講演番号 390

Articulated Vehicle Detection via Learning Inter-Object Relationships using GNN in LiDAR Point Clouds

Riku Kagohashi [1)] Keigo Hariya [1)] Keisuke Yoneda [1)] Yukiya Fukuda [1)] Naoki Suganuma [1)]

1) Kanazawa University, Graduate School of Natural Science and Technology, Division of Mechanical Science and Engineering
Kakumamachi, Kanazawa-shi, Ishikawa, 920-1164, Japan (E-mail:kagohashi@stu.kanazawa-u.ac.jp)

KEY WORDS: **Software and its Underlying Technologies, Autonomous driving system platform, Machine learning,** Inter-Object Relationship [E3]

Accurate LiDAR-based 3D object detection is essential for autonomous driving systems. Among dynamic objects, articulated trucks present a complex structure comprising two 3D boxes: a Connected_Truck (C_truck) and a Trailer. This configuration frequently causes self-occlusion and information loss, making them difficult to distinguish from single-unit trucks. Consequently, PointPillars, a representative 3D object detection model utilizing LiDAR point clouds, struggles with misclassification, missed detections, and false positives. However, the strong spatial proximity between these two boxes provides crucial relational information. In this study, we propose extending PointPillars into a two-stage detection framework that explicitly incorporates inter-object relationships. Specifically, we construct a graph from first-stage detection proposals and introduce a Graph Neural Network (GNN) module to learn their associations. By integrating individual shape features with inter-vehicle spatial and contextual information, our approach aims to improve 3D detection accuracy by leveraging the inherent structural constraints of articulated trucks.

Fig.1 illustrates our proposed network, which extends PointPillars into a two-stage framework. Initial proposals are used to extract Region of Interest (RoI) features, from which inter-object relationship graphs are constructed. A Graph Neural Network (GNN) then updates node features by aggregating neighborhood information. We compare two GNN aggregation strategies: EdgeConv and GATv2Conv. Although EdgeConv captures local geometries via max-pooling, it is vulnerable to background noise. To address this, GATv2Conv utilizes a multi-head attention mechanism to dynamically weight critical nodes, mitigating noise. Finally, to preserve hierarchical representations, outputs from all GNN layers are concatenated with the initial RoI features. These refined features are fed into a second detection head to recalculate 3D bounding box parameters and object classes.

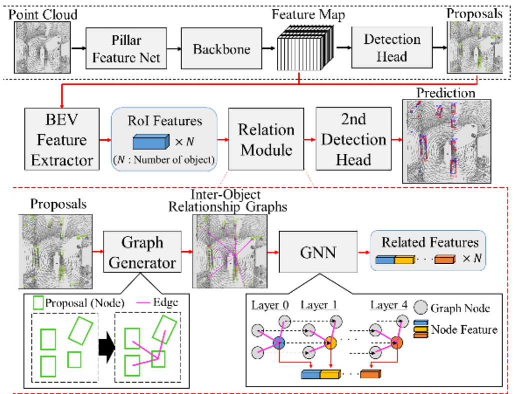

Fig.1 Overview of Network

Table 1 presents the Precision, Recall, and F1-Score for the Trailer and C_truck classes across the baseline PointPillars and the proposed method. The results demonstrate that the proposed method, which explicitly utilizes inter-object relationships, improves Precision for both classes compared to PointPillars. Furthermore, employing GATv2Conv—which adaptively learns critical relationships—for feature aggregation proves more effective in suppressing false positives. Qualitatively, Figure 2 illustrates that the proposed method successfully mitigates false positive detections of Trailers against the background, a common failure case in the baseline PointPillars. Conversely, Table 1 indicates a degradation in Recall and F1-Score for both Trailer and C_truck under the proposed method. We hypothesize that this increase in missed detections (false negatives) stems from noise contamination in the refined features during the GNN aggregation process. Specifically, EdgeConv is prone to indiscriminately aggregating irrelevant noise from surrounding backgrounds and distant objects. Similarly, in GATv2Conv, if the computed attention coefficients lack sufficient variance to clearly distinguish between target objects and noise, the aggregated features remain corrupted, ultimately leading to an increased number of missed detections.

Future work will focus on mitigating noise contamination within the GNN module. To achieve this, we plan to investigate restricting background edge formation during graph construction and optimizing GATv2Conv attention using a ground-truth-based loss function.

Table1 Results for Precision[%]], Recall[%], and F1-Score[%]

model		Precision[%]		Recall[%]		F1-Score[%]	
		Trailer	C_truck	Trailer	C_truck	Trailer	C_truck
PointPillars		54.88	52.37	**43.00**	**42.67**	**48.22**	**47.02**
Ours	EdgeConv	59.44	62.85	34.50	31.96	43.66	42.37
	GATv2Conv	**60.59**	**64.00**	31.99	28.04	41.87	39.00

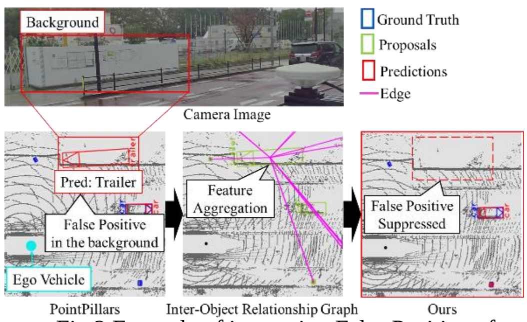

Fig.2 Example of improving False Positive of Trailer in the background