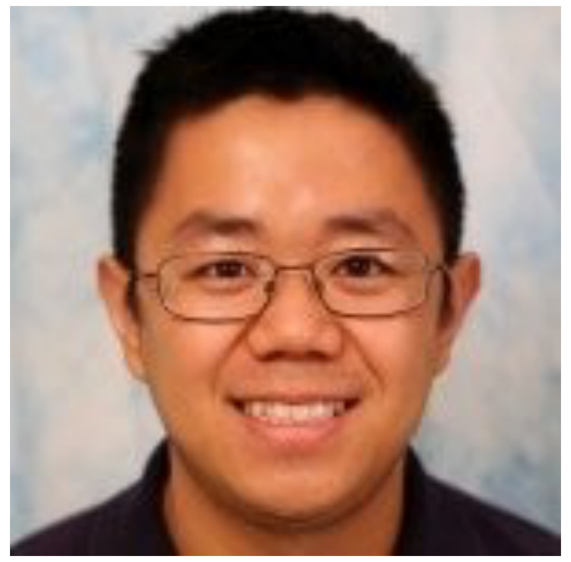
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Point Cloud Forecasting as a proxy for 4D Occupancy Forecasting


Tarasha Khurana*


Peiyun Hu*

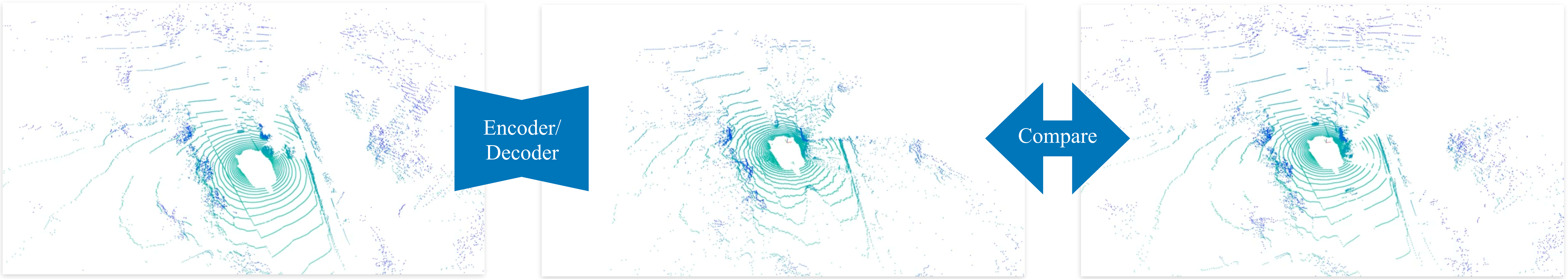

David Held


Deva Ramanan

Carnegie Mellon University

Overview: Learning to forecast 4D occupancy

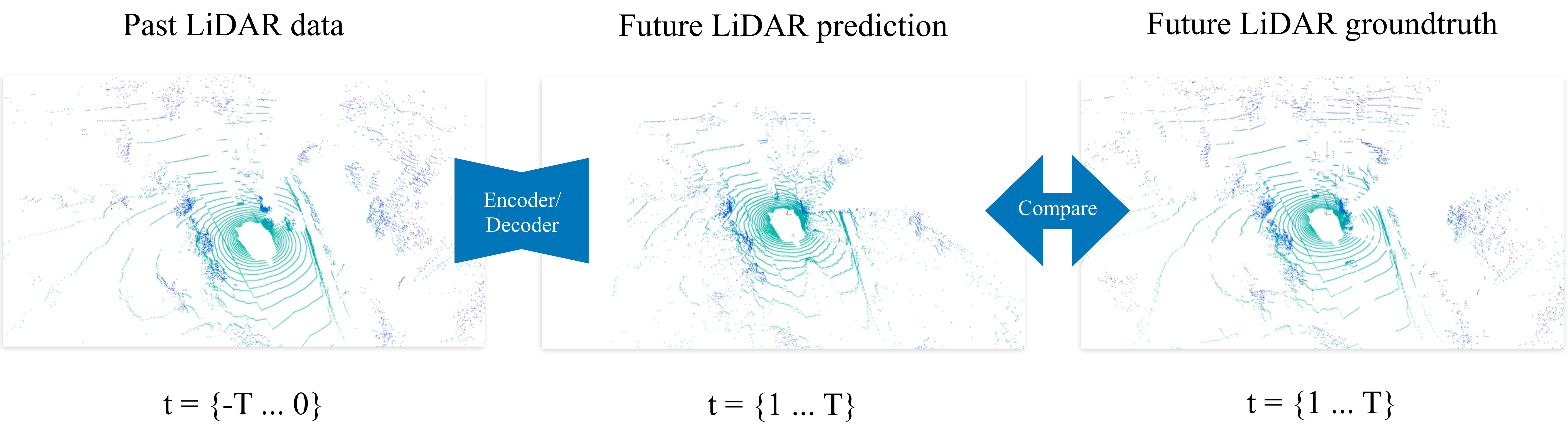
Past LiDAR data | Future LiDAR prediction | Future LiDAR groundtruth



t = {-T ... 0} | t = {1 ... T} | t = {1 ... T}

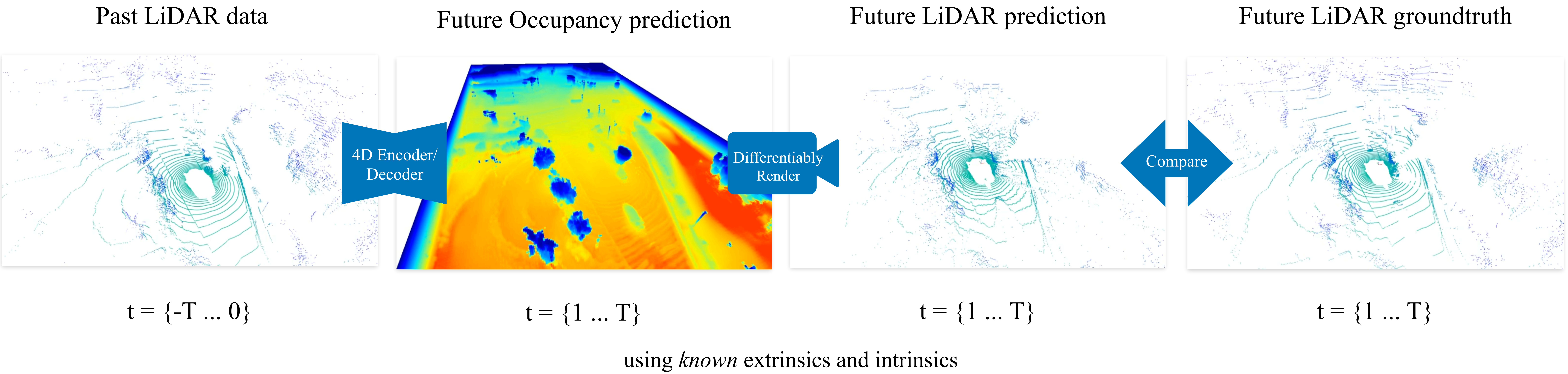
- Point cloud forecasting

Overview: Learning to forecast 4D occupancy



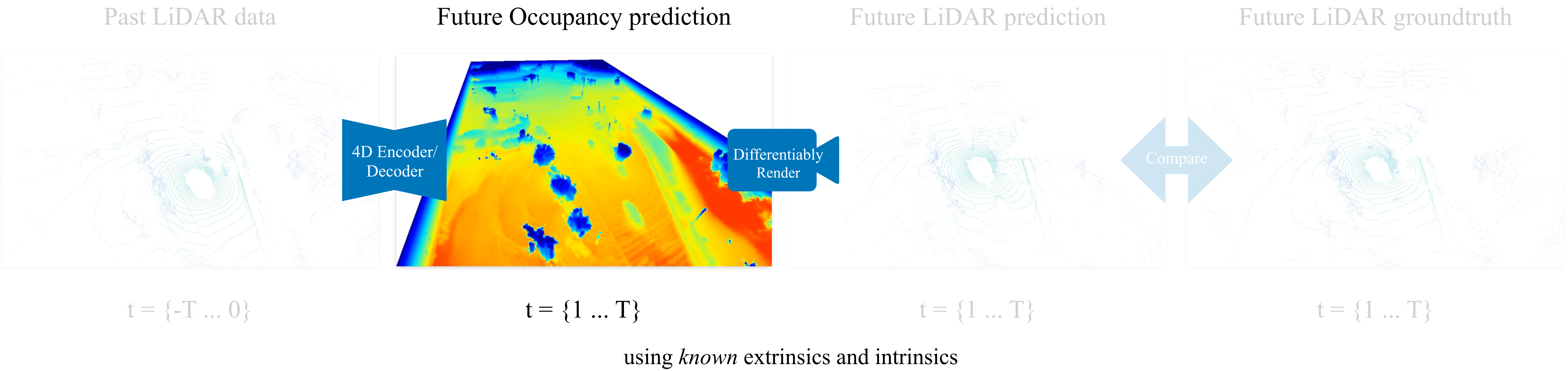
- Point cloud forecasting = sensor extrinsics and intrinsics

