

도메인 갭 해결을 통한 라이다 객체인식 모델 백본 특성 유지

장지윤, 장민철, *김진규
고려대학교 컴퓨터학과

e-mail : fick17@korea.ac.kr, m1ncheoree@korea.ac.kr, jinkyukim@korea.ac.kr

Retaining Backbone Representation under Limited Data in 3D LiDAR-based Detection by Addressing Domain Gap

Jiyun Jang, Mincheol Chang, *Jinkyu Kim
Department of Computer Science and Engineering
Korea University

Abstract

LiDAR-based 3D object detection has greatly improved through deep learning, but those detectors require repetition of collecting and annotating data in a new city or environment for adaption. Hence, to build an autonomous driving system in real-world scenarios, a fine-tuning strategy with few annotated data from target domain is essential.

In this paper, we propose an effective fine-tuning strategy leveraging pre-trained backbone models with limited labeled data, maintaining representation power of backbones. Our extensive experiments on various datasets show our method's superiority.

I. 서론

라이다 기반 3D 객체 인식 모델들은 자율주행, 로봇 등 다양한 분야에서 중요하게 사용이 된다. 같은 도메인에서 학습되고 테스트 될 때는 좋은 성능을 보이나 다른 도메인에 적용 될 경우 취득 센서의 세팅 (라이다 빔 수, FOV, 해상도 등) 이나 위치적 변화에

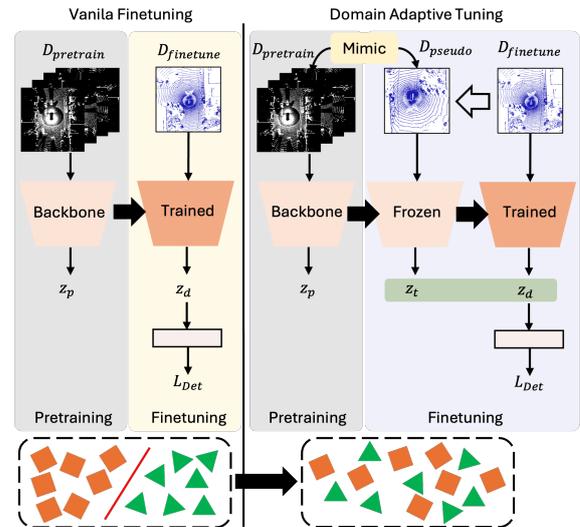


그림 1. 기존 파인 튜닝 방식과 도메인 적응 파인 튜닝 방식 비교

따라 인식 모델 성능 저하를 야기한다. 그러므로 새로운 지역에서 인식 모델을 적응시킬 때 타겟 도메인에서 추가적인 데이터 수집 및 레이블링이 요구된다.

이를 해결 하기 위해 AD-PT [1]와 같은 연구는 대량의 레이블링이 안된 라이다 데이터를 다양한 라이다 증강법과 자기주도학습을 통하여 여러

도메인에서 적용가능한 백본 모델을 만들었다. 비록 여러 벤치마크 데이터셋에서 좋은 성능을 보였지만 여전히 백본 모델을 타겟 도메인에 파인 튜닝(Fine-tuning)할 때 타겟 도메인에서 많은 양의 레이블 된 데이터를 필요로 하는 문제가 있다.

본 연구에서는 새로운 도메인에서 데이터 레이블링 소요를 최소화하는 동시에 백본 모델이 가지고 있는 특성 및 성능을 최대한 유지할 수 있는 파인 튜닝 방식을 제시한다. 이는 백본 모델이 학습된 도메인과 타겟 도메인 사이의 차이를 줄여주는 방법으로 실현된다. 그림 1에서 흔히 사용되는 기존 파인 튜닝 방식(왼쪽)과 본 논문에서 제시한 도메인 적응 파인 튜닝 방식(오른쪽)간의 비교를 보여준다. 도메인 적응 파인 튜닝기법은 먼저 타겟 도메인의 라이더 밀도를 백본 도메인의 라이더 밀도와 비슷하게 맞춰주고 거기서 나오는 특성을 큰 모델 (Teacher), 기존 밀도를 유지한 데이터를 학습하는 모델을 작은 모델 (Student)로 두고 큰 모델로부터 증류한 지식을 작은 모델로 전이(transfer)시킨다. 여러 데이터셋에서 기존 방식과 본 연구에서 제시한 파인 튜닝의 성능 비교를 통하여 도메인 적응 방식의 유용성을 보여준다.

II. 본론

기존에 주어진 백본 모델을 새로운 타겟 도메인에서 적은 학습 데이터로 파인 튜닝하는 방식을 다음과 같이 제시한다.

