Supplementary Material for: Gaussian Activated Neural Radiance Fields for High Fidelity Reconstruction & Pose Estimation

Shin-Fang Chng, Sameera Ramasinghe, Jamie Sherrah, and Simon Lucey

Australian Institute for Machine Learning
University of Adelaide
{shinfang.chng,sameera.ramasinghe,jamie.sherrah,simon.lucey}@adelaide.edu.au

A Additional Results

A.1 Hyperparameters for coarse-to-fine scheduling

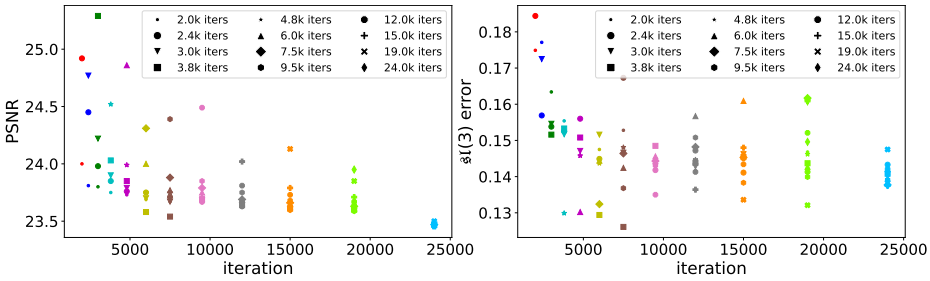

Fig. 1. We exhaustively sampled through the log-space for the optimal coarse-to-fine scheduler (t_{start} and t_{end}) for BARF [12] in Sec. 4.1. x-axis corresponds to different hyperparameters for t_{end} and each marker symbol corresponds to different hyperparameters for t_{start}. *Left*: PSNR. *Right*: $\mathfrak{sl}(3)$ error. These hyperparameters had failed to find the correct warping parameter for all patches.

Lin *et. al* [12] had proposed a coarse-to-fine scheduler, where each k^{th} frequency band of γ in Eq. (5) in main text (Sec. 3.2) is weighed as

$$\gamma_k(\mathbf{x}) = \omega_k(\alpha).[\sin(2^k\mathbf{x}), \cos(2^k\mathbf{x})], \tag{1}$$

where the weight ω_k is defined as

$$\omega_k(\alpha) = \frac{(1 - \cos(\text{clamp}(\alpha - k, 0, 1))}{2}, \tag{2}$$

and $\alpha(t)$ is computed as

$$\alpha(t) = \frac{t - t_{start}}{t_{end} - t_{start}} D, \tag{3}$$

where t is the current training iteration, and t_{start} and t_{end} are hyperparameters for when α should start enabling the positional encoding and reach the maximum number of frequencies D.

A.2 Ablation Study on the Network Architectures

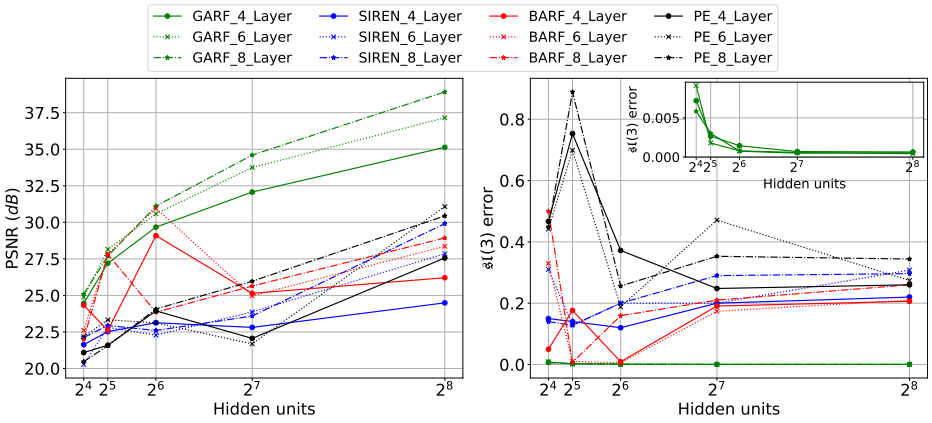

Fig. 2. We analyzed the performance of PE-MLP, BARF, SIREN and GARF on the 2D planar alignment instance in Sec. 4.1 when varying the number of hidden layers and hidden units of the MLP architecture. For all the different network architectures, PE-MLP, BARF and SIREN struggled to accurately estimate the camera pose. GARF was able to converge to the optimal pose, with increasingly better neural reconstruction when the number of hidden layers and the hidden units increase.

A.3 Quantitative and Qualitative Comparison of GARF and PE-MLP(without coarse-to-fine scheduler) on Real-World scenes [17]

Table 1. Quantitative comparison of GARF (Ours) and PE-MLP on real-world scenes [17] given *unknown* camera poses.

