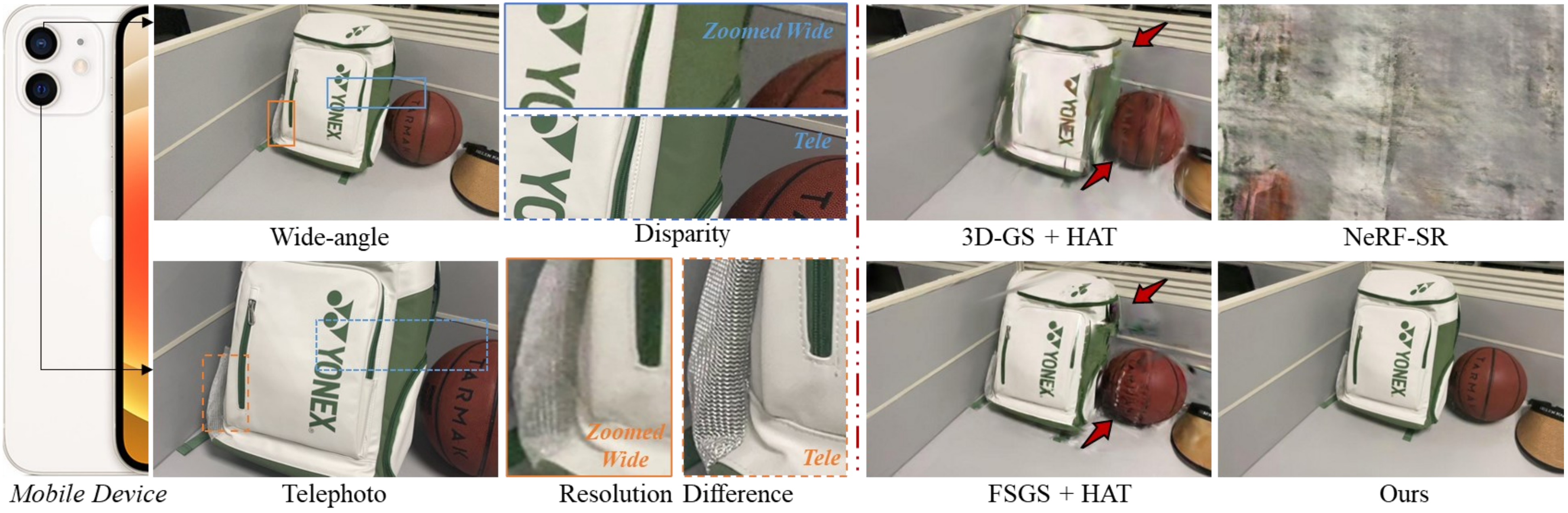
High-Resolution and Few-shot View Synthesis from Asymmetric Dual-lens Inputs

Ruikang Xu · Mingde Yao · Yue Li · Yueyi Zhang · Zhiwei Xiong

University of Science and Technology of China

Asymmetric Dual-lens System for Novel View Synthesis

1) Combining the wide-angle and telephoto images forms an asymmetric stereo configuration, which stores the geometric information to facilitate the few-shot training.
2) The telephoto images have higher resolution than the wide-angle ones, naturally providing additional HR guidance to improve the resolution of newly synthesized views.

