DriveGEN: Generalized and Robust 3D Detection in Driving via Controllable Text-to-Image Diffusion Generation

Supplementary Material

In the supplementary, we first provide more related work and discussions to clarify existing vision-centric 3D object detection methods. In addition, we provide more experimental details, visualizations, and results of DriveGEN. We organize our supplementary materials as follows.

- In Appendix A, we review vision-centric 3D object detection and provide more discussions.
- In Appendix B, we provide more details of KITTI-C and the real-world scenarios of nuScenes.
- In Appendix C, we provide more details and results of DriveGEN based on Stable Diffusion 2.1 and XL.
- In Appendix D, we show more experimental results to demonstrate the effectiveness of the proposed DriveGEN.
- In Appendix E, we show more qualitative results of our DriveGEN.

A. More Related Work and Discussions

In this section, we first provide more related work and discussions to clarify existing solutions to 3D object detection.

Vision-Centric 3D Object Detection. In autonomous driving, vision-centric 3D object detection is essential for accurate environment understanding. Traditional methods have often relied on LiDAR data, which provides precise depth information but necessitates costly hardware. Recently, there has been a shift toward using monocular and stereo cameras to reduce hardware dependency, but these methods struggle with depth accuracy, particularly at longer distances. Based on advances in transformer architectures, current approaches [7, 13, 35, 38, 40, 48] attempt to bridge the gap between 2D images and 3D spatial reasoning using feature extraction and spatial alignment. However, these models often demand high computational resources, making them challenging for real-time application. Diffusion-based models contribute to this task by offering a robust generation of 3D scene layouts capable of simulating diverse environments [45, 55], while advances in multimodal integration [18, 23, 26] enable the use of additional sensory data to enrich 3D object detection frameworks. The increasing trend of vision-centric 3D perception systems of autonomous vehicles further proves the effectiveness of model robustness.

However, they still fall short in effectively capturing and aligning spatiotemporal features, which reduces overall accuracy and contextual understanding in complex environments. As discussed in the manuscript, existing 3D detectors often fail to maintain stable performance in OOD scenarios, which raises concerns about safety risks.

Figure 7. Illustration of the real-world scenario Daytime, Night and Rainy of the nuScenes dataset. Given a pre-trained multi-view 3D detector, we enhance the detector with the augmented data from DriveGEN. Even if the augmented data never appears in the validation set (*e.g.,* Snow), DriveGEN still improves the model performance, which shows the robustness and generalizability improvement of the augmented detector.

B. More Details on Dataset Construction

In this section, we first provide more visualizations of the KITTI-C dataset to illustrate the OOD scenarios. Then, we offer more details of the real scenarios (*i.e.,* Daytime, Night and Rainy) of the nuScenes dataset.

For the KITTI-C dataset, as shown in Figure 8, we follow MonoTTA to build 13 OOD scenarios based on the original KITTI validation set [22], which is able to fully verify the effectiveness of each method in addressing dataset distribution shifts for Monocular 3D Object Detection.

As for the nuScenes dataset, we split images into Night and Rainy scenarios according to their descriptions, following [24], as shown in Figure 7. To be specific, given a pre-trained multi-view 3D detector, DriveGEN first augments the original training data into various OOD scenarios and then mixes the augmented data with the original training data for the model retraining. It is worth mentioning that even if the augmented scenarios never appear in the validation set, DriveGEN still consistently improves the model performance, demonstrating that DriveGEN improves the robustness and generalizability of the augmented detector by injecting the knowledge from diffusion models.

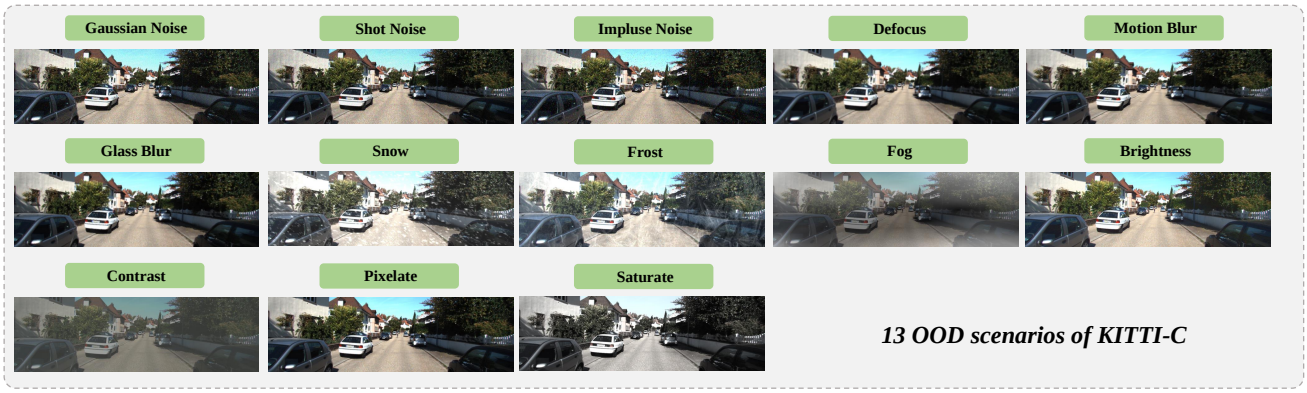

Figure 8. Illustration of 13 distinct OOD scenarios of the KITTI-C [22] dataset.

