NEURAL INFORMATION PROCESSING SYSTEMS

DiffE2E: Rethinking End-to-End Driving with a Hybrid Diffusion-Regression-Classification Policy

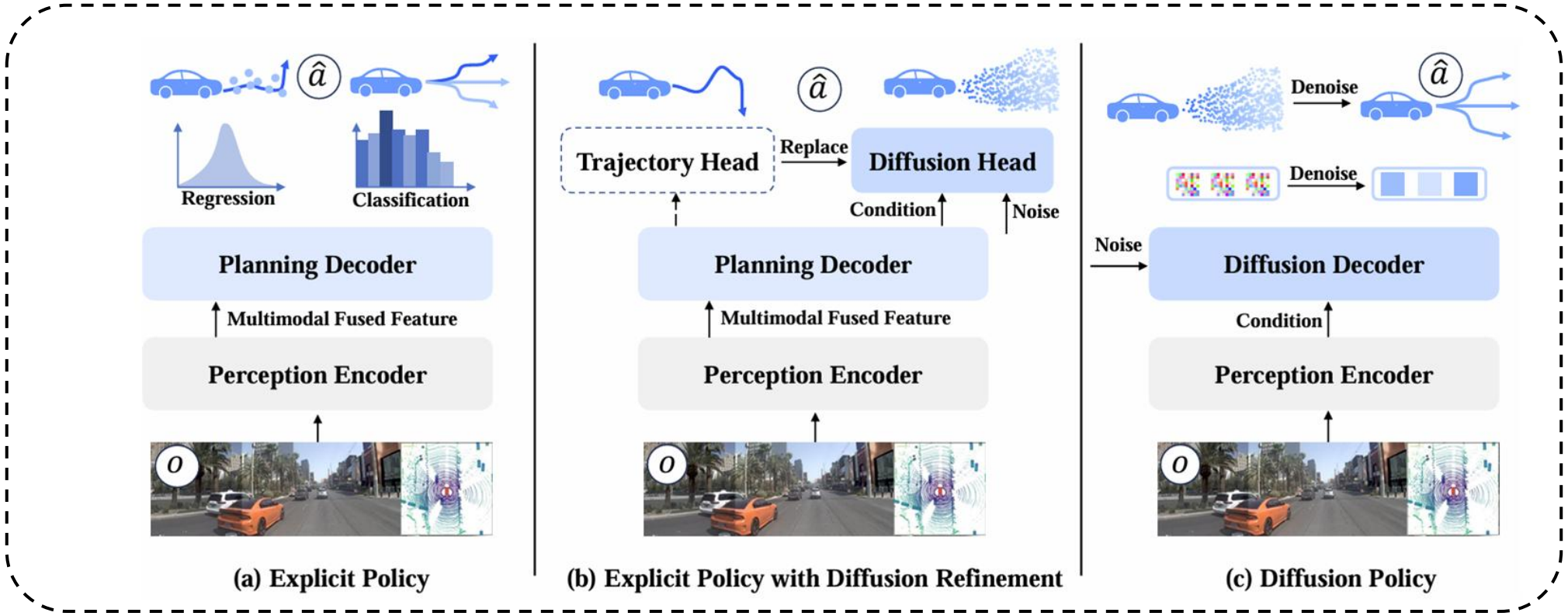
Rui Zhao, Yuze Fan, Ziguo Chen, Fei Gao*, Zhenhai Gao

College of Automotive Engineering, Jilin University

National Key Laboratory of Automotive Chassis Integration and Bionics, Jilin University

