

# 시뮬레이션과 실제 환경 사이의 연결: 자율 주행을 위한 객체 탐지 성능 향상

김성식, 조현아, 이명학, 이재구\*

국민대학교

\*jaekoo@kookmin.ac.kr

## Bridging the Gap between Simulation and Real-world: Enhancing Object Detection in Autonomous Vehicles

Sungsik Kim, Hyunae Cho, Myunghak Lee, Jaekoo Lee\*

College of Computer Science, Kookmin University.

### 요약

안정적인 자율 주행 기술을 위해 주변 객체를 탐지할 수 있는 능력은 필수적이다. 이를 위해 공개된 데이터셋을 사용하는 것은 합리적이지만, 때때로 영역 격차나 극단적인 상황의 부족으로 인해 실패를 겪을 수 있다. 하지만 이를 위해 모든 영역과 상황에 대해 데이터셋을 만드는 것은 매우 어려운 일이다. 이러한 문제를 해결하고자, 시뮬레이션을 통해 주행 환경에서 객체 탐지 일반화 성능을 높이는 프레임워크를 제안한다. 우선, 실제 환경의 디지털 트윈을 시뮬레이션 상에서 제작하여 데이터셋을 제작한다. 다음으로 시뮬레이션과 실제 환경 간의 전이 학습을 수행하여 성능이 증가함을 실험을 통해 보였다. 마지막으로, 비지도 학습 기반으로도 모델이 준수한 성능을 보장할 수 있음을 입증하였다.

### I. 서론

자율 주행(autonomous driving) 차량은 운전자의 개입 없이 주행을 해야 하므로 넓은 범위의 기술과 능력을 갖고 있어야 한다. 이를 위한 가장 중요한 과업 중 하나는 객체 탐지(object detection)이며, 객체 탐지 모델을 학습시키기 위한 주행 데이터셋은 많이 공개되어 있다. 하지만, 이러한 데이터셋들은 보통 제한된 지역과 조건에서 수집되고 상황이 제한적이기 때문에, 학습된 모델이 새로운 환경과 상황에서는 강건하지 않을 수 있다. 그렇다고 해서 수많은 환경과 상황마다 각각 별도로 데이터셋을 수집하는 것은 비용이 매우 많이 드는 일이다.

그와 반면에, [그림 1]과 같이 특정 환경을 시뮬레이션 상에서 표현할 수 있다면 그 안에서 데이터를 무한하게 생성하고, 여러 시나리오를 가정하며, 라벨 또한 사람의 노동 없이 자동으로 생성할 수 있다. 따라서 시뮬레이션을 활용하는 것은 그렇지 않은 경우와 비교했을 경우 엄청난 장점을 가지고 있기 때문에 활용 가치는 풍부하다. 본 논문은 시뮬레이션을 자율 주행을 위한 객체 탐지 분야에 적용하여 그 성능을 향상할 수 있는 방법을 모색하고 이를 프레임워크로 제시한다.

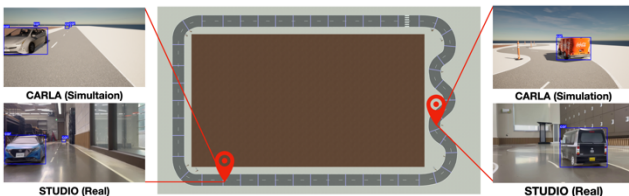


그림 1. 실제 환경과 디지털 트윈한 시뮬레이션의 장면 비교

### II. 본론

[그림 2-(a)]는 프레임워크를 위한 두 가지 데이터셋을 생성하는 과정을 표현한 것이다. 먼저, [그림 2-(a)] 위쪽에 해당하는 부분은 시뮬레이션으로 소스 데이터셋을 만드는 과정이다. 실제 주행 환경이 있을 경우 이를 시뮬레이션

