

인공지능 데이터 구축 가이드라인

- 차량파손 이미지 데이터 -

담당 역할	기관명
데이터 구축 총괄	한국전기차리빌딩협회
데이터 설계	한국전기차리빌딩협회, 쏘카, Superb AI, 꿀비
데이터 수집	한국전기차리빌딩협회, 쏘카
데이터 정제	Superb AI
가공(라벨링, 어노테이션)	Superb AI
데이터 검수(자체 검수)	한국전기차리빌딩협회, SuperbAI, 꿀비

구축 가이드라인 작성	한국전기차리빌딩협회	김현식
	꿀비	정상현, 안수현
가이드라인 버전 (제작일자)	ver 2.0 ('22.6.28)	

목 차

1. 데이터 구축 개요	1
2. 문제 정의	4
2.1 임무 정의	4
2.2 데이터 구축 유의사항	4
3. 데이터 수집·정제	5
3.1 원시데이터 선정	5
3.2 수집·정제 절차 및 방법	6
3.3 수집·정제 기준	12
3.4 수집·정제 도구	15
4. 데이터 가공	18
4.1 라벨링 절차 및 방법	18
4.2 라벨링 기준	19
4.3 라벨링 규격	28
4.4 라벨링 조직	30
4.5 라벨링 도구	30
5. 검수	34
5.1 검수 절차 및 방법	34
5.2 검수 기준	37
5.3 검수 조직	39

1. 데이터 구축 개요

자동차 정비소 및 쏘카 등에 축적되어 있는 차량사고 이미지에 대해 정제, 가공을 통하여 AI 데이터 셋 구축하고 사고차량 견적을 위한 AI 데이터 셋을 기반으로 사고 견적을 도출할 수 있는 알고리즘을 개발하여 보험사, 공공기관, 자동차 관련 기관 등에서 서비스 개발에 활용 할 수 있도록 함.

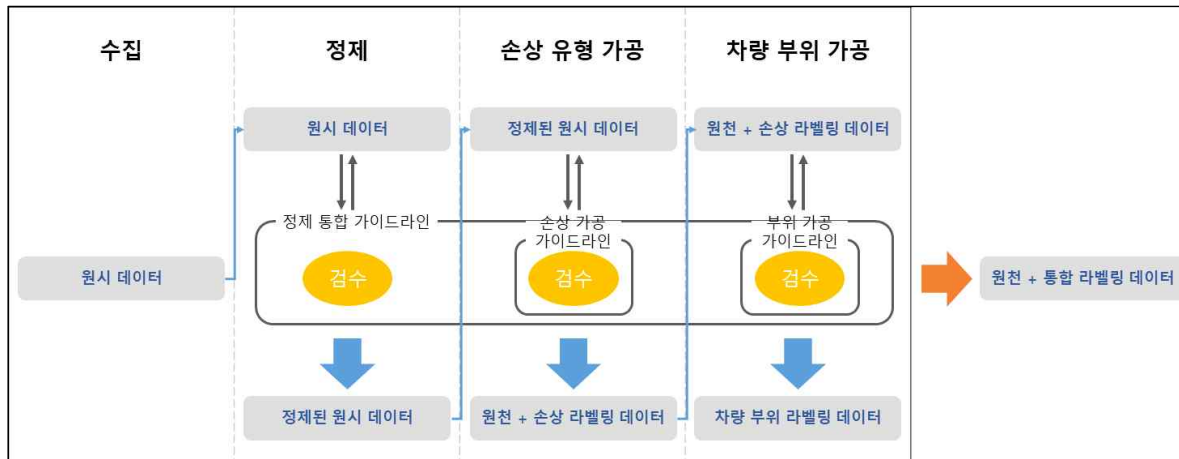
○ 데이터셋 명: 차량파손 이미지 데이터 셋

○ 활용분야

- 사고차량에 대한 수리비 자동산출을 통해 보험업계, 손해사정인, 자동차정비소, 부품유통업 등에서 활용 가능

○ 목적 및 필요성

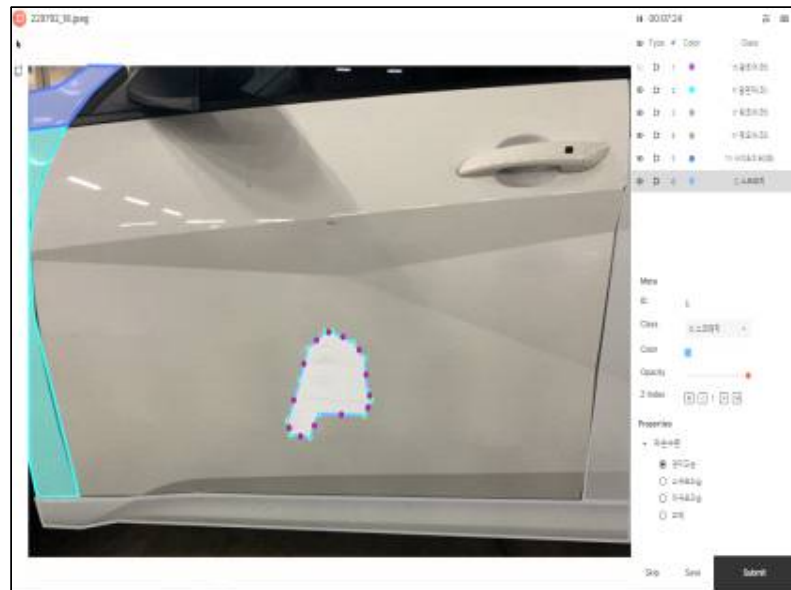
- 교통사고 등 차량 손상시 손상 부위와 수리비용에 대해 보험사와 정비소 간에 증빙을 위하여 이미지를 촬영하고 수리견적서를 첨부하여 주고받음으로서 데이터가 축적되고 있으며, 쏘카 등 렌터카 업체에서도 차량의 손상 부위 입증을 위하여 렌터전과 렌터 후의 비교를 위해 촬영을 한 후 관리시스템에 업로드하고 있음
- 따라서 이러한 축적된 데이터를 활용하기 위하여 사고차량의 손상부위 이미지를 촬영하고 시스템에 적용함으로써 자동으로 수리비용을 산출할 수 있도록 인공지능 데이터셋을 구성하고자 함
- 이러한 데이터셋을 활용하여 자동으로 수리비를 산출하는 시스템개발을 통하여 자동차 사고현장에서 즉시 수리비를 예측할 수 있게 함으로써 사고 당사자 간의 분쟁 해결 및 수리에 대한 신속한 판단을 통하여 정비업소, 보험사, 손해사정인, 차량 소유주들에 불필요한 업무 제거, 신속한 업무처리 등으로 비용을 절감할 수 있음



