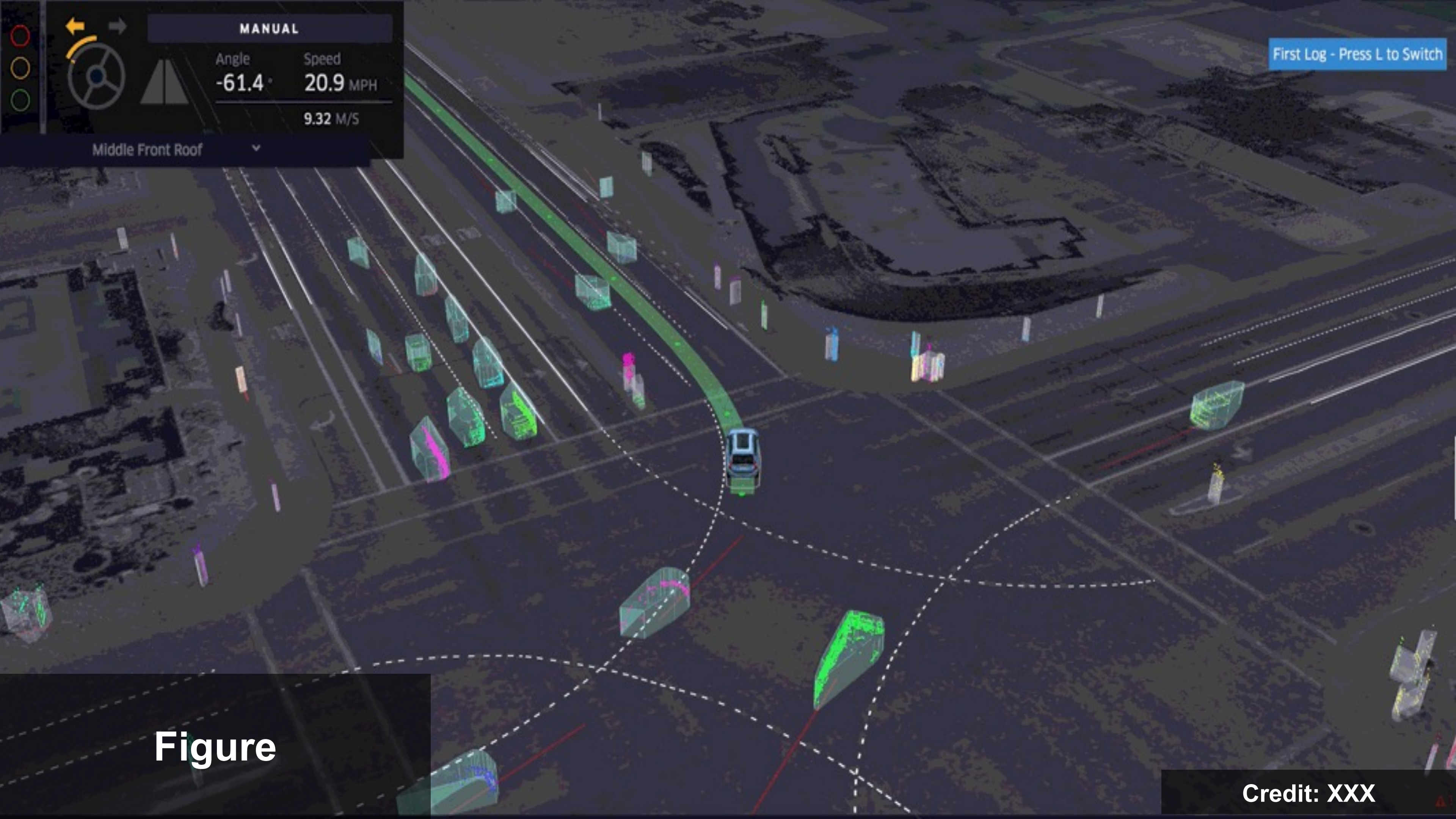
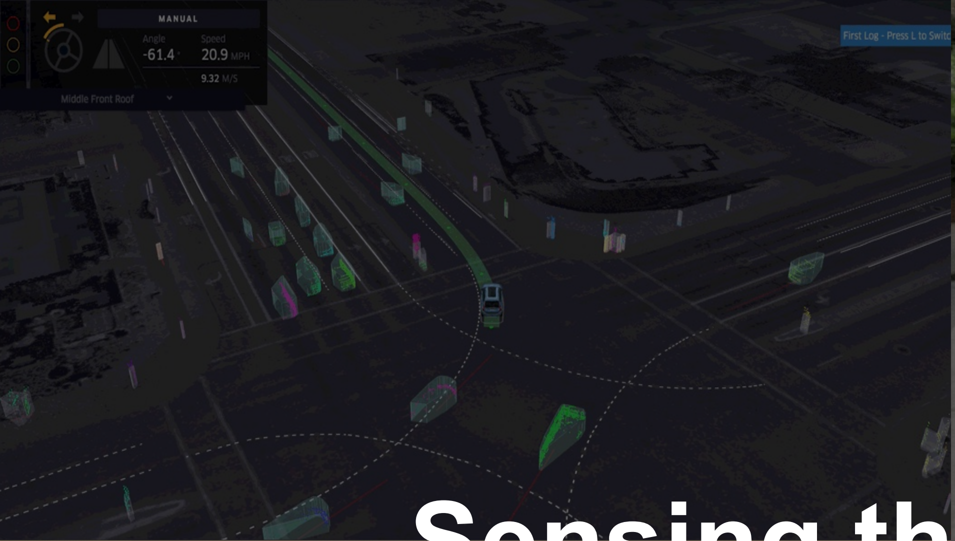
Learning to Compress 3D Visual Data

Shenlong Wang
UIUC


MANUAL
Angle
-61.4
Speed
20.9 MPH
9.32 M/S
Middle Front Roof
First Log - Press L to Switch
Figure
Credit: XXX

Sensing the 3D World



3D Sensing

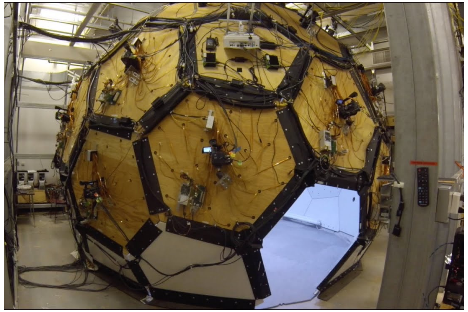
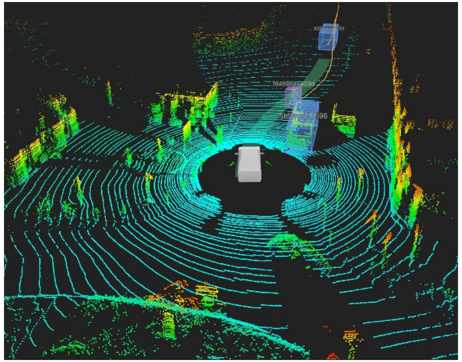


Image credit: CMU, Velodyne, Davis, Imaging Science, Apple, Microsoft

3D Data

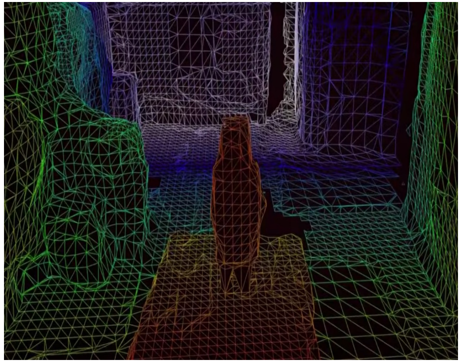
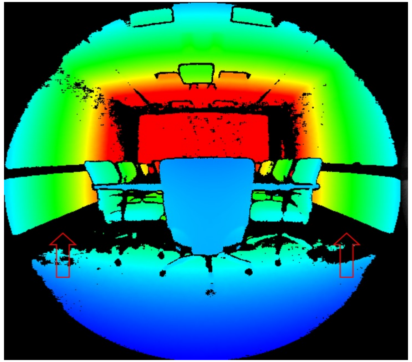
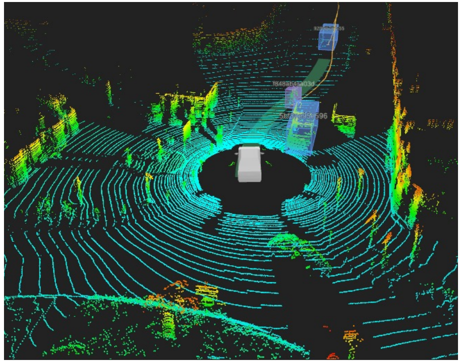
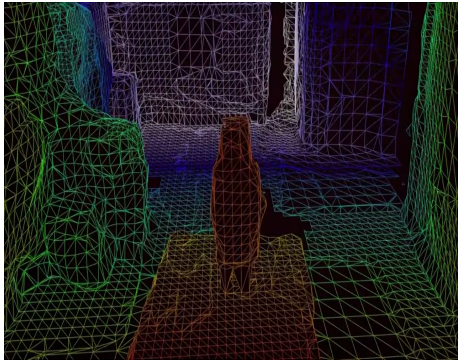



Image credit: Stanford, Velodyne, Davis, Imaging Science, Apple, Microsoft

3D Data





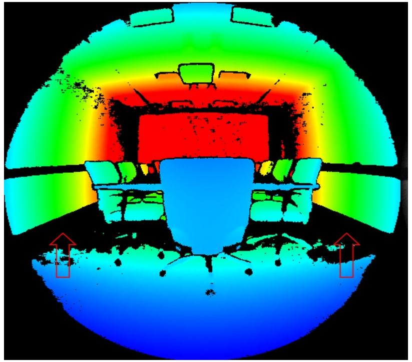
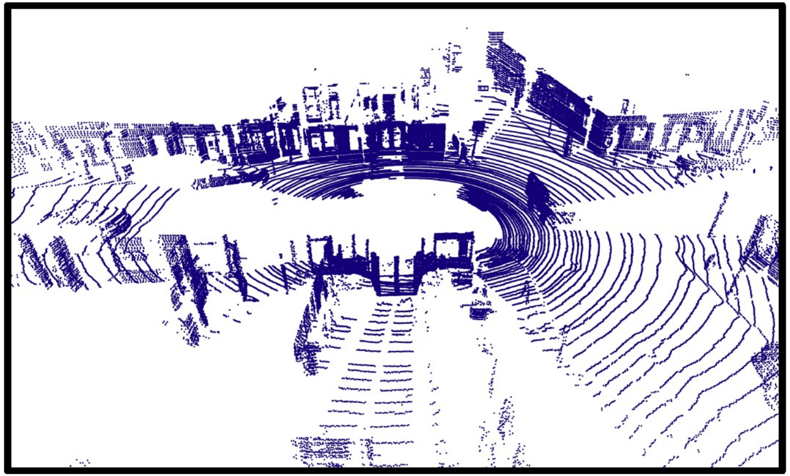


Image credit: Stanford, Velodyne, Davis, Imaging Science, Apple, Microsoft

3D Requires Storage



> 1TB /
8hr

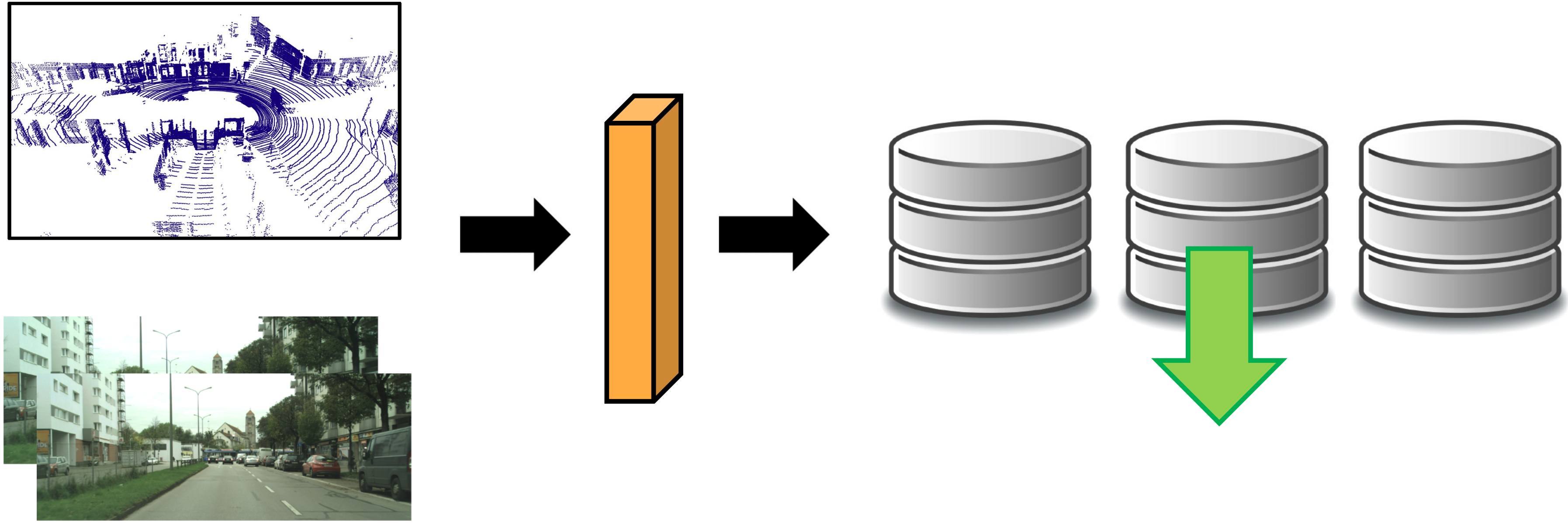
3D Requires Storage

Left Cam

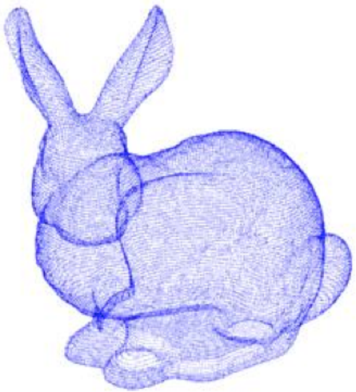
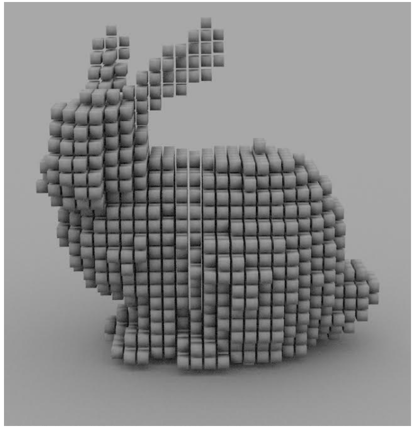
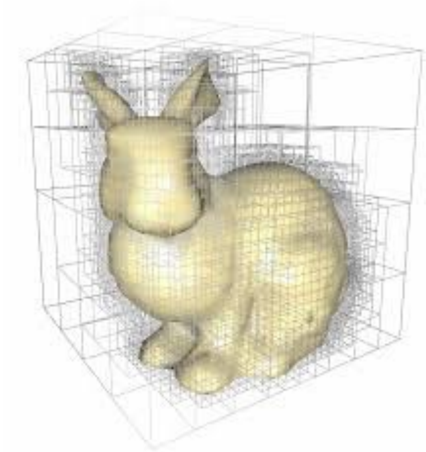
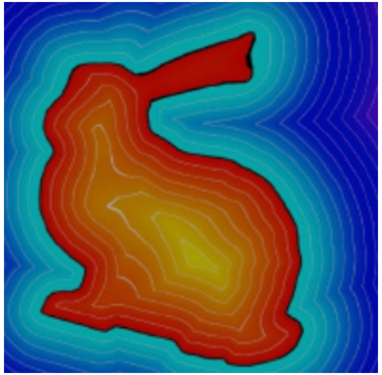
Right Cam

> 2TB /
8hr

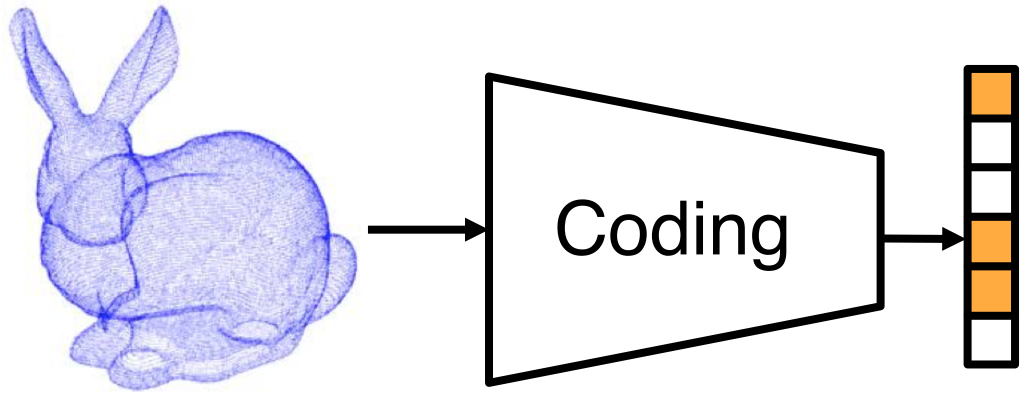
Overview



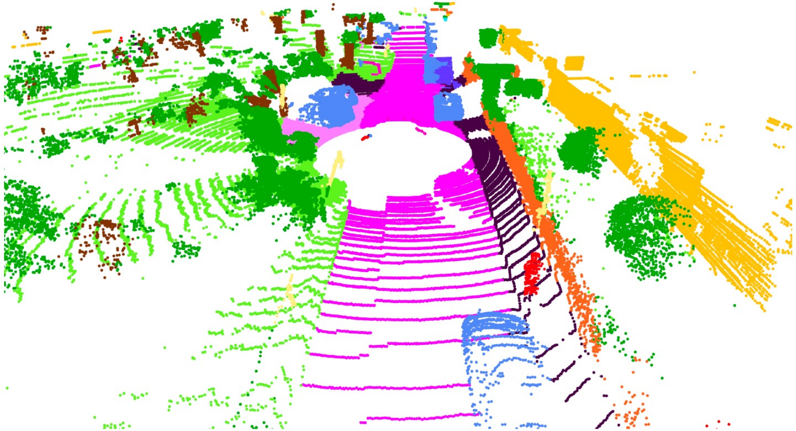
Key Challenge I: 3D Data Structure

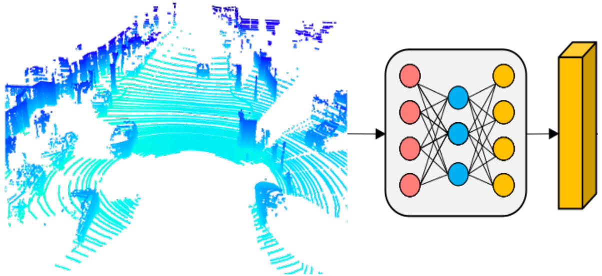
Key Challenge II: Coding



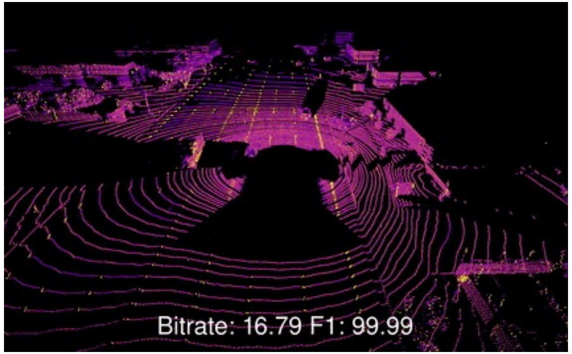
Key Challenge III: Varying Downstream Task



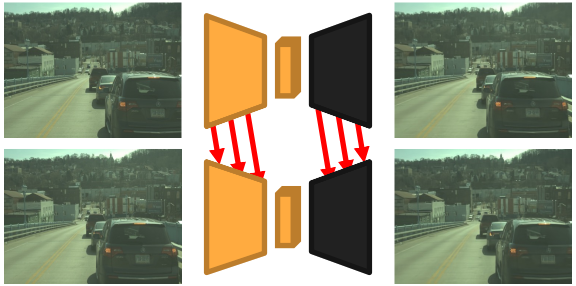
Overview



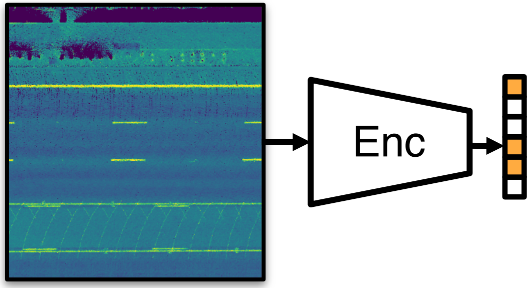
Learning to Compress
LiDAR



Learning to Compress
LiDAR Stream



Learning to Compress
Stereo Cameras



Learning to Compress
Maps for Localization

Stereo Compression

- Compress these stereo pairs to **reduce storage** while **maximizing quality**.

Left Cam

Right Cam

Key Challenge

- How can we use **shared information** between stereo pairs?

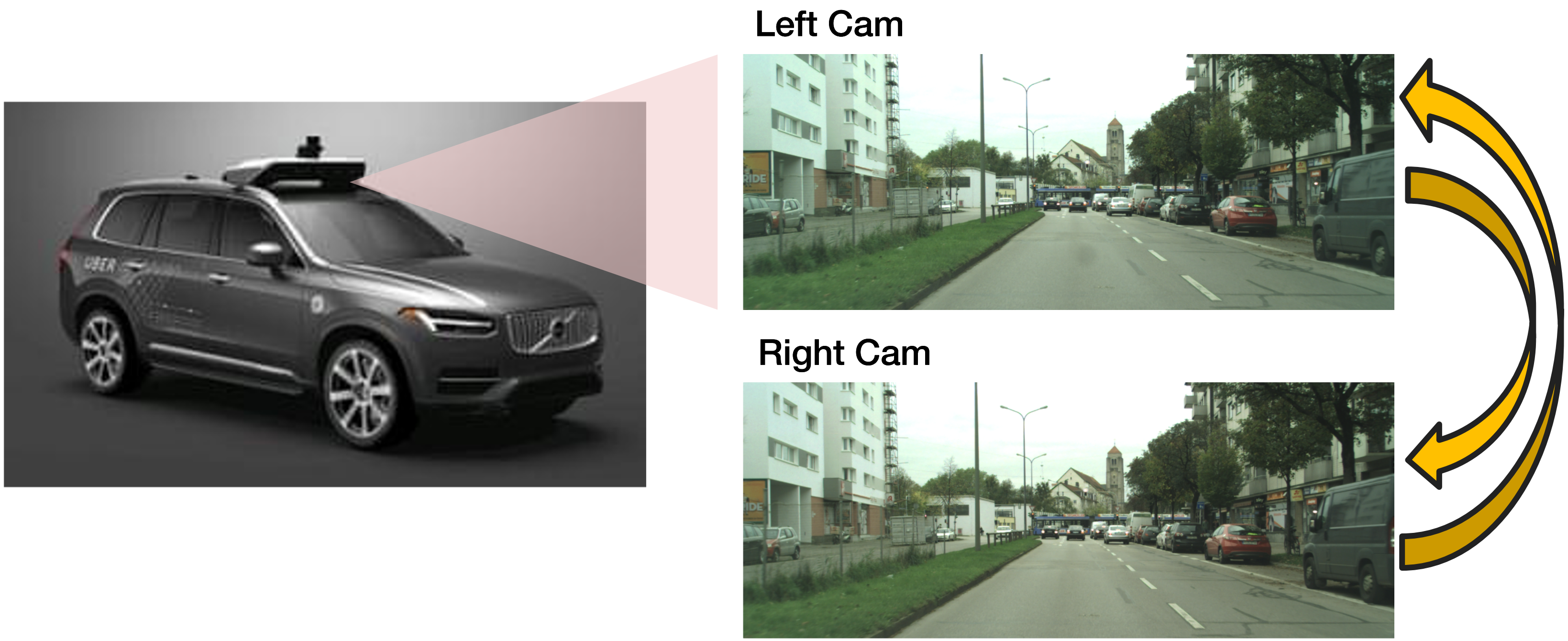
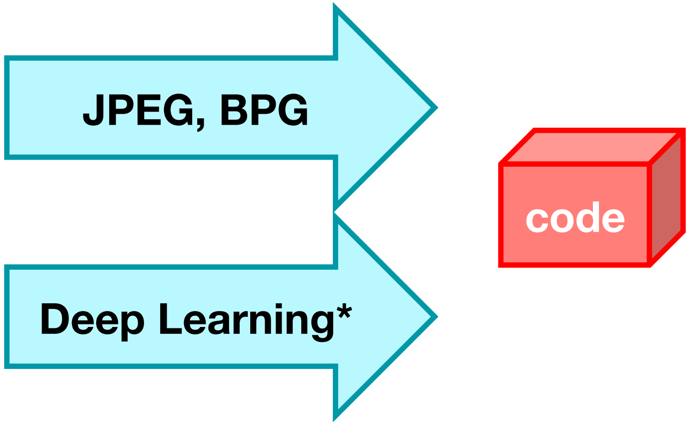


Image Compression

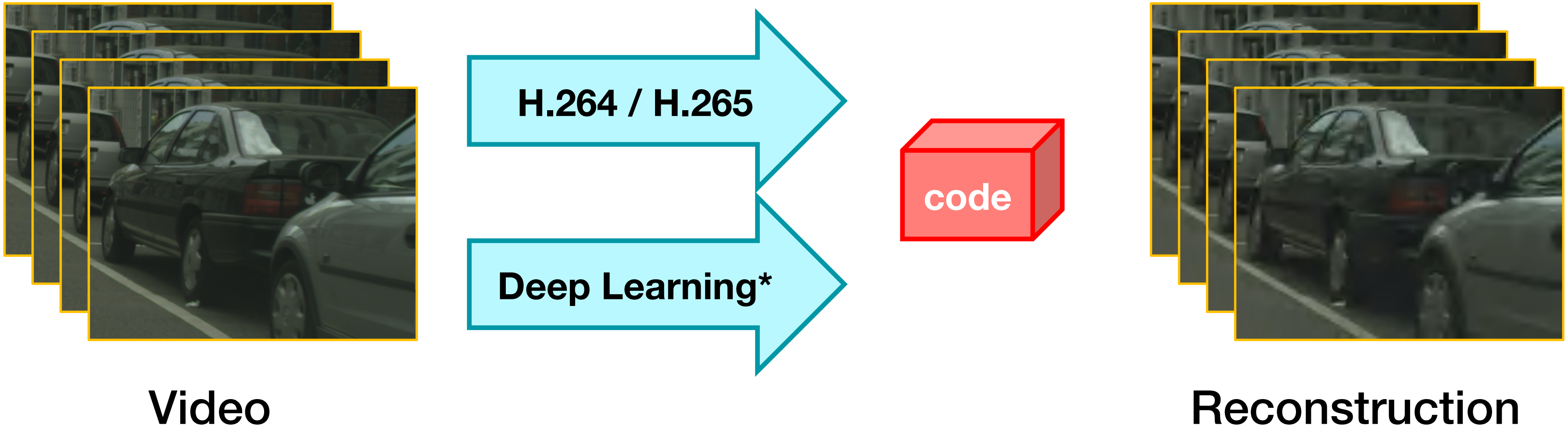


Image

Reconstruction

Transform coding to exploit redundancy in image

Video Compression



Estimate motion + **coding residual** to exploit temporal dependencies

Our Approach

- We design a neural architecture for **compressing stereo images**.
- Achieve a **30-50%** reduction in the second image bitrate at low bitrates, compared to single-image compression.

