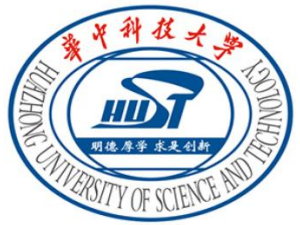
Rethinking Driving World Model as Synthetic Data Generator for Perception Tasks

Kai Zeng, Zhanqian Wu, Kaixin Xiong, Xiaobao Wei, Xiangyu Guo,
Zhenxin Zhu, Kalok Ho, Lijun Zhou, Bohan Zeng, Ming Lu,
Haiyang Sun, Bing Wang, Guang Chen, Hangjun Ye, Wentao Zhang

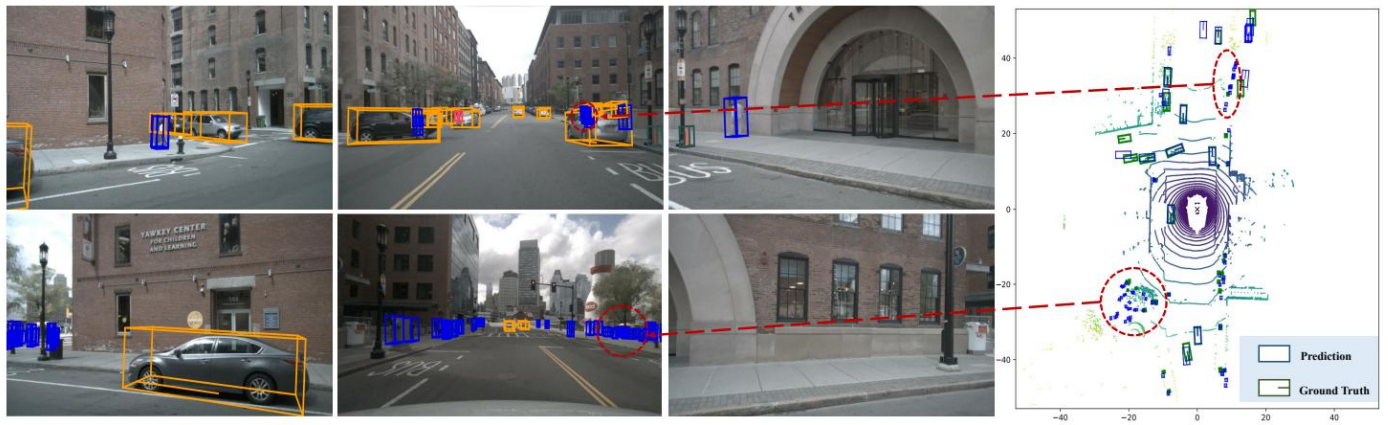
The Fourteenth International Conference on Learning Representations
ICLR 2026