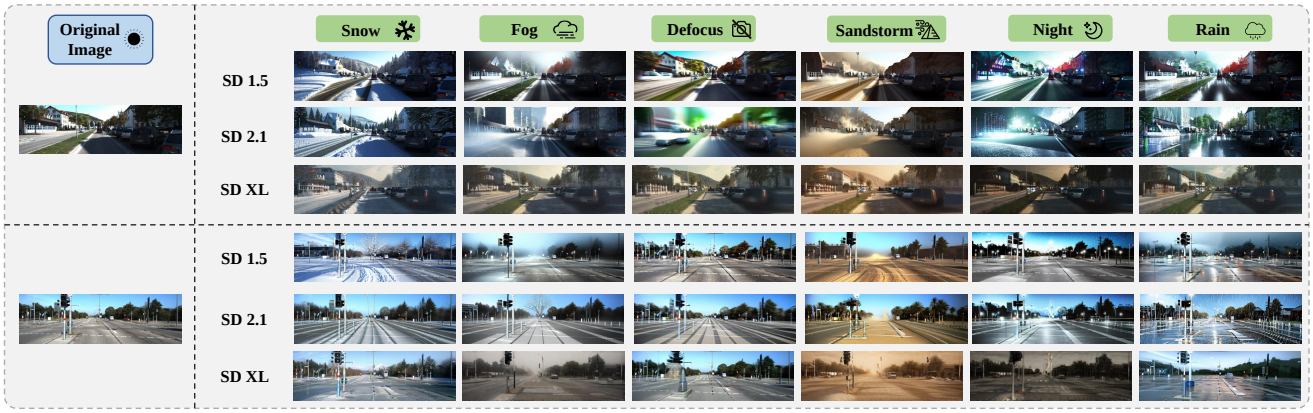

Figure 9. Qualitative results of DriveGEN based on Stable Diffusion 2.1 and Stable Diffusion XL.

C. More Details and Results with Stable Diffusion 2.1 and Stable Diffusion XL

More Implementation Details. Based on PyTorch [29], we conduct experiments with NVIDIA A100 (80GB of memory) GPUs and each method is executed on a single GPU. To be specific, we adopt keys from the first self-attention layer of the U-Net decoder as the features following Freecontrol [25]. Besides, the number of steps for DDIM inversion and DDIM sampling is 200. We set s and τ to 7.5 and 0.3 following existing method [25], while σ is set to 0.1. The corresponding ablation studies are put in Appendix **D**. All images of KITTI and nuScenes have the fixed generation size of 368×1240 and 896×1600, respectively.

As for the training of 3D detectors, we follow their original settings without any hyper-parameter modification. Specifically, given the augmented training data, we mix the original training data with all augmented OOD scenario for model training. Then, we validate all models on the original validation set to choose the best model because the original training data predominantly represents the most commonly occurring scenarios.

Stable Diffusion 2.1 and XL 1.0 In this section, we present additional quantitative results based on Stable Diffusion 2.1 and Stable Diffusion XL. Since DriveGEN requires no additional diffusion model training, it is straightforward to extend DriveGEN to the U-Net architecture [36] based diffusion model, as shown in Figure 9. These results demonstrate that DriveGEN effectively supports Stable Diffusion models across different versions (*i.e.,* U-Net-based), showcasing its flexibility and scalability for integrating with vision-centric 3D detection methods.

D. More Experimental Results

In this part, we conduct ablation studies to show the effectiveness of DriveGEN, like enhancing training data via various single OOD scenarios as shown in Table 3, as well as enhancing training data by three additional OOD scenarios as shown in Table 4. Meanwhile, we provide more detailed experimental results of the Cyclist category and the results regarding Moderate $AP_{3D|R_{40}}$ on the KITTI-C dataset. Eventually, we provide more results of nuScenes with the BEVFormer-small [20].

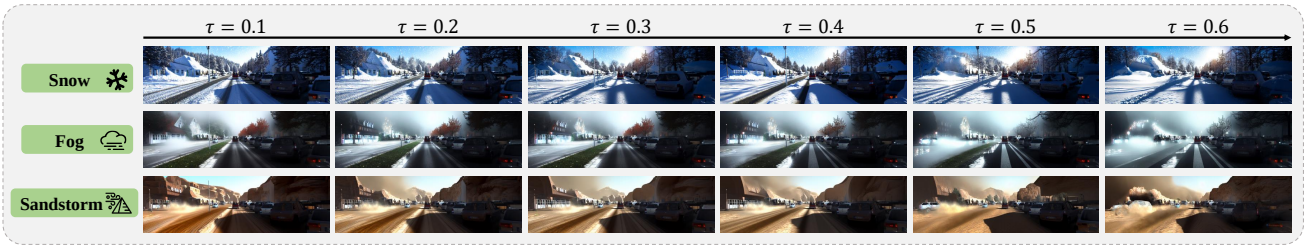

Figure 10. DriveGEN is able to control the severity of corruptions while still preserving all objects.

Table 3. Comparison on the KITTI-C dataset, severity **level 1** regarding Mean $AP_{3D|R_{40}}$. Each scenario represents the training of the 3D detector, which is enhanced with corresponding OOD data. The **bold** number indicates the best result.

