DyLiN: Making Light Field Networks Dynamic
Supplementary Material

Heng Yu[1] Joel Julin[1] Zoltán Á. Milacski[1] Koichiro Niinuma[2] László A. Jeni[1]
[1]Robotics Institute, Carnegie Mellon University [2]Fujitsu Research of America
{hengyu, jjulin, zmilacsk}@andrew.cmu.edu kniinuma@fujitsu.com laszlojeni@cmu.edu

1. Overview

In this supplementary material, we provide detailed quantitative and additional qualitative results, showcasing the benefits of our proposed DyLiN and CoDyLiN methods. Furthermore, we also provide the training times one should expect given our current setup.

2. Per-Scene Quantitative Results

For the sake of completeness, we provide the detailed per-scene quantitative results for reconstruction quality (PSNR, SSIM, MS-SSIM, LPIPS) on the synthetic (Tab. 5) and real (Tab. 6) dynamic scenes, extending Tab. 1 and Tab. 2 in the main paper that average these numbers across the scenes. Accordingly, we found that our DyLiN performs the best with respect to the SSIM and LPIPS metrics, generating perceptually better images, yet it sometimes falls behind in terms of PSNR and MS-SSIM that may prefer blurred results. Knowledge distillation improves a lot, our deformation and hyperspace MLPs yield slightly better results, while fine-tuning on the original training data gives a considerable boost.

3. More Qualitative Results

We provide additional qualitative results for 3 experiments.

First, Fig. 9 depicts more qualitative results for reconstruction quality on synthetic dynamic scenes, extending Fig. 6 in the main paper. Specifically, the Standup scene includes buttons on the shirt of the avatar (Fig. 9a), and the baselines are all missing them (Figs. 9b and 9c), whereas our full method is capable of reconstructing such details (Fig. 9e). Furthermore, the Bouncing Ball scene involves shadow casting (Fig. 9f). Inside the shadowed area, D-NeRF [29] produces horizontal artifacts (Fig. 9g), while TiNeuVox [8] predicts an inaccurate boundary (Fig. 9h). Again, our full model outputs the correct shadow (Fig. 9j).

Second, Fig. 10 shows qualitative results for ablation on the synthetic Standup scene using a D-NeRF teacher model, complementing Fig. 8 in the main paper that is restricted to real scenes and distilling from HyperNeRF [28]. D-NeRF gives an oversmoothed prediction (Fig. 10b), whereas the two MLPs of our DyLiN gradually reduce the blurriness of the face (Figs. 10c to 10e).

Lastly, Fig. 11 illustrates qualitative results for the real controllable Transformer scene, complementing the numbers of Tab. 4 in the main paper. We portray the effects of altering the attribute input $\alpha_i \in [-1, 1]$, which encodes the body pose of the character. We found that the CoNeRF [13] teacher model produces yellow color artifacts outside the boundary of the character (see, e.g., top row 1^{st} inset), whereas our CoDyLiN student model captures the boundary well.

4. Training Times

On a single NVIDIA A100 GPU, the full process takes $\approx$ 38–43 h, including 5–10 h to train the teacher, 13 h for drawing $S = 10{,}000$ training samples for KD, and 20 h for training the student via KD.

Table 5. Per-scene quantitative results on synthetic dynamic scenes. Notations: Multi-Layer Perceptron (MLP), PD (pointwise deformation), FT (fine-tuning). We utilized D-NeRF as the teacher model for our DyLiNs. The winning numbers are highlighted in bold.