Overview: Learning to forecast 4D occupancy



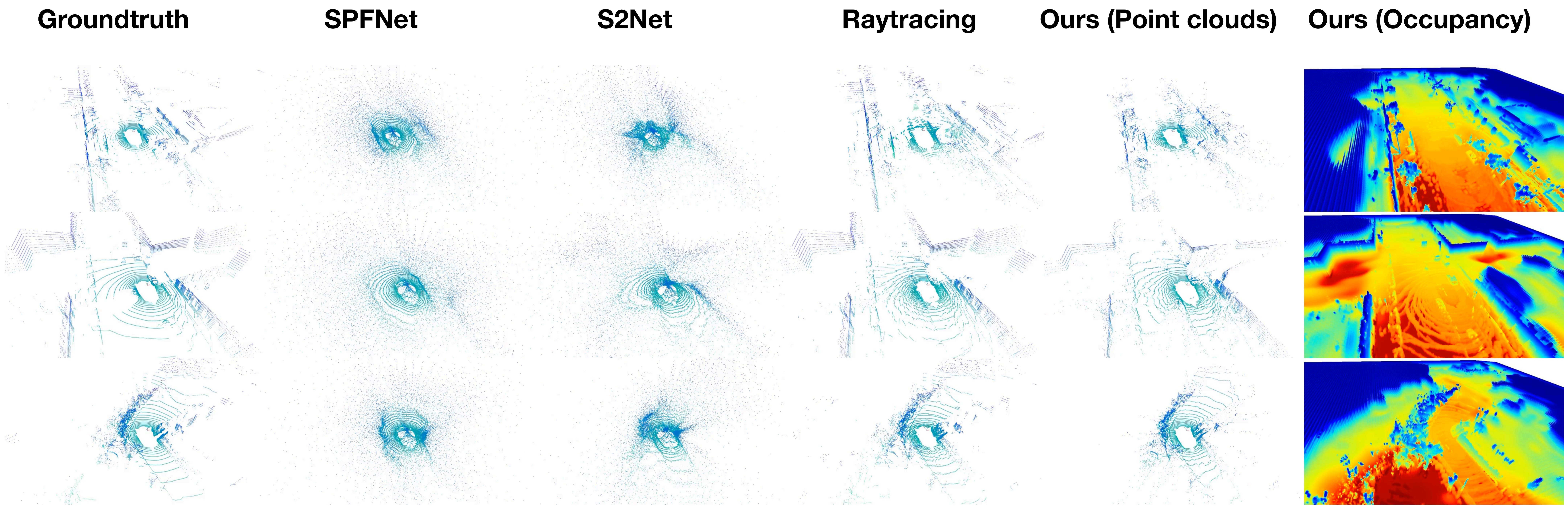
- Point cloud forecasting = sensor extrinsics and intrinsics + **4D occupancy forecasting**

Overview: Learning to forecast 4D occupancy



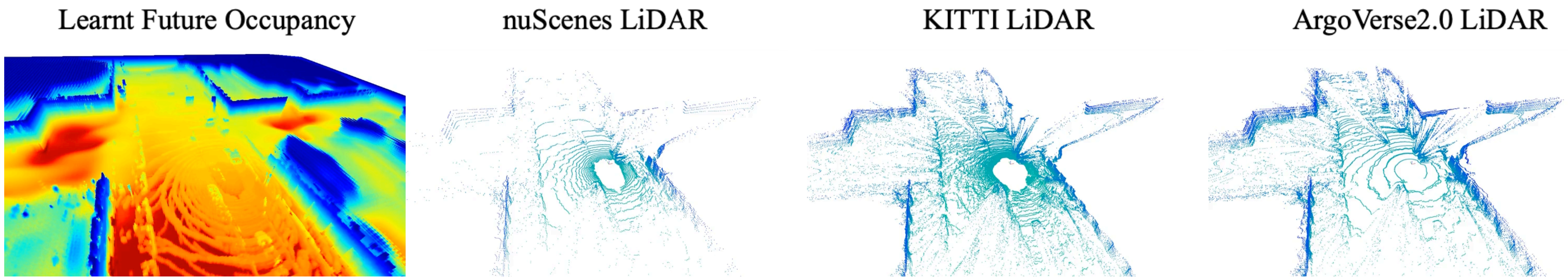
- Point cloud forecasting = sensor extrinsics and intrinsics + **4D occupancy forecasting**

Overview: Learning to forecast 4D occupancy



- Point cloud forecasting = sensor extrinsics and intrinsics + **4D occupancy forecasting**
- Dramatically improved performance on point cloud forecasting as compared to SOTA

Overview: Learning to forecast 4D occupancy

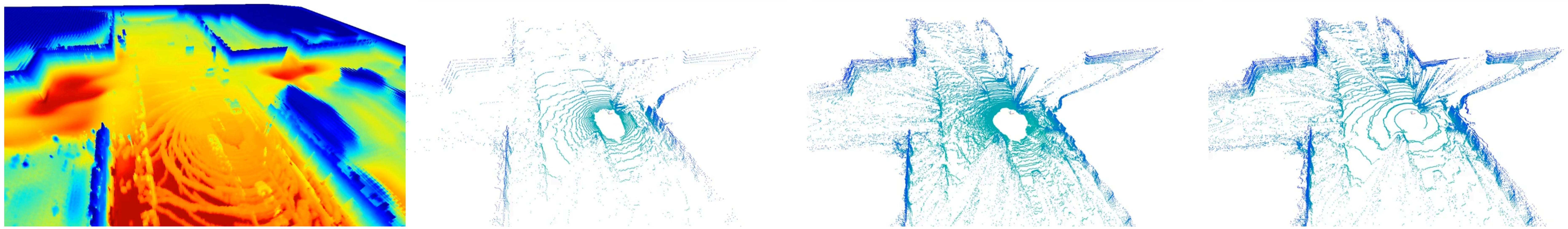


Novel intrinsic-view synthesis

- Point cloud forecasting = sensor extrinsics and intrinsics + **4D occupancy forecasting**
- Dramatically improved performance on point cloud forecasting as compared to SOTA
- Disentaglement allows for cross-sensor training and generalization

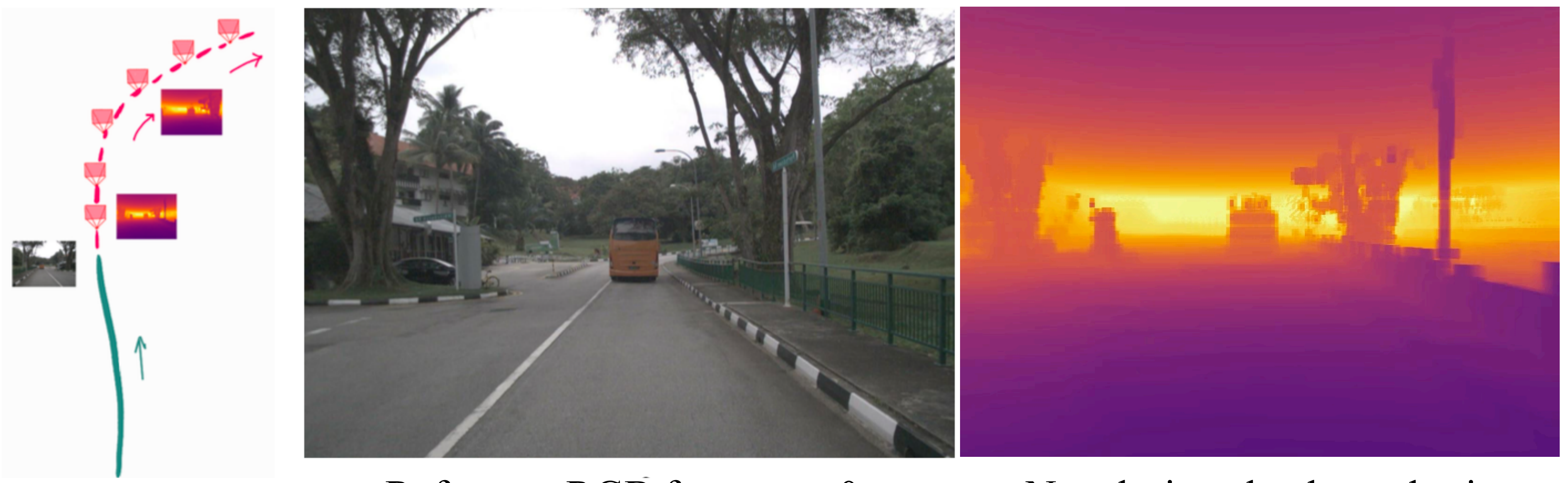
Overview: Learning to forecast 4D occupancy