Rethinking Driving World Model as Synthetic Data Generator for Perception Tasks

Background

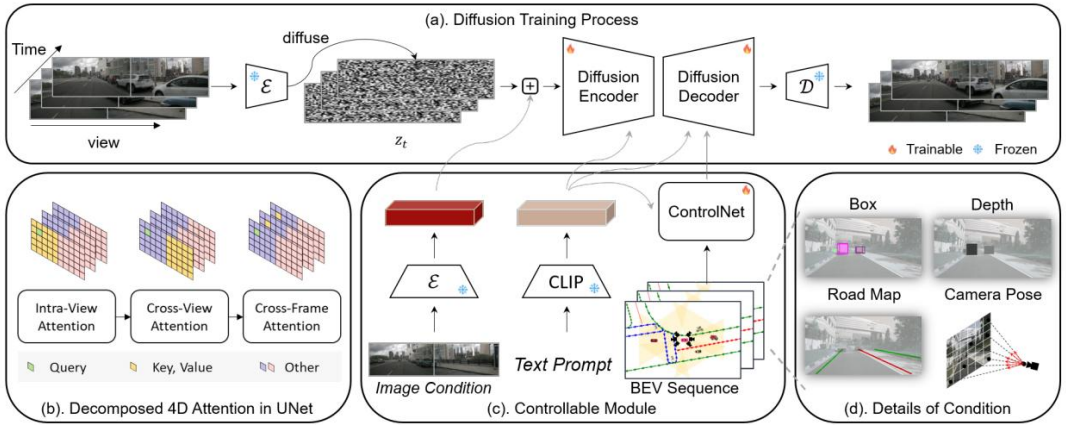
Synthetic Data



Downstream Tasks

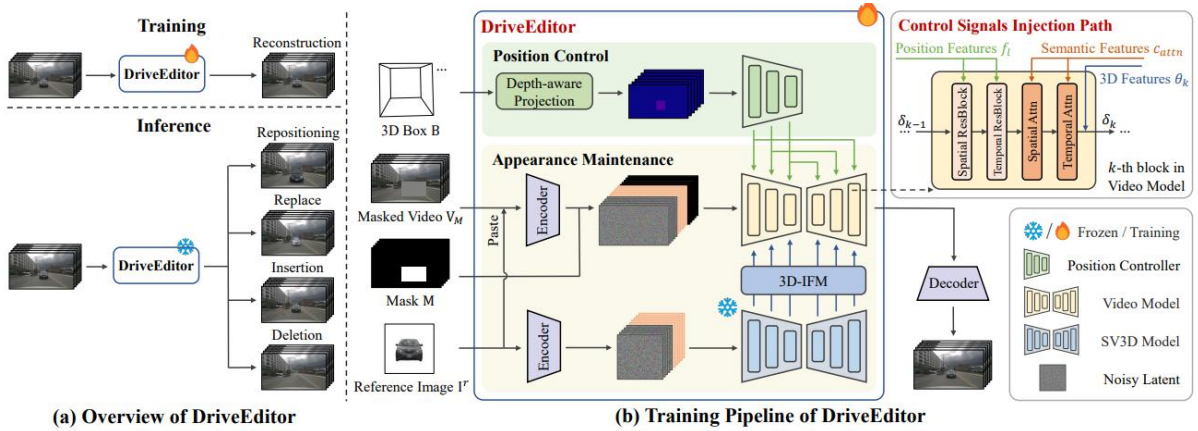
Motivation & Previous methods

(a) Full-Scene Video Generation



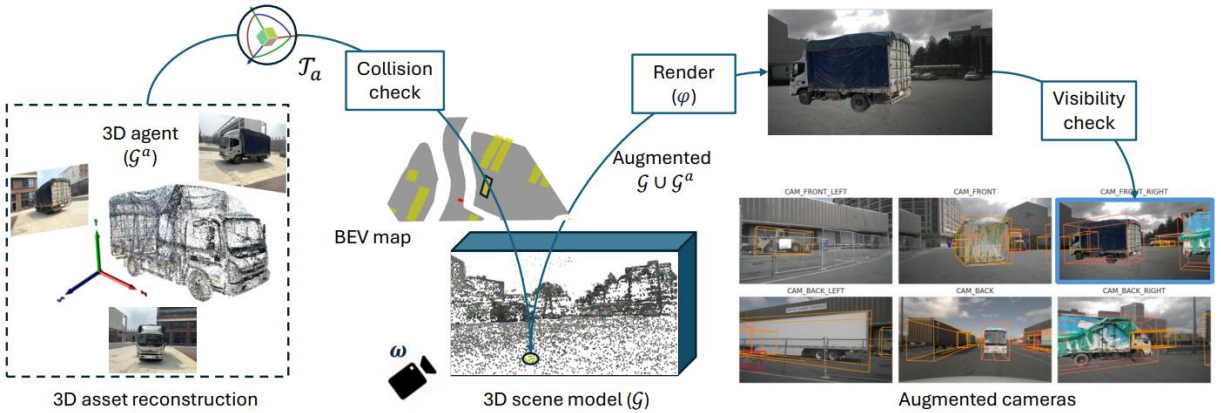
limit geometric diversity and struggle to generate high-quality long-tail corner cases ☹

