

인공지능 데이터 구축·활용 가이드라인

- [분야35] 100 교통문제 해결을 위한 CCTV 교통 데이터 (고속도로) -

인공지능 데이터 구축	사업 총괄	LAON PEOPLE
	데이터 설계	KE-TI 한국전자기술연구원 Korea Electronics Technology Institute TESTWORKS
	원천데이터 수집	ex 한국도로공사 m 메타빌드(주)
	데이터 정제	CRESPLIT
	데이터 가공	TESTWORKS
	데이터 검수	TESTWORKS
	클라우드 소싱	TESTWORKS
	저작도구 개발	TESTWORKS
	AI모델 개발	KE-TI 한국전자기술연구원 Korea Electronics Technology Institute
	응용 서비스 개발	m 메타빌드(주) KE-TI 한국전자기술연구원 Korea Electronics Technology Institute A (주)알트에이 ALT-A Inc.
가이드라인 작성	한국전자기술연구원	장수현 책임
	메타빌드	김영희 책임
	크레스프리	전근홍 이사
	테스트웍스	박효원 책임, 박지혜 연구원
가이드라인 버전	2.2 2021.04.20	

목 차

1. 데이터 명세 정보 1

1.1 데이터 정보 요약 1

1.2 데이터 포맷 1

1.3 어노테이션 포맷 4

1.4 데이터 구성 5

1.5 데이터 통계 6

1.6 원시데이터 특성 10

2. 데이터 구축 가이드 32

2.1 데이터 구축 개요 32

2.2 문제정의 33

2.3 획득·정제 34

2.4 어노테이션/라벨링 40

2.5 검수 49

2.6 활용 52

[부 록]

VDS 속도추정 GT 데이터 60

1. 데이터 명세 정보

1.1 데이터 정보 요약

데이터 이름	[데이터 100] 교통문제 해결을 위한 CCTV 교통 데이터(고속도로)	
활용 분야	차량 속도와 교통량 측정용 AI 데이터셋 구축, 고속도로 교통상황 분석	
데이터 요약	고속도로 교통량과 차량 속도를 자동 측정하는 AI 영상데이터 구축을 목표로 고속도로 CCTV 영상 49개소 500시간 이상의 영상과 이미지를 데이터로 수집	
데이터 출처	<ul style="list-style-type: none"> • 수도권본부(영동선, 경부선), 지방부(강원본부, 대전충남본부, 부산경남본부, 부산항대교, 전북본부, 충북본부) 본부별 관할 고속도로 • 수도권본부 CCTV: 14개소 • 지방부 CCTV: 35개소 <ul style="list-style-type: none"> - CCTV 위치, 도로 개수, 날씨 등 다양하게 수집 • 동영상: 총 504.3시간 • FHD(1920*1080), HD(1280*720) 영상 수집 • 수집한 CCTV 영상데이터 기반 Detection, segmentation 학습데이터 이미지 가공 	
데이터 이력	배포버전	1.0
	개정이력	2021.03.05
	작성자/배포자	

1.2 데이터 포맷

1.2.1 CCTV영상 데이터 포맷

- ▶ 교통정보 CCTV 영상은 수도권본부, 지방부 중 9개의 본부를 선정하여 총 49개소에서 504.3 시간의 영상데이터를 획득
- ▶ 다양한 환경에서 데이터를 수집하기 위하여, 시간대(동통/주간/야간), 차로 수(2, 3, 4차로), 도로 형태(일반도로, 교량, 터널), 지역별(수도권, 지방부)로 구분하여 수집. 또한, 기후조건에 강건한 모델을 만들 수 있도록 약조건 기후(눈, 비, 안개)에서 약 10% 정도의 원천 영상데이터를 획득
- ▶ 수집한 CCTV 영상데이터는 영상분석에 적합한 길이인 60초 단일 파일로 파싱하여 총 30,259개로 관리
- ▶ CCTV 영상 및 이미지 파일 이름은 메타데이터가 들어가도록 정의
- ▶ 메타데이터 정의
 - 수집지점, 날씨, 해상도, 도로 유형 등 원천영상 수집환경 정보인 메타데이터를 파일명에 표기하여 제공하며, 해당 메타데이터는 별도의 파일로도 제공됨
 - 메타데이터는 국문 및 영문을 함께 제공됨

[표 1] 메타데이터 정의 방안

국문	영문
<ul style="list-style-type: none"> 동영상 파일명에 한글로 메타정보가 포함되도록 하여 테스트웍스에서 빠르고 편하게 가공 추후 데이터셋 제공 시, 각 데이터가 어디서 수집되었는지 구분될 수 있도록 수집지점에 대한 위치명을 하위 폴더로 구분 예시) 강원/강릉/동해선/01_20201020_1200_0001.png 폴더명에는 한글이 포함되어도 상관 없으나, png파일명과 xml 어노테이션 파일 안에 들어가는 파일명은 한글이 포함되지 않도록 파일명 변경하여 제공 	<ul style="list-style-type: none"> 기본적으로 한글의 메타정보 규칙을 따르며, 다음의 규칙을 추가함. 한글 지역/지점명 변환 과정에 발생하는 불임표 '-' 삭제, 지점명은 띄어쓰기 안함 요일 표기는 영문 대문자 3개로 통일 첨두 = RH(Rush Hour), 비첨두 = NH(Normal Hour) 편도 = OW(One Way), 왕복 = TW(Two Way) **교 = **brdg, **육교 = **overpass, **터널 = **tunnel

[표 2] 메타데이터 예시

본부	지사	설치위치	국문	영문
수 권 본 부	수원	영동선 상행선 44.5K	수원_CH10_20201010_1200_월_9m_비첨두_고속도로_편도5_맑음_FHD.avi	suwon_CH10_20201010_1200_MON_9m_NH_highway_OW5_sunny_FHD.avi
강 원 본부	강릉	동해선 강동2교 48K	강릉_강동2교_20201010_1200_월_15m_비첨두_고속도로_왕복2_맑음_FHD.avi	Gangneung_Gangdong2brdg_20201010_1200_MON_15m_NH_highway_TW2_sunny_FHD.avi
충 북 본부	진천	중부선 남이육교 248	진천_남이육교_20201010_1200_월_15m_비첨두_고속도로_편도2_맑음_FHD.avi	Jincheon_namioverpass_20201010_1200_MON_15m_NH_highway_OW2_sunny_FHD.avi
전 북 본부	부안	서해안선 울촌 62	부안_울촌_20201010_1200_월_15m_비첨두_고속도로_왕복2_맑음_FHD.avi	Buan_yulchon_20201010_1200_MON_15m_NH_highway_TW2_sunny_FHD.avi
대 전 충 남 본부	천안	경부선 옥산 309	천안_옥산_20201010_1200_월_15m_비첨두_고속도로_왕복3_맑음_FHD.avi	Cheonan_ogsan_20201010_1200_MON_15m_NH_highway_TW3_sunny_FHD.avi
부 산 대 교	부 산	부산항대교 감만 방면	부산_부산항1_20201029_0900_목_8m_비첨두_고속도로_왕복2_맑음_FHD.mp4	Busan_busanportbrdg1_20201029_0900_THU_8m_NH_highway_TW2_sunny_FHD.mp4

1.2.2 이미지 데이터 포맷

▶ Bounding box 데이터

- 수집된 영상 클립에서 정제(프레임 이미지 추출)한 인공지능 학습 데이터용 이미지에서 객체(차량)이 존재하는 이미지만 학습 데이터셋의 가공 이미지로 사용
- 개발용 학습 데이터셋은 train : validation = 8 : 2로 분할
- 수집된 CCTV 동영상 파일에서 일정 프레임 간격을 파싱하여 png 파일로 저장
- 이미지의 해상도는 원시 동영상과 동일하게 가로형 FHD(1920x1080) 세로형 FHD(1080x1920)이 전체 약 97%를 구성하며 약 3% 가 가로형 HD(1280 x 768)

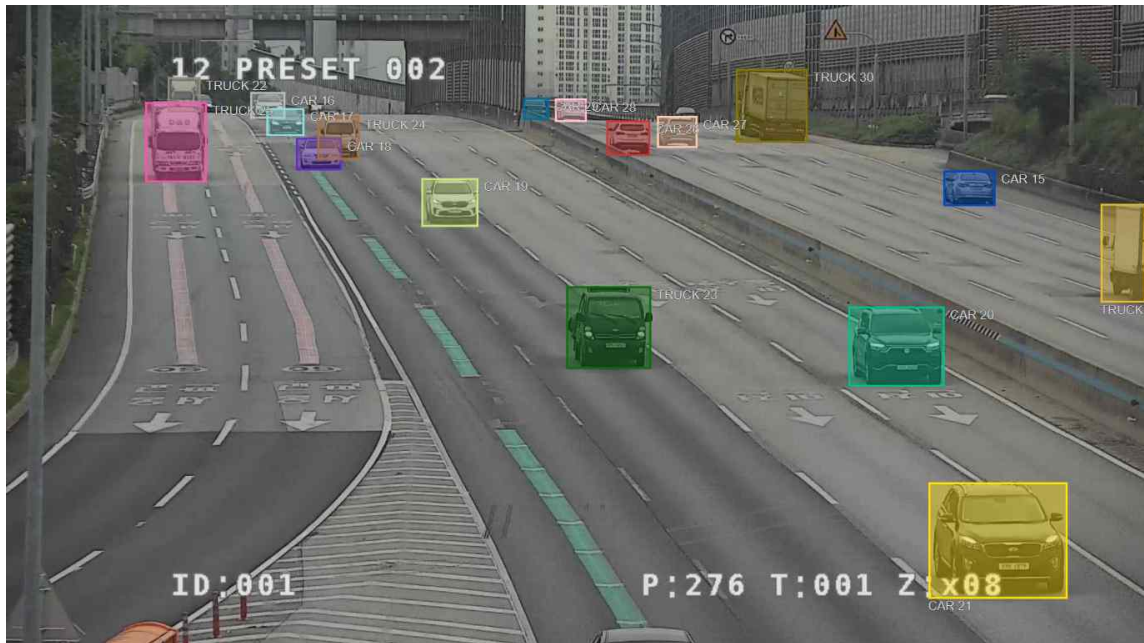


그림 2 FHD (1920x1080) png 이미지 Bounding Box 데이터 가공 예시

► Polygon Segmentation 데이터

- Bounding Box와 동일함
- 수집된 영상 클립에서 정제(프레임 이미지 추출)한 인공지능 학습 데이터용 이미지에서 객체(차량)이 존재하는 이미지만 학습 데이터셋의 가공 이미지로 사용
- 개발용 학습 데이터셋은 train : validation = 8 : 2로 분할
- 수집된 CCTV 동영상 파일에서 일정 프레임 간격을 파싱하여 png 파일로 저장
- 이미지의 해상도는 원시 동영상과 동일하게 가로형 FHD(1920x1080) 세로형 FHD(1080x1920)이 전체 약 97%를 구성하며 약 3% 가 가로형 HD(1280 x 768)



그림 3 FHD (1920x1080) png 이미지 Polygon Segmentation 데이터 가공 예시

1.3 어노테이션 포맷

- ▶ 이미지 가공에 대한 어노테이션 데이터는 가공 정보가 직관적인 xml 포맷을 사용함

[표 3] Bounding Box 어노테이션 파일 내부 구조 설명

항목	설명	xml에 입력된 어노테이션 데이터 예시
<name>	이미지 가공 작업의 이름, 메타데이터가 적용된 동영상 파일 이름과 같음	<name>Suwon_CH01_20200720_1700_MON_9m_NH_highway_TW5_sunny_FHD</name>
<label>	객체(Class)정보 정의	<labels> <label> <name>car</name> </label> <label> <name>truck</name> </label> <label> <name>bus</name> </label> </labels>
<image>	가공 대상인 원천 이미지 데이터 파일명 및 해상도 등의 정보 정의	...(중략)... <image id="0" name="Suwon_CH01_20200720_1700_MON_9m_NH_highway_TW5_sunny_FHD_001.png" width="1080" height="1920"> ...(중략)...
<box>	이미지 위에 가공된 바운딩박스 객체의 정보, 라벨 정보와 box의 좌측 상단, 우측하단 좌표 정보를 포함	...(중략)... <box label="car" occluded="0" xtl="507.64" ytl="747.49" xbr="582.19" ybr="798.56" z_order="3"> </box> ...(중략)...

```

1_Suwon_CH01_20200720_1700_MON_9m_NH_highway_TW5_sunny_FHD.xml
51 <image id="0" name="Suwon_CH01_20200720_1700_MON_9m_NH_highway_TW5_sunny_FHD_001.png" width="1080" height="1920">
52 <box label="truck" occluded="0" xtl="862.22" ytl="420.16" xbr="910.65" ybr="467.05" z_order="27">
53 </box>
54 <box label="car" occluded="0" xtl="507.64" ytl="747.49" xbr="582.19" ybr="798.56" z_order="3">
55 </box>
56 <box label="car" occluded="0" xtl="126.41" ytl="923.45" xbr="250.52" ybr="1015.94" z_order="1">
57 </box>
58 <box label="car" occluded="0" xtl="371.28" ytl="914.43" xbr="486.94" ybr="1001.31" z_order="2">
59 </box>
60 <box label="truck" occluded="0" xtl="738.62" ytl="654.51" xbr="784.12" ybr="706.47" z_order="15">

```

그림 4 Bounding Box 가공 어노테이션 파일 실제 예시

[표 4] Polygon Segmentation 어노테이션 파일 내부 구조 설명

항목	설명	xml에 입력된 어노테이션 데이터 예시
<name>	이미지 가공 작업의 이름, 메타데이터가 적용된 동영상 파일 이름과 같음	<name>Suwon_CH01_20200721_1500_TUE_9m_NH_highway_TW5_sunny_FHD</name>
<label>	객체(Class)정보 정의	<labels> <label> <name>car</name> </label> <label> <name>truck</name> </label> <label> <name>bus</name> </label> </labels>
<image>	가공 대상인 원천 이미지 데이터 파일명 및 해상도 등의 정보 정의	...(중략)... <image id="0" name="Suwon_CH01_20200721_1500_TUE_9m_NH_highway_TW5_sunny_FHD_001.png" width="1080" height="1920"> ...(중략)...
<polygon>	이미지 위에 가공된 바운딩박스 객체의 정보, 라벨 정보와 polygon 다각형을	...(중략)... <polygon label="car" occluded="0" points="68.25,669.1

