GameIR: A Large-Scale Synthesized Ground-Truth Dataset for Image Restoration over Gaming Content

ACCV HANOI VIETNAM 2024 DEC 8-12

Lebin Zhou[1,2], Kun Han[2], Nam Ling[1], Wei Wang[2], Wei Jiang[2]
[1] Santa Clara University, Santa Clara, CA, [2] Futurewei Technologies Inc. , San Jose, CA

Introduction

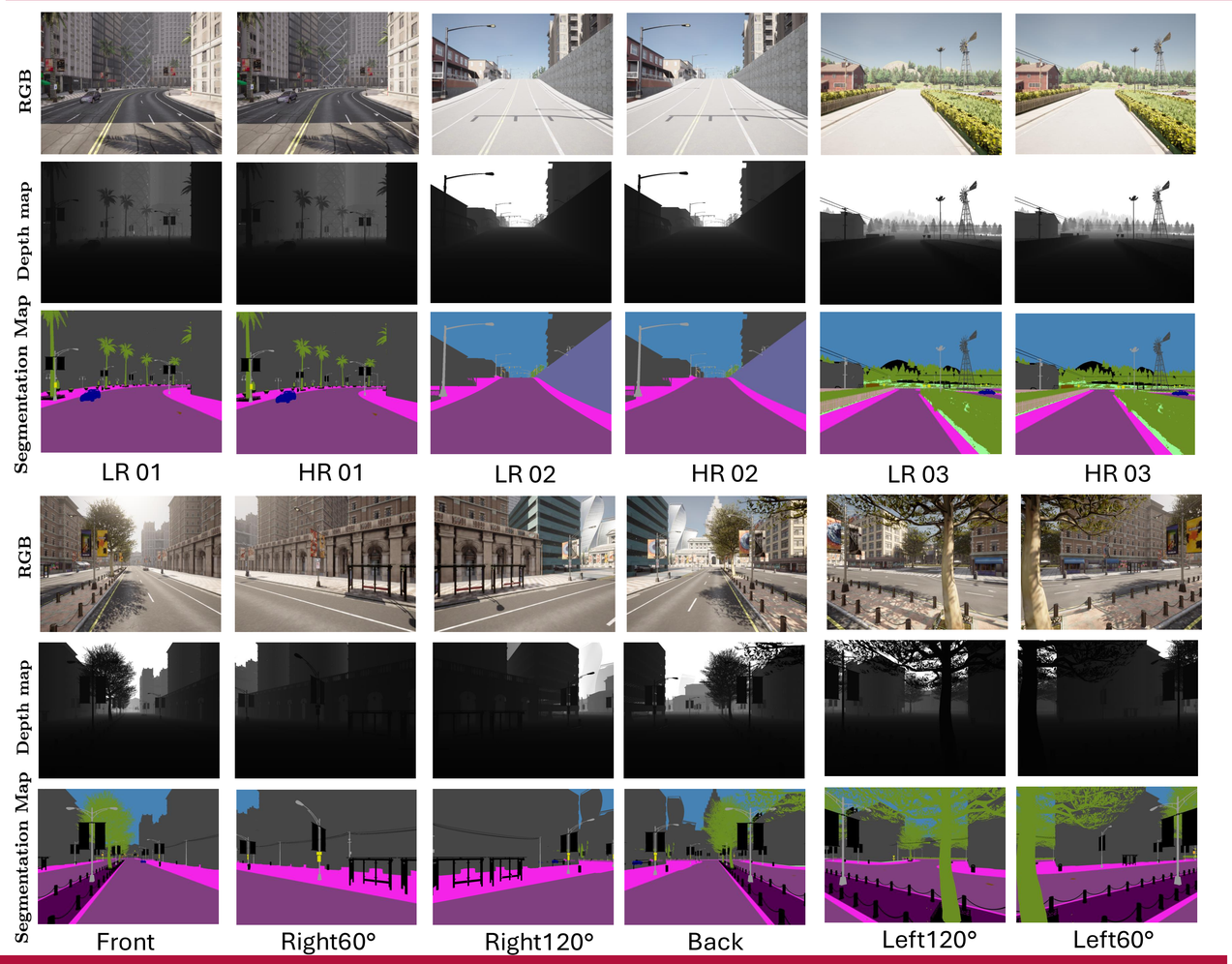
Motivation

- Current SR research relies on pseudo training data. The pseudo LR frames often contain noise and blur.
- Obtaining real LR-HR or multi-view paired data is significant challenges for research and applications.

