QCRAFT

Data-Efficient Perception for Autonomous Driving

Xiaodong Yang
Principal Scientist & Head of Research

Bring Autonomous Driving into Real Life

Bring Autonomous Driving into Real Life

Driven-by-QCraft

Rapid Growth and Expansion

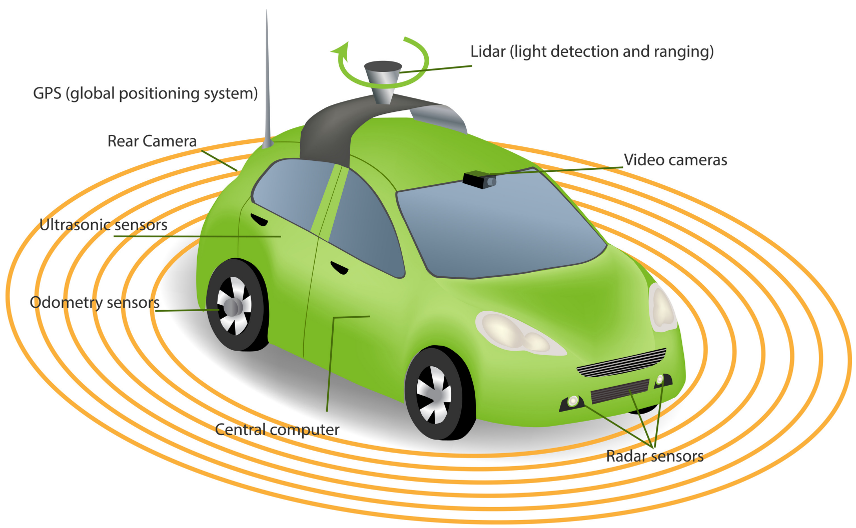
Our business applications

Since October 2020, the Longzhou One, the autonomous minibus using QCraft's technology, has breezed through its trials on open roads in Suzhou, Shenzhen, Jiaxing, Xi'an, Wuhan and the other cities.

In Suzhou, more than 11,000 passengers have taken a ride around Robobus Line Q1 in the driverless vehicle as of March 31, with an average daily passenger count of about 116.

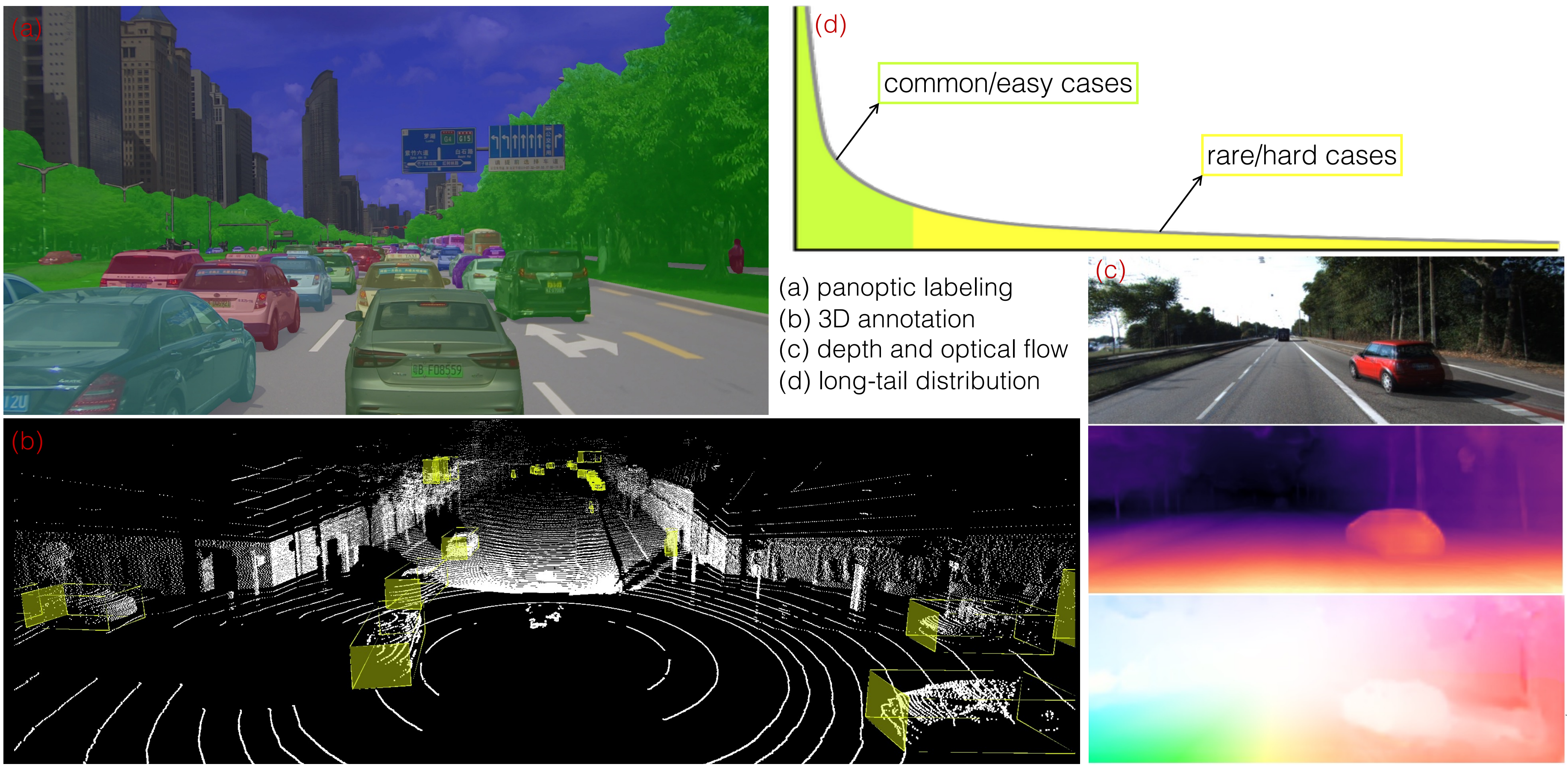
Data Flood of Intelligent Vehicles

- Over 4 TB data produced by a single vehicle per day
- Less than 5% of the data accessed for development
- Growing fleet of self-driving cars



QCRAFT

Adversity in Data Mining and Labeling



(a) panoptic labeling
(b) 3D annotation
(c) depth and optical flow
(d) long-tail distribution