Car, IoU @ 0.7, 0.5, 0.5														
Method	Noise			Blur			Weather				Digital			Avg.
	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Snow	Frost	Fog	Brit.	Contr.	Pixel	Sat.	
Monoflex [63]	13.06	20.91	14.09	20.17	28.59	30.34	33.64	30.31	19.58	45.22	20.01	29.07	38.85	26.45
• Snow	17.07	26.78	23.78	32.89	37.52	39.06	40.61	34.91	25.29	46.21	27.12	38.25	44.45	33.38
• Fog	17.98	29.72	20.66	34.10	39.25	38.47	39.49	38.48	30.05	47.90	30.71	37.02	45.19	34.54
• Rainy	19.49	31.10	**27.44**	33.34	**39.55**	39.41	**41.42**	**41.10**	**37.11**	47.59	**38.69**	41.05	45.35	37.13
• Night	23.28	**35.21**	26.82	**35.13**	39.24	39.62	40.69	40.77	34.68	47.46	36.06	**42.40**	**45.55**	**37.46**
• Defocus	22.06	32.16	26.64	34.45	37.71	**40.69**	40.37	37.94	31.63	**48.49**	34.88	41.19	44.23	36.34
• Sandstorm	**24.46**	33.96	26.09	31.50	37.18	37.07	40.79	38.83	32.65	43.85	33.44	37.98	42.60	35.41

Table 4. Comparison on the KITTI-C dataset, severity **level 1** regarding Mean $AP_{3D|R_{40}}$. Each setting represents the training of the 3D detector is enhanced with corresponding mixed OOD data. The **bold** number indicates the best result.

Car, IoU @ 0.7, 0.5, 0.5														
Method	Noise			Blur			Weather				Digital			Avg.
	Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Snow	Frost	Fog	Brit.	Contr.	Pixel	Sat.	
Monoflex [63]	13.06	20.91	14.09	20.17	28.59	30.34	33.64	30.31	19.58	45.22	20.01	29.07	38.85	26.45
• Snow & Fog & Rain	25.63	37.04	29.13	**34.13**	**39.15**	36.81	38.58	37.98	33.93	45.39	34.66	39.36	43.81	36.58
• Defocus & Night & Sandstorm	**28.93**	**37.91**	**32.12**	31.59	37.66	**39.27**	**40.34**	**40.88**	**37.72**	**47.15**	**38.00**	**42.34**	**45.58**	**38.42**

Figure 11. More ablation studies of DriveGEN regarding hyper-parameters s and σ.

Ablation Study on Augmented OOD Scenarios. To fully validate the effectiveness of DriveGEN, we provide more results of enhancing training data via various single OOD scenarios as shown in Table 3 and enhancing training data by three additional OOD scenarios as shown in Table 4. On the one hand, compared with the Snow augmentation. setting in the manuscript, DriveGEN improves the 3D detection model with stable performance improvement by the other 5 OOD scenarios. In Table 3, even with a single augmentation, DriveGEN improves the 3D detector [63] up to **11.01 mAP** (*i.e.,* Night) across 13 OOD scenarios. On the other hand, if we enhance the 3D detector with another three scenarios (*i.e.,* Defocus, Night, Sandstorm), DriveGEN still achieves significant performance improvement compared with the pre-trained 3D detector as shown in Table 4.

Ablation Study on Hyper-parameters. One intuitive concern is whether the severity of corruptions can be controlled by DriveGEN while still preserving all annotated objects with precise geometry. To this end, we provide more qualitative results regarding various values of τ as shown in Figure 10. It is clearly observed that the corresponding corruption progressively exerts a more severe impact on the background with the increasing of τ. However, DriveGEN maintains all objects well with their precise 3D geometry, thereby demonstrating the effectiveness of the proposed method. Meanwhile, we also analyze the effects of s and σ as shown in Figure 11. As s increases, corruptions intensify while larger σ retains more original object details. It also shows that DriveGEN consistently preserves objects across various settings, validating its insensitivity to hyper-parameters.

Table 5. Comparison on the Cyclist category of the KITTI-C dataset regarding Mean $AP_{3D|R_{40}}$. **Bold** number indicates the best result.

