Analysis of Point Cloud Domain Gap Effects for 3D Object Detection Evaluation

Aitor Iglesias[1,2][a], Mikel García[1,2][b], Nerea Aranjuelo[1][c], Ignacio Arganda-Carreras[2,3,4,5][d] and Marcos Nieto[1][e]

[1] *Fundación Vicomtech, Connected and Cooperative Automated Systems, Spain*
[2] *University of the Basque Country (UPV/EHU), Donostia - San Sebastian, Spain*
[3] *IKERBASQUE, Basque Foundation for Science, Bilbao, Spain*
[4] *Donostia International Physics Center (DIPC), Donostia - San Sebastian, Spain*
[5] *Biofisika Institute, Leioa, Spain*

Keywords: Point Clouds, Deep Learning, Domain Gap, Object Detection, Simulation.

Abstract: The development of autonomous driving systems heavily relies on high-quality LiDAR data, which is essential for robust object detection and scene understanding. Nevertheless, obtaining a substantial amount of such data for effective training and evaluation of autonomous driving algorithms is a major challenge. To overcome this limitation, recent studies are taking advantage of advancements in realistic simulation engines, such as CARLA, which have provided a breakthrough in generating synthetic LiDAR data that closely resembles real-world scenarios. However, these data are far from being identical to real data. In this study, we address the domain gap between real LiDAR data and synthetic data. We train deep-learning models for object detection using real data. Then, those models are rigorously evaluated using synthetic data generated in CARLA. By quantifying the discrepancies between the model's performance on real and synthetic data, the present study shows that there is indeed a domain gap between the two types of data and does not affect equal to different model architectures. Finally, we propose a method for synthetic data processing to reduce this domain gap. This research contributes to enhancing the use of synthetic data for autonomous driving systems.

1 INTRODUCTION

In the current context of the automotive industry, road safety and autonomous driving have emerged as critical areas of research and development. Accurate and reliable detection of objects in a vehicle's near environment is essential to ensure the safety of passengers, pedestrians, and other road users. One of the key challenges in achieving this level of accuracy and reliability lies in the development of advanced sensor technologies that can provide real-time data about the surrounding environment. As technology advances, LiDAR point clouds have arisen as a promising source of three-dimensional data, offering a detailed representation of the environment around the vehicle (Li and Ibanez-Guzman, 2020; Li et al., 2021a). These representation presents an opportunity to significantly enhance 3D object detection systems compared to traditional image-based methods (Li et al., 2021b).

Data acquisition is crucial for developing and validating automotive object detection algorithms. For example, Kalra and Paddock (2016) estimate that achieving a 1.09 fatalities per 100 million miles rate with 95% confidence for autonomous vehicles would require 275 million miles. However, obtaining large, diverse, and representative real-world driving datasets can be costly and limiting in terms of time and resources. To address these limitations, the generation of synthetic data is gaining popularity as a practical solution. Simulators such as Dosovitskiy et al. (2017) and Rong et al. (2020) enable the creation of virtual environments that accurately replicate driving conditions, offering the ability to generate substantial volumes of data in a controlled and diverse manner.

Generated data is often used to train deep learning models or to validate automotive functions. However,

[a] https://orcid.org/0009-0007-0828-991X
[b] https://orcid.org/0000-0002-3973-7267
[c] https://orcid.org/0000-0002-7853-6708
[d] https://orcid.org/0000-0003-0229-5722
[e] https://orcid.org/0000-0001-9879-0992

Iglesias, A., García, M., Aranjuelo, N., Arganda-Carreras, I. and Nieto, M.
Analysis of Point Cloud Domain Gap Effects for 3D Object Detection Evaluation.
DOI: 10.5220/0012357200003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 4: VISAPP*, pages 278-285
ISBN: 978-989-758-679-8; ISSN: 2184-4321

Studies (Dworak et al., 2019; Huch et al., 2023) reveal a domain gap between real and synthetic data, which can negatively affect the results. The domain gap refers to the mismatch or disparity between the data distribution. This paper examines the domain gap in point cloud-based object detection models trained with real data and presents a method to mitigate it.

