




인공지능 데이터 구축·활용 가이드라인

- 자율주행버스 개발을 위한 노선 주행 이미지-

인공지능 데이터 구축	사업 총괄	
	데이터 설계	 
	원천데이터 수집 및 정제	  STARMATA
	데이터 가공	 STARMATA
	데이터 검수	  STARMATA SPACE
	클라우드 소싱	  STARMATA SPACE
	저작도구 개발	 STARMATA
	AI모델 개발	 
	응용 서비스 개발	
가이드라인 작성	건국대학교 산학협력단	김상권
	건국대학교 산학협력단	정지민
가이드라인 버전	Ver 1.1 2021. 02. 06	

목 차

1. 데이터 명세 정보	1
1.1 데이터 정보 요약	1
1.2 데이터 포맷	1
1.3 어노테이션 포맷	2
1.4 데이터 구성	3
1.5 데이터 통계	4
1.6 원시데이터 특성	5
1.7 기타 정보	5
2. 데이터 구축 가이드	6
2.1 데이터 구축 개요	6
2.2 문제정의	6
2.3 획득·정제	6
2.4 어노테이션/라벨링	7
2.5 검수	10
2.6 활용	10

1. 데이터 명세 정보


1.1 데이터 정보 요약

데이터 이름	자율주행버스 개발을 위한 노선 주행 이미지	
활용 분야	자율주행, 버스	
데이터 요약	일반 도로에서 고정 노선으로 운행하는 대중 교통 서비스에 적용할 수 있는 인공지능 학습을 위한 데이터	
데이터 출처	<ul style="list-style-type: none"> ● 데이터 수집 방법 <ul style="list-style-type: none"> - 10개 이상 노선을 운행하는 버스에 단안카메라, GNSS를 부착하여 다양한 환경 데이터 수집 ● 데이터 구축량 <ul style="list-style-type: none"> - 300시간 이상 다양한 환경에서의 버스 노선 주행 영상, 인공지능 학습데이터 100만장 이상 ● 데이터 분류 체계 <ul style="list-style-type: none"> - 동적 객체, 교통표지, 버스 정류장, 도로 환경 	
데이터 이력	배포버전	1.0
	개정이력	신규
	작성자/ 배포자	건국대학교 산학협력단/ 건국대학교 산학협력단

1.2 데이터 포맷

- 이미지 기반 객체 검출 및 세그멘테이션을 위한 단안카메라, GNSS 데이터로 구성

	예시	데이터 항목	데이터 형식
영상		원천 데이터	*.mp4
이미지		객체 영역 객체 종류 도로 영역 도로 종류	*.png *.json <pre data-bbox="1129 1525 1350 2029"> { "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, "lickr_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, } </pre>

			
GNSS	\$GPRMC,083559.00,A,4717.11437,N,00833.91522,E,0.004,091202	시간 상태 위도 North/South 경도 East/West 속도 날짜	*.txt

1.3 어노테이션 포맷

No	항목명	항목설명	타입	필수구분	단위
1	데이터셋 정보	데이터셋 전체에 관한 전반적인 정보를 포함하는 메타데이터	Object		
1	1-1 데이터셋 상세설명		str	Y	
	1-2 데이터셋 담당기관		str		
	1-3 데이터셋 생성일자		datetime	Y	
2	이미지 정보	데이터셋을 구성하는 각 이미지에 대한 메타데이터 및 학습 데이터	List		
2	2-1 이미지 식별자		int	Y	
	2-2 이미지 너비		int	Y	pixel
	2-3 이미지 높이		int	Y	pixel
	2-4 이미지 파일명		str	Y	
3	이미지 어노테이션 정보	데이터셋의 어노테이션에 대한 메타데이터 및 학습 데이터	List		
3	3-1 어노테이션 식별자		int	Y	
	3-2 연관 이미지 식별자		int	Y	
	3-3 어노테이션 카테고리		int	Y	
	3-4 세그멘테이션 정보		RLE or [polygon]	Y	
	3-5 세그멘테이션 넓이		float		
	3-6 바운딩박스 정보		[x,y,width,height]	Y	
	3-7 단일 개체/개체군 정보		0 or 1		
4	카테고리 정보	어노테이션의 카테고리 정보에 대한 메타데이터 및 학습 데이터	List		
4	4-1 카테고리 식별자		int	Y	
	4-2 카테고리 이름		str	Y	
	4-3 상위 카테고리		str		

1.4 데이터 구성

1.4.1 데이터 폴더 구성

○ 시간, 주행 코스, 방향, in-out 주행여부, 날씨에 따라 폴더를 생성하여 데이터 분류

- 예) 20201229_11_CW_in_D_S

1.4.2 데이터 취득에 관한 메타데이터

- 이미지 취득 장비, 포인트 클라우드 취득 장비 등 취득에 관한 기타 정보를 메타데이터로서 작성하여 '데이터셋상세설명'에 첨부하고자 함

1.4.3 데이터 Naming

- <영상촬영일>_<수정여부>_<코스>_<코스형태>_<촬영 시간>_<날씨>_<원본순서>_<캡처프레임> 순서로 표기
 - 예) 200819_R_16_CW_in_D_B_004_00006 (2020년 8월 19일에 촬영한 4번째 영상 6번째 Frame)

