

센서 융합 기반 3차원 객체 검출 네트워크의 도메인 일반화 성능 개선

윤지수, 이준호, 최준원*
한양대학교, *서울대학교

jsyoon@spa.hanyang.ac.kr, jhlee@spa.hanyang.ac.kr, *junwchoi@snu.ac.kr

Domain Generalization for Sensor Fusion-based 3D Object Detection

Jisu Yoon, Junho Lee, Jun Won Choi*
Hanyang Univ., *Seoul National Univ.

요약

본 논문은 자율주행 환경에서 사용되는 LiDAR 센서의 포인트 클라우드 증강 기법을 통해 LiDAR-카메라 융합 기반 3차원 객체 검출 모델의 도메인 일반화 능력을 개선시키는 기법을 제안하였다. 제안하는 방법은 3차원 객체 검출 모델이 센서 설정에 따라 달라지는 다양한 도메인에 대해 학습하는 것을 목표로 한다. 먼저 LiDAR로부터 얻은 포인트 클라우드 데이터에 채널 샘플링을 수행한다. 그 다음, 학습 과정에서 샘플링된 포인트 클라우드와 원본 포인트 클라우드 중 랜덤하게 선택하여 입력 데이터로 활용한다. 실험은 소스 도메인으로써 nuScenes 데이터셋을 사용했으며, 타겟 도메인으로 Waymo, KITTI 데이터셋을 사용하여 성능 평가를 진행했다. 제안한 방식을 통해 기존 방법 대비 모델의 일반화 능력이 향상됨을 확인하였다.

I. 서론

본 논문에서는 LiDAR-카메라 융합 기반 3차원 객체 검출 모델의 도메인 일반화 능력을 개선시키는 데이터 증강 기법을 제안한다. 성능 향상을 목적으로 하는 대부분의 연구는 학습 데이터셋과 동일한 데이터셋 상에서 성능 평가를 수행한다. 이 경우 학습 데이터셋과 다른 데이터 분포를 갖는 데이터셋에서 성능 평가가 이루어졌을 때, 도메인 이동이 발생하며 성능이 하락한다. 도메인 일반화 작업은 소스 도메인 데이터만을 학습에 사용하여 이전에 보지 못한 타겟 도메인에서도 모델이 강건한 성능을 유지하는 것을 목표로 한다[1,2]. 이때 소스 도메인에서의 성능 역시 유지되어야 한다.

자율주행을 위한 데이터셋에서도 이러한 도메인 이동이 발생한다. 가장 큰 원인으로 데이터셋이 수집된 지역 차이와 사용된 센서 설정의 차이 등이 있다. 대표적인 자율주행 데이터셋 KITTI, nuScenes, Waymo를 비교했을 때, 데이터가 수집된 지역은 독일, 미국, 싱가포르로 상이하며 지역에 따른 차의 평균 크기에 차이가 존재한다. 또한 사용된 LiDAR 센서 기종에 따라 채널 수, 시야 범위, 장면에 포함된 포인트 수에서 큰 차이가 발생한다.

제안하는 기법은 LiDAR의 채널 수 차이에 초점을 맞춘 채널 샘플링 기법이며, 실험을 위한 기본 모델로 LiDAR-카메라 센서 융합 모델인 BEVFusion[3]을 채택하였다. 32 채널 LiDAR를 사용하는 nuScenes 데이터셋을 소스 도메인으로 설정하여, 제안하는 채널 샘플링 기법을 통해 다양한 채널로 증강된 포인트 데이터를 생성한다. 매 반복마다 원본 데이터와 증강된 데이터 중 랜덤하게 모델의 입력 데이터를 선택하여 학습을 진행한 결과, nuScenes 데이터셋에서의 성능을 유지함과 동시에 기존 모델에 비해 타겟 도메인인 Waymo, KITTI 데이터셋에서 성능을 향상시켰다.

II. 본론

2.1) 제안하는 LiDAR 포인트 채널 샘플링 기법

제안하는 기법의 목표는 소스 도메인 nuScenes 데이터셋에서 모델 성능을 유지하는 동시에 타겟 도메인 Waymo, KITTI 데이터셋에서 성능을 향상시키는 것이다. Fig. 1-(a)에 제안하는 채널 샘플링 기법과 Fig. 1-(b)에 이를 적용한 전체 구조를 나타냈다. 채널 샘플링 기법은 크게 LiDAR 채널 사이 보간법을 통해 포인트 클라우드의 밀도를 높이는 업 샘플링과, 기존의 채널을 삭제하여 최소화하게 만드는 다운 샘플링으로 나뉜다.

먼저 업 샘플링은 포인트의 zenith 각도를 기준으로 K-means Clustering을 수행하여 포인트가 포함된 채널의 인덱스를 구한다. 이후 i 번째 채널에 속하는 임의의 포인트 p^i 를 기준으로 K-Nearest Neighbors 알고리즘을 통해 동일한 채널 인덱스를 갖는 이웃 포인트를 찾는다. 포인트 p^i 와 이웃 포인트 사이의 방향 벡터와 포인트 세기 값을 이용하여 동일 채널 내의 새로운 포인트를 생성한다. 이때, 채널 수 증가를 위해서는 생성된 포인트의 z 좌표 조정이 추가적으로 필요하다. 현재 i 번째 채널과 $i+1$ 번째 채널의 zenith 각도를 보간 하는 zenith 각도를 계산한다. 해당 각도를 이용하여 생성된 포인트를 3차원 공간에 매핑함으로써 새로운 채널을 생성할 수 있다.