In the present study, we have designed a workflow as shown in Figure 1 to improve the understanding of the domain gap between synthetic and real point cloud data by performing the following contributions:

- **Synthetic Data Generation Pipeline.** We propose a synthetic data generation pipeline to generate data in a standardized format, which can be easily adapted to different sensor setups. We used this pipeline to generate our own synthetic dataset: `https://datasets.vicomtech.org/di21-carla-point-cloud-data/carla-point-cloud-data.md`.
- **Analysis of the Generalization of Models Against Synthetic Data.** We evaluate different architectures of LiDAR-based 3D object detection models with the generated synthetic dataset. We then analyze the ability of each model to adapt to this domain.
- **Investigation of the Domain Gap Between Real and Synthetic Data.** We compare the generated synthetic point clouds with point clouds of a dataset with real data to identify the differences between both domains.
- **Domain Gap Reduction.** We propose a method to reduce the domain gap between real and synthetic data. This method is able to reduce the precision difference of the models between both domains.

The rest of this paper is organized into five sections. The first section presents the state of the art of point cloud datasets and deep learning models, especially emphasizing the existing knowledge gap regarding the response of models to synthetic data. The second section describes the methodology used during this research. The third section details the experiments done during the investigation. In the fourth section, an exhaustive analysis of the results can be found. Finally, the last section compiles the conclusions and future work of the research.

2 RELATED WORK

The development of autonomous driving has relied heavily on the availability of high-quality datasets for training and validating autonomous driving algorithms. Among the most popular datasets are KITTI (Liao et al., 2021), Waymo (Sun et al., 2019), and nuScenes (Caesar et al., 2019). However, the datasets used in autonomous driving present some challenges, such as data limitations, particularly in acquiring data for rare or extreme driving scenarios that are difficult to obtain.

Given the considerable expense and practical difficulties associated with obtaining only real data to meet the necessary quantity, variability, and diversity, synthetic data are widely utilized as supplementary resources (Qiao and Zulkernine, 2023; Wang et al., 2019; Inan et al., 2023). These data are generated to cover hazardous or unusual situations that may not be easily found in real data and avoid the process and cost associated with obtaining and annotating real data. For the generation of these data, simulation environments such as CARLA (Dosovitskiy et al., 2017) or LGSVL (Rong et al., 2020) are used and even datasets exclusively with synthetic data have been published (Kloukiniotis et al., 2022; Sekkat et al., 2022; Xu et al., 2022).

Synthetic data are often used to study functions such as object detection. Object detection in point clouds is a constantly evolving area of research in the fields of machine perception, computer vision, and robotics (Li et al., 2021b). This task is very challenging due to the complexity and variability of point clouds, which are sparse and unordered and may contain noise, occlusions, and objects with very different shapes and sizes. In recent years, several deep learning models have been proposed for the detection of objects in point clouds (Lang et al., 2018; Zhu et al., 2020; Yin et al., 2020; Bai et al., 2022). However, these models are still far from achieving the results needed for a fully autonomous vehicle as described by Martínez-Díaz and Soriguera (2018). However, these models are usually trained on public datasets where capturing the full spectrum of real-world edge scenarios is often unfeasible, resulting in a knowledge gap regarding how the trained models will perform in critical edge scenarios. Consequently, the use of synthetic data has emerged as a pivotal strategy to supplement the available datasets.

Although different research proves the effectiveness of synthetic data, different works (Dworak et al., 2019; Huch et al., 2023) demonstrate that the impact of a domain gap may diverge across deep learning models depending on their architecture. This is especially accentuated when using synthetic data. Furthermore, the state-of-the-art LiDAR-based 3D detection models are not analyzed in terms of generalization or sensitivity to this domain gap.

