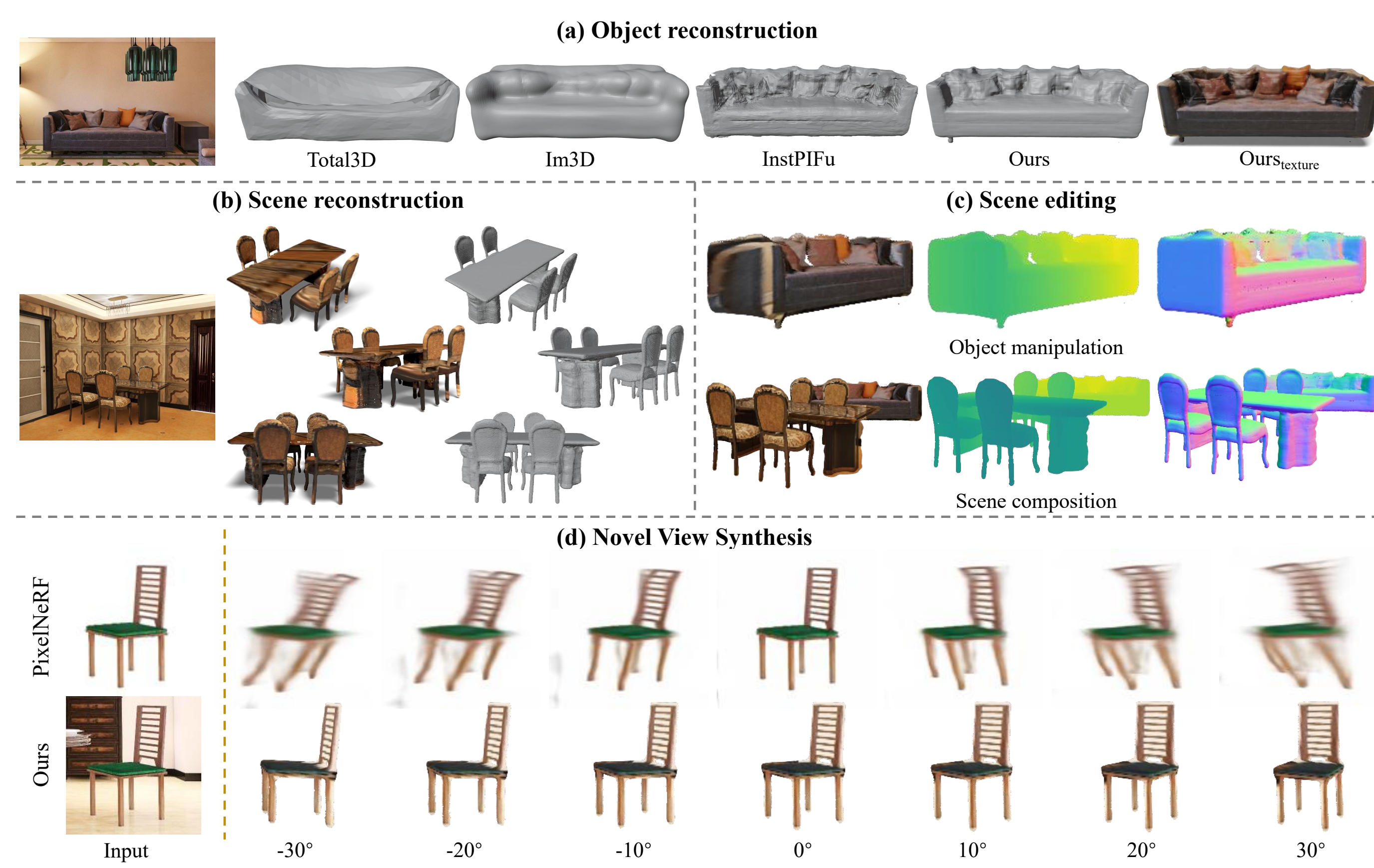
\*Equal contributors †Work done during an internship at BIGAI<sup>1</sup>

<sup>1</sup>National Key Laboratory of General Artificial Intelligence, BIGAI <sup>2</sup>Tsinghua University <sup>3</sup>Peking University



## Single-view 3D Reconstruction

Single-view 3D reconstruction is a challenging task in computer vision that aims to recover a scene's 3D geometry and appearance from a single monocular image.



