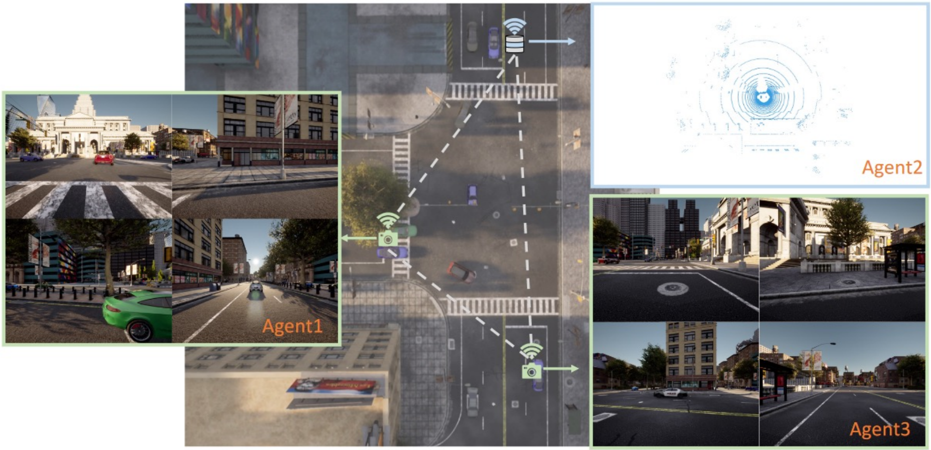
RCDN: Towards Robust Camera-Insensitivity Collaborative Perception via Dynamic Feature-based 3D Neural Modeling

Tianhang Wang, Fan Lu, Zehan Zheng, Guang Chen, Changjun Jiang
Robotics & Embodied AI Lab, Tongji University

Lab Page: https://www.embodiment.ai

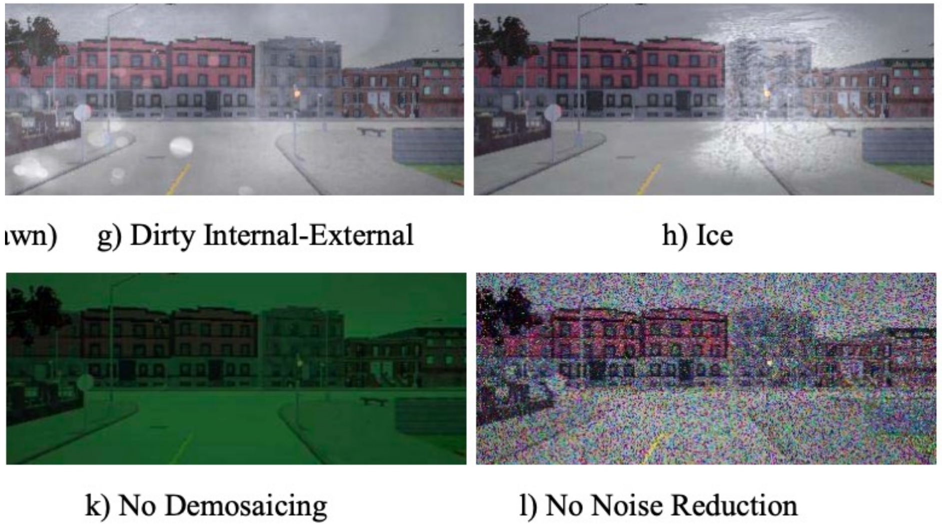
Existing Problem

- Harsh realities of real-world sensors in collaboration
  - **Blurred**
  - **High noise**
  - **Interruption and even failure**



The number of cameras used in the collaborative process

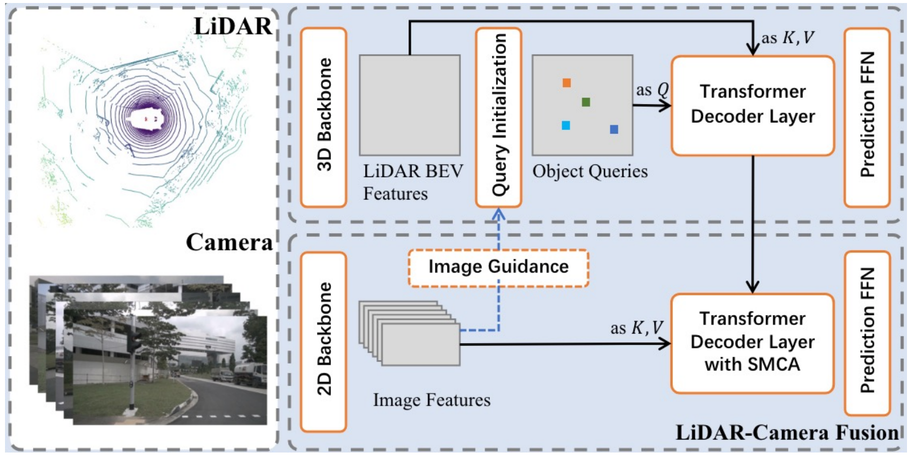
Typical Camera Fault Analysis in Realistic Scenarios