(그림) 구축 프로세스 흐름도

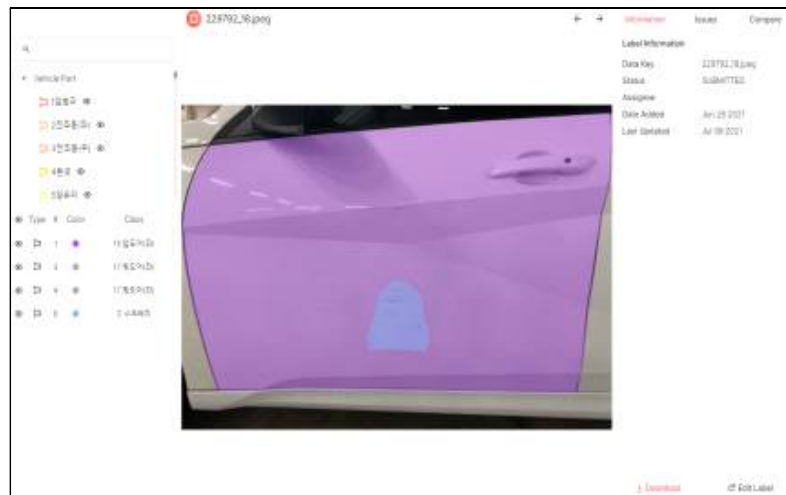
<p>1. 데이터 획득</p>	<div data-bbox="496 293 868 607"> </div> <div data-bbox="628 622 735 645"> <p>쏘카 보유 데이터</p> </div> <div data-bbox="884 286 1362 645"> </div> <div data-bbox="1098 622 1278 645"> <p>자동차 정비업소 보유 데이터</p> </div> <div data-bbox="788 667 1054 712"> <p><원시 데이터 수집></p> </div>
<p>2. 데이터 저장</p>	<div data-bbox="536 815 1307 1328"> </div> <div data-bbox="715 1346 1129 1384"> <p><Cloud Storage(AWS) 와 연동></p> </div>
<p>3. 데이터 정제</p>	<div data-bbox="501 1420 1315 1865"> </div> <div data-bbox="692 1883 1129 1921"> <p>< 정제기준에 따라 Image 선택></p> </div>

4. 데이터 라벨링



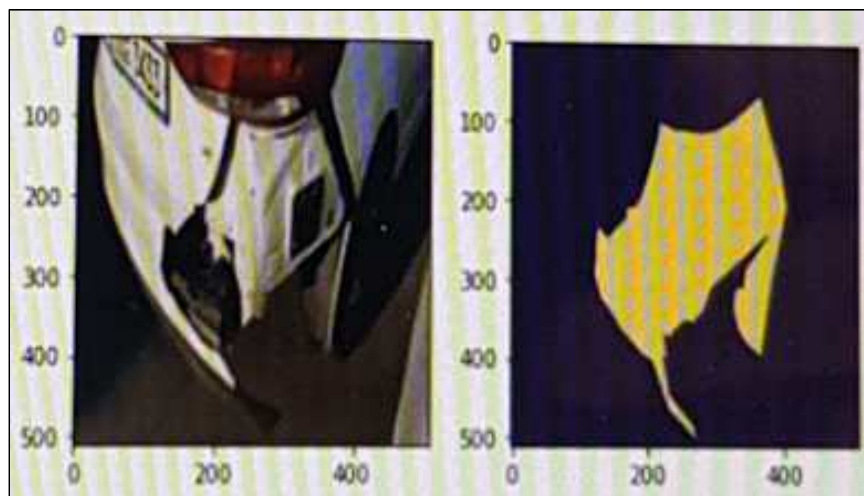
<가공 기준에 맞게 라벨링 수행>

5. 데이터 검수



<전문가 검수>

6. AI모델 개발 및 모델/데이터 검증



<AI 예측 결과값 확인>

2. 문제 정의

2.1. 임무 정의

- 인공지능이 자동차의 손상종류 및 손상심각도를 이해하고 교체/수리 등의 정비 방법 및 이를 위한 수리내역을 도출하기 위해 AI SW가 해당 이미지의 세그멘테이션과 손상범위를 이해할 수 있는 형태로 가공된, 다양한 차종 및 부품에 대한 손상 데이터 구축이 요구됨
- 손상된 차량부위와 손상심각도를 파악하여 수리비 자동 산출을 위한 차량과 손 이미지 데이터셋 구축
- 인공지능이 사고차량 사진을 통해 손상 차량부위와 손상심각도를 파악하고 수리비를 자동으로 산출해주는 기술 개발 등에 활용 목적
- 인공지능 학습용 데이터 구축량 : 사고차량 외관부품별 근거리 촬영 및 원거리 촬영 손상 이미지 각각 최소 2,000장 이상 총 50만장 데이터셋 구축

2.2. 데이터 구축 유의사항

- 저작권 침해, 데이터 활용 범위 제한과 같은 법적 문제 발생 소지를 제거하기 위해 촬영된 이미지와 데이터에 대한 재산권 확보
- 개인정보 침해 방지
 - 정제/가공/검증에 참여하는 클라우드 소싱/인하우스 인력은 비밀유지/개인정보처리 동의에 서명한 인력에 한해서 참여함.
 - 이미지 내 안면 정보 유출을 방지하기 위해, 직접 안면이 보이는 이미지나 표면에 비춰 식별 가능한 안면이 있는 이미지는 원천 데이터에서 제외함.
 - 신체 정보로 인해 개인을 특정할 가능성이 있는 이미지는 데이터에서 제외함.
 - 수집된 이미지에서 개인을 특정할 수 있는 개인정보 (이름, 전화번호, 구체적 주소 등 기타 개인을 특정할 수 있는 정보)가 포함된 경우 원천 데이터에서 제외함.
 - 차량 번호판은 개인정보이지만, 차량파손 이미지 데이터셋 구축에 있어 필연적으로 수집됨. 따라서 원천 데이터의 차량 번호판에 대해 비식별화를 통해 개인정보 보호

3. 수집·정제

3.1. 원시데이터 선정

○ 데이터 특성

- 차량 외관 부품에서 나타날 수 있는 다양한 손상종류를 학습하여 손상 심각도를 판단할 수 있는 데이터
- 사고데이터 수집을 위해 인위적으로 발생시킬 수 없는 관계로 인해 이미 보유하고 있는 사고차량에 대한 이미지를 확보하는 방법 채택



(그림) 차량 파손 이미지 예

○ 대상 데이터

- 자동차 사고현장에서 촬영한 사고 차량 이미지
- 정비공장에 사고 차량 입고 시 다각도에서 촬영하는 사고 차량 손상부품 이미지
- 파일 형식 : JPG 파일
- 해상도 : 800×600 이상

○ 데이터 규모 및 분포

- 목표 수집 수량: 이미지 데이터 90만장(수리비 전적 데이터 제외)
- 구축 목표수량: 사고차량 이미지 50만장
- 차종, 근/원거리에 따라 데이터 파일 분류. 각각의 파일 분류기준에 따라 이미지 최소 2,000장 이상 수집
- 대상 차종 및 차종별 수량은 적절한 비율로 제시
 - 손상부위 촬영거리에 따라 나타나는 결함종류별 특징이 다르므로 근거리 촬영사진과 원거리 촬영사진이 적절한 비율로 구성
- 객체 클래스별 데이터 편향 없는 데이터 구축을 위해 기준 수립

클래스	클래스 속성	비고
차종	경형, 소형, 중형, 대형	버스, 화물차 등 제외
연식	2015 ~ 2021	수집량 부족시, 전체 데이터의 10%는 2015년 이전 연식 수집 허용

색상	흰색, 검은색, 회색, 빨강 계열, 파랑 계열, 기타	-
----	-------------------------------------	---