- Prior work: lack geometry details, without texture
- Our goal: fine-grained object shapes, high-fidelity textures

## Key Challenges

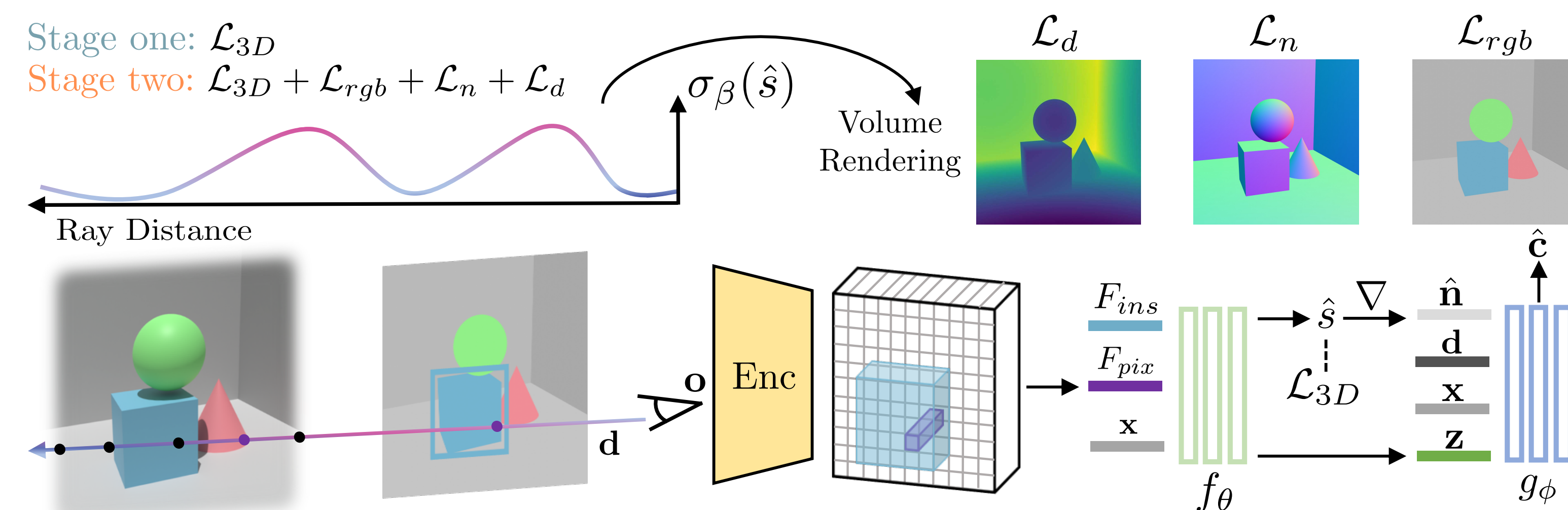
- **Inherent ambiguities** resulting from the **limited observations** in a single image.
- **Shape-appearance ambiguity** for complex textured images.

## Contributions

- We propose a novel framework that recovers **high-fidelity object shapes and textures** from single-view images.
- We introduce a **two-stage learning curriculum** that improves 2D-3D supervision coordination. and addresses shape-appearance ambiguity.
- Experiments and ablations show the benefits of the proposed method and its components.

