Dynamic LiDAR Re-simulation using Compositional Neural Fields

1 ETH zürich 2 TUM 3 NVIDIA 4

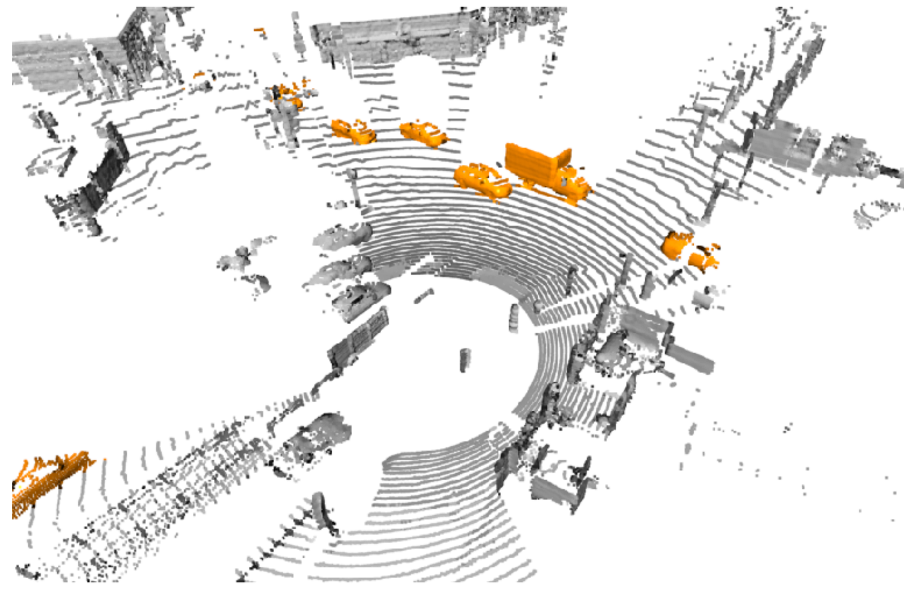
Hanfeng Wu[1,2] , Xingxing Zuo[2] , Stefan Leutenegger[2], Or Litany[3,4], Konrad Schindler[1], Shengyu Huang[1]

CVPR JUNE 17-21, 2024 SEATTLE, WA

Project page

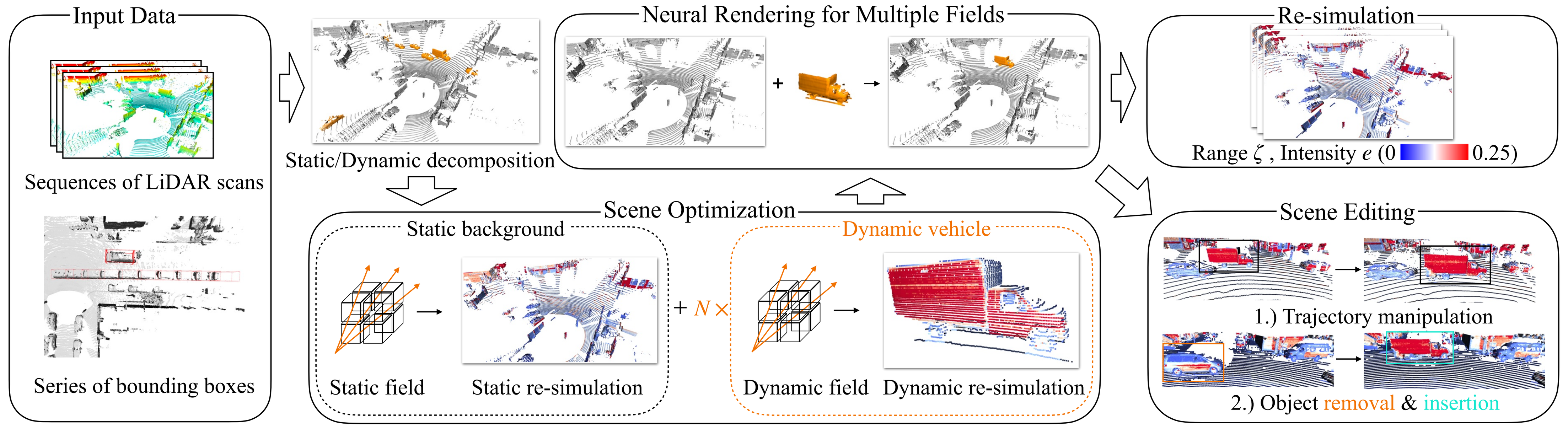
Problem statement

- **Input:** Posed LiDAR scans $\mathbf{X} = \{\mathbf{X}_v\}_{v=1}^{n_v}$, along with tracked bounding boxes for dynamic vehicles $\mathbf{B} = \{\mathbf{B}_t^v\}_{v=1}^{N}$. Every ray $\mathbf{r}(\mathbf{o}, \mathbf{d})$ records measurements (ζ, e, p_d) as:
  - ζ : range of the first return
  - e : intensity of the first return
  - $p_d \in \{0,1\}$: ray drop mask
- **Goal:** Render virtual LiDAR scans $\mathbf{X}_{tgt}$ from novel sensor poses $\mathbf{T}_{tgt}$



Method Overview

- Dynamic Neural LiDAR fields: $F_{static} \cup \{F^v\}_{v=1}^{N}$.