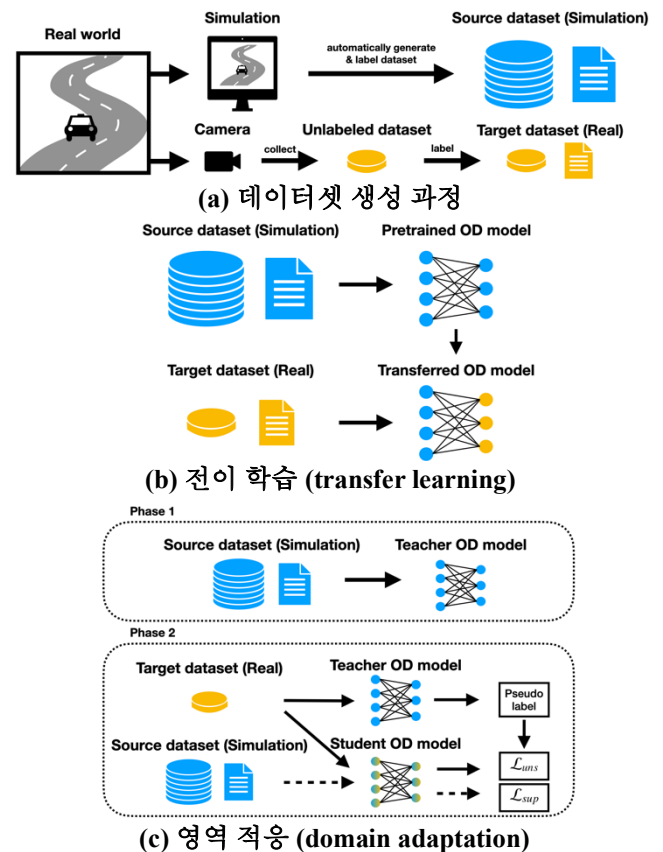


그림 2. 제작한 자율 주행 시뮬레이션 학습 프레임워크

환경에서 최대한 동일하게 표현한다면, 데이터셋을 자동으로 대량 생성할 수 있다. 다음으로 [그림 2-(a)] 아래쪽은 모델이 학습하고자 하는 타겟 데이터셋을 생성하는 과정이다. 이를 위해서 카메라를 이용해 주행 영상을 촬영해야 하며, 수작업으로 라벨을 다는 과정까지 수행하면 라벨을 포함한 데이터셋을 최종적으로 얻을 수 있다.

[그림 2-(b)] 는 생성한 시물레이션 데이터셋으로 객체 탐지 모델을 사전 학습(pre-training) 한 후 라벨링 된 타겟 데이터 셋으로 전이 학습(transfer learning)을 수행한 것이다.

[그림 2-(c)]는 영역 적응(domain adaptation) 방식 중 교사-학생(teacher-student) 프레임워크 과정을 그림으로 나타낸 것이다. 우선 Phase 1에서는 소스 데이터셋을 활용하여 교사 모델을 사전 학습한다. 다음으로 Phase 2는 먼저 타겟 데이터셋을 사전 학습된 교사 모델에 입력하여 수도 라벨을 얻는다. 학생 모델은 소스 데이터셋에 대한 출력과 라벨을 비교하여  $\mathcal{L}_{sup}$ 를 구하며, 타겟 데이터셋에 대한 출력을 통해 앞서 구한 수도 라벨과 비교하여  $\mathcal{L}_{uns}$ 를 계산한다. 두 가지 손실 함수를 통해 학생 모델을 학습하여 최종적으로 사용하게 된다.

### III. 실험

본 논문에서는 데이터를 수집하고 시물레이션 환경으로 재구성하기 쉽도록 하기 위해, 트랙이 배치된 실내 스튜디오에서 데이터를 수집하였다. 이를 STUDIO 데이터셋이라 하며 총 1,010 장으로 구성하였다. 또한, CARLA[1] 시뮬레이터에서, 실내 스튜디오와 동일한 비율로 맵을 재구성하여 실제 환경 데이터셋을 제작하였다. 이를 CARLA 데이터셋이라 부르며 총 37,501 장의 사진들과 라벨을 자동으로 추출하였다.

성능 평가 기준으로는 IoU(Intersection over Union) 기준이 0.5 인 AP<sub>50</sub>(Average Precision)과 같이 IoU 기준을 0.5 부터 0.95 까지 점차 증가시킨 후 측정된 AP의 평균값인 AP<sub>50:95</sub>를 사용하였다.

