라이다 객체 검출을 위한 몬테카를로 드롭아웃 효과에 관한 연구

이종록, 박민철*

한국전자기술연구원, *한국전자기술연구원

jongrok@keti.re.kr, *mincheol.p@keti.re.kr

Rethinking Monte Carlo Dropout for LiDAR Object Detection

Jongrok Lee, Mincheol Park*

Korea Electronics Technology Institute, *Korea Electronics Technology Institute

요 약

본 연구는 라이다 객체 검출(LiDAR Object Detection) 분야에서 널리 사용되는 CenterPoint 모델을 기반으로, Monte Carlo Dropout(MC 드롭아웃) 기법의 새로운 활용 가능성을 탐구한다. MC 드롭아웃은 완전연결(FC) 계층에서 과적합 방지를 위해 널리 사용되었으나, 이미지 기반 비전 모델에서는 거의 활용되지 않는 기법이다. 그러나 본 연구에서는 MC 드롭아웃이 라이다 기반 3D 객체 검출에서 Spatial Reasoning 능력을 향상시킬 수 있음을 실험적으로 확인하였다. 구체적으로, 백본 네트워크의 특징맵 단계에 MC 드롭아웃을 적용하여 학습과 추론 과정 모두에서 불확실성을 반영하도록 하였으며, 그 결과 보행자(Pedestrian) 클래스에서 평균 정밀도(AP)가 2~3포인트 향상되는 성과를 얻었다. 제안 방법은 구현이 간단하여 CenterPoint뿐 아니라 다양한 LiDAR 객체 검출 네트워크에도 손쉽게 적용 가능하다. 본 연구는 드롭아웃의 활용을 재고찰함으로써, 3D 객체 검출의 성능 향상을 달성할 수 있는 실용적인 접근법을 제시한다.

Ⅰ. 서 론

LiDAR 기반 3D 객체 검출은 자율주행과 로봇 분야에서 핵심 기술로, BEV(Bird's-Eye View) 표현을 활용하는 CNN 백본 구조가 널리 사용되고 있다. 그러나 포인트 클라우드의 희소성(sparsity)으로 인해 BEV 백본이 확보할 수 있는 공간적 특징(Spatial Feature)의 밀도가 제한되며, 이는 특히 보행자나 자전거와 같이 작은 객체의 인식 성능 저하로 이어질 수 있다. 기존의 CNN 연구에서는 과적합 방지를 위해 드롭아웃(Dropout) 기법이 활용되었으나, 개별 픽셀 단위로 무작위 제거를 수행하는 Monte Carlo Dropout(MC 드롭아웃)은 공간적 특징을 훼손하는 부작용이 있어 주로 채널 단위로 제거하는 Dropout2D나 Spatial Dropout이 사용되어 왔다.[1] 하지만 LiDAR 도메인에서는 제한적인 Spatial Feature를 오히려 다양한 형태로 학습하게 함으로써 공간 추론(Spatial Reasoning) 능력을 강화할 가능성이 있다. 본 연구는 이러한 가설을 검증하기 위해, 학습 시 BEV 백본의 특징맵 단계에 MC 드롭아웃을 적용하여 제한적인 공간 특징의 활용을 다양화함으로써 객체 인식 성능을 향상시키는 접근을 제안한다. 또한 최근 제안된 BoS(Bounding Box Stability)[2] 연구에서 착안하여, MC 드롭아웃을 통한 특징 변동성이 모델의 일반화에 긍정적인 영향을 줄 수 있음을 고려하였으며, 간단한 드롭아웃 레이어 추가만으로도 보행자(Pedestrian) 클래스에서 AP가 약 2~3포인트 향상되는 결과를 확인하였다.

Ⅱ. 본론

본 연구에서는 제안한 Monte Carlo Dropout(MC 드롭아웃) 기법의 효과를 검증하기 위해 KITTI 데이터셋[3]을 활용하여 3D 객체 검출 실험을 수행하였다. 검출 모델로는 라이다 기반 객체 탐지 모델 중, BEV 구조를 사용하는 CenterPoint-PointPillar[4] 구조를 선택하였으며, 입력 포인트 클라우드를 BEV 형태로 변환하기 위해 두 가지 다른 보셀 크기(voxel size)를 설정하였다. 구체적으로, 0.16m와 0.32m 두 가지 보셀 크기를 적용하여 BEV 상 Spatial Feature 해상도 차이에 따른 드롭아웃 성능 변화를 비교하였다. 제안 기법은 백본 네트워크의 특징맵 단계 1과 2에 MC 드롭아웃 레이어를 삽입하는 방식으로 구현하였다. 학습 단계에서는 일반적인 드롭아웃과 동일하게 무작위 feature를 제거하며, 추론 단계에서도 드롭아웃을 유지하여 불확실성을 반영하는 Monte Carlo 추론을 수행하였다. 이를 통해 희소성이 큰 BEV 특징맵에서 다양한 공간적 표현을 학습할 수 있도록 하였다.

