Elastic Deformable Augmentation for Autonomous Driving

Jungwook Shin[1*], Jaeill Kim[1*], Wonjong Rhee[1,2]
[1] Department of Intelligence and Information, Seoul National University
[2] Interdisciplinary Program in Artificial Intelligence (IPAI), Seoul National University
Seoul, Republic of Korea
jungwook.shin@snu.ac.kr, jaeill0704@snu.ac.kr, wrhee@snu.ac.kr

***Abstract*—Unlocking the information concealed in 3D point clouds by LiDAR is a key mission for autonomous driving. However, this task is challenging due to the sparsity, continuity, and unordered nature of point clouds. In addition, creating annotated data is time-consuming and expensive. This emphasizes the crucial necessity of data augmentation. Conventional object data augmentation methods such as rotation, scaling, and translation are not fully effective for 3D data. Therefore, we propose a novel augmentation method, Elastic Deformable Augmentation (EDA), which enhances data diversity for better model robustness. EDA applies deformation methods from 3D graphics to the objects, diversifying their shapes without violating occlusion or intensity properties. We demonstrate EDA on the KITTI dataset, where it improves object detection performance, particularly increasing the mean AP scores by 1.55% and 1.00%, respectively. Consequently, our research provides compelling evidence that EDA is a promising approach for augmenting 3D object detection tasks in autonomous driving.**

***Index Terms*—Deformation, Data Augmentation, 3D Object Detection, Autonomous Driving**

I. Introduction

Object detection, using LiDAR-generated 3D point clouds in autonomous driving, is crucial for tasks such as path planning. However, the nature of these point clouds (unordered, sparse, and continuous) presents unique challenges [9]. Additionally, deep neural networks for 3D object detection require large volumes of annotated data for accuracy. Nevertheless, the annotation of 3D data is both time-intensive and costly. Moreover, with the advancement of LiDAR technologies, there's often a need for periodic re-labeling. Therefore, data augmentation becomes essential for object detection in autonomous driving.

In 3D autonomous driving object detection, foreground diversity has a significant impact, making both the number and balanced distribution of object classes crucial elements [2]. However, the KITTI dataset, one of the most popular datasets in the autonomous driving domain, is abundant in vehicle data but critically lacks pedestrian and cyclist data. Unlike 2D images, point clouds are inherently sparse and become even more so as the distance increases. For pedestrians, their representation in point clouds is often minimal. Cyclists, especially at distances exceeding 10 meters, are typically represented by fewer than 50 points. These challenges contribute to the complexities models face during their learning processes.

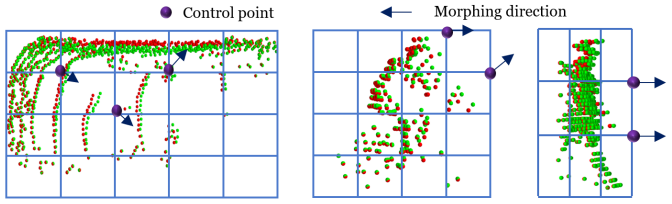
In the task of 3D object detection, various factors, including self and external occlusion, signal miss, and scale with distance, lead to a diverse range of object shapes. These transformed shapes occur in a wide range of locations. Because 3D data differs significantly from 2D data, augmentation through simple rotation, translation, and scaling has its limitations, and there are disputes over the effectiveness of local augmentation, which is a concept derived from 2D [5].

The previous studies [1] have addressed this issue by creating data with CAD, generating intensity using a depth map, and applying a copy-and-paste method to the background to augment objects for learning. However, restoring intensity, which is a unique feature of LiDAR, is challenging. In DR.CPO [2], objects were constructed as whole bodies, and occlusions were generated according to position to ensure diversity and augment objects. But as mentioned in the paper, additional work is necessary when motorcycles and motorcyclists are labeled similarly to vehicles, and the focus is on securing diversity generated by occlusion. We aim to augment objects through object deformation, which existing augmentations do not adequately address, such as vehicle changes and human movements. While occurrences of damaged vehicles are rare in real-world scenarios, their detection can significantly improve performance [8]. Additionally, pedestrian movements exhibit a wide range of variations, and the movements of cyclists and motorcyclists are equally diverse. Using conventional data augmentation methods to enhance these aspects is considerably limited.

Therefore, we propose Elastic Deformable Augmentation (EDA), an augmentation method suitable for 3D object structures. Our proposal is based on deformation methods used in 3D graphics and diversifies the shape of objects without violating occlusion and intensity. EDA extracts the ground truth from a scene based on a bounding box, designates the object's centroid as the control point, and reproduces various movements of the object through deformation based on 3D graphic techniques. The augmented object data is then repositioned in each scene for learning. Our approach is described in Figure 1 and Figure 2. The principal contributions of our work are:

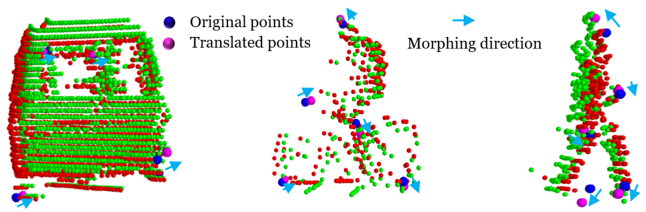
- The introduction of a new pipeline, to the best of our

[*]Equal contribution.



