Transformation and Augmentation for Improving Perception of LiDAR Point Cloud in Adverse Conditions

Yuxiao Zhang

Abstract

With the advancements in future society technologies such as intelligent vehicles, smart cities, and the Internet of Things (IoT), the performance issues linked to environmental perception remain a substantial barrier to their widespread deployment. This thesis aims at advancing the perception ability of LiDAR in the realm of adverse conditions, through transformation and augmentation of LiDAR point cloud data.

Central to this research is the recognition that adverse conditions such as rain, fog, snow, and contamination significantly affect the performance of exteroceptive sensors. These conditions can distort signal reception, leading to critical safety challenges by reducing the efficacy of object detection and increasing the risk of false detections. To address these challenges, the thesis presents two main approaches: a deep-learning architecture for transforming adverse condition LiDAR point clouds into clear condition data and a novel data augmentation model to expand LiDAR datasets with simulated adverse effects.

The transformation aspect is embodied in the development of the 'L-DIG' model, a deep-learning framework that employs unpaired data and depth-image priors to accurately remove the noise caused by adverse effects while keeping structure integrity. Innovative structures for both the generator and discriminator within the Generative Adversarial Network (GAN) have been specially tailored, incorporating specific loss functions and training techniques. This customization is aimed at establishing an understanding of the scene and optimizing the transformation performance of the model to its fullest potential. The model effectively removes adverse effects, whether they are in scattered or clustered forms, and additionally compensates for occluded objects, thereby enhancing the overall quality of the point cloud. Complementing this, a designated 3D clustering algorithm has been developed to facilitate a quantifiable assessment of these effects, enabling the generation of precise conditional guides for further data augmentation as well.

Furthermore, this thesis addresses the critical gap in existing datasets, which often lack diversity in terms of adverse environmental conditions. The classification of

adverse effects is achieved through the creation of segmentation maps, which are based on cluster units. These maps then function as conditional guides for the generative models. By employing advanced data fusion techniques, the thesis demonstrates the ability to replicate and augment adverse conditions in a natural way within LiDAR point clouds, significantly improving the training and performance of autonomous driving detectors. This includes the deployment of unique data fusion methods to generate paired datasets that effectively bridge significant domain gaps, such as those found in traffic layouts and environmental contexts.

Through extensive testing and validation across varied datasets, the thesis demonstrates that LiDAR-based perception can be significantly improved to detect and classify objects with greater accuracy, even in the presence of adverse conditions. This leads to the conclusion that LiDAR sensors, when equipped with the advanced processing methods developed in this research, can achieve the levels of precision necessary for safe perception.

This thesis contributes to the autonomous driving field by providing novel methods that enable a clearer assessment of the potential of LiDAR sensors for environment perception, identifying current limitations, and paving the way for future sensor and algorithm development to overcome these challenges. The resulting advancements from this research hold the promise of making sensor-based intelligent modalities more reliable and safer in all conditions, thereby moving closer to the goal of widespread adoption.

Acknowledgments

This research was supported by Nagoya University and JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2120.

I would like to express my sincere gratitude to all the professors who have provided valuable instruction and guidance throughout my Ph.D. studies: Prof. Kazuya Takeda, Ming Ding, Kento Ohtani, Alexander Carballo, and Keisuke Fujii. My thanks also go to the secretaries of Takeda Lab for their unconditional support and assistance during my entire study period, especially Chika Ando-san. During the application stage, Ando-san was immensely helpful in managing overseas matters, despite the 14-hour time difference. Upon my arrival at Nagoya University, the secretaries, including Isobe-san, Ienaga-san, and Ikai-san, assisted me in settling down in the lab with patience and kindness, which was incredibly meaningful to me.

As an interdisciplinary student, I feel fortunate to have close friends (Hanting, Yingjie, and Yan), from whom I have learned so much about technical knowledge. I also had a great time with my lab mates (Maoning, Kuribayashi-kun, Kyle, Pat, Jacob, Satomi, Bito-san, Nakahara-san). Life in Nagoya becomes colorful and easier with you guys.

Having lived outside of my homeland for over six years since my Master's study, half of which was during the pandemic era in three different countries across the globe, I am thankful for the understanding and support of my parents and family. Endeavoring to climb the summit of Ph.D. has been a precious experience in my life, one that has helped shape who I am today.

Finally, I extend my deepest gratitude to the significant other in my life, Qi Sun, who has stood by my side all the time. I could not have navigated the darkest days of this journey without her unwavering support and encouragement and I will always be grateful.

Contents

List of Figures

List of Tables

List of Acronyms

ADAS . . . . . Advanced Driver-Assistance System

ADS . . . . . Autonomous Driving Systems

AP . . . . . Average Precision

AV . . . . . Autonomous Vehicle

BEV . . . . . Bird's Eye View

BPNN . . . . . Back-Propagation Neural Network

CADC . . . . . Canadian Adverse Driving Conditions

CMOS . . . . . Complementary Metal Oxide Semiconductor

CNN . . . . . Convolutional Neural Network

CUT . . . . . Contrastive Unpaired Translation

DBI . . . . . Davies-Bouldin Index

DBSCAN . . . . . Density-Based Spatial Clustering of Applications with Noise

DNN . . . . . Deep Neural Network

DROR . . . . . Dynamic Radius Outlier Removal

DSOR . . . . . Dynamic Statistical Outlier Removal

FN . . . . . False Negative

FOV . . . . . Field of View

FP . . . . . False Positive

GAN . . . . . Generative Adversarial Networks

HOG . . . . . Histogram of Oriented Gradient

ICD . . . . . Inter-Cluster Distances

IoT . . . . . Internet of Things

IV . . . . . Intelligent Vehicle

JARI . . . . . Japan Automobile Research Institute

L-DIG . . . . . LiDAR Depth Images GAN

LiDAR . . . . . Light Detection And Ranging

MPE . . . . . Maximum Permissible Exposure

OPTICS . . . . . Ordering Points To Identify the Clustering Structure

PCA . . . . . Principal Component Analysis

PGM . . . . . Polar Grid Map

R-CNN . . . . . Region-based Convolutional Neural Network

RGB . . . . . Red Green and Blue

ROR . . . . . Radius Outlier Removal

RSU . . . . . Road Side Unit

SOR . . . . . Statistical Outlier Removal Filter

SOTA . . . . . State-of-the-art

SNR . . . . . Signal-to-Noise Ratio

SSIM . . . . . Structural Similarity Index

TOF . . . . . Time of Flight

TP . . . . . True Positive

UAV . . . . . Unmanned Aerial Vehicles

V2I . . . . . Vehicle-to-Infrastructure

V2V . . . . . Vehicle-to-Vehicle

V2X . . . . . Vehicle-to-Everything

YOLO . . . . . You Only Look Once

Chapter 1

Introduction

Perception and sensing are at the core of future society, including intelligent vehicles, smart cities, and the Internet of Things (IoT) [3], etc. Instruments employed for the detection and interpretation of ambient conditions are perception sensors, commonly involving LiDAR (Light Detection And Ranging), camera, radar, ultrasonic, and so on. One example of comprehensive integration of perception sensors is Intelligent Vehicles (IVs), or Autonomous Vehicles (AVs), as shown in Fig. 1.1, which are capable of sensing their external environment and navigating securely with minimal or no human intervention. These vehicles have the potential to fundamentally transform the transportation of people and goods, offering significant societal benefits for the future [4] [5]. However, the challenges associated with sensor perception are still of concern under adverse conditions such as precipitation, inclement weather, seasonal variations, and secondary adverse effects influenced by man-made structures. To facilitate the evolution of intelligent society, it is imperative to address the pressing issue of enhancing perception and sensing capabilities in these adverse conditions [6].

Figure 1.1: The Toyota Prius test AV from Nagoya University. LiDAR sensors, and cameras, alongside other sensors, are bolted on a plate mounted firmly on top of the car.

1.1 Adverse Condition Influences on Perception Sensors

Adverse environmental conditions manifest predominantly through weather phenomena. On average, global precipitation occurs 11.0% of the time [7]. Factors like water wetness can absorb and disperse sensor signals, while atmospheric conditions such as fog and haze reduce visibility. Obstructions like snowflakes and sandstorms not only impede the transmission path of signals but can also result in false detections [8]. Secondary challenges arise from intense sunlight, reflections from tall structures, extreme temperatures, and sensor hardware contamination, each contributing to unpredictable or deleterious effects on sensor perception abilities.

In general, the degradation of perception is the main limitation of perception sensors in adverse conditions. A table and a radar chart are summarized to more intuitively show the strengths and weaknesses of each sensor in adverse conditions, as in Table 1.1 and Fig. 1.2. It is worth noticing that level 3 influences (moderate), that cause perception error up to 30% of the time in this table, could also mean up

to 30% of the LiDAR point cloud is affected by noise, or up to 30% of the pixels in the camera images are affected by distortion or obscure. The same applies to level 4 influences (serious), as well.

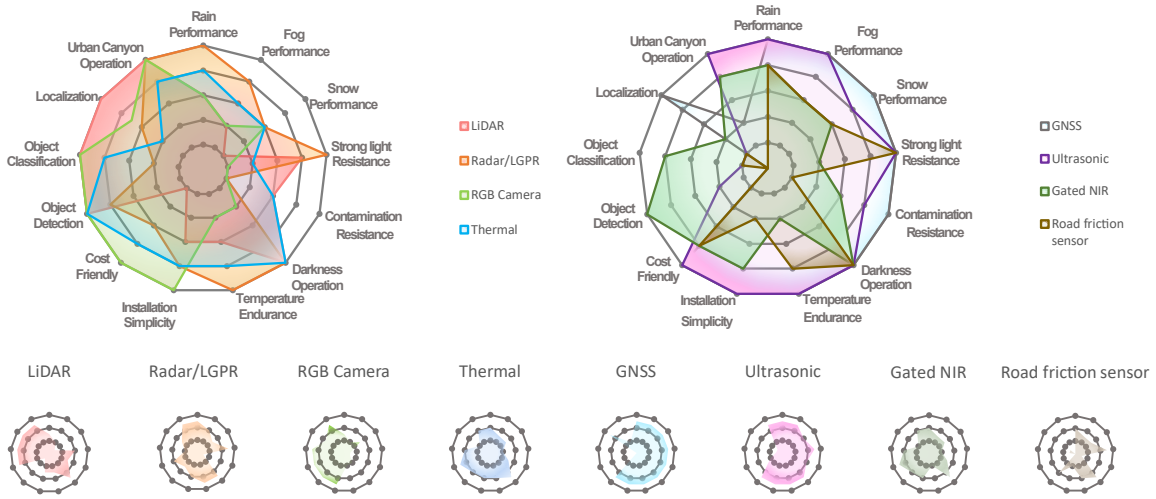


Figure 1.2: Sensor performance and characteristics radar map.

To comprehensively understand the impact of adverse conditions on sensor performance in detail, this section will individually introduce the major types of perception sensors currently deployed in AVs and discuss the specific challenges that adverse conditions pose for each.

1.1.1 LiDAR

LiDAR is one of the core perception sensors in the autonomous driving field. The use of 3D LiDAR on cars hasn't exceeded much more than a decade and has already demonstrated its indispensability in the Advanced Driver-Assistance System (ADAS) with high measurement accuracy and illumination independent sensing capabilities [9]. This 3D laser scanning technology has some key attributes: measurement range, measurement accuracy, point density, scan speed and configuration ability, wavelength, robustness to environmental changes, form factor, and cost [10]. LiDAR's measurement range, measurement accuracy, and point density are among the key factors that could be interfered with by adverse conditions. Researchers have

Table 1.1: The influence level of various weather conditions on sensors

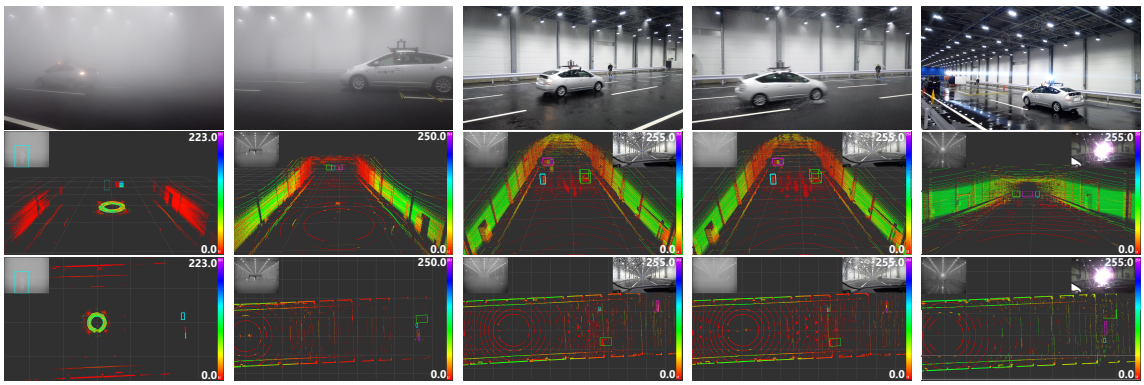
Modality	Light rain <4mm/hr	Heavy rain >25mm/hr	Dense smoke /Mist vis<0.1km	Fog vis<0.5km	Haze /Smog vis>2km	Snow	Strong light	Contamination (over emitter)	Operating Temperature (℃)	Installation complexity	Cost
LiDAR (λ 850-950nm and 1550nm)	2	3	5	4	1	5	2	3	-20 +60	easy	high
Radar (24, 77 and 122 GHz)	0	1	2	0	0	2	0	2	-40 to +125	easy	medium
Camera (RGB, Stereo, Fisheye)	3	4	5	4	3	2 (dynamic) 3 (static)	5	5	-20 to +40	easiest	lowest
Gated NIR Camera (λ 800-950nm)	2	3	2	1	0	2	4	3	normally 0 to +65 for InGaAs cameras	easy	low
Thermal FIR Camera* (λ 2-10μm)	2	3	3	1	0	2	4	3	-40 to +60	easy	low
Road-friction sensor (infrared)	2	3	3	2	1	2	1	5	-40 to +60	medium	low

The effect level each phenomenon causes to sensors:
0 - negligible: almost negligible
1 - minor: barely cause detection error
2 - slight: cause small errors on special occasions
3 - moderate: cause perception error up to 30% of the time
4 - serious: cause perception error more than 30% but lower than 50% of the time
5 - severe: causes false detection or detection failure
*Considered to be installed outside of cabin and without glass housings

done tests and validations on LiDAR under adverse weather conditions [11] in artificial environments like fog chambers [10], real-world snowfields [12], or simulation environments [13].

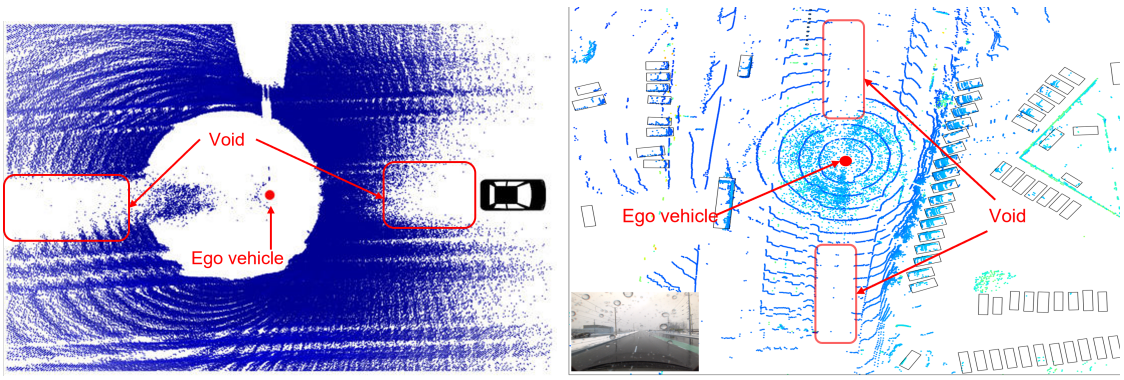
Drizzling and light rain don't affect LiDAR functions very much on small aperture LiDAR sensors because the power attenuation due to scattering by direct interaction between laser beam and raindrops of comparable is almost negligible [13]. However, rains with a high and non-uniform precipitation rate would most likely form lumps of agglomerate fog and create fake obstacles to the LiDARs. Studies proved that the signal reflection intensity drops significantly at a rain rate of 40 mm/hr and 95 mm/hr [14]. Upon examining the LIBRE Dataset [10] [15] collected from the Japan Automobile Research Institute's (JARI) weather experimental facility, the LiDAR point clouds depicted in Fig. 1.3 reveal discouraging performance under conditions such as fog, rain, and wet surfaces. During the fog test, human presence was only detectable by the LiDAR at a distance of 13 meters in the dense setting, with an insufficient number of points for reliable recognition. In a less dense setting, detection was possible from 47 meters. In rain tests, objects were discernible at a distance of 24 meters from the LiDAR, with varying levels of noise attributable to different rain settings. Additionally, as evidenced in Fig. 1.3e, strong light sources exert a negative effect on LiDAR data. Specifically, when the vehicle was positioned 40 meters away from a Xenon light source with a peak illuminance of 200 klx, object detection was largely impaired, resulting in an unclassifiable dark region.

Different from rain, snow consists of solid objects, snowflakes, and could easily shape themselves into much larger solid objects and become obstacles that either cause false detection of LiDAR or block the line of sight for useful detection. Fig. 1.4a shows the point cloud of accumulating multiple 3D scans as the ego vehicle moved behind the preceding vehicle. The turbulent snow caused by car motions creates significant snow clusters beside the ego vehicle and voids in the front and back view in the point cloud. Similar clusters and voids both in front of and behind the ego vehicle can also be observed in the Canadian adverse driving conditions (CADC) dataset [16], as shown in Fig. 1.4b. Additionally, a common LiDAR's designed lowest operating



(a) Dense fog (b) Light fog (c) Rain 30 mm/hr (d) Rain 80 mm/hr (e) Strong light

Figure 1.3: Adverse condition influences on LiDAR point clouds. Top row depicts sample conditions, middle and bottom rows show the 3D LiDAR point cloud, thermal camera image and RGB camera image (not available for fog experiments), targets of interest (human/mannequin, car, and reflective targets) are highlighted. (a) dense fog with visibility of 17 m. (b) light fog with visibility 162 m. (c) rainfall setting of 30 mm/hr and average humidity of 89.5 %. (d) rainfall setting of 80 mm/hr and average humidity of 93 %. (e) strong light at 200 klx at 155 A.



(a) Ouster OS1-64 point cloud in snow swirl. Snow clusters and voids are observed [12].

(b) Snow swirl effects in CADC Dataset [16]. Points color depth indicates intensities.

Figure 1.4: LiDAR point clouds with swirl effects in snow conditions.

temperature is $-10°C$, which brings further challenges in cold environments. When the temperature change is at a large scale, such as from an extremely cold ($-20°C$) to an extremely hot ($+60°C$) environment, the time delay of LiDAR measurement would increase about 6.8 ns, which widens the LiDAR ranging by over a 1 meter and lowers the precision at near field [17].

Sandstorms and smog, as rare as they might appear, could be more serious problems because particles from road dirt attached to the outer surface of the emitter window could worsen the LiDAR signal attenuation [18]. Tests with near-homogeneous dust particles being distributed on the surface of a scanner show a 75% reduction in LiDAR maximum range [19]. Such adverse conditions are not exclusive to terrestrial settings; they also present significant challenges for airborne and space-based LiDAR sensors. Systems like the CALIPSO high spectral resolution LiDAR and single-photon LiDAR are commonly employed in satellite and aircraft platforms for atmospheric monitoring of Earth and 3D terrain mapping [20]. These applications necessitate the capability to penetrate atmospheric obstructions such as haze and sandstorms [21]. Future aerial LiDARs and Unmanned Aerial Vehicles (UAVs) are facing additional weather challenges including wind-affected and time-varying refractive gradients produced by atmosphere turbulence, which lead to scintillation, beam spreading, and wander [22]. The particular effect of such adversarial conditions on aerial LiDARs and UAVs hasn't been studied in a methodically way as they are still at the stage of developing in the autonomous driving area, but it's safe to assume that they are going to need to overcome this problem to be able to serve the intelligent transportation system under hazy and turbulent conditions in the future.

Currently, the majority of the market uses 905 nm wavelength LiDAR deployment. while others argue that 1550 nm LiDAR can overcome adverse conditions because higher optical power is allowed to be emitted at this wavelength [23]. There are two critical design considerations in LiDAR selection: eye safety and ambient suppression. Most civilian or commercial LiDARs are used in an environment where human eyes are exposed, so the infrared laser of LiDAR must not exceed the maximum permissible exposure (MPE) or cause any damage to retinas, according to the international laser

product safety standard (IEC 60825-1:2014) class 1 [24]. Therefore, the selection of laser wavelength is pretty much narrowed down to two ranges: 800 nm - 1000 nm and 1300 nm - 1600 nm. That's why current LiDARs made for AVs have the selection of 850 nm, 905 nm, and 1550 nm wavelengths, and they also all fall into the window of low solar irradiance, which helps suppress the ambient light for the signal receiver with a lower signal-to-noise ratio (SNR) [25]. A water extinction coefficient chart is plotted as shown in Fig. 1.5 in order to show that the 1550 nm wavelength is more likely to be absorbed by water. The extinction coefficient α_{water} is also known as the Lambert absorption coefficient, which is acquired from:

$$\alpha(\lambda) = 4\pi k(\lambda)/\lambda \tag{1.1}$$

in which λ is wavelength and $k(\lambda)$ is the extinction coefficient of water at 25°C. The detail of the acquisition of $k(\lambda)$ can be found in [26]. Therefore, a 1550 nm laser can be largely absorbed in the crystalline lens or the vitreous body of an eye so more energy is allowed than 905 nm, which seems to be a good thing considering the power attenuation predicament in weather [27].

However, based on the research of Wojtanowski et al. [28] on the comparison of 905 nm and 1550 nm performance deterioration due to adverse environmental conditions, 905 nm reaches two times further than 1550 nm in a rain rate of 25 mm/hr. There are opinions arguing that light propagation at 1550 nm might suffer less attenuation than at shorter wavelengths, but Kim et al. [29] suggested that this rule only applies to haze condition (visibility $>$ 2 km), while in fog (visibility $<$ 500 m) the attenuation is independent of wavelength and 905 nm still measures 60% longer than 1550 nm. What's more, 1550 nm waves have approximately 97% worse reflectance in snow compared to 905 nm [30]. Less interference from snow doesn't make up for the insufficiency in object detection. Still, 1550 nm has great potential in further solid-state LiDAR development and better compatibility with Complementary Metal Oxide Semiconductor (CMOS) technology.

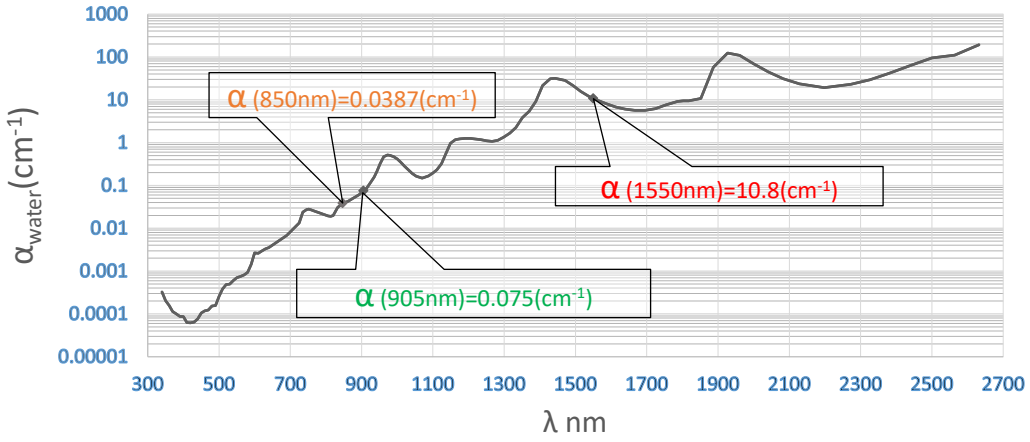


Figure 1.5: Water extinction coefficient spectrum. Laser energy absorption by water of 1550 nm is over 100 times larger than that of 905 nm. [1] [2]

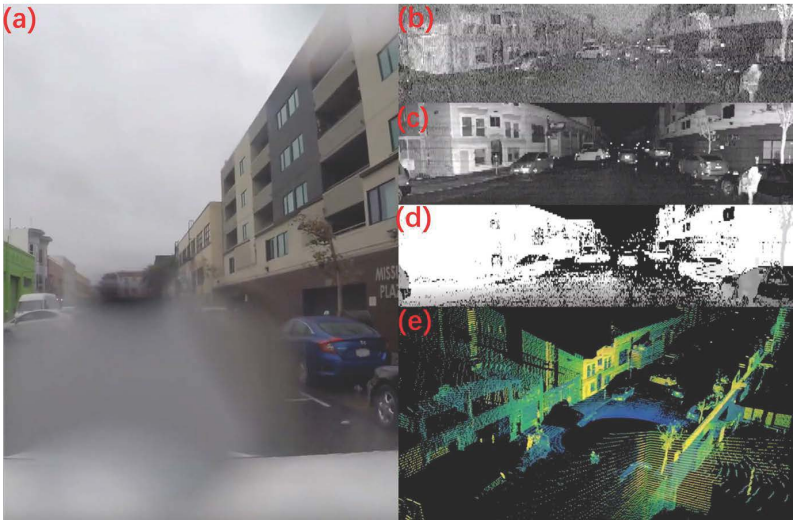


Figure 1.6: Camera vs LiDAR in rain condition. (a) camera perspective; (b)intensity; (c) reflectivity; (d) noise; (e) 3D point cloud colored by intensity. Image courtesy of Ms. Kim Xie, Ouster Inc.

1.1.2 Camera

Camera is one of the widest-used sensors in perception tasks, while also one of the most vulnerable in adverse conditions. Unlike LiDAR's relative robustness against mild wetness, a camera in precipitation, regardless of how high resolution, can be easily incapacitated by a single water drop on the emitter or lens, as shown in the comparison of Fig. 1.6. The blockage and distortion in the image would instantly make the system lose the sense of input data and fail to process correctly. Water mists induced by heavy precipitation conditions, based on their density, would create near-homogeneous blockages which is a direct deprivation of information to cameras. Study shows that camera-only perception suffers up to 40% rise in miss rate in night or fog conditions [31].

Winter conditions like snow could affect the camera when the snowflakes touch the lens or the camera's optical window and melt into ice-slurry immediately. What's worse, those ice water mixtures might freeze up again quickly in low temperatures and form an opaque blockage. Heavy snow or hail could fluctuate the image intensity and obscure the edges of the pattern of a certain object in the image or video which leads to detection failure [11]. Besides the dynamic influence, snow can extend itself to a static phenomenon by accumulating on the surface of the earth and blocking road marks or lane lines. Under such situations, the acquisition of data sources is compromised for cameras, and the process of perception would be interrupted at the very beginning. In another area, too high an illumination such as direct sun glare, skyscrapers' light pollution, bright beam light of other cars, and tunnel exits can degrade the visibility of a camera down to almost zero and make the camera exposure selection a difficult task [32].

1.1.3 Radar and Ultrasonic

Automotive radar typically operates at bands between 24 GHz and 77 GHz which are known as mm-wave frequencies, while some on-chip radar also operates at 122 GHz. Radar can be used in the detection of objects and obstacles like in the parking

assistance system, and also in detecting positions, and speed relative to the leading vehicle as in the adaptive cruise control system [33]. Radar seems to be more resilient in weather conditions. In order to intuitively see the difference, a chart of electromagnetic power attenuation in different precipitation rates is plotted [34] [35]. From Fig. 1.7, it can be observed that the attenuation for radar at 77 GHz is at the level of 10 dB/km in a 25 mm/hr heavy rain, while 905 nm LiDAR's attenuation is about 35 dB/km under the same visibility below 0.5 km condition [36] [37]. In snowfall, a higher snow rate yields larger attenuation, and wet snow shows higher attenuation because of the higher water absorption and larger snowflakes [38]. Given that a snowfall with 10 mm/hr has quite low visibility ($<$ 0.1 km) [39], the specific attenuation for a 77 GHz radar in a 10 mm/hr snow is estimated at about 6 dB/km.

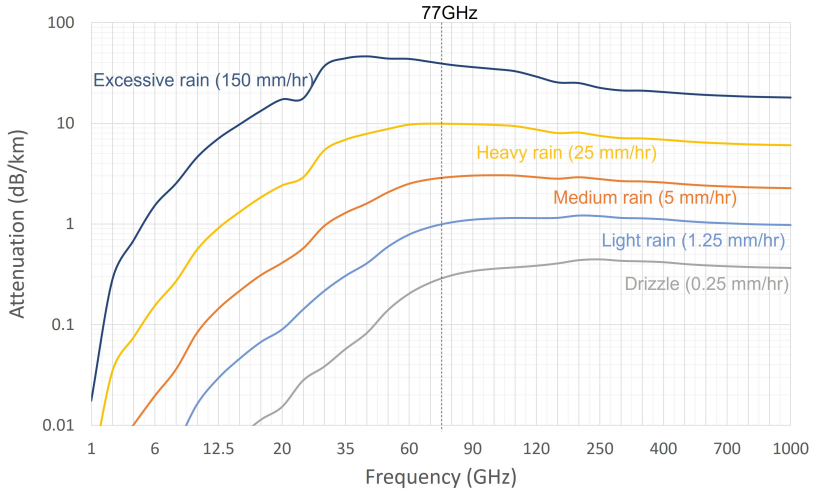


Figure 1.7: Electromagnetic power attenuation vs frequency in different precipitation rates [34] [35].

Operating with a similar principle at ultrasound band, around 40 to 70 kHz, ultrasonic sensors commonly serve as parking assisting sensors and blindspot monitors at close ranges [40]. The speed of sound traveling in air is affected by air pressure, humidity, and temperature [41], causing concern about accuracy fluctuation. However, the return signal of an ultrasonic wave does not get decreased due to the target's dark color or low reflectivity, so it's more reliable in low visibility environments than cameras, such as high-glare or shaded areas beneath an overpass.

No doubt that radar and ultrasonic sensors are objectively better adaptive to wet weather, but when compared with LiDAR, they often receive criticism for their insufficient ability in pedestrian detection, object shape, and size information classification due to low spatial resolution. Nevertheless, they still hold an important position in potential sensor fusion configurations in terms of adverse condition solutions.

1.2 Research Goal and Approaches

Acknowledging the influences that adverse conditions impose on perception sensors, it becomes clear that the degradation of perception performance and reliability under adverse conditions is the fundamental problem that the perception sensors are facing, and it is the goal of this research to enhance the perception capabilities of LiDAR point clouds in such environments. Drawing from the comprehensive tests and validations conducted by Mercedes-Benz AG on commercial autonomous driving detectors [42] [43] [44], it has been demonstrated that involving and enriching adverse effects in training datasets can raise the average precision of detectors by at least 2.1% compared to those trained solely on clear baselines. Furthermore, the availability of paired data encompassing adverse effects and classifications in the training datasets significantly boosts the detectors' ability to capture smaller elements in adverse driving scenarios, such as pedestrians and cyclists [45]. Consequently, the core objective of this thesis is to advance LiDAR point cloud perception in the realm of adverse conditions. This will be achieved by strategically involving and enriching such conditions in datasets, utilizing learning-based approaches and point cloud data manipulation techniques, and concurrently enabling the generation of paired adverse conditions datasets. As shown in the flow diagram of Fig. 1.8, research approaches mainly contain two aspects: transforming point cloud with adverse effects into clear conditions; and expanding adverse data by LiDAR point cloud augmentation. Though these approaches process LiDAR point clouds from divergent perspectives, both facilitate the generation of paired datasets that include adverse conditions. This, in turn, significantly contributes to the enhancement of perception model training and

the establishment of robustness.

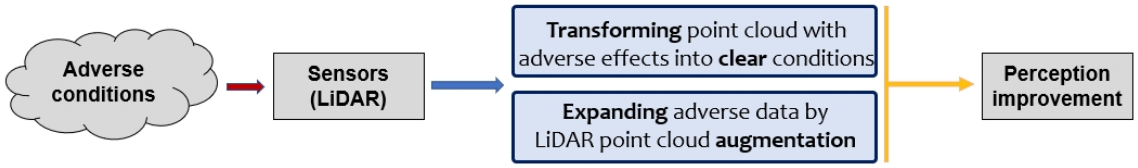


Figure 1.8: The flow diagram of this thesis. Upon acknowledging the influences that adverse conditions have on LiDAR sensors, two approaches are employed: transforming adverse point clouds into clear and expanding adverse data through augmentation to achieve the research target of perception improvement.

Among adverse conditions, rain, fog, snow, and contamination exhibit the closest associations with perception sensor performance. As indicated in Table 1.1 and Fig. 1.2, snow represents one of the most significant challenges to sensor fidelity. Notably, snow conditions exhibit the greatest overlap with other adverse environmental factors. False or inaccurate detection induced by elements such as water droplets, surface wetness, and aggregated water mist can all be found under snowfall and snow swirl conditions. Moreover, snow presents substantial difficulties in obtaining paired data, largely because common experimental facilities are not outfitted with artificial snow generators appropriate for driving scenario tests. This situates snow conditions as a suitable representation of perception challenges in adverse settings. In this thesis, snow conditions under driving scenarios are selected in LiDAR point clouds as the major research subject.

1.3 Thesis Structure and Contributions

The remainder of this thesis is structured into five parts. Chapter 2 introduces prior research related to the topics of this research. Chapter 3 discusses 3D point cloud analysis and advanced adaptive filter explorations, and also explains the methodology of transforming LiDAR point clouds from adverse to clear conditions with experiment results presented. Chapter 4 focuses on the details of the expansion of datasets under adverse conditions through conditional point cloud augmentation models. Finally, Chapter 5 provides a comprehensive conclusion of the research conducted in this

thesis and outlines directions for future work.

The main contributions of this thesis are as follows:

- Holistic presentation of the impacts on perception sensors arising from or circumstantially influenced by adverse conditions in Sec. 1.1.
- A deep-learning architecture for the transformation of LiDAR point clouds adverse effects into clear ones, built upon unpaired data and depth-image priors, supplemented by new loss functions to preserve scene integrity is presented in Sec. 3.3.
- Advanced feature modeling and quantifiable assessment of adverse effects in LiDAR point clouds, facilitated by a custom 3D clustering algorithm presented in Sec. 3.4.
- Deployment of a conditional generative model proficient in synthesizing natural adverse effects with an in-depth understanding of driving scenarios, enabled by novel data fusion techniques, presented in Sec. 4.2.1.
- A method of employing cluster-based segmentation maps of adverse scenes to serve as conditional guides in the generative model, presented in Sec. 4.2.2.
- The attainment of robustness against significant domain gaps, particularly in the context of scene layouts, presented in Sec. 4.3.
- Enabling the generation of paired datasets incorporating adverse effects, thereby enriching the training data available for enhancing perception models, presented in Sec. 4.3.

Chapter 2

Related Works

The enhancement of LiDAR perception largely relies upon the manipulation of point cloud data to elevate data quality. Initially, the focus of academic inquiry highly centered on the task of noise removal within point clouds, giving rise to a huge emergence of conventional outlier-removal-based filters. Evolving from 2D to 3D spatial contexts, an array of adaptive parameter configurations have been deployed to improve both the accuracy and recall of noise elimination. However, this trajectory of advancement plateaued once all geometric and intensity metrics were exhaustively explored, plus the quest for an optimal balance between noise removal efficacy and the preservation of the original structural integrity remained elusive. Concurrently, attempts at learning-based transformation targeting the removal of adverse effects or the restoration of detection were made. The mechanisms and attributes of adaptive filters and learning-based transformation models will be introduced in Section 2.1.

