# Gaussian Semantics: Lifting 2D Semantics to 3D
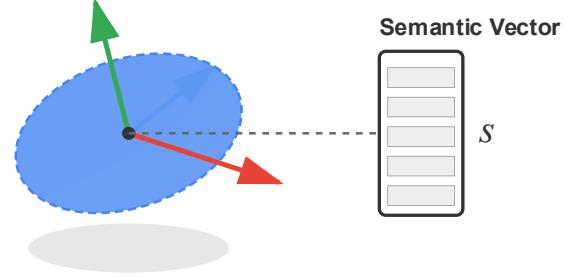
Noora Andreev-Ahmed   Serife Damla Konur
Technical University of Munich

Motivation: 3D Segmentation

- **Goal**: view-consistent 3D semantic labels
- **Bottleneck**: High computational cost & data scarcity for 3D segmentation
- 3D Gaussian Splatting is SOTA for 3D scene reconstruction
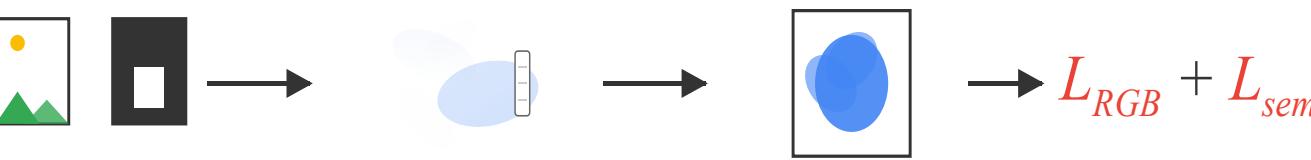
- **Idea**: inject semantics into optimization to obtain semantic label for each Gaussian

Methodology: Joint Optimization

Input
- Images
- Masks

3D Gaussians → **Rasterize** → **Compute Loss** → $L_{RGB} + L_{sem}$
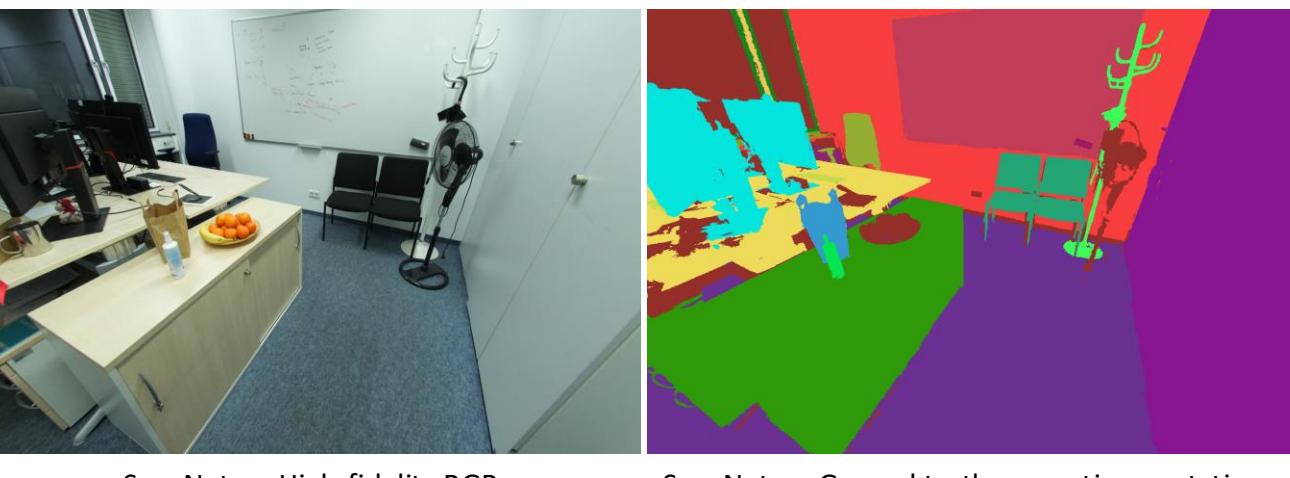
Dataset & Supervision