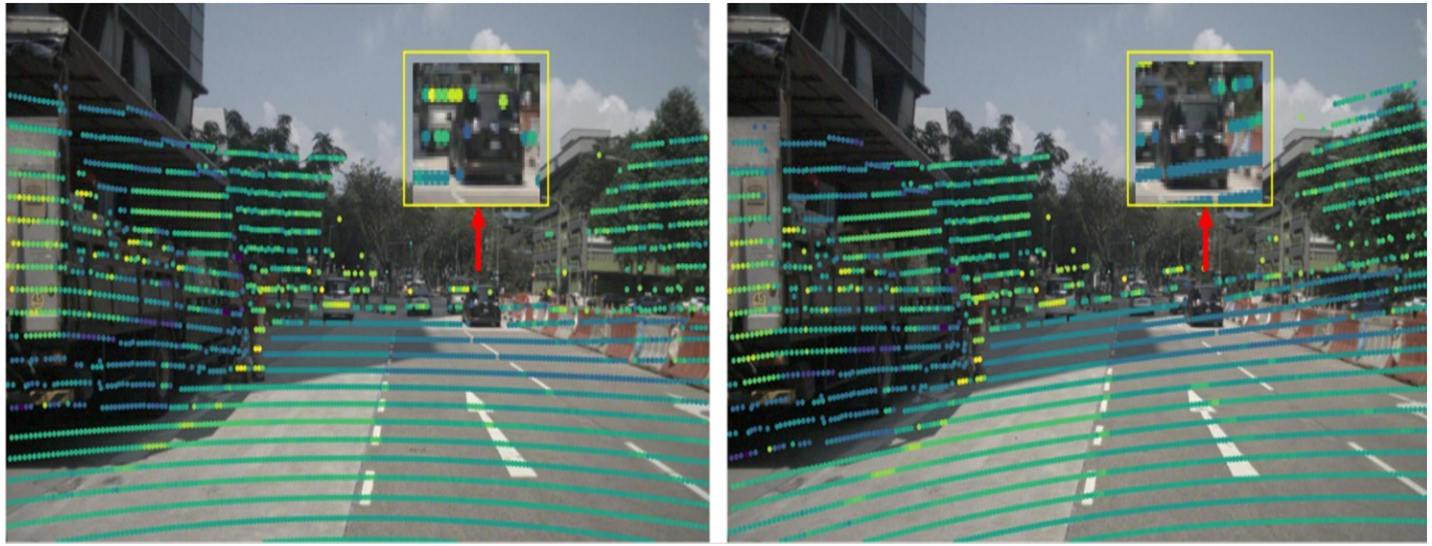
Summary of Camera Malfunctions

Existing Technology Routine

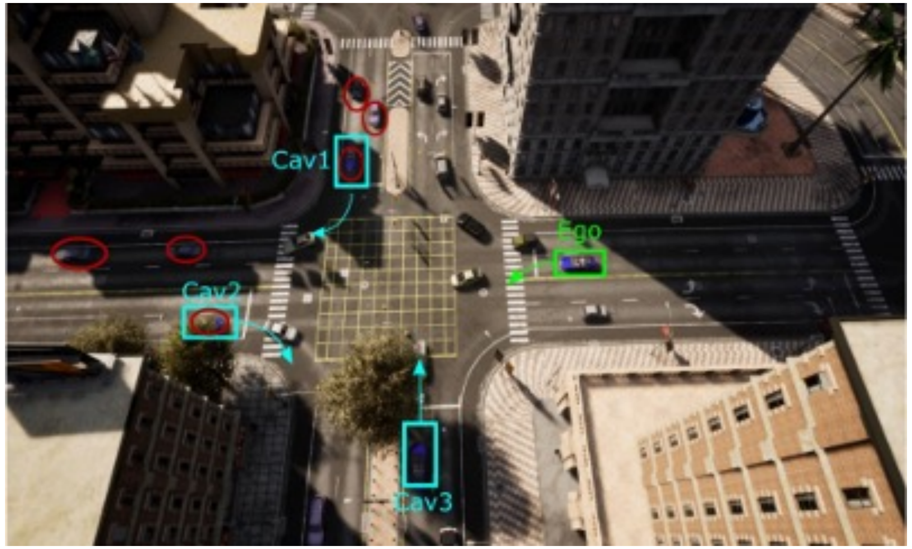
- For Single perception: introducing LiDAR
  - Spatial misalignment effect
- For collaborative perception
  - One possible solution: introducing LiDAR too.
  - Why not utilize the unique attributes of multi-view.



LiDAR camera mechanism based on confidence complementarity



Spatial misalignment effect



The complementary characteristics of multiple perspectives in the collaborative process

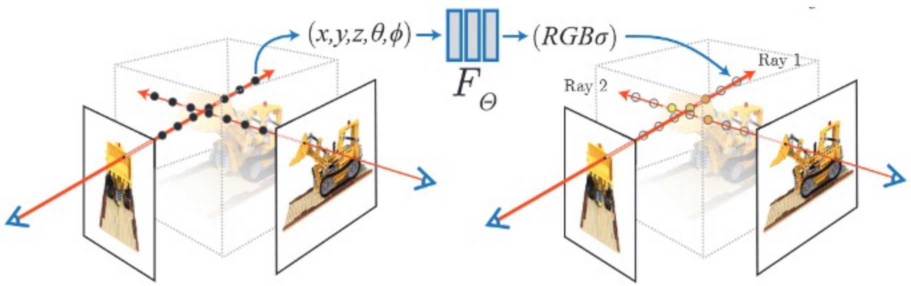
❑ Proposed: RCDN

Input:

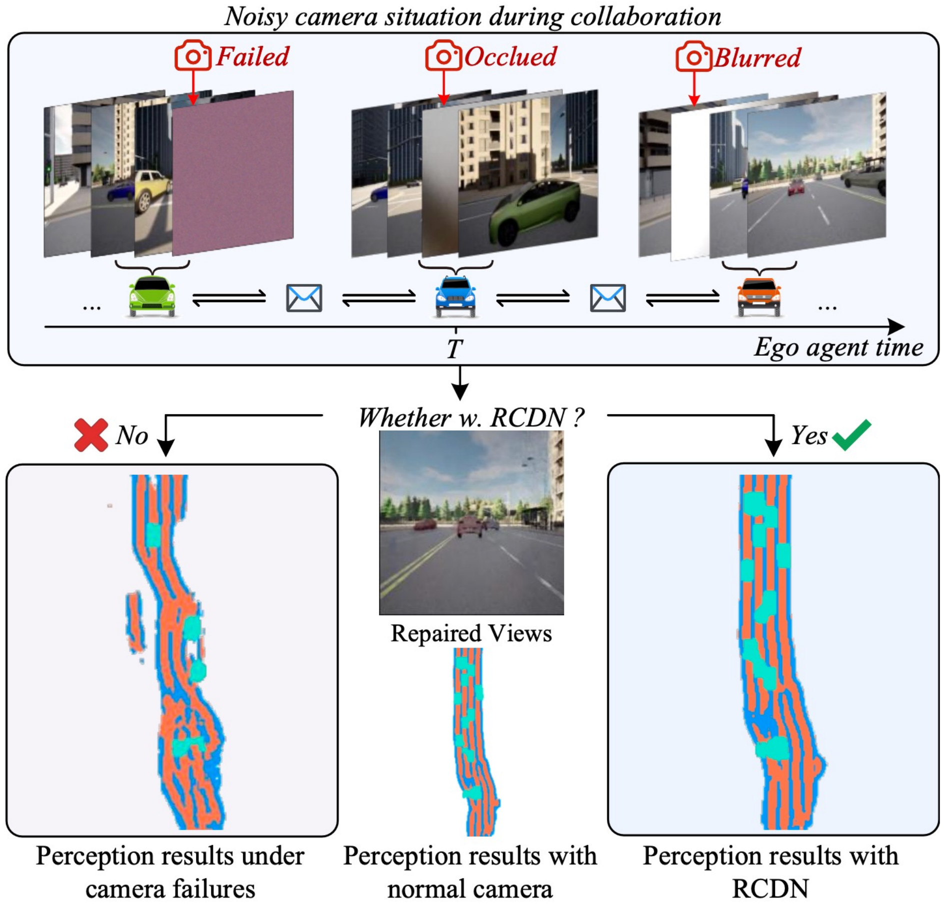
- Raw camera data sequence C = $\{C_0, C_1, \ldots, C_M\}$ including unpredictable noise signal.
- Sensor poses $P = \{P_0, P_1, \ldots, P_M\}(P_0 \in SE(3))$
- Timestamps $T = \{t_0, t_1, \ldots, t_{n-1}\}(t_i \in \mathbb{R})$

Output:

- Repaired camera data sequence $\hat{C} = \{\hat{C}_0, \hat{C}_1, \ldots, \hat{C}_M\}$