Data-Efficient Learning

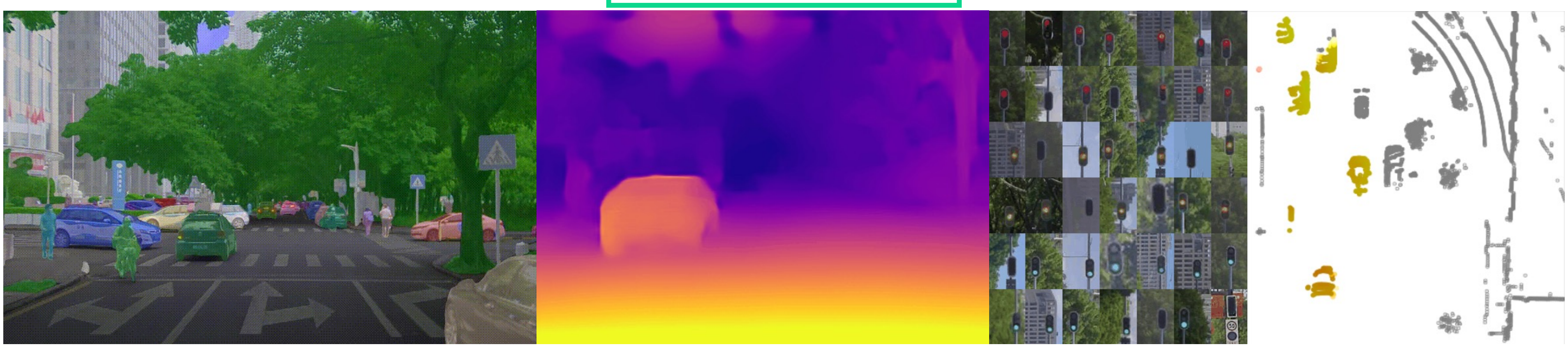
data synthesis | self-supervised learning | semi-supervised learning

↓

self-driving data with incomplete labels

↓

perception models



Semi-Supervised Panoptic Segmentation

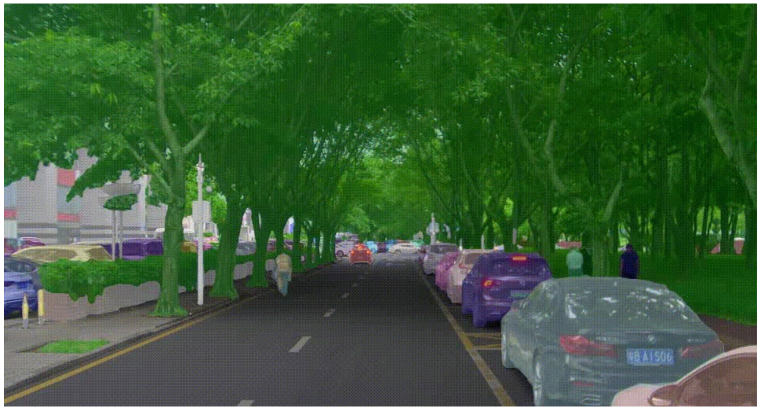
data synthesis

self-supervised learning

semi-supervised learning

self-driving data with incomplete labels

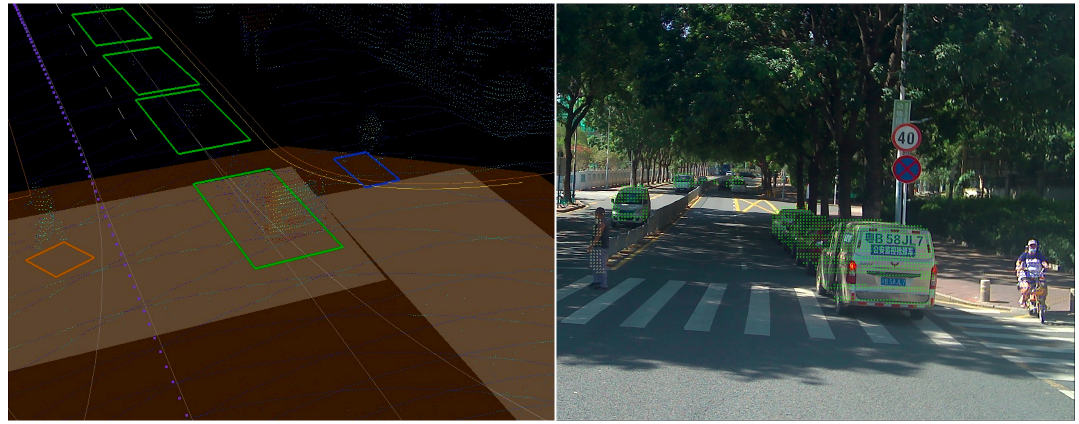
panoptic segmentation model



QCRAFT

Semi-Supervised Panoptic Segmentation

spatial semi-supervised learning



cross-sensor semi-supervised learning

temporal semi-supervised learning

QCRAFT

Semi-Supervised Panoptic Segmentation

QCRAFT

Self/Semi-Supervised Depth Estimation

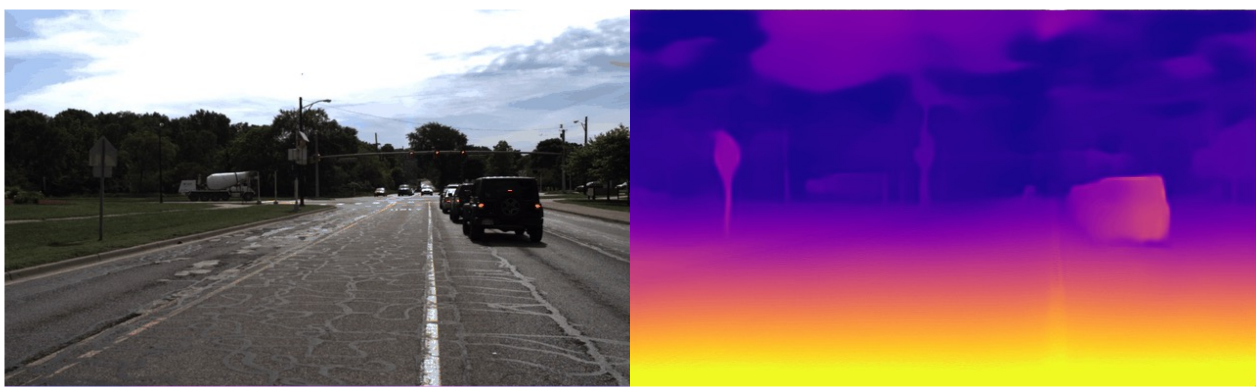
data synthesis | self-supervised learning | semi-supervised learning

