

Gen3DSR: Generalizable 3D Scene Reconstruction via Divide and Conquer from a Single View

Andreea Ardelean
Dogaru

Mert Özer

Bernhard Egger



Summary

Most single-image scene-level reconstruction methods require 3D supervised end-to-end training and suffer from poor generalization capabilities. We propose a modular system where each module performs well by focusing on specific tasks that are easier to supervise. Our approach is compositional: the scene is *divided* into entities which are reconstructed (*conquered*) individually and composed into the final 3D scene using the unprojected depth as a layout reference.



Input Image



DreamGaussian (non-compositional)



Gen3DSR (ours, compositional)



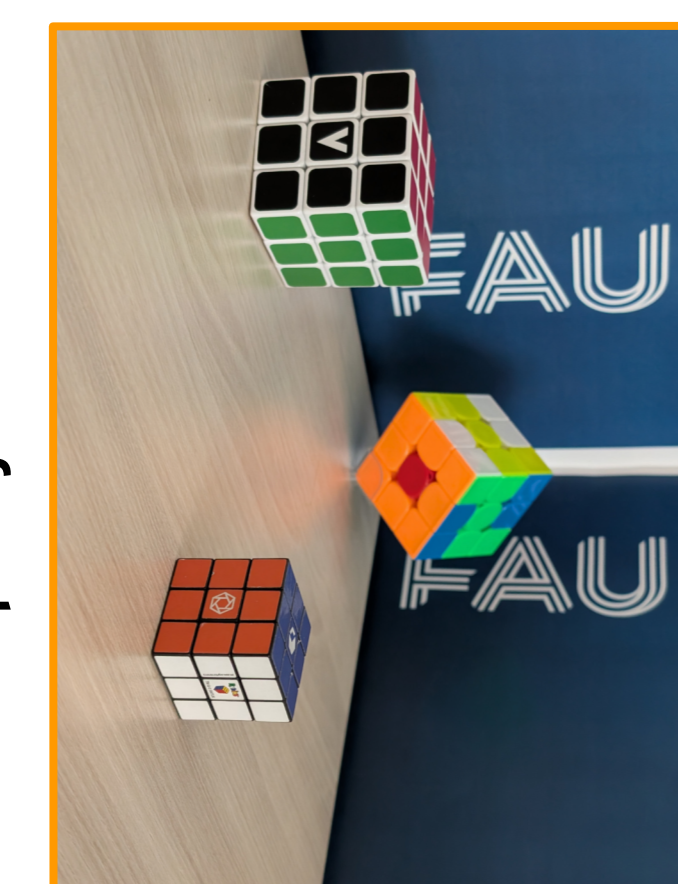
3D Scene Editing

Insights

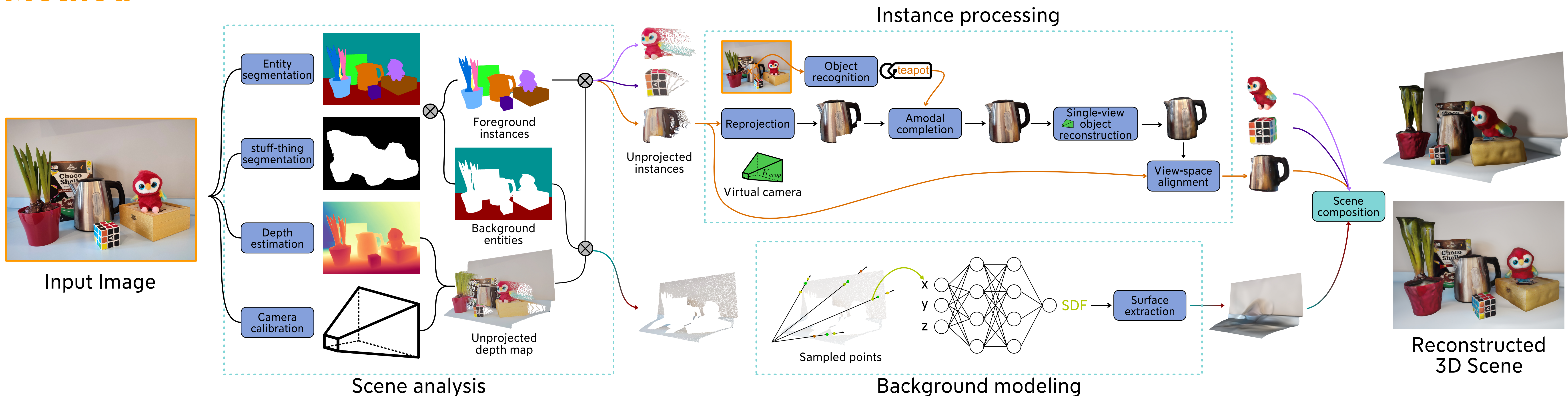
Amodal Completion



Reprojection



Method



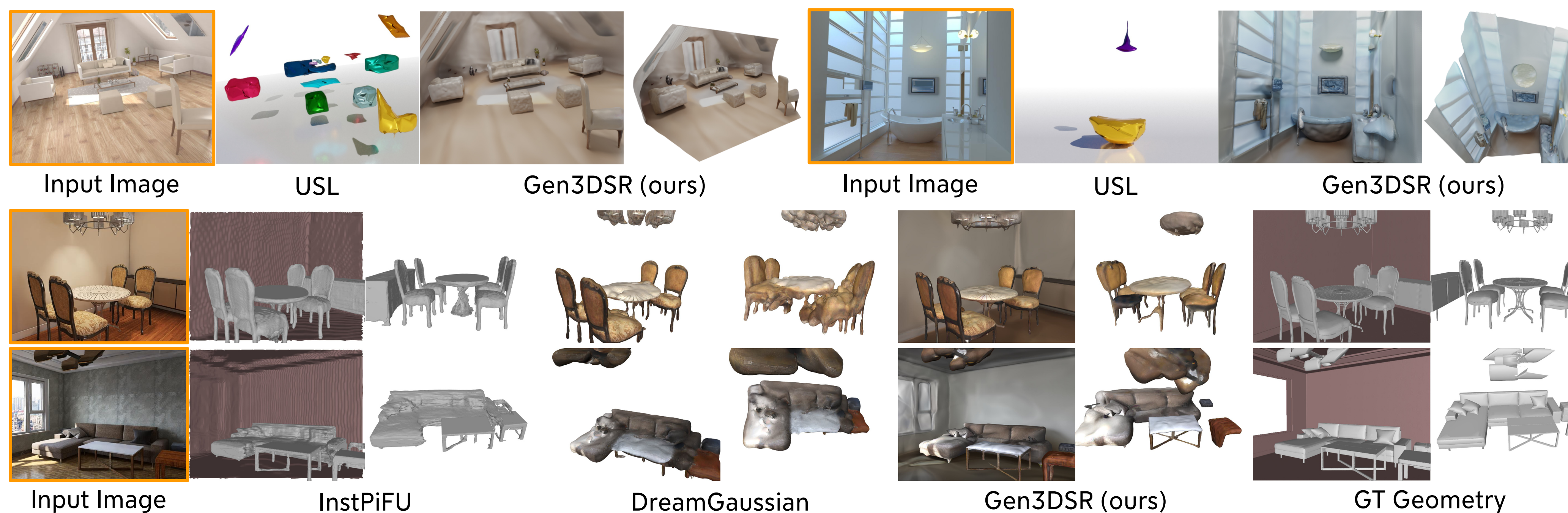
Evaluation

Method	3D-FRONT	
	Chamfer ↓	F-Score ↑
InstPIFu	0.119	70.63
BUOL	0.294	37.04
Uni-3D	0.448	32.97
Gen3DSR	0.099	75.33

Table 1: Full scene reconstruction

Method	3D-FRONT		HOPE-Image	
	Chamfer ↓	F-Score ↑	Chamfer ↓	F-Score ↑
InstPIFu	0.124	74.14	-	-
DreamG	0.207	41.09	3.038	29.57
DreamG + reprojection	0.187	49.98	4.059	30.62
Gen3DSR	0.120	68.82	1.446	54.82

Table 2: Foreground instances reconstruction



Real-world results