ScanNet++
- indoor dataset with high-resolution RGB images
- semantic annotations for over 2700 classes
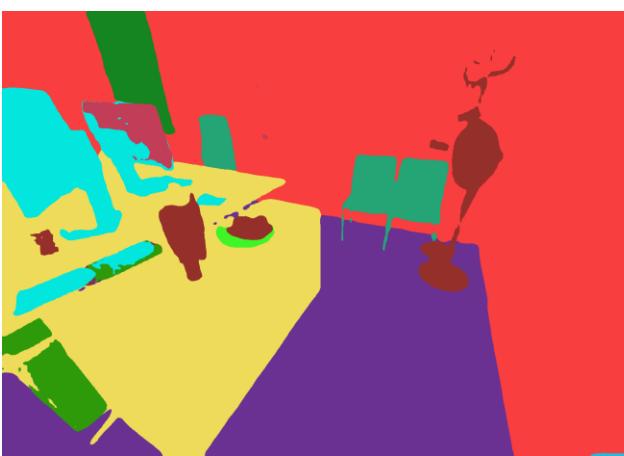
ScanNet++: High-fidelity RGB.

ScanNet++: Ground truth semantic annotation.
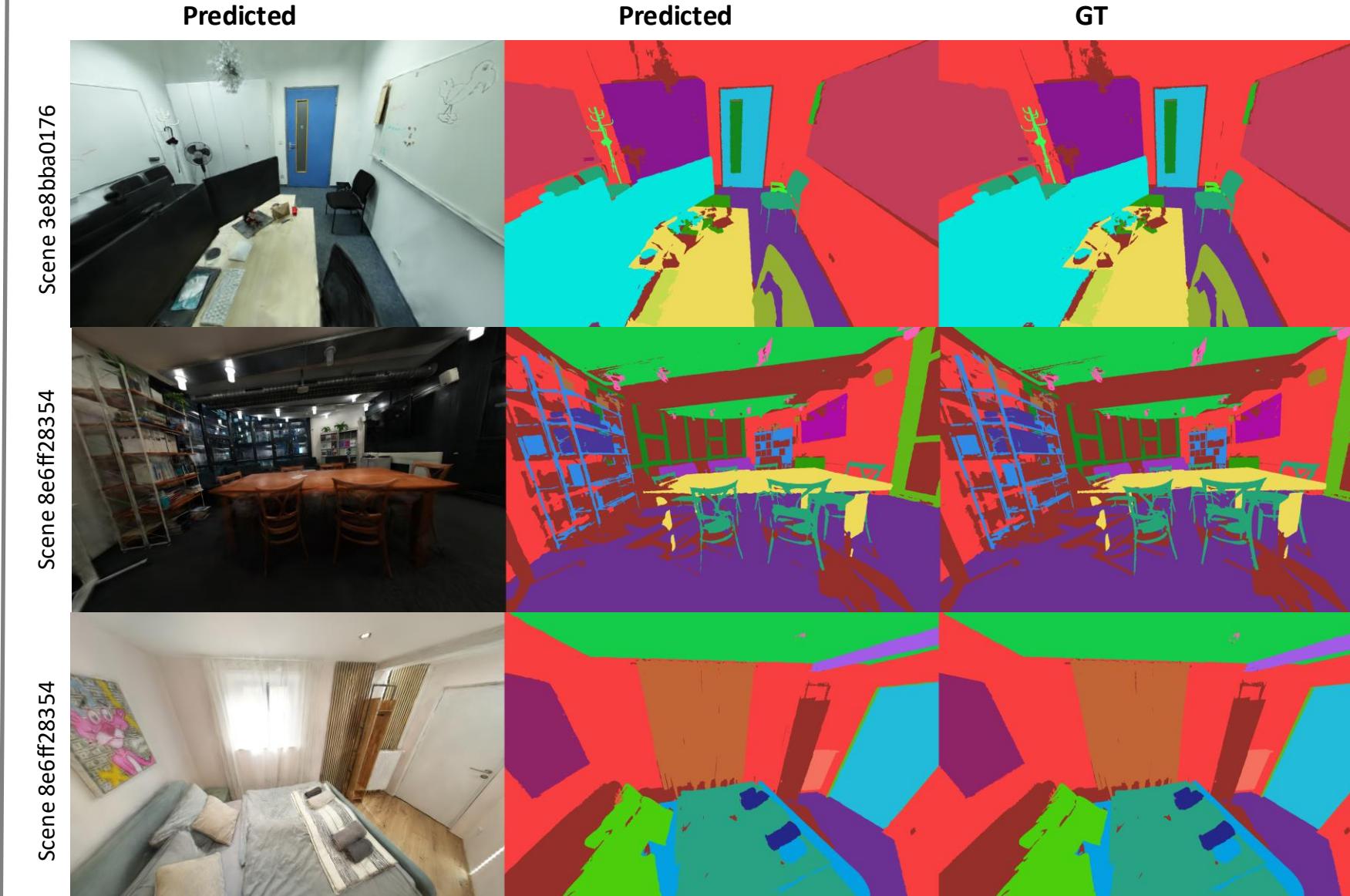
OneFormer
- SOTA image segmentation model
- suffers from temporal inconsistencies
