


인공지능 데이터 구축·활용 가이드라인

- 차로워반 영상 데이터 -

인공지능 데이터 구축	사업 총괄	D-TEG
	데이터 설계	D-TEG
	원천데이터 수집	D-TEG
	원천데이터 정제	D-TEG
	데이터 가공	 crowdworks
	데이터 검수	 crowdworks
	저작도구 개발	D-TEG  crowdworks
	AI모델 개발	<i>Agile SoDA</i>
	응용 서비스 개발	<i>Agile SoDA</i>
가이드라인 작성	디텍	성제남
가이드라인 버전	차로워반 영상 데이터 가이드라인 Ver.1 ('22.06.20)	

목 차

1. 데이터 명세 정보	1
1.1 데이터 정보 요약	1
1.2 데이터 포맷	1
1.2.2 원천 데이터 포맷	2
1.2.3 원시/원천/학습용 데이터 라벨링 JSON 포맷	2
1.3 어노테이션 포맷	3
1.3.1 영상데이터 획득 시 라벨링 공통항목	3
1.3.2 영상데이터 획득 시 라벨링 선택항목	3
1.4 데이터 구성	4
1.4.1 구축 데이터 클래스 분류	4
1.4.2 어노테이션 포맷 및 저장 형식	4
1.4.3 라벨링 저장구조 정의	5
1.4.4 원천 데이터 저장 구조	6
1.5 데이터 통계	7
1.5.1 데이터 구축 규모	7
1.5.2 데이터 분포	7
1.6 원시데이터 특성	9
1.6.1 원시 데이터 명세	9
1.6.2 원시 데이터 정의	11
1.6.3 원시 데이터 획득 방법	11
1.6.4 원시 데이터 포맷	12
1.6.5 원시 데이터 획득 규모	12
1.7 기타정보	13

2. 데이터 구축 가이드	14
2.1 데이터 구축 개요	14
2.2 문제정의	15
2.2.1 임무 정의	15
2.3 획득·정제	16
2.3.1 원시데이터 선정	16
2.3.2 데이터 획득 · 정제 절차	16
2.3.3 데이터 획득 항목	16
2.3.4 데이터 저장 및 관리	17
2.3.5 원시 데이터 정제 방식	17
2.3.6 원시 데이터 정제 기준	18
2.3.7 획득 도구 및 정제 도구	19
2.3.8 획득/정제 시 고려사항	19
2.4 어노테이션/라벨링	21
2.4.1 어노테이션/라벨링 절차	21
2.4.2 어노테이션/라벨링 기준	22
2.4.3 어노테이션/라벨링 조직	23
2.4.4 어노테이션/라벨링 도구	24
2.5 검수	27
2.5.1 검수 절차	27
2.5.2 검수 기준	28
2.5.3 단계별 검수 방법	28
2.5.4 검수 도구 및 작업방법	29
2.6 활용	33
2.6.1 활용 모델	33
2.6.2 유형 정의	33
2.6.3 인공지능 모델 처리 프로세스	34
2.6.4 인공지능 모델 적용 알고리즘	34

1. 데이터 명세 정보

1.1 데이터 정보 요약

데이터 명		차로 위반 영상 데이터
데이터 포맷		차량용 DVR 동영상
활용 분야		차로 위반 교통단속 도로 안전시설 점검 자율주행 차량의 안전 자율주행 동적 정보 업데이트
데이터 요약		주행 중인 버스에서 촬영된 영상으로 차로 위반 차량 영상의 객체 및 행위를 검출할 수 있는 인공지능 데이터셋
데이터 출처		서울 시내버스 주행 영상 서울 마을버스 주행 영상
데이터 이력	배포버전	LaneViolationImageDataSet_ver1.
	개정이력	신규
	작성자/배포자	수행기관 / (주)디텍
데이터 통계	데이터 구축 규모	버스 주행영상이 녹화된 HDD 약 200개 x 2TB (예상)
	데이터 분포	차로구분 분포: 백색(30%), 청색(20%), 황색(30%), 갓길차로(20%) 차량 분포: 승용차(60%), 승합차(15%), 화물차(15%), 이륜차(10%)
기타 정보	대표성	해당 사항 없음
	독립성	
	유의사항	해당 사항 없음
	관련 연구	

[표 1] 차로위반 영상 데이터 정보

1.2 데이터 포맷

```

1  {} nia_sample_1108.json > ...
2  {
3    "dataID": 데이터 고유 ID,
4    "data_set_info": {
5      "sourceValue": 원천데이터 파일명(imagefile name),
6      "data": [
7        {
8          "objectID": 객체 고유 ID,
9          "value": {
10             "metainfo": 원천데이터 정보,
11             "annotation": 오노테이션(바운딩) 방법,
12             "points": 폴리곤 바운딩 좌표값,
13             "object_Label": {
14               객체정보
15               "vehicle_type"/"lane_type": 차량/차선 타입,
16               "vehicle_attribute"/"lane_attribute": 위반여부/차량종류,
17               "vehicle_shown": 차량 오브젝트 완전성 여부
18             },
19             "extra": 객체 추가 정보(차량과 차선 바운딩 영역의 구분 색상)
20           }
21        }
22      ]
23    }
24  },

```

[그림 1] 데이터 포맷

1.2.1 원천 데이터 포맷

- 정제된 원천 데이터는 영상데이터로서 MP4 및 JPG 형식으로 구축



[그림 2] 원천 데이터 구축 형식

1.2.2 원시/원천/학습용 데이터 라벨링 JSON 포맷

포맷	MS COCO 스타일	
labelling	한 이미지에 세 가지 task 라벨링을 동시에 진행	
data ID	데이터 고유 ID	
sourceValue	원천데이터 파일명	
objectID	객체를 구분하기 위한 고유 ID	
metainfo	원천 데이터 관련 정보	
	violation_type	위반차선종류 4종
	video_id	파일번호
	camera_channel	카메라채널 3종
	time_info	시간정보
	camera_number	관리번호
annotation	어노테이션 방법	
point	polygon 좌표값 (x, y)	
extra	객체유형 추가정보(차량과 차선 색상구분)	
	value	vehicle/lane
	label	차량/차선
	color	#096ecd/#e0182d
object_label	객체정보	
	vehicle_type	차량 타입 : 승용차, 승합차/버스, 트럭/특수차량, 이륜차
	lane_type	차선 타입 : 백색, 청색, 황색, 갓길
	vehicle_attribute	위반 여부 : 정상, 위험 위반 (차량 객체에 해당)
	lane_attribute	차선 종류 : 1줄 실선, 2줄 실선, 1줄 점선, 좌점선, 우실선, 좌실선, 우점선 (차선 객체에 해당)
	vehicle_shown	차량 완전성 : 완전/불완전

[표 2] 차로위반 영상 데이터 라벨링 JSON 포맷

1.3 어노테이션 포맷

1.3.1 영상데이터 획득 시 라벨링 공통항목

NO	속성명	항목설명	Type	필수여부	작성예시
1	dataID	데이터셋 식별자	number	필수	61085960
2	object value	객체 분류	string	필수	vehicle, lane
3	point	폴리곤 좌표값	number	필수	x: [0-1920] y: [0-1080]
4	camera_channel	카메라 설치 위치	string	필수	A:전방,B:좌측방,C:우측방
5	Object type	객체 유형	string	필수	car, bus, truck 등 white, blue 등
6	Object attribute	객체 속성	string	필수	normal, violation 등 solid, dashed 등

[표 3] 데이터 라벨링 공통항목

1.3.2 영상데이터 획득 시 라벨링 선택항목

NO	속성명	항목설명	Type	필수여부	작성예시
1	info.framerate	영상 프레임레이트	number	선택	15
2	info.GPS	획득 위치. GPS좌표	string	선택	10,20,10,0
3	info.Gyro	Gyro값	string	선택	0,0,0
4	info.GSensor	GSensor값	string	선택	-31,35,975
5	info.speed	주행속도	number	선택	50
6	info.date	획득 일자	string	선택	2021.05.30 12:00:30
7	info.length	영상 재생시간	number	선택	20(초)
8	info.width	영상 너비	number	선택	1280
9	info.height	영상 높이	number	선택	720
10	info.busnumber	버스 번호	string	선택	1005-1

[표 4] 데이터 라벨링 선택항목

1.4 데이터 구성

1.4.1 구축 데이터 클래스 분류

채널 정의	카메라 설치 위치
A 채널	전방 카메라 영상
B 채널	좌측방 카메라 영상
C 채널	우측방 카메라 영상

위반 종류	설 명
백색 차로 위반	양옆의 같은 방향으로 주행하는 차량의 차로 위반(일반 차로 위반)
청색 차로 위반	버스 전용차로와 같은 특수용도로 사용되는 차로의 위반
황색 차로 위반	반대 방향으로 주행하는 차량의 차로 위반(중앙선)
갓길 차로 위반	도로주정차로 주정차 위반

위반 객체	설 명
승용자동차	10인 이하를 운송하기에 적합하게 제작된 자동차
승합자동차	11인 이상을 운송하기에 적합하게 제작된 자동차
화물자동차/특수자동차	화물을 운송하기에 적합하게 제작된 자동차 / 특수 작업을 수행하기에 적합하게 제작된 자동차
이륜자동차	1인 또는 2인의 사람을 운송하기에 적합하게 제작된 이륜의 자동차