↓

self-driving data with incomplete labels

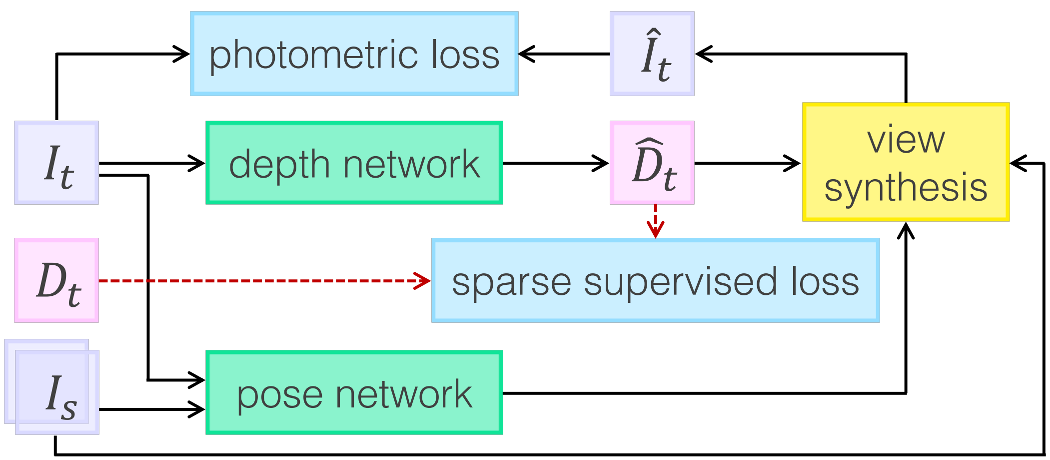
↓

depth estimation model

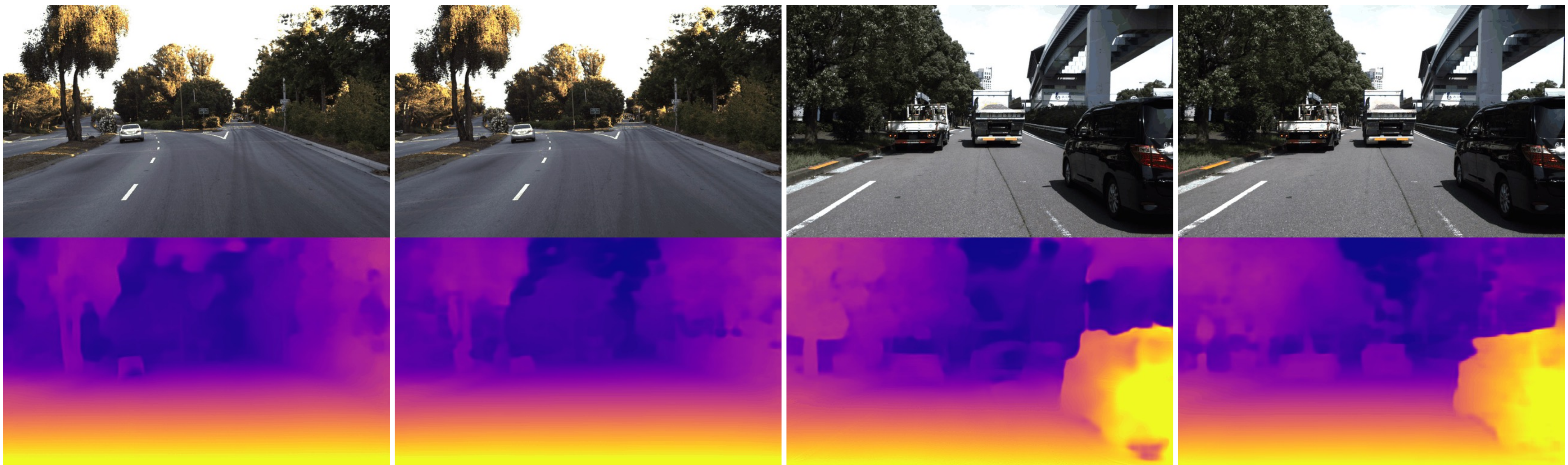


QCRAFT

Self/Semi-Supervised Depth Estimation



Self/Semi-Supervised Depth Estimation



scene-1 self-supervised	scene-1 semi-supervised	scene-2 self-supervised	scene-2 semi-supervised

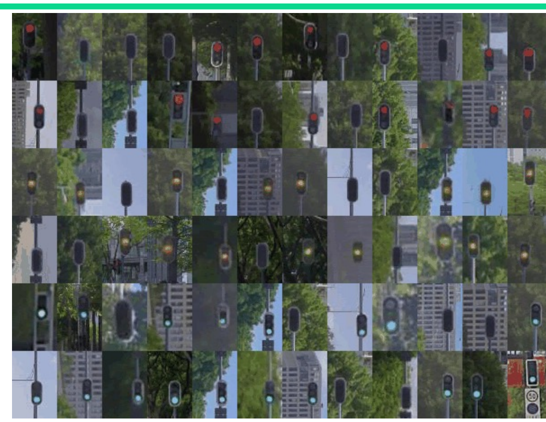
QCRAFT

Data Synthesis for Traffic Light Recognition

data synthesis | self-supervised learning | semi-supervised learning

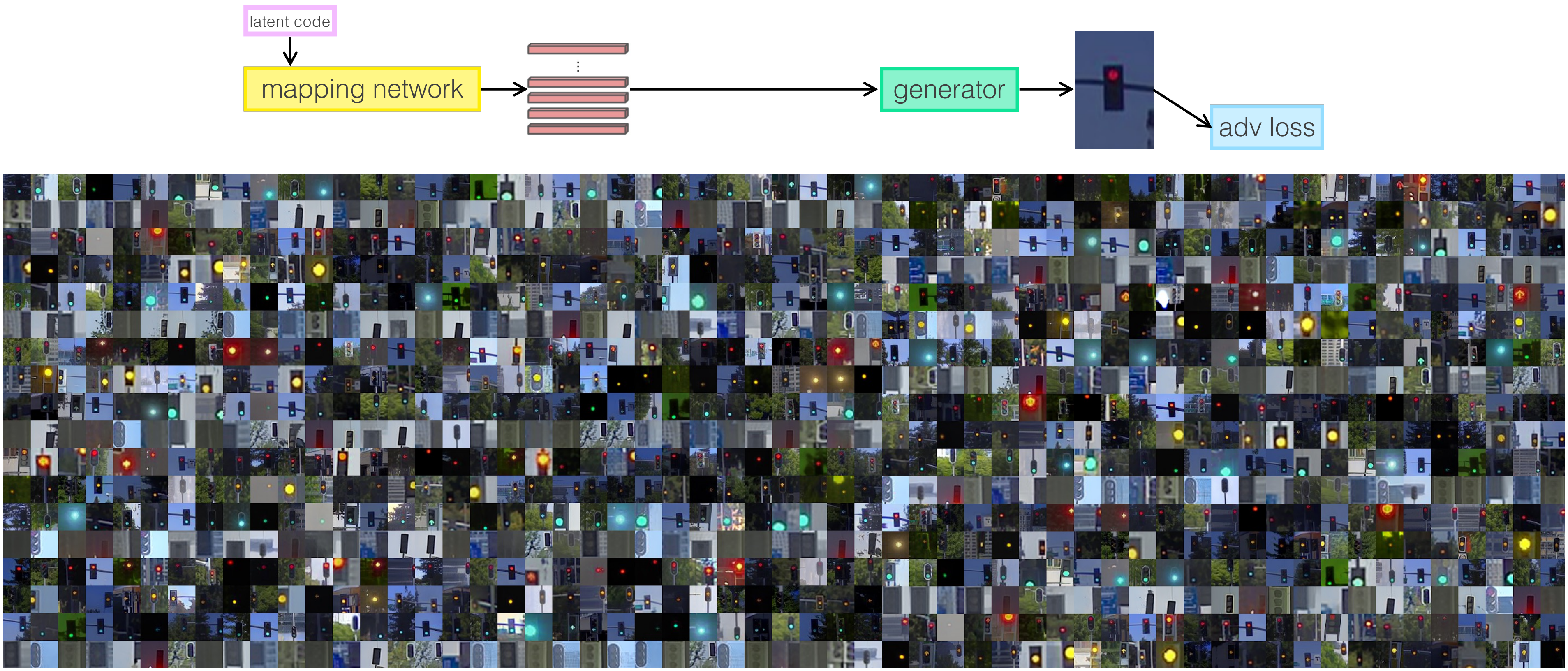
self-driving data with incomplete labels