Introduction

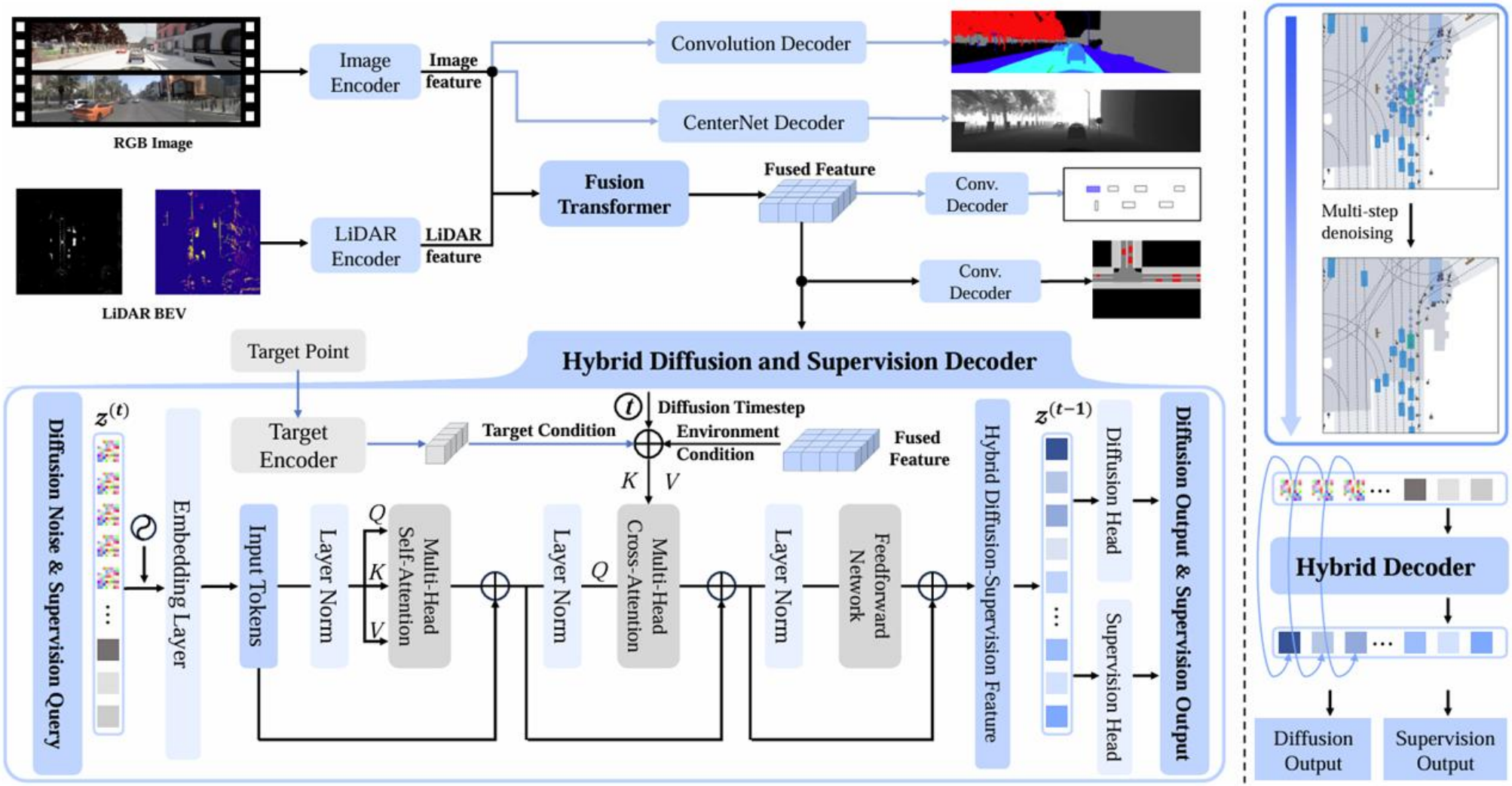
◆ Comparison of different end-to-end driving training paradigms



(a) Explicit Policy. Directly predicts trajectories through regression after sensor input processing. (b) Explicit Policy with Diffusion Refinement. Uses diffusion models to replace traditional explicit policy trajectory output heads. (c) Diffusion Policy. Uses diffusion models to directly generate trajectories based on perception encoder features.