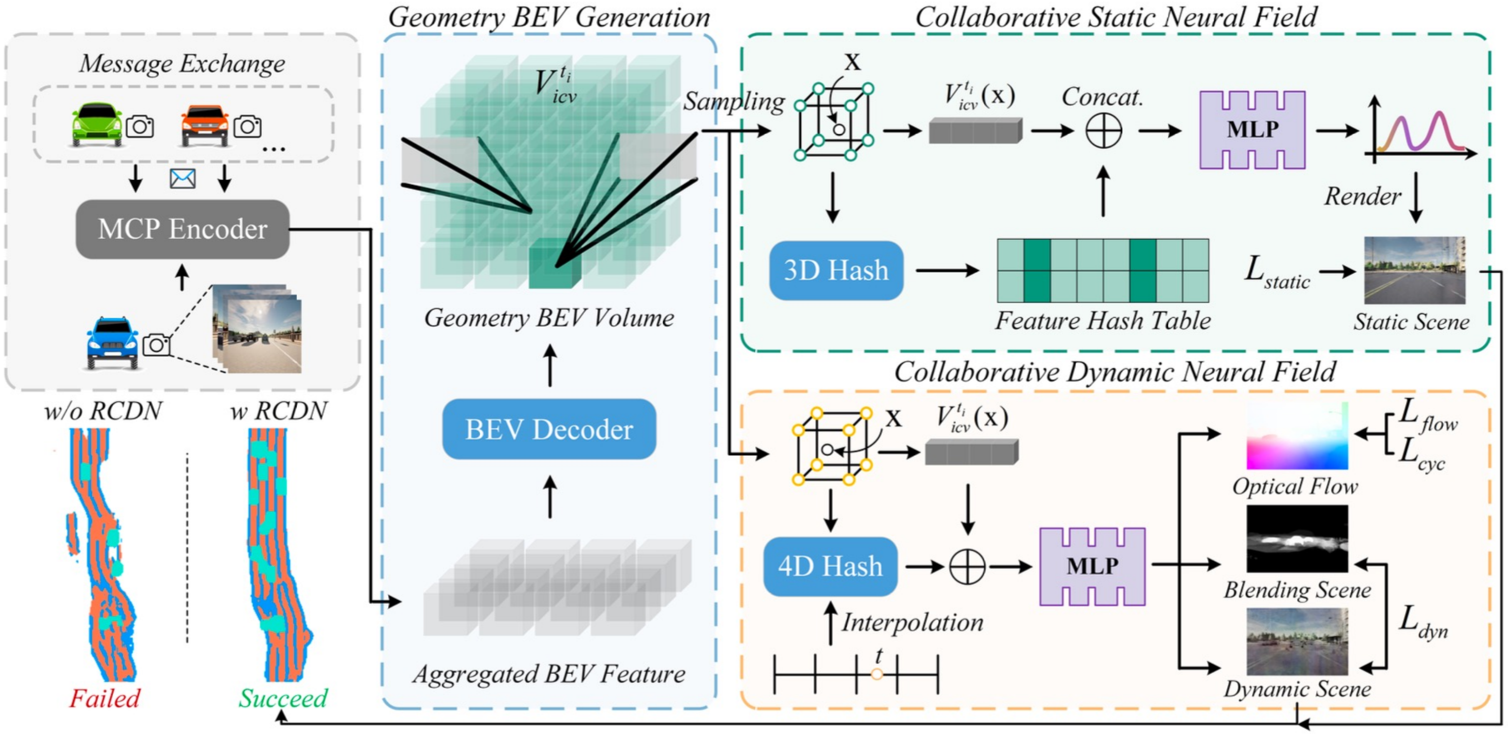


Multi view complementary characteristics in NeRF



❑ RCDN

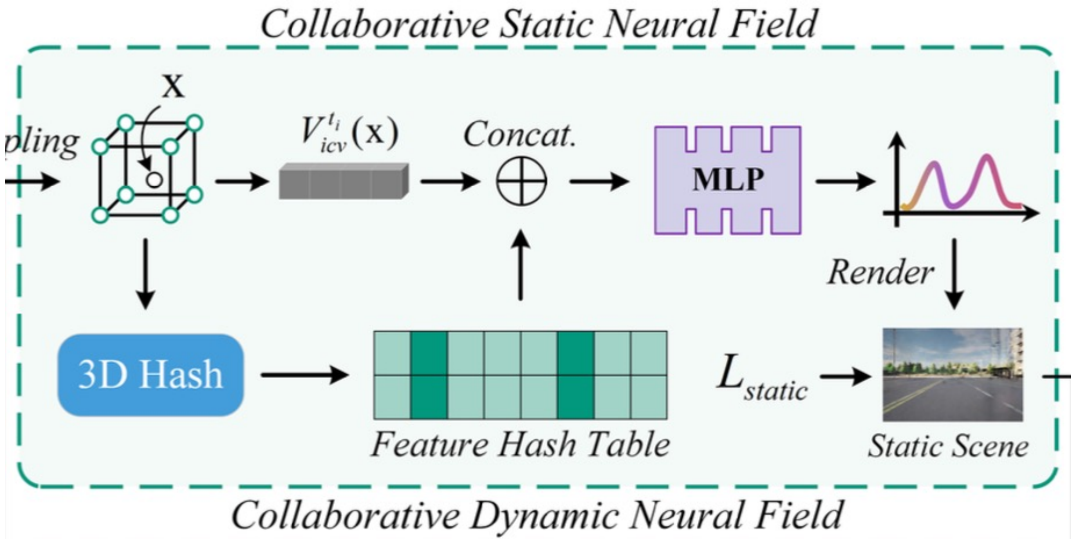
- ➢ Expand the 2D bird's-eye view features, establish spatial sampling based on 3D geometric bird's-eye view features, and optimize scene representation
- ➢ Propose a dynamic static decoupling neural field and design a hash grid rendering module based on generalizable features to improve reconstruction quality
- ➢ Annotate the robust dataset of collaborative cameras, manually label and design different camera fault scenarios, and assist in the research of collaborative camera robustness