traffic light recognition model



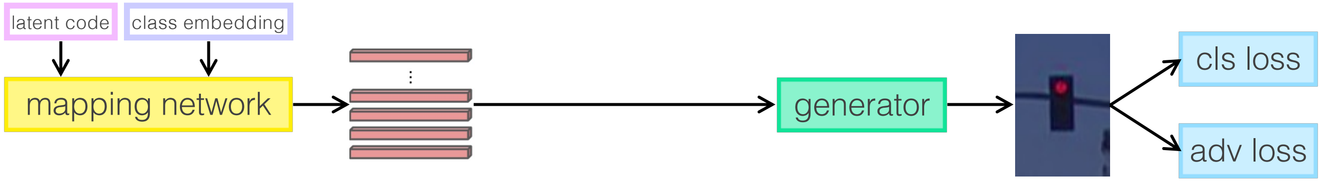
QCRAFT

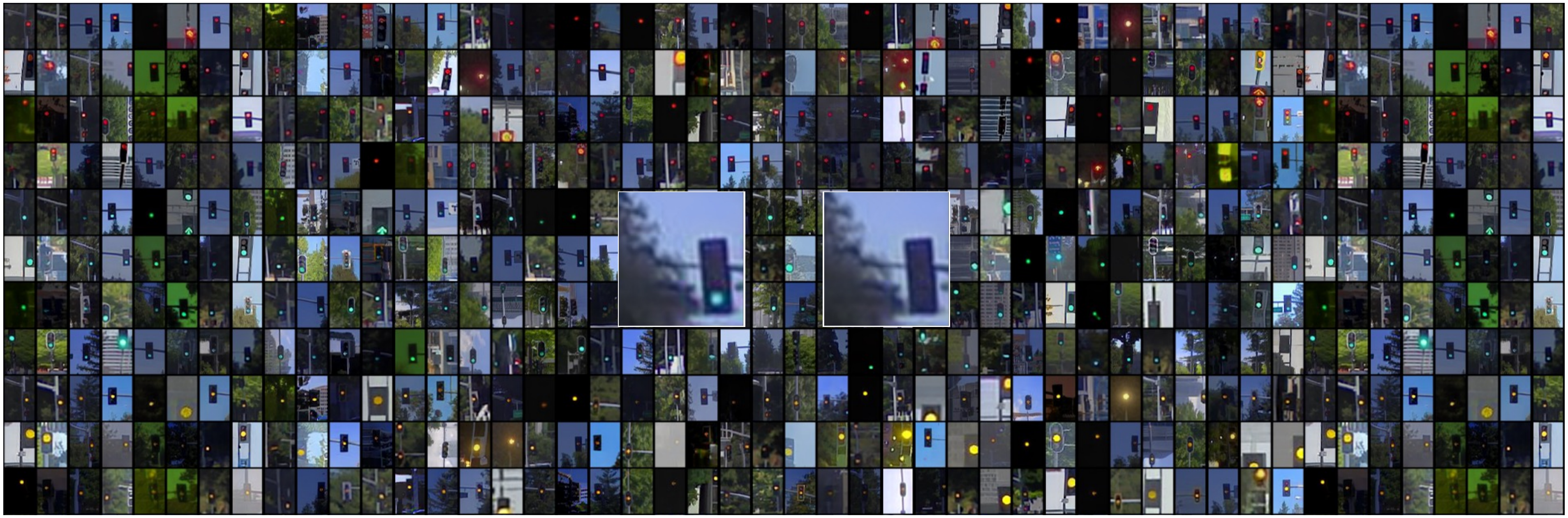
Data Synthesis for Traffic Light Recognition



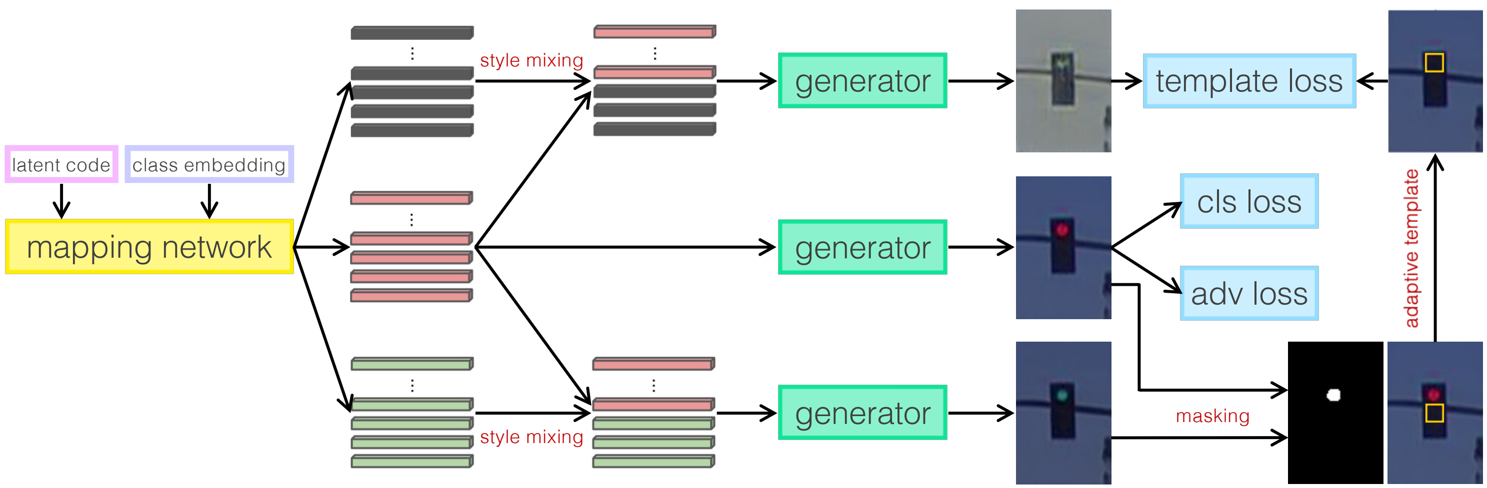
A Style-Based Generator Architecture for Generative Adversarial Networks. CVPR 2019.

