

●○ 도로상태 및 자율버스 과제

자율주행버스 개발을 위한 노선 주행 이미지



●○ 개요: 자율주행버스 개발을 위한 노선 주행 이미지 데이터셋이란?

본 데이터셋은 일반 도로에서 고정 노선으로 운행하는 대중 교통 서비스에 적용 가능한 학습데이터로 건국대학교 산학협력단에서 주관하여 구축하였다. 이는 주행 영상 데이터와 GNSS 데이터의 결합된 형태로, 10개 이상의 노선에 대한 300시간의 동영상 데이터를 구축한 것이며 100만장 이상의 AI 학습용 데이터를 제공한다.

영상 인식 기술은 주변 객체, 도로 상태 등의 주행환경을 스스로 인지하여 차량을 안전하게 주행하게 하기 위한 기술이다. 차량이 직면할 수 있는 다양한 주행환경 (주간/야간, 맑음/흐림/비·눈 등) 혹은 악조건 속에서도 일정 수준 이상의 인식성능을 유지하기 위해 다양한 주행 데이터셋을 통해 AI를 학습하여 해결하려는 기술연구 및 상용화가 진행되고 있다.

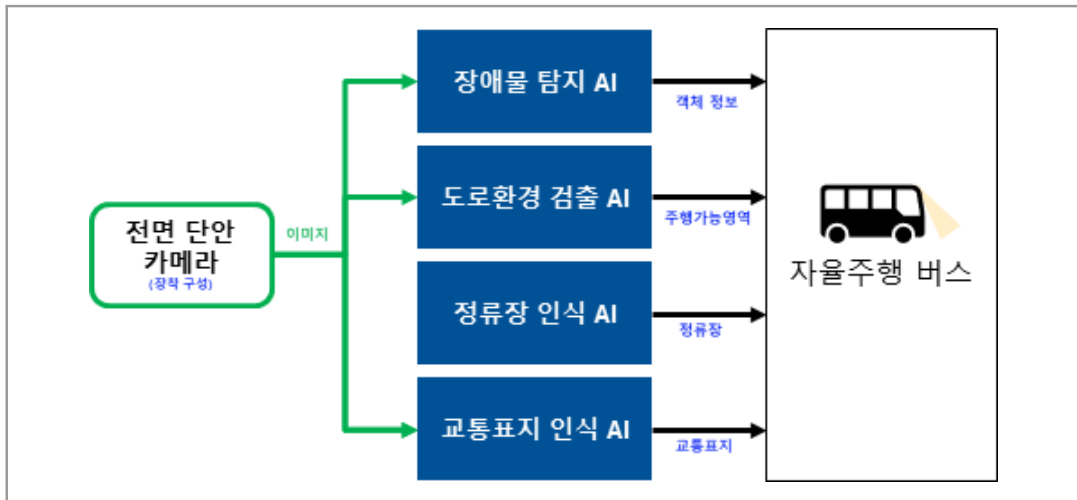


그림1 | 이미지 기반 객체 인식 AI를 활용한 자율주행 시스템 구조

●○ 데이터셋의 구성

본 데이터셋 구성에 사용된 센서는 이미지 학습을 위한 단안 카메라와 기준 위치정보 제공을 위한 GPS이다. 이 중 학습 데이터로 구성되는 부분은 이미지 프레임으로, 이미지에 대한 객체 바운딩 박스와 객체 시멘틱 세그멘테이션이다. 자세한 구성은 아래의 표와 같다.

이미지 프레임		
객체 바운딩박스	동적 객체	차
		트럭
		버스
		자전거
		오토바이
		보행자
		자전거 탑승자
		오토바이 탑승자
		기타 객체
	교통 표지	교통표지판
	신호등	
버스 정류장	셸터형 정류장	
	표지판형 정류장	
객체 시멘틱 세그멘테이션	도로 환경	버스전용차로
		전용차로 외 주행 가능 영역
		주행 불가 영역

●○ 데이터셋의 설계 기준과 분포

데이터셋을 설계할 때 가장 중요하게 고려했던 점은 데이터 밸런스이다. 버스노선 주행에서 검출 가능한 다양한 수집항목 중 기 구축 데이터셋의 기준 항목을 고려하여 적절한 수집항목을 만들었고, 외부환경(날씨)과 촬영시간대를 다양화하여 학습 모델이 환경적인 요인에 영향을 덜 받을 수 있도록 데이터 편향성을 최소화하도록 했다.



