2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

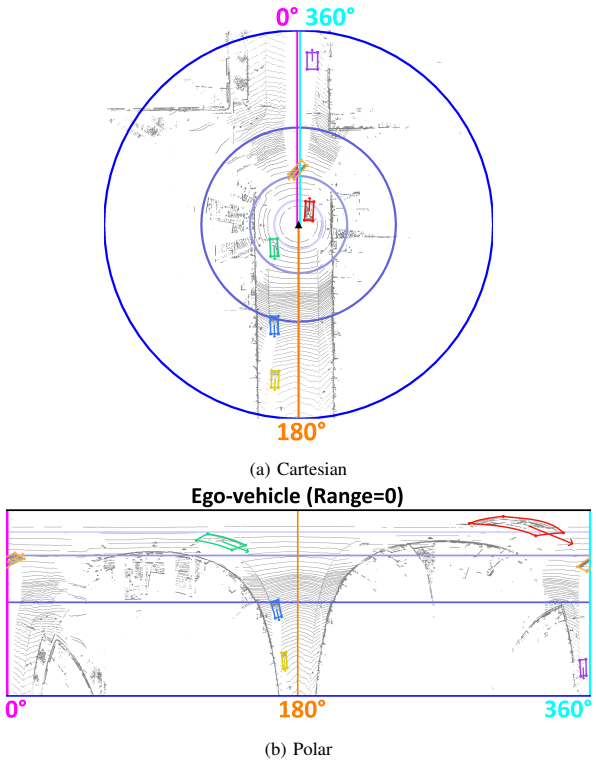
Fully Polar Coordinate Object Detection: A Constraint-Based Polar Bounding Box Approach for LiDAR and Scanning Radar

Shu-Heng Lin, Chieh-Chih Wang, and Wen-Chieh Lin

***Abstract*— Polar coordinates are widely used in segmentation tasks for range sensors such as LiDAR and radar, owing to their ability to naturally align with point cloud sparsity and distribution. However, their use in detection is limited by feature distortion. Existing polar-based detection works focused on undistorting features from the polar coordinates back to canonical Cartesian representations, but their results remain unsuccessful. In this work, we propose fully polar coordinate object detection, performing training and evaluation entirely in polar coordinates without relying on Cartesian metrics. To achieve this, we design a constraint-based polar bounding box representation, that enables the direct conversion of Cartesian bounding boxes via a constrained minimum bounding rectangle (MBR). Using the state-of-the-art polar-based detector as our baseline, we conduct experiments on the Boreas dataset. The results demonstrate that our approach improves the LiDAR detection AP_{30} metric by 2.88%, and yields a 2.17% gain over Cartesian-based detection methods. On more challenging scanning radar detection experiments, our method achieves an 13.11% improvement in AP_{30} compared to Cartesian-based detection methods. These findings validate the feasibility of fully polar coordinate object detection and demonstrate its robustness and generalizability across multiple range sensor modalities.**

I. Introduction

In autonomous vehicles, detection and segmentation are the core components of the perception system, allowing the vehicle to estimate and interact with its surroundings. Among various sensing modalities, range sensors such as LiDAR and radar play a crucial role in autonomous vehicle perception. These sensors capture measurements in range-based formats (e.g., spherical or cylindrical coordinates), which can be naturally projected onto 2-D polar coordinates that are parameterized by range and azimuth. Figure 1 illustrates how LiDAR point clouds can be represented in both polar and Cartesian coordinates, which facilitates the visualization of the correspondence between the two coordinates. However, most existing methods LiDAR-based and radar-based detection and segmentation first convert raw sensor measurements into Cartesian coordinates, disregarding the geometric characteristics inherent to the sensing modality. In addition, while Cartesian-based evaluation metrics are intuitive, they implicitly constrain the entire pipeline to a specific coordinates.



(a) Cartesian

(b) Polar

Fig. 1: Visualization of LiDAR data in Cartesian (a) and polar (b) coordinates, with different objects distinguished by color. In the polar coordinates map (b), the horizontal axis represents azimuth and the vertical axis represents range. The center point (range = 0) is marked in black, while varying shades of blue indicate different range values. These range intervals form concentric circles in the Cartesian coordinates, but appear as horizontal lines with constant range values in the polar coordinates. Azimuth is color-coded, with purple representing 0°, orange 180°, and cyan 360°.