(b) Video Editing



Single perspective cannot be used for downstream task training ☹

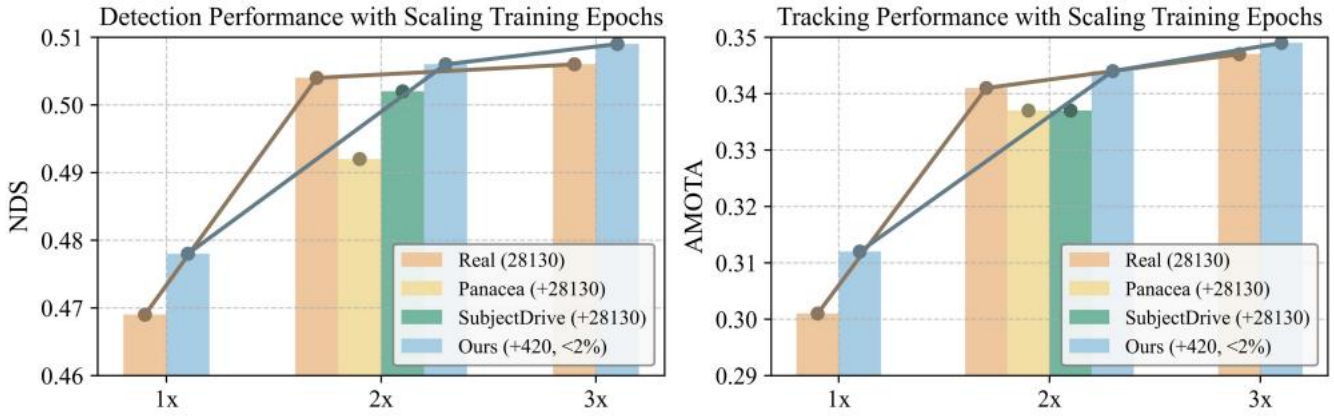
(c) Gaussian Splatting



inconsistent lighting, and restricted background diversity ☹

Figure 2. Overview of multi-camera 3D data augmentation through object placement in 3D field of Gaussian Splatting.

