TriNeRFLet: A Wavelet Based Triplane NeRF Representation
Supplementary Materials

Rajaei Khatib and Raja Giryes

Tel Aviv University
rajaeikhatib@mail.tau.ac.il and raja@tauex.tau.ac.il

1 Ablation

Wavelet type. For all the experiments we conducted in the paper, we applied the Biorthogonal 6.8 (Bior6.8) wavelet. We present here an ablation experiment that checks the impact of other wavelet filters. We compare the reconstruction performance of 4 scenes from the Blender dataset. The setups we tested are vanilla Triplane (no wavelet), Haar, Biorthogonal 2.2 (Bior2.2), Biorthogonal 2.6 (Bior2.6), Biorthogonal 4.4 (Bior4.4) and Biorthogonal 6.8 (Bior6.8). The results are presented in Figure 1. We use the TriNeRFLet Base Light setting. Vanilla Triplane has the same size and structure as this setting. Note that all wavelets achieve a significant performance advantage over vanilla Triplane (as shown also in the paper). Generally, higher order wavelets provide a better representation of smooth functions [3]. Thus, it is not surprising that Bior6.8, Bior4.4 and Bior 2.6 achieve better performance than Bior 2.2 and Haar. Note though that in terms of training time, training with Haar wavelet is faster by up to 30% compared to the other wavelets as Haar complexity is $O(N)$ and the complexity of applying the other wavelets is $O(N\log(N))$.
Coarse To Fine. As mentioned in the paper, coarse to fine (c2f) accelerates TriNeRFLet framework and is presented as a mitigation for the wavelet inverse transform overhead. To further check the additional contributions of c2f, we conduct the Blender reconstruction experiment (Base Light version) but without c2f. Results are presented in Table 2, and as noticed, c2f indeed brings a contribution in the reconstruction quality, alongside the contribution in training time.
Wavelet Regularization. To check the importance of the wavelet regularization loss, we extend the c2f experiment above to be without wavelet regularization. The results in Table 2 illustrate the important contribution of wavelet regularization and that it is an essential component of TriNeRFLet performance.

2 Technical Details

3D Reconstruction. Table 1 contains the parameters for the reconstruction versions.

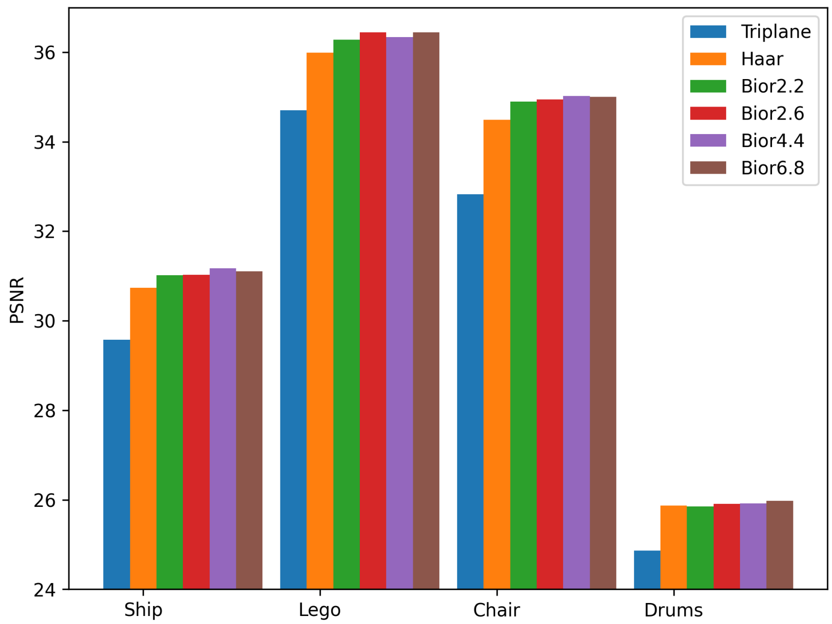

Fig. 1: Performance comparison between different wavelet filters and no wavelet at all (Triplane).

Blender Super-Resolution. In both Blender settings ($100 \rightarrow 400$ & $200 \rightarrow 800$), we use $C = 16$ (channels), $s_{LR} = 6000$ (LR only steps), $s = 16000$ (total steps), $s_{refresh} = 500$ (refresh steps), $T_{min} = 0.02$ and T_{max} is initialized with 0.98 and decreases linearly to reach 0.25 at the end. We use [1] as a base for implementing the super-resolution scheme, as it is more suitable for this task. We run the SR method on a single A6000 RTX GPU, and it runs for almost 7 hours.

LLFF Super-Resolution. We use similar settings to Blender $200 \rightarrow 800$ experiment except for, $C = 32$ (channels), $s = 20000$ (total steps), $s_{refresh} = 250$ (refresh steps). Like [4], we also operate in NDC space coordinates, as it fits this type of scenes better. Similar to the Blender experiment, we ran LLFF SR method on a single A6000 RTX GPU, where each scene runs almost for 12 hours.