Cyclist, IoU @ 0.7, 0.5, 0.5															
Method	Training-free diffusion	Noise			Blur			Weather				Digital			Avg.
		Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Snow	Frost	Fog	Brit.	Contr.	Pixel	Sat.	
Monoflex [63]		0.43	2.41	0.64	2.76	8.30	9.14	12.85	11.09	5.73	17.44	4.84	3.25	9.89	6.83
• Color Jitter (Traditional aug.)		0.63	3.15	1.91	1.62	3.43	7.92	11.03	10.09	4.60	12.41	4.61	1.43	10.23	5.62
• Brightness (Traditional aug.)		0.21	1.16	0.25	1.33	3.45	6.14	9.67	8.81	4.89	13.66	5.82	2.02	7.93	5.03
• ControlNet (Only Snow aug.)	✗	0.00	0.30	0.00	0.00	3.77	4.29	7.27	6.47	6.97	15.79	6.49	1.67	2.54	4.27
• ControlNet (3 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.50	1.50	1.82	1.35	0.00	0.00	0.55
• ControlNet (6 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
• Freecontrol (Only Snow aug.)	✓	1.58	4.43	1.72	0.00	0.39	0.94	3.97	1.52	0.54	5.26	0.68	1.07	7.50	2.28
• Freecontrol (3 scenarios aug.)	✓	0.00	0.00	0.00	0.00	2.50	0.00	1.25	1.04	1.91	3.82	1.78	0.00	2.47	1.14
• Freecontrol (6 scenarios aug.)	✓	0.19	0.24	0.45	0.00	0.31	0.00	0.55	0.52	1.01	2.13	2.12	0.61	0.81	0.69
• DriveGEN (Only Snow aug.)	✓	0.70	1.27	0.61	1.34	5.26	5.27	10.90	7.12	3.73	15.14	4.37	1.74	11.24	5.28
• DriveGEN (3 scenarios aug.)	✓	1.04	3.42	1.53	2.36	5.62	7.69	8.14	4.17	5.20	13.24	5.23	4.58	11.07	5.64
• DriveGEN (6 scenarios aug.)	✓	0.53	0.93	0.54	3.07	10.95	9.38	11.12	12.60	9.07	15.39	10.81	1.99	8.05	**7.26**
MonoGround [31]		0.21	1.86	1.34	0.83	2.93	2.23	5.00	3.43	0.94	11.48	1.21	2.04	5.92	3.03
• Color Jitter (Traditional aug.)		0.39	2.67	2.11	0.31	2.03	2.19	5.38	4.63	1.12	13.64	1.67	2.89	5.00	3.39
• Brightness (Traditional aug.)		0.06	0.61	0.22	0.36	1.33	1.06	4.72	2.32	1.41	6.87	0.78	0.90	2.81	1.80
• ControlNet (Only Snow aug.)	✗	0.00	0.00	0.52	0.00	0.77	1.33	0.44	1.10	0.14	6.77	0.30	0.50	0.54	0.95
• ControlNet (3 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.30	0.00	0.83	1.50	0.00	1.70	0.00	0.37	0.00	0.36
• ControlNet (6 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
• Freecontrol (Only Snow aug.)	✓	0.46	0.70	0.32	1.07	0.17	0.50	1.60	0.91	0.21	4.48	0.17	1.93	4.50	1.31
• Freecontrol (3 scenarios aug.)	✓	0.19	0.33	0.43	0.00	0.52	0.50	0.51	0.33	0.58	1.32	0.42	0.38	0.38	0.45
• Freecontrol (6 scenarios aug.)	✓	0.00	0.34	0.62	0.00	0.42	0.56	0.00	0.00	1.00	2.07	0.83	1.25	0.74	0.60
• DriveGEN (Only Snow aug.)	✓	0.13	0.81	0.38	0.31	2.23	3.66	3.96	2.02	0.90	8.46	1.74	2.07	4.58	2.40
• DriveGEN (3 scenarios aug.)	✓	0.61	2.17	1.51	1.09	3.24	2.73	5.57	5.30	1.05	9.79	1.33	4.66	6.10	3.47
• DriveGEN (6 scenarios aug.)	✓	1.49	2.16	1.66	3.30	5.97	5.55	5.64	5.49	2.49	9.37	3.48	3.79	5.65	**4.31**

Table 6. Detection results on nuScenes-C and real-world scenarios of nuScenes, regarding mAP and NDS.

Metric	Method	nuScenes-C									Real-world Scenarios
		Brightness	CameraCrash	ColorQuant	Fog	FrameLost	LowLight	MotionBlur	Snow	Avg.	nuScenes-Night
mAP	BEVFormer-small	36.12	23.25	36.05	32.70	**32.16**	23.61	**32.03**	13.66	28.70	19.59
	• DriveGEN (3k Snow)	**37.99**	**24.20**	**37.66**	**34.60**	31.74	**25.82**	**32.89**	**17.63**	**30.32**	**22.39**
NDS	BEVFormer-small	47.36	39.49	47.28	45.01	**44.05**	38.62	44.46	28.45	41.84	27.27
	• DriveGEN (3k Snow)	**48.90**	**40.44**	**48.63**	**46.56**	43.72	**39.67**	**45.47**	**32.32**	**43.21**	**28.86**

To summarize, DriveGEN enhances the generalizability and robustness of vision-centric 3D detectors with diverse augmented OOD scenarios, achieving stable performance improvement within various training data settings. These experimental results further demonstrate the effectiveness of DriveGEN.

Model Performance on Cyclist. We further provide more results of the cyclist category of the KITTI-C dataset as shown in Table 5. As mentioned in MonoTTA [22], even Fully Test-Time Adaptation methods (*i.e.,* allowed to access test data for model adaptation) only gain limited performance improvement on the Cyclist category. Table 5 gives a similar observation: even if DriveGEN achieves the best performance in these extremely difficult cases, all methods only gain limited performance improvement, which indicates that the challenge of minority-class object detection still requires further investigation.

Model Performance Regarding the Moderate Level. In 3D object detection, the performance for the *Moderate* difficulty level of the KITTI dataset is one of the most significant indicators of model effectiveness. To this end, we provide more experimental results as shown in Table 9. This table shows that DriveGEN still achieves the best average performance within various augmentation settings and base models, demonstrating the effectiveness of our method.