다운 샘플링 역시 업 샘플링 과정과 비슷하게 포인트의 zenith 각도를 기준으로 포인트들의 채널 인덱스를 구한다. 그 다음 목표 채널 수에 맞도록 일정 간격에 따라 해당 인덱스 채널에 속하는 포인트들을 제거한다. 위 과정을 통해 기존 32 채널로부터 64 채널, 16 채널의 다양한 밀도를 갖는 포인트 클라우드 데이터를 획득할 수 있다.

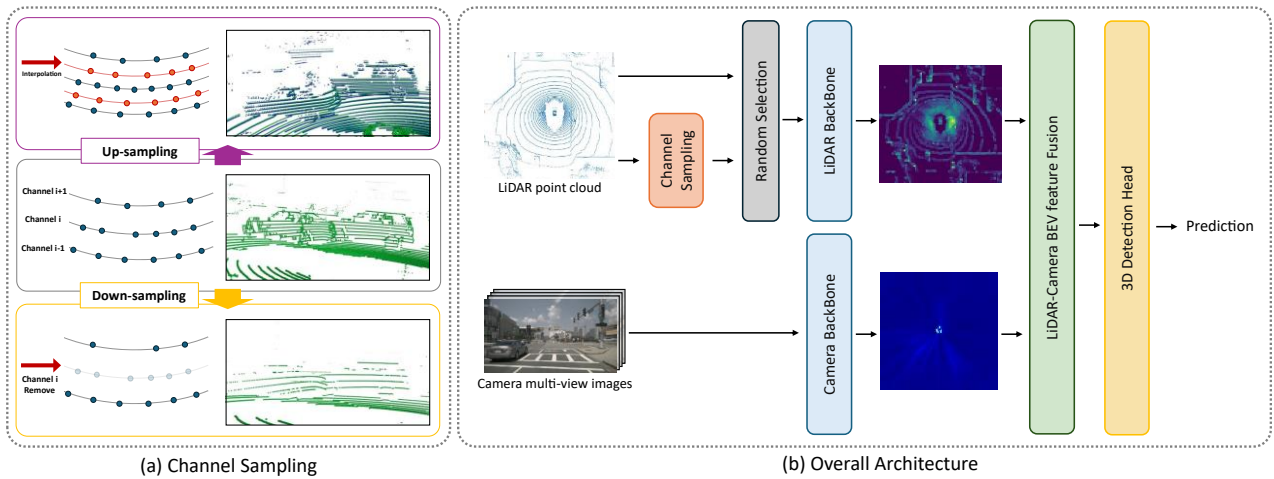


Figure 1. 제안하는 기법과 전체 구조

2.2) 랜덤 채널 선택 학습

채널 샘플링 기법을 통해 생성된 여러 채널의 포인트 클라우드 데이터는 모델 학습에 활용된다. 학습 과정에서 매 반복마다 원본 포인트 클라우드와 샘플링 된 포인트 클라우드 중 하나를 랜덤하게 선택하고, 선택된 포인트 클라우드 데이터는 LiDAR Backbone 모델에 입력된다. 이러한 학습 방법은 원본 데이터를 통해 소스 도메인에서의 모델 성능을 해치지 않으면서, 증강된 데이터를 통해 모델이 다양한 채널 수에 대해 학습할 수 있도록 돕는다.

Table 1. nuScenes, Waymo, KITTI 데이터셋 실험 결과

Method	nuScenes AP(%)	Waymo AP(%)	KITTI AP(%)
BEVFusion (baseline)	85.3	53.27	63.06
Up-sampling	85.1	57.4	72.94
Down-sampling	85.5	46.55	63.93
Up / Down sampling	84.9	56.38	74.02

2.3) 실험 결과

제안하는 기법을 적용하여 nuScenes 데이터셋의 차량 클래스에 대해 모델을 학습하였고, 해당 모델 가중치를 이용하여 nuScenes, Waymo, KITTI 데이터셋에 대해 성능 평가를 실시했다. 단일 클래스의 성능 비교를 위해 객체 검출 평가 지표인 AP 를 사용하였다. 실험은 업 샘플링, 다운 샘플링 그리고 두 샘플링 기법을 모두 사용하는 총 3 가지 시나리오에 대해 진행했다. 성능 평가 결과는 Table. 1 과 같이 나타난다. 베이스 모델인 BEVFusion[3]에 비해 세가지 시나리오 모두 nuScenes 에서의 성능을 거의 유지했으며, Waymo 의 경우 업 샘플링만 사용했을 때 AP 가 4.13% 가량 상승했다. KITTI 의 경우 업/다운 샘플링을 모두 사용했을 때 AP 가 10.96% 상승하며 도메인 이동에도 강한 일반화 능력을 보였다.

III. 결론

본 논문에서는 3 차원 객체 검출 모델의 일반화 능력을 개선하기 위한 데이터 증강 및 학습 기법을 제안한다.

다양한 자율주행 데이터셋에서 LiDAR 채널 개수 차이에 따른 도메인 이동 문제를 해결하기 위해 포인트 클라우드 데이터에 채널 샘플링 기법을 적용하였다. 채널 업/다운 샘플링을 통해 소스 도메인의 포인트 클라우드 데이터에 증강을 적용했고, 이를 원본 데이터와 함께 학습에 활용하였다. 실험 결과 소스 도메인 nuScenes 데이터셋에서 성능을 유지하고, 여러 타겟 도메인 Waymo, KITTI 데이터셋에서 성능 향상을 보이며 다양한 도메인 상에서 모델의 일반화 능력이 개선됨을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2020-0-01373,인공지능대학원 지원(한양대학교))과 2024 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2020R1A2C2012146)

참 고 문 헌

- [1] 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..
- [2] Wang, Shuo, et al. "Towards domain generalization for multi-view 3d object detection in bird-eye-view." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.
- [3] Liu, Zhijian, et al. "Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation." *2023 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2023.