Framework

◆ DiffE2E Framework



Experiment Demonstration

◆ Experiment Demonstration

Comparison on the CARLA Longest6 Benchmark

Method	Traj. Decoder	Img. Encoder	Input	DS ↑	RC ↑	IS ↑
Expert	-	-	-	81 ± 3	90 ± 1	0.91 ± 0.04
WOR [4]	-	ResNet-34 [17]	C	21 ± 3	48 ± 4	0.56 ± 0.03
LAV v1 [2]	Explicit Policy	ResNet-18 [17]	C&L	33 ± 1	70 ± 3	0.51 ± 0.02
InterFuser [42]	Explicit Policy	ResNet-50 [17]	C&L	47 ± 6	74 ± 1	0.63 ± 0.07
TransFuser [9]	Explicit Policy	RegNetY-3.2GF [39]	C&L	47 ± 6	93 ± 1	0.50 ± 0.06
TCP [53]	Explicit Policy	ResNet-34 [17]	C	48 ± 3	72 ± 3	0.65 ± 0.04
ThinkTwice [26]	Explicit Policy	ResNet-50 [17]	C&L	51 ± 4	64 ± 4	0.80 ± 0.03
LAV v2 [2]	Explicit Policy	ResNet-18 [17]	C&L	58 ± 1	83 ± 1	0.68 ± 0.02
Perception PlanT [40]	Explicit Policy	RegNetY-3.2GF [39]	C&L	58 ± 5	88 ± 1	0.65 ± 0.06
DriveAdapter [25]	Explicit Policy	ResNet-50 [17]	C&L	58 ± 2	73 ± 3	0.79 ± 0.04
TF++ [23]	Explicit Policy	RegNetY-3.2GF [39]	C&L	69 ± 0	94 ± 2	0.72 ± 0.01
DiffE2E (Ours)	Diffusion Policy	RegNetY-3.2GF [39]	C&L	**79 ± 4**	94 ± 2	**0.84 ± 0.02**

Comparison on the Navtest Benchmark

Method	Traj. Decoder	Img. Enc.	Input	PDMS↑	NC↑	DAC↑	EP↑	TTC↑	C↑
Human	-	-	-	94.8	100	100	87.5	100	99.9
AD-MLP [58]	Explicit Policy	-	S	65.6	93.0	77.3	62.8	83.6	**100**
VADv2 [6]	Explicit Policy	ResNet-34 [17]	C&L	80.9	97.2	89.1	76.0	91.6	**100**
UniAD [21]	Explicit Policy	ResNet-34 [17]	C	83.4	97.8	91.9	78.8	92.9	**100**
LTF [9]	Explicit Policy	ResNet-34 [17]	C	83.8	97.4	92.8	79	92.4	**100**
PARA-Drive [52]	Explicit Policy	ResNet-34 [17]	C	84.0	97.9	92.4	79.3	93	99.8
TransFuser [9]	Explicit Policy	ResNet-34 [17]	C&L	84.0	97.7	92.8	79.2	92.8	**100**
LAW [32]	Explicit Policy	ResNet-34 [17]	C	84.6	96.4	95.4	81.7	88.7	99.9
DRAMA [56]	Explicit Policy	ResNet-34 [17]	C&L	85.5	98	93.1	80.1	94.8	**100**
Hydra-MDP [33]	Explicit Policy	ResNet-34 [30]	C&L	86.5	98.3	96.0	78.7	94.6	**100**
DiffusionDrive [34]	Diffusion Policy*	ResNet-34 [17]	C&L	88.1	98.2	96.2	82.2	94.7	**100**
GoalFlow [54]	Diffusion Policy*	V2-99 [30]	C&L	90.3	98.4	98.3	85	94.6	**100**
Hydra-MDP° [33]	Explicit Policy	ViT-L [15] + V2-99 [30]	C&L	91.0	98.7	98.2	**86.5**	95.0	**100**
Hydra-MDP++ [31]	Explicit Policy	V2-99 [30]	C	91.0	98.6	**98.6**	85.7	95.1	**100**
DiffE2E†	Diffusion Policy	ResNet-34 [17]	C&L	89.8	99.2	96.8	83.6	96.7	**100**
DiffE2E‡	Diffusion Policy	V2-99 [30]	C	90.9	99.7	97.1	84.2	98.2	99.9
DiffE2E (Ours)	Diffusion Policy	V2-99 [30]	C&L	**92.7**	**99.9**	**98.6**	85.3	**99.3**	99.9