Previous **segmentation studies** have used **polar coordinates** to address the non-uniform sparsity and spatial distribution of LiDAR point clouds [1], [2]. As shown in Figure 2, the polar voxel grids naturally align with the density characteristics of point clouds, providing finer resolution in near-field regions where points are densely distributed and coarser resolution in far-field regions that are sparsely distributed. This adaptive allocation helps preserve fine-grained geometric details close to the sensor while reducing computational overhead in distant, less informative regions. In contrast, **Cartesian voxel grids** impose uniform discretization across the entire space, which often compresses dense regions and allocates unnecessary computation to distant regions. The effectiveness of polar coordinates in segmentation tasks can

Shu-Heng Lin is with the Graduate Degree Program of Artificial Intelligence, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. shunycu.ee10@nycu.edu.tw

Chieh-Chih Wang is with the College of Electrical and Computer Engineering, National Yang Ming Chiao Tung University, and with the Mechanical and Mechatronics Systems Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan. bobwang@ieee.org

Wen-Chieh Lin is with the Institute of Multimedia Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. wclin@cs.nctu.edu.tw

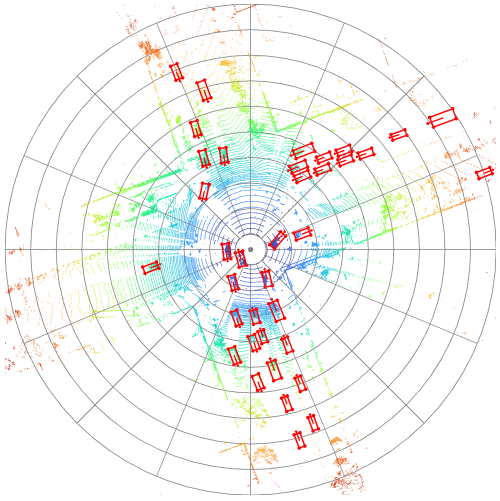


Fig. 2: An example of a polar-based voxel grid layout for LiDAR voxelization illustrates how denser grids are allocated to near-field regions, while sparser grids are used in the far-field regions. Furthermore, the LiDAR input exhibits a ring-shaped pattern that resembles the structure of the polar voxel grid.

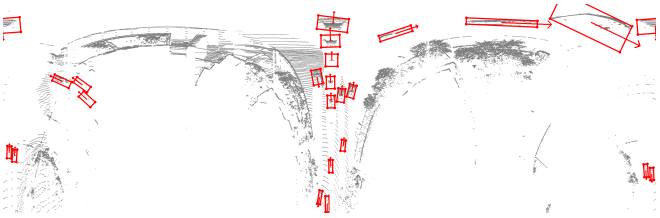


Fig. 3: Illustration of the proposed constraint-based polar bounding box.

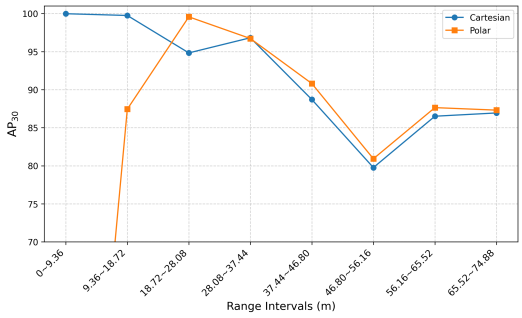


Fig. 4: AP_{30} performance across range intervals for detections in polar and Cartesian coordinates. The experimental range is 74.88 m, uniformly divided into eight intervals of 9.36 m each. Each dot indicates the AP_{30} computed from objects within the corresponding range interval. Polar-based detection underperforms in the near-field but consistently surpasses Cartesian-based detection as the range increases.

be attributed to its reliance on point-wise evaluations rather than object-level structures, unlike detection, which evaluates objects as a whole. This property allows segmentation to directly benefit from the higher resolution of polar grids in densely distributed regions, thereby enhancing accuracy without incurring additional computational cost in distant regions.

Although polar coordinates have proven to be effective in segmentation tasks, detection tasks are still predominantly implemented in Cartesian coordinates and thus still suffer from the aforementioned challenges of non-uniform sparsity and spatial distribution of LiDAR point clouds. While several works have explored polar-based detection, they have not consistently outperformed their Cartesian counterparts under the same network architectures [3]–[5]. This performance gap has been attributed to **feature distortion**, where objects at different locations exhibit substantially different features in polar coordinates. As illustrated in Figure 1(b), such distortions often appear in curvilinear forms, which complicate alignment with their Cartesian representations. Therefore, prior works [3]–[5] have explored various undistortion approaches, aiming to restore distorted polar features into canonical object representations.

Our main contribution is the design of a **constraint-based polar bounding box**, which projects Cartesian bounding boxes into polar coordinates and applies a minimum bounding rectangle (MBR) with two additional constraints. The box constraint preserves the Cartesian box dimensions to prevent overly small boxes caused by sparse projected points, while the orientation constraint aligns the box with the object's direction in polar coordinates. With this design, polar bounding boxes can be automatically generated for training and evaluation, are illustrated in Figure 3. We explore three different variants of polar bounding boxes, each defined by different combinations of the proposed constraints, and analyze their impacts on detection tasks. To ensure a fair comparison, we align the the number of pixels between Cartesian and polar voxel maps to represent physically equivalent regions, and compute IoU for both training and evaluation using the proposed polar box formulation.