Learnt Future Occupancy | nuScenes LiDAR | KITTI LiDAR | ArgoVerse2.0 LiDAR



Novel intrinsic-view synthesis

Novel extrinsic-view synthesis



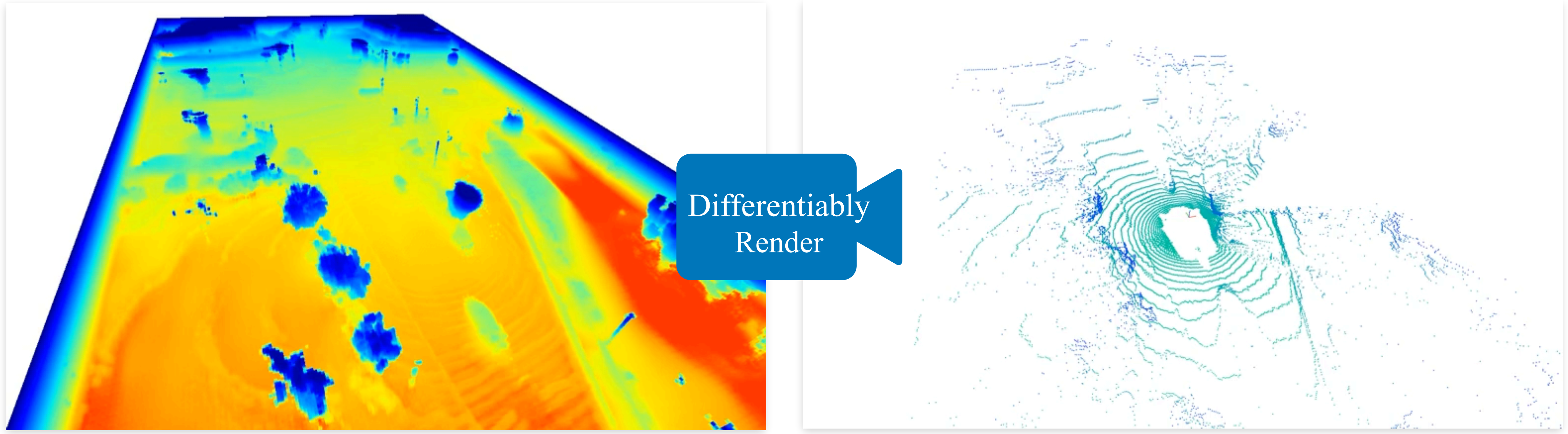
Reference RGB frame, t = 0s | Novel-view depth synthesis

- Point cloud forecasting = sensor extrinsics and intrinsics + **4D occupancy forecasting**
- Dramatically improved performance on point cloud forecasting as compared to SOTA
- Disentaglement allows for cross-sensor training and generalization