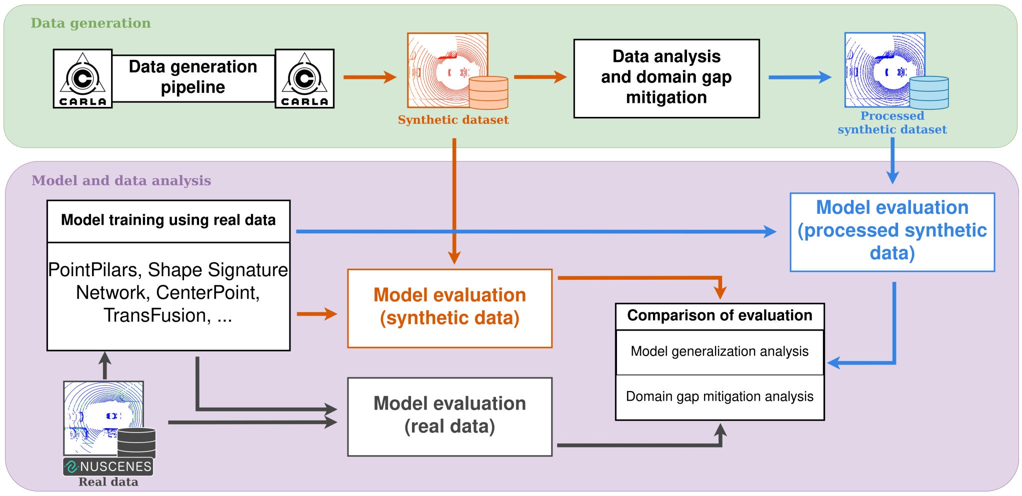

Figure 1: Methodology of the research. First, we generate a synthetic dataset, analyze it, and propose a domain gap mitigation strategy. Then, we train different architectures of LiDAR-based 3D object detection models and evaluate their performance with real data, synthetic data, and processed synthetic data.

3 METHODOLOGY

The methodology proposed in this research is divided into two parts, as can be seen in Figure 1. The first part is focused on data generation and processing, and the second one is about model training and evaluation. In the first part, we first generate a synthetic dataset using the CARLA simulator, thenceforth we study the peculiarities of these data and propose a strategy to mitigate the potential domain gap. In the second part, we first train state-of-the-art LiDAR-based 3D object detection models with different architectures. Subsequently, we evaluate these models with a real-world dataset (the nuScenes dataset in this case), the synthetic dataset, and the processed synthetic dataset. Finally, the results of the evaluations for the different domains are compared and analyzed.

3.1 Data Generation Pipeline

In this work, we design a pipeline for the creation of synthetic datasets using the CARLA simulator (Dosovitskiy et al., 2017). This dataset will later be used to evaluate object detection models in point clouds. We use CARLA due to its extensive usage in state-of-the-art research (Kloukiniotis et al., 2022; Sekkat et al., 2022; Xu et al., 2022) and its versatile capabilities, but other alternatives could be used as well.

3.1.1 Data Acquisition

To generate point clouds using the CARLA simulator we use two virtual sensors attached to a vehicle of the simulator. First, we use a LiDAR sensor to generate point clouds. Second, we use a semantic LiDAR, which does not represent a real sensor but provides the type of object the simulated LiDAR rays collide with. Both LiDARs are needed to generate the data and the corresponding automatic annotations. The data generation process is conducted as follows.

First, before data generation, we define the sensor information of our setup in the standardized OpenLABEL[1] format. This format defines how information on the extrinsic and intrinsic parameters of the sensors of the vehicle must be indicated. The flexibility of this format allows easily defining different sensor setups to be simulated. Once we have the sensor information defined in an OpenLabel format file, the CARLA map where the scene will take place is selected and loaded. Then, we load the ego-vehicle and attach the sensors specified in the file. After loading the map, the sensors, and the vehicle, several actors are generated at different points of the scene, these actors are selected from a distribution of cars, trucks, vans, buses, bikes, motorcycles, and pedestrians. After the loading of different actors, all of them are activated in auto-pilot mode, enabling independent movement in compliance with traffic regulations.

Before starting to generate and annotate data we define a margin of 200 frames (20 seconds) for each vehicle to reach the appropriate speed for the position they are in, as when they are loaded, they are stationary. Within frame, a point cloud is stored. Then, we use the semantic LiDAR to automatically generate an-

[1] https://www.asam.net/standards/detail/openlabel/

notations for the actors appearing in the point cloud. Generated data can be seen in Figure 2.

3.1.2 Data Structure

The generated dataset can be categorized into two primary components: point clouds and their corresponding annotations, with the annotations being stored in OpenLABEL format. Each scene is defined by an OpenLABEL file, containing the timestamp of every frame. Each frame encapsulates the following data:

- Transformation matrix: This field holds a matrix detailing the vehicle's position and orientation in relation to the ego vehicle's initial position within the frame.
- Objects: This section comprises annotations for the objects depicted in the frame, each consisting of: 3D bounding box, velocity vector, acceleration vector, velocity, acceleration and class.