구분	이름	설명
Date	YY, MM, DD	년도, 월, 일
수정 여부	R	Raw(원본)
	E	Edited(수정)
코스	N	N번 코스
코스형태	CW or CCW	코스 주행 형태
	in or out	
원본 영상순서	N	N번째 영상
캡처 프레임	N	N번째 프레임
촬영 시간	D	Day(낮)
	E	Evening/Morning(일몰/출)
	N	Night(밤)
날씨	B	Bright (맑음)
	C	Cloudy (흐림)
	S	Snowy (눈)
	F	Foggy (안개)
	R	Rainy (비)

1.5 데이터 통계

1.5.1 데이터 구축 규모

- 영상 및 이미지 데이터의 규모 표현
 - 도로 및 전방 촬영 영상 300시간의 분량
 - 연속 촬영 영상(sequence) 형태로 구성된 촬영 이미지(1920*1080 png) 100만 장에 대해 객체 bounding box 및 segmentation mask 부착
- GNSS 데이터의 규모 표현
 - 도로 및 전방 촬영 이미지와 동기화된 절대 위치 정보 100만 건

1.5.2 데이터 분포

- 객체별 분포

이미지 프레임 바운딩박스 및 폴리곤	
동적 객체	차
	트럭
	버스
	이륜차 (자전거/오토바이)
	보행자
	기타 객체
교통 표지	교통표지판
	신호등
버스 정류장	쉘터형 정류장
	표지판형 정류장
도로 환경	버스전용차로
	전용차로 외 주행 가능 영역

○ 날씨별 분포

- 맑음, 흐림, 우천, 눈, 안개에 대해 각 30%, 30%, 20%, 10%, 10% 수집
 - 아래의 첨부된 첨부자료 (1), (2)를 바탕으로 날씨를 분류하고, 수집한 날씨 분포를 설정
 - 기상조건이 충족되지 않아 불가피하게 데이터 수집이 어려울 경우에는 충족 가능한 날씨 환경에서 최대한 편향되지 않은 데이터 취득

날씨 분류	기존용어	운량	비고
맑음	맑음	0~2할 또는 상층운 0~4할	
	구름조금	3~5할 또는 상층운 5~7할	대체로 맑음
흐림	구름많음	6~8할 또는 상층운 8~10할	대체로 흐림
	흐림	9~10할	

표 8 첨부자료 (1) - 강수가 없는 날에 대한 날씨 분류

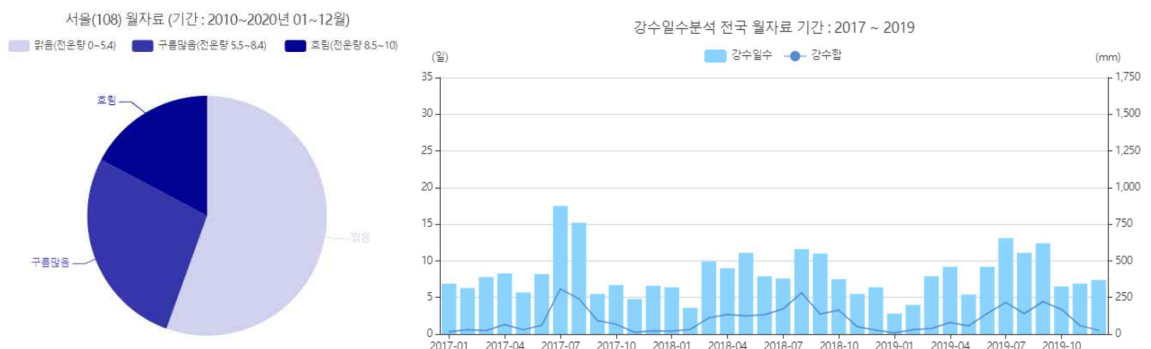


그림 20 첨부자료 (2) - 강수가 없는 날의 날씨 통계자료(왼쪽), 강수일 분석(오른쪽)

1.5.3 기타 활용 통계

(출처: Encoder-Decoder with Atrous SeparableConvolution for Semantic Image Segmentation, Liang-Chieh Chen et al. 2018,

RangeNet++: Fast and accurate LiDAR semantic segmentation, Milioto et al, 2019)

Backbone Decoder ASPP Image-Level	mIOU
X-65	77.33
X-65 ✓	78.79
X-65 ✓	79.14
X-71 ✓	79.55

(a) val set results

Method	Coarse	mIOU
ResNet-38 [83]	✓	80.6
PSPNet [24]	✓	81.2
Mapillary [80]	✓	82.0
DeepLabv3	✓	81.3
DeepLabv3+	✓	82.1

(b) test set results

Table 7. (a) DeepLabv3+ on the Cityscapes *val* set when trained with *train.fine* set. (b) DeepLabv3+ on Cityscapes *test* set. **Coarse:** Use *train.extra* set (coarse annotations) as well. Only a few top models are listed in this table.

TABLE I: IoU [%] on test set (sequences 11 to 21). RangeNet21 and RangeNet53 represent the new baselines with augmented Darknet backbones (21 and 53 respectively), and the versions with (++) are treated with our fast point cloud post-processing based on range.