More Results on NuScenes and NuScenes-C. Based on BEVFormer-small [20], we apply the Snow augmentation (3k Snow) to enhance the detector as mentioned in the manuscript. Table 6 shows that DriveGEN consistently enhances BEVFormer-small across 8 OOD scenarios with an average of 1.62 mAP and 1.37 NDS, and across the real-world scenario of nuScenes (*i.e.,* Night) with 2.80 mAP and 1.59 NDS. These results further demonstrate our effectiveness and superiority.

E. More Qualitative Results

As shown in Figure 12, we first provide more qualitative results based on the training images of the KITTI dataset. In addition, we also offer more qualitative results based on the training images of the nuScenes dataset as shown in Figure 13. It is evident that DriveGEN supports existing vision-

Figure 12. More qualitative results of DriveGEN for the training data of the KITTI dataset.

centric 3D detection tasks (*i.e.,* monocular 3D object detection and multi-view 3D object detection) since our method only requires input images and corresponding object annotations without any additional diffusion model training, demonstrating that DriveGEN can preserve all objects and enhance training data, thus achieving superior results even in challenging multi-view tasks.

Table 7. The quality comparisons of object regions based on PSNR and SSIM.

Metric	Method	Defocus	Snow	Fog	SandStorm	Night	Rainy	Avg.
PSNR	ControlNet	7.671	7.551	8.130	7.764	7.825	8.129	7.845
	FreeControl	11.883	10.601	12.529	11.957	11.515	12.119	11.767
	DriveGEN	**19.584**	**18.963**	**19.528**	**19.551**	**18.906**	**19.308**	**19.306**
SSIM	ControlNet	0.077	0.069	0.081	0.073	0.074	0.080	0.075
	FreeControl	0.119	0.067	0.144	0.143	0.106	0.097	0.113
	DriveGEN	**0.641**	**0.616**	**0.632**	**0.633**	**0.614**	**0.627**	**0.627**

More Discussions on Image Quality. We first report PSNR and SSIM for object regions between generated and original images. Table 7 reveals DriveGEN stably preserves object geometries within all scenarios without compromising the quality. Note that preserving object information is crucial for reusing annotations, as misaligned objects may introduce bias as shown in Figure 2 in the manuscript.

Table 8. Comparisons of the detector performance enhanced by DriveGEN with various diffusion steps.

KITTI-C	MonoFlex	FreeControl(200 steps)	DriveGEN-50	DriveGEN-100	DriveGEN-200
Time (hour)	36.85	43.97	**8.69**	32.40	52.97
Car (mAP)	26.45	22.44	32.48	33.93	**34.07**

Besides, to explore the computation cost and improve efficiency, we provide further analysis of DriveGEN with different diffusion step settings. As shown in Table 8, we conduct experiments with 50 and 100 generation steps (8×3090 GPUs). With fewer steps, DriveGEN drastically reduces time consumption while still achieving sufficient gains. To further validate DriveGEN, we also present additional qualitative results obtained with varying numbers

