H-V2X: A Large Scale Highway Dataset for BEV Perception

Chang Liu, Mingxu Zhu, Cong Ma, Kuifeng Su

AD-Lab Tencent

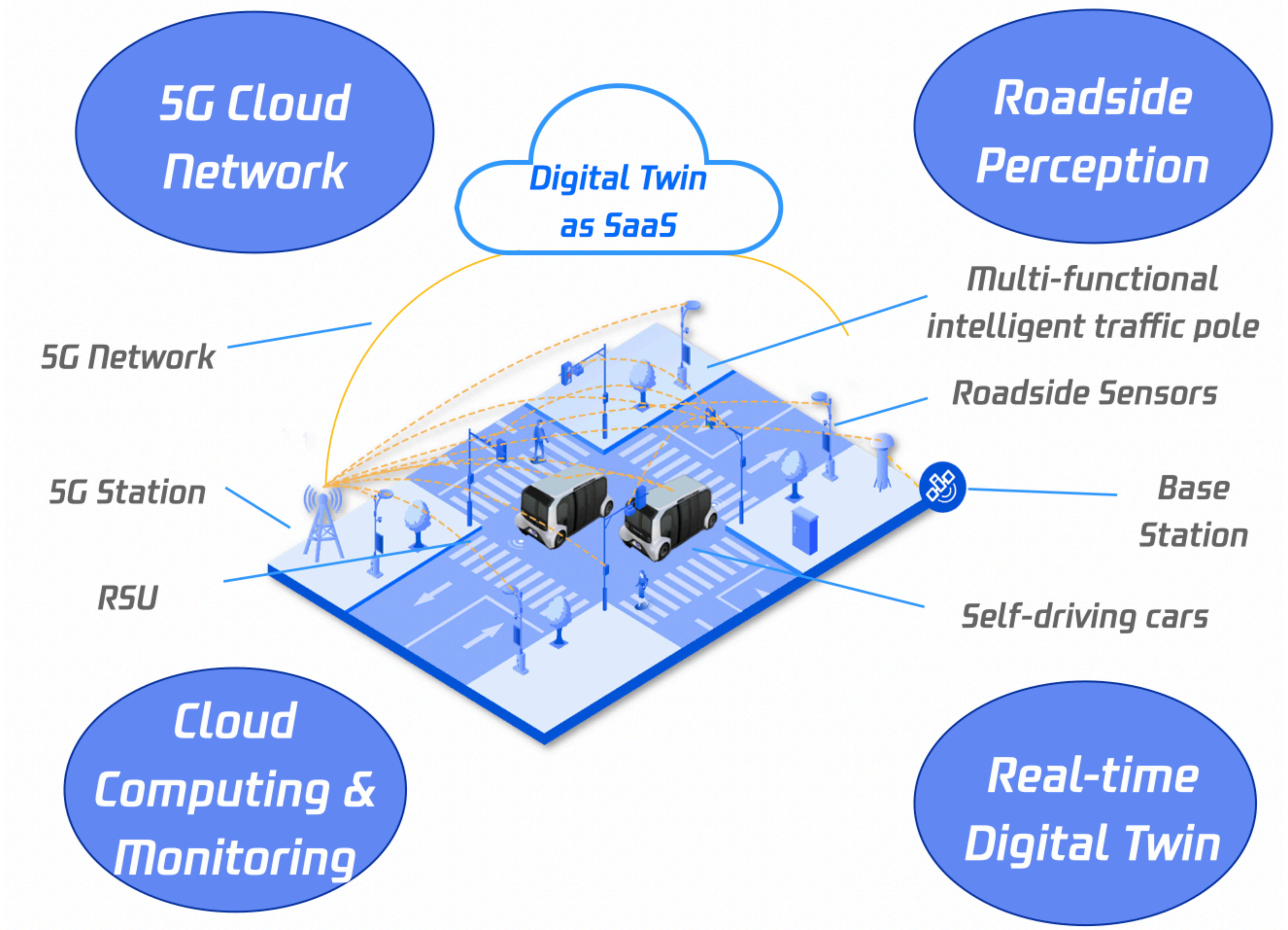
腾讯自动驾驶
Tencent Autonomous Driving

ECCV
EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO
2024

Tencent 腾讯

Background

- Safety Concerns in transportation.
- Vehicle intelligence penetration rate is low.
- Ability of intelligent driving is of concern:
  - Abilities of driving vary.
  - Obstructed sensor views.
  - Sudden appearances of pedestrians.