Scene	**Pose accuracy**				**View synthesis**					
	Rotation (°)		Translation (10^{-2})		PSNR ↑		SSIM ↑		LPIPS ↓	
	PE-MLP	Ours	PE-MLP	Ours	PE-MLP	Ours	PE-MLP	Ours	PE-MLP	Ours
flower	7.71	**0.46**	2.91	**0.22**	15.89	**26.40**	0.30	**0.79**	0.54	**0.11**
fern	0.84	**0.47**	1.38	**0.25**	19.82	**24.51**	0.59	**0.74**	0.40	**0.29**
leaves	63.80	**0.13**	11.68	**0.23**	9.51	**19.72**	0.06	**0.61**	0.79	**0.27**
horns	34.17	**0.03**	25.84	**0.21**	9.28	**22.54**	0.22	**0.69**	0.80	**0.33**
trex	173.22	**0.66**	59.43	**0.48**	10.55	**22.86**	0.17	**0.80**	0.87	**0.19**
orchids	68.00	**0.43**	23.83	**0.41**	10.04	**19.37**	0.09	**0.57**	0.82	**0.26**
fortress	144.58	**0.03**	51.04	**0.27**	12.34	**29.09**	0.33	**0.82**	0.79	**0.15**
room	101.98	**0.42**	50.04	**0.32**	11.41	**31.90**	0.31	**0.94**	0.91	**0.13**

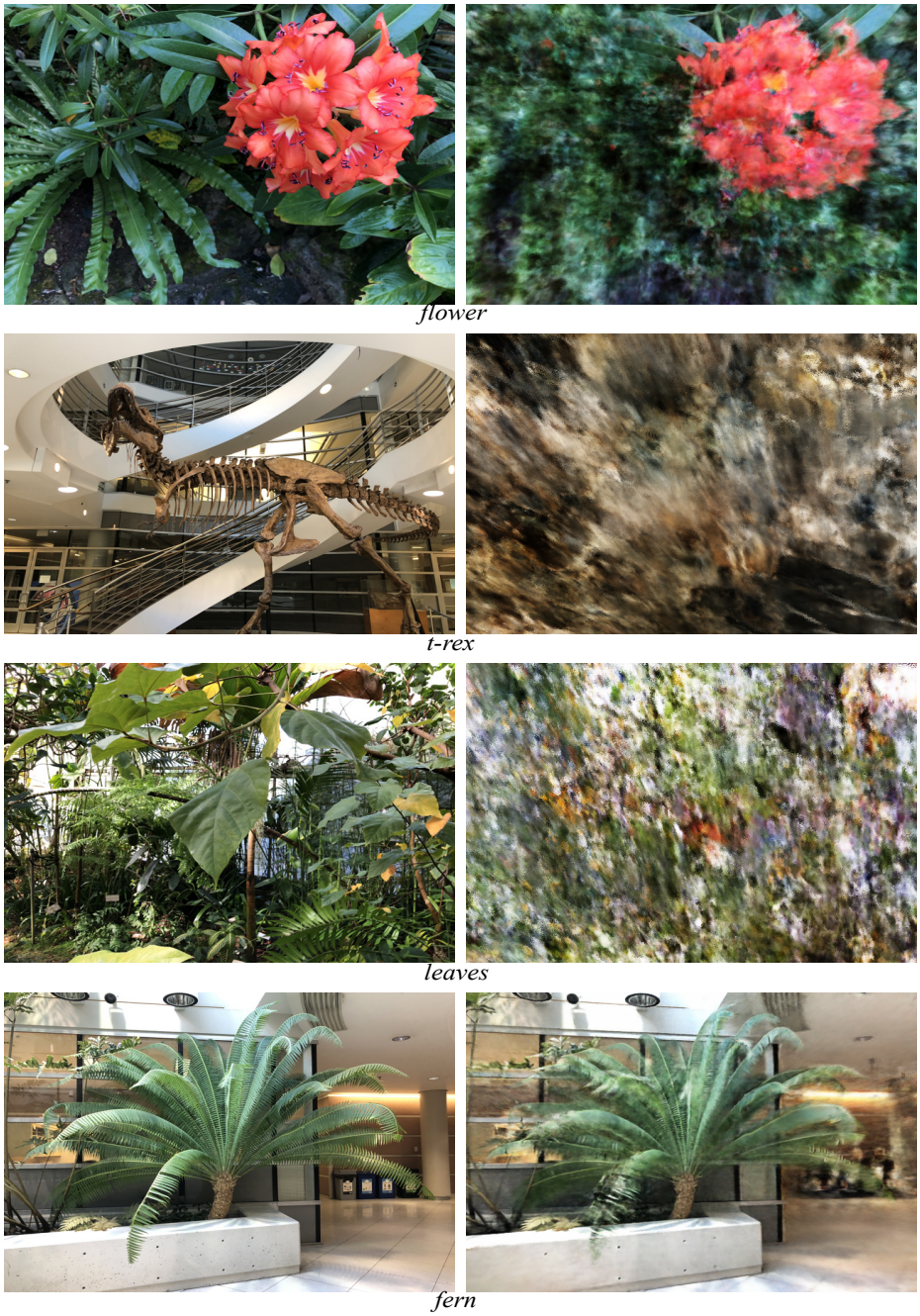

Fig. 3. Qualitative results on test-views of real-world scenes [17]. *Left*: Groundtruth. *Right*: PE-MLP(without coarse-to-fine scheduler).