Experiment Demonstration

◆ Experiment Demonstration

Comparison on the CARLA Town05 Long Benchmark

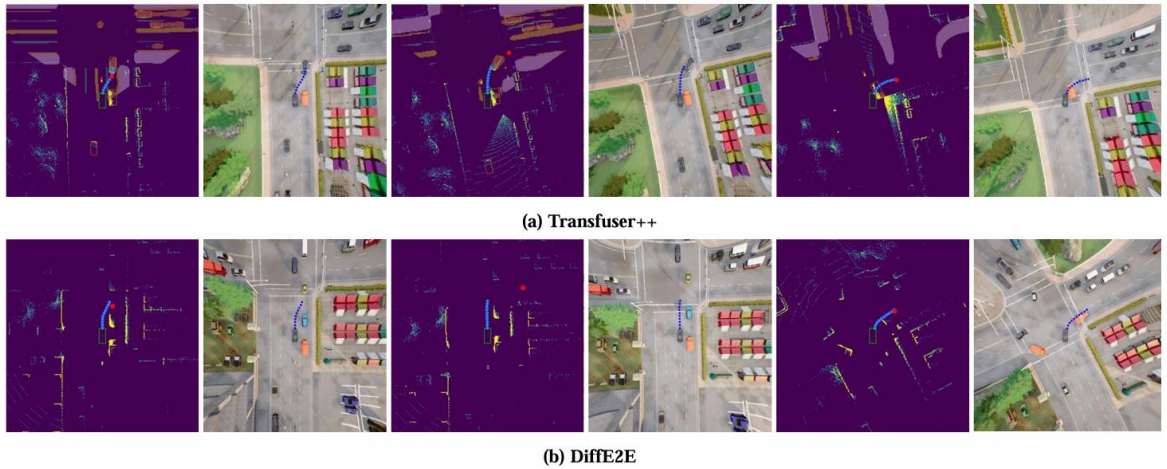
Method	Traj. Decoder	Img. Encoder	Input	DS ↑	RC ↑	IS ↑
CILRS [10]	Explicit Policy	ResNet-34 [17]	C	7.8	10.3	0.75
LBC [3]	Explicit Policy	ResNet-18,34 [17]	C	12.3	31.9	0.66
Roach [60]	Explicit Policy	ResNet-34 [17]	C	41.6	96.4	0.43
ST-P3 [20]	Explicit Policy	EfficientNet-B4 [49]	C	11.5	83.2	-
VAD [27]	Explicit Policy	ResNet-50 [17]	C	30.3	75.2	-
MILE [19]	Explicit Policy	ResNet-18 [17]	C	61.1	97.4	0.63
DriveMLM [51]	Explicit Policy	ViT-G [59]	C&L	76.1	98.1	0.78
VADv2 [6]	Explicit Policy	ResNet-50 [17]	C	85.1	98.4	0.87
DiffE2E (Ours)	Diffusion Policy	RegNetY-3.2GF [39]	C&L	**90.8**	**100**	**0.91**

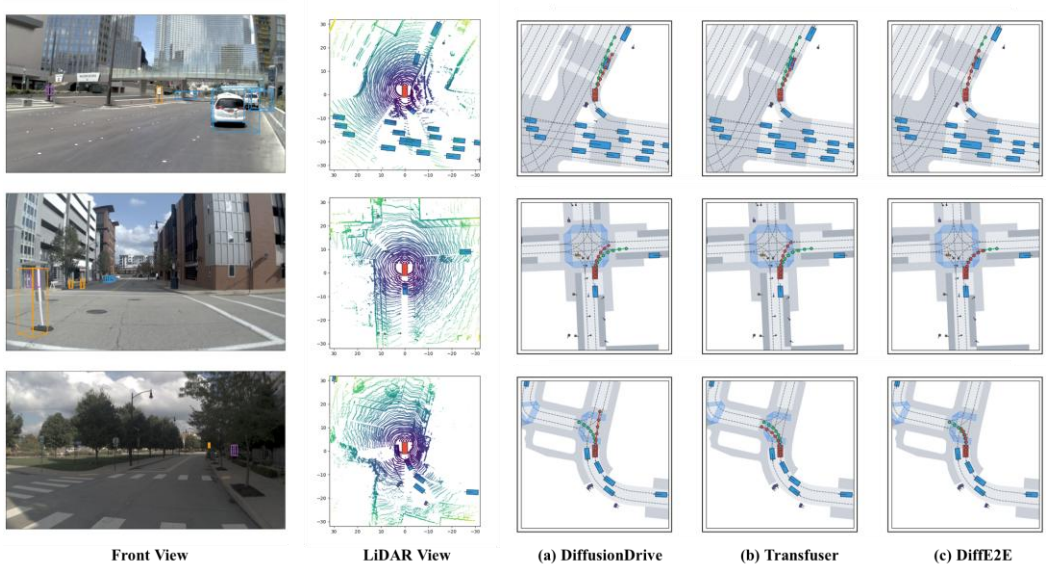
Comparison on the CARLA Town05 Short Benchmark