Conversely, the augmentation of data pertinent to adverse conditions addresses the issue from an alternate vantage point. An expanded collection of adverse effect data, particularly if paired, can significantly facilitate the specialized training of perception models under challenging conditions. Methodologies for synthesizing adverse effects through physical model construction, as well as augmenting pre-existing adverse datasets, have both undergone scholarly exploration. Section 2.2 will introduce literature concerning the expansion of adverse condition data.

2.1 Adverse Effects Removal

2.1.1 Adaptive point cloud filters

Traditional median filters operate within a 2D plane and are not well-suited for handling the dynamic nature of adverse conditions like snowfall. The data are quite sparse in the vertical field of view above ground and the 2D filter couldn't handle the noise point removal and edge smoothing properties well. To address these issues, the 3D point cloud radius outlier removal (ROR) filter is employed. This 3D ROR filter iteratively examines each point within the point cloud and identifies neighboring points within a defined search radius. If the number of neighboring points is less than a specified minimum threshold (k_{min}), the point is labeled as noise and subsequently removed. This method aligns well with the characteristics of snowfall, where snowflakes manifest as small, isolated, solid objects.

However, a straightforward application of the 3D ROR filter can lead to the unintended removal of distant environmental points, thus undermining the LiDAR's anticipatory perception capabilities. Charron et al. [46] introduced the dynamic radius outlier removal (DROR) filter, where the search radius for each point (SR_p) is adaptively set based on its intrinsic geometric properties, as defined in Eq.2.1. The DROR filter effectively retains critical points located at a distance (6 m - 18 m) from the sensor while eliminating close-range noise (within 6 m). It has been shown to improve precision by nearly 4 times compared to standard ROR filters.

$$SR_p = \beta_d(r_p\alpha) \tag{2.1}$$

r_p is the range from the sensor to the point p, α is the horizontal angular resolution of the LiDAR, and the product of $(r_p\alpha)$ represents point spacing, which is expected to be computed assuming that the laser beam is reflecting off a perpendicular surface. The multiplication factor β_d is meant to account for the increase in point spacing for surfaces that are not perpendicular to the LiDAR beams [46].

In another approach, the statistical outlier removal (SOR) filter [47] removes any

point whose mean distance to its k nearest neighbors exceeds the threshold when iterating each point. The threshold T for filtering is computed as:

$$T = \mu + \sigma\beta_s \tag{2.2}$$

where μ is the global mean of the distances from all points to their k nearest neighbors; σ is the global standard deviation of the distances; and β_s is a specified multiplier parameter. The selection of k is the consideration of the approximate points number of an average actual object's points cluster based on the signal density of the LiDAR device used. Too big a k value makes a certain point's mean distance to its k neighbors inappropriately large, which could result in falsely accusing this point of noise, and vice versa. The rise of β_s results in the rise in threshold T, which makes the filter's tolerance higher, ending up in a weaker ability to remove noise points. On the other hand, too small a β_s makes the filter remove more valid points.

Building on the strengths of both the DROR and SOR, the Dynamic Statistical Outlier Removal (DSOR) [48] has been proposed. The filter threshold from SOR is now dynamically changed with range, and the dynamic threshold T_d is set by:

$$T_d = \beta_d s(Tr_p) \tag{2.3}$$

where T is from Eq. 2.2 and r_p is the distance of every point from the sensor, same as in Eq. 2.1. $\beta_d s$ serves as a multiplicative factor for point spacing. Larger $\beta_d s$ leads to a milder filter.

Subsequently, a variety of optimized filters based on adaptive parameters such as intensity-based [49], and density-based [50] were put forward. While adaptive filters have made strides in de-noising, they are inherently constrained by predefined removal rules. This limitation becomes particularly evident when considering the diverse and infinite ways of noise clusters, a complexity that cannot be fully captured by such rigid algorithms. Despite these challenges, adaptive filters have laid a solid foundation for the early stages of LiDAR point cloud processing and transformation.

Another filtering approach was introduced by Pfennigbauer et al. [51], focusing

on online waveform processing under fog and dense smoke conditions. Different from the traditional mechanism of time-of-flight (TOF), they identify the targets by the signatures of laser reflection properties (reflectivity and directivity), size, shape, and orientation with respect to the laser beam, which means, this echo-digitizing LiDAR system is capable of recording the waveform of the targets which makes it possible to identify the nature of the detected target, i.e. fog and dense smoke by recognizing their waveforms. Furthermore, since the rate of amplitude decay caused by the fog follows a certain mathematical pattern with regard to the density of the fog, they realized visibility range classification and thusly were able to filter out false targets that don't belong in this range. Even though their experiments were confined within a critically close range (30 m), they paved the way for recovering targets hidden inside fog and smoke, regardless of the attenuation and scattering effects as long as the signal power stays above the designated floor level, because too low visibility could block the detection almost entirely. Most importantly, the concept of waveform identification brought the multi-echo technique to the commercial LiDAR markets.

Wallace et al. [52] explored the possibility of implementing Full Waveform LiDAR in fog conditions. This system records a distribution of returned light energy, and thus can capture more information compared to discrete return LiDAR systems. They evaluated 3D depth image performance using FWL in a fog chamber at a 41 m distance. This type of LiDAR can be classified as a single-photon LiDAR and 1550 nm wavelength, which Tobin et al. [53] also used to reconstruct the depth profile of moving objects through fog-like high-level obscurant at a distance up to 150 m. The high sensitivity and high-resolution depth profiling that single-photon LiDAR offers make it appealing in remote, complex, and highly scattering scenes. But this brings back the question of 1550 nm wavelength as stated in Section 1.1.1, which still leaves room for further exploration.

2.1.2 Learning-based adverse effects removal models

Although the input data formats may differ within images and point clouds, the underlying principles governing these models remain largely similar. Take de-raining

as an example. The detection and removal of raindrops can be divided into falling raindrops and adherent raindrops that accumulate on the protective covers of sensors [54]. For rain streaks removal, several training and learning methods have been put to use including Quasi-Sparsity-based training [55] and continual learning [56]. Quan et al. [57] proposed a cascaded network architecture to remove rain streaks and raindrops in a one-go while presenting their own real-world rain dataset. Their raindrop removal and rain streak removal work in a complementary way and the results are fused via an attention-based fusion module. They effectively achieved de-raining on various types of rain with the help of neural architecture search and their designated de-raining search space. Ni et al. [58] introduced a network that can realize both removal and rendering. They constructed a Rain Intensity Controlling Network (RIC-Net) that contains three sub-networks: background extraction, high-frequency rain streak elimination, and main controlling. Histogram of oriented gradient (HOG) and auto-correlation loss are used to facilitate the orientation consistency and repress repetitive rain streaks. They trained the network all the way from drizzle to downpour rain and validation using real data shows superiority.

Like common de-noising methods, a close loop of both generation and removal can present better performance. Wang et al. [59] handled the single image rain removal task by first building a full Bayesian generative model for rainy images. The physical structure is constructed by parameters including direction, scale, and thickness. The good part is that the generator can automatically generate diverse and non-repetitive training pairs so that efficiency is ensured. Similar rain generation is proposed by Ye et al. [60] using disentangled image translation to close the loop. Furthermore, Yue et al. [61] surpassed image frames and achieved semi-supervised video de-raining with a dynamic rain generator. The dynamical generator consists of both an emission and transition model to simultaneously construct the rain streaks' spatial and dynamic parameters like the three mentioned above. They use deep neural networks (DNNs) for semi-supervised learning to help the generalization for real cases.

Zhang et al. introduced a Deep Dense Multi-Scale Network (DDMSNet) for snow removal from camera images [62]. The process initiates with a coarse removal network

comprising three distinct modules: a pre-processing module, a core module, and a post-processing module. These modules use various combinations of dense blocks and convolutional layers to preliminarily mitigate the impact of falling snow. The coarse result then feeds into another network to obtain semantic and geometric labels. Leveraging semantic and geometric priors through self-attention, DDMSNet produces clean, snow-free images. Intriguingly, the team employed Photoshop to artificially snowify images from the Cityscapes and KITTI datasets for evaluation. While this approach demonstrates state-of-the-art snow removal capabilities, there remains a need for more advanced techniques to simulate photo-realistic snowy images.

The initial implementations of point cloud transformation using deep learning were primarily focused on mitigating the effects of rain and fog, which introduce obscurity or diffusion of LiDAR signals due to the presence of small water droplets. Additionally, external disturbances such as wind or spray can lead to the formation of clustered fog and water mist, thereby causing false obstacles for LiDAR [63]. Lin et al. [64] implemented the nearest neighbor segmentation algorithm and Kalman filter on the point cloud with an improvement rate of less than 20% within the 2 m range. Shamsudin et al. [65] developed algorithms for removing fog from 3D point clouds after detection using intensity and geometrical distribution to separate and target clusters of points, which were then removed from the point cloud. The restriction is that their environment is an indoor laboratory and the algorithms are designed for building search and rescue robots whose working condition has too low visibility to be adapted into outdoor driving scenarios where beam divergence and reflectance are significantly larger in the far field than in the near field.

Given the scarcity of paired data for training, unpaired data training methods have gained prominence, often utilizing generative models like GANs (generative adversarial networks). Built upon the architecture of CycleGAN [66], DiscoGAN [67], and DualGAN [68], which introduce a cycle-consistency constraint to establish connections between the inputs, de-weather frameworks have proven effective in removing multiple adverse conditions including haze and snow [69] [70]. Engin et al. [71] proposed Cycle-Dehaze which is an improved version of CycleGAN that combines cycle

consistency and perceptual losses in order to improve the quality of textural information. Shao et al. [72] proposed a domain adaptation paradigm that introduces an image translation module that translates haze images between the real and synthesis domains. These methods' de-hazing results often fall short of expectations due to the presence of artifacts. However, their ability to operate without the need for paired images holds promise for the development of more robust models.

Consequently, the utilization of GAN models for LiDAR point cloud manipulation has emerged as a logical next step. Early efforts by Sallab et al. [73] [74] pioneered the translation from simulated CARLA driving environments to synthetic KITTI [75] point clouds using the CycleGAN framework. Similarly, Lee et al. [76] undertook point cloud translations across sunny, rainy, and foggy weather states, employing the depth and intensity channels of the 2D Polar Grid Map (PGM) for CycleGAN processing. However, it should be noted that the challenge of inter-domain translations persists, particularly when transitioning from fog chamber scenarios to real-world settings. This arises because artificial precipitation produced by sprinklers in fog chambers is detected by LiDAR sensors as vertical cylinders rather than as natural rainfall [10], a discrepancy that may compromise the interpretability of weather reflection features in point clouds and thereby potentially degrade the performance of translations. The inability of chambers to simulate snowy conditions has also resulted in a lack of advancements in the processing of LiDAR point cloud data in snow environments.

Heinzler et al. [45] proposed a CNN-based (Convolutional Neural Network) approach capable of weather segmentation and de-noising with road data augmentation to enhance the model's capability of comprehending real-world elements that are challenging to simulate in fog chambers. The authors first collected both clear and foggy datasets within the CEREMA's climatic chamber, enabling controlled precipitation rates and visibility conditions. Leveraging the paired datasets, they facilitated an automated labeling procedure that annotated weather data, enabling the model to acquire knowledge regarding the distinctive attributes of rainfall and fog reflections in point clouds. When augmented with additional road data, the trained model exhibited the proficiency to distinguish between fog and rain-induced point clusters, and to

remove noise while maintaining the integrity of original objects, such as pedestrians and cyclists. This further ascertains that a closed-loop system encompassing both generation and removal of adverse conditions offers enhanced perception capabilities, thereby paving the way for the advancement of data augmentation techniques focused on adverse conditions.

Aldibaja et al. [77] [78] described a particular issue in adverse conditions: Accumulated snow on the roadside creates sharp intensity peaks with irregular distribution for LiDARs while wet elements from snow-rain weather leave a track of line with low reflectivity on the road. The wearing of old roads and vegetation whose branches reach into the road space also create anomalies sometimes. These confuse the LiDAR about the actual whereabouts of the lane lines and the boundaries of the road area which leads to wrong lateral movements. Aldibaja's group explained the reason for lateral localization drifting by converting map images into edge profiles to represent the road marks in a series of LiDAR signal reflectivity peaks and proposed to use the Principal Component Analysis (PCA) method to extract dominant edge profile distribution patterns and eliminate the "fake" lane lines via edge profile matching [79]. Also by patching the missing LiDAR elements based on leading eigenvectors (eigenroads), a reliable LiDAR profile was reconstructed. The error in the lateral movement was reduced to 15 cm with a localization accuracy of 96.4% in critical environments [78].

2.1.3 Position of proposed research

Upon an extensive examination of all geometric parameters associated with LiDAR signals, alternative adaptive variables that could potentially act as metrics for a denoising filter are subject to be explored. The search for a novel adaptive variable is motivated by the desire to improve both the precision and recall of the filter, while ensuring minimal impact on the crucial structural details within the point cloud. The optimal metric should enable the filter to adjust effectively to different noise levels and characteristics, especially those resulting from challenging adverse conditions such as snowfall. This exploration leads to the consideration of non-geometric parameters, including intensity [50] [80] and entropy [81] [82].

Table 2.1: Comparisons between conventional transformation methods and the proposed method

Point Cloud Transformation	Filters	Proposed learning-based model
Mechanism	k-d-tree neighbor searching remove outliers	Process the point cloud as a whole towards a clear condition template
Merits	Adaptive threshold: radius, intensity, neighbor numbers, etc.	Maintain strucutre integrity while effectively removing adverse effects
Problems	Trade-off between: Effectively removing adverse effects and damaging structure integrity	Sensitivity to domain gaps in lack of paired data

More importantly, to facilitate deep learning, an approach originated from GANs that leverages depth image priors for transforming LiDAR point clouds in challenging environments is proposed in Chapter 3. This model, trained on unpaired data, is designed to process point clouds from driving scenarios captured during actual adverse events, transforming these adverse condition point clouds into clear, undisturbed views, so as to eliminate the possibility of false detection to the greatest extent. The proposed learning-based model is designed to develop an understanding of environmental scenarios and the patterns of adverse effects as they appear in point clouds. This understanding is cultivated through training on authentic datasets collected in adverse conditions. The model's transformation effectiveness stems from its specially designed generator and discriminator structures. Importantly, this approach not only ensures the preservation of the scene's overall integrity but also facilitates the compensation of obstructed objects to a considerable degree. A comparison between conventional transformation methods and the proposed method is shown in Table 2.1.

2.2 Adverse Condition Data Augmentations

2.2.1 Adverse effect synthesis

Compared to adverse effect classification-removal-based methods, adverse data augmentation based methods show better effectiveness and accuracy because the annotations or labelings on the effects are time-consuming and unreliable [83]. However,

samples with adverse conditions in current driving datasets take up only a small proportion. Rasshofer et al. [80] reproduced the optical returns of the signals measured under weather conditions to analyze the influences of weather on LiDAR sensors. This is one of the primary attempts at data augmentation when weather datasets were not common. The replication of the signal returns points out the essence of LiDAR data augmentation: Synthesizing perception results in the same way that real weather conditions show to the LiDAR sensor. However, the realization of the adverse effects in point clouds still depends on the collections in weather chambers [45] because of the requirement for paired data. The low domain similarities between chambers and real roads aside, common experimental facilities with controllable precipitation rates across the world can hardly simulate complicated weather conditions such as dynamic snowfall [6] [84]. On the other hand, targeted collections of weather datasets in certain areas such as snowy regions suffer huge domain gap issues with other datasets which leads to poor model generality. Recent works started to collect weather datasets utilizing a professional haze/fog generator that imitates the real conditions of haze scenes [85], or multiple weather stacking architecture [86] which generates images with diverse weather conditions by adding, swapping out and combining components. Therefore, the augmentation of weather conditions in clear conditions is among the most useful and popular ways to help deep-learning-based models address adverse condition problems.

Initial approaches start with establishing a physical or geometrical model of adverse effects as precise quantitative reflections of such effects and inserting them into clear scenes. Von Bernuth et al. [87] simulated and evaluated snowflakes in three steps: first, reconstruct the 3D real-world scene with depth information in OpenGL; then snowflakes are distributed into the scene following physical and meteorological principles, including the motion blur that comes from wind, gravitation or the speed of vehicle displacement; finally, OpenGL renders the snowflakes in the realistic images. While this modeling successfully captures the physical characteristics of adverse weather conditions, it often lacks the inclusion of dynamic environmental variables for dispersing these effects across the scene. For example, Guo et al. treated rain and fog

as uniformly distributed noise points and snow as normal distribution arrangements in the translations of a whole scene for testing platform [88]. This method, although systematic, reveals a limitation in capturing the variability and nuance in the distribution of adverse effects in real-world environments. Piroli et al. synthesized the frost condensed from the exhaust pipe of a car when encountering cold air with α-Shape algorithm [89]. The frost entities were well-reconstructed and their positioning was highly accurate due to the directional nature of the frost emission, thus avoiding the distribution problem.

2.2.2 Existing adverse effects enrichment

Chen et al. [90] find that de-hazing models trained on synthetic images usually generalize poorly to real-world hazy images due to the domain gap between synthetic and real data. As a result, models derived from pre-existing adverse effects within the dataset were developed for the best preservation of effect authenticity. Hahner et al. [43] developed a method that simulates snow particles in a 2D space corresponding to each LiDAR line and adjusts each LiDAR beam's measurements based on the resulting geometry. In addition, they factored in ground wetness, a common occurrence during snowfall, in their LiDAR point clouds as a supplement of the augmentation. The notable enhancement observed in the performance of 3D object detection subsequent to training on semi-synthetic snowy training data substantiates the successful simulation of snowfall. It is of particular importance to acknowledge that their snow augmentation approach predominantly focuses on light snowfall conditions under the rate of 2.5 mm/hr, wherein the prevalent snow effects in LiDAR point clouds manifest as dispersed noise points rather than snow clusters.

Strong light and contamination pose unique challenges in the field of adverse effects, particularly because of the inherent difficulty in acquiring paired data sets of both affected and clear images under realistic driving conditions. This makes them a prime candidate for synthetic generation and data augmentation techniques. Yahiaoui et al. developed a sunshine glare dataset in autonomous driving called WoodScape [91], including situations like direct sunlight in the sky or sun glares on

dry roads, road marks being wiped off by sun glares on wet roads, sun glares on reflective surfaces, etc. The glare is detected by an image processing algorithm with several processing blocks including color conversion, adaptive thresholding, geometric filters, and blob detection, and trained with CNN network.

Uřičář et al. [92] created a dataset called SoilingNet having both opaque and transparent soiling [93], and developed a GAN-based data augmentation for camera lens soiling detection in autonomous driving. Different from rain or snow, the general soiling is normally considered opaque or semi-transparent, so a complementary sensing method might not be able to perform with enough accuracy. Their network would generate an image with a random soiling pattern, which provides a blurred mask obtained from the semantic segmentation network applied with a Gaussian smoothing filter on the generated soiled image, and finally, the synthetic version of the soiled image is composed of the original image and the soiled pattern estimated via the mask. Although strong light and contamination augmentation on LiDAR point cloud have yet to be explored, the potential for synthetic generation and data augmentation in addressing contamination therefore also suggests a broader applicability for these techniques in tackling other hard-to-acquire adverse conditions in LiDAR point cloud data.

Training networks to cope with adverse conditions leverages a variety of neural network architectures, such as CNN [45], R-CNN (Region-based CNN) [94], DNN [61], and BPNN (Back-propagation Neural Network) [95], each employing a range of advanced algorithms. Despite the challenges in obtaining high-quality datasets for these adverse conditions, artificially synthesized weather effects are increasingly becoming the go-to approach in this research area and have shown to be effective.

2.2.3 Position of proposed research

In this thesis, a model for augmenting LiDAR point cloud data is proposed in Chapter 4, focusing on expanding datasets under adverse conditions through conditional guidance. This model generates segmentation maps with unique labels that identify the presence of adverse effects in point clouds, serving as conditional guides.

Table 2.2: Comparisons between conventional augmentation methods and the proposed method

Point Cloud Augmentation	Adverse effect synthesis	Proposed augmentation model
Mechanism	Synthesize adverse effects out of physical models	Learn the pattern and expand adverse effects
Merits	Precise quantitative reflections of adverse effects	Data expansion with understanding of both adverse effects and distribution
Problems	Lack understanding of effect distributions	Relies on pre-existed dataset

The proposed approach involves developing efficient fusion techniques to integrate these guides with raw data into the generative model. This integration facilitates the classification and synthesis of adverse effects. The proposed method effectively incorporates realistic adverse conditions into point clouds, taking into account scene context and striving to preserve structural integrity, which is crucial for creating reliable paired datasets. Furthermore, by leveraging the conditional guide, the aim is to bridge the domain gap related to traffic patterns and environmental layouts, which often pose challenges in realistically augmenting adverse data. This approach enables the creation of quasi-natural adverse effects in clear datasets, thereby enhancing the robustness and diversity of adverse data collection. A comparison between conventional augmentation methods and the proposed method is shown in Table 2.2.

Chapter 3

LiDAR Point Cloud Transformation From Adverse Conditions to Clear Conditions

3.1 Overview

As one of the core elements of autonomous vehicles and robots, perception in autonomous driving systems has experienced serious impediments due to the presence of adverse weather conditions. As established in the previous context, the challenges imposed by adverse conditions on perception sensors in autonomous vehicles are multifaceted. They range from signal degradation and intensity attenuation to the introduction of spurious noise, as shown in the red boxes of the left figure of Fig. 3.1, all of which hinder the sensor's core functionalities. These impediments not only reduce the efficacy of object detection and classification algorithms but can also introduce false positives—commonly referred to as 'ghost objects'. Such inaccuracies have immediate safety implications, increasing the likelihood of unnecessary braking or stops, thereby compromising both the safety and comfort aspects of autonomous driving.

To mitigate these challenges, recent research on perception in adverse conditions

has primarily centered on noise reduction and quality enhancement techniques, reaching near-saturation in methods such as de-noising [96]. However, the domain has witnessed limited exploration in the application of deep learning-based models. Unlike traditional mathematical approaches, which offer limited insights, learning-based models possess the capability to uncover both manifest and latent features inherent to adverse driving conditions. With the rapid development of machine learning technology in recent years, a number of deep learning models and algorithms have been developed to 'de-weather' or 'de-noise' the sensor data [97] [98], to achieve the transformation towards clear in the point cloud, as shown in the right figure in Fig. 3.1.

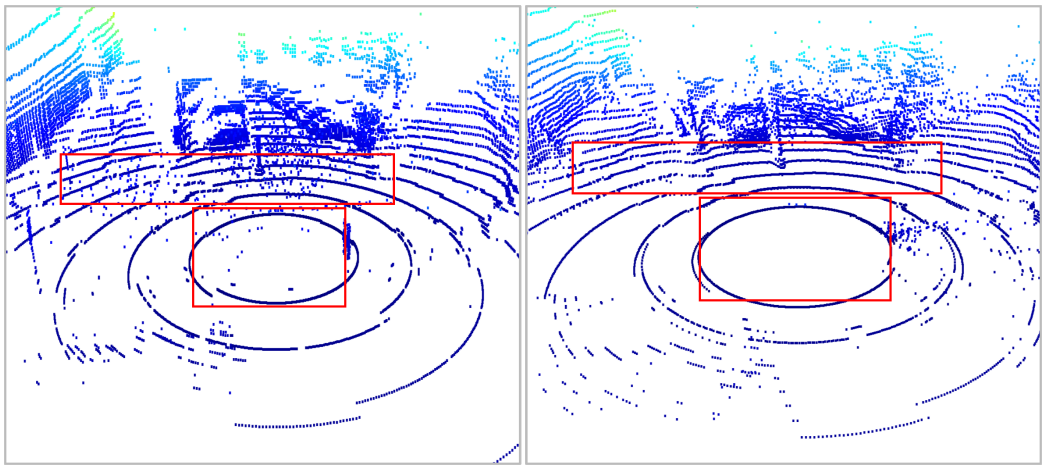


Figure 3.1: Point cloud collected while driving in adverse snow conditions (left), and corresponding transformed clear results (right) acquired from the proposed method. Color encoded by height. Red boxes denote areas of snow points before and after transformation.

Point cloud de-noising initially targeted fog conditions, largely because sensors exhibit a greater resilience to rain than to fog. Later, various filters and transformation algorithms that can restore the point cloud to a state as close to the original, noise-free condition as possible have been employed. These initial efforts laid the groundwork for more sophisticated approaches, including machine learning algorithms that can adapt to varying levels of fog density and other environmental variables. Over time, the methodologies for point cloud de-noising have evolved to include deep learning models tailored for this specific task. However, unlike other precipitation and wet

weather, snow is rarely uniformly distributed across a given driving scene. Instead, snowflakes show characteristics of forming clusters of crystal aggregates and irregular snow swirls [12] due to wind and vehicles swirling on accumulated road snow, causing irregular noise and anomalies in LiDAR point clouds. The study on snow in point cloud has been focusing on k-d-tree-based neighbor searching outlier filters in recent years as stated in previous context and the de-noising performance has almost reached saturation [96]. In the meantime, few attempts at the implementation of deep-learning-based models have been made in adverse conditions. Unlike filters, learning-based models have the potential to grasp both the surface and hidden features of noise clusters in a certain driving scene.

The development of robust models benefits from training on paired data, i.e. a pair of corrupted data and clear data with the rest of elements identical, which are commonly obtained via artificially synthesizing realistic weather effects in previously clear driving scene images [99] [100] [93]. Such an approach has been proven highly effective in rain [61] [60], fog [85] [71], and snow [101] weather conditions in camera images, plus contaminations on the camera lens [92]. However, due to the relatively low data density, the realization of the adverse effect in point clouds still largely depends on the collections in weather chambers before the mature realization of weather data augmentation in point clouds. Therefore, it's necessary to develop a way to work with few paired data or unpaired data. In terms of disentangled data processing, CycleGAN [66] demonstrates a high ability in style conversion and object generation [102] based on datasets with different backgrounds and from different domains, and its implementation in weather models has been proven feasible [69] [101].

In order to implement deep learning with unpaired data, the 'L-DIG' (LiDAR Depth Images GAN), a GAN-based method using depth image priors for LiDAR point cloud transformation under adverse conditions is proposed. The proposed model trained with unpaired datasets aims to perform point cloud transformation on driving datasets collected under real adverse conditions and to convert point clouds from adverse conditions to clear. The main contributions of this work are as follows:

1. A 3D clustering algorithm is customized as the means of analysis of 3D point

clouds with adverse conditions. Adaptive parameters to cluster different forms and levels of adverse effects are set to calculate multiple indexes, and distribution changes are presented as quantitative evaluation metrics to reflect the conditions of the whole dataset in a comprehensive way.

2. A range of adaptive filters on genuine adverse data are rigorously tested and validated, and their geometric parameters are thoroughly evaluated. Building on this, an advanced adaptive filter centered on entropy as the key metric for filtering is designed and implemented.

3. A deep-learning-based LiDAR point cloud transformation model with unpaired data and depth image priors is built. A new discriminator structure to better remove adverse noise and new loss functions including depth and SSIM (Structural Similarity Index Measure) losses to maintain the driving scene integrity have been designed in the proposed model.

4. The proposed model is able to perform LiDAR point cloud transformation from adverse conditions to clear conditions in driving scenes. The model demonstrates a certain level of understanding of the adverse effect features and performs effective adverse effects removal, which could help create paired datasets for training or simulation in autonomous driving applications.

This chapter is structured as follows. The related research and background reviewing have already been introduced in the current section and Section 2.1. Section 3.2 provides experimental validations on adaptive filters and a novel filter design based on entropy parameters. The architecture and methodology of the proposed 'L-DIG' model will be presented in Section 3.3. Section 3.4 lays out the construction of the 3D clustering algorithm. The experiments and both qualitative and quantitative results will be given in Section 3.5 and Section 3.6 provides a discussion and limitations on this model.

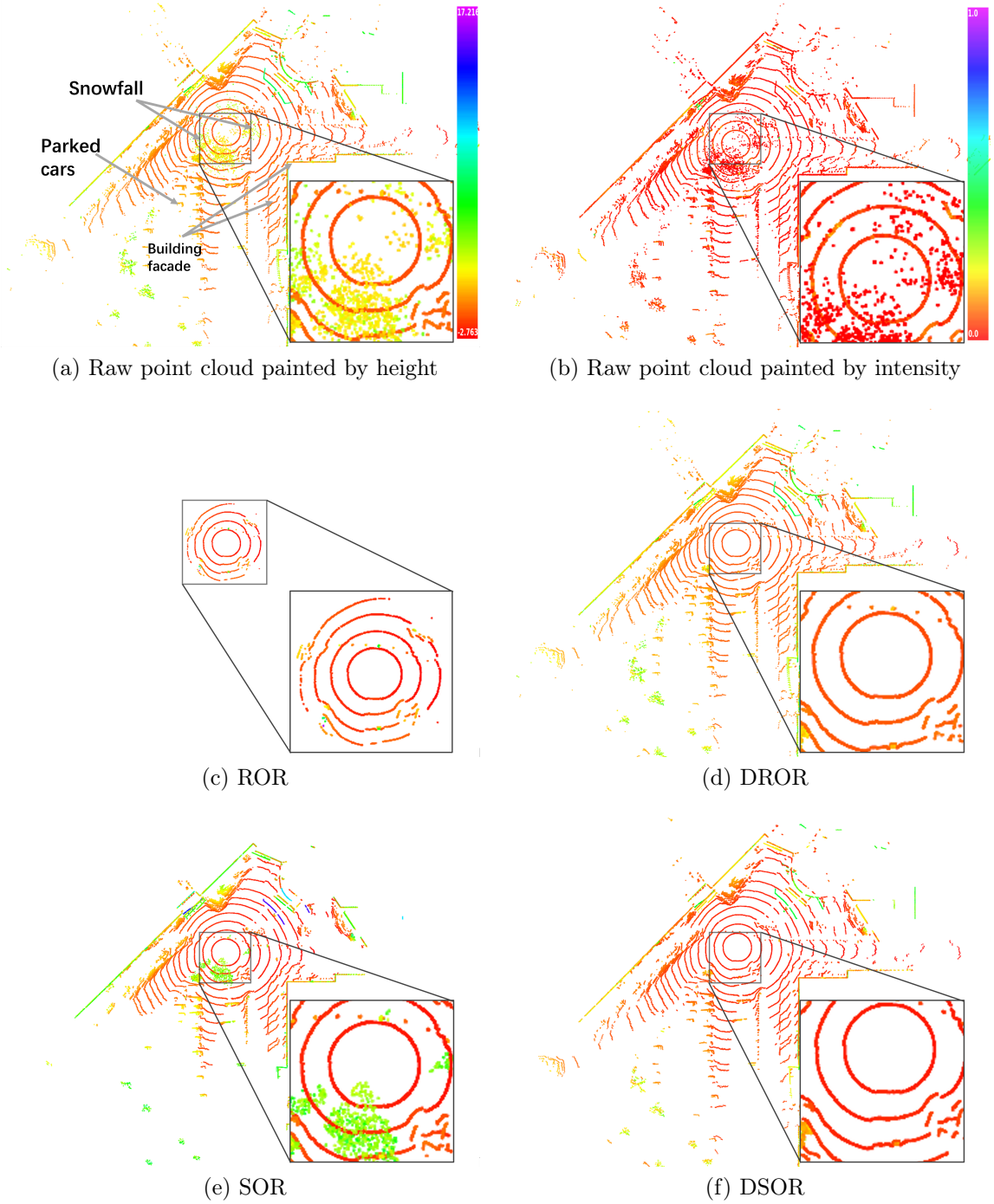
3.2 Advanced Adaptive Filters Validations and Exploration

3.2.1 Adaptive geometric filters

In this study, the performances of different adaptive filters for transforming LiDAR point clouds from adverse conditions to clearer states are evaluated. A scene from the CADC dataset is used, which includes various features like buildings, trees, and parked cars, as the test environment. This scene is shown in Fig. 3.2a and Fig. 3.2b, where the raw point cloud is displayed with height and intensity scales. Fig. 3.2 presents results from four filters: ROR, DROR, SOR, and DSOR. The focus is on removing noise clusters and individual noise points, marked by arrows, while also checking if the filters preserve the integrity of structural elements like parked cars and building facades. In each set of figures, an inset is added at the lower right corner that enlarges the central area. This is designed to provide a clearer visualization of the clusters and scattered noise points.

The ROR filter, when applied in 3D as seen in Fig. 3.2c, removes too many points, leading to a loss of pretty much all the essential environmental details and reducing the LiDAR's effective perception of the scene. The DROR filter, however, shown in Fig. 3.2d, manages to maintain key points at a distance and removes closer noise, improving upon the ROR's approach.

SOR's performance, displayed in Fig. 3.2e, is a step up from ROR, but it compromises de-noising effectiveness. This is evident from the remaining snow clusters, particularly those in green color below the center. The DSOR filter, configured with a radius of 0.05 in accordance with Eq. 2.3, as illustrated in Fig. 3.2f, demonstrates its effectiveness. This filter adeptly eliminates snow clusters while preserving a greater number of structural features than the SOR filter. Additionally, it offers the advantage of quicker computational processing [48].



(a) Raw point cloud painted by height

(b) Raw point cloud painted by intensity

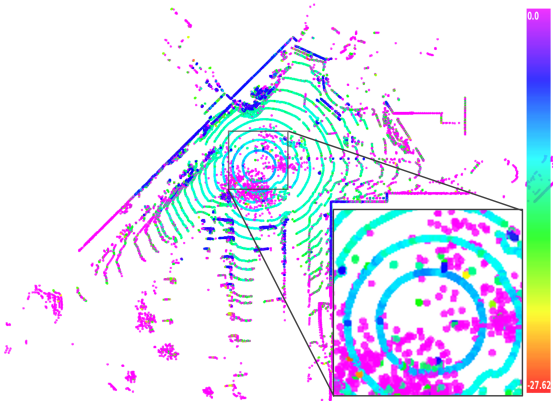
(c) ROR

(d) DROR

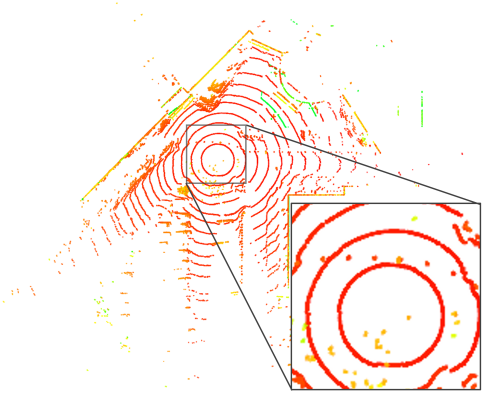
(e) SOR

(f) DSOR

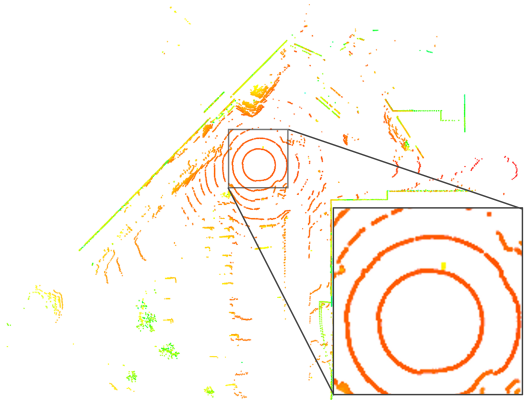
Figure 3.2: LiDAR point clouds in adverse snowfall conditions with different adaptive filters applied, produced using CADC dataset [16]. (a) Raw point cloud painted by height; (b) Raw point cloud painted by intensity; (c) to (f) are painted by height (Z axis), and share the same color scale as (a).