(a) Car (b) Cyclist (c) Pedestrian

Fig. 1: Illustration of FFD with ground truth objects: Control points are set at the lattice of a regularly spaced grid, and morphing is carried out based on arbitrarily chosen directions and weights. The points before deformation are shown in red, and the points after deformation are displayed in green.



(a) Car (b) Cyclist (c) Pedestrian

Fig. 2: Illustration of RBFD with ground truth objects: Control points are set using the Farthest Point Sampling (FPS) method, and deformation is carried out in an arbitrary direction by the set weight. The blue circles represent the original points from FPS, and the deformation through RBF is carried out based on the arbitrarily moved points, indicated by pink circles.

knowledge the first in the autonomous driving field, that utilizes Elastic Deformable Augmentation for 3D object detection.

- The proposal of a method that integrates 3D graphics techniques with ground-truths, and empirical evidence demonstrating its positive effects on 3D object detection performance.

II. Methods

We present a unified pipeline for data augmentation in 3D Object Detection systems, designed specifically for autonomous vehicles. Our approach employs morphing methods to introduce variations in detected 3D objects. By adding diversity to the dataset, we aim to improve model performance across diverse real-world scenarios. The pipeline comprises two core methodologies: Free-Form Deformation [3](FFD) and Radial Basis Function based Deformation [4](RBFD). It proceeds through four primary stages: Control Point Extraction, Morphing, Bounding Box Adjustment and Inserting the Transformed Object.

A. Free Form Deformation

Freeform Deformation (FFD) [3] deforms an object based on a lattice surrounding object as shown Figure 1. When the points of the lattice are moved, the object deforms according to the transformation of those lattice points. Each point within the lattice gets a new position as a result of the deformation. The transformation of a point inside the lattice can be defined by trilinear interpolation using the surrounding lattice points. The mathematical formulation of FFD is:

$$\boldsymbol{P}(x,y,z) = \sum_{i=0}^{l}\sum_{j=0}^{m}\sum_{k=0}^{n} B_i(x)B_j(y)B_k(z)\boldsymbol{P}_{ijk} \qquad (1)$$

In this equation, $\boldsymbol{P}(x,y,z)$ represents the position of a point inside the lattice. The functions B_i, B_j, and B_k are the Bernstein basis functions for the x, y, and z directions, respectively. The point $\boldsymbol{P}_{ijk}$ stands as a specific control point on the lattice, while the summations iterate over the lattice dimensions denoted by l, m, and n.

B. Radial Basis Functions Deformation

The Radial Basis Functions (RBF) [7] interpolation technique, in contrast to FFD, allows localized control points refinements without the constraints of a regular lattice. The RBF shape parametrization technique is defined as:

$$\boldsymbol{P}' = \boldsymbol{P} + \sum_{i=1}^{N} \lambda_i \phi(\|\boldsymbol{P} - \boldsymbol{P}_i\|) \qquad (2)$$

In the deformation process, the symbol $\boldsymbol{P}$ denotes the original position of a point, while $\boldsymbol{P}'$ represents the deformed position. The transformation is influenced by control points, specifically $\boldsymbol{P}_i$, which has an associated weight λ_i. The Radial Basis Function [7], represented by ϕ, determines the degree of this influence. We use Gaussian spline as this function. The distance between $\boldsymbol{P}$ and the control point $\boldsymbol{P}_i$ measures the influence based on their proximity. Figure 2 illustrates our proposal of RBFD.

C. Pipeline for Elastic Deformation in Autonomous Driving

Control Point Extraction: This stage involves extracting control points using different methods for FFD and RBFD. For FFD, a cubic lattice encapsulating the object is created with dimensions(l, m, n) varying based on the object type: Cars/Pedestrian/Cyclist use a (5,4,3)/(4,3,3)/(3,3,3) lattice. We designate the intersections of the lattice as control points. In contrast, for RBFD, we adopt Farthest Point Sampling (FPS) [6] algorithm with a default of n control points.

Morphing: After the control points have been selected, a random deformation based on the control points is applied to objects which are extracted from a various of scenes, using either the FFD or RBFD method. For FFD, the number of selected control points and the axis (l, m, n) along which deformation occurs are chosen randomly. The deformation magnitude is determined by a random value within the defined weight range, typically between -0.5 and 0.5. Figure 1 illustrates our FFD method. For RBFD, a random jitter is applied to the control points using a random jitter function as illustrated by Figure 2. This jittered points are then used to deform the original object using RBF interpolation.