Data Synthesis for Traffic Light Recognition

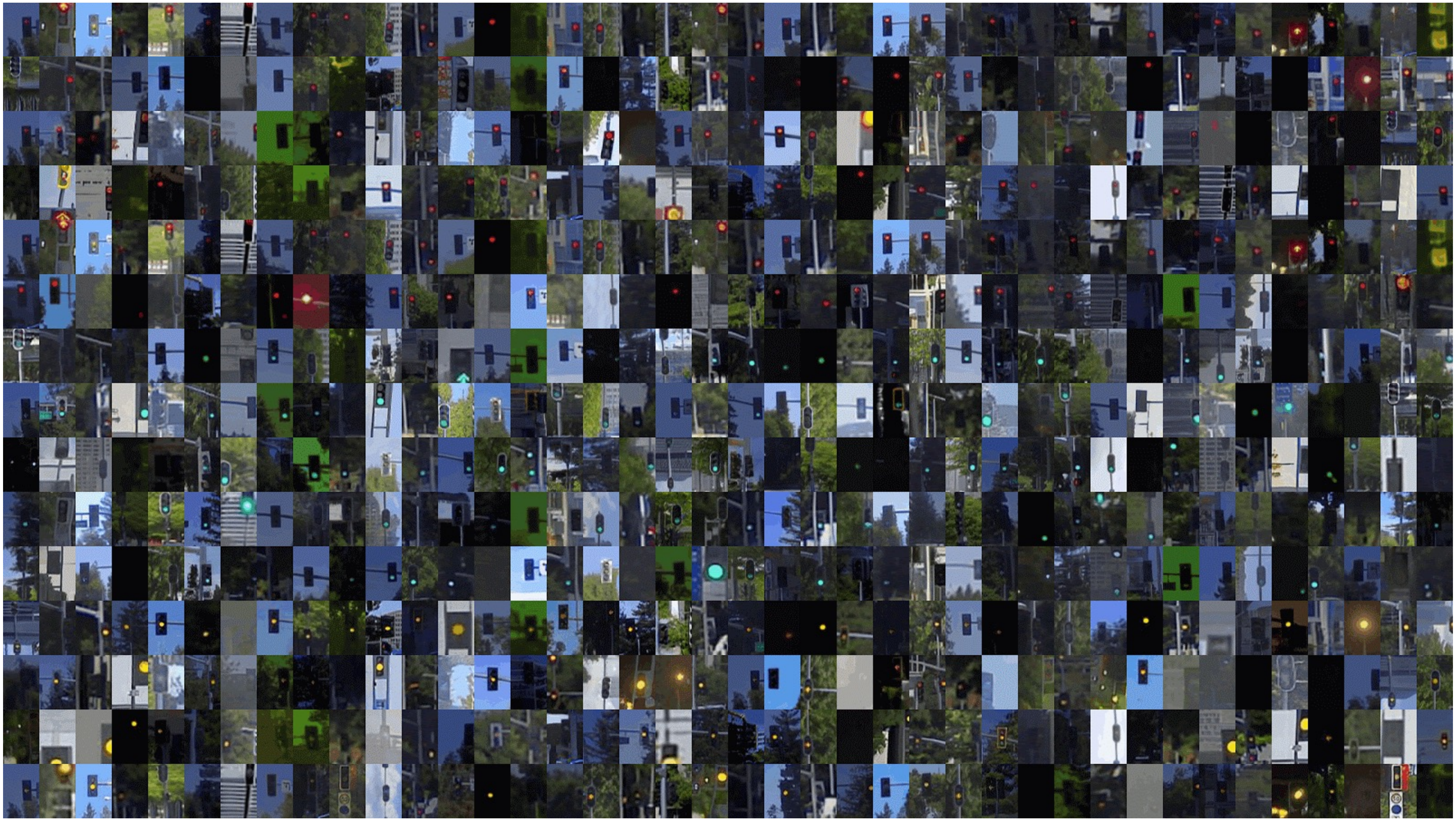




Data Synthesis for Traffic Light Recognition



QCRAFT



Self-Supervised Point Cloud Motion Prediction

data synthesis | self-supervised learning | semi-supervised learning

↓

self-driving data with incomplete labels

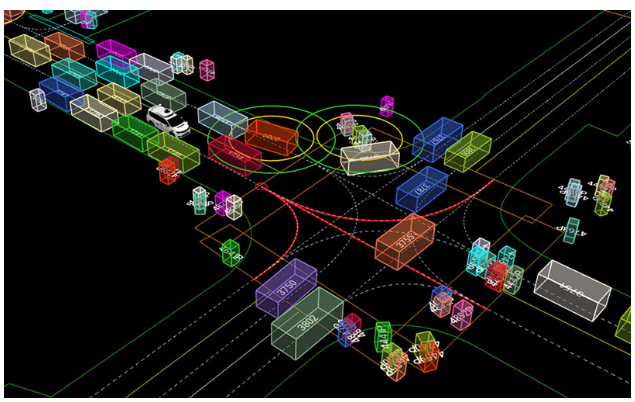
↓

pillar motion prediction model

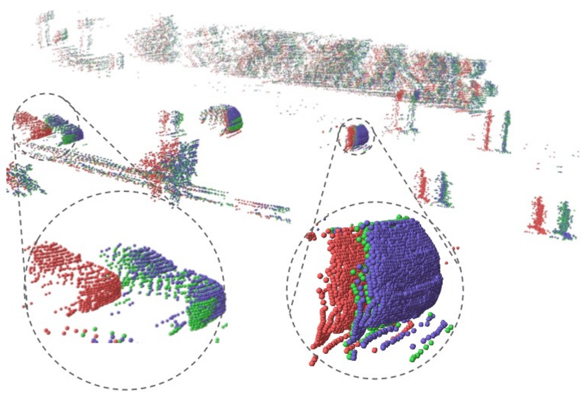
QCRAFT

Self-Supervised Point Cloud Motion Prediction

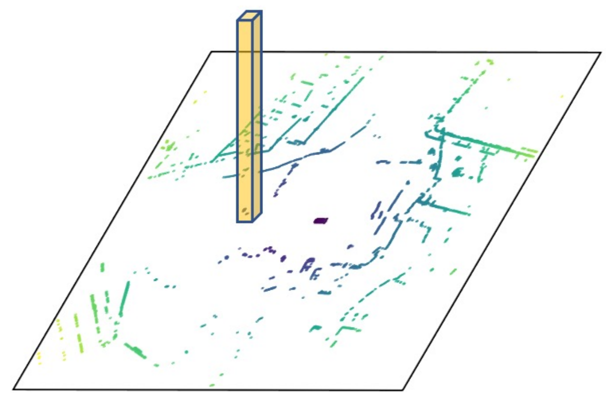
- Understanding point cloud motion
  - Facilitates various self-driving modules in highly dynamic environments
  - Existing: closed-set, computationally prohibitive, labeling extensive
  - Ours: open-set, real-time, self-supervised