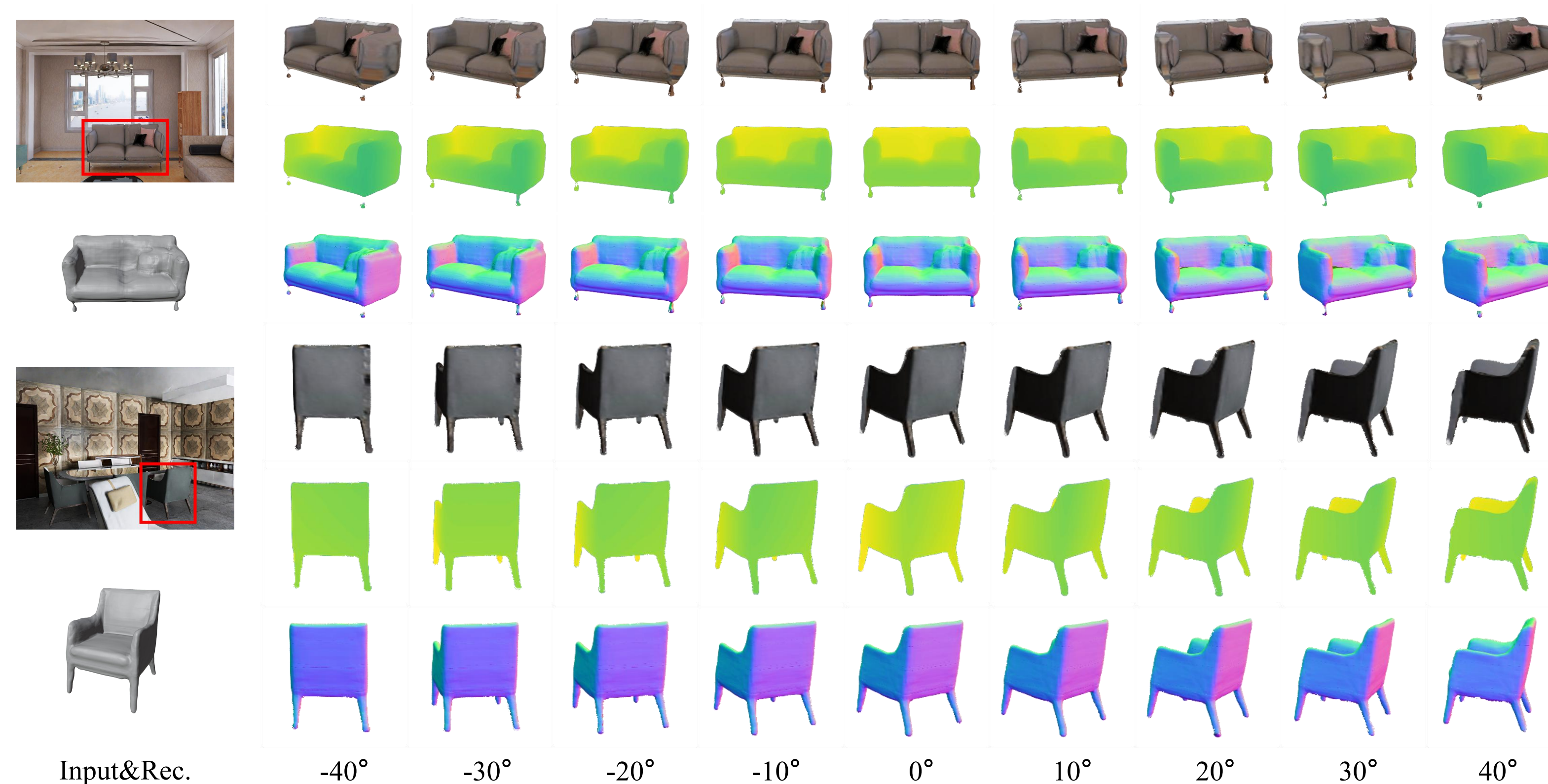
References:  
 [1] Nie Y, et al. Total3D, CVPR 2020 [2] Zhang C, et al. Im3D, CVPR 2021  
 [3] Liu H, et al. InstPIFu, ECCV 2022 [4] Yu A, et al. PixelNeRF, CVPR 2021