- pre-trained on ADE20K dataset
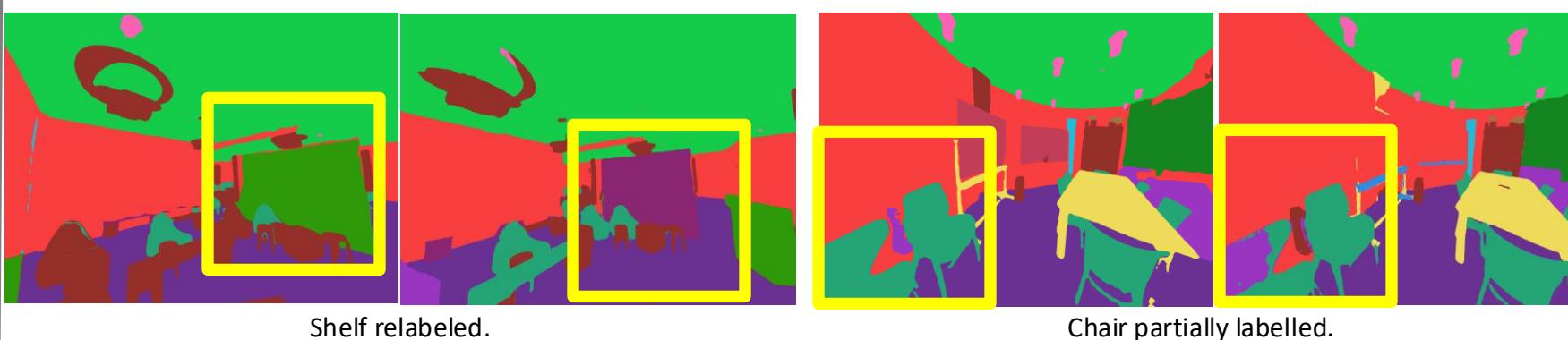
OneFormer prediction (noisy).
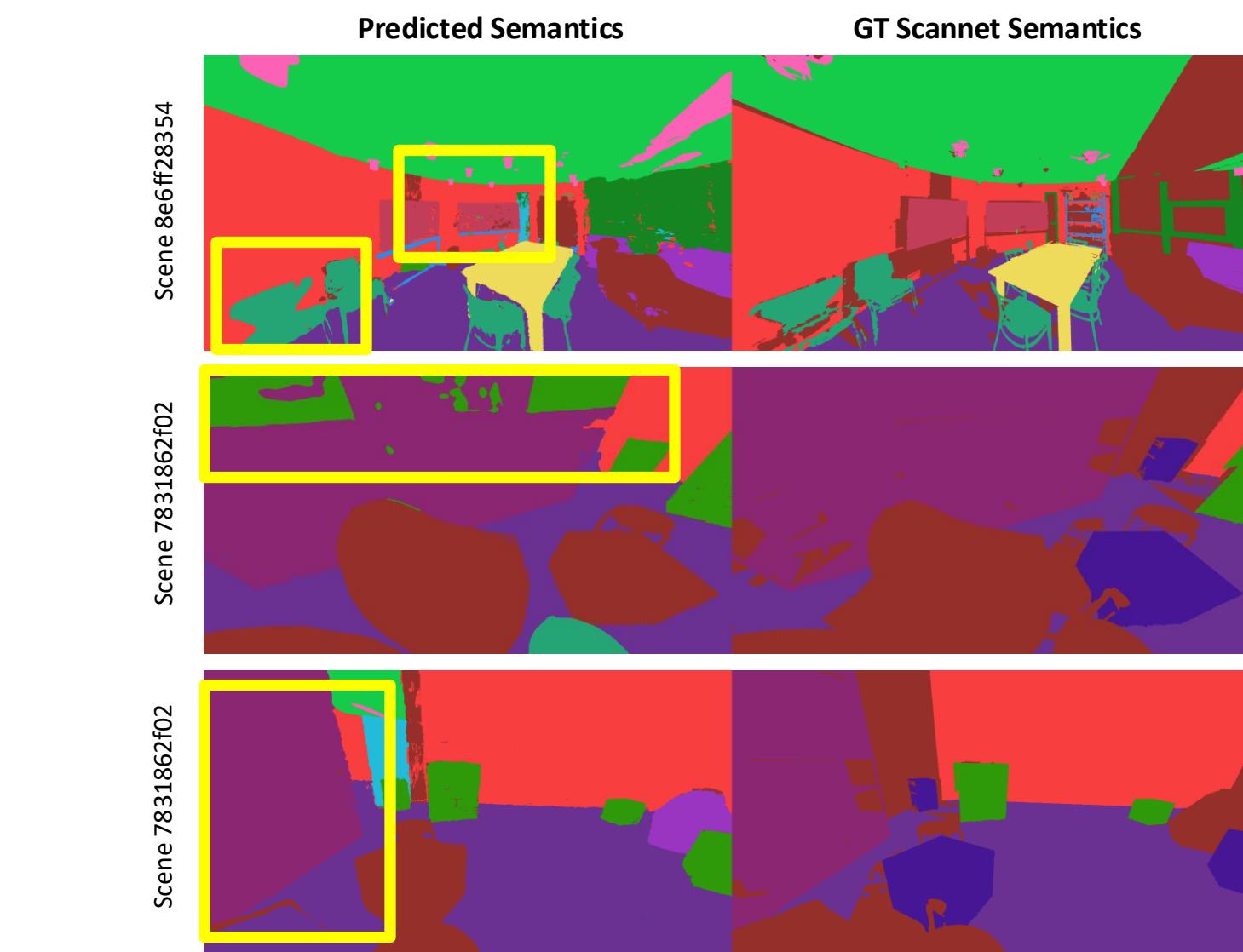
Qualitative Results with ScanNet++ Input Masks

