C^3-GS: Learning Context-aware, Cross-dimension, Cross-scale Feature for Generalizable Gaussian Splatting

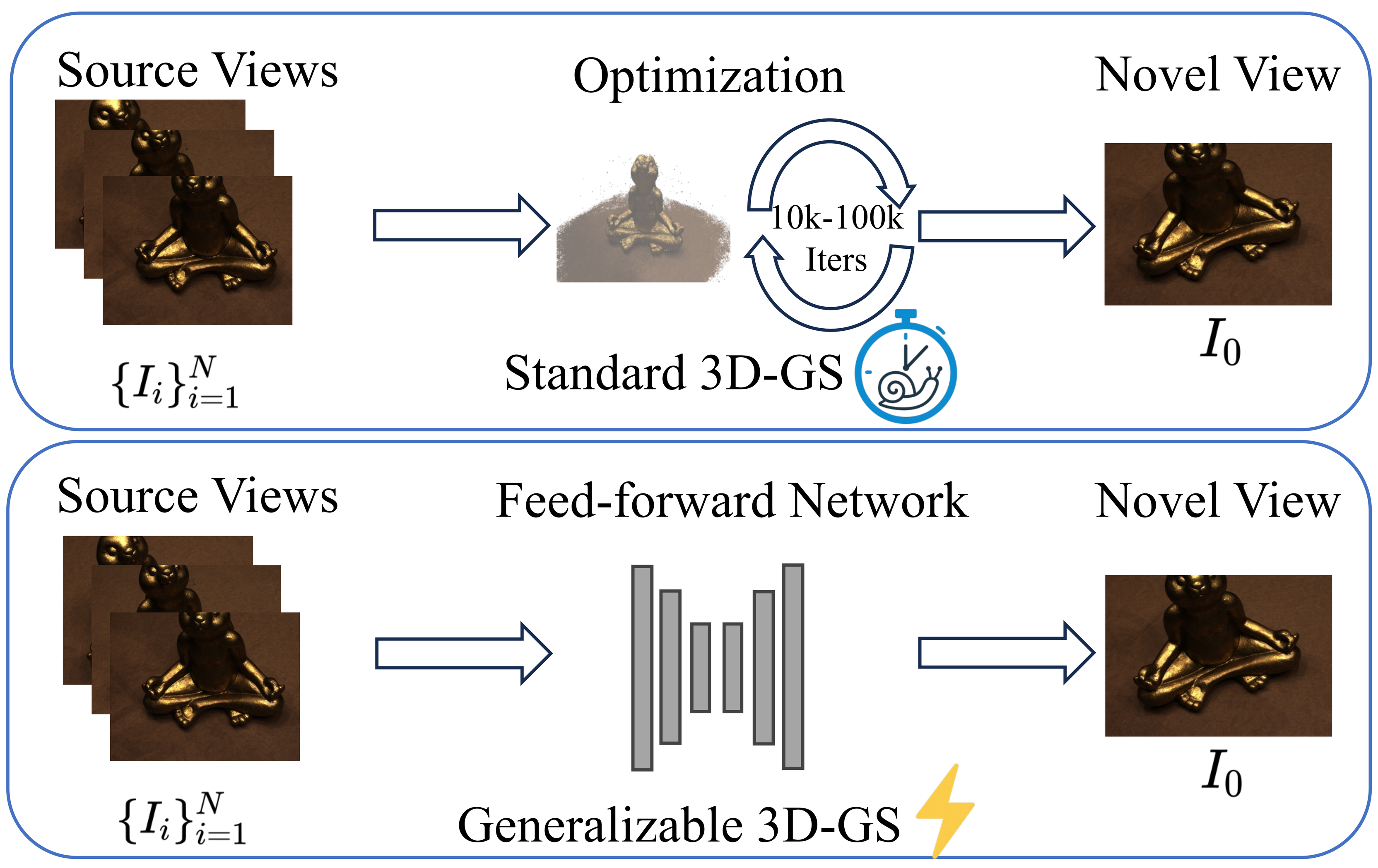
Yuxi Hu[1] Jun Zhang[1] Kuangyi Chen[1] Zhe Zhang[2] Friedrich Fraundorfer[1]