## Method

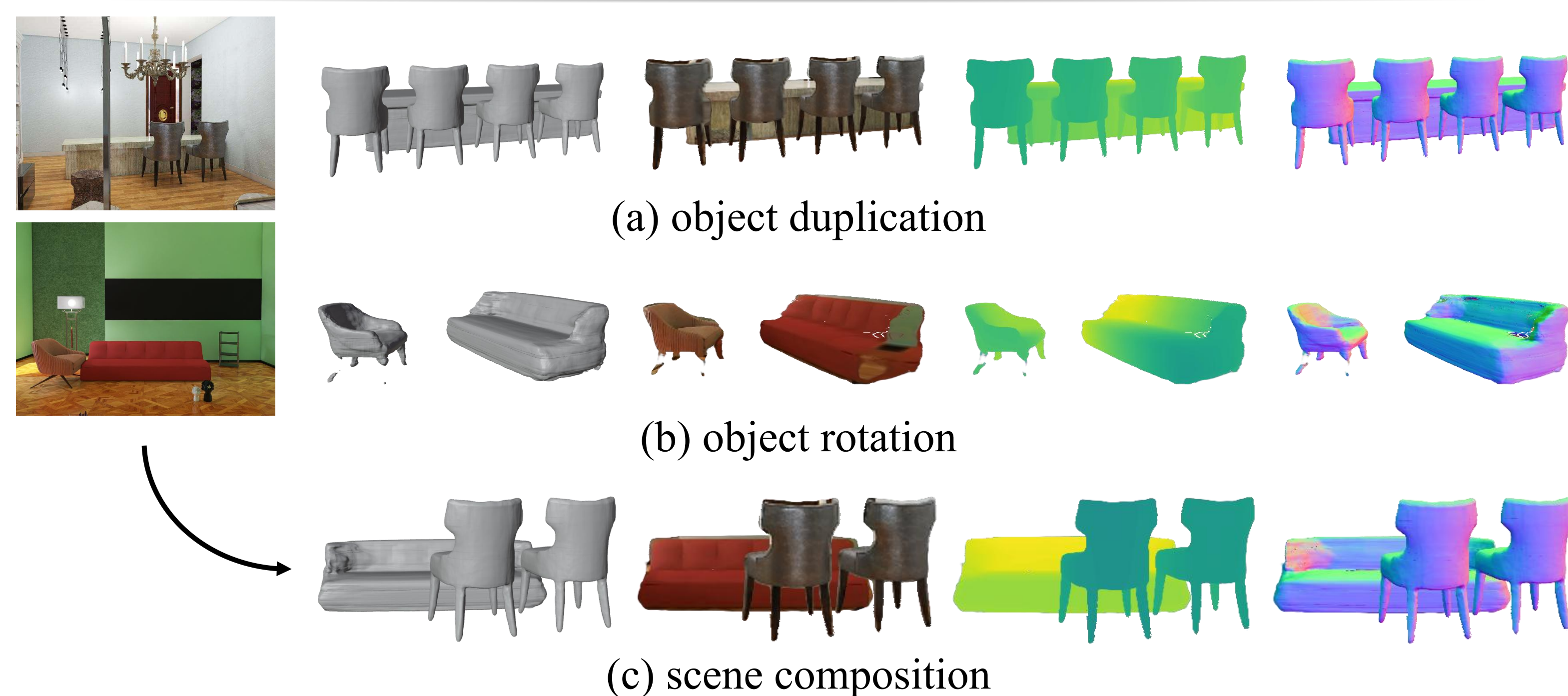


Our model can be trained end-to-end with both 3D shape supervision and volume rendering of color, depth, and surface normal images.