Method	Traj. Decoder	Img. Encoder	Input	DS ↑	RC ↑
CILRS [10]	Explicit Policy	ResNet-34 [17]	C	7.5	13.4
LBC [3]	Explicit Policy	ResNet-18,34 [17]	C	31.0	55.0
TransFuser [9]	Explicit Policy	RegNetY-3.2GF [39]	C&L	54.5	78.4
ST-P3 [20]	Explicit Policy	EfficientNet-B4 [49]	C	55.1	86.7
NEAT [8]	Explicit Policy	ResNet-34 [17]	C	58.7	77.3
Roach [60]	Explicit Policy	ResNet-34 [17]	C	65.3	88.2
WOR [4]	-	ResNet-18,34 [17]	C	64.8	87.5
VAD [27]	Explicit Policy	ResNet-50 [17]	C	64.3	87.3
VADv2 [6]	Explicit Policy	ResNet-50 [17]	C	89.7	93.0
InterFuser [42]	Explicit Policy	ResNet-50 [17]	C&L	95.0	95.2
DiffE2E (Ours)	Diffusion Policy	RegNetY-3.2GF [39]	C&L	**95.2**	**99.7**

Experiment Demonstration

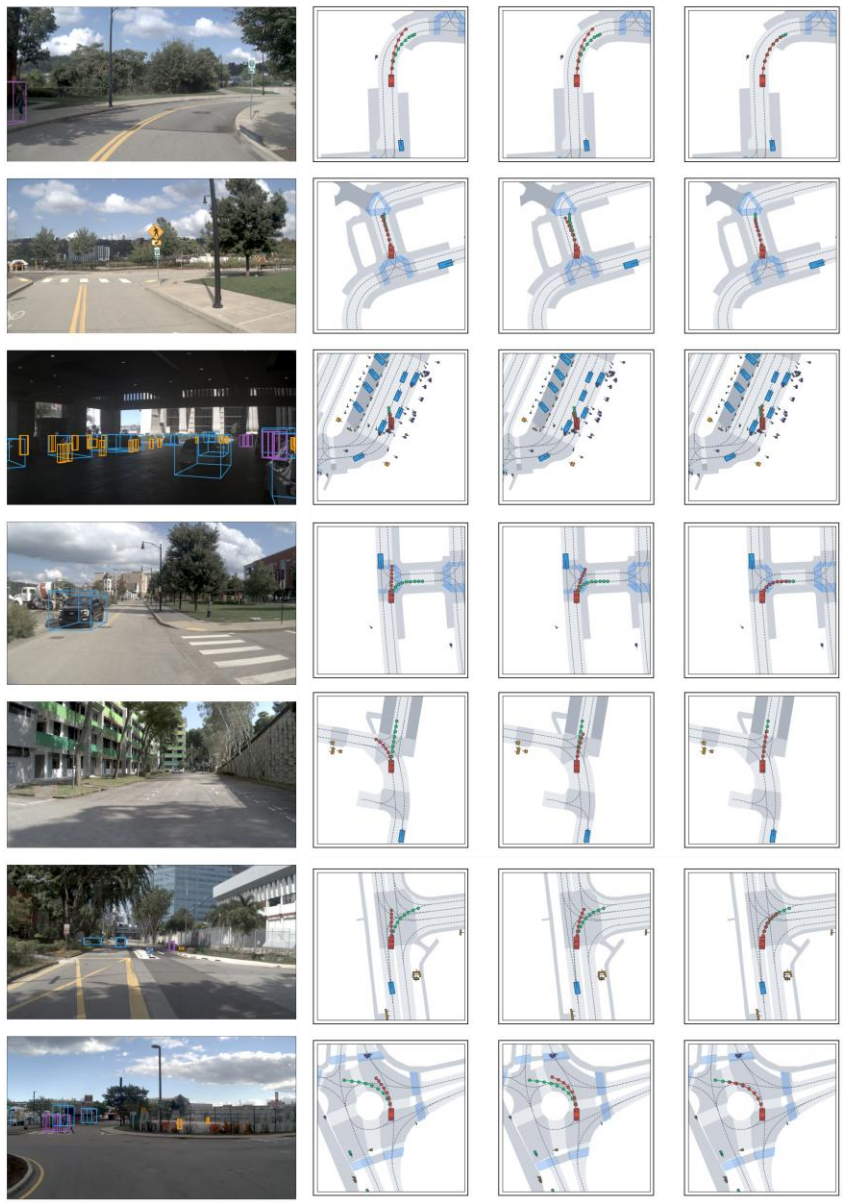
◆ Visualization





Left: Visualization in CARLA Simulator. In both the LiDAR and scene visualizations, blue points represent the predicted trajectory, while red points in the LiDAR view denote the target waypoints.