1.5 데이터 통계

1.5.1 데이터 구축 규모

- ▶ 1분 단위 500시간 영상데이터 구축
- ▶ 30분 길이의 원시데이터 영상 중 시점이 변하지 않는 1분 단위로 영상을 파싱하여 총 504.3시간 분량의 데이터를 구축함
- ▶ 날씨의 경우 악천후의 영상과 데이터는 10% 이상 확보

[표 5] 1분 단위 500시간 데이터 통계분석

항목	내용	
1. 지역별(CCTV개수)	수도권	14개
	지방부	35개
2. 시간대	동틀 (06:00~08:59)	61.71시간 (12.24%)
	주간 (09:00~17:00)	393.31시간 (77.99%)
	야간 (17:01~06:00)	49.28시간 (9.77%)
3. 차로 별(CCTV개수)	2차로	24개 (48.98%)
	3차로	11개 (22.45%)
	5차로	14개 (28.57%)
4. CCTV 설치높이(CCTV개수)	4.8m	1개 (2%)
	7m	9개 (18.37%)
	8m	6개 (12.24%)
	9m	10개 (20.41%)
	10m	2개 (4%)
	15m	21개 (42.86%)
5. 날씨	• 전체 10% 이상, 악천후 최소비율 눈(1% 이상), 비(5% 이상), 안개(1% 이상)	
	눈	10.5시간 (2.08%)
	비	29.36시간 (5.82%)
	안개	19.06시간 (3.78%)
6. 교통특성	첨두	49.35시간 (9.79%)
	비첨두	454.96시간 (90.21%)
7. 품질특성	FHD(1920x1080)	491.31시간 (97.42%)
	HD(1280x720)	13시간 (2.58%)

1.5.2 데이터 분포

▶ Bounding Box

- Bounding Box 가공 데이터셋은 아래 표 6~12와 같은 분포도(데이터 다양성)를 포함

[표 6] 가공 형태별 이미지 수량 및 객체 class 분포도

[30만장 가공시]					[객체 수 보완을 위한 초과 가공 수행]				
이미지수	객체 수 (Class)				이미지수	객체 수 (Class)			
	car	bus	truck	합계		car	bus	truck	합계
301,329	1,413,751	55,028	637,579	2,106,358	378,206	2,566,293	153,728	870,639	3,590,660
-	67.12%	2.61%	30.27%	100%	-	71.47%	4.28%	24.25%	100%

[표 7] 교통 특성별 데이터 분포도

(RH : Rush Hour, NH : Normal Hour)

구분	RH	NH	합계
이미지 수	21,141	244,477	265,618
비율	7.96%	92.04%	100%

[표 8] 데이터 수집 지점 목록

연번	수집 지점 메타데이터 (이름 순)	수집 지점 설명
1~6	busanportbrdg1~5	부산광역시 부산항대교 1지점~6지점
7~16	CH01~10	영동선 수원-신갈 구간 CCTV 채널01번~10번
17	chujeomtunnel(changwon)	충청북도 충주시사 추점터널 창원 방향
18	chujeomtunnel(yangpyeong)	충청북도 충주시사 추점터널 양평 방향
19	daemog	전라북도 부안지사 대목
20	danpyeong1brdg	충청북도 충주시사 단평1교
21	deungcheon	전라북도 전주지사 둔촌
22	dongjisan	전라북도 부안지사 동지산
23	gangdong2brdg	강원도 강릉지사 강동2교
24	gangjeong	충청남도 천안지사 강정
25	gudeog	전라북도 전주지사 구덕
26	guljitunnel	강원도 홍천지사 굴지터널
27	gwanggyeog1brdg	강원도 홍천지사 광격 1교
28	hyangho1brdg	강원도 강릉지사 향호1교
29	hyangho2brdg	강원도 강릉지사 향호2교
30	injetunnel2	강원도 인제터널 2지점
31	jangam	충청북도 진천지사 장암리
32	jeongeub	전라북도 전주지사 정읍
33	jogogli	충청북도 충주시사 조곡리
34	jongsin	전라북도 부안지사 종신
35	namioverpass	충청북도 진천지사 나미육교
36	nodong	강원도 강릉지사 노동
37	nodong1brdg	강원도 강릉지사 노동1교
38	ogsan	충청남도 천안지사 옥산
39	sajeong	충청남도 천안지사 사정
40	samhobrdg	충청북도 진천지사 삼호교
41	sangpyeongoverpass	전라북도 전주지사 상평육교
42	sangwoo1brdg	충청북도 충주시사 상우1교
43	seomgangbrdg	강원도 홍천지사 섬강교
44	seongnam2	충청남도 천안지사 성남2지점
45	soju2brdg	전라북도 부안지사 소주2교
46	taeinshelter(seoul)	전라북도 전주지사 태인터널(서울방향)
47	wichon2brdg	강원도 강릉지사 위촌2교
48	wonchang4brdg	강원도 홍천지사 원창4교
49	yeji	전라북도 부안지사 예지
50	yulchon	전라북도 부안지사 율촌
51~61	gyeongbu n	수도권 경부고속도로 판교-용인 구간 10개 지점

[표 9] 시간대별 데이터 분포

시간대	가공수량	비율
동틈(06:00~08:59)	29,130	9.67 %
주간(09:00~17:00)	204,875	67.99 %
야간(17:01~06:00)	67,324	22.34 %
합계	301,329	100 %

[표 10] 차로별 데이터 분포

메타데이터	메타데이터 설명	가공 수량	비율
OW2	단방향 2차선	21,510	7.14%
OW5	단방향 5차선	53,478	17.75%
TW2	왕복 2차선(총4차선)	139,424	46.27%
TW3	왕복 3차선(총6차선)	68,732	22.81%
TW5	왕복 5차선(총10차선)	18,185	6.03%
합계		301,329	100%

[표 11] 도로 형태별 데이터 분포

구분	설명	가공 수량	비율
일반도로	수집 지점 메타데이터에 'brdg', 'tunnel' 이 포함되지 않은 모든 지점	181,548	60.25%
교량	수집 지점 메타데이터에 'brdg' 포함 교량 위가 아닌 진-출입로까지 포함	104,324	34.62%
터널	수집 지점 메타데이터에 'tunnel' 포함 터널 내부가 아닌 진-출입로까지 포함	15,457	5.13%
합계		301,329	100%

[표 12] 날씨별 데이터 분포

날씨	메타데이터	가공 수량	비율
맑음	sunny	239,126	79.36%
비	rainy	38,069	12.63%
안개	fog	13,567	4.50%
눈	snow	9,549	3.17%
터널내부(식별불가)	tunnel	1,018	0.34%
합계		301,329	100%

► Polygon Segmentation

- Polygon Segmentation 가공 데이터셋은 아래 표 13~19와 같은 분포도(데이터 다양성)를 포함

[표 13] 가공 형태별 이미지 수량 및 객체 class 분포도

[20만장 가공시]					[객체 수 보안을 위한 초과 가공 수행]				
이미지 수	객체 수 (Class)				이미지 수	객체 수 (Class)			
	car	bus	truck	합계		car	bus	truck	합계
265,618	1,053,640	47,370	509,846	1,610,856	313,098	1,409,640	127,370	585,846	2,122,856
	65.41%	2.94%	31.65%	100%		66.40%	6.00%	27.60%	100%

[표 14] 교통 특성별 데이터 분포도

(RH : Rush Hour, NH : Normal Hour)

구분	RH	NH	합계
이미지 수	21,141	244,477	265,618
비율	7.96%	92.04%	100%

[표 15] 데이터 수집 지점 목록

연번	수집 지점 메타데이터 (이름 순)	수집 지점 설명
1~6	busanportbrdg1~5	부산광역시 부산항대교 1지점~6지점
7~16	CH01~10	영동선 수원-신갈 구간 CCTV 채널01번~10번
17	chujeomtunnel(changwon)	충청북도 충주시사 추점터널 창원 방향
18	chujeomtunnel(yangpyeong)	충청북도 충주시사 추점터널 양평 방향
19	daemog	전라북도 부안지사 대목
20	danpyeong1brdg	충청북도 충주시사 단평1교
21	deungcheon	전라북도 전주지사 둔촌
22	dongjisan	전라북도 부안지사 동지산
23	gangdong2brdg	강원도 강릉지사 강동2교
24	gangjeong	충청남도 천안지사 강정
25	gudeog	전라북도 전주지사 구덕
26	guljitunnel	강원도 홍천지사 굴지터널
27	gwanggyeog1brdg	강원도 홍천지사 광격 1교
28	hyangho1brdg	강원도 강릉지사 향호1교
29	hyangho2brdg	강원도 강릉지사 향호2교
30	injetunnel2	강원도 인제터널 2지점
31	jangam	충청북도 진천지사 장암리
32	jeongeub	전라북도 전주지사 정읍
33	jogogli	충청북도 충주시사 조곡리
34	jongsin	전라북도 부안지사 종신
35	namioverpass	충청북도 진천지사 나미육교
36	nodong	강원도 강릉지사 노동
37	nodong1brdg	강원도 강릉지사 노동1교
38	ogsan	충청남도 천안지사 옥산
39	sajeong	충청남도 천안지사 사정
40	samhobrdg	충청북도 진천지사 삼호교
41	sangpyeongoverpass	전라북도 전주지사 상평육교
42	sangwoo1brdg	충청북도 충주시사 상우1교
43	seomgangbrdg	강원도 홍천지사 섬강교
44	seongnam2	충청남도 천안지사 성남2지점
45	soju2brdg	전라북도 부안지사 소주2교
46	taeinshelter(seoul)	전라북도 전주지사 태인터널(서울방향)
47	wichon2brdg	강원도 강릉지사 위촌2교
48	wonchang4brdg	강원도 홍천지사 원창4교
49	yeji	전라북도 부안지사 예지
50	yulchon	전라북도 부안지사 율촌
51~61	gyeongbu n	수도권 경부고속도로 판교-용인 구간 10개 지점

[표 16] 시간대별 데이터 분포

시간대	가공수량	비율
동틈(06:00~08:59)	23,907	9.00%
주간(09:00~17:00)	181,806	68.45%
야간(17:01~06:00)	59,905	22.55%
합계	265,618	100%

[표 17] 차로별 데이터 분포

메타데이터	메타데이터 설명	가공 수량	비율
OW2	단방향 2차선	21,194	7.98%
OW5	단방향 5차선	40,986	15.43%
TW2	왕복 2차선(총4차선)	126,321	47.56%
TW3	왕복 3차선(총6차선)	62,847	23.66%
TW5	왕복 5차선(총10차선)	14,270	5.37%
합계 :		265,618	100%

[표 18] 도로 형태별 데이터 분포

구분	설명	가공 수량	비율
일반도로	수집 지점 메타데이터에 'brdg', 'tunnel' 이 포함되지 않은 모든 지점	157,569	59.32%
교량	수집 지점 메타데이터에 'brdg' 포함 교량 위가 아닌 진-출입로까지 포함	92,729	34.91%
터널	수집 지점 메타 데이터에 'tunnel' 포함 터널 내부가 아닌 진-출입로까지 포함	15,320	5.77%
합계		265,618	100%

[표 19] 날씨별 데이터 분포

날씨	메타데이터	가공 수량	비율
맑음	sunny	232,838	87.66%
비	rainy	19,788	7.45%
안개	fog	6,268	2.36%
눈	snow	5,808	2.19%
터널내부(식별불가)	tunnel	916	0.34%
합계		265,618	100%

1.6 원시데이터 특성

1.6.1 대상분류

- ▶ 수도권 본부(영동선, 경부선), 지방부(강원본부, 대전충남본부, 부산경남본부, 부산항대교, 전북본부, 충북본부) 각 본부별 관할 고속도로 총 49개소 CCTV영상 수집
- ▶ 수도권본부 (14개소)
 - 영동선 10개소, 경부선 4개소
- ▶ 지방부 (35개소)
 - 강원본부 1개소, 대전충남본부 4개소, 부산경남본부 5개소, 부산항대교 6개소, 전북본부 11개소, 충북본부 8개소

1.6.2 지점별 샘플 이미지

▶ 강원본부

[표 20] 강원본부 수집지점

설치위치	이미지
injetunnel2	

▶ 대전충남본부

[표 21] 대전충남본부 수집지점

설치위치	이미지
gangjeong	
ogsan	
sajeong	

	
seongnam2	

▶ 부산경남본부


[표 22] 부산경남본부 수집지점

설치위치	이미지
busanoegwag1	<p>12-10-2020 Thu 10:50:08</p> 
busanoegwag2	<p>12-10-2020 Thu 14:04:02</p> 
busanoegwag3	<p>12-10-2020 Thu 15:52:22</p> 
busanoegwag4	

	
busanoegwag5	

▶ 부산항대교

[표 23] 부산항대교 수집지점

설치위치	이미지
busanportbrdg1	
busanportbrdg2	

	
busanportbrdg3	
busanportbrdg4	
busanportbrdg5	



▶ 수도권본부

[표 24] 수도권본부 수집지점




설치위치	이미지
gyeongbu2	
gyeongbu6	
gyeongbu10	

	
gyeongbu11	

▶ 영동선

[표 25] 영동선 수집지점

설치위치	이미지
CH01	
CH02	
CH03	

CH04	 <p>PRESET 001</p> <p>An aerial view of a multi-lane highway. A pink lane marking is visible on the left side of the road, and a green lane marking is visible on the right side. Several vehicles are traveling on the highway. The background shows a city skyline and a bridge.</p>
CH05	 <p>PRESET 002</p> <p>An aerial view of a multi-lane highway. A red lane marking is visible on the right side of the road. Several vehicles are traveling on the highway. The background shows a city skyline and a bridge.</p>
CH06	 <p>PRESET 003</p> <p>An aerial view of a multi-lane highway. A red lane marking is visible on the right side of the road. Several vehicles are traveling on the highway. The background shows a city skyline and a bridge.</p>