Our experiments reveal that while polar representation inherently complicates annotation of near-field objects, it improves detection for medium- and far-field objects, as illustrated by the AP-range curves in Figure 4. This occurs because Cartesian grids uniformly partition space without considering the distance-dependent decrease in point density, resulting in fewer points per cell at longer ranges and thus degraded performance. In contrast, polar representation alleviates this effect, resulting in improved detection at medium- and far-field. However, near-field objects are heavily distorted and occupy large pixel regions, making them difficult to represent accurately with a single rectangular bounding box. This mismatch leads to inaccurate IoU estimation and directly degrades AP. Accordingly, our problem scope is defined as enhancing detection performance while excluding the near-field objects, which we define as those within 11.5 m based on our observations of distortion. Beyond this range, polar-based detection consistently outperforms Cartesian-based detection, highlighting the strengths of a fully polar-based approach for safety-critical long-range perception and motion planning at higher speeds. Although this study primarily evaluates the "Car" class, the proposed constraint-based formulation is class-agnostic and directly applicable to other object classes.

We build upon the state-of-the-art polar-based detector PARTNER [3], replacing its Cartesian ground truth with our

proposed constraint-based polar bounding boxes and introducing several architectural modifications, including polar-based data augmentation and an azimuthal shift strategy to address the continuity issue along the azimuth axis.

Compared to the original PARTNER, our method achieves a 2.88% improvement in AP_{30} on LiDAR-based detection. Furthermore, it outperforms a fully Cartesian configuration by 2.17%, suggesting that our polar-based approach offers advantages over Cartesian configurations. On the more challenging range-sensing modality of scanning radar, which is more difficult than LiDAR and where the baseline PARTNER model fails to produce usable results, our method achieves an 13.11% increase in AP_{30} over Cartesian-based approaches. These results demonstrate the robustness and generalizability of our approach across different sensor modalities, underscoring the potential of fully polar coordinate detection pipelines.

II. Related Work

Polar-based LiDAR Segmentation. A considerable number of LiDAR segmentation studies [1], [2], [6], [7] have adopted polar bird's-eye view (BEV) and cylindrical space partitioning as voxelization strategies. Among them, the pioneering work PolarNet [1] is a representative example. It identified the imbalanced spatial distribution of LiDAR points in Cartesian coordinates, and proposed the use of polar coordinates to alleviate this issue. PolarNet also introduced ring convolution, which preserves azimuthal continuity and enables effective feature learning across azimuthal boundaries. Through comprehensive experiments on multiple segmentation backbones, PolarNet demonstrated that the polar representation achieves comparable or better performance than Cartesian representations.

Polar-based Object Detection in LiDAR. While polar-based representations have been extensively studied in LiDAR segmentation, their adoption in LiDAR-based object detection remains limited. Detection pipelines are still predominantly implemented in Cartesian coordinates, where LiDAR detection continue to suffer from imbalanced point distribution. This gap highlights the potential benefits of polar representations for detection, and motivates subsequent works.

One of the earliest works to address this gap is by Rapoport-Lavie and Raviv, who analyze the ring-shaped patterns produced by rotating LiDAR scanners and propose an end-to-end perception architecture in cylindrical coordinates [4]. Their work also identifies several unique challenges: voxel sizes vary with range, orientations are viewpoint-dependent, and azimuthal continuity issues, all of which further complicate effective representation learning.

PolarStream [5] extends this line of research by exploring polar coordinates in the context of streaming-based LiDAR perception. It formally identifies **feature distortion**, caused by unfolding the polar grid into a rectangular array for convolution, as a major limitation for object detection in polar coordinates. To mitigate this, PolarStream introduces a feature undistortion module along with range-stratified convolution and normalization layers to compensate for varying feature scales across different ranges.

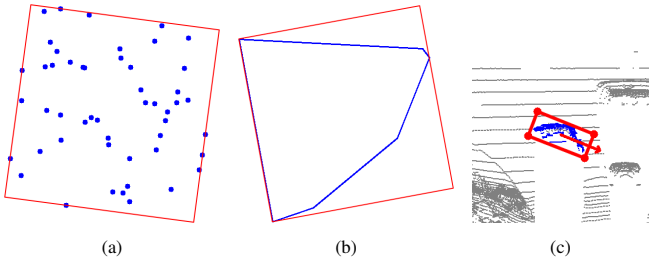


Fig. 5: Examples of minimum bounding rectangles (MBRs) applied to (a) a set of discrete points, (b) a closed arbitrary shape, and (c) a LiDAR car object in polar coordinates. In (c), the MBR fails to reflect the object's actual orientation, indicating that unconstrained MBRs are unsuitable as bounding box definitions.