Right: Visualization in Navtest benchmark. Red trajectories denote the predicted paths of each method, while green trajectory corresponds to the ground truth.

Visualization



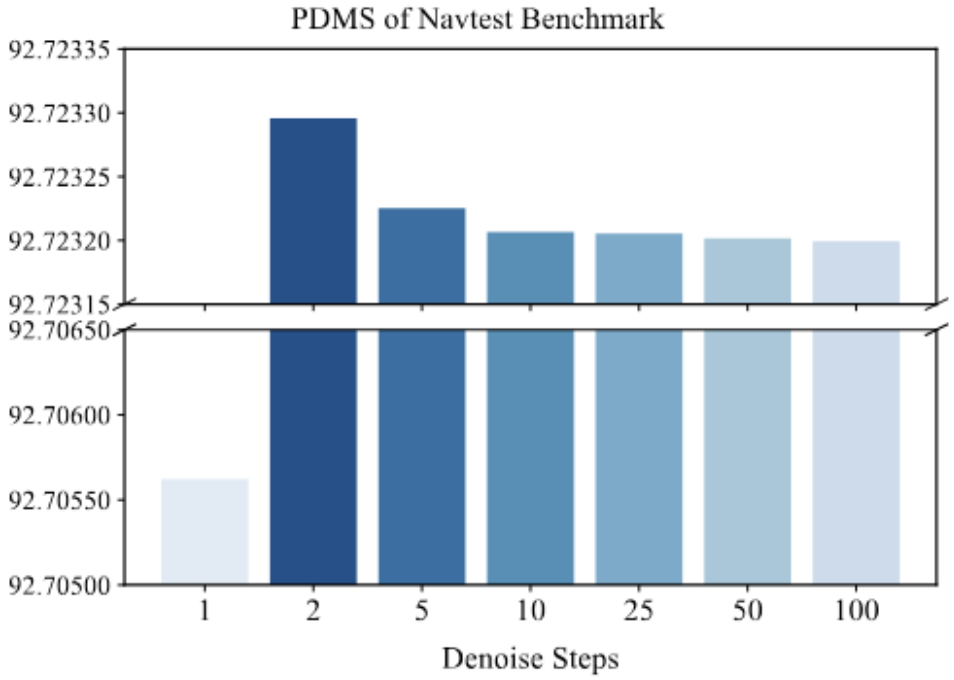
Front View (a) DiffusionDrive (b) Transfuser (c) DiffE2E

Front View (a) DiffusionDrive (b) Transfuser (c) DiffE2E

Experiment Demonstration

◆ Ablation Study

Type	Method	DS	RC	IS
Base	**DiffE2E**	**82.9**	**96.2**	**0.86**
Input	w/o ego state	68.9	88.8	0.81
	w/o command	69.6	93.8	0.77
Component	w/o GRU	66.8	75.5	0.88
Training Paradigm	Full Diffusion	70.1	91.1	0.76
	Full Discrimination	70.3	83.5	0.86
	One-Stage Training	18.2	21.8	0.79
	Two-Stage Training	**82.9**	**96.2**	**0.86**
Output	Noise Prediction	20.1	27.2	0.72
	Trajectory Prediction	**82.9**	**96.2**	**0.86**