○ 원시데이터 명세서

데이터 명		차량 파손 부품 이미지 데이터
데이터 포맷		JPG(이미지 데이터), PDF(텍스트 데이터)
활용 분야		차량 손상부품과 유형을 출력하고 그에 따라 사고 심각도 및 예상 수리비용을 예측할 수 있는 모델 개발에 활용
데이터 요약		다양한 차량 손상부품과 손상유형이 보이는 이미지 데이터셋과 매칭되는 수리비 견적서 파일이 존재하는 차량손상 Binary classification 인공지능 데이터셋
데이터 출처		쏘카 사고관리시스템 및 정비업체로부터 확보
데이터 이력	배포버전	Accident_repair_dataset_v1
	개정이력	신규
	작성자/배포자	데이터 획득 수행기관(쏘카)
데이터 통계	데이터 구축 규모	차량 손상 이미지 총 65만 건 ~ 75만 건 차량 수리 견적서 총 10만 건 ~ 15만 건
	데이터 분포	차량별 분포: 경차(8.0~14.0%), 소형(10.0~18.0%), 중형(40.0%), 대형(28.0~30.0%) 거리별 분포: 근거리 60~80%, 원거리 20~40%

○ 지적 재산권 및 개인정보 보호 침해 관련

- 데이터에 대한 재산권 확보
 - 쏘카 보유 데이터: 자체 공개
 - 정비소 보유 데이터: 구매시 계약서에 재산권 및 활용계획 명시
- 개인정보 노출 등 아래의 경우 제외
 - 이름, 개인을 특정 지을 수 있는 주소, 전화번호 등
 - 기업 번호 및 특정 지을 수 있는 주소 등(상호명은 라벨링 가능)
- 차량번호판이 부착된 이미지에 대해서는 번호판 부위를 폴리곤 영역으로 라벨링하여 비식별화 처리

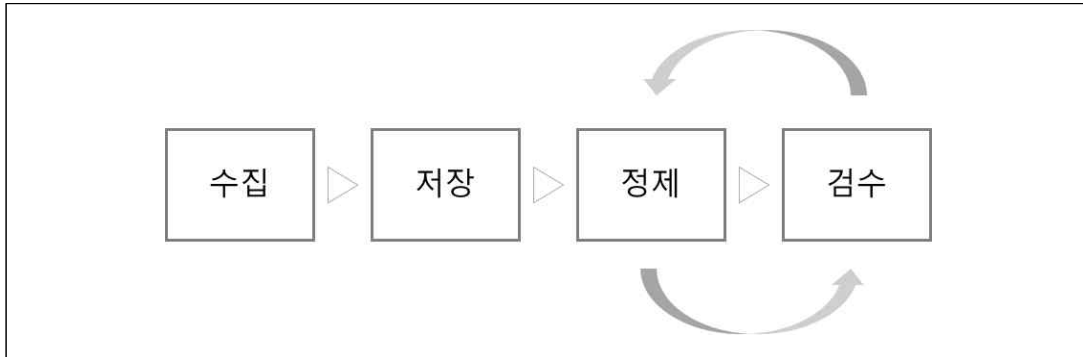
3.2. 수집·정제 절차 및 방법

(가) 데이터 수집 절차 및 방법

○ 데이터 수집 절차

- 손상 예측 모델 및 부위 분류 모델, 원근 추정 모델을 적용하여 이미지 데이터에 1차 태깅 수행
- class 불균형을 해결하기 위해 획득 기간을 이원화
- 추가 획득 기간에서 수집 기준량에 못 미치는 이미지를 추가 획득

- 획득 이후, 파일명 재정의 및 데이터 파일 분류 시행
- 데이터의 획득을 위한 기준을 수립하고, 그 기준에 맞는 이미지를 수집 하되 확보가 어려운 상황이 발생할 시 그 내역을 기록함으로써 데이터 이용자 및 향후 추가 수집을 위한 레퍼런스로 삼도록 함



(그림) 데이터 수집, 정제, 검수 절차

○ 데이터 수집 방법

- 국내 최대의 공유 차량 서비스 운영중인 쏘카로부터 수집
 - 차량 손상 이미지 수집 Pipeline 구축 및 관련 서비스 운영 중 (클라우드 기반)
 - 차량에 대한 정상 데이터 및 손상데이터 확보 용이
- 자동차 정비업소를 통해 데이터 활용 및 공개에 대한 동의를 득하여 이미 촬영된 사고차량 손상부품 이미지 구매
- 재산권 확보
 - AIHub 공개시 데이터 권리 문제로 인한 이슈가 없도록 구매계약서에 자세히 기술

○ 수집 이미지 속성 정의

No	속성명	항목 설명	데이터 타입	필수 여부	작성 예시
1	filename	파일명	String	Y	210511_4_000001_1_01.jpg
2	date	촬영날짜(년,월), 시간	Date	N	21-05-11-18-20
3	file format	파일 형식(포맷)	String	Y	jpg
4	imsize	이미지 파일 크기	String	N	50.9KB
5	images_Photographer	촬영자	String	N	홍길동
6	device(camera, lidar)	장비정보	String	N	스마트폰
7	images_location	촬영 위치	String	N	전남 목포시
8	copyright	저작권 정보	String	N	쏘카
9	width, height	이미지 사이즈	String	N	800X600

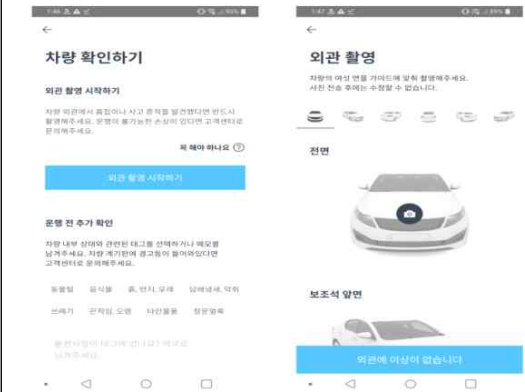
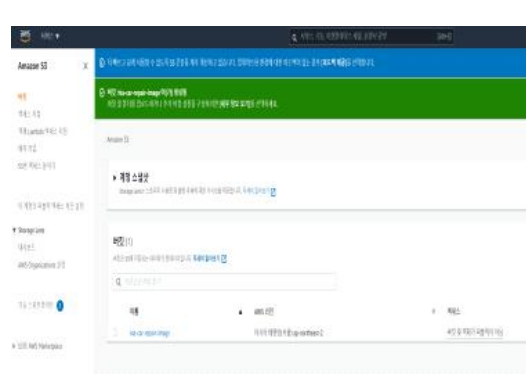
10	Aspect ratio	비율(중형비)	String	N	4:3
11	resolution	해상도(DPI)	String	N	96X96
12	bit	비트값	Integer	N	컬러색상/기본 24bit
13	depth	RGB 여부	String	N	RGB