Left Cam

Right Cam

Our Approach

Key Insight: The information from the first image should both improve reconstruction and reduce bitrate for second image.

Left Cam

Right Cam

Our Approach

Key Insight: The information from the first image should both improve reconstruction and reduce bitrate for second image.

Proposed Solution:

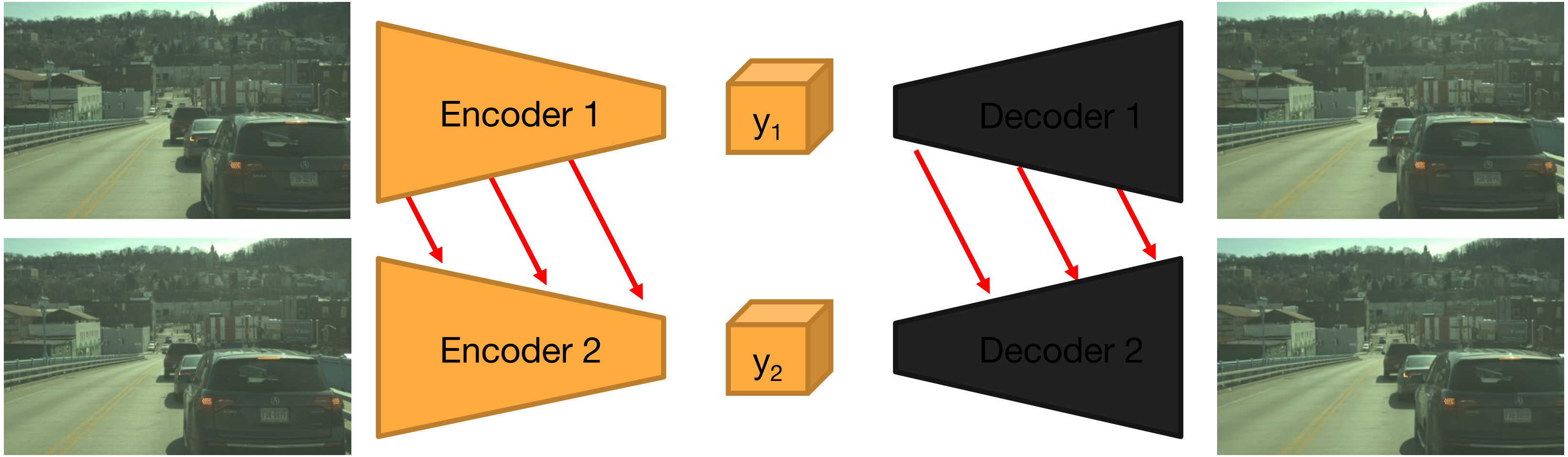
- Feature level: **parametric skip function**
- Entropy level: **conditional entropy model**

Left Cam

Right Cam

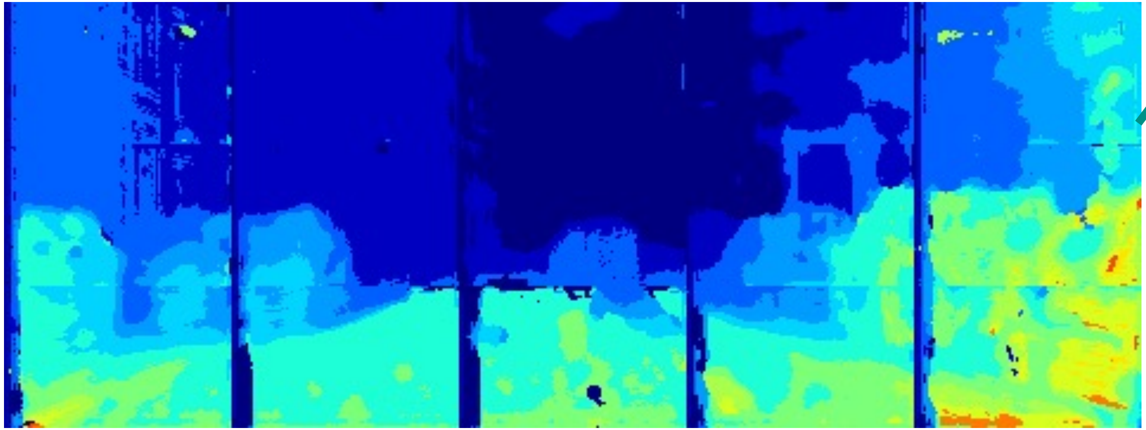
Architecture

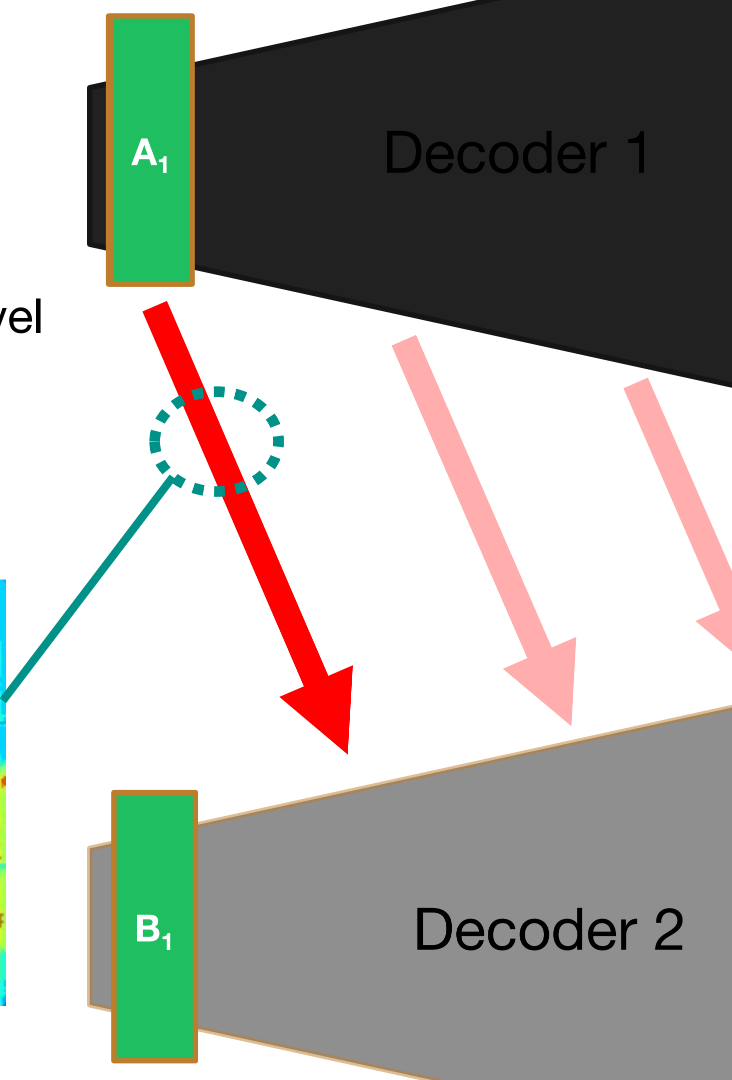
- Two encoder/decoder copies (inspired from Ballé et al. 2018)
- Feed features with **parametric skip functions**



Parametric Skip Functions

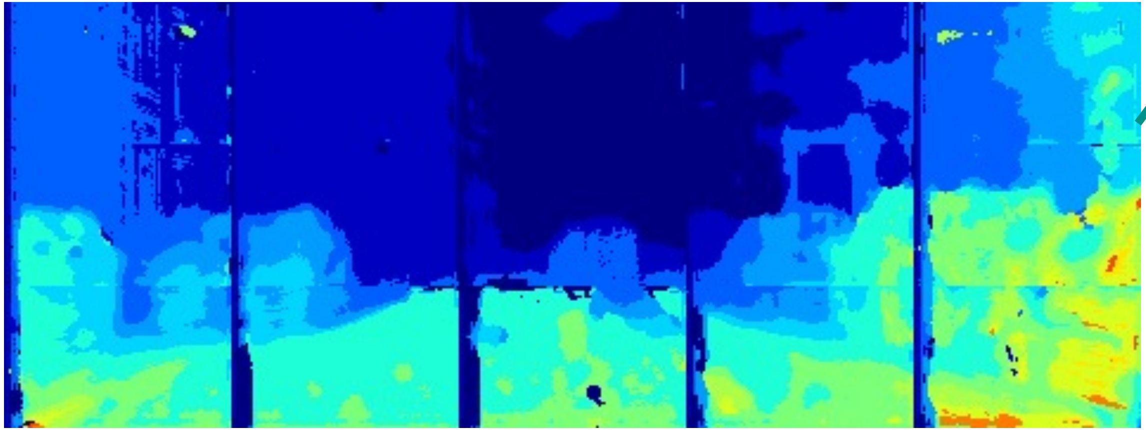
- Predict a **disparity map** at every single feature level

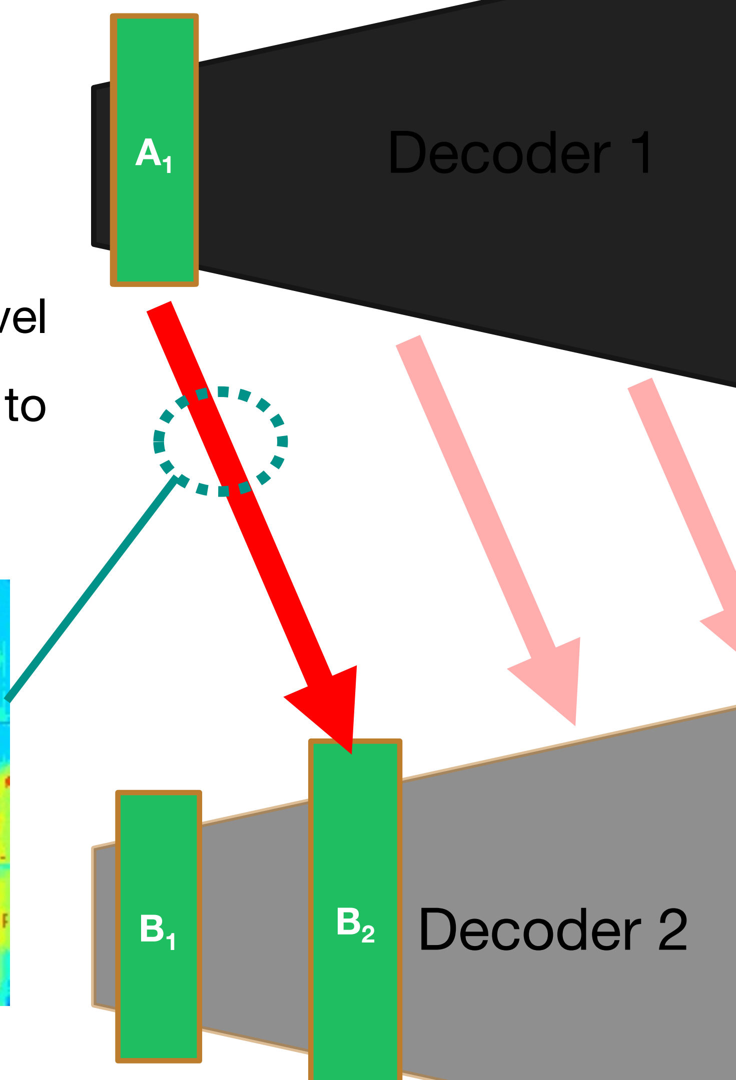




Parametric Skip Functions

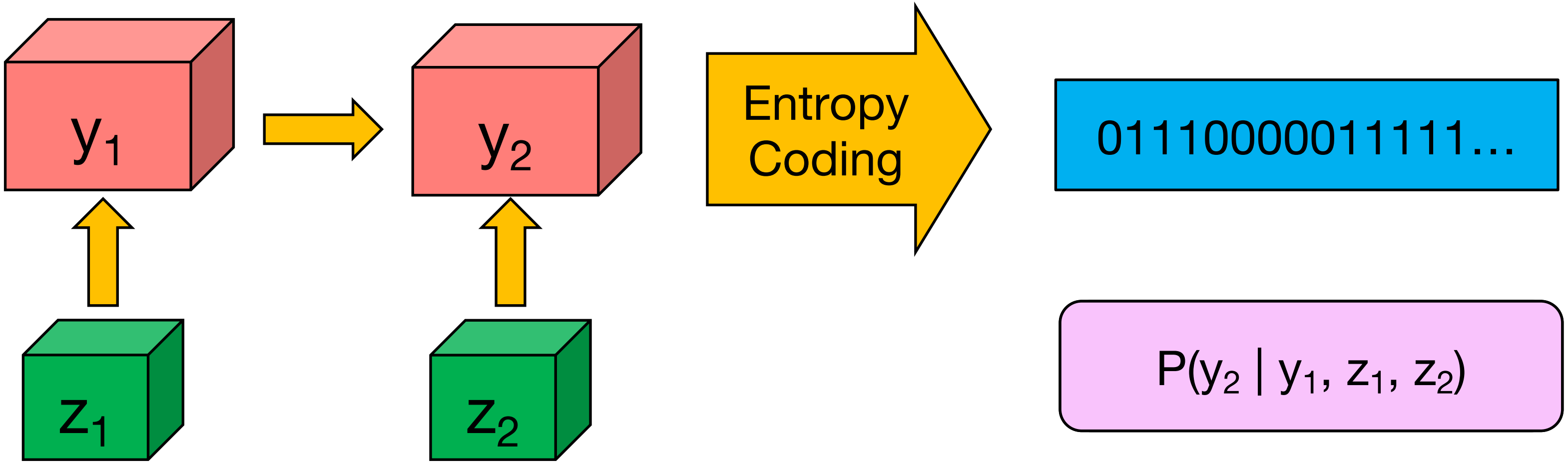
- Predict a **disparity map** at every single feature level
- Use the disparity map to **warp** camera 1 features to camera 2





Conditional Entropy Model

- **Entropy coding / Arithmetic coding:** losslessly encode symbols into bitstream
- **Fit probability model:** high probability symbols → low entropy → low bitrate



Dataset

NorthAmerica

Cityscapes

Quantitative Results

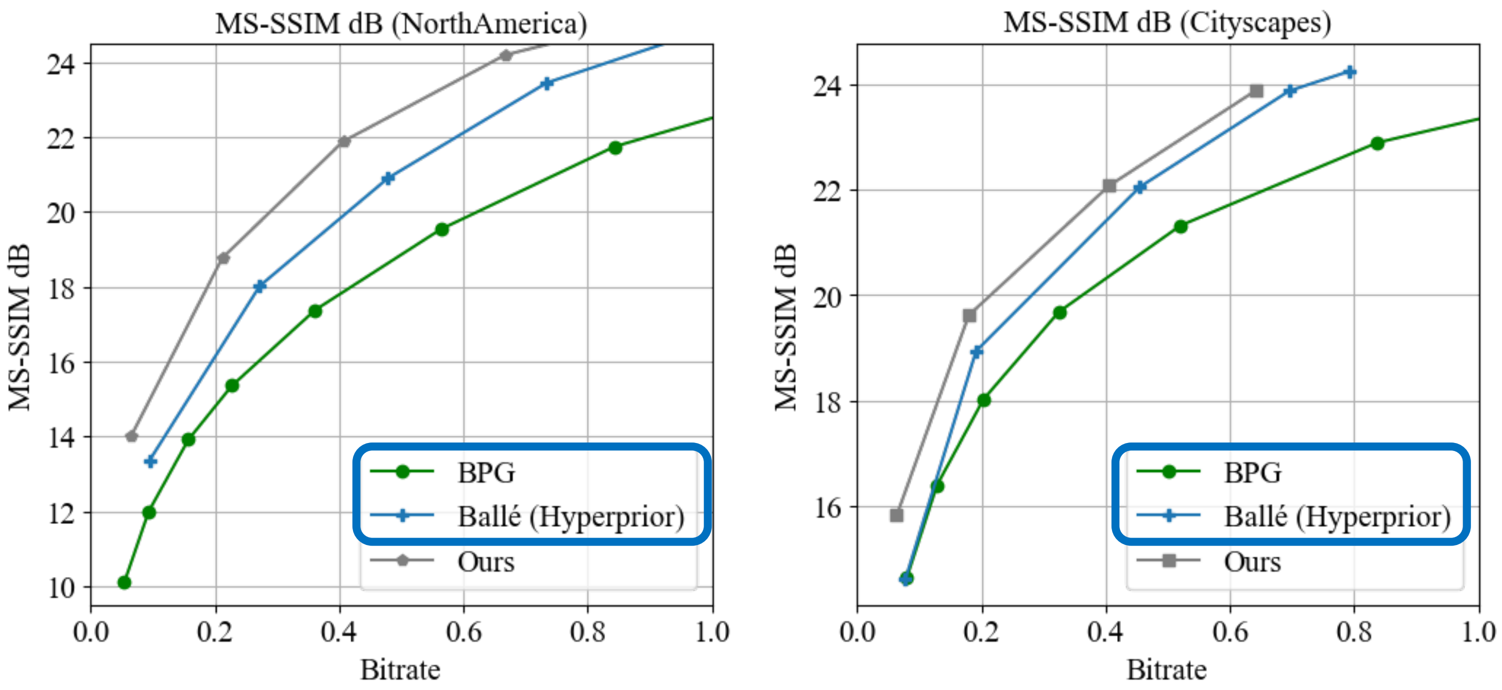
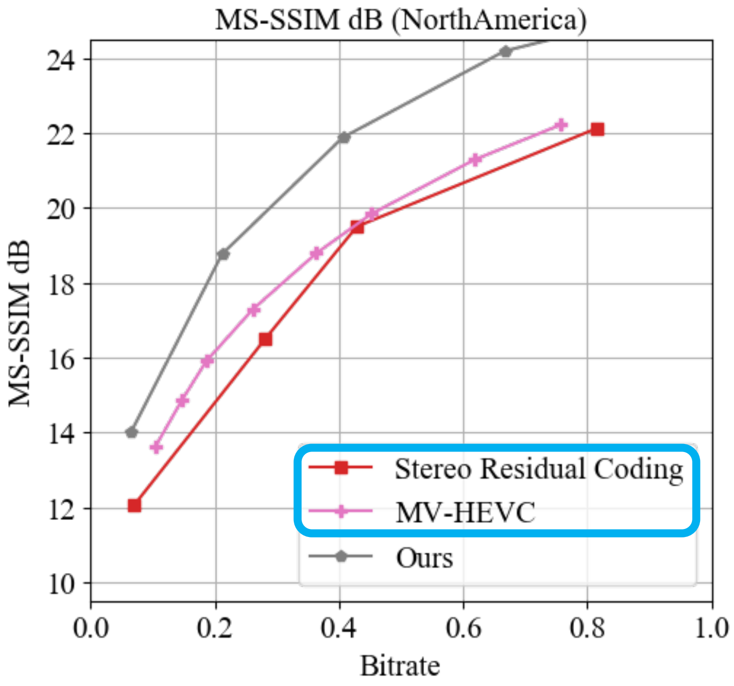
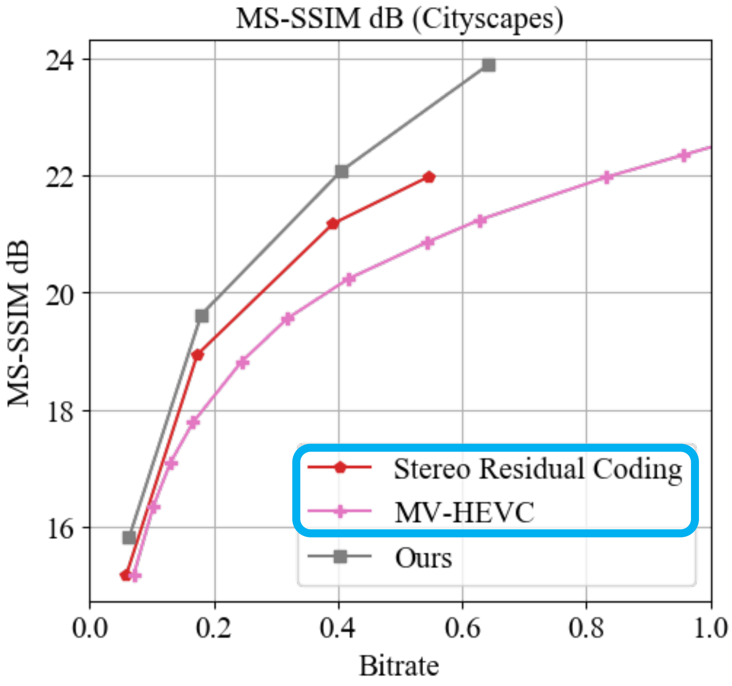


Image Encoding Baselines

MS-SSIM – takes into account structural quality of the reconstruction

Quantitative Results





Residual Encoding Baselines

MS-SSIM – takes into account structural quality of the reconstruction

Ballé (single-image)

Ours

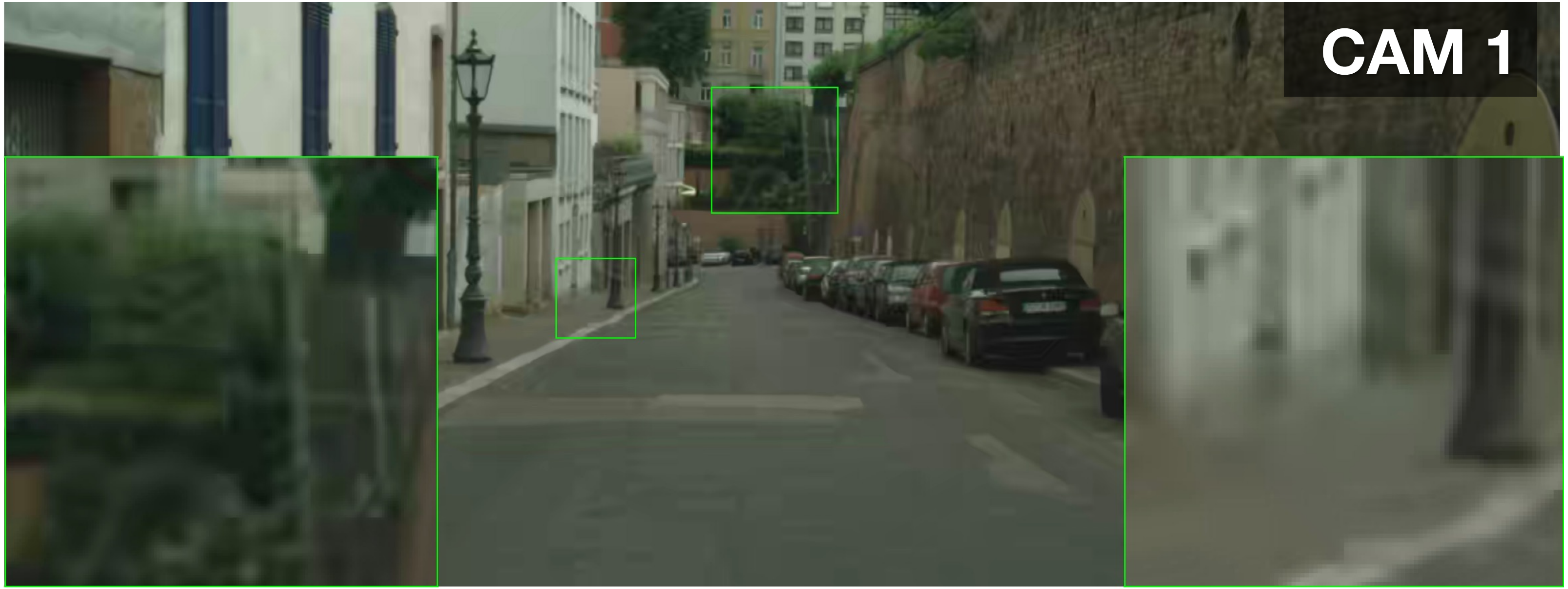
~18% bitrate reduction with higher MS-SSIM

Ballé (single-image)