3.2 Data Analysis and Domain Gap Mitigation

Once we generate the data, we compare the synthetic point clouds with point clouds of a real-world dataset (more specifically, the nuScenes dataset). We identify the following differences:

1. All the lasers used by a simulated LiDAR always obtain information. This does not happen with a real LiDAR, because either by the surface or the material with which the laser collides or by the weather, many of the points are lost. However, this does not happen in a simulated environment.
2. The points form perfect circles in synthetic point clouds. This is a result of the simulation environment. Real LiDAR cannot obtain the same accuracy as sensors used in simulators, due to sensor errors, surface irregularities, weather, or even motion distortion. This is why real point clouds do not have perfect rings like synthetic clouds.
3. Finally, another factor that causes these domains to be different is the intensity value of the points. Intensity calculation of CARLA lacks realism as it solely relies on point distance, it does not take into account the material with which the laser collides or the angle of incidence.

These differences can be seen in Figure 3, which illustrates variations between a point cloud from the nuScenes dataset and a synthetic dataset. The color of the points varies based on the intensity of the points. Based on the detected data disparities, we propose to process the virtually generated clouds to mitigate the difference between the two domains as follows.

1. To solve the problem of point loss in the real environment, some points are randomly removed from the cloud.
2. To solve the problem of perfect virtual point clouds, in comparison with the inherent noise obtained from the measurements of real LiDAR, we add a random noise distribution to the points.
3. Finally, to deal with the problem of intensity values, we propose to change all intensity values to a common value.

3.3 Model and Data Analysis

To study the differences between real and synthetic data, and the performance of object detection models on real and synthetic point clouds. We train different architectures of these models, along with models having various configurations in terms of point cloud accumulation. In this way, a more exhaustive study of the domain gap between real and synthetic data can be performed. We train the models with real data and then evaluate them with real data as well as with synthetic and processed synthetic data.

4 EXPERIMENTS

In the upcoming section, we delve into the experimental phase of our study, where our first objective is to generate a synthetic dataset. We list the parameters used in data processing and summarize the obtained dataset. We also present the LiDAR-based 3D models we utilized, as well as their configuration.

4.1 Synthetic Data

The generated dataset consists of 9,600 point clouds, which correspond to 96 scenes with 100 point clouds each, at a frame rate of 0.1 seconds. Each point cloud is labeled with information about the objects present in it, encompassing 6 different classes: car, pedestrian, bicycle, motorcycle, bus, and truck. Detailed quantitative information can be found in Table 1.

4.1.1 Data Generation

For the generation of these data, four maps from the CARLA simulation environment are used to give more variety to the dataset scenes these maps have rural and urban environments. In addition, the default weather and time settings are used, because, in the simulator, the LiDAR sensor is not affected by these changes. In each scene, 30 cars, 10 trucks,

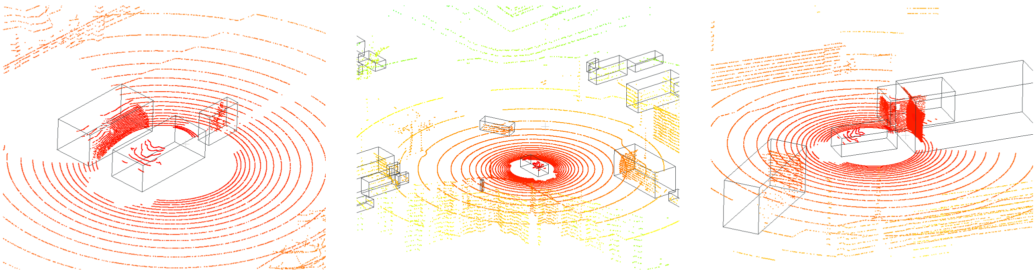

Figure 2: Example of generated point clouds. Each image is a point cloud of a different scene.

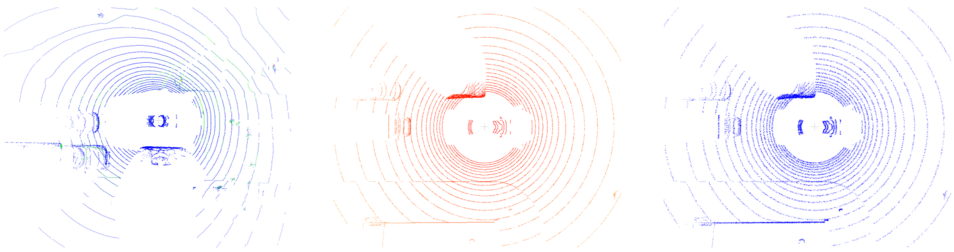

Figure 3: Comparison between point clouds of different domains. In the first column a point cloud of the nuScenes dataset, in the second column a point cloud of the synthetic dataset and in the third column the point cloud of the synthetic dataset after applying our processing on it.