Motivation & Previous methods



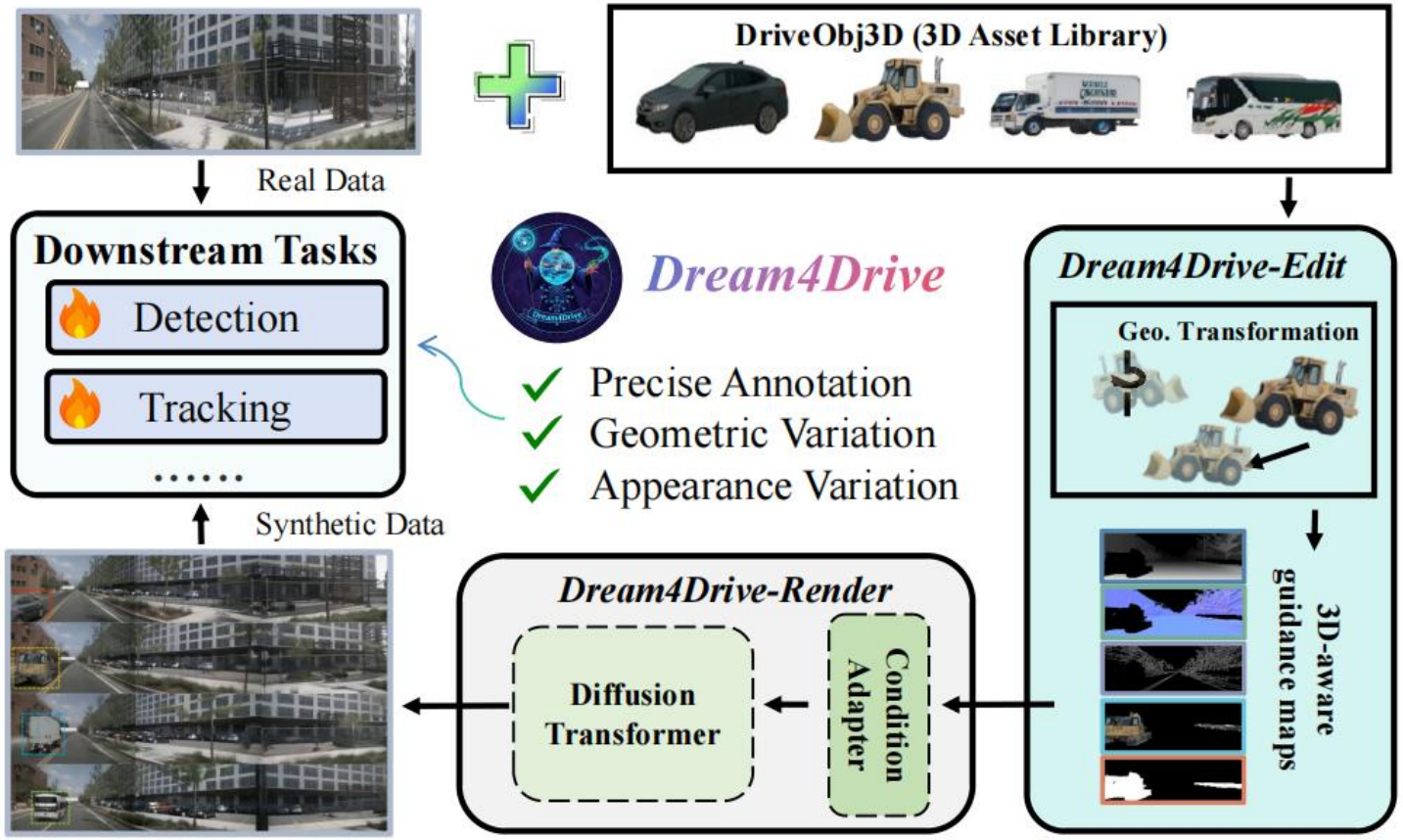
Under the same training epochs, hybrid datasets do not show any advantage over real data alone !

```
Evaluating bboxes of pts_bbox
mAP: 0.3462
mATE: 0.6666
mASE: 0.2742
mAOE: 0.5944
mAVE: 0.2892
mAAE: 0.2043
NDS: 0.4702
Eval time: 148.6s

Per-class results:
Object Class            AP      ATE     ASE     AOE     AVE     AAE
car                     0.546   0.486   0.152   0.116   0.273   0.203
truck                   0.289   0.708   0.216   0.200   0.236   0.206
bus                     0.302   0.779   0.219   0.134   0.583   0.311
trailer                 0.084   1.046   0.245   0.709   0.181   0.054
construction_vehicle    0.098   0.909   0.492   1.209   0.129   0.394
pedestrian              0.399   0.704   0.293   0.647   0.385   0.195
motorcycle              0.346   0.610   0.268   0.921   0.363   0.260
bicycle                 0.330   0.523   0.259   1.241   0.163   0.011
traffic_cone            0.567   0.424   0.312   nan     nan     nan
barrier                 0.501   0.477   0.287   0.171   nan     nan
```

Some categories have extra low indicators.

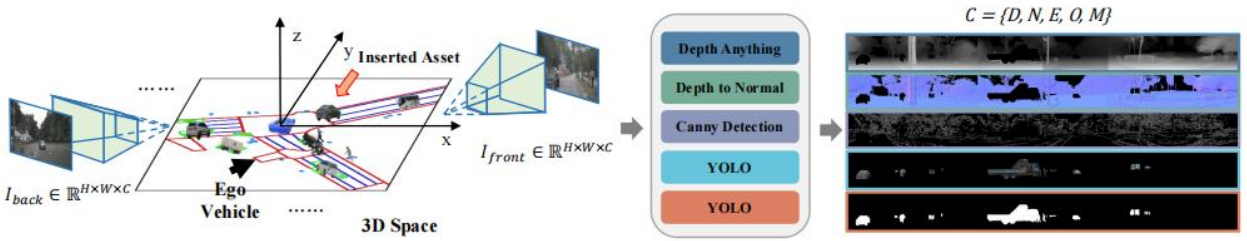
Method


DriveObj3D (3D Asset Library)
Real Data
Downstream Tasks
Detection
Tracking
.....
Synthetic Data
Dream4Drive
Precise Annotation
Geometric Variation
Appearance Variation
Dream4Drive-Edit
Geo. Transformation
3D-aware guidance maps
Dream4Drive-Render
Diffusion Transformer
Condition Adapter