[표 5] 구축 데이터 클래스

1.4.2 어노테이션 포맷 및 저장 형식

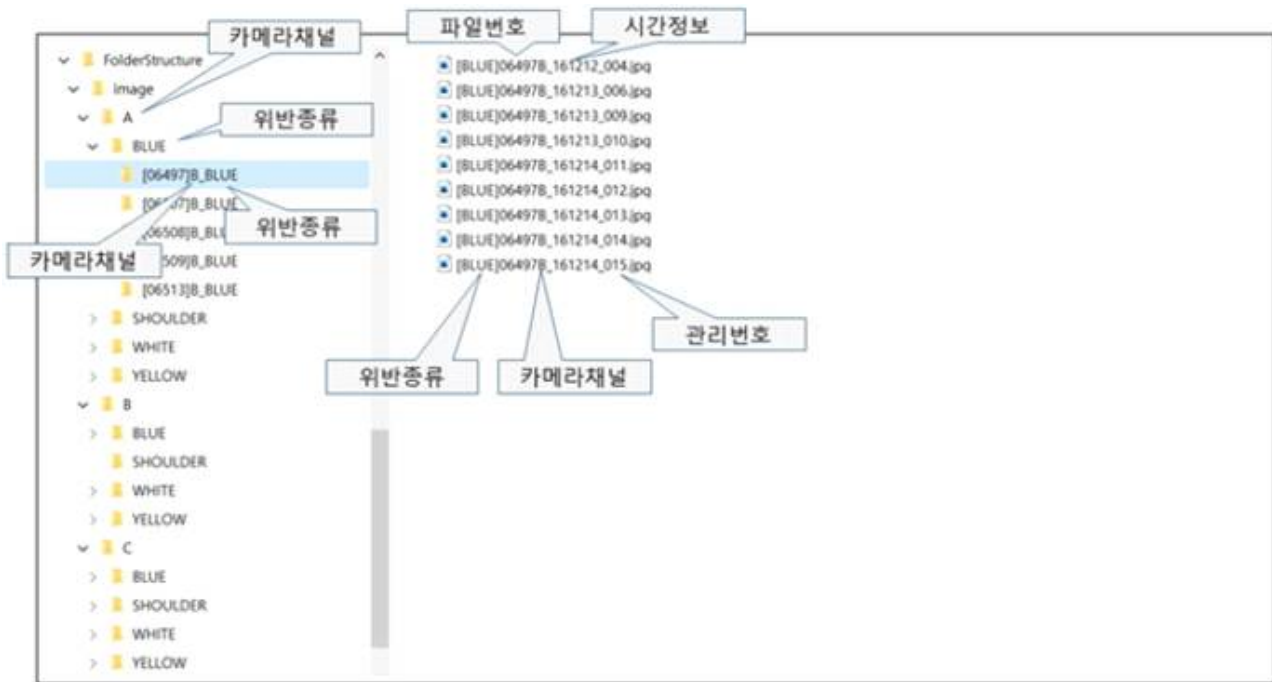
- 구축하고자 하는 데이터는 CCTV 영상에서 차로위반 전/후를 포함하여 건당 평균 10초 Video Sequence를 초당 5장 추출한 원천데이터 이미지에서 차로위반 차량 객체와 차로를 어노테이션 함. 차량 객체와 차로는 폴리곤 형태로 어노테이션 하기 때문에 학습용 데이터로서 가치를 부여하는 어노테이션 정보를 저장할 수 있는 별도의 데이터 구조와 파일 포맷을 정의함
- 어노테이션 파일 포맷은 특정 소프트웨어에 종속되지 않고 쉽게 열고 편집할 수 있고 구조화된 어노테이션 정보를 저장하기 적합한 json 포맷으로 함
- 일반적으로 대상이 잘리지 않으면서 불필요한 공간이 최소화 되도록 대상의 크기에 맞춰 라벨링하고 이미지 내 대상의 경계가 분명하지 않아 영역을 설정하기 모호한 상황이 발생하지 않도록 작업 상세가이드를 작성해 작업자가 헛갈리지 않고 일관된 기준을 갖고 작업

1.4.3 라벨링 저장구조 정의

메타 데이터	라벨링 데이터
<pre> "metainfo": { "violation_type": "white", "video_id": "06492", "camera_channel": "B", "time_info": "102743", "camera_number": "001" } </pre>	<pre> "data": [{ "objectID": "data_set_info_61085960_1", "value": { "metainfo": { "violation_type": "white", "video_id": "6492", "camera_channel": "B", "time_info": "102743", "camera_number": "1" }, "annotation": "POLYGONS", "points": [{ "x": 716.7516454392573, "y": 829.0556045895852 }, /*(폴리곤 좌표 생략).... */], "extra": { "value": "vehicle", "label": "차량", "color": "#096ecd" }, "object_Label": { "vehicle_type": "vehicle_car", "vehicle_attribute": "normal", "vehicle_shown": "full" } } }] </pre>

[표 6] 라벨링 저장 구조

1.4.4 원천 데이터 저장 구조 (Annotation, Image)



[그림 3] 원천 및 최종 데이터 저장 구조

규칙	예시	위반종류 파일번호카메라채널 시간정보_관리번호.jpg
완성 예문		[WHITE]00001A_101123_001.jpg
위반 종류	WHITE	WHITE/BBLUE/YELLOW/SOULDER
파일번호	00001	파일번호(5자리) - 최초 비디오 소스파일번호
카메라 채널	A	전방카메라/좌측방카메라/우측방카메라(A/B/C)
시간정보	101123	HHMMSS(시분초)
관리번호	001	관리번호

[표 7] 원천 데이터 저장 규칙

1.5 데이터 통계

1.5.1 데이터 구축 규모

- 수집하고자 하는 데이터의 규모는 이미지 환산 기준 80만 장 이상

지역별 버스	버스 수
서울 시내버스	약 3,200 대
서울 마을버스	약 1,700 대
광주광역시 시내버스	약 1,000 대

[표 8] 지역별 데이터 규모

데이터 정의	수량
Video Sequence	20,000 개 이상
Video 건당 길이(평균)	10 초
초당 Frame	5 장
이미지 환산	800,000 장

[표 9] 데이터 정의 및 수량

- 원천 데이터의 다양성을 확보하기 위해 버스에 장착된 3개 카메라 채널, 촬영 시간대, 차선 종류, 차량 종류를 반영하여 총 48종의 서로 다른 카테고리에 따라 데이터를 취득
- 원천 데이터는 3개의 채널에서 개별 수집된 영상임

1.5.2 데이터 분포

- 카테고리별 데이터 분포

다양성 카테고리	다양성 조건	비고
카메라 채널(3종)	전방	<ul style="list-style-type: none"> 총 다양성 조건: 48종
	좌측방	
	우측방	
차선 종류(4종)	청색	
	백색	
	황색	
	갓길차선(주정차)	
차량종류(4종)	승용자동차	
	승합자동차	
	화물(특수)자동차	
	이륜자동차	

[표 10] 데이터의 다양성

○ 데이터 종류 및 수량

No.	카메라 채널(3종)	비율	구축 수량 (장)
1	전방	60%	480,000
2	좌측방	20%	160,000
3	우측방	20%	160,000

No.	차선 종류(4종)	비율	구축 수량 (장)
1	백색	30%	240,000
2	청색	20%	160,000
3	황색	30%	240,000
4	갓길차선(주정차)	20%	160,000

No.	차량종류(4종)	비율	구축 수량 (개)
1	승용자동차	60%	480,000 이상
2	승합자동차	15%	120,000 이상
3	화물(특수)자동차	15%	120,000 이상
4	이륜자동차	10%	80,000 이상

[표 11] 데이터 종류별 및 수량

○ 라벨링은 자동차 객체, 차선, 차선/차로 위반 여부 등에 대해 진행

라벨링 대상	라벨링 방법
자동차 객체	Polygon Segmentation
차로	Polygon Segmentation
차로 위반 여부	위반
	위험(주정차)
	정상

[표 12] 데이터 라벨링 대상

○ 데이터 수집 조건과 라벨링 선정을 위한 사전에 검토한 자료는 다음과 같음

객체 : 일반차량(승용차 등), 이륜차(오토바이, 전동 킥보드, 자전거 등 종별), 사업용 차량(버스, 택시, 화물차 등 종별), 사람, 동물, 낙하물(보행자, 동물 등 종별)

○ 상황 구분

구분	상세 분류
도로	스쿨존
	주택지 일반도로
	상가 인접 일반도로
	고속도로
	교차로
	차로수
시간대	주간 / 야간
기상	맑음 / 비

[표 13] 상황별 데이터 상세 분류

구분	상세 분류
차선 위반	중앙선 침범
	정지선 위반
통행 위반	부당한 회전
	주정차 위반
	진로 변경 위반
	직진 및 우회전 통행 방해

[표 14] 위반별 데이터 상세 분류

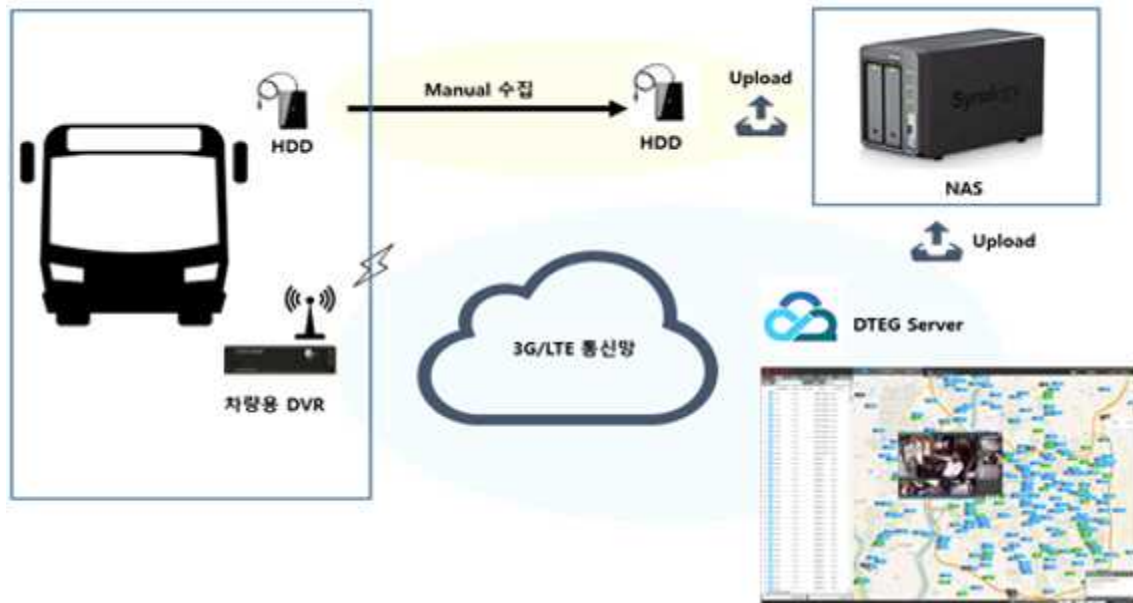
1.6 원시 데이터 특성

1.6.1 원시 데이터 명세

데이터 명		차로 위반 영상 데이터
데이터 포맷		차량용 DVR 동영상
활용 분야		차로 위반 교통단속 도로 안전시설 점검 자율주행 차량의 안전 자율주행 동적 정보 업데이트
데이터 요약		주행 중인 버스에서 촬영된 영상으로 차로 위반 차량 영상의 객체 및 행위를 검출할 수 있는 인공지능 데이터셋
데이터 출처		서울 시내버스 주행 영상 서울 마을버스 주행 영상
데이터 이력	배포버전	LaneViolationImageDataSet_ver1.
	개정이력	신규
	작성자/배포자	수행기관(주)디텍
데이터 통계	데이터 구축 규모	버스 주행영상이 녹화된 HDD 약 200개 x 2TB (예상)
	데이터 분포	차로구분 분포: 백색(30%), 청색(20%), 황색(30%), 갓길차로(20%) 차량 분포: 승용차(60%), 승합차(15%), 화물차(15%), 이륜차(10%)
기타 정보	대표성	해당 사항 없음
	독립성	
	유의사항	
	관련 연구	해당 사항 없음