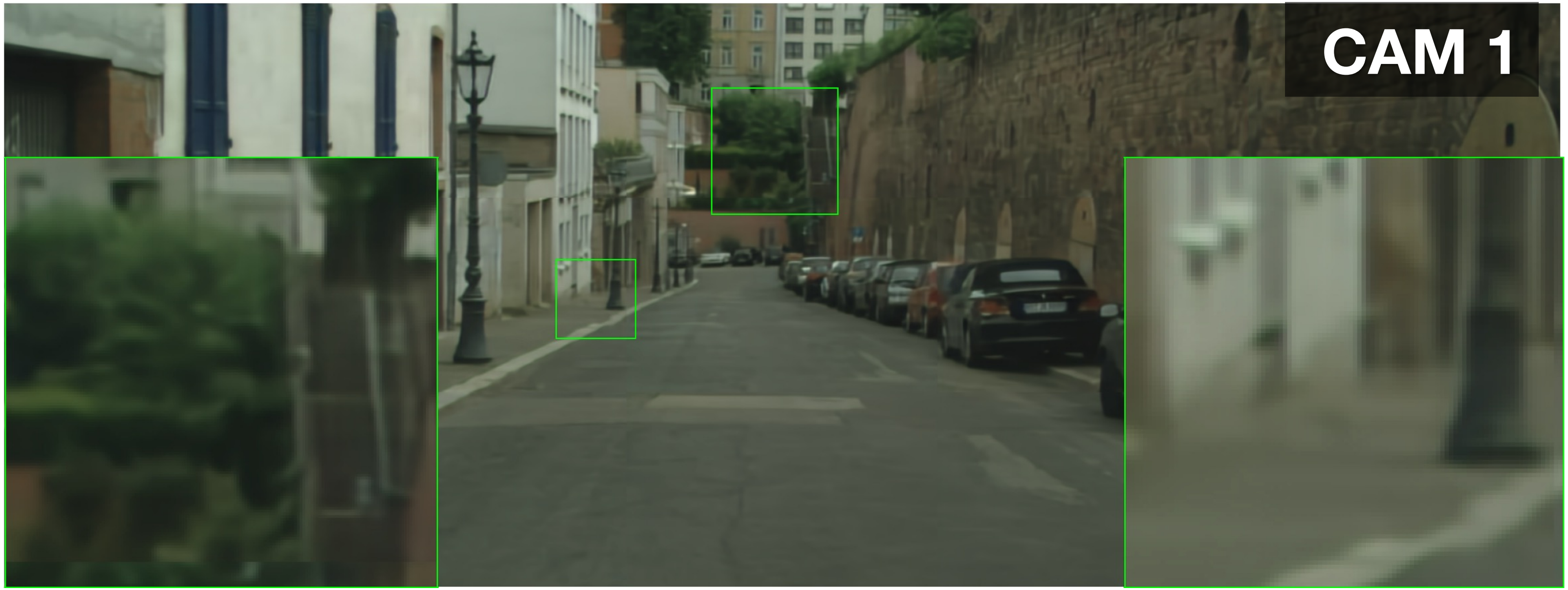
Ours

~24% bitrate reduction with higher MS-SSIM

BPG

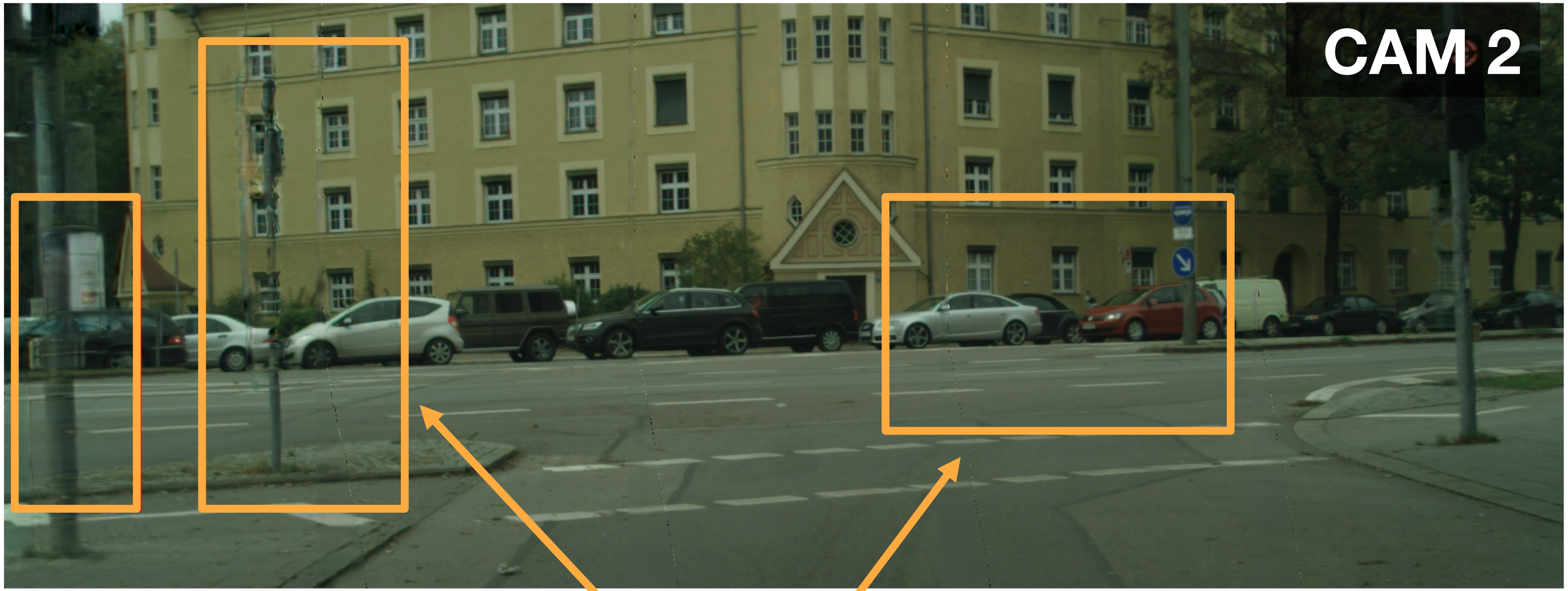


Ours



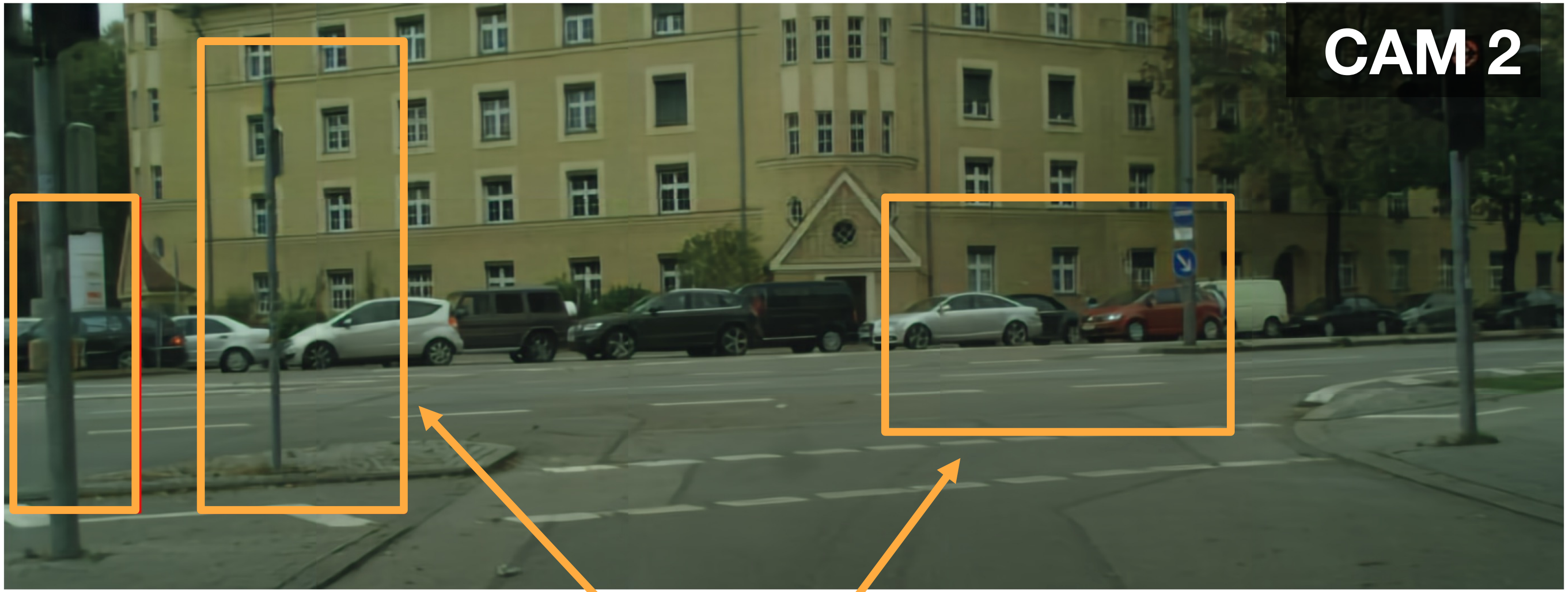
~24% bitrate reduction with much higher MS-SSIM

Stereo Residual Coding



Artifacts/tears

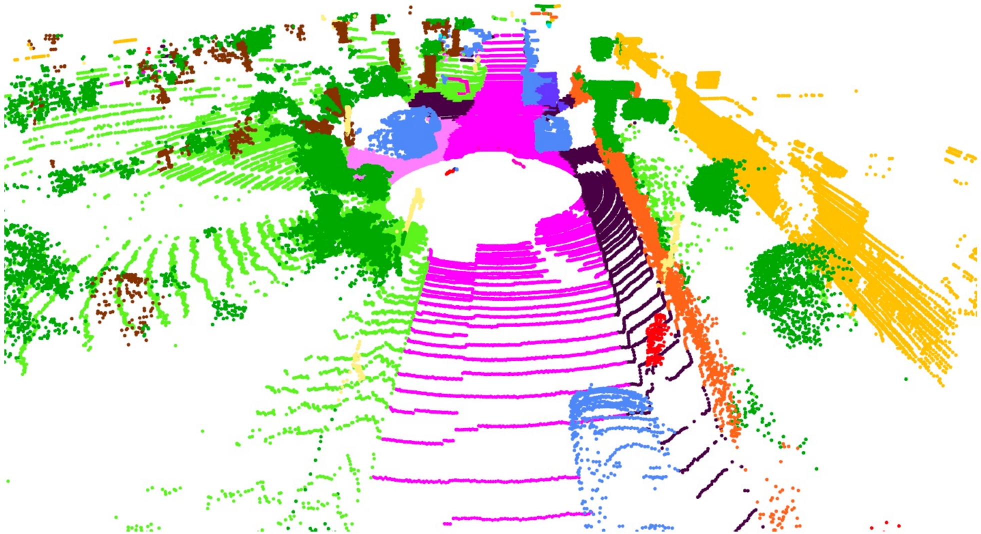
Ours



No artifacts/tears

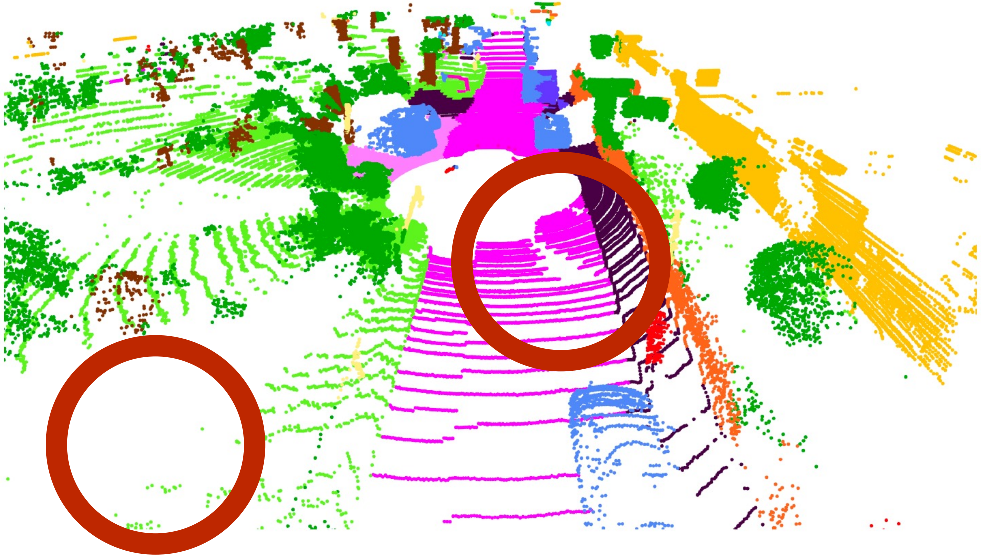
Goal

- Reduce **storage size** without compromising **downstream tasks**



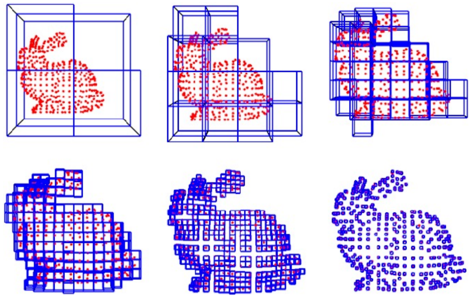
Challenges

- LiDAR point clouds are **sparse** and spatially **non-uniform**

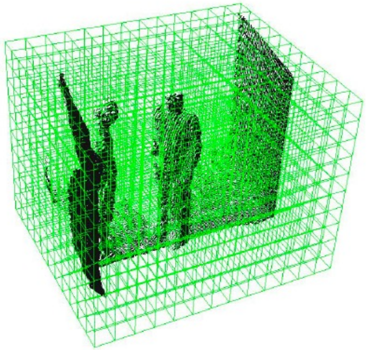


Prior Work

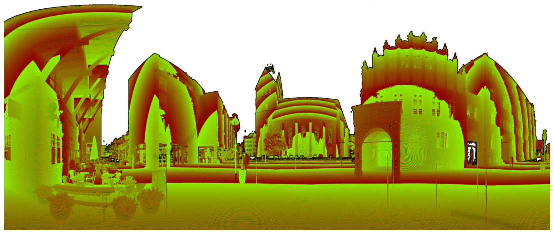
- Memory-efficient data structure
- Powerful density estimation model



Kammerl et al. 2012

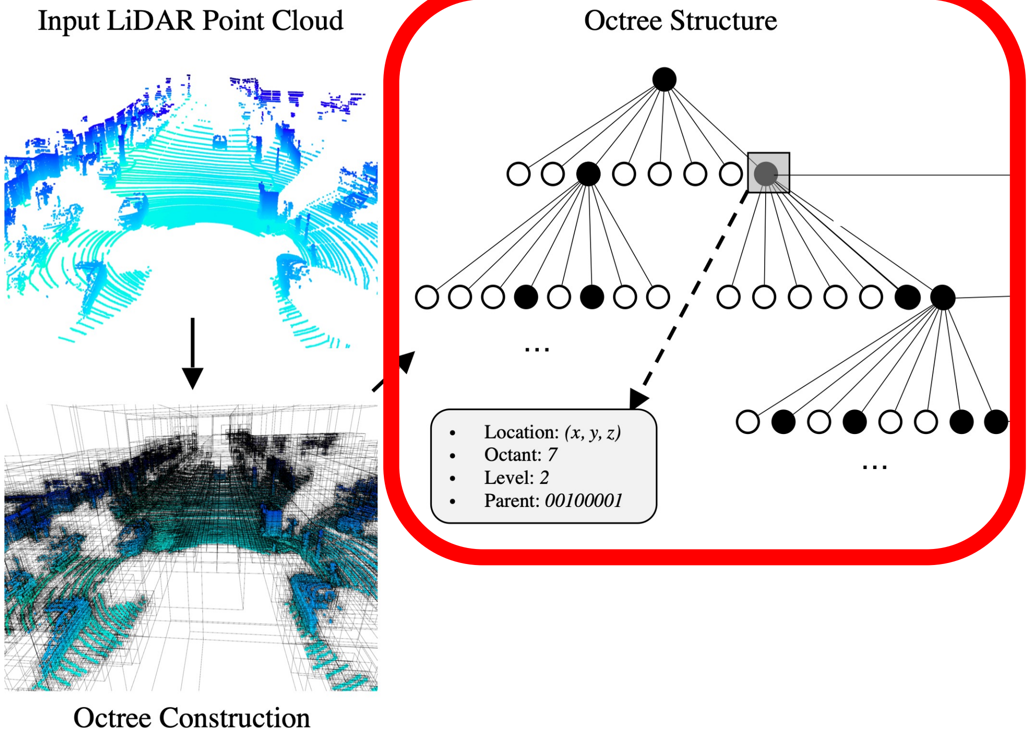


Wang et al. 2012

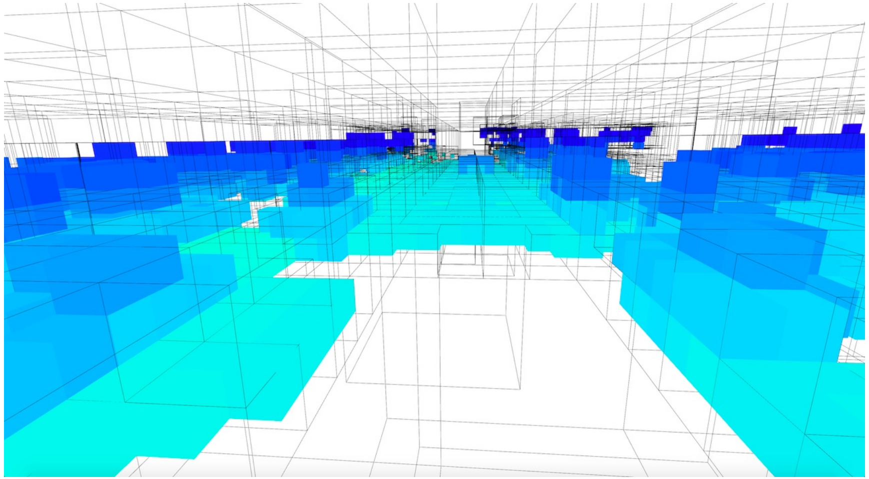


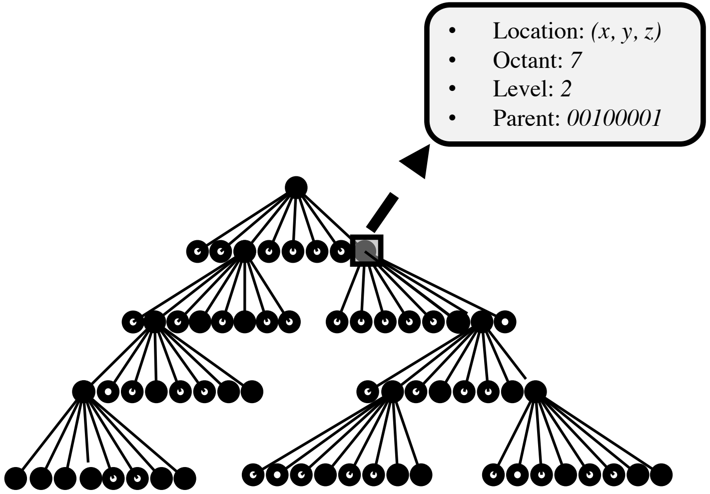
Houshiar et al. 2015

Our Method

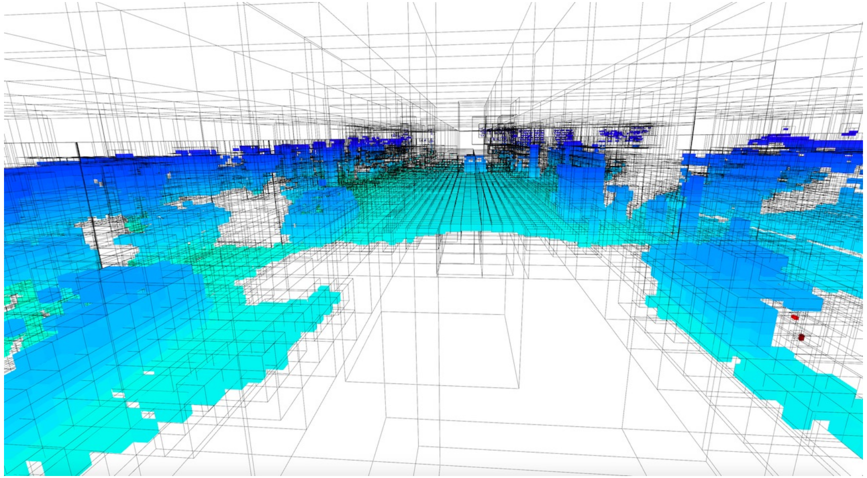

Input LiDAR Point Cloud
Octree Structure
Location: (x, y, z)
Octant: 7
Level: 2
Parent: 00100001
...
...
Octree Construction