우리는 객체 탐지 모델인 Yolo[2]와 Faster R-CNN[3]을 가지고 4 가지 공통 실험을 진행하였다. 우선 베이스라인 모델의 성능을 평가하기 위해 CARLA 와 STUDIO 데이터셋에 대해 학습하고 성능을 측정하였다. 다음으로, 두 데이터셋 간의 영역 격차를 측정하기 위해 CARLA 에서 학습한 모델을 추가적인 학습 없이 STUDIO 데이터셋에 대해 시험하였다. 또한, 전이 학습의 효과를 보기 위해 CARLA 에서 사전 학습하고 STUDIO 에서 전이 학습한 모델을 STUDIO 데이터셋에 대해 성능을 확인하였다. 마지막으로, Faster R-CNN 의 경우, 영역 적응 방식을 실험하고자 Adaptive Teacher[4] 을 적용하여 실험을 진행하였다.

[표 1]에서 볼 수 있듯이, Yolo[2]의 경우 전이 학습 여부에 따라 AP<sub>50</sub>은 2.2%, AP<sub>50:95</sub>는 15.9% 상승하였다. 상승폭이 크지 않게 느껴질 수 있지만, AP<sub>50</sub>의 경우 99.4%로 오탐이 거의 없었으며, AP<sub>50:95</sub> 또한 높은 수치를 보였다. 주변 객체를 탐지하지 못하는 경우가 매우 치명적인 자율 주행과 같은 분야에서는 오탐을 최대한 줄이는 것이 중요하다. 따라서 전이 학습을 통한 성능 향상은 충분히 의미가 있으며, 추후 넓은 활용 가능성을 입증했다고 할 수 있다.

[표 2]에서 CARLA 데이터셋에서 사전 학습을 충분히 진행한 Faster R-CNN[4] 모델도 STUDIO 데이터셋에 대한 어떠한 정보를 얻지 못한 경우에는 AP<sub>50</sub>은 18.4%, AP<sub>50:95</sub>는 7.5%로 매우 낮은 성능을 기록하였다. 반면, 영역 적응 기법을 사용한 결과 대략 AP<sub>50</sub>은 303%, AP<sub>50:95</sub>는 484% 대폭 증가하였다. 데이터의 정답을 알 수 없는

표 1. Yolo 모델 실험 결과

Method	Train	Test	AP <sub>50</sub>	AP <sub>50:95</sub>
Baseline	CARLA	CARLA	99.0	90.6
Baseline	STUDIO	STUDIO	97.2	71.6
Baseline	CARLA	STUDIO	6.4	3.7
Transfer learning	CARLA → STUDIO	STUDIO	99.4	83.0

표 2. Faster R-CNN 실험 결과

Method	Train	Test	AP <sub>50</sub>	AP <sub>50:95</sub>
Baseline	CARLA	CARLA	95.0	76.8
Baseline	STUDIO	STUDIO	98.0	75.0
Baseline	CARLA	STUDIO	18.4	7.5
Transfer learning	CARLA → STUDIO	STUDIO	99.6	78.2
Domain adaptation	CARLA → STUDIO	STUDIO	74.2	43.3

상황에서도 큰 발전을 이끌어냈으며, 영역 적응 분야가 발전함에 따라 해당 방식을 활용한 연구도 활발해질 것이다.

### IV. 결론

본 논문에서는 시물레이션 환경 상에서 원하는 영역과 주행 상황을 직접 구축하고 데이터셋을 제작하여 객체 인식 성능을 높일 수 있는 프레임워크를 제시한다. 해당 프레임워크에 대한 실험을 통해 시물레이션이 긍정적인 효과를 가져다 줄 수 있음을 증명하였다. 또한, 타겟 영역에서 라벨이 전혀 없는 상황에서도 현존하는 영역 적응 기법을 활용하는 것도 고려할만한 선택지임을 실험적으로 입증하였다. 추후 우리는 해당 프레임워크를 객체 탐지 뿐만 아니라 차선 인식(lane segmentation) 과업에도 적용하는 연구를 진행할 것이다.

### ACKNOWLEDGMENT

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.RS-2022-00167194,미션 크리티컬 시스템을 위한 신뢰 가능한 인공지능)

### 참 고 문 헌

- [1] Dosovitskiy, Alexey, et al. "CARLA: An open urban driving simulator." Conference on robot learning. PMLR, 2017.
- [2] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767, 2018.
- [3] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.
- [4] Li, Yu-Jhe, et al. "Cross-Domain Adaptive Teacher for Object Detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.