[표 15] 원시 데이터 명세

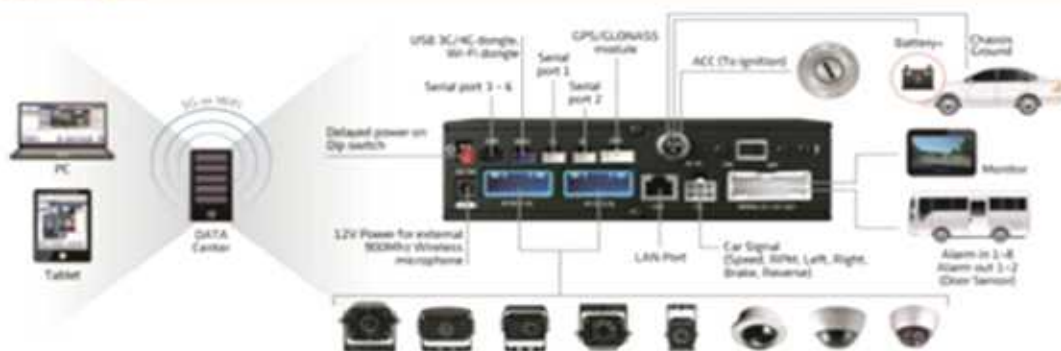
- 서울, 부산, 광주 시내버스 등 ㈜디텍에서 기 설치/운영 중인 차량용 DVR에서 설치된 Camera들을 이용하여 주행 동영상 수집
- 버스 1대당 6개 ~ 8개의 Camera가 설치되어 있으며, 이 중 전방, 좌측방(운전석 쪽), 우측방(입구 쪽)에 설치된 3개의 Camera에서 수집된 외부 차량 주행 동영상 수집
- 시내버스 및 마을버스의 경우 이른 아침부터 늦은 저녁까지 매일 장시간 운행하며 평균 한번 운행에 50개 이상의 버스 정류장을 거치기 때문에 방대한 양질의 Data 수집 가능
- 개인정보의 문제제기 가능성이 있는 보행자 얼굴과 차량 번호판은 비식별화 처리를 하여 데이터 공개



[그림 4] 원시 데이터 수집 프로세스

- 데이터 획득을 위한 작업도구

CRX3108 System Diagram



[그림 5] 데이터 수집 장치 구성 및 사양

1.6.2 원시 데이터 정의

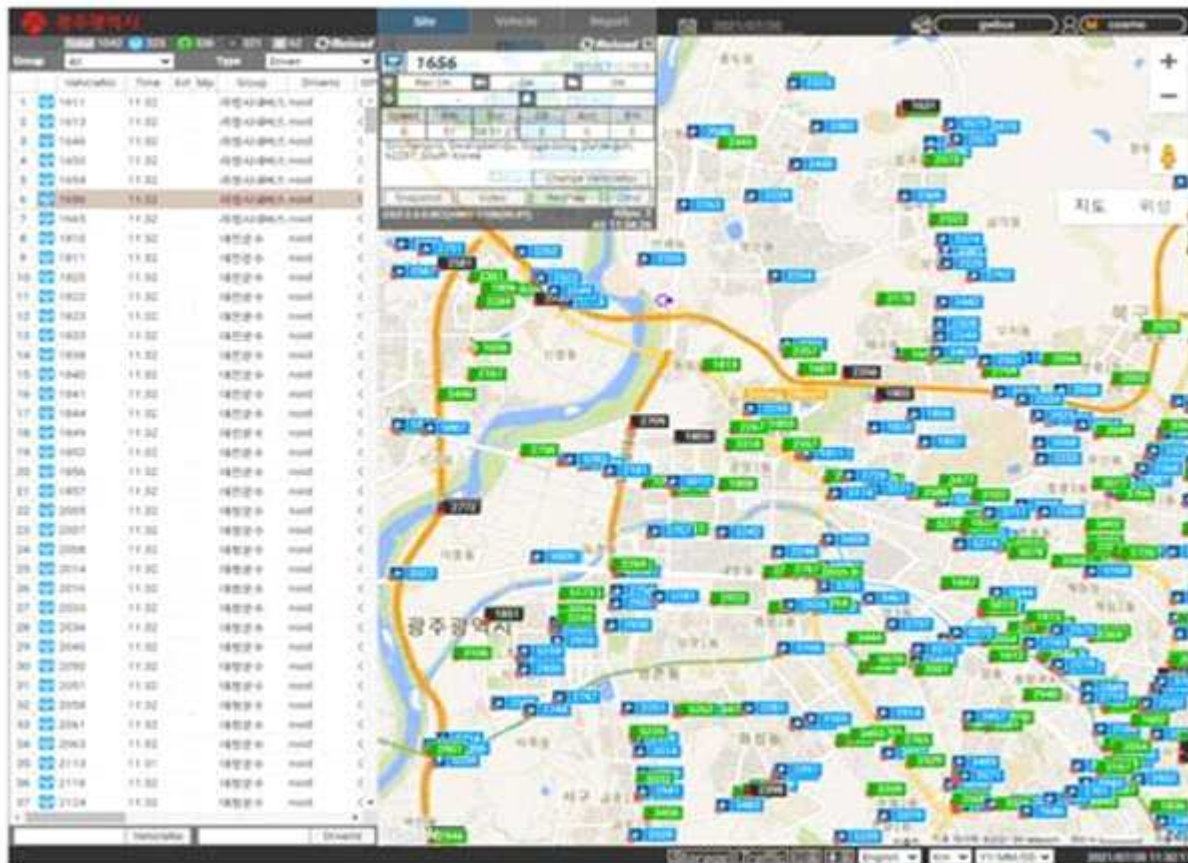
- 차로 위반 영상 데이터셋 구축을 위한 원시 데이터는 주행 중인 버스로부터 촬영된 도로 교통 상의 동영상 데이터로서 차로위반 객체 및 차로위반 행위를 검출할 수 있는 정보가 포함되어 있는 동영상 데이터
- 버스에 설치된 차량용 DVR의 다채널 영상 중 전방 카메라(A 채널), 좌측방 카메라(B 채널), 우측방 카메라(C 채널)에서 입력된 동영상 데이터를 사용하여 최종 학습용 데이터로 변환
- 카메라 설치 방향



[그림 6] 카메라 설치 방향

1.6.3 원시 데이터 획득 방법

- 원시 데이터는 이미 설치가 완료되어 운영 중인 서울시내버스와 서울마을버스의 차량용 DVR로부터 HDD를 직접 교체하여 취득함
- 버스 영상데이터의 사용을 위해 미리 서울시내버스/서울마을버스 업체와 데이터 사용에 대한 협약을 체결하고, 협약이 체결된 버스 업체 내에서 취득 대상 버스를 선정함
- 원시 데이터 취득은 (주)디텍에서 수행하며 HDD 교체 작업은 해당 버스의 차고지에서 실시
- HDD(2TB)에는 설치된 카메라 개 수, 해상도 등에 따라 가변적이거나 통상 2~3 주 정도의 영상데이터를 보관할 수 있음. 따라서 한 번의 HDD 교체를 통해 약 2~3주 분량의 영상 데이터를 획득할 수 있음. 만약 동일한 노선버스의 영상이 추가로 필요하다면 약 2~3주 정도의 주행 녹화 기간 이후에 취득함
- 기 구축된 데이터 수집 플랫폼(DCS, 디텍 자체 서버)을 이용하여 원시 데이터 수집



[그림 7] 데이터 수집 플랫폼

1.6.4 원시 데이터 포맷

- 원시 원시 데이터는 (주)디텍 자체의 파일 및 폴더 구조를 따르고 있으며, 정제단계에서 파일 변환 작업을 통해 MP4 형식으로 변환됨

	데이터 유형	압축방식	해상도	Frame Rate	수집장비
원시 데이터	동영상 (주)디텍 자체 파일포맷)	H.264	FHD 이상 (1920x1080)	10 FPS 이상	차량용 DVR

[표 16] 원시 데이터 포맷

1.6.5 원시 데이터 획득 규모

- 학습용 데이터의 목표 구축 수량은 800,000건(image+json)
- 목표 수량 달성을 위한 원시 데이터(HDD)의 규모는 200개 × 2TB 정도로 예상됨(하루 위반 건수 12건, 한 개의 영상은 10초 기준)
- AI 학습을 위해, 정상 주행(비위반) 데이터는 전체 수집 수량 중 약 30%



[그림 8] 원시 데이터 폴더 구조(HDD)

1.7 기타정보

○ 원시 데이터 수집과 개인 정보 보호

- ㈜디텍에서 이미 설치하여 운용 중인 버스 차량용 DVR 시스템을 이용하면 방대한 영상 데이터를 안정적으로 수집할 수 있으며 자체 서버시스템 또한 구축되어있어 통신 네트워크를 통한 데이터 수집도 가능함
- 본 사업과 관련하여 버스 사업자(또는 버스 조합)와의 협약을 통해 버스 영상 데이터 사용에 관한 동의를 확보함
- 개인정보보호법 등 법적 이슈에 관해서 개인정보보호위원회의 가이드라인을 준수하고 비식별화 등의 적절한 조치를 취하면 데이터 취득 및 대외 공개에 문제가 없을 것으로 판단됨. 필요 시 법률전문기관에 본 사업에 관해 법률검토를 의뢰함

2. 데이터 구축 가이드

2.1 데이터 구축 개요

○ 데이터 구축 인프라 구성방안

- 유동적이고 실시간으로 연계 가능한 온라인 플랫폼 인프라를 활용하여 가공과 검수를 실시간으로 진행하며, 이를 통해 클라우드소싱 작업 인력 퍼포먼스 모니터링 대시보드에서 확인 가능한 환경 구성을 갖추고 있음
- 작업과 검수 건수 및 진척율 확인 통해 가공 데이터에 대한 재작업 요청 및 오류 피드백 과정 가능, 클라우드 워커를 교육하여 대량의 데이터를 안정적으로 구축 가능한 환경을 갖추고 있음