(a) Raw point cloud map entropy



(b) Entropy filter



(c) Intensity filter

Figure 3.3: Raw point cloud entropy and non-geometric filters. Entropy filter and intensity filter are painted by height (Z axis), and share the same color scale as Fig. 3.2a

3.2.2 Non-geometric filters

Having thoroughly explored all the geometric parameters of the LiDAR signal, alternative adaptive variables that could serve as the metric for a de-noising filter are examined. The quest for a new adaptive variable is guided by the aim to enhance the filter's precision and recall while minimizing the compromise on important structural details in the point cloud. The ideal metric would allow the filter to adapt to varying noise levels and types, particularly those induced by adverse conditions like snow. The exploration of non-geometric parameters including intensity and entropy are presented in Fig. 3.3, where an inset that enlarges the central area has also been added to the lower right corner of each set like before.

Fig. 3.3a shows the entropy representation of the original raw point cloud scene. The entropy h of a certain point q_k in the point cloud is computed by:

$$h(q_k) = \frac{1}{2}\ln\left|2\pi \mathrm{e}\Sigma(q_k)\right| \tag{3.1}$$

in which $\Sigma(q_k)$ is the sample covariance of mapped points in a local radius r (r = 0.25 m in this research) around q_k. While entropy seems like a good candidate for classifying snow points, Fig. 3.3b shows the result after the solitary points (points with less than 15 neighbors) with high entropy being filtered out [103]. As observed, the trade-off between de-noising performance and structural integrity is approaching a balanced state in this new design. However, the entropy filter still struggles to achieve 100% satisfactory filtering under conditions of dense snowfall.

Fig. 3.3c is a direct intensity filter where all the points with intensity values outside of the interval of [0.03, 0.15] are filtered out (intensity varies from [0,1]). This interval is set based on several trial-and-error attempts. It can be seen that the result is hardly acceptable due to huge element loss and the problem of how to determine the exact interval in different scenes that can both filter out snowfall and keep objects. Considering the intensity loss of laser signals as stated in Fig. 1.5, the practical use of the intensity filter is limited in adverse conditions.

This outcome prompts the redirection of the attention towards learning-based

methods for point cloud de-noising. Given the intricate and dynamic nature of environmental noise like dense snowfall, machine learning algorithms, particularly deep learning models, offer the adaptability and complexity required to tackle such challenges effectively. The aim is to explore architectures and algorithms capable of discerning more subtle patterns in the point cloud data, patterns that simpler geometric or entropy-based metrics might miss. This could pave the way for more robust and adaptive filtering techniques that are better suited for real-world applications where conditions can vary widely.

3.3 Transformation Model Architecture and Methodology

3.3.1 LiDAR Depth Images GAN

In this research, the LiDAR Depth Images GAN (L-DIG) model is proposed, an enhanced learning-based model using GAN as the foundational structure. As illustrated in Fig. 3.4, the blueprint of the proposed model is depicted. The model uses a generator to convert real snow input to synthetic clear output while referring to a real clear input, which is completed in a discriminator. The synthetic output also goes through a reconstruction process during training to maintain transformation stability. Subsequently, the procedure is inverted and the model is enriched by incorporating a cycle process, as shown in Fig. 3.5, leveraging the strengths of CycleGAN [66].

In this architecture, A and B symbolize two sets of data flow in the forward and backward cycle respectively, and Snow A and Clear B are the inputs in the form of depth images. C subscripts correspond to clear weather conditions, whereas S subscripts are used to indicate snowy conditions. The generators responsible for transitioning between snowy to clear and clear to snowy states are represented by G_{SC} and G_{CS} respectively while the discriminators are expressed as D_A. The reconstruction process is a feature of maintaining transformation stability. Snow A and Clear B are input to the model in the depth images form and the raw snow conditions are

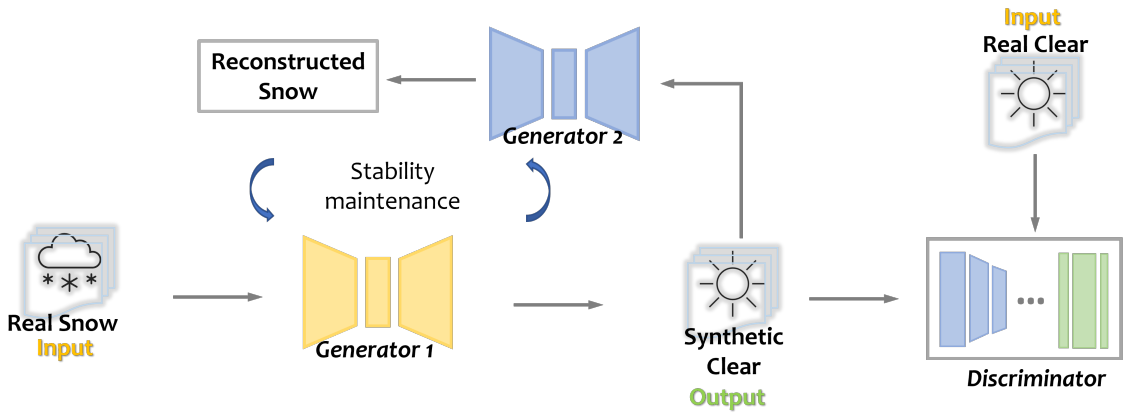


Figure 3.4: Proposed LiDAR transformation model architecture. Real adverse and real clear data are inputs in the form of depth images. The synthetic clear is the desired output.

transformed into clear conditions, with reference to the input Clear B data. Then the generated Clear A will be reconstructed back to snow with reference to Snow A to fix any possible overfitting during transformation. Utilizing the same Generators but in reverse sequence, Clear B input is converted into synthetic Snow B. This transformation is guided by the original Snow A and assessed by an additional discriminator D_B composed of D_{nB} and D_{pB}. The generated synthetic Snow B is then reconstructed back to a clear state, using Clear B as a reference for stabilization. The primary motivation for completing this cycle is to multiply the training opportunities for both Generators, particularly when dealing with limited unpaired data. The pseudo-code of the complete L-DIG is provided in Algorithm 1.

3.3.2 Pixel-attention Discriminators

A new discriminator structure is designed to enable the model to recognize the adverse noise points more accurately. The discriminators D_A and D_B are each composed of two parts: N-layer Discriminators D_{nA} and D_{nB} with 3 convolutional layers, and pixel Discriminators D_{pA} and D_{pB}. The N-layer Discriminators concentrate on relevant objects within the scene while the Pixel Discriminators scrutinize each pixel individually to ascertain its authenticity through a 1×1 patch. This approach contributes to a minor disturbance to the binary discriminator's threshold, elevating the criteria to

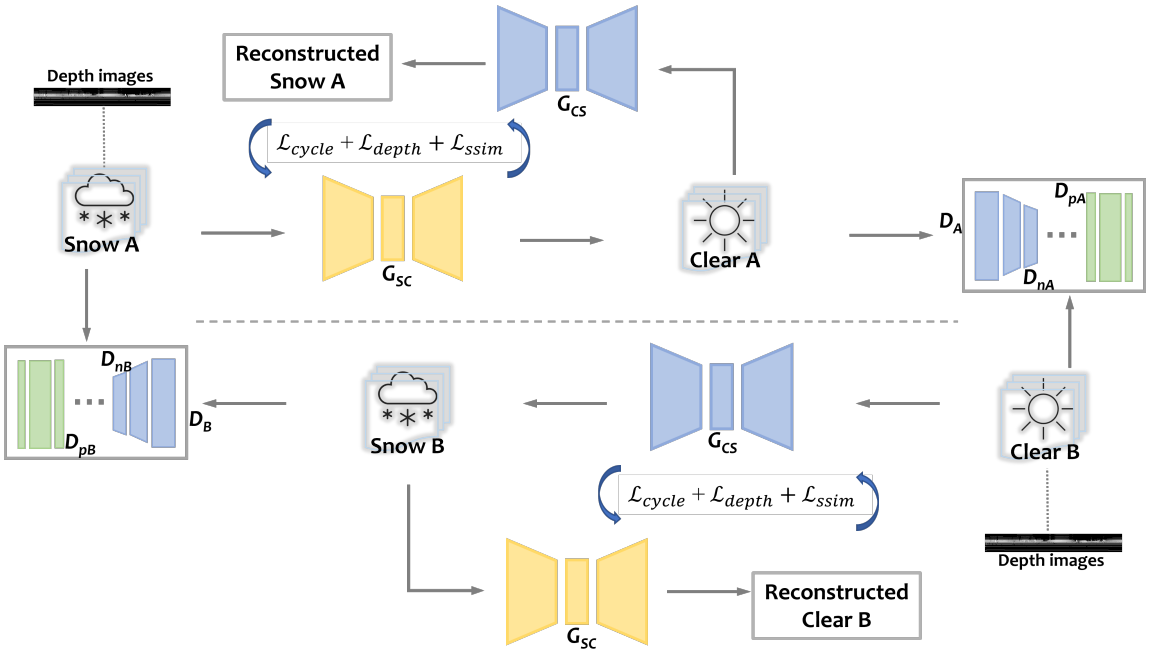


Figure 3.5: Proposed LiDAR transformation model architecture. Datasets A and B are in the form of depth images. G_{SC} and G_{CS} are the Generators. D_{nA} and D_{nB} are the N-layer Discriminators, and D_{pA} and D_{pB} are the Pixel-attention Discriminators.

achieve a 1 (approved) rather than a 0 (rejected) to a more stringent level for isolated noise points. This strategy markedly enhances the de-noising effect, particularly in areas surrounding the ego vehicle, where dispersed snow points are densely packed. However, pixel-attention discriminators cannot be introduced from the onset due to their stringent criteria, which carry the risk of causing model instability or breakdown. Hence, the pixel-attention discriminators are designed to undergo training for several epochs subsequent to the stabilization of the model training with N-layer Discriminators, as shown in Fig. 3.5 and Algorithm 1.

3.3.3 Loss Function

Depth loss

Scale ambiguity poses a challenge for depth images, necessitating the use of loss functions resilient to rough estimations [104]. Taking a cue from the *Fine* network [105], a depth loss, $\mathcal{L}_{depth}$, is crafted, that is integrated into the training cycles to

Algorithm 1 LiDAR Depth Images GAN (L-DIG)

Input: Training data pairs (A, B) $\quad \triangleleft$ Snow A and Clear B
Output: Generator networks G_{SC} and G_{CS}, N-Layer Discriminators D_{n_A}, D_{n_B}, and Pixel Discriminators D_{p_A}, D_{p_B}

Initialize generators G_{SC}, G_{CS} and N-Layer Discriminators D_{n_A}, D_{n_B}
Define loss functions including GAN, cycle, depth, and SSIM loss
Define optimizers for generators and discriminators
while epoch $\leq$ (total_epochs - continued_epochs) **do**
  for each data pair (A, B) in data_loader **do**
    Generate fake images: $F_A = G_{CS}(B)$, $F_B = G_{SC}(A)$
    Generate reconstructed images: $Rec_A = G_{CS}(F_B)$, $Rec_B = G_{SC}(F_A)$
    Compute GAN, cycle, depth, and SSIM loss
    Update discriminators D_{n_A}, D_{n_B} and generators G_{SC}, G_{CS}
  end for
end while
Initialize Pixel Discriminators D_{p_A}, D_{p_B}
while (total_epochs - continued_epochs) $<$ epoch $\leq$ total_epochs **do**
  for each data pair (A, B) in data_loader **do**
    Use the same generators to produce fake images as in previous training
    Compute Adversarial, Cycle, Depth, and SSIM loss
    Update discriminators D_{p_A}, D_{p_B}, D_{n_A}, D_{n_B} and generators G_{SC}, G_{CS}
  end for
end while

uphold consistency in the scale of depth images, as represented in (3.2),

$$\mathcal{L}_{depth} = \frac{1}{n}\sum_{i}\left(\hat{d}_i - d_i\right)^2 - \frac{\lambda_{depth}}{n^2}\left(\sum_{i}\left(\hat{d}_i - d_i\right)\right)^2 \tag{3.2}$$

where n denotes the sample amount, $\hat{d}_i$ and d_i symbolize the reconstructed and initial depth respectively, and the hyperparameter λ_{depth} governs the scale invariance. The assignment of $\lambda_{depth} = 1$ is to achieve complete scale invariance. This is due to the objective of preserving the transformed point cloud as similar to the original as possible and safeguarding the relevant objects and environmental elements from distortions in shape and size.

SSIM loss

As the point cloud transformation occurs on the scale of the entire scene, it sometimes involves objects and structures that are partially obscured or incomplete, leading to the model's suboptimal comprehension of these elements. This can cause distortions or alterations in the original forms, particularly in environmental features. To mitigate this, an SSIM (Structural Similarity Index Measure) loss [106], depicted in (3.3) and (3.4), has been incorporated into the training cycle to aid in preserving structural consistency.

$$\mathrm{SSIM}(N, \hat{N}) = \frac{(2\mu_N \mu_{\hat{N}} + c_1)(2\sigma_{N\hat{N}} + c_2)}{(\mu_N^2 + \mu_{\hat{N}}^2 + c_1)(\sigma_N^2 + \sigma_{\hat{N}}^2 + c_2)} \tag{3.3}$$

$$\mathcal{L}_{\mathrm{ssim}} = 1 - \mathrm{SSIM}(N, \hat{N}) \tag{3.4}$$

where N is the normalized image tensor (color, depth, texture, etc.) of the original real image, $\hat{N}$ is the normalized image tensor of the reconstructed image, $\mu_{\hat{N}}$ is the average of $\hat{N}$, μ_N is the average of N, $\sigma_{\hat{N}}^2$ is the variance of $\hat{N}$, σ_N^2 is the variance of N, $\sigma_{N\hat{N}}$ is the covariance of $\hat{N}$ and N, c_1 and c_2 are two variables to stabilize the division with a weak denominator.

The SSIM loss computation is performed post a subtraction by 1, due to the fact that SSIM loss gauges similarity while the training mechanism is geared towards attaining minimum values. Consequently, the difference is aimed to be lowered by training 1 minus the SSIM function. This also elucidates the necessity for prior normalization on the image tensor between [0, 1]. Meanwhile, it's essential to maintain a relatively low λ_s weight setting on the SSIM loss to prevent the model from becoming overly rigid, thereby obstructing any desired transformation.

Adversarial Losses

The adversarial losses between the clear and adverse data can be formulated as follows:

$$\mathcal{L}_{\mathrm{GAN}}(G_{SC}, D_A, S, C) = \mathbb{E}_{s\sim S}[\log(1 - D_A(G_{SC}(s)))] + \mathbb{E}_{c\sim C}[\log(D_A(c))] \tag{3.5}$$

$$\mathcal{L}_{\mathrm{GAN}}(G_{CS}, D_B, C, S) = \mathbb{E}_{c\sim C}[\log(1 - D_B(G_{CS}(c)))] + \mathbb{E}_{s\sim S}[\log(D_B(s))] \tag{3.6}$$

where $\mathbb{E}$ denotes the expectation. G_{SC} and G_{CS} are the generators responsible for the transformations from adverse to clear and clear to adverse, respectively. D_A and D_B are the discriminators for clear and snowy images, respectively. The generator aims to minimize the first term of each loss while the discriminator aims to minimize the second term.

Cycle consistency loss

The cycle consistency loss (3.7) derived from CycleGAN is employed, aiming at maintaining the created depth images closely aligned with the original domain. Provided a minimal variation in the background during the transformation, the weight λ_c of cycle consistency loss can be set as equal to that of the customized depth loss.

$$\mathcal{L}_{cyc} = \| G_{SC}(G_{CS}(S_B)) - S_B \| + \| G_{CS}(G_{SC}(C_A)) - C_A \| \tag{3.7}$$

Overall loss function

Upon integrating the conventional GAN adversarial losses between the clear and snowy data, denoted by $\mathcal{L}_{GAN}(G_{SC}, D_A, S, C)$ and $\mathcal{L}_{GAN}(G_{CS}, D_B, C, S)$, the comprehensive objective loss arrives at:

$$\begin{aligned}\mathcal{L}(G_{SC}, G_{CS}, D_A, D_B) = \lambda_g \mathcal{L}_{GAN}(G_{SC}, D_A, S, C) + \lambda_g \mathcal{L}_{GAN}(G_{CS}, D_B, C, S) \\ + \lambda_c \mathcal{L}_{cyc} + \lambda_d \mathcal{L}_{depth} + \lambda_s \mathcal{L}_{ssim}\end{aligned} \tag{3.8}$$

where λ_g, λ_c, λ_d, and λ_s denote the weight coefficients of adversarial losses, cycle consistency loss, depth loss, and SSIM loss respectively. The higher the weight, the larger the influence the corresponding loss function has on the model.

3.4 3D Point Cloud Analysis and Evaluation Metrics

3.4.1 3D clustering algorithm

In meteorology, precipitation rate, such as 10mm/hr, serves as a linear indicator for rain or snow levels. However, the performance degradation of sensors and the associated driving risks in adverse conditions may not strictly correlate with this objective metric. ADS rely on sensor perception, and even moderate precipitation rates can sometimes severely impair sensor functionality. For instance, water spray or snow swirls generated by the interaction between vehicles can significantly degrade perception capabilities. In traditional approaches, snow conditions have often been oversimplified as uniformly distributed noise with linear variations in both LiDAR and image-based perception systems. This model, although convenient, is fundamentally flawed; it fails to capture the intricate and highly variable nature of snow's presence within point clouds. The assumptions of linearity and uniformity break down when faced with real-world complexities. Specifically, the dynamics introduced by the movement of the ego vehicle, as well as interactions among multiple vehicles in proximity, contribute to air turbulence that can both redistribute atmospheric snow and agitate snow that has already accumulated on the ground, as demonstrated in Fig. 1.4. This results in snow manifesting as points within the LiDAR point cloud that defy simplistic models. The snow points display unpredictable density and non-uniform distribution patterns within the three-dimensional space defined by the LiDAR point cloud.

To better capture the complexities of snow conditions affecting driving scenarios, a divided classification for snow impact is proposed.

1. Mild: This is the initial phase of a snowfall event. The road has not yet accumulated a full layer of snow, and both precipitation rate and wind speed are moderate. In the point cloud, snow points manifest as isolated noise points, resembling 'salt and pepper' noise, distributed sporadically across the scene.

2. Severe: At this level, either the precipitation rate or the wind speed is considerably high. There is substantial snow accumulation on the ground. The ego vehicle's

movement, as well as that of nearby vehicles, generates significant snow dust and swirls. In the point cloud, snow points form a large, chaotic cluster near the center, accompanied by more densely packed noise clusters in the surrounding areas.

While this division may not capture all the manifestations of snow conditions in point clouds due to the inherently volatile nature of weather, it offers a relatively quantitative and adaptable framework. This tiered approach is particularly effective for addressing the challenges posed by snow swirls and provides a generalized basis to aid deep learning models in identifying snow-related features in LiDAR point clouds.

Given that a LiDAR point cloud is essentially a complex distribution of points within a 3D space, the nature of snow's representation within this 3D space is equally intricate. As such, the use of advanced 3D point clustering algorithms could offer an understanding of the physical characteristics of snow in various driving conditions. These algorithms could also serve as evaluative benchmarks for assessing the efficacy of point cloud processing methods in filtering or handling snow-induced noise and distortions.

The OPTICS (Ordering Points To Identify the Clustering Structure) algorithm [107] is chosen for its proficiency in handling clusters of varying densities. Originating from the DBSCAN algorithm [108], this clustering approach identifies distinct clusters by grouping together data points that are densely connected within a predefined radius. OPTICS operates with two essential parameters: ε, which denotes the maximum radius to be considered for neighborhood search, and $MinPts$, which indicates the minimal quantity of points necessary to constitute a cluster. Unlike DBSCAN, which requires a single threshold for density across the whole dataset, OPTICS deals with varying densities by computing two main values for each point, core distance (c-dist):

$$\text{c-dist}_{\varepsilon,MinPts}(p) = \begin{cases} \text{UNDEFINED} & \text{if } |N_\varepsilon(p)| < MinPts \\ \text{distance to the } MinPts^{\text{th}} & \\ \text{nearest point within } N_\varepsilon(p) & \text{otherwise} \end{cases} \tag{3.9}$$

where $MinPts$ describing the number of points required to form a cluster, $N_\varepsilon(p)$

being the set of points that are within a radius ε of the point p, and reachability distance (r-dist):

$$\text{r-dist}_{\varepsilon,MinPts}(o,p) = \begin{cases} \text{UNDEFINED} & \text{if } |N_\varepsilon(p)| < MinPts \\ \max(\text{core-dist}_{\varepsilon,MinPts}(p), \text{dist}(p,o)) & \text{otherwise} \end{cases} \tag{3.10}$$

which means the reachability distance of another point o from a point p is either the distance between o and p, or the core distance of p, whichever is bigger.

Using these calculated distances, OPTICS sorts the data points in a manner that reflects their density-based spatial relationships. Through this ordered sequence, dense groups of points, indicative of clusters, are revealed as contiguous segments where points have low mutual reachability distances. The undefined eventually remains as individual points whose cluster size equals to 1.

In the scope of this study, the algorithm has been adapted to autonomously identify the optimal $MinPts$ value for each cluster by leveraging the DBSCAN methodology. Moreover, the parameter ε is assigned based on the dimensions of the object with the smallest reflective surface within the analyzed context. For instance, in the case of a road intersection, the minimal reflective surface is typically presented by a speed limit sign. Consequently, ε is determined to be 0.6 meters, correlating to the width of this sign.

The algorithm offers several advantages: it is more adaptable to variable densities, exhibits reduced sensitivity to parameter choices, elucidates hierarchical relationships between clusters, and excels in outlier detection. The output from OPTICS facilitates easier interpretation and extraction of clusters without necessitating extensive manual parameter tuning. These attributes render OPTICS particularly well-suited for managing the variable behaviors exhibited by snow points in point clouds.

An example of the OPTICS clustering result is shown in Fig. 3.6. It can be seen that despite the density variation, everything has been classified into separated groups of clusters, each distinguished by unique colors, while the environmental structures are also well segmented. The conglomeration of snow swirl points, positioned at the

lower left of the center, are collectively assigned to large blue and purple clusters. Minor snow clusters, such as those in the immediate right vicinity of the center, along with individual scattered snow points spread across the scene, are categorized into smaller, uniquely colored clusters.

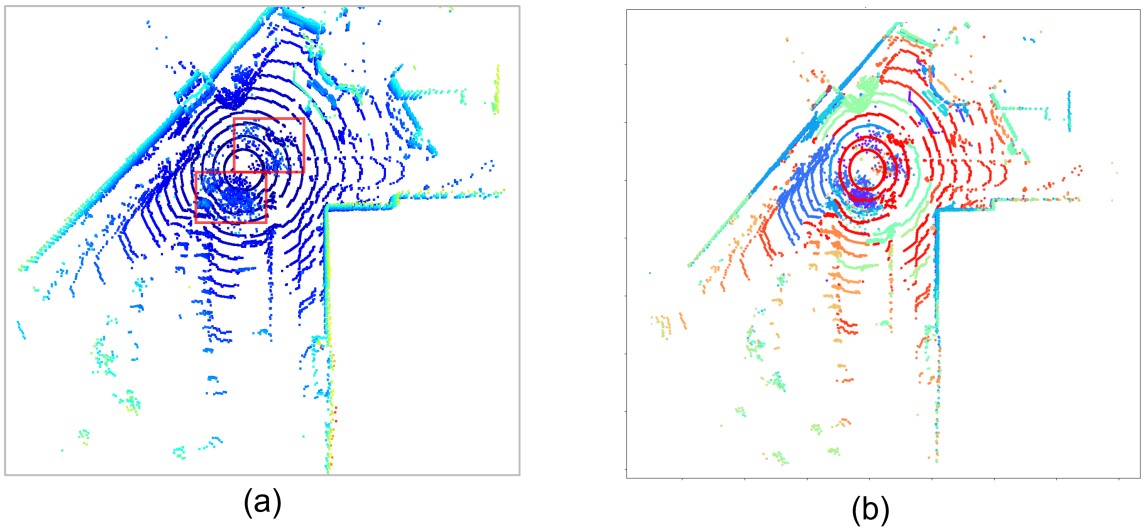


Figure 3.6: A frame of the point cloud featuring both dispersed noise points and snow clusters. (**a**) The original point cloud from the Canadian Adverse Driving Conditions (CADC) dataset [16], with colors representing height. Red boxes annotate scattered snow and snow swirl points. (**b**) The clustering result of the same point cloud based on the OPTICS algorithm, where varying colors signify different cluster groups. The prominent purple cluster at the center represents the snow swirl.

3.4.2 Quantitative evaluation metrics establishment

The quantitative evaluation of LiDAR point cloud processing with adverse effects has always been a tricky task. Researchers used to select a small number of samples and manually determine if a point is a snow point, in order to calculate the precision and recall of the removal of snow noise points [46, 48]. Even though straightforward, it has two downsides: For one, it consumes large amount of time and manpower to manually annotate a whole dataset while a small portion of samples suffers the risk of bias. Secondly, the accuracy of human annotation on point clouds with over 20,000 points in each frame can be as low as 85% and not satisfied enough to support the subsequent calculations on precision and recall [83]. Therefore, the OPTICS algorithm

is selected to reflect the quantitative characteristics of snow conditions and provide evaluation metrics in the research of snow problems in point clouds. Statistically, seven metrics based on the adaptive clustering algorithm are produced:

- Noise Number: Points without any neighbor points within a designated range (solitary points) are considered noise points, mostly snowflakes. A decrease in noise number is one of the most direct indicators of an effective snow removal performance.

  The count N of the points that have fewer than P_{min} neighbors within a given radius ϵ is shown as:

  $$N = |\{x \in X \mid \mathrm{P}(x, \epsilon) < P_{min}\}| \tag{3.11}$$

  Where X is the set of all points, and P(x,ϵ) returns the count of points within ϵ-radius of x.

- Cluster Number: A main output of the algorithm, representing groups of data points that are closely related based on their reachability. The cluster number can be simply denoted as C.

- Reachability Distance: The smallest distance required to connect point A to point B via a path of points that satisfy the density criteria. Normally, the average reachability distance would rise along with larger cluster numbers.

  For points A and B, the reachability distance $R(A, B)$ could be defined as:

  $$R(A, B) = \max\left(\mathrm{CD}(A), d(A, B)\right) \tag{3.12}$$

  where CD(A) (core distance) is the minimum distance required to separate A from its neighbors, and $d(A, B)$ is the Euclidean distance between A and B.

- Inter-Cluster Distances (ICD): The concept here involves identifying the centroid, or the average point, of each cluster, and subsequently computing the

distance between every possible pair of centroids. Should there be an increase in the average of these distances, it would suggest a reduction in the number of clusters and a more dispersed cluster distribution. In the context of this study, such a pattern could be interpreted as an effect of de-snowing.

For clusters i and j with centroids C_i and C_j:

$$\text{ICD} = \frac{1}{\binom{C}{2}} \sum_{i \neq j} d(C_i, C_j) \tag{3.13}$$

where $\binom{C}{2}$ denotes the number of unique pairs of clusters.

- Size of Clusters: This is essentially determined by the number of points each cluster holds. Under conditions dominated by scattered snow, the snow noise points tend to form numerous small-scale clusters. Their elimination, consequently, leads to an increase in the size of the clusters.

  For cluster i with n_i points, the size S could be:

$$S = \frac{1}{C} \sum_{i=1}^{C} n_i \tag{3.14}$$

- Silhouette Score: Measures the cohesion within clusters and the separation between clusters. A silhouette score close to 1 indicates a good clustering quality, while a score close to -1 indicates poor clustering. A lower silhouette score is commonly observed in snowy conditions due to the more overlap between clusters.

  For a point x, in cluster A, the silhouette score $s(x)$ is calculated as:

$$s(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} \tag{3.15}$$

where $a(x)$ is the average distance from x to the other points in the same cluster A, and $b(x)$ is the smallest average distance from x to points in a different

cluster, minimized over clusters. The overall silhouette score is then the average $s(x)$ over all points.

- Davies-Bouldin Index (DBI): Measures the ratio of within-cluster scatter to between-cluster separation and assesses the quality of the overall cluster separation. A lower Davies-Bouldin index indicates better clustering, with zero being the ideal value. Snow conditions with many noise points or swirl clusters exhibit higher values of DBI.

  The Davies-Bouldin index DBI is calculated as:

$$\text{DBI} = \frac{1}{C} \sum_{i=1}^{C} \max_{i \neq j} \left(\frac{t_i + t_j}{d(C_i, C_j)} \right) \tag{3.16}$$

  where t_i is the average distance of all points in cluster i to centroid C_i.

In summary, the above metrics serve as the quantitative evaluation criteria for assessing the effectiveness of snow removal in point cloud data, all of which are derived from the OPTICS clustering algorithm. These metrics include the Noise Number, Cluster Number, Reachability Distance, Inter-Cluster Distances, Size of Clusters, Silhouette Score, and the Davies-Bouldin Index. Each metric provides unique insights into the spatial characteristics of the point cloud, enabling a comprehensive understanding of snow conditions. For instance, metrics like Noise Number directly indicate the presence of snowflakes, while the Davies-Bouldin Index gives an overall quality measure of the clustering, thereby indirectly indicating snow conditions. Collectively, these metrics offer a robust framework for evaluating and comparing snow removal and generation performance across various datasets and conditions.

3.4.3 Analysis on snow point clouds

Some example scenes are presented in Fig. 3.7, including both a mild snowfall condition and a severe snow swirl condition to illustrate how the OPTICS algorithm evaluates snow conditions, with their metrics summarized in Table 3.1. A condition with both scattered noise points and snow swirls is provided as well for better

understanding. To ensure the authenticity of the snow effect, the first and one of the best datasets focusing on snow conditions in autonomous driving areas, which is CADC [16] is used. This dataset contains over 7000 frames of LiDAR point cloud collected during the winter in Waterloo, Ontario. The driving scenes cover both urban and suburban environments, high and low speed conditions, and different levels of snowfall and heavy snow accumulation conditions.

From the statistics, one can tell that with the increase in snow level, the noise number and cluster number of the driving scene are gradually rising, along with the average reachability distances as expected. With fewer noise points and fewer snow clusters, the average inter-cluster distances and the average sizes of clusters decrease correspondingly. The tendencies toward deteriorated clustering and increased overlaps, as indicated by the DBI and silhouette score, are also consistent with conditions of heavier snow conditions. This validation proves the OPTICS algorithm's capability of evaluating the change of snow conditions in a LiDAR point cloud.

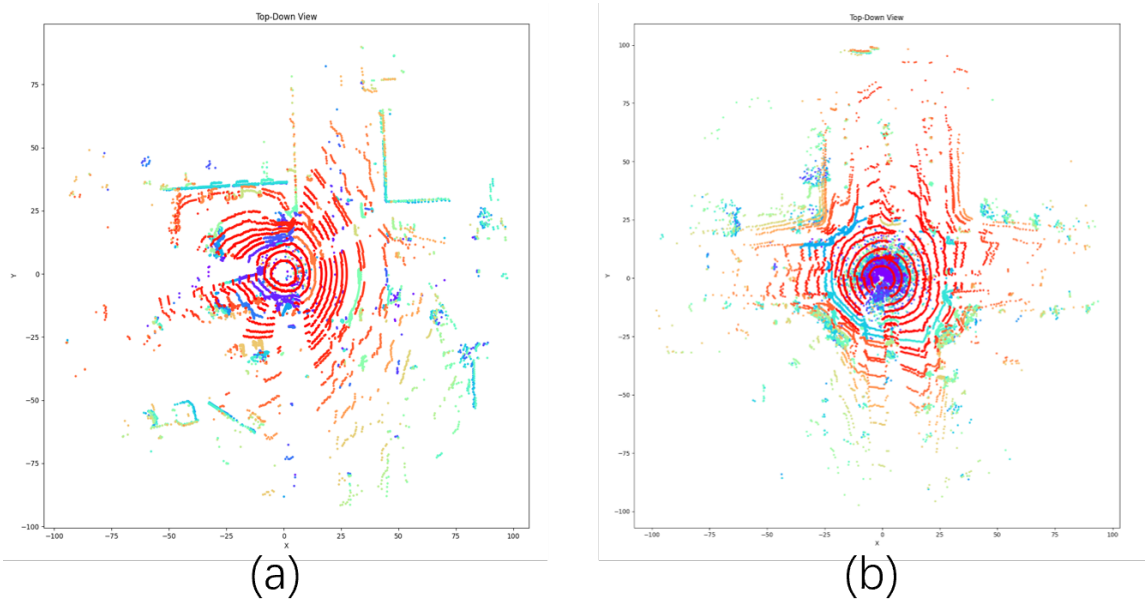


Figure 3.7: The clustering results of three example scenes from the OPTICS algorithm, where varying colors signify different cluster groups. (**a**) Mild snowfall condition with bits and pieces of snow points. (**b**) Severe snow swirl condition with huge snow swirl clusters surrounding the ego vehicle.

In order to more effectively illustrate the differences between mild and heavy snow

Table 3.1: The corresponding 3D clustering metrics in Fig. 3.7 from the OPTICS algorithm for ascending snow levels.

Items	Mild snowfall	Severe snowfall
Noise Number	1861	2942
Cluster Number	761	1165
Reachability Distance	0.3336	0.4166
Inter-Cluster Distances	54.0279	46.1090
Size of Cluster	17.6373	12.3494
Davies-Bouldin Index	4.0110	18.4671
Silhouette Score	-0.1653	-0.2772

conditions in terms of reachability distances, inter-cluster distances, and cluster sizes, violin plots for these three metrics are provided in Figure 3.8, positioning mild snow on the left and heavy snow on the right in each set. These plots not only show their quartiles but more crucially, delineate the differences in distribution. It can be observed that as snow conditions get heavier, the distribution in all three metrics exhibits somewhat abrupt curves with sharper edges and sudden shifts, indicating the inherent disarray characteristic of heavy snow. Generally, scenes with less snow presence display a relatively uniform and smooth distribution, as depicted in the violin plots, with a lower skewness value [109], offering another angle to assess the changes in snow conditions.

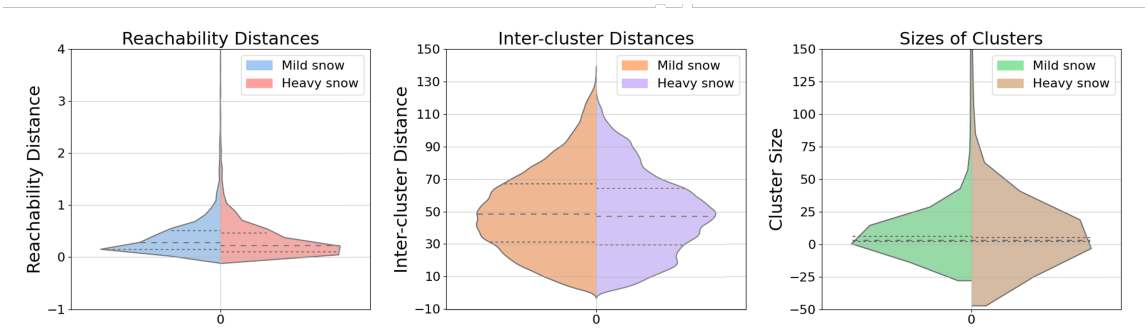


Figure 3.8: Violin plots for the comparison between mild snow and heavy snow. Limits on the y-axes are set for a better illustration of the distributions.

Furthermore, the whole CADC dataset is processed and divided into two groups as defined in the previous section and particularly extracted the snow points. The same

filter built within the CADC dataset (DROR) was used to complete the snow point extraction because manually labeling each snow point not only consumes unimaginable time and manpower but lacks satisfying accuracy and the amount of points is too large (at the magnitude of several thousand) with points sometimes overlapping in 3D space. In order to minimize potential bias, the filter's parameters are tuned towards high leniency in the case of extracting non-snow points. Points deemed as snow points by the filter do not exactly reflect ground truth but it is qualified for a snow cluster analysis when the possibility of object points being counted as snow points has been controlled to a minimum level by a lenient filter.