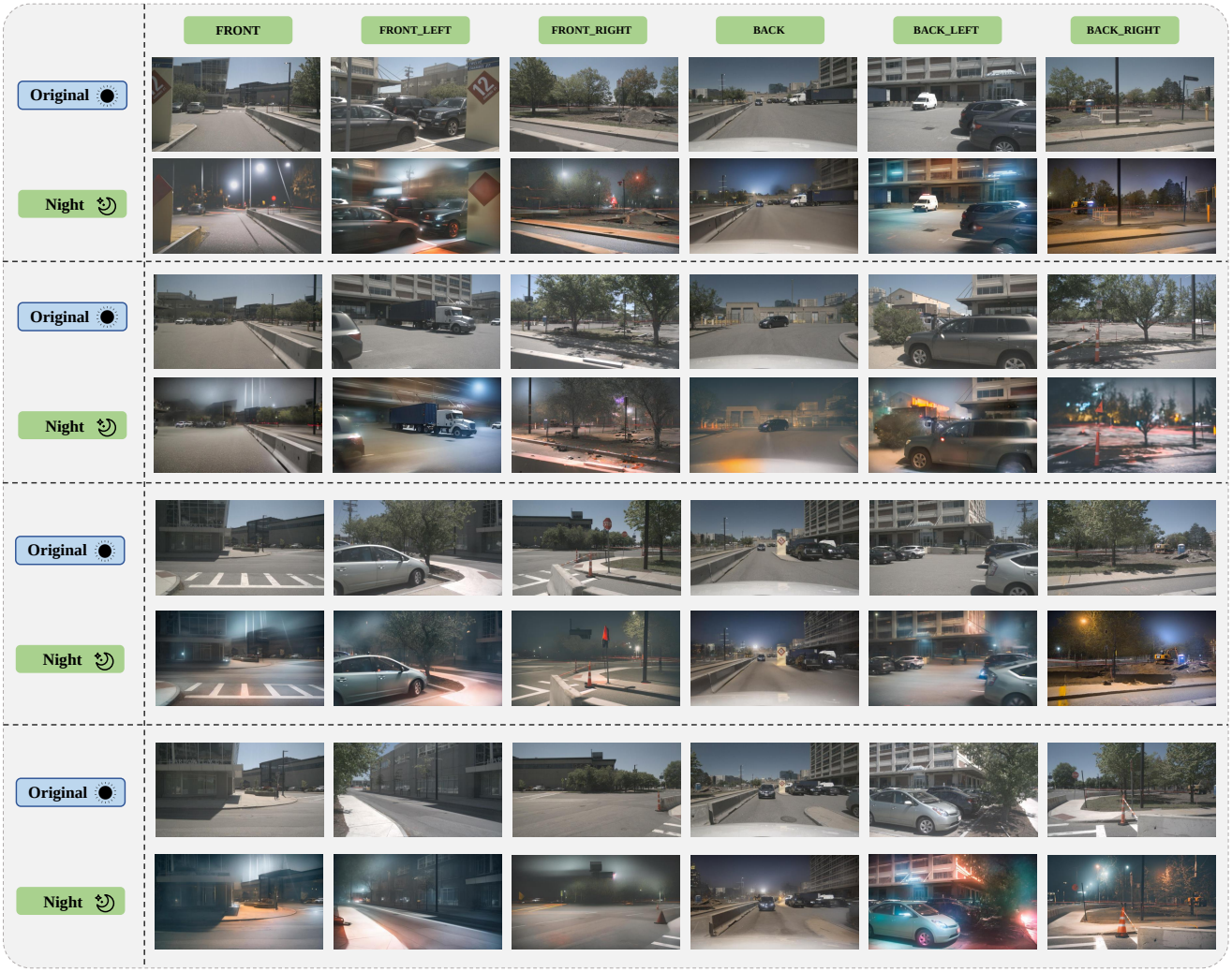

Figure 13. More qualitative results of DriveGEN for multi-view training images of the nuScenes dataset.

Figure 14. More qualitative results of DriveGEN with varying numbers of diffusion steps.

of diffusion steps as shown in Figure 14. It is observed that DriveGEN consistently preserves objects even when only 50 steps ($N_t = 50$) are used. Moreover, increasing the number of steps leads to progressively enhanced image quality. These results reveal the inherent trade-off between image quality and efficiency, thereby enabling users to tailor the approach according to their application demands.

Table 9. Comparison on the KITTI-C dataset, severity **level 1** regarding **Moderate** $AP_{3D|R_{40}}$. The **bold** number indicates the best result.