| Predicted | Predicted | GT |

Scene 3e8bba0176

Scene 8e6ff28354

Scene 8e6ff28354

The Problem: 2D Inconsistency

Shelf relabeled.

Chair partially labelled.

Qualitative Results with OneFormer Input Masks

| Predicted Semantics | GT Scannet Semantics |

Scene 8e6ff28354

Scene 7831862f02

Scene 7831862f02
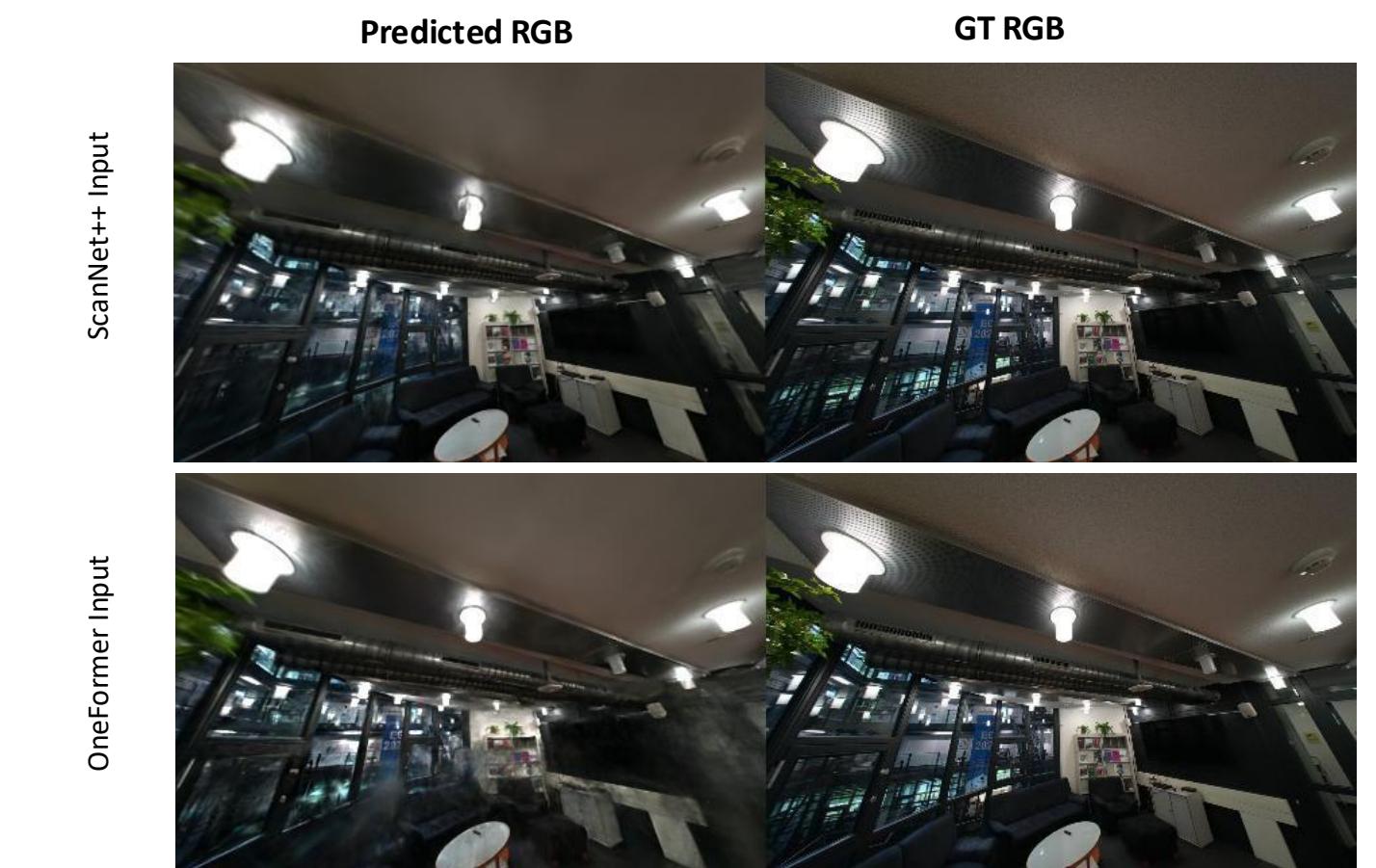
Quantitative Results

Table 1: Per-scene results for two mask sources. RGB: PSNR↑, SSIM↑, LPIPS↓. Semantics: mIoU↑. Gaussian Grouping (GG) is only evaluated for OneFormer inputs.