⬇

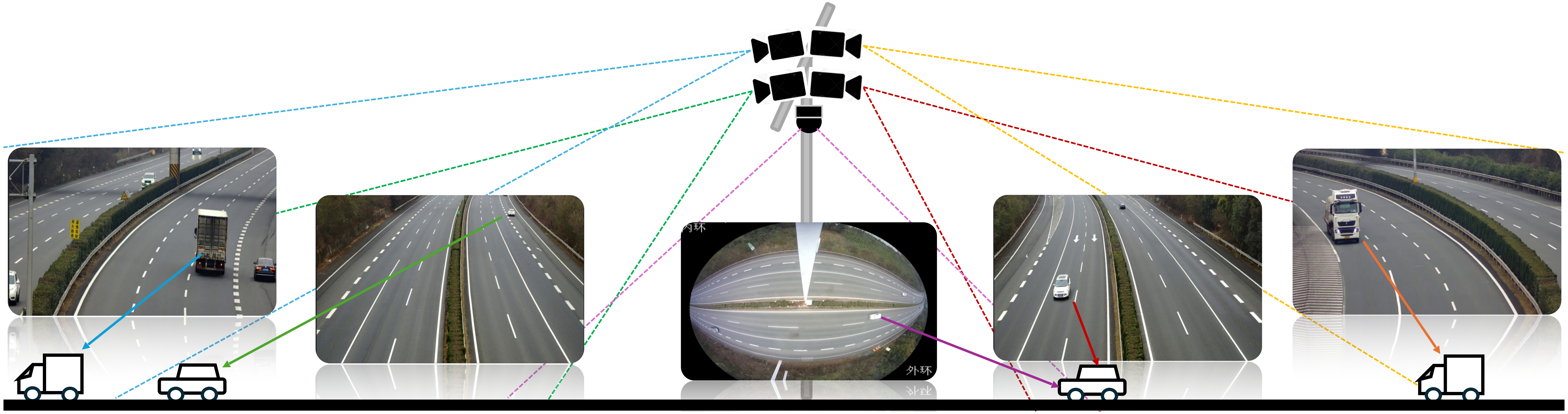
V2X (Vehicle to Everything), an emerging research direction solving above issues by introducing the ability of roadside perception and communication.