CH07	
CH08	
CH09	

CH10	
------	--

▶ 전북본부

[표 26] 전북본부 수집지점

설치위치	이미지
daemog	
dongjisan	

jongsin	
soju_2brdg	

yeji	
yulchon	

deungcheon	
gudeog	

sangpyeongover pass	
jeongeub	

<p>taeinshelter (seoul)</p>	
---------------------------------	--

▶ 충북본부

[표 27] 충북본부 수집지점

지사	이미지
<p>danpyeong1brd 9</p>	

	
jogogli	 
sangwoo1brdg	

	
jangam	
namioverpass	
samhobrdg	

<p>chujeomtunnel (changwon)</p>	
<p>chujeomtunnel (yangpyeong)</p>	

1.6.3 제약조건

- ▶ 다양한 날씨 정보 반영을 위해 악천후(눈, 비, 안개) 비율 10% 이상 확보

[표 28] 악천후 이미지

설치위치	이미지
눈	
비	
안개	

- ▶ 날씨별 효과적인 데이터 수집을 위하여 악천후 영상(10%)을 포함하기로 함

1.6.4 속성

- ▶ 1분 단위 500시간 영상데이터
 - 30,259개 영상 파일, 504.3시간 분량
 - AVI, MP4 파일 (FHD 1920x1080, 1080x1920), (HD 1280x720), 30fps

1.6.5 포괄성

- ▶ 1분 단위 500시간 영상은 수도권본부 14개소, 지방부 35개소 총 49개소 지점을 선정하여 시점이 변하지 않는 구간으로 영상을 파싱하여 구축함
- ▶ 영동선의 경우 연속류 구간 5개 폴에 총 10개 CCTV 채널을 설치
 - 해당 위치의 CCTV는 자동으로 카메라 프리셋이 변경되는 투어링 기능이 설정되어 있으나 1분 단위 영상에선 시점이 변하는 구간이 포함되지 않도록 파싱함
 - 영동선 CCTV는 가로형(1920x1080) 세로형(1080x1920) 두 가지 해상도를 포함함
- ▶ 역관, 빛 번짐, 저조도 등 원시데이터 품질이 열악한 경우 수집 및 가공 대상에서 제외함

2. 데이터 구축 가이드

2.1 데이터 구축 개요

- ▶ 본 인공지능 학습 데이터셋 구축은 고속도로 CCTV 영상 내 차량 속도와 교통량을 자동으로 측정하는 AI기술 개발에 활용을 목표로 함
- ▶ 차량 속도와 교통량 측정을 위한 학습용 데이터셋을 구축하여 고속도로의 교통상황을 분석하는데 핵심적인 지표로 활용하여 교통문제 해결이 최우선 목표임

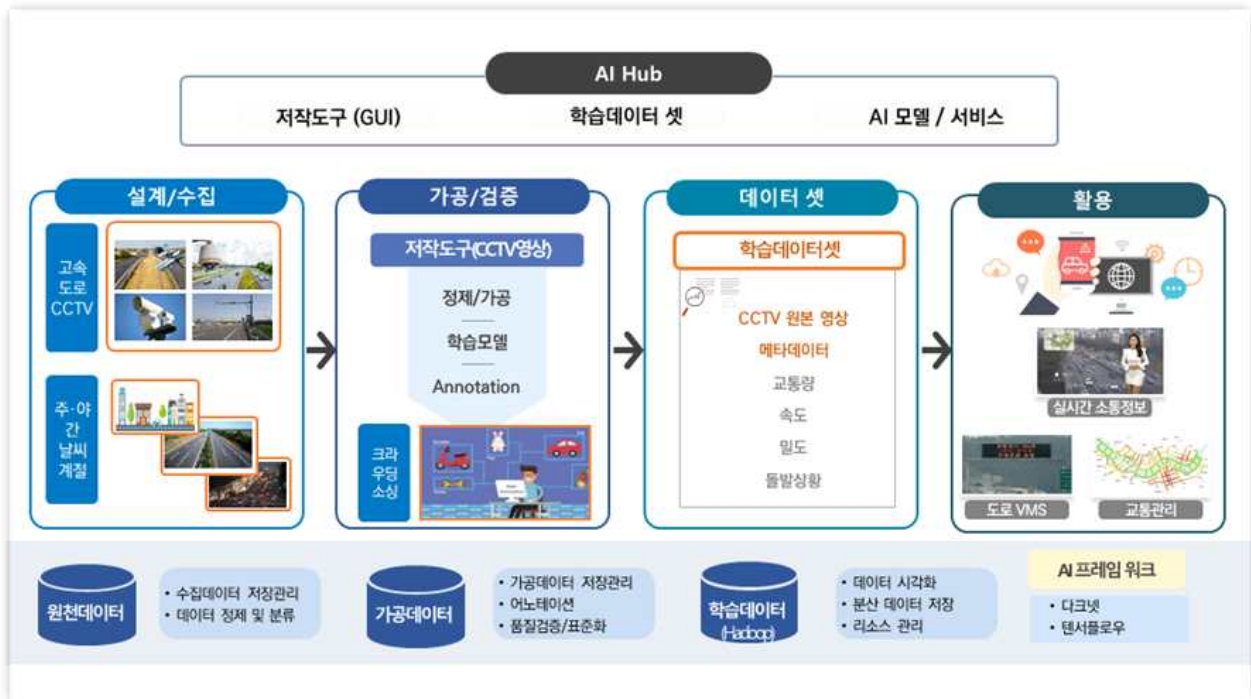


그림 7 인공지능 학습 데이터셋 구축 전체 프로세스

- ▶ 본 '교통문제 해결을 위한 CCTV 교통 데이터 (고속도로)' 인공지능 학습 데이터셋 구축은 참여기업의 크라우드소싱 플랫폼인 aiworks를 통해, 작업자를 모집하고 데이터 가공을 수행함.



그림 8 크라우드소싱 방식의 데이터 가공 플랫폼 개요



그림 9 클라우드소싱 방식의 사용자 참여

2.2 문제정의

2.2.1 임무 정의

- ▶ 지능형 도로, 자율 주행 등 인공지능 융합 도로 교통 산업 분야의 기술 및 시장 경쟁력 강화
 - 고속도로 CCTV 영상의 수집/정제/저장/가공/운용 시스템의 개발 및 적용을 통한 대용량 학습/가공데이터 구축과 개방, 차량 속도/교통량 검지 인공지능 알고리즘 개발 및 출연연·중소·벤처·스타트업 등의 기술개발 촉진, 인공지능 산업 육성, 데이터 기반 신 서비스 사업모델 발굴 등을 통한 산업 경쟁력 강화 지원을 목적으로 함
- ▶ 지능형 교통흐름분석 서비스를 위해 고속도로 CCTV 원천영상 수집, 정제와 도로 객체검출/분할 모델 개발을 위한 학습데이터(각 30만장, 20만장) 구축, 이를 통해 참조용 AI기반 교통흐름분석 모델을 개발하고, 분석된 교통파라미터를 통한 교통흐름분석용 서비스 개발과 공개를 목표로 함.



그림 10 데이터 수집, 구축 및 활용 프로세스

2.2.2 데이터 구축 유의사항

- ▶ 개인 정보 유출 및 확산 방지를 위한 민감 정보 비식별화
 - 원시 데이터 수집 과정에서 의도하지 않은 개인 정보가 수집될 수 있고 이는 차후 공개 데이터로 사용할 때 초상권이나 기타 개인 권리를 침해하는 법적 분쟁 소지가 있어서 비식별화 기술을 활용하여 데이터 사용자와 개인의 권리를 보호하고 법적 분쟁을 미연에 방지
 - 정보 비식별화 대상은 운전자 얼굴을 포함한 사람 얼굴과 차량 번호판
 - 비식별화는 학습 데이터셋 구축을 위한 어노테이션과 마찬가지로 1단계는 자동화, 2단계를 작업자들의 수동 검수를 통해 시행할 예정

2.3 획득·정제

2.3.1 원시데이터 선정 및 획득

- ▶ 원시 데이터의 형태 및 규모
 - 60초 단위 총 500시간 영상클립
 - 영동고속도로(상행선) 5개소 : 수원-신갈 JCT부터 약 3.1Km 연속류 구간을 5개의 폴에 총 10개의 CCTV 채널이 설치되어 모니터링 중
 - CCTV 카메라의 시점 및 배율을 변경하는 프리셋은 총 23개이며, 자동으로 카메라의 프리셋을 변경하는 투어링 (Touring) 기능이 설정
 - 영상의 화질은 FHD급 1920x1080(가로형), 1080x1920(세로형) 해상도
 - 각 카메라 시간대별 AVI 및 MP4등 범용 동영상 미디어 포맷으로 추출
 - 한국도로공사 관내 50개소
 - 강원본부 강릉지사, 홍천지사 10개소(동해안고속도로 6, 중앙고속도로 4)
 - 충북본부 진천지사, 충주지사 17개소(중부고속도로 7, 중부내륙고속도로 10)
 - 전북본부 부안지사, 전주지사 11개소(서해안고속도로 6, 호남고속도로 5)
 - 대전충남본부 천안지사, 공주지사 12개소(경부고속도로 5, 당진영덕고속도로 7)
 - 상기 동영상에서 추출한 인공지능 학습 데이터용 전처리 이미지 50만장 (JPG 포맷)
- ▶ 수집지점 선정을 위해서 아래 표와 같은 지표들이 기본적으로 고려되어야 하며, 세부적인 선정 기준은 수요기관의 요구사항과 참조용 교통흐름분석 모델 및 응용 서비스의 요구사항을 반영

[표 29] 데이터 수집 지점 선정 기준

1. 구조적 특성	2. 교통 특성	3. 품질 특성	4. 효과분석 용이성	5. 환경
<ul style="list-style-type: none"> • 수도권/지방부 • CCTV 설치형태 ✓ 설치높이 ✓ 터널/교량/진출입구 ✓ 검지방향 : 다가움/멀어짐 	<ul style="list-style-type: none"> • 침두 • 비침두 • 위험구간 	<ul style="list-style-type: none"> • CCTV 영상 해상도 (1080p@ 30fps) • 야간 조명 여부 • 시간대 따른 빛 번짐 여부 	<ul style="list-style-type: none"> • 학습데이터 구축 효과분석이 용이한 구간 	<ul style="list-style-type: none"> • 검지 및 제공이 용이한 도로환경 • 수요기관 및 제안사의 접근 용이

[표 30] 데이터 수집구간 검토

1. 시간대	2. 차로수	3. 도로 형태	4. 지역별	5. 날씨
<ul style="list-style-type: none"> • 동통 • 주간 • 야간 	<ul style="list-style-type: none"> • 2차로 • 3차로 • 4차로 	<ul style="list-style-type: none"> • 일반도로 • 교량 • 터널 	<ul style="list-style-type: none"> • 수도권 • 지방부 	<ul style="list-style-type: none"> • 맑음 • 눈 • 비 • 안개

▶ 그림11의 예와 같이 원시데이터의 품질이 열악한 경우 가공 대상에서 제외



그림 11 원시 데이터 품질이 열악한 경우

▶ 저작권은 이미지 추출 및 전처리 이전 동영상의 저작권을 따름

▶ 영동선 상행선 CCTV 설치지점(5개소)

[표 31] 영동선 상행선 CCTV 설치지점

지역본부	지사	기기명	설치위치(이정)	도로선형	차로수	장애물	높이
수도권본부	수원	CCTV	영동선 상행선 44.5K	좌로굽은	단선 5차로	무	9m
수도권본부	수원	CCTV	영동선 상행선 44.5K	직선	단선 5차로	무	9m
수도권본부	수원	CCTV	영동선 상행선 44.0K	직선	단선 5차로	무	9m
수도권본부	수원	CCTV	영동선 상행선 44.0K	직선	단선 5차로	무	9m
수도권본부	수원	CCTV	영동선 상행선 43.3K	직선	단선 5차로	무	9m
수도권본부	수원	CCTV	영동선 상행선 43.3K	좌로굽은	단선 5차로	무	9m
수도권본부	수원	CCTV	영동선 상행선 42.7K	우로굽은	단선 5차로	무	9m
수도권본부	수원	CCTV	영동선 상행선 42.7K	<ul style="list-style-type: none"> • 좌로굽은 • 오르막 	단선 5차로	무	9m
수도권본부	수원	CCTV	영동선 상행선 41.94K	내리막	왕복 5차로	표지판	9m
수도권본부	수원	CCTV	영동선 상행선 41.94K	직선	왕복 5차로	무	9m

▶ 한국도로공사 관내 50개소

- 강원본부 강릉지사, 홍천지사 10개소(동해안고속도로 6, 중앙고속도로 4)

- 충북본부 진천지사, 충주지사 17개소(중부고속도로 7, 중부내륙고속도로 10)
- 전북본부 부안지사, 전주시사 11개소(서해안고속도로 6, 호남고속도로 5)
- 대전충남본부 천안지사, 공주시사 12개소(경부고속도로 5, 당진영덕고속도로 7)

[표 32] 한국도로공사 관내(50개소)