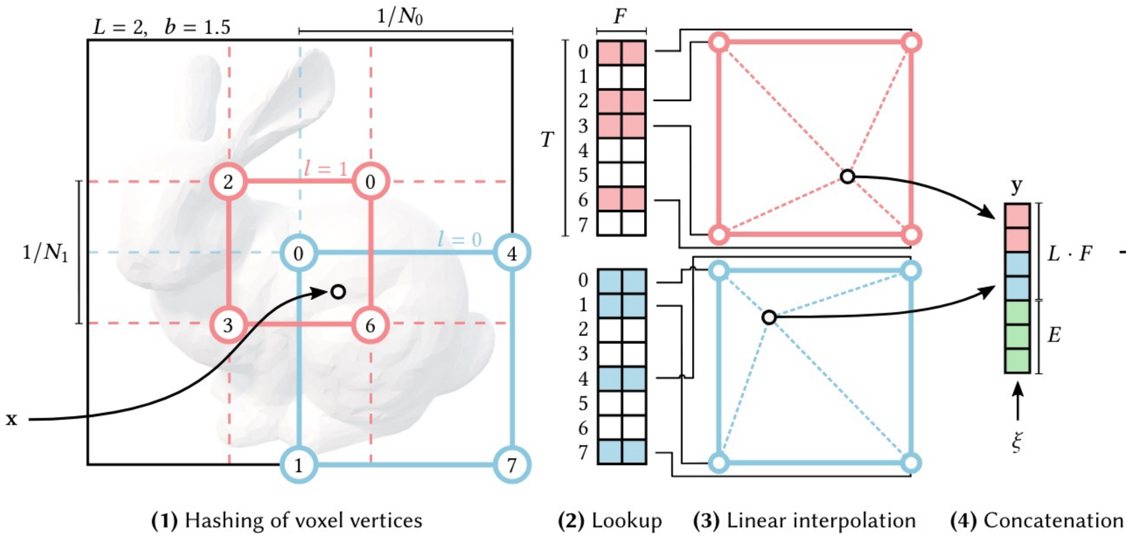


RCDN

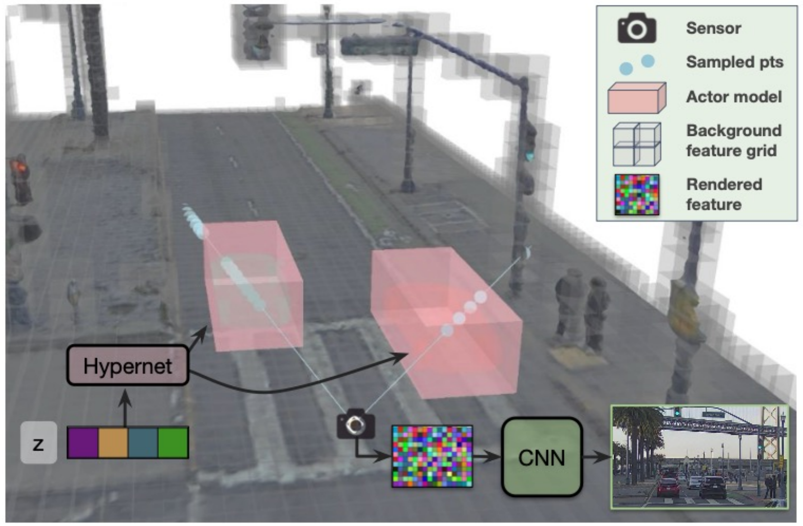
Collaborative Static Neural Field

- Geometry BEV Features & Hash Grids
- Static & Dynamic Decomposition
- Objected-based Modeling





Hash grid representations

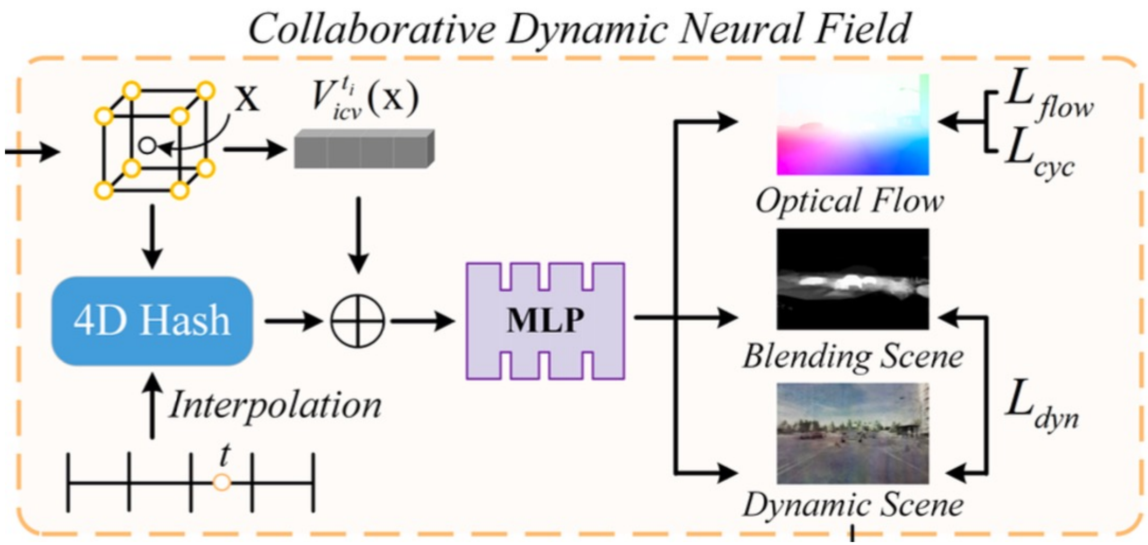


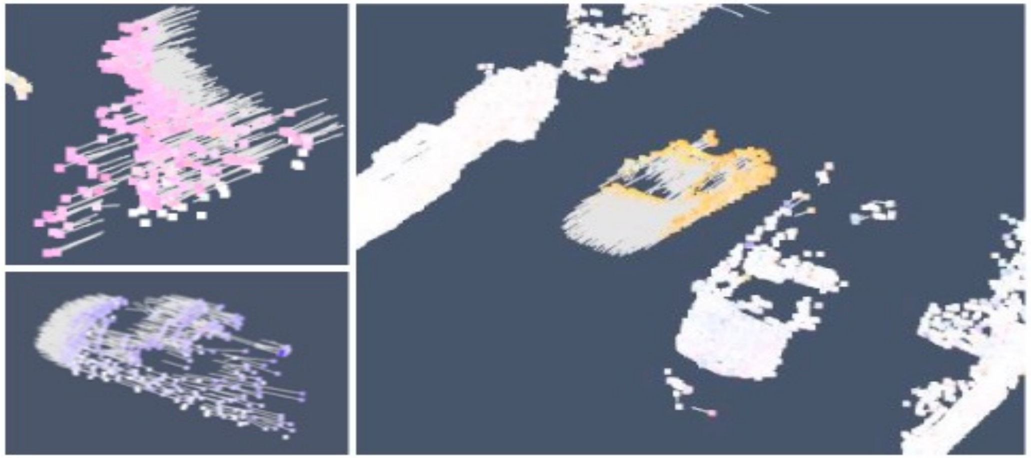
Object-based hash grid

❑ RCDN

Collaborative Dynamic Neural Field

- Flow MLP
- Object-based Movement Consistency
- 4D (+ Temporal Interpolation) Hash Grid