Approach	Size	car	bicycle	motorcycle	truck	inter-vehicle	person	bicyclist	motorcyclist	total	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mean IoU	Standard	
Pointnet [14]	30000pts	186.2	1.3	0.3	0.1	0.8	0.2	0.2	0.0	61.6	15.8	35.7	1.4	41.4	12.9	31.0	4.6	17.6	2.4	3.7	14.6	2	
Pointnet++ [15]		23.7	1.9	0.2	0.9	0.2	0.9	1.0	0.0	72.9	18.7	41.8	5.6	62.3	16.9	36.5	13.8	30.0	6.0	8.9	20.1	0.1	
SPLatNet [10]		68.3	0.9	1.5	0.9	0.8	1.0	6.0	0.0	45.5	1.7	24.2	0.3	68.2	22.5	29.2	27.2	17.0	18.3	10.5	20.0	0.2	
SPLatNet [19]		69.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	79.4	0.8	41.2	0.0	68.7	27.8	24.3	35.9	35.8	13.8	0.0	22.8	1	
TangentConv [20]		86.8	1.3	12.7	11.6	10.2	17.1	20.2	0.5	82.9	15.2	61.7	9.0	82.8	44.2	75.5	42.5	55.5	30.2	22.2	35.9	0.3	
SqueezeSeg [21]		68.8	16.0	2.1	3.3	3.6	12.9	13.1	0.9	85.4	26.9	54.3	4.5	57.3	29.0	60.0	24.3	53.7	17.5	24.5	25.5	66	
SqueezeSeg-CRF [21]		68.3	18.1	5.1	4.1	4.8	16.5	17.3	1.2	84.9	28.4	54.7	0.4	61.2	29.2	59.6	25.5	54.7	11.2	36.3	30.8	55	
SqueezeSegV2 [22]		81.5	18.5	17.9	13.4	14.0	20.1	25.1	3.9	88.6	45.8	67.6	17.7	73.7	41.1	71.8	35.8	60.2	20.2	36.3	39.7	50	
SqueezeSegV2-CRF [22]		82.7	21.0	22.9	14.5	15.9	20.2	24.3	2.9	88.5	42.8	65.3	18.7	73.8	41.0	68.5	36.9	58.9	12.9	41.0	39.6	40	
RangeNet21 [Ours]		85.4	26.2	26.5	18.6	15.6	31.8	33.0	4.0	91.4	57.0	74.0	26.4	81.9	52.3	77.6	48.4	63.6	38.0	50.0	47.4	20	
RangeNet53 [Ours]		64 × 2048 px	186.4	24.5	32.7	25.5	22.6	36.2	33.6	4.7	91.8	64.8	74.6	27.9	84.1	55.0	78.3	50.1	64.0	38.9	52.2	49.9	13
		64 × 1024 px	184.6	20.0	25.3	24.5	17.3	27.5	27.7	7.1	90.4	61.8	72.1	22.8	80.4	50.9	75.1	46.0	62.7	33.4	45.4	45.4	25
		64 × 512 px	81.0	9.9	11.7	19.3	7.9	16.8	25.8	2.5	90.1	49.9	60.4	2.0	76.0	45.5	74.2	38.8	62.7	25.5	38.1	39.3	52
RangeNet53++ [Ours+kNN]		64 × 2048 px	91.4	25.7	34.4	25.7	23.0	38.3	38.8	4.8	91.8	65.0	75.2	27.8	87.4	58.6	80.5	55.1	64.6	47.9	55.9	52.2	12
		64 × 1024 px	90.5	20.6	27.1	25.2	17.6	29.6	34.2	7.1	90.4	52.2	72.7	22.8	85.9	53.3	71.7	52.0	63.7	43.8	47.4	48.0	21
		64 × 512 px	87.4	9.9	12.4	19.6	7.9	18.1	29.5	2.5	90.0	50.7	70.0	2.0	80.2	48.9	77.1	45.7	64.1	37.1	42.0	41.9	38

1.6 원시데이터 특성

1.6.1 대상분류

- 실제

1.6.2 제약조건

○ 일부 제약 있음

- 도로 주행 영상 수집 시 특정 객체가 존재하는 경우
- 특정 객체: 동적 객체, 교통 표지, 버스 전용 차선 등

1.6.3 속성

- 이미지
 - 1920*1080(FHD), 15FPS

1.7 기타정보

1.7.1 포괄성

- 10개 이상의 버스 전용도로 및 일반 도로 운행 노선 대상
- 시내, 교외 등을 모두 포함한 지역에서 취득하여 다양한 지역에 대한 데이터를 포함
- 상암 자율주행 테스트베드, 판교 제로시티 자율주행 실증단지 등의 자율주행 시범 서비스 지역 포함

1.7.2 독립성

- 사람 얼굴, 번호판 등 사생활 침해의 소지가 있는 민감 정보 정제 필요

1.7.3 유의사항

- 자율주행 시스템 적용을 통한 관련 기업들의 기술력 제고 및 AI 산업 관련 국가경쟁력 강화
- 데이터·AI 혁신기술 중심으로 산업구조 전환되는 경제적 파급효과
- 공공 데이터 제공을 통한 국민의 다양한 활용 및 사회 전반의 참여기회 제공

1.7.4 관련 연구

- GS3D: An Efficient 3D Object Detection Framework for Autonomous Driving, Buyu Li, Wanli Ouyang, Lu Sheng, Xingyu Zeng, Xiaogang Wang; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1019-1028
- A. Ćorović, V. Ilić, S. Đurić, M. Marijan and B. Pavković, "The Real-Time Detection of Traffic Participants Using YOLO Algorithm," 2018 26th Telecommunications Forum (TELFOR), Belgrade, 2018, pp. 1-4, doi: 10.1109/TELFOR.2018.8611986.
- H. Jeong, K. Park and Y. Ha, "Image Preprocessing for Efficient Training of YOLO Deep Learning Networks," 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, 2018, pp. 635-637, doi: 10.1109/BigComp.2018.00113.