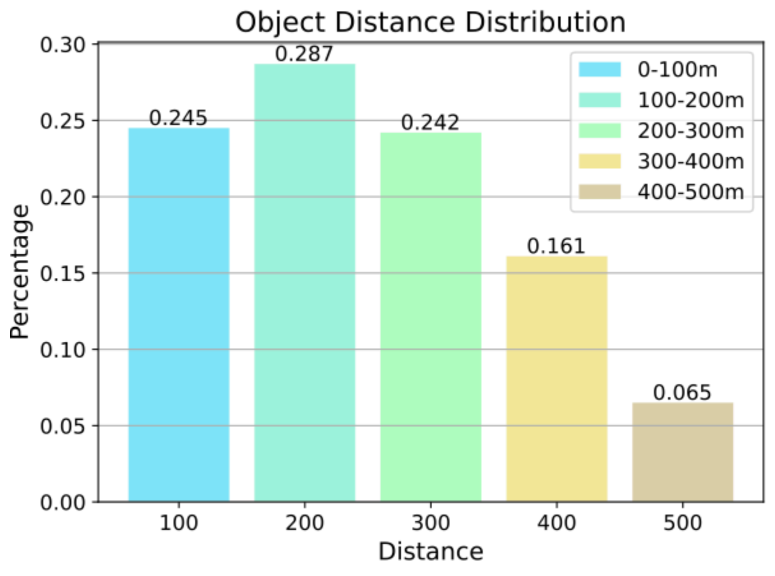
Dataset — Overview

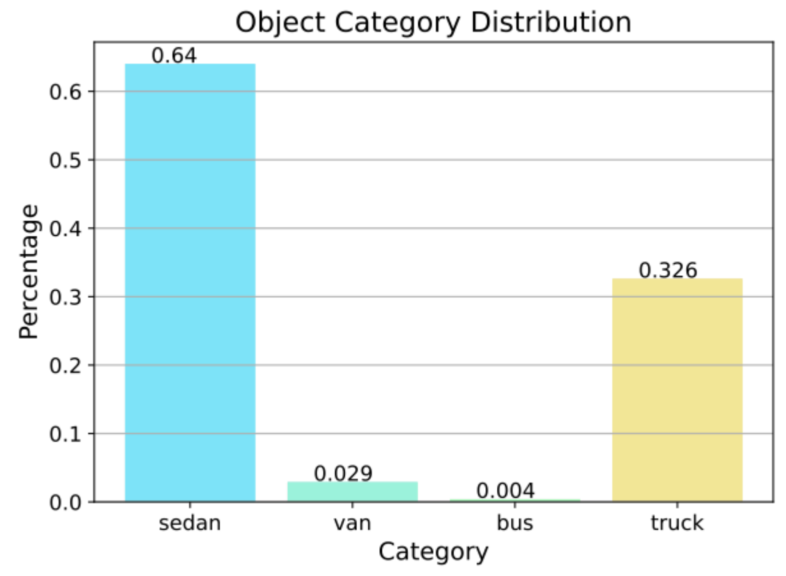
腾讯自动驾驶
Tencent Autonomous Driving

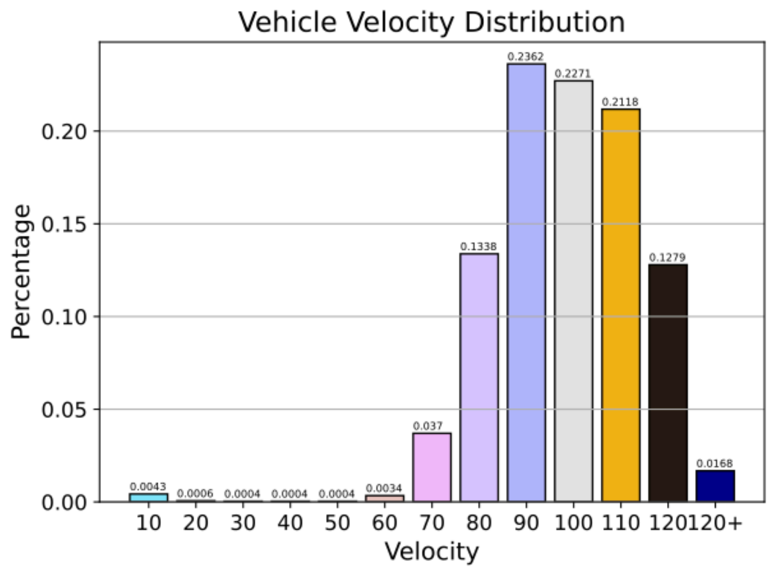
- We introduce H-V2X,

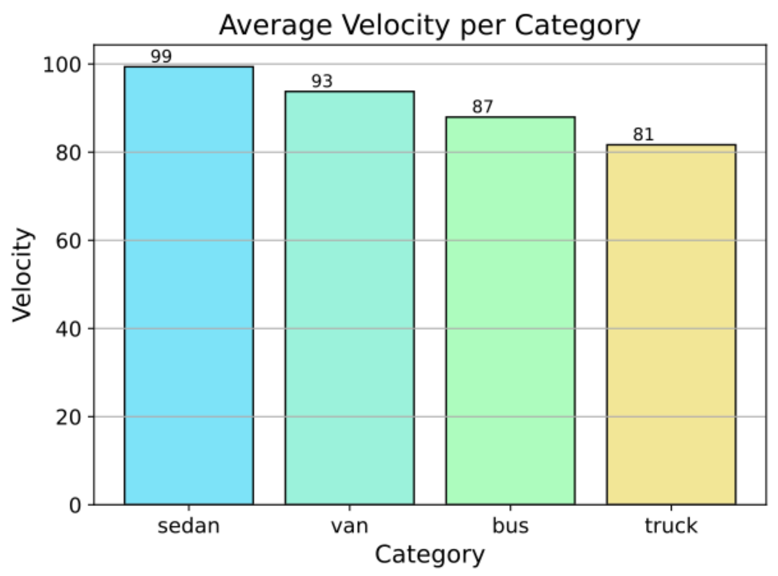
  1. Gound truths are constructed in BEV space across multiple **time and space aligned** sensors, making end-to-end BEV learning possible.
  2. Original sensor data, along with ground truths constructed by algorithms and human labor are provided, over **1.94M** samples.
  3. Three tasks with benchmarks, i.e., BEV detection, BEV tracking, and trajectory prediction are proposed.

