

해외포커스

자율주행 기술 고도화를 위한 국내외 자율주행 데이터셋 현황

김형주

차세대융합기술연구원 경기도자율주행센터 선임연구원
hyungjoo@snu.ac.kr

서론

정부는 ‘2030 미래차 산업 발전 전략’을 통해 27년 전국 주요 도로에서 레벨 4 자율주행차 상용화를 목표로 하고 있다. 이를 성공적으로 달성하기 위해서는 자율주행 차량의 학습과 성능 향상을 위한 학습 데이터셋이 중요하다. 자율주행은 라이다, 카메라, 레이더 등 다양한 센서 기반으로 주변 상황을 인지·판단하고 제어함으로써 차량 스스로 목적지까지 운행하는 기술로 각 단계별로 레벨 0(비자동화), 레벨 1(운전자 보조), 레벨 2(부분 자동화), 레벨 3(조건부 자율주행), 레벨 4(고도 자율주행), 레벨 5(완전 자율주행)까지 구분된다(Society of Automotive Engineers). 각 레벨의 차이점은 주로 제어의 주체와 주행 책임에서 구분되며 레벨 0~2는 운전자가 주행의 주요 책임을 지고 시스템은 주행을 보조하는 역할을 한다. 반면, 레벨 3부터는 차량이 주행의 상당 부분 또는 전부를 자동으로 처리할 수 있으며, 레벨이 올라갈수록 인간 운전자의 개입이 점점 더 줄어들게 된다.

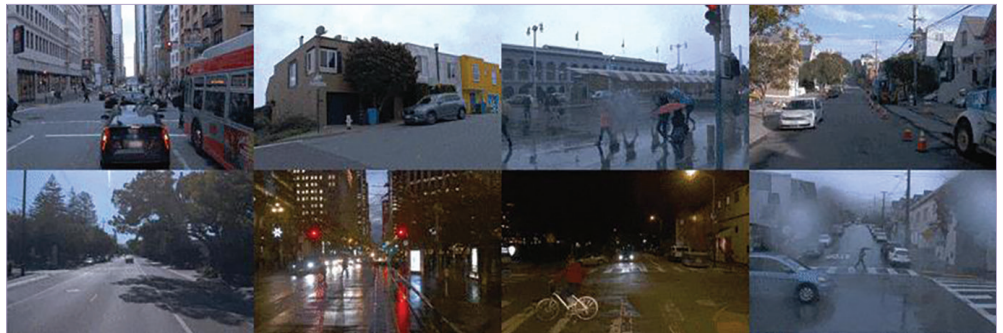
이렇듯 자율주행 기술 레벨이 고도화됨에 따라 차량이 복잡한 도로 환경에서 더욱 안전하고 효과적으로 주행할 수 있도록 다양한 센서 기반의 인공지능 기술이 요구된다. 라이다, 카메라, 레이더 등의 센서는 차량 주변 환경의 정밀한 정보를 수집하여, 차량 주변 환경을 인지하고 적절한 주행 결정을 내린다. 이러한 고도화된 사물 인식 기능을 개발하고 최적화하기 위해서는 방대한 양의 학습 데이터셋이 필요하다. 학습 데이터셋은 자율주행 차량의 인공지능 시스템이 인지, 판단, 제어 등의 과정을 통해 학습하고 개선할 수 있도록 다양한 환경과 주행 조건에서 수집된 데이터의 집합이다. 이러한 데이터셋은 주로 차량의 센서에 의해 수집되며, 차량과 주변 환경에 대한 정보를 포함한다. 대표적으로 차량위치, 주변차량 및 보행자위치, 도로상황, 교통신호상태, 기상조건 등이 포함된다. 따라서, 본고에서는 국내외 자율주행 데이터셋에 대한 사례조사를 토대로 자율주행 차량의 학습과 성능 향상을 위한 학습 데이터셋의 중요성을 제시하고자 한다.