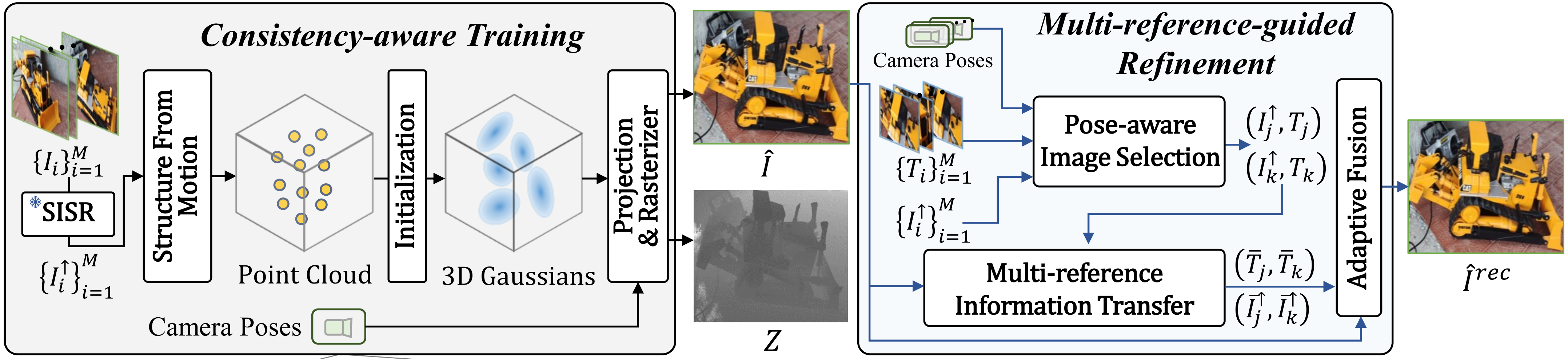


(a) Characteristics of Dual-lens System (b) Visual Comparison

中国科学技术大学 University of Science and Technology of China

DL-GS: Consistency-aware Training

- Dual-lens-consistent Loss: Enforce the view consistency between the newly synthesized view and the corresponding telephoto image, implicitly exploiting the geometric information of dual-lens system.
- Depth-wise Loss: Monocular depth as supplementary supervision.



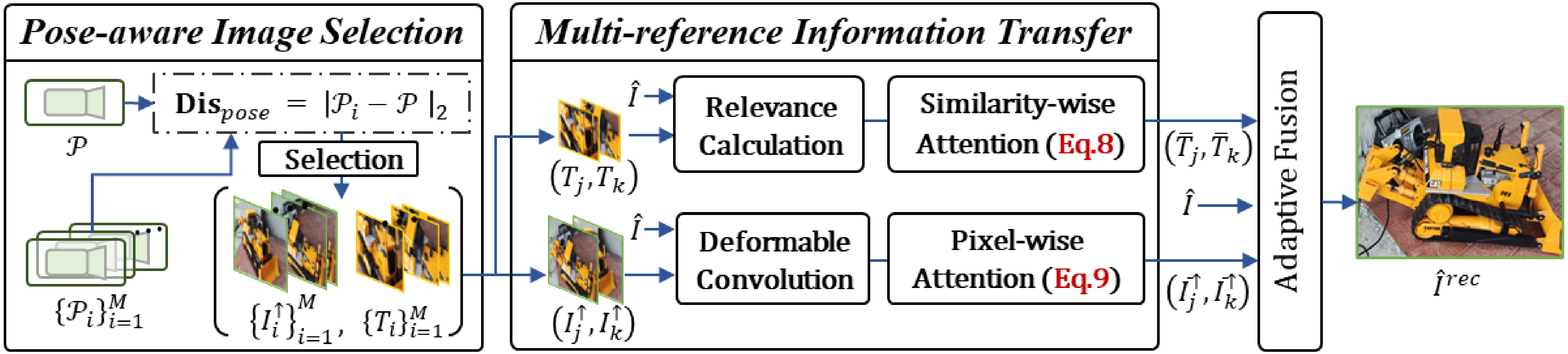
Overview of Pipeline

$$\mathcal{L} = \mathcal{L}_{GS}(\hat{I}, I^{\uparrow}) + \beta_1 \mathcal{R}_c + \beta_2 \mathcal{R}_d \qquad \mathcal{R}_c = \mathrm{Mask}_v \|\mathbf{Warp}_c(\hat{I}) - T\|_1 \qquad \mathcal{R}_d = \frac{\mathbf{Cov}\,(Z, D)}{\sqrt{\mathbf{Var}\,(Z)}\sqrt{\mathbf{Var}\,(D)}}$$

Loss Functions for Consistency-aware Training

DL-GS: Multi-reference-guided Refinement

- ➢ Using camera positions relationship (R-T matrices) to select the ***Reference*** images from training samples and perform information Transfer;
- ➢ HR-LR pairs from telephoto and wide-angle images for ***self-supervised training***.



Multi-reference-guided Refinement

$$\mathcal{L}_{DL} = \lambda_1 \|\mathbf{Crop}(\hat{I}^{rec}) - T^{align}\|_2 + \lambda_2 \|\hat{I}^{rec} - I^{\uparrow}\|_2 + \lambda_3 \mathcal{L}_{cx}(\mathbf{Crop}(\hat{I}^{rec}), T)$$