	Hell Warrior			Mutant			Hook			Bouncing Balls		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF [23]	13.52	0.8100	0.2500	20.31	0.9100	0.0900	16.65	0.8400	0.1900	20.26	0.9100	0.2000
DirectVoxGo [33]	13.51	0.7500	0.2500	19.45	0.8900	0.1200	16.16	0.8000	0.2100	20.20	0.8700	0.2200
Plenoxels [9]	15.19	0.7800	0.2700	21.44	0.9100	0.0900	17.90	0.8100	0.2100	21.30	0.8900	0.1800
T-NeRF [29]	23.19	0.9300	0.0800	30.56	0.9600	0.0400	27.21	0.9400	0.0600	37.81	0.9800	0.1200
D-NeRF [29]	25.10	0.9500	0.0600	31.29	0.9700	0.0200	29.25	0.9600	0.1100	38.93	0.9800	0.1000
TiNeuVox-S [8]	27.00	0.9500	0.0900	31.09	0.9600	0.0500	29.30	0.9500	0.0700	39.05	0.9900	0.0600
TiNeuVox-B [8]	**28.17**	0.9700	0.0700	33.61	0.9800	0.0300	**31.45**	0.9700	0.0500	40.73	0.9900	0.0400
DyLiN, w/o two MLPs, w/o FT (ours)	26.81	0.9885	0.0363	32.13	0.9961	0.0186	29.89	0.9922	0.0297	39.78	0.9997	0.0099
DyLiN, w/o two MLPs (ours)	27.73	0.9893	0.0317	33.26	0.9971	0.0101	30.20	0.9928	0.0187	41.13	**0.9998**	0.0064
DyLiN, PD MLP only, w/o FT (ours)	26.82	0.9886	0.0362	32.13	0.9963	0.0185	29.94	0.9923	0.0296	39.70	0.9996	0.0096
DyLiN, PD MLP only (ours)	27.75	0.9896	0.0302	33.47	0.9972	0.0102	30.39	0.9930	**0.0186**	41.52	**0.9998**	**0.0062**
DyLiN, w/o FT (ours)	26.90	0.9887	0.0360	32.17	0.9963	0.0182	29.99	0.9923	0.0289	40.02	0.9997	0.0098
DyLiN (ours)	27.79	**0.9898**	**0.0298**	**33.80**	**0.9974**	**0.0086**	30.49	**0.9931**	**0.0186**	**41.59**	**0.9998**	**0.0062**

	Lego			T-Rex			Stand Up			Jumping Jacks		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF [23]	20.30	0.7900	0.2300	24.29	0.9300	0.1300	18.19	0.8900	0.1400	18.28	0.8800	0.2300
DirectVoxGo [33]	21.13	0.9000	0.1000	23.27	0.9200	0.0900	17.58	0.8600	0.1600	17.80	0.8400	0.2000
Plenoxels [9]	21.97	0.9000	0.1100	25.18	0.9300	0.0800	18.76	0.8700	0.1500	20.18	0.8600	0.1900
T-NeRF [29]	23.82	0.9000	0.1500	30.19	0.9600	0.1300	31.24	0.9700	0.0200	32.01	0.9700	0.0300
D-NeRF [29]	21.64	0.8300	0.1600	31.75	0.9700	0.0300	32.79	0.9800	0.0200	32.80	0.9800	0.0300
TiNeuVox-S [8]	24.35	0.8800	0.1300	29.95	0.9600	0.0600	32.89	0.9800	0.0300	32.33	0.9700	0.0400
TiNeuVox-B [8]	**25.02**	0.9200	0.0700	32.70	0.9800	0.0300	35.43	0.9900	0.0200	**34.23**	0.9800	0.0300
DyLiN, w/o two MLPs, w/o FT (ours)	22.11	0.9747	0.0612	31.35	0.9978	0.0290	33.98	0.9973	0.0140	33.24	0.9981	0.0260
DyLiN, w/o two MLPs (ours)	22.42	0.9761	0.0493	32.80	0.9984	0.0170	35.31	0.9980	0.0084	33.67	0.9984	0.0155
DyLiN, PD MLP only, w/o FT (ours)	22.13	0.9748	0.0618	32.18	0.9982	0.0282	33.97	0.9973	0.0140	33.19	0.9982	0.0257
DyLiN, PD MLP only (ours)	22.76	0.9775	0.0452	32.77	**0.9985**	0.0176	35.56	0.9981	0.0082	33.68	0.9984	0.0152
DyLiN, w/o FT (ours)	22.24	0.9754	0.0600	32.24	0.9982	0.0276	34.15	0.9974	0.0141	33.23	0.9983	0.0256
DyLiN (ours)	23.10	**0.9791**	**0.0443**	**32.91**	**0.9985**	**0.0168**	**35.95**	**0.9983**	**0.0074**	33.84	**0.9985**	**0.0151**

Table 6. Per-scene quantitative results on real dynamic scenes. Notations: Multi-Layer Perceptron (MLP), PD (pointwise deformation), FT (fine-tuning), N/A (not available in the cited research paper). We utilized HyperNeRF as the teacher model for our DyLiNs. The winning numbers are highlighted in bold.