ALL RUNNING CHECK_ONLY HOLDING ENDED						
총합당 건수	작업 건수	작업 진척율	검수 건수	검수 진척율	작업자 현황	가이드
1,449	942	58.11%	0	0%	보기	PC M C
50	0	0%	0	0%	보기	PC M C
200	7	3.50%	0	0%	보기	PC M C
179,293	179,292	100.00%	177,354	98.92%	보기	PC M C
7,600	1,695	22.30%	987	58.23%	보기	PC M C
3,790	10	0.26%	9	90%	보기	PC M C
3,790	10	0.26%	10	100%	보기	PC M C
10,000	2,660	26.60%	2,537	88.71%	보기	PC M C
161,148	161,148	100%	159,143	98.76%	보기	PC M C
159,209	159,209	100%	156,923	98.56%	보기	PC M C

[그림 9] 실시간 프로젝트 모니터링 대시보드



[그림 10] 일별 작업 건수 등 통계 화면

2.2 문제정의

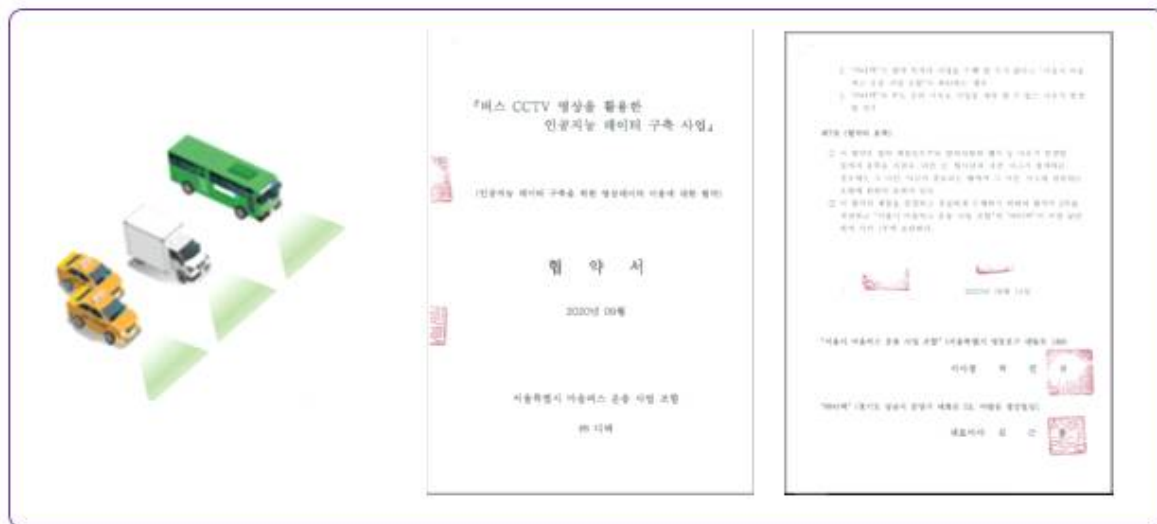
2.2.1 임무 정의

- 본 과제 컨소시엄의 주관기업은 전국 버스에 차량 감시 및 모니터링용 DVR시스템을 설치 및 운용 중인 기업으로 현재 기보급 된 단말기에 통신 환경을 구축하거나 혹은 HDD 교체만으로도 전국 분포의 방대한 데이터를 확보 가능

구분	설치수량	도입연도
서울 시내버스	3,200대	2019년
서울 마을버스	1,650대	2020년
부산 시내버스	2,500대	2016년
부산 마을버스	400대	2019년
광주 시내버스	1,044대	2021년

[표 17] 주관사의 차량용 DVR 운영현황

- DCS에 Connection된 통신형 수집 장치(CRX) 활용
 - 일반주행 데이터 및 이벤트 동영상 자동 전송 : 별도 수집을 위한 인력 불필요
 - 요금제/이벤트설정에 따라 수집 데이터의 양을 다양하게 조정 가능
 - 영상 수집에 관하여 (주)디텍은 본사업을 포함한 포괄적 빅데이터 취득에 관하여 버스 사업자들과의 협약을 체결



[그림 11] 사업용 차량 운영사와의 협약서

2.2.2 원시 데이터 선정

- 데이터 수집 장치인 차량용 DVR은 이미 설치 완료되어 운용 중이므로 추가 설치작업 없이 바로 데이터 수집이 가능하며, HDD 교체 또는 기 구축된 통신 네트워크/서버를 이용하여 비교적 저렴한 비용으로 원시 데이터의 취득이 가능

2.3 획득·정제

2.3.1 원시 데이터 선정

- 데이터 수집 장치인 차량용 DVR은 이미 설치 완료되어 운용 중이므로 추가 설치작업 없이 바로 데이터 수집이 가능하며, HDD 교체 또는 기 구축된 통신 네트워크/서버를 이용하여 비교적 저렴한 비용으로 원시 데이터의 취득이 가능

2.3.2 데이터 획득 · 정제 절차

	데이터 획득 형태	수집 장비	데이터 형식	데이터 처리
1	차량 영상데이터 수집(버스 차고지)	차량용 DVR	HDD(RAW)	데이터 수집 (HDD 직접 교체)
2	차량 영상데이터 업로드	(주)디텍 NAS	HDD → NAS(RAW)	데이터 업로드 (디텍 NAS)
3	1차 데이터 정제	(주)디텍 NAS	RAW -> MP4, JPG	차로위반 영상 검색 영상파일 추출 파일변환(RAW->MP4, JPG) 파일분류
4	2차 데이터 정제	(주)디텍 NAS	JPG -> JPG	비식별화 (얼굴, 차량 번호판)
5	원천데이터 업로드	공용 NAS	MP4, JPG	데이터 업로드 (공용 NAS)

※ RAW: (주)디텍 자체 동영상 파일형식

[표 18] 데이터 획득 및 정제 절차

2.3.3 데이터 획득 항목

No.	데이터 획득항목	항목 설명	작성예시
1	BusType	버스 구분	시내버스/마을버스
2	BusNumber	버스 노선번호	문자열 + 숫자
3	Date	녹화날짜	2020.10.28. ~ 2020.11.10
4	Resolution	영상 해상도	FHD (1920x1080)
5	FPS	Frame Rate	10 fps

[표 19] 데이터 획득 항목

2.3.4 데이터 저장 및 관리

- 데이터 저장소 및 저장 데이터

 저장소	 저장 데이터	 Comment
HDD	 원시 데이터(원본)	차량용 DVR 수집(교체) NAS Upload
디텍 Storage Server	 원시데이터(복사본, 정제 작업용) 원천 데이터 1(1차 정제 완료) 원천 데이터 2(2차 정제 완료)	주행영상 검색/추출/변환 비식별화
컨소시엄 Storage Server	 원천 데이터 2(복사본, 가공 작업용)	데이터 가공

[그림 12] 데이터 저장 및 관리

- 수집된 원시 데이터(HDD)에 고유번호를 부여하고 각각의 HDD에 대한 획득정보를 별도의 파일 및 폴더구조로 관리함

Index	버스종류	버스회사	노선번호	차량번호	녹화설정						영상 녹화 기간		입고일	저장경로 [Index_차량번호]
					A 채널		B 채널		C 채널		시작	종료		
					해상도	FPS	해상도	FPS	해상도	FPS				
001	마을버스	일원교통	강남01	3380	FHD	10	FHD	10	FHD	10	2021.04.05	2021.04.20	2021.04.21	001_3380
002	마을버스	일원교통	강남01	3333	FHD	10	FHD	10	FHD	10	2021.04.05	2021.04.20	2021.04.21	002_3333
003	시내버스	아진교통	141	6504	FHD	10	FHD	10	FHD	10	2021.04.12	2021.04.27	2021.04.28	003_6504
004	시내버스	아진교통	141	6523	FHD	10	FHD	10	FHD	10	2021.04.12	2021.04.27	2021.04.28	004_6523

[표 20] 원시 데이터 관리 대장

2.3.5 원시 데이터 정제 방식

- 원시 데이터 정제 프로세스



[그림 13] 단계별 정제 내역

- 원시 데이터 업로드
 - 취득된 원시 데이터(HDD)를 버스 구분(서울시내버스/서울마을버스), 버스 노선번호별로 구분하여 주관수행기관의 자체 서버에 업로드. 원시 데이터 획득 정보(녹화날짜, 해상도, Frame Rate)를 별도의 파일로 작성하여 관리함
- 데이터 정제 (1)
 - 전용 Viewer Software를 사용하여 정제 기준에 맞게 작업자가 직접 위반영상을 검색. 작업자는 변환툴(Viewer Software)을 사용하여 검색된 영상을 원천데이터 포맷에 맞게 추출 및 파일변환(상용 파일포맷, MP4, JPG). 변환된 파일은 원천 데이터의 분류 기준에 따라 폴더 단위로 분류됨
- 데이터 정제 (2)
 - 1차 정제된 영상 파일은 비식별화 전문업체를 통해 얼굴, 차량 번호판의 비식별화과정 (Blur 처리)을 거침
- 원천 데이터 업로드
 - 비식별화 처리가 완료된 원천 데이터를 컨소시엄 공유 서버로 업로드

2.3.6 원시 데이터 정제 기준

카메라 채널	항목 설명	비고
A 채널	전방 카메라 영상	
B 채널	좌측방 카메라 영상	
C 채널	우측방 카메라 영상	

[표 21] 카메라 채널별 분류 기준

위반 종류	항목 설명	비고
백색 차로 위반	백색 실선/복선/이중실선 차로 변경	
청색 차로 위반	청색 실선/복선 차로 변경	
황색 차로 위반	황색 실선/복선/이중실선 차로 변경	
갓길 차로 위반	갓길 차로 주정차	

[표 22] 위반 종류별 분류 기준

객체 종류	항목 설명	비고
승용자동차	아반떼, 쏘나타, 그랜저, 제네시스 코나, 투싼, 싼타페, 팰리세이드 등	
승합자동차	스타렉스, 카니발, 스타리아, 버스 등	
화물자동차/특수자동차	포터 등	
이륜자동차	오토바이	