Point Cloud Forecasting as a proxy for 4D Occupancy Forecasting

Future Occupancy prediction

Future LiDAR prediction



t = {1 ... T}

t = {1 ... T}

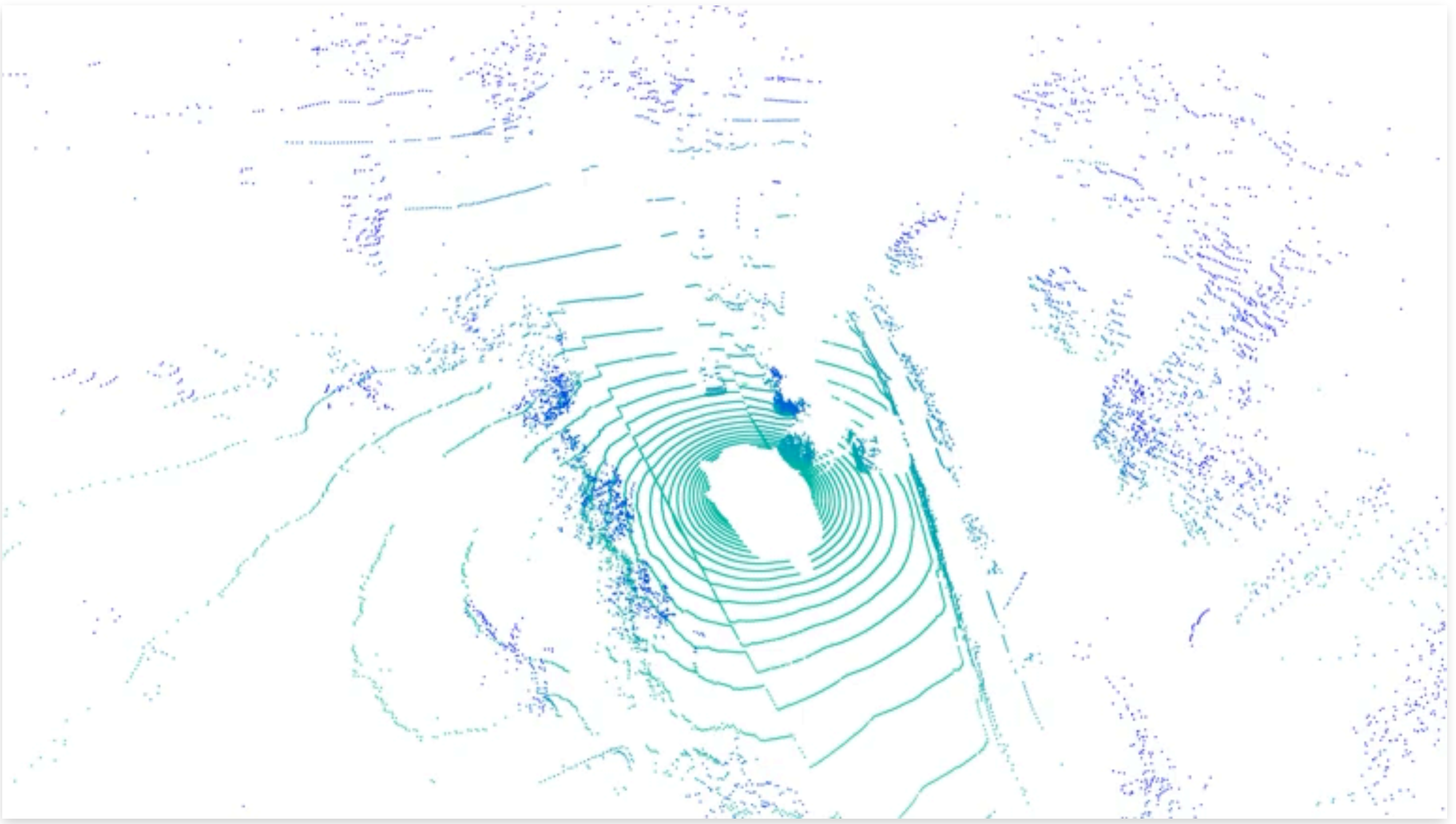
using *known* extrinsics and intrinsics

Standard perception and prediction requires costly labels

Credit: Waymo Open Challenges

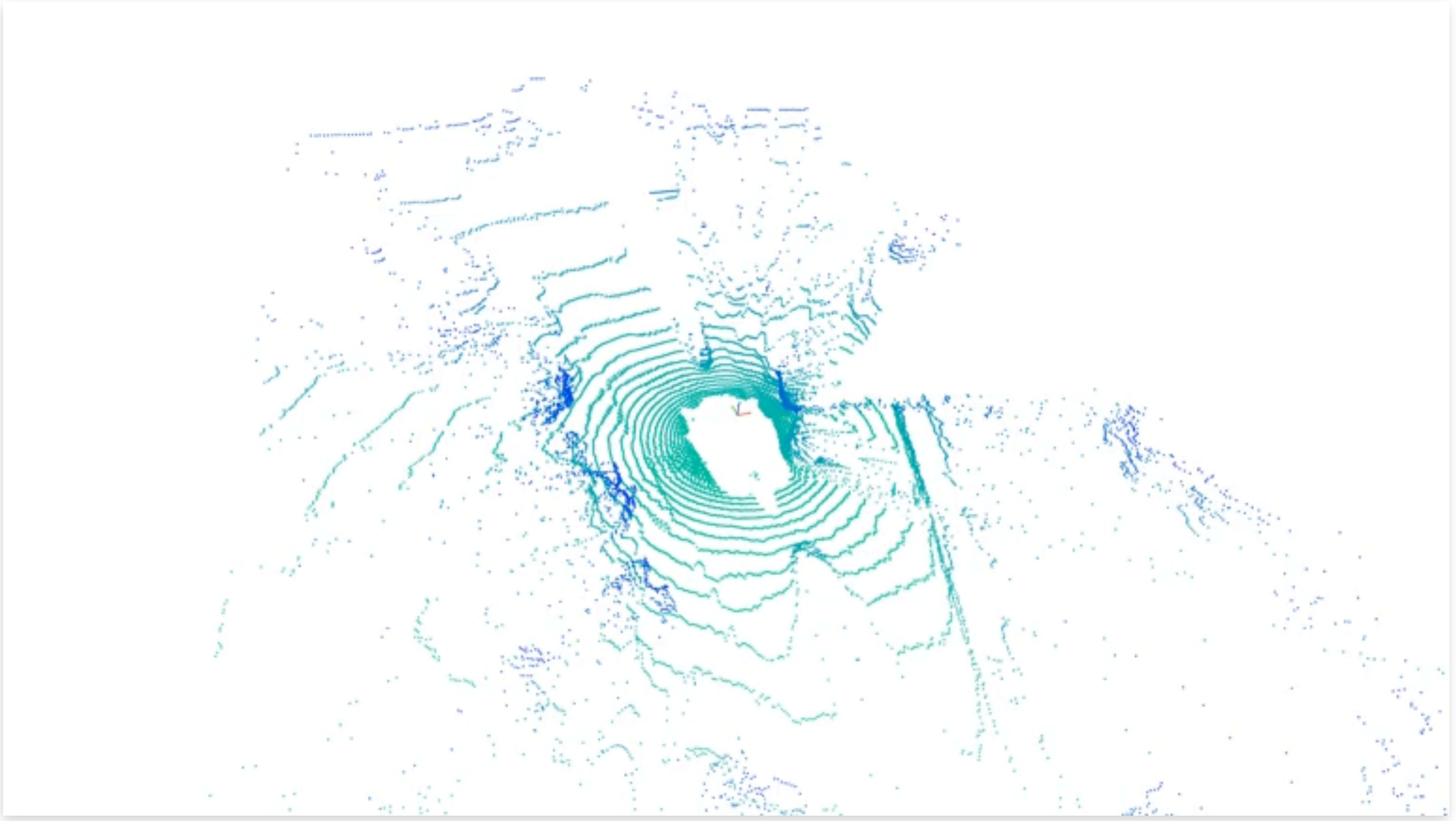
Point Cloud Forecasting

Historical LiDAR Sweeps



Predict

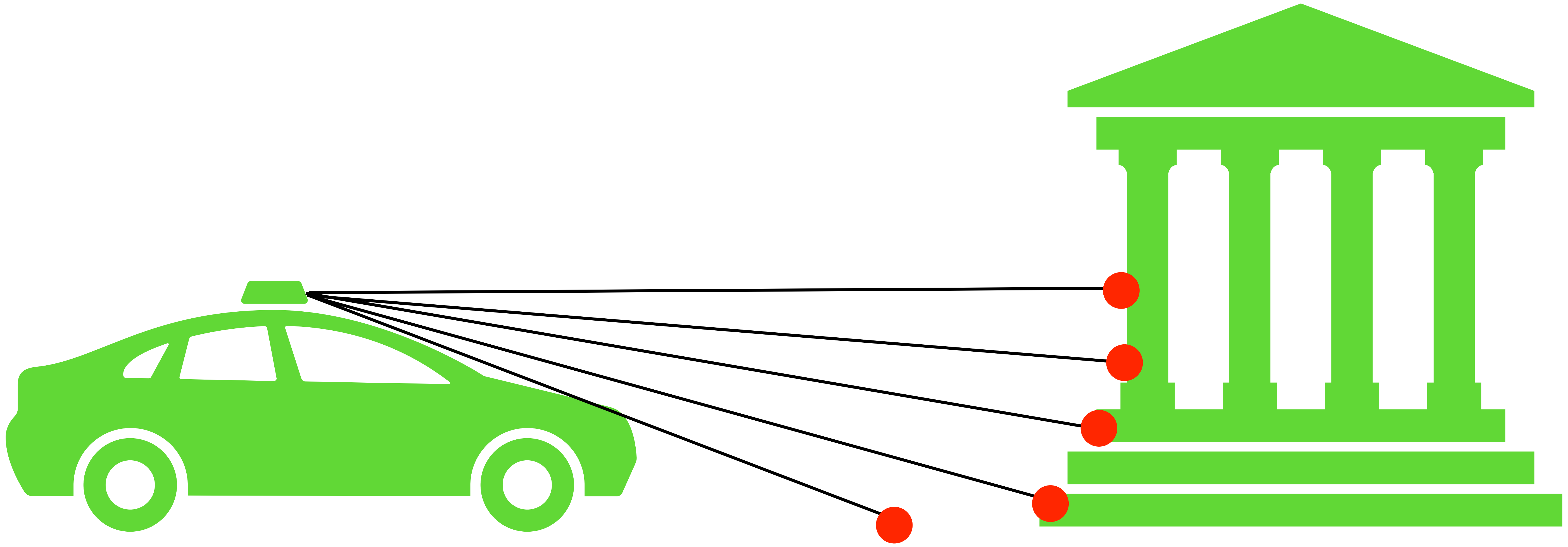
Future Point Clouds



4D Forecasting: Sequential Forecasting of 100,000 Points
Weng et al., CVPR'21

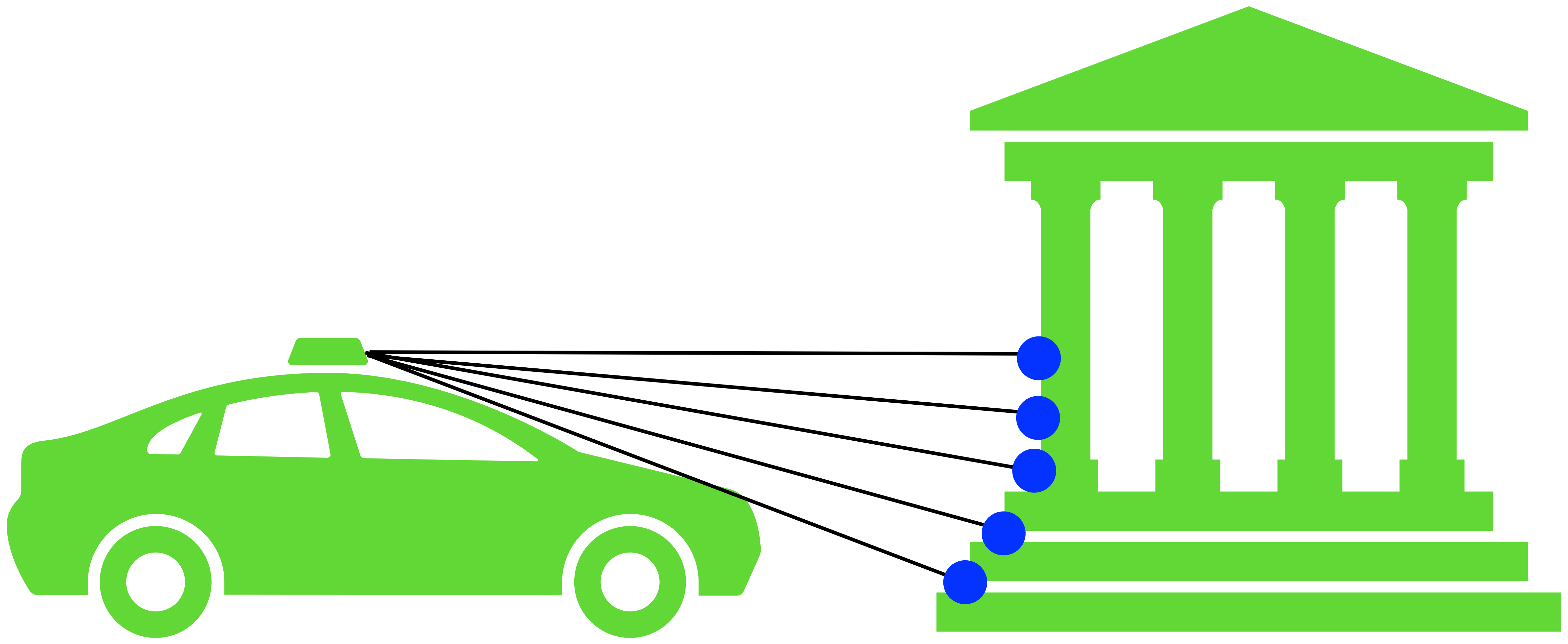
Self-supervised Point Cloud Prediction using 3D Spatial-temporal Convolutional Networks
Mersch et al., CORL'22