	Broom		3D Printer		Chicken	
Method	PSNR↑	MS-SSIM↑	PSNR↑	MS-SSIM↑	PSNR↑	MS-SSIM↑
NeRF [23]	19.90	0.653	20.70	0.780	19.90	0.777
NV [19]	17.70	0.623	16.20	0.665	17.60	0.615
NSFF [16]	**26.10**	**0.871**	**27.70**	**0.947**	26.90	0.944
Nerfies [27]	19.20	0.567	20.60	0.830	26.70	0.943
HyperNeRF [28]	19.30	0.591	20.00	0.821	26.90	0.948
TiNeuVox-S [8]	21.90	0.707	22.70	0.836	27.00	0.929
TiNeuVox-B [8]	21.50	0.686	22.80	0.841	**28.30**	0.947
DyLiN, w/o two MLPs, w/o FT (ours)	21.98	0.808	22.99	0.899	26.89	0.948
DyLiN, w/o two MLPs (ours)	22.04	0.811	23.16	0.905	27.35	0.954
DyLiN, PD MLP only, w/o FT (ours)	22.02	0.805	23.04	0.903	26.88	0.948
DyLiN, PD MLP only (ours)	22.14	0.815	23.19	0.906	27.53	0.955
DyLiN, w/o FT (ours)	22.04	0.809	23.06	0.902	26.91	0.948
DyLiN (ours)	22.14	0.823	23.21	0.906	27.62	**0.956**
	Peel Banana		Americano		Expressions	
Method	PSNR↑	MS-SSIM↑	PSNR↑	MS-SSIM↑	PSNR↑	MS-SSIM↑
NeRF [23]	20.00	0.769	N/A	N/A	N/A	N/A
NV [19]	15.90	0.380	N/A	N/A	N/A	N/A
NSFF [16]	24.60	0.902	N/A	N/A	N/A	N/A
Nerfies [27]	22.40	0.872	N/A	N/A	N/A	N/A
HyperNeRF [28]	23.30	0.896	18.42	0.720	25.40	0.958
TiNeuVox-S [8]	22.10	0.780	N/A	N/A	N/A	N/A
TiNeuVox-B [8]	24.40	0.873	N/A	N/A	N/A	N/A
DyLiN, w/o two MLPs, w/o FT (ours)	23.38	0.872	18.45	0.722	25.36	0.950
DyLiN, w/o two MLPs (ours)	24.35	0.906	30.85	0.977	26.33	0.967
DyLiN, PD MLP only, w/o FT (ours)	23.70	0.882	18.47	0.722	25.55	0.960
DyLiN, PD MLP only (ours)	25.72	0.936	31.01	0.978	26.33	0.967
DyLiN, w/o FT (ours)	23.97	0.886	18.48	0.722	26.51	0.969
DyLiN (ours)	**27.36**	**0.952**	**31.56**	**0.982**	**26.91**	**0.974**