Building on these insights, PARTNER [3] proposes a transformer-based 3D object detector tailored for polar coordinates. As one of the current state-of-the-art methods in this domain, it effectively alleviates the feature distortion in polar coordinates and achieves strong performance in streaming-based LiDAR perception.

Owing to its effectiveness, we adopt PARTNER as our baseline. Unlike prior works that primarily tackle feature distortion in polar coordinates, our approach takes a different perspective: we directly annotate bounding boxes in polar coordinates. Our approach is not limited to streaming-based settings and aims to extend the applicability of polar coordinates to general object detection tasks.

III. Method

A. Constraint-based Polar Bounding Box Definitions

Objects at different locations exhibit features with varying shapes and areas when represented in polar coordinates. In this representation, objects often appear in curvilinear forms, as illustrated in Figure 1(b). To enable fully polar coordinate object detection, it is essential to establish a principled approach for bounding box annotation directly within the polar coordinates.

A straightforward solution is to adopt minimum bounding rectangles (MBR), which refers to the smallest possible rectangle that completely encloses a set of points or a given shape. Figure 5(a)(b) present examples of MBRs applied to both cases. However, this approach disregards the relationship between the object's geometric features and its orientation. In addition, expanding curvilinear shapes into MBRs can produce overly large boxes with irrelevant features, leading to inaccurate regression, as shown in Figure 5(c).

To address these limitations, we incorporate two additional constraints into the MBR: a box constraint and an orientation constraint. The **box constraint** computes the MBR by projecting each Cartesian bounding box into polar coordinates and using the resulting curvilinear shape (Figure 6(a)), rather than directly from the enclosed points. The **orientation constraint** enforces the ground-truth orientation by projecting the original Cartesian bounding box orientation onto the

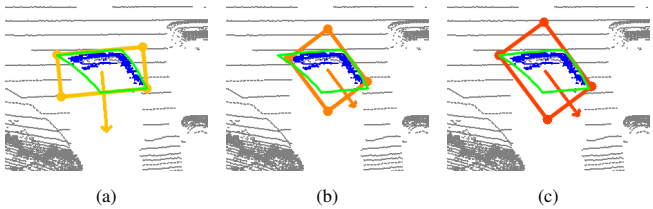


Fig. 6: Example of bounding box definitions under different constraints in polar coordinates: (a) box constraint, (b) orientation constraint, and (c) both constraints. Box definitions under the three constraints are shown in varying shades of orange. The green curvilinear shape represents the projection of a Cartesian bounding box into polar coordinates, while the blue points denote object points, defined as those lying within the original Cartesian bounding box.

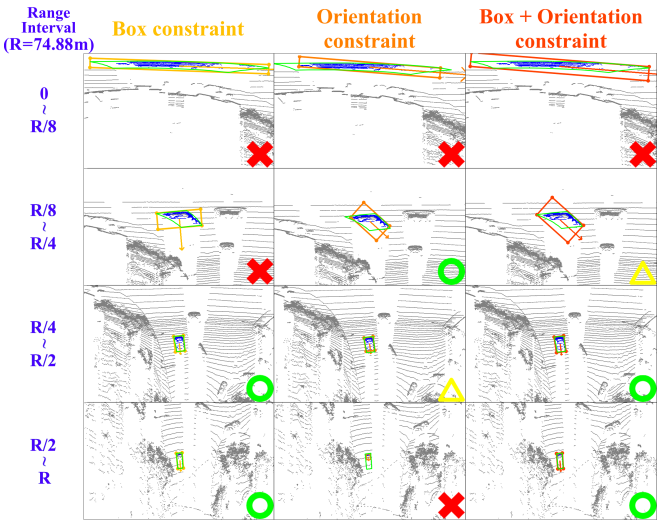


Fig. 7: An example demonstrating how polar bounding boxes defined by different combinations of constraints, vary with increasing range. The green curvilinear shape represents the projection of a Cartesian bounding box into polar coordinates, while the blue points denote object points, defined as those lying within the original Cartesian bounding box. From left to right, each column shows polar bounding boxes defined by different constraint combinations, with varying color shades indicating the respective constraints. From top to bottom, each row corresponds to increasing range from the sensor, with the maximum range denoted as R. Boxes that are unacceptable, marginally acceptable, or optimal are marked with a red cross, yellow triangle, and green circle, respectively.

polar coordinates (Figure 6(b)), instead of relying on the orientation induced by the MBR itself. We analyze three combinations of these constraints, resulting in three distinct bounding box definitions. For each definition, we examine its properties across objects at different ranges, as illustrated in Figure 7, where R denotes the maximum range. The range intervals are set at $R/8$, $R/4$, $R/2$, and R.