2. 데이터 구축 가이드

2.1 데이터 구축 개요

2.1.1 데이터 수집

- 취득, 저장, 동기화 검수

2.1.2 데이터 정제

- 사용불가 데이터 분류
- 개인정보 비식별화

2.1.3 데이터 가공

- 단계별 어노테이션

2.1.4 데이터 검수

- 어노테이션 검사
- 유효성 검사

2.2 문제정의

2.2.1 임무 정의

- 고정 노선으로 운행하는 대중교통 서비스에 적용 가능한 학습 데이터 구축
 - 자율주행 버스 운영 지자체 등의 딥러닝 개발 인력들이 자유롭게 DB를 사용할 수 있도록 함
 - 주변 환경을 인지하며 정해진 노선을 자율적으로 주행하는 대중교통 서비스 개발 가능

2.2.2 데이터 구축 유의사항

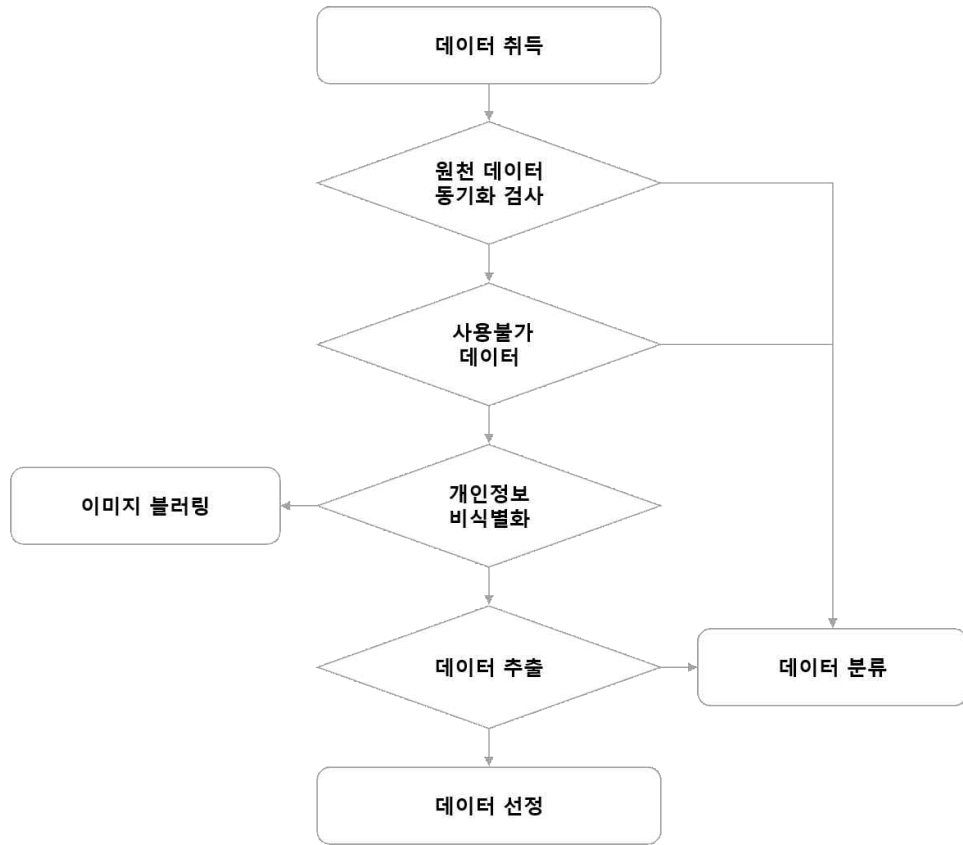
- 개인정보 비식별화
 - 모든 수집된 원천데이터 영상 및 학습데이터 영상에 대한 개인 초상권 및 차량번호 등 개인정보 비식별화
 - 육안으로 식별 가능한 얼굴 정보 비식별화
 - 육안으로 식별 가능한 차량 번호판 및 간판(상표 등) 비식별화

2.3 획득·정제

2.3.1 원시데이터 선정

- 원천 데이터의 수집대상은 일반 도로에서 고정 노선으로 운행하는 버스로 설정
- 원천 데이터의 수집은 수집대상인 버스에 데이터 수집용 센서를 부착한 후, 기존의 정해진 고정 노선을 반복적으로 운행하며 데이터를 취득함으로써 형성됨
- 원천 데이터 수집용 센서에는 RGB 영상 수집을 위한 전면 단안 카메라, 위성기반 절대위치 좌표 수집을 위한 GNSS 센서 등이 있음
- 원천 데이터 수집 시에는 버스를 운행하는 동안 발생할 수 있는 여러 가지 시나리오에 대한 학습용 데이터를 수집
 - 시간, 날씨 등
 - 도로 형태, 수집 구간 등