Contributions

- Introduced a large-scale, high-quality, computer-generated ground-truth dataset tailored for 2 applications: SR and NVS.
- Evaluated multiple SOTA algorithms on the proposed dataset and enhanced them by incorporating GBuffers as additional inputs, establishing a baseline for assessing the performance.

GameIR-SR & GameIR-NVS



Evaluation of Super-Resolution

Performance without Gbuffer

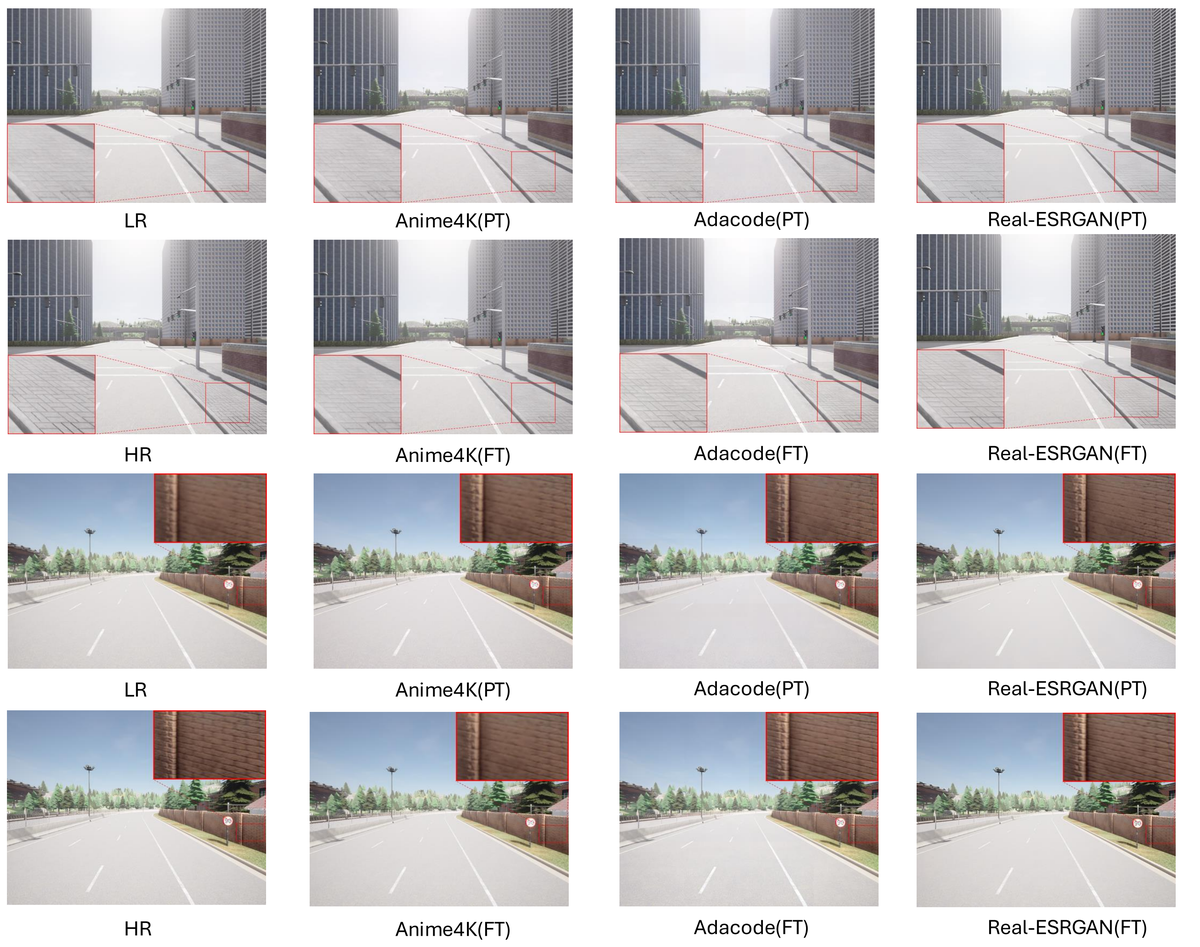
	PSNR↑			SSIM↑			FID↓			LPIPS↓		
	PT	FT pseudo	FT (GT)	PT	FT pseudo	FT (GT)	PT	FT pseudo	FT (GT)	PT	FT pseudo	FT (GT)
Anime4K	30.9274	30.5923	**31.2974**	0.9008	0.8874	**0.9057**	4.7365	6.7408	**4.6172**	0.0866	0.1069	**0.0788**
Adacode	28.6791	27.9637	**29.5762**	0.8382	0.8146	**0.8741**	14.8789	17.6222	**11.4608**	0.0884	0.0899	**0.0451**
Real-ESRGAN	29.1421	29.2743	**30.2517**	0.8639	0.8623	**0.8916**	17.7232	14.7961	**8.3153**	0.0905	0.0706	**0.0379**

