Depth-Supervised Dynamic NeRF

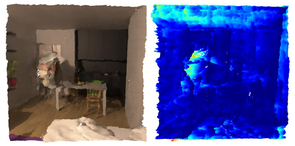
Technical University Munich
Chang Luo

ADL4CV
WiSe 22 / 23
01.02.2023

Introduction

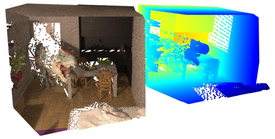
Novel view synthesis on dynamic scenes is under-constrained. The color output of D-NeRF [1] is blurry in non-rigid regions and the depth estimates reveal reconstruction failure.

Our method leverages depth information to improve novel view synthesis on dynamic scenes.



D-NeRF [1]

Depth $[1.47, 1.83]$ $\sigma^2 = 0.0021$



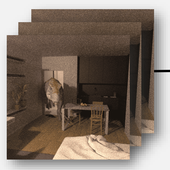
Groundtruth

Depth $[1.70, 8.39]$ $\sigma^2 = 1.5931$

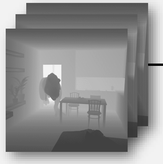
Contributions:

+ 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

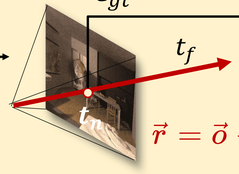
$\{T_i\}_{i=0}^{N-1}$

Time steps

Dataset

Ray Sampler

Get Rays

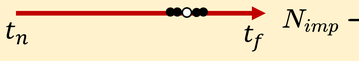


Rays Casting

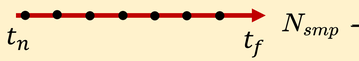
$\vec{r} = \vec{o} + t_k \cdot \vec{d}$

Two Pass Samping

D_{gt}

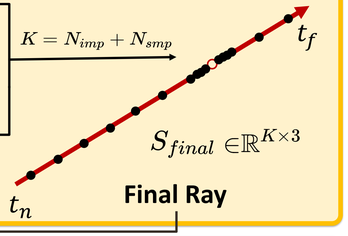


Depth Guided Sampling (DGS)



Stratified Sampling

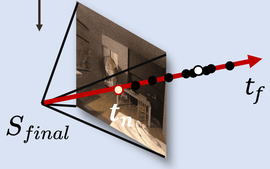
$K = N_{imp} + N_{smp}$



Final Ray

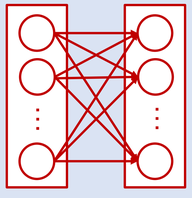
D-NeRF

At timestep $T = T_i$



Deformed Space

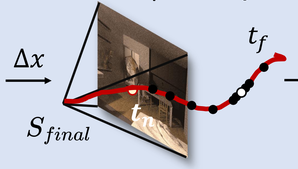
(x, T)



MLP Ψ_t

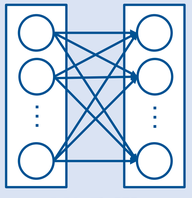
Δx

At timestep $T = T_0$



Canonical Space

x_0

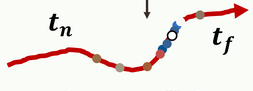


MLP Ψ_x

$\{\boldsymbol{c}_i\}_{i=0}^{K}$

$\{\sigma_i\}_{i=0}^{K}$

Renderer



Volume Rendering

$$C_{pred} = \sum_{k=1}^{K} w_k \boldsymbol{c}_k$$

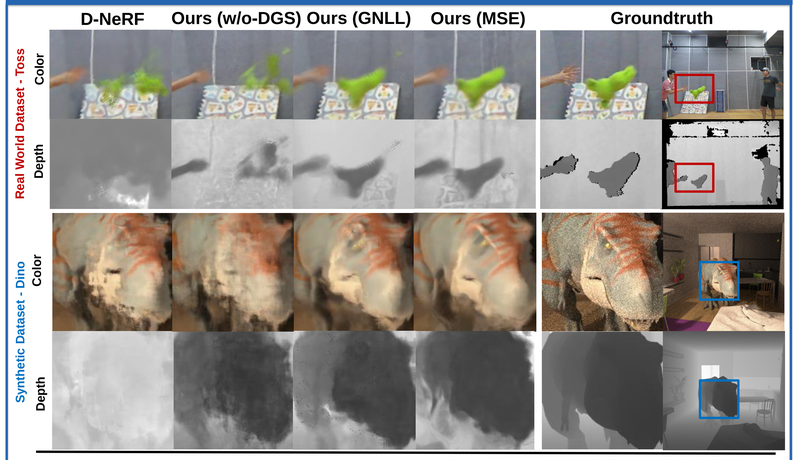
$$D_{pred} = \sum_{k=1}^{K} w_k t_k$$

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

Depth-Supervised Dynamic NeRF takes advantage of depth information to better constrain the optimization, which improves novel view synthesis in terms of color and depth.

The large decrease in depth error shows that our method better recovers the geometry of dynamic scenes.

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