Method	Training-free diffusion	Noise			Blur			Weather				Digital			Avg.
		Gauss.	Shot	Impul.	Defoc.	Glass	Motion	Snow	Frost	Fog	Brit.	Contr.	Pixel	Sat.	
Car, IoU @ 0.7, 0.5, 0.5															
Monoflex [63]		12.69	19.65	13.88	17.81	25.86	27.44	31.53	28.77	18.90	42.36	18.94	26.48	35.51	24.60
• Color Jitter (Traditional aug.)		9.75	15.31	11.84	20.55	23.02	27.19	31.55	28.69	18.62	39.22	19.32	15.85	32.95	22.61
• Brightness (Traditional aug.)		10.76	17.60	11.99	10.06	17.14	19.25	25.22	20.77	12.55	36.73	12.38	19.35	27.79	18.58
• ControlNet (Only Snow aug.)	✗	0.33	1.21	1.44	3.99	10.07	15.37	21.99	19.16	9.20	32.35	8.86	1.30	16.00	10.87
• ControlNet (3 scenarios aug.)	✗	1.01	1.10	0.48	0.39	0.41	0.84	4.04	3.20	1.24	9.03	1.11	0.46	3.60	2.07
• ControlNet (6 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.00	1.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14
• Freecontrol (Only Snow aug.)	✓	18.31	25.11	19.97	11.37	19.31	18.88	25.29	16.56	9.30	31.85	9.13	25.37	31.73	20.17
• Freecontrol (3 scenarios aug.)	✓	13.39	18.37	11.69	14.06	17.39	14.96	19.65	15.63	14.17	21.49	15.40	22.21	27.72	17.39
• Freecontrol (6 scenarios aug.)	✓	11.48	16.64	14.79	11.25	14.73	11.51	16.06	14.93	13.18	21.69	13.83	20.03	21.74	15.53
• DriveGEN (Only Snow aug.)	✓	15.82	24.72	23.33	28.17	32.42	34.85	38.00	35.39	25.90	**45.99**	27.51	35.55	**40.41**	31.39
• DriveGEN (3 scenarios aug.)	✓	**23.71**	**33.55**	**26.80**	30.26	35.20	33.15	35.60	34.68	31.59	41.10	31.70	35.50	39.92	33.29
• DriveGEN (6 scenarios aug.)	✓	23.24	31.24	26.40	**33.48**	**36.20**	**37.00**	**37.21**	**37.10**	**35.16**	41.00	**35.93**	**38.92**	40.30	**34.86**
MonoGround [31]		12.30	20.27	17.08	18.41	27.22	29.04	32.16	25.56	13.34	43.41	14.14	31.21	32.55	24.36
• Color Jitter (Traditional aug.)		11.96	22.17	17.69	20.67	27.17	28.47	32.93	28.37	19.10	41.85	19.58	27.59	33.42	25.46
• Brightness (Traditional aug.)		13.08	21.53	18.18	21.98	28.58	26.12	32.29	30.67	17.75	39.47	16.92	24.05	34.25	24.99
• ControlNet (Only Snow aug.)	✗	1.81	2.72	4.08	5.02	12.18	13.18	16.51	11.40	2.68	33.23	2.69	6.52	12.53	9.58
• ControlNet (3 scenarios aug.)	✗	0.00	0.00	0.27	1.49	1.76	1.86	5.16	5.29	0.59	17.31	1.29	7.63	6.00	3.74
• ControlNet (6 scenarios aug.)	✗	0.00	0.00	0.00	1.83	1.05	0.40	1.44	0.53	0.45	3.47	0.38	2.17	1.52	1.02
• Freecontrol (Only Snow aug.)	✓	11.09	19.54	14.50	15.84	18.97	19.18	29.09	19.34	13.10	32.39	13.07	24.27	34.87	20.40
• Freecontrol (3 scenarios aug.)	✓	14.22	18.36	15.44	11.34	14.79	12.63	15.89	14.80	11.07	22.18	12.62	19.93	21.82	15.78
• Freecontrol (6 scenarios aug.)	✓	13.30	19.87	13.29	19.34	18.40	16.58	15.59	13.14	13.22	21.52	15.49	20.72	23.59	17.24
• DriveGEN (Only Snow aug.)	✓	15.53	24.35	21.67	29.67	33.88	35.40	**37.81**	32.55	23.75	**43.15**	25.75	34.80	41.17	30.73
• DriveGEN (3 scenarios aug.)	✓	18.12	29.12	25.30	33.18	36.06	35.28	36.08	33.67	26.57	41.40	27.03	39.13	**41.57**	32.50
• DriveGEN (6 scenarios aug.)	✓	**21.78**	**29.21**	**27.16**	**34.57**	**36.52**	**35.91**	34.70	**34.91**	**29.74**	39.94	**31.94**	**40.06**	40.69	**33.63**
Pedestrian, IoU @ 0.7, 0.5, 0.5															
Monoflex [63]		0.98	3.79	0.81	7.91	17.25	14.60	13.23	9.09	4.99	19.51	5.34	1.63	8.64	8.29
• Color Jitter (Traditional aug.)		0.88	3.59	1.61	9.86	14.48	12.12	14.86	11.94	7.51	18.39	9.29	0.97	11.11	8.97
• Brightness (Traditional aug.)		0.79	1.88	1.08	4.80	13.74	12.14	7.54	5.85	1.70	16.75	2.57	0.62	3.69	5.62
• ControlNet (Only Snow aug.)	✗	0.00	0.00	0.00	1.74	7.92	5.23	3.62	3.52	1.41	10.79	1.54	0.00	0.66	2.80
• ControlNet (3 scenarios aug.)	✗	0.00	0.00	0.00	1.19	2.78	1.15	1.67	1.32	0.83	4.11	2.05	0.00	0.00	1.16
• ControlNet (6 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00	2.50	0.00	0.00	0.00	0.22
• Freecontrol (Only Snow aug.)	✓	3.87	5.18	3.52	3.20	5.07	5.66	4.75	3.44	3.94	7.42	3.77	8.07	7.44	5.03
• Freecontrol (3 scenarios aug.)	✓	2.69	3.74	3.06	3.07	5.01	4.22	7.44	4.62	7.18	10.45	6.07	6.82	7.28	5.51
• Freecontrol (6 scenarios aug.)	✓	5.17	8.71	8.56	4.18	6.82	5.08	6.98	7.18	6.35	10.19	9.18	9.45	9.32	7.47
• DriveGEN (Only Snow aug.)	✓	1.10	3.17	4.32	16.14	**19.36**	**20.31**	**19.45**	14.44	8.77	**24.07**	8.86	9.59	17.64	12.86