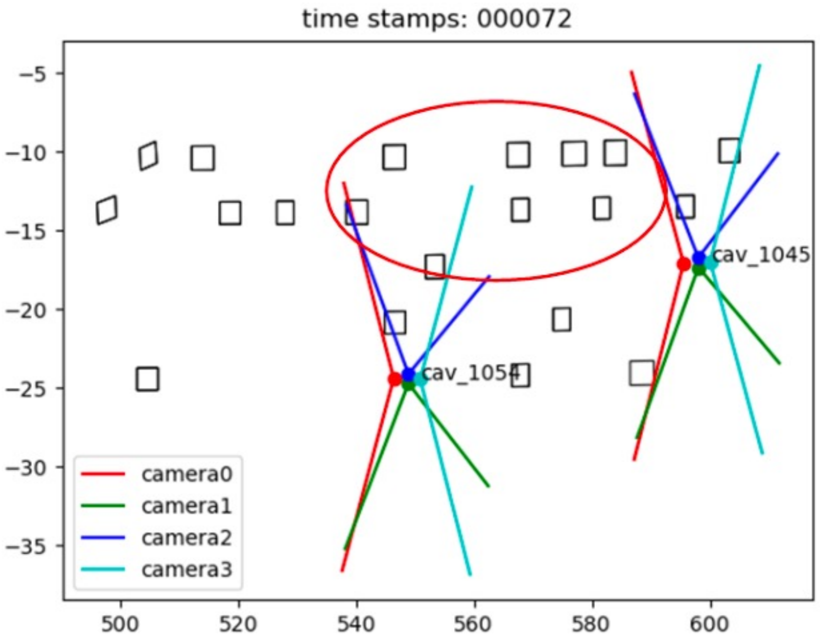


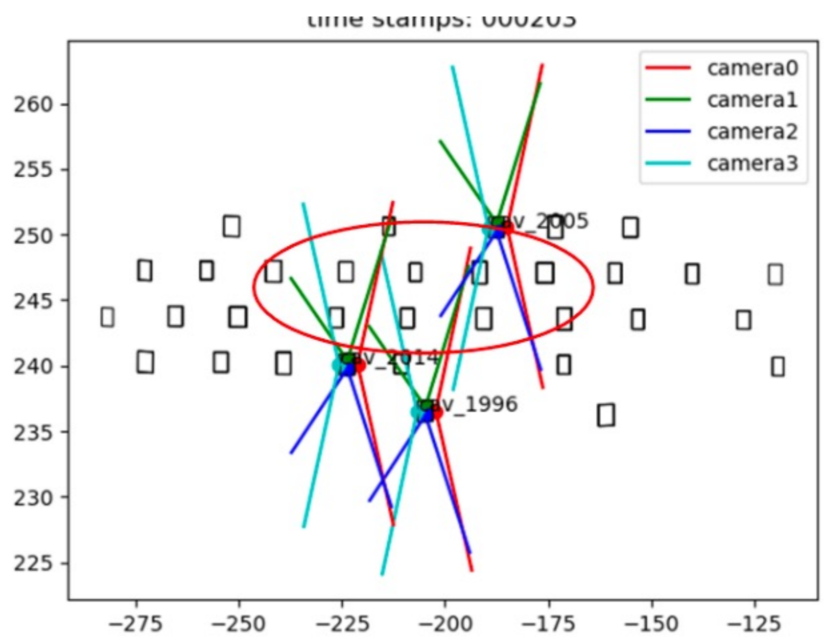
Object Movement

Effeteness of 4D Hash Grid

❑ OPV2V-N for RCDN

➢ Data Recording logic



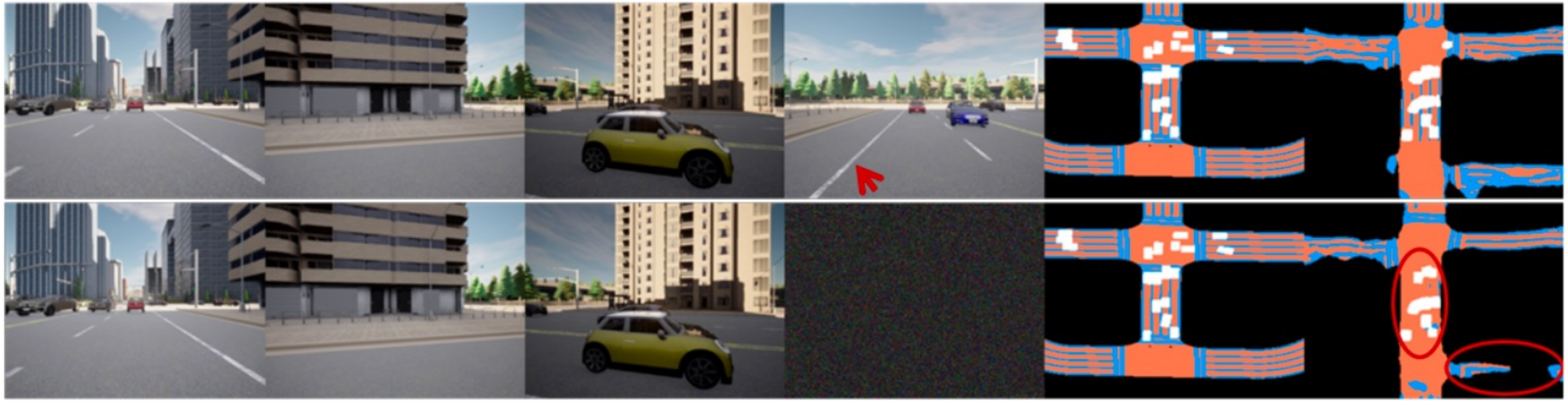


➢ Data Modal

(a) camera input (b) forward flow (c) backward flow (d) masks

❑ OPV2V-N for RCDN

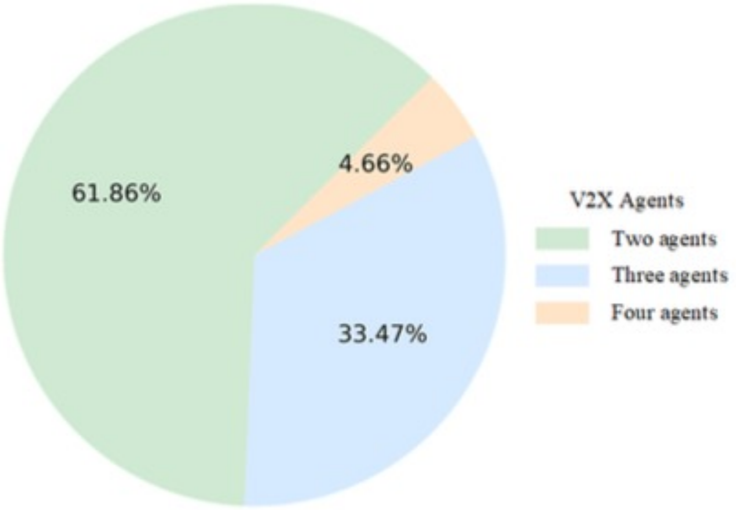
➢ Qualitative Evaluation



➢ Quantitative Evaluation

OPV2V-N	w. random noisy	w.o. random noisy
Dr. Area	49.37	52.64
Lanes	34.80	37.96
Dynamic Veh.	39.81	47.49