The analysis of snow points is shown in Table 3.2. Considering only the snow points are evaluated and most of them are noise points (solitary points), the statistics regarding cluster groups are not presented. Based on human annotation, 2583 samples of mild snowfall, and 1223 samples of fierce snow swirl conditions are grouped. As the severity of snow conditions escalates from mild to severe, an inverse trend in the average reachability distance can be observed compared to complete point clouds. This trend is consistent with the increasing density and disorder of the snow, reflecting a more complex and cluttered points environment. The two similar DBIs of around 1.7 - 1.8 indicate that the OPTICS algorithm has provided comparable moderate clustering quality and separation for mild, and fierce snow conditions, further affirming the choice of the OPTICS algorithm. The Silhouette Score also reflects the quality of the clustering but at a smaller scope down to each individual point. The decrease of the Silhouette Score below 0 indicates more overlaps and poor separations at the individual data point level while the severity of snow increases. These results match the observation and the behavior of snow points extracted from real snow scene point clouds.

As it turns out, the OPTICS algorithm shows high adaptability to variable densities and is less sensitive to parameter settings. It is capable of revealing hierarchical relationships among clusters and offers enhanced outlier detection capabilities. These attributes align well with the behaviors of all levels of snow conditions in point clouds, making OPTICS an effective evaluation method for snow datasets. Through the use

Table 3.2: The OPTICS clustering algorithm on the snow points extracted from the CADC dataset

Items	Mild Snowfall	Severe Snow Swirl
Snow points number avg.	1340.13	7713.03
Sample amount	2583	1223
Reachability distance avg.	0.6548	0.4328
Davies-Bouldin Index avg.	1.7032	1.7756
Silhouette score avg.	-0.2038	-0.3810

of adaptive clustering principles, the process of cluster interpretation and extraction is considerably streamlined, minimizing the need for extensive manual parameter tuning. In the following contents, this 3D clustering algorithm will be used to conduct quantitative evaluations on point clouds in adverse conditions.

3.5 Experiments and Results

3.5.1 Experiments

Experiments with the trained models on two different conditions are conducted: (1) Mild snow conditions: snowfalls only without snow swirls; (2) Fierce snow conditions: both snowfalls and snow swirls. The experiment on snowfall-only conditions was first conducted to examine the performance of scattered noise point capture. In the meantime, this less occlusion condition provides a better opportunity to check how well the original environmental structures have been maintained, so as to affirm the modal's ability of accurate snow capturing. Then the same experiment was conducted on conditions with both snowfalls and snow swirls, to comprehensively present the modal's ability to handle highly adverse conditions. The CADC dataset is used to conduct the experiments, spanning both urban and suburban settings.

To adapt LiDAR point cloud data to fit within the structure of the GAN-based model, pre-processing to the point clouds was initially applied and a 2D visualization of the point clouds was yielded, namely, depth images, which signify the orthographic projection of the point clouds while keeping the depth information (i.e. the dis-

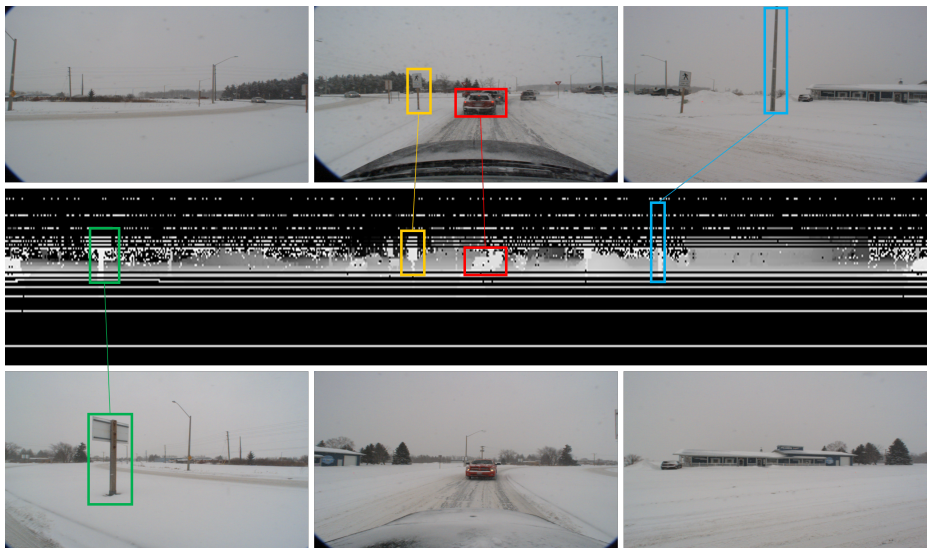


Figure 3.9: An illustration of a specific frame of the depth image under heavy snow conditions. The middle row displays the depth image, while the top and bottom rows depict corresponding camera images derived from the CADC dataset [16]. Images captured from multiple cameras targeting different directions around the ego vehicle. Color-coded boxes represent matching objects in both the camera and depth images. Green - The back side of a road sign. Yellow - The front side of another road sign. Red - The leading vehicle. Blue - A pole. The images do not denote the original resolutions but are adjusted for better demonstration purposes.

tances from points to the ego vehicle) within its pixel value. There is another form of point cloud 2D representation, Bird's Eye View (BEV) images, that represent the horizontal projection of point clouds and store the height information (i.e. the distances from points to the ground) within the point pixel values. Given the primary data of LiDAR signals are acquired in the front-view position, depth images are selected as the representative of point clouds in order to reduce cumulative errors in processing and reconstructing point clouds. By unrolling the exterior surface of the LiDAR's horizontal field of view (FOV) cylinder and mapping each point from the point cloud onto this frontal-view plane, a rectangular image encompassing all points within the LiDAR's FOV is obtained. The horizontal field is partitioned evenly into w columns and the vertical field is distributed uniformly into h rows. Consequently, post-projection, a depth image bearing a resolution of $w \times h$ can be secured, where the horizontal resolution is proportional to the sensor's rotation rate, and the vertical resolution is proportional to the number of physical layers [110].

An illustration of a specific frame of the depth images under noticeably snowy conditions, along with the corresponding camera images featuring identical objects, can be found in Fig. 3.9. A close observation reveals that the relevant objects are well reflected and the snow noises, bearing resemblance to 'salt and pepper' speckles, are prominently displayed in the depth image. The position of each pixel reflects the frontal-view projection of the points within the point clouds, while the pixel value from 0 to 255 signifies the distance between the points and the observing vehicle. Smaller pixel values (darker) imply greater distances, whereas higher pixel values (brighter) suggest closer distances. It's important to note that although the depth images appear monochromatic, they are actually in RGB format. In this format, the pixel values in the red, green, and blue channels are identical. This setup minimizes variable alterations during data pre-processing, mitigating the risk of information compression or loss in the learning stage.

To obtain depth images under clear conditions, the DSOR filter [48] is applied to snow datasets. In this context, the typical approach of creating synthetic weather conditions in datasets for training is inverted but generates an artificial clear dataset. Given that filters cannot guarantee absolute precision and recall in the de-snowing process, the filtered result cannot be considered equivalent to the ground truth. However, the filter still provides a valuable sample pool for unpaired training. The main reasons for choosing DSOR over alternative filters are its rapid processing speed and its excellent capacity to retain as many environmental elements as possible when the filter parameters are set to an uncompromising level [48].

Training, testing, and data processing are conducted utilizing the Pytorch framework. All possible combinations of ResNet residual blocks (ranging from 4 to 9) in G_{SC} and G_{CS}, and convolutional layers (ranging from 1 to 4) in D_{nA} and D_{nB} are initially examined, and the most optimal combination is identified. When variables are kept constant, a combination of 4 ResNet residual blocks in G_{CS}, G_{SC} and 2 downsampling convolutional layers in D_{nA} and D_{nB} produce the most superior transformation result. In the model used to present the optimal results, the weights of adversarial losses, cycle consistency loss, and depth loss were set to 10, and the weight of SSIM

loss was set to 0.5.

Square-shaped samples randomly cropped from the depth images are input to two NVIDIA RTX 3090Ti graphics cards with a batch size of 8 for training. In the second half of the N-Layer Discriminator stage training, a linearly declining learning rate schedule starting from 0.02 until convergence is adhered to until the process converges.

The entire dataset which consists of 7000 samples was split in a 7.5:1:1.5 ratio for training, validation, and testing purposes. The quantitative analysis is conducted based on 500 samples under mild snowfall conditions and the other 500 samples under severe snow swirl conditions out of the testing dataset. The reported metrics in the following results all mean the average values.

3.5.2 Results

Qualitative results

Fig. 3.10 and Fig. 3.11 show the transformation results of proposed model under mild snow conditions, which means the majority of the snow is scattered noise points without the snow swirl phenomenon. (a) and (b) sets show 2 scenarios with the left column being the original snow scene from CADC, and the right column being the L-DIG clear results. Each scenario features an overall BEV in the top row, the clustered results showing the changes of snow clusters in the middle row; and the bottom rows showing magnified third-person views of the point cloud's central region, where the ego vehicle is situated. The same arrangement applies to the following qualitative results.

As indicated by the red arrows and encircled by red boxes, it's clear that the 'salt-and-pepper' noise points have been largely erased, with key environmental features left unaltered. Essential components like vehicles (outlined in green) are not only well preserved but also exhibit a level of point enhancement, as demonstrated in the Clear (a) set. Moreover, the road sign enclosed in the red box of (a) which was partially obscured by snow points in the earlier image, seems to be better defined, a testament

to the deep scene comprehension facilitated by the proposed model.

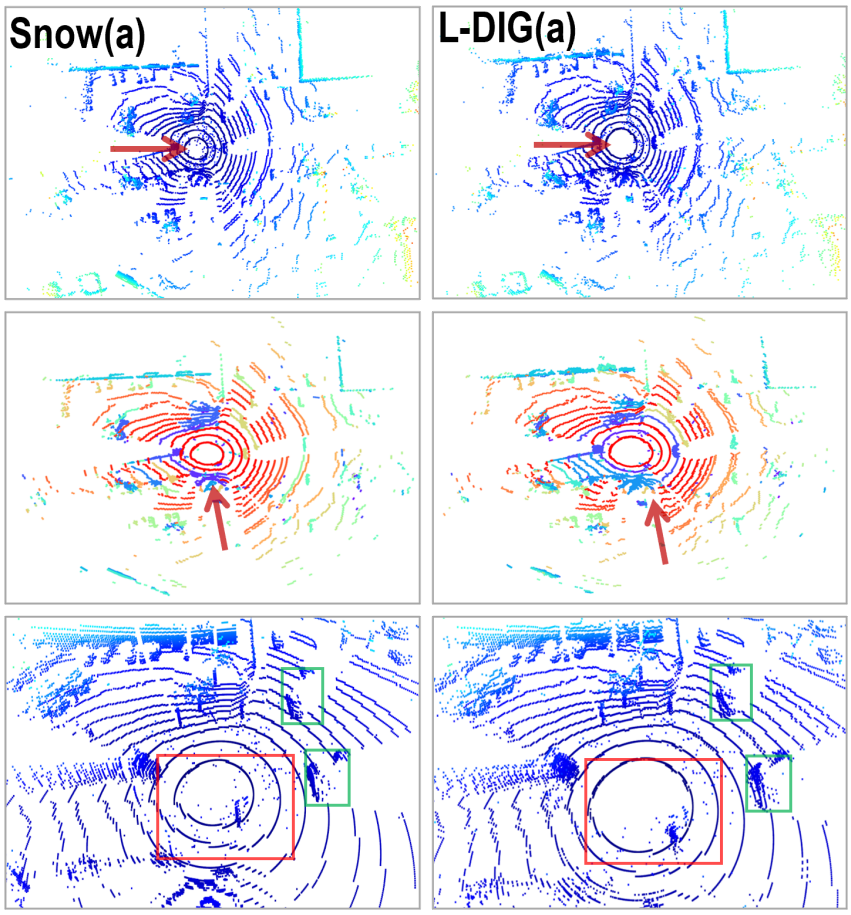


Figure 3.10: 1st set of point cloud transformation results from mild snow conditions to L-DIG clear, with colors encoded by height. First row - BEV scenes; middle row - clustered results; bottom row - enlarged third-person view center part around the ego vehicle. Red boxes and arrows - locations where snow's effects are alleviated; green boxes - vehicles

Fig. 3.12 and Fig. 3.13 demonstrate the transformation outcomes of proposed model under fierce snow conditions, characterized by the presence of snow swirls around the ego vehicle. Two distinctive scenarios (c) and (d) have been chosen for illustration, and are presented in the same format as in the previous. In these harsh snowy conditions where the snowfall has dramatically increased, it becomes easier to observe that the light-colored airborne snowdrifts (highlighted in shades of red, green, yellow, and cyan) have been substantially mitigated, as indicated by the red arrows. Also, the snow swirl clusters are largely mitigated as shown in the reduced purple

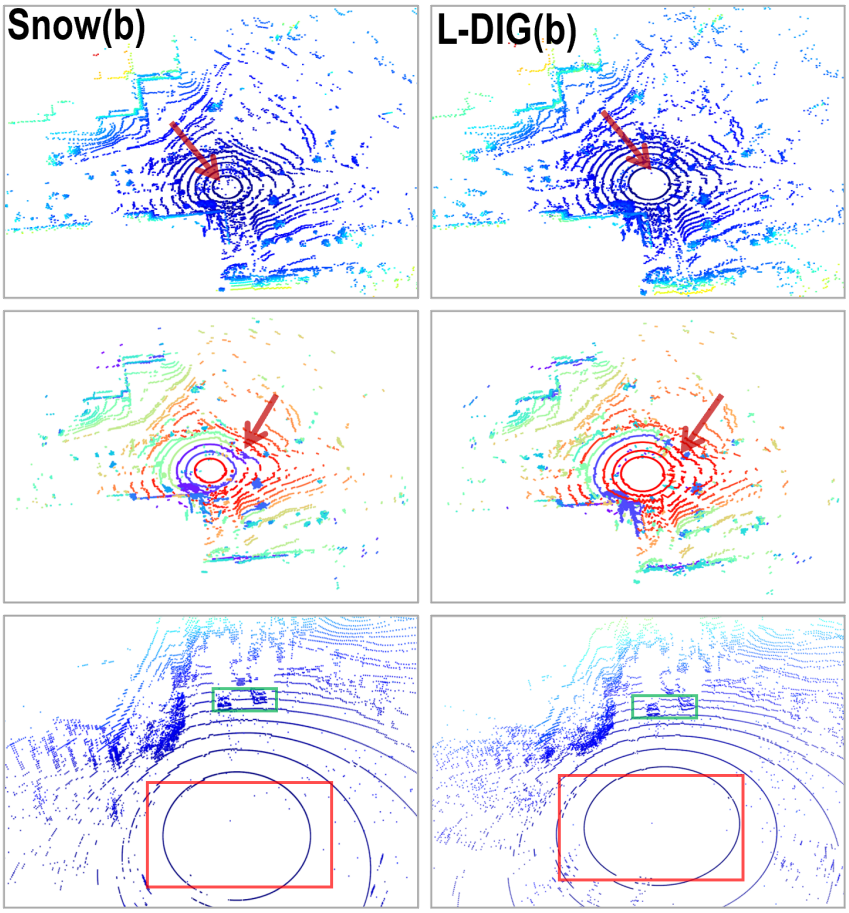


Figure 3.11: 2nd set of point cloud transformation results from mild snow conditions to L-DIG clear, with colors encoded by height. First row - BEV scenes; middle row - clustered results; bottom row - enlarged third-person view center part around the ego vehicle. Red boxes and arrows - locations where snow's effects are alleviated; green boxes - vehicles

snow clusters around the center.

Under these severe snow circumstances featuring dense snow swirl clusters, the attention is more on the noise reduction near the ego vehicle, as indicated by the red boxes, instead of entirely eradicating the snow swirls, as this could lead to a loss of important environmental elements. A balance between significant snow removal and effective preservation of objects is strove for, like the vehicles shown in the green boxes. Simultaneously, a certain degree of point cloud restoration can also be observed near the central ground rings, as shown in the enlarged part of set (c). This can be credited to the profound comprehension of the scene by the transformation model.

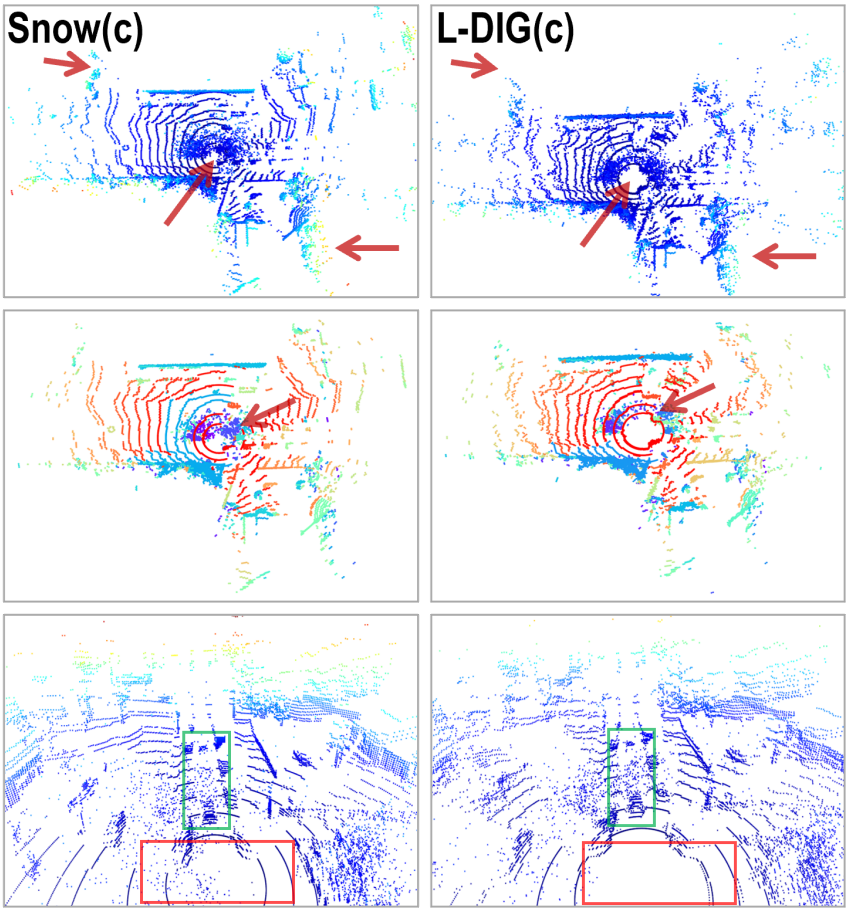


Figure 3.12: 1st set of point cloud transformation results from fierce snow conditions to L-DIG clear, with colors encoded by height. First row - BEV scenes; middle row - clustered results; bottom row - enlarged third-person view center part around the ego vehicle. Red boxes and arrows - locations where snow's effects are alleviated; green boxes - vehicles.

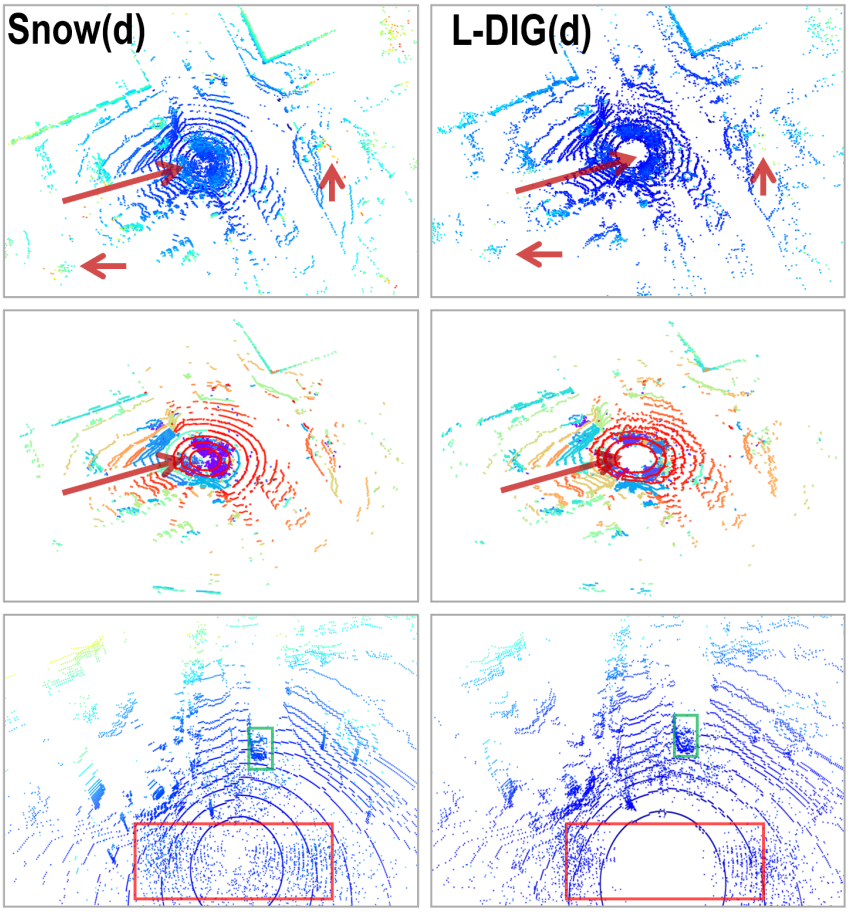


Figure 3.13: 2nd set of point cloud transformation results from fierce snow conditions to L-DIG clear, with colors encoded by height. First row - BEV scenes; middle row - clustered results; bottom row - enlarged third-person view center part around the ego vehicle. Red boxes and arrows - locations where snow's effects are alleviated; green boxes - vehicles.

Table 3.3: The 3D clustering metrics (avg.) comparison between transformed L-DIG clear conditions and different snow conditions.

Snow conditions	**Mild Snow**		**Fierce Snow**	
Items	CADC	L-DIG	CADC	L-DIG
Noise Number	2865.07	2689.46	2598.17	1606.23
Cluster Number	964.23	792.23	954.24	639.99
Reachability Distances	0.4076	0.3766	0.3805	0.3293
Inter-Cluster Distances	51.4262	53.9163	45.9245	47.4392
Size of Clusters	14.1603	15.8366	14.9750	24.2110
Davies-Bouldin Index	4.0279	3.7427	5.4138	3.9073
Silhouette Score	-0.3011	-0.2111	-0.3699	-0.3109

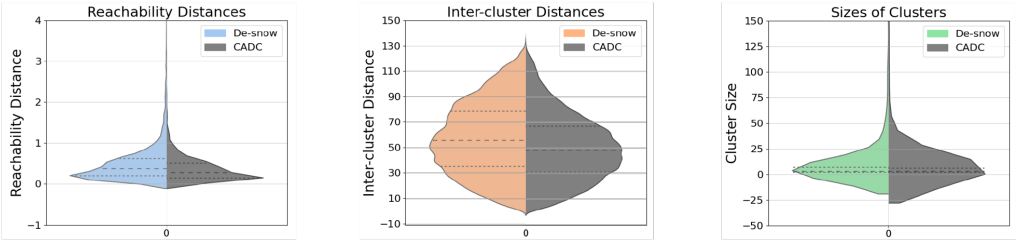


Figure 3.14: Violin plots for L-DIG transformation results under **mild snow** conditions. Limits on the y-axes are set for a better illustration of the distributions.

3D clustering results

The 3D clustering results are presented in Table 3.3. For mild snow conditions, the noticeable reduction in the average noise number, cluster count, and overall reachability distances in the clear results strongly suggests the effectiveness of the transformation process. As the majority of clusters now comprise object points and environmental features that are more densely and uniformly packed, the average inter-cluster distances, and average cluster sizes naturally increase. This shift in cluster characteristics is a byproduct of fewer, but more meaningful, clusters primarily representing substantive elements of the environment rather than scattered snow points. Similarly, the declines in the DBI and Silhouette Score are in line with the expectations for the de-snowing process.

In the violin plots of Fig. 3.14, the colored data on the left represents de-snowed clear data, while the gray data on the right serves as a comparison from the CADC

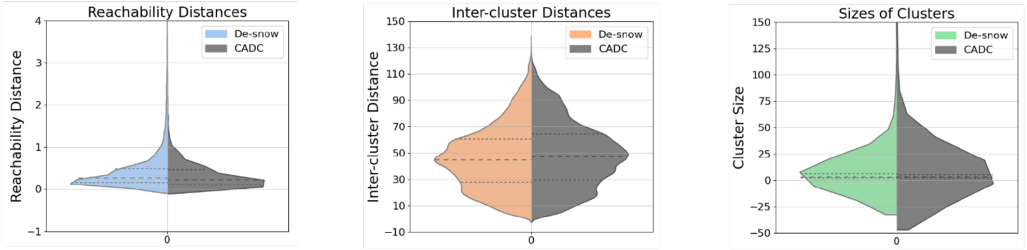


Figure 3.15: Violin plots for L-DIG transformation results under fierce snow conditions. Limits on the y-axes are set for a better illustration of the distributions.

dataset. This arrangement is consistent across all subsequent violin plots. A glance at the better evenness within the cluster distribution on the left half of each violin plot reveals the improvement of the de-snowing process compared to the slightly skewed distribution on the right. This observation is further substantiated by the lower skewness of the de-snowed distributions. Calculations show that for the reachability distances, inter-cluster distances, and sizes of clusters, the skewness values for the de-snow data are 8.11, 0.23, and 16.10, respectively, while for the CADC data, these values are 9.64, 0.30, and 21.49. Note that the median reachability distance of the de-snow is a little bit higher than with snow. This small anomaly originates from a few detached clusters at a remote distance after de-snowing, which can be seen from very few sample points exceeding the upper limit of the y-axis.

For the fierce snow condition with snow swirls, the transformation effects are applied to more clusters spanning the entire scene during heavy snowfall, all metrics largely veer towards less noise and tidier clustering results in the process. From Fig. 3.15, one can tell that the shifts in quartile lines are less prominent, which can be attributed to the fact that snow swirls typically have capacities similar to those of object clusters. Nevertheless, the efficacy of the de-snowing process is evidenced by the smoother and more consolidated distributions in the violin plots. This assertion is additionally validated by the slightly improved skewness of the de-snowed data which stand at 8.87, 0.38, and 28.04 respectively. Conversely, for the CADC data, these values are 10.45, 0.42, and 32.76.

Table 3.4: Ablation study on the transformation model from adverse to clear conditions.

Ablation conditions	Proposed model	Without D_p	Without SSIM loss	Without depth loss	Basic CycleGAN	Domain gap
Noise Number	**2689.46**	2882.94	2935.19	3879.13	2891.98	5031.85
Cluster Number	792.23	**761.25**	772.25	1097.3	833.47	1132.72
Reachability Distances	**0.3766**	0.4373	0.4511	0.5597	0.4863	0.7663
Inter-cluster Distances	**53.9163**	52.4850	53.1700	46.1819	49.9774	47.2124
Size of Clusters	15.8366	**16.3089**	15.7794	9.9985	14.0490	10.1896
Davies-Bouldin Index	**3.7427**	4.0880	4.1745	5.4330	5.1194	7.1288
Silhouette Score	**-0.2111**	-0.3532	-0.3394	-0.2990	-0.3638	-0.2825

Ablation study

To affirm the significance of the model's key components, an ablation study is conducted using the transformation model under mild snow conditions. This study investigates the impact of the absence of the pixel-attention Discriminator, SSIM loss, depth loss, and the basic CycleGAN. Additionally, a training pair with a considerable domain gap is examined. For this purpose, 6000 frames from the LIBRE dataset [10] were selected, which was collected under clear conditions in Nagoya, Japan's urban area. This choice serves as a representative due to the substantial domain disparity between Canada and Japan in terms of scenario layouts and traffic patterns. The CADC dataset contains a large portion of suburban scenarios with fewer buildings and more vegetation, which hardly appears in the LIBRE dataset. Table 3.4 presents the results, using the proposed model as a reference.

The absence of the pixel-attention Discriminator results in an immediate degradation in the performance, as evidenced by the increased noise number and reachability distance. Failing to remove a certain amount of solitary noise points substantiates the importance of the pixel-attention Discriminator in de-snowing.

More noise points are observed in the scenario without SSIM loss. Apart from the slightly reduced cluster number, other metrics especially the elevated reachability distance indicate a breakdown in structural integrity during the transformation process. A primary objective of the proposed model is to maintain the crucial objects

and environmental elements as effectively as possible, thus affirming the critical role of SSIM loss.

The scenario without depth loss indicates a complete failure in de-snowing, as evidenced by the significant plummeting in all metrics toward noisy and poor clustering. The cause of this failure lies in the unique properties of depth images, which are highly sensitive to non-linear scale changes during the conversion back to point clouds. Consequently, the depth loss forms the cornerstone of the transformation model based on depth images.

In the basic CycleGAN model, the mediocre statistics could be interpreted as an utter ineffectiveness in point cloud transformation, without managing to preserve the original states either. This result underscores the necessity of all the components in the proposed model for achieving successful transformation outcomes.

Finally, when trained on datasets with a substantial domain gap, the model does not yield satisfactory transformation performance. This is suggested by the exceedingly high noise number, reachability distances, and low cluster sizes, at least under the same parameter settings as before. The unjustifiably high noise and cluster numbers are the result of poor clustering, which is corroborated by the exceedingly high DBI. This result, derived under extreme conditions, serves to confirm the judicious decision to generate unpaired clear point cloud data with filters. However, it does not necessarily suggest that the proposed model lacks generality. Despite this, the model's robustness against domain gaps does stand as a major limitation of the current transformation model.

3.5.3 Comparisons

Adverse effects removal efficiency

To evaluate the overall capability of the L-DIG model in removing adverse effects compared to other adaptive methods, a precision-recall analysis was carried out, as detailed in (3.17) and (3.18). Diverging from the conventional approach of manually inspecting each point in the point clouds used in prior studies, the point cloud is

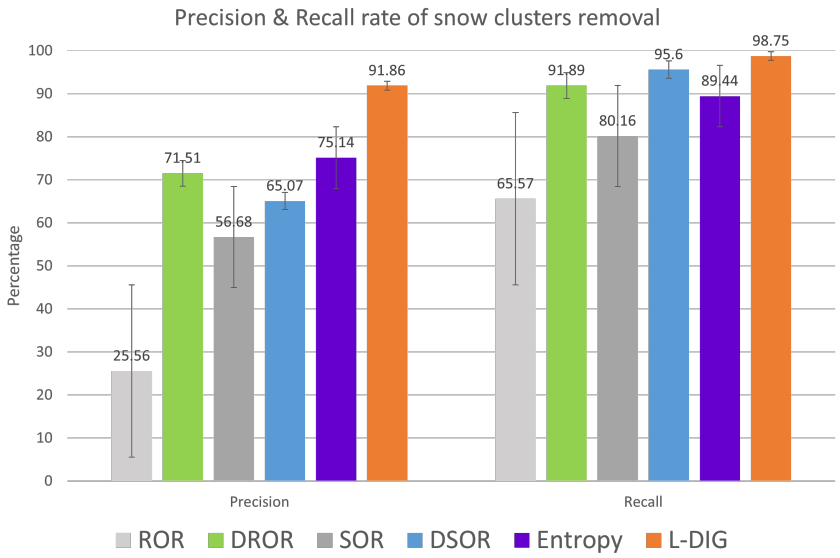


Figure 3.16: Precision and recall rates comparisons of adverse effects removal based on snow clusters.

examined based on clusters. This approach significantly reduces the need for manpower and time, while ensuring accuracy beating the 85% reliability rate typically associated with human annotation.

$$\text{Removal Precision} = \frac{\text{removed clusters} \cap \text{labeled as snow}}{\text{total removed clusters}} \tag{3.17}$$

$$\text{Removal Recall} = \frac{\text{removed clusters} \cap \text{labeled as snow}}{\text{total labeled snow clusters}} \tag{3.18}$$

A subset of samples from the test dataset was selected impartially, and each cluster within these samples was manually labeled to distinguish whether it consisted of snow clusters or not. It is important to note, however, that achieving a recall rate of 1 is always theoretically possible, but comes at the cost of losing vital environmental features, leading to considerably reduced precision. Consequently, the parameters of all methods involved in this study have been calibrated to prioritize maintaining scene integrity as much as possible. As illustrated in Fig. 3.16, both the DROR and DSOR methods significantly outperform their foundational methods. However, they plateau at a precision of 70% and a recall of 90% level, constrained by the challenging balance between effective removal and preservation of scene integrity. Similarly, the Entropy-

based adaptive filter struggled to find a more optimal balance. In contrast, the L-DIG transformation model stands out as the only method capable of transcending this trade-off, achieving notably high precision and recall rates with minimal error tolerance.

Perceptual improvement

For sensor perception result quality, the ultimate benchmark often lies in its plausibility to human observers. Thus, a Human Perceptual Examination (HPE) is conducted to compare the raw point cloud with outputs from the proposed model and other transformation models. This exercise aims to validate the effectiveness of proposed model in improving LiDAR perception.

In alignment with established protocols in image translation quality evaluation [111] [112], 20 human annotators with adequate practice in point cloud were divided into 5 groups and were provided with a sequence of trials featuring original and transformed point clouds from the CADC dataset. In each trial, annotators first engaged in object detection labeling on 10 frames of the original point cloud with adverse conditions from the CADC dataset. The labeling action is simplified into the confirmations of whether an object including vehicles and pedestrians has been detected to ensure reliability. The detection results of the annotators will undergo the process of looking for matches with the ground truth provided by CADC and then be used for precision and recall calculations. The precision and recall of the HPE are calculated as follows:

$$\text{Detection Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\text{TP}}{\text{all detections}} \tag{3.19}$$

$$\text{Detection Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{all ground truth}} \tag{3.20}$$

where True Positive (TP): a correct detection, False Positive (FP): a wrong detection, and False Negative (FN): a ground truth not detected. The results were presented as percentages, which can be found detailed in Table 3.5.

Table 3.5: Human Perceptual Examination on transformation models

Model	Precision (%)	Recall (%)
CADC (untransformed)	92.3% ± 6.5%	72.2% ± 8.5%
DROR	92.6% ± 2.8%	71.3% ± 3.6%
DSOR	93.1% ± 2.1%	67.5% ± 6.3%
Basic CycleGAN	**97.5**% ± 3.9%	77.5% ± 2.3%
CUT	66.4% ± 11.9%	67.2% ± 9.1%
Proposed model	94.4% ± **1.9%**	**98.1%** ± **1.8%**

Subsequently, they worked on another 10 frames from different scenarios where the adverse conditions had been converted to clear conditions using various models. To ensure unbiased results, each session focused on evaluating a single model, and no annotator was exposed to more than one model, thereby maintaining a 100% bias-free environment in the assessment process. All point clouds were displayed in the Open3D environment and each model was tested by more than one annotator so that average statistics with tolerance was obtained.

In the analysis of the CADC dataset, a precision rate of 92.3%, tempered by a few false detections attributed to snow clusters can be observed. However, the recall rate was considerably lower at 72.2%, primarily due to the challenging adverse conditions which also contributed to a high degree of tolerance in the measurements. Despite state-of-the-art (SOTA) filters enhancing the point cloud quality by removing most snow clusters—often misinterpreted as 'ghost objects'—they were not effective in improving the recall rate. In fact, the wrongful elimination of pertinent points by these filters often led to a decrease in recall, as it made the recognition of missed objects even more challenging. This limitation largely stems from the filters' sole emphasis on removing snow points, without adequately addressing the loss of points due to snow cluster occlusion, nor effectively preserving the integrity of the original point cloud.