For near-field objects (within $[0, R/8]$, first row of Figure 7), the point density is high and the objects occupy a large number of pixels, making it difficult to circumscribe their shapes using a single bounding box. As a result, none of the definitions provides a satisfactory approximation.

For near-to-mid-field objects (within $[R/8, R/4]$, second row of Figure 7), the network tends to rely more on features at the object level to infer orientation. Under these conditions, incorporating the orientation constraint leads to more

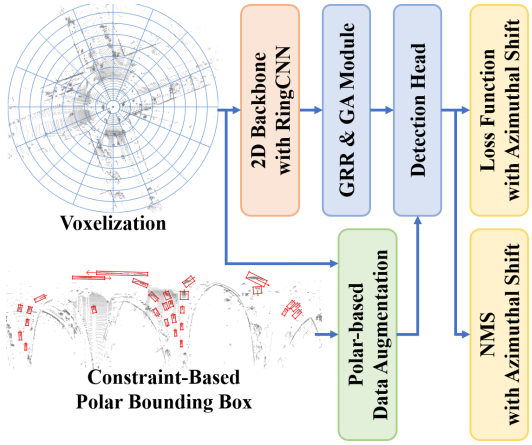


Fig. 8: Overview of the proposed polar-based detection pipeline. Our main modifications to PARTNER [3] include replacing the input Cartesian ground truth with constraint-based polar bounding boxes, adapting convolution padding and masking, and incorporating an azimuthal shift, which is applied to the loss during the training phase and to NMS during inference.

appropriate bounding box definition.

For mid-field objects (within $[R/4, R/2]$, third row of Figure 7), the distortion is minimal, the object areas in Cartesian and polar coordinates are similar, and their shapes closely resemble those of Cartesian bounding boxes. Therefore, applying either the box or orientation constraint can produce reasonable bounding boxes. However, when partially occluded, these objects may suffer from sparse point observations, leading to challenges similar to those discussed later for far-field objects.

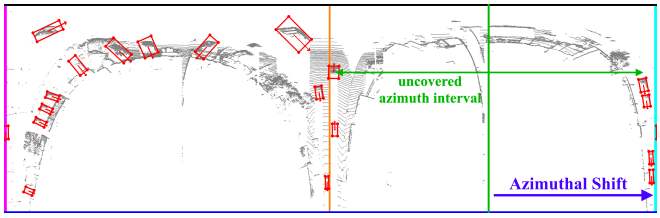
For far-field objects (within $[R/2, R]$, fourth row of Figure 7), the number of points is typically low. Without the box constraint, deriving an MBR directly from the object points may produce an overly small bounding box, which can significantly affect the IoU calculation during training and evaluation.

Based on the analysis, we adopt two bounding box definitions for further use. The configuration with both box and orientation constraints (third column of Figure 7) is generally more effective for mid- and far-field objects, while the one with only the orientation constraint (second column of Figure 7) better captures near-to-mid-field objects.

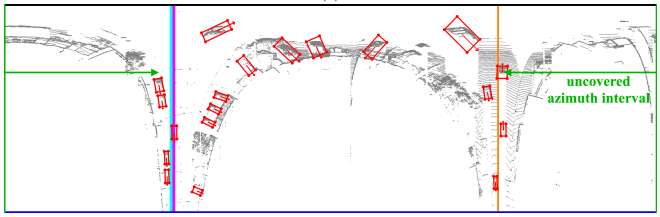
B. Proposed Architecture

We build our baseline on PARTNER [3], a transformer-based object detector that achieves state-of-the-art performance in polar coordinates. PARTNER mitigates the feature distortion in polar representations through two key components: the Global Representation Re-alignment (GRR) module and the Geometry-Aware Adaptive (GA) module.

We extend PARTNER with three modifications, which together form our complete detection pipeline illustrated in Figure 8. First, we replace the Cartesian ground truth with our proposed constraint-based polar bounding boxes and introduce a **polar-based data augmentation** strategy tailored to both ground truth and polar voxels (Section III-C). Second, we replace convolution and attention mechanisms by



(a)



(b)

Fig. 9: Illustration of the proposed azimuthal shift for addressing the continuity issue along the azimuth axis in polar coordinates. An azimuth interval not covered by any ground truth bounding box is first identified, with its center indicated by green line (a). All bounding boxes are then cyclically shifted based on this reference azimuth (b), ensuring that no object crosses the azimuthal boundary during loss computation, NMS, or evaluation. The input image is shifted only for visualization; in practice, feature continuity is preserved by ring convolution, while the azimuthal shift is applied solely to bounding boxes during training and evaluation.

incorporating the **ring convolution** from PolarNet [1], which applies circular padding along the azimuth axis, and by removing unnecessary masking in the Swin Transformer [8] to better preserve azimuthal continuity. Third, we apply an **azimuthal shift** that cyclically aligns ground truth and predictions during loss computation, non-maximum suppression (NMS) [9] and evaluation, thereby preventing objects from being split across azimuthal boundaries (Figure 9).