2.3.2 획득·정제 절차



2.3.3 획득·정제 기준

○ 사용할 수 없는 데이터 제거

- 햇빛 반사에 의한 영상
- 빗물에 의한 식별 불가 영상
- 신호 대기 등으로 인한 반복적으로 나타나는 동일 데이터 제거
 - 차량이 오래 멈춰있는 상태를 GPS 데이터 기반으로 판단하여 제거할 예정
 - 연속/중복된 데이터를 피하고자 최소 1m 이상의 프레임 이동거리를 가지도록 데이터 선별
- 신호 대기 등으로 인한 반복적으로 나타나는 동일 데이터 제거

기반으로 판단하여 제거할 예정

데이터 선별

○ 버스 등 상용 차량 외 일반 차량의 개인 식별 정보 비식별화

- 번호판 인식 툴 사용
 - 번호판의 색깔로 상용 차량 구분 (노란색)
 - 번호판의 한글 글씨로 상용 차량 구분 (바, 사, 아, 자 등 사용)
- 사람이 직접 각 데이터에 대해 모자이크 처리

○ 이미지 내 사람 얼굴 개인 식별 정보 비식별화

- 얼굴인식 알고리즘 사용
- 사람이 직접 각 데이터에 해 모자이크 처리

○ 데이터 중복 방지를 위한 적절한 프레임 분할 기준

- 도구를 통하여 자동적으로 영상의 중복성, 동일 위치에 해당하는 데이터 제거
- GPS 주행 속도를 확인하고 이동거리에 따라 후보 데이터 추천
- 수작업을 통하여 중복되거나 변화가 거의 없는 데이터 제거
- 수작업을 통하여 품질이 낮거나 오류가 있는 데이터 정제
- GPS 데이터를 활용하여 지역적 분포를 주기적으로 분석하여 지역적 편향성을 가지지 않도록 정제

2.4 어노테이션/라벨링

2.4.1 어노테이션/라벨링 절차

- 후처리 S/W를 통해 보정된 고정밀 GNSS 기준위치 데이터, 정제 단계에서 확인 가능한 왜곡된 이미지 제거 등을 통해 원천데이터에서 어노테이션 진행할 데이터 선정

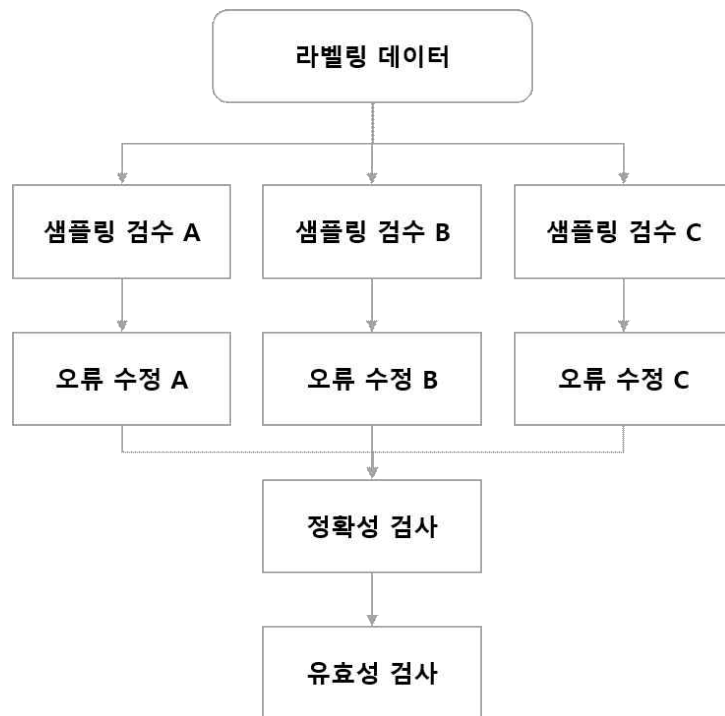
2.4.2 어노테이션/라벨링 기준

	레이블	세부내용
객체 바운딩 박스	동적 객체	▶ 차, 트럭, 버스, 보행자, 이륜차 등의 동적 객체 
	교통 표지	▶ 교통 표지판, 신호등 등의 교통 표지 
	버스 정류장	▶ 쉘터형 정류장, 표지판형 정류장(정류장은 지역마다 외관이 다양하고 특색이 있을 수 있으므로, 비슷한 노선이 아닌 다양한 지역을 주행하는 노선 선정)