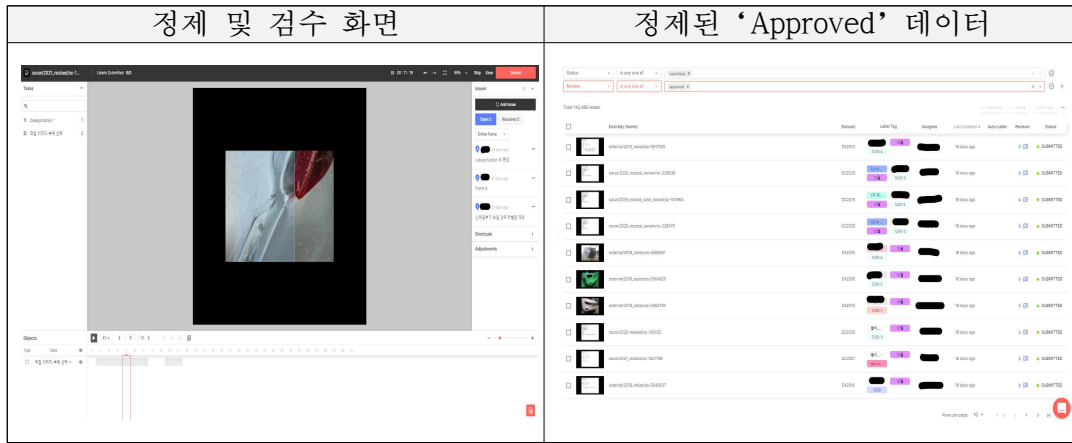
(나) 데이터 정제 절차 및 방법

- 보유 데이터의 특성상 하나의 사고(수리) 단위로 묶어 차량의 수리전/수리중/수리후 영상이 혼재되어 있음.
- 수집 이미지에 대해 정제 기준에 따라 이미지 선별
 - 촬영 각도/구도/거리에 따라 손상의 형태와 유형이 달라 보일 수 있어, 수리 전, 수리 중 수리 후의 영상을 비교하며 정확한 손상 위치를 유추.
 - 따라서 차량 이미지를 사고 단위로 묶어 정제를 진행.
 - 하나의 사고 단위 데이터를 정제하는 과정은 아래의 표와 같음.

원시 데이터 폴더 리스트			폴더 내부 데이터	
이름	수정된 날짜	유형		
as-0000006	2022-01-11 오후 2:59	파일 폴더		
as-0000007	2022-01-11 오후 2:59	파일 폴더		
as-0000032	2022-01-11 오후 2:59	파일 폴더		
as-0000039	2022-01-11 오후 2:59	파일 폴더		
as-0000046	2022-01-11 오후 2:59	파일 폴더		
as-0000055	2022-01-11 오후 2:59	파일 폴더		
as-0000059	2022-01-11 오후 2:59	파일 폴더		
as-0000075	2022-01-11 오후 2:59	파일 폴더		
as-0000081	2022-01-11 오후 3:00	파일 폴더		
as-0000092	2022-01-11 오후 3:00	파일 폴더		
as-0000093	2022-01-11 오후 3:00	파일 폴더		
as-0000094	2022-01-11 오후 3:00	파일 폴더		
as-0000143	2022-01-11 오후 3:00	파일 폴더		
as-0000147	2022-01-11 오후 3:00	파일 폴더		
as-0000151	2022-01-11 오후 3:00	파일 폴더		
as-0000155	2022-01-11 오후 3:00	파일 폴더		

(그림) 원시데이터 구조

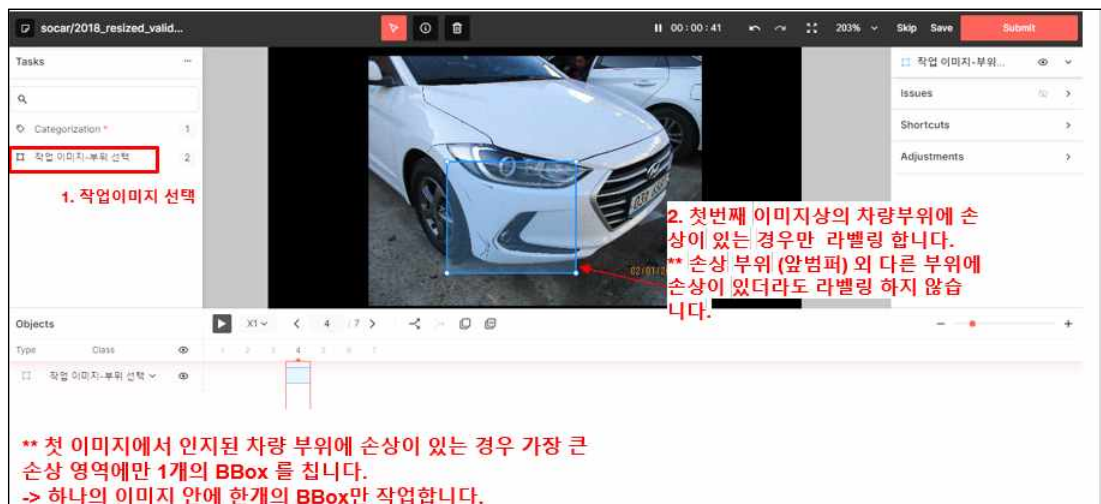
쏘카 어플리케이션	쏘카 서버와 컨소시엄 AWS와 연동
	



(그림) 데이터 수집, 정제 과정 예시

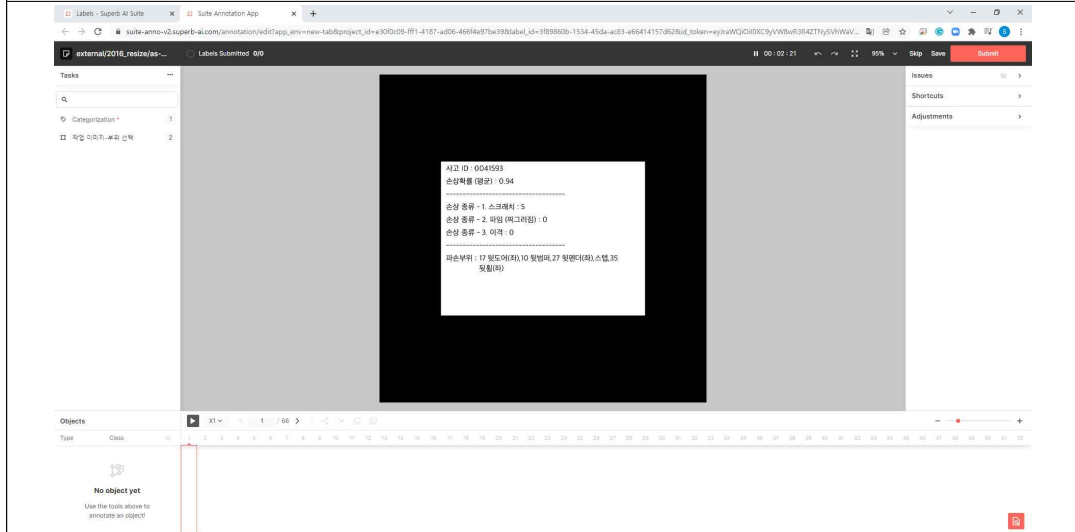
○ 정제 방법 및 절차

- 쏘카 사고관리 시스템에 업로드된 공업사 이미지 및 전문 사고관리 매니저의 등록 사진의 손상 이미지, 정비협회에서 협약한 공업사의 차량 손상 이미지를 수집
- 원시 데이터를 Cloud Storage(AWS)와 데이터 구축 소프트웨어인 “Superb AI Suite” 의 서버에 저장.
- 비밀유지/개인정보처리 동의에 서명한 클라우드 소싱/인하우스 인력은 각각 정제 작업자와 정제 검수자로 구분됨.
- 각각의 작업자에게 데이터를 사고 단위로 묶인 데이터를 할당함.
- 작업자가 정제 가이드라인에 따라 원천 데이터가 될 이미지를 Bounding Box를 이용해 선별하면, 검수자는 Bounding box를 보고 정제 가이드라인 준수 여부에 따라 “Approve” 나 “Reject” 를 진행.
- “Approve” 가 된 이미지는 손상유형 가공 단계로 넘어가고, “Reject” 된 이미지는 검수자의 “Reject” 사유와 함께 할당된 작업자에게 다시 전달됨.