Table 1: Number of annotations of the synthetic dataset.

Object Class	Number of annotations
car	48,878
truck	26,273
bus	13,097
pedestrian	29,373
motorcycle	20,526
bicycle	18,177
total	156,324

10 vans, 10 buses, 20 bikes, 20 motorcycles and 60 pedestrians are loaded. We employ this distribution in an attempt to simulate the distribution of the dataset against which the models are evaluated, the nuScenes dataset (Caesar et al., 2019).

4.1.2 Synthetic Data Processing

The values for the reduction and translation of the points (Section 3.2) are selected trying to obtain clouds as close to reality as possible. We empirically found the optimal percentage of points to remove is 10% and the optimal value for the translation is a range of $[-0.05, 0.05]$ meters in each axis. As for the intensity value, we decide to take the value 1 as it is one of the most common in real point clouds since the car is surrounded by poorly reflecting objects such as the road. Note that these changes do not completely eliminate the domain gap, but they do reduce it. In Figure 3 it can be seen a comparison of the change of the cloud before and after applying the processing.

4.2 Model Configuration and Training

For our experiments, we employ the MMDetection3D library (Contributors, 2020) due to its rich assortment of point cloud object detection models. We use the nuScenes (Caesar et al., 2019) dataset. In this research, we employ two categories of models of the state-of-the-art with different architectures and properties. The first category includes PointPillars (Lang et al., 2018) and Shape Signature Network (Zhu et al., 2020), which convert the point cloud into an intermediate representation like voxels or pillars. The second category comprises CenterPoint (Yin et al., 2020) and TransFusion (Bai et al., 2022) (LiDAR only), which not only transform the point cloud into an intermediate representation but also incorporate the top-view of the point cloud. To assess the impact of point cloud accumulation, we conduct model training in two distinct configurations: one where no sweeps[2] are incorporated (denoted as ”0 sweeps”), and another where 10 sweeps are integrated into the training process. Furthermore, due to the limitations of the CARLA simulator, only 6 classes are considered for this study (car, truck, bus, pedestrian, motorcycle, bicycle) even though nuScenes contains 10 classes. The models are trained replicating the state-of-the-art configuration, we use the usual state-of-the-art 3D object detection

[2]Sweeps in the nuScenes benchmark refer to the intermediate frames without annotations, while samples refer to those key frames with annotations. Sweeps are used to accumulate point clouds and thus make the main cloud denser.

Table 2: Comparison of model evaluation (average precision) across 0 sweep models using three types of data: real data, synthetic data, and processed synthetic data.

Model	PointPillars (Lang et al., 2018)			SSN (Zhu et al., 2020)			CenterPoint (Yin et al., 2020)			TransFusion (Bai et al., 2022)		
Dataset	Real	Synth	P. synth	Real	Synth	P. synth	Real	Synth	P. synth	Real	Synth	P. synth
car	0.733	0.593	0.605	0.722	0.656	0.669	0.741	0.125	0.588	0.806	0.500	0.771
truck	0.306	0.584	0.582	0.372	0.633	0.630	0.399	0.007	0.428	0.426	0.500	0.831
bus	0.395	0.280	0.258	0.435	0.304	0.293	0.570	0.000	0.216	0.696	0.000	0.271
pedestrian	0.569	0.537	0.620	0.481	0.592	0.578	0.654	0.053	0.312	0.783	0.000	0.983
motorcycle	0.232	0.717	0.708	0.239	0.676	0.677	0.317	0.000	0.024	0.531	0.000	0.057
bicycle	0.034	0.432	0.428	0.027	0.528	0.502	0.180	0.053	0.119	0.285	0.079	0.351
Mean	0.378	0.524	0.533	0.379	0.565	0.558	0.477	0.040	0.281	0.588	0.180	0.544

Table 3: Comparison of model evaluation (average precision) across 10 sweep models using three types of data: real data, synthetic data, and processed synthetic data.

