LidaRF: Delving into Lidar for Neural Radiance Field on Street Scenes

Shanlin Sun[1], *Bingbing Zhuang*[3], *Ziyu Jiang*[3], *Buyu Liu*[3], *Xiaohui Xie*[1], *Manmohan Chandraker*[2,3]

[1]University of California, Irvine [2]University of California, San Diego [3]NEC Labs America

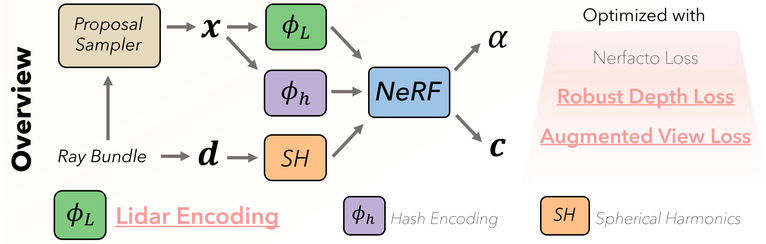
Introduction: Appearance Simulation

Input: Front RGB camera, Lidar, sensor poses

Goal: Photorealistic appearance simulation of street scenes

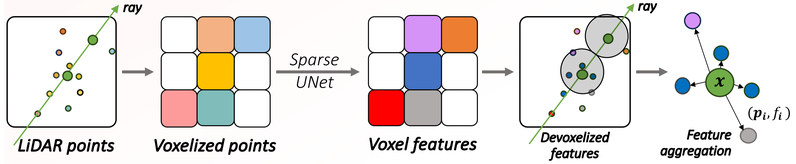
Challenge: Sparse view coverage, low-texture road surface …

Motivation: Modern autonomous systems are often equipped with Lidar. How can we use it more than just as a depth loss?



Contrib #1: Lidar Encoding

- Lidar holds strong potential for geometric guidance
- Lidar encoding through 3D sparse CNN has proven powerful in 3D detection, but is underexplored in NeRF
- Fuse Lidar encoding and hash grid feature



Contrib #2: Robust Depth Supervision

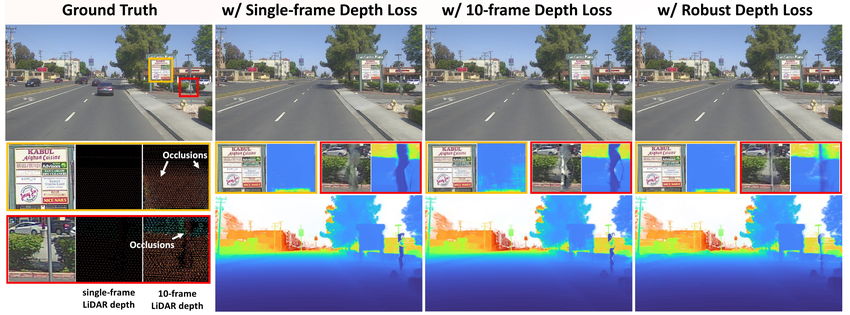
- Accumulate adjacent Lidar frames for denser supervision
- Inter-points occlusion due to the camera-Lidar displacement
- For samples in $\mathcal{D}^m_{\text{reliable}}$, we adopt the pixel-level depth loss

$$\mathcal{D}^m_{\text{reliable}} = \left\{ \mathcal{D}_i \mid \mathcal{D}_i \le \epsilon^m_t, \mathcal{D}_i \le \hat{\mathcal{D}}_i + \epsilon^m_o, \mathcal{D}_i \in \mathcal{D} \right\}$$

$$\epsilon^m_t = \min\{\alpha_t \epsilon^{m-1}_t, \epsilon_t\}, \quad \alpha_t > 1$$

$$\epsilon^m_o = \max\{\alpha_o \epsilon^{m-1}_o, \epsilon_o\}, \quad \alpha_o < 1$$

Intuition: count on near points initially and gradually add far points congruent along with NeRF training



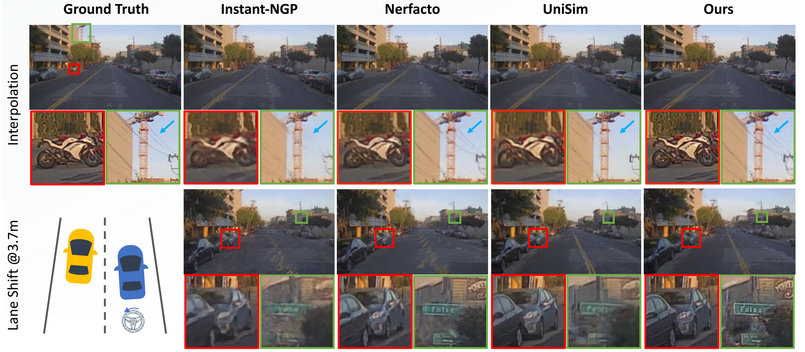
Contrib #3: Augmented View Supervision

- Project Lidar points to "render" more training views
- Handle occlusions using the same scheme as "contribution #2"

Future work: in-paint sparse projection with diffusion priors

Experiment: Comparisons with SOTA

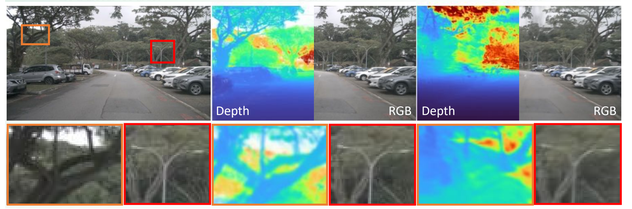
PandaSet Dataset:



Argoverse Dataset:

Methods	Interpolation			Lane Shift	
	PSNR↑	SSIM↑	LPIPS↓	FID↓ @ 2m	FID↓ @ 3.7m
Instant-NGP	24.282	0.733	0.408	140.3	173.2
Mip-NeRF 360	23.693	0.691	0.496	189.4	231.1
Nerfacto	27.122	0.804	0.268	116.7	151.0
UniSim	26.014	0.768	0.342	118.5	141.3
Ours	27.255	**0.812**	**0.224**	**106.5**	**126.0**

NuScenes Dataset:



Methods	S-NeRF	Ours			
		w/o $\mathcal{L}_{ds}$	w/o ϕ_L	w/o $\mathcal{L}_{aug}$	Full
PSNR↑	29.377	30.629	31.001	31.133	**31.162**
SSIM↑	0.859	0.871	0.873	0.883	**0.884**
LPIPS↓	0.349	0.278	0.237	0.222	**0.211**

Project page