The difficulty of predicting points



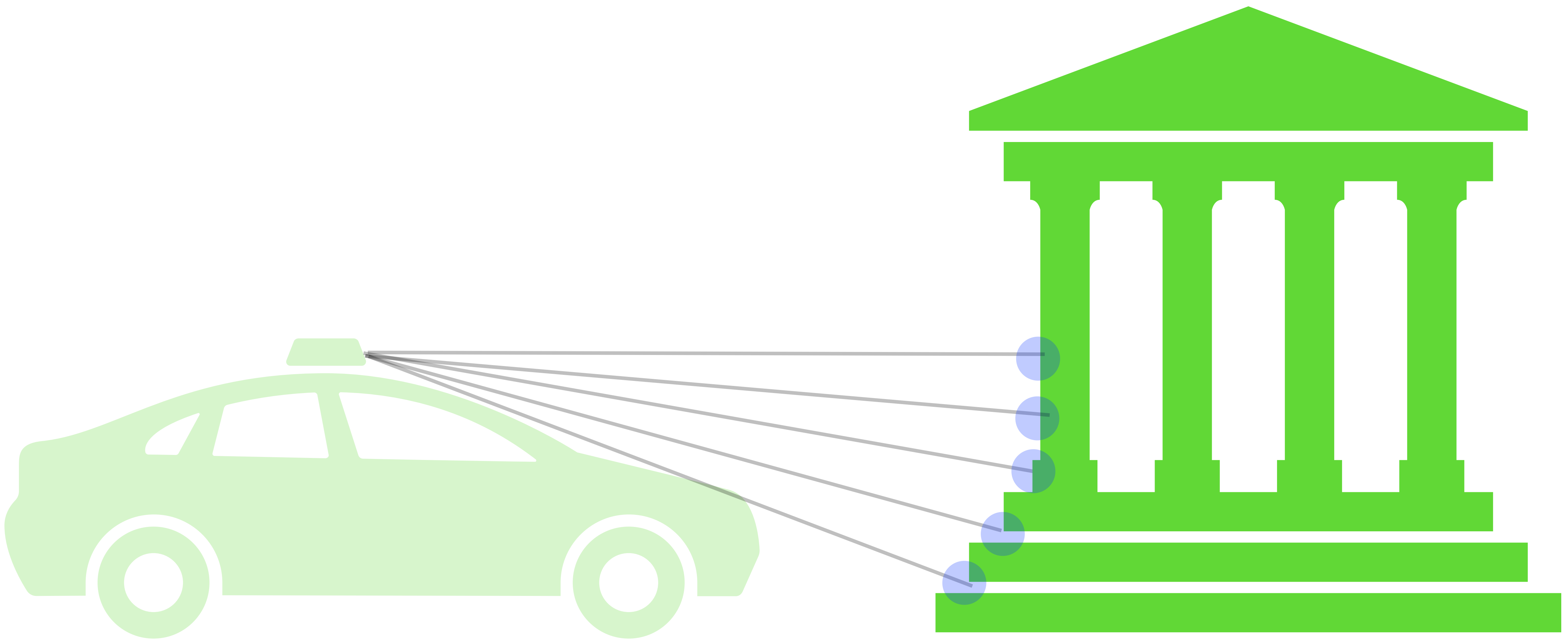
Points lie at the intersection of sensor rays and environment

The difficulty of predicting points



Rays change with change in sensor extrinsics *and* intrinsics

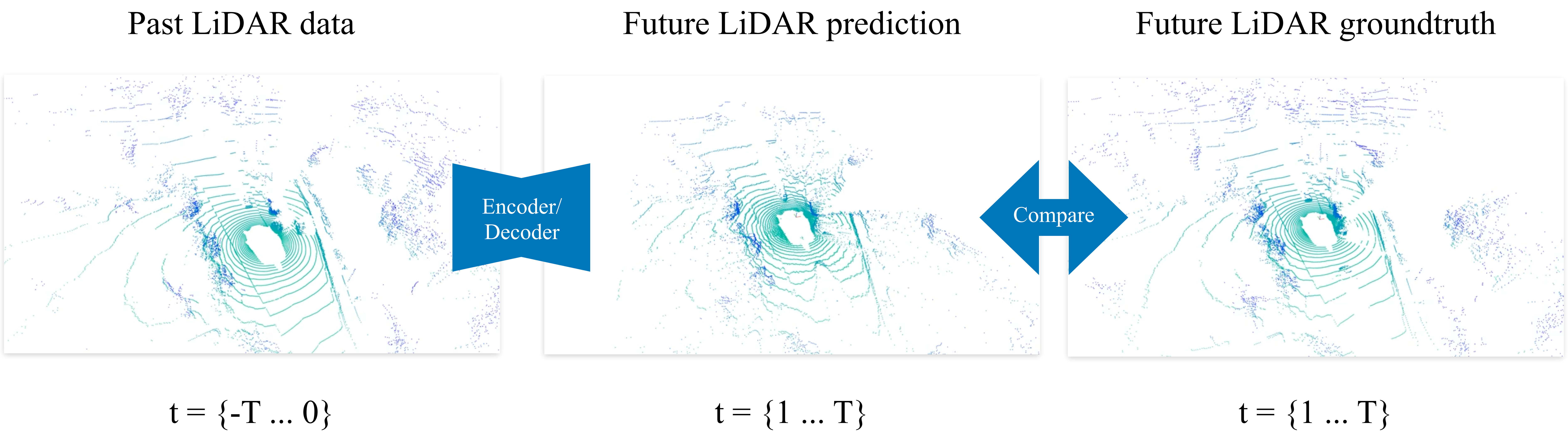
The difficulty of predicting points



We should make predictions about our environment, not our sensor!

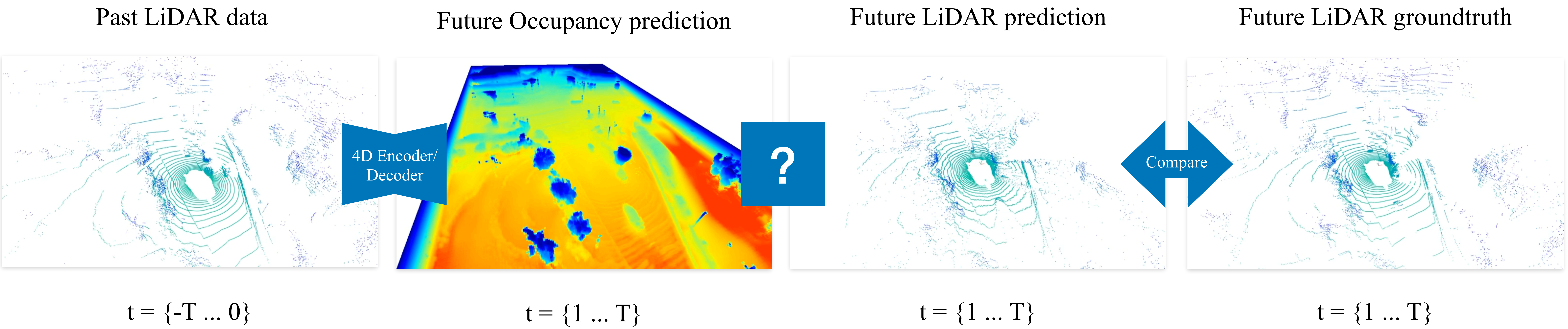
4D Occupancy Forecasting

w/ differentiable volumetric rendering



4D Occupancy Forecasting

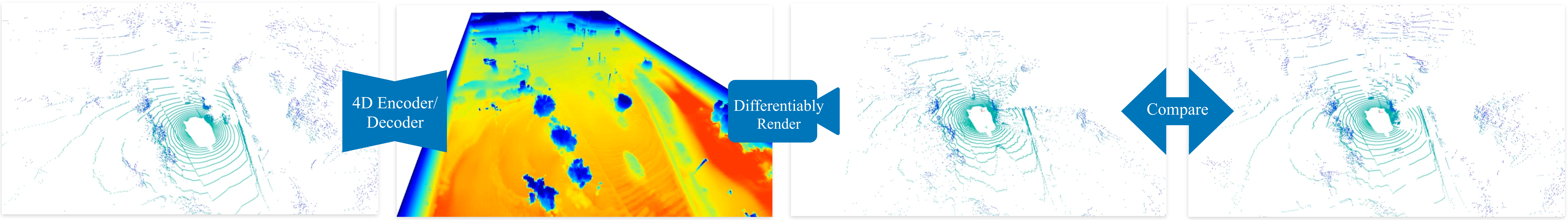
w/ differentiable volumetric rendering



4D Occupancy Forecasting

w/ differentiable volumetric rendering

Past LiDAR data | Future Occupancy prediction | Future LiDAR prediction | Future LiDAR groundtruth



t = {-T ... 0} | t = {1 ... T} | t = {1 ... T} | t = {1 ... T}

using *known* extrinsics and intrinsics

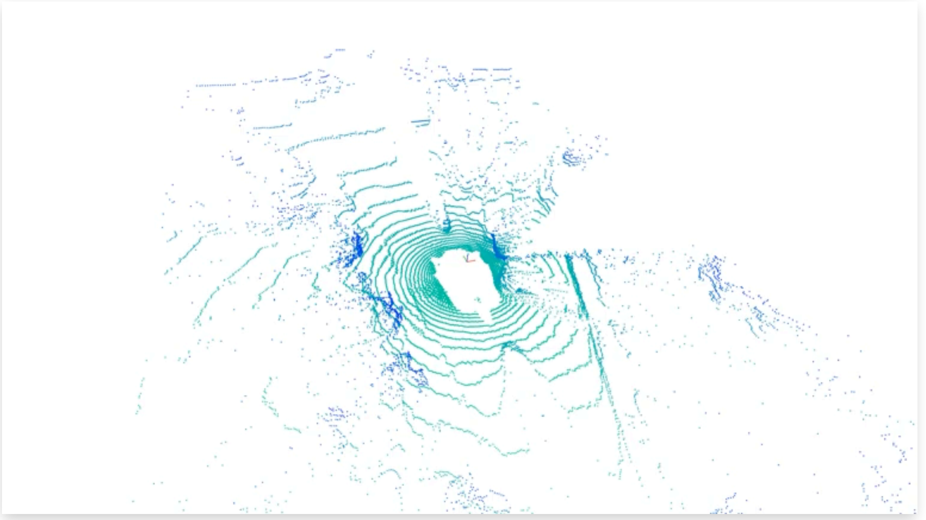
Qualitative results on nuScenes

Groundtruth | **SPFNet** | **S2Net** | **Raytracing** | **Ours (Point clouds)** | **Ours (Occupancy)**

Non-learned raytracing baseline is much stronger than SOTA. We improve upon it by recovering dynamic/evolving scene elements.