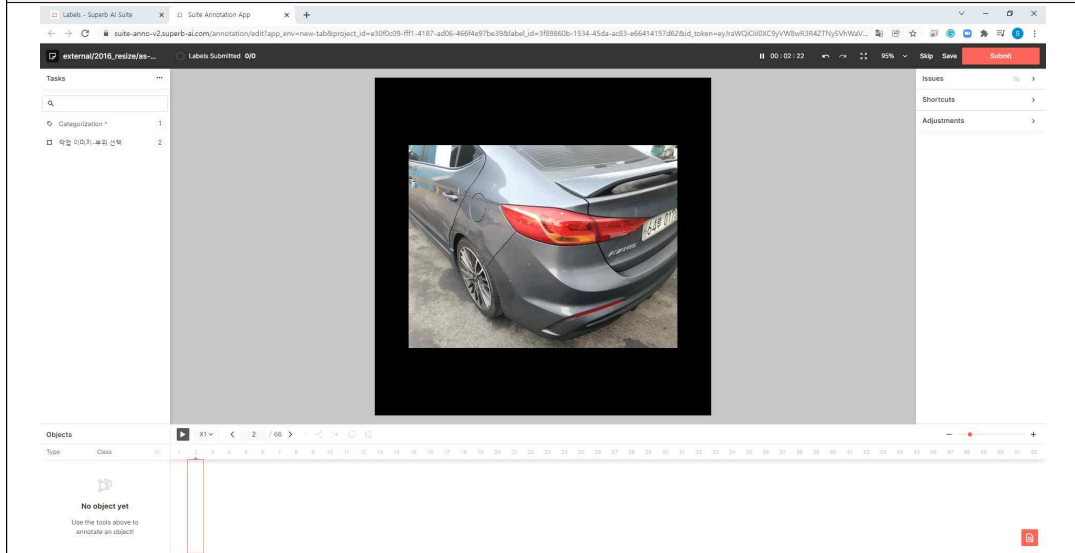


(그림) 도구를 이용한 정제 예시 과정

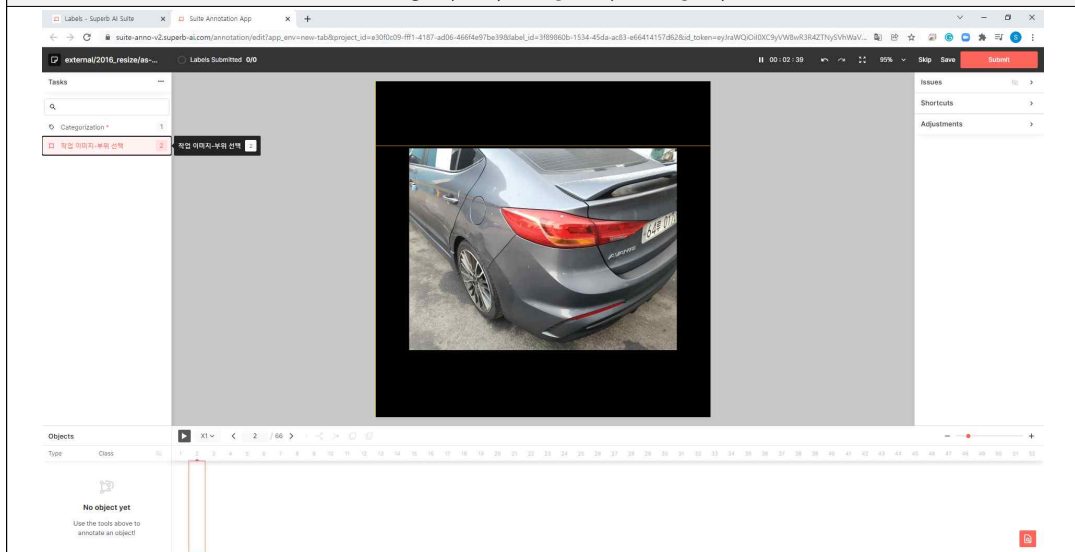
1. 사고 단위 데이터에 대한 정제 시작



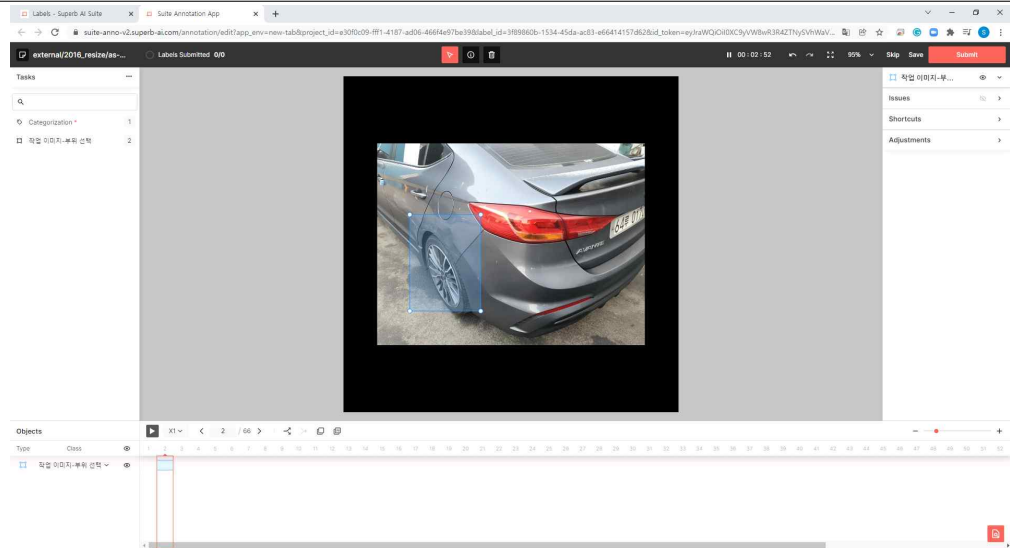
2. 정제 가이드라인에 기반을 두어 이미지 확인



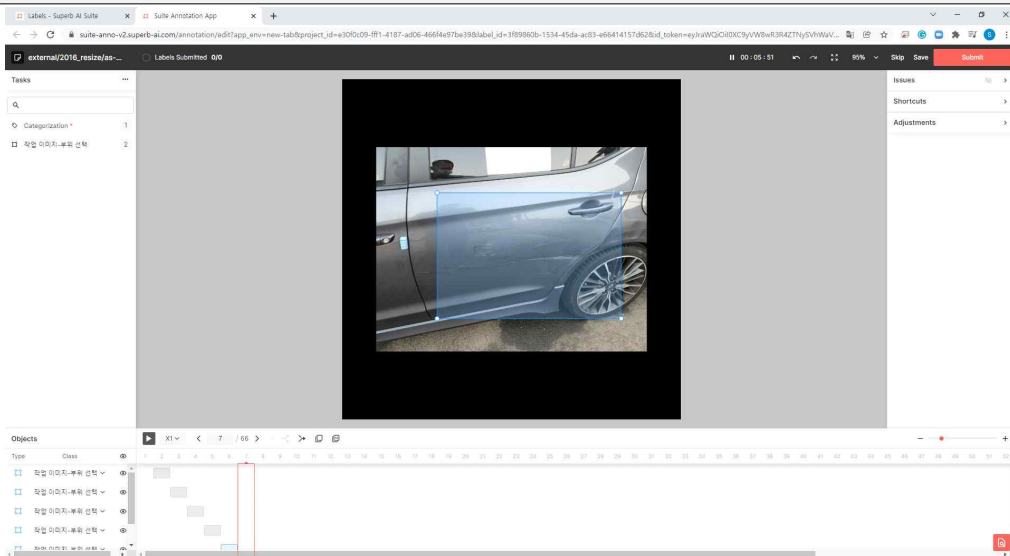
3. 정제 라벨링 기준 통과



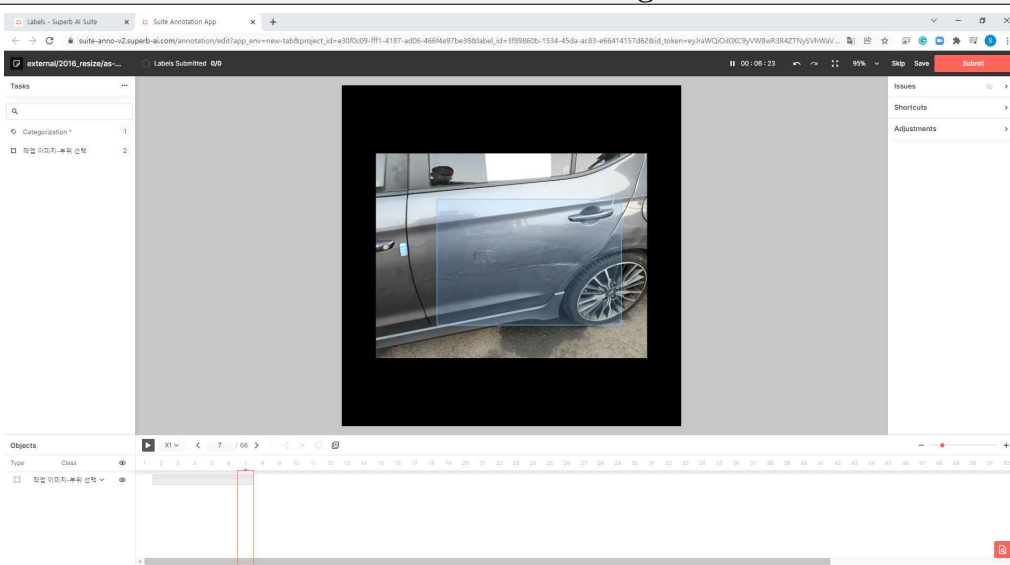
4. 정제 라벨링 진행

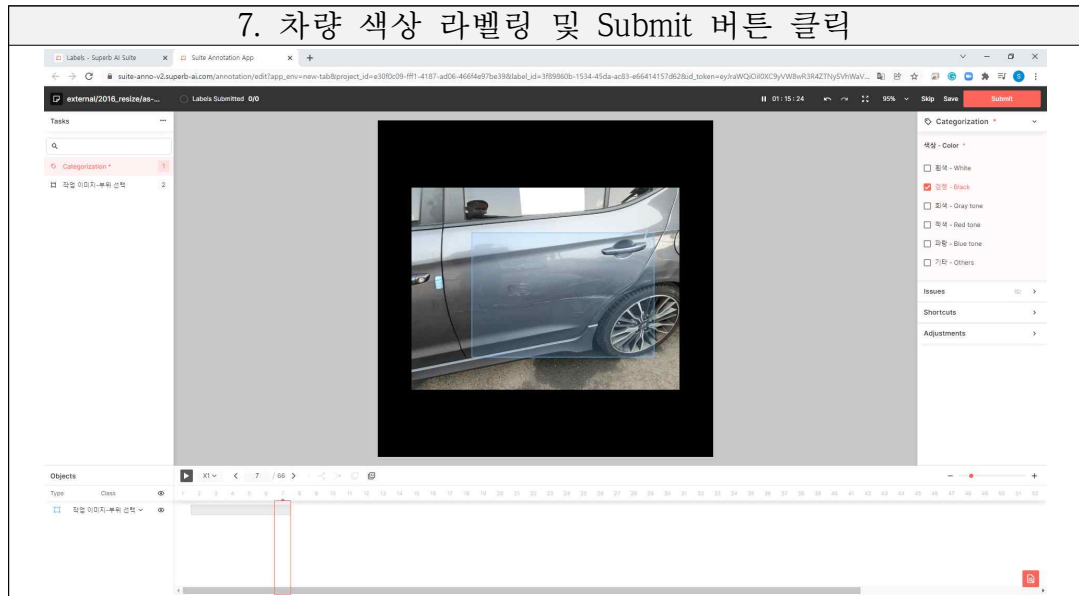


5. 한 데이터 내 모든 이미지 정제



6. 정제 라벨링 Merge





(그림) 사고 단위 데이터 정제 과정

3.3. 수집·정제 기준

○ 사고데이터 수집을 위한 기준

- 최종 제출 데이터 50만 이미지의 경우, 시장의 평균 수치/목표수치/최소 수치로 제시함. (수치별 +/- 1%의 예외 폭을 두고자 함)
- 차량연식의 경우, 최신데이터 ~ 2015년 연식까지를 중심으로 수집할 예정
 - 불가피하게 이전 차량을 수집할 경우, 전체의 10% 이내로 수집 (2012~2014년)

○ 수집시 고려사항

- 학습모델에 적합하지 않은 데이터는 수집에서 제외할 수 있도록 가이드 라인을 제시하여 적합한 데이터만 구매하는 것이 바람직함

수집시 고려사항	설명	비고