A.4 Real World Demo

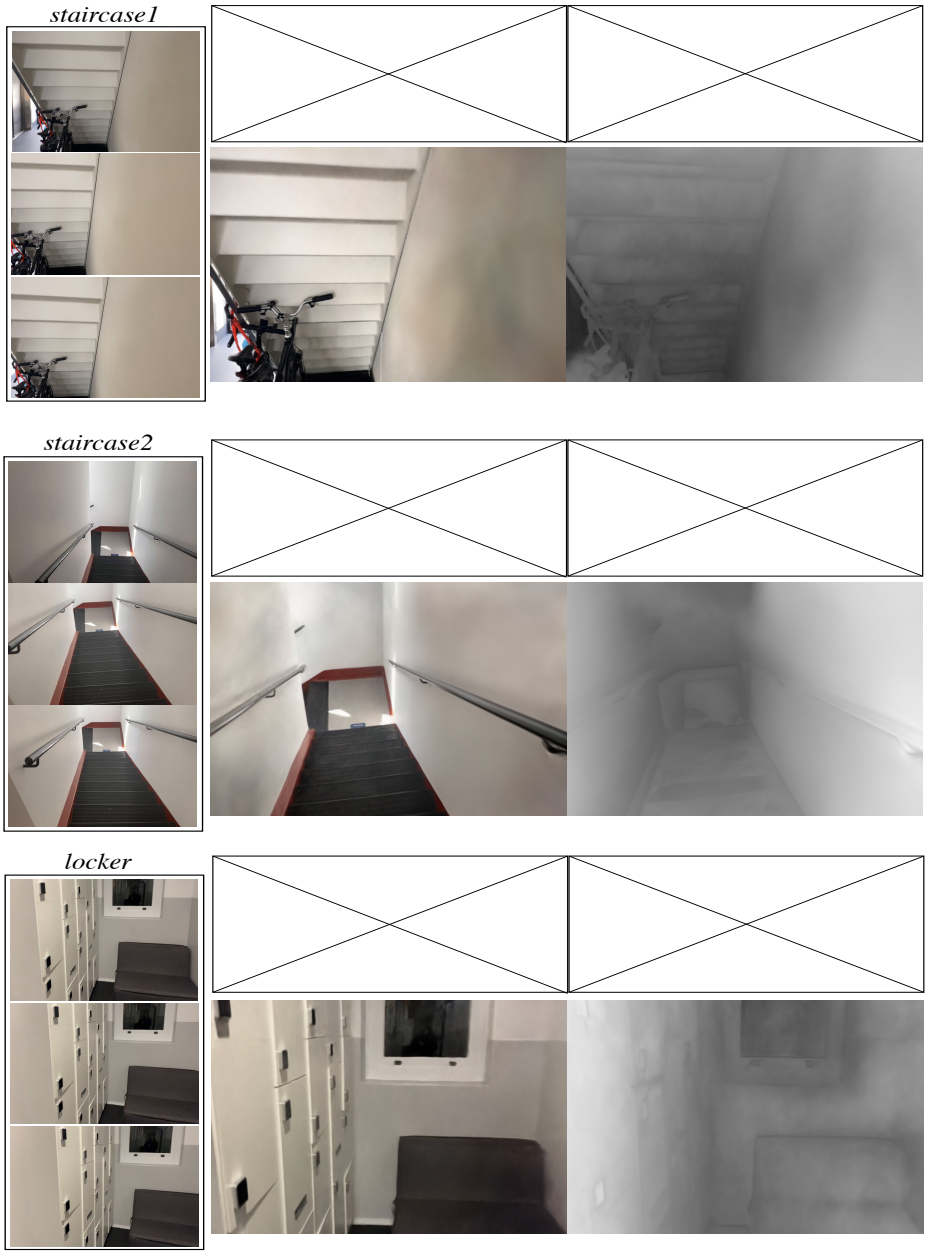

Fig. 4. Novel view synthesis on multiple low-textured scenes captured using iPhone. *Left*: Training images. *Top row*: No rendered image and depth from ref-NeRF as Colmap has failed to reconstruct the scene. *Bottom row*: Rendered image and depth using GARF.