해외 자율주행 데이터셋 사례

대표적인 해외 자율주행 학습 데이터셋으로는 Waymo Open Dataset이 존재한다. 알파벳 자회사인 웨이모가 2019년에 공개한 자율주행 데이터셋은 25개 도시에서 약 1,000만 마일에 달하는 자율주행 테스트를 진행하며 데이터를 수집하였다.

데이터셋에는 인지 데이터와 모션 데이터로 구성되어 있으며, 인지 데이터는 10Hz로 수집된 각 20개의 세그먼트 2,030개로 구성되어 있다. 모션 데이터의 경우 10Hz에서 20초 동안의 객체 트랙과 세그먼트가 포함하는 영역에 대한 지도 데이터를 포함하는 103,354개의 세그먼트(2천만 프레임 이상), 574시간의 데이터로 구성되어 있다. 이러한 연속주행 영상은 다양한 도로 환경에서 객체의 행동을 추적하고 예측하는 모델을 개발하는 데 활용될 수 있다. 특히 Waymo 데이터셋을 활용하여 20~23년까지 Waymo Open Dataset Challenge를 개최하여 자율주행 인지·판단·제어 분야별 기술 고도화에 기여하고 있다.

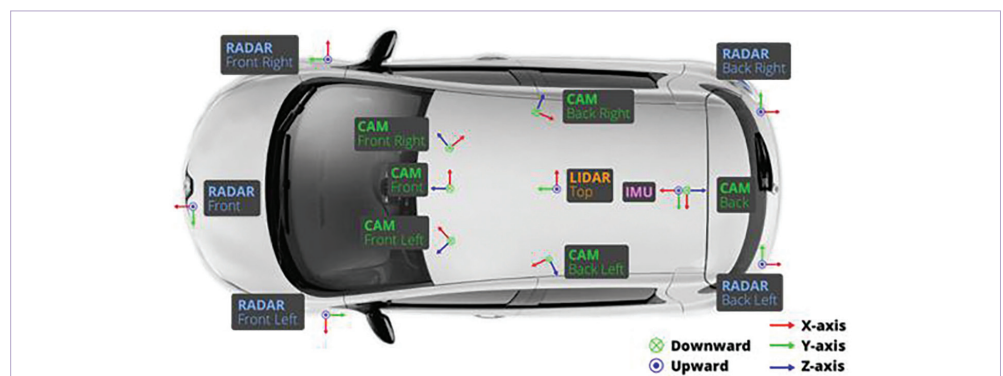
다양한 시공간적 범위를 포함하는 Waymo Open Dataset



자료: <https://waymo.com/open/>

두 번째로 Motional에서 개발한 nuScenes 자율주행 데이터셋은 3D 객체 주석을 갖춘 대규모 공개 자율주행 데이터셋이다. 데이터셋에는 Motional 테스트 차량으로 교통이 혼잡한 미국 보스턴과 싱가포르 도로구간에서 수집한 140만 개의 카메라 이미지, 39만 개의 라이다 정보, 140만 개의 레이더 정보, 140만 개의 object bounding box가 포함되어 있다. 주행 장면은 센서 당 2Hz로 수집된 20초 영상으로 물체 단위로 23개의 클래스와 8개의 속성으로 주석이 달려 있으며 다양한 교통상황과 위험상황을 포함한다. 수집된 데이터셋은 교통량이 많고 위험상황이 많이 포함되어 복잡한 시나리오 상황을 연구하기 위해 사용되고 있다. 특히 nuPlan Planning Challenge를 개최하여 nuScenes 데이터셋을 활용한 자율주행 차량별 인지·판단·제어 기술을 평가하여 자율주행 시스템의 안전성 검증에 기여하고 있다.