획득 가능성	<ul style="list-style-type: none"> 이미 축적된 정비소의 데이터를 활용하기 때문에 확보에는 문제가 없으나, 인공지능 데이터셋 구축을 상정하고 촬영한 것이 아니므로 획득 품질에 대한 가이드 라인을 작성하여 그 기준에 맞는 데이터를 획득해야 함 	가능성 검토

데이터 정확성	<ul style="list-style-type: none"> • 서비스의 활용목적에 세부 항목이 정확히 존재하는가 검토 필요 • 획득 목적에 맞는 데이터를 획득하기 위해서는 사전, 사후 처리 방안 필요 • 사업계획서 당시 정의한 구축요건에 맞춰 데이터를 획득, 정제함 • 그 밖의 사례를 참고하여 데이터 획득, 정제에 필요한 사항을 가이드에 반영함 	정확성 검토
보안사항 개인정보 및 저작권	<ul style="list-style-type: none"> • 데이터 획득시 저작권에 대한 라이선스 확보 및 가공된 데이터의 라이선스에 대한 협약 및 협약서 • 획득 데이터에 대한 차량정보/보험사 직원명, 정비소 정보는 비식별화 처리 	공개 데이터를 위한 필수 사항
데이터 균형	<ul style="list-style-type: none"> • 차종, 세그먼테이션, 손상유형, 원근거리별 데이터 균형을 고려하여 획득 	활용분야 검토
신뢰성	<ul style="list-style-type: none"> • 데이터 획득대상, 획득방법이 법, 제도를 저촉하거나 사회윤리에 어긋나지 않도록 해야 함 • 이미지 데이터 획득 시 획득된 데이터의 종류에 따라 검증을 진행하여 데이터에 대한 신뢰성 부여 • 획득된 이미지 데이터의 중복 여부를 검사하고 신뢰할 수 있는 데이터인지를 확인 	신뢰성 검토