그림2 | 데이터셋 구성 개요

데이터 카테고리, 외부 환경, 촬영시간대의 설계기준과 분포는 다음과 같다.

- 데이터 카테고리: Cityscapes 등 기존 구축 데이터셋의 기준 항목을 바탕으로, 크게 동적 객체(차량, 사람 등), 교통 표지(교통표지판, 신호등 등), 버스 정류장(쉘터형, 표지판형), 도로환경(주행 가능/불가능 영역) 등 고루 포함되도록 했다.
- 외부 환경: 다양한 날씨 정보를 포함하고 있어야 학습 모델이 환경적인 요인에 영향을 덜 받으므로, 맑음 50%, 흐림 40%, 강우/강설 10% 으로 구성하였다.
- 촬영시간대: 주간 및 야간을 적절한 비율로 촬영해야 하므로 주간 최소 50% 이상, 야간 최소 10% 이상으로 구성하였다.

●○ 데이터 구조

데이터의 샘플 구조는 아래와 같다.

- FHD급 카메라의 MP4 데이터 샘플 구조

구조	내용
Video Data	H.264
해상도	1920x1080
Format	YUV420(NV12)

• GNSS 데이터의 샘플 구조

구조	형식	내용
Time	hhmmss.ss	UTC 시간
status	character	A(Data유효함)
latitude	ddmm.mmmm	위도(각도&분)
NS	character	North/South
longitude	ddmm.mmmm	경도(각도&분)
EW	character	East/West
spd	numeric	속도
date	ddmmyy	일,월,년

●○ 데이터 예시

객체 바운딩박스(Bounding box) 데이터에 대한 예시이다.

- 각 파일마다 속성 값으로 데이터를 취득한 시간, 촬영시간대, 설명 등 기록한다.

```

{
  "info" : info,
  "images" : [image],
  "annotations" : [annotation],
  "licenses" : [license],
}

info{
  "year" : int,
  "version" : str,
  "description" : str,
  "contributor" : str,
  "url" : str,
  "date_created" : datetime,
}

image{
  "id" : int,
  "width" : int,
  "height" : int,
  "file_name" : str,
  "license" : int,
  "flickr_url" : str,
  "url" : str,
  "date_captured" : datetime,
}
    
```

```

}

license{
  "id" : int,
  "name" : str,
  "url" : str,
}

annotation{
  "id" : int,
  "image_id" : int,
  "category_id" : int,
  "segmentation" : RLE or [polygon],
  "area" : float,
  "bbox" : [x,y,width,height],
  "iscrowd" : 0 or 1,
}

categories[[
  "id" : int,
  "name" : str,
  "supercategory" : str,
]]

```

객체 시맨틱 세그멘테이션(Segmentation)에 대한 예시이다.

- 각 이미지 안의 분할된 객체들에 대한 class id를 픽셀로 갖는다.



그림3 | 이미지 형식의 segmentation mask예시

●○ 데이터 구축 과정

데이터 구축 과정은 데이터 적재, 데이터 전처리, 데이터 가공, 데이터 검증 절차를 통해 구축된다. 데이터 적재는 원천 데이터를 취득 및 저장 이후, 검수 절차를 통해 주행 중 정지 등 이유로 중복되는 데이터 또는 빗물에 의해 보이지 않는 사진 등 사용 불가능한 데이터 제거 절차를 거친다. 이후 개인 정보 비식별화 전처리 과정과 신뢰할만한 데이터를 얻기 위한 어노테이션 다중 수행을 통해 데이터를 가공하고, 데이터의 유효성 검사 및 학습 적용 테스트를 통해 데이터를 검증한다.



●○ 검수와 품질 확보

데이터셋의 품질을 보장하기 위해 아래 그림과 같이 검수 절차, 기준을 수립하고, 3개의 품질관리 조직을 구성하였다. 총 4차 샘플링 검수를 통해 데이터의 완성도를 증가시키고, 자동 입력, 작업자 검증 및 품질검사의 반복을 통해 기준에 따라 일관성 있는 품질을 확인한다. 데이터셋의 품질을 보장하기 위해 구축 단계별 검증, 데이터셋 유효성 확인, 품질보증 활동을 수행하는 3개의 팀을 구성하였으며 주요 활동은 모두 품질 검수를 담당하는 한국정보통신기술협회(TTA)와 상시 협업체계를 구축하여 데이터 가이드라인을 준수한다.

