NeurIPS #443

Dataset Resolution Avg. number of 3D Gaussians	Argoverse 2 1550 × 2048 8.02M	Waymo Open 640 × 960 2.75M	Mean (ms)	Percentage (%)
1. Scene graph evaluation: retrieve ω, $[\mathbf{R}\|\mathbf{t}]$, 3D Gaussians at (s, t)	38.5	13.0	25.75	52.4
2. Scene composition: apply $[\mathbf{R}\|\mathbf{t}]$ to 3D Gaussians	2.3	1.5	1.90	3.9
3. 3D Gaussian projection	2.0	3.5	2.75	5.6
4. Query neural fields ϕ and ψ	9.5	3.6	6.55	13.3
5. Rasterization	21.3	3.0	12.15	24.8
Total	73.6	24.6	49.1	100
FPS	13.6	40.7	20.4	-

Table 1. **Inference runtime analysis**. We divide our algorithm into its main components (left), report the runtimes on each dataset (middle), and the average across datasets (right). We observe that the smaller-scale scenes with lower-resolution images in Waymo Open render significantly faster than the high-resolution images of larger-scale scenes in Argoverse 2. On average, we observe that steps 1 and 5 dominate the runtime of our method, owing to the complexity of rasterizing millions of 3D Gaussians across a high-resolution image and handling hundreds to thousands of dynamic objects across one or multiple dynamic captures. Still, our method achieves interactive rendering speeds on both datasets and 20.4 FPS on average.

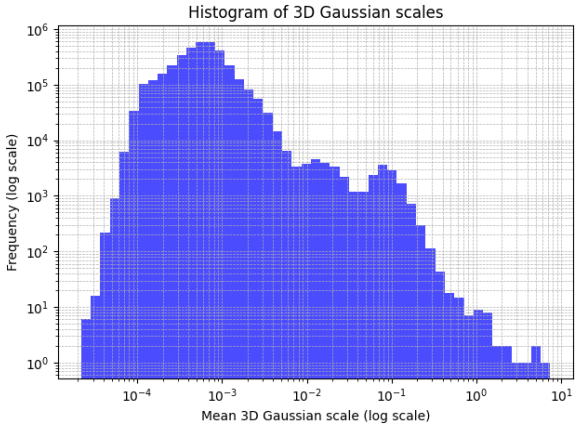

Figure 1. **Histogram of mean scale per 3D Gaussian of our model trained on Argoverse 2 residential.** Note that both axes in this plot are in logarithmic scale. The vast majority of 3D Gaussians have a small scale of less than 10^{-3}, while there are a few outliers with scales exceeding 10^{0}. The scene is approximately within [-1, 1] not including background and sky regions.

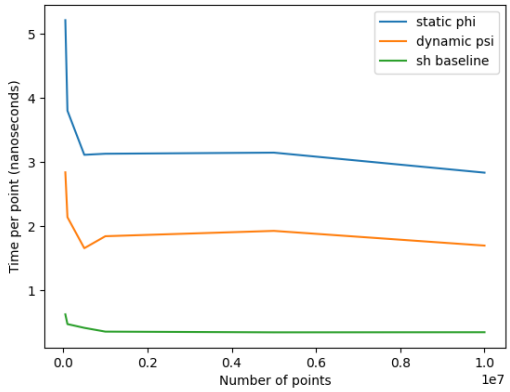

Figure 2. **Runtime comparison of neural fields vs. spherical harmonics.** We compare the runtime of querying neural fields of different sizes (equivalent to the sizes of ϕ and ψ in our paper) versus querying a spherical harmonics function of degree 3. We report time-per-point in nanoseconds. While querying the neural fields is slower, we note that **this does not lead to a critical increase in overall runtime as shown in Tab. 1**.

		PSNR ↑	SSIM ↑	LPIPS ↓
KITTI MOT	SUDS† [2, 3]	23.12	0.821	0.135
	MARS [3]	24.00	0.801	0.164
	NeuRAD [1]	27.00	0.795	0.082
	NeuRAD-2x [1]	<u>27.91</u>	<u>0.822</u>	<u>0.066</u>
	4DGF (Ours)	**30.01**	**0.913**	**0.052**

Table 2. **Additional comparison.** We follow the experimental protocol in [1] to compare with NeuRAD [1]. †baseline from [3].

References

[1] Adam Tonderski, Carl Lindström, Georg Hess, William Ljungbergh, Lennart Svensson, and Christoffer Petersson. Neurad: Neural rendering for autonomous driving. In *CVPR*, 2024. 1

[2] Haithem Turki, Jason Y Zhang, Francesco Ferroni, and Deva Ramanan. Suds: Scalable urban dynamic scenes. In *CVPR*, 2023. 1

[3] Zirui Wu, Tianyu Liu, Liyi Luo, Zhide Zhong, Jianteng Chen, Hongmin Xiao, Chao Hou, Haozhe Lou, Yuantao Chen, Runyi Yang, et al. Mars: An instance-aware, modular and realistic simulator for autonomous driving. In *CAAI International Conference on Artificial Intelligence*, pages 3–15. Springer, 2023. 1