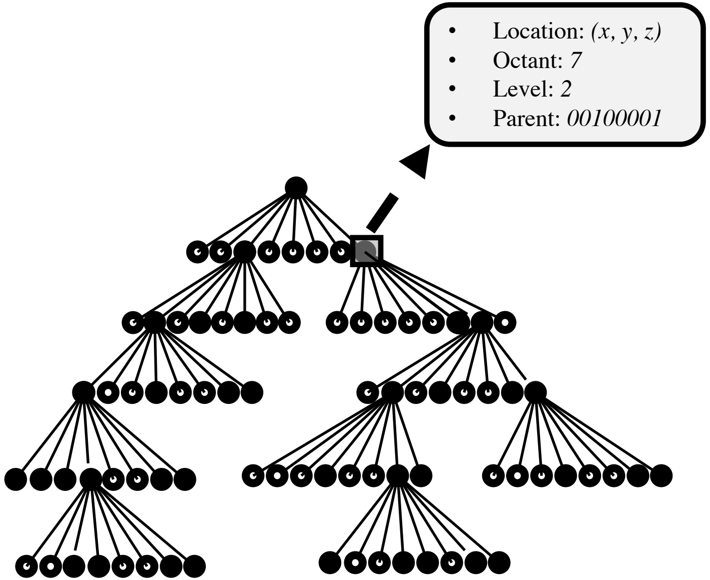
Octree



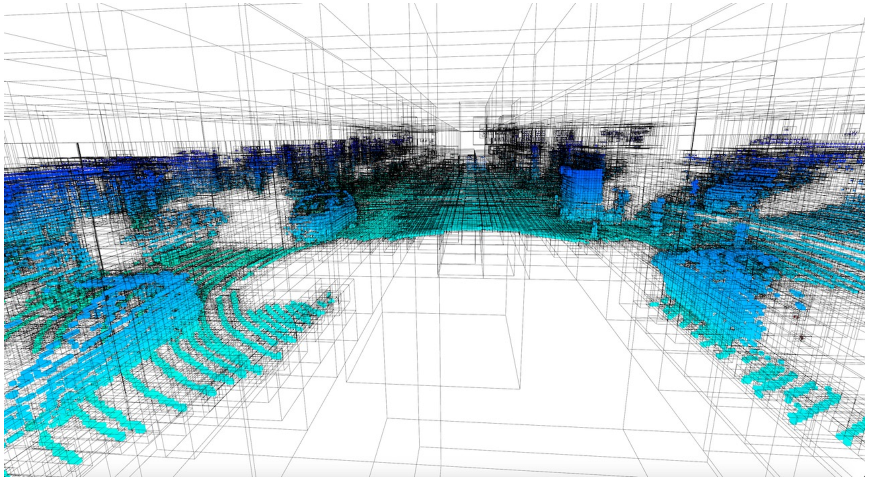


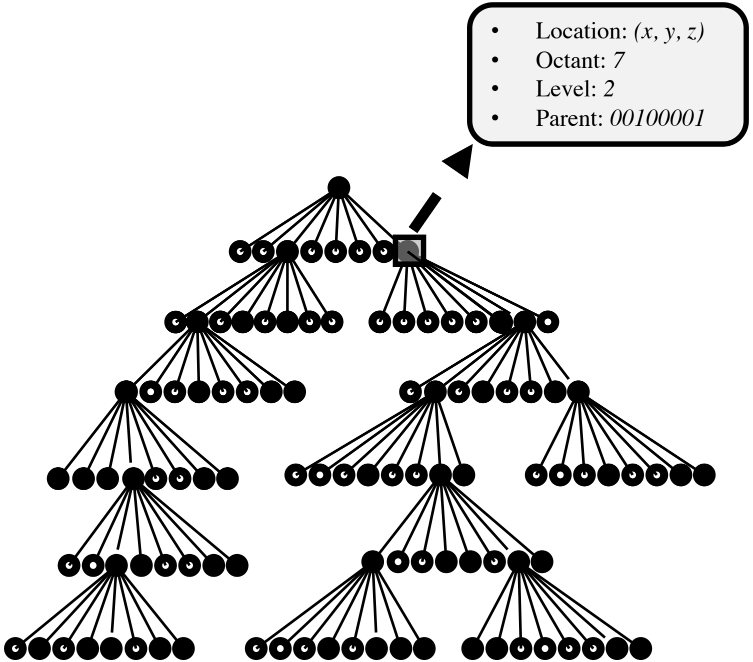
Octree



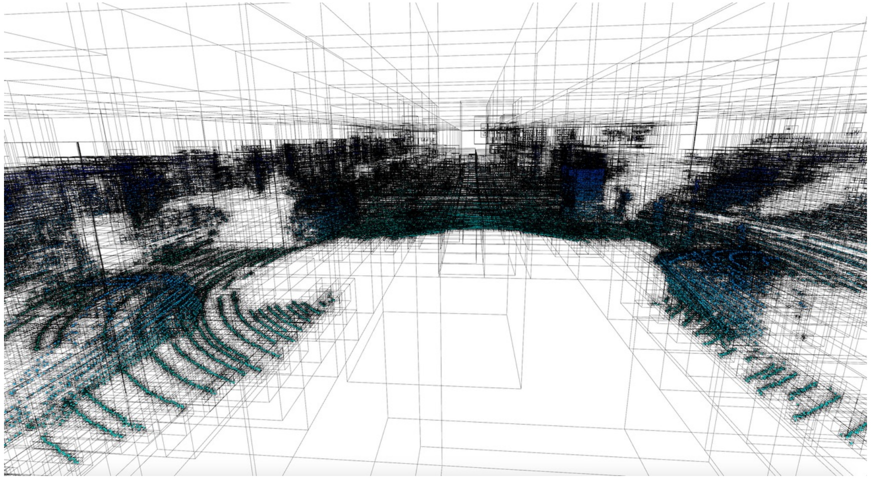


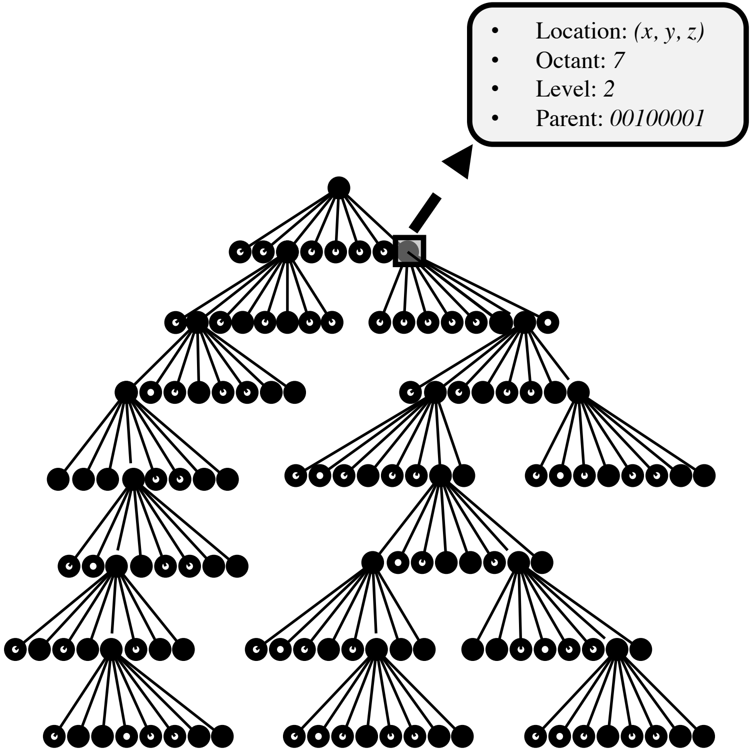
Octree



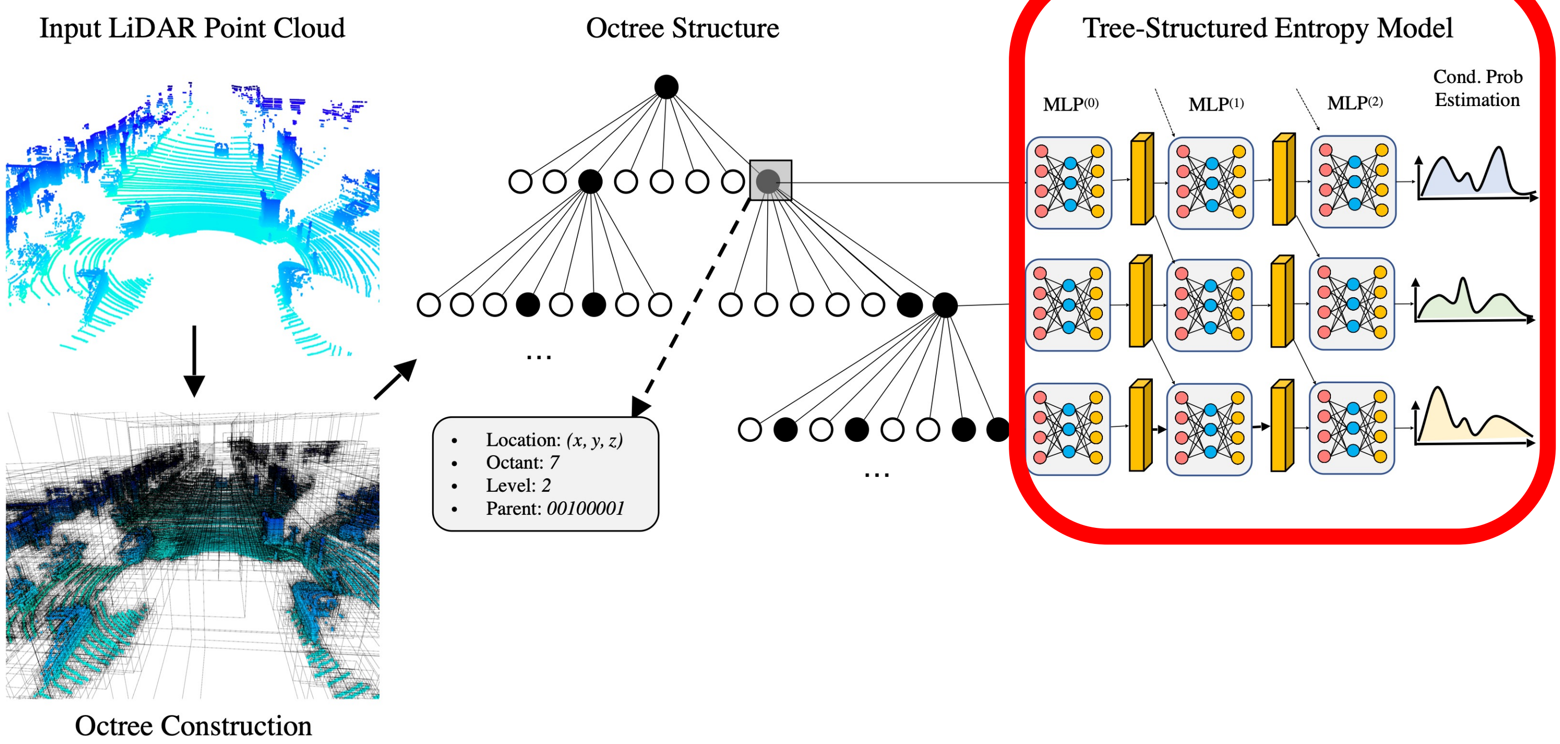


Octree



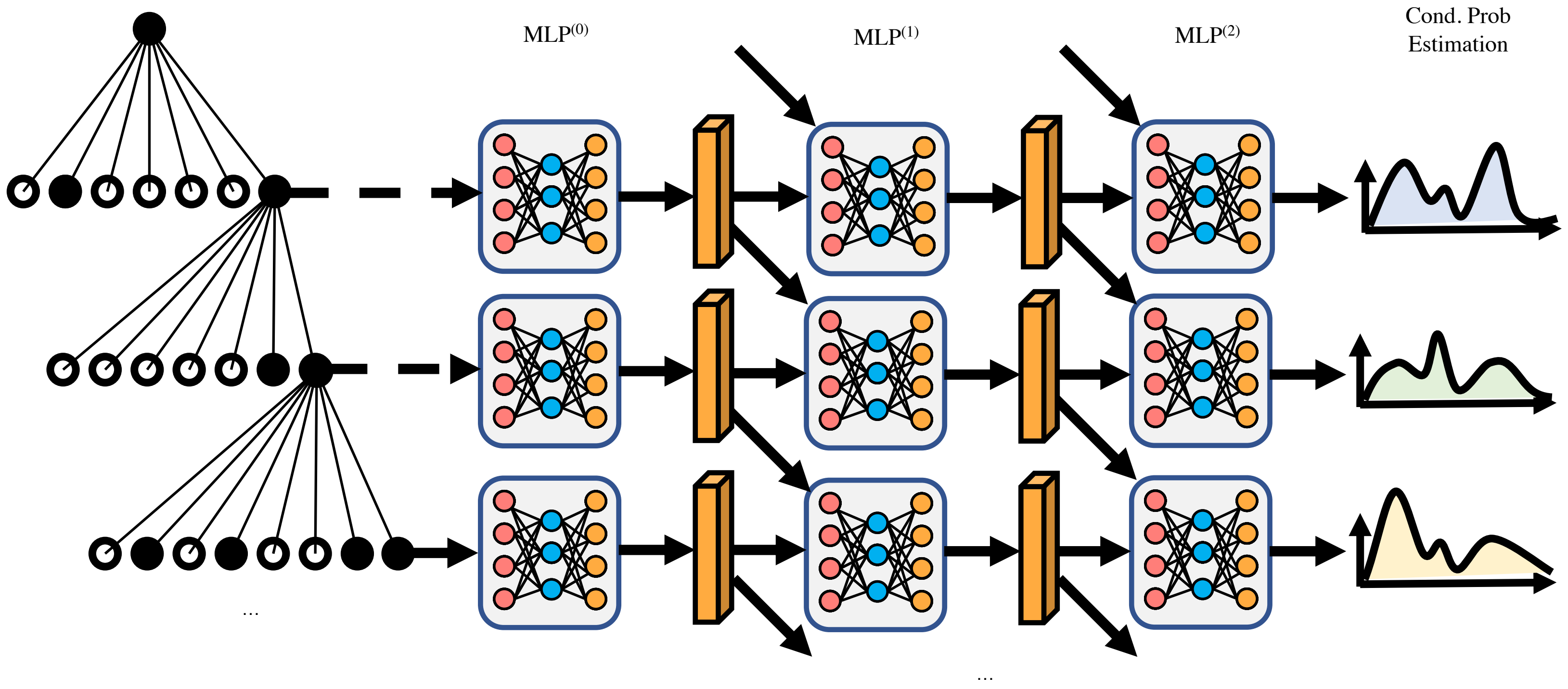


Octree

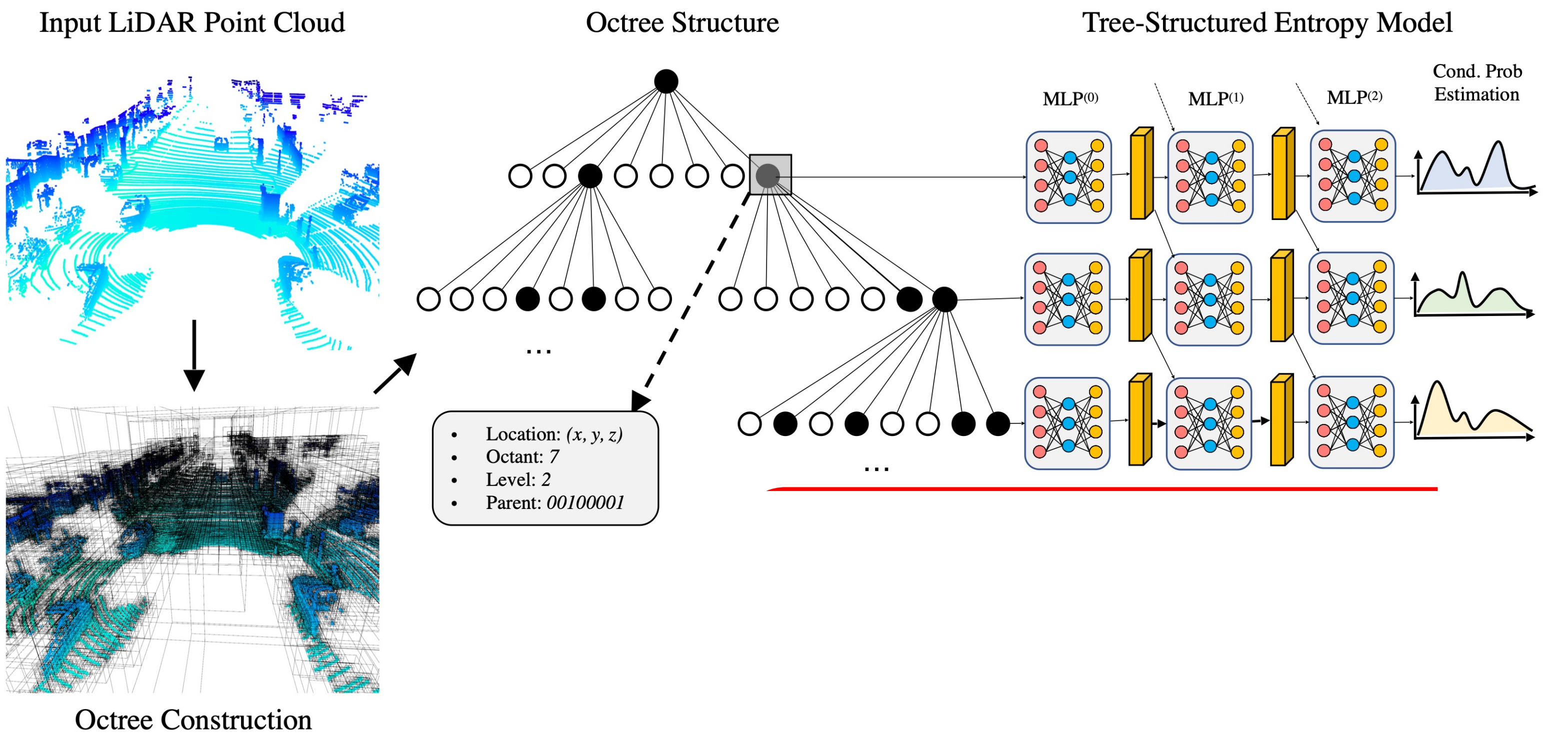


Input LiDAR Point Cloud
Octree Construction
Octree Structure
...
Location: (x, y, z)
Octant: 7
Level: 2
Parent: 00100001
...
Tree-Structured Entropy Model
MLP(0)
MLP(1)
MLP(2)
Cond. Prob
Estimation

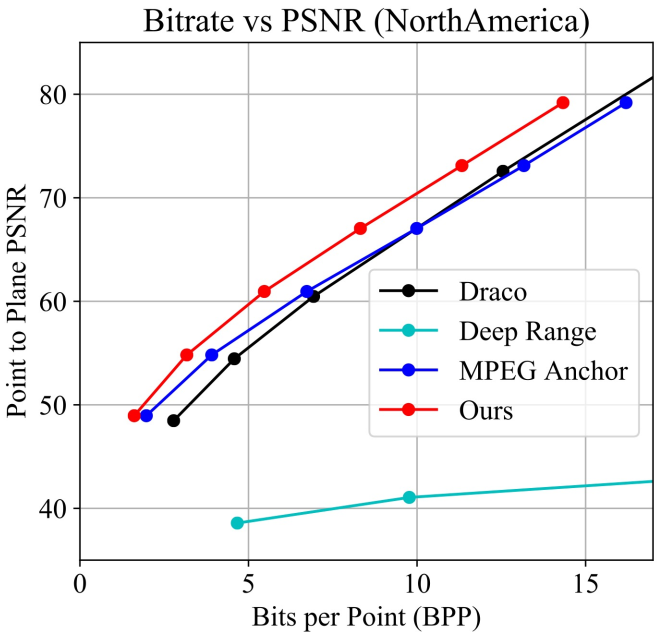
Tree-Structured Entropy Model

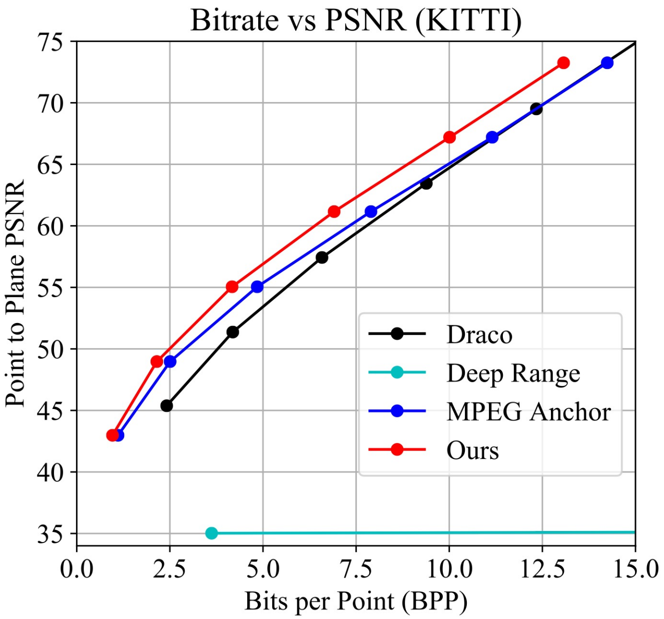


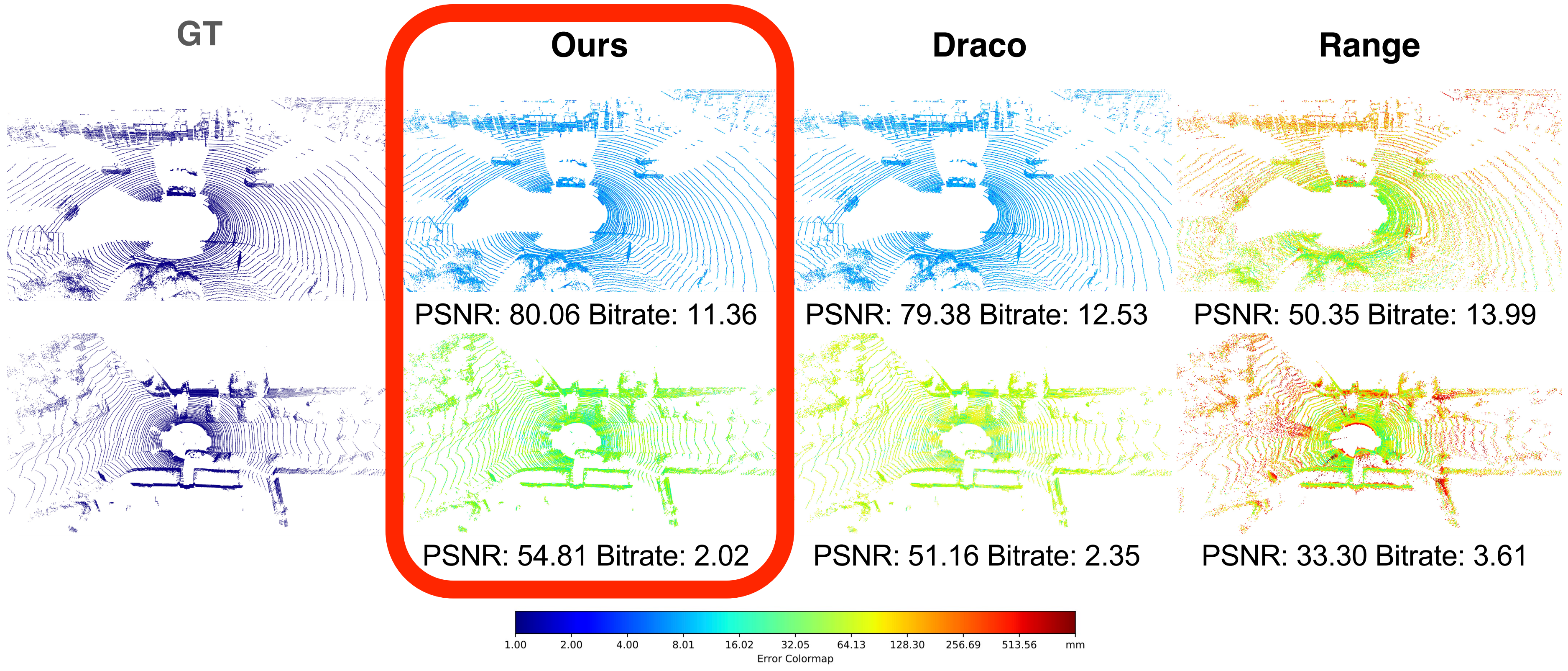
Our Method

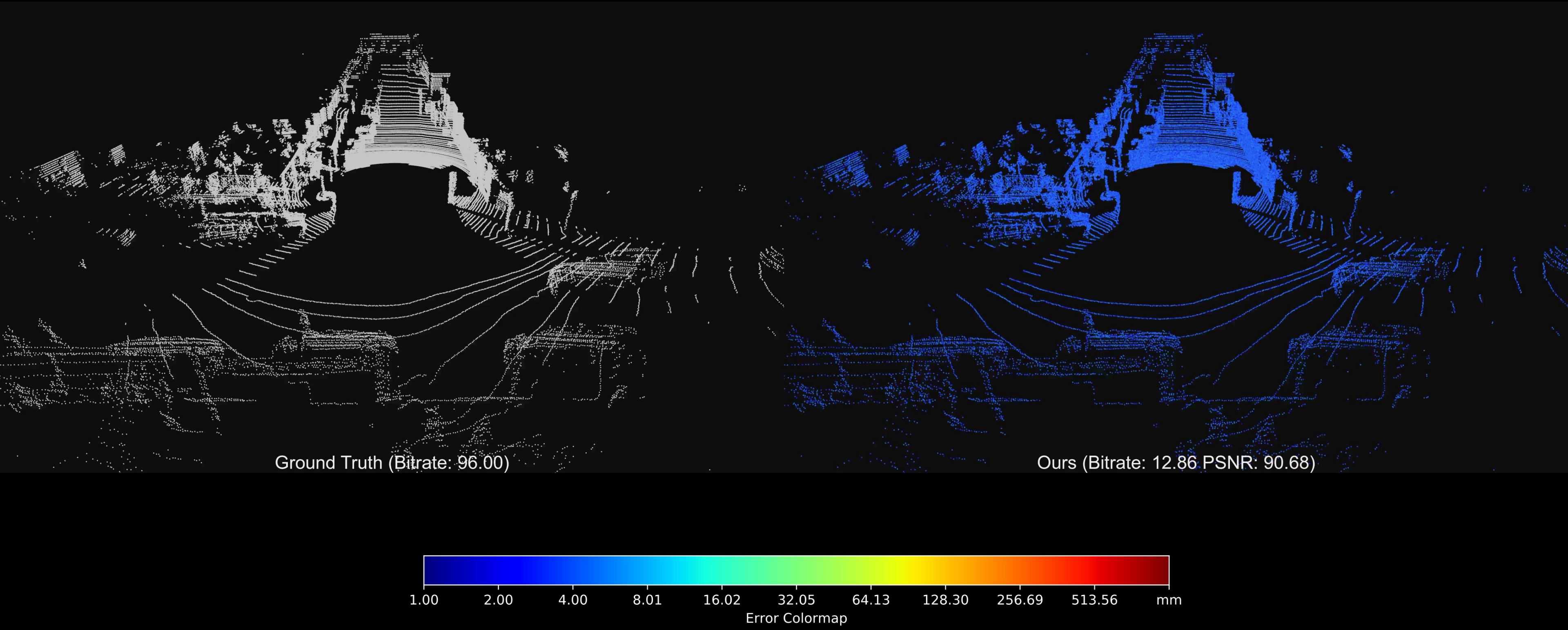

Input LiDAR Point Cloud
Octree Construction
Octree Structure
...
Location: (x, y, z)
Octant: 7
Level: 2
Parent: 00100001
Tree-Structured Entropy Model
MLP(0)
MLP(1)
MLP(2)
Cond. Prob Estimation

Reconstruction

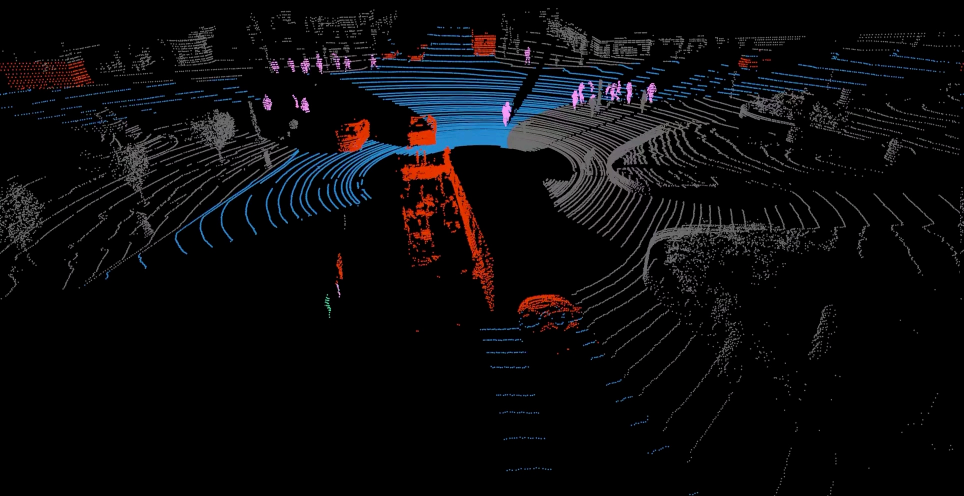




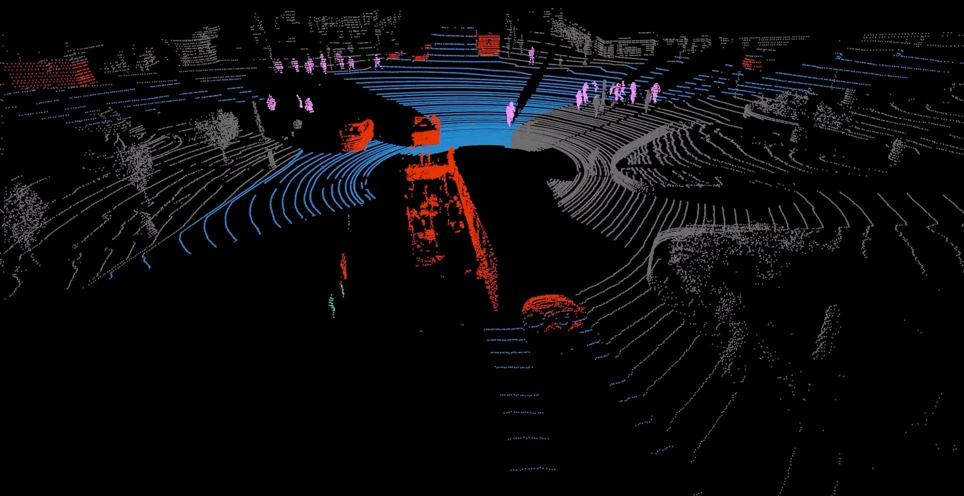

GT
Ours
Draco
Range
PSNR: 80.06 Bitrate: 11.36
PSNR: 79.38 Bitrate: 12.53
PSNR: 50.35 Bitrate: 13.99
PSNR: 54.81 Bitrate: 2.02
PSNR: 51.16 Bitrate: 2.35
PSNR: 33.30 Bitrate: 3.61
1.00 2.00 4.00 8.01 16.02 32.05 64.13 128.30 256.69 513.56 mm
Error Colormap