3 Detailed Results

Tables 3,4 and 5 contain the detailed results of the SR experiments described in the paper.

	N_{LL}	L	N_{base}	N_{final}	C	γ	W	D_{dens}	D_{col}	train steps	trainable parameters
Small	64	4	512	1024	16	0.2	64	1	2	$6k$	$17M$
Base Light	64	5	512	2048	32	0.4	64	1	2	$10k$	$134M$
Base	64	5	512	2048	32	0.4	64	1	2	$43k$	$134M$
Large	64	5	512	2048	48	0.6	128	1	2	$83k$	$201M$

Table 1: The configurations used in the different TriNeRFLet versions

	Mic	Chair	Ship	Materials	Lego	Drums	Ficus	Hotdog	Avg.	Train Time
Base Light	35.77	35.00	31.10	29.35	36.44	25.98	33.96	36.93	33.07	1.5 hours
Base Light w/o c2f	35.52	35.08	31.11	29.14	36.48	25.76	33.46	36.88	32.93	2 hours
Base Light w/o c2f & wavelet reg.	34.14	33.33	30	28.17	35.02	25.01	30.76	35.95	31.55	2 hours
Triplane	33.85	32.83	29.58	28.15	34.70	24.86	30.35	35.80	31.26	2 hours

Table 2: Coarse to fine (c2f) and wavelet regularization importance ablation. We compare Base Light version with the versions without c2f and without c2f and wavelet regularization. The results indicate the important contribution of the c2f and wavelet regularization.

	Mic	Chair	Ship	Materials	Lego	Drums	Ficus	Hotdog	Avg.
PSNR	31.54	29.4	28.11	27.9	27.65	24.18	27.7	31.92	28.55
LPIPS	0.05	0.051	0.118	0.048	0.073	0.067	0.044	0.039	0.061
SSIM	0.914	0.946	0.844	0.921	0.914	0.896	0.936	0.939	0.913

Table 3: Blender 100 $\rightarrow$ 400 detailed results.

	Mic	Chair	Ship	Materials	Lego	Drums	Ficus	Hotdog	Avg.
PSNR	31.11	29.86	28.14	28.02	30.46	24.17	30.6	33.58	29.49
LPIPS	0.032	0.059	0.114	0.046	0.032	0.071	0.021	0.034	0.051
SSIM	0.966	0.944	0.844	0.93	0.922	0.913	0.96	0.96	0.93

Table 4: Blender 200 $\rightarrow$ 800 detailed results.

4 Qualitative Results

Figures 2 and 3 provide qualitative results for reconstruction and LLFF SR experiments respectively. For all results, the reader is referred to the project webpage `https://rajaeekh.github.io/trinerflet-web`. Furthermore, TriNeRFLet-SR can be incorporated into a generative model. To show that, we use the dif-

	Flower	Fern	Leaves	Fortress	Horns	Orchids	Room	Trex	Avg.
PSNR	26.6	24	20.01	28.88	24.93	21.38	28.56	25.57	25.00
LPIPS	0.214	0.205	0.239	0.175	0.26	0.244	0.125	0.163	0.203
SSIM	0.808	0.766	0.639	0.832	0.737	0.668	0.901	0.819	0.771

Table 5: LLFF $378 \times 504 \rightarrow 1512 \times 2016$ detailed results.

fusion process in [2] that generates a triplane and then apply SR to it using our scheme. Fig. 4 presents examples of triplanes rendered with resolution 128 that are upscaled to 512 by TriNeRFLet SR.