지역본부	지사	기기명	설치위치(이정)	도로선형	차로수	장애물	높이
강원	강릉	CCTV	동해선 강동2교 48K	우로굽은	왕복2	무	15m
강원	강릉	CCTV	동해선 위촌2교 60K	직선	왕복2	무	15m
강원	강릉	CCTV	동해선 노동1교 63k	좌로굽은	왕복2	표지판	15m
강원	강릉	CCTV	동해선 노동 67k	직선	왕복2	무	15m
강원	강릉	CCTV	동해선 향호1교 75k	교량	왕복2	무	15m
강원	강릉	CCTV	동해선 향호2교 76k	교량	왕복2	무	15m
강원	홍천	CCTV	중앙선 섬강교 327k	직선	왕복2	표지판	15m
강원	홍천	CCTV	중앙선 광격1교 329k	좌로굽은	왕복2	무	15m
강원	홍천	CCTV	중앙선 굴지터널 370k	우로굽은	왕복2	무	15m
강원	홍천	CCTV	중앙선 원창4교 378K	교량	편도2	무	15m
충북	진천	CCTV	중부선 남이육교 248	직선도	편도2	무	15m
충북	진천	CCTV	중부선 진천터널2 274	우로굽은	왕복2	무	15m
충북	진천	CCTV	중부선 삼호교 292	직선도	왕복2	무	15m
충북	진천	CCTV	중부선 화봉육교1 301	직선도	왕복2	무	15m
충북	진천	CCTV	중부선 화봉육교 302	직선도	왕복2	무	15m
충북	진천	CCTV	중부선 장암1 304	좌로굽은	왕복2	무	15m
충북	진천	CCTV	중부선 장암 305	직선도	왕복2	표지판	15m
충북	충주	CCTV	중부내륙선 오가1교 205	우로굽은	왕복2	무	15m
충북	충주	CCTV	중부내륙선 추첨터널(창원)205.8	직선도	편도2	무	15m
충북	충주	CCTV	중부내륙선 추첨터널(양평)205.8	우로굽은	편도2	무	15m
충북	충주	CCTV	중부내륙선 추첨리 207	좌로굽은	왕복2	무	15m
충북	충주	CCTV	중부내륙선 괴산 209	직선도	왕복2	무	15m
충북	충주	CCTV	중부내륙선 조곡리 210	직선도	왕복2	무	15m
충북	충주	CCTV	중부내륙선 조곡터널 213.6	직선도	왕복2	무	15m
충북	충주	CCTV	중부내륙선 용전터널 226	우로굽은	왕복2	표지판	15m
충북	충주	CCTV	중부내륙선 상우1교 252	직선도	왕복2	무	15m
충북	충주	CCTV	중부내륙선 단평1교 254	직선도	왕복2	무	15m
전북	부안	CCTV	서해안선 울촌 62	직선도	왕복2	무	15m
전북	부안	CCTV	서해안선 예지 73	직선도	왕복2	무	15m
전북	부안	CCTV	서해안선 소주2교 102	직선도	왕복2	무	15m
전북	부안	CCTV	서해안선 종신 120	좌로굽은	왕복2	무	15m
전북	부안	CCTV	서해안선 대목 123	교량	왕복2	무	15m
전북	부안	CCTV	서해안선 동지산 132	좌로굽은	왕복2	무	15m
전북	전주	CCTV	호남선 등천 117	우로굽은	왕복2	무	10m
전북	전주	CCTV	호남선 상평육교 126	직선도	왕복2	무	15m
전북	전주	CCTV	호남선 정읍 128	직선도	왕복2	무	15m
전북	전주	CCTV	호남선 태인줄음섬터(서울)142	좌로굽은	왕복2	무	15m
전북	전주	CCTV	호남선 구덕 177	좌로굽은	왕복6	무	15m
대전충남	천안	CCTV	경부선 옥산 309	직선도	왕복3	무	15m
대전충남	천안	CCTV	경부선 천수천교 313	우로굽은	왕복3	무	15m
대전충남	천안	CCTV	경부선 사정 318	직선도	왕복3	표지판	15m
대전충남	천안	CCTV	경부선 강정 319	직선도	왕복3	무	15m
대전충남	천안	CCTV	경부선 성남2 328	우로굽은	왕복3	무	10m
대전충남	공주	CCTV	당진대전선 용리 17	직선도	왕복2	무	15m
대전충남	공주	CCTV	당진대전선 삽교천교 18	직선도	왕복2	무	15m
대전충남	공주	CCTV	당진대전선 양막 20.5	우로굽은	왕복2	무	15m
대전충남	공주	CCTV	당진대전선 좌방리 22	직선도	왕복2	무	15m
대전충남	공주	CCTV	당진대전선 월곡 25	직선도	왕복2	무	15m
대전충남	공주	CCTV	당진대전선 탄방교 31	직선도	왕복2	무	15m
대전충남	공주	CCTV	당진대전선 신양 37	우로굽은	왕복2	가로등	15m

2.3.2 메타데이터 정의

▶ 메타데이터 정의

- [표 33]과 같이 수집지점, 날씨, 해상도, 도로유형 등 원천영상 수집환경 정보인 메타데이터를 파일명에 표기하여 제공하며, 해당 메타데이터는 별도의 json 파일로도 제공됨

▶ json 메타데이터 파일의 구조 및 예시

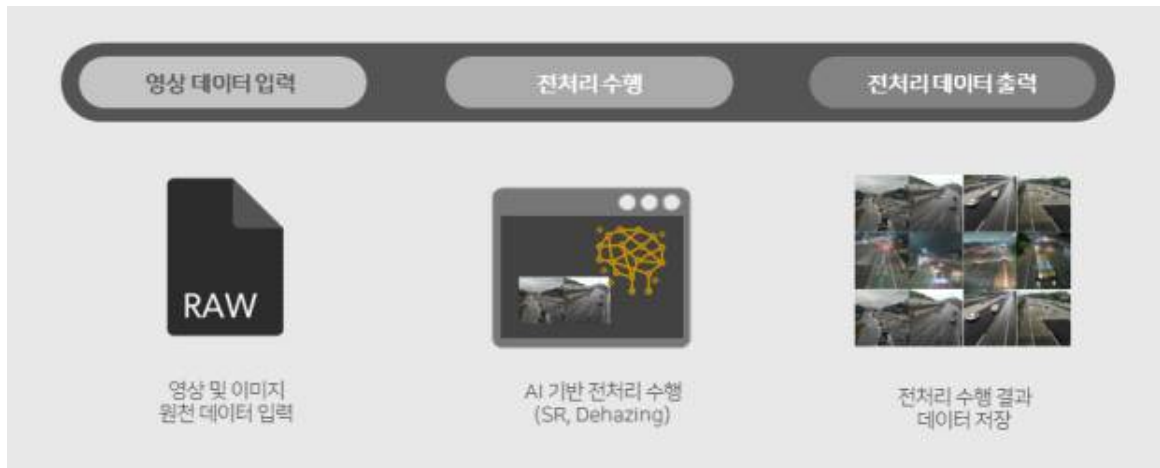
- 본 가이드라인에서 제시하는 메타데이터 파일 저장 포맷은 json을 사용함
- 본부별로 json 파일이 하나씩 존재하며, 각 json 파일은 영상 제목을 key 값으로 가짐
- json 파일 안에는 수많은 tag가 존재하며, VDL_distance, VDL을 제외한 나머지 tag는 원천영상 수집환경 정보이고, VDL_distance, VDL은 속도 추정할 때 이용하는 정보이며, json 파일의 구조 예시는 아래 표와 같음

[표 33] json 파일 예시 (Metadata)

json 파일 내부구조 예시	json 파일 태그 정보 설명									
<pre>{ "Inje_injetunnel2_20201209_1537_WED_4.8m_NH_highway_OW2_tunnel_FHD-00-of-30.avi": { "branch": "Inje", "installationLocation": "injetunnel2", "date": "20201209", "time": "1537", "dayOfTheWeek": "WED", "height": "4.8m", "peak": "NH", "roadType": "highway", "lane": "OW2", "weather": null, "tunnel": "True", "resolution": "FHD", "VDL_distance": "20", "VDL": "[547, 732, 124, 484, 894, 135, 879, 123]" }, "Inje_injetunnel2_20201209_1537_WED_4.8m_NH_highway_OW2_tunnel_FHD-01-of-30.avi": { "branch": "Inje", "installationLocation": "injetunnel2", "date": "20201209", "time": "1537", "dayOfTheWeek": "WED", "height": "4.8m", "peak": "NH", "roadType": "highway", "lane": "OW2", "weather": null, "tunnel": "True", "resolution": "FHD", "VDL_distance": "20", "VDL": "[547, 732, 124, 484, 894, 135, 879, 123]" }, ... }</pre>	<p>[원천영상 수집환경 정보] branch: 지점명 installationLocation: 설치위치 date: 날짜 time: 시간 dayOfTheWeek: 요일 height: 높이 peak: 첨두 roadType: 도로유형 lane: 차로수 weather: 날씨 tunnel: 터널 (True, False) resolution: 해상도</p> <p>[속도 추정용 정보] VDL_distance: 두 VDL 사이의 간격 VDL: Virtual Detection Line(VDL) 픽셀 좌표</p> <table><tr><th>lane type</th><th>OW</th><th>TW</th></tr><tr><th>Order of points</th><td></td><td></td></tr><tr><th>"VDL" example</th><td>"VDL": [$x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4$]</td><td>"VDL": [$x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, x_5, y_5, x_6, y_6, x_7, y_7, x_8, y_8$]</td></tr></table>	lane type	OW	TW	Order of points			"VDL" example	"VDL": [$x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4$]	"VDL": [$x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, x_5, y_5, x_6, y_6, x_7, y_7, x_8, y_8$]
lane type	OW	TW								
Order of points										
"VDL" example	"VDL": [$x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4$]	"VDL": [$x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, x_5, y_5, x_6, y_6, x_7, y_7, x_8, y_8$]								

2.3.3 정제 절차 및 방법

- ▶ 본 단계에서는 수집된 데이터에 대하여 저화질 영상 및 선명도가 저하된 영상을 보정하는 과정을 수행



정제 방법	작업
1. Super Resolution	<ul style="list-style-type: none"> 저해상도 영상 데이터 입력 Super Resolution 전처리 툴을 통해 원하는 해상도로 확대 선택된 출력 해상도로 영상 출력
2. DeHazing	<ul style="list-style-type: none"> 빛의 반사 또는 산란 등에 의해 선명도가 저하된 영상 데이터 입력 Dehazing 전처리 툴을 통해 화질 개선 입력 포맷과 동일하게 화질 개선된 영상 출력

▶ Super Resolution

- 저해상도 CCTV를 통해 수집된 영상의 해상도 증대를 위해 기존의 영상처리 기법에 의한 방식 뿐만 아니라, 최근의 신경망을 이용한 방법을 통해 확대 영상 화질을 개선함
- GPU 기능이 탑재되지 못한 시스템에서의 실행을 위하여 CPU 만을 이용하는 방법 또한 포함
- 연산 수행 시 GPU 사용 여부 등을 설정하며, 출력을 위한 해상도 변환 비율을 선택함
- 전처리 수행 후, 결과 데이터는 지정된 폴더에 저장됨

항목	지원사항
입력 형태	<ul style="list-style-type: none"> 이미지 또는 영상
픽셀 규격	<ul style="list-style-type: none"> 8-bit unsigned int B(blue)-G(green)-R(red)
출력 해상도 변환 비율	<ul style="list-style-type: none"> 1.5배, 2배, 3배, 4배 (기본: 2배)
출력 형태	<ul style="list-style-type: none"> 이미지 또는 영상 단, array 의 크기는 해상도 변환 비율에 따라 달라짐.
전처리 방식	<ul style="list-style-type: none"> Interpolation 기반 영상 크기 변환 또는 deep learning 기반 영상 크기 변환 적용알고리즘: Bi-cubic Interpolation, SRCNN, FSRCNN, ESPCN

(예) Deep Learning 기반 Super Resolution 적용



그림 15 Super Resolution 적용 전후

▶ Dehazing

- 영상 획득 시 빛의 반사 또는 산란 등에 의하여 선명도가 저하된 영상에 대한 dehazing 기능을 통해 영상 화질을 개선함.
- 입력 영상 조건에 따라 dehazing 기능을 효과적으로 사용하기 위해 전통적인 영상처리 방식 또는 Deep Learning 방식을 선택하여 사용하도록 제공함. Dehazing 기능 외 일부 화질 개선 방법 적용.
- GPU 기능이 탑재되지 못한 시스템에서의 실행을 위하여 CPU만을 이용하는 방법 또한 포함
- 연산 수행 시 GPU 사용 여부 등을 설정하여 GPU 장착 시 수행 속도 개선
- 전처리 수행 후, 결과 데이터는 지정된 폴더에 저장됨

항목	지원사항
입력 형태	<ul style="list-style-type: none"> 이미지 또는 영상
픽셀 규격	<ul style="list-style-type: none"> 8-bit unsigned int B(blue)-G(green)-R(red)
출력 형태	<ul style="list-style-type: none"> 이미지 또는 영상
전처리 방식	<ul style="list-style-type: none"> 이미지 프로세싱 또는 Deep learning 기반 Dehazing 처리 적용알고리즘: Adaptive Histogram Equalization, Dark Channel, AOD-net

(예) Deep Learning 기반 Dehazing 적용



그림 16 Dehazing 적용 전



그림 17 Dehazing 적용 후

2.4 어노테이션/라벨링

2.4.1 어노테이션/라벨링 절차 및 방법

▶ 데이터 어노테이션 및 데이터셋 구축을 위한 상세 프로세스

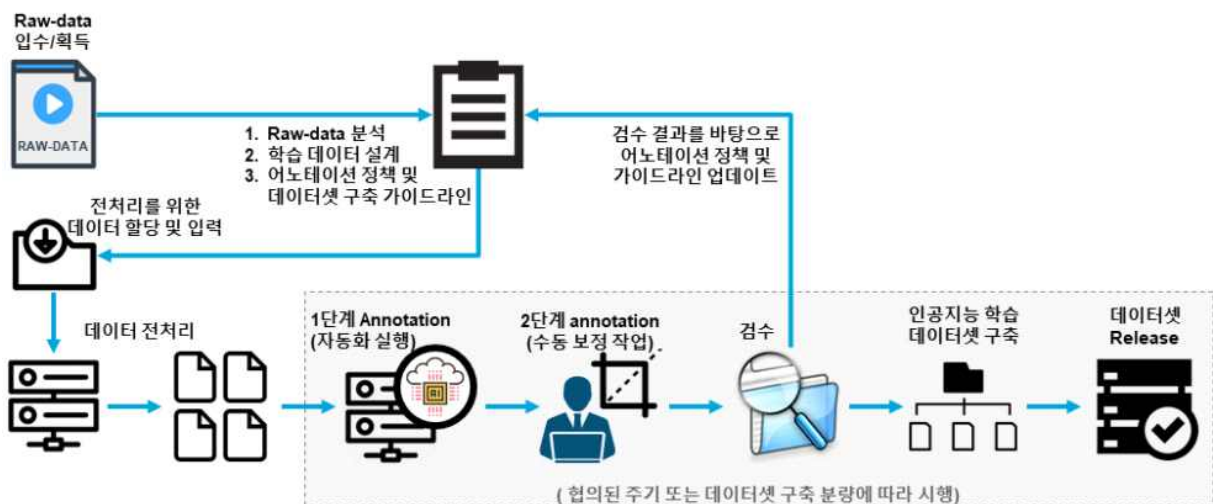


그림 18 학습 데이터 가공 및 제작 프로세스

- 체계화된 프로세스를 통해 높은 정확도의 가공 작업을 수행
- 테스트웍스 자체 인공지능 학습 데이터셋 가공 및 제작 도구인 블랙올리브(Blackolive)의 자동화 기능을 활용하여 1단계 Annotation 작업 수행