Validation on whether random noise will affect collaborative perception system

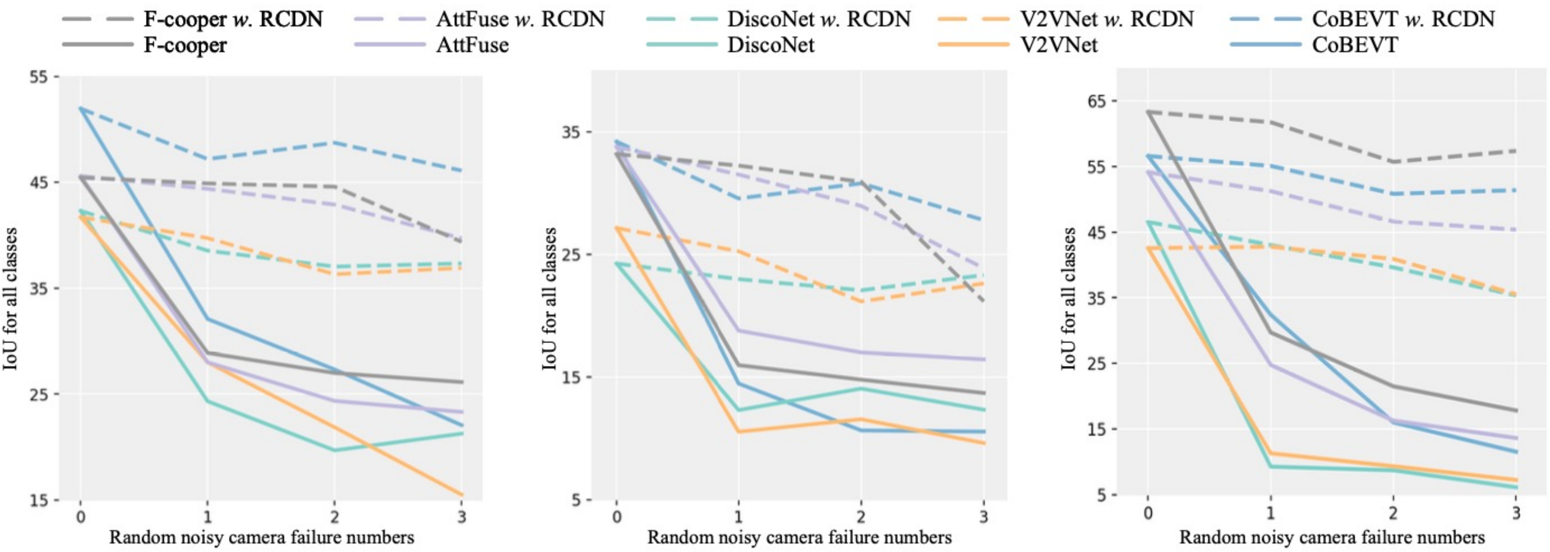


Distributions of V2X collaborative agents

❑ Experiments

❑ Experiments

Model / Metric	Static Part (*Perf. Comparison*)				Dynamic Part Vehicle	
	Drivable Area		Lane			
	Normal	Failure *w.o/w.* RCDN	Normal	Failure *w.o/w.* RCDN	Normal	Failure *w.o/w.* RCDN
F-Cooper[1]	45.44	28.87/44.89(↑55.49%)	33.17	15.95/32.23(↑102.07%)	63.33	29.70/61.76(↑107.95%)
AttFuse[16]	45.59	27.99/44.38(↑58.56%)	33.76	18.77/31.50(↑67.82%)	54.14	24.76/52.15(↑110.62%)
DiscoNet[42]	42.30	24.31/38.54(↑58.54%)	24.24	12.29/22.97(↑86.90%)	46.56	9.25/43.03(↑365.19%)
V2VNet[35]	41.70	27.99/39.72(↑41.91%)	27.14	10.52/25.24(↑139.92%)	42.57	11.28/42.76(↑279.08%)
CoBEVT[6]	51.96	32.08/47.19(↑47.10%)	34.19	14.45/29.55(↑104.50%)	56.61	32.41/55.10(↑70.01%)



(a) Performance on static part Dr. Area (b) Performance on static part Lane (c) Performance on dynamic part Veh.

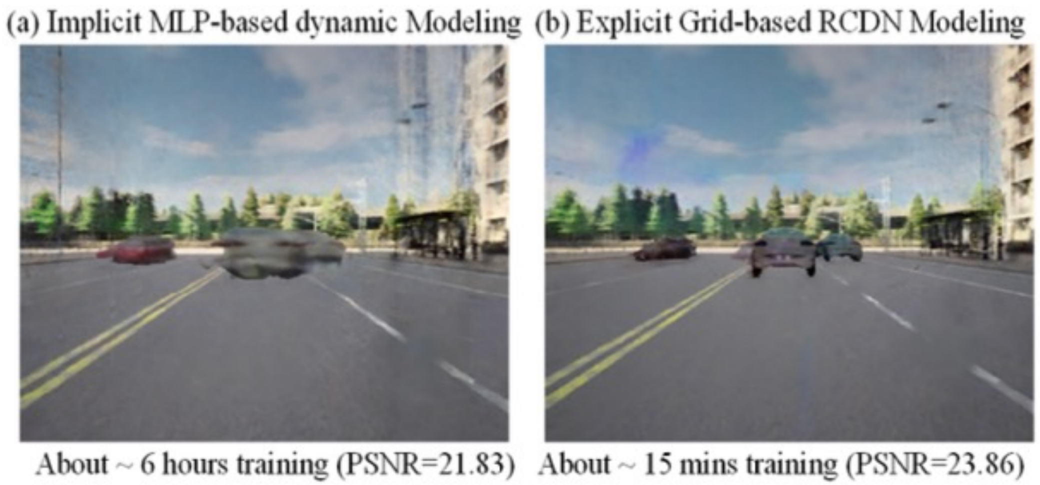
- ✓ The proposed RCDN can stabilize the performance of all benchmark methods in both static and dynamic parts of map view segmentation under all camera fault settings.