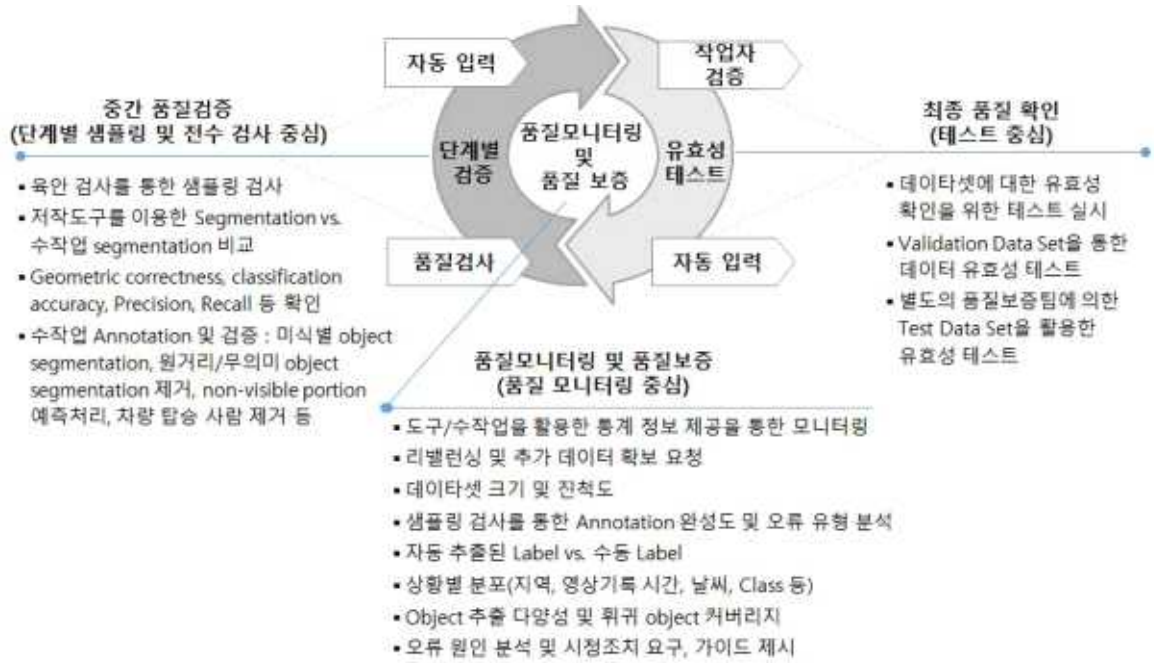
		
<p>객체 폴리곤</p>	<p>도로 환경</p>	<p>▶ 버스전용차로, 전용차로 외 주행가능 영역(버스 전용차로 및 비전용차로를 나타내는 노면 표시들이 지워진 구역도 있으므로 다양한 지역을 주행하는 노선을 선정하여 노면 표시 데이터 취득)</p> 

2.5 검수

2.5.1 검수 절차



2.5.2 검수 기준



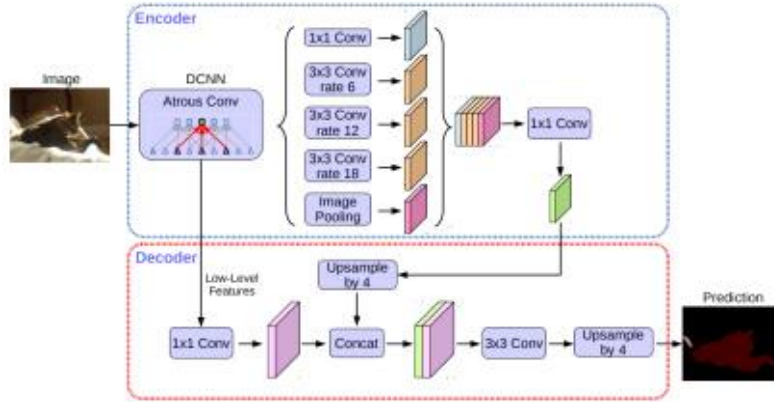
2.5.3 검수 조직

- 품질관리 조직은 데이터셋의 품질을 보장하기 위하여 구축 단계별 검증, 데이터 셋 유효성 확인, 품질보증 활동의 3가지 활동을 수행하는 3개의 팀(검수팀, 품질평가팀, 품질보증팀)으로 구성
 - 자체 가공된 데이터 및 클라우드 소싱을 통해 가공된 데이터의 검수 및 최종 annotation을 책임지는 검수팀
 - 구축된 데이터의 유효성과 모델의 정확도를 테스트하고 평가하는 품질평가팀
 - 전체적인 사업의 관리 및 품질관리 활동을 리드하고 플랫폼 기반의 인력별 작업배정, 공정관리, 모니터링, 품질측정을 통괄하는 품질보증팀
- 과제 추진 기간 동안 품질관리를 위해 품질 검수를 담당하는 한국정보통신기술협회(TTA)와 상시 협업체계 구축, 품질관리·검증 활동 협조, 데이터 가이드라인 수립·제출 등을 준수함
 - 사업 전반의 품질관리 및 검증을 위해 품질 책임자 지정 등의 품질관리 계획 수립
 - 인공지능 데이터 중간 및 최종산출물의 객관적 품질 지표와 정량 목표를 정의하고 TTA로부터 검증받음
 - 사업 종료 전까지 한국정보통신기술협회(TTA)의 데이터 품질검증 수행 및 품질검증 결과서 획득(검증에 필요한 환경, 도구 등을 제공) 예정임