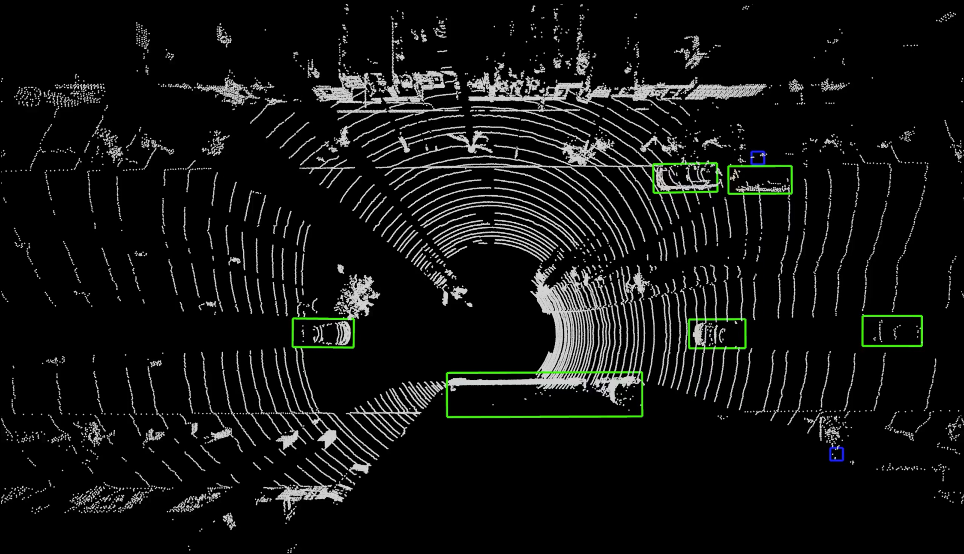
Ground Truth (Bitrate: 96.00)
Ours (Bitrate: 12.86 PSNR: 90.68)
1.00
2.00
4.00
8.01
16.02
32.05
64.13
128.30
256.69
513.56
mm
Error Colormap



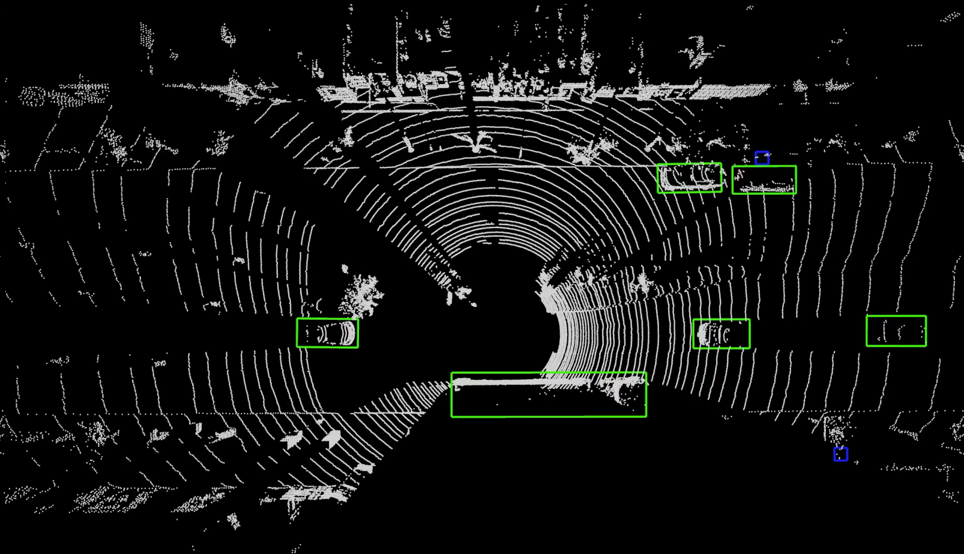
Oracle (Bitrate: 96.00 IOU: 73.99)



Ours (Bitrate: 14.61 IOU: 70.94)



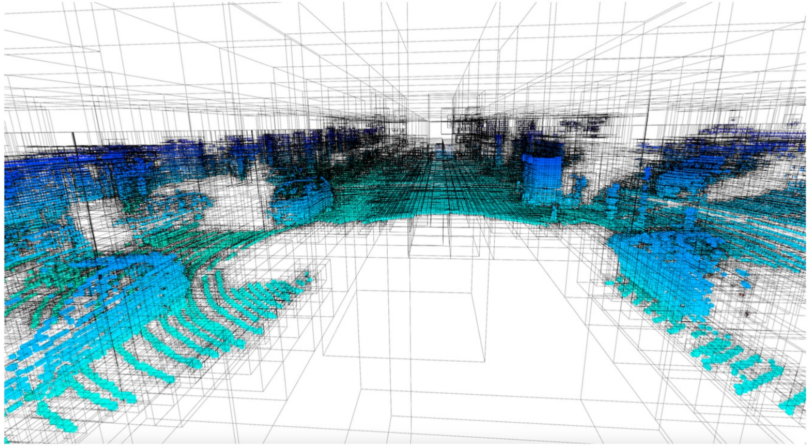
Oracle (Bitrate: 96.00 AP: 72.54)



Ours (Bitrate: 12.91 AP: 71.49)

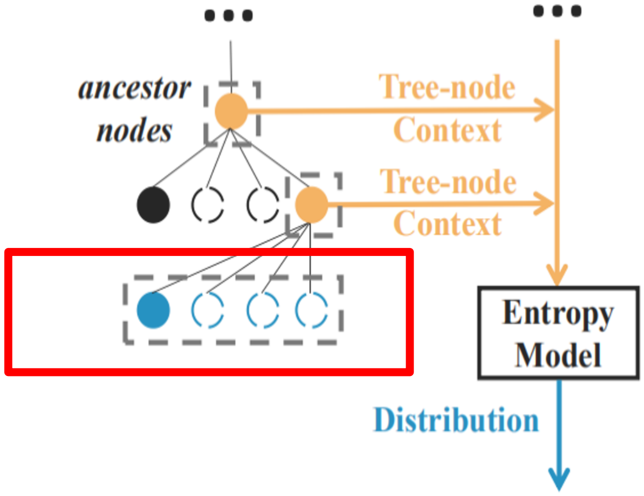
Learning from Sparse Voxel Octree

- Compressed representation
- Coarse-to-fine structured
- Controllable geometry loss

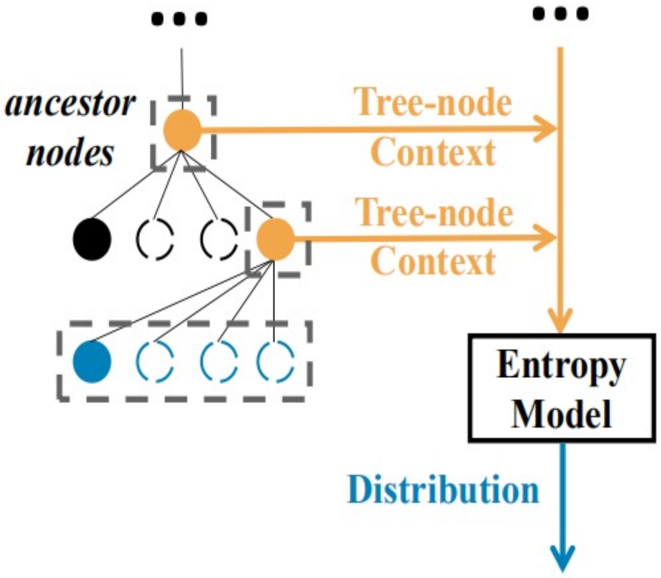


What is missing?

- Could we exploit richer context over space and time?
- How would we compress attributes associated with each point?



Tree-node context



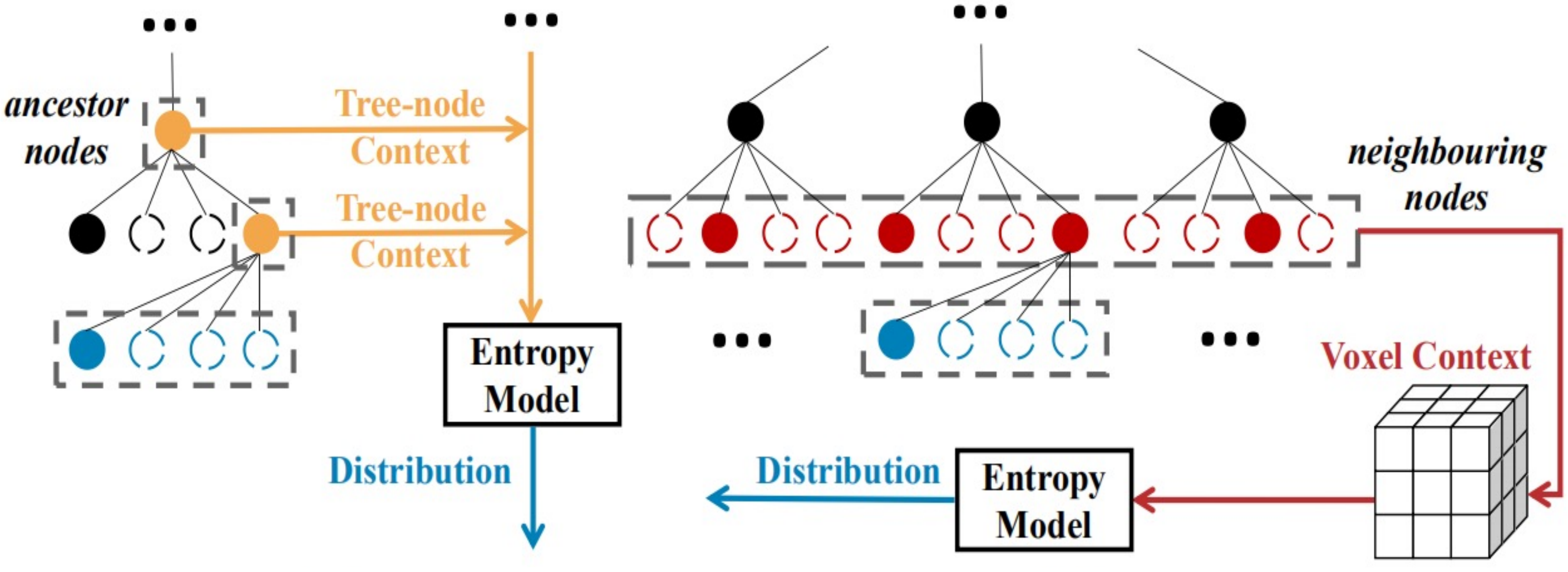
(a) tree-node context in OctSqueeze

- Tree-node context:
  - Exploit context information from ancestor nodes
  - Ignore the strong prior information between *spatial neighbouring nodes* at the same depth level

Huang L, Wang S, Wong K, et al. OctSqueeze: Octree-structured entropy model for LiDAR compression. In *CVPR*, 2020.

Que, Lu, Xu, CVPR 2021

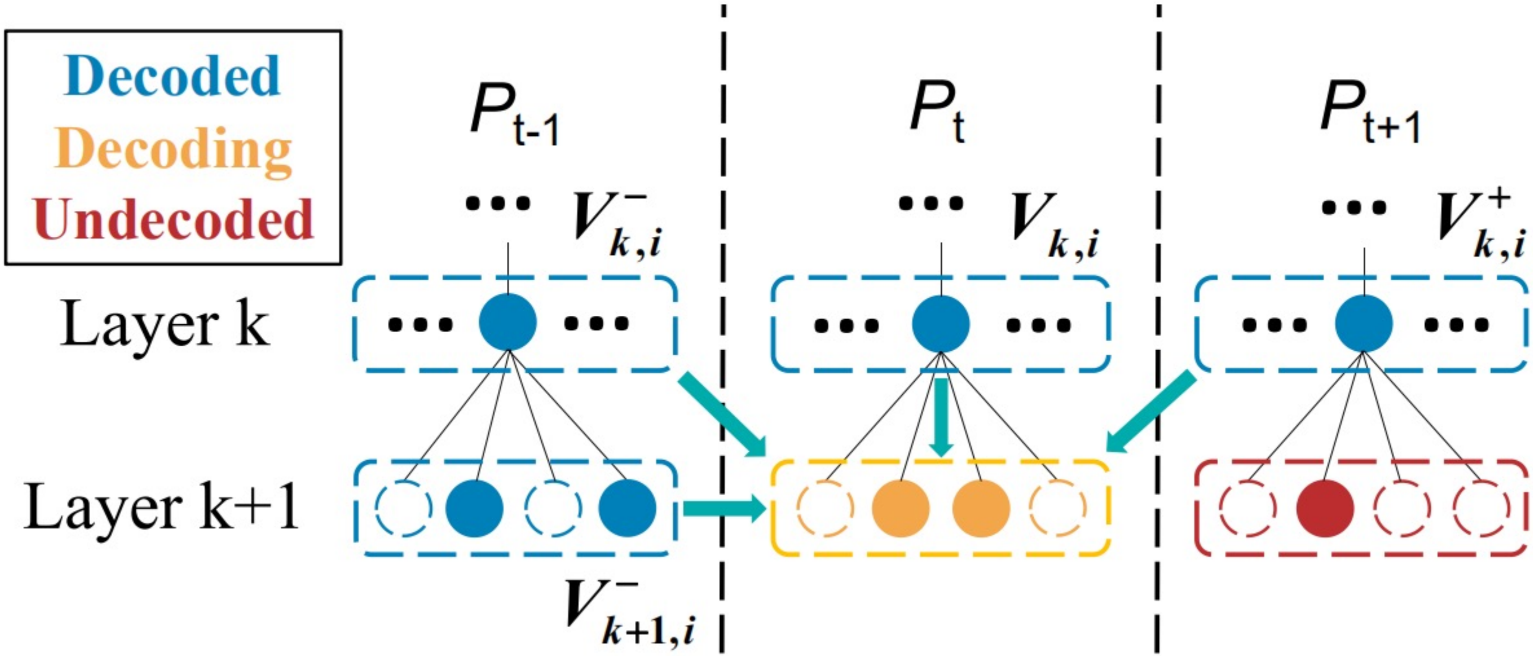
Local voxel context



(a) tree-node context **(b) proposed local voxel context**

Que, Lu, Xu, CVPR 2021

Local Voxel Context for **Dynamic** Point Clouds

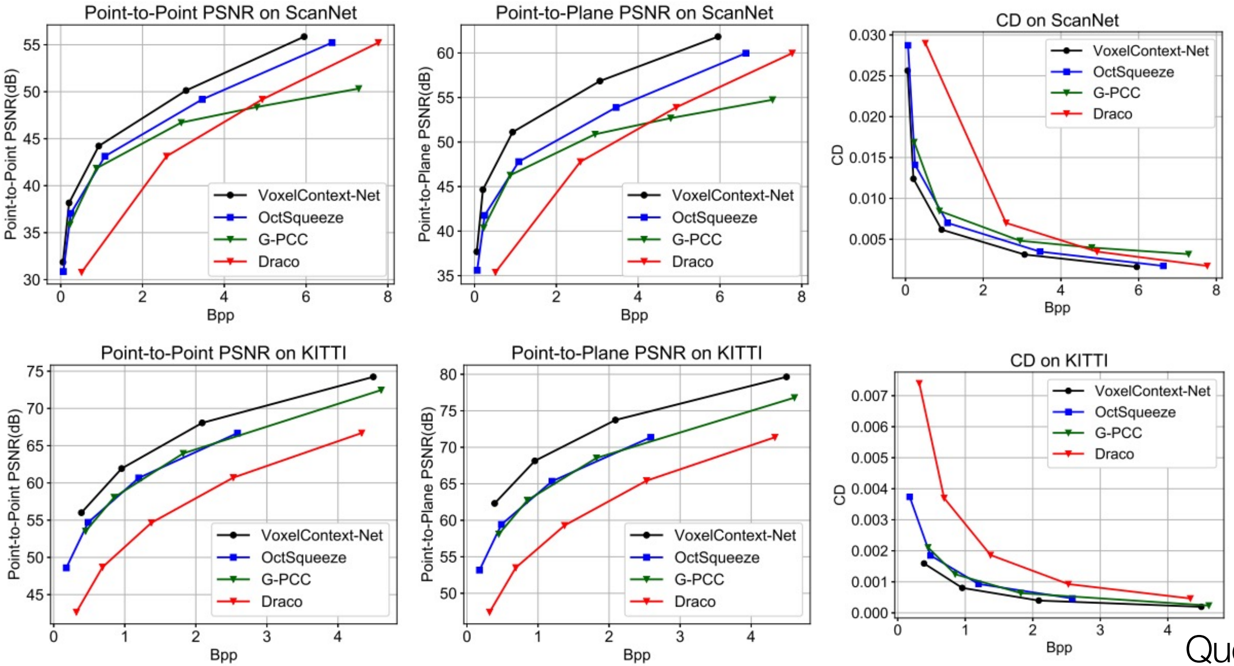


- align point clouds to the same coordinate system based on their pose information
- **concatenate the features extracted from four different voxel representations**
- all point clouds from a sequence are decoded in a depth by depth fashion

Que, Lu, Xu, CVPR 2021

Experimental Results

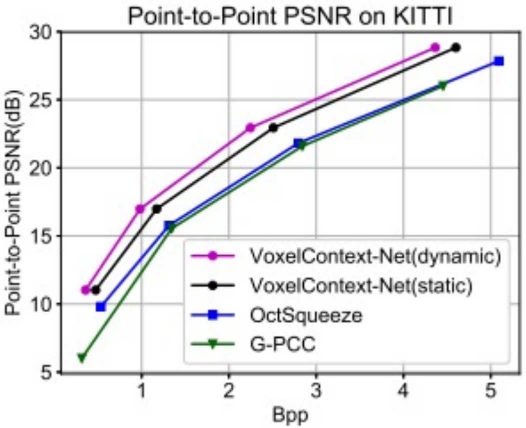
- Static Point Cloud Compression
- Evaluation metrics: Bpp & Point-to-Point(Plane) PSNR / CD

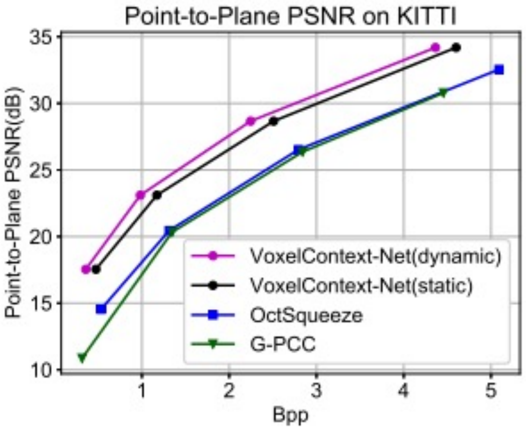


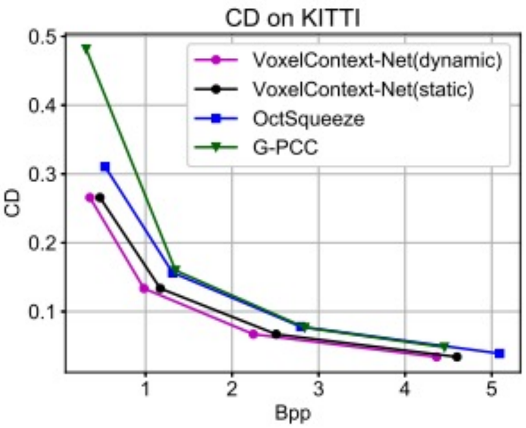
Que, Lu, Xu, CVPR 2021

Experimental Results

- Dynamic Point Cloud Compression
- Evaluation metrics: Bpp & Point-to-Point(Plane) PSNR / CD

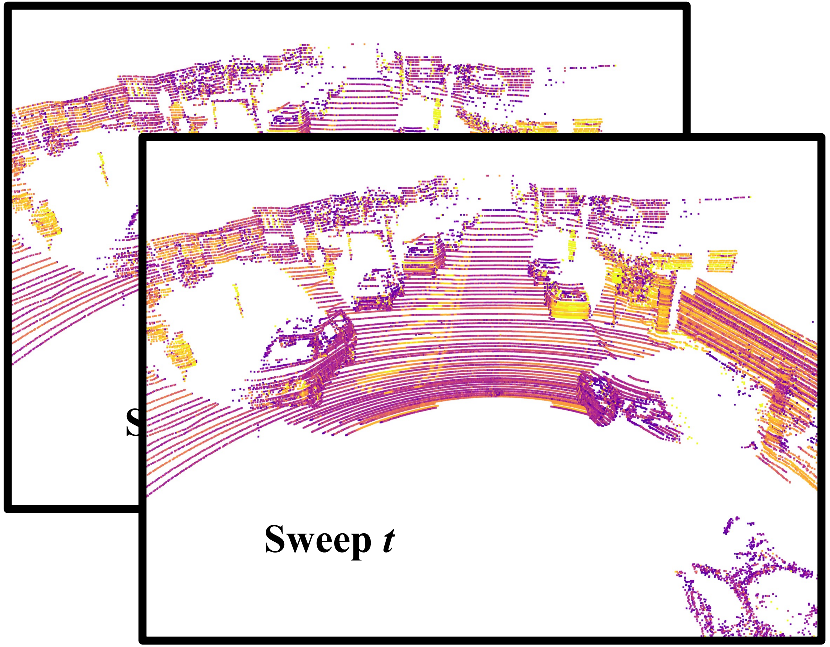






Que, Lu, Xu, CVPR 2021

How About Attributes?