Evaluation protocol

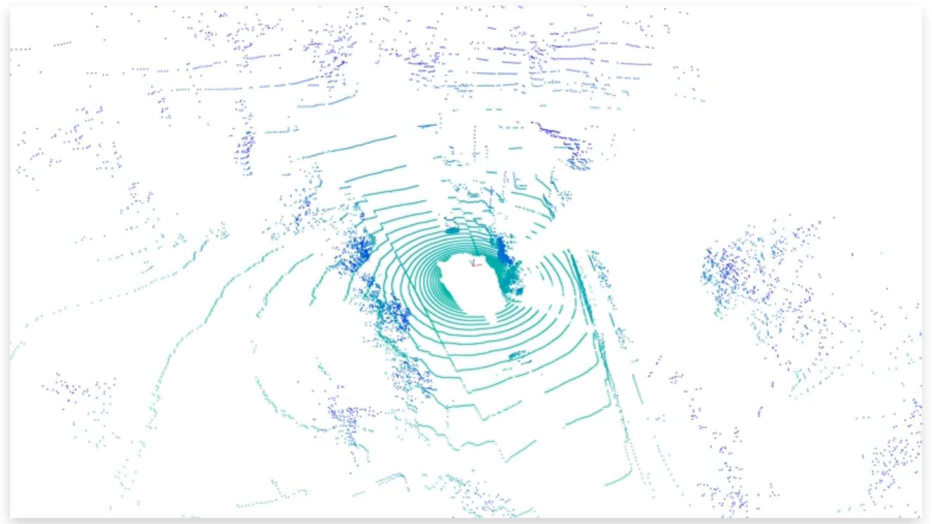
Future Point Clouds



t = {1 ... T}

Chamfer Distance

Groundtruth Point Clouds

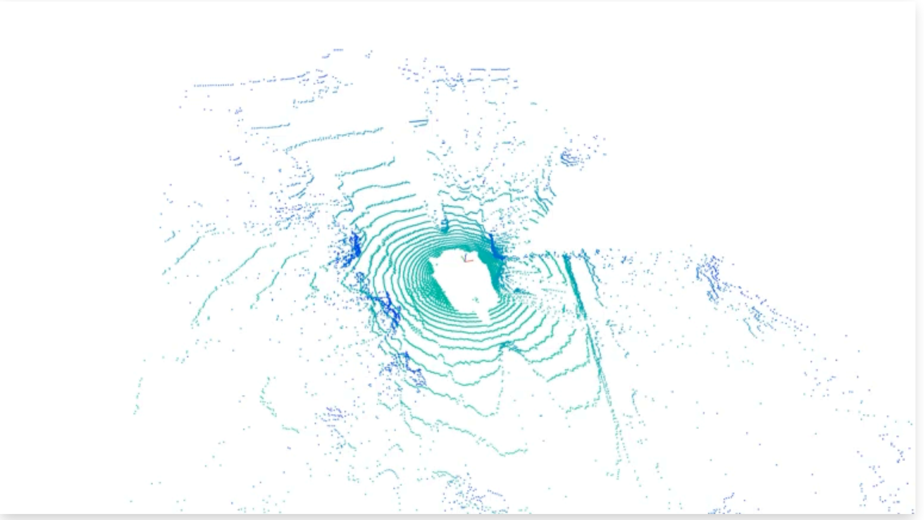


t = {1 ... T}

Traditional setup

Evaluation protocol

Future Point Clouds

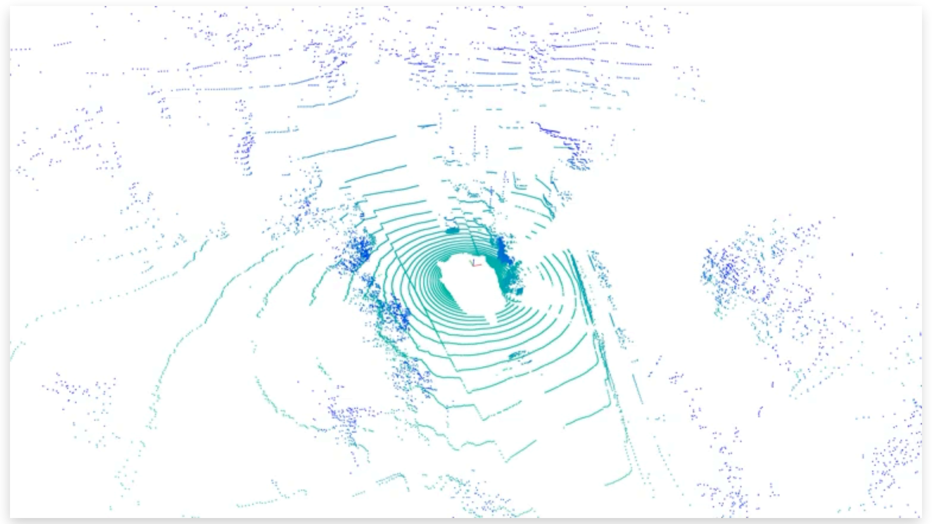


$t = \{1 \dots T\}$

Chamfer Distance

$\longleftrightarrow$

Groundtruth Point Clouds



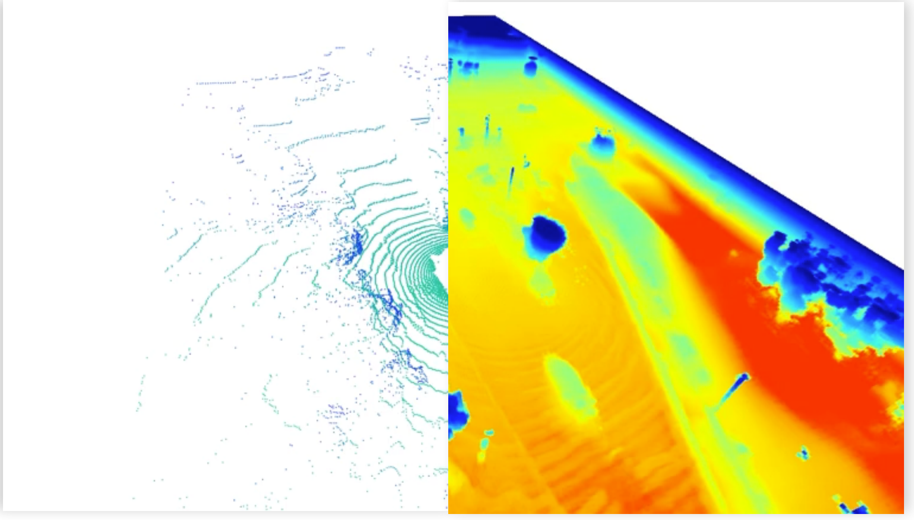
$t = \{1 \dots T\}$

Traditional setup

Proposed setup

Future Scene Representation

(a)