❑ Experiments

➢ Ablation

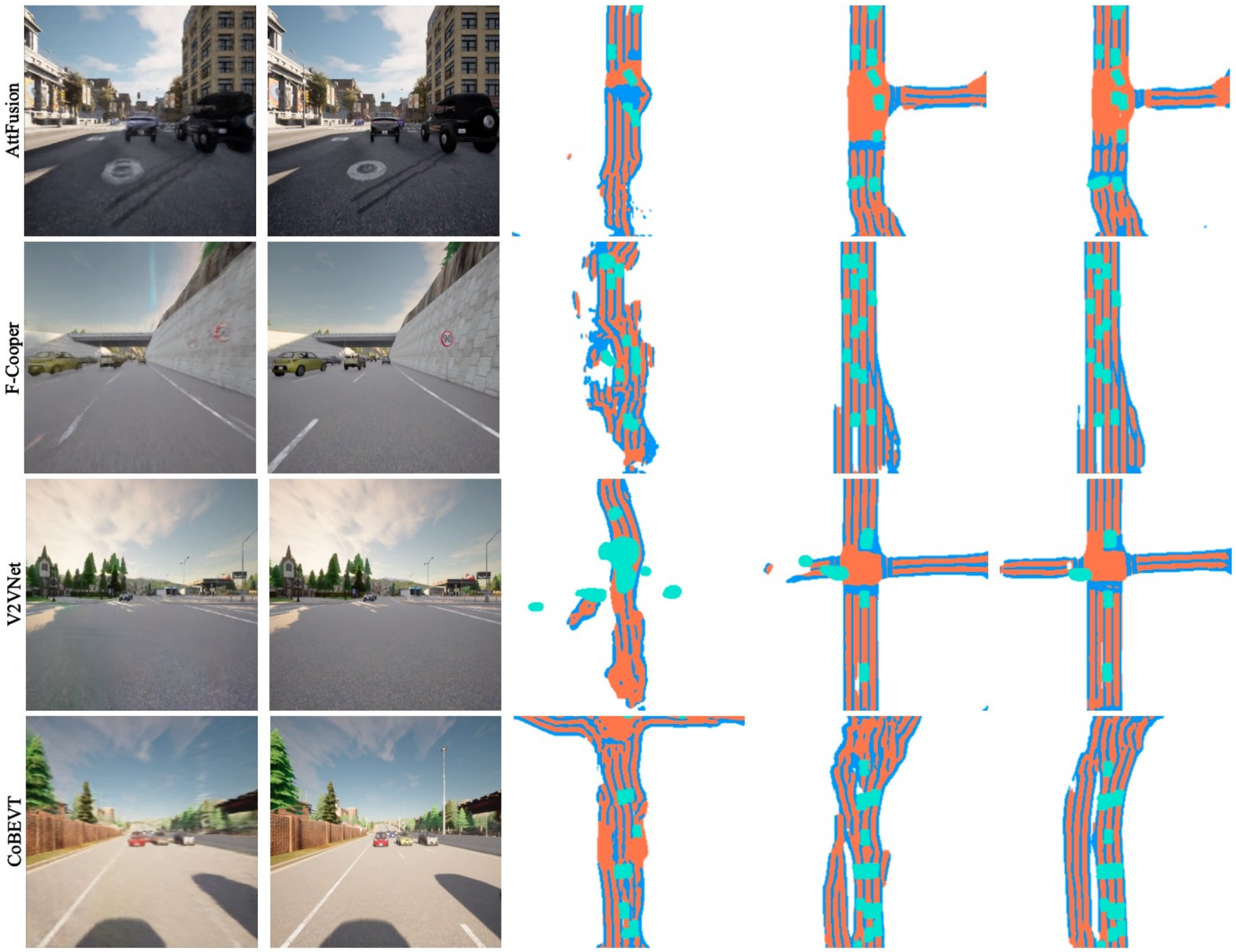
Modules		Dr. Area	Lanes	Dynamic Veh.
Neural Field	Time Model			
✗	✗	24.55	10.07	30.67
✔	✗	24.47	11.71	41.55
✔	✔	27.37	10.63	46.65