On the other hand, the CUT (Contrastive Unpaired Translation) [113] model failed on both ends compared to the original data, as it overly disrupted the scene's structure. In contrast, the basic CycleGAN model showed a better recall rate and the highest original precision, but at a limited level and a larger margin of error. It is the

proposed model that outperformed all, achieving not only the highest precision (tolerance included) but also a huge leap in recall rate with the lowest tolerances. This superior performance can be attributed to its effective transformation capabilities, exceptional maintenance of structural integrity, and ability to reconstruct occluded objects. The achieved precision rate of 94.4% seems to be a trivial improvement, but it can be attributed to the overall increase of 'all detections' and instances where certain objects were excessively occluded, leading to mislabeling. Even though these objects were captured by the annotators because of the regular edges forming by the compensating effects of the proposed model, it sometimes still couldn't help confirm their original categories. Therefore, although the substantial improvement in recall greatly reduces the likelihood of missing ground truth elements, it increases the difficulty of achieving high TP as well. The current results represent the upper limit of precision achievable in adverse conditions with transformation models.

The HPE further validates the findings of this research, indicating that the proposed transformation model significantly enhances LiDAR point cloud perception under adverse conditions. This improvement is critical for developing more reliable and accurate autonomous driving systems capable of navigating through challenging environmental scenarios.

3.6 Conclusion

In this chapter, an innovative 3D clustering algorithm has been introduced, customized specifically for analyzing the complexities that adverse conditions introduce into 3D point clouds. This algorithm not only adapts to different forms of adverse effects but also offers a suite of quantitative metrics for a holistic evaluation of data conditions. Furthermore, various adaptive filters on authentic adverse datasets have been rigorously assessed and an advanced filter utilizing entropy as the primary metric has been subsequently designed.

Building upon these foundational techniques, the L-DIG model has been developed, a GAN-based approach that uses depth image priors for point cloud transforma-

tion. The proposed model exhibits an in-depth understanding of adverse condition characteristics and shows proficiency in the effective removal of noise, maintaining the integrity of the driving scene. The unique discriminator structure and loss functions incorporated into the model further underscore its robustness and adaptability. Extensive quantitative analysis and comparisons have been conducted among the proposed model and others. The statistics indicate that the proposed approach not only excels in efficiently removing adverse effects but also significantly enhances perception, achieving superior levels of both precision and recall.

The methodologies and findings documented in this chapter lay the groundwork for further in-depth studies, potentially setting new benchmarks in point cloud processing and autonomous vehicle operation under adverse conditions. This research opens the door to potential advancements in both machine learning and sensor-based technologies for autonomous vehicles operating in adverse weather conditions. While strides in this area have been made, there are still several aspects that warrant further investigation and development. Consequently, enhancing LiDAR perception in adverse conditions may require a different approach, specifically focusing on the expansion of datasets that include adverse conditions. This alternative strategy could provide a more effective means of improving accuracy under challenging environmental scenarios.

LiDAR point cloud transformation under adverse conditions has consistently faced the challenge of lacking reliable adverse data. Given the difficulty in acquiring paired or quasi-paired data under both adverse and clear conditions, the current model must strike a balance between the strength of translation and model stability, which subsequently leads to domain sensitivity. Moreover, the limited resources of the CADC dataset intensify the adversity for training and testing. To address these limitations, the future goal of this study is to develop the capability to generate high-quality paired data under adverse conditions. This aim is to augment the LiDAR point cloud with adverse effects based on a deep understanding of the driving scene, with the ultimate intention of preserving the original state of the scene to the greatest extent possible.

Chapter 4

Expanding Adverse Condition Data by LiDAR Point Cloud Augmentation

4.1 Overview

4.1.1 Adverse dataset status

Adverse conditions perception and classification research can't be done without adverse datasets. Many features used in object detection tasks need to be extracted from datasets and almost every algorithm needs to be tested and validated on datasets. In order to better solve the adverse weather problems in autonomous driving, it's essential to have enough data covering each kind of weather. Unfortunately, the majority of the datasets commonly used for training do not contain too many conditions different from clear weather. Some famous datasets that were collected in tropical areas like nuScenes [114] contain some rain conditions in Singapore, A*3D [115] has rain conditions at night, and ApolloScape [116] includes some strong light and shadow conditions.

A summary of the weather conditions coverage and the sensors used for collection in each dataset is shown in Table 4.1. The first half of the table shows datasets featuring LiDAR point cloud data. It becomes apparent that when contrasted with camera-based datasets, those involving LiDAR data under adverse conditions are no-

tably less. Based on the weather support status collected in Table 4.1, rain conditions can be considered adequate in current autonomous driving datasets, while fog and snow are not so much. Fog or haze is not time-sustained weather that is easy to encounter during data collection, so normally fog datasets are acquired from test facilities or simulators. As for snow, due to the difference between falling snow and accumulated snow, the qualities of the snow conditions contained in current datasets vary largely. And since the obvious difficulty of constructing an artificial snow environment compared to rain, experimental facilities' snow condition supports are very much limited. Furthermore, the strong light and contamination supports are seriously lacking in datasets, even rarer in simulators and facilities, which makes the research in this area relatively short. Therefore, as rich as the dataset resources are getting, the limitations on weather support are still realistic problems for perception and sensing research in adverse weather.

4.1.2 Adverse conditions classification

Perception enhancement fundamentally enables ADS to navigate through various inclement conditions, but it mainly focuses on how to ignore the interference or compensate for the negative effects. At some point, it's also important to do adverse conditions classification as a way to sense the surrounding conditions. Karlsson et al. [144] did an estimation on the real-time rainfall rate out of automotive LiDAR point cloud under both static and dynamic conditions in a weather chamber using probabilistic methods. At first, adverse conditions classification was limited to binary classification like distinguishing clear or not [145] on single images. Further machine learning techniques like kernel learning achieved multi-class condition classifications including sunny, rain, fog, and static snow. At this stage, the classification task is realized by setting classifiers with the unique features of each kind of condition. Sunny features come from the clear sky region of a picture and form a highly multi-dimensional feature vector; when sky elements are not included in the picture, a strong shadow region with confident boundaries becomes the indicator of sunny conditions. Rain streak is hard to capture in images so HOG features are extracted

Table 4.1: Coverage of adverse conditions in common autonomous driving datasets

Dataset	Synthesis	Rain	Fog/Haze /Smog	Snow	Strong Light/ Contamination	Night	Sensors
LIBRE [10]	-	✓	✓	-	✓Strong light	-	10 LiDARs, Camera, IMU, GNSS, CAN, 360° 4K cam, Event cam, Infrared cam
CADCD [16]	-	-	-	✓	-	-	1 LiDAR, 8 Cameras, GNSS, IMU
Oxford RobotCar [117]	-	✓	-	✓	-	✓	3 LiDARs, 3 Cameras, Stereo cam, GPS
nuScenes [114]	-	✓	-	-	-	✓	1 LiDAR, 6 Cameras, 5 Radars, GNSS
Argoverse [118]	-	✓	-	-	-	✓	2 LiDARs,7 Cameras ring, 2 Stereo cams, GNSS
Waymo Open [119]	-	✓	-	-	-	✓	5 LiDARs, 5 Cameras
A*3D [115]	-	✓	-	-	-	✓	1 LiDAR, 2 Cameras
ApolloScape [116]	-	✓	-	-	✓Strong light	✓	2 LiDARs, Depth Images, GPS/IMU
NCLT [120]	-	-	-	✓	-	-	2 LiDARs, 1 Camera, GPS, IMU
KAIST multispectral [121]	-	-	-	-	✓Strong light	✓	1 LiDAR, 2 Cameras, 1 Thermal (infrared) cam,IMU, GNSS
DENSE [42]	-	✓	✓	✓	-	✓	1 LiDAR, Stereo Camera, Gated Camera, FIR Camera, Weather Station
A2D2 [122]	-	✓	-	-	-	-	5 LiDARs, 6 Cameras, GPS, IMU
Radiate [123]	-	✓	✓	✓	-	✓	1 LiDAR, 1 stereo camera, 360°radar, GPS
EU [124]	-	-	-	✓	-	✓	4 LiDARs, 2 stereo cameras, 2 fish-eye cameras, radar, RTK GPS, IMU
WADS [125] [126]	-	✓	-	✓	-	✓	3 LiDARs, 1 camera, 1 NIR camera, 1 LWIR camera GNSS, IMU, 1550 nm LiDAR
Boreas [127]	-	✓	-	✓	-	✓	1 LiDAR, 1 camera, 1 360 °radar, GNSS-INS
GROUNDED [128]	-	✓	-	✓	-	-	1 LiDAR, 1 camera, RTK-GPS, LGPR
Berkley DeepDrive [129]	-	✓	✓	✓	-	✓	Cameras
Mapillary [130]	-	✓	✓	✓	-	✓	Mobile phones, Tablets, Action cameras,
EuroCity [131]	-	✓	✓	✓	-	✓	2 Cameras
D2-City [132]	-	✓	✓	✓	✓Contamination	-	Dashcams
DDD17 [133]	-	✓	-	-	-	✓	Dynamic and active-pixel vision Camera
Snowy Driving [134]	-	-	-	✓	-	-	Dashcams
ACDC [135]	-	✓	✓	✓	-	✓	1 Camera
4Seasons [136]	-	✓	-	-	-	✓	1 Stereo Camera, GNSS, IMU
Raincouver [137]	-	✓	-	-	-	✓	Dashcam
WildDash [138]	-	✓	✓	✓	-	✓	Cameras
SoilingNet [93]	-	-	-	-	✓Contamination	-	Cameras
HSI-Drive [139]	-	✓	✓	-	-	✓	1 Photonfocus 25-band hyperspectral camera
DAWN [140]	-	✓	✓	✓	✓Sandstorms	-	Google and Bing search engines
Foggy Cityscape [99]	✓	-	✓	-	-	-	-
SYNTHIA [141]	✓	-	-	✓	-	-	-
P.F.B [142]	✓	✓	-	✓	-	✓	-
ALSD [143]	✓	✓	-	✓	-	✓	-

from the image to be the rain feature vector. Falling snow is considered noise, while pixels with certain gray levels are defined as snowflakes. Haze is determined by dark channels, where some pixels have very low intensities in at least one color channel which is the dark channel [146]. With the development in AI technologies, machine learning neural networks such as deep CNN are used by Elhoseiny et al. [147] in this task to enhance feature extraction and learning performance.

In meteorology, rain is observed and measured by weather radar and stationary rain gauges. Considering carrying a weather station on a car like [42] is not practical for commercial generalization, people started in an early stage to realize vehicle-based binary (wet/dry) precipitation observations [148] [149]. Goodin et al. [150] tried to establish the relationship between the two parameters: rain rate, as manifested by the rain scattering coefficient, and the max range of the LiDAR sensor for a 90% reflective target in clear conditions, and successfully generated a quantitative equation between rain rate and sensor performance. Bartos et al. [151] raised the idea of producing high-accuracy rainfall maps using windshield wipers measurement on connected vehicles in 2019. It's a very leading concept considering the network of connected vehicles has not been constructed on a large scale. Simply the status (on/off) of windshield wipers serves as the perfect indicator of binary rainfall state compared to traditional sensing methods like rain gauges. This work is supposed to help city flash flood warnings and facilitate stormwater infrastructure's real-time operation, but the involvement of cars provides a line of thought on vehicle-based rain sensing.

Heinzler et al. [152] achieved a pretty fine classification with a multi-echo LiDAR sensor only. The point cloud is first transformed into a grid matrix and the presence of rain or fog can be easily noticed by the occurrence of secondary echoes on objects. Then, different from recording the echoes of each kind of condition, the mean distance of each echo and their mathematical properties like variance are used for detailed classification as the covariance matrices are influenced by different levels of rain or fog and the change in the point cloud or to say the matrix is visible. It can be imagined that the test result might not be as good when using a LiDAR sensor with a smaller vertical FOV due to the insufficient number of points and dynamic

scenarios compared to static scenes. That means this method still has its reliance on controlled environments and the robustness might not meet level 4 or higher autonomy requirements. Dannheim et al. [153] proposed to use the fusion data from both LiDAR and camera to do adverse conditions classification several years before. Their main classifier was based on the intensity difference generated by the backscattering effect of rain and fog and no neural network was mentioned in their image processing.

4.1.3 Adverse data augmentation

Compared to methods based on classification, those employing adverse conditions data augmentation demonstrate superior effectiveness and accuracy. This is largely due to the inherent difficulty and unreliability of annotating or labeling adverse conditions [83]. However, within current driving datasets, samples featuring adverse conditions represent only a small fraction. Furthermore, datasets specifically targeted at collecting adverse conditions in certain areas like snowy regions, often face significant domain gap issues when compared to other datasets. This discrepancy can lead to diminished model generality. As a result, augmenting clear condition datasets with simulated adverse conditions has become one of the most effective and popular approaches to enhance the capability of deep-learning-based models in addressing adverse challenges.

Precipitations are the most common weather conditions that can be encountered in a driving situation. Raindrops/streaks, fog clusters, haze, and snowflakes are all among the elements to be removed in a driving scene image. Generative models synthesize repetitive rain elements based on the rain's physical structures including direction, scale, and thickness [59], or translate the whole scene into rainy images [60] or videos [61] as the corrupted pairs of the existed clean data for training purpose, in order to eventually achieve the removal of rain. Fog or haze conditions are less easy to be treated as solid objects, hence image translations are widely applied to acquire adverse data. The hazy effects of an image can be realized with a professional fog/haze generator that is able to imitate real conditions of hazy scenes [85]. The quality or the fidelity of the hazy effects are either guided by physical priors [90] or ensured

by the parameters constraints in the translation models [71] such as Cycle-GAN [66]. With the support of data diversity, different levels and various combinations of adverse conditions can be generated [86] [101]. More diversity in the adverse data means more training samples, which ultimately serves the robustness of learning-based perception models.

The augmentation for snow conditions is currently at a less mature level than rain and fog due to the scarcity of snow presence in datasets. Attempts at the generation of snowflakes and realistic rendering across the entire driving scene have been made [87] in images and synthetic snow scenes have also been used for perception models evaluation purposes [62]. But the translation of a snow scene and the embodiment of different snow levels are still ongoing. Besides the common precipitation conditions, contamination of the sensor emitters is also a critical corruption to perception. Mud or water stains could obscure the camera lens and leave invalid images with large portions blank. The difficulty of acquiring such datasets is almost beyond description. As a result, models have been developed to generate these soiling effects on camera and augment the contamination datasets [92]. Not only the augmentation effects are almost as genuine as the real cases, but downstream removal models also rely on such data greatly for training purposes.

There are two main aspects of point cloud augmentation, object augmentation, and scene augmentation. The first mainly serves the completion of occluded objects in indoor environments, and the latter mostly is for the transformation of an entire scene, commonly involving jittering, rotation, shearing, etc. Such augmentations aim at improving the accuracy and robustness of downstream tasks, mostly focusing on semantic segmentation and object detection. For adverse conditions in driving scenes, it is a complicated mix of the two aspects of augmentation mentioned above. On one hand, the presence of weather phenomena such as water mist, agglomerated fog, and snow swirls sometimes manifest as solid "ghost" objects in the point cloud. On the other hand, scattered noise point distribution is the result formed from ego vehicle motion, vehicle interactions, environmental layouts, wind, etc. instead of a fixed pattern or predictable mathematical distribution. As a result, a designated aug-

mentation method is needed to better adapt to the features of the adverse conditions with a deep understanding of the driving scenario, thereby enabling the generation of paired adverse data.

LiDAR suffers the same influences that happen to cameras in terms of adverse conditions. Rasshofer et al. [80] reproduced the optical returns of the signals measured under weather conditions to analyze the influences of adverse conditions on LiDAR sensors. This is one of the primary attempts at data augmentation when adverse datasets were not common and laid a foundation for signal waveform processing in the matter of adverse condition problems [52]. The replication of the signal returns points out the essence of LiDAR data augmentation: Synthesizing perception results in the same way that real adverse conditions show to the LiDAR sensor. However, the realization of the adverse effects in point clouds still depends on the collections in weather chambers [45] because of the requirement for paired data. The low domain similarities between chambers and real roads aside, common experimental facilities with controllable precipitation rates across the world can hardly simulate complicated adverse conditions such as dynamic snowfall [6]. Therefore, it's necessary to develop a way to realize the adverse data augmentation in LiDAR point clouds similar to the methods in camera images mentioned above to provide robustness to the LiDAR perception ability in autonomous driving.

The integration of paired datasets that include both adverse effects and corresponding classifications is proven to significantly improve the detectors' proficiency in identifying smaller entities, like pedestrians and cyclists, in challenging driving conditions [45]. Hahner et al. [43] developed a method that simulates snow particles in a 2D space corresponding to each LiDAR line and adjusts each LiDAR beam's measurements based on the resulting geometry. In addition, they factored in ground wetness, a common occurrence during snowfall, in their LiDAR point clouds as a supplement of the augmentation. Based on testing and validation on commercial autonomous driving detectors conducted by the DENSE group [42], it has been established that the inclusion and enrichment of adverse effects within training datasets can enhance the detectors' average precision by a minimum of 2.1%, in contrast to training with

only clear condition data. The notable enhancement observed in the performance of 3D object detection subsequent to training on semi-synthetic snowy training data substantiates the successful simulation of snowfall. It is of particular importance to acknowledge that their snow augmentation approach predominantly focuses on light snowfall conditions under the rate of 2.5 mm/hr, wherein the prevalent snow effects in LiDAR point clouds manifest as dispersed noise points rather than snow clusters. Snow clusters pose a greater challenge to LiDAR perception in actual driving conditions and the primary focus revolves around the study of snow clusters as one of the main representatives of adverse conditions.

In this chapter, a LiDAR point cloud augmentation model based on conditional guides is proposed to realize the expansion of adverse condition data. Segmentation maps capturing adverse effect presences in the point cloud as conditional guides are produced, and effective fusion methods are designed to input both the raw data and conditional guide into the generative model for adverse effect classification and generation. Natural adverse effects are managed to be added into point clouds with understanding of the scene while maintaining structural integrity as much as possible for paired data consideration. The main contributions of this work are as follows:

1. Segmentation maps of adverse effects are produced via a designated 3D clustering algorithm. With partial human annotations and weak supervision, all the noise clusters and scattered noise points are automatically labeled in the segmentation maps. The adverse effects classification and conditional guides for generative models are provided through the labeled segmentation maps, providing the basis for learning-based conditional generative models.

2. Both an early and a late data fusion are designed for the integration of the raw data and segmentation data. Each fusion method emphasizes different stages of the conditional learning model and presents generation results with different strengths.

3. Natural adverse effects can be reproduced compared to real conditions and quasi-natural effects can be generated across huge domain gaps in terms of traffic

layouts and environments, presenting high-level robustness in the generation of paired adverse data.

4. Experiments conducted on 3D detection methods incorporating the proposed data augmentation scheme have demonstrated a notable enhancement in detection performance compared to the baseline model that lacks augmentation. This finding reinforces the importance of expanding adverse datasets as proposed, highlighting its crucial role in advancing the goal of perception improvement.

This chapter is structured as follows. The related research and background reviewing have already been introduced in the current section and Section 2.2. Section 4.2 explains the construction process of the segmentation map and conditional guide, and presents the methodology of the proposed conditional generative model architecture, including the details of early and late data fusion techniques. Section 4.3 shows the qualitative and quantitative results of generated adverse effects, including both the reproduction of real adverse effects and the quasi-natural generation of adverse conditions in Nagoya. Finally, Section 4.4 summarizes this work and provides discussions.

4.2 Conditional Generative Model

4.2.1 Clusters classification and segmentation map

In the task of adverse effects augmentation, the CADC dataset continues to be used for its excellent representation of snow conditions. To establish a solid baseline for classifying snow swirl clusters, 2691 samples that prominently feature snow clusters were meticulously chosen from the entire dataset. These samples consist of 120 different sequences, corresponding to 120 unique driving scenarios. These scenarios comprehensively cover aspects such as the ego vehicle's movement, interactions with other traffic participants, and wind influence.

For each of these scenarios, one representative frame is selected for manual annotation at the cluster level, as demonstrated in the previous Fig. 3.6. All snow clusters

Algorithm 2 Cluster-Based Snow Condition Classification

Input: Cluster Size, Center X, Y, Z coordinates, and Average Reachability Distance for each cluster.

Output: Labels for each cluster representing different snow conditions.

for each cluster in the dataset **do**
  if Cluster Size $\in [2, 500]$ **then**
    if Center $X \in [x_1, x_2]$ **and** Center $Y \in [y_1, y_2]$ **and** Center $Z \in [z_1, z_2]$ **and** Avg. Reachability Distance $\in [0.5, 2]$ **then**
      Label all points within this cluster as 1 (snow swirls).
    else
      Label all points within this cluster as 3 (objects).
    end if
  else if Cluster Size $\in [1, 1]$ **then**
    Label all points within this cluster as 2 (scattered snow).
  else
    Label all points within this cluster as 3 (objects).
  end if
end for

that were discernible to the human eye were manually identified and labeled. The 3D clustering algorithm was then employed to aggregate and analyze their spatial and clustering characteristics. Based on this weakly supervised approach, an automatic tool for classifying snow effects was developed. This tool operates according to Algorithm 2.

In Algorithm 2, X means the horizontal axis in BEV and Y means the vertical axis; while Z denotes the height axis (from ground to sky). The threshold values for the spatial XYZ coordinates and other metrics are determined through a comprehensive evaluation by the proposed algorithm. To ensure absolute accuracy in cluster identification, the thresholds are further corroborated by a secondary layer of human verification. Notably, the XYZ coordinates represent the central point of each cluster, thus eliminating the risk of misidentifying clusters on the margins.

CADC dataset was collected in Canada where vehicles drive on the right side of the road and oncoming vehicles will approach the ego vehicle from the left. This justifies the slightly closer threshold for Center X on the left (-7.7646) compared to the right (8.4354), relative to the ego vehicle. Additionally, as illustrated in Fig. 1.4b, the movement of the ego vehicle tends to disturb snow on the ground, creating whirls

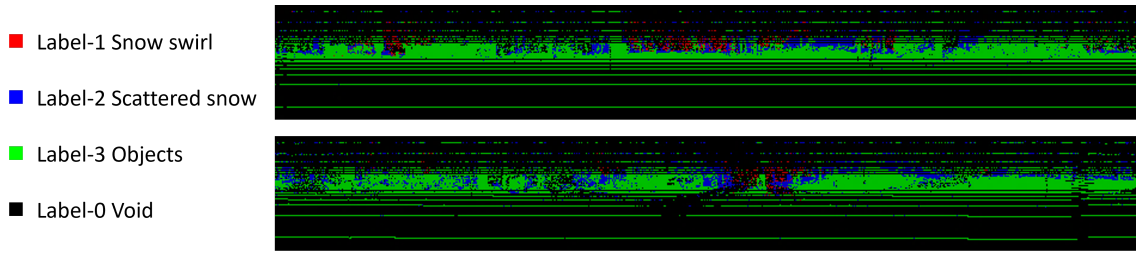


Figure 4.1: Examples of segmentation maps of CADC dataset. Images are rendered with pixel values multiplied by 64 under the OpenCV BGR environment for illustration purposes. Red points denote snow clusters; blue denotes scattered snow points; green denotes all the objects; black means void (no signal).

of snow clusters in its wake. Consequently, the threshold for Center Y at the rear (-10.4947) is found further from the vehicle compared to the front (7.5053). Outside of the interval of [0.5165, 1.9999] on the height direction, no clusters fit the criteria of snow within the given X and Y range, hence the Z thresholds. This self-adaptive approach is designed to accurately capture all snow clusters, theoretically eliminating the possibility of errors.

The outcomes from said classification tool are depicted in Fig. 4.1. For clarity in illustration, these images are rendered in the OpenCV BGR format, with pixel values amplified by a factor of 64. In these visualizations, red points indicate snow clusters, blue points represent scattered snow particles, green points denote all objects, and black signifies void areas. This classification equates to a segmentation map with four distinct labels, effectively illuminating the presence of various adverse conditions. This segmentation map not only provides a fundamental insight into the characteristics of adverse effects but also aids in the generation of such conditions in data augmentation.

Fig. 4.2 shows the architecture of the segmentation map alongside the workflow of the condition-guided augmentation model. This presentation offers a comprehensive view of how adverse conditions are identified, classified, and subsequently utilized in guiding data generation for model training purposes.

The methodology begins with raw data that includes adverse conditions, which is then processed using the 3D clustering algorithm to create a segmentation map. In

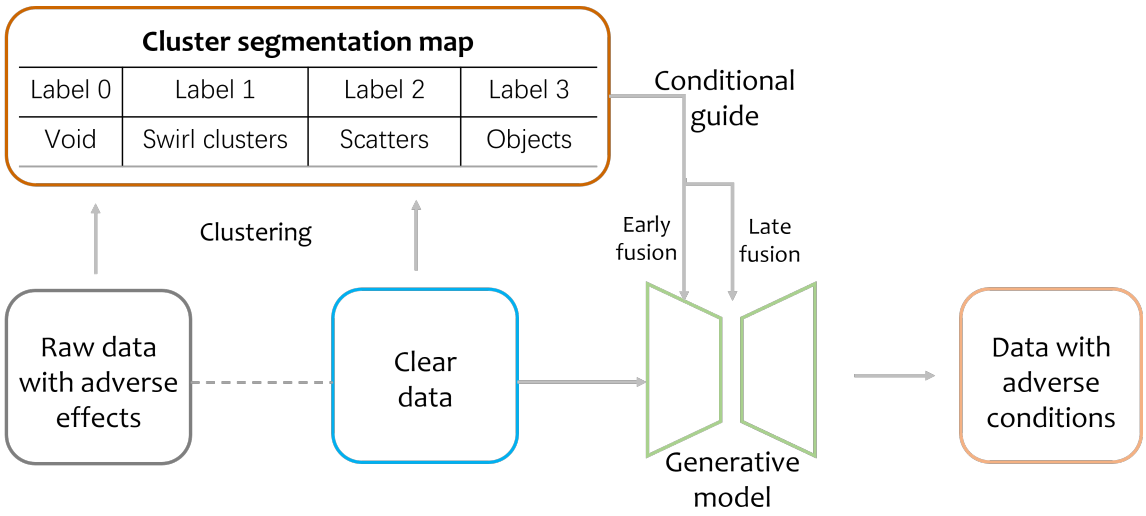


Figure 4.2: Architecture of the condition-guided adverse effects augmentation model based on segmentation maps. The clear data input is obtained from filtered raw adverse data to establish intrinsic correlation for optimal training. The cluster segmentation map serves as a conditional guide, which can be input into the generative model through both early and late data fusion. Data with adverse conditions are generated under the guidance of the segmentation map.

this map, four distinct labels are categorized. Labels 1 and 2 are assigned to noise clusters and individual noise points, stored in the Red and Blue channels respectively. Label 3 encompasses all other elements, such as objects and structures, essentially anything not classified as an adverse effect, stored in the Green channel. Label 0 represents void areas, characterized by the absence of reflective signals in the point clouds, stored in all three channels.

This segmentation map is then utilized as a conditional guide in the proposed generative model, alongside clear data. For optimal training outcomes, it is advantageous for this clear data to have some correlation with the raw data. This correlation facilitates a more intuitive understanding of the underlying logic in adverse effect classification. Therefore, filtered raw data are employed in this context.

The fusion of this data can be implemented at either early or late stages of the process, each requiring different structural techniques. Experiments with both approaches were conducted, and the results demonstrate varying strengths, which will be elaborately discussed in the subsequent sections.

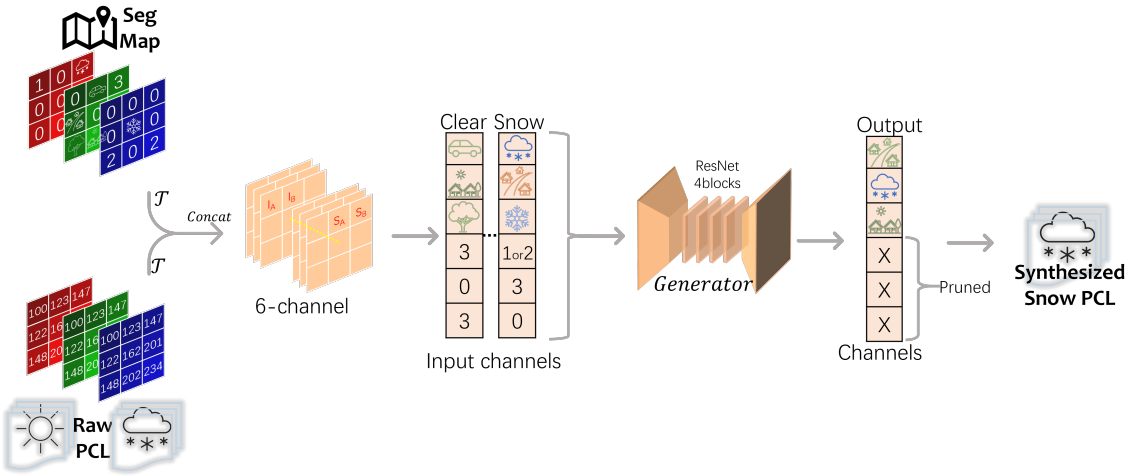


Figure 4.3: Diagram of the early fusion process for conditional augmentation in point clouds. A standardized transformation, $\mathcal{T}$, is applied to both point cloud inputs I_A, I_B and their segmentation maps S_A, S_B to ensure alignment and compatibility for concatenation. The process results in a 6-channel dataset for each domain, blending the original point cloud with segmentation labels. The generator G, equipped with a 4-ResNet-block structure, processes these datasets, producing snow-augmented point clouds O_A and O_B. These outputs retain the structure of the original datasets, with post-processing to revert to standard RGB format.

4.2.2 Conditional guide data fusion

Early fusion with dimension expansion

The core principle of early fusion with the segmentation map is to activate the effectiveness of the conditional guide right from the beginning. This ensures that the entire generation process is monitored and constrained within the parameters set by the guide. Leveraging the advantage of a shared RGB data format, the possibility of merging the segmentation map with the raw inputs is explored, as shown in Fig. 4.3.

The fusion process is initiated by setting a random transforming seed, $\mathcal{T}$, a standardized protocol to ensure that both the point cloud inputs from domains A and B — I_A, I_B — and their corresponding segmentation maps S_A, S_B undergo identical transforming process prior to being fed into the generator. This consistent transforming process is vital for aligning and rendering the datasets compatible for concatenation. Following this, a 6-channel dataset is constructed for each domain, with the first three channels comprising the original image and the latter three filled with the labels from

Algorithm 3 Early fusion process for conditional augmentation model

Input: Point cloud inputs I_A, I_B and segmentation maps S_A, S_B from domains A and B.

Output: Scene-adaptive augmented snowy point cloud O_A and O_B.

Define transforming $\mathcal{T}$ for clear inputs and segmentation maps.
Define generator G with input dimension of 6-channel (3 + 3).
for each input i in dataset **do**
    Apply consistent transforming $\mathcal{T}$ to I_A^i, I_B^i, S_A^i, S_B^i using a fixed random seed.
    Concatenate transformed inputs and segmentation maps along the channel dimension:
    $Input_A^i \leftarrow Concat(\mathcal{T}(I_A^i), \mathcal{T}(S_A^i))$
    $Input_B^i \leftarrow Concat(\mathcal{T}(I_B^i), \mathcal{T}(S_B^i))$
    Feed 6-channel $Input_A^i$ and $Input_B^i$ to the generator G.
    Generate outputs O_A^i and O_B^i with the same dimensionality as I_A^i and I_B^i.
    Prune additional channels from O_A^i and O_B^i to retrieve 3-channel output.
end for

the segmentation maps. This process is uniformly applied to both the clear and snowy datasets.

Incorporating a 4-ResNet-block structure, as detailed in Section 3.5, the input channels of the generator G are adjusted to accommodate the six-channel input. This modification facilitates the integration of the concatenated data. After the encoding-decoding phase within the generator, the output comprises snow-augmented point clouds, O_A and O_B, for each domain. These outputs retain the format of the original datasets, with the exception that the latter three channels, dedicated to labels, are subsequently pruned to revert the data back to the standard RGB format. This results in the generation of the synthesized snow point cloud.

Throughout this entire fusion process, the segmentation map remains integrally connected with the raw data, imparting directional guidance at every stage of the conditional generative model. This method ensures optimal guidance with minimal supervision, effectively circumventing potential issues of deviation or overfitting, which could arise from the relatively lower data weight of the labels compared to the pixels. The efficiency of this early fusion approach is thus underscored, as is its depiction in Algorithm 3.

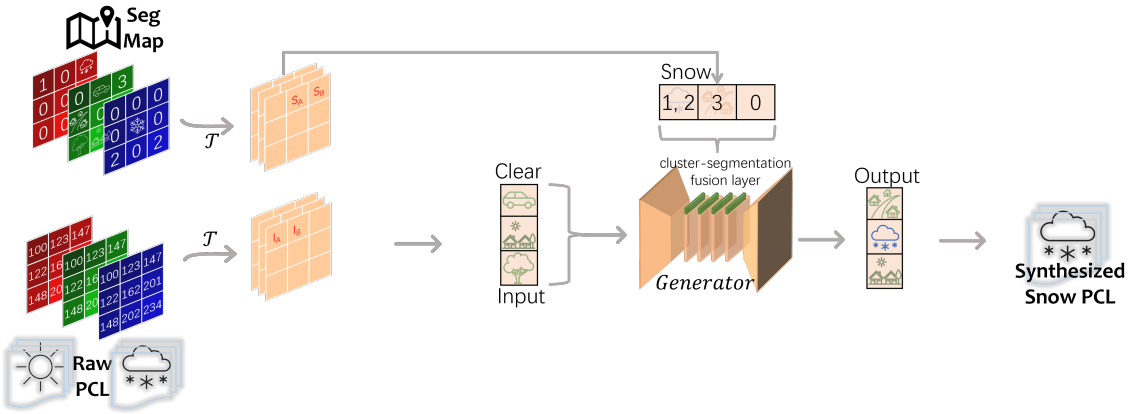


Figure 4.4: Diagram of the late fusion process for conditional augmentation in point clouds. Input I_A, I_B and their segmentation maps S_A, S_B go through the same standardized transformation, $\mathcal{T}$. Depth-conditioned features are extracted from the segmentation map S and sent to the cluster-segmentation fusion layer, to be further modulated with the feature map obtained from the original input of raw point clouds. The modulated feature map passes through the forward layers of the generator, resulting in the production of the scene-adaptive augmented snowy point cloud.

Late fusion with cluster-segmentation layer

In the realm of conditional augmentation, the late fusion approach delineated in Fig. 4.4 and Algorithm 4, stands in contrast to the early fusion methodology. This technique preserves the separation of input data and segmentation information until a pivotal moment in the generative process, allowing for a refined and adaptive integration of conditional guidance.

The late fusion technique for conditional augmentation begins with the distinct processing of the clear input C and the segmentation map S. The clear input undergoes initial processing through the shared layers of the generator G_{shared}, while the segmentation map is uniquely transformed through the cluster-segmentation fusion layer $\mathcal{F}$. This layer is critical as it allows for the discrete and specialized handling of the segmentation map, setting the foundation for the subsequent fusion process.

The segmentation map S is processed by $\mathcal{F}$ to yield depth-conditioned features D. These features are crucial in directing the augmentation process, as they encapsulate vital information about the segmentation that will guide the augmentation. From these features, scaling $Scale$ and shifting $Shift$ factors are derived, which become

instrumental in modulating the initial feature map extracted from the clear input.

Here, the clear input's feature map F, obtained through G_{shared}, is modulated through $F_{modulated}$ using the scaling and shifting factors derived from $\mathcal{F}$. This process effectively integrates the depth-conditioned segmentation information into the generative pathway, ensuring that the augmentation is aligned with the segmentation map's guidance.

The late fusion culminates as the modulated feature map $F_{modulated}$ passes through the forward layers of the generator $G_{forward}$, resulting in the production of the scene-adaptive augmented snowy point cloud O. This output not only reflects the original data's structure but also integrates the depth-conditioned segmentation features, providing a dynamic and adaptive augmentation. This late-stage fusion allows the initial features of the clear input to develop more independently, reducing risks associated with early overfitting or deviation. Moreover, the controlled and nuanced influence of the segmentation map, exerted through $\mathcal{F}$, potentially leads to more accurate and realistic augmentations.