○ 정제기준

- 수리가 진행 중인 차량의 경우 제외함
 - 차량을 비닐로 덮어놓거나, 테이프/케이블타이로 차량 수리작업을 진행한 경우
 - 수리를 위해 부품을 탈거(脫去)한 경우
 - 차량 내부를 촬영한 경우(* 차량 내부/부품 내부)
 - 차량 내부에서 외관을 촬영한 경우(* 차량 내부에서 앞 유리를 촬영한 경우)
 - 수리 도구가 차량에 맞닿은 경우(* 리프트는 라벨링 가능)
- *단 수리를 위한 전(前) 단계는 라벨링 함(차량 수리 전에 준비를 위해 차량에 마킹한 경우)
- 실사 차량 이미지가 아닌 경우 등 제외함
 - 휴대전화 이미지를 캡처한 경우
 - 모니터 화면을 촬영한 경우
 - 후(後)처리된 이미지의 경우(* 동그라미 등. 단 카메라나 휴대폰에서 입력되는 촬영정보나 워터마크는 라벨링 가능 이미지)
 - 흔들림이 심해 객체 또는 객체의 외각을 구분하기 힘든 경우
 - 트레일러/트럭에 승차되어 있는 경우

- 개인정보 노출 등 아래의 경우 제외
 - 식별 가능한 안면이 비치는 경우(* 자동차 및 거울)
 - 개인정보가 보이는 경우(* 이름, 특정 지을 수 있는 주소, 전화번호 등)
 - 회사 정보가 보이는 경우(* 기업 번호 및 특정 지을 수 있는 주소 등. 상호명은 라벨링 가능)
 - 흔들림이 심해 객체 또는 객체의 외각을 구분하기 힘든 경우
 - 트레일러/트럭에 승차되어 있는 경우
- 차량번호판이 부착된 이미지에 대해서는 번호판 부위를 폴리곤 영역으로 라벨링하여 비식별화 처리



(그림) 정제 가이드라인 예시



(그림) 정제를 통하여 제외되는 이미지 예시

○ 정제시 고려사항

대항목	고려사항	비고
정제 기준의 명확성	이미지 데이터내의 손상부위를 명확히 구분할 수 있도록 이미지 내에서 차량의 비중, 손상부위의 비중 등에 대해 정확한 기준을 수립하고 있는지 여부	기준 검토
중복성 방지	작업자별, 폴더별, 이미지별 중복성 검사	중복이미지 정제
정제작업 매뉴얼	유형별 사례별 정제 기준 및 가이드라인 작성	작업 방법 정밀성
정제 도구	정제 작업에 사용될 SW 도구를 확보 및 사용 방법 숙지 여부	활용 툴
정제작업 방식	차량 손상 데이터 특성 및 수리비 산정에 맞는 적절한 정제방식 선정 여부 및 선정 기준 타당성 여부	작업 방식의 적절성
수리비 추출	수리비 견적서에서 가장 기준이 되는 항목 추출	핵심 수리방법 도출

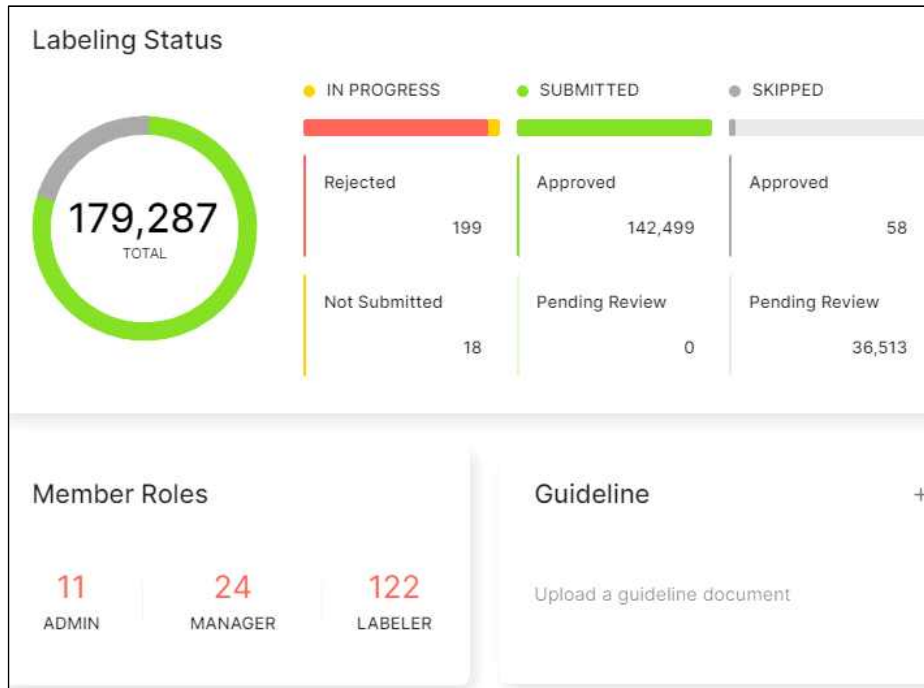
3.4. 수집·정제 도구

○ 도구명: Superb AI Suite

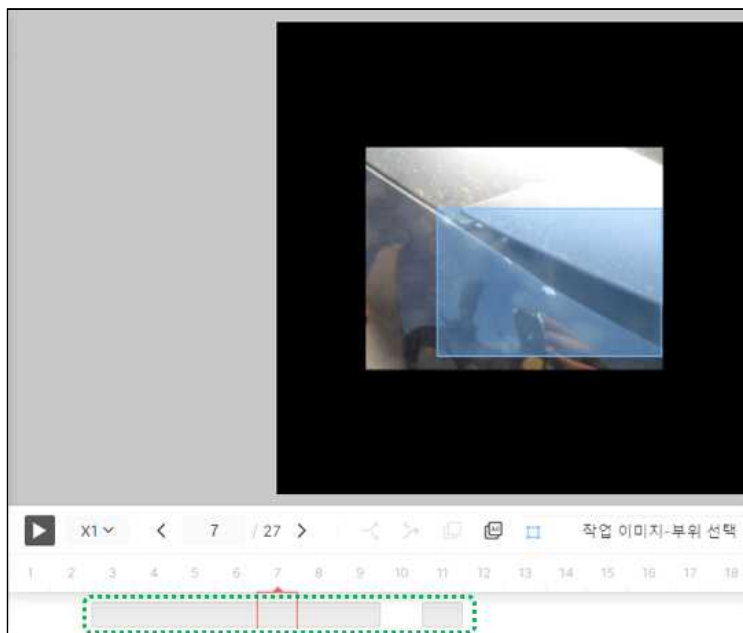
○ 기능

- 이미지 선별 기능
- 작업 요구 부분에 바운딩 박스 지정 기능
- 이미지의 원하는 위치에 검수자의 검수의견 지정 기능
- 정제자의 의견 지정 기능
- 검수자의 ‘Approve’, ‘Reject’ 지정 기능