2.1 라이더 데이터 증강

먼저 타겟 도메인의 라이더 데이터를 백본 모델의 도메인 데이터 밀도의 맞추는 과정을 진행한다. AD-PT 백본은 대규모 레이블링이 안된 ONCE [2] 데이터셋에서 학습이 진행되었다. ONCE 데이터셋은 빔(beam) 수가 40으로 낮은 편이므로 이를 위하여 높은 빔 수의 라이더 데이터를 낮은 빔 수로 내리는 모듈: PLBG (Pseudo Low Beam Generation)을 구현하였다.

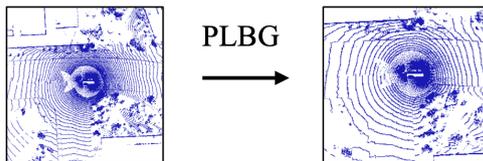


그림 2. PLBG 예시

먼저 x, y, z 의 Cartesian 좌표를 가지고 있는 라이더 포인트들을 r, θ, ϕ 의 Spherical 좌표로 변환을 한다.

그 후 좌표 값을 라이더 세팅의 FOV에 맞춰 이미지 좌표로 변환을 한다. 이후 K-means Clustering을 통하여 비슷한 vertical 좌표를 가지고 있는 포인트 cluster들을 만들고 cluster들에서 downsampling을 진행함으로 타겟 빔 수의 라이더 데이터를 생성한다. 그림 2에서 볼 수 있듯이 왼쪽에 촘촘하고 고밀도의 모델 input 라이더 썬이 PLBG를 통하여 오른쪽의 낮은 빔 수의 라이더 썬으로 바뀌었다. 이 pseudo 데이터를 백본에 제공함으로써 타겟 도메인과 백본 도메인의 밀도 갭을 줄여주고 백본 모델의 특성을 유지한다.

2.2 Teacher-Student 증류

기존 타겟 도메인 라이더 인풋 D_t 와 증강된 D_t 가 각각 Student와 Teacher 백본에 들어간다. 이 때 Teacher의 백본은 Frozen, 즉 gradient를 업데이트 해주지 않으면서 증류를 진행한다.

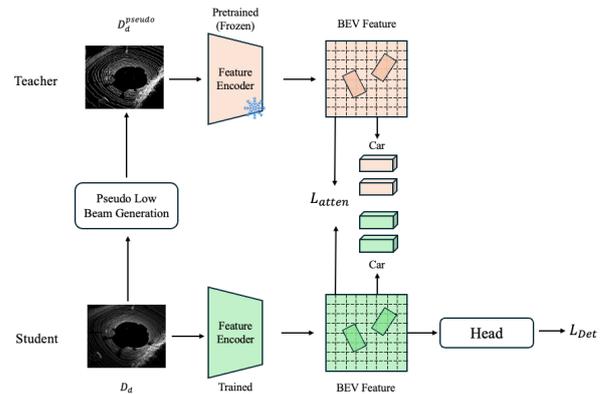


그림 3. Teacher-Student 증류 구조

그림 3에서 볼 수 있듯이 먼저 빔 수를 내려준 라이더 데이터 D_{pseudo} 에서 Teacher 백본을 통해 Bird Eye's View (BEV) 특성을 뽑아낸다. 그리고 원래 데이터 D_t 를 Student 백본을 통해 BEV 특성을 뽑아낸다. 그 다음 Teacher, Student BEV에서 각각 자동차에 해당하는 특성들을 뽑고 둘 사이에 거리를 L_{atten} 손실함수로 계산한다. 마지막으로 Teacher에서 지식을 증류 받은 Student BEV는 객체 인식을 진행하며 L_{Det} 을 계산한다. 계산된 Loss는 더해주고 gradient 역전파에 참여한다.

III. 실험 결과

3.1 실험 환경

Baseline 모델 및 본 연구 방법론의 성능평가는

Waymo[3] 및 KITTI[4] 공개 자율주행 데이터셋에서 진행되었다. 본 연구에서 다른 데이터셋들에 세부 내용은 다음과 같다:

Dataset	VFOV	# of Beams	# of Avg. Points	Location
ONCE	[-25°, -15°]	40	64,981	China
Waymo	[-17.6°, -2.4°]	64	160,887	USA
KITTI	[-25°, -15°]	64	119,748	Germany

표 1. 데이터셋 별 세부 내용

표 1에서 볼 수 있듯이 타겟 도메인인 Waymo, KITTI의 빔 수가 ONCE보다 많으므로 PLBG를 적용하여 데이터 증강을 진행하였다. 모델 구현은 PyTorch로 구현했으며 Adam Optimizer[5]를 사용해 모델 Loss를 최적화 하였다. 모든 객체 인식 모델의 실험은 KITTI에서 80, Waymo에서 30 epoch동안 진행 됐으며 RTX 3090에서 진행되었다. Waymo와 KITTI의 타겟 도메인의 학습 데이터는 각각 144장, 185장으로 제한하였다. 본 연구에서 객체 인식 모델로는 SECOND[6] 구조를 선택하였다.