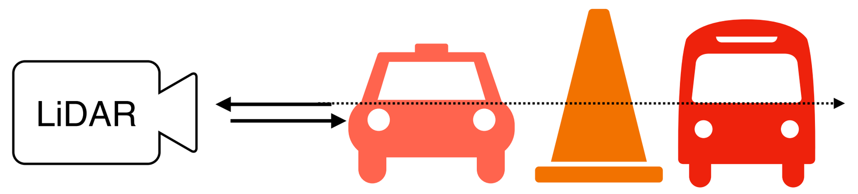
- Loss terms: $L = w_\zeta L_\zeta + w_s L_s + w_{eik} L_{eik} + w_e L_e + w_{drop} L_{drop}$

Active sensing versus passive sensing

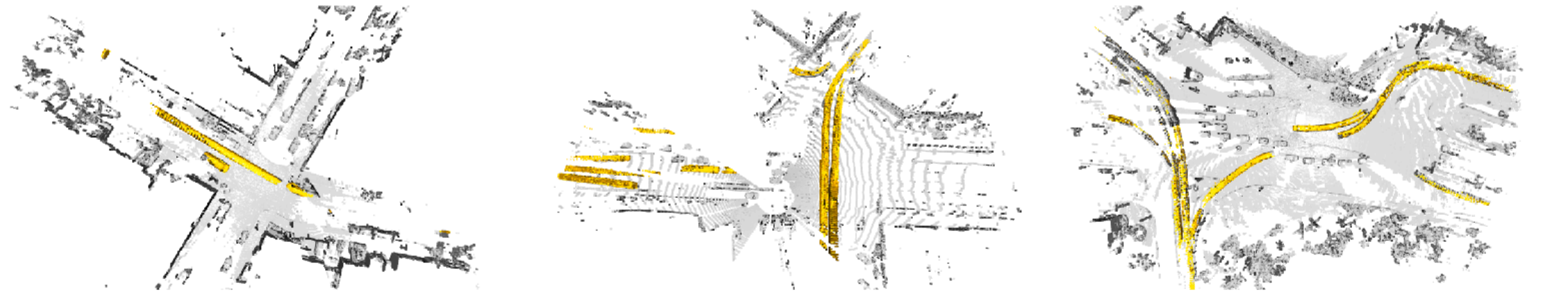
- $$w_j = 2\alpha_j \cdot \prod_{k=1}^{j-1} (1 - 2\alpha_k),$$ $$w_j = \alpha_j \cdot \prod_{k=1}^{j-1} (1 - \alpha_k)$$
- $$\alpha_n = max\{\frac{\Phi_s(f(\mathbf{p}(n)))^2 - \Phi_s(f(\mathbf{p}(n+1)))^2}{2\Phi_s(f(\mathbf{p}(n)))^2}, 0\}$$ $$\alpha_n = max\{\frac{\Phi_s(f(\mathbf{p}(n))) - \Phi_s(f(\mathbf{p}(n+1)))}{\Phi_s(f(\mathbf{p}(n)))}, 0\}$$

Neural Fields Composition

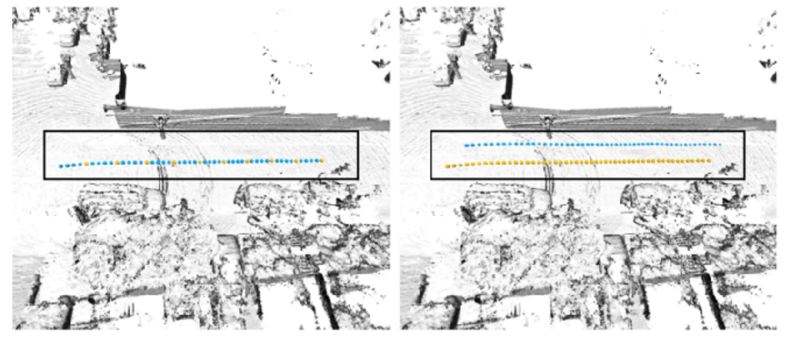
- Perform volume rendering for each **intersected** neural field, yielding k+1 LiDAR measurements $\{(\zeta, e, p_d)_i\}_{i=0}^{k+1}$
  - The Ray is classified as ***dropped*** if $p_{d,i} > 0.5$
  - Determine the result based on the closest ***non-dropped*** neural field.