Bounding Box Adjustment: Following the morphing phase, we adjust the bounding box of the object to account for the changes in its shape. We recalibrate the box by calculating the minimum and maximum points in the morphed object, and subsequently recalculating the box's range and center shift. This recalibration ensures that the bounding box accurately encloses the morphed object, facilitating the 3D detection model to correctly interpret the newly augmented data.

Inserting the Transformed Object: As a final step, we insert the transformed object back into the scene for training. In order to maintain the shape of self-occlusion according to the LiDAR beam direction, the transformed object keeps its original position. If there is an existing object at the intended insertion location, to avoid collisions, the insertion is omitted.

This stochastic pipeline effectively integrates FFD and RBFD to augment ground truth objects for 3D object detection models in autonomous driving, enhancing diversity and resistance to overfitting. It holds potential to improve model performance across various scenarios.

III. Experiments

We evaluated our proposed Free-Form Deformation (FFD) and Radius Basis Function Deformation (RBFD) augmentation methods using object detection metrics for Car, Pedestrian, and Cyclist. We utilized PV-RCNN++ as our baseline model for this experiment, and compared its performance to that of the PV-RCNN++ model trained with our FFD and RBFD augmentation on the KITTI validation dataset. We measured mean average precision (mAP) with IOU thresholds of 0.7 for cars and 0.5 for pedestrians and cyclists, following the standards typically applied in other 3D object detection performance evaluations. The AP performances are calculated using 40 recall positions (R40). The 'Diff.' column indicates the performance enhancement over the baseline.

A. The result of Freeform Deformation augmentation

TABLE I: The impact on performance by Freeform Deformation augmentation.

Method	Car	Pedestrian	Cyclist	Mean	Diff.
PV-RCNN++ [2]	86.19	60.25	76.15	74.20	
FFD	**86.63**	**63.26**	**77.37**	**75.75**	+1.55

As shown in Table I, the FFD-augmented model outperforms the baseline across all categories. Particularly notable improvements were observed in the Pedestrian and Cyclist categories. The overall Mean score also saw a significant enhancement, showing an increase by 1.55% compared to the baseline. These results convincingly demonstrate the effectiveness of our FFD augmentation approach in enhancing the performance of 3D object detection model.

B. The result of RBF Deformation augmentation

Table II illustrates the performance of our RBFD augmentation approach, displaying a superior performance over the baseline in almost all categories except for Car, with particularly noteworthy improvements in the Pedestrian and Cyclist categories. The overall Mean score also showed a significant enhancement, increasing by 1.00% compared to the baseline. This performance outcome substantiates the potential of our RBFD augmentation method to significantly boost the efficacy of 3D object detection model.

TABLE II: The impact on performance by Radius Basis Function Deformation augmentation.

Method	Car	Pedestrian	Cyclist	Mean	Diff.
PV-RCNN++ [2]	**86.19**	60.25	76.15	74.20	
RBFD	85.99	**61.98**	**77.63**	**75.20**	+1.00

IV. Conclusion

In this research, we introduce Elastic Deformable Augmentation (EDA) for enhanced 3D object detection. EDA utilizes 3D graphic transformations to diversify object shapes, increasing data variety and model resilience. The approach yields a significant performance boost on the KITTI validation dataset, with gains of 1.55% and 1.00% in the overall performance metrics. Future work may explore additional transformations to further optimize 3D detection models.

Acknowledgements

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1A2C2007139) and in part by IITP grant funded by the Korea government (MSIT) [NO.2021-0-01343, Artificial Intelligence Graduate School Program (Seoul National University)].

References

[1] Fang, J., Zuo, X., Zhou, D., Jin, S., Wang, S., & Zhang, L. (2021). Lidar-aug: A general rendering-based augmentation framework for 3d object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4710-4720).

[2] Shin, J., Kim, J., Lee, K., Cho, H., & Rhee, W. (2023, June). Diversified and Realistic 3D Augmentation via Iterative Construction, Random Placement, and HPR Occlusion. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 2, pp. 2282-2291).

[3] Sederberg, T. W., & Parry, S. R. (1986, August). Free-form deformation of solid geometric models. In Proceedings of the 13th annual conference on Computer graphics and interactive techniques (pp. 151-160).

[4] De Boer, A., Van der Schoot, M. S., & Bijl, H. (2007). Mesh deformation based on radial basis function interpolation. Computers & structures, 85(11-14), 784-795.

[5] Reuse, M., Simon, M., & Sick, B. (2021). About the ambiguity of data augmentation for 3d object detection in autonomous driving. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 979-987).

[6] Moenning, Carsten; Dodgson, Neil A. Fast marching farthest point sampling. University of Cambridge, Computer Laboratory, 2003.

[7] Buhmann, M. D. (2000). Radial basis functions. Acta numerica, 9, 1-38.

[8] Lehner, Alexander, et al. "3D-VField: Adversarial augmentation of point clouds for domain generalization in 3D object detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[9] Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 652-660).