Model	PointPillars (Lang et al., 2018)			SSN (Zhu et al., 2020)			CenterPoint (Yin et al., 2020)			TransFusion (Bai et al., 2022)		
Dataset	Real	Synth	P. synth	Real	Synth	P. synth	Real	Synth	P. synth	Real	Synth	P. synth
car	0.803	0.662	0.673	0.816	0.615	0.636	0.821	0.331	0.563	0.871	0.000	0.770
truck	0.362	0.522	0.515	0.464	0.564	0.584	0.484	0.320	0.590	0.387	0.000	0.693
bus	0.426	0.194	0.180	0.561	0.316	0.309	0.627	0.037	0.247	0.734	0.000	0.246
pedestrian	0.742	0.873	0.841	0.665	0.769	0.732	0.748	0.455	0.949	0.866	0.000	0.974
motorcycle	0.332	0.540	0.533	0.427	0.640	0.626	0.433	0.000	0.054	0.620	0.000	0.000
bicycle	0.094	0.333	0.337	0.158	0.721	0.740	0.302	0.094	0.204	0.294	0.081	0.311
Mean	0.460	0.521	0.513	0.515	0.604	0.604	0.569	0.206	0.434	0.629	0.013	0.499

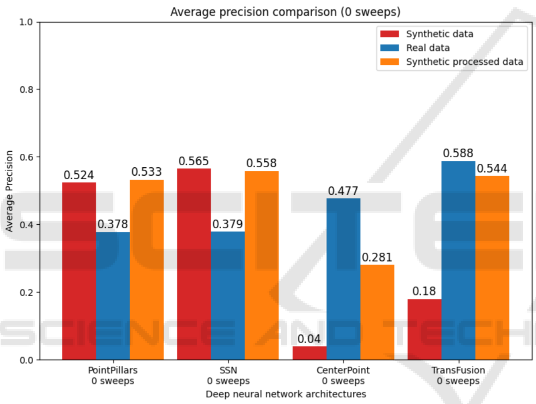

Figure 4: Average precision values for each data type: synthetic (red), real (blue) and synthetic processed (orange), without point cloud accumulation for each model.

evaluation metrics. We focus mainly on the Average Precision (AP). Model evaluation results can be found in Tables 2 and 3.

4.3 Analysis of the Results

The trained models are evaluated with real, synthetic, and processed synthetic data. Results can be seen in Tables 2 (0 sweeps) and 3 (10 sweeps). Notably, PointPillars and SSN exhibit superior performance on synthetic data, while CenterPoint and TransFusion have a better performance on real data. We think that the reason for the accuracy drop in CenterPoint and TransFusion may be due to the use of the heatmap obtained from the top view of the point cloud, a representation seemingly more prone to domain gap effects. Although the difference in accuracy is low for the first two models, the fact that it is so high for the other two models implies that some models cannot generalize properly and that there is indeed a significant domain gap between synthetic and real data.

After applying our proposed processing to synthetic point clouds, a comparison of model performance reveals that those achieving superior precision on synthetic data exhibit slightly lower precision across most classes, aligning them more closely with results obtained from real data. The precision of CenterPoint and TransFusion has increased drastically, reducing the precision gap between both kinds of data. This can be seen especially in Figure 4, where the mean AP of each model with different data types can be found. The loss of accuracy does not imply that our processing does not reduce the domain gap, on the contrary, the accuracy gap is reduced in most of the cases. The difference between the AP in the real and synthetic domain decreases since the processing applied to the clouds is effective, making synthetic and real clouds more similar.

While all models get a notable improvement when accumulating point clouds on real data, this is not always the case with synthetic data, the only model that behaves the same way is CenterPoint. PointPillars and SSN hardly notice the change between single-scan and accumulated point clouds. The accuracy of TransFusion is reduced when accumulating point clouds, increasing the domain gap. We conclude that in the case of synthetic point clouds, the accumulation can have different results depending on the model used, as it can be either detrimental or favorable.

The TransFusion model exemplifies a notable decrease in accuracy differences across domains. Without point cloud accumulation, our processing reduces the difference by 36.4% and with accumulation it is