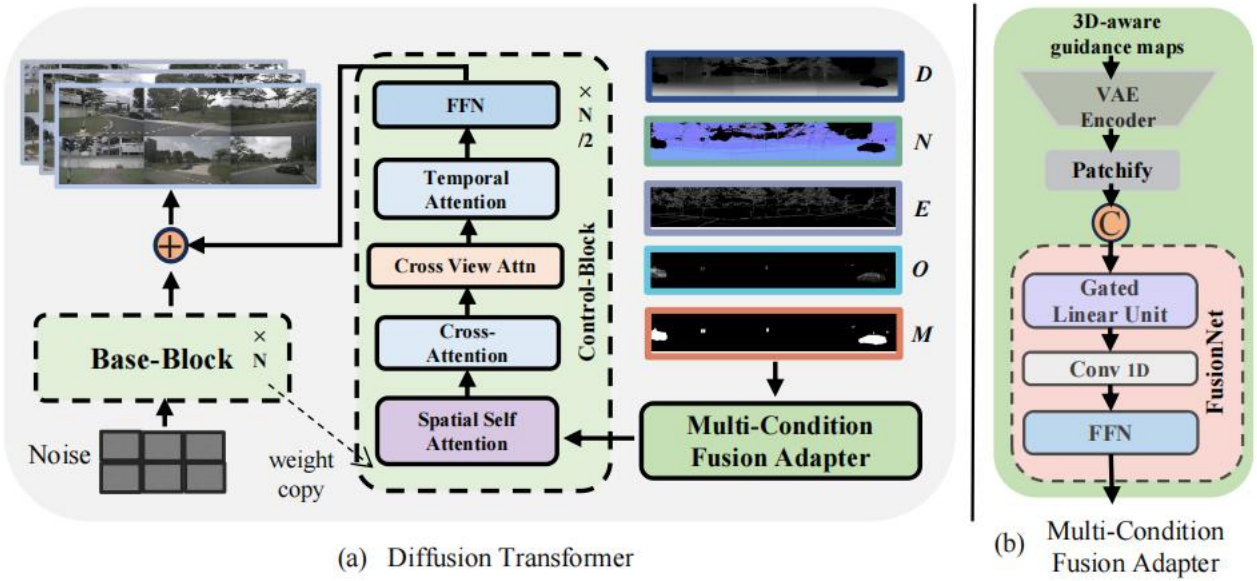
RAIN: A Reusable Asset Infusion Paradigm for Generating Multimodal Autonomous Driving Data