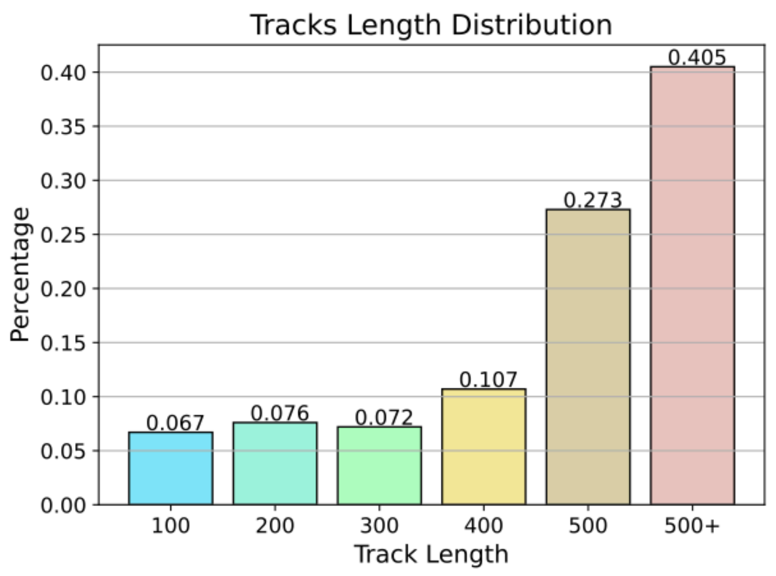


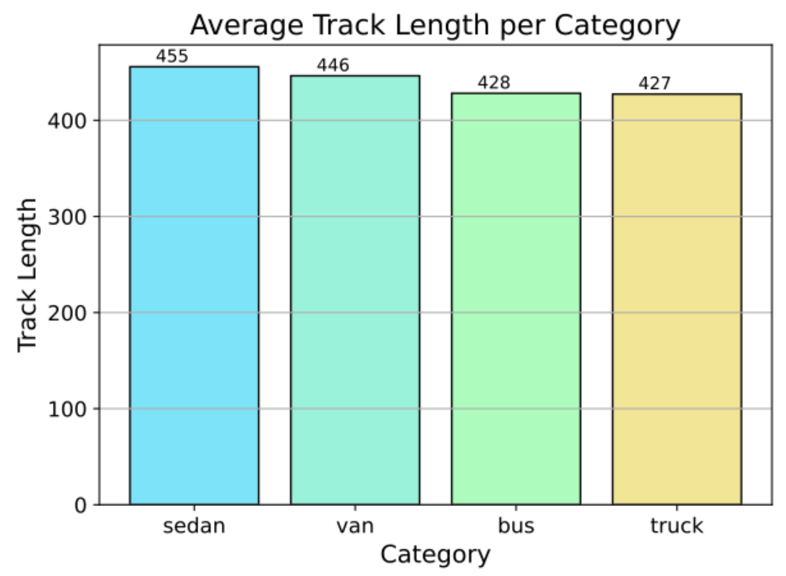












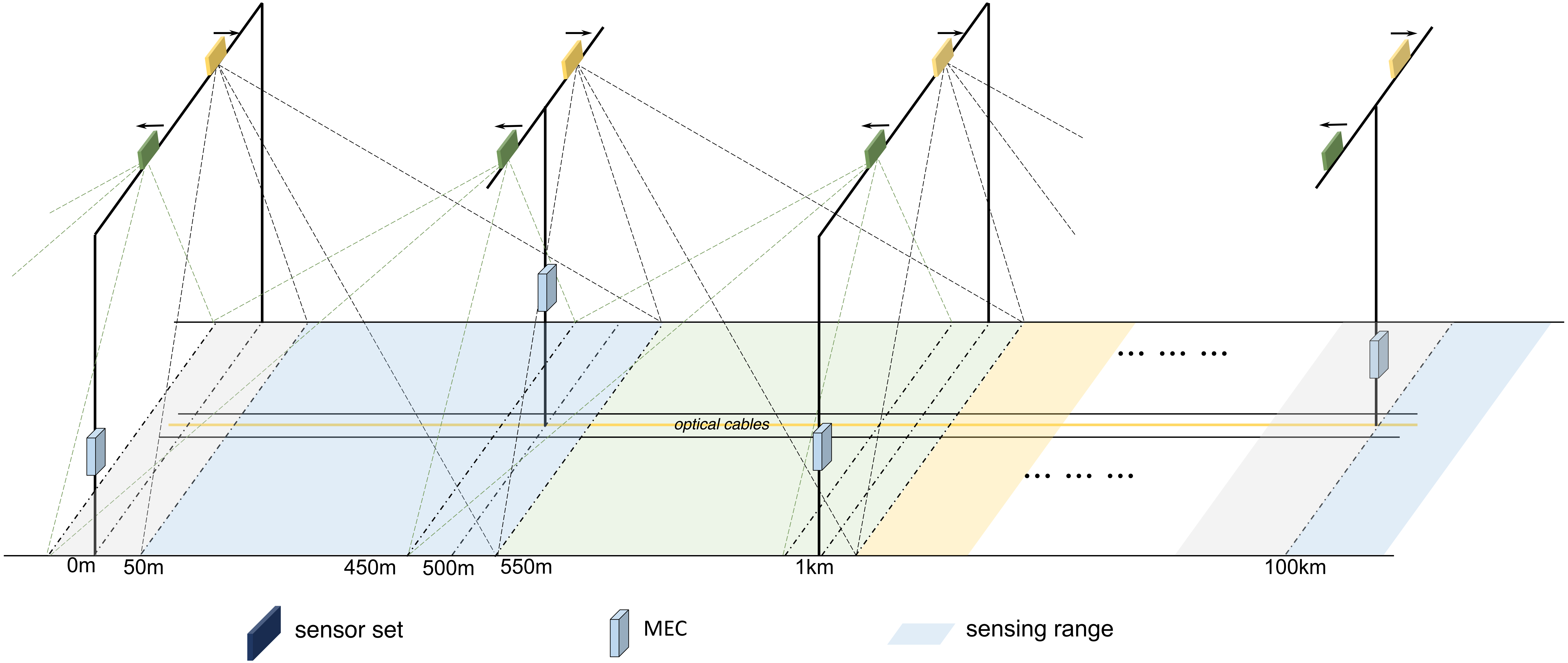
Dataset — Comparison

- Compared to previous roadside perception datasets, H-V2X demonstrates competitive statistics across multiple aspects.

Dataset	Year	Senario	Sim Real	Num Samples	Num Classes	With Map	Provided Sensors	Sequential Trajectories	Supported Tasks	Range | Coverage
WIBAM [11]	2021	Urban	Real	33092	1	X	C	X	Mono3D	1 intersection
Rope3d [35]	2022	Urban	Real	50009	12	X	C	X	Mono3D	200m
DAIR-V2X-C [36]	2022	Urban	Real	12424	10	X	C&L	X	Mono3D	20km
DAIR-V2X-I [36]	2022	Urban	Real	7058	10	X	C&L	X	Mono3D	20km
V2X-Seq [38]	2023	Urban	Real	11275	9	VectorMap	C&L	✓	Tracking 3D & Pred	28 intersections
A9-I [46]	2023	Urban	Real	4800	10	X	C&L	✓	Mono3D	1 intersection