[표 23] 위반 객체별 분류 기준

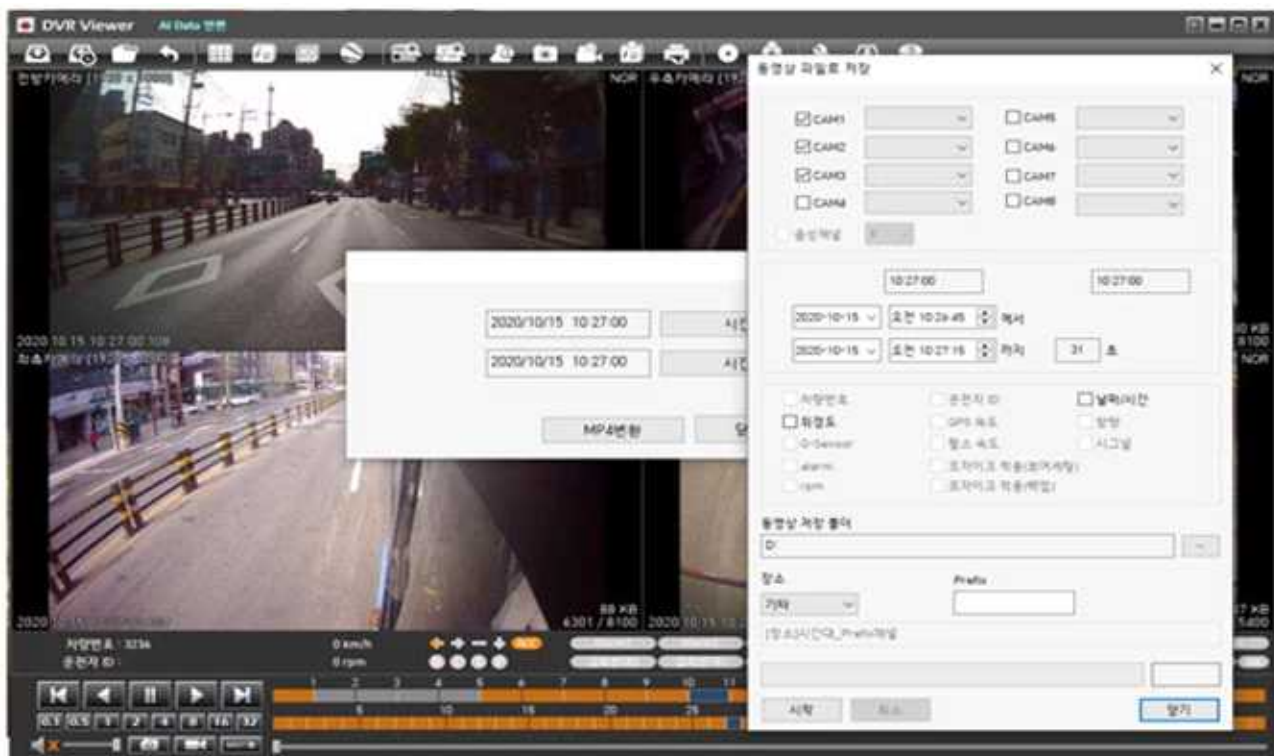
2.3.7 획득 도구 및 정제 도구

○ 획득 도구

- 버스 차량에 설치된 차량용 DVR(Digital Video Recorder)을 사용하여 원시 데이터를 획득함. 차량용 DVR은 다채널 영상 촬영이 가능하며 주행 영상과 함께 날짜, 시간 및 GPS/G-Sensor/Gyro 등의 기타 메타 데이터를 함께 수집함
- 주요 녹화 사양: FHD(1920x1080) 해상도, 10 FPS 이상

○ 정제 도구

- 전용 정제 프로그램을 사용하여 영상 검색, 영상 및 추출, 영상 변환/저장 작업을 수행
- 작업자가 정제 프로그램을 사용하여 영상을 재생하면서 차로 위반 상황을 시간 단위로 선택하면 설정된 원천 데이터 형식(카메라 채널, 상황 전후 시간, 상용 파일포맷 등)으로 자동 추출/변환/저장됨



[그림 14] 정제 도구 화면

2.3.8 획득/정제 시 고려사항

○ 보안사항, 개인정보 및 저작권

- 개인정보보호법 등 법적 이슈에 관해서 개인정보보호위원회의 가이드 라인을 준수하여 획득/정제 작업 및 데이터 저장관리를 수행
- 모든 작업자에게 보안서약서를 작성하도록 하고 개인정보보호 및 보안 등의 이슈가 발생하지 않도록 관리

- 영상 취득 예정인 버스 업체와 영상 사용에 대한 협약서를 체결
- 주행 영상 내의 비식별화 범위: 얼굴, 차량 번호판
- 정제 저작 도구는 컨소시엄사에서 자체 개발하여 소스를 공개할 예정
- 획득 가능성
 - 이미 운용 중인 버스의 차량용 DVR 시스템을 이용하여 데이터를 취득할 예정이고 버스 업체와 자료 수집 협약을 기 체결함.
- 데이터 균형
 - 전방 카메라뿐 아니라 일반적으로 취득하기 어려운 좌/우 측방 카메라를 활용하여 주행 영상을 취득하고, 위반 감지를 위한 차로, 위반 객체(차량 종류), 시간대(주/야)를 고려하여 데이터를 분류함
 - 획득/정제되는 데이터 분포 비율을 매주 측정하고 부족한 데이터가 예상되면 이를 보완하여 데이터의 다양성을 확보함.
- 데이터 정확성
 - 차로 위반 상황 인지를 위해서는 중앙선, 버스전용 차로, 백색 실선 등의 구분이 필요한데, 이를 위해 실선, 점선, 색상 등의 항목으로 데이터를 분류함.
 - 차로 위반에 관한 법규와 빈번한 위반 행태를 데이터 라벨러들에게 교육하여, 데이터의 정확성을 높임

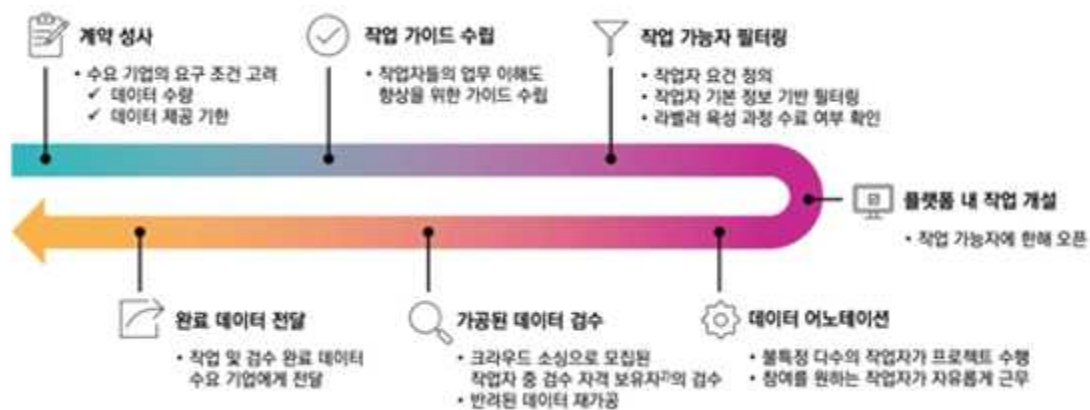
2.4 어노테이션/라벨링

2.4.1 어노테이션/라벨링 절차

- 획득→정제 과정을 통해 도출된 원천데이터를 라벨링하여 학습 데이터를 생성하기 위한 과정으로 다음의 사항들을 고려하여 라벨링 작업 진행

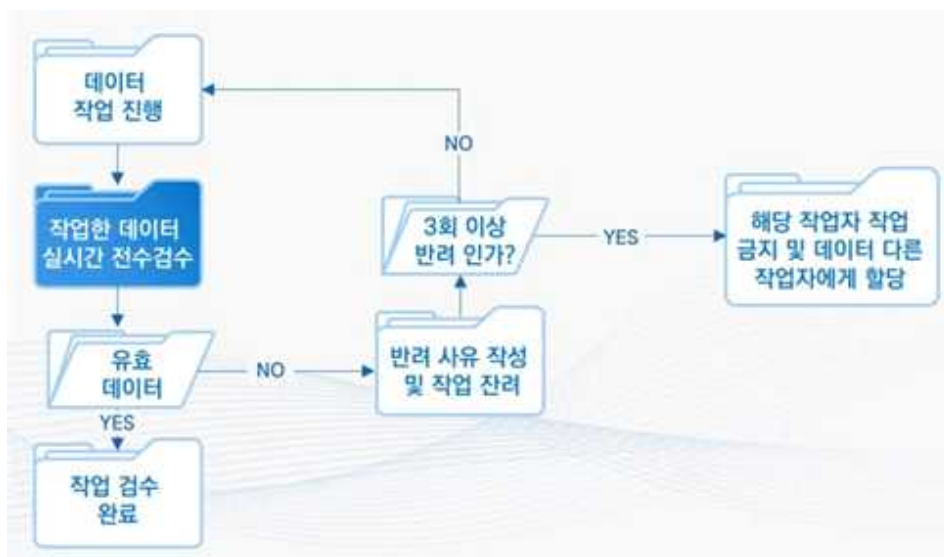
데이터획득형태	수집장비	데이터형식	데이터 처리	담당 인원
어노테이션데이터	클라우드소싱 서버	JSON	어노테이션 툴	고용 인원

[표 24] 라벨링 사항



[그림 15] AI 학습 데이터 가공 전문 기관 활용

- 작업의 배분 기준
 - 온라인 플랫폼을 통해 클라우드 워커들에게 데이터를 순차적·자동적으로 할당하며, 할당된 데이터를 가공 완료한 클라우드 워커들에게는 다음 시점의 데이터가 할당되는 방식으로 작업 배분 관리



[그림 16] 작업 배분

2.4.2 어노테이션/라벨링 기준

작업 기준	작업 방식
라벨링 작업 대상	<ul style="list-style-type: none"> 차량 : 원천데이터 획득차량 주행차로 및 좌우 차로 내 [표]에 제시한 차량 (4종) 폴리곤 세그멘테이션 <ul style="list-style-type: none"> 단, 차량객체의 전방, 좌/우 측방, 후방 중 최소 한 개이상의 단면이 가려 지거나 잘림이 없는 차, 100*50 픽셀 이상인 차량이 대상임 차선 : 원천데이터 획득차량 주행차선의 좌/우 차선이 작업대상으로 폴리곤 세그멘테이션 <ul style="list-style-type: none"> 단, 아래 차선은 작업 제외 <ul style="list-style-type: none"> : 2줄 점선, 복합선, 교차로의 갓길, 횡단보도, 정지선
라벨링 작업 기준	<ul style="list-style-type: none"> 오태깅 <ul style="list-style-type: none"> : 차량 가로/세로 100*50 픽셀 미만 태깅 대상 아님 (단, 이륜은 세로 100 픽셀 미만), √ 차량객체의 전방, 좌/우 측방, 후방 중 단 하나의 단면도 정확하게 표시되지 않는 (단면이 가려지거나 잘림이 있는) 차량은 태깅 대상 아님 - (차선) 라벨링 범대상/위가 아닌 차선은 태깅 대상 아님 과태깅 <ul style="list-style-type: none"> : (차량) 승용자동차, 승합자동차, 화물(특수)자동차, 이륜자동차가 아닌 대상은 태깅 대상 아님 : (차선) 정의되지 않은 클래스 차선은 태깅 대상 아님 미태깅 <ul style="list-style-type: none"> : (차량) 라벨링 대상/범위에 준하는 차량이 태깅되지 않은 경우 : (차선) 라벨링 대상/범위에 준하는 차선이 태깅되지 않은 경우
라벨링 범위	<ul style="list-style-type: none"> 차로위반 차량 <ul style="list-style-type: none"> : 위반차량이 있는 경우, 근접 위반 차량 1대 : 위반차량이 없는 경우, 해당 차로의 근접 차량 최대 3대 (측면) 바로 옆 차선에 위치한 차량 (정면) 주행차로와 좌우 차선에 위치한 차량 차선 <ul style="list-style-type: none"> : 위반차량이 있는 경우, 위반 차선 1개 : 위반차량이 없는 경우, (측면) 바로 옆 차선 1개 / (정면) 주행차로의 좌우 차선