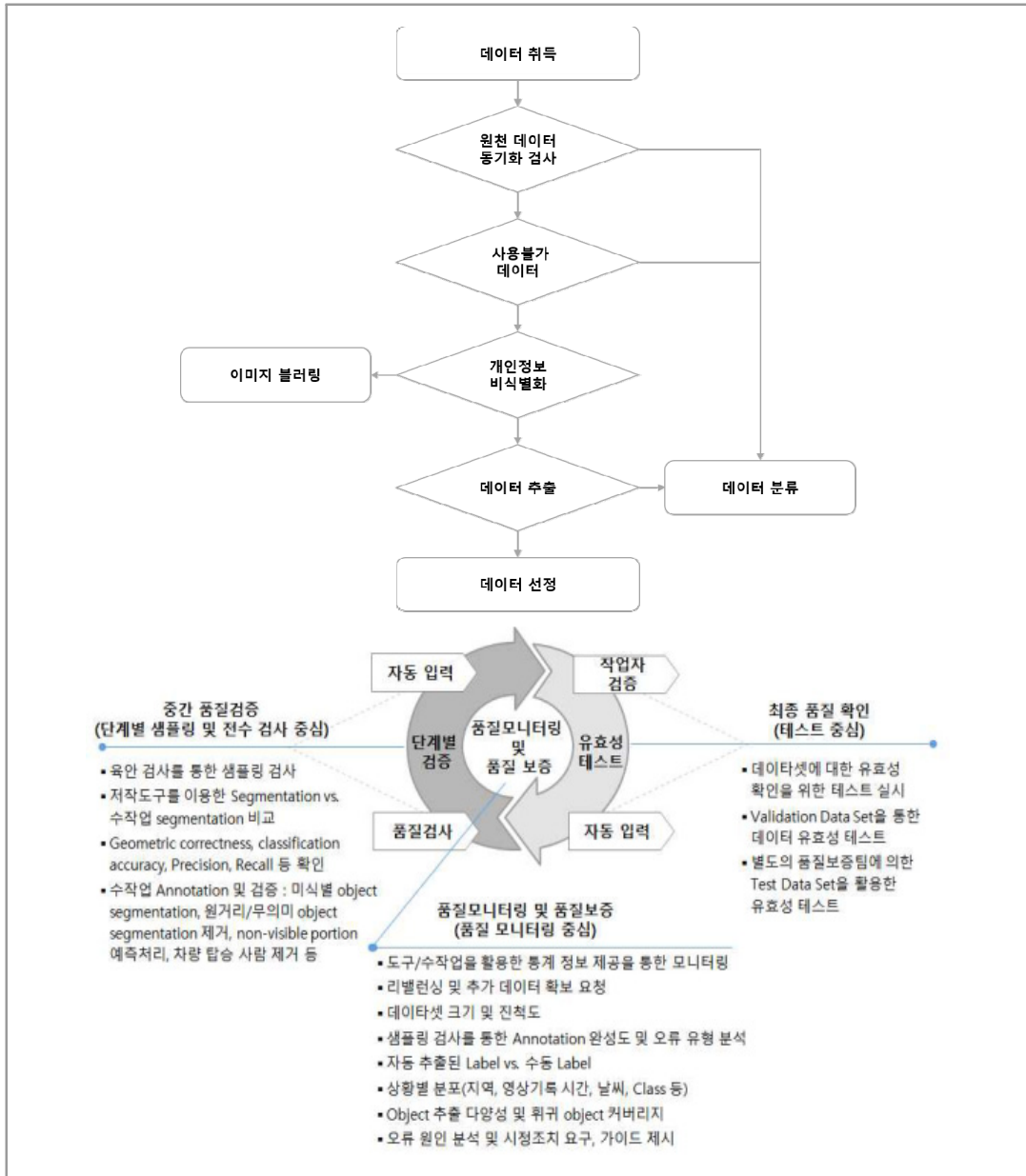


그림4 | 품질 확보를 위한 품질 검수 절차 및 체계

●○ 데이터 구축 담당자

수행기관(주관) : 건국대학교 산학협력단

(전화: 02-2049-6265, 이메일: kichun@konkuk.ac.kr)