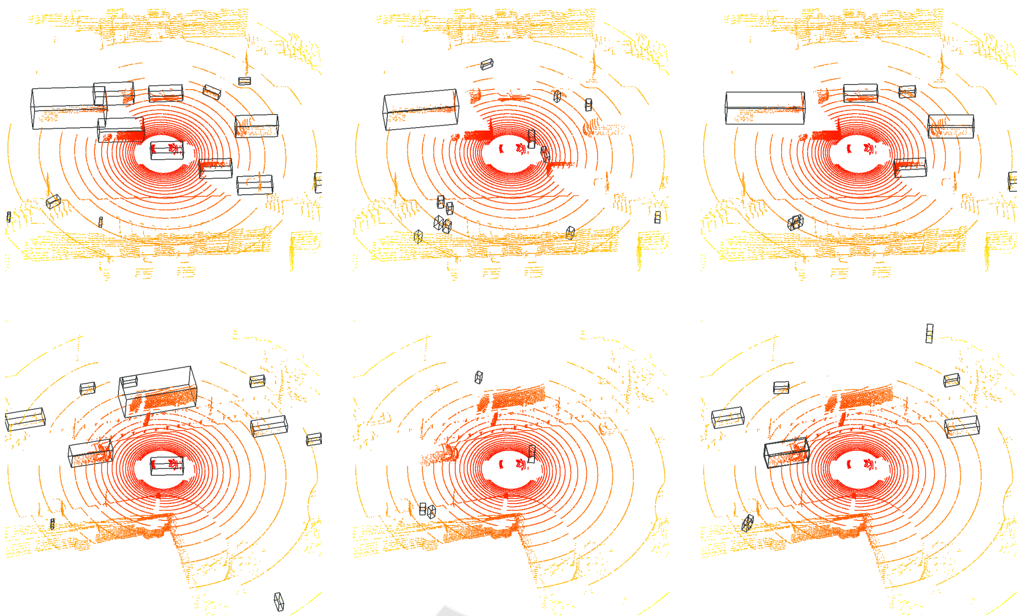

Figure 5: Comparison of detections in synthetic point clouds obtained with the CenterPoint model without point cloud accumulation. In each row, there is a frame of a different scene, the left column (a) contains the ground truth bounding boxes, the middle column (b) the detections without cloud processing and the right column (c) the detections with our cloud processing.

reduced by 48.6%. CenterPoint also reduces the accuracy difference with and without point cloud accumulation by 24.1% and 22.8%, respectively. With PointPillars, we have only achieved a reduction of 0.8% with point cloud accumulation and with SSN we have only archived a reduction of 0.7%. Our method effectively minimizes differences across all models, particularly in the car and pedestrian classes. The impact extends to other classes in most models. Based on these results, we conclude that our method reduces the difference between real and synthetic point clouds, even though high-performing models may not be as noticeably affected by our processing proposal.

In this paper, we also present qualitative results in Figure 5, illustrating the detection performance of our models on synthetic data and processed synthetic data in comparison to ground truth annotations. These images show the effectiveness of our proposed point cloud processing techniques. By comparing the model's detections on raw synthetic data to those on processed synthetic data, it is remarkable how our point cloud processing method significantly enhances the detection results, as a consequence of the domain gap reduction. These images highlight the improvements achieved by our approach, demonstrating the enhanced accuracy and robustness of the detection system when applied to processed point cloud data.

5 CONCLUSIONS AND FUTURE WORK

Recent advancements in driving simulators have transformed them into powerful tools for data generation. However, despite the easy access to data through these tools, there remains a question about their suitability for validating Automated Driving (AD) functions. Our study shows that there is a domain gap between real and simulated point cloud data.

Our study introduces an innovative synthetic data generation pipeline that creates standardized data adaptable to diverse sensor setups.

We study the evaluation of different LiDAR-based 3D object detection architectures concerning their performance with both real and synthetic data. Our findings reveal distinct responses across the models. PointPillars and SSN exhibited robustness to domain gap effects, while CenterPoint and TransFusion exhibited higher challenges in object detection when exposed to synthetic data. These observations emphasize the algorithmic influence on domain gap effects, underscoring the need for a prior analysis to assess how synthetic data may affect a specific algorithm.

Furthermore, we examined the domain gap between real and synthetic point clouds, identifying key distinctions. Our analysis revealed variations in reflected point numbers, structural differences in ring patterns, and fluctuations in point intensity values.

In response to these insights, we propose a domain gap reduction process for point clouds. This process proves its effectiveness through clear qualitative enhancements and a substantial reduction in accuracy gaps among various models when comparing their performance on real and synthetic data. Notably, this reduction is particularly prominent in the case of point cloud accumulation, where the CenterPoint and TransFusion models exhibit accuracy differences that are reduced by 22.8% and 48.6%, respectively. This approach can be applied to more reliably validate AD functions using synthetic point clouds.