INT2 [29]	2023	Urban	Real	106.8M	1	VectorMap	C&L	✓	Trajectory Prediction	16 intersections
HighD [15]	2018	Highway Drone	Real	1.48M	2	X	C	X	2DBbox	420m
DAIR-V2X-V [36]	2022	Urban Highway	Real	15627	10	X	C&L	X	Mono3D	20km
A9-Dataset [6]	2022	Highway Ramp	Real	1098	9	X	C&L	✓	Mono3D	3km
OPV2V [27]	2022	Urban Rural	Sim	11464	1	X	C&L	X	BEV	70 Scenes
V2X-Sim [18]	2022	Urban	Sim	57200	1	X	C&L	X	BEV & 3D Tracking	100Scenes
Roadside-Opt [14]	2023	Urban	Sim	37641	1	X	L	X	BEV	10 Scenes
H-V2X**(ours)**	2024	Highway	Real	1.94 Million	4	VectorMap	C&R	✓	BEV Det MOT Tracking Trajectory Prediction	> 100km

Dataset — Setups

- Sensors are mounted on masts on the highway, around 12meters high.
- Both side of the highway are covered using long-range cameras, short-range cameras and fisheye camera.
- Each mast is equipped with one MEC (Mobile Edge Computer) to collect sensor data.