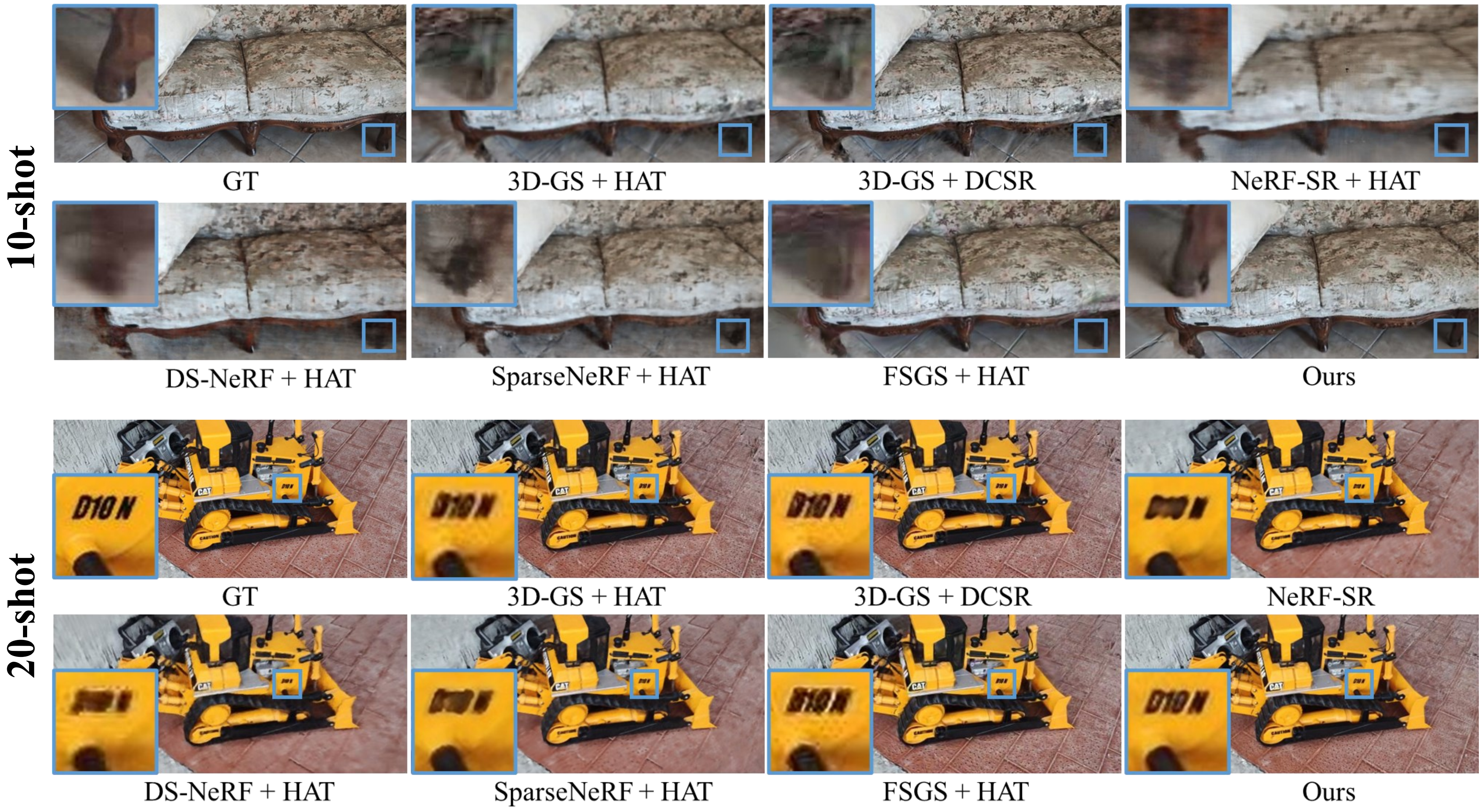
Self-training Loss Function

Quantitative comparisons on Simulated Data

Method	10-shot			20-shot			90-shot		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
3D-GS [19] + Bicubic	17.63	0.4979	0.4411	20.75	0.5905	0.3624	24.03	0.7113	0.2677
3D-GS [19] + SwinIR [25]	17.84	0.4988	0.4387	20.87	0.5924	0.3613	24.52	0.7192	0.2628
3D-GS [19] + HAT [9]	17.89	0.4995	0.4402	20.89	0.5931	0.3615	24.57	0.7196	0.2650
3D-GS [19] + DCSR [51]	17.92	0.5037	<u>0.4314</u>	20.92	0.5964	<u>0.3592</u>	24.60	0.7265	0.2582
NeRF-SR [49]	17.40	0.5032	0.4830	20.84	<u>0.5967</u>	0.3726	<u>24.89</u>	<u>0.7394</u>	<u>0.2422</u>
DS-NeRF [13] + HAT [9]	19.05	<u>0.5546</u>	0.4598	<u>21.54</u>	0.5801	0.4366	22.47	0.6002	0.4395
RegNeRF [33] + HAT [9]	18.78	0.5539	0.4573	20.18	0.5613	0.4476	22.34	0.6259	0.4040
SparseNeRF [50] + HAT [9]	<u>19.12</u>	0.5441	0.4482	21.31	0.5724	0.4398	22.49	0.6329	0.4006
FSGS [65] + HAT [9]	19.09	0.5511	0.4321	20.68	0.5897	0.3637	24.42	0.7183	0.2526
Ours	**19.67**	**0.5772**	**0.3877**	**21.77**	**0.6366**	**0.3366**	**25.61**	**0.7692**	**0.2076**

