

효율적인 3 차원 물체 검출을 위한 포인트 필라 네트워크 성능 개선

박건율, 김예철, 김지송, 최광진, 최준원*
한양대학교

konyulpark@spa.hanyang.ac.kr, yckim@spa.hanyang.ac.kr, jskim@spa.hanyang.ac.kr,
kjchoi@spa.hanyang.ac.kr, junwchoi@hanyang.ac.kr

Enhanced PointPillar Network for Efficient 3D Object Detection

Konyul Park, Yecheol Kim, Jisong Kim, Kwangjin Choi, Jun Won Choi*
Hanyang University

요 약

본 논문은 자율주행 환경에서 사용되고 있는 LiDAR 센서 사용 모델인 PointPillars 모델의 성능을 개선하는 기술을 제안하였다. 우선 기존 모델에서 사용하는 필라에 대해 수직 방향으로의 복셀화를 진행한 후 수직 방향으로의 복셀들에 대한 연산을 추가 진행하여 수직 방향으로의 지역적 정보를 추출한다. 이후 2D Sparse Convolution을 사용한 특징 지도와 앞서 만든 복셀화된 필라에 Residual Network를 통과시켜 수직 방향으로의 지역적 정보와 수평 방향으로의 지역적 정보를 취합한다. 본 논문에서는 nuScenes 데이터셋을 이용하여 실험을 진행하였으며, 제안한 방식을 사용하여 기존 방법에 비해 높은 성능을 달성하였다.

I. 서 론

최근 라이다 포인트 클라우드 데이터를 활용한 3 차원 물체 검출연구는 성능향상에 초점을 맞춰 꾸준한 발전을 이뤄냈다[1, 2, 3, 6]. 라이다 데이터를 사용한 3 차원 물체 검출 모델들은 라이다 포인트 인코딩 방식에 따라 포인트 기반 방식과 복셀 기반 방식으로 나눌 수 있다. 이들 중에서 복셀 기반 모델들을 활용하여 실제 자율주행 상황에서 사용할 수 있도록 모델 동작속도를 향상시키는 연구가 최근 활발히 진행되고 있다. PointPillars[1]는 기존 3 차원 형식의 복셀 형태를 z 축 방향에 대해 통합하여 수직 열 기둥 모양(필라)으로 그리드를 나누고 2 차원 CNN을 사용하여 특징 벡터를 뽑고 물체를 검출한다. PointPillars 모델은 2 차원 CNN을 사용하기 때문에 다른 모델들에 비해 속도가 매우 빠르지만 SECOND 모델에 비해 낮은 정확도를 가진다. PillarNet[4]은 공간-의미적 특징 융합을 위한 Encoder 네트워크를 구성하여 SECOND 보다 빠르지만 비슷한 정확도를 얻었다. 본 논문에서는 복셀과 필라 기반 모델의 정확도 차이를 줄이고 더 빠른 3 차원 물체 검출 알고리즘을 개발하였다. PointPillars 모델을 기본 모델로 사용하여 라이다 포인트를 필라 형식으로

변환하고 Z 축에 대해 부족한 정보를 강화하기 위해 Z 축 방향의 Aggregation을 통해 모델의 연산 속도는 거의 유지하고 성능을 향상시켰다.

II. 본론

2.1) 제안하는 3 차원 물체 검출 최적화 기술

제안하는 모델의 설계 목적은 PointPillars 모델의 연산 속도를 유지하면서 성능을 향상시키는 것이다. 본 논문에서 소개하는 3 차원 물체 검출 최적화 알고리즘은 두가지 모듈을 새롭게 제안하며, 모델 전체 구조는 Fig. 1과 같다.

Z-aware Aggregation 모듈은 필라 인코딩 정보를 Z 축 방향에 대해 여러 구획으로 나누고 이를 CNN 연산을 수행해 Z 축 정보를 강화하는 모듈이다. Bin-wise residual network 모듈은 2D 희소 합성곱을 사용하여 XY 평면상의 지역적 특징 벡터를 얻은 후 Z 평면의 지역적 특징 벡터를 얻는다. 이후 물체 검출 모듈을 통하여 최종적인 3 차원 물체 박스를 얻는다.