Method

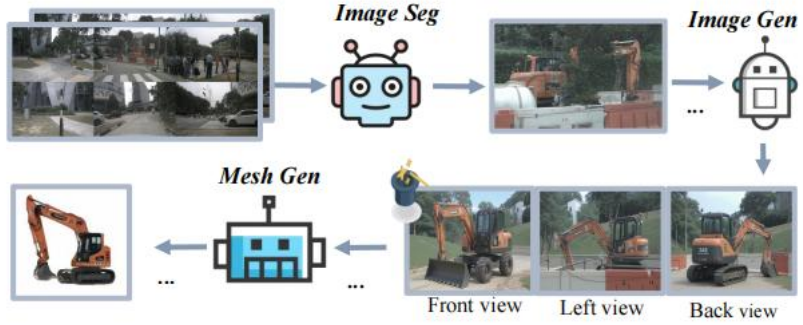
(a) 3D-aware Scene Editing

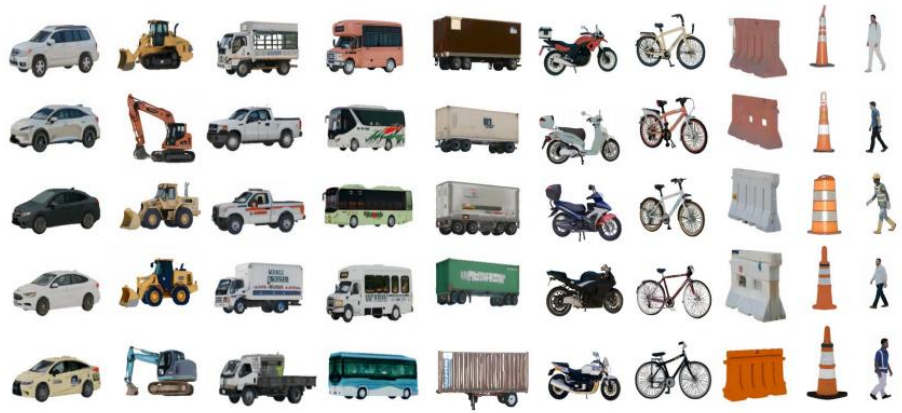


(b) 3D-aware Video Rendering



(c) DriveObj3D





Experiments

1. Effectiveness for Downstream Tasks

	1x Epochs		2x Epochs						
	Real (28130)	Dream4Drive (+420, <2%)	Real (28130)	DriveDreamer (Wang et al., 2024a)	WoVoGen* (Lu et al., 2024)	MagicDrive (Gao et al., 2023)	Panacea (Wen et al., 2024)	SubjectDrive (Huang et al., 2024a)	Dream4Drive (Ours)
mAP ↑	34.5	**36.1**	38.4	35.8	36.2	35.4	37.1	38.1	**38.7**
mAVE ↓	29.1	**28.9**	27.7	–	123.4	–	27.3	**26.4**	26.8
NDS ↑	46.9	**47.8**	50.4	39.5	18.1	39.8	49.2	50.2	**50.6**

	1x Epochs		2x Epochs			
	Real (28130)	Dream4Drive (+420, <2%)	Real (28130)	Panacea (Wen et al., 2024)	SubjectDrive (Huang et al., 2024a)	Dream4Drive (Ours)
AMOTA ↑	30.1	**31.2**	34.1	33.7	33.7	**34.4**
AMOTP ↓	137.9	**135.4**	134.1	135.3	135.3	**133.5**

2. Effectiveness for Various Resolutions

	1x Epochs			2x Epochs			3x Epochs		
	Real	Naive Insert	Dream4Drive	Real	Naive Insert	Dream4Drive	Real	Naive Insert	Dream4Drive
mAP ↑	36.1	40.1	**40.7**	42.2	42.9	**43.6**	43.1	43.1	**44.5**
mATE ↓	69.2	64.7	**64.2**	61.6	62.4	**61.6**	60.5	61.5	**59.8**
mAOE ↓	56.7	49.0	**48.0**	43.2	**37.5**	39.4	45.7	**38.9**	40.1
mAVE ↓	28.5	28.4	**27.1**	27.5	**27.3**	27.4	27.4	27.4	**27.2**
NDS ↑	47.9	51.3	**52.0**	53.2	54.0	**54.3**	53.6	54.2	**55.0**

	1x Epochs			2x Epochs			3x Epochs		
	Real	Naive Insert	Ours	Real	Naive Insert	Ours	Real	Naive Insert	Ours
AMOTA ↑	32.8	36.5	**37.9**	39.7	42.2	**42.6**	41.3	42.2	**43.5**
AMOTP ↓	134.0	128.7	**128.0**	125.1	124.0	**123.3**	124.1	123.7	**121.3**
MOTA ↑	28.1	31.7	**33.1**	35.6	37.3	**37.4**	36.8	37.5	**38.5**
RECALL ↑	44.0	45.4	**46.9**	50.7	51.1	**51.8**	52.4	51.8	**52.5**

Method	car	truck	bus	trailer	construction_vehicle	pedestrian	motorcycle	bicycle	traffic_cone	barrier
Real	0.562	0.323	0.296	0.067	0.111	0.422	0.377	0.371	0.569	0.515
Ours	0.600	0.354	0.341	0.106	0.135	0.468	0.402	0.411	0.626	0.565

Table 6: AP comparison across different categories for Real and Ours (+420) at 1× training epoch.