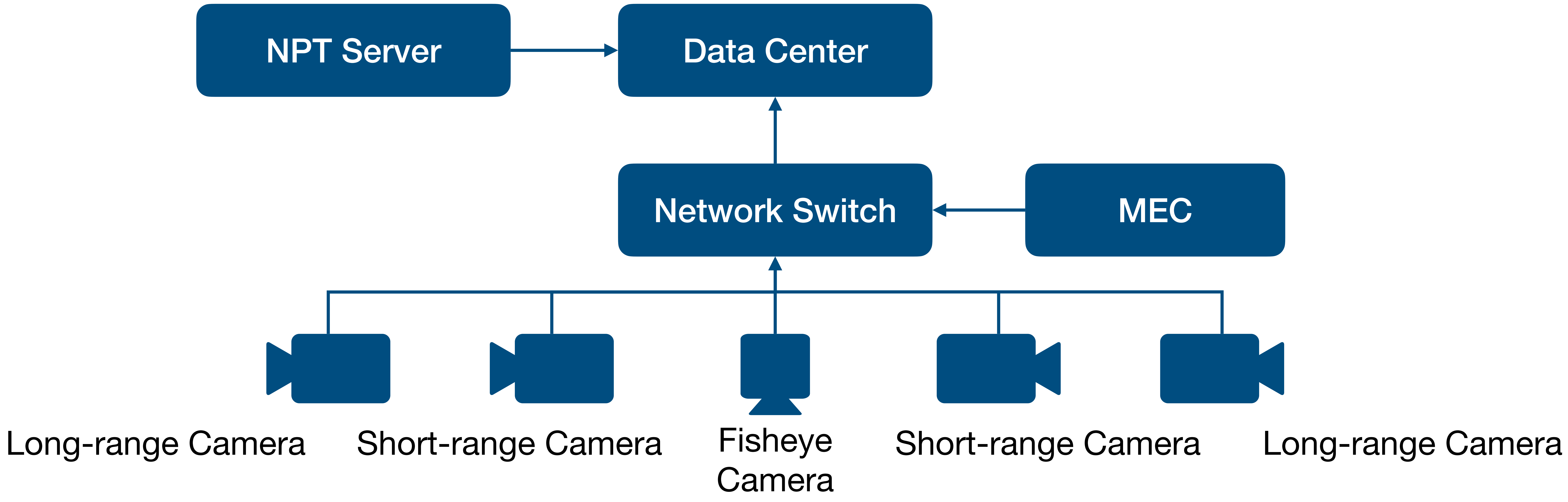


Dataset — Sensors & Topology

腾讯自动驾驶
Tencent Autonomous Driving

Table 2: Sensors Specifications and Configurations

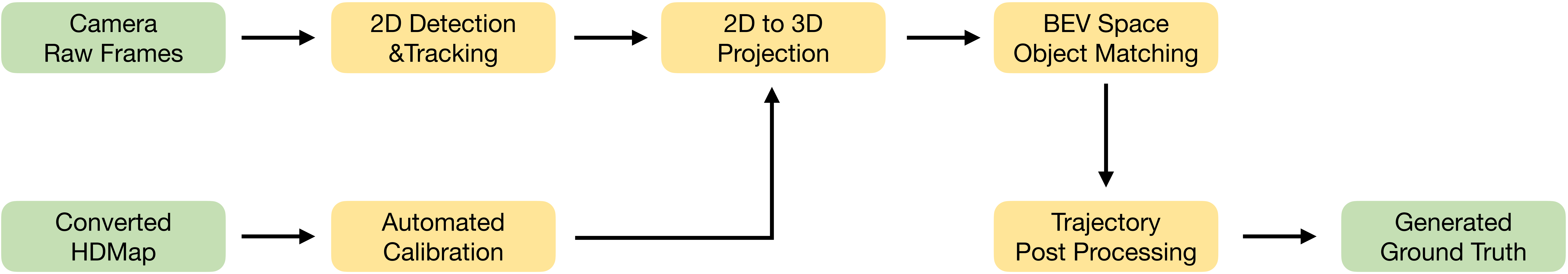
Sensor	Focus	Frequency	Resolution	Format	FOV	BandWidth
Short-range Camera	12mm	10hz	1920x1080	JPEG	49.78	4Mbps
Long-range Camera	70mm	10hz	1920x1080	JPEG	9.1	4Mbps
Fisheye Camera	1.27mm	10hz	1280x1280	JPEG	180	4Mbps