In this approach, where the data weights of both layers are more closely matched, the model may encounter a level of effectiveness that challenges its stability. Given this heightened intensity, the possibility of the model becoming overwhelmed or even collapsing is not insignificant. To address this potential issue, an additional phase of conditional training, implemented after the initial training process, presents itself as a viable solution. This subsequent training stage can provide a safety net, offering the model a chance to stabilize and adapt to the high degree of effectiveness without compromising its overall performance.

4.2.3 Identity controlled generator

In the endeavor to augment snow effects onto clear point clouds, a key focus is placed on preserving the intrinsic structural integrity of the input data. To achieve this, besides the Depth loss and SSIM loss adopted from the previous research in Section 3.3.3, an approach similar to identity loss is incorporated , a mechanism that ensures the augmented output retains the fundamental characteristics of the original point

Algorithm 4 Late fusion process for conditional augmentation model

Input: Clear input C, segmentation map S.
Output: Scene-adaptive augmented snowy point cloud O.

Define cluster-segmentation fusion layer $\mathcal{F}$.
Define shared initial layers of generator G_{shared}.
Define forward layers of generator $G_{forward}$.
Process segmentation map S through $\mathcal{F}$ to obtain depth-conditioned features D.
$D \leftarrow \mathcal{F}_{condition}(S)$
Generate scaling factors $Scale$ and shifting factors $Shift$ from D.
$Scale \leftarrow \mathcal{F}_{scale}(D)$
$Shift \leftarrow \mathcal{F}_{shift}(D)$
Process input I through shared layers to get initial feature map F.
$F \leftarrow G_{shared}(I)$
Modulate initial feature map F with scaling and shifting factors.
$F_{modulated} \leftarrow (Scale \cdot F) + Shift$
Process modulated feature map $F_{modulated}$ through forward layers to get final output O.
$O \leftarrow G_{forward}(F_{modulated})$
return Output O with cluster-segmentation features.

cloud. This aspect is crucial for maintaining realism and accuracy in the synthesized snowy scenes. The identity loss function is shown below:

$$\mathcal{L}_{id} = \| G(I_A) - I_B \| + \| G(I_A) - I_B \| \tag{4.1}$$

where G is the generator, I_A and I_B represent the clear and snow input point clouds, respectively.

Central to this methodology is the strategic input of the target dataset into the model. By feeding the clear point clouds as inputs in their original form, the model learns a mapping that minimizes alterations to the inherent structure of these point clouds. The identity loss function acts as a regulatory mechanism, guiding the model to respect and preserve the original data's topology and spatial configuration.

During the training phase, the model processes the unaltered clear point clouds alongside the primary task of generating snow-augmented point clouds. This dual processing enables the model to compare the output against the original input, ensuring that the augmentation process does not compromise the essential structural

elements. The identity loss quantifies the deviation of the augmented output from the original structure, pushing the model to generate outputs where the addition of snow effects is seamless and natural, without distorting the underlying point cloud architecture.

The integration of identity loss in the proposed model serves to ensure structural fidelity, as it maintains the geometric and spatial characteristics of the original point cloud. It enhances the realism of the augmentation by constraining the model to respect the original data's structure, aligning the snow effects with real-world dynamics. Furthermore, this approach enhances the model's robustness, reducing the risk of overfitting to specific snow patterns or anomalies in the training data and thereby improving its generalizability.

4.2.4 Violations and solutions in LiDAR data augmentation

There are two primary types of discrepancies in augmented data: the physical realism of the data compared to actual real-world scenarios and the consistency of the data with the output formats of LiDAR sensors. Let's explore these two aspects:

First, real-world reflection. In data augmentation, synthetic data aim to replicate a specific target form. In LiDAR point cloud augmentation, this target is not the physical presence of adverse conditions but rather their representation within point clouds. For instance, a large cluster of snowflakes in mid-air might create a 'ghost object' and leave voids in the point cloud due to significant signal absorption. Therefore, the augmented representation of such a snowflake cluster should focus on replicating this 'ghost object' and the accompanying voids, rather than the actual physical structure of the snowflakes. While there is an inherent discrepancy between the augmented and actual physical forms, the primary objective of augmentation is not physical accuracy but replicating the perceptual effects as they would be captured by LiDAR sensors. As long as the augmentation accurately reflects these effects, it is deemed successful, irrespective of its fidelity to the physical reality.

Secondly, the violation against LiDAR sensor properties. Violations concerning LiDAR sensor properties are a frequent occurrence in data augmentation, and various

solutions have been proposed to address them. LiDAR sensors perceive the world in a polar coordinate system, with the sensor at the center, as opposed to a Cartesian coordinate system. Initially, this difference wasn't adequately considered in scene transformations, because added noise was uniformly distributed throughout the scene [88]. However, as demands for accuracy increased, the physical characteristics of noise points began to be incorporated. For instance, a normal distribution was used for arranging synthetic snow points to more accurately reflect their physical properties [88], but the violation remains. A more refined approach involves resampling the entire dataset in a resolution that corresponds to the sensor's horizontal turning rate and the number of vertical channels. This method has been validated for its effectiveness in multiple studies [89] [154] [155]. Such resampling ensures that the augmented data more accurately mirror the way LiDAR sensors capture and interpret the world, leading to more realistic and useful augmentation outcomes.

Lastly, occlusion poses a potential risk in point cloud data augmentation, where the addition of extra points could obscure the trace of an original signal, rendering the original point's presence illogical. While this topic is not extensively covered in current academic literature, two perspectives can be offered:

1. Spherical coordinates filtering:

Most points introduced into the point clouds are intended to be synthetic noise, and their existence typically remains minimal, often as low as 1%. Consequently, the impact of occlusion by these points is generally not seen as a significant threat to the point cloud's integrity. However, in certain conditions, like severe snow swirls observed in adverse datasets, the proportion of noise points can escalate to as much as 35%. This substantial increase occurs approximately 10% of the time and is too significant to overlook. To address this, a filtering method based on spherical coordinates is suggested.

By converting the LiDAR point cloud's Cartesian coordinates (x, y, z) into spherical coordinates (θ, ϕ, r), it becomes feasible to identify points sharing the same angular coordinates (θ, ϕ) but differing in the radial distance (r). Points with identical angular coordinates but larger radial distances would be recognized as occluded

by the artificially added points. This conversion and subsequent filtering ensure an accurate representation of the occlusion effects caused by augmented data.

2. Unsupervised deep-learning insights:

Deep-learning models, as demonstrated in previous works, have shown an intrinsic, unsupervised ability to understand adverse effect features and driving scene dynamics. This is exemplified in Fig. 3.7, where the color coding within the adverse effects indicates varying heights: clusters near the ego vehicle are slightly elevated due to swirl effects, and those more sparse clusters on the sidewalks are even higher, influenced by wind.

In the transformation results showcased in Fig. 3.10 and Fig. 3.11, it can be observed that objects originally obscured by adverse clusters are reconstructed into complete forms by the model. This indicates that these deep-learning models are not only recognizing effect features but also comprehending the broader scene, including the occlusion aspects of objects. Such insights are crucial as they suggest that these models can inherently account for and adapt to the complexities of occlusion in augmented point cloud data.

4.3 Experiments and Results

4.3.1 Reproduction of adverse conditions

Experiment settings

Experiments using the trained model were carried out on the CADC dataset which was split in a 6:1:3 ratio for training, validation, and testing purposes. With available ground truth for CADC, the goal is to replicate the adverse effects within this dataset and assess the model's generation capabilities through a comparison between the synthesized and original data. Additionally, experiments using both early and late fusion models were conducted on the CADC dataset. This shows the strengths and limitations of each model respectively. The model used for presenting the optimal results in this thesis adheres to the same loss function configurations as in Sec. 3.3.3

with the weight of the identity loss set to 1.

Training, testing, and data processing were executed using the Pytorch framework. The proposed model incorporates four ResNet residual blocks within the generator, optimized for training on two NVIDIA RTX 3090Ti graphics cards. A batch size of 4 was set, a decision influenced by the complexity inherent in the conditional generative models. For the continued training phase in the late fusion approach, a linearly declining learning rate schedule starting from 0.02 was employed, until the model reached convergence.

Qualitative results

Fig. 4.5, Fig. 4.6, and Fig. 4.7 present the adverse effect reproduction of the CADC dataset based on the early fusion model. For sets (a) (b) and (c) in each figure, the clear data, fake snow generation, and the real snow from CADC are placed at the left, middle, and right columns respectively. Each of the scenarios features an overall BEV in the top row, the clustered results showing the changes of snow clusters in the middle row; and the bottom rows showing magnified third-person views of the point cloud's central region, where the ego vehicle is situated. The same arrangement applies to the following qualitative results.

Observations from the areas highlighted by red arrows and enclosed within red boxes indicate a notable reproduction of scattered snow features. Snow clusters, particularly those in purple and dark blue around the ego vehicle appear enhanced. A detailed examination of the clustering in set (a) from Fig. 4.5, and set (c) in Fig. 4.7, reveals that the synthesized snow clusters are denser and more extensive than in the actual scenarios. However, in set (b) from Fig. 4.6, the synthetic clusters do not extend across the central region as broadly as in the real case. This outcome is attributed to the early fusion method's characteristic of capturing both snow clusters and objects via segmentation maps. As a result, while the model efficiently replicates snow clusters, it simultaneously accentuates relevant objects with minimal interference from surrounding minor snow clusters. This effect is evident in the bottom row of Fig. 4.6, where an area with multiple pedestrians is shown to be enlarged. Here,

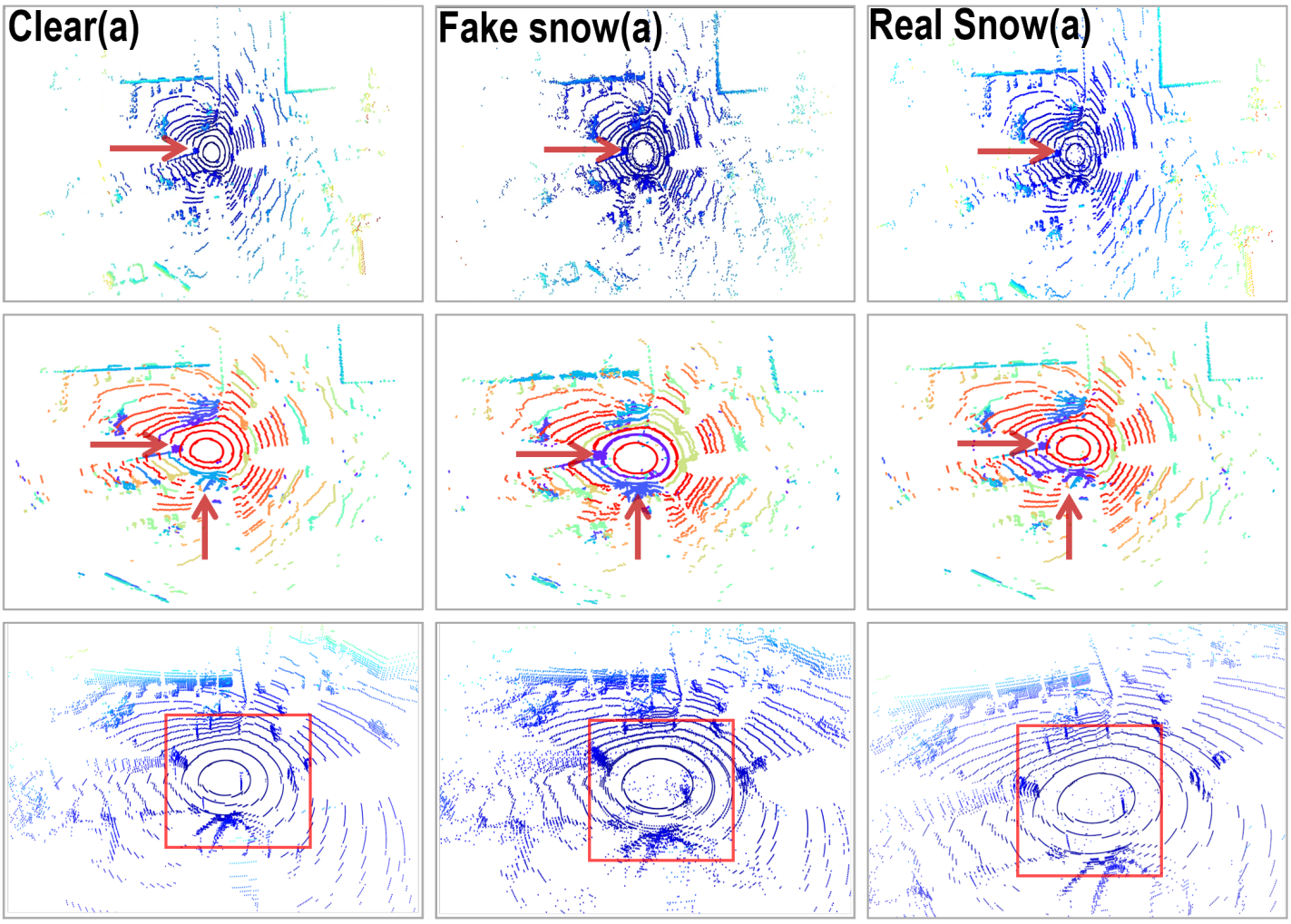


Figure 4.5: 1st set of point cloud augmentation results with **early fusion** model. First row - BEV scenes, colored by height; middle row - clustered results, colored by cluster groups; bottom row - enlarged third-person view center part around the ego vehicle, colored by height. Red boxes and arrows - locations where snow's effects are reproduced.

snow clusters are generated without significantly dispersing the signal on pedestrians, but rather enhancing it. Similar patterns are observable in Fig. 4.7.

In summary, the early fusion model effectively replicates adverse weather effects in the CADC dataset while ensuring that pertinent objects remain clearly identifiable, maintaining the integrity of their structure.

Fig. 4.8 and Fig. 4.9 show the reproduction of adverse effects in the CADC dataset using the late fusion model, arranged similarly to the earlier figures. From the BEV and clustering results, indicated by red arrows, it can be discerned that the generation of adverse effects, particularly snow clusters, does not overflow the level observed in actual snow conditions. Consequently, the enhancement of cars and pedestrians is less pronounced, as highlighted in the orange and green boxes in the bottom rows.

Focusing on the red boxes, which zoom into the central areas, the replication of scattered snow points is observed. However, their presence is noticeably less extensive compared to the actual scenarios. This could suggest a different kind of resemblance to the CADC dataset, but the shortfall in generating snow clusters and scattered snow points highlights a limitation of the late fusion model. It appears to struggle with fully leveraging the conditional guide function of the segmentation maps. This finding aligns with the expectations for the late fusion model, particularly considering the continued training techniques employed. It suggests that the late fusion approach may not match the generative capabilities of the early fusion model.

3D clustering results

The 3D clustering performance of synthesized snow data, augmented using the early fusion technique, is compared with the CADC dataset, as detailed in Table 4.2 and visualized in Fig. 4.10. This alignment demonstrates the effectiveness of the synthesis approach in replicating real-world snowy conditions, while also introducing certain enhancements.

The synthesized snow dataset exhibits a modest increase in the number of noise points (3025.06) compared to the CADC dataset (2865.07). This higher noise level

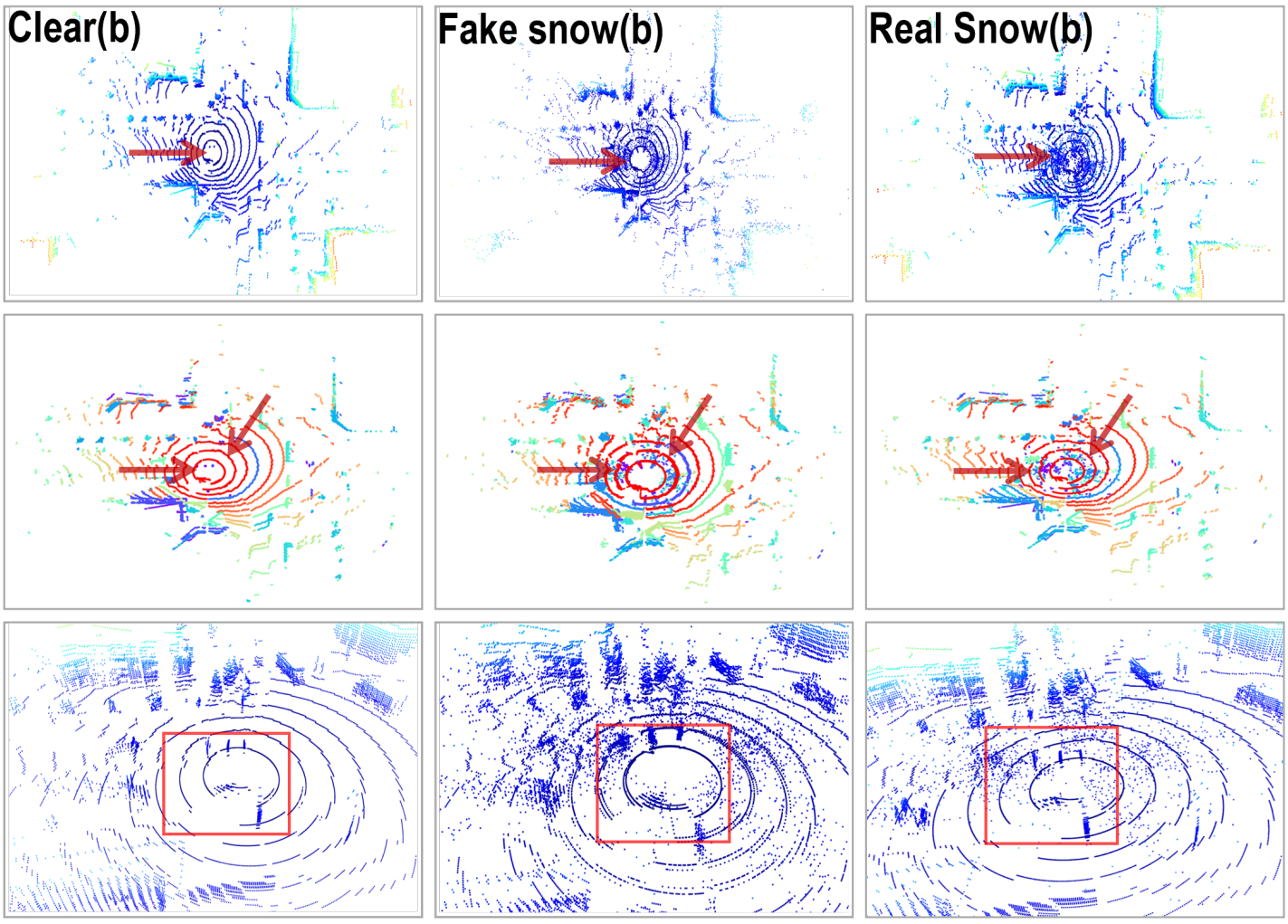


Figure 4.6: 2nd set of point cloud augmentation results with **early fusion** model. First row - BEV scenes, colored by height; middle row - clustered results, colored by cluster groups; bottom row - enlarged third-person view center part around the ego vehicle, colored by height. Red boxes and arrows - locations where snow's effects are reproduced.

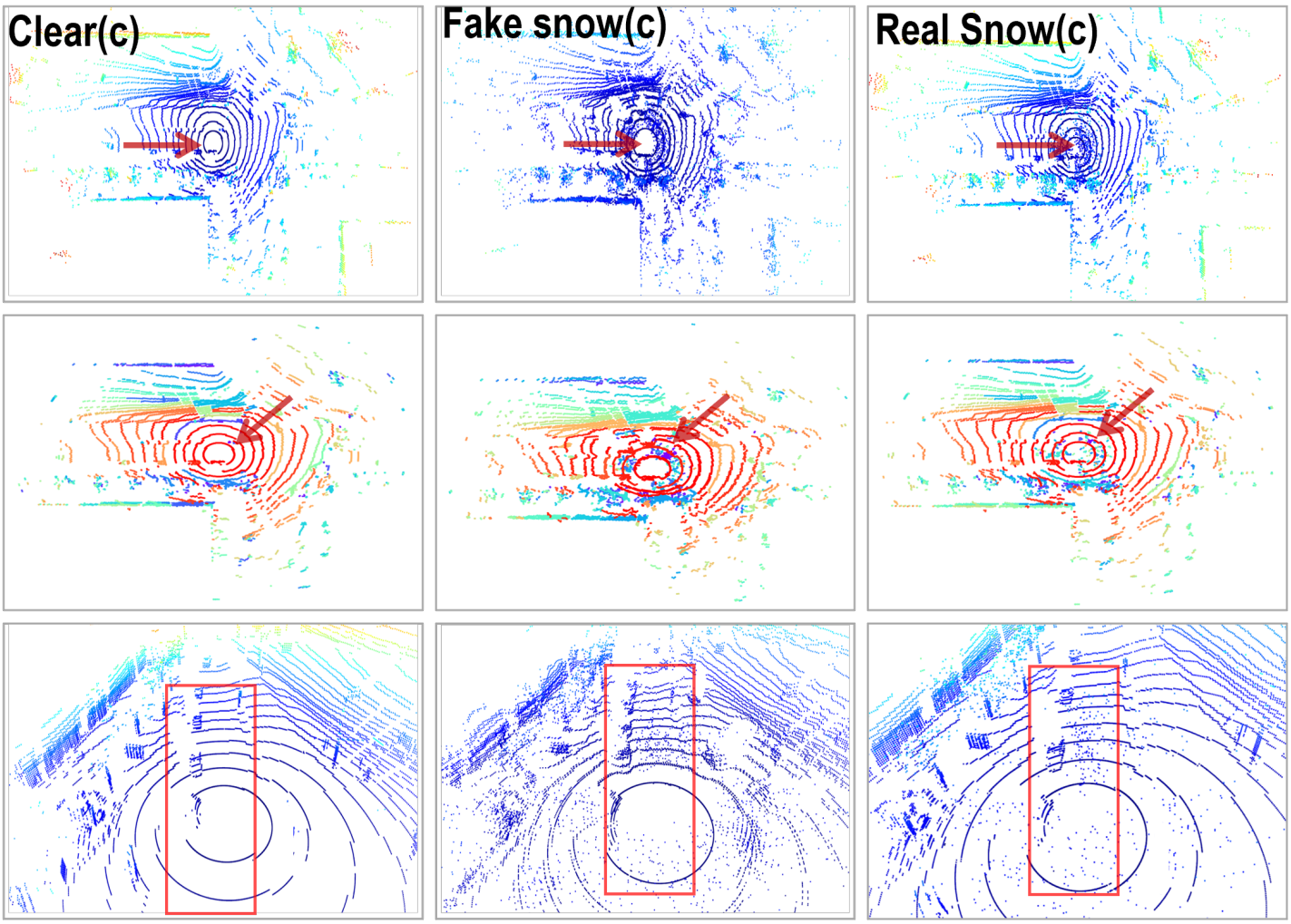


Figure 4.7: 3rd set of point cloud augmentation results with **early fusion** model. First row - BEV scenes, colored by height; middle row - clustered results, colored by cluster groups; bottom row - enlarged third-person view center part around the ego vehicle, colored by height. Red boxes and arrows - locations where snow's effects are reproduced.

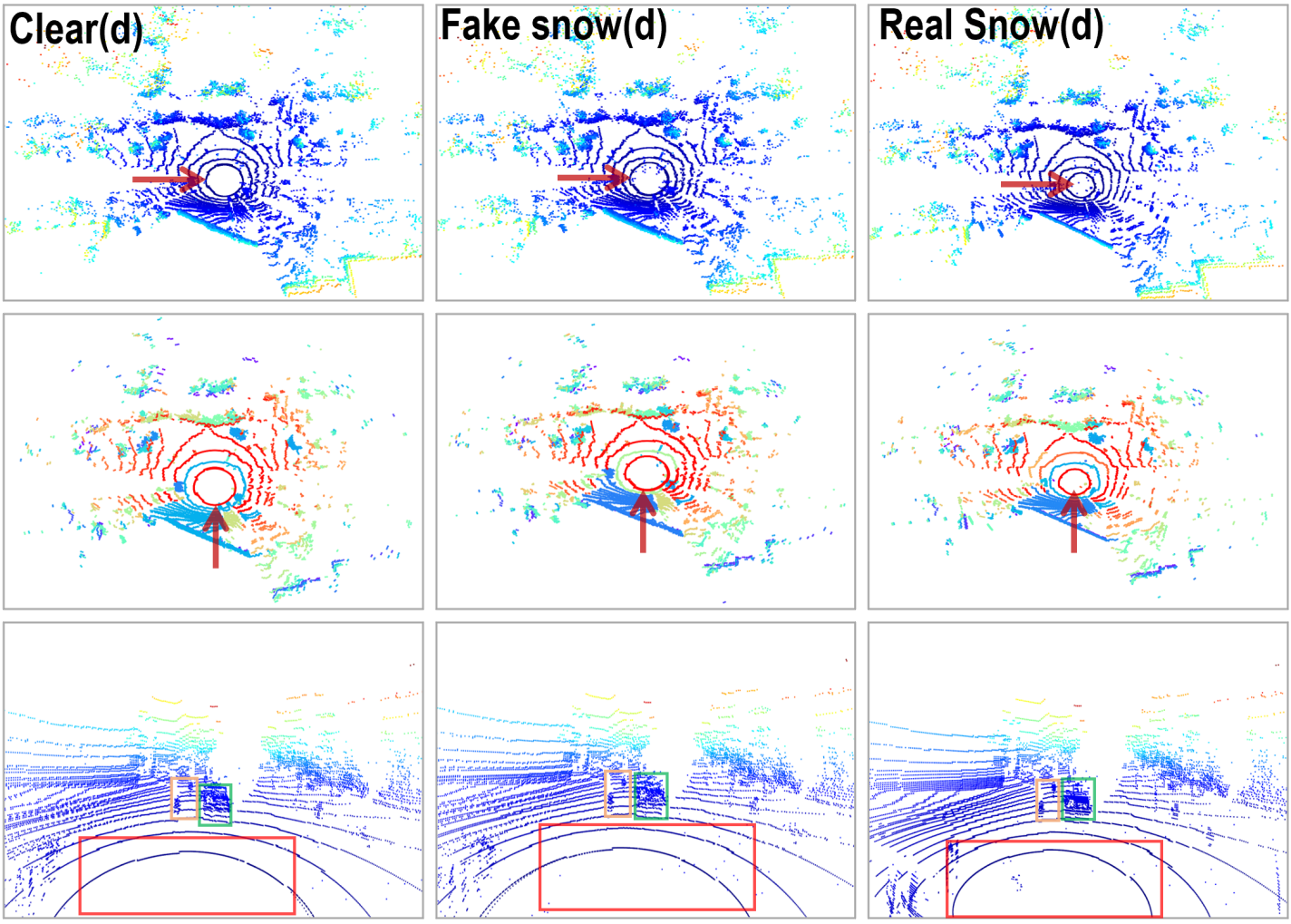


Figure 4.8: 1st set of point cloud augmentation results with **late fusion** model, with colors encoded by height. First row - BEV scenes; middle row - clustered results; bottom row - enlarged third-person view center part around the ego vehicle. Red boxes and arrows - locations where snow's effects are reproduced; green boxes - vehicles; orange boxes - pedestrians.

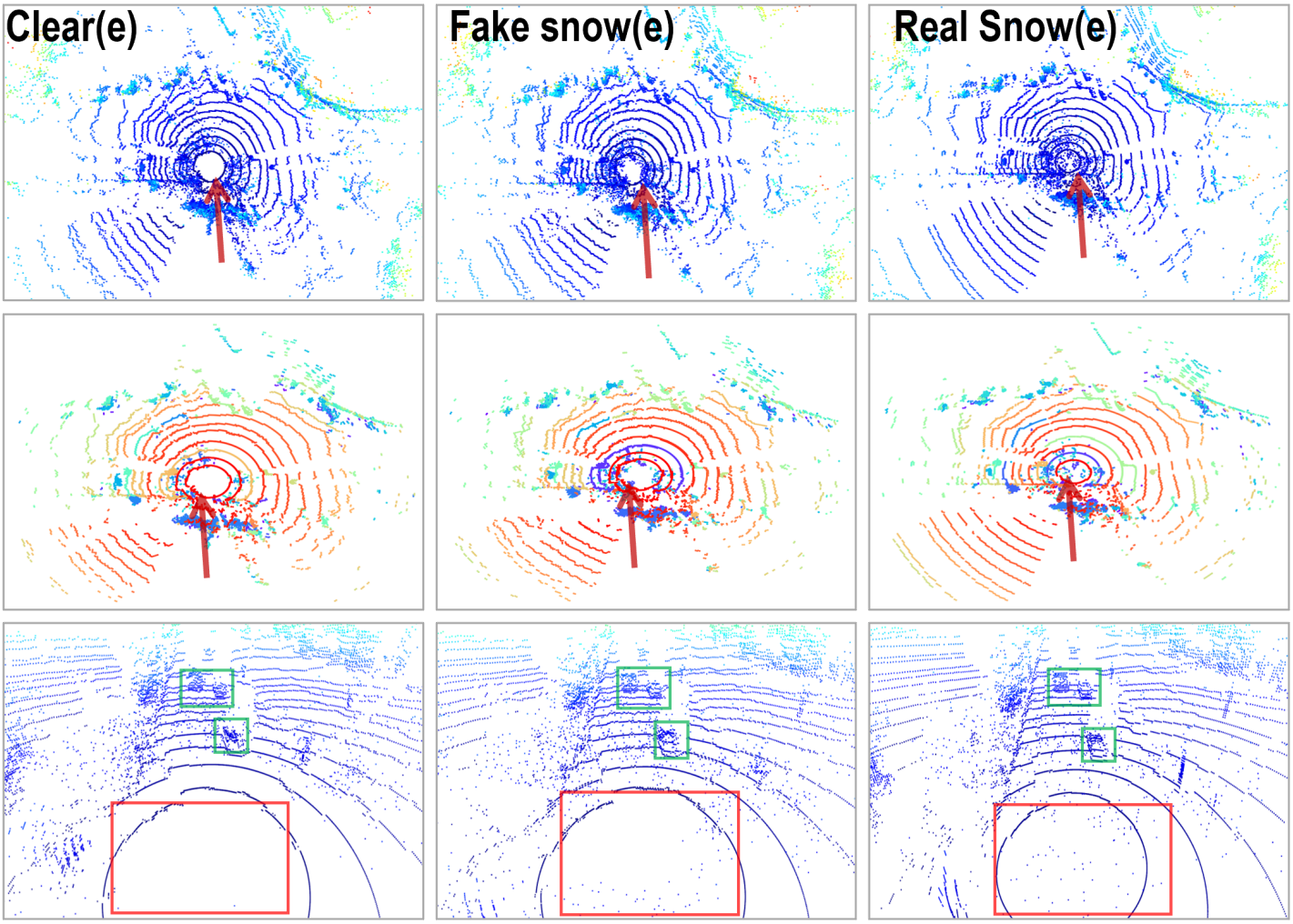


Figure 4.9: 2nd set of point cloud augmentation results with **late fusion** model, with colors encoded by height. First row - BEV scenes; middle row - clustered results; bottom row - enlarged third-person view center part around the ego vehicle. Red boxes and arrows - locations where snow's effects are reproduced; green boxes - vehicles.

in the synthesized dataset can be seen as an improvement, potentially offering a more robust training environment for machine learning models by simulating more challenging real-world conditions.

Similarly, the number of clusters is slightly higher in the synthesized dataset (1011.04) than in the CADC dataset (964.23). This increase suggests that the synthesis process adds complexity to the point cloud data, thus providing a more detailed representation of snowy environments.

In terms of reachability distance, inter-cluster distance and Davies-Bouldin Index (DBI), the synthesized snow dataset is quite comparable to the CADC dataset, with only minor deviations. These similarities indicate that the augmentation process maintains the essential spatial relationships between points, a crucial aspect for realistic scene representation.

The slight decrease in the average size of clusters in the synthesized data (12.9028) compared to CADC (14.1603) further emphasizes the finer granularity achieved through the proposed augmentation technique. This characteristic might prove advantageous in scenarios where detailed differentiation of features is required.

Finally, the Silhouette Score shows a marginal improvement in the synthesized snow dataset (-0.2730) compared to the CADC dataset (-0.3011). This enhancement suggests that the synthesized data maintains clear cluster differentiation, which is benificial for effective machine learning model training.

Overall, the artificially generated snow demonstrates a remarkable replication capacity, as evidenced by the highly alike violin plots (left and right) in Fig. 4.10, including the quartile lines. The degree of skewness (8.87, 0.33, and 22.43) is remarkably close to the previously mentioned CADC snow skewness (9.64, 0.30, and 21.49), further attesting to the model's ability to accurately reproduce snow effects.

The 3D clustering metrics for the synthesized snow dataset, generated using the late fusion technique, are presented in Table 4.3, offering a comparative perspective against the CADC dataset. This comparison not only sheds light on the late fusion model's capabilities but also allows for an understanding of its differences from the early fusion approach.

Table 4.2: The 3D clustering metrics (avg.) comparisons between CADC and synthesized snow with early fusion technique.

Items	CADC	Synthesized snow
Noise Number	2865.07	3025.06
Cluster Number	964.23	1011.04
Reachability Distance	0.4076	0.4339
Inter-Cluster Distance	51.4262	51.1447
Size of Clusters	14.1603	12.9028
Davies-Bouldin Index	4.0279	4.7175
Silhouette Score	-0.3011	-0.2730

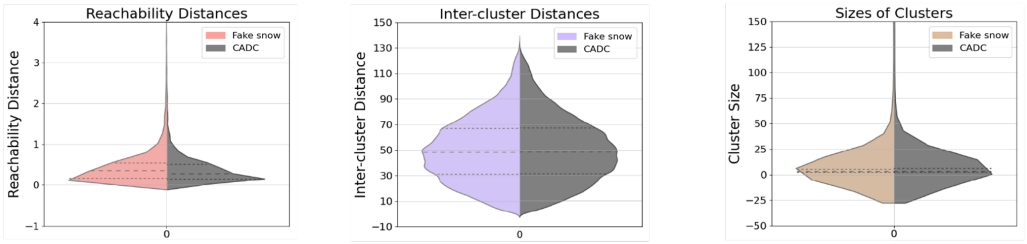


Figure 4.10: Violin plots for fake snow results with **early fusion** model. Limits on the y-axes are set for a better illustration of the distributions.

The synthesized snow dataset with late fusion shows a notable increase in the number of noise points (3124) compared to the CADC dataset (2598.17). This increase is more pronounced than that observed in the early fusion model, indicating a tendency of the late fusion approach to generate more noise. While this could be beneficial in creating challenging scenarios for model training, it also suggests a potential over-emphasis on noise generation instead of cluster, as evidenced in the the number of clusters in the late fusion synthesized dataset (998.77) being closer to the CADC dataset (954.24) than what was observed with the early fusion technique. This closer alignment might imply a more conservative generation of clusters, aligning with your previous observation of the late fusion model's less pronounced enhancement of cars and pedestrians.

In examining the reachability distances and cluster sizes in the late fusion synthesized dataset, metrics that closely mirror those of the CADC dataset were observed. The reachability distances (0.4358) and inter-cluster distances (44.3044) in the synthesized data exhibit a similarity to the CADC dataset's figures (0.3805 and 45.9245,

respectively). Additionally, the size of clusters in the late fusion data (13.7164) also approximates the CADC dataset (14.9750). While these measures suggest a level of similarity to the real-world dataset, it's important to note that this resemblance is not as pronounced as that achieved by the early fusion technique. The late fusion approach, in this aspect, shows a slightly less effective replication of the CADC's intricate clustering characteristics.

The Davies-Bouldin Index and Silhouette Score show a slight increase in the late fusion synthesized dataset compared to CADC, indicating a minor reduction in clustering quality. This could be reflective of the model's struggle to fully leverage the conditional guide function of the segmentation maps, as noted in the previous discussion.