실험 결과, Car 클래스에서는 성능 차이가 거의 나타나지 않았으며, Cyclist 클래스의 경우 소폭 성능 저하가 관찰되었다. 반면, BEV상 크기가 작아 가장 제한적인 Spatial Feature를 가지는 Pedestrian 클래스에서 두 보셀 크기(0.16, 0.32) 모두 평균 정밀도(AP)가 약 2~3 포인트 향상되는 개선 효과가 확인되었다. 이는 포인트 클라우드의 희소성으로 인해 BEV 백본이 충분한 공간 정보를 확보하기 어려운 상황에서, MC 드롭아웃이 오히려 다양한 공간적 표현 학습을 유도하여 Pedestrian 검출의 Spatial Reasoning 능력을 강화한 결과로 해석할 수 있다.

표 1. (Moderate) MC 드롭아웃에 따른 모델 별 성능 변화 (AP)
Table 1. Performance Variation by Model with MC Dropout (AP)

	Car	Pedestrian	Cyclist
0.32	63.41	30.29	45.67
0.32_MD	62.58	**33.19**	45.89
0.16	72.24	40.18	59.85
0.16_MD	72.57	**42.33**	58.49

표 2. (Hard) MC 드롭아웃에 따른 모델 별 성능 변화 (AP)
Table 2. Performance Variation by Model with MC Dropout (AP)

	Car	Pedestrian	Cyclist
0.32	58.99	28.21	42.92
0.32_MD	58.34	**30.49**	42.56
0.16	66.54	37.95	55.83
0.16_MD	**69.53**	**39.39**	54.42

Ⅲ. 결론

본 연구에서는 LiDAR 기반 3D 객체 검출에서 Monte Carlo Dropout(MC 드롭아웃)의 가능성을 재고찰하였다. 기존 CNN 이미지 도메인에서는 공간적 특징(Spatial Feature)을 훼손하는 문제로 인해 거의 사용되지 않던 MC 드롭아웃을, 포인트 클라우드의 희소성으로 인해 Spatial Feature가 부족한 BEV 백본에 적용함으로써 오히려 Spatial Reasoning 능력을 강화할 수 있음을 실험적으로 확인하였다.

이는 제안 방식이 단순히 드롭아웃 레이어를 추가하는 것만으로 구현 가능하다는 점에서, 다양한 LiDAR 객체 검출 네트워크에도 손쉽게 확장될 수 있는 장점이 있다. 다만 본 연구에서는 특정 백본 구조와 제한된 데이터셋에 국한하여 검증하였으므로, 향후에는 PV-RCNN, BEVDet 등 다양한 3D 검출 모델로의 확장 실험과 Waymo, nuScenes와 같은 대규모 데이터셋에 대한 검증이 필요하다.

ACKNOWLEDGMENT

※ 본 연구는 산업통상자원부의 제조 기반 생산 시스템 사업(No. 00442974, 대규모 토석 운반 자동화를 위한 덤프트럭용 자율작업 및 운영 시스템 개발)의 지원을 통해 연구가 수행되었음

※ 본 연구는 과학기술정보통신부가 주최하고 정보통신기획평가원 및 한국정보통신기술협회가 주관하는 「2025년도 인공지능 챔피언(AI Champion) 대회」의 지원을 받아 수행되었음.

참 고 문 헌

[1] Tompson, J., Goroshin, R., Jain, A., LeCun, Y., & Bregler, C. (2015). Efficient object localization using convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 648-656).

[2] Yang, Y., Wang, W., Chen, Z., Dai, J., & Zheng, L. (2024). Bounding box stability against feature dropout reflects detector generalization across environments. arXiv preprint arXiv:2403.13803.

[3] Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The kitti dataset. The international journal of robotics research, 32(11), 1231-1237.

[4] Yin, T., Zhou, X., & Krahenbuhl, P. (2021). Center-based 3d object detection and tracking. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11784-11793).