Depth-Supervised Dynamic NeRF

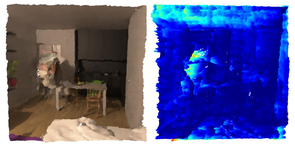
Technical University Munich
Chang Luo

ADL4CV
WiSe 22 / 23
01.02.2023

Introduction

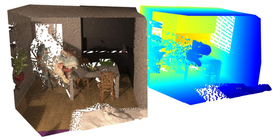
The novel view synthesis is under-constrained on dynamic scenes. The color outcome of the D-NeRF [1] is blurry in non-rigid region, while it can only output a very flat depth map.

Our work focus on introduction of depth towards more reliable color and depth prediction.



D-NeRF [1]

Range $[1.47\sim1.83]$ $\sigma^2 = 0.0021$



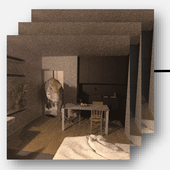
Groundtruth

Range $[1.70\sim8.39]$ $\sigma^2 = 1.5931$

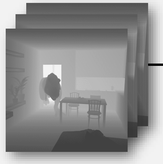
Contributions:

+ Depth Supervision with *MSE* & *GNLL* Loss function
+ *Depth Guided Sampling* Strategy
+ Benchmarking on both *Real and Synthetic* Datasets

Method



RGB Frames



Depth Frames

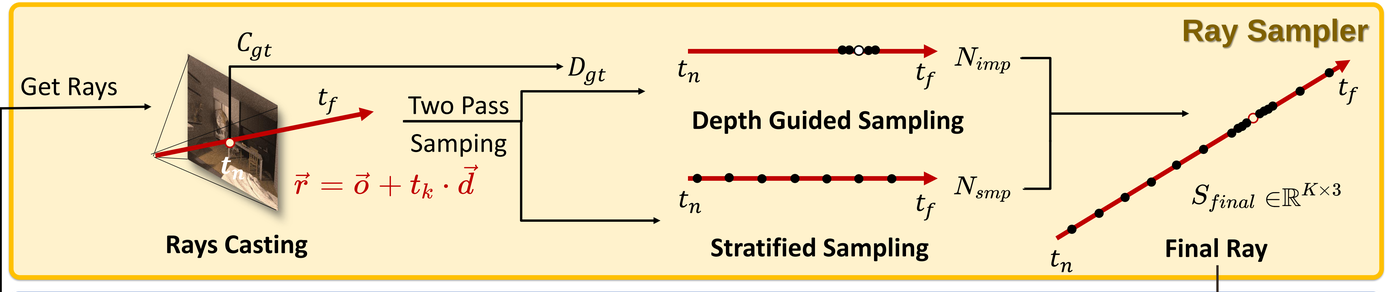
$\{P_i\}_{i=0}^{N-1} = [R_i|t_i]$

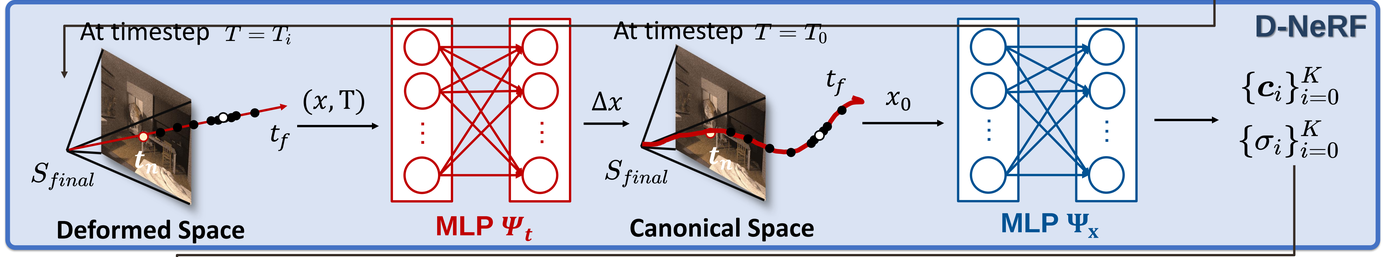
Camera Poses

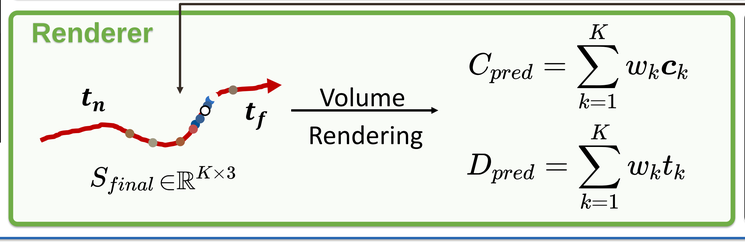
$\{T_i\}_{i=0}^{N-1}$

Time steps

Dataset





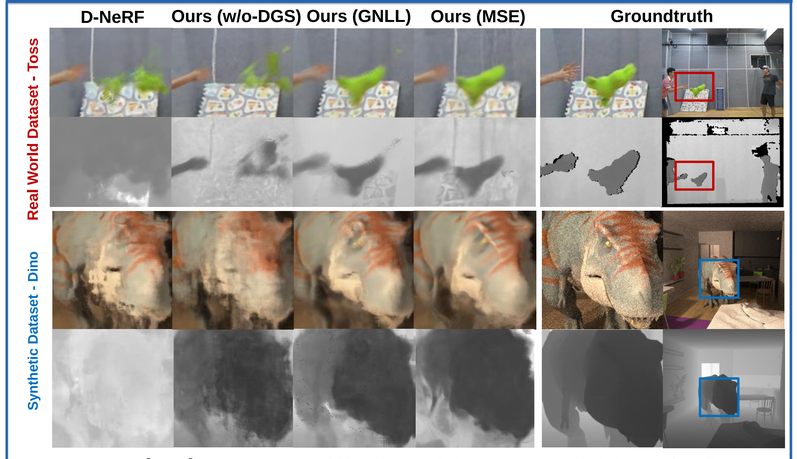


Loss function

$$\mathcal{L}_{color} = ||C_{pred} - C_{gt}||_2^2 \quad \mathcal{L}_{depth} = ||D_{pred} - D_{gt}||_2^2$$

$$\mathcal{L} = \mathcal{L}_{color} + \lambda \cdot \mathcal{L}_{depth}$$

Result



	Method	PSNR↑	SSIM↑	LPIPS↓	RMSE↓
Toss	D-NeRF [1]	20.31	0.703	0.315	2.348
	Ours (w/o DGS)	20.45	**0.715**	0.318	0.501
	Ours (w/ GNLL)	20.54	0.710	**0.313**	**0.415**
	Ours (w/ MSE)	**20.87**	0.703	0.335	0.468
Dino	D-NeRF [1]	22.13	0.330	0.692	1.298
	Ours (w/o DGS)	22.27	0.328	0.703	0.088
	Ours (w/ GNLL)	22.26	0.336	0.688	**0.022**
	Ours (w/ MSE)	**22.73**	**0.339**	**0.685**	0.094

Conclusion

We presented **DSD-NeRF** for taking advantage of depth information for better constraining the optimization with improved quality of both color and depth observed.

The decrease in depth error is especially significant, which helps one to obtain a more meaningful geometric structure in space.

Reference: [1] D-NeRF: Neural Radiance Fields for Dynamic Scenes, Pumarola et. al