[1]Institute of Visual Computing, Graz University of Technology, Austria

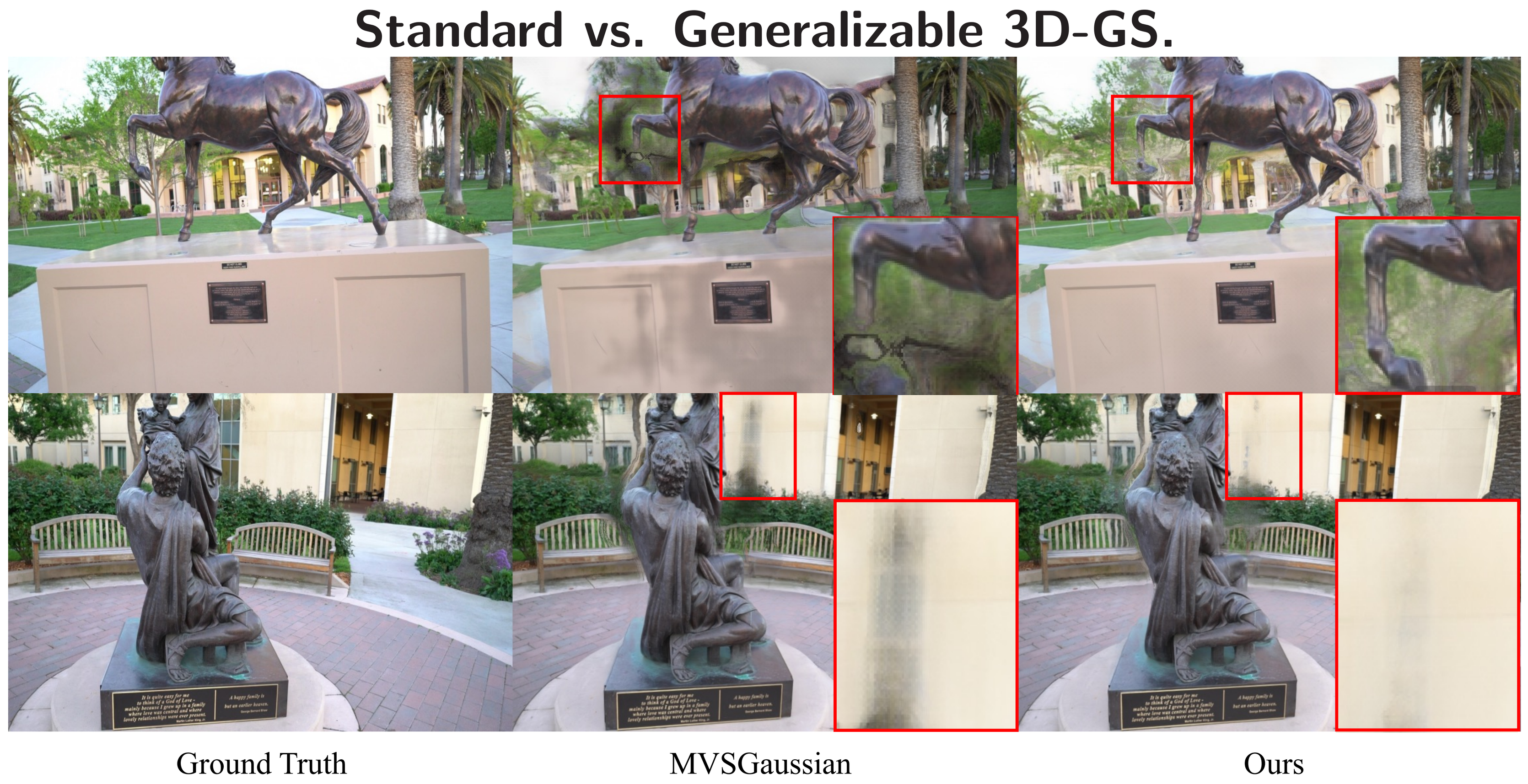
[2]School of Electronic and Computer Engineering, Peking University, China

Contributions

Generalizable Gaussian Splatting aims to synthesize novel views for unseen scenes without per-scene optimization. We propose C^3-GS, a framework that enhances feature learning by incorporating context-aware, cross-dimension, and cross-scale constraints.