trajectory prediction



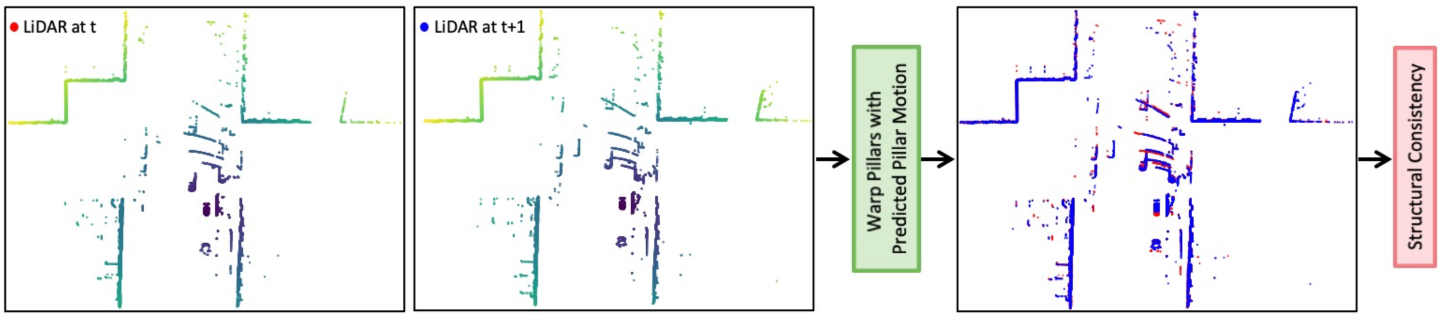
point cloud scene flow



point cloud pillar motion

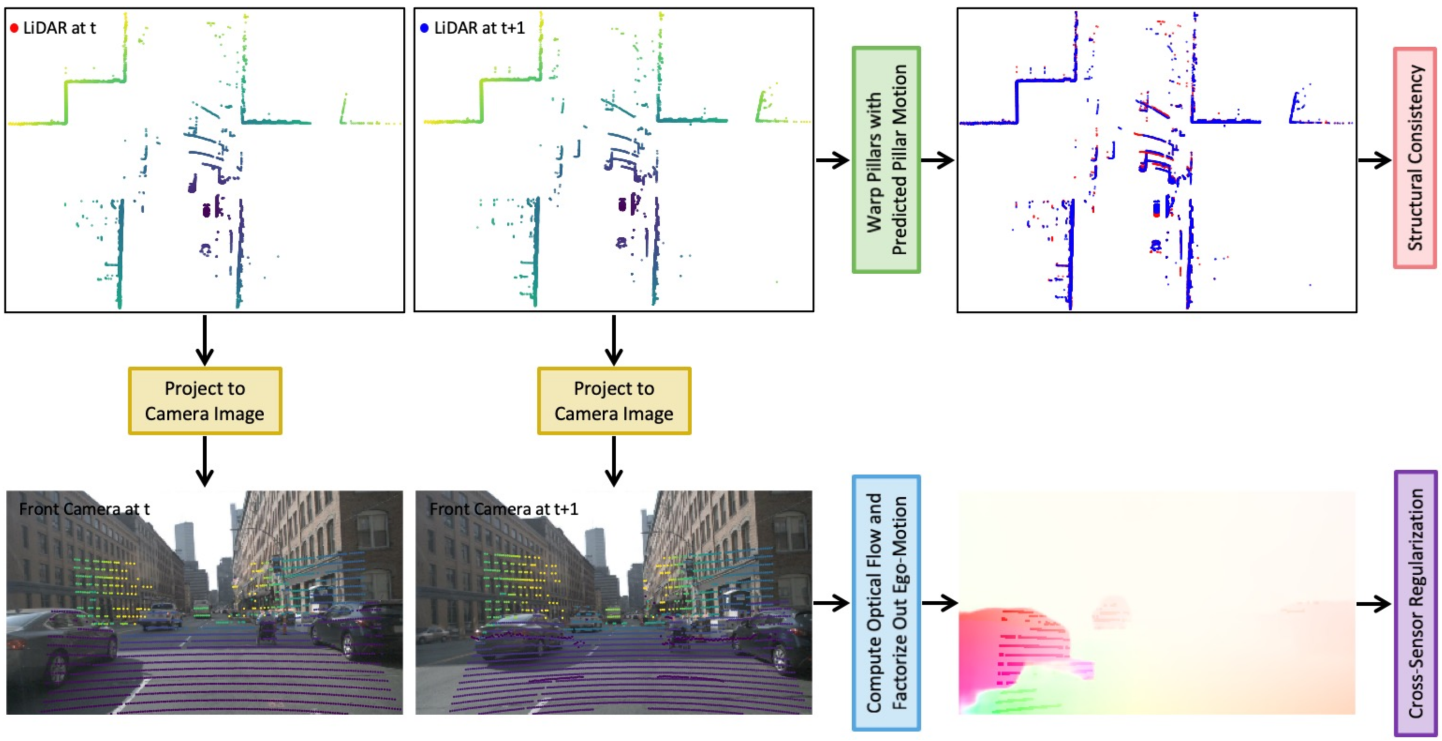
Self-Supervised Point Cloud Motion Prediction

- LiDAR based structural consistency



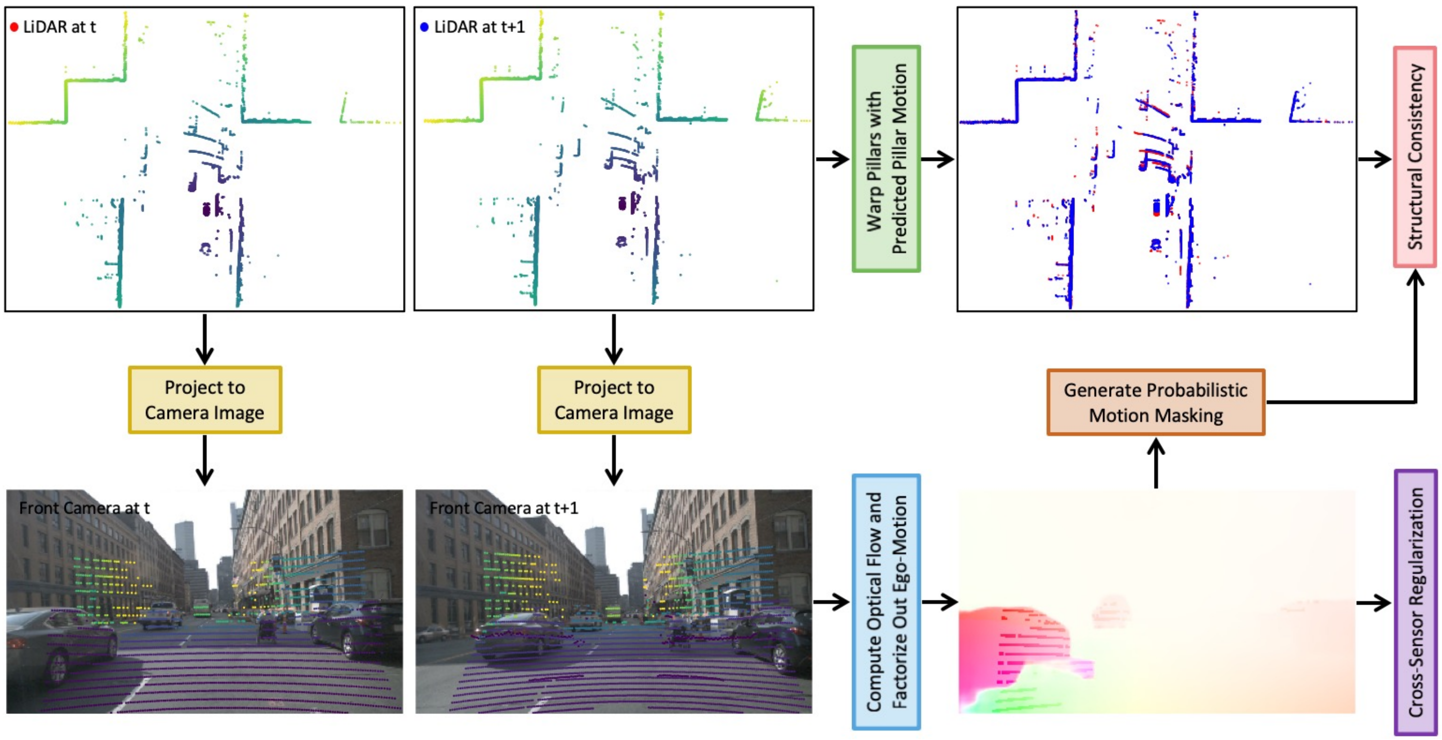
Self-Supervised Point Cloud Motion Prediction