C. Data Augmentation in Polar Coordinates

Existing polar-based detectors typically voxelize LiDAR inputs in polar coordinates but still apply data augmentation in Cartesian coordinates. This is because their ground truth labels remain defined in Cartesian coordinates. In contrast, our method conducts fully polar coordinate object detection, with both inputs and labels represented in polar coordinates. Traditional Cartesian augmentations such as random horizontal and vertical flipping, rotation, and translation are not directly applicable due to geometric discrepancies, motivating the design of augmentation strategies tailored to polar coordinates.

While convolutional operations in Cartesian coordinates exhibit translational equivariance, those in polar coordinates instead exhibit rotational equivariance along the azimuthal axis. This distinction shows how common geometric transformations map differently across the two representations. For example, a horizontal flip in Cartesian space (i.e., across the x-axis) corresponds to a horizontal flip along the azimuth axis in polar coordinate. A vertical flip in Cartesian coordinate (i.e., across the y-axis) maps to a horizontal flip along the azimuth axis followed by a 180° azimuthal shift. Applying both horizontal and vertical flips in Cartesian space is thus equivalent to a 180° azimuthal shift in polar space.

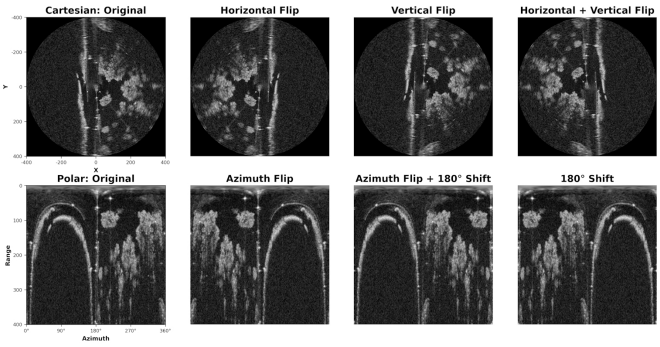


Fig. 10: An example set of various data augmentation applied to a sample scanning radar image. The top row shows the image in Cartesian coordinates, while the bottom row are the corresponding representations in polar coordinates. Each pair in the same column represents equivalent augmentations across the two coordinates.

As illustrated in Figure 10, these geometric correspondences highlight the need to adapt data augmentation strategies to coordinate systems. In polar coordinates, augmentation can be effectively achieved by applying a horizontal flip along the azimuth axis together with a random azimuthal translation. This approach leverages the inherent rotational properties of polar coordinates and ensures the consistency of augmentation effects across coordinates.

IV. Experiments

A. Experimental Setup

a) Datasets: The Boreas dataset [10] includes both a 128-beam LiDAR and a 360° Navtech CIR304-H scanning radar. The scanning radar provides a range resolution of 0.0596 m and an azimuth resolution of 0.9° at a frequency of 4 Hz, with a maximum operational range of 200 m. We use the Boreas-Objects-V1 subset, the only subset of Boreas officially provided with 3-D bounding box annotations. It contains 37 training scenes collected in sunny weather, each with a duration ranging from 20 to 70 seconds. LiDAR annotations are available at 5 Hz, despite the LiDAR sensor operates at 10 Hz. Radar annotations are derived by interpolating the LiDAR labels, which are then propagated up to 100 m using tracking, and transformed into the radar frame. These annotations are represented as 2-D bounding boxes due to the lack of height information in the radar data. This dataset offers annotated high-density LiDAR and high-resolution radar data, which is the primary reason for its adoption in our work.

b) Training and Testing Settings: We conduct experiments on LiDAR and scanning radar data, training models separately using Cartesian and polar coordinate representations under the same network architecture. Due to the difficulty of reliably representing near-field objects in polar coordinates, these objects are excluded from both training and evaluation in polar-based experiments. In Cartesian-based experiments, all objects are retained for training, but near-field objects are excluded from predictions and evaluation during testing. Consequently, the number of training objects is lower in the polar-based setting, while the number

of testing frames remains unchanged. To ensure fair evaluation, both configurations use the same set of testing objects for each sensor modality. For simplicity, our experiments consider only objects from the "Car" class. The total number of frames and annotated objects is summarized in Table I.

Modality	Coordinates	Training Frames	Training Objects	Testing Frames	Testing Objects
LiDAR	Cartesian Polar Box	3,812	68,174 60,815	1,130	11,259
Radar	Cartesian Polar Box	3,595	64,956 58,414	473	3,574

TABLE I: The number of frames and annotated objects used for training and testing in different coordinates and sensor modalities. Among all 37 sequences, to ensure sufficient training data, 7 sequences were selected as the test set for LiDAR-based experiments, and 4 sequences were selected as the test set for radar-based experiments. The remaining sequences were used for training.