Method	car	truck	bus	trailer	construction_vehicle	pedestrian	motorcycle	bicycle	traffic_cone	barrier
Real	0.622	0.371	0.375	0.154	0.132	0.493	0.430	0.400	0.651	0.589
Ours	0.633	0.383	0.389	0.168	0.147	0.497	0.446	0.434	0.647	0.604

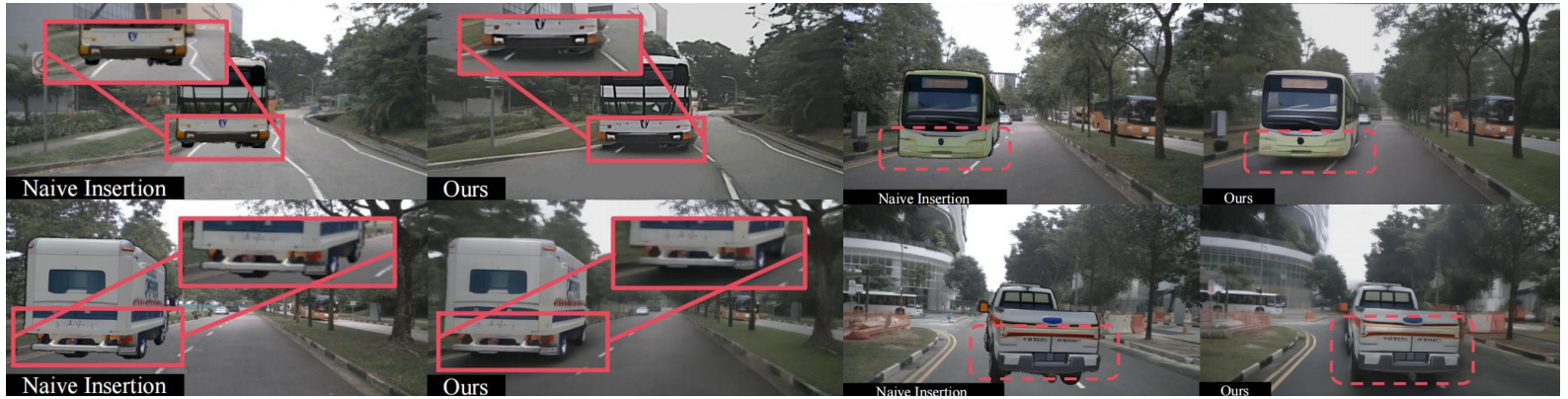
Table 7: AP comparison across different categories for Real and Ours (+420) at 2× training epoch.

Method	car	truck	bus	trailer	construction_vehicle	pedestrian	motorcycle	bicycle	traffic_cone	barrier
Real	0.627	0.362	0.367	0.154	0.132	0.488	0.436	0.454	0.656	0.632
Ours	0.639	0.377	0.408	0.190	0.164	0.501	0.446	0.439	0.654	0.621

Table 8: AP comparison across different categories for Real and Ours (+420) at 3× training epoch.

Experiments

3. Quantitative and Qualitative Comparison with Naive Insertion.



4. Ablation Studies

	Views				Distances			3D Asset Generation Methods		
	Front	Back	Left	Right	Close	Mid	Far	Trellis	Hunyuan3D	Ours
mAP ↑	**40.2**	**40.2**	**40.2**	39.8	39.7	40.3	**40.5**	39.8	40.2	**40.7**
mATE ↓	66.2	66.0	**64.6**	66.2	65.7	65.4	**65.1**	65.6	65.1	**64.2**
mAOE ↓	51.2	55.1	**45.7**	51.4	52.2	51.7	**49.7**	51.8	50.8	**48.0**
mAVE ↓	27.7	**27.5**	28.5	27.9	28.1	28.0	**27.9**	27.8	28.0	**27.1**
NDS ↑	51.0	50.6	**51.6**	50.7	50.5	50.9	**51.3**	50.8	50.9	**52.0**