- Cross-sensor motion regularization



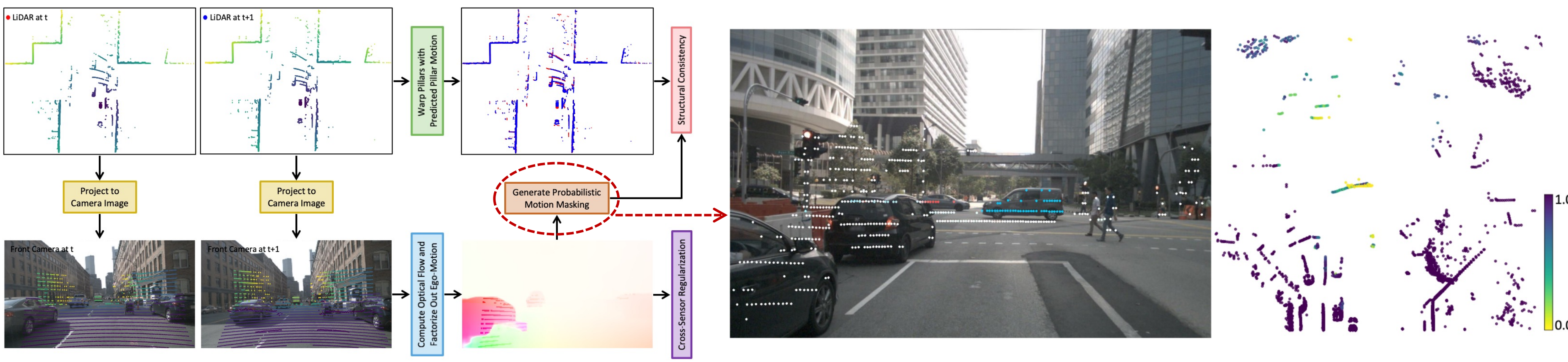
Self-Supervised Point Cloud Motion Prediction

- Probabilistic motion masking



Self-Supervised Point Cloud Motion Prediction

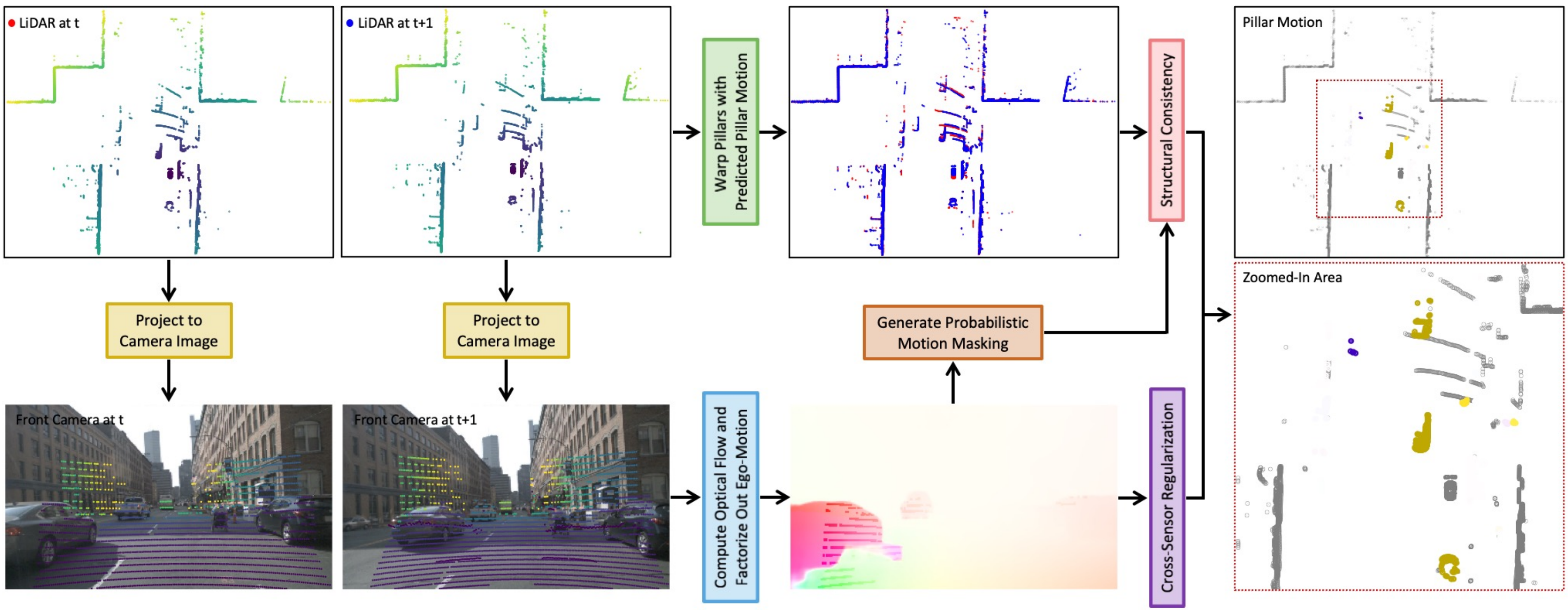
- Probabilistic motion masking



QCRAFT

Self-Supervised Point Cloud Motion Prediction

○ Method



Self-Supervised Point Cloud Motion Prediction

Results

	$\mathcal{L}_{\text{consist}}$	$\mathcal{L}_{\text{regular}}$	Mask	Static		Speed $\leq$ 5m/s		Speed > 5m/s		Nonempty		Foreground		Moving	
				Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
(a)	✓			0.3701	0.0063	0.5014	0.1352	1.9405	1.2760	0.3437	0.0081	0.5936	0.1139	0.7516	0.2359
(b)		✓		**0.0285**	**0.0002**	0.3733	**0.0719**	4.2954	3.9788	0.0897	0.0020	0.7914	0.0656	1.1267	0.3948
(c)	✓	✓		0.1688	0.0389	0.4277	0.1694	1.7603	1.2021	0.3133	0.0062	0.5667	0.1017	0.7064	0.1980
(d)	✓		✓	0.0738	0.0038	0.4017	0.1214	1.9384	1.2931	0.1085	0.0007	0.5416	0.0767	0.8064	0.2279
(e)	✓	✓	✓	0.0619	0.0004	**0.3438**	0.1196	**1.7119**	**1.1438**	**0.0846**	**0.0001**	**0.4494**	**0.0507**	**0.5953**	**0.1612**