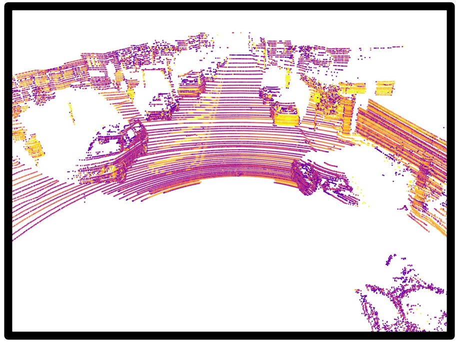


LiDAR Intensity

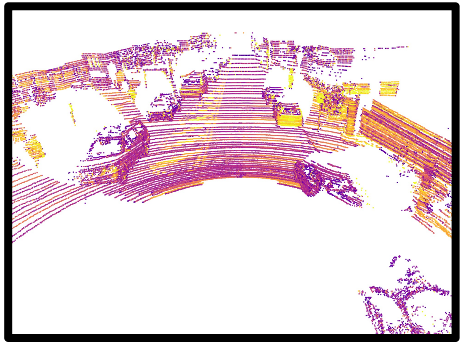
RGB Color

Motivation

Sweep *t*-1

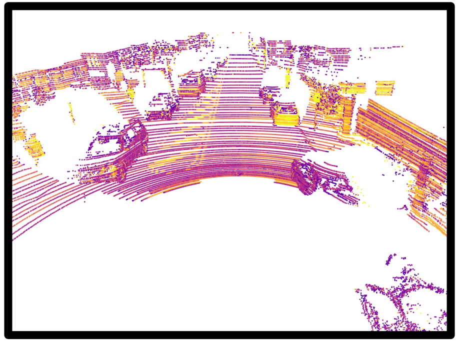


Bitrate: 104.00

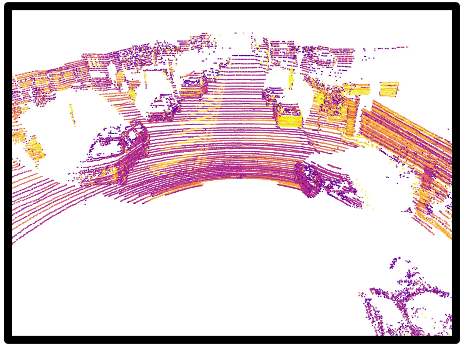


Bitrate: **13.9**

Sweep *t*

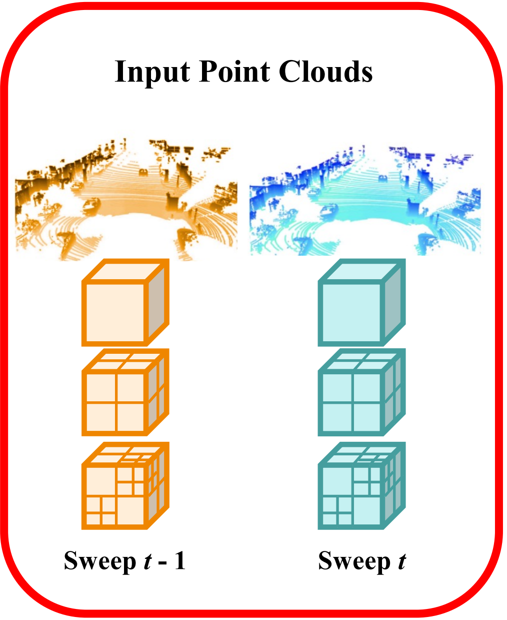


Bitrate: 104.00

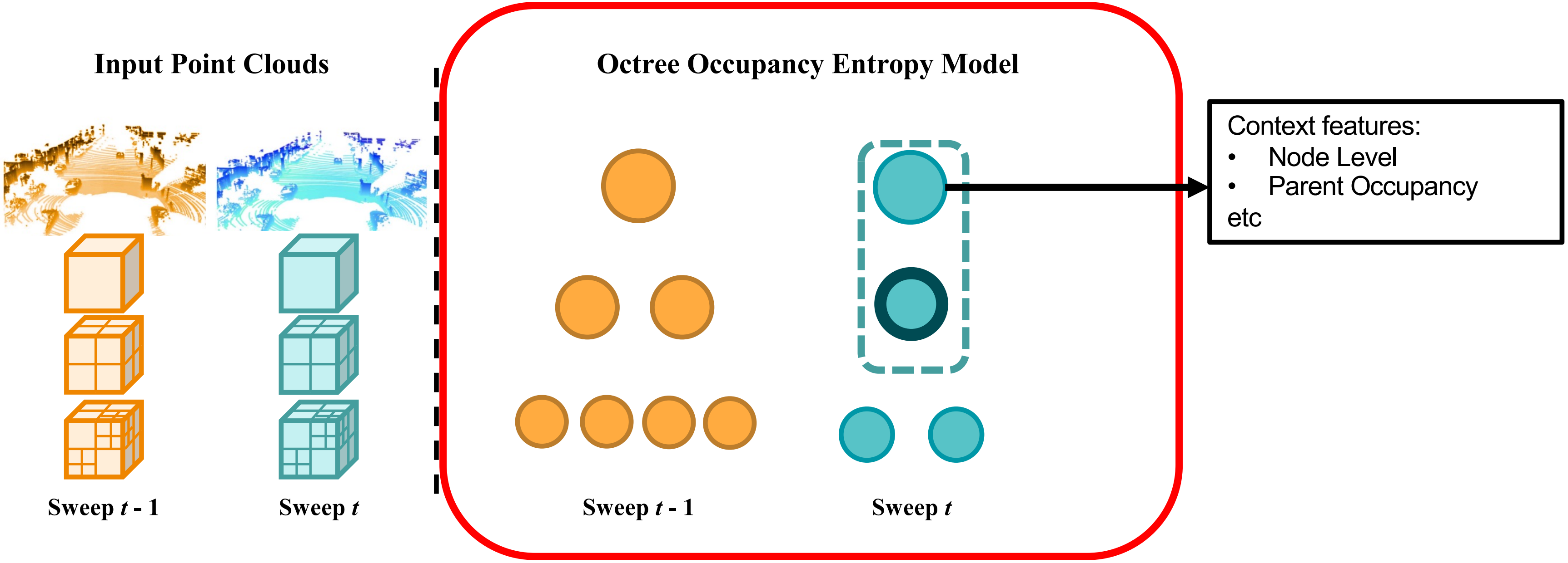


Bitrate: **13.7**

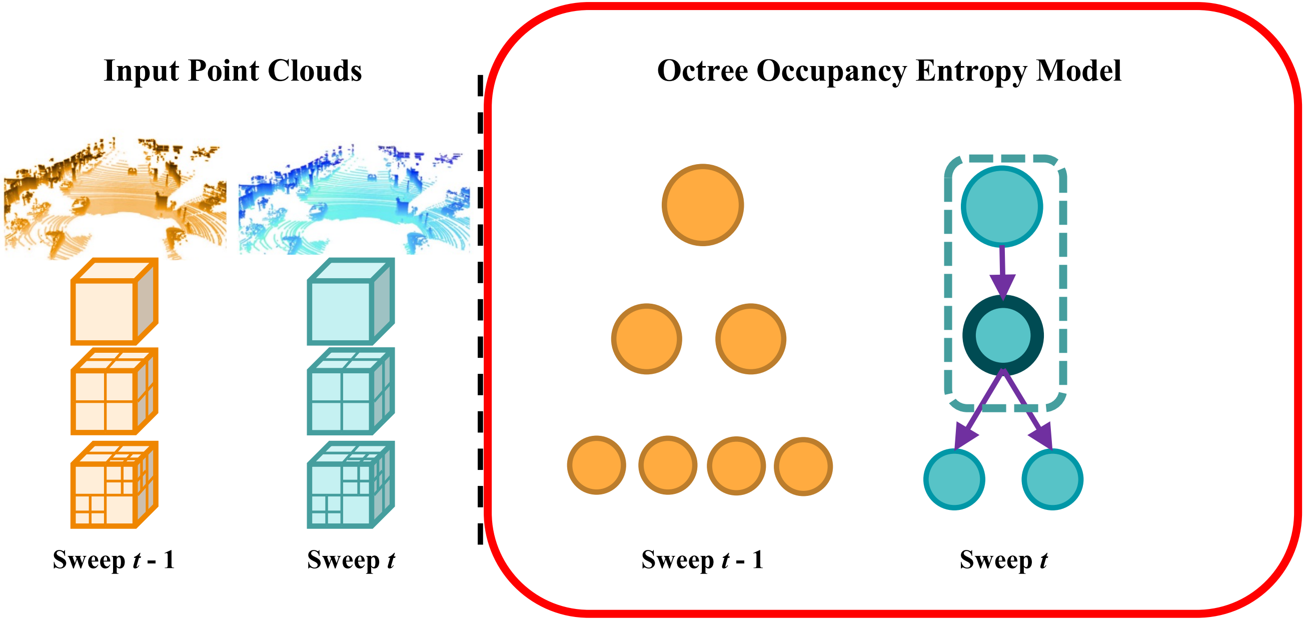
Our Method



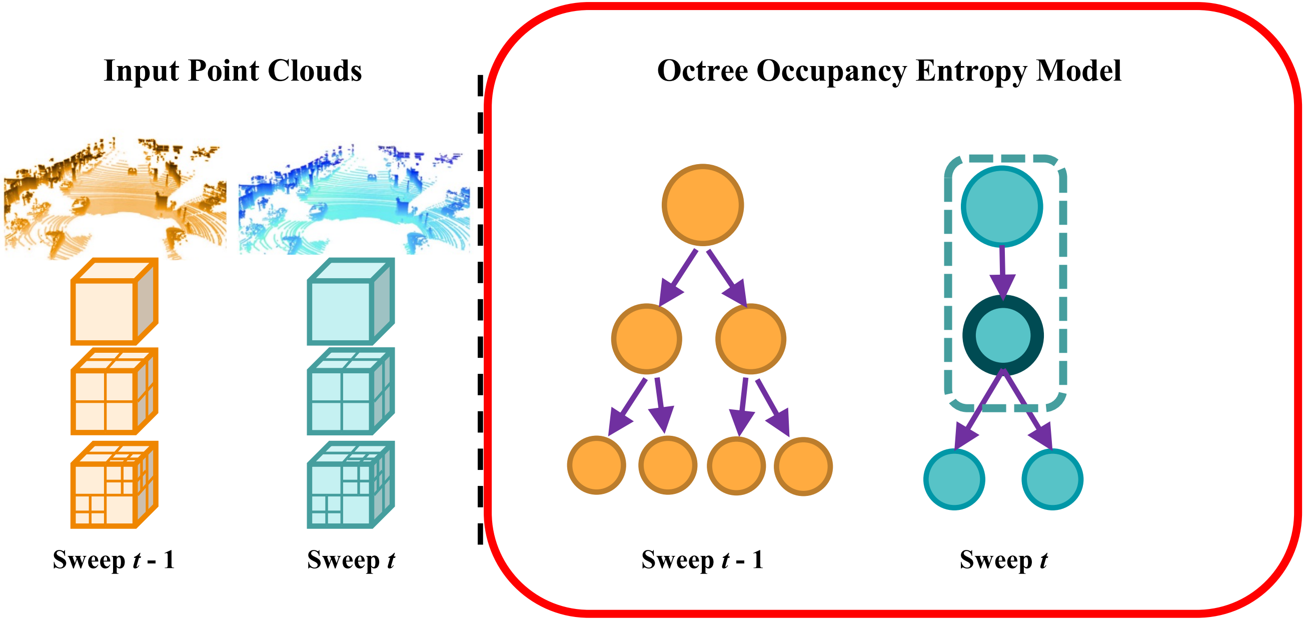
Our Method


Input Point Clouds
Sweep t - 1
Sweep t
Octree Occupancy Entropy Model
Sweep t - 1
Sweep t
Context features:
• Node Level
• Parent Occupancy
etc

Our Method

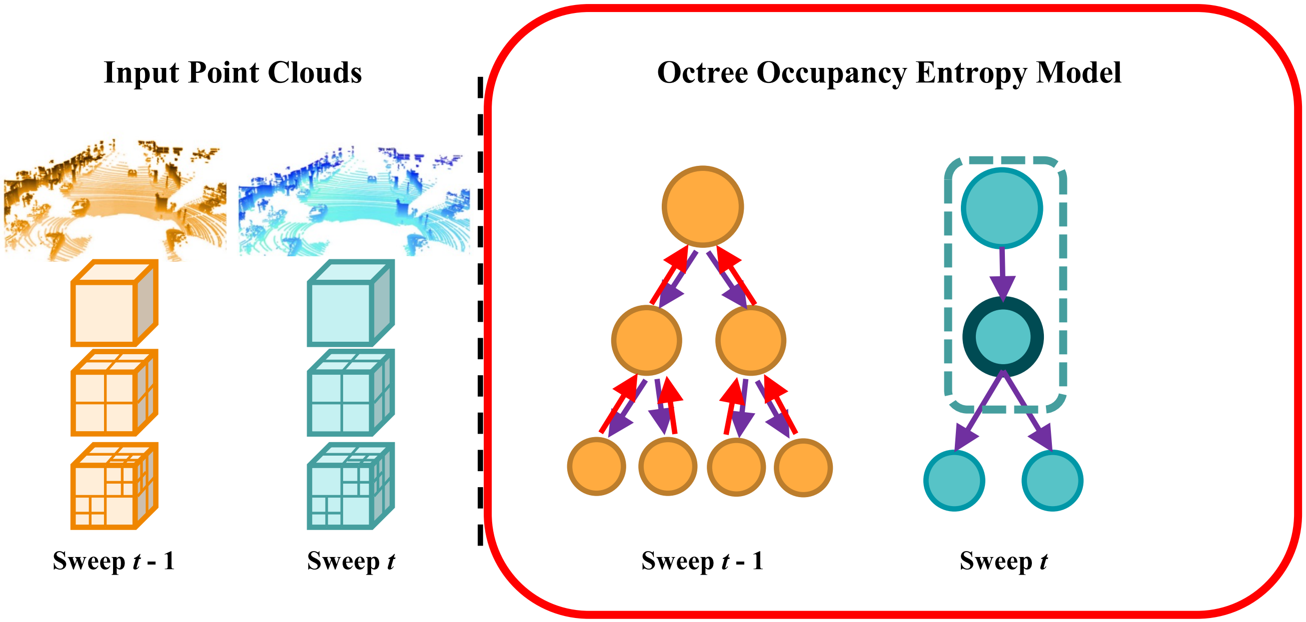

Input Point Clouds
Octree Occupancy Entropy Model
Sweep t - 1
Sweep t
Sweep t - 1
Sweep t

Our Method



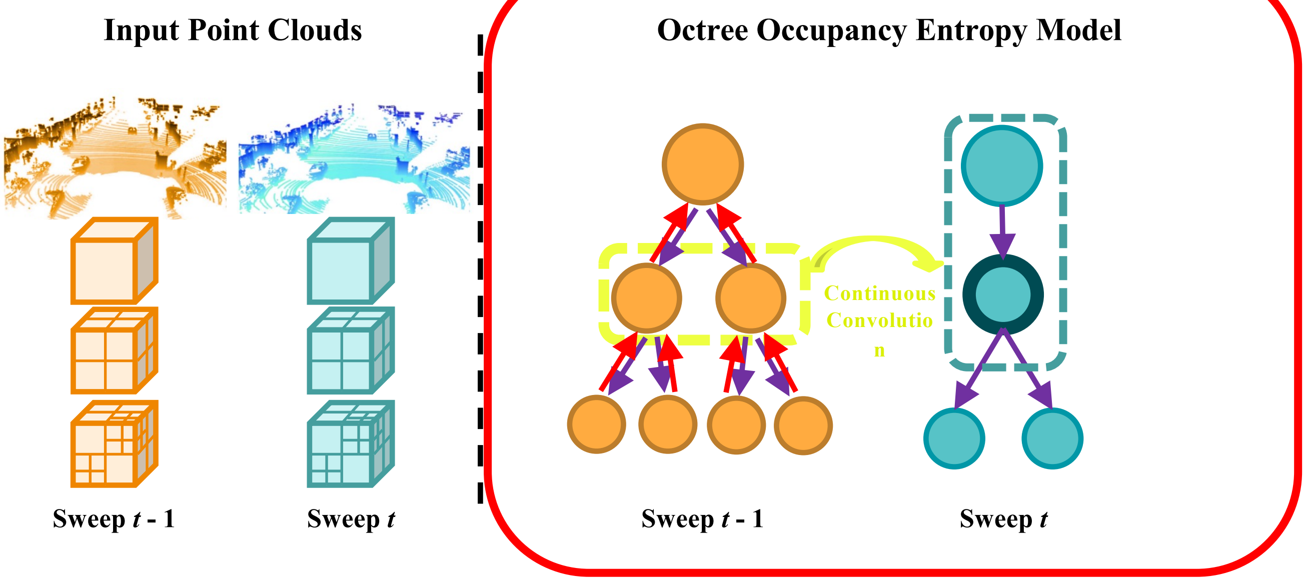
Input Point Clouds
Octree Occupancy Entropy Model
Sweep t - 1
Sweep t
Sweep t - 1
Sweep t

Our Method

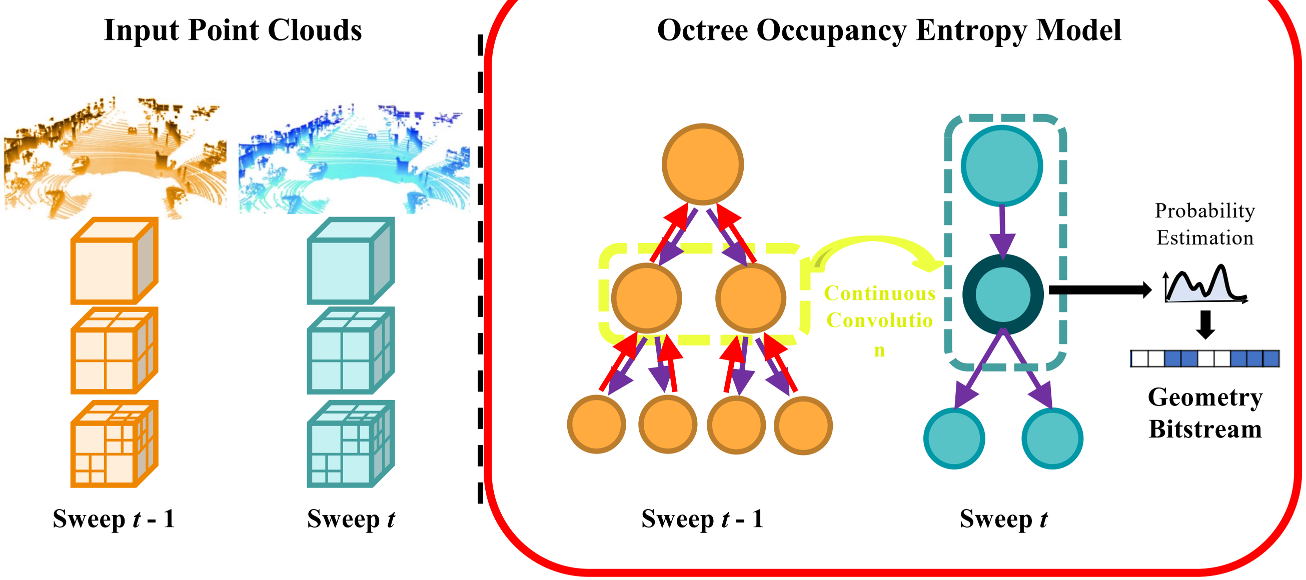


Input Point Clouds
Octree Occupancy Entropy Model
Sweep t - 1
Sweep t
Sweep t - 1
Sweep t

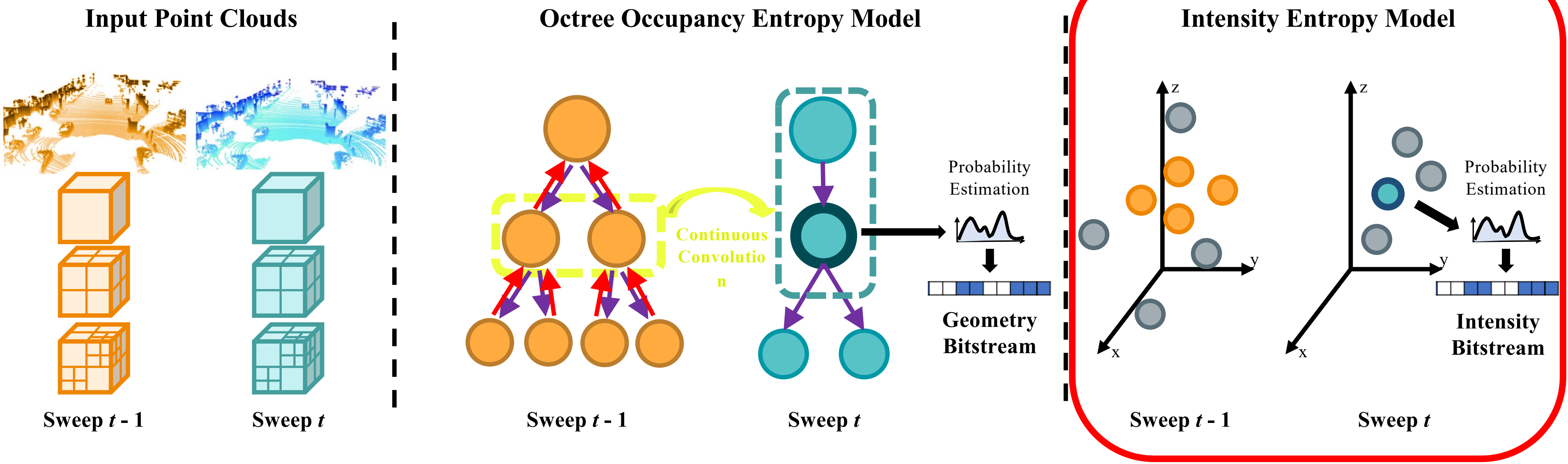
Our Method


Input Point Clouds
Octree Occupancy Entropy Model
Continuous
Convolutio
n
Sweep t - 1
Sweep t
Sweep t - 1
Sweep t

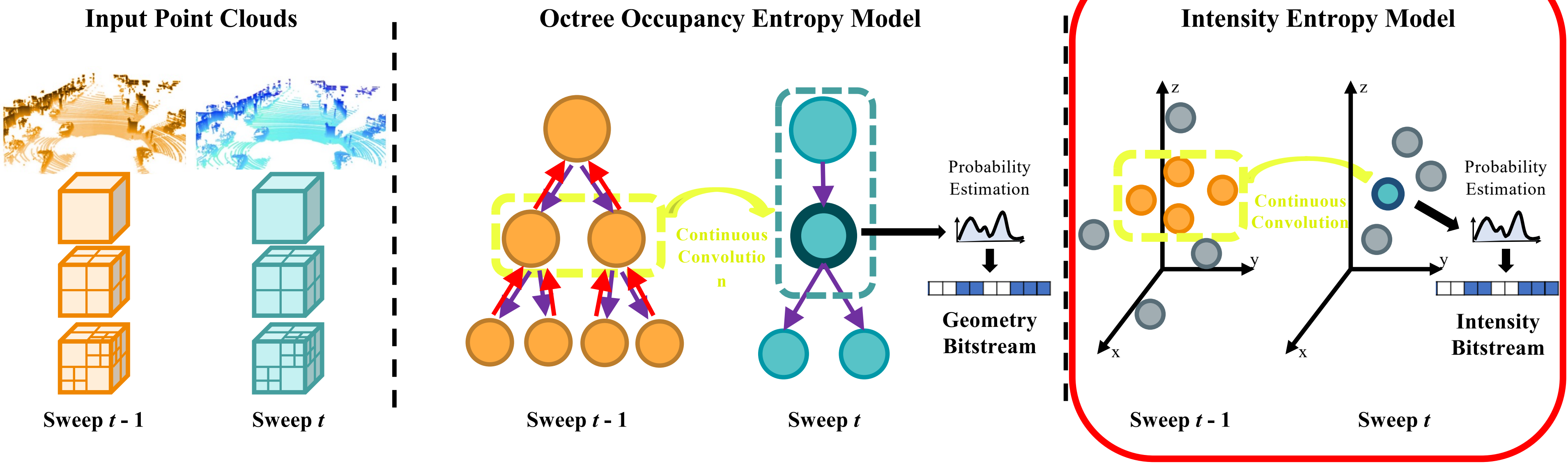
Our Method


Input Point Clouds
Octree Occupancy Entropy Model
Continuous
Convolutio
n
Probability
Estimation
Geometry
Bitstream
Sweep t - 1
Sweep t
Sweep t - 1
Sweep t

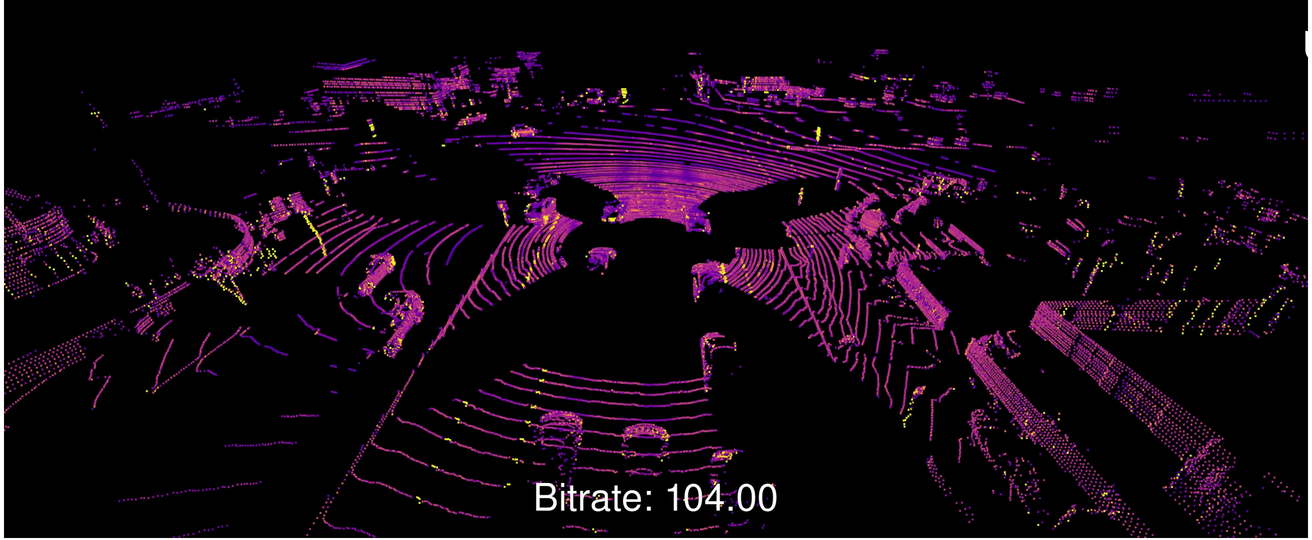
Our Method

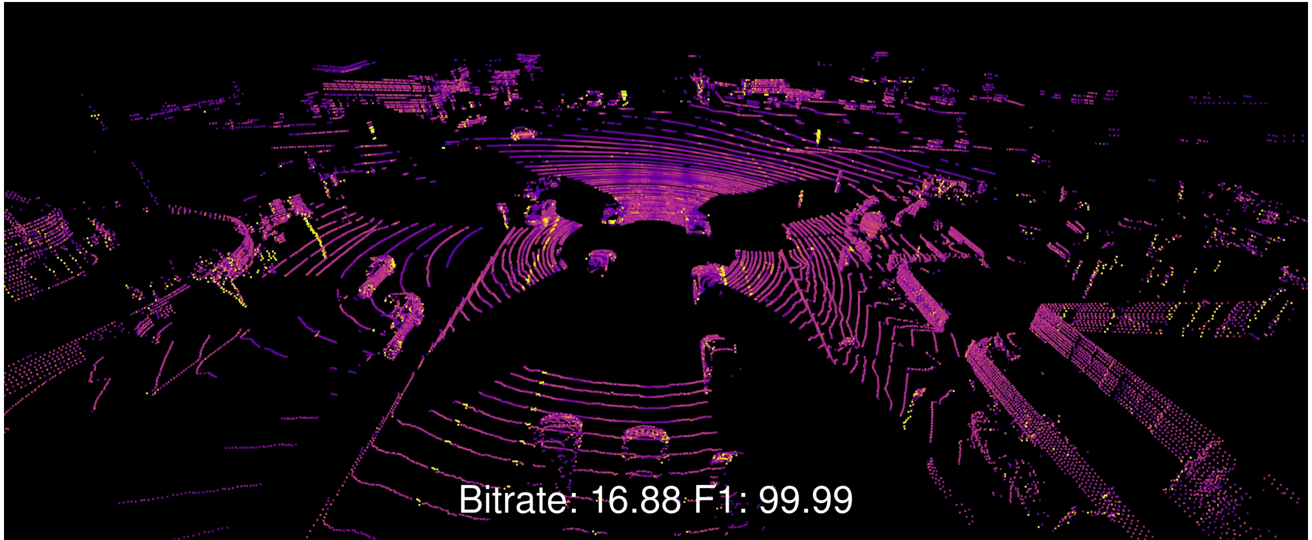

Input Point Clouds
Sweep t - 1
Sweep t
Octree Occupancy Entropy Model
Continuous Convolutio n
Probability Estimation
Geometry Bitstream
Sweep t - 1
Sweep t
Intensity Entropy Model
z
y
x
z
y
x
Probability Estimation
Intensity Bitstream
Sweep t - 1
Sweep t