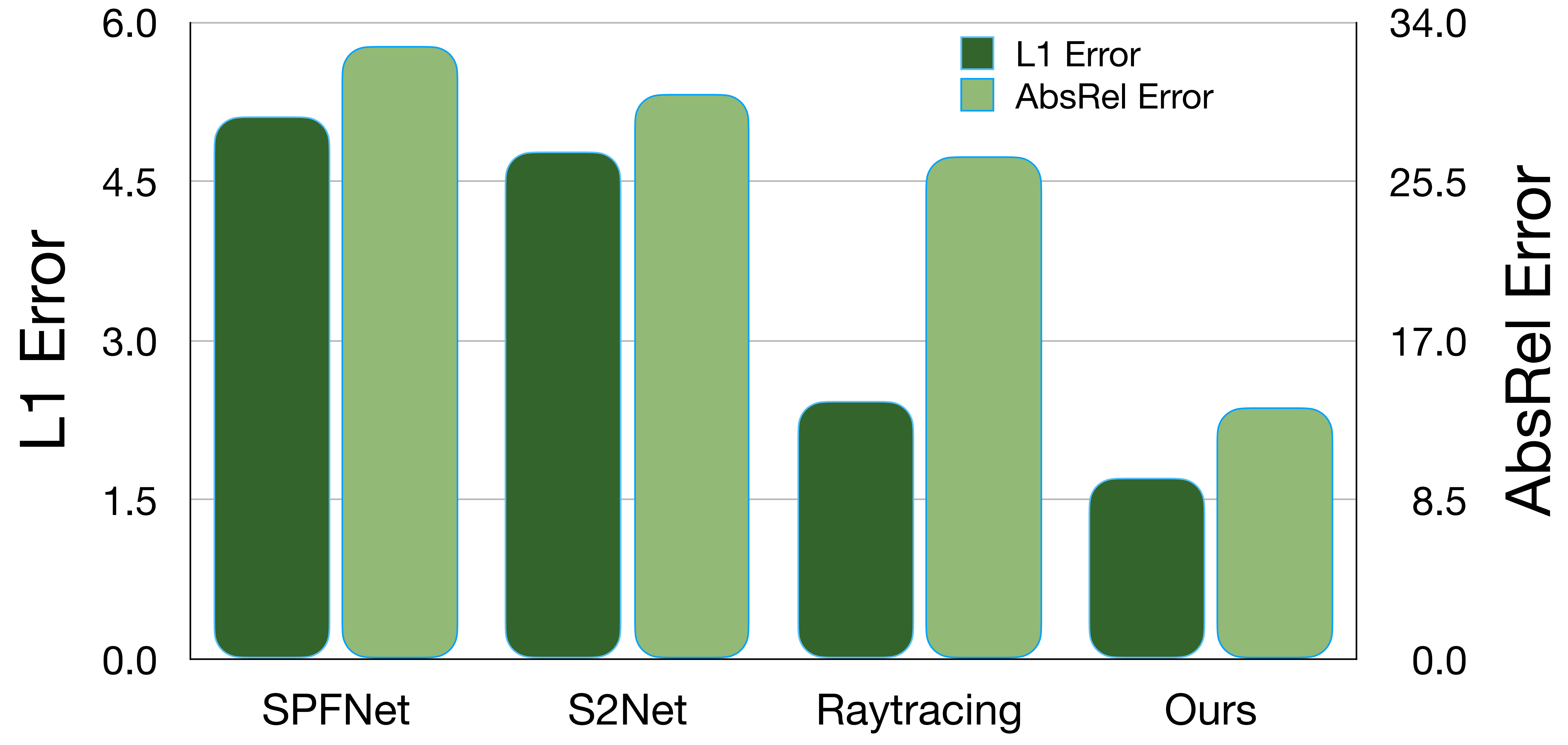
(b) Ray queries

Depth for each query

(L1, AbsRel)

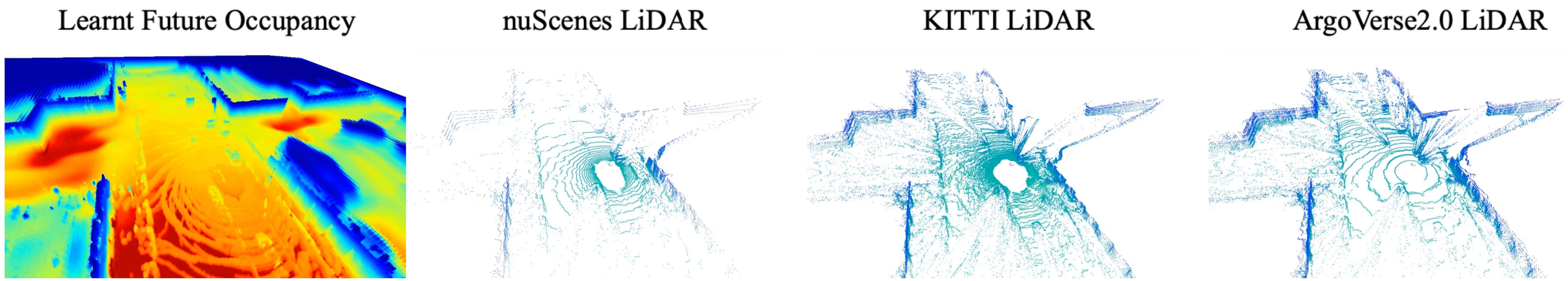
Evaluation on nuScenes

Significant performance improvement upon SOTAs



Metrics in-line with qualitative results: SOTA << Raytracing < Ours

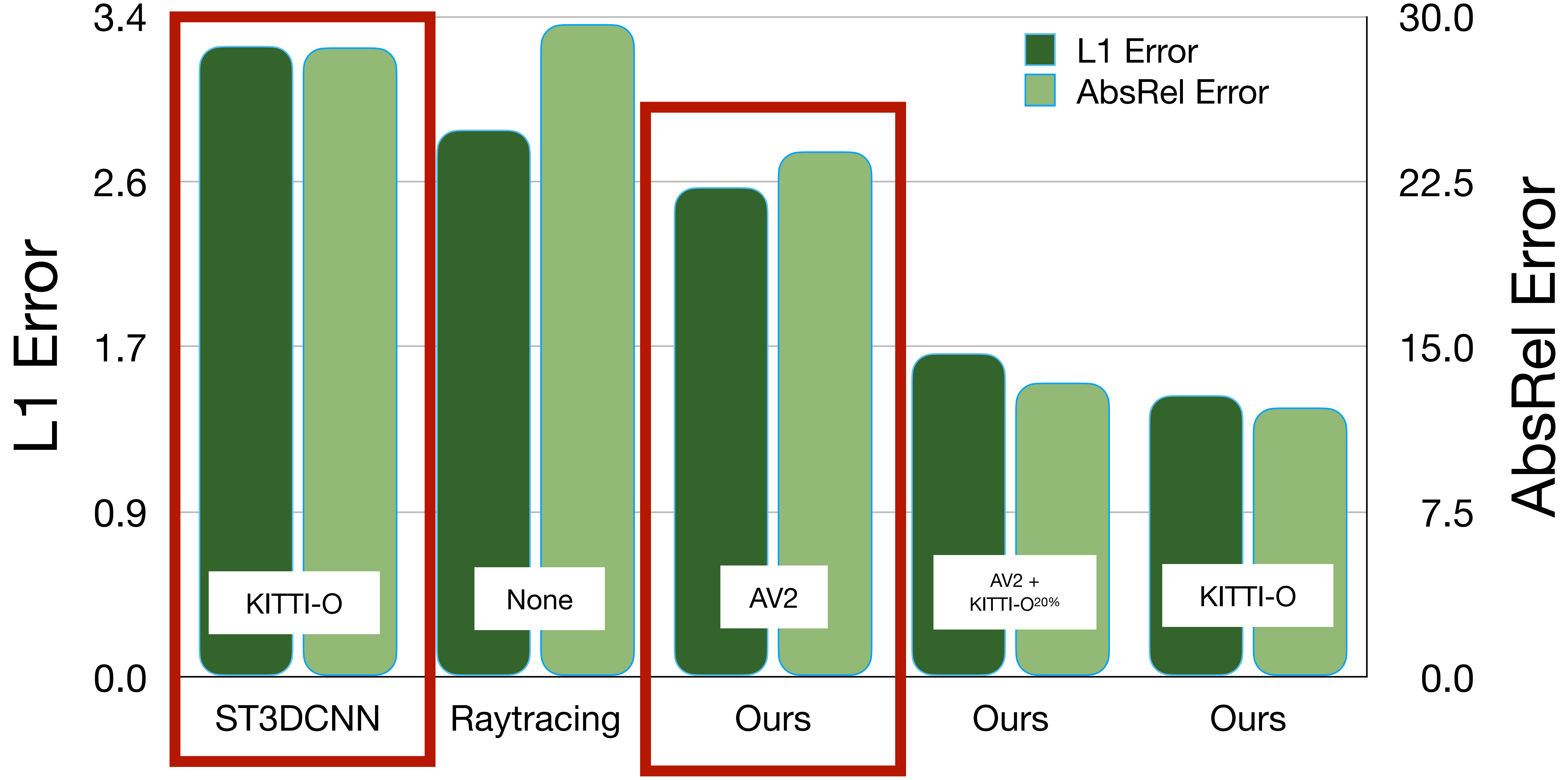
Potential applications: Changing intrinsics



Use predicted occupancy to render point clouds for different sensors.