Datasets and evaluation protocol



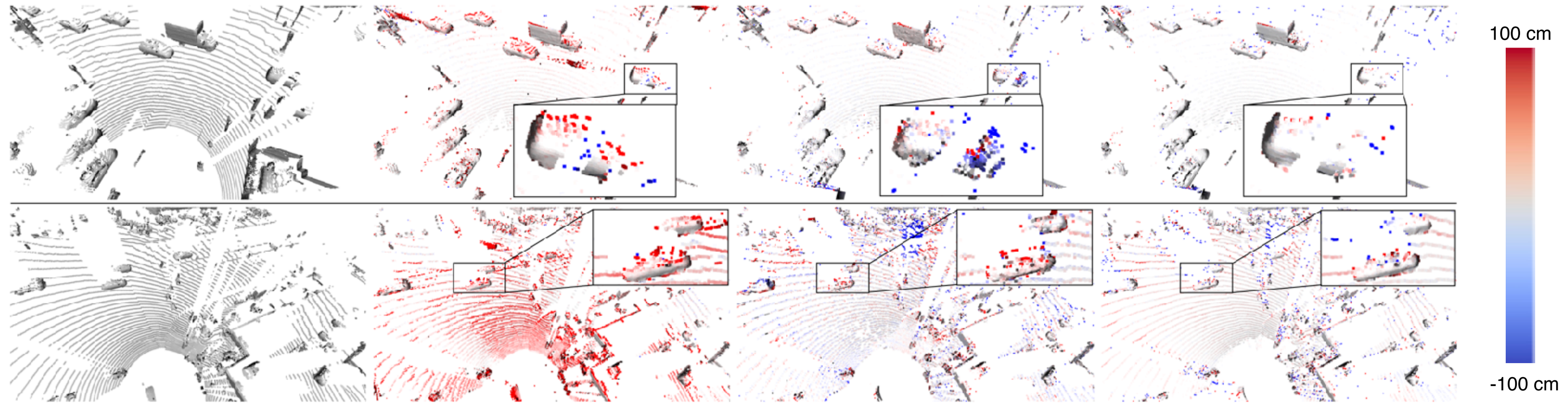
4 dynamic scenes from Waymo Open datasets



(a) Waymo Dynamic (b) Waymo Dynamic NVS

Experimental results

- Comprehensive ray measurement evaluation of LiDAR NVS on *Waymo Dynamic* dataset.



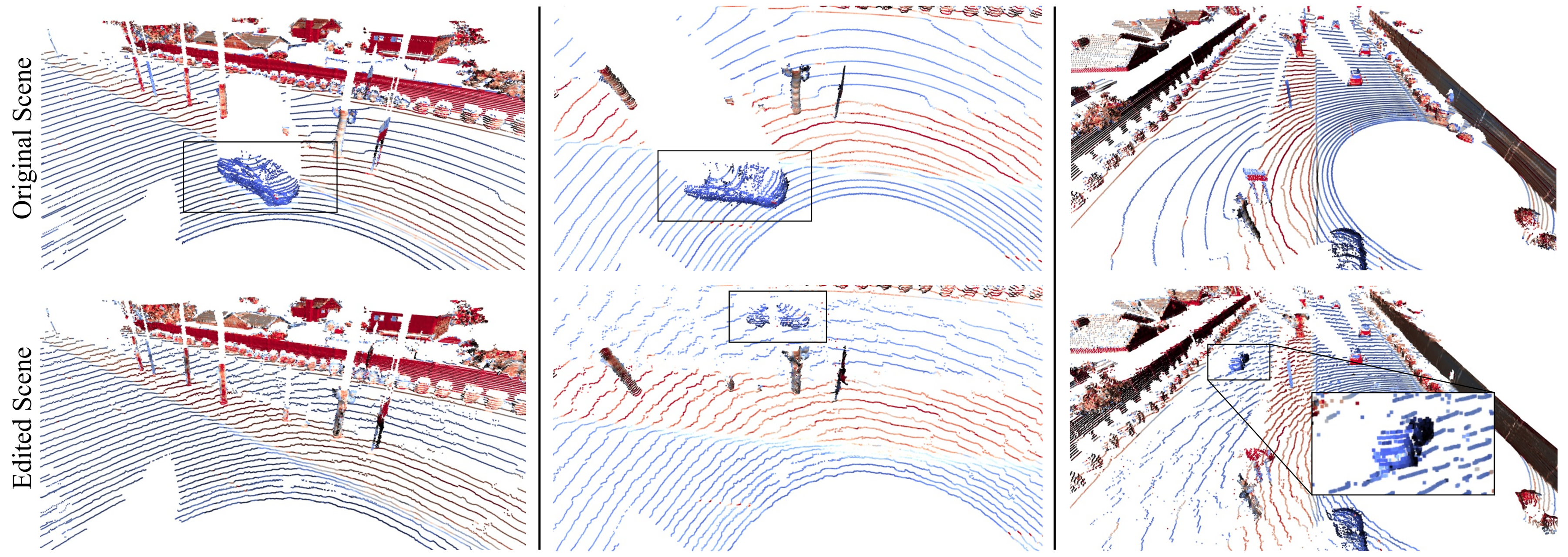
Method	MAE ↓	MedAE ↓	CD ↓	MedAE Dyn ↓	Intensity RMSE ↓
LiDARsim [27]	170.1	11.5	31.1	16.0	0.10
Unisim [55]	35.6	6.1	14.3	14.3	**0.05**
Ours	**30.8**	**3.0**	**10.9**	**8.5**	**0.05**