- Baselines: 1) vanilla 3D-GS followed with SISR; 2) vanilla 3D-GS followed with dual-lens SR; 3) HR NVS method; 4) few-shot NVS methods followed with HAT.
- DL-GS shows superior performance over the previous methods by leveraging the characteristics of the dual-lens system.

Qualitative comparisons on Simulated Data



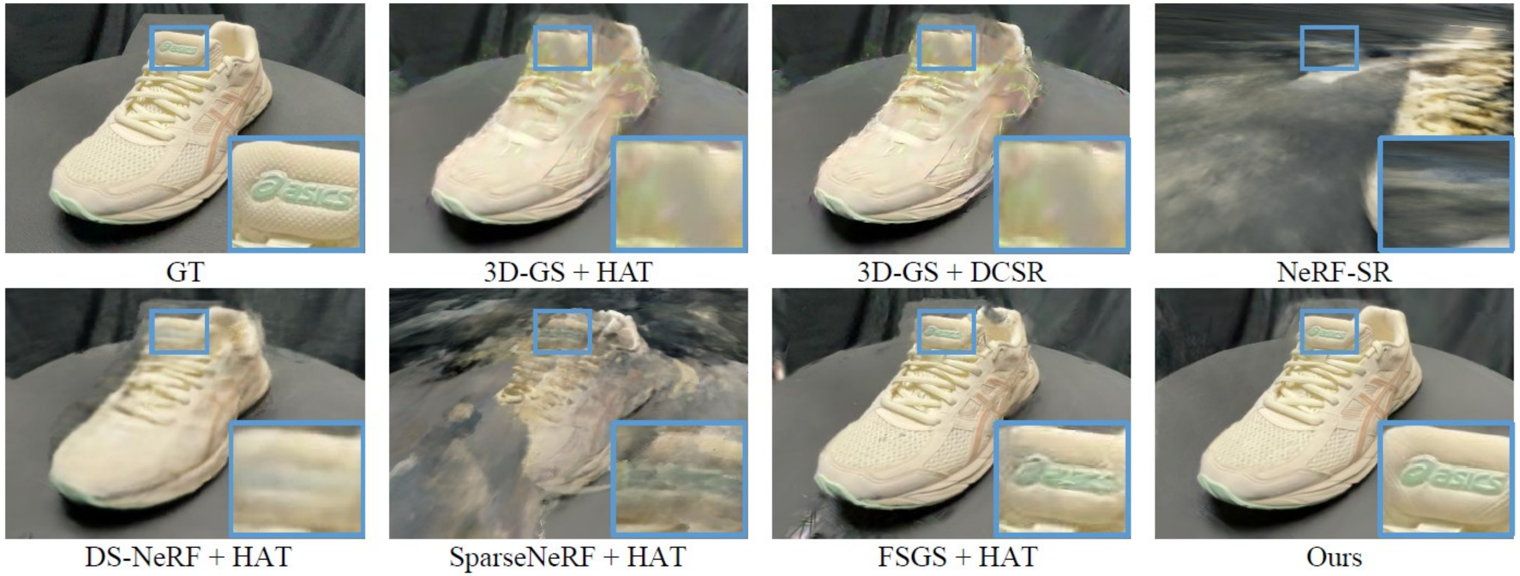
中国科学技术大学
University of Science and Technology of China