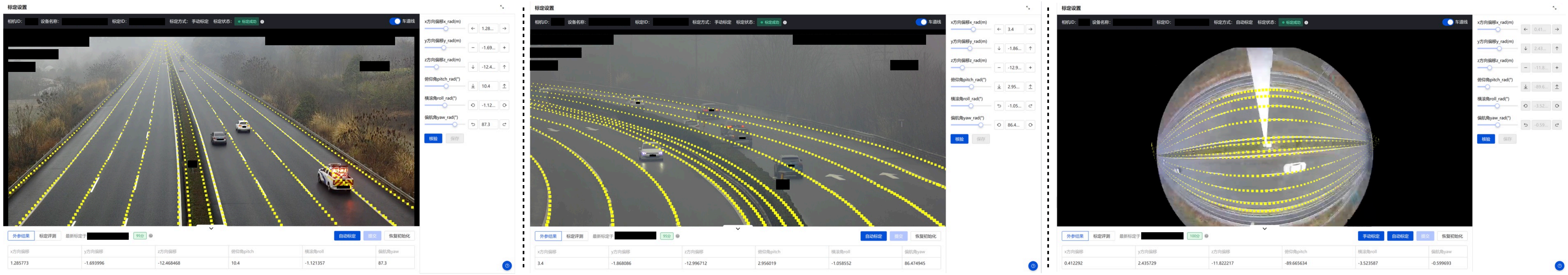
Dataset — Ground Truth Generation

腾讯自动驾驶
Tencent Autonomous Driving

- Vector map and cameras are calibrated using autmoated calibration algorithms and human adjustments.
- BEV ground truths are generated using below pipeline and with human verification.



BEV GT Generation Pipeline



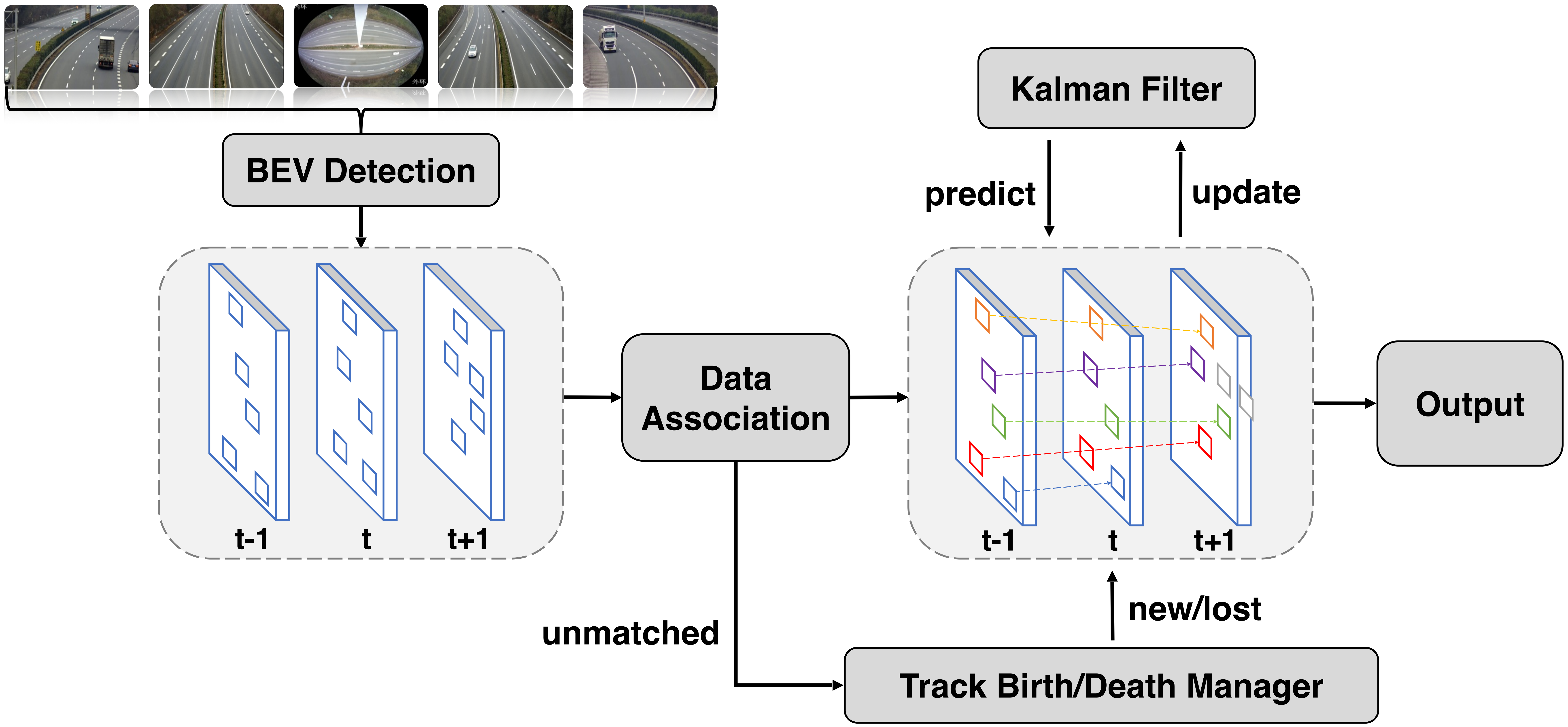
Calibration Software

Benchmarks — BEVDetection

- We propose H-BEV (Highway BEV) model, a BEV detection neural net incorporating fisheye camera and vector map. The model is built on BEVDet, learns depth distribution and construct frustum point cloud, illustrated in Fig. 6. To create the projection coordinates of each camera's pixel points in the BEV perspective, we used 3D lane information for interpolation. We generated the frustum point cloud for each camera through a table lookup and added a attribute branch to predict vehicle attribute information.