Experiments

5. More Visualization

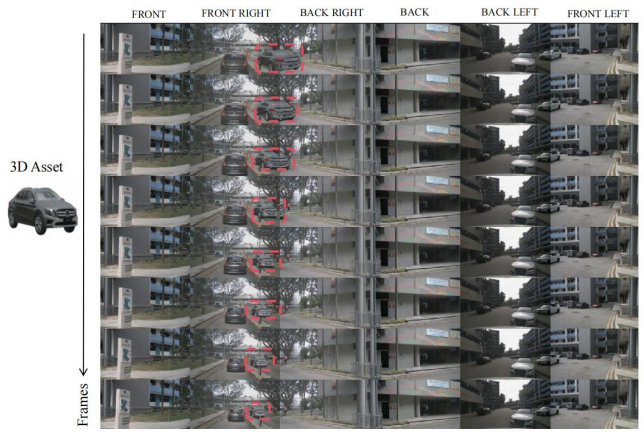


Figure 11: Insertion of a car in the right-side region.

Figure 12: Insertion of a truck in the left-side region.

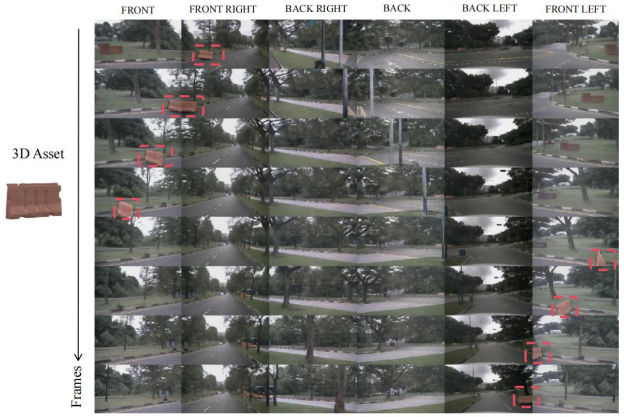


Figure 13: Insertion of a barrier in the left-side region.

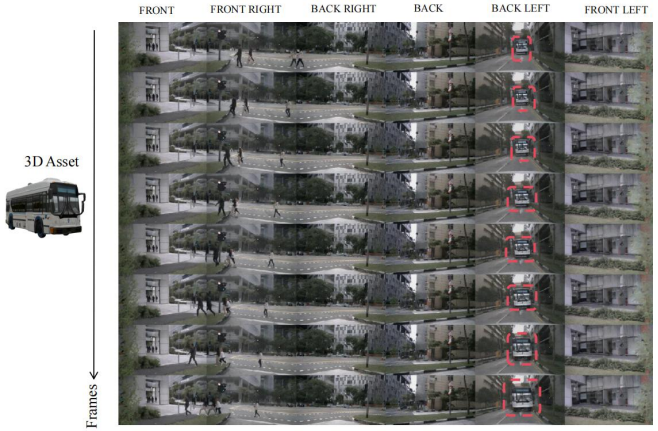


Figure 14: Insertion of a bus in the back-side region.

Figure 15: Insertion of a traffic cone in the left-side region.

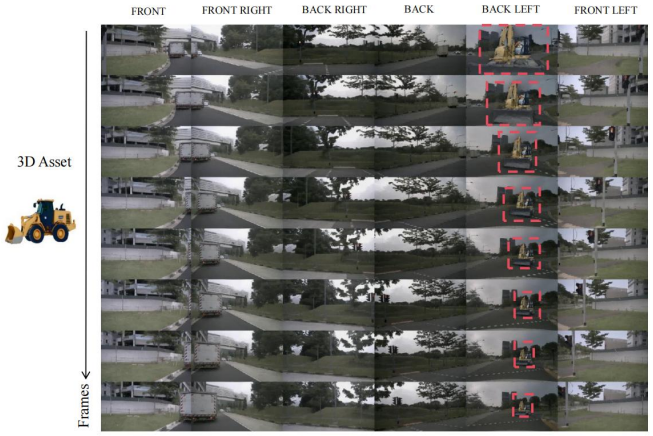


Figure 16: Insertion of a construction vehicle in the back-side region.