[표 25] 라벨링 기준

2.4.3 어노테이션/라벨링 조직

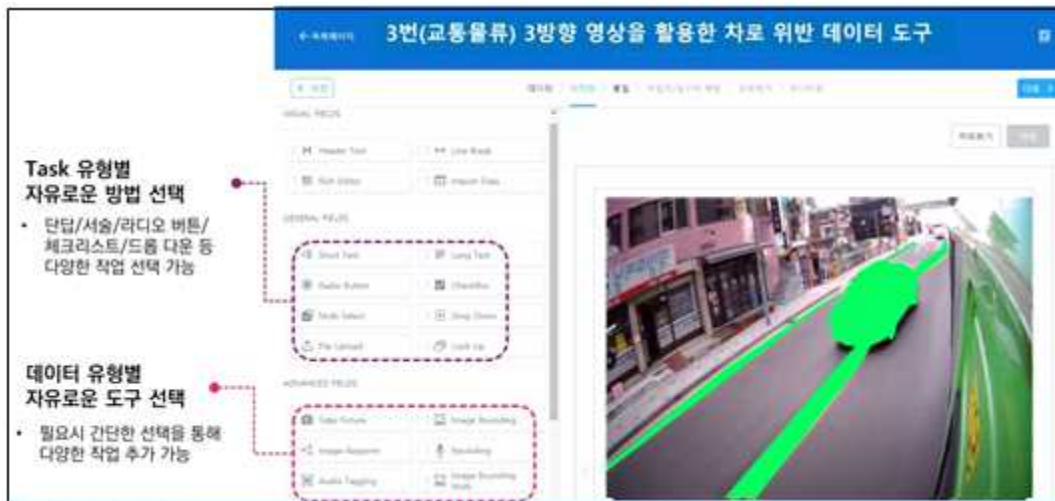
- 클라우드 소싱 활용
 - 모집 대상 : 학생, 취약계층, 경력단절 여성 등 나이, 성별, 학력, 직업 등의 제한 없이 누구나 참여하고 수행할 수 있음
 - 모집 방법 : 컨소시엄의 클라우드 소싱 플랫폼을 활용하고 수행기관 홈페이지 홍보 및 언론사 등을 통한 온라인 홍보를 진행할 예정이며, (주)클라우드웍스 플랫폼에 현재 등록되어있는 회원들을 중심으로 모집
 - 작업 참여 이전 교육 시행
 - : 이론, 실습, 점검 프로세스로 이루어진 라벨러 양성 교육을 통해 작업을 원하는 불특정 다수가 작업에 필요한 역량을 갖추도록 함
 - : 교육 이수자를 대상으로 수수료 배부
 - 작업 가이드 수립 : 작업자들의 업무 이해도 향상을 위한 가이드 수립
 - 작업 가능자 필터링
 - : 작업자 요건 정의 : 프로젝트 난이도 및 유형별 요건 정의
 - : 작업자 기본 정보 기반 필터링
 - : 수수료 보유 여부 기반 필터링
 - 플랫폼 내 작업 개설 : 작업 가능자에 한해 오픈
 - 데이터 어노테이션 실시
 - : 불특정 다수의 작업자가 프로젝트 수행
 - : 참여를 원하는 작업자가 자유롭게 근무
 - : 데이터 수요 기업의 요구에 따른 데이터 구축량 달성까지 작업 오픈
 - 가공된 데이터 검수
 - : 클라우드 소싱으로 모집된 작업자 중 검수 자격 보유자의 검수 시행
 - : 데이터 검수 후 승인 및 반려 판단
 - : 반려된 데이터 재가공 시행
- 클라우드 소싱 인력 지원 및 교육
 - 클라우드 워커(작업자, 검수자)의 전문성을 높이기 위한 데이터 라벨링에 특화된 커리큘럼 구성 및 교육 시행



[그림 17] 작업자 육성 교육 / 검수자 교육 과정

2.4.4 어노테이션/라벨링 도구

- 다수의 프로젝트에서 검증된 자체 보유 저작 도구 활용
- 작업 편리성을 극대화한 UI구성 및 요구에 따라 쉽고 빠르게 가능하도록 설계
- 객체에 대한 Drawing과 메타데이터 값을 동시에 진행할 수 있는 기능을 갖추고 있음
- 라벨링 저작도구 커스터마이징 : (주)클라우드웍스는 드롭&드래그 방식으로 저작도구를 생성할 수 있는 시스템 보유하고 있고 이 시스템을 이용하여 구축데이터 유형 맞춤 라벨링 저작 도구 디자인 가능
- 인터넷 환경에서 PC만 있으면 시간과 장소에 정함이 없이 컨소시엄사의 플랫폼에 접속해 언제든지 작업 가능한 클라우드 소싱 방식 플랫폼을 통해 데이터 라벨링 진행



[그림 18] 작업 Task 유형별 작업도구 커스터마이징 가능

- 데이터 라벨링 프로세스



[그림 19] 단계별 Task 및 품질 통계

○ 데이터 어노테이션 포맷 및 저장 형식

- 어노테이션 포맷

: 구축데이터는 원시 데이터인 CCTV 영상 파일과 원천데이터인 이미지 데이터는 비정형 데이터로 인공지능 모델 학습을 위하여 meta data가 기록된 json으로 구성

: 어노테이션 포맷의 형태는 이미지 파일(jpg)과 json 파일이 하나의 set

: json은 텍스트 형식의 개방형 표준 포맷으로, 구축된 데이터가 AI-hub를 통해 공개된다는 점을 감안하면 포맷의 사용성 및 개방성이 크다는 장점이 있음

○ 어노테이션 정보 저장 형식

- 컨소시엄사의 플랫폼을 통해 작업 진행

- 실시간으로 검수가 진행되며, 데이터는 모두 수행사의 GCS(Google Cloud Storage)에 저장됨. 저장된 데이터는 아래와 같은 구조의 json으로 추출

- 결과 값 외에 해당 데이터를 검수하는 과정에서 진행되는 반려 횟수, 반려 사유 등의 작업 이력 역시 DB에 저장됨

○ 저장 방법

- 구축데이터 생애주기별 저장 및 관리 방식

생애주기	1단계	2단계	3단계
획득 · 정제	(주)디텍 Storage Server	GCS 보조 데이터 센터	NCP ¹⁾ Storage
라벨링 · 학습	GCS ²⁾		
최종 데이터셋			인공지능허브 FTP

[표 26] 생애주기별 저장 관리

: 데이터 라벨링/검사 단계 : 컨소시엄사의 GCS에는 라벨링된 result 값과 json 파일, 원천 data가 저장되며, 해당 data는 GCS 내부에서 replica(복사본)를 유지하는 이중화 백업으로 데이터 유실에 대하여 대응

: GCS 데이터 센터 장애 발생할 경우를 대비 데이터를 보조 데이터 센터에 복제하여 백업

: 추가적으로 NCP 스토리지에 저장하여 3중 백업 방식으로 데이터 유실에 대응

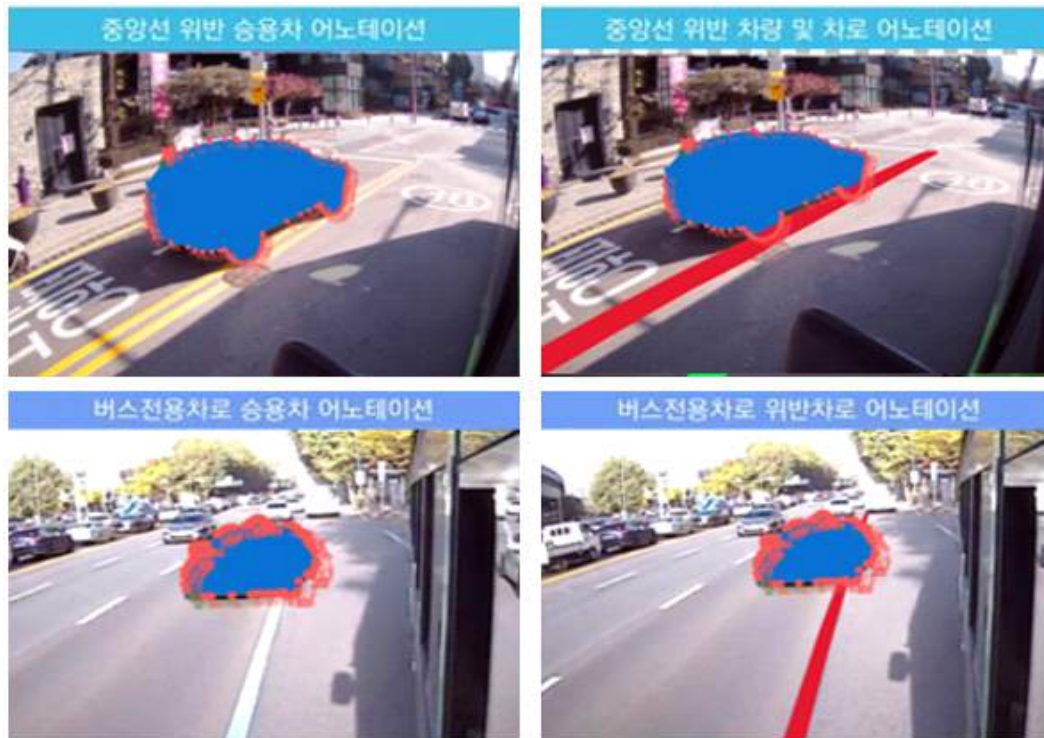
1) NCP : Naver Cloud Platform

2) GCS : Google Cloud Storage

폴더구조	파일명구조
<ul style="list-style-type: none"> ▼ 00과제_루트폴더 <ul style="list-style-type: none"> 00_초기데이터(10%) ▼ 01_데이터 <ul style="list-style-type: none"> 00_구축가이드 ▼ 01_샘플(Sample) <ul style="list-style-type: none"> 00_원천데이터 01_라벨링데이터 ▼ 02_시험(Test) <ul style="list-style-type: none"> 00_원천데이터 01_라벨링데이터 ▼ 03_검증(Validation) <ul style="list-style-type: none"> 00_원천데이터 01_라벨링데이터 ▼ 04_학습(Training) <ul style="list-style-type: none"> 00_원천데이터 01_라벨링데이터 02_저작도구 03_시모델 04_원시데이터 	<div style="display: flex; justify-content: space-around; margin-bottom: 10px;"> ①②③④ </div> <div style="text-align: center; font-size: 24px; margin-bottom: 10px;">A_2108261513_01_04.jpg</div> <div style="text-align: left;"> <p>① 카메라 채널 (A:전방, B:우측, C:좌측)</p> <p>② 타임스탬프 (예: 2021년8월26일15시13분)</p> <p>③ 차로 색상 (01:백색, 02:청색, 03:노랑색)</p> <p>④ 차로 유형 (01:실선, 02:점선, 03,이중실선 등등)</p> </div>