In our experiments, the detection ranges are set to [0.0 m, 74.88 m] for LiDAR-based experiments and [0.0 m, 100.88 m] for radar-based experiments. We generate radar point clouds following [11], where the filtering threshold is defined as the sum of the mean and standard deviation along the azimuth axis. Our MBR estimation is then applied to the resulting point cloud within each object.

c) Voxel Map Configuration: To ensure a fair comparison, we align the number of pixels representing physically equivalent regions in both coordinates. Specifically, the polar voxel map covers a circular region in Cartesian coordinates (depicted as blue concentric circles in Figure 1(a)), with a radius denoted as R pixels, resulting in a total area of πR^2. For a 1920×1920 Cartesian voxel map, this corresponds to $R = 960$ pixels. The polar voxel map is configured with height $R = 960$ and width $\pi R \approx 3008$ pixels, yielding approximately the same total area of πR^2 pixels and ensuring comparability with the Cartesian representation. Both 960 and 3008 are multiples of 32, which facilitates compatibility with convolutional neural networks.

B. Comparison Results

We re-implemented the original PARTNER following the official codebase and use it as our baseline. To validate the effectiveness of our fully polar approach, we compare it with two configurations: the original PARTNER with Cartesian bounding box regression and PARTNER-C, a fully Cartesian configuration. PARTNER-C uses Cartesian coordinates for both input representation and regression, reflecting the conventional design in existing detectors. Our method builds upon PARTNER with the modifications described in Section III.

Experimental results for LiDAR and scanning radar data are reported in Table II and Table III. On the LiDAR dataset, our method outperforms both the original PARTNER model and its Cartesian-based configuration (PARTNER-C), highlighting the benefits of avoiding Cartesian-based metrics and leveraging native polar representations. A similar trend is observed on the more challenging radar data with longer detection ranges, where our approach demonstrates greater robustness under sparser and noisier measurements. In contrast, the baseline PARTNER model fails under these conditions, likely due to its difficulty in reconstructing object features from polar representations under such degraded conditions.

Experiment	Ground Truth Coordinates	Input Coordinates	AP_{30}
PARTNER [3]	Cartesian	Polar	92.77
PARTNER-C	Cartesian	Cartesian	93.42
Ours	Polar Box	Polar	**95.45**

TABLE II: Detection results on LiDAR data. Due to differences in the coordinates used for ground truth annotations, results are grouped accordingly. Our method achieves the highest AP_{30} compared to both PARTNER-C and PARTNER [3] baselines.

Experiment	Ground Truth Coordinates	Input Coordinates	AP_{30}
PARTNER-C	Cartesian	Cartesian	38.99
Ours	Polar Box	Polar	**44.10**

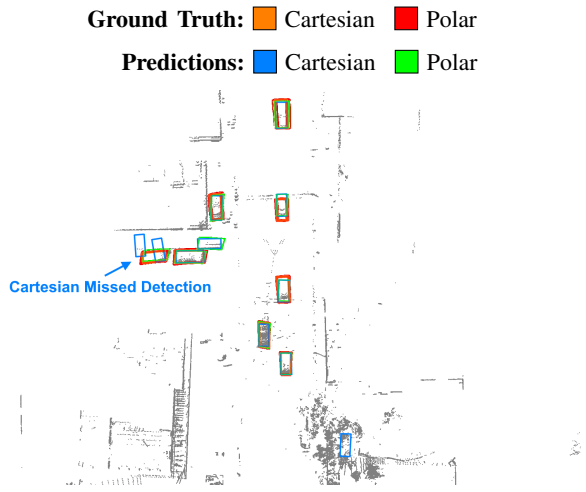
TABLE III: Detection results on scanning radar data. Results are grouped based on the coordinates used for ground truth annotations. Our method outperforms the PARTNER-C in AP_{30}, demonstrating better robustness under sparse and noisy measurements.

The qualitative results on both LiDAR and radar data are presented in Figure 11(a) and (b), respectively. These consistent gains across two distinct range-sensing modalities validate the effectiveness of our fully polar object detection pipeline and underscore its potential as a robust, modality-adaptive solution for object detection.