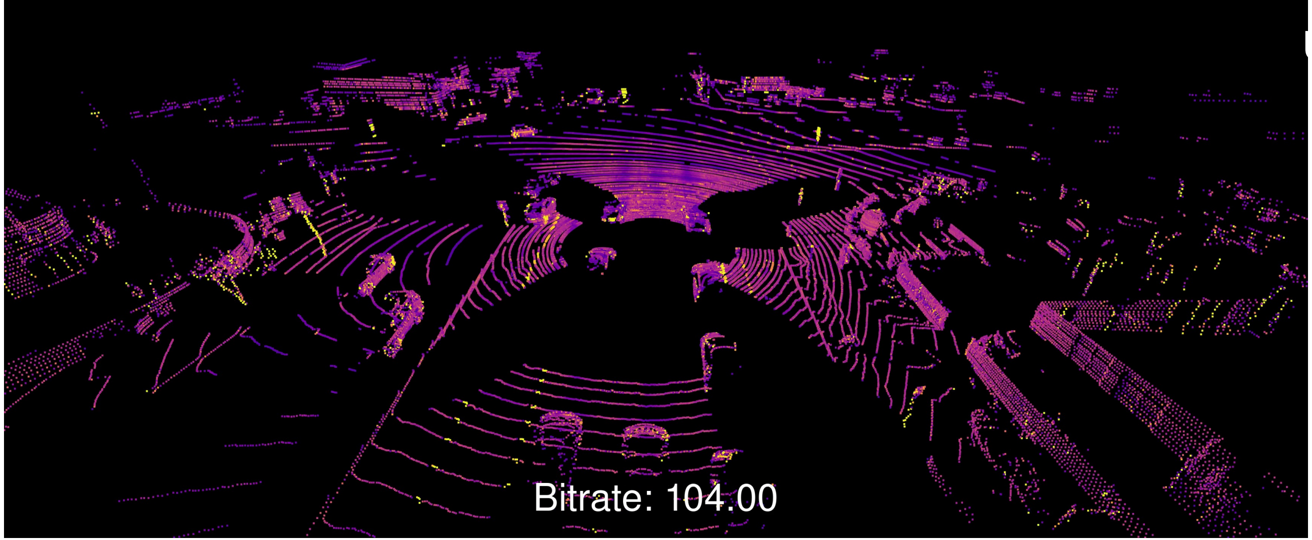
Our Method

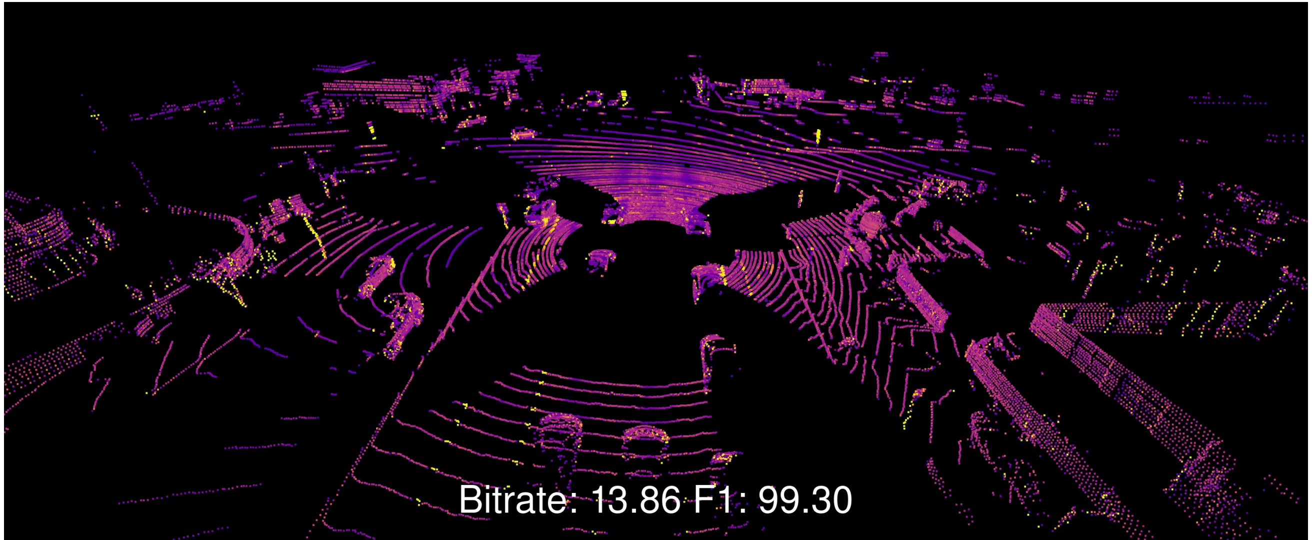


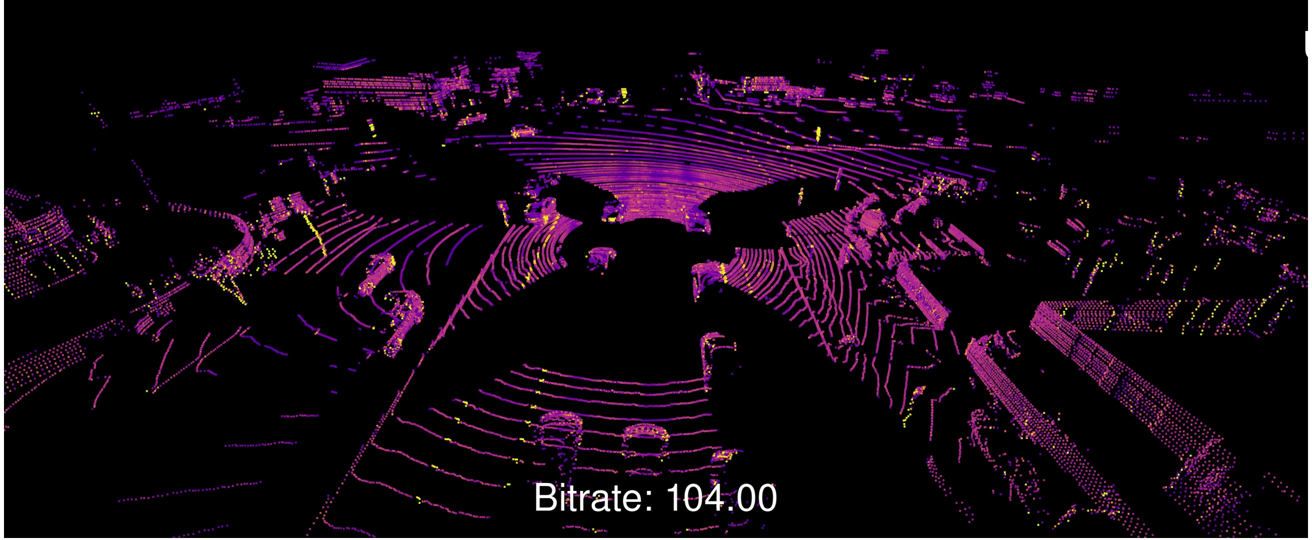
Improves LiDAR stream compression 15 – 35%

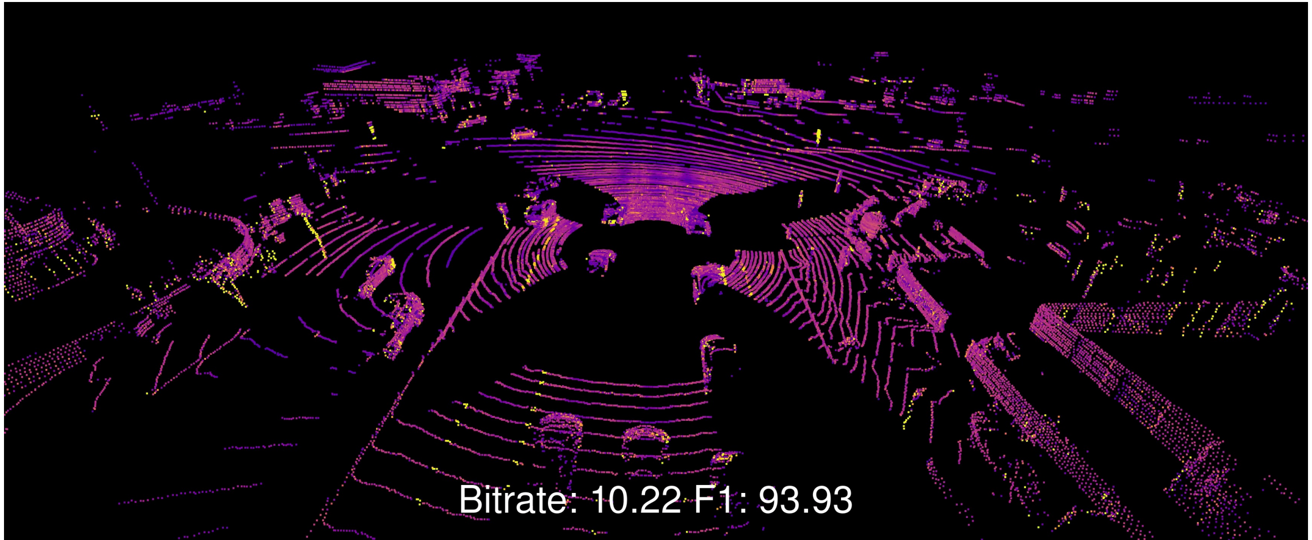

Bitrate: 104.00

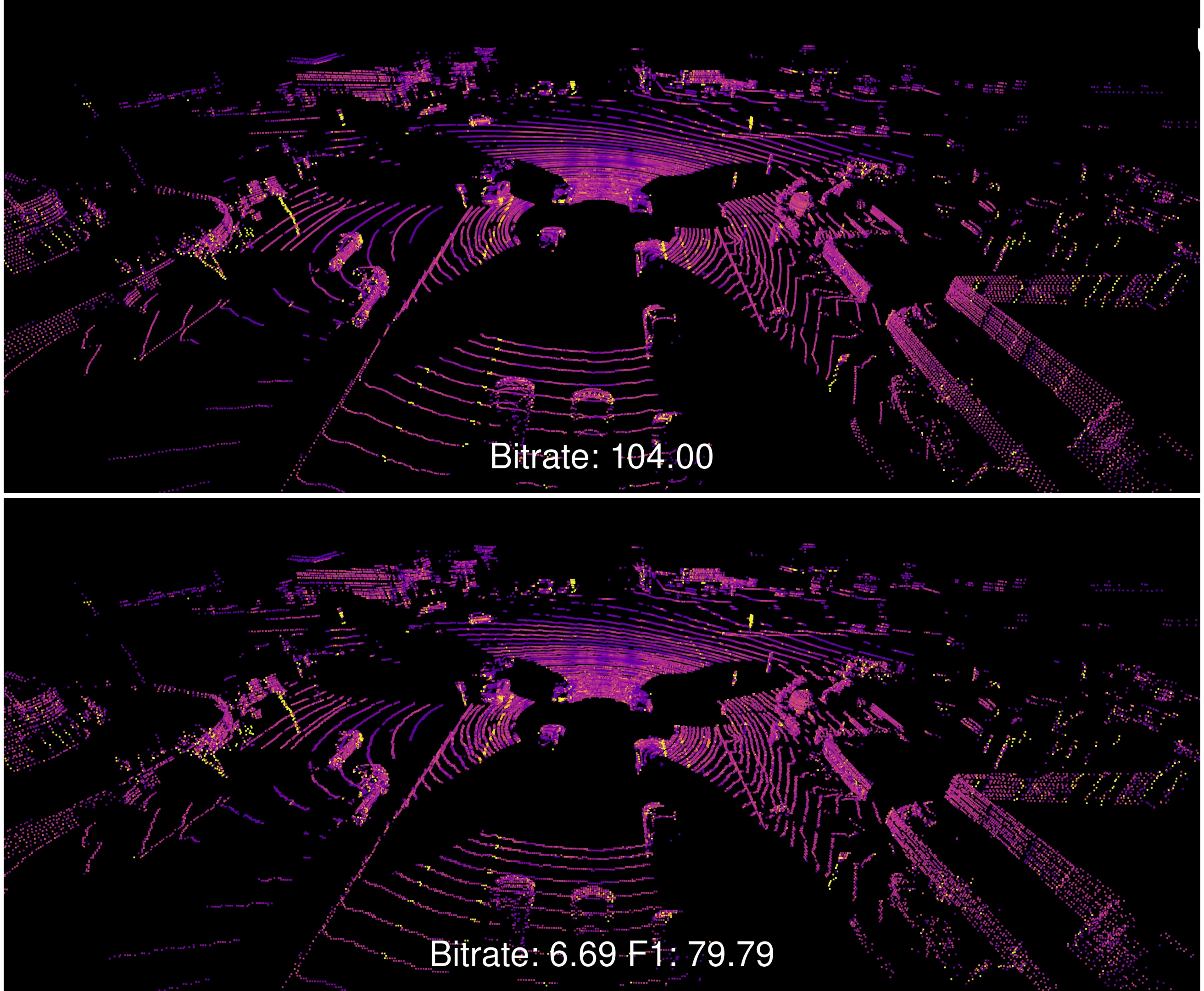

Bitrate: 16.88 F1: 99.99

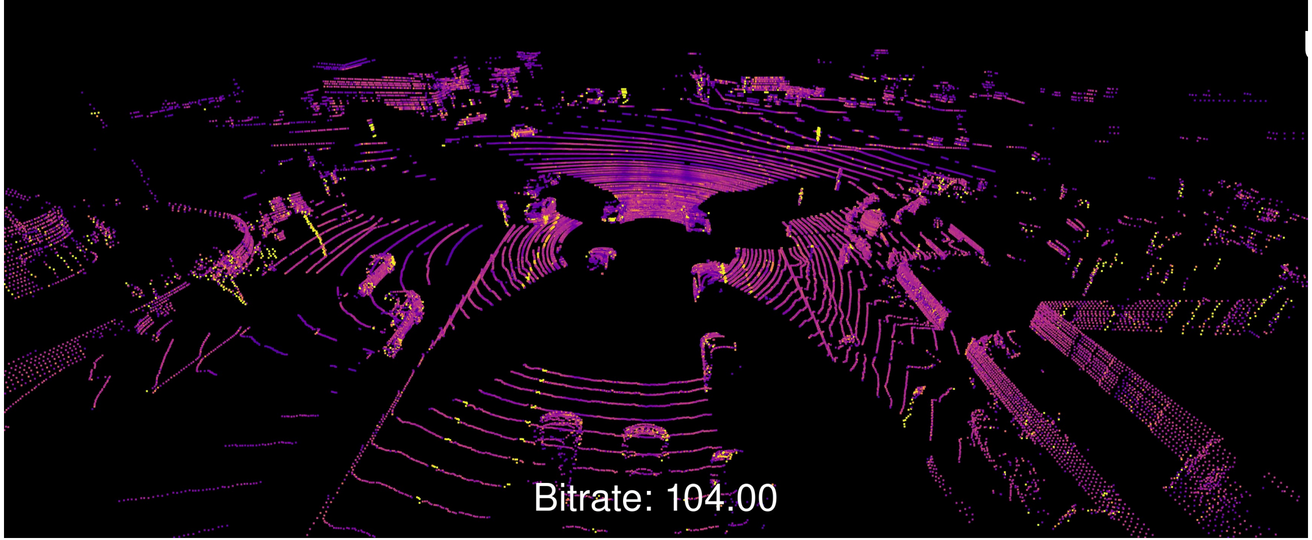

Bitrate: 104.00

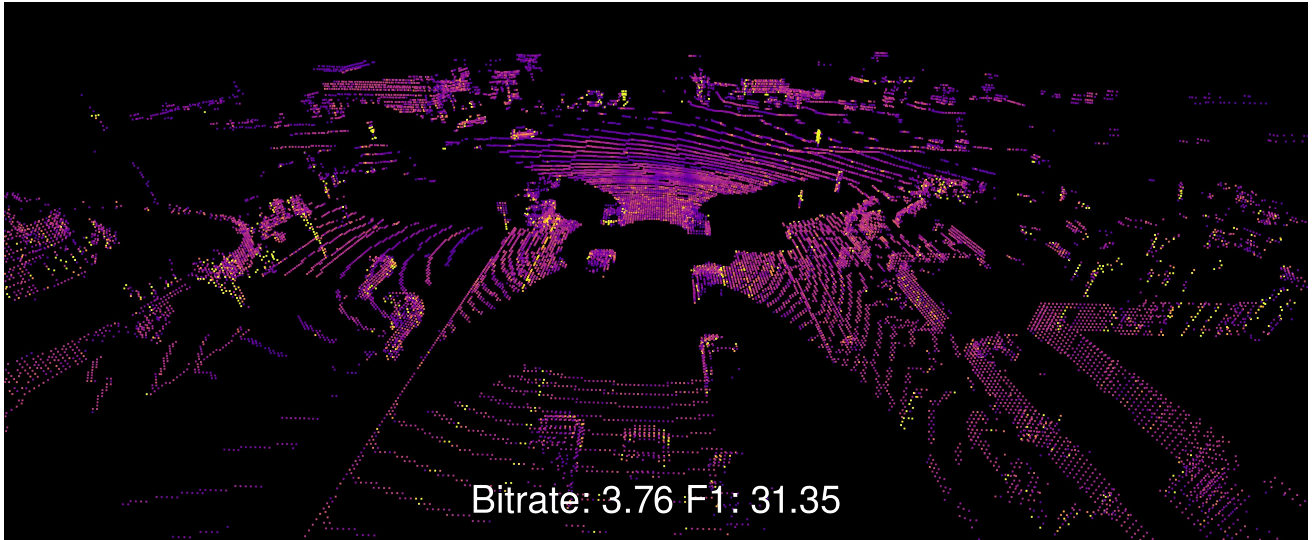

Bitrate: 13.86 F1: 99.30

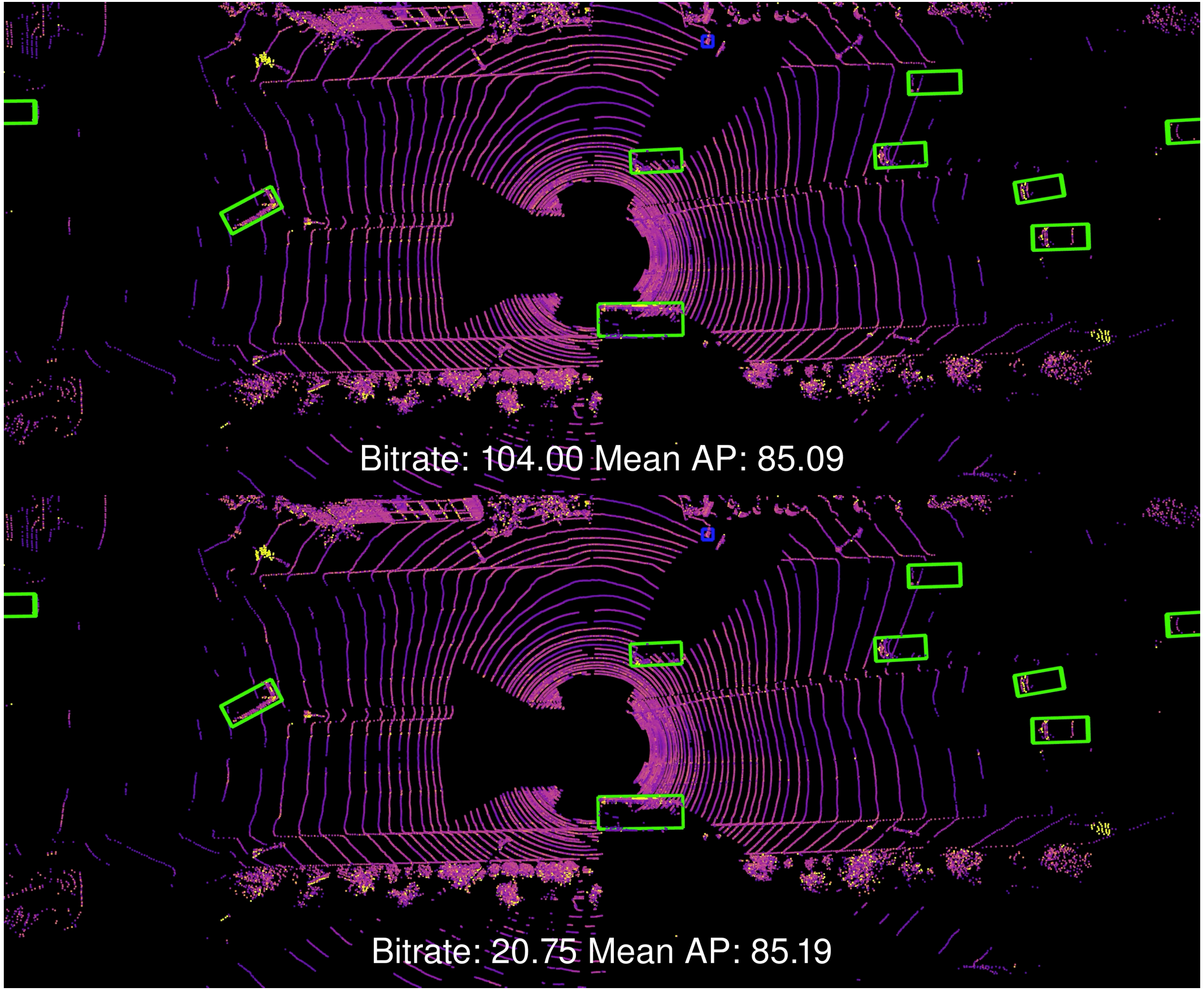

Bitrate: 104.00

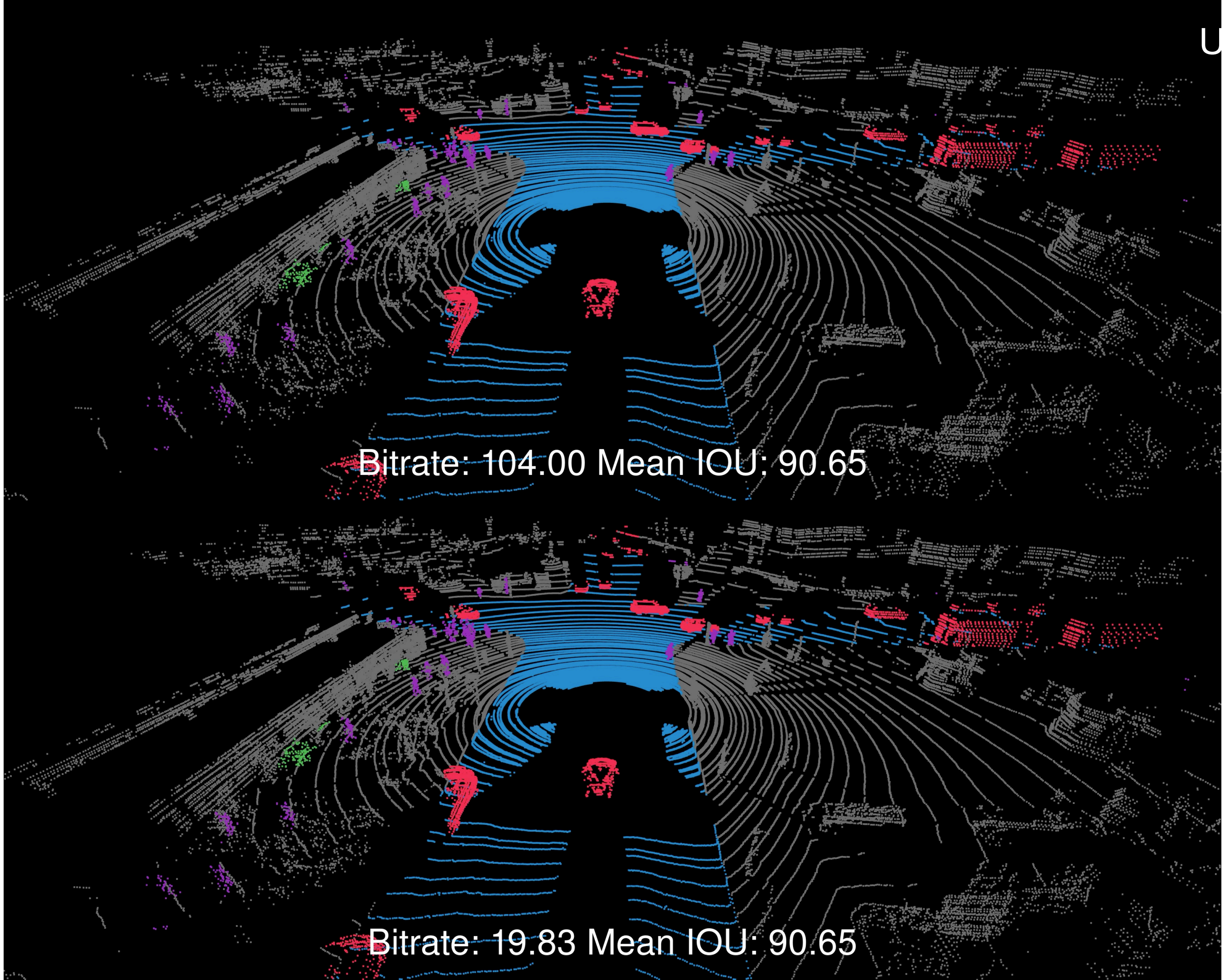

Bitrate: 10.22 F1: 93.93


Bitrate: 104.00
Bitrate: 6.69 F1: 79.79


Bitrate: 104.00


Bitrate: 3.76 F1: 31.35


Bitrate: 104.00 Mean AP: 85.09
Bitrate: 20.75 Mean AP: 85.19


U
Bitrate: 104.00 Mean IOU: 90.65
Bitrate: 19.83 Mean IOU: 90.65

Should we always optimize distortion?