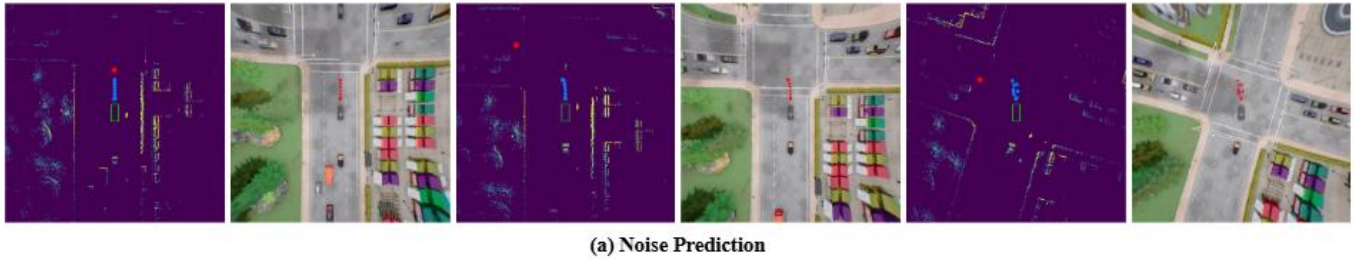
Component & Training Ablation: Removing ego state, navigation command, or GRU sharply reduces performance, confirming their necessity. The hybrid decoder and two-stage training are essential—single-stage training collapses and leads to unstable driving. Direct trajectory prediction works best.

Denoising Step Ablation: Two denoising steps give the best PDMS, while fewer or more degrade performance, showing that our hybrid modeling allows high-quality trajectories with minimal diffusion iterations.

Experiment Demonstration

◆ Ablation Study

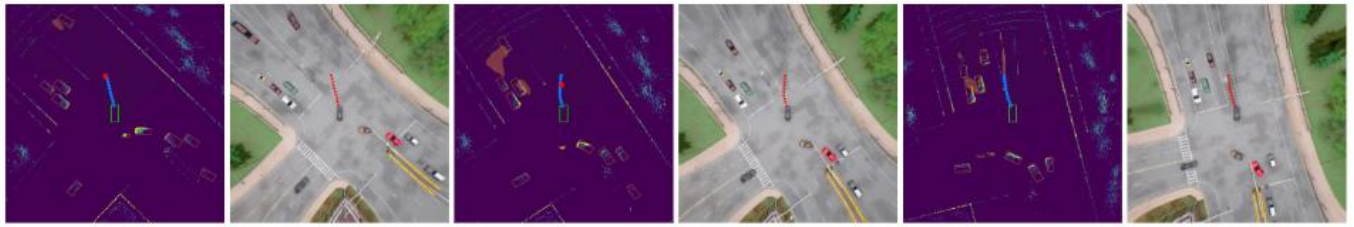
Ablation Study of Output Type on the CARLA Longest6 Benchmark



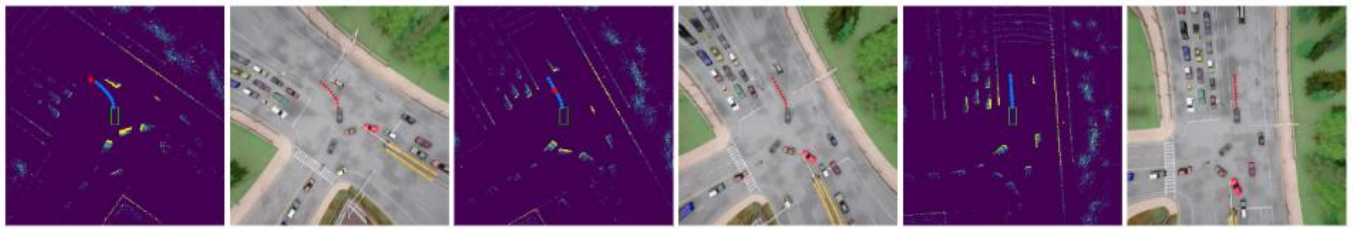
(a) Noise Prediction

(b) Trajectory Prediction

Ablation Study of Training Paradigms on the CARLA Longest6 Benchmark



(a) One-Stage Training



(b) Two-Stage Training