Evaluation on KITTI-Odometry

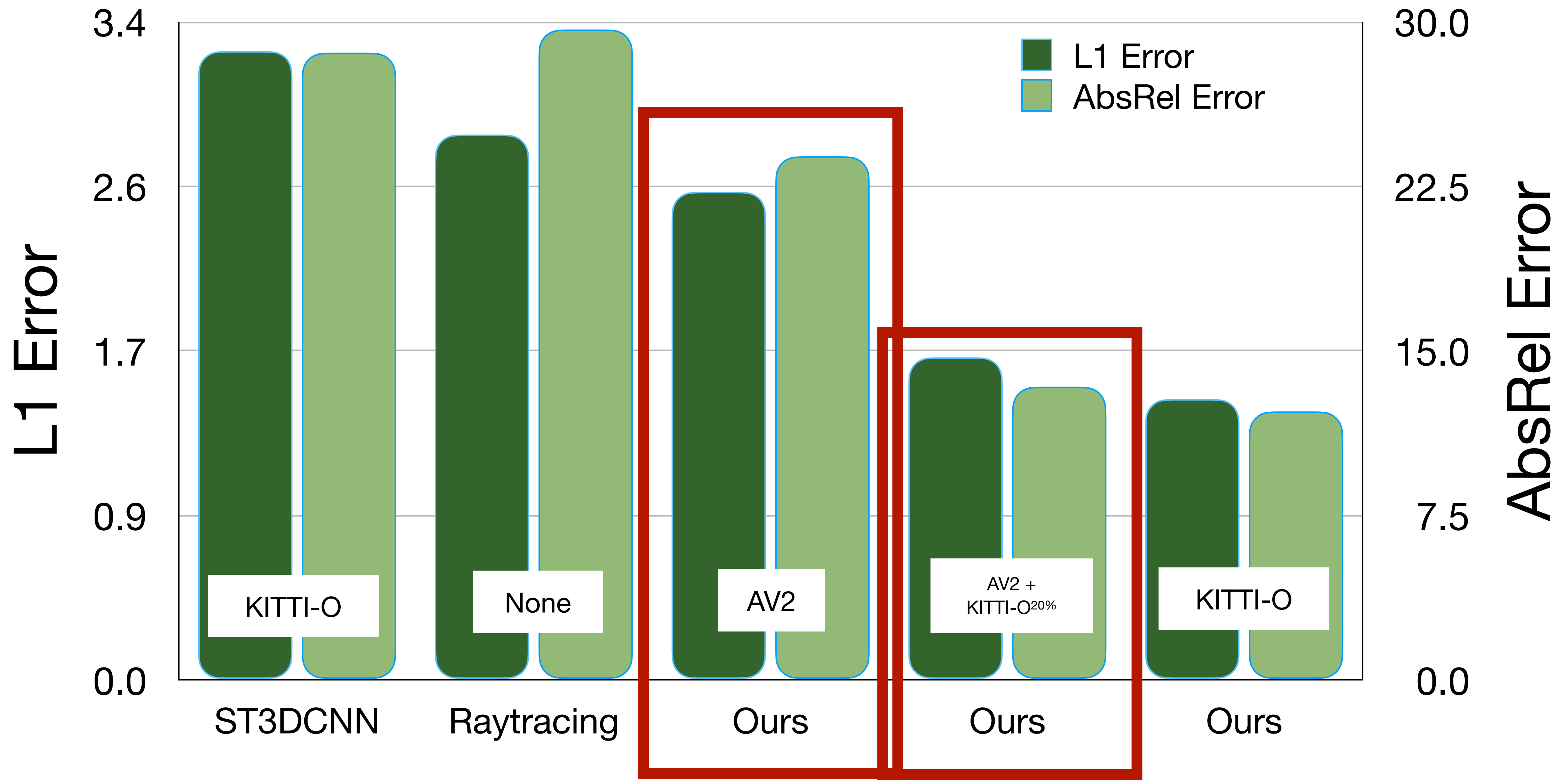
Significant performance improvement upon SOTA



Multi-domain (AV2 + KITTI-O) training does remarkably better than SOTA

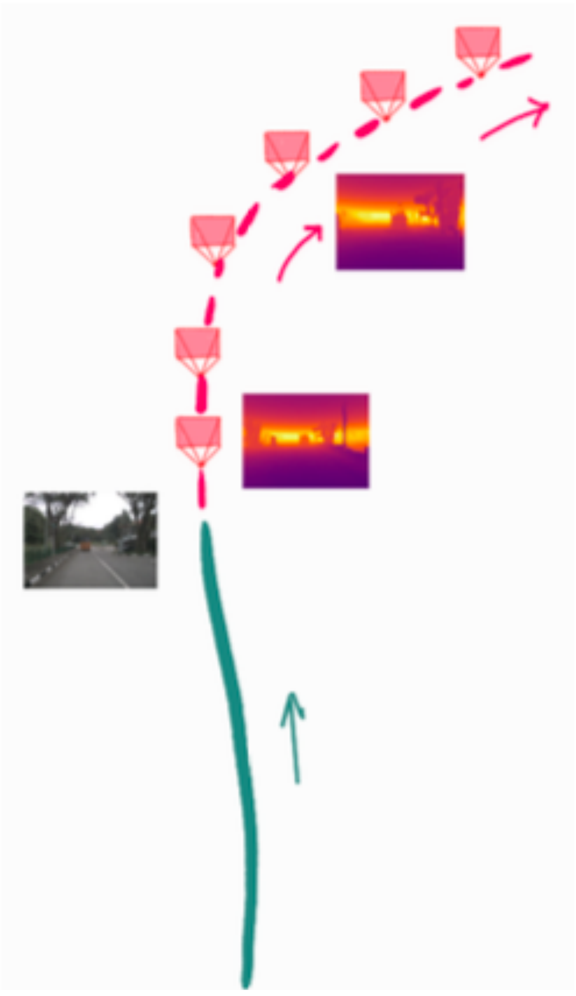
Evaluation on KITTI-Odometry

Significant performance improvement upon SOTA

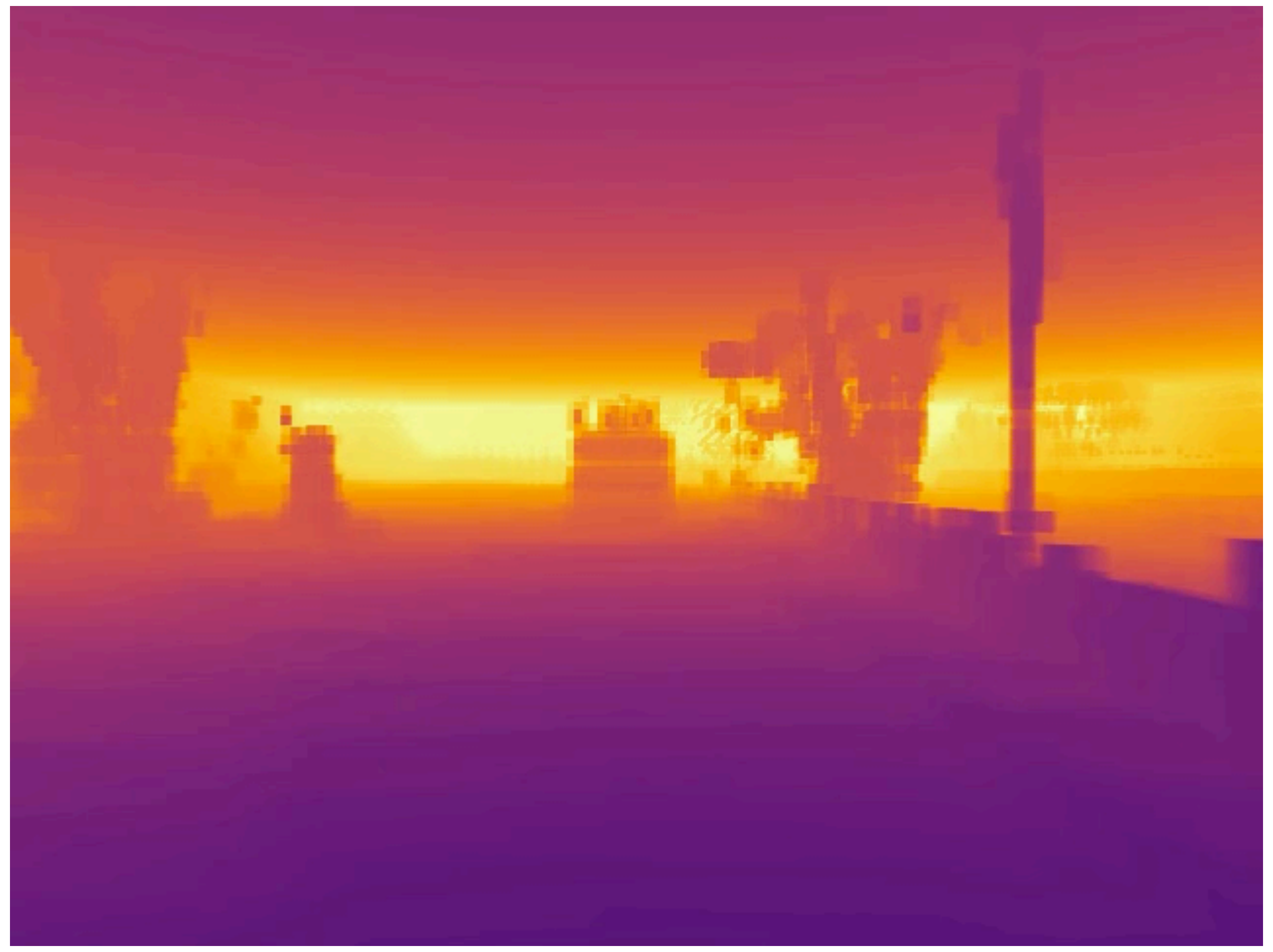


Multi-domain (AV2 + KITTI-O) training does remarkably better than SOTA

Potential applications: Changing extrinsics



Reference RGB frame, t = 0s



Novel-view depth synthesis

Use predicted occupancy to render dense depth maps from novel views (camera).

Summary

- Point cloud forecasting = **4D occupancy forecasting** + sensor extrinsics and intrinsics
- Disentanglement results in dramatic improvement, while also opening up cross-sensor applications
- Benchmarking protocol should evaluate underlying geometry with rays, not uncorrelated points

tarashakhurana/4d-occ-forecasting