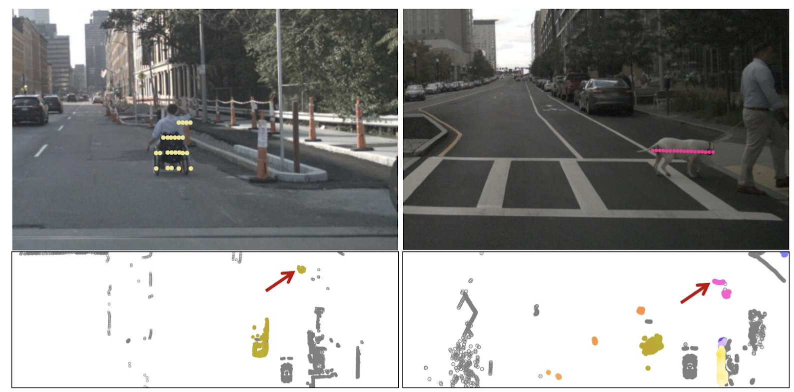
Comparison of our models using different combinations of the proposed structural consistency, cross-sensor regularization, and probabilistic motion masking.

Amount	Self-Supervised	Static		Speed $\leq$ 5m/s		Speed > 5m/s	
		Mean	Median	Mean	Median	Mean	Median
0%	✓	0.1620	0.0010	0.6972	0.1758	3.5504	2.0844
20%	✗	0.0473	0.0001	0.4635	0.1400	2.0946	1.1676
	✓	0.0394	0.0001	0.2970	0.1309	1.0280	0.6055
40%	✗	0.0459	0.0001	0.3712	0.1385	1.7060	0.8950
	✓	0.0329	0.0000	0.2813	0.1280	0.8923	0.5287
60%	✗	0.0412	0.0001	0.3082	0.1338	1.0912	0.6830
	✓	0.0352	0.0000	0.2801	0.1297	0.8499	0.5148
80%	✗	0.0347	0.0001	0.2930	0.1322	0.9824	0.6110
	✓	0.0247	0.0000	0.2301	0.0933	0.7788	0.4700

Benefits of our self-supervised pre-training under different amounts of training data.

Method	Static		Speed $\leq$ 5m/s		Speed > 5m/s		Time
	Mean	Median	Mean	Median	Mean	Median	
FlowNet3D (pre-trained) [20]	2.0514	**0.0000**	2.2058	0.3172	9.1923	8.4923	0.434s
HPLFlowNet (pre-trained) [11]	2.2165	1.4925	1.5477	1.1269	5.9841	4.8553	0.352s
Ours (self-supervised)	**0.1620**	0.0010	**0.6972**	**0.1758**	**3.5504**	**2.0844**	0.020s
FlowNet3D [20]	0.0410	0.0000	0.8183	0.1782	8.5261	8.0230	0.434s
HPLFlowNet [11]	0.0041	0.0002	0.4458	0.0969	4.3206	2.4881	0.352s
PointRCNN [31]	0.0204	0.0000	0.5514	0.1627	3.9888	1.6252	0.201s
LSTMEncoderDecoder [30]	0.0358	0.0000	0.3551	0.1044	1.5885	1.0003	0.042s
MotionNet [39]	0.0239	0.0000	0.2467	0.0961	1.0109	0.6994	0.019s
MotionNet (pillar-based) [39]	0.0258	-	0.2612	-	1.0747	-	0.019s
MotionNet+MGDA [39]	**0.0201**	0.0000	0.2292	0.0952	0.9454	0.6180	0.019s
Ours (fine-tuned)	0.0245	**0.0000**	**0.2286**	**0.0930**	**0.7784**	**0.4685**	0.020s

Comparison with the state-of-the-art results.



Examples of perceiving rare objects by pillar motion: wheelchair and dog, which are not seen in the training of 3D object detection.

Pillar Motion Prediction

by Our Self-Supervised Model

Pillar Motion Prediction

by Our Self-Supervised Model with Fine-Tuning

Self-Supervised Point Cloud Motion Prediction

Chenxu Luo, Xiaodong Yang, Alan Yuille

Self-Supervised Pillar Motion Learning for Autonomous Driving. CVPR 2021.

To learn more about this paper, please visit our project page at https://github.com/qcraftai/pillar-motion.

Summary

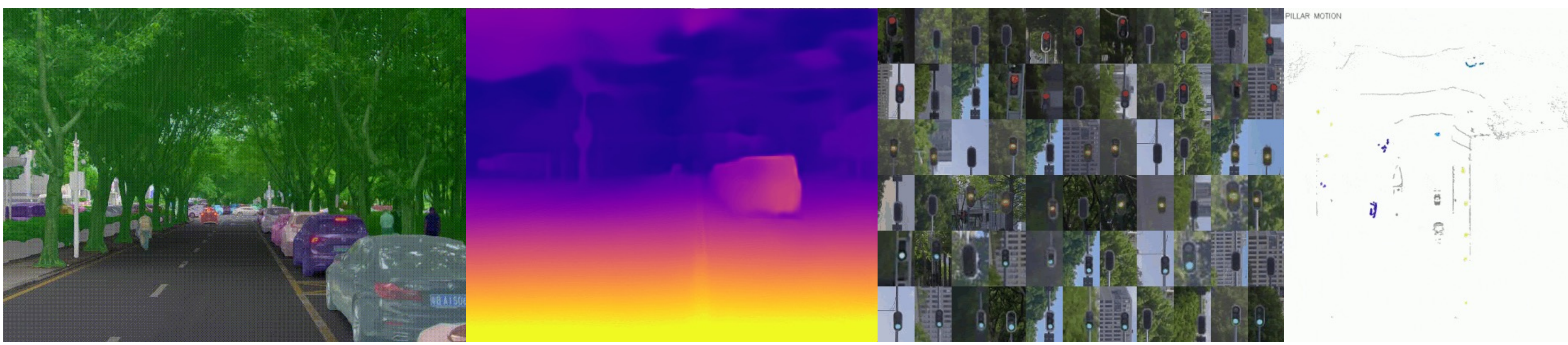
data synthesis | self-supervised learning | semi-supervised learning

↓

self-driving data with incomplete labels

↓

perception models



Bring Autonomous Driving into Real Life.

Thanks!

Questions?

QCRAFT