(a) Waymo Dynamic

Method	MAE ↓	MedAE ↓	CD ↓	MedAE Dyn ↓	Intensity RMSE ↓
LiDARsim [3]	448.4	55.1	77.0	38.7	0.13
Unisim [10]	115.1	9.7	33.5	24.3	0.19
Ours	**72.9**	**3.8**	**22.9**	**14.0**	**0.07**

(b) Waymo Dynamic NVS

- Scene editing capabilities. The scenes are color coded by intensity values (0 0.25).



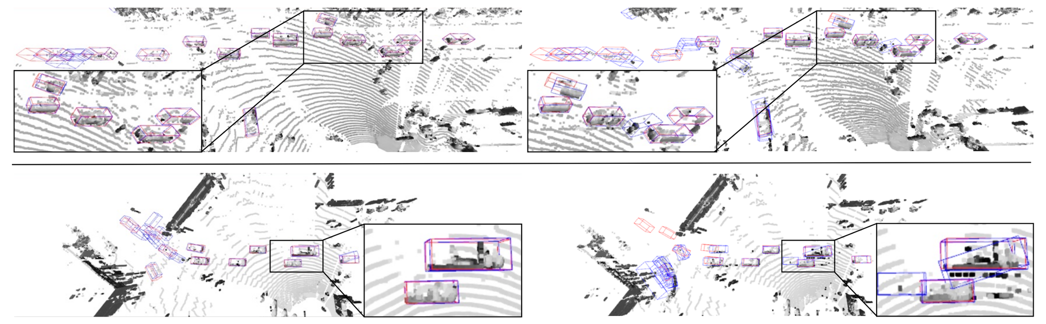
Vehicle removal Trajectory manipulation Vehicle insertion

- Effectiveness of volume rendering for active sensing.

Datasets	MAE ↓	MedAE ↓	CD ↓
TownClean	26.7(-1.5)	0.7(-0.2)	6.7(-0.5)
Waymo Interp	28.3 (0.1)	4.7 (-0.2)	12.5 (-0.1)
Waymo Dynamic	30.8 (-0.3)	3.0 (-0.2)	10.9 (-0.3)

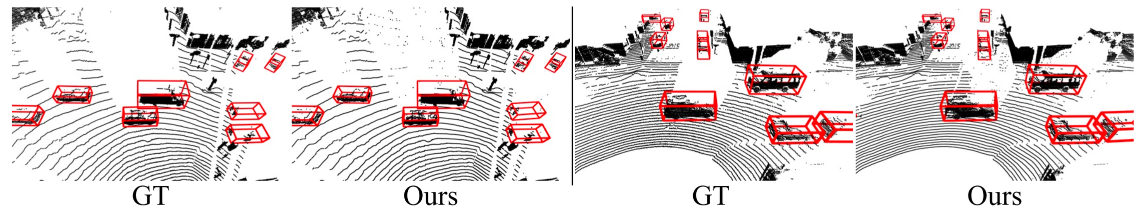
- Object detection agreement. Results from GT and synthetic LiDAR scans are marked with red and blue, respectively.

	GT	Ours			LiDARSim[27]		
Threshold	AP↑	AP↑	Agg.↑	Dyn. Agg.↑	AP↑	Agg.↑	Dyn. Agg.↑
IoU>0.7	0.85	0.86	**0.77**	**0.71**	**0.90**	0.76	0.68
IoU>0.5	**0.98**	0.96	**0.87**	**0.76**	0.95	0.86	**0.76**



Ours LiDARsim

- Object detection on noisy re-simulated LiDAR scans



GT Ours GT Ours

Our contribution

- Derived SDF-based volume rendering formulation for active sensors;
- Developed a novel neural fields composition method for Dynamic LiDAR re-simulation;
- Provided powerful scene editing capabilities for assets manipulation.

Future work

- Integrate shape priors for unseen regions;
- Refine object b-box based on rendering loss;
- A generative model with learned scene layout.