Quantitative comparisons on Real-captured Data

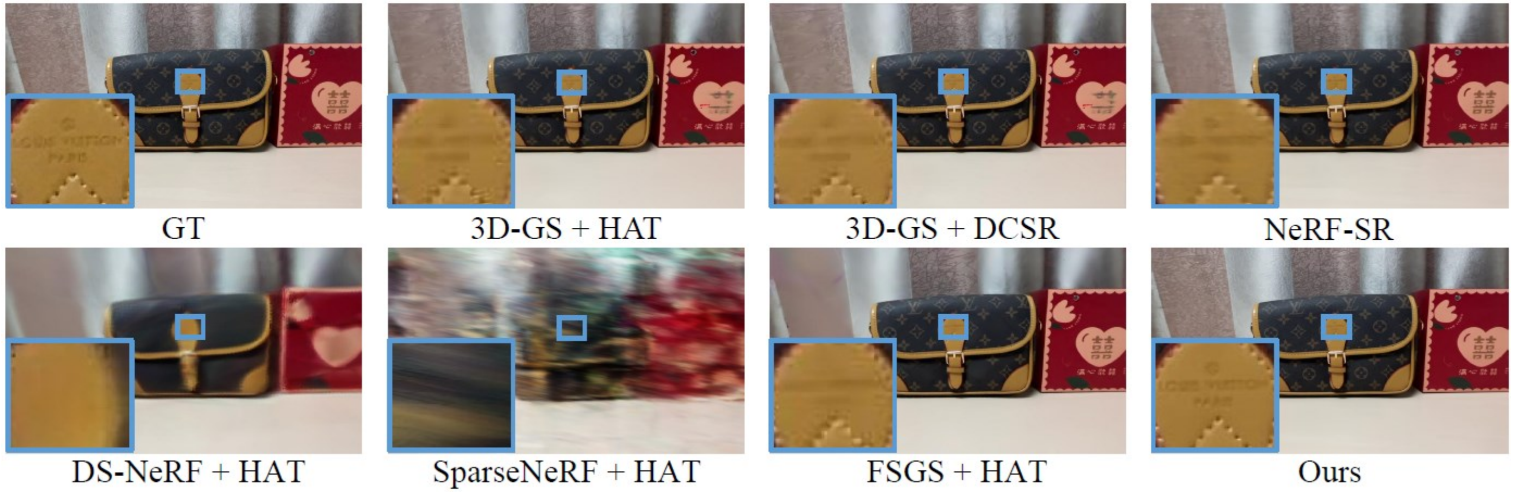
Method	Forward-facing						Inward-facing (360°)					
	5-shot			50-shot			15-shot			50-shot		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
3D-GS [19] + Bicubic	20.87	0.7064	0.3690	28.38	0.8467	0.2769	21.00	0.7036	0.4570	29.15	0.8278	0.3625
3D-GS [19] + SwinIR [25]	20.94	0.7088	0.3681	28.46	0.8485	0.2750	21.02	0.7043	0.4479	29.19	0.8280	0.3609
3D-GS [19] + HAT [9]	20.91	0.7086	0.3670	28.52	0.8492	0.2763	21.01	0.7050	0.4477	29.21	0.8283	0.3611
3D-GS [19] + DCSR [51]	21.03	0.7099	0.3636	28.29	0.8421	0.2674	21.00	0.7051	0.4540	29.05	0.8248	0.3646
NeRF-SR [49]	12.53	0.5389	0.5606	<u>30.53</u>	<u>0.8687</u>	**0.2372**	15.80	0.6300	0.5638	<u>29.76</u>	<u>0.8419</u>	<u>0.3479</u>
DS-NeRF [13] + HAT [9]	19.18	0.7019	0.4516	27.02	0.7951	0.3594	22.38	0.7307	0.4685	26.81	0.7711	0.4502
RegNeRF [33] + HAT [9]	22.39	0.7075	0.3679	24.31	0.7541	0.4160	20.20	0.6958	0.4949	23.55	0.7321	0.4829
SparseNeRF [50] + HAT [9]	22.98	0.7127	0.3787	24.57	0.7641	0.4091	20.31	0.7048	0.4767	23.72	0.7497	0.4794
FSGS [65] + HAT [9]	<u>23.08</u>	<u>0.7322</u>	<u>0.3595</u>	29.90	0.8319	0.2996	<u>22.96</u>	<u>0.7461</u>	<u>0.4415</u>	27.77	0.8023	0.4038
Ours	**24.05**	**0.7525**	**0.3249**	**31.28**	**0.8823**	<u>0.2435</u>	**24.07**	**0.7601**	**0.4172**	**30.72**	**0.8597**	**0.3224**

- We collect a set of dual-lens image pairs, captured from different viewpoints of static scenes by an off-the-shelf smartphone (i.e., iPhone12).
- DL-GS shows superior performance over the previous methods in most cases, which verifies the effectiveness of our method on the real-captured data.

Qualitative comparisons on Real-captured Data



15-shot on inward-facing scene.



5-shot on forward-facing scene.

Conclusion

- *New 3D-GS-based solution for HR and few-shot views synthesis by leveraging the characteristics of the asymmetric dual-lens system.*
- *Consistency-aware training strategy to exploit the geometric information of dual-lens pairs for regularizing optimization.*
- *Multi-reference-guided refinement module to enhance newly synthesized views by making the best use of dual-lens training samples.*
- *Effective on simulated and real-captured experiments.*