Comparing these results with the early fusion technique, it becomes evident that the late fusion model exhibits a different kind of resemblance to the CADC dataset. While it manages to generate a closer number of clusters and maintains a reasonable level of noise, it falls short in replicating the exact nature and extent of snow clusters and scattered snow points. The violin plots in Fig. 4.11 demonstrate smoother edges of reachability distances and the more concentrated distribution in cluster sizes of the imitation snow, hinting at a feature of reduced noise at the center. However, the skewness values associated with the artificially generated heavy snow are 9.95, 0.42, and 31.75, respectively. These values start to leave a distance from the actual data, which are 10.45, 0.42, and 32.76, respectively, compared to early fusion.

The disparity is likely due to the late fusion model's approach to integrating the segmentation map information, which occurs later in the generative process and may not be as effective in guiding the generation of realistic snow effects. In summary, the late fusion synthesized snow dataset demonstrates a unique approach to augmenting point cloud data with snow effects. While it shows certain improvements in noise generation and cluster numbers, it also reveals limitations in accurately replicating the detailed structure of snowy environments. This analysis suggests that while the late fusion technique offers a viable alternative, it may not quite match the early fusion model in terms of generative capabilities for reproducing complex snow effects.

Table 4.3: The 3D clustering metrics (avg.) comparisons between CADC and synthesized snow with late fusion technique.

Items	CADC	Synthesized snow
Noise Number	2598.17	3124
Cluster Number	954.24	998.77
Reachability Distances	0.3805	0.4358
Inter-Cluster Distances	45.9245	44.3044
Size of Clusters	14.9750	13.7164
Davies-Bouldin Index	5.4138	6.0746
Silhouette Score	-0.3699	-0.3661

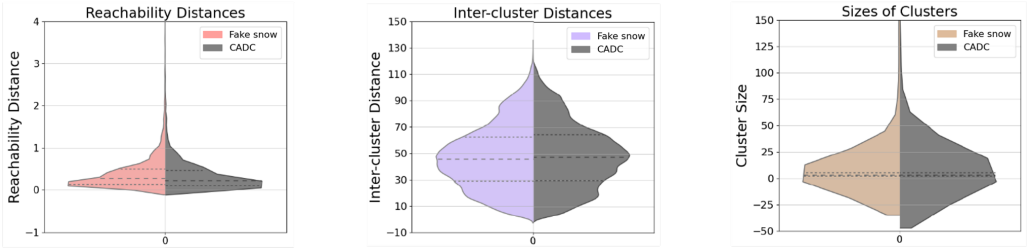


Figure 4.11: Violin plots for fake snow results with **late fusion** model. Limits on the y-axes are set for a better illustration of the distributions.

Precision and recall

To assess the effectiveness of the proposed conditional generative model in synthesizing adverse effects, particularly in comparison with other methods, a precision-recall analysis was conducted along with an ablation study on the model without the conditional guide, as outlined in (4.2) and (4.3). Similar to the previous approach, the point clouds were analyzed on a cluster basis. This analysis, focusing on replicating snow effects within the CADC dataset, bases its calculations on the CADC ground truth to ensure reliability.

For this evaluation, a subset of samples from the test dataset, which had not been included in training, was selected without bias. Each cluster in these samples underwent manual labeling to identify the presence of snow clusters. Unlike the previous task, where a trade-off was a key consideration, here the primary focus was on maintaining scene integrity only. Consequently, the parameters for all comparative methods were uniformly set, ensuring equal treatment in parameter application

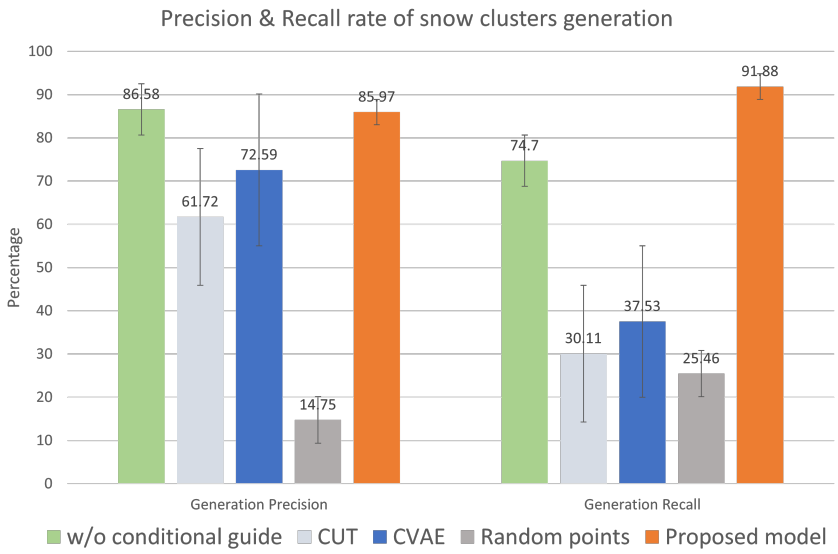


Figure 4.12: Precision and recall rates comparisons of adverse effects generation based on snow clusters.

across all methods under evaluation. This consistent parameter setting allows for a fair comparison of each method's ability to generate realistic adverse effects while preserving the integrity of the driving scene.

$$\text{Precision} = \frac{\text{generated clusters} \cap \text{labeled as snow}}{\text{total generated clusters}} \tag{4.2}$$

$$\text{Recall} = \frac{\text{generated clusters} \cap \text{labeled as snow}}{\text{total labeled snow clusters in CADC}} \tag{4.3}$$

Looking at Fig. 4.12, it's evident that both the model without conditional guide and CUT struggle to accurately reproduce snow clusters found in the original dataset. This challenge stems from their limited understanding of adverse effects. Intriguingly, the precision of the model without conditional guide slightly exceeds the proposed model by less than 1%. This minor advantage primarily arises from its handling of occluded objects, a residual aspect in this model. As the proposed model endeavors to regenerate snow clusters, it inadvertently reconstructs or augments some partially composed non-snow objects, leading to a precision just shy of 90%, but an apparent improvement in recall rate.

Conversely, the CVAE model, despite conditional guidance, fails to effectively

address driving scenarios. Its focus is predominantly on generating snow clusters around the vehicle's immediate vicinity, neglecting the broader scene. This results in a significant loss of the original point structure and tiny snow clusters in other areas, contributing to its notably low recall rate. As for randomly generated points with normal distribution, the 3D clustering algorithm has a hard time identifying distinct clusters, leading to a substantial increase in noise with a corresponding decrease in cluster identification, as reflected in the 14.75% precision rate.

In summary, the proposed conditional guided generative model stands out with the highest recall in reproducing snow effects from the CADC dataset, achieving this with satisfactory precision. This underscores the efficacy of the proposed model in generating realistic adverse weather conditions for LiDAR point clouds.

Detection rate improvement

For evaluating improvements in detection, the 3D object detection metrics from the KITTI evaluation framework [75] were utilized. In line with [156], the average precision (AP) across 40 recall positions is reported to ensure a balanced comparison. The investigation focuses on the impact of the proposed augmentation scheme for adverse conditions on two widely used 3D object detection methods [157] [158]. The proposed approach against a baseline model without any augmentation is assessed, and then the results of using augmentation data generated by the proposed model are considered. In addition, the results with those obtained from de-noising point clouds using the DROR method [46] are compared. The DROR filter is selected as the representative of noise removal algorithms to better show the difference between clear-data-based and adverse-data-based augmentations, given its balanced performance and wild references in literature [43].

For training the detection models, the mmdetection [159] was employed and the default training configurations specified for each method were adhered to, ensuring that the only variable in this comparison was our augmented data. And all methods are trained from scratch. Augmentation data consists of 2691 frames of point cloud with adverse effects. The training and testing datasets follow the distribution of 7:

Table 4.4: Comparison of augmentation methods for 3D object detection in snowfall on CADC.

Detection method	PV-RCNN [157]			SECOND [158]		
Augmentation method	None	DROR	Ours	None	DROR	Ours
3D average precision (AP)	43.11	38.69	**45.57**	37.08	35.31	**38.23**

3, in which the testing dataset remains the same as in previous sections.

Detection rates are detailed in Table 4.4. 3D average precision (AP) is reported on baseline (no augmentation) and the proposed augmentation model with 2691 samples. Augmentation with DROR filters is also provided for reference. A key observation from this data is that the comprehensive augmentation, featuring both scattered and clustered adverse effects, markedly enhances performance in the most challenging test scenario, specifically during fierce snowfall. This enhancement is significant when compared to both the baseline approach and the de-noising filter. Here, the complete augmentation outperforms the baseline by a noteworthy 2.46% increase in AP.

An example of visual results from PV-RCNN showing the data augmentation scheme is presented in Fig. 4.13. The top row displays a composite of RGB images from three cameras, oriented left-front, directly ahead, and right-front, collectively representing a 180° frontal view of the ego vehicle. The subsequent BEV point clouds compare detection outcomes using the proposed augmentation method and the DROR filter against a baseline with no augmentation, alongside the ground truth. Pedestrians are denoted in red dots while cars (and trucks) are denoted in black bounding boxes with red dots in the center.

From the RGB images and ground truth, a scenario with many moving pedestrians and parked cars under heavy snow is observed. The baseline model, lacking augmentation data, struggles to detect pedestrians and cars beyond a certain distance or those partially obscured, and often fails to accurately gauge the gesture and dimensions of detected objects. In contrast, the proposed augmented data method enables more precise detection of cars and pedestrians, particularly those nearer to the ego vehicle, with accurate parameters. As for the augmentation scheme with DROR fil-

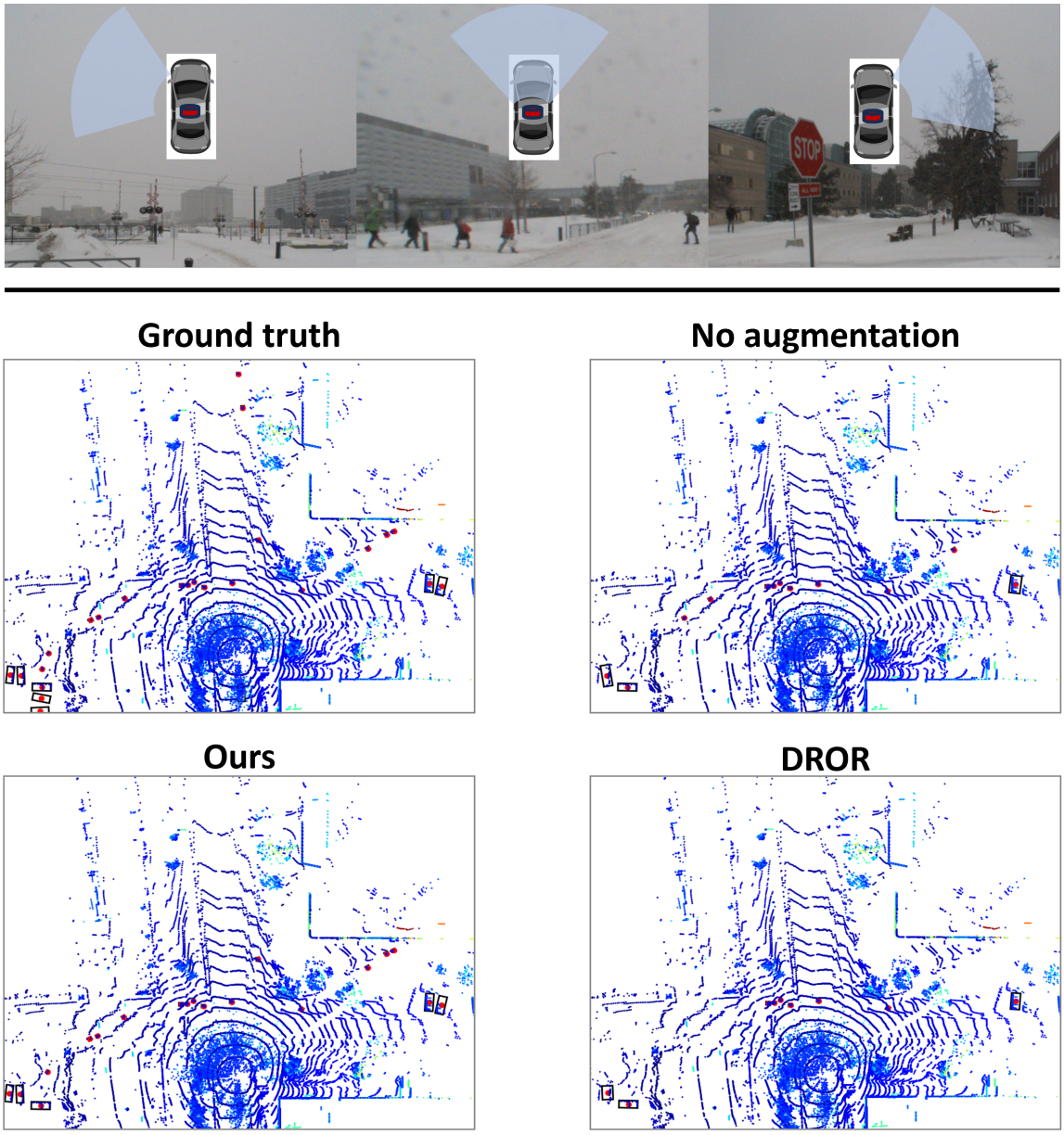


Figure 4.13: Qualitative comparison of detection results on samples from CADC containing fierce adverse conditions. The top row shows the corresponding forward 180° RGB images. The rest show the LiDAR point clouds with ground-truth boxes and predictions using the baseline ("no augmentation"), the proposed augmentation, and DROR. Red dots denote pedestrians, and black boxes with red dots in the center denote cars (or trucks). Point cloud colors encoded by height.

ters, the performance turns out to be worse than the baseline. This is attributed to the removal of critical points necessary for object detection.

However, it's important to note that two undetected parked cars at the bottom left and two undetected pedestrians far ahead, as shown in the ground truth, are identified through camera assistance due to their minimal LiDAR signal presence. This is a common issue in adverse condition datasets and partly explains the generally low average detection rate observed in Table 4.4.

There is a car near the rear of the ego vehicle in the ground truth, barely noticeable due to snow swirl occlusion, which none of the methods detect due to severe occlusion. This might hint at the limitation of current learning-based perception improvement methods and suggest the need for advancements in sensor hardware to further overcome such challenges.

4.3.2 Synthetic adverse conditions

Experiment settings

For the LIBRE Nagoya dataset [10], the objective is to evaluate the model's performance in overcoming domain gaps. The pretrained model based on CADC was directly tested on the 6000 frames of the Nagoya dataset which was collected under clear conditions in the urban area of Nagoya, Japan. This selection is particularly representative due to the significant domain differences between Canada and Japan, especially regarding scenario layouts and traffic behaviors. The CADC dataset predominantly features suburban environments with sparse buildings and abundant vegetation, a contrast to the urban settings commonly found in the Nagoya dataset.

Since the Nagoya Dataset lacks ground truth, the assessment will provide 3D clustering results without precision and recall in this task. These evaluations were essential for a quantitative assessment of the model's effectiveness in generating synthetic snow on a clean dataset. The metrics reported in the subsequent results sections represent average values derived from these samples.

Qualitative results

Fig. 4.14, Fig. 4.15, and Fig. 4.16 present the adverse effect synthesis on the Nagoya dataset where domain gap exists. The early fusion model is implemented given its overall advantages. For sets (a) (b) and (c) in each figure, the clear data and the synthetic snow generation from the proposed model are placed in the left and right columns respectively. Each of the scenarios features an overall BEV in the top row, the clustered results showing the changes of snow clusters in the middle row; and the bottom rows showing magnified third-person views of the point cloud's central region, where the ego vehicle is situated. The same arrangement applies to the following qualitative results.

The red arrows in the figures highlight the areas where snow clusters have been generated. Upon examining these indicated locations, it's evident that a number of snow clusters have formed near the central area, resembling typical patterns observed in snow-affected driving scenarios. However, it's important to note that the level of generated snow is not an exact match for the snow conditions found in the CADC dataset, particularly in terms of the density of the purple and dark blue clusters depicted in the middle rows. This discrepancy partly arises from a deliberate decision to reduce the extent of snow generation by tuning down the weights in loss functions. Additionally, the model is operating across a significant domain gap, characterized by different environmental layouts and traffic patterns, which necessitates a more cautious approach.

As a result, and as illustrated by the red boxes, the snow generation in the vicinity of the ego vehicle and around the sidewalks predominantly features scattered snow points, rather than extensive clusters. This cautious approach is particularly appropriate given the Nagoya dataset's inherent composition, which includes numerous small-sized clusters, often resulting from vegetation cover and occlusions. Therefore, generating snow in a more restrained manner is the most effective strategy to maintain the original structural integrity of the dataset while successfully synthesizing quasi-natural adverse effects.

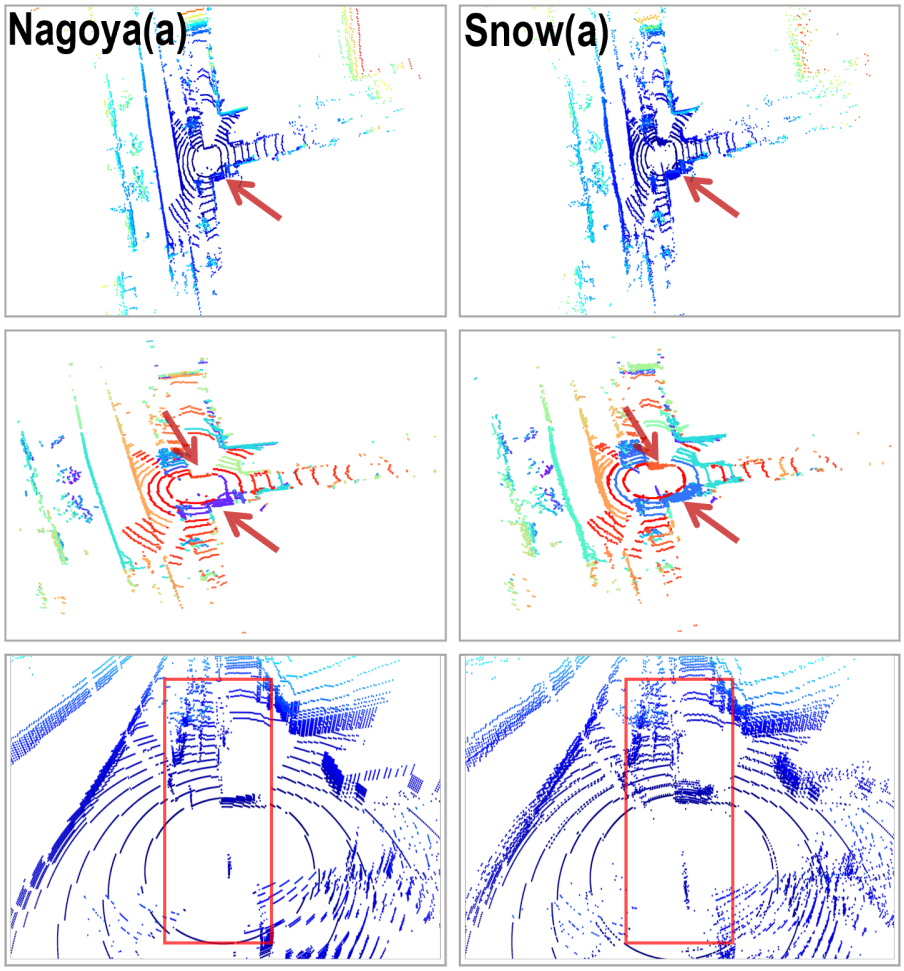


Figure 4.14: 1st set of adverse condition synthesis on Nagoya dataset. First row - BEV scenes, colored by height; middle row - clustered results, colored by cluster groups; bottom row - enlarged third-person view center part around the ego vehicle, colored by height. Red boxes and arrows - locations where adverse effects are synthesized.

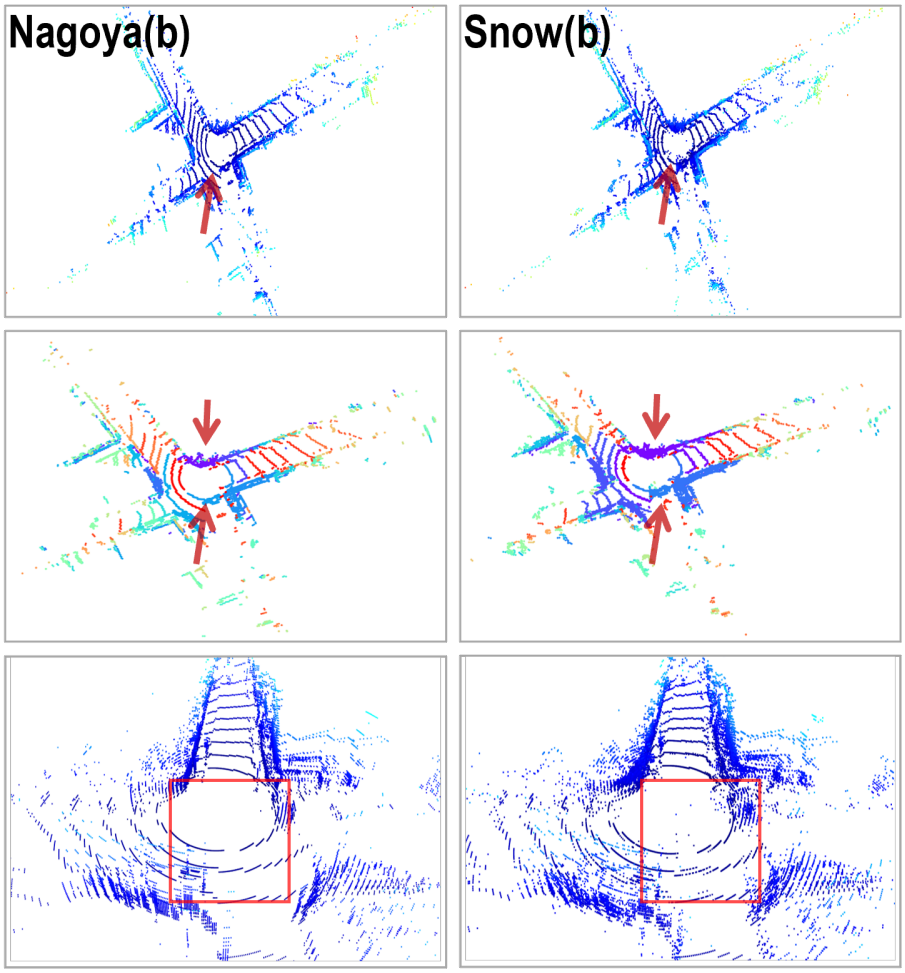


Figure 4.15: 2nd set of adverse condition synthesis on Nagoya dataset. First row - BEV scenes, colored by height; middle row - clustered results, colored by cluster groups; bottom row - enlarged third-person view center part around the ego vehicle, colored by height. Red boxes and arrows - locations where adverse effects are synthesized.

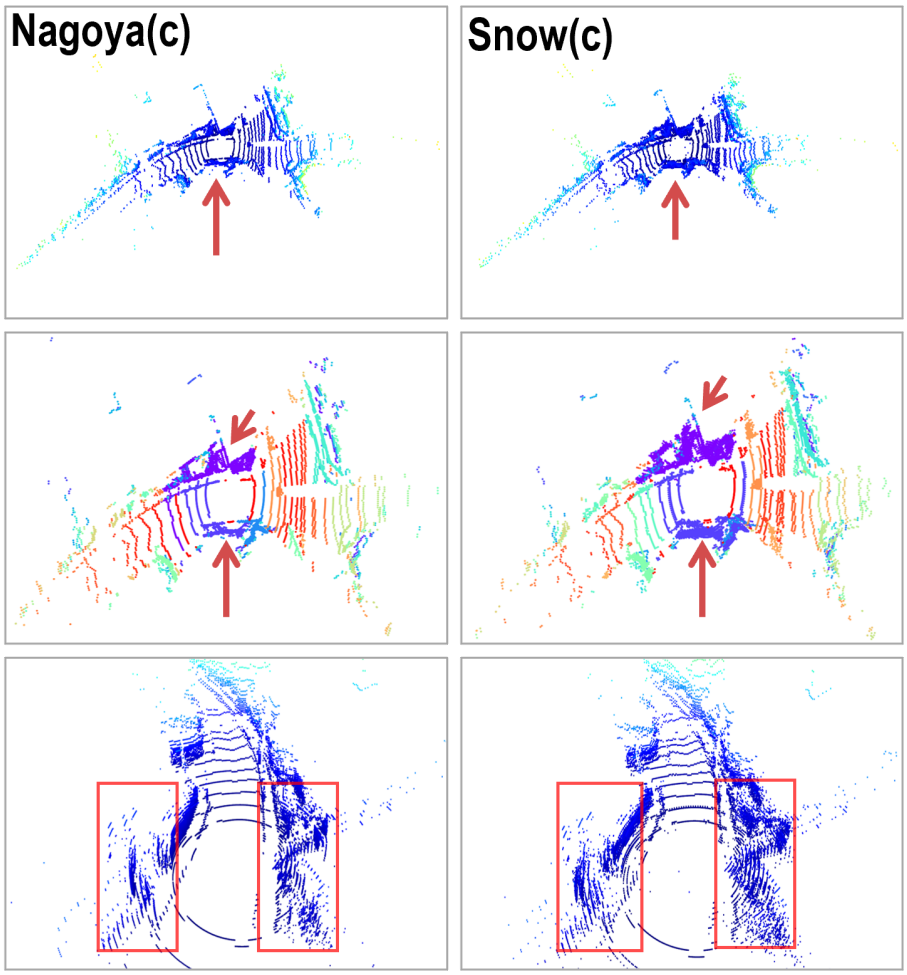


Figure 4.16: 3rd set of adverse condition synthesis on Nagoya dataset. First row - BEV scenes, colored by height; middle row - clustered results, colored by cluster groups; bottom row - enlarged third-person view center part around the ego vehicle, colored by height. Red boxes and arrows - locations where adverse effects are synthesized.

Table 4.5: The 3D clustering metrics (avg.) comparisons between the original Nagoya dataset and synthesized snow in Nagoya.

Items	Nagoya	Synthesized snow
Noise Number	1204.67	2631.46
Cluster Number	480.25	1073.52
Reachability Distance	0.2470	0.3952
Inter-Cluster Distance	59.30	49.45
Size of Clusters	28.2638	13.3046
Davies-Bouldin Index	2.3653	4.4149
Silhouette Score	-0.2170	-0.2927

3D clustering results

Table 4.5 offers a quantitative comparison between the original Nagoya dataset and its augmented version with synthesized snow, demonstrating the effects of the snow synthesis process. Given the absence of ground truth data for snow conditions in the Nagoya dataset, the violin plots were not included.

The synthesized snow dataset shows a notable increase in the number of noise points, rising from 1204.67 in the original Nagoya dataset to 2631.46. This increase is indicative of the additional complexity introduced by the synthetic snow, representing a more challenging scenario for processing and analysis. This statistic is one of the most direct indicators of a successful adverse effect synthesis.

The cluster number more than doubles in the synthesized dataset (1073.52) compared to the original dataset (480.25), a direct consequence of the snow synthesis process. The additional clusters likely represent the snowflakes or snow clusters, providing a more realistic representation of a snowy environment.

Reachability distances in the synthesized dataset show an increase to 0.3952 from 0.2470 in the original dataset. This change suggests that the introduction of snow creates a more scattered distribution of data points, emulating the disperse nature of snowfall and its impact on the visibility and distinguishability of objects within the dataset.

The average inter-cluster distance decreases in the synthesized dataset (49.45) compared to the original dataset (59.30). This reduction might reflect the additional

clusters formed due to snow, which are closer together, simulating the dense and overlapping nature of snow in the environment.

There is a notable decrease in the average size of clusters in the synthesized data (13.3046) compared to the original dataset (28.2638). This decrease can be attributed to the synthesized snow creating smaller, more numerous clusters, which is consistent with the physical characteristics of snow affecting spatial data.

The Davies-Bouldin Index increases in the synthesized dataset (4.4149) compared to the original (2.3653), indicating less compact but more separated clusters, a likely result of snow altering the spatial relationships within the data. Similarly, the Silhouette Score drops slightly in the synthesized dataset (-0.2927) compared to the original (-0.2170), suggesting a decrease in the separation distance between neighboring clusters due to the added snow.

Overall, these metrics demonstrate the significant impact of synthesized snow on the Nagoya dataset, successfully adding quasi-natural adverse effects to the scenario, which is crucial for testing and improving algorithms in snowy conditions.

4.4 Conclusion

In this chapter, an innovative approach has been presented to augment adverse weather condition data in autonomous driving applications. The proposed model, leveraging conditional guides, has shown proficiency in generating natural adverse effects in LiDAR point cloud data. This augmentation is critical given the scarcity of diverse weather conditions in existing datasets, which hinders the development of robust perception systems for autonomous vehicles.

The proposed methodology involved the development of a segmentation and classification system for adverse weather effects using a designated 3D clustering algorithm. This system effectively guides the generative model in creating precise and realistic augmentations. Both the early and late data fusion methods have been explored, each providing unique benefits in integrating raw and segmented data. These methods were validated across two datasets with significant differences, the CADC

and LIBRE Nagoya datasets, demonstrating the model's effectiveness in replicating adverse weather conditions and its capability to bridge domain gaps.

A noteworthy outcome of this research is the observed improvement in detection rates when perception models are trained with augmented data. This enhancement in performance under adverse conditions is a significant stride towards ensuring safer and more reliable autonomous driving technologies.

However, it's important to acknowledge that the current work primarily derives its adverse condition data from a single dataset, the CADC. While this dataset has provided a robust foundation for initial investigations, particularly with snow as a test subject, the diversity of adverse weather conditions experienced globally suggests a need for a broader data source. Moving forward, a key area of development will be to extend the data sources beyond the singular dataset. This expansion will enable the capture of a more comprehensive range of adverse conditions, reflecting the varied and unpredictable nature of weather around the world. By integrating data from multiple datasets, each representing different environmental and weather conditions, the proposed model can be further refined and its applicability broadened. This approach will enhance the model's ability to simulate a wider array of adverse scenarios, thereby improving the robustness and reliability of perception systems in autonomous vehicles under diverse environmental conditions.

In conclusion, this work addresses a vital need in the field of autonomous driving by enriching the diversity and realism of adverse weather data available for training perception systems. The progress made in this research not only contributes significantly to the field but also lays the groundwork for further advancements in enhancing autonomous vehicle safety and reliability in challenging environmental conditions.

Chapter 5

Summary and Future Works

5.1 Thesis Summary

This thesis has embarked on an ambitious journey to address the critical challenges that adverse weather conditions pose to LiDAR-based perception systems in autonomous driving. Recognizing the limitations in current datasets and perception models under such conditions, this research has focused on enhancing LiDAR point cloud data's robustness and reliability in adverse environments. By combining in-depth analysis with innovative computational techniques, significant strides have been made in advancing the field of autonomous vehicle perception.

This research commenced with a detailed exploration of the impacts that various adverse conditions have on perception sensors, informing the understanding of how these conditions compromise sensor performance. This understanding was crucial to developing strategies for counteracting these challenges. The specifics of how each type of adverse weather have been dug into — rain, fog, snow, and contamination — uniquely affects sensors like LiDAR, cameras, radar, and ultrasonic systems.

The focus on LiDAR sensors revealed the complexities of how point clouds are distorted in adverse weather where both false and missing detections occur. For camera systems, the challenge lay in the lens blur caused by these conditions, drastically reducing visibility and the effectiveness of image processing. In the case of radar sensors, their diminished reliability due to factors like wetness and heavy precipita-

tion was analyzed, while ultrasonic sensors showed compromised performance due to acoustic signal attenuation in conditions such as heavy rain or snow. This comprehensive sensor analysis was instrumental in directing the research towards enhancing perception under these challenging conditions. It highlighted the necessity of a holistic approach that addresses the collective impacts of adverse conditions on multiple sensor systems, guiding the focus on improving LiDAR point cloud data for more reliable autonomous vehicle perception.

Addressing these challenges, this research focuses on the potential of deep learning-based models, a path less traveled in the domain of perception under adverse conditions, after a range of adaptive filters on authentic adverse data were rigorously evaluated, an advanced adaptive filter centered on non-geometric variables was eventually designed. Unlike traditional mathematical approaches that offer limited scope, learning-based models unveil a broader spectrum of both apparent and latent features inherent to adverse driving scenarios. In light of the rapid development of machine learning technology, the L-DIG model was developed, a GAN-based method utilizing depth image priors for the transformation of LiDAR point clouds under various adverse conditions. This innovative model, trained with unpaired datasets, is proficient at transforming point clouds from adverse to clear conditions, demonstrating a profound understanding of the unique characteristics of adverse effects. This transformation is crucial, considering adverse conditions' tendency to form clusters of crystal aggregates and irregular swirls, which introduce irregular noise and anomalies in LiDAR point clouds. The proposed unique discriminator structure and loss functions, including depth and SSIM losses, play a pivotal role in maintaining the integrity of the driving scene while effectively removing noise.

Complementing the L-DIG model is the customized 3D clustering algorithm, specifically tailored for analyzing and quantifying the adverse effects in LiDAR point clouds. This algorithm enables adaptive clustering of various forms of adverse effects, presenting a comprehensive quantitative evaluation of the conditions within the dataset. This strategic combination of a deep-learning model and advanced filtering techniques signifies a major leap forward in the ability to transform and refine LiDAR

point cloud data, thereby enhancing the perception capabilities of autonomous driving systems under adverse weather conditions. The proposed approach marks a significant departure from the traditional de-noising methods, offering a more dynamic and insightful solution to the challenges posed by adverse weather. The outcomes of this research not only contribute to the field of autonomous vehicle technology but also set the stage for future advancements in machine learning and sensor-based technologies, geared towards overcoming the hurdles of operating in challenging environmental conditions.

Another cornerstone of this research was the development of a conditional generative model proficient in synthesizing natural adverse effects in LiDAR point cloud data. This model plays a crucial role in addressing the scarcity of varied weather conditions in existing datasets, a significant hurdle in the advancement of autonomous driving technologies. Utilizing advanced data fusion techniques and cluster-based segmentation maps as conditional guides, the proposed model has demonstrated an exceptional ability to generate realistic and precise adverse weather effects. This achievement is particularly noteworthy considering the challenges posed by the limited representation of adverse conditions in current driving datasets.

The effectiveness of the proposed model is underscored by the superior performance of methods employing adverse conditions data augmentation compared to traditional classification-based methods. The inherent difficulty in annotating or labeling adverse conditions, coupled with the sparsity of samples featuring such conditions in existing datasets, necessitates a more innovative approach. The proposed model addresses these challenges by augmenting clear condition datasets with simulated adverse conditions, thereby enhancing the deep learning models' capability to navigate and interpret complex driving scenarios under diverse weather conditions.

Key contributions of this work include the development of segmentation maps of adverse effects using the specially designed 3D clustering algorithm, the provision of classification and conditional guides for generative models, and the design of both early and late data fusion methods. Each of these elements contributes to the model's capability to not only reproduce natural adverse effects but also to bridge the gap

between different traffic layouts and environmental conditions, thus offering high-level robustness in the generation of paired adverse data.

The research in this area opens up new avenues for enhancing the perception models used in autonomous driving. By successfully integrating paired datasets that include both adverse effects and corresponding classifications, it becomes possible to significantly improve detection rates, especially in identifying smaller entities like pedestrians and cyclists, under challenging driving conditions. This enhancement is a testament to the potential of the proposed approaches in elevating the safety and reliability of autonomous driving technologies.

In summary, this thesis has contributed significantly to the advancement of autonomous vehicle technologies under challenging environmental conditions. The proposed innovative approaches in transforming point clouds and expanding adverse condition data have laid a solid groundwork for future research. Moving forward, the insights and methodologies developed in this thesis will pave the way for further advancements, aiming to fortify the resilience and adaptability of autonomous driving technologies in the face of dynamic and diverse adverse environmental challenges.