2.6 활용

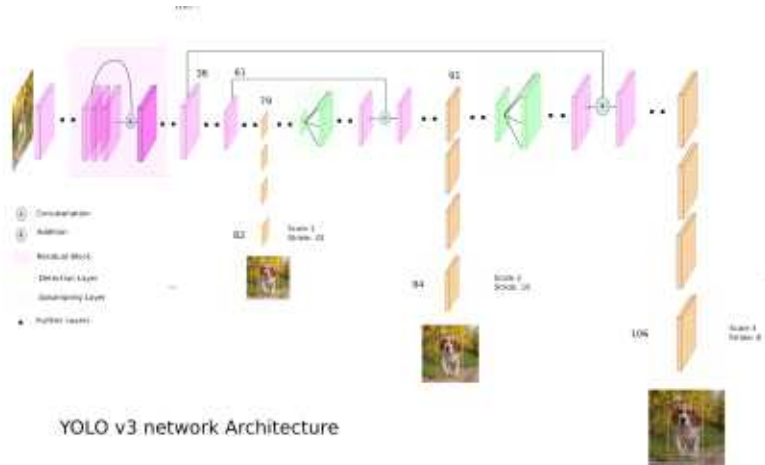
2.6.1 활용 모델

- 모델 학습
 - DeepLab V3+ 모델을 활용한 차선/차로 검출 알고리즘
 - 이미지 정보를 입력으로 받아 시맨틱 세그멘테이션 된 이미지 결과를 출력
 - 출력된 시맨틱 세그멘테이션을 바탕으로 차선을 구분하고 버스전용차로 구분 가능



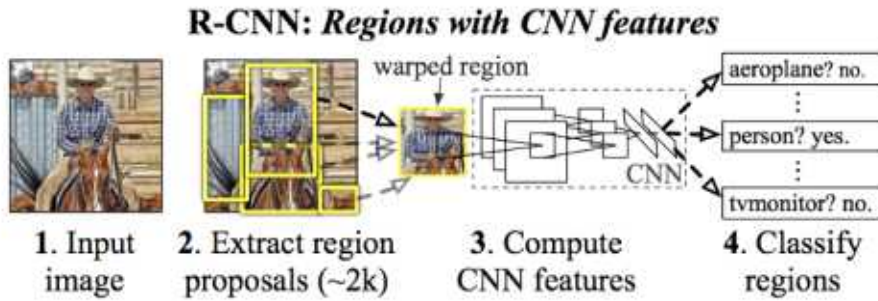
<DeepLabv3+ 모델 구조도>

- YOLOv3 모델을 활용한 정류장 인식 알고리즘
 - 이미지 정보를 입력으로 받아 물체의 클래스와 바운딩박스 정보를 출력
 - 정류장의 운영형태를 클래스로 나누어 구분
 - 현재 위치 정보를 결합하여 정차 위치 등 파악



< YOLOv3 모델 구조도 >

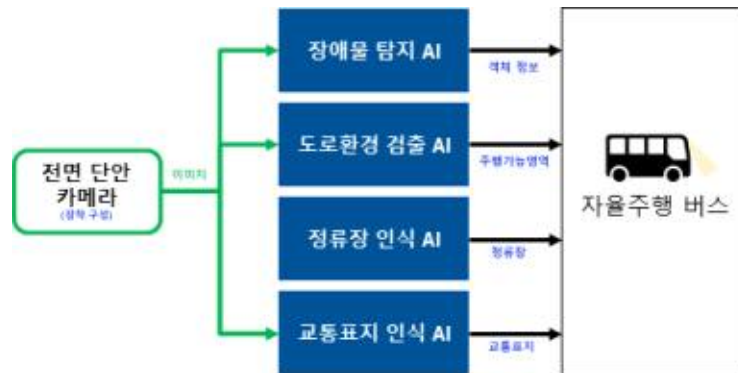
- R-CNN 모델을 활용한 교통안전표지 인식 알고리즘
 - 이미지 정보를 입력으로 받아 물체의 클래스 정보를 출력
 - 노면표지, 표지판, 신호등 등 각종 규제 정보 획득 가능
 - 현재 위치 정보를 입력하여 먼 거리에서 인식된 정보를 처리 가능
 - 정밀도로지도를 활용하여 사전 예측 가능



<R-CNN 모델 구조도>

○ 서비스 활용 시나리오







- 자율주행 버스 서비스
 - 주변 환경을 인지하며 정해진 노선을 자율적으로 주행하는 대중교통 서비스
 - 수요가 적은 지역에 도입하여 이용 형평성 확보
 - 대중교통 사고 중 상당수를 대중교통 자율주행 도입을 통해 예방
 - 짧은 차두거리로 군집주행을 통한 수송능력 확대 및 도시 공간 효율화
 - 학습 데이터 활용을 위한 다양한 크기의 모델별 샘플 코드 제공