Motional 데이터 수집 차량



자료: <https://www.nuscenes.org/nuscenes>



경기도 판교 자율주행 데이터셋 사례

자율주행 인공지능 학습 데이터셋의 국내 대표사례는 경기도 판교제로시티 자율주행차 시범운행지구를 들 수 있다. 판교 자율주행 데이터셋은 국내 기업 및 연구기관의 자율주행 기술 개발에 공공데이터로 활용되고 있으며 국내 자율주행 산업생태계 조성 및 기술 상용화에 기여하고 있다. 국내 최초 자율주행 실증단지인 판교제로시티는 제1, 2판교 테크노밸리 지역으로 일반차와 자율주행차가 혼재되어 운영되고 있다. 특히 통합관제센터를 중심으로 자율주행 차량과 도로인프라 정보 공유를 위한 통신환경이 구축되어 있으며, 대표적으로 신호현시 서비스, 보행자케어 서비스, 자율주행 도로감시 서비스, 도로환경 감시 서비스 등이 포함된다.

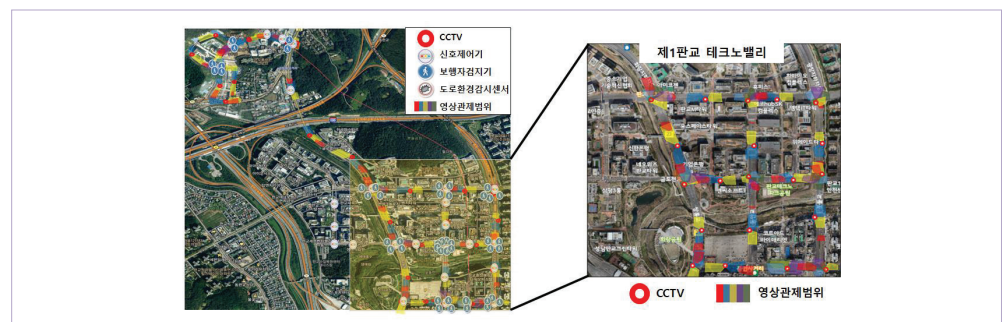
자율주행 데이터셋 구축을 위한 차량플랫폼은 제로셔틀, 스타렉스 개조차량, 판타G버스 등이 있으며 각 차량플랫폼은 라이다, 카메라, 레이더 등의 센서가 장착되어 다양한 도로환경, 교통상황, 날씨조건 등의 판교 데이터셋을 구축 중이다. 인프라 관점에서는 CCTV, 신호제어기, 보행자감지기, 도로환경감시센서 등이 구축되어 실시간으로 데이터를 수집하고 있다.

판교 자율주행 차량 데이터 수집

구분	자율주행 제로셔틀	스타렉스 개조차량	자율주행 판타G버스
차량			
실증내용	일반인 탑승 운행 및 자율주행 실증 연구	자율주행 데이터 수집 및 국산센서 개발지원	일반인 탑승 운행
센서구성	LiDAR(16ch 1개, 32ch 1개, 2D 4개), Radar 2개, Camera 4개 등	LiDAR(16ch 4개, 32ch 1개, 128ch 1개), Radar 2개, Camera 6개 등	LiDAR(16ch 1개, 32ch 4개), Radar 2개, Camera 8개 등