[표 27] 데이터 폴더 및 파일명 구조

○ 데이터 라벨링 적용 예시



[그림 20] 차량 바운딩 예시

2.5 검수

- 인공지능 학습용 데이터 구축사업을 통해 구축되는 데이터는 인공지능 학습 모델뿐만 아니라 데이터가 필요한 이들에게 공개되는 공공 데이터로 구축에 있어 품질의 매우 중요하며 이에 데이터 전 생애주기별 프로세스 측면과 데이터 측면의 검사절차와 품질관리 활동 필요

2.5.1 검수 절차

- 검사 프로세스는 학습용 데이터 구축 공정 단계별로 검사를 수행
- 데이터 구축 전 과정에 걸친 클라우드소싱 및 기술 기반 내부 검사 진행하고 데이터 표준 및 가공 가이드 정의를 위해, 전문가를 활용하여 구축데이터 전반 품질 향상
- 원천데이터 획득부터 검사 단계까지 모든 단계에서 적합한 검사 규모로 진행
- 수집/가공별 검사 가이드 작성
- 획득 및 가공 단계별로 가이드(체크리스트)를 작성하여, 해당 가이드 기반으로 데이터가 적합하게 획득 및 자동 가공되었는지를 모니터링



[그림 21] 구축 데이터 검수 절차

2.5.2 검수 기준

검사절차	검사 기준	요구 사항
1차 검사 (획득)	법/제도 준수	데이터 구축 시 관련 법/제도를 검토하고 해결방안을 제시
	사실적 환경 구성	학습데이터는 실제 특성이 반영되어야 함
	편향성 방지	편향된 데이터 구축 방지를 위해 고르게 수집을 진행 - 최소 20개 이상의 버스노선에서 원시데이터를 취득
2차 검사 (정제)	정제 기준의 명확성	정제 기준을 구체화하여 기준에 맞게 정제 작업 진행 - 카메라 각도, 해상도, 프레임 수, 영상화질 - 품질 목표치: 제외되는 데이터 비율 10% 이하
	중복 및 누락 방지	데이터가 중복 및 누락 없이 수집되었는지 확인
	작업 매뉴얼	정제 방법 및 기준에 대한 가이드라인을 제공
3차 검사 (라벨링)	정제 도구	정제 작업을 위한 적절한 도구를 마련하여 제공
	필드 항목 정리	mapping되어야 하는 필드 항목들을 기준에 맞추어 정리
	라벨링 검사 도구	meta 정보를 mapping하는 라벨링 작업을 시스템적으로 자동 진행
4차 검사 (전수)	라벨링 작업 모니터링	시스템을 통해 라벨링 된 학습 데이터가 기준에 맞게 필드 값이 채워졌는지 모니터링 시스템을 통해 확인
	부적합 판정 데이터 분포 확인	정의된 품질기준 미달로 학습데이터 활용가치가 없는 부적합 데이터를 판정
	외부 검사자	외부 검수 전문기관에서 수립된 가이드 기준에 따라 검수