This paper, has not investigated the potential of training models using solely synthetic data or in conjunction with real data, and it does not assess the contribution of these data in training, with or without processing. The exploration of this task is deferred to future work. Regarding the domain gap, although it has been possible to reduce it, it still exists and has not yet been completely reduced; continuing with the quantification and reduction of the domain gap in point clouds is still a pending and developing task.

ACKNOWLEDGEMENTS

This work has received funding from Basque Government under project AutoTrust of the program ELKARTEK-2023. This work is partially supported by the Ministerio de Ciencia, Innovación y Universidades, AEI, MCIN/AEI/10.13039/501100011033.

REFERENCES

Bai, X., Hu, Z., Zhu, X., Huang, Q., and et al. (2022). Transfusion: Robust lidar-camera fusion for 3d object detection with transformers.

Caesar, H., Bankiti, V., Lang, A. H., and et al. (2019). nuscenes: A multimodal dataset for autonomous driving.

Contributors, M. (2020). MMDetection3D: OpenMMLab next-generation platform for general 3D object detection. `https://github.com/open-mmlab/mmdetection3d`.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V. (2017). Carla: An open urban driving simulator.

Dworak, D., Ciepiela, F., Derbisz, J., Izzat, I., and et al. (2019). Performance of lidar object detection deep learning architectures based on artificially generated point cloud data from carla simulator. In *2019 24th International Conference on MMAR*, pages 600–605.

Huch, S., Scalerandi, L., Rivera, E., and Lienkamp, M. (2023). Quantifying the lidar sim-to-real domain shift: A detailed investigation using object detectors and analyzing point clouds at target-level. *IEEE Transactions on Intelligent Vehicles*, 8(4):2970–2982.

Inan, B. A., Rondao, D., and Aouf, N. (2023). Enhancing lidar point cloud segmentation with synthetic data. In *2023 31st MED*, pages 370–375.

Kalra, N. and Paddock, S. M. (2016). *Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?* RAND Corporation, Santa Monica, CA.

Kloukiniotis, A., Papandreou, A., Anagnostopoulos, C., and et al. (2022). Carlascenes: A synthetic dataset for odometry in autonomous driving. In *CVPR Workshops*, pages 4520–4528.

Lang, A. H., Vora, S., Caesar, H., and et al. (2018). Pointpillars: Fast encoders for object detection from point clouds.

Li, Y. and Ibanez-Guzman, J. (2020). Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Processing Magazine*, 37(4):50–61.

Li, Y., Ma, L., Zhong, Z., and et al. (2021a). Deep learning for lidar point clouds in autonomous driving: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8):3412–3432.

Li, Y., Ma, L., Zhong, Z., Liu, F., Chapman, M. A., and et al. (2021b). Deep learning for lidar point clouds in autonomous driving: A review. *IEEE TNNLS*, 32(8):3412–3432.

Liao, Y., Xie, J., and Geiger, A. (2021). Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d.

Martínez-Díaz, M. and Soriguera, F. (2018). Autonomous vehicles: theoretical and practical challenges. *Transportation Research Procedia*, 33:275–282. CIT2018.

Qiao, D. and Zulkernine, F. (2023). Adaptive feature fusion for cooperative perception using lidar point clouds. In *Proceedings of the IEEE/CVF WACV*, pages 1186–1195.

Rong, G., Shin, B. H., Tabatabaee, H., Lu, Q., Lemke, S., and et al. (2020). LGSVL simulator: A high fidelity simulator for autonomous driving. *CoRR*, abs/2005.03778.

Sekkat, A. R., Dupuis, Y., Kumar, V. R., and et al. (2022). SynWoodScape: Synthetic surround-view fisheye camera dataset for autonomous driving. *IEEE Robotics and Automation Letters*, 7(3):8502–8509.

Sun, P., Kretzschmar, H., and Dotiwalla, Xerxes, e. a. (2019). Scalability in perception for autonomous driving: Waymo open dataset.

Wang, F., Zhuang, Y., Gu, H., and Hu, H. (2019). Automatic generation of synthetic lidar point clouds for 3-d data analysis. *IEEE TIM*, 68(7):2671–2673.

Xu, R., Xiang, H., Xia, X., Han, X., and et al. (2022). Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication.

Yin, T., Zhou, X., and Krähenbühl, P. (2020). Center-based 3d object detection and tracking. *CVPR*.

Zhu, X., Ma, Y., Wang, T., Xu, Y., and et al. (2020). Ssn: Shape signature networks for multi-class object detection from point clouds.