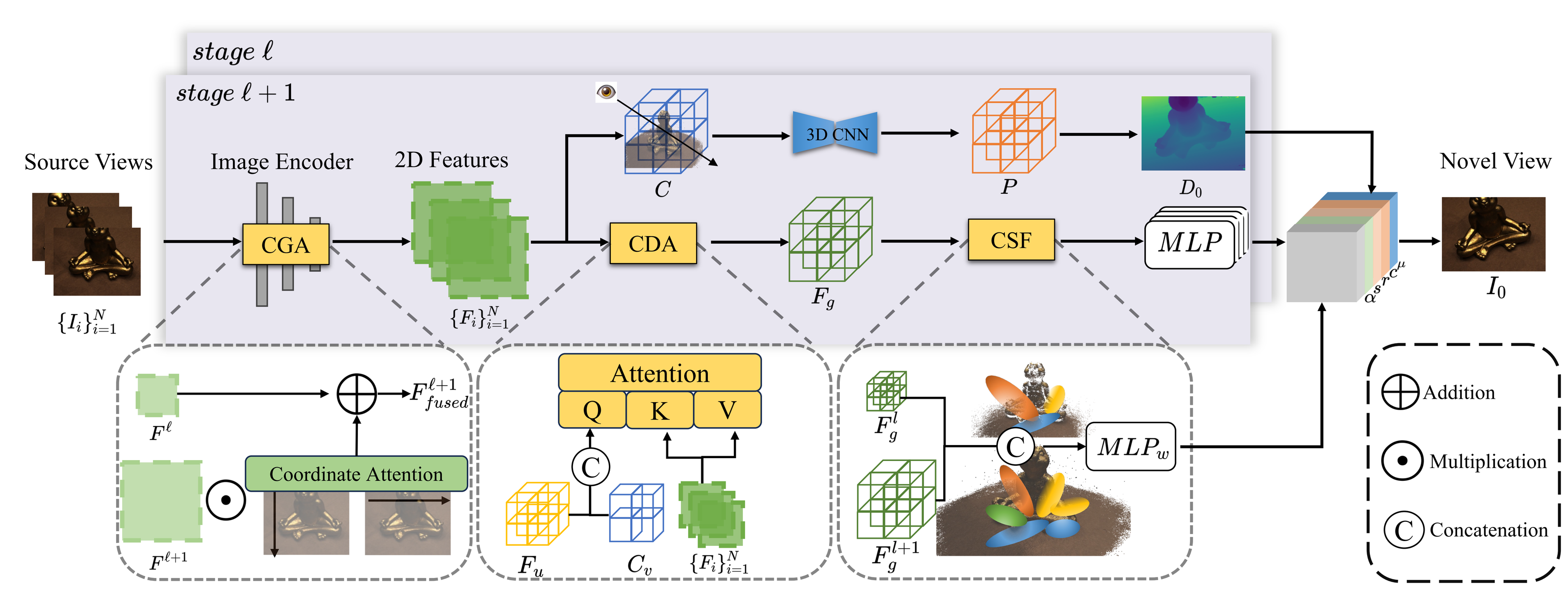


Standard vs. Generalizable 3D-GS.



Synthesized images on "Horse" and "Family" scenes.

Pipeline



The overall architecture. The proposed method is a coarse-to-fine framework that estimates depths and Gaussian representations from low resolution ($stage\ \ell$) to high resolution ($stage\ \ell+1$).

