

SimGen: Simulator-conditioned Driving Scene Generation

Rebuttal supplementary materials: Figures and tables

Table 1: Evaluation of annotated data quality. We assess the quality levels of the annotated data in terms of text, depth, and segmentation. The pseudo-labels of text are derived from GPT4V and are checked by manual annotators. REL: absolute relative error of depth estimation. †: we randomly sample 500 images from this dataset.

Dataset	Text (ROUGE-L↑)	Depth (REL ↓)	Segmentation (mIoU ↑)
DIVA-Real†	84.4	-	-
nuScenes	85.2	0.118	82.4

Table 2: Per-class results of the generation controllability for perception tasks. Oracle: a single-frame version of BEVFusion. Constr.: Construction vehicles; Ped.: Pedestrian; Motor.: Motorcycle.

Method	Car	Truck	Bus	Trailer	Constr.	Ped.	Motor.	Bicycle	Cone	Barrier
Oracle	47.0	21.4	45.1	14.4	1.7	33.5	27.2	25.4	48.7	53.0
SimGen	41.0 (-6.0)	19.6 (-1.8)	40.3 (-4.8)	13.5 (-0.9)	1.5 (-0.2)	28.7 (-4.8)	24.7 (-2.5)	22.5 (-2.9)	45.2 (-3.5)	49.7 (-3.3)

Table 3: Per-class comparison involving data augmentation using synthetic data. Constr.: Construction vehicles; Ped.: Pedestrian; Motor.: Motorcycle.

Method	Car	Truck	Bus	Trailer	Constr.	Ped.	Motor.	Bicycle	Cone	Barrier
Baseline	47.0	21.4	45.1	14.4	1.7	33.5	27.2	25.4	48.7	53.0
SimGen	49.1	23.6	48.6	15.8	4.9	35.7	30.1	26.9	50.3	54.7

Table 4: Per-class results of the ablation on designs in SimGen. All proposed designs contribute to the final performance. Constr.: Construction vehicles; Ped.: Pedestrian; Motor.: Motorcycle.

Method	Car	Truck	Bus	Trailer	Constr.	Ped.	Motor.	Bicycle	Cone	Barrier
Baseline	45.7	20.8	43.9	12.4	0.0	32.5	26.0	24.7	47.4	49.5
+ Cascaded Pipeline	46.3	21.4	45.2	12.8	0.3	33.2	27.7	25.5	47.6	51.1
+ ExtraCond	47.6	22.0	46.5	14.5	2.8	33.4	29.2	26.1	48.2	51.9
+ Unified Adapter	48.2	22.2	47.5	15.0	3.5	33.8	29.7	26.7	49.4	54.1

Table 5: Distribution of nuScenes and DIVA-Real on location, time period, and weather. N. A.: North America; S. A.: South America.

Dataset	Continent					Period				Weather				
	N. A.	S. A.	Europe	Asia	Africa	Daytime	Dawn	Dusk	Nighttime	Normal	Rainy	Cloudy	Foggy	Snowy
nuScenes	44.1%	0.0%	0.0%	55.9%	0.0%	88.4%	0.0%	0.0%	11.6%	80.5%	19.5%	0.0%	0.0%	0.0%
DIVA-Real	56.9%	8.5%	16.9%	14.6%	3.1%	55.8%	16.3%	10.1%	17.8%	58.2%	1.0%	28.6%	2.1%	10.2%

Table 6: Layout distribution of nuScenes and DIVA-Sim.

Dataset	Forward	Left Turn	Right Turn	Left Lane Change	Right Lane Change	Intersection Passing	U-Turn	Stop
nuScenes	47.1%	18.0%	10.2%	5.0%	2.5%	13.1%	0.0%	4.1%
DIVA-Sim	36.2%	14.3%	10.0%	10.7%	17.4%	6.6%	1.2%	3.6%

Table 7: Behavior distribution of corner cases in DIVA-Sim. A single scenario may involve multiple behaviors.

Dataset	Accelerate	Decelerate	Close Following	Sudden Breaking	Crash Front	Crash Back	Crash Left	Crash Right	Cutting In
Proportion	30.4%	15.2%	7.3%	8.4%	29.4%	41.2%	16.5%	13.9%	19.0%

Table 8: Scalability analysis for data augmentation.

Proportion	50%	100%	150%	200%	250%
AP _{Car}	47.7	49.1	50.2	51.2	51.9

Table 9: Analysis for classifier-free guidance (CFG) scale.

CFG	5	7.5	9.0	11.5	14
D _{Pix}	21.3	24.5	26.6	28.0	28.2