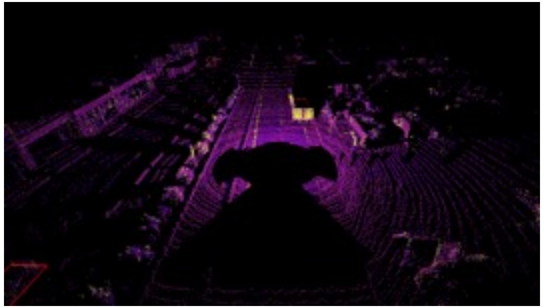
$$\ell(\boldsymbol{x}, \hat{\boldsymbol{x}}) + \beta R(\hat{\boldsymbol{y}})$$

distortion

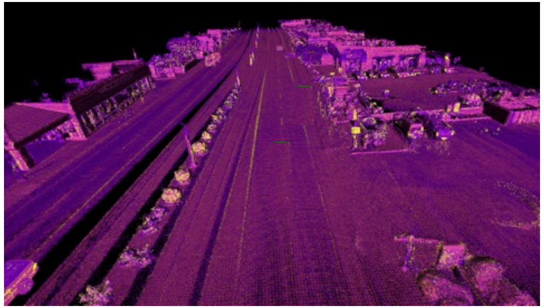
rate

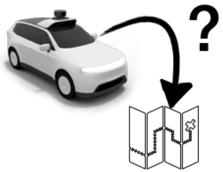
Localization

Online Sensor

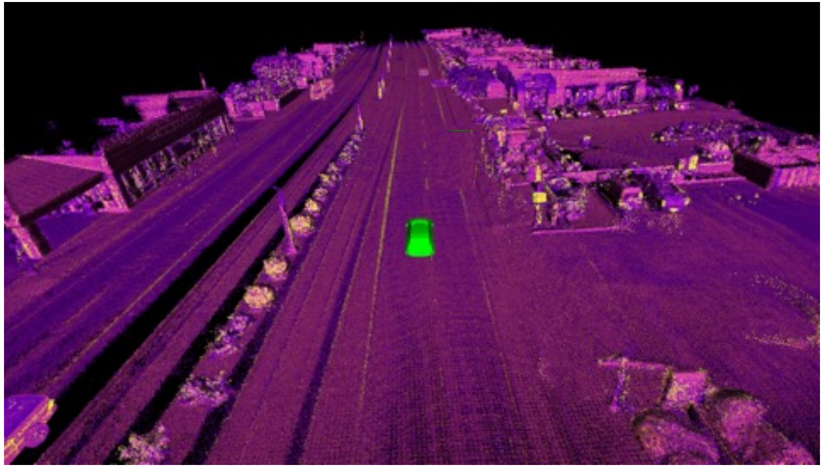


Map





Robot Location wrt the Map

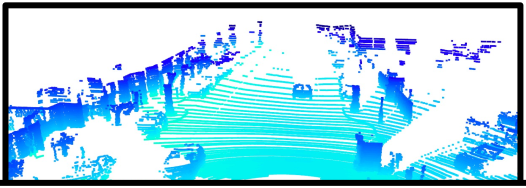


Desiderata

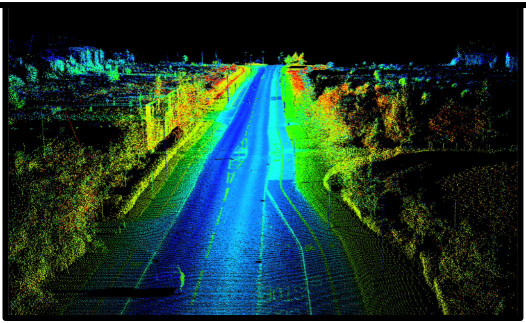
- Robust: localization should work everywhere with 99.99%
- Accurate: self-driving requires cm-level accuracy
- Efficient: localization needs to reflect real-time
- Low storage: maps for localization should be small

Prior Work

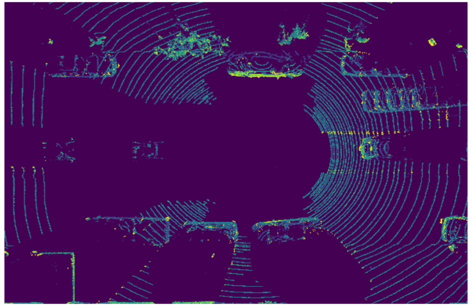
3D Geometry

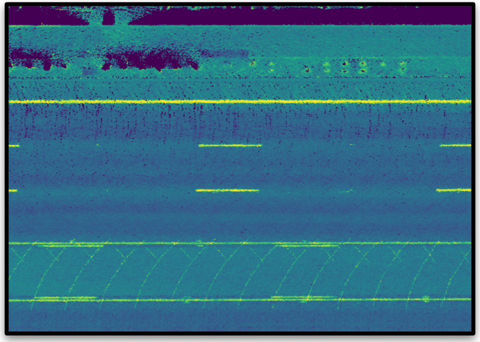


- High cost for storage & computation
- Geometry degenerated scene



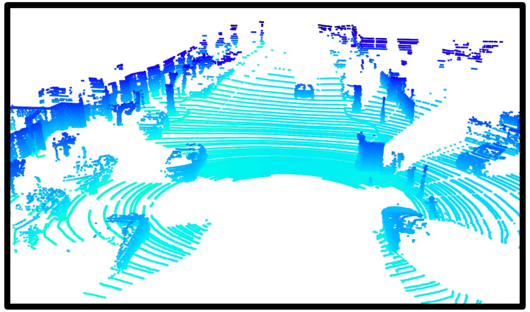
Rasterized Birds-Eye View Reflectance





Birds-Eye-View Rasterization

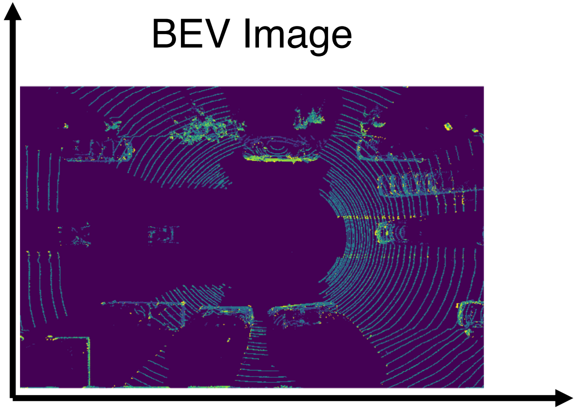
3D Geometry



Ortho-projection from
Top-down

Left

BEV Image

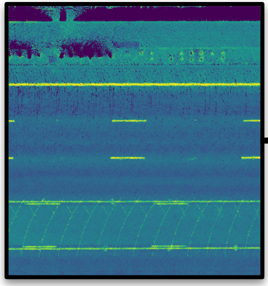


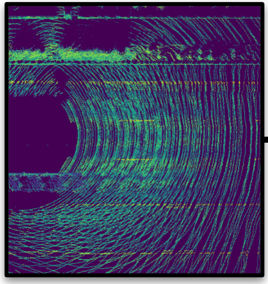
Front

Learning to Match

Learn representation to match

Map



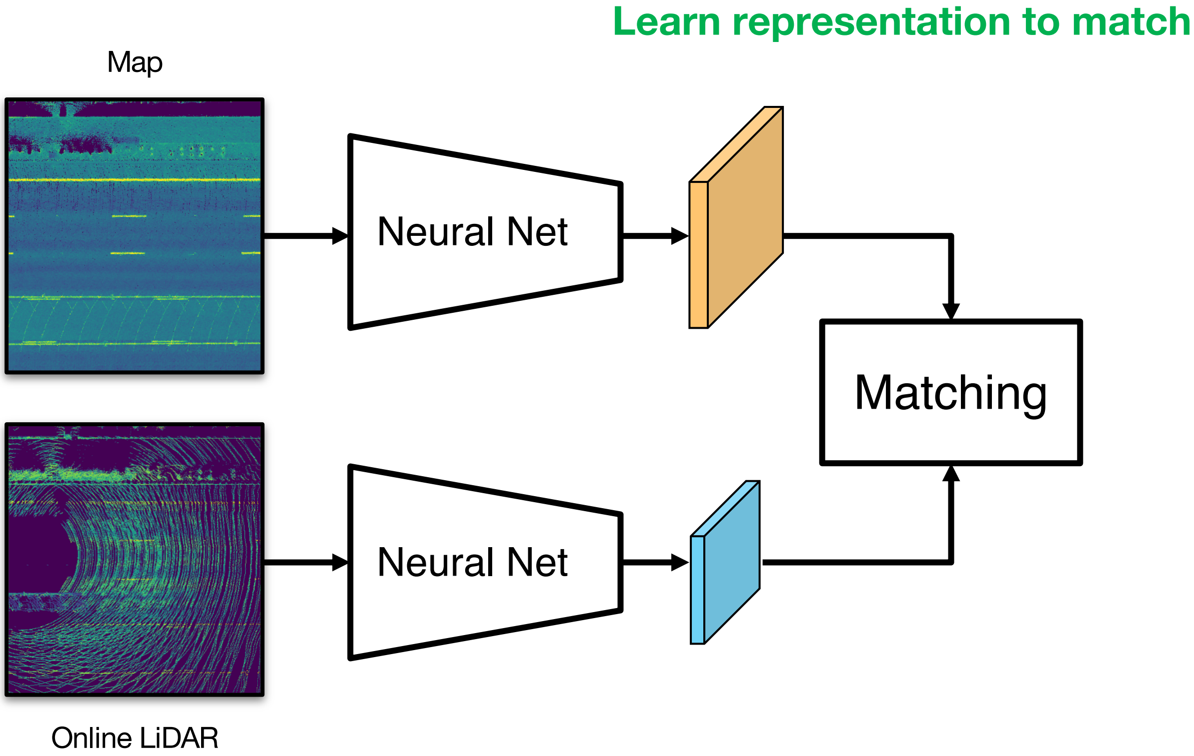


Online LiDAR

- Not robust to weather changes, intensity miscalibration
- Difficult to handle dynamic objects
- Subpar accuracy

Barsan*, **Wang***, Porovsky, Urtasun, CoRL 2018

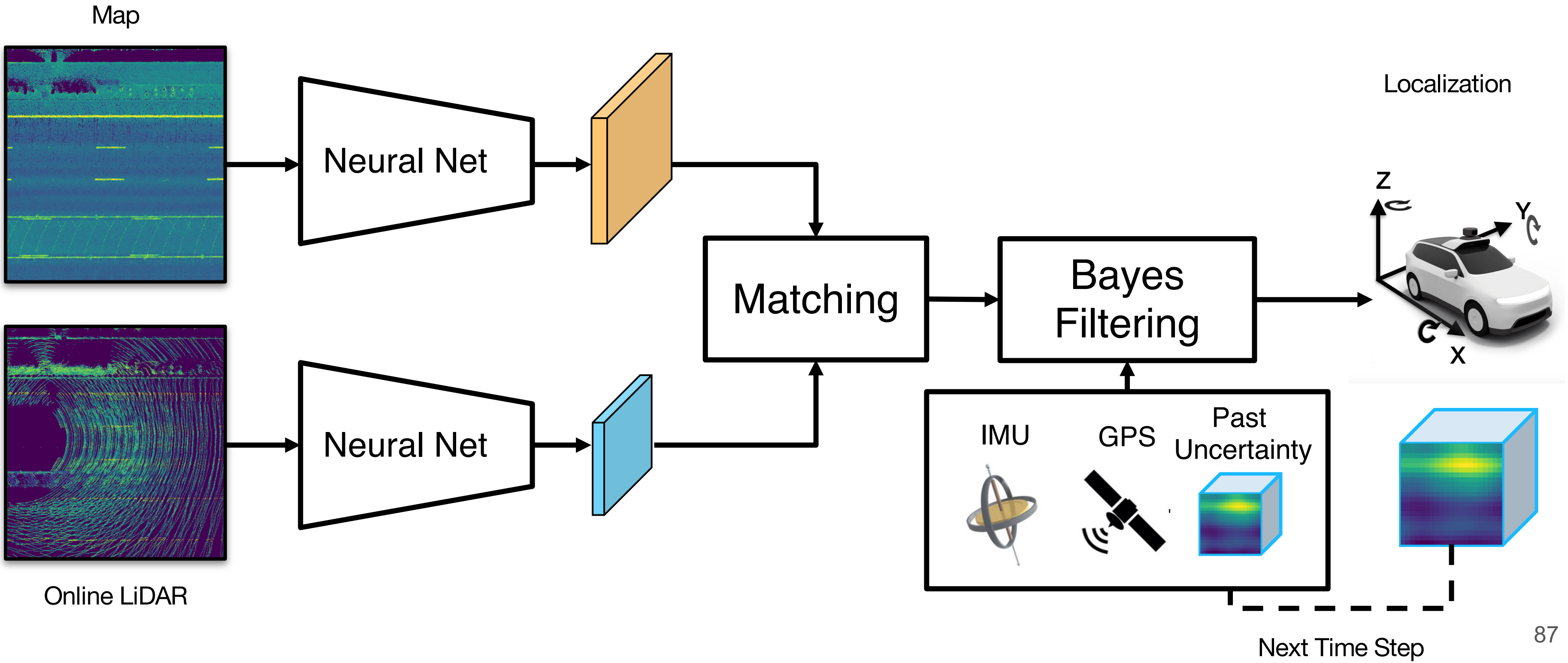
Learning to Match



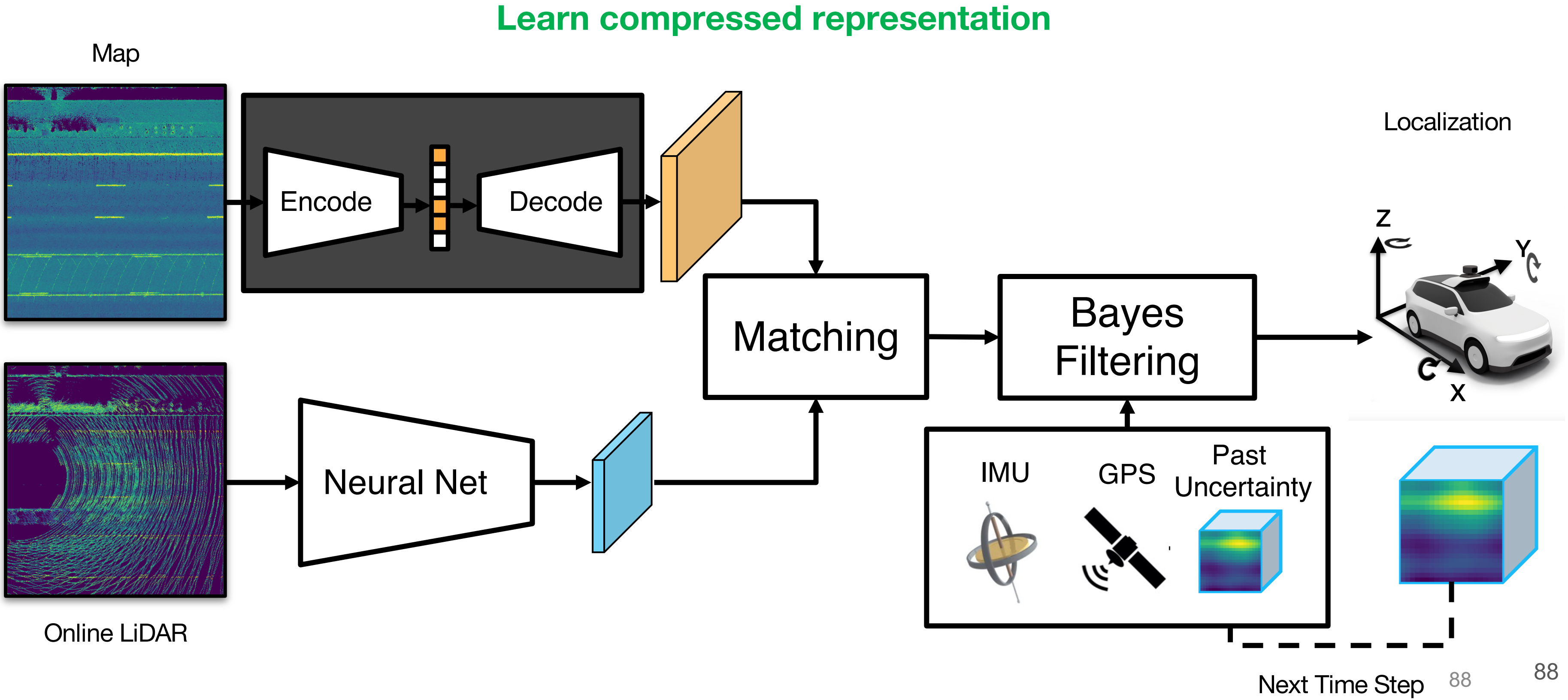
Barsan*, **Wang***, Porovsky, Urtasun, CoRL 2018

Learning to Match

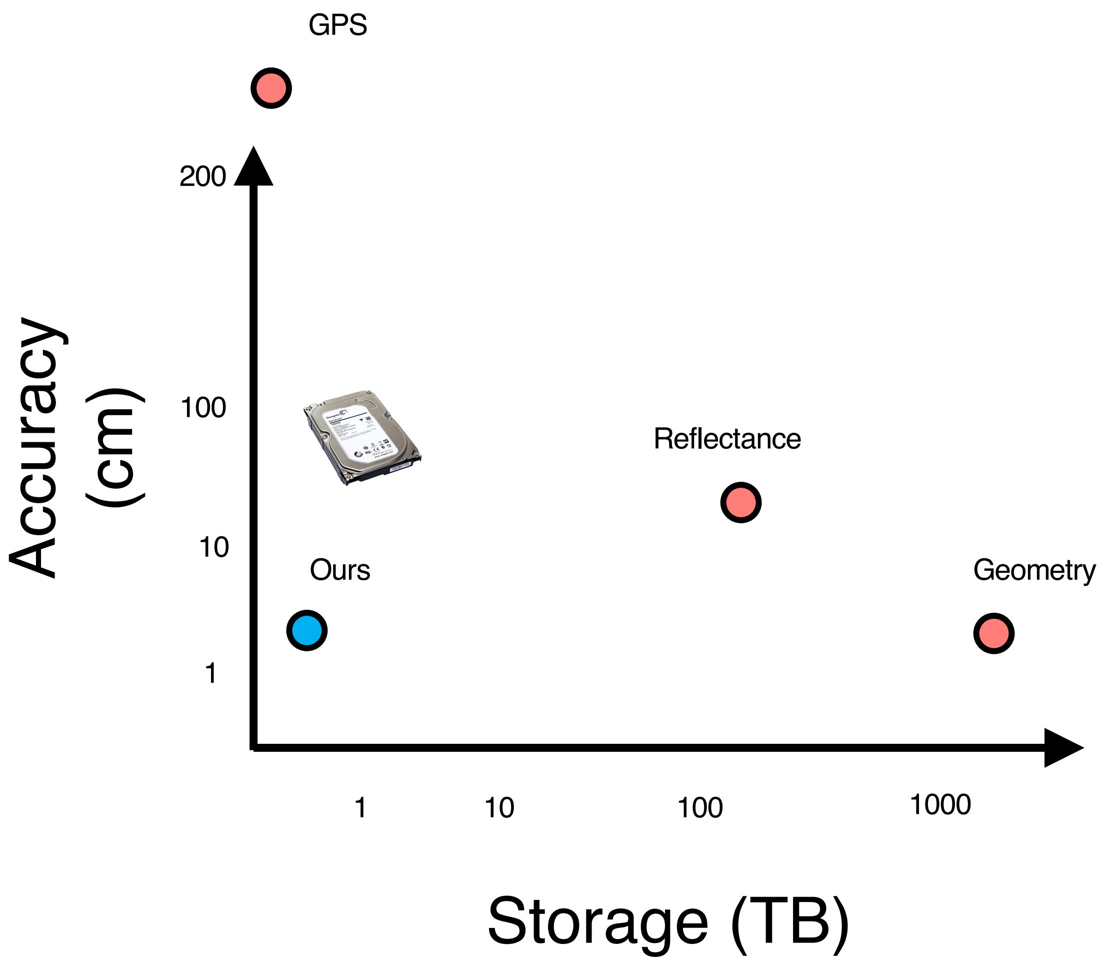
Learn representation to match



Learning to Compress and Match

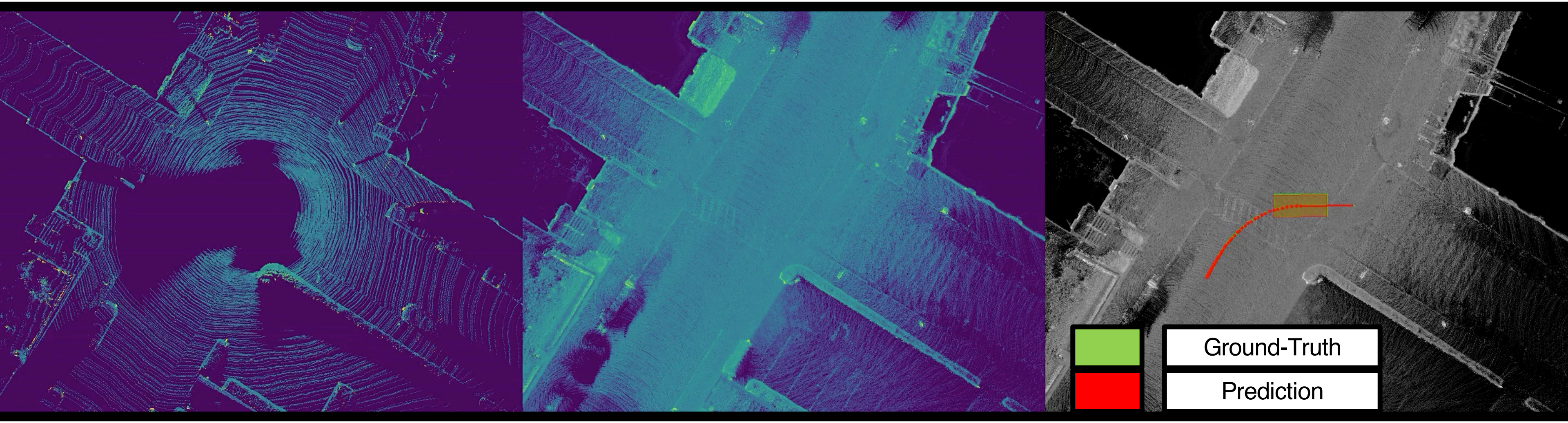


88



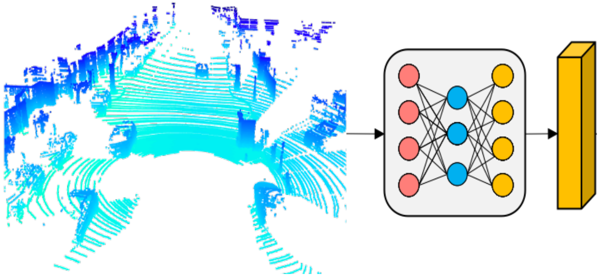
Wei*, Barsan*, **Wang***, Martinez, Urtasun, CVPR 2019

Online Lidar Map Results

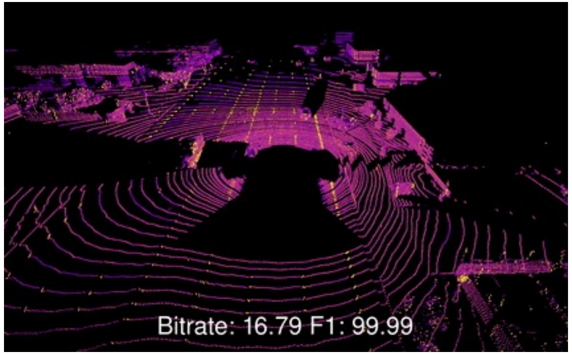


Wei*, Barsan*, **Wang***, Martinez, Urtasun, CVPR 2019

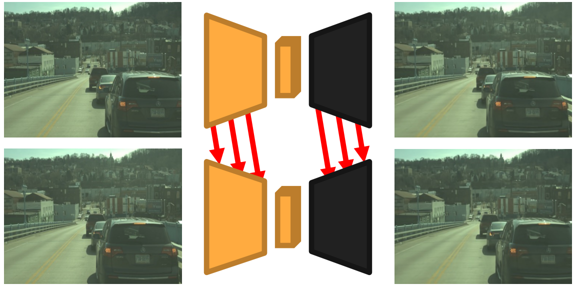
Summary



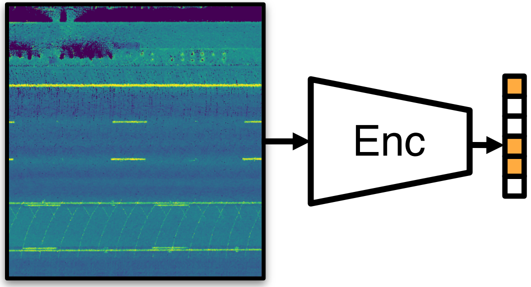
Learning to Compress
LiDAR



Learning to Compress
LiDAR Stream



Learning to Compress
Stereo Cameras



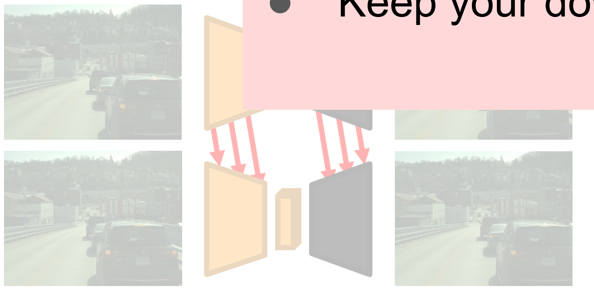
Learning to Compress
Maps for Localization

Summary

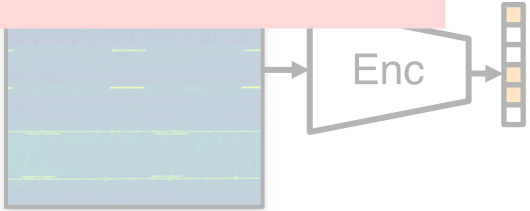
Learning

Take-Home Message

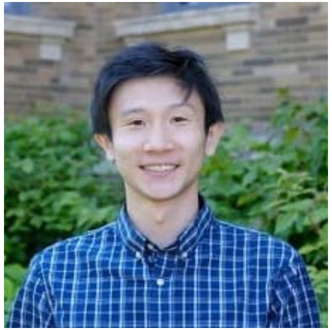
- Memory-efficient structures
- Train deep entropy models
- Keep your downstream task in mind



Learning to Compress
Stereo Cameras



Learning to Compress
Maps for Localization


Jerry Liu


Lila Huang


Sourav Biswas

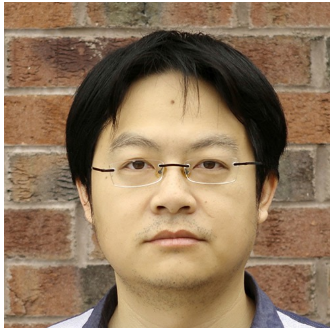

Xinkai Wei


Kelvin Wong


Raquel Urtasun


Andrei Barsan


Wei-Chiu Ma


Rui Hu


Julieta Martinez

References

[1] **MuSCLE: Multi Sweep Compression of LiDAR using Deep Entropy Models**, Sourav Biswas, Jerry Liu, Kelvin Wong*, Shenlong Wang, Raquel Urtasun, Neural Information Processing Systems (NeurIPS), 2020

[2] **Conditional Entropy Coding for Efficient Video Compression** Jerry Liu, Shenlong Wang, Wei-Chiu Ma, Meet Shah, Rui Hu, Pranaab Dhawan, Raquel Urtasun European Conference on Computer Vision (ECCV), 2020

[3] **OctSqueeze: Octree-Structured Entropy Model for LiDAR Compression** Lila Huang, Shenlong Wang, Kelvin Wong, Jerry Liu, and Raquel Urtasun International Conference on Computer Vision and Pattern Recognition (CVPR), 2020 (**Oral**)

[4] **DSIC: Deep Stereo Image Compression** Jerry Liu, Shenlong Wang, and Raquel Urtasun International Conference on Computer Vision (ICCV), 2019 (**Oral**)

[5] **Learning to Localize through Compressed Binary Maps** Xinkai Wei*, Ioan Andrei Barsan*, Shenlong Wang*, Julieta Martinez, and Raquel Urtasun International Conference on Computer Vision and Pattern Recognition (CVPR), 2019

[6] **VoxelContext-Net: An Octree based Framework for Point Cloud Compression**, Zizheng Que, Guo Lu, Dong Xu, Urtasun International Conference on Computer Vision and Pattern Recognition (CVPR), 2021