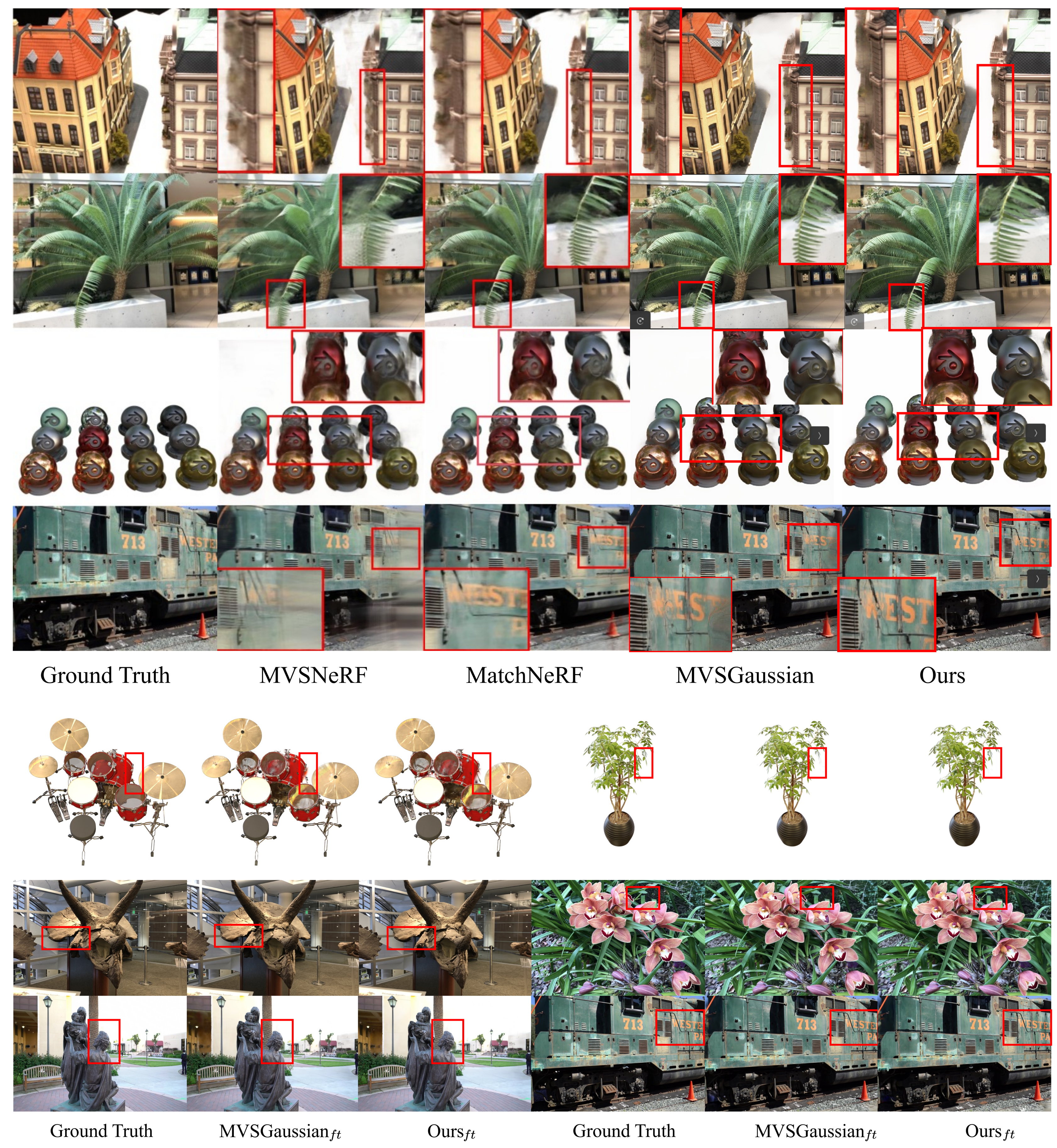
- We propose **Coordinate-Guided Attention (CGA)**, which models the relative importance of spatial positions between features to capture long-range, feature-channel dependencies.
- We introduce **Cross-Dimensional Attention (CDA)**, to encode multi-view features for complete Gaussian representation. CDA enhances the interaction between 2D features and 3D scene information.
- We propose **Cross-Scale Fusion (CSF)**, which modulates the opacity of Gaussians across scales to enhance consistency. The output Gaussians are used to render the novel view I_0.