• DriveGEN (3 scenarios aug.)	✓	5.60	**9.66**	**9.01**	14.73	18.29	16.45	15.30	**15.81**	15.30	21.73	16.81	13.84	**17.73**	14.63
• DriveGEN (6 scenarios aug.)	✓	**6.44**	9.32	7.10	**17.46**	19.25	19.89	16.89	15.51	**15.50**	23.48	**16.64**	**14.37**	17.12	**15.31**
MonoGround [31]		2.81	3.40	5.76	17.46	18.40	17.40	12.64	9.06	4.01	23.33	5.51	3.08	7.06	9.99
• Color Jitter (Traditional aug.)		2.44	3.35	3.67	14.85	18.04	15.69	14.63	12.06	9.01	24.26	9.34	2.06	7.76	10.55
• Brightness (Traditional aug.)		2.85	4.23	7.50	13.84	13.78	13.96	11.52	12.34	5.60	20.39	5.27	2.25	10.40	9.53
• ControlNet (Only Snow aug.)	✗	1.88	1.03	0.80	7.44	9.68	8.30	0.91	2.93	1.25	12.78	1.42	0.26	0.96	3.82
• ControlNet (3 scenarios aug.)	✗	0.00	0.00	0.00	3.32	3.85	2.32	1.17	1.05	1.50	4.69	1.25	0.26	0.77	1.55
• ControlNet (6 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19
• Freecontrol (Only Snow aug.)	✓	**9.63**	12.34	10.78	14.08	12.11	12.79	15.08	12.35	7.72	15.06	11.47	15.99	14.16	12.58
• Freecontrol (3 scenarios aug.)	✓	0.82	1.80	1.84	2.30	3.14	3.77	3.75	0.40	2.84	3.73	3.43	3.78	5.49	2.85
• Freecontrol (6 scenarios aug.)	✓	6.48	7.49	7.75	5.51	5.44	4.85	2.72	1.36	4.16	8.23	4.88	9.02	6.30	5.71
• DriveGEN (Only Snow aug.)	✓	6.11	6.69	8.96	14.75	16.80	18.38	16.23	12.99	9.64	22.73	11.25	11.78	14.07	13.11
• DriveGEN (3 scenarios aug.)	✓	6.53	9.20	10.90	15.78	**19.10**	**20.51**	**18.01**	15.79	12.98	**24.46**	13.92	19.26	**19.29**	15.83
• DriveGEN (6 scenarios aug.)	✓	9.49	**13.28**	**13.02**	**16.33**	17.10	19.67	16.70	**18.02**	**15.35**	22.42	**14.99**	**19.72**	17.80	**16.45**
Cyclist, IoU @ 0.7, 0.5, 0.5															
Monoflex [63]		0.48	1.59	0.53	2.05	6.41	7.45	**9.93**	**8.61**	4.70	**14.08**	3.54	2.80	7.82	5.38
• Color Jitter (Traditional aug.)		0.52	2.45	**1.51**	1.29	2.79	6.36	8.72	8.08	3.55	10.00	3.62	0.84	7.92	4.43
• Brightness (Traditional aug.)		0.16	0.83	0.19	1.26	2.63	4.81	7.67	6.91	3.64	10.67	4.30	1.68	6.22	3.92
• ControlNet (Only Snow aug.)	✗	0.00	0.26	0.00	0.00	3.13	3.20	5.95	4.85	5.67	12.72	4.88	1.67	1.68	3.39
• ControlNet (3 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.50	1.50	1.58	1.35	0.00	0.00	0.53
• ControlNet (6 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
• Freecontrol (Only Snow aug.)	✓	1.13	3.22	1.34	0.00	0.39	0.67	2.89	1.06	0.29	4.07	0.33	0.65	6.16	1.71
• Freecontrol (3 scenarios aug.)	✓	0.00	0.00	0.00	0.00	2.50	0.00	0.91	1.04	1.43	3.03	1.33	0.00	2.05	0.95
• Freecontrol (6 scenarios aug.)	✓	0.20	0.18	0.22	0.00	0.31	0.00	0.52	0.52	0.76	1.61	1.82	0.60	0.61	0.57
• DriveGEN (Only Snow aug.)	✓	0.43	0.99	0.50	1.02	4.09	4.70	8.48	5.43	2.69	12.12	3.35	1.42	8.52	4.13
• DriveGEN (3 scenarios aug.)	✓	**0.87**	**2.70**	1.07	1.66	4.01	5.98	6.22	2.87	3.72	10.14	3.53	3.59	8.40	4.21
• DriveGEN (6 scenarios aug.)	✓	0.50	1.62	0.64	**3.06**	**6.00**	**7.73**	7.46	8.11	**8.14**	11.55	**9.11**	**4.16**	**8.82**	**5.92**
MonoGround [31]		0.13	1.39	1.01	0.63	2.51	1.82	3.90	2.81	0.66	8.91	0.90	1.84	4.33	2.37
• Color Jitter (Traditional aug.)		0.30	**2.00**	**1.59**	0.25	1.60	1.73	4.00	3.60	0.74	**10.33**	1.30	2.19	3.96	2.58
• Brightness (Traditional aug.)		0.04	0.47	0.17	0.36	1.05	0.84	3.67	1.69	1.12	5.30	0.55	0.67	2.15	1.39
• ControlNet (Only Snow aug.)	✗	0.00	0.00	0.67	0.00	0.56	1.25	0.39	0.44	0.12	5.12	0.28	0.37	0.51	0.75
• ControlNet (3 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.30	0.00	0.83	1.50	0.00	1.22	0.00	0.31	0.00	0.32
• ControlNet (6 scenarios aug.)	✗	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
• Freecontrol (Only Snow aug.)	✓	0.33	0.42	0.29	1.07	0.16	0.26	1.42	0.63	0.20	3.59	0.17	1.34	3.65	1.04
• Freecontrol (3 scenarios aug.)	✓	0.14	0.37	0.32	0.00	0.37	0.50	0.37	0.33	0.58	1.09	0.42	0.27	0.29	0.39
• Freecontrol (6 scenarios aug.)	✓	0.00	0.32	0.65	0.00	0.42	0.54	0.00	0.00	1.00	1.70	0.83	1.25	0.74	0.57
• DriveGEN (Only Snow aug.)	✓	0.10	0.54	0.29	0.31	1.45	2.89	2.92	1.48	0.74	6.44	1.04	1.39	3.66	1.79
• DriveGEN (3 scenarios aug.)	✓	**0.43**	1.41	1.12	0.80	2.37	2.10	4.13	4.12	0.78	7.71	0.85	**3.22**	**4.72**	2.60
• DriveGEN (6 scenarios aug.)	✓	1.17	1.52	1.25	**2.88**	**4.62**	**4.25**	**4.54**	**4.37**	**1.79**	7.43	**2.65**	2.63	4.17	**3.33**