## Novel View Synthesis

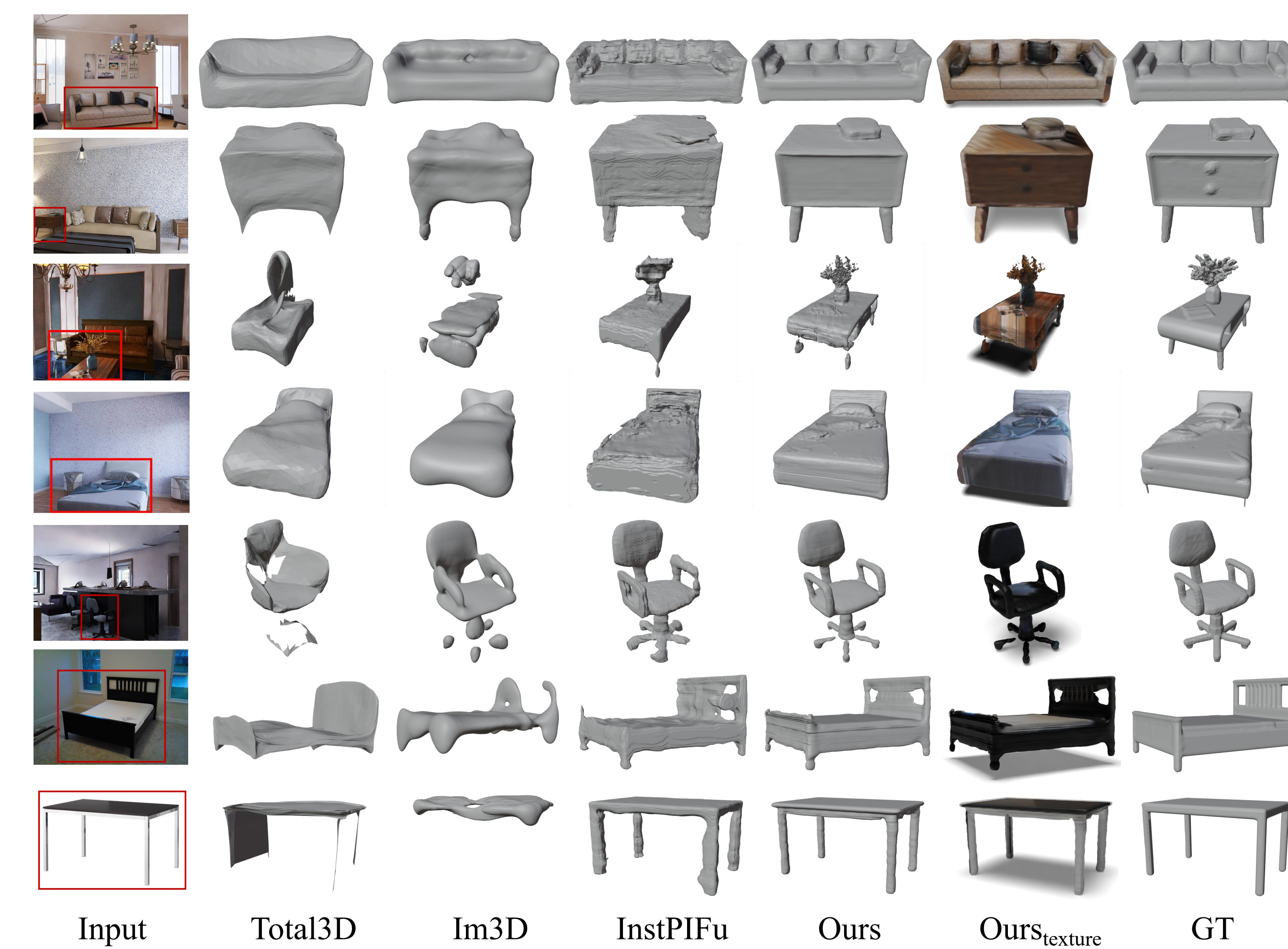


## 3D Scene Editing



## Experiments on 3D-FRONT

Category	bed	chair	sofa	table	desk	nightstand	cabinet	bookshelf	mean	
CD ↓	MGN	15.48	11.67	8.72	20.90	17.59	17.11	13.13	10.21	14.07
	LIEN	16.81	41.40	9.51	35.65	26.63	16.78	7.44	11.70	28.52
	InstPIFu	18.17	14.06	7.66	23.25	33.33	<b>11.73</b>	<b>6.04</b>	8.03	14.46
	Ours	<b>4.96</b>	<b>10.52</b>	<b>4.53</b>	<b>16.12</b>	<b>25.86</b>	17.90	6.79	<b>3.89</b>	<b>10.45</b>
F-Score ↑	MGN	46.81	57.49	64.61	49.80	46.82	47.91	54.18	54.55	55.64
	LIEN	44.28	31.61	61.40	43.22	37.04	50.76	69.21	55.33	45.63
	InstPIFu	47.85	59.08	67.60	56.43	<b>48.49</b>	57.14	<b>73.32</b>	66.13	61.32
	Ours	<b>76.34</b>	<b>69.17</b>	<b>80.06</b>	<b>67.29</b>	47.12	<b>58.48</b>	70.45	<b>85.93</b>	<b>71.36</b>
NC ↑	MGN	-	-	-	-	-	-	-	-	-
	LIEN	-	-	-	-	-	-	-	-	-
	InstPIFu	-	-	-	-	-	-	-	-	-
	Ours	<b>0.896</b>	<b>0.833</b>	<b>0.894</b>	<b>0.838</b>	<b>0.764</b>	<b>0.897</b>	<b>0.856</b>	<b>0.862</b>	<b>0.854</b>



## Single-view 3D Scene Reconstruction