[표 28] 구축 데이터 검수 기준

2.5.3 단계별 검수 방법

- 계획 수립 단계부터 학습데이터셋 구축을 완료하는 단계까지 전 생애주기 단계별 핵심 Task 수행으로 양질의 인공지능 학습용 데이터를 구축

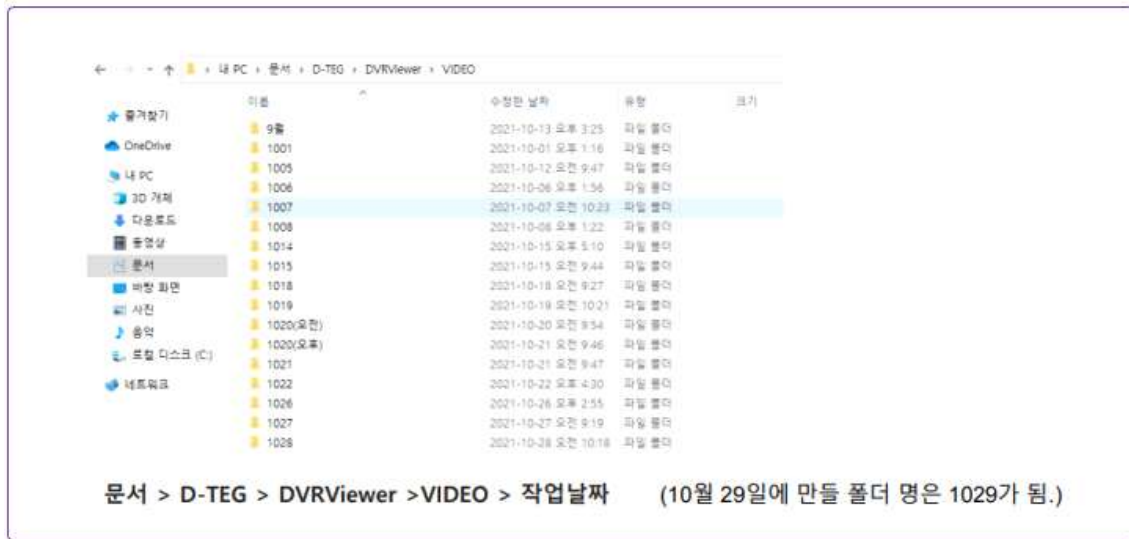


[그림 22] 단계별 검사 및 개선을 통한 모델의 품질 향상 항목

2.5.4 검수 도구 및 작업방법

○ 작업방법

1) 폴더 생성

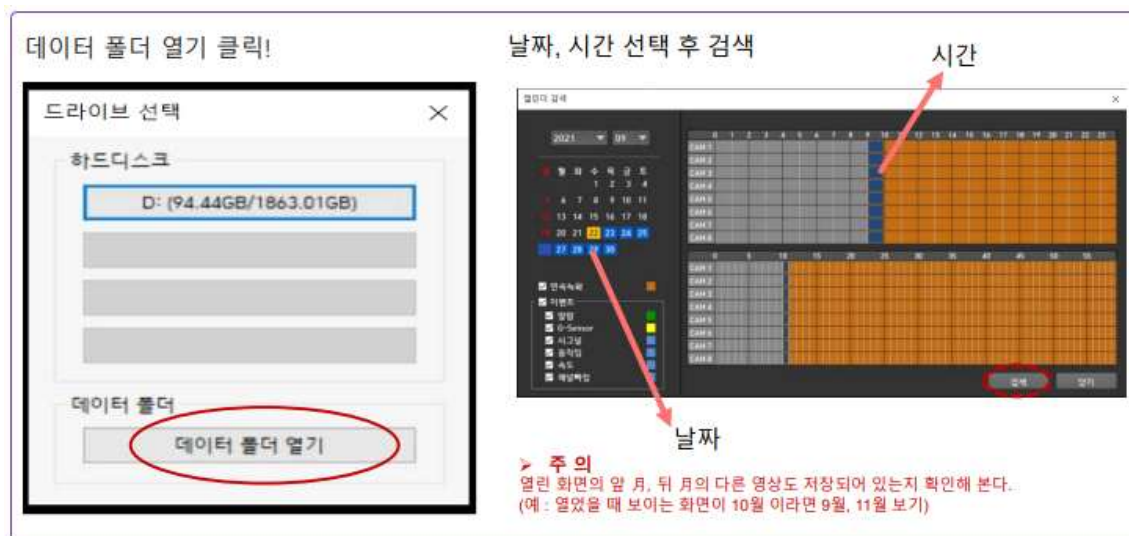


[그림 22] 작업 폴더 생성 예시

2) 프로그램 실행



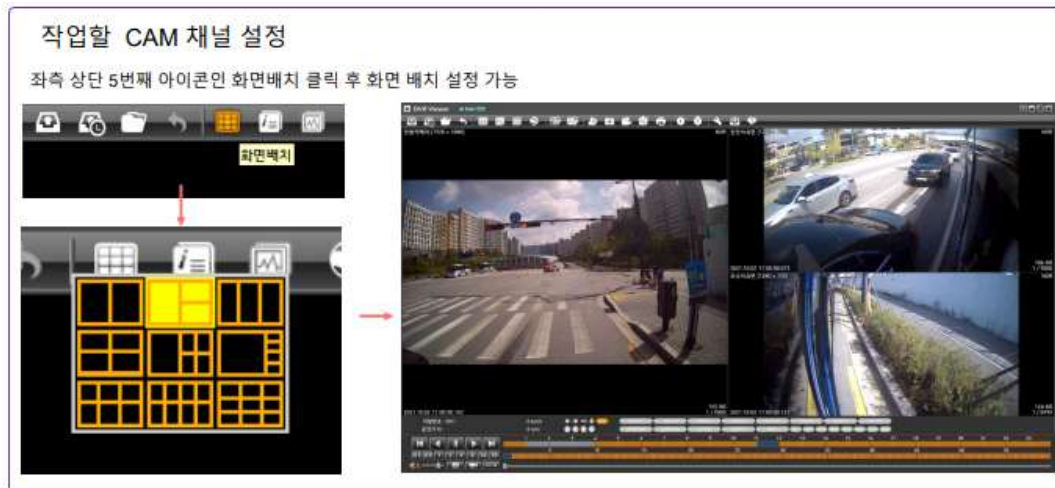
[그림 23] 단계별 검수 방법 - 프로그램 실행 1



[그림 24] 단계별 검수 방법 - 프로그램 실행 2



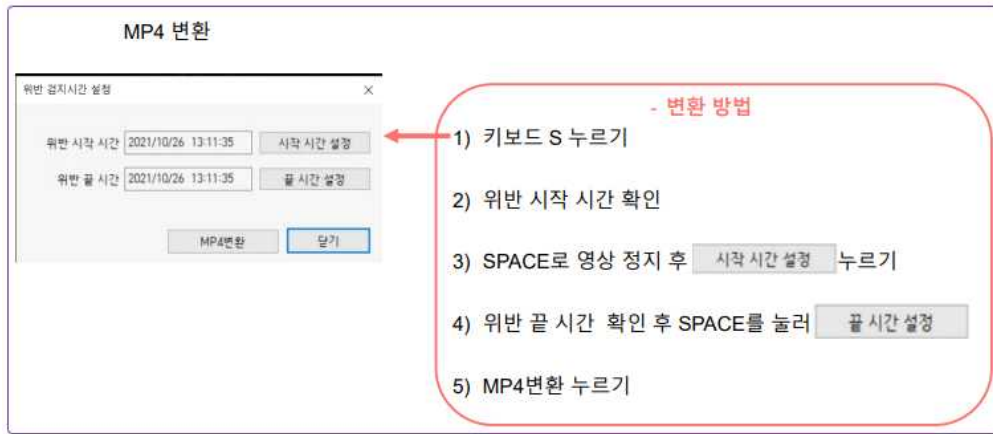
[그림 25] 단계별 검수 방법 - 프로그램 실행 3



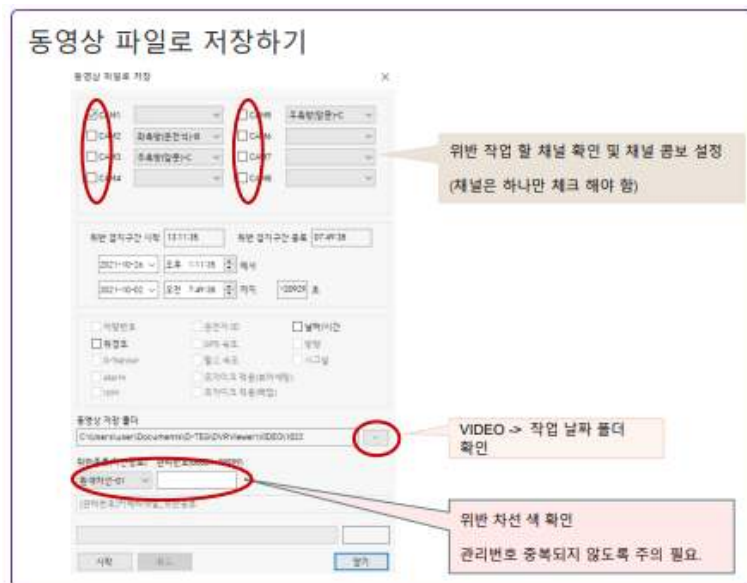
[그림 26] 단계별 검수 방법 - 프로그램 실행 4



[그림 27] 단계별 검수 방법 - 프로그램 실행 5



[그림 28] 단계별 검수 방법 - 프로그램 실행 6



[그림 29] 단계별 검수 방법 - 프로그램 실행 7



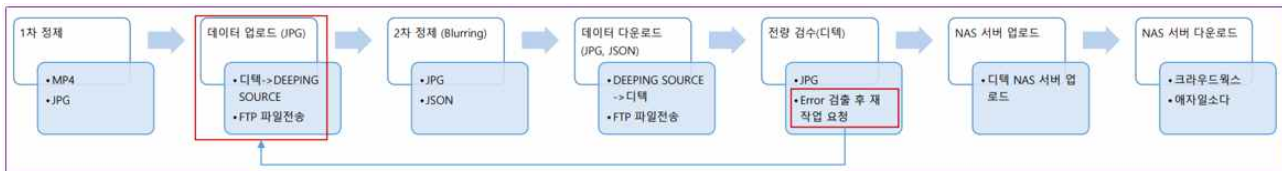
[그림 30] 단계별 검수 방법 - 프로그램 실행 8

○ 데이터 2차 정제(비식별화) 기준



[그림 31] 2차 정제 기준

○ 데이터 2차 정제(비식별화) 프로세스



[그림 32] 2차 정제 프로세스

2.6 활용

2.6.1 활용 모델

데이터명	AI 모델	모델 성능 지표	응용서비스(예시)
차선 위반 영상 데이터	차량 인지 모델	mAP(loU 0.5 기준, 최소 사이즈 90x50px, 이륜차는 세로 100px) 85%	<ul style="list-style-type: none"> 차선 위반 차량 인식 전용 차로 위반 인식
	차선 인지 모델	mAP(loU 0.5 기준, 최소 사이즈 90x50px) 85%	
	교통법규 위반탐지 모델	F1 score 85%	

[표 29] 인공지능 데이터 활용 모델 개발 요약

2.6.2 유형 정의

<p>중앙버스 전용차로 청색 평일 토·공휴일 24시간 운영</p>	<p>전일제버스 전용차로 청색2줄 평일 07:00~21:00 토요일·공휴일은 단속하지 않음</p>	<p>시간제 버스전용차로 청색1줄 평일 07:00~10:00 17:00~21:00 토요일·공휴일은 단속하지 않음</p>	<p>차선위반 유형</p> <p>지정 차로 위반</p> <ul style="list-style-type: none"> 허용되지 않은 구역 (Zone)의 침범(예) 버스 전용 차선 위반 <p>차선 위반</p> <ul style="list-style-type: none"> 불법 차선 변경+지정 차선 위반 예) 실선 차선에서의 변경 등 (터널 내의 변경도 차선으로 구분) 중앙선 침범 위반 	<p>지정차로 위반</p> <ul style="list-style-type: none"> 지정 차선 위반 인식 대형트럭 등의 1차선 주행 인식 <p>불법 차선변경</p> <ul style="list-style-type: none"> 실선차선 변경 인식 중앙선 침범 인식 <p>버스 전용차선 위반</p> <ul style="list-style-type: none"> 버스전용차선 인식 청색 이중실선 + 중앙선 (황색 이중실선) 청색 이중실선 + 도로표 버스전용차선 구역에서 버스 외 차량 인식 중앙선 또는 끝차선 인식 청색 차선 인식 청색 차선 기준, 중앙선 또는 끝차선과 가까운 차선이 버스전용차로
---	---	---	---	--

[그림 33] 비병차로 위반(왼쪽) 및 차로 위반 유형의 정의(오른쪽)

1) 지정차로 위반

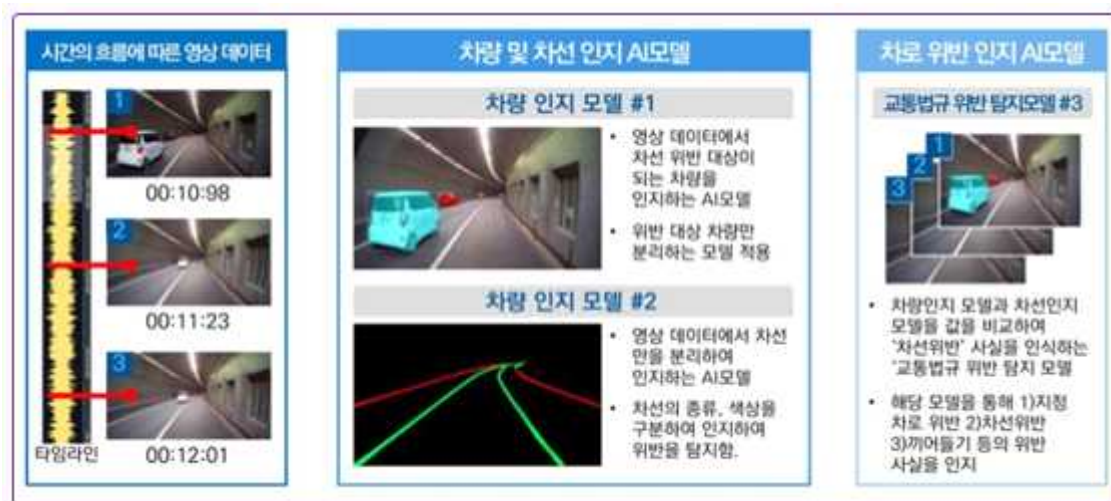
- 허용되지 않은 구역의 침범 : (예) 버스 전용차로 위반

2) 차로 위반

- 불법 차로 변경, 지정차로 위반(터널 내 실선 차선 변경)
- 중앙선 침범 위반
- 실선 차선 변경 인식

2.6.3 인공지능 모델 처리 프로세스

- 1) 스틸 이미지 한장 입력
 - 단속 대상 차량의 이미지 입력
- 2) 차량 인지 모델
 - 영상에서 대상 차량을 인식하는 인공지능 모델
 - 위반 대상 차량만 분리하여 모델 적용
- 3) 차선 인지 모델
 - 영상 데이터에서 차선만 분리하여 인지하는 인공지능 모델
 - 차선의 종류(실선, 점선 등), 차선의 색상(백색, 황색, 청색)을 인지하는 모델
- 4) 차로 위반 인지 모델
 - 차량 인지 모델과 차선 인지 모델을 결합 비교하여 차로 위반 사실을 인식함
 - 이를 통해 대상 차량이 차선을 위반했는지 여부에 대해 판단



[그림 34] 인공지능 모델 처리 프로세스

2.6.4 인공지능 모델 적용 알고리즘

- 1) 차량 인지 모델
 - Segmentation 기법을 적용하여 차량을 인식
 - Mask R-CNN, Cascade Mask R-CNN 등의 알고리즘 사용
- 2) 차선 인지 모델
 - Segmentation 기법을 적용하여 차선을 인식
 - Mask R-CNN, Cascade Mask R-CNN 등의 알고리즘 사용

3) 교통법규 위반 탐지 모델

- 이미지를 보고 위반 여부를 판단할 수 있는 분류 모델
- ResNet, ViT 등의 알고리즘 활용

모델 구분	구현 방안	적용알고리즘
차량 인지 모델	<ul style="list-style-type: none"> Segmentation 기법을 적용하여 승용차, 버스, 이륜차를 인식 	<ul style="list-style-type: none"> Mask R-CNN Cascade Mask R-CNN
차선 인지 모델	<ul style="list-style-type: none"> Segmentation 기법을 적용하여 차선의 색상과 종류를 인식 	<ul style="list-style-type: none"> Mask R-CNN Cascade Mask R-CNN
교통법규 위반 탐지 모델	<ul style="list-style-type: none"> 차량과 차선 정보를 활용하여 이미지 속 차량의 위반 여부에 대한 직물을 계산해 분류 	<ul style="list-style-type: none"> ResNet ViT

[그림 35] 인공지능 모델 적용 알고리즘