✓ The effectiveness of dynamic and static decoupling

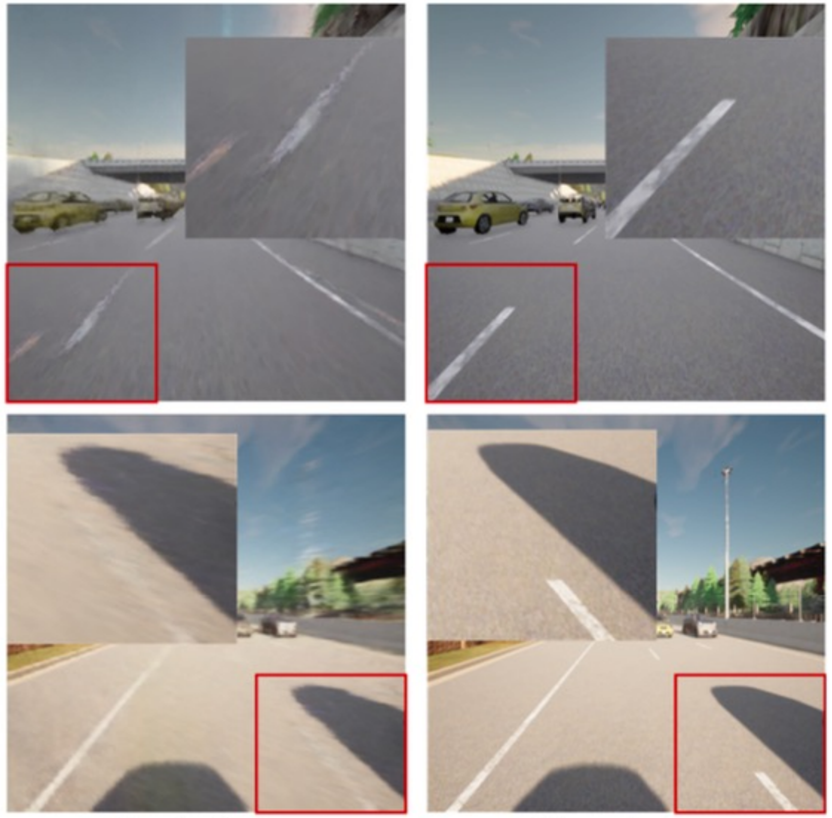


✓ Hash grid rendering module with generalizable features

❑ Experiments



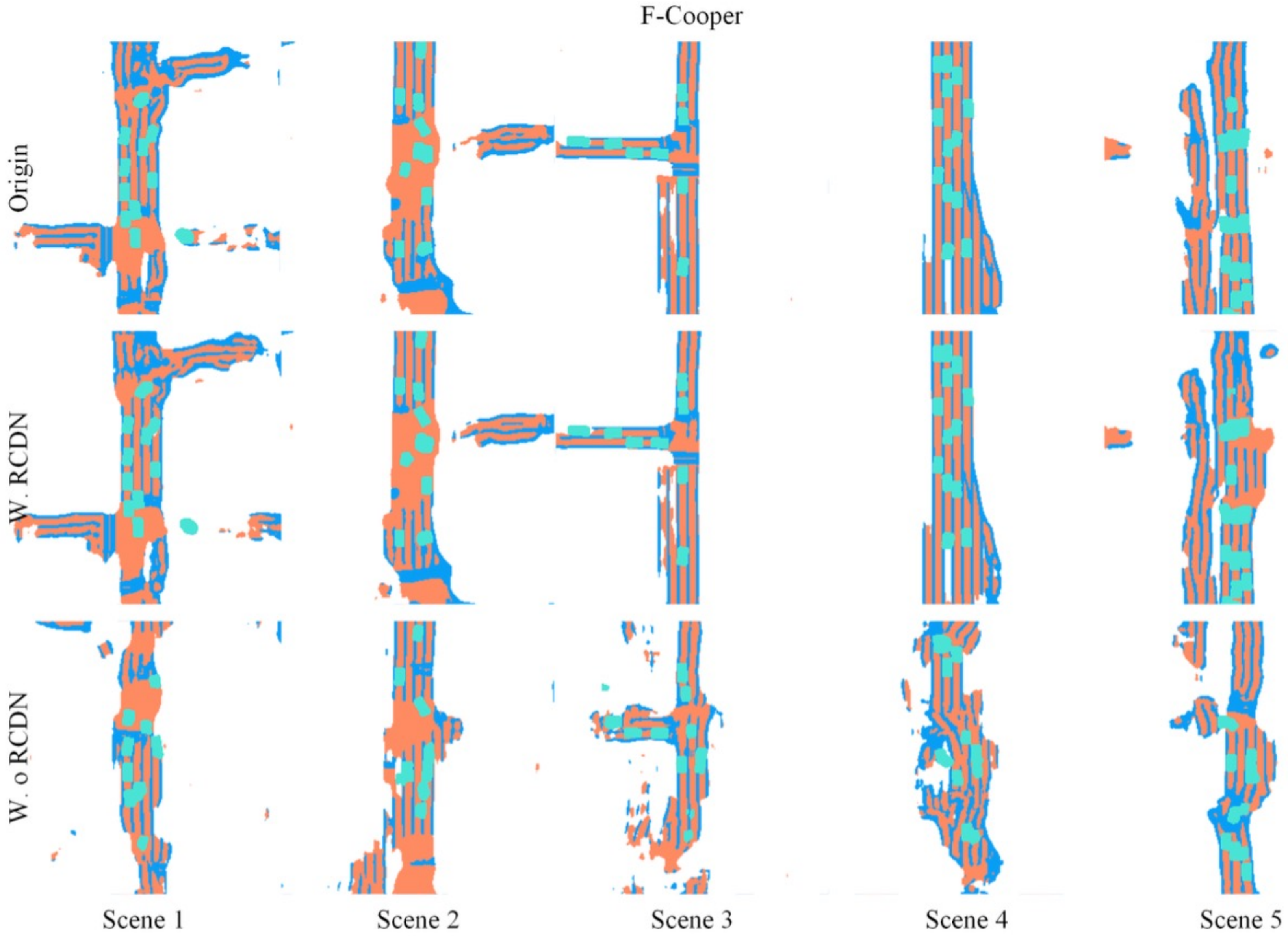
(a) Repaired Views (b) Origin Views (c) *w.o.* RCDN (d) *w.* RCDN (e) Origin segmentation map



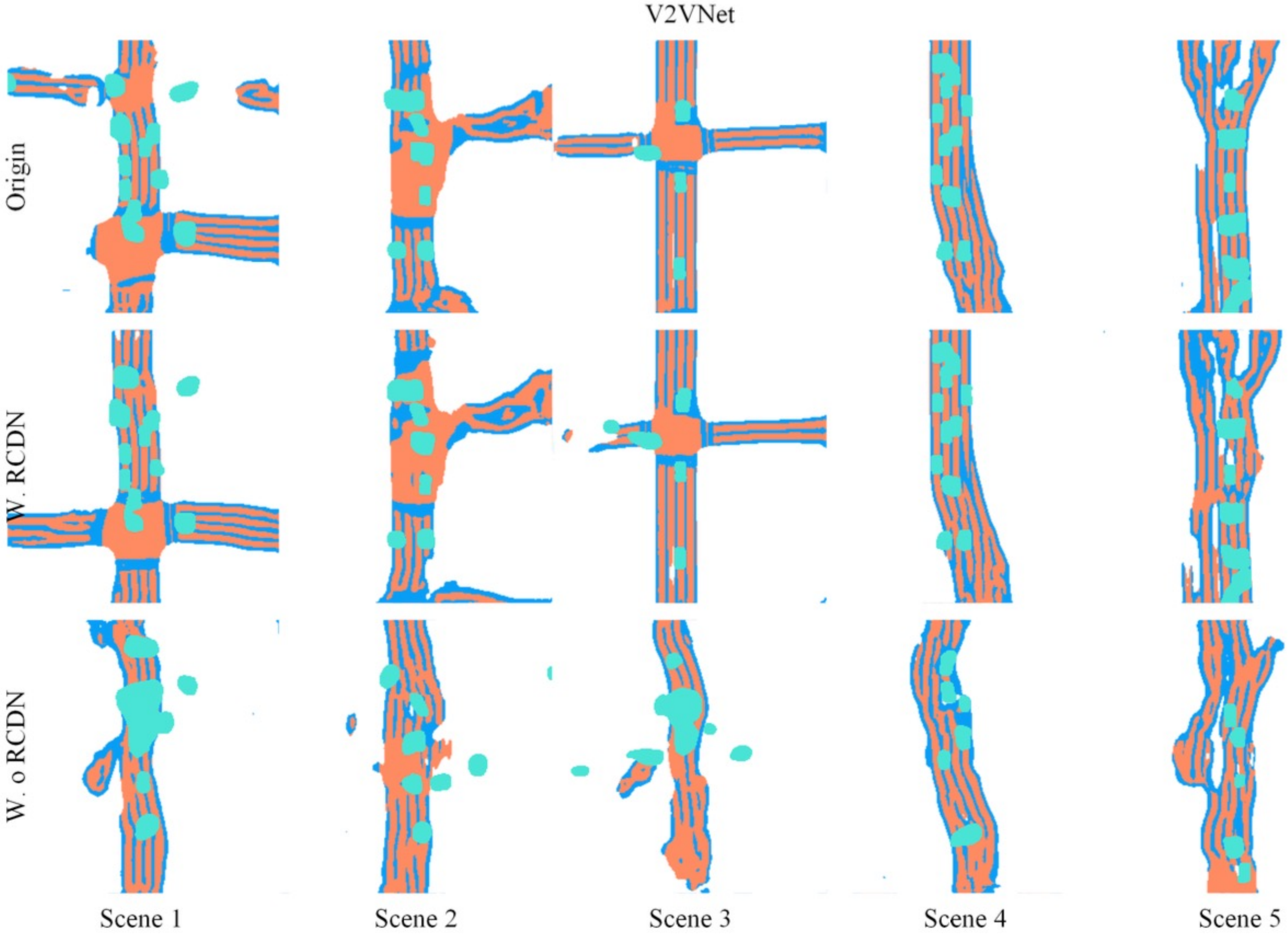
Repaired view Normal view

Modules	Time Cost
f_{static}	4.47±0.11ms
$f_{dynamic}$	3.94±0.21ms
f_{render}	20.98±0.22ms

❑ Visualization of baselines: F-Cooper w/w.o RCDN



❑ Visualization of baselines: V2VNet w/w.o RCDN



Thank you for listening!