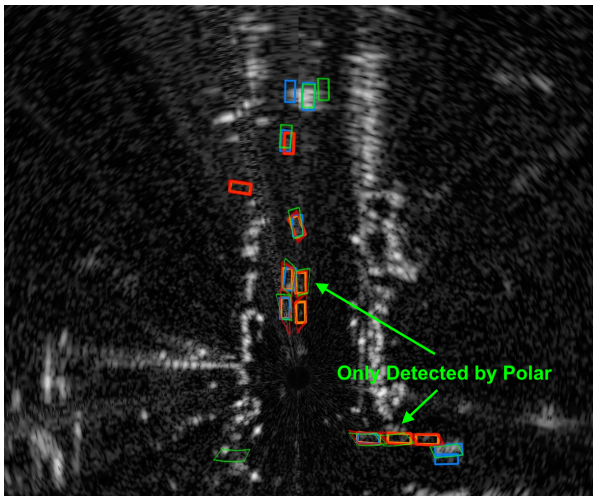
C. Ablation Study

In Section III-A, we introduce the box and orientation constraints that define our proposed constraint-based polar bounding box. We follow a consistent guideline in which mid- and far-field objects use both constraints, while near-to-mid-field objects apply only the orientation constraint, as the box constraint provides limited benefit in high-resolution regions. To best approximate the polar curvilinear projection, we compute the IoU between each object's polar curvilinear shape and both box definitions, selecting the one with the highest IoU as the final box. An ablation study is conducted to examine how different constraint combinations in box definition affect detection performance.

As shown in Table IV, the configuration without constraints performs the worst, as the box definition is based solely on the MBR, resulting in poor feature alignment. Applying only the box constraint improves AP for mid- and far-field objects, while using both constraints further enhances performance for near-to-mid-field objects, as previously shown in the Figure 7. Following a consistent guideline, mid- and far-field objects use both constraints, whereas near-to-mid-field objects apply only the orientation constraint. This adaptive design achieves the highest AP.



(a) LiDAR-based detection results



(b) Radar-based detection results

Fig. 11: Qualitative results on the LiDAR (a) and radar (b) datasets. The figure presents both Cartesian-based and polar-based detections, with the polar bounding boxes projected into Cartesian coordinates and superimposed on their Cartesian counterparts for direct comparison.

V. Conclusion

In this work, we propose a constraint-based polar bounding box representation that enables object detection entirely in polar coordinates, aiming to reduce reliance on Cartesian-based evaluation. This design naturally aligns with the geometry of range sensor measurements and supports a fully polar coordinate object detection pipeline. Experiments on LiDAR and scanning radar data show consistent improvements over baselines, validating the robustness and generalizability of our approach. Nevertheless, annotating near-range objects remains challenging due to high point density and large spatial footprint, which hinder accurate representation using a single bounding box in polar coordinates. Future research may explore alternative geometric descriptors, such as trapezoidal shapes, to more effectively capture object contours in polar coordinates. Network-level designs that account for distance-induced scale variations may serve as a potential direction to alleviate related detection errors.

Experiment	AP_{30}
No Constraints	92.81
Box Constraint Only	94.39
Orientation Constraint Only	93.50
Box + Orientation Constraints	95.38
Ours	**95.45**

TABLE IV: Ablation study on different box definitions. The no-constraint configuration performs the worst due to poor feature alignment. Adding the box constraint improves by better handling mid- and far-field objects. Applying both box and orientation constraints yields further gains, and our adaptive design achieves the best performance.

VI. Acknowledgment

This work is partially supported by the National Science and Technology Council in Taiwan via NSTC 113-2221-E-A49-148-MY3 and 113-2221-E-A49-159-MY3, and under the "Top Research Centers in Taiwan Key Fields Program" of the Ministry of Education (MOE), Taiwan.

References

[1] Y. Zhang, Z. Zhou, P. David, X. Yue, Z. Xi, B. Gong, and H. Foroosh, "Polarnet: An improved grid representation for online lidar point clouds semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[2] X. Zhu, H. Zhou, T. Wang, F. Hong, Y. Ma, W. Li, H. Li, and D. Lin, "Cylindrical and asymmetrical 3d convolution networks for lidar segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 9939–9948.

[3] M. Nie, Y. Xue, C. Wang, C. Ye, H. Xu, X. Zhu, Q. Huang, M. B. Mi, X. Wang, and L. Zhang, "Partner: Level up the polar representation for lidar 3d object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 3801–3813.

[4] M. Rapoport-Lavie and D. Raviv, "It's all around you: Range-guided cylindrical network for 3d object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2021, pp. 2992–3001.

[5] Q. Chen, S. Vora, and O. Beijbom, "Polarstream: Streaming object detection and segmentation with polar pillars," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 26 871–26 883.

[6] Z. Zhou, Y. Zhang, and H. Foroosh, "Panoptic-polarnet: Proposal-free lidar point cloud panoptic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 13 194–13 203.

[7] X. Wang, W. Feng, L. Kong, and L. Wan, "Nuc-net: Non-uniform cylindrical partition network for efficient lidar semantic segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

[8] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 10 012–10 022.

[9] 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.

[10] 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.

[11] Z. Hong, Y. Petillot, A. Wallace, and S. Wang, "Radar slam: A robust slam system for all weather conditions," *arXiv preprint arXiv:2104.05347*, 2021.