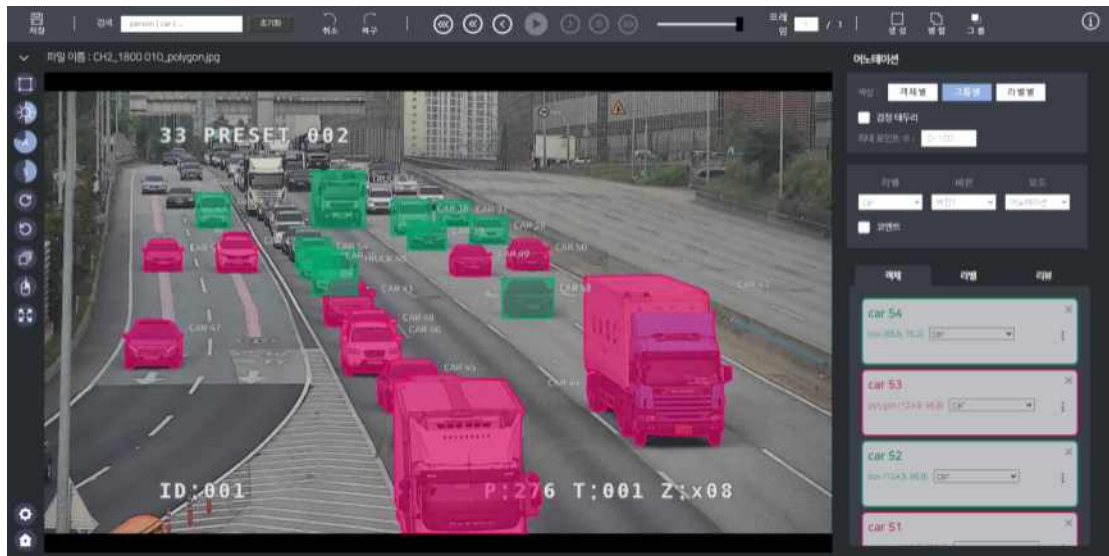
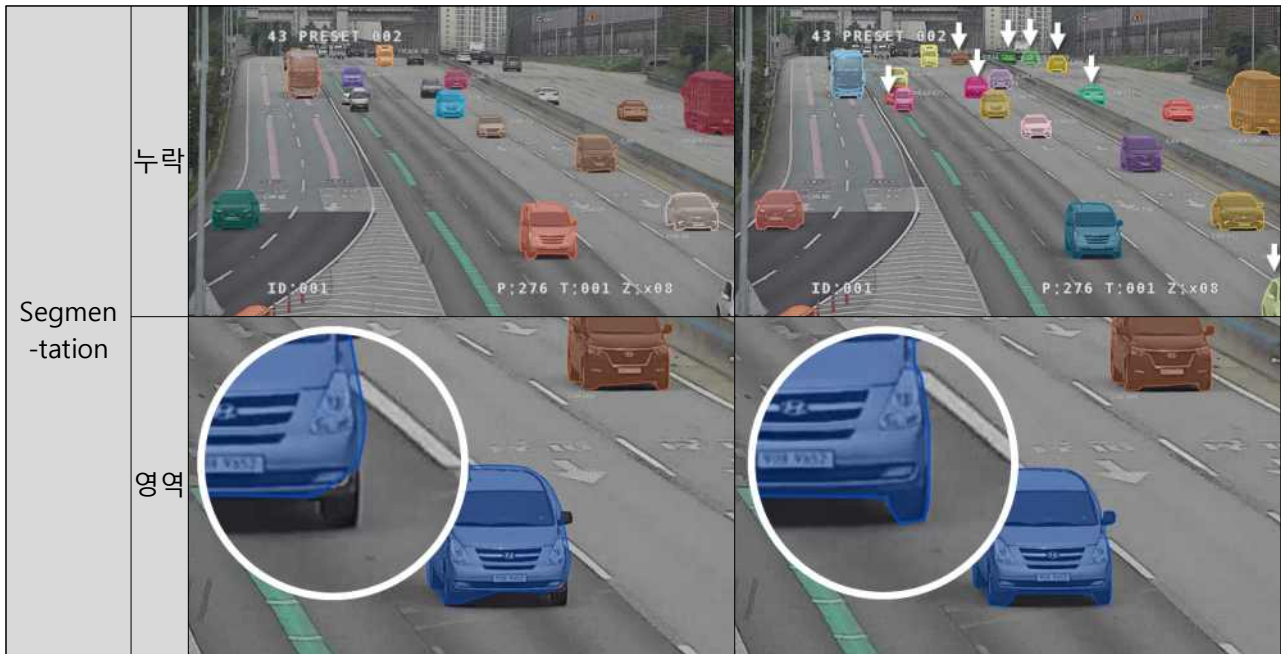


그림 19 테스트웍스의 인공지능 학습 데이터 제작 및 가공 플랫폼

- 클라우드 소싱으로 작업자를 모집하고 온라인-클라우드 기반 작업 환경에서 2단계 Annotation 및 보정 작업을 수행하며 학습 데이터의 정확성과 품질을 높임.
- 2단계 Annotation 과정에서는 학습 데이터 가공 자동화 단계에서 탐지하지 못했거나, 라벨링 및 태깅 오류가 있는 객체들의 보정 작업을 수행

[표 34] Annotation 보정 작업 예시

Annotation Type	오류 사례	1단계 (+자동화)	2단계 (보정)
Bounding Box	누락		
	영역		



▶ 이미지 비식별화

- 원시 데이터 수집 과정에서 의도하지 않은 개인정보가 수집될 수 있고 이는 차후 공개 데이터로 사용할 때 초상권이나 기타 개인 권리를 침해하는 법적 분쟁 소지가 있어서 비식별화 기술을 활용하여 데이터 사용자 및 개인의 권리를 보호하고 법적 분쟁을 미연에 방지
- 정보 비식별화 대상은 운전자 얼굴을 포함한 사람 얼굴과 차량 번호판

[표 35] 비식별화 절차

1. 개인정보 비식별화 대상 객체 (사람 얼굴 및 자동차 번호판)에 대해 Bounding Box 가공 수행	2. 영상 내 얼굴과 차량 번호판의 바운딩 박스에 해당하는 부분에 대해 OpenCV Library를 활용한 Blur 처리 프로그램으로 비식별화



그림 20 자동차 번호판 비식별화 예시 (좌 : 비식별화 전, 우 : 비식별화 후)

- 클라우드워커의 비식별화 참여시 전원 보안서약서 작성 필수

[별지 제10호 서식] 보안서약서

보안서약서

본인은 " 교통안전 AI 데이터 "사업 수행의 일원으로서 참여하면서 다음 사항을 준수할 것을 서약합니다.

1. 본 사업을 수행하는 과정에서 알 수 있었던 정보통신망상 분야 기밀에 대하여 사업 수행 중은 물론 종료 후에도 수행기관장 또는 전담기관장의 허락 없이 자신 또는 제3자를 위하여 사용하지 않는다.
2. 본 사업 추진결과가 적법하게 공개된 경우라고 하여도 미공개 부분에 대해서는 제1항에서와 같이 반드시 비밀유지 의무를 부담한다.
3. 본 사업이 완료되거나 사업을 수행할 수 없게 된 경우, 그 시점에서 본인이 보유하고 있는 기밀을 포함한 관련 자료를 즉시 총괄책임자에게 반납하며 제1항 및 제2항에서와 같이 비밀유지 의무를 부담한다.
4. 또한 퇴직 시 본인은 직무상 지득한 핵심기술 및 정보, 정보통신망상 분야 관련 제반 비밀사항 및 중요 기술비밀을 퇴직 후에도 일체 누설하지 않는다.
5. 상기 사항을 위반할 경우 본인은 정보통신망상사업에서 불이익을 받을 수 있으며, 관련 법률에 따른 민형사상 책임을 질 것을 서약합니다.

서약인 성명 : 김 [REDACTED] (인)

2021. 01. 11 .

한국정보화진흥원장 귀하

※ PMS의 참여인력에 기재하는 모든 참여인력에 대해 작성하여 제출

그림 21 실제 보안서약서 샘플

2.4.2 어노테이션/라벨링 기준

▶ 검출 클래스 정의

- 검출 Class의 종류는 car, truck, bus 3종으로 분류함
- car : 일반 세단형 및 해치백 승용차와 SUV, RV, 픽업트럭, 12인 승급 승합차 (카니발, 스타렉스, 봉고차)등 개인이 가정에서 운용할 수 있는 차량의 범주를 모두 포함



그림 22 car 객체 class의 예시 1 (Bounding Box)



그림 23 car 객체 class의 예시 2 (Polygon Segmentation)

- bus : 미니버스, 마을버스를 포함하는 승합용 버스
- truck : SUV에서 파생된 픽업 트럭 차종을 제외한 모든 화물 적재가 가능한 차량을 truck class로 구분

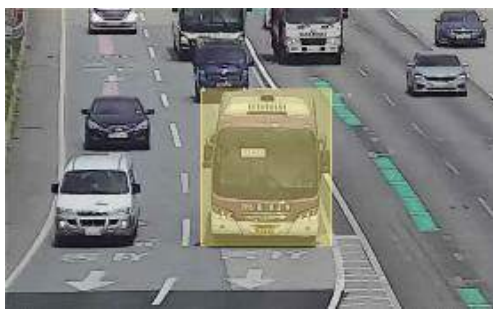


그림 24 bus 객체 class의 예시 1
(Bounding Box)

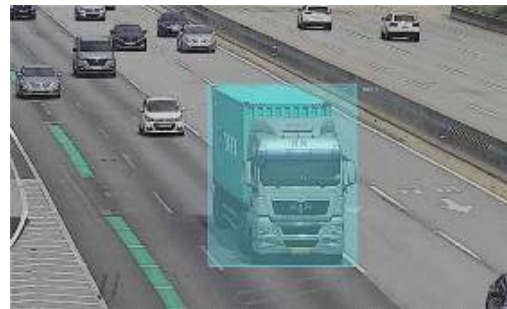


그림 25 truck 객체 class의 예시 1
(Bounding Box)



그림 26 bus 객체 class의 예시 2
(Polygon Segmentation)



그림 27 bus 객체 class의 예시 2
(Polygon Segmentation)

- 총 이미지 데이터셋 50만장 (100%)에 대하여 검출 Class별 학습 데이터 가공 방식별 구축 비율은 아래 표와 같이 계획

[표 36] 검출 Class 및 학습 데이터 가공 방식별 구축 비율

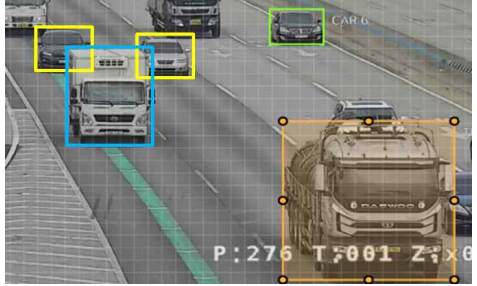
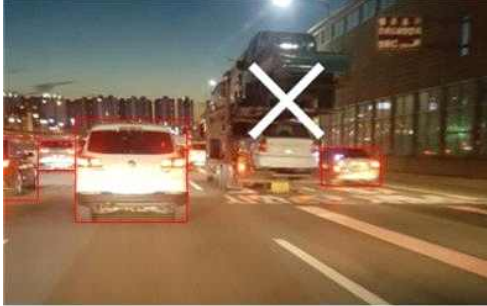

Class 구분	Bounding Box	Polygon Segmentation	합계
Car	1,500,000	1,000,000	2,500,000
Bus	150,000	100,000	250,000
Truck	150,000	100,000	250,000

▶ 어노테이션 정책 및 가이드 라인

- 데이터 가공을 위한 세부 가이드라인은 표 37, 38과 같으며 데이터 검수 시에도 품질 평가 체크리스트로 활용


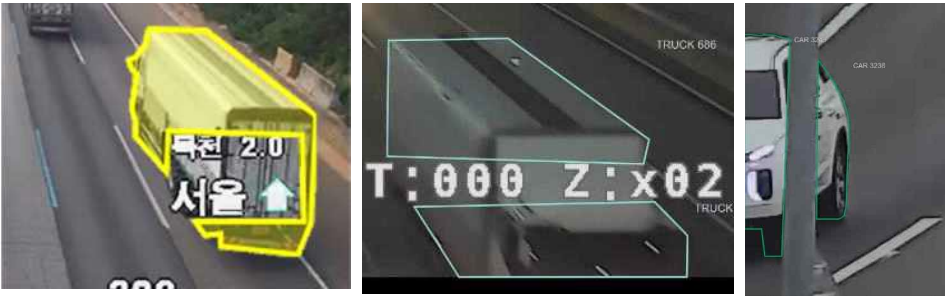
[표 37] Bounding Box 어노테이션 세부 가이드라인 및 예외처리 방안

항목	가이드라인
차량 최소 크기	<ul style="list-style-type: none"> 단축 기준 30 pixel 이상
차량 바퀴 부분	<ul style="list-style-type: none"> 차량 바퀴가 보이는 경우 바퀴 부분을 포함하여 라벨링 야간, 악천후, 그림자 등으로 바퀴가 명확히 보이지 않는 경우, 차량의 윤곽선만 따라서 라벨링
화면 밖 잘린 차량	<ul style="list-style-type: none"> 차량이 50% 이상 보이는 경우, bbox 좌표가 이미지 영역 안에서 존재하도록 라벨링
견인차(렉카), 사고차량	<ul style="list-style-type: none"> 견인하는 차량이 명확히 보이면 라벨링 견인되는 차량(사고차량) 라벨링 제외 견인하는 차량이 명확히 보이지 않으면 라벨링 제외
짐 실은 차량	<ul style="list-style-type: none"> 적재함에 물건을 포함하여 라벨링 컨테이너를 싣고 가는 차량, 트럭은 컨테이너 포함하여 라벨링 뒤에서 봤을 때, 적재함 물건이 차량보다 지나치게 높게 쌓여 차량 식별이 힘든 경우 라벨링 제외
캠핑카, 카라반	<ul style="list-style-type: none"> 카라반을 끌고 가는 모체 차량은 해당 차량의 기준 클래스로 라벨링 뒤에 연결된 카라반은 라벨링 제외
안테나 및 외부 구조물	<ul style="list-style-type: none"> 사이드미러 포함하여 라벨링 순찰차, 택시 위에 설치된 표시등 포함하여 라벨링 안테나는 라벨링 제외