자료: 경기도자율주행센터

판교 자율주행 인프라 데이터 수집

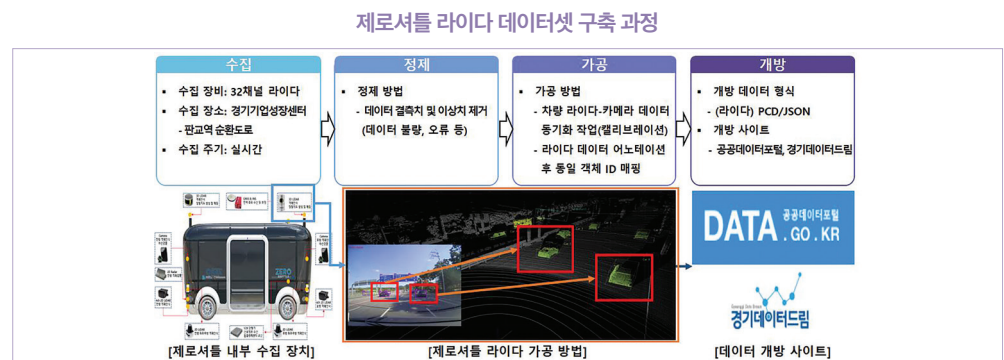


자료: 경기도자율주행센터

※ 경기도 판교 자율주행 데이터 개방 현황 : 약 549만 건 데이터 개방

- 2021년 개방데이터 약 4.5만 건
 - 공공데이터 포털(<https://www.data.go.kr/>) 약 4.5만 건
- 2022년 개방데이터 약 215만 건
 - 공공데이터 포털 약 190만 건
 - 공간융합 빅데이터 플랫폼(<https://bigdata-geo.kr/>) 약 25만 건
- 2023년 개방데이터 약 330만 건
 - AI Hub(<https://www.aihub.or.kr/>) 약 39만 건
 - 공간융합 빅데이터 플랫폼 약 291만 건

판교 자율주행 데이터셋은 판교제로시티에서 수집되는 자율주행차 및 인프라 센서데이터를 가공 및 표준화하여 공공데이터로 개방되고 있다. 차량 관점에는 라이다, 카메라, GPS/INS, 객체위치인식정보, 차량제어정보, 차량상태정보, 기본안전메세지 등이 포함되며, 수집/정제/가공 과정을 통해서 공공데이터로 개방된다. 예를 들어 라이다 데이터 가공과정에서는 차량 라이다-카메라 데이터 캘리브레이션 및 어노테이션 작업 등이 포함된다. 인프라 관점에는 CCTV, 라이다, 레이더, 도로노면 감시정보, 자율주행 도로감시 서비스 결과, 보행자케어 감지내역 등이 포함된다. 인프라 데이터의 경우도 수집/정제/가공 과정을 거치며 CCTV 데이터 내 개인정보 비식별화와 객체 어노테이션 작업 등이 포함된다. 특히 판교 자율주행 데이터를 활용한 5개 도전분야(2D 객체검지, 3D 객체검지, CCTV 혼잡도 분류, 자율주행차량 경로추정, 자유주제)에 대한 경진대회("23.9-12월)를 개최하여 실제 현장데이터를 활용하는 경험과 자율주행 기술에 대한 이해를 심화시키는데 기여했다.



자료: 경기도자율주행센터

결론

자율주행 차량이 실도로의 복잡하고 예측 불가능한 도로환경에서 안전하게 주행하는데 신뢰할 수 있고 다양한 시나리오를 포함하는 학습 데이터셋은 필수적이다. 자율주행 기술 고도화를 위한 다양한 도로 및 교통상황, 위험상황 등에 대한 데이터셋이 요구됨과 동시에 자율주행 시스템의 학습 및 테스트에 필요한 데이터의 양과 품질이 점차 중요해지는 시점이다. 특히 국내에서 부족한 실도로 환경의 도로종류별/교통상황별 위험상황 데이터셋 구축과 오픈소스 데이터셋의 접근성 개선, 그리고 경진대회를 통한 자율주행 데이터 활용도 제고 등이 필요하다. 이를 위해서 정부, 산업체, 연구기관 등이 협력하여 자율주행 데이터셋의 수집, 구축, 공유, 활용방안 수립이 요구된다.

참고문헌

1. 미국 자동차공학회 SAE(Society of Automotive Engineers) J3016
2. Waymo Open Dataset, <https://waymo.com/open/>
3. nuScenes 자율주행 데이터셋, <https://www.nuscenes.org/nuscenes>
4. 경기도자율주행센터, <https://ggzerocity.or.kr/>