Performance of pretrained (PT) and finetuned (FT) super-resolution algorithm with pseudo low-resolution (LR) data or ground-truth (GT) LR data

Performance with Additional Gbuffer

Input Channels	PSNR↑	SSIM↑	FID↓	LPIPS↓
RGB pretrained models	30.1798	0.9040	10.0374	0.0535
w/ S(concat)	30.1065	0.9026	9.9637	**0.0518**
w/ D(concat)	30.2236	0.9038	10.2756	**0.0499**
w/ SD(concat)	30.1639	0.9033	9.9164	**0.0524**
w/ S(SFT)	30.0750	0.9025	10.2182	**0.0519**
w/ D(SFT)	30.1049	0.9020	9.5882	**0.0492**
w/ SD(SFT)	30.1996	0.9052	9.8511	**0.0507**

Performance of using different GBuffer : segmentation map (S), depth map (D), or both (SD), by different methods: as additional input (concat), or as guidance condition (SFT).



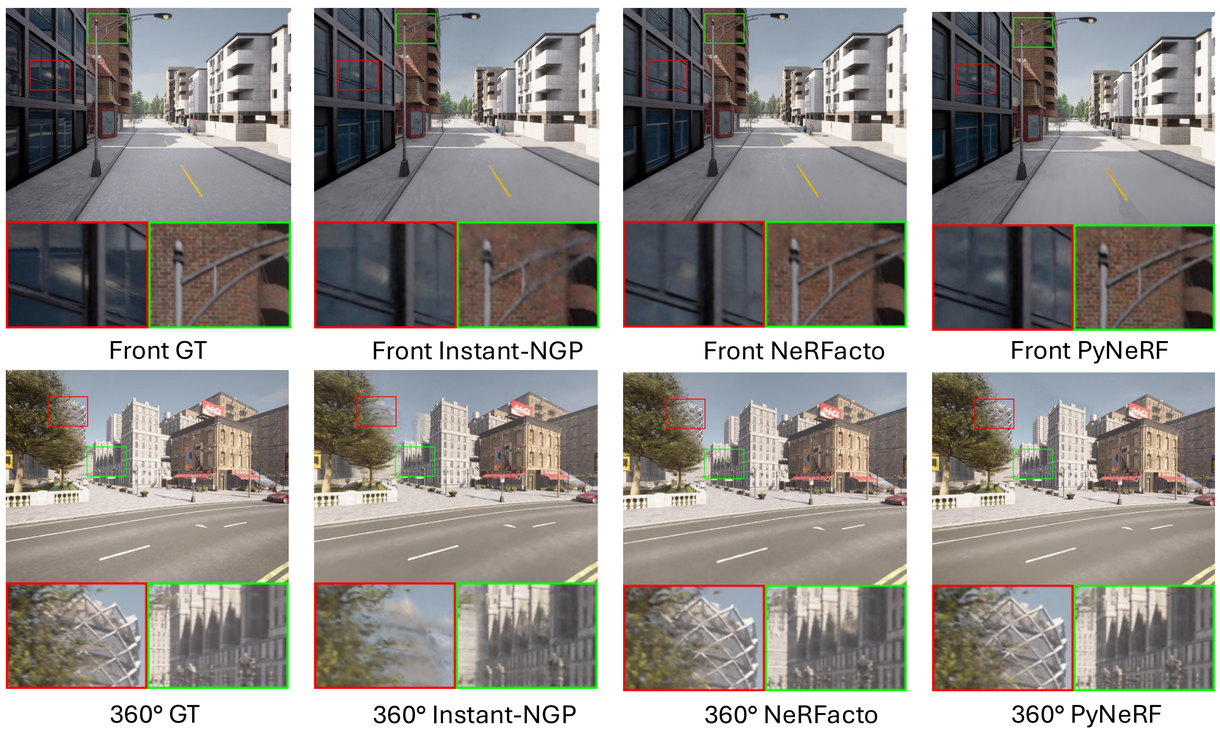
Qualitative comparison of pretrained (PT) and finetuned (FT) super-resolution methods.