차량 간 겹침으로 뒷 차량이 보이지 않는 경우	<ul style="list-style-type: none"> 객체간 가려지거나 겹쳐진 영역에 대해서는 예측되는 범위만큼 추론하여 가공 동영상 데이터에 정보를 표시하는 텍스트와 가공 대상 객체가 겹쳐지는 경우 텍스트를 무시하고 가공 차량의 위치를 가늠하기 어려운 경우 라벨링 제외 	
장애물(가로등, 표지판, 텍스트)로 인해 차량이 가려지는 경우		
기타사항	<ul style="list-style-type: none"> 포크레인(굴삭기)와 오토바이는 가공하지 않음 (고속도로 주행 대상 아님) 차량이 트레일러에 실린 경우 싣고 가는차, 싣려가는 차 모두 가공하지 않음 견인 상황 발생시 끌고가는 견인차만 가공하며 끌려가는 고장 차량은 가공 제외 	 

[표 38] Polygon Segmentation 어노테이션 세부 가이드라인 및 예외처리 규칙 명세

항목	가이드라인
차량 최소 크기	<ul style="list-style-type: none"> bbox 기준으로 라벨링되는 차량은 모두 라벨링
차량 바퀴 부분	<ul style="list-style-type: none"> 바퀴가 명확히 보이는 경우 바퀴 윤곽선을 따라 라벨링 날씨(악천 후, 맑은 날)와 상관없이 그림자 등으로 바퀴가 명확히 보이지 않는 경우, 차량의 바퀴를 유추하여 라벨링
화면 밖 잘린 차량	<ul style="list-style-type: none"> 차량이 50% 이상 보이는 경우, polygon 좌표가 이미지 영역 안에서 존재하도록 라벨링
견인차(렉카), 사고차량	<ul style="list-style-type: none"> 견인하는 차량이 명확히 보이면 라벨링 견인되는 차량(사고차량) 라벨링 제외 견인하는 차량이 명확히 보이지 않으면 라벨링 제외
짐 실은 차량	<ul style="list-style-type: none"> 적재함에 물건을 포함하여 라벨링 컨테이너를 싣고 가는 차량, 트럭은 컨테이너 포함하여 라벨링 뒤에서 봤을 때, 적재함 물건이 차량보다 지나치게 높게 쌓여 차량 식별이 힘든 경우 라벨링 제외
캠핑카, 카라반	<ul style="list-style-type: none"> 카라반을 끌고 가는 모체 차량은 해당 차량의 기준 클래스로 라벨링 뒤에 연결된 카라반은 라벨링 제외
안테나 및 외부 구조물	<ul style="list-style-type: none"> 사이드미러 포함하여 라벨링 순찰차, 택시 위에 설치된 표시등 포함하여 라벨링 안테나는 라벨링 제외

<p>차량 간 겹침으로 뒷 차량이 보이지 않는 경우</p>	<ul style="list-style-type: none"> 객체 간 가려지거나 겹쳐진 영역에 대해서는 사람의 추론에 의한 가공이 아닌 최대한 사람 시각에 보이는 부분만 가공 	
<p>이미지 내 장애물(가로등, 표지판, 텍스트)로 인해 차량이 가려지는 경우</p>	<ul style="list-style-type: none"> 가로등, 표지판으로 분할된 차량 영역은 그룹핑하여 라벨링 동영상 데이터에 정보를 표시하는 텍스트와 가공 대상 객체가 겹쳐지는 경우 텍스트 영역을 우회하여 가공 기둥과 같은 구조물에 차량이 부분적으로 가려진 경우, 가려진 영역을 추론하여 가공하지 않고, 보이는 부분을 여러 조각으로 가공하여 그룹 지정을 해야 함 	
<p>기타사항</p>	<ul style="list-style-type: none"> Bounding Box와 다른 중요한 차이점으로 Polygon은 겹침, 가림, 사람의 추론과 판단이 반영된 가공을 지양하고 가공 대상 객체에서 최대한 사람의 시각으로 인지된 부분만을 가공해야 함 차량이 아닌 영역은 폴리곤 영역에 포함되면 안됨 차량이 전체 면적의 50% 이상이 화면 밖으로 벗어난 (반대로 화면 안에 있는 차량의 면적이 50% 미만인 경우) 가공 제외함 Bounding Box의 기타사항을 포함 	

2.4.3 어노테이션 조직

▶ 어노테이션 조직 구성

- 어노테이터 : 최초 어노테이션을 수행하거나 자동화 처리 후 2단계 어노테이션 보정
- 리뷰어 : 어노테이터의 작업 결과를 리뷰하고, 미세/소수 오류에 대해서 직접 보정작업을 수행하거나, 리뷰 대상 데이터셋의 품질이 아주 나쁜 경우 보고서 작성 및 어노테이터에서 재작업 지시
- 매니저 : 리뷰어의 리뷰 결과에 대한 최종 검수 수행

▶ 작업자에 대한 실무 전문 교육 시행

- 인공지능 학습용 데이터 가공 서비스 품질 향상을 위해 데이터 매니저 양성 커리큘럼을 다음과 같이 구성

[표 39] 청년 작업자 교육 커리큘럼 구성 방안

회차	교과목명	세부 교과 내용	시간
1	오리엔테이션, 어노테이션 이해	- 어노테이션 소개 및 기본 교육 - 블랙올리브 Tool 매뉴얼 교육	4
2	어노테이션 교육	- OD 매뉴얼 교육 - 바운딩 박스 샘플 작업	4
3	기초 OA(엑셀, 파워포인트, 워드)	기초 OA 활용 교육	4
4	어노테이션 실습 및 피드백	어노테이션 작업	4
5	어노테이션 매뉴얼 해석, 제작, 적용 실습	샘플 고객사 요구사항 매뉴얼 해석, 수정, 제작	8
6	-어노테이션 최종 평가 실습 - 직장문화이해 교육	- 어노테이션 최종 평가(작업 1, 리뷰 1 각 1건씩) - 직장 문화 이해 교육(사내 매너 교육)	8

2.4.3 어노테이션 관리 도구

- ▶ 본 가이드라인의 어노테이션 방식은 Bounding Box와 Polygon Segmentation 2가지 기능으로 작업 환경 GUI 예시는 본 가이드라인 2.4.1 참고
- ▶ 데이터셋 가공 관리 도구
 - 효율적인 일정, 진척도 및 품질 관리를 위해 아래 기능을 포함하고 있는 웹 기반의 관리 도구 활용 권장
 - 업무 분배, 프로젝트 및 작업자 관리, 검수 프로세스 관리 과정 모니터링
 - 게시판을 통해 작업자, 검수자, 관리자 간 이슈 공유 및 내용 트래킹
 - 작업자와 검수자는 할당받은 모든 작업물을 리스트 형식으로 모니터링
 - 작업의 진행률, 반려율, 검수 현황에 대한 대시보드
 - 검수 요청 현황 확인, 검수자와 작업자 변경, 작업물 다운로드 등
 - 작업자별, 프로젝트별 통계를 추출 및 그래프 출력



그림 35 데이터셋 가공 관리 도구 예시

2.5 검수

2.5.1 검수 절차 및 방법

▶ 단계별 품질 관리 활동



그림 36 단계별 품질 관리 프로세스

- ▶ 인공지능 학습용 데이터를 구축하면서 수행하는 전체 공정 단계를 기획, 설계, 수집, 정제 및 가공, 검수, 검수 후 단계로 구분하여 단계별로 데이터 품질을 유지 및 개선하기 위한 활동을 수행
- ▶ 데이터셋 정제 및 가공 단계의 품질 관리
 - 수집된 데이터를 가공 과정의 프로세스, 작업 가이드, 오류 개선 활동을 중심으로 품질 관리 활동을 수행함
- ▶ 데이터셋 검수 단계의 품질 관리
 - 가공된 데이터를 검수하는 단계에서 프로세스, 도구, 조직 체계를 검토하여 일정한 품질의 데이터셋 구축 결과를 확인함
- ▶ 데이터셋 검수 후 단계의 품질 관리
 - 최종 결과물의 전달 과정에 관한 확인

2.5.2 검수 기준

- ▶ 전문 인력에 의한 검수 시 어노테이션 불량 판단 기준
 - 본 문서 [표 37~38] 어노테이션 세부 가이드라인 및 예외처리 규칙 명세에 따라 검수하며, 해당 가이드라인을 준수하지 않은 객체를 어노테이션 불량으로 판단함.
 - Bounding Box 어노테이션 범위 불량 판단 기준 : 어노테이션 영역이 태깅 대상 객체 상하좌우 모든 부분을 10픽셀 안쪽으로 들어오거나, 바깥쪽으로 10픽셀 이상 벗어난 경우
 - Polygon Segmentation 어노테이션 범위 불량 판단 기준 : 자동차임을 특징할 수 있는 사이드 미러의 형상과 바퀴의 형상이 어노테이션 범위에 포함하지 않는 경우

▶ Bounding Box 정확도 품질 측정 방법

- 최종 구축 및 Release 완료된 학습 데이터셋에서 이미지(+xml) 100장 샘플 추출
- 이미지 1장당 아래의 품질 검사 수식을 적용하고, 이미지 100장에 대한 평균값 도출

$$(\text{Bounding Box 정밀도 품질}) = \frac{x - (y_1 + y_2 + y_3 + y_4)}{x} \times 100$$

- x = 이미지 내 어노테이션 해야 할 객체의 수
- y_1 = 오검출된 객체의 수
- y_2 = 미검출된 객체의 수
- y_3 = 과 검출된 객체의 수
- y_4 = 어노테이션 범위 불량인 객체의 수

▶ Polygon Segmentation 정확도 품질 측정 방법

- 최종 구축 및 Release 완료된 학습 데이터셋에서 이미지(+xml) 100장 샘플 추출
- 이미지 1장당 아래의 품질 검사 수식을 적용하고, 이미지 100장에 대한 평균값 도출

$$(\text{Polygon Segmentation 정밀도 품질}) = \frac{x - (y_1 + y_2 + y_3 + y_4 + y_5)}{x} \times 100$$

- x = 이미지 내 어노테이션해야 할 객체의 수
- y_1 = 오검출된 객체의 수
- y_2 = 미검출된 객체의 수
- y_3 = 과검출된 객체의 수
- y_4 = 어노테이션 범위 불량인 객체의 수
- y_5 = 객체 간 그룹화 오류의 수

▶ 객체검출/객체분할 정확도 평가 방법

- Average Precision(AP)을 활용하여 모델의 정밀도를 평가하며, AP는 컴퓨터 비전의 객체검출 및 이미지 분류 모델의 성능을 평가하는 주요 척도로 AP가 1에 가까울수록 모델의 성능이 우수

2.5.3 검수 조직

▶ 품질 관리 조직 구성

- 전반의 품질 관리 및 검증을 위한 품질 책임자와 품질관리위원회 담당으로 품질 관리 수행
- 각 역할에 따른 품질 관리 역할 수행

▶ 데이터셋 3단계 검수

- 3단계의 검수 절차를 거쳐 오검출, 미검출, 과 검출, 어노테이션 범위 불량인 객체가 없는지 검수하여 데이터 품질을 관리
- Peer Review : 클라우드 소싱으로 참여한 작업자 중 높은 수행 성과를 달성한 작업자 간 리뷰 진행, 데이터 품질 목표치의 95% 확보
- Manager Review : 작업자와 프로젝트 매니저 간에 리뷰 진행, 품질 기준에 미달한 데이터셋은

재작업 수행으로 데이터 품질 목표치의 98% 확보

- Final Review : 최종 전달 이전에 2단계 검수 절차에서 품질 기준을 충족한 데이터 중 샘플 데이터 추출 및 검수를 통해 데이터 품질 목표치의 100% 확보



그림 37 인공지능 학습 데이터 3단계 수동 검수 프로세스

▶ 데이터셋 오류 유형 및 예시

유형	내용
오검출	가공된 데이터를 검수하는 과정에서 잘못된 레이블을 붙여 태깅하거나 레이블에 포함되지 않은 객체를 태깅한 경우의 오류를 범하여 객체가 발견되는 경우
미검출	가공된 데이터를 검수하는 과정에서 태깅해야 할 객체를 태깅하지 않은 오류가 발견된 경우
과검출	가공된 데이터를 검수하는 과정에서 정의된 객체에 해당하는 객체이지만 객체의 크기, 선명도 등에 의해 태깅에서 제외해야 할 객체를 태깅하는 오류가 발견된 경우
어노테이션 범위 불량	객체를 어노테이션할 때 가이드라인에서 규정한 범위에 못미치거나 벗어나서 어노테이션하는 오류가 발견된 경우

2.5.4 기타 품질 관리 활동

▶ TTA 품질점검

- 한국정보통신기술협회(TTA)에서 마련한 AI 데이터 구축 분야별 표준/가이드라인을 통해 운영 단계 품질검증 실행

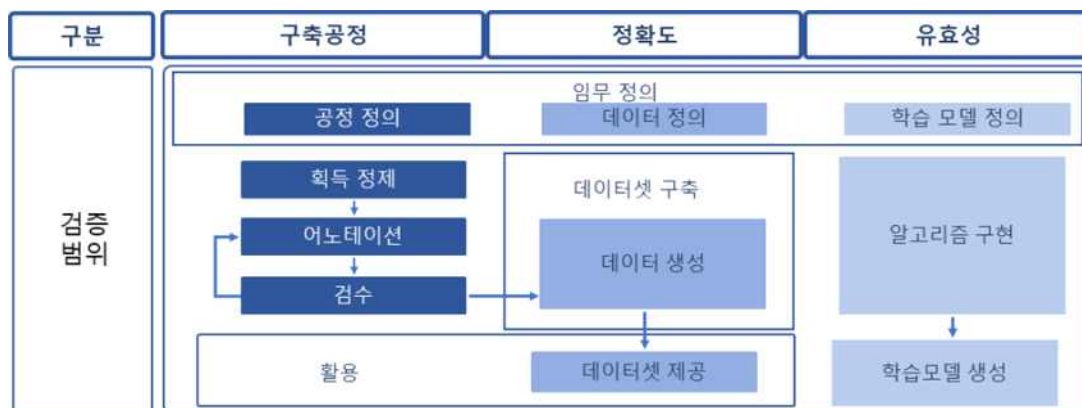


그림 38 TTA 인공지능 데이터 품질검증 체계

- 검증대상: 공정 전주기, 데이터 및 저장소, 학습모델

- 검증기준: 데이터 구축활용 가이드(예시, 안)
- 구축공정: 체크리스트 이행 여부
- 정확도: 통계정보, 데이터, 어노테이션, 저장소 구조, 참값(Ground Truth)
- 유효성: 학습 성능
- 검증방법: 검증 범위, 항목 선정 → 검증 기준, 절차, 방법 → 검증 결과 확인 및 분석
- 구축공정: 문서검토, 수행기업 인터뷰, 현장점검 및 근거자료 확인
- 정확성: 전수 또는 샘플링 검사, 자동화 검수도구, 검증 데이터 분석
- 유효성: 학습 조건 설정 및 수행(데이터 구분, 반복 횟수 등)
- 검증결과: 품질검증 결과서(구축공정, 정확도, 유효성)