Mask Source	Scene (#cls)	Gaussian Grouping (Baseline)				Ours (Joint Opt)			
		PSNR	SSIM	LPIPS	mIoU	PSNR	SSIM	LPIPS	mIoU
ScanNet++	3e8bba0176 (28)			—		26.37	0.889	0.23	0.92
	8e6ff28354 (22)			—		24.74	0.820	0.30	0.90
	7831862f02 (14)			—		30.62	0.900	0.18	0.99
	f36e3e1e53 (16)			—		26.69	0.910	0.20	0.94
OneFormer	3e8bba0176 (17)	15.09	0.59	0.58	0.13	25.75	0.790	0.24	0.56
	8e6ff28354 (15)	15.57	0.49	0.61	0.12	21.99	0.750	0.38	0.51
	7831862f02 (11)	15.49	0.54	0.59	0.17	23.49	0.850	0.27	0.49
	f36e3e1e53 (11)	16.94	0.65	0.54	0.21	22.53	0.830	0.34	0.54

The Trade-off: RGB Quality

| Predicted RGB | GT RGB |

ScanNet++ Input

OneFormer Input

Ablations

Ablation Study: Softmax & Entropy Regularization

- **complex scene**: softmax and entropy regularization improve performance (+0.9-2.4% mIoU)
- **simple scene**: standard optimization sufficient

- **softmax**: probabilistic modeling of class assignments
- **entropy**: regularization weight to sharpen semantic boundaries
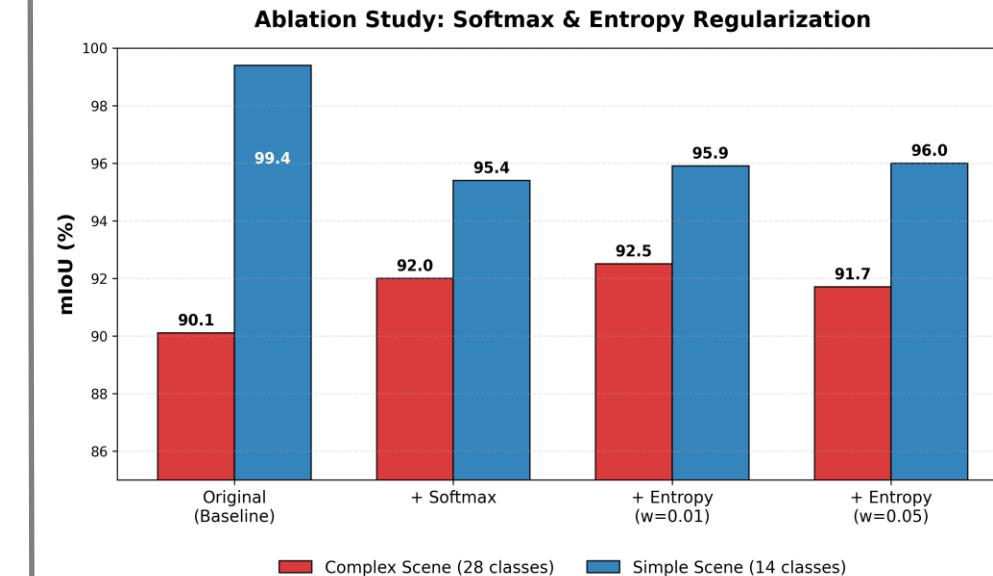
Table 2: Ablation Study: Softmax and Entropy Regularization

Scene	Method	RGB PSNR	RGB SSIM	mIoU (%)	Pixel Acc. (%)
38bba0176	Original (Baseline)	**24.74**	**0.822**	90.1	**97.9**
	+ Softmax	26.39	0.890	92.0	98.8
	+ Entropy ($w=0.01$)	26.28	0.888	92.5	98.8
	+ Entropy ($w=0.05$)	26.28	0.887	91.7	98.8
7831862f02	Original (Baseline)	30.62	**0.903**	99.4	99.9
	+ Softmax	27.42	0.828	95.4	99.2
	+ Entropy ($w=0.01$)	27.32	0.827	95.9	99.1
	+ Entropy ($w=0.05$)	27.33	0.826	96.0	99.1