앞서 서술한 방법들은 기존의 PointPillars 알고리즘의 취약점인 Z 방향으로의 연산을 가능하게 함으로써

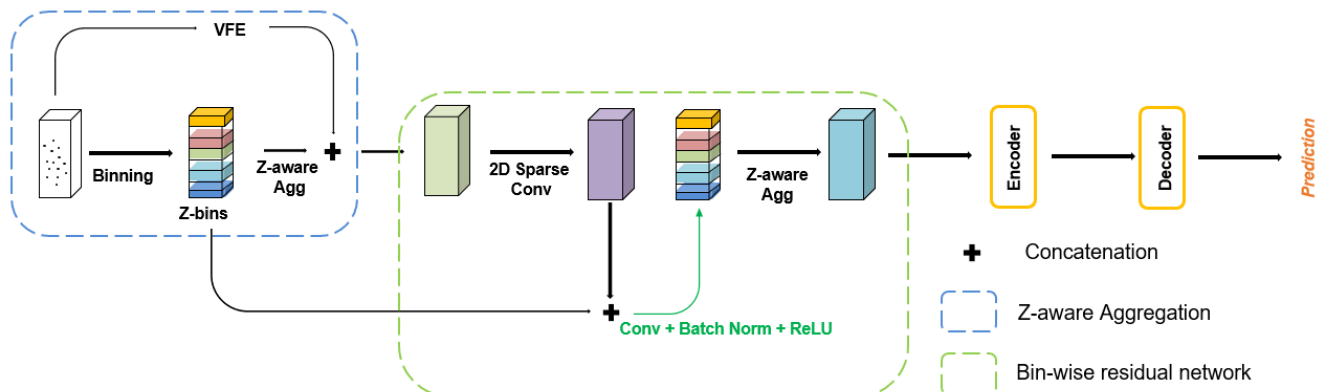


Figure 1. 제안하는 네트워크 전체 구조

연산속도를 최대한 유지함과 동시에 3 차원 물체 검출의 성능을 향상시킨다.

2.2) Z-aware Aggregation

Z-aware Aggregation 은 기존 필라 인코딩 방식에 부족한 Z 축 방향으로의 지역적 정보를 얻게 해주는 모듈이다. 필라를 Z 방향에 대해 8 개의 구획으로 나눈 후 각 구획을 대표하는 특징을 해당 구획 내부 포인트 클라우드의 평균값으로 정의한다. 이후 8 개의 구획에 대한 특징 벡터들을 Z 축 방면으로 연속적으로 붙여준 뒤, 1-channel CNN 을 통과시켜 각 8 개의 구획에 해당하는 정보를 얻는다. 이후 기존 PointPillars 에서 사용하는 Voxel Feature Encoding 모듈이 적용된 특징을 더해준다.

2.3) Bin-wise residual network

각 필라 내부에 대한 연산이 끝난 이후 2D 회소 CNN 을 사용하여 XY 평면으로의 연산을 적용한다. 이때 해당 네트워크는 XY 평면으로 연산이 적용된 특징에 대해 Z 방향으로의 연산을 추가로 진행한다. Z-aware Aggregation 방법에서 Z 축 연산을 수행하기 위해 나눈 pillar 를 Fig. 1 에서 보이듯 Z-bins 라 한다. 이 Z-bins 와 앞서 2D 회소 CNN 을 사용하여 XY 평면으로의 연산이 적용된 특징을 통합한다. 이렇게 통합된 특징 정보에 1-channel CNN, Batch Normalization 을 통과시키고 이후 Z-aware Aggregation 을 진행한다. 이러한 방법을 사용하게 되면 XY 평면에서의 2D 회소 CNN 연산으로 얻어진 지역적 정보를 가진 특징 정보에 대해서 Z 축 방향으로의 추가 연산을 진행할 수 있는 장점이 있다. Bin-wise network 이후의 연산 과정은 PointPillars 모델의 Region Proposal Network(RPN) 모듈과 동일하다.

Table 1. nuScenes 1/7 데이터셋을 사용한 실험 결과

Method	mAP (%)	NDS (%)	FPS
Centerpoint (pillar-based)	42.53	54.6	10.77
Centerpoint (voxel-based)	52.98	61.41	6.09
PillarNet	49.33	58.87	8.27
Ours	49.60	59.52	7.22

2.4) 실험

제안하는 네트워크는 nuScenes 1/7 데이터셋을 이용하여 실험을 진행하였다. 제안하는 방법의 성능 평가는 아래 Table 1, Table 2 와 같이 나타났다. 기존의 방식보다 더 높은 NDS, mAP 를 보이는 것을 확인할 수 있다. 또한 FPS 의 경우에도 기존 방법과 비교하여 큰 차이를 보이지 않음을 확인할 수 있다.

본논문에서는 새로운 3 차원 물체 검출 최적화 알고리즘을 제안한다. 자율 주행 환경에서 활발히 사용하는 PointPillars 모델에 두가지 모듈을 추가하여 Z 축으로의 연산을 진행하였다. Z-aware Aggregation 을 사용하여 기존 PointPillars 모델의 부족한 Z 축 정보를 채웠으며, Bin-wise residual network 를 사용하여 XY 평면에 대한 연산에 Z 방향으로의 연산을 residual connection 방법으로 추가하여 Z 방향 정보를 더욱 강화하였다. nuScenes 데이터 셋을 활용하여 실험을 진행하였으며, 비교모델인 PillarNet 모델에 비해 mAP 0.27, NDS 0.65 높은 성능을 보였다.

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2022-0-00957, 엣지용 분산형 온칩메모리-연산기 융합 PIM 반도체 기술개발)

참 고 문 헌

- [1] Lang, Alex H., et al. "Pointpillars: Fast encoders for object detection from point clouds." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
- [2] Yin, Tianwei, Xingyi Zhou, and Philipp Krahenbuhl. "Center-based 3d object detection and tracking." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
- [3] Yan, Yan, Yuxing Mao, and Bo Li. "Second: Sparsely embedded Aggregational detection." Sensors 18.10 (2018): 3337.
- [4] Shi, Guangsheng, Ruifeng Li, and Chao Ma. "PillarNet: High-Performance Pillar-based 3D Object Detection." *arXiv preprint arXiv:2205.07403* (2022).
- [5] Zhou, Yin, and Oncel Tuzel. "Voxelnet: End-to-end learning for point cloud based 3d object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

III. 결론