2.6 활용

2.6.1 활용 모델

▶ 교통흐름분석 모델 개발

- CCTV에서 수집된 영상정보를 기반으로 실시간 차로단위 교통흐름(교통량, 속도) 및 돌발상황 및 대기행렬을 분석하고, 도로상황정보를 모니터링
- 교통파라미터 : 속도(차량ID별), 평균속도(차로별), 교통량(차로별) 등



그림 39 고속도로 교통흐름 분석 모델 개발 절차

- 참조용 학습 모델 개발 안
- 객체(차량)검출을 위한 Bounding box기반 객체검출(Detection) 모델과 Mask 기반 객체분할(Segmentation) 모델 개발 2종
- 참조용 학습 모델의 목표 성능지표
- 객체 검출(Detection) 모델 : Box mAP@0.5, 60%
- 객체 분할(Segmentation) : Mask mAP@0.5, 40%

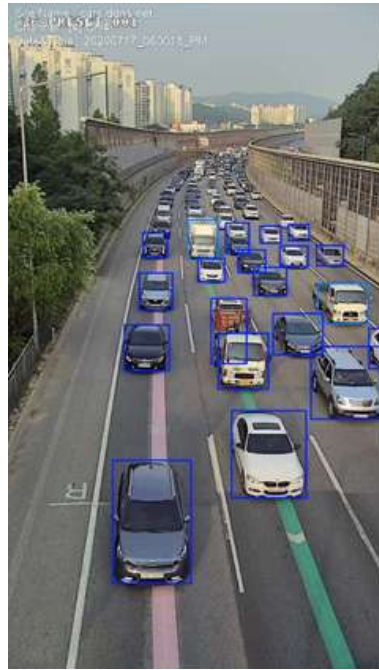


그림 40 Detection(좌) 및 Segmentation(우) 알고리즘 적용 예시

- 속도 추정 알고리즘 개발
 - CCTV 영상에서 교통량 및 속도 추정을 위한 차로단위 영역 분할, Virtual Line 생성
 - 영상 내 Virtual Line을 생성하여 차량이 Virtual Line을 지나가는 Time-stamp를 측정 후 개별 차량의 속도를 계산
 - 검출 후 추적(Tracking by Detection) 알고리즘을 이용한 실시간 차로단위 교통흐름 분석



그림 41 (left) 차로 단위 영역 분할, (right)속도 추정을 위한 Virtual Line

- 트래킹 알고리즘 개발
 - 객체검출모델로 검출된 차량의 Bounding Box 좌푯값을 이용하여, 차량 추적 알고리즘 수행
 - 차량 추적 알고리즘으로 DeepSORT 또는 IOU-tracker를 고려하고 있으며, 중간 프레임에서 차량 검출이 실패하더라도 차량을 지속해서 추적할 수 있는 추적 알고리즘 개발 예정

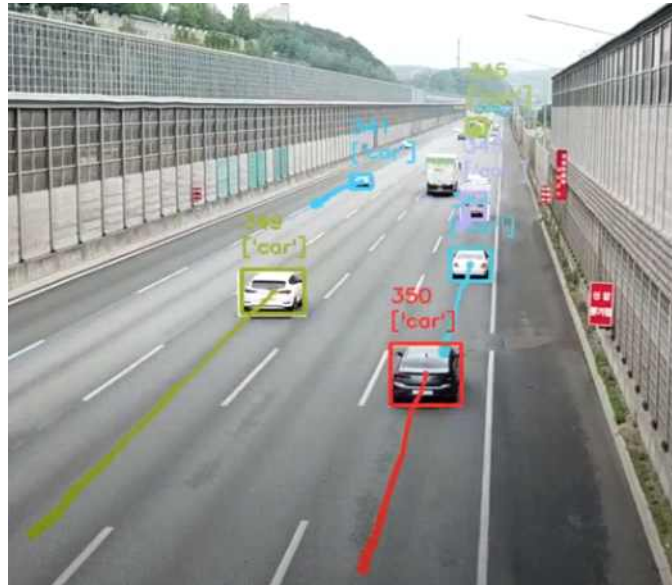


그림 42 차량 검출 및 추적 결과

- 객체검출 성능개선을 위한 Instance Segmentation을 이용한 교통흐름 분석
- 고속도로 내 교통 정체 시 객체검출 모델의 성능이 현저히 열화되며, 이는 Tracking by Detection 시스템에서 Tracking 모델의 성능 열화를 유발
- 이는 정체로 인해 차량이 겹쳐져 Bounding Box로는 식별이 어려워 발생하는 문제이며, 픽셀단위로 검출을 하는 Instance Segmentation 모델을 적용하여, 차량 간 occlusion이 큰 상황에서도 효과적으로 차량 검출이 가능
- 단, Segmentation 모델은 객체검출 모델에 비해 연산량이 크며, 실시간 영상분석에는 한계가 있어, 교통흐름 모니터링 요구사항에 따라, 선택적으로 적용되어야 할 것으로 판단됨
- YOLACT, DeepSnake, CenterMask와 같은 실시간 Instance Segmentation 모델을 사용하여 실시간성을 보장하는 Instance Segmentation 모델 적용



그림 43 Occlusion 환경에서 Instance Segmentation(예시)

2.6.2 서비스 개발

- ▶ - 교통흐름분석 모델기반 교통 파라미터 활용 서비스 개발
- ▶ AI 기반 교통흐름분석 모델을 활용하여 아래의 2가지 교통파라미터를 생성할 수 있으며, 교통파라미터를 활용하여, 교통 흐름 모니터링 서비스를 개발할 수 있음.
 - cctv-traffic.json: 차량별 속도, 위치 정보
 - cctv-traffic-vds.json: 차로별 속도, 교통량 정보

[표 40] cctv-traffic.json 파일 예시

cctv-traffic.json 파일 내부구조 예시	cctv-traffic.json 파일 태그 정보 설명
<pre>{ "traffic": [{ "id": "1", "label": "car", "lane": "2", "direction": "down", "speed": 63, "bbox": [[926, 651], [1013, 719]] }], "highway": "2", "cctv": "1", "preset": "1", "timestamp": 1611646274 }</pre>	<p>[개별 차량 정보] id: 차량별 아이디 label: 차량 종류 (car, bus, truck) lane: 주행 차로 (1,2,3 차로) direction: 차량 진행 방향 (up, down) speed: 차량 속도 bbox: [[left_up x, left_up y], [right_down x, right_down y]]</p> <p>[CCTV 정보] highway: cctv 설치 지사명 cctv: cctv 설치 위치 preset: cctv의 현재 프리셋 timestamp: json 파일 생성 시간(code time)</p>

[표 41] cctv-traffic-vds.json 파일 예시

cctv-traffic-vds.json 파일 내부구조 예시	cctv-traffic-vds.json 파일 태그 정보 설명
<pre>{ "highway": "2", "cctv": "1", "preset": "1", "timestamp": 1611646351, "upVolume": [3, 2, 0], "downVolume": [</pre>	<p>[CCTV 정보] highway: cctv 설치 지사명 cctv: cctv 설치 위치 preset: cctv의 현재 프리셋 timestamp: json 파일 생성 시간(code time)</p> <p>[차로별 정보] upVolume: 상행 차로별 교통량 (상행 1, 2, 3 차로) downVolume: 하행 차로별 교통량 (하행 1, 2, 3 차로) upAvgSpeed: 상행 차로별 평균속도 (상행 1, 2, 3 차로) downAvgSpeed: 하행 차로별 평균속도 (하행 1, 2, 3 차로)</p>

<pre> 1, 1, 0], "upAvgSpeed": [93, 93, -1], "downAvgSpeed": [90, -1, -1] } </pre>	
--	--

[표 42] highway, cctv 값 정의

highway	지사	CCTV	설치위치
1	영동선	1	CH01
		2	CH02
		3	CH03
		4	CH04
		5	CH05
		6	CH06
		7	CH07
		8	CH08
		9	CH09
		10	CH10
2	부산항대교	1	부산항대교1 감만 방면
		2	부산항대교2 감만 방면
		3	부산항대교3 영도 방면
		4	부산항대교4 영도 방면
		5	부산항대교5 감만 방면
		6	부산항대교6 감만 방면
3	강릉지사	1	노동교
		2	노동1교
4	충주지사	1	단평1교
		2	상우1교
5	부안지사	1	종신
		2	대목
6	전주지사	1	정읍
		2	삼평육교
7	천안지사	1	강정
		2	사정



그림 44 교통파라미터 활용 서비스 개발(예시)

2.6.3 데이터 제공

- ▶ NIA AIHub 를 통한 기술혁신 지원
 - 개발된 인공지능 모델을 정보화진흥원 AIHub 를 통해 제공
 - 또한, AI 모델을 적용한 응용서비스를 공개하여 국내 AI기업 및 연구소, 학교 등에서 고속도로 CCTV 관련 AI 서비스 활성화 지원
 - 구축된 학습 데이터셋을 산학연에서 인공지능 학습 및 AI 서비스 개발 등에 활용할 수 있도록 활용 매뉴얼 제공

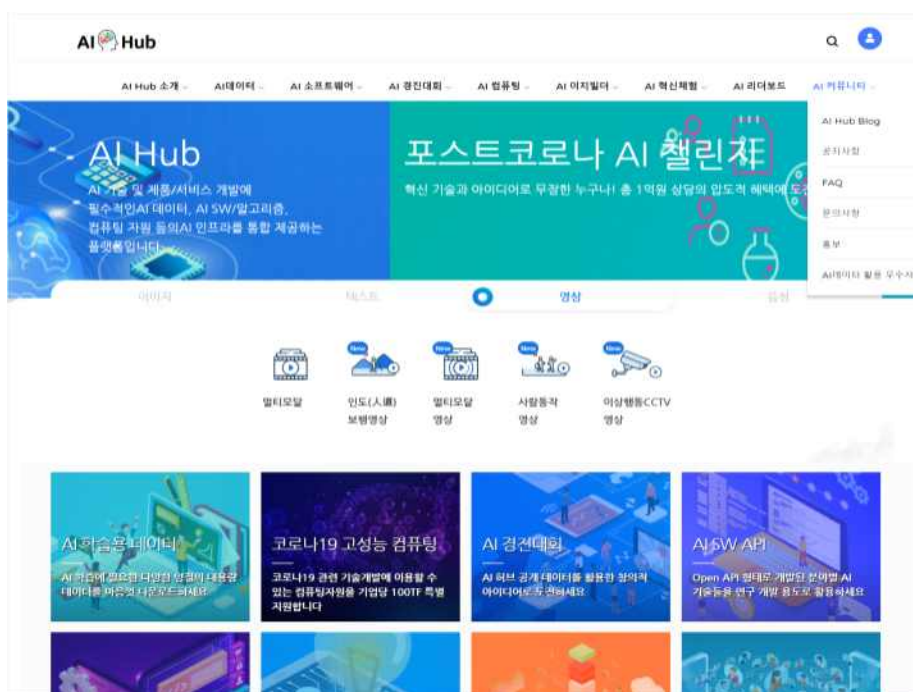


그림 45 NIA AIHub

- ▶ 외부활용 기업 및 기관 기술적 지원 방안
 - 지능형 도로 및 자율주행 등 유사 인공지능 기술 활용 분야의 주요 서비스 기업의 수요 조사 활동
 - 모델의 성능 수준 정보 제공 및 가능 시나리오 정보 등을 제공하고 실제 서비스 개발에 대한 수요가 있을 시, 관련 데이터셋 추가 구축 및 API를 통한 모델 연동 제공

- 모든 구축 공정별로 산출되는 데이터셋, 저작도구, 모델 등은 개발 및 구축 완료 후 공개를 원칙으로 하며 공개 후 활용이 용이하도록 각 산출물에 대한 활용 방안을 매뉴얼화하여 함께 공개
 - 지적재산권 이슈 등 직접 공개가 어려우면, 공개 가능 범위 확인 및 활용 가능한 형태로 재가공 후 공개
 - 기술 지원 팀을 구성하고 데이터셋 활용 관련 기업, 기관 등 문의 시 기술 지원 수행
- ▶ 자체 사업을 통한 지속적 데이터 확대 및 고도화 방안
- 예산과 시간, 자원, 원시데이터 접근성 등 현실적인 제약사항으로 인해 유관 산업 분야의 많은 중소기업들이 고속도로 교통 정보를 인공지능 학습용 데이터셋으로 단기간 내 구축하는 데 어려움이 있음
 - 이에, 향후 데이터의 증가 및 활용 분야 확장이 가능하도록 고속도로 교통 정보 데이터셋 구축을 위한 가장 효율적, 효과적 방안을 실험하고 기틀을 구축하고자 함
 - 인공지능 데이터셋 구축 전문기업인 주관사 테스트웍스는 단순히 기업의 이윤뿐 아니라 지속적인 소셜 임팩트 창출을 추구하는 사회적 기업
- ▶ 산출물의 AIHub 공개를 통한 자생적 데이터 확산 생태계 마련
- 원시 데이터 영상, 가공 데이터셋, 저작도구, 매뉴얼, 모델 등을 AIHub을 통해 공개함으로써 인공지능 연구 및 서비스 개발 등에 자유롭게 활용될 수 있도록 기여
 - 데이터셋 활용이 쉽도록 공정별 수행 내용을 상세히 매뉴얼화 하여 공개
 - 주관기업의 홈페이지 내 AIHub 링크를 통해 공개 데이터셋 접근성 확대
 - 본 데이터셋 구축 및 모델 학습 내용을 논문 형태로 산출 및 GitHub 등에 게시하고 AIHub으로 링크를 제공함으로써 데이터 확산에 기여