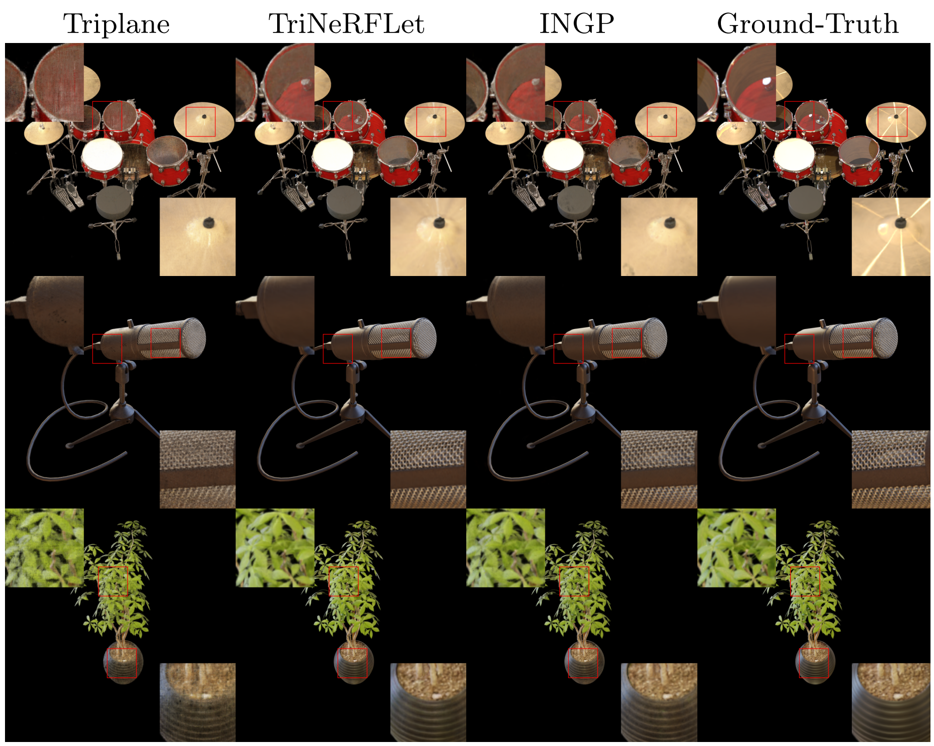

Fig. 2: *NeRF reconstruction qualitative results.* Notice the improvement in reconstruction quality of TriNeRFLet compared to Triplane. More visual results appear in the project page.

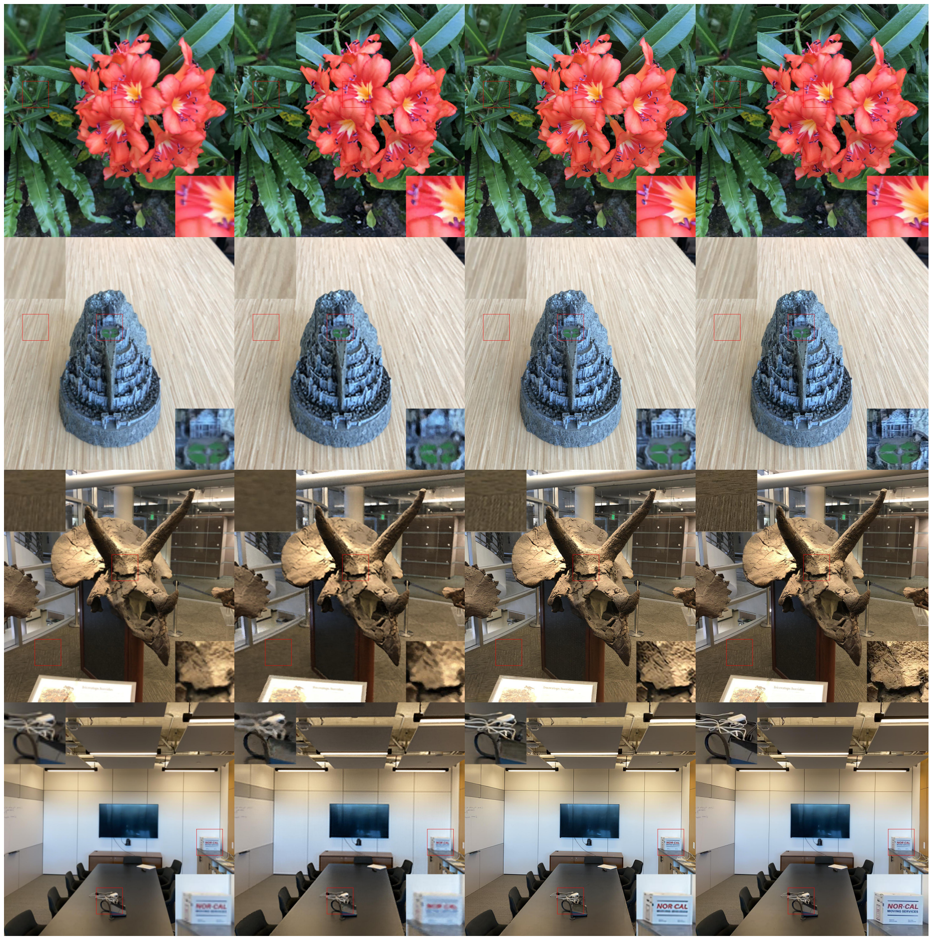

Fig. 3: *NeRF LLFF super-resolution qualitative results.*

Fig. 4: SR of low-res objects generated by triplane diffusion with text "Astronaut suit and helmet" and "Colorful electric scooter". First row is low-res and second row is TriNerFLet-SR.

References

1. Guo, Y.C., Liu, Y.T., Shao, R., Laforte, C., Voleti, V., Luo, G., Chen, C.H., Zou, Z.X., Wang, C., Cao, Y.P., Zhang, S.H.: threestudio: A unified framework for 3d content generation. `https://github.com/threestudio-project/threestudio` (2023)
2. Hong, F., Tang, J., Cao, Z., Shi, M., Wu, T., Chen, Z., Wang, T., Pan, L., Lin, D., Liu, Z.: 3dtopia: Large text-to-3d generation model with hybrid diffusion priors. arXiv preprint arXiv:2403.02234 (2024)
3. Mallat, S.: A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way. Academic Press, Inc. (2008)
4. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)