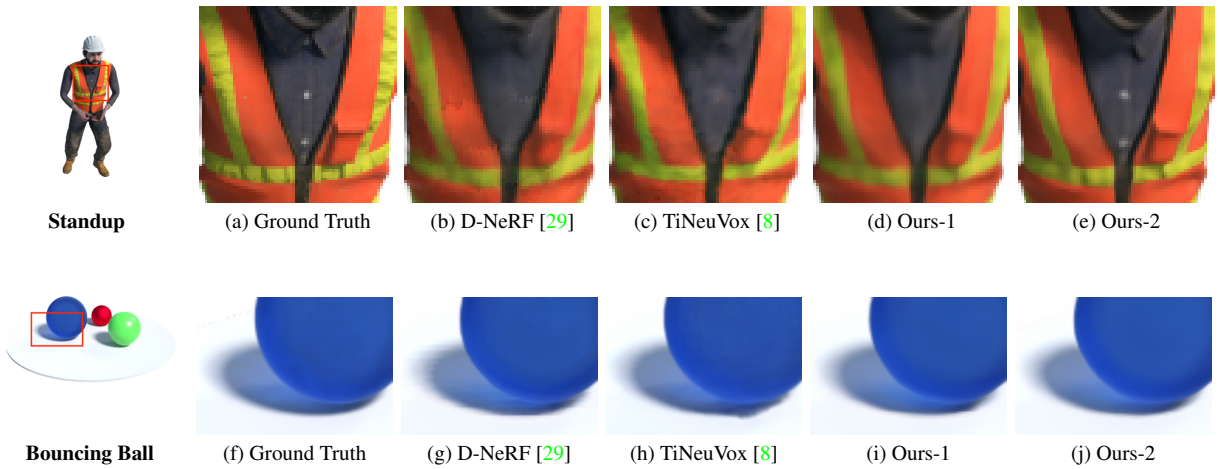

Figure 9. More qualitative results on synthetic dynamic scenes. We compare our DyLiN (Ours-1, Ours-2) with the ground truth, the D-NeRF teacher model and TiNeuVox. Ours-1 and Ours-2 were trained without and with fine-tuning on the original data, respectively.

Figure 10. Qualitative results for ablation on the synthetic Standup scene. We compare our DyLiN (Ours-1, Ours-2, Ours-3) with the ground truth and the D-NeRF teacher model. Ours-1 was trained without our two MLPs. Ours-2 was trained with pointwise deformation MLP only. Ours-3 is our full model with both of our proposed two MLPs.

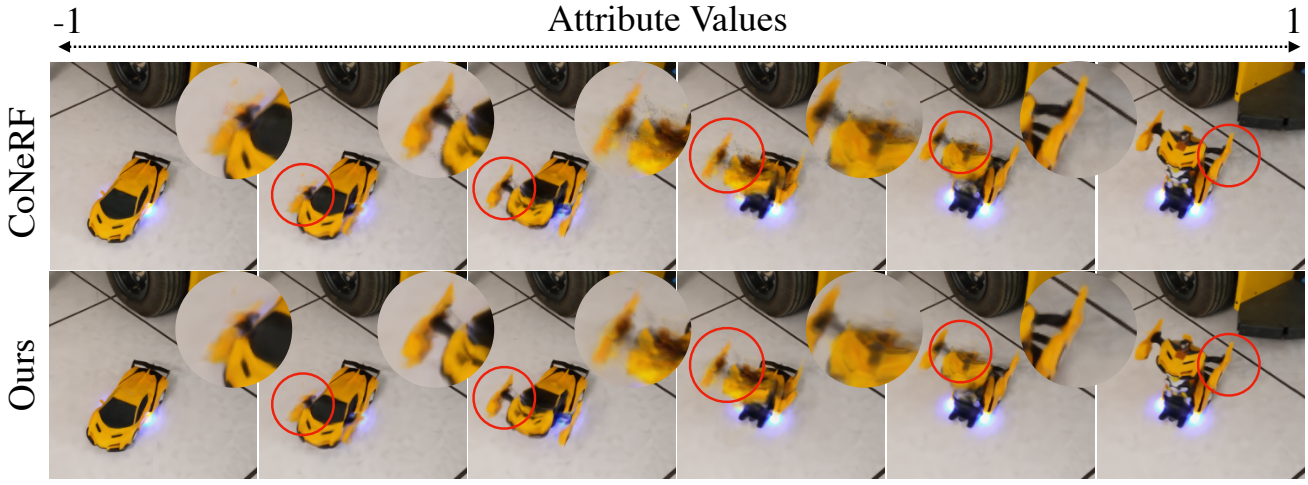

Figure 11. Qualitative results on the real controllable Transformer scene. We utilized CoNeRF [13] as the teacher model for our CoDyLiN. Red circles indicate regions enlarged in insets. Best viewed zoomed in.