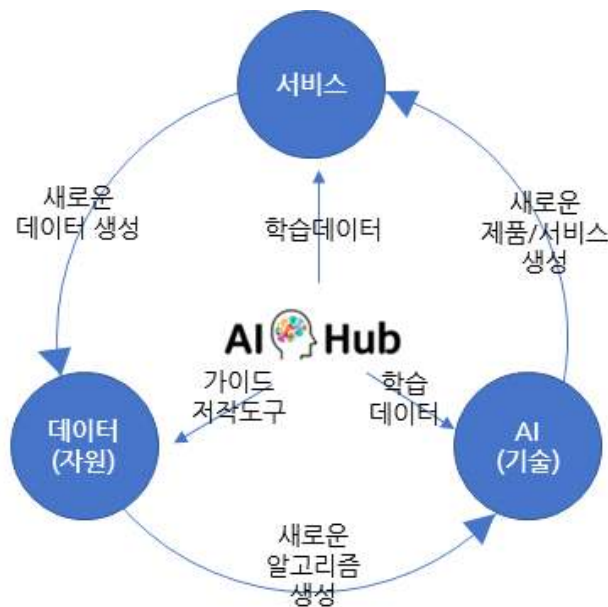


그림 46 데이터 확산 생태계 구현 방안

2.6.4 데이터 유지보수

- ▶ 공개 데이터로 제공된 인공지능 학습 데이터셋 유지 보수

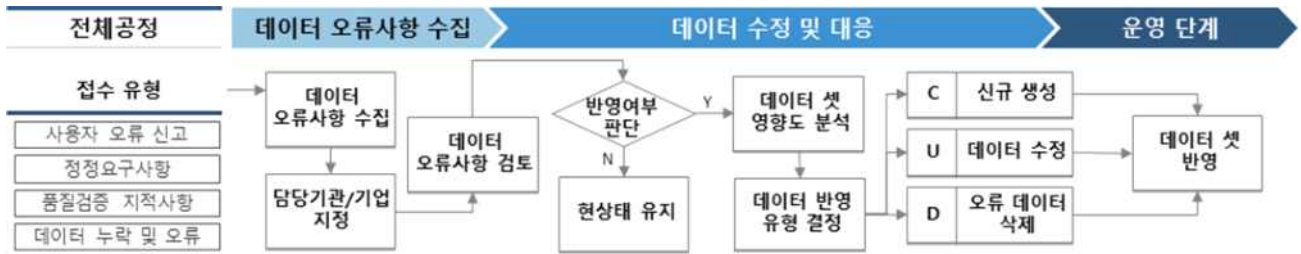


그림 47 데이터 오류 사항 처리 절차

부록

1. VDS 속도추정 GT 데이터

1.1. VDS 개요

▶ VDS(Vehicle Detection System)

- VDS란 도로의 특정 지점의 교통정보를 검지하여 해당 지점의 속도, 교통량을 분석하는 시스템
- 본 VDS 데이터는 radar를 이용한 교통량, 차량 속도를 검지하여 사용자가 차로별 교통량, 차로별 평균속도를 ground truth 데이터로 활용할 수 있도록 영상데이터를 함께 제공함

1.2. 원시데이터 수집 및 가공


▶ 원시데이터의 형태 및 규모




- 원시데이터 영상은 5개 지점에서 30분 단위로 영상을 획득
- 총 10시간 분량의 영상 데이터와 xlsx(엑셀) 데이터를 제공함
- 영상의 화질은 FHD 1920x1080
- 개인정보 비식별화를 위해 차량 번호판 영역은 블러(blur) 처리함

▶ VDS 검지 영역

- virtual detection line을 이용하여 영상에서 차량의 인식 구간을 표시함
- 차량이 두 번째 VDL에 진입한 시간을 기록함
- 해당 영역에서 차선을 침범하는 경우 검지 대상에서 제외됨
- 차량의 전면부가 보이는 방향이 VDS 검지 영역으로 반대 방향은 검지하지 않음

[표 43] VDS 인식구간 VDL

설치 위치	VDL pixel 좌표	지점 샘플 이미지
VDS-01	[(412, 514), (794, 498), (388, 776), (941, 746)]	

VDS-02	[(486, 520), (764, 493), (470, 755), (885, 707)]	
VDS-03	[(869, 407), (1196, 387), (847, 700), (1344, 676)]	
VDS-04	[(810, 433), (1252, 426), (715, 747), (1388, 754)]	



VDS-09	[(387, 540), (851, 507), (376, 861), (1093, 815)]	
--------	---	--

▶ 검지 제외 대상

- 이륜차, 보행자는 VDS 검지 대상에서 제외
- 검지 영역에서 차선을 침범하는 차량의 경우 검지 대상에서 제외됨
- 단, 검지 영역 밖에서 차선을 변경하고 검지 영역 안에서 차선침범 없이 주행한 차량은 정상적으로 검지함

[표 44] 검지 제외 케이스

인식영역 안에서 차선침범	
------------------	---

	
이륜차	

2. VDS GT 데이터

2.1. VDS GT 데이터 활용

- ▶ Radar 인식결과를 기준 검지기 ground truth 데이터로 사용
 - xlsx 파일 형태로 차로별 교통량, 1분 단위 평균속도 정보를 제공함
 - 주행 중인 차량이 검지 영역을 통과하는 시간 값을 '기준 검지기' 컬럼에 기록됨
 - 시간 정보는 영상의 좌측 상단에 표기되는 시간 정보를 이용함



그림 48 검지 기준 시간

- ▶ 인식결과 입력
 - 사용자의 인식 모델을 통해 차량이 검지 영역에서 인식된 시간 값을 RowData Table의 '대상 검지기' 컬럼에 공백 없이 오름차순 순으로 기록해야 함
 - VDS 각 지점마다 1차로, 2차로 등 차로별 시트가 구분되어 있으므로 차로 정보를 구분한 뒤 동일한 차로에 값을 입력해야 함
 - 사용자는 '대상 검지기' 컬럼을 제외한 다른 컬럼 및 셀에 임의로 데이터를 입력해서는 안 됨
 - '대상 검지기' 컬럼에 기록된 시간 값은 기준 검지기(ground truth)값과 비교 분석 후, 그 결과가 분석 Table 및 결과 Table에 자동으로 기록됨
 - 분석 Table의 검지구분 컬럼에서 'Un' : 미검지, 'Over' : 과검지를 의미함
 - 미검지란 기준 검지기에 기록된 시간 값이 있지만 대상 검지기에 매칭되는 시간 값이 없을 때 차량을 검지하지 못한 것으로 판단한 결과
 - 과검지란 기준 검지기에 기록된 시간 값이 없는데 대상 검지기에서 특정 시간대에 차량이 지나갔다고 잘못 검지한 결과

상행 1차로

※ (RowData Table)은 반드시 검지시간을 오름차순으로 정렬한 후 입력

RowData Table			분석 Table								
구분	기준 검지기	대상 검지기	순번	기준 검지기	검지 구분	대상 검지기	순번	시간차	오검내용 및 유효성	이륜차 여부	
1	13:00:04		0	Over	0:00:00	1					
2	13:00:16		1	13:00:04	Un		1				
3	13:00:41		2	13:00:16	Un		1				
4	13:00:47		3	13:00:41	Un		1				
5	13:00:48		4	13:00:47	Un		1				
6	13:00:57		5	13:00:48	Un		1				
7	13:01:10		6	13:00:57	Un		1				
8	13:01:11		7	13:01:10	Un		1				
9	13:01:23		8	13:01:11	Un		1				
10	13:01:25		9	13:01:23	Un		1				
11	13:01:42		10	13:01:25	Un		1				
12	13:01:43		11	13:01:42	Un		1				
13	13:01:47		12	13:01:43	Un		1				
14	13:02:09		13	13:01:47	Un		1				
15	13:02:20		14	13:02:09	Un		1				
16	13:02:37		15	13:02:20	Un		1				
17	13:02:40		16	13:02:37	Un		1				
18	13:02:54		17	13:02:40	Un		1				
19	13:02:57		18	13:02:54	Un		1				
20	13:02:58		19	13:02:57	Un		1				
21	13:03:01		20	13:02:58	Un		1				
22	13:03:08		21	13:03:01	Un		1				
23	13:03:10		22	13:03:08	Un		1				
24	13:03:14		23	13:03:10	Un		1				
25	13:03:16		24	13:03:14	Un		1				
26	13:03:18		25	13:03:16	Un		1				
27	13:03:33		26	13:03:18	Un		1				
28	13:03:42		27	13:03:33	Un		1				
29	13:03:47		28	13:03:42	Un		1				
30	13:03:49		29	13:03:47	Un		1				

상행 1차로

※ (RowData Table)은 반드시 검지시간을 오름차순으로 정렬한 후 입력

RowData Table			분석 Table								
구분	기준 검지기	대상 검지기	순번	기준 검지기	검지 구분	대상 검지기	순번	시간차	오검내용 및 유효성	이륜차 여부	
1	13:00:04	13:00:05	1	13:00:04	-	13:00:05	1	0:00:01			
2	13:00:16	13:00:16	2	13:00:16	-	13:00:16	2	0:00:00			
3	13:00:41	13:00:41	3	13:00:41	-	13:00:41	3	0:00:00			
4	13:00:47	13:00:47	4	13:00:47	-	13:00:47	4	0:00:00			
5	13:00:48	13:00:48	5	13:00:48	-	13:00:48	5	0:00:00			
6	13:00:57	13:00:59	6	13:00:57	Un		5				
7	13:01:10		6	Over	13:00:59	6					
8	13:01:11		7	13:01:10	Un		6				
9	13:01:23		8	13:01:11	Un		6				
10	13:01:25		9	13:01:23	Un		6				
11	13:01:42		10	13:01:25	Un		6				
12	13:01:43		11	13:01:42	Un		6				
13	13:01:47		12	13:01:43	Un		6				
14	13:02:09		13	13:01:47	Un		6				
15	13:02:20		14	13:02:09	Un		6				
16	13:02:37		15	13:02:20	Un		6				
17	13:02:40		16	13:02:37	Un		6				
18	13:02:54		17	13:02:40	Un		6				
19	13:02:57		18	13:02:54	Un		6				
20	13:02:58		19	13:02:57	Un		6				
21	13:03:01		20	13:02:58	Un		6				
22	13:03:08		21	13:03:01	Un		6				
23	13:03:10		22	13:03:08	Un		6				
24	13:03:14		23	13:03:10	Un		6				
25	13:03:16		24	13:03:14	Un		6				
26	13:03:18		25	13:03:16	Un		6				
27	13:03:33		26	13:03:18	Un		6				
28	13:03:42		27	13:03:33	Un		6				
29	13:03:47		28	13:03:42	Un		6				
30	13:03:49		29	13:03:47	Un		6				

그림 49 인식결과 입력 전/후

▶ 이륜차 여부

- VDS 시스템에서 이륜차는 검지 대상이 아니므로 이륜차에 해당하는 검지 결과가 기록된 경우 분석 Table에서 이륜차 여부 컬럼에 '1'을 기록해야 함
- '이륜차 여부' 컬럼에 1을 기록하는 경우 자동으로 검지 대상에서 제외하여 계산됨
- '이륜차 여부' 컬럼에 1을 기록하지 않는 경우 과검지로 기록됨

상행 1차로

※ [RowData Table]은 반드시 검지시간을 오름차순으로 정렬한 후 입력

RowData Table			분석 Table								
구분	기준 검지기	대상 검지기	순번	기준 검지기	검지 구분	대상 검지기	순번	시간차	오검내용 및 유효성	이륜차 여부	
1	13:00:04	13:00:04	1	13:00:04	-	13:00:04	1	0:00:00			
2	13:00:16	13:00:16	2	13:00:16	-	13:00:16	2	0:00:00			
3	13:00:41	13:00:41	3	13:00:41	-	13:00:41	3	0:00:00			
4	13:00:47	13:00:47	4	13:00:47	-	13:00:47	4	0:00:00			
5	13:00:48	13:00:48	5	13:00:48	-	13:00:48	5	0:00:00			
6		13:00:59	5	Over	13:00:59	6					

결과 Table											
구분	시간대	표본수		오검		정검차량		평균속도			
		기준	대상	마감	과검	상행X	상행O				
13:00											
1	13:00:00 ~ 13:01:00	5	6	0	1	80.0	80.0			61	

그림 50 이륜차 여부를 체크 하지 않는 경우

상행 1차로

※ (RowData Table)은 반드시 검지시간을 오름차순으로 정렬한 후 입력

RowData Table			분석 Table								
구분	기준 검지기	대상 검지기	순번	기준 검지기	검지 구분	대상 검지기	순번	시간차	오검내용 및 유효성	이륜차 여부	
1	13:00:04	13:00:04	1	13:00:04	-	13:00:04	1	0:00:00			
2	13:00:16	13:00:16	2	13:00:16	-	13:00:16	2	0:00:00			
3	13:00:41	13:00:41	3	13:00:41	-	13:00:41	3	0:00:00			
4	13:00:47	13:00:47	4	13:00:47	-	13:00:47	4	0:00:00			
5	13:00:48	13:00:48	5	13:00:48	-	13:00:48	5	0:00:00			
6		13:00:59	5	Over	13:00:59	6				1	

결과 Table											
구분	시간대	표본수		오검		정검차량		평균속도			
		기준	대상	마감	과검	상행X	상행O				
13:00											
1	13:00:00 ~ 13:01:00	5	6	0	0	100.0	100.0			61	

그림 51 이륜차 여부를 체크 하는 경우

▶ 결과 Table

- 결과 Table에는 기준 검지기와 대상 검지기 컬럼에 입력된 시간 값에 따라 1분 단위 오검지를 분석함
- 결과 Table에서 차로의 1분 단위 평균속도 정보를 함께 제공함 (개별 차량의 속도는 제공하지 않음)