5.2 Future Works

The journey of advancing LiDAR sensor perception in adverse conditions is ongoing, and the future work within this domain holds immense potential for innovation and enhancement of autonomous driving systems. The research avenues that lie ahead are both challenging and exciting, promising significant contributions to the field of IoT society.

5.2.1 Addressing special adverse conditions

One of the specialized adverse effects that demand attention is the influence of strong light on sensor performance. The issue of strong light and its impact on sensor performance in autonomous vehicles is a tricky area that requires dedicated attention in future research endeavors. Strong light presents a unique challenge in the

realm of sensor perception, primarily due to its non-physical nature. Unlike other adverse conditions that have tangible attributes, strong light is a phenomenon that can significantly distort sensor readings without a corresponding physical object. This distortion manifests as an overwhelming signal that can saturate sensors, particularly LiDAR, leading to loss of data fidelity and potentially causing safety-critical errors.

To address this challenge, future research must move towards innovative filtering techniques that not only reduce the noise induced by strong light but also retain the structural integrity of the sensor data. Signal-pattern-attention filtering emerges as a promising direction in this regard. Such filtering methods would involve sophisticated algorithms capable of identifying the specific signatures of noise created by intense light sources and attenuating their effects on the sensor's signal. The goal would be to enable the system to maintain high signal quality even when confronted with conditions that would traditionally result in signal loss or corruption.

Beyond filtering techniques, enhancing the robustness of perception models against the effects of strong light is crucial. This enhancement can be achieved through designated data augmentation strategies that introduce simulated strong light scenarios into training datasets [91]. By creating and training on these augmented datasets, perception models can learn to recognize and compensate for the aberrations caused by such lighting conditions. The data augmentation process would involve generating a diverse range of scenarios that a perception sensor might encounter, from the low-angle glare of sunrise and sunset to the artificial brightness of urban illumination.

In essence, the future work aimed at mitigating the effects of strong light on sensor performance will involve two aspects. On one hand, it will leverage advanced signal processing and filtering methods to improve immediate sensor data quality. On the other hand, it will employ strategic data augmentation to build foundational resilience in perception models, allowing sensors to operate safely and effectively in a world where lighting conditions are ever-changing and often less than ideal.

5.2.2 Exploring other modalities and sensor fusion

Another promising direction for future research is the exploration of new modalities of LiDAR technology, such as Frequency Modulated Continuous Wave (FMCW) and aerial LiDARs, along with the integration of radar and various cameras. With the rapid development of RSU (Roadside Units) and UAVs like drones, it's becoming realistic to do weather transportation perception from the top view, which sees what couldn't be seen from the ground [160]. These technologies present a rich opportunity to capture a broader spectrum of environmental data, but they also introduce the challenge of fusing different data formats. Overcoming this challenge will involve creating new conditional guides that can leverage additional information provided by these diverse sensors [42]. By synthesizing this information, a learning-based model with multiple input channels that can process and integrate a variety of sensor data can be developed, enhancing the perception system's ability to operate cohesively in complex environments.

This holistic approach to sensor fusion will enable a more comprehensive understanding of the vehicle's surroundings, taking advantage of the strengths of each sensing modality. For instance, while LiDAR provides high-resolution distance measurements, cameras offer rich color and texture information, and radar adds robustness under adverse weather conditions. The fusion of these modalities will lead to a synergistic model that is greater than the sum of its parts, ensuring more reliable navigation and decision-making by the autonomous driving system.

In conclusion, the path ahead for improving LiDAR sensor perception in adverse conditions is marked by innovative research strategies aimed at overcoming current limitations. By tackling the unique challenges presented by special adverse effects and harnessing the power of sensor fusion, more sophisticated and resilient autonomous driving technologies can be looked forward to being developed. These advancements will not only improve the safety and reliability of autonomous vehicles but also pave the way for their widespread adoption in various driving scenarios.

Bibliography

[1] K. F. Palmer and D. Williams, "Optical properties of water in the near infrared," *Journal of the Optical Society of America*, vol. 64, no. 8, pp. 1107–1110, 1974.

[2] F. M. Sogandares and E. S. Fry, "Absorption spectrum (340-640 nm) of pure water. i. photothermal measurements." *Applied optics*, vol. 36 33, pp. 8699–709, 1997.

[3] K. Rose, S. Eldridge, and L. Chapin, "The internet of things: An overview," *The internet society (ISOC)*, vol. 80, pp. 1–50, 2015.

[4] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE Access*, vol. 8, pp. 58 443–58 469, 2020.

[5] Y. Dong, Y. Zhong, W. Yu, M. Zhu, P. Lu, Y. Fang, J. Hong, and H. Peng, "Mcity data collection for automated vehicles study," *arXiv preprint* arXiv:1912.06258, 2019.

[6] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 146–177, 2023.

[7] K. E. Trenberth and Y. Zhang, "How often does it really rain?" *Bulletin of the American Meteorological Society*, vol. 99, no. 2, pp. 289–298, 2018.

[8] A. Mehra, M. Mandal, P. Narang, and V. Chamola, "Reviewnet: A fast and resource optimized network for enabling safe autonomous driving in hazy weather conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4256–4266, 2021.

[9] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann *et al.*, "Stanley: The robot that won the darpa grand challenge," *Journal of field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.

[10] A. Carballo, J. Lambert, A. Monrroy, D. Wong, P. Narksri, Y. Kitsukawa, E. Takeuchi, S. Kato, and K. Takeda, "LIBRE: The multiple 3D LiDAR dataset," in *Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 1094–1101.

[11] S. Zang, M. Ding, D. Smith, P. Tyler, T. Rakotoarivelo, and M. A. Kaafar, "The impact of adverse weather conditions on autonomous vehicles: how rain, snow, fog, and hail affect the performance of a self-driving car," *IEEE vehicular technology magazine*, vol. 14, no. 2, pp. 103–111, 2019.

[12] M. Jokela, M. Kutila, and P. Pyykönen, "Testing and validation of automotive point-cloud sensors in adverse weather conditions," *Applied Sciences*, vol. 9, no. 11, p. 2341, 2019.

[13] T. Fersch, A. Buhmann, A. Koelpin, and R. Weigel, "The influence of rain on small aperture LiDAR sensors," in *German Microwave Conference (GeMiC)*. IEEE, 2016, pp. 84–87.

[14] S. Hasirlioglu, I. Doric, C. Lauerer, and T. Brandmeier, "Modeling and simulation of rain for the test of automotive sensor systems," in *Intelligent Vehicles Symposium (IV)*. IEEE, 2016, pp. 286–291.

[15] J. Lambert, A. Carballo, A. M. Cano, P. Narksri, D. Wong, E. Takeuchi, and K. Takeda, "Performance analysis of 10 models of 3D LiDARs for automated driving," *IEEE Access*, vol. 8, pp. 131 699–131 722, 2020.

[16] M. Pitropov, D. E. Garcia, J. Rebello, M. Smart, C. Wang, K. Czarnecki, and S. Waslander, "Canadian adverse driving conditions dataset," *The International Journal of Robotics Research*, vol. 40, no. 4-5, pp. 681–690, 2021.

[17] T. Gao, F. Gao, G. Zhang, L. Liang, Y. Song, J. Du, and W. Dai, "Effects of temperature environment on ranging accuracy of LiDAR," in *Tenth International Conference on Digital Image Processing (ICDIP 2018)*, vol. 10806. SPIE, 2018, pp. 1915–1921.

[18] M. Trierweiler, T. Peterseim, and C. Neumann, "Automotive LiDAR pollution detection system based on total internal reflection techniques," in *Light-Emitting Devices, Materials, and Applications XXIV*, vol. 11302. SPIE, 2020, pp. 135–144.

[19] M. Trierweiler, P. Caldelas, G. Gröninger, T. Peterseim, and C. Neumann, "Influence of sensor blockage on automotive LiDAR systems," in *IEEE SENSORS*, 2019, pp. 1–4.

[20] R. Rogers, M. Vaughan, C. Hostetler, S. Burton, R. Ferrare, S. Young, J. Hair, M. Obland, D. Harper, A. Cook *et al.*, "Looking through the haze: evaluating the CALIPSO level 2 aerosol optical depth using airborne high spectral resolution LiDAR data," *Atmospheric Measurement Techniques*, vol. 7, no. 12, pp. 4317–4340, 2014.

[21] J. Rapp, J. Tachella, Y. Altmann, S. McLaughlin, and V. K. Goyal, "Advances in single-photon LiDAR for autonomous vehicles: Working principles, challenges, and recent advances," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 62–71, 2020.

[22] G. R. Osche and D. S. Young, "Imaging laser radar in the near and far infrared," *Proceedings of the IEEE*, vol. 84, no. 2, pp. 103–125, 1996.

[23] M. Kutila, P. Pyykönen, H. Holzhüter, M. Colomb, and P. Duthon, "Automotive LiDAR performance verification in fog and rain," in *International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 1695–1701.

[24] I. E. Commission *et al.*, "Safety of laser products-part 1: Equipment classification and requirements," *IEC 60825-1*, 2007.

[25] X. Sun, L. Zhang, Q. Zhang, and W. Zhang, "Si photonics for practical LiDAR solutions," *Applied Sciences*, vol. 9, no. 20, p. 4225, 2019.

[26] G. M. Hale and M. R. Querry, "Optical constants of water in the 200-nm to 200-μm wavelength region," *Applied optics*, vol. 12, no. 3, pp. 555–563, 1973.

[27] M. E. Warren, "Automotive lidar technology," in *Symposium on VLSI Circuits*, 2019, pp. C254–C255.

[28] J. Wojtanowski, M. Zygmunt, M. Kaszczuk, Z. Mierczyk, and M. Muzal, "Comparison of 905 nm and 1550 nm semiconductor laser rangefinders' performance deterioration due to adverse environmental conditions," *Opto-Electronics Review*, vol. 22, no. 3, pp. 183–190, 2014.

[29] I. I. Kim, B. McArthur, and E. J. Korevaar, "Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications," in *Optical Wireless Communications III*, vol. 4214. International Society for Optics and Photonics, 2001, pp. 26–37.

[30] V. Lidar, "A guide to LiDAR wavelengths for autonomous vehicles and driver assistance," *Article updated March*, 2021.

[31] F. Reway, W. Huber, and E. P. Ribeiro, "Test methodology for vision-based adas algorithms with an automotive camera-in-the-loop," in *IEEE International Conference on Vehicular Electronics and Safety (ICVES)*. IEEE, 2018, pp. 1–7.

[32] P. Radecki, M. Campbell, and K. Matzen, "All weather perception: Joint data association, tracking, and classification for autonomous ground vehicles," *arXiv preprint* arXiv:1605.02196, 2016.

[33] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 22–35, 2017.

[34] I. R. P. Series, "Specific attenuation model for rain for use in prediction methods," *Rec. P. 838-3, ITU-R*, 2005.

[35] R. Olsen, D. V. Rogers, and D. Hodge, "The aR^b relation in the calculation of rain attenuation," *IEEE Transactions on antennas and propagation*, vol. 26, no. 2, pp. 318–329, 1978.

[36] M. Ijaz, Z. Ghassemlooy, H. Le Minh, S. Rajbhandari, and J. Perez, "Analysis of fog and smoke attenuation in a free space optical communication link under controlled laboratory conditions," in *International Workshop on Optical Wireless Communications (IWOW)*. IEEE, 2012, pp. 1–3.

[37] I. Gultepe, "Measurements of light rain, drizzle and heavy fog," in *Precipitation: advances in measurement, estimation and prediction*. Springer, 2008, pp. 59–82.

[38] F. Norouzian, E. Marchetti, E. Hoare, M. Gashinova, C. Constantinou, P. Gardner, and M. Cherniakov, "Experimental study on low-THz automotive radar signal attenuation during snowfall," *IET Radar, Sonar & Navigation*, vol. 13, no. 9, pp. 1421–1427, 2019.

[39] R. M. Rasmussen, J. Vivekanandan, J. Cole, B. Myers, and C. Masters, "The estimation of snowfall rate using visibility," *Journal of Applied Meteorology*, vol. 38, no. 10, pp. 1542–1563, 1999.

[40] A. Carullo and M. Parvis, "An ultrasonic sensor for distance measurement in automotive applications," *IEEE Sensors journal*, vol. 1, no. 2, p. 143, 2001.

[41] J. Z. Varghese, R. G. Boone *et al.*, "Overview of autonomous vehicle sensors and systems," in *International Conference on Operations Excellence and Service Engineering*, 2015, pp. 178–191.

[42] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 682–11 692.

[43] M. Hahner, C. Sakaridis, M. Bijelic, F. Heide, F. Yu, D. Dai, and L. Van Gool, "LiDAR snowfall simulation for robust 3D object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 364–16 374.

[44] T. Yang, Q. Yu, Y. Li, and Z. Yan, "Learn to model and filter point cloud noise for a near-infrared tof LiDAR in adverse weather," *IEEE Sensors Journal*, 2023.

[45] R. Heinzler, F. Piewak, P. Schindler, and W. Stork, "CNN-based LiDAR point cloud de-noising in adverse weather," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2514–2521, 2020.

[46] N. Charron, S. Phillips, and S. L. Waslander, "De-noising of LiDAR point clouds corrupted by snowfall," in *Conference on Computer and Robot Vision (CRV)*. IEEE, 2018, pp. 254–261.

[47] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, "Towards 3d point cloud based object maps for household environments," *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 927–941, 2008.

[48] A. Kurup and J. Bos, "DSOR: A scalable statistical filter for removing falling snow from LiDAR point clouds in severe winter weather," *arXiv preprint* arXiv:2109.07078, 2021.

[49] J.-I. Park, J. Park, and K.-S. Kim, "Fast and accurate desnowing algorithm for LiDAR point clouds," *IEEE Access*, vol. 8, pp. 160 202–160 212, 2020.

[50] W. Wang, X. You, L. Chen, J. Tian, F. Tang, and L. Zhang, "A scalable and accurate de-snowing algorithm for LiDAR point clouds in winter," *Remote Sensing*, vol. 14, no. 6, p. 1468, 2022.

[51] M. Pfennigbauer, C. Wolf, J. Weinkopf, and A. Ullrich, "Online waveform processing for demanding target situations," in *Laser Radar Technology and Applications XIX; and Atmospheric Propagation XI*, vol. 9080. SPIE, 2014, pp. 142–151.

[52] A. M. Wallace, A. Halimi, and G. S. Buller, "Full waveform LiDAR for adverse weather conditions," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7064–7077, 2020.

[53] R. Tobin, A. Halimi, A. McCarthy, P. J. Soan, and G. S. Buller, "Robust real-time 3D imaging of moving scenes through atmospheric obscurant using single-photon LiDAR," *Scientific reports*, vol. 11, no. 1, pp. 1–13, 2021.

[54] Y. Hamzeh and S. A. Rawashdeh, "A review of detection and removal of raindrops in automotive vision systems," *Journal of Imaging*, vol. 7, no. 3, p. 52, 2021.

[55] Y. Wang, C. Ma, and B. Zeng, "Multi-decoding deraining network and quasi-sparsity based training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13 375–13 384.

[56] M. Zhou, J. Xiao, Y. Chang, X. Fu, A. Liu, J. Pan, and Z.-J. Zha, "Image de-raining via continual learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4907–4916.

[57] R. Quan, X. Yu, Y. Liang, and Y. Yang, "Removing raindrops and rain streaks in one go," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE/CVF, 2021, pp. 9147–9156.

[58] S. Ni, X. Cao, T. Yue, and X. Hu, "Controlling the rain: From removal to rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6328–6337.

[59] H. Wang, Z. Yue, Q. Xie, Q. Zhao, Y. Zheng, and D. Meng, "From rain generation to rain removal," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 791–14 801.

[60] Y. Ye, Y. Chang, H. Zhou, and L. Yan, "Closing the loop: Joint rain generation and removal via disentangled image translation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2053–2062.

[61] Z. Yue, J. Xie, Q. Zhao, and D. Meng, "Semi-supervised video deraining with dynamical rain generator," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 642–652.

[62] K. Zhang, R. Li, Y. Yu, W. Luo, and C. Li, "Deep dense multi-scale network for snow removal using semantic and depth priors," *IEEE Transactions on Image Processing*, vol. 30, pp. 7419–7431, 2021.

[63] M. Bijelic, T. Gruber, and W. Ritter, "A benchmark for LiDAR sensors in fog: Is detection breaking down?" in *Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 760–767.

[64] S.-L. Lin and B.-H. Wu, "Application of kalman filter to improve 3D LiDAR signals of autonomous vehicles in adverse weather," *Applied Sciences*, vol. 11, no. 7, p. 3018, 2021.

[65] A. U. Shamsudin, K. Ohno, T. Westfechtel, S. Takahiro, Y. Okada, and S. Tadokoro, "Fog removal using laser beam penetration, laser intensity, and geometrical features for 3D measurements in fog-filled room," *Advanced Robotics*, vol. 30, no. 11-12, pp. 729–743, 2016.

[66] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[67] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1857–1865.

[68] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2849–2857.

[69] H. Yang, A. Carballo, and K. Takeda, "Disentangled bad weather removal gan for pedestrian detection," in *2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring)*. IEEE, 2022, pp. 1–6.

[70] D.-W. Jaw, S.-C. Huang, and S.-Y. Kuo, "Desnowgan: An efficient single image snow removal framework using cross-resolution lateral connection and GANs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 4, pp. 1342–1350, 2020.

[71] D. Engin, A. Genç, and H. Kemal Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 825–833.

[72] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2808–2817.

[73] A. E. Sallab, I. Sobh, M. Zahran, and N. Essam, "LiDAR sensor modeling and data augmentation with GANs for autonomous driving," *arXiv preprint* arXiv:1905.07290, 2019.

[74] I. Sobh, L. Amin, S. Abdelkarim, K. Elmadawy, M. Saeed, O. Abdeltawab, M. Gamal, and A. El Sallab, "End-to-end multi-modal sensors fusion system for urban automated driving," *NIPS Workshop on Machine Learning for Intelligent Transportation Systems 2018*, 2018.

[75] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3354–3361.

[76] J. Lee, D. Shiotsuka, T. Nishimori, K. Nakao, and S. Kamijo, "GAN-based LiDAR translation between sunny and adverse weather for autonomous driving and driving simulation," *Sensors*, vol. 22, no. 14, p. 5287, 2022.

[77] M. Aldibaja, N. Suganuma, and K. Yoneda, "Improving localization accuracy for autonomous driving in snow-rain environments," in *International Symposium on System Integration (SII)*. IEEE/SICE, 2016, pp. 212–217.

[78] M. Aldibaja, R. Yanase, A. Kuramoto, T. H. Kim, K. Yoneda, and N. Suganuma, "Improving lateral autonomous driving in snow-wet environments based on road-mark reconstruction using principal component analysis," *IEEE Intelligent Transportation Systems Magazine*, vol. 13, no. 4, pp. 116–130, 2021.

[79] M. Aldibaja, N. Suganuma, and K. Yoneda, "Robust intensity-based localization method for autonomous driving on snow–wet road surface," *IEEE Transactions on industrial Informatics*, vol. 13, no. 5, pp. 2369–2378, 2017.

[80] R. H. Rasshofer, M. Spies, and H. Spies, "Influences of weather phenomena on automotive laser radar systems," *Advances in Radio Science*, vol. 9, no. B. 2, pp. 49–60, 2011.

[81] A. Carballo, A. Monrroy, D. Wong, P. Narksri, J. Lambert, Y. Kitsukawa, E. Takeuchi, S. Kato, and K. Takeda, "Characterization of multiple 3d lidars for localization and mapping performance using the ndt algorithm," in *2021 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*. IEEE, 2021, pp. 327–334.

[82] N. Nishizawa and M. Yamanaka, "Characteristics of spectral peaking in coherent supercontinuum generation," in *2021 Conference on Lasers and Electro-Optics (CLEO)*. IEEE, 2021, pp. 1–2.

[83] J. Bergius, "LiDAR point cloud de-noising for adverse weather," Ph.D. dissertation, Halmstad University, 2022.

[84] H. Yang, M. Ding, A. Carballo, Y. Zhang, K. Ohtani, Y. Niu, M. Ge, Y. Feng, and K. Takeda, "Synthesizing realistic snow effects in driving images using GANs and real data with semantic guidance," in *2023 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2023, pp. 1–6.

[85] C. Ancuti, C. O. Ancuti, and R. Timofte, "Ntire 2018 challenge on image dehazing: Methods and results," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 891–901.

[86] V. Muşat, I. Fursa, P. Newman, F. Cuzzolin, and A. Bradley, "Multi-weather city: Adverse weather stacking for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2906–2915.

[87] A. Von Bernuth, G. Volk, and O. Bringmann, "Simulating photo-realistic snow and fog on existing images for enhanced CNN training and evaluation," in *Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 41–46.

[88] A. Guo, Y. Feng, and Z. Chen, "LiRTest: augmenting LiDAR point clouds for automated testing of autonomous driving systems," in *Proceedings of the 31st ACM SIGSOFT International Symposium on Software Testing and Analysis*, 2022, pp. 480–492.

[89] A. Piroli, V. Dallabetta, M. Walessa, D. Meissner, J. Kopp, and K. Dietmayer, "Robust 3D object detection in cold weather conditions," in *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022, pp. 287–294.

[90] Z. Chen, Y. Wang, Y. Yang, and D. Liu, "PSD: Principled synthetic-to-real dehazing guided by physical priors," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE/CVF, 2021, pp. 7180–7189.

[91] L. Yahiaoui, M. Uřičář, A. Das, and S. Yogamani, "Let the sunshine in: Sun glare detection on automotive surround-view cameras," *Electronic Imaging*, vol. 2020, no. 16, pp. 80–1, 2020.

[92] M. Uřičář, G. Sistu, H. Rashed, A. Vobecky, V. R. Kumar, P. Krizek, F. Burger, and S. Yogamani, "Let's get dirty: GAN based data augmentation for camera lens soiling detection in autonomous driving," in *Winter Conference on Applications of Computer Vision (WACV)*. IEEE/CVF, 2021, pp. 766–775.

[93] M. Uřičář, P. Křížek, G. Sistu, and S. Yogamani, "Soilingnet: Soiling detection on automotive surround-view cameras," in *Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 67–72.

[94] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[95] V. Vaibhav, K. R. Konda, C. Kondapalli, K. Praveen, and B. Kondoju, "Real-time fog visibility range estimation for autonomous driving applications," in *International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–6.

[96] M.-H. Le, C.-H. Cheng, and D.-G. Liu, "An efficient adaptive noise removal filter on range images for LiDAR point clouds," *Electronics*, vol. 12, no. 9, p. 2150, 2023.

[97] D. Steinhauser, P. Held, B. Thöresz, and T. Brandmeier, "Towards safe autonomous driving: Challenges of pedestrian detection in rain with automotive radar," in *European Radar Conference (EuRAD)*. IEEE, 2021, pp. 409–412.

[98] G. Li, Y. Yang, and X. Qu, "Deep learning approaches on pedestrian detection in hazy weather," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 10, pp. 8889–8899, 2019.

[99] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *International Journal of Computer Vision*, vol. 126, no. 9, pp. 973–992, 2018.

[100] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3194–3203.

[101] H. Yang, A. Carballo, Y. Zhang, and K. Takeda, "Framework for generation and removal of multiple types of adverse weather from driving scene images," *Sensors*, vol. 23, no. 3, p. 1548, 2023.

[102] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[103] J. Razlaw, D. Droeschel, D. Holz, and S. Behnke, "Evaluation of registration methods for sparse 3D laser scans," in *2015 European Conference on Mobile Robots (ECMR)*. IEEE, 2015, pp. 1–7.

[104] A. Mertan, D. J. Duff, and G. Unal, "Single image depth estimation: An overview," *Digital Signal Processing*, p. 103441, 2022.

[105] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," *Advances in neural information processing systems*, vol. 27, 2014.

[106] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[107] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "Optics: Ordering points to identify the clustering structure," *ACM Sigmod record*, vol. 28, no. 2, pp. 49–60, 1999.

[108] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[109] S. Kokoska and D. Zwillinger, *CRC standard probability and statistics tables and formulae*. Crc Press, 2000.

[110] V. Zyrianov, X. Zhu, and S. Wang, "Learning to generate realistic LiDAR point clouds," in *European Conference on Computer Vision*. Springer, 2022, pp. 17–35.

[111] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[112] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 649–666.

[113] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*. Springer, 2020, pp. 319–345.

[114] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.

[115] Q.-H. Pham, P. Sevestre, R. S. Pahwa, H. Zhan, C. H. Pang, Y. Chen, A. Mustafa, V. Chandrasekhar, and J. Lin, "A*3D dataset: Towards autonomous driving in challenging environments," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2267–2273.

[116] X. Huang, P. Wang, X. Cheng, D. Zhou, Q. Geng, and R. Yang, "The apolloscape open dataset for autonomous driving and its application," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 10, pp. 2702–2719, 2019.

[117] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," *International Journal of Robotics Research*, vol. 36, no. 1, pp. 3–15, 2017.

[118] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan *et al.*, "Argoverse: 3d tracking and forecasting with rich maps," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 8748–8757.

[119] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2446–2454.

[120] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of michigan north campus long-term vision and LiDAR dataset," *The International Journal of Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2016.

[121] Y. Choi, N. Kim, S. Hwang, K. Park, J. S. Yoon, K. An, and I. S. Kweon, "KAIST multi-spectral day/night data set for autonomous and assisted driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 934–948, 2018.

[122] J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn *et al.*, "A2D2: Audi autonomous driving dataset," *arXiv preprint* arXiv:2004.06320, 2020.

[123] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "RADIATE: A radar dataset for automotive perception in bad weather," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 1–7.

[124] Z. Yan, L. Sun, T. Krajník, and Y. Ruichek, "EU long-term dataset with multiple sensors for autonomous driving," in *International Conference on Intelligent Robots and Systems (IROS)*. IEEE/RSJ, 2020, pp. 10 697–10 704.

[125] J. P. Bos, D. Chopp, A. Kurup, and N. Spike, "Autonomy at the end of the earth: an inclement weather autonomous driving data set," in *Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure 2020*, vol. 11415. SPIE, 2020, pp. 36–48.

[126] J. P. Bos, A. Kurup, D. Chopp, and Z. Jeffries, "The michigan tech autonomous winter driving data set: year two," in *Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure 2021*, vol. 11748. SPIE, 2021, pp. 57–65.

[127] K. Burnett, D. J. Yoon, Y. Wu, A. Z. Li, H. Zhang, S. Lu, J. Qian, W.-K. Tseng, A. Lambert, K. Y. Leung *et al.*, "Boreas: A multi-season autonomous driving dataset," *The International Journal of Robotics Research*, vol. 42, no. 1-2, pp. 33–42, 2023.

[128] T. Ort, I. Gilitschenski, and D. Rus, "GROUNDED: The localizing ground penetrating radar evaluation dataset." in *Robotics: Science and Systems*, vol. 2, 2021.

[129] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2636–2645.

[130] G. Neuhold, T. Ollmann, S. Rota Bulo, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in *International conference on computer vision (ICCV)*. IEEE, 2017, pp. 4990–4999.

[131] M. Braun, S. Krebs, F. Flohr, and D. M. Gavrila, "Eurocity persons: A novel benchmark for person detection in traffic scenes," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1844–1861, 2019.

[132] Z. Che, G. Li, T. Li, B. Jiang, X. Shi, X. Zhang, Y. Lu, G. Wu, Y. Liu, and J. Ye, "D^2-city: A large-scale dashcam video dataset of diverse traffic scenarios," *arXiv preprint* arXiv:1904.01975, 2019.

[133] J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, "DDD17: End-to-end DAVIS driving dataset," *arXiv preprint* arXiv:1711.01458, 2017.

[134] Y. Lei, T. Emaru, A. A. Ravankar, Y. Kobayashi, and S. Wang, "Semantic image segmentation on snow driving scenarios," in *International Conference on Mechatronics and Automation (ICMA)*. IEEE, 2020, pp. 1094–1100.

[135] C. Sakaridis, D. Dai, and L. Van Gool, "Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 765–10 775.

[136] P. Wenzel, R. Wang, N. Yang, Q. Cheng, Q. Khan, L. von Stumberg, N. Zeller, and D. Cremers, "4Seasons: A cross-season dataset for multi-weather SLAM in autonomous driving," in *DAGM German Conference on Pattern Recognition (GCPR)*. Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM), 2021, pp. 404–417.

[137] F. Tung, J. Chen, L. Meng, and J. J. Little, "The raincouver scene parsing benchmark for self-driving in adverse weather and at night," *Robotics and Automation Letters (RA-L)*, vol. 2, no. 4, pp. 2188–2193, 2017.

[138] O. Zendel, K. Honauer, M. Murschitz, D. Steininger, and G. F. Dominguez, "Wilddash-creating hazard-aware benchmarks," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 402–416.

[139] K. Basterretxea, V. Martínez, J. Echanobe, J. Gutiérrez-Zaballa, and I. Del Campo, "Hsi-drive: A dataset for the research of hyperspectral image processing applied to autonomous driving systems," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 866–873.

[140] M. A. Kenk and M. Hassaballah, "Dawn: vehicle detection in adverse weather nature dataset," *arXiv preprint* arXiv:2008.05402, 2020.

[141] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez, "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3234–3243.

[142] S. R. Richter, Z. Hayder, and V. Koltun, "Playing for benchmarks," in *International Conference on Computer Vision (ICCV)*. IEEE, 2017, pp. 2213–2222.

[143] D. Liu, Y. Cui, Z. Cao, and Y. Chen, "A large-scale simulation dataset: Boost the detection accuracy for special weather conditions," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[144] R. Karlsson, D. R. Wong, K. Kawabata, S. Thompson, and N. Sakai, "Probabilistic rainfall estimation from automotive LiDAR," *arXiv preprint arXiv:2104.11467*, 2021.

[145] C. Lu, D. Lin, J. Jia, and C.-K. Tang, "Two-class weather classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3718–3725.

[146] Z. Zhang and H. Ma, "Multi-class weather classification on single images," in *International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 4396–4400.

[147] M. Elhoseiny, S. Huang, and A. Elgammal, "Weather classification with deep convolutional neural networks," in *International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 3349–3353.

[148] U. Haberlandt and M. Sester, "Areal rainfall estimation using moving cars as rain gauges–a modelling study," *Hydrology and Earth System Sciences*, vol. 14, no. 7, pp. 1139–1151, 2010.

[149] D. J. Hill, "Assimilation of weather radar and binary ubiquitous sensor measurements for quantitative precipitation estimation," *Journal of Hydroinformatics*, vol. 17, no. 4, pp. 598–613, 2015.

[150] C. Goodin, D. Carruth, M. Doude, and C. Hudson, "Predicting the influence of rain on LiDAR in ADAS," *Electronics*, vol. 8, no. 1, p. 89, 2019.

[151] M. Bartos, H. Park, T. Zhou, B. Kerkez, and R. Vasudevan, "Windshield wipers on connected vehicles produce high-accuracy rainfall maps," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

[152] R. Heinzler, P. Schindler, J. Seekircher, W. Ritter, and W. Stork, "Weather influence and classification with automotive lidar sensors," in *2019 IEEE intelligent vehicles symposium (IV)*. IEEE, 2019, pp. 1527–1534.

[153] C. Dannheim, C. Icking, M. Mäder, and P. Sallis, "Weather detection in vehicles by means of camera and LiDAR systems," in *International Conference on Computational Intelligence, Communication Systems and Networks*. IEEE, 2014, pp. 186–191.

[154] Y. Zhang, M. Ding, H. Yang, Y. Niu, Y. Feng, M. Ge, A. Carballo, and K. Takeda, "LiDAR point cloud translation between snow and clear conditions using depth images and GANs," in *2023 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2023, pp. 1–7.

[155] Y. Zhang, M. Ding, H. Yang, Y. Niu, Y. Feng, K. Ohtani, and K. Takeda, "L-DIG: A GAN-based method for LiDAR point cloud processing under snow driving conditions," *Sensors*, vol. 23, no. 21, p. 8660, 2023.

[156] A. Simonelli, S. R. Bulo, L. Porzi, M. López-Antequera, and P. Kontschieder, "Disentangling monocular 3D object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1991–1999.

[157] S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang, and H. Li, "PV-RCNN: Point-voxel feature set abstraction for 3D object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 529–10 538.

[158] Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, p. 3337, 2018.

[159] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu *et al.*, "MMDetection: Open mmlab detection toolbox and benchmark," *arXiv preprint* arXiv:1906.07155, 2019.

[160] J. R. Vargas Rivero, T. Gerbich, V. Teiluf, B. Buschardt, and J. Chen, "Weather classification using an automotive LiDAR sensor based on detections on asphalt and atmosphere," *Sensors*, vol. 20, no. 15, p. 4306, 2020.

List of Publications

Journal Publications

1. **Yuxiao Zhang**, Alexander Carballo, Hanting Yang, and Kazuya Takeda. "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey." ISPRS Journal of Photogrammetry and Remote Sensing, 196 (2023): 146-177. DOI: https://doi.org/10.1016/j.isprsjprs.2022.12.021

2. **Yuxiao Zhang**, Ming Ding, Hanting Yang, Yingjie Niu, Yan Feng, Kento Ohtani, and Kazuya Takeda. 2023. "L-DIG: A GAN-Based Method for LiDAR Point Cloud Processing under Snow Driving Conditions" Sensors 23, no. 21: 8660. DOI: https://doi.org/10.3390/s23218660

3. Hanting Yang, Alexander Carballo, **Yuxiao Zhang**, and Kazuya Takeda. 2023. "Framework for Generation and Removal of Multiple Types of Adverse Weather from Driving Scene Images" Sensors 23, no. 3: 1548. DOI: https://doi.org/10.3390/s23031548

4. Hanting Yang, Alexander Carballo, **Yuxiao Zhang**, and Kazuya Takeda. 2023. "Controllable Unsupervised Snow Synthesis by Latent Style Space Manipulation" Sensors 23, no. 20: 8398. DOI: https://doi.org/10.3390/s23208398

Conference Publications

1. **Yuxiao Zhang**, Ming Ding, Hanting Yang, Yingjie Niu, Yan Feng, Maoning Ge, Alexander Carballo, and Kazuya Takeda. "LiDAR Point Cloud Translation

Between Snow and Clear Conditions Using Depth Images and GANs." In 2023 IEEE Intelligent Vehicles Symposium (IV), pp. 1-7. IEEE, 2023. DOI: https://doi.org/10.1109/IV55152.2023.10186814

2. Yingjie Niu, Ming Ding, **Yuxiao Zhang**, Maoning Ge, Hanting Yang, and Kazuya Takeda. "Open-world driving scene segmentation via multi-stage and multi-modality fusion of vision-language embedding." In 2023 IEEE Intelligent Vehicles Symposium (IV), pp. 1-6. IEEE, 2023. DOI: https://doi.org/10.1109/IV55152.2023.10186652

3. Hanting Yang, Ming Ding, Alexander Carballo, **Yuxiao Zhang**, Kento Ohtani, Yingjie Niu, Maoning Ge, Yan Feng, and Kazuya Takeda. "Synthesizing Realistic Snow Effects in Driving Images Using GANs and Real Data with Semantic Guidance." In 2023 IEEE Intelligent Vehicles Symposium (IV), pp. 1-6. IEEE, 2023. DOI: https://doi.org/10.1109/IV55152.2023.10186565

4. Yingjie Niu, Ming Ding, **Yuxiao Zhang**, Kento Ohtani, and Kazuya Takeda. "Auditory and visual warning information generation of the risk object in driving scenes based on weakly supervised learning." In 2022 IEEE Intelligent Vehicles Symposium (IV), pp. 1572-1577. IEEE, 2022. DOI: https://doi.org/10.1109/IV51971.2022.9827382