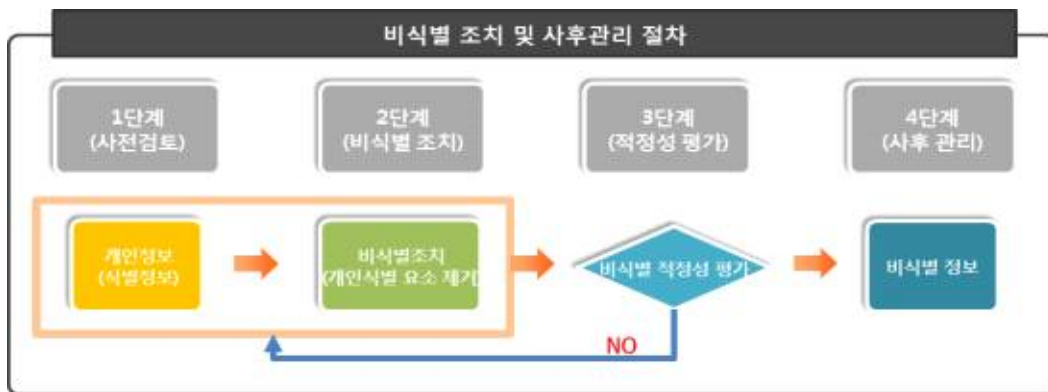
<자율주행 버스 서비스 구조도>

2.6.2 데이터 제공

- 구축된 데이터는 일반적인 오픈소스 공개 방식인 Github을 이용하여 데이터 활용 방식 공개 예정
 - 공개된 Github 사이트를 통해 데이터 정합 방법 등 데이터 활용에 필요한 안내 제공
 - 데이터 이용자를 위한 데이터 저장 형식-key feature 및 사용 가이드 라인 제공
- 구축된 학습용 데이터 자료의 공개·개방 시 저작권은 CC 라이선스를 따름
 - 제시된 CC 라이선스 중 CC BY로 가장 활용성이 높은 라이선스를 채택 예정

라이선스	이용조건	문자표기
	• 저작자표시 저작자의 이름, 저작물의 제목, 출처 등 저작자에 관한 표시를 해주어야 함	CC BY
	• 저작자표시-비영리 저작자를 밝히면 자유로운 이용이 가능하지만 영리목적으로 이용할 수 없음	CC BY-NC
	• 저작자표시-변경금지 저작자를 밝히면 자유로운 이용이 가능하지만, 변경 없이 그대로 이용해야 함	CC BY-ND
	• 저작자표시-동일조건변경허락 저작자를 밝히면 자유로운 이용이 가능하고 저작물의 변경도 가능하지만, 2차적 저작물에는 원 저작물에 적용된 것과 동일한 라이선스를 적용해야 함	CC BY-SA
	• 저작자표시-비영리-동일조건변경허락 저작자를 밝히면 이용이 가능하며 저작물의 변경도 가능하지만, 영리목적으로 이용할 수 없고 2차적 저작물에는 원 저작물과 동일한 라이선스를 적용해야 함	CC BY-NC-SA
	• 저작자표시-비영리-변경금지 저작자를 밝히면 자유로운 이용이 가능하지만, 영리목적으로 이용할 수 없고 변경 없이 이용해야 함	CC BY-NC-ND

- 공개된 데이터 활용을 필요로 하는 새로운 산업과 기술 발전으로 개인정보 침해 위험도 증가하고 있고, 개인정보 침해가능성을 최소화하면서 데이터의 활용을 높이기 위해 개인정보 비식별화가 필수임
- (개인정보) 살아 있는 개인에 관한 정보로서 개인을 알아볼 수 있는 정보이며, 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보를 포함



<비식별 조치 및 사후관리 순서도>

단계	내용	비고
1단계 사전 검토	개인 정보에 해당하는지 여부 검토후, 개인정보가 아닌 것이 명백한 경우 법제 규제 없이 활용	
2단계 비식별 조치	정보집합물(데이터셋)에서 개인을 식별할 수 있는 요소를 전부 또는 일부 삭제나 대체 등 활용, 개인을 알아 볼 수 없도록 하는 조치	
3단계 적정성 평가	다른 정보와 쉽게 결합하여 개인을 식별할 수 있는지 비식별 조치 적정성 평가	재조취
4단계 사후 관리	비식별 정보 안전조치, 재식별 가능성 모니터링 비식별 정보 활용 과정에서 재식별 방지를 위한 조치	

- 본 사업에서 구축 되는 데이터는 4차 산업혁명의 근간이 되는 AI 학습용 데이터 중 하나로 국내 자율주행 등 AI분야 산업체, 대학, 연구 대상으로 개방
 - 국내 대학, 산업체, 연구소가 관련분야의 AI 기술 개발을 할 수 있도록 시기적절하게 DB를 공급하여 중소·중견 기업의 머신러닝 기반의 기술개발경쟁력 강화
- 공개·개방의 정책적 기술적 방안
 - 정부는 데이터·AI경제 선도 국가로 도약을 위해 데이터·AI경제 활성화 계획 발표
 - 인공지능 혁신생태계를 조성하기 위해 AI 허브*를 구축하고 지속적인 지원 진행 중
 - AI 서비스 개발 필수 인프라인 AI 학습용 데이터, 알고리즘 개발 지원, 컴퓨팅 파워를 온라인 일괄 제공 (<http://www.aihub.or.kr>)
 - 본 사업에서 생산된 도로상태 및 자율버스 AI 데이터는 AI 학습용 데이터셋 구축 분야 로 AI 허브를 통해 공개·개방하고 관련 소스코드 및 결과 자료는 관련 포럼이나 웹사이트를 통하여 공개하여 관련 사업의 연구 개발자들이 활용할수 있게 지원함
- 기존 데이터셋과의 연계 방안
 - KITTI 등에서 기존에 공개한 자율주행 연구용 데이터와의 연계를 위해 원천데이터의 데이터 제공 형식을 기존 공개 데이터셋과 동일하거나 유사한 형태로 제공
 - 카메라 데이터: *.png 형식의 이미지 데이터 및 calibration 정보
 - GNSS/INS 데이터: *.txt 형식의 데이터 및 설명 파일