Link

Code LinkedIn

Scan the QR code for the full paper, code, and LinkedIn contact.

Qualitative Comparison



Qualitative comparison on DTU, Real Forward-facing, NeRF Synthetic, and Tanks and Temples datasets.

Quantitative Comparison

Method	2-view			3-view		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
MVSNeRF	24.03	0.914	0.192	26.63	0.931	0.168
MatchNeRF	25.03	0.919	0.181	26.91	0.934	0.159
PixelSplat	14.01	0.662	0.389	-	-	-
MVSplat	13.94	0.473	0.385	14.29	0.508	0.370
MVPGS	17.55	0.823	0.147	20.65	0.877	0.099
MVSGaussian	23.83	0.918	0.121	27.03	0.959	0.084
Ours	24.27	0.936	0.113	27.87	0.962	0.077

Generalization on DTU dataset.

Method	Novel view		
	Abs err ↓	Acc(2)↑	Acc(10)↑
PixelNeRF	47.8	0.039	0.187
IBRNet	324	0.000	0.866
MVSNeRF	7.00	0.717	0.866
MVSGaussian	4.06	0.820	0.936
Ours	3.79	0.832	0.943

Depth on DTU dataset.

Method	Views	Real Forward-facing			NeRF Synthetic			Tanks and Temples		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
MVSNeRF	2	20.22	0.763	0.287	20.56	0.856	0.243	18.92	0.756	0.326
MatchNeRF		20.59	0.775	0.276	20.57	0.864	0.200	19.88	0.773	0.334
MVSplat		15.32	0.370	0.422	12.59	0.665	0.282	16.35	0.490	0.355
MVPGS		18.53	0.607	0.280	-	-	-	16.33	0.687	0.373
MVSGaussian		22.58	0.817	0.192	23.56	0.929	0.122	21.05	0.830	0.222
Ours		22.32	0.818	0.201	23.83	0.930	0.109	21.58	0.832	0.221
MVSNeRF	3	21.93	0.795	0.252	23.62	0.897	0.176	20.87	0.823	0.260
MatchNeRF		22.43	0.805	0.244	23.20	0.897	0.164	20.80	0.793	0.300
MVPGS		19.91	0.696	0.229	-	-	-	18.60	0.607	0.467
MVSGaussian		23.87	0.856	0.169	25.31	0.943	0.098	22.70	0.872	0.187
Ours		23.97	0.857	0.167	26.14	0.946	0.079	23.30	0.879	0.178
MVPGS	4	21.28	0.750	0.180	-	-	-	20.27	0.679	0.371
MVSGaussian		24.01	0.864	0.175	24.81	0.938	0.106	21.85	0.860	0.209
Ours		24.16	0.867	0.174	25.29	0.941	0.087	22.35	0.866	0.204

Generalization on Real Forward-facing, NeRF Synthetic, and Tanks and Temples datasets.

Ablation Study

	CGA	CDA	CSF	PSNR↑	SSIM↑	LPIPS↓
baseline	-	-	-	27.03	0.959	0.084
No.1	✓	-	-	27.07	0.959	0.084
No.2	-	✓	-	27.34	0.959	0.084
No.3	-	-	✓	27.31	0.960	0.080
No.4	✓	✓	-	27.42	0.960	0.078
No.5	✓	-	✓	27.58	0.960	0.079
No.6	-	✓	✓	27.79	0.961	0.077
No.7	✓	✓	✓	**27.87**	**0.962**	**0.077**

Ablation studies on the combination of modules on DTU dataset. The baseline method is MVSGaussian.

Method	Views	DTU		
		PSNR ↑	SSIM ↑	LPIPS ↓
Ours	2	24.27	0.936	0.113
Ours*		25.14	0.939	0.102
Ours*	3	27.87	0.962	0.077
Ours	4	27.63	0.962	0.081
Ours*		27.95	0.962	0.079

Ablation studies on the number of views on DTU. "Ours*" denotes identical training and testing views.