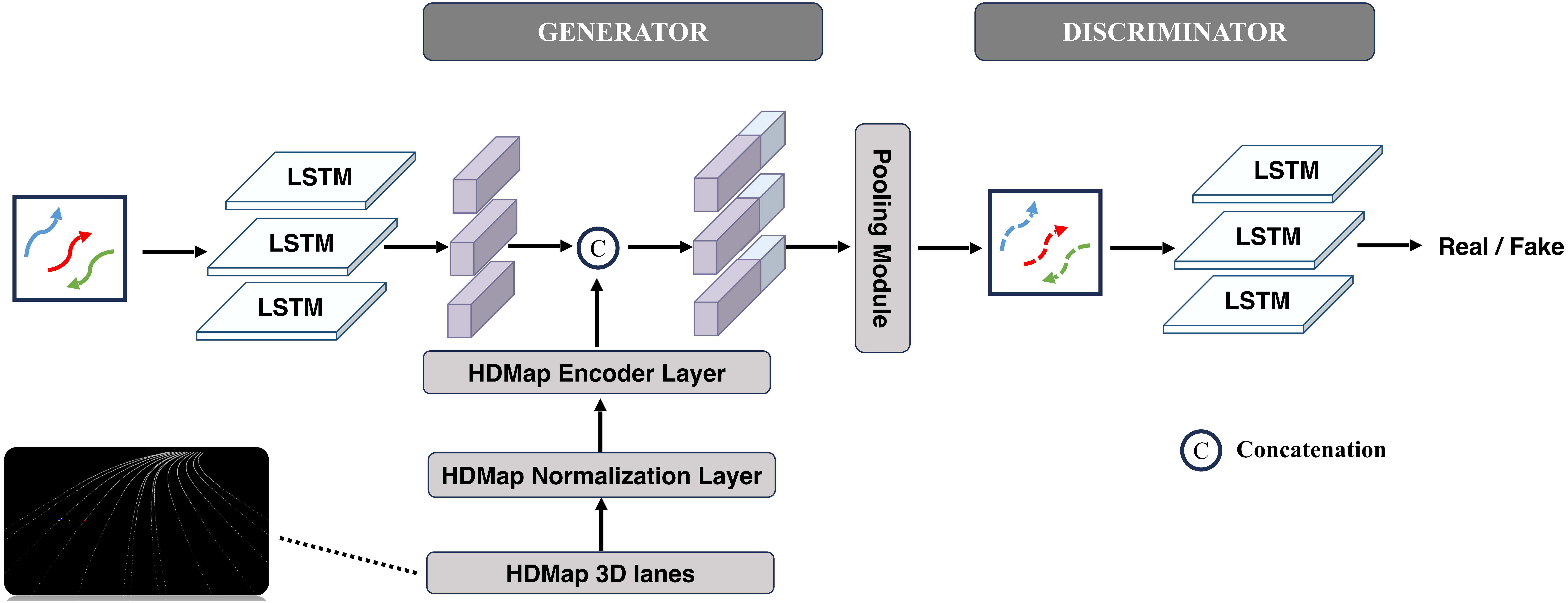
Benchmarks — BEVTracking

- We present a tracking-by-detection MOT algorithm as the baseline approach. We utilize BEV detection results from multiple cameras to trace the trajectory of each object within the perception region. We adapt a variant of the SORT algorithm to establish associations between objects across successive frames. This leads to the assignment of consistent IDs to individual objects, resulting in continuous object trajectories.



Benchmarks — Trajectory Prediction

- We propose HD-GAN, a network that uses global map information to extract the global positional feature of an object. The HDMap Layer takes in lane information and calculates the object's global position via interpolation, normalizing it to [0,1] in the HDMap Normalization Layer. The HDMap Encoder Layer encodes the object's global map positional features, which are then combined with the local positional features of the observed trajectory.



Thank You !