Evaluation of Novel View Synthesis

Performance without Gbuffer

	PSNR↑		SSIM↑		FID↓		LPIPS↓	
	front	360°	front	360°	front	360°	front	360°
Instant-NGP	30.9481	27.7004	0.9373	0.9089	30.8528	44.7195	0.1095	0.1703
NeRFacto	26.6457	26.7710	0.8539	0.8774	36.1718	43.0672	0.1206	0.1648
PyNeRF	**36.1329**	**32.0542**	**0.9612**	**0.9357**	**20.2392**	**36.3312**	**0.0758**	**0.1301**

Performance of different NVS methods, using front views or 360° views training data.



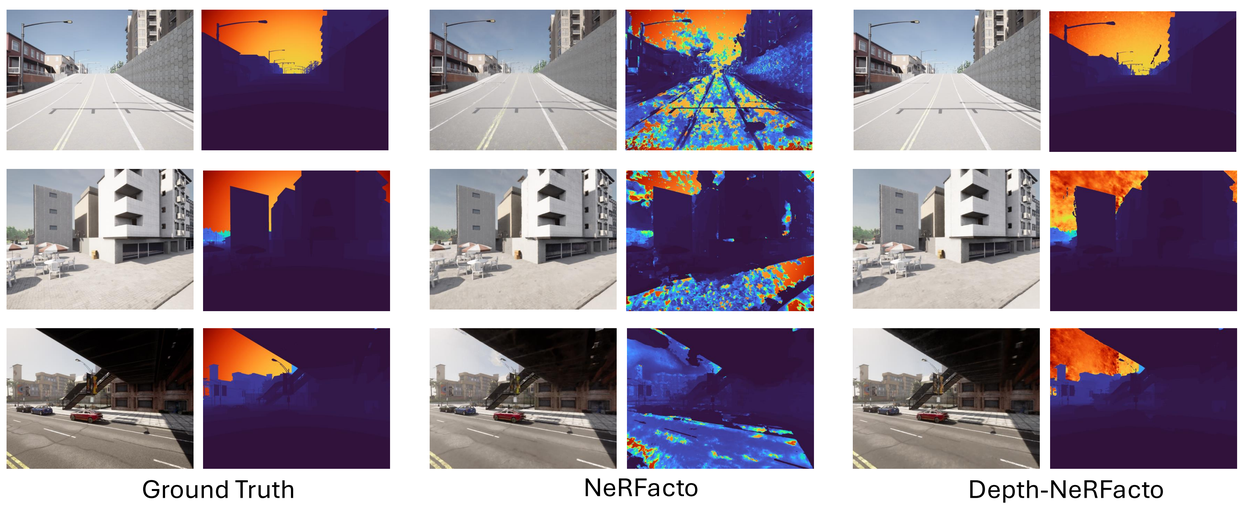
Qualitative comparison of 3 NVS methods on both Front and 360° views.

Performance with Additional Gbuffer

$$Depth\ Error = \frac{\sqrt{\frac{1}{M}\sum_{j=1}^{M}(I_j - \hat{I}_j)^2}}{\max(I) - \min(I)}$$

	PSNR↑		SSIM↑		FID↓		LPIPS↓		Depth Error %	
	front	360°	front	360°	front	360°	front	360°	front	360°
NeRFacto	22.9597	23.3168	0.8133	0.8276	60.3013	40.8520	0.1654	0.1830	32.8053	30.5997
Depth-NeRFacto	**25.3482**	**24.4990**	**0.8425**	**0.8306**	**29.3219**	**28.7427**	**0.1249**	**0.1736**	**5.4212**	**7.2538**

Performance of NeRFacto and Depth-NeRFacto, using 10% training views.



Qualitative comparison of NeRFacto and Depth-NeRFacto. The RGB (left) and depth map (right) generated by Depth-NeRFacto better match the ground truth compared to NeRFacto.