3.2 정량적 비교

객체 인식의 평가 지표로는 주로 mAP (Mean Average Precision)을 사용한다. 이 지표는 먼저 모델이 예측한 박스와 정답인 Ground Truth 박스 간에 IoU로 계산된다:

$$IoU = \frac{Intersection}{A + B - Intersection}$$

IoU는 두가지 영역의 전체 집합에서 교집합에 대한 확률로 정의된다. 두가지 영역에 대한 겹침 부분이 많을수록 성능이 나음을 의미한다. 이 IoU값들에 threshold에 따른 precision recall값들을 사용해 평균 AP를 구한 mAP를 정량적 지표로 사용하였다.

모델/성능	자동차	보행자	자전거
Vanilla	37.2	21.8	37.6
DAT	47.4	26.6	41.1

표 2. Waymo mAP 성능 비교

표 2는 AD-PT 백본 모델을 Waymo 데이터셋에서 144개의 랜덤하게 추출된 적은 양의 학습데이터로 파인 튜닝한 실험들의 성능 비교이다. Vanilla는 기존처럼 그대로 파인 튜닝하는 방식, DAT (Domain Adaptive Tuning)은 본 연구에서 제시한 도메인 적응 방식으로 백본을 파인 튜닝한 결과이다. 표 2에서 볼 수 있듯이 DAT 방식이 Vanilla에 비해 모든 객체에서 인식 성능이 향상됐다.

추가적으로 KITTI에서도 DAT의 성능 향상을

확인하였다. KITTI에서도 185장의 적은 양의 라이더 데이터를 가지고 파인 튜닝 실험을 진행했다.

모델/성능	자동차	보행자	자전거
Vanilla	50.1	30.3	31.9
DAT	59.6	31.2	38.8

표 3. KITTI mAP 성능 비교

표 3은 KITTI 데이터셋에서의 Vanilla와 DAT 방식의 성능 비교이다. 적은 데이터가 주어진 상황에서 DAT는 Vanilla 파인 튜닝 방식에 비해 여러가지 객체에서 월등한 mAP 성능을 보여준다. 두 데이터셋에서 진행한 실험을 통해 DAT 방식이 적은 데이터 상황에서 도메인 갭을 줄여주므로 백본 모델을 더 잘 활용하고 준수한 성능을 보여줌을 입증하였다.

3.2 정성적 비교

정량적인 비교 뿐만 아니라 정성적으로 Vanilla와 DAT 파인 튜닝을 비교하였다.

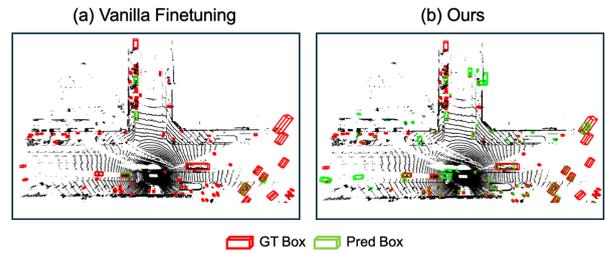


그림 4. 라이더 객체 인식 시각화 비교

위 그림은 Waymo 데이터셋에서 학습된 Vanilla (a)와 DAT (b) 객체 모델을 가지고 샘플 씬에서 예측한 결과 박스 (빨강)와 Ground Truth Box(초록)를 시각화한 결과이다. 그림 4에서 볼 수 있듯 DAT를 활용한 파인 튜닝 방식이 더 많이 예측을 진행하고 더 객체를 잘 잡는 것을 확인 할 수 있다.

IV. 결론 및 향후 연구 방향

본 연구에서는 타겟 도메인에서 파인 튜닝 시 레이블된 양질의 학습 데이터가 적은 상황에서 백본 모델의 특성을 유지하고 활용을 최대화하는 DAT 파인 튜닝 기법을 제시한다. DAT는 Teacher Student 증류 프레임워크에 기반하여 백본과 타겟 도메인 사이의 갭을 완화해준다. 여러 실험을 통하여 벤치마크 데이터셋들에서 성능 향상을 이뤄내므로 본 연구 방법의 유용성을 보여주었다.

참고문헌

- [1] Yuan, Jiakang, et al. "Ad-pt: Autonomous driving pre-training with large-scale point cloud dataset." *Advances in Neural Information Processing Systems* 36 (2024).
- [2] Mao, Jiageng, et al. "One million scenes for autonomous driving: Once dataset." *arXiv preprint arXiv:2106.11037* (2021).
- [3] Sun, Pei, et al. "Scalability in perception for autonomous driving: Waymo open dataset." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
- [4] Geiger, Andreas, et al. "Vision meets robotics: The kitti dataset." *The International Journal of Robotics Research* 32.11 (2013): 1231-1237.
- [5] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [6] Yan, Yan, Yuxing Mao, and Bo Li. "Second: Sparsely embedded convolutional detection." *Sensors* 18.10 (2018): 3337.