- 각 현황별 통계 기능



- 위의 통계 현황에 대해 유형별, 조건별 조회 기능
 - 정제자별 작업 현황 조회
 - 검수자별 검수 현황 조회
- 정제를 통하여 선정 또는 비선정된 사례
- 바운딩 박스를 이용하여 이미지를 선정한 사례
 - 예: 아래 사례는 27개의 이미지중 초록색 점선안에 회색으로 표시된 이미지는 가공할 이미지로 선정된 것임(현재는 7번 이미지)



- 이미지를 선택하지 않고 제외한 사례(가공 단계에 사용되지 않음)

- 예: 아래 사례는 하나의 사고차량을 촬영한 36개의 이미지 중 초록색 점선안에 회색으로 표시된 것이 선정된 이미지임
- 본 예시에서는 15번 프레임이 선택되지 않았음: 페인트 칠(타원 표시)을 해 놓은 상태로서 수리중인 이미지라서 제외 대상임



4. 데이터 가공

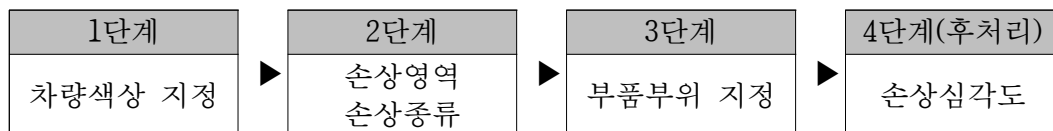
4.1. 라벨링 절차 및 방법

- 사고 차량 이미지 및 수리내역에 대한 정제과정을 통해 산출된 원천 데이터에 대해 부품부위, 손상종류를 제시하기 위해 최적화된 항목을 도출하고 이를 라벨링에 활용할 수 있도록 방법과 절차를 제시함

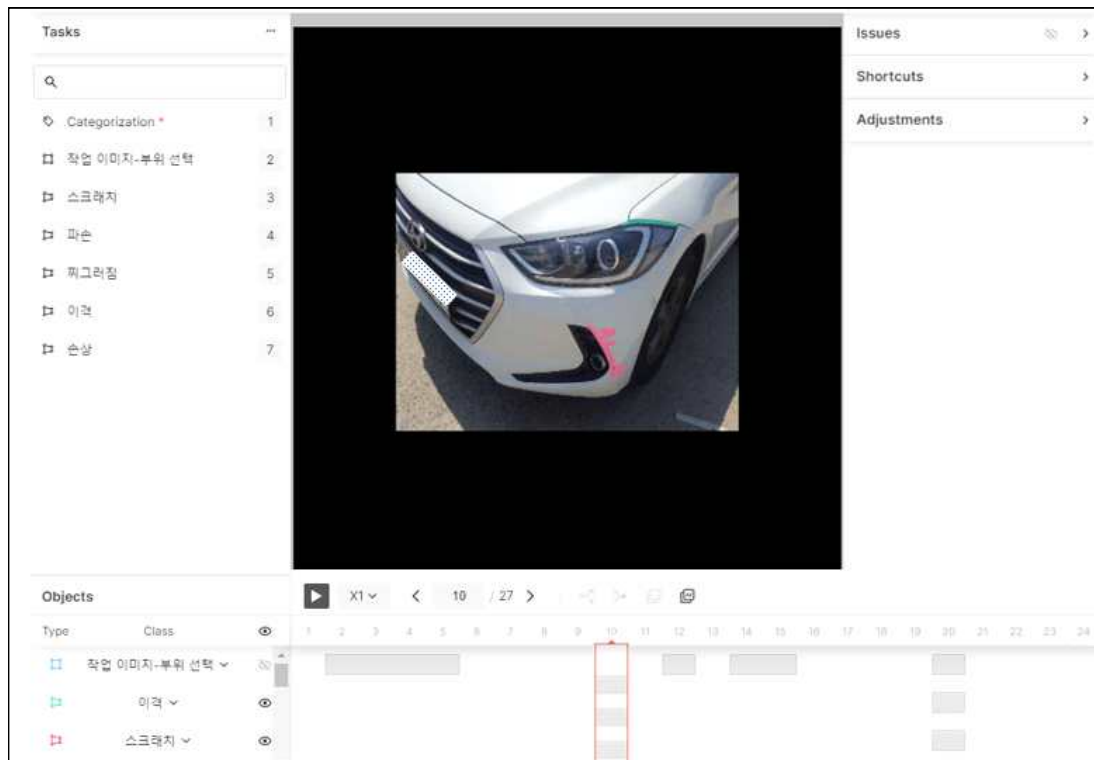
구분	주요내용	비고
데이터 메타정보 입력	<ul style="list-style-type: none"> 원시데이터에서 입수한 메타정보를 기계적으로 자동 입력 - 수집 장소, 일시 등의 기타 정보 	Python을 통한 입력
데이터 구축 (AI자동화 및 사람라벨링 작업)	<ul style="list-style-type: none"> 1차 가공 (전문가를 통한 정교한 GT 구축) - 최소 6개월 이상 근무자 투입 	AI가공을 위한 소수 인력의 1차 가공
	<ul style="list-style-type: none"> 2차 가공 - Custom AI 생성 및 AI Labeling - 구축된 데이터에 특화된 AI 개발/활용 	단순한 작업의 경우, AI로 자동화 시킴
	<ul style="list-style-type: none"> 3차 가공 - 클라우드소싱 - AI Labeling 결과 중 오류사항 수정 작업 	클라우드소싱 가공
	<ul style="list-style-type: none"> 4차 가공(1차 검수) - 전문가 검수 - 전문가의 1차 검수/수정 	전문가 검수 (가공 유 경험자)

- 라벨링 진행방법 / 평균 예상 객체 (Image 당 최소 2개~최대 10개 작업)

- 라벨러는 Image에 표시된 손상 부위를 Segmentation 작업 진행
 - 손상된 차량부위의 Class 분류 작업 진행
- 손상된 차량 부위 Segmentation 작업 진행(전체 부위가 보일 경우 시행)
- 손상 Class에 대한 손상심각도는 손상면적을 이용하여 프로그램으로 후처리
- 라벨링 진행 단계 및 도구를 이용한 작업 단계



- 라벨링(폴리곤 작업) 화면 예시



4.2. 라벨링 기준

○ 기본사항

- 사고차량에 대하여 손상된 모든 종류에 대하여 모두 폴리곤 작업을 수행함
- 즉 하나의 이미지 내에 스크래치, 찌그러짐, 이격, 파손이 존재 할 경우 이들에 대해 각각, 폴리곤 작업을 수행함,
- 스크래치, 찌그러짐, 이격, 파손 부위가 중첩될 경우 각각의 손상에 대해 폴리곤 작업을 수행함, 즉 각 손상종류에 대한 영역이 중첩도리 수 있음
- 동일 종류의 손상이 흩어져 나타날 때는 그루핑하여 작업한다.(단, 너무 멀리 떨어져 있지 않는 경우)

○ 어노테이션 카테고리

▪ 차량 종류 정의

소분류	정의	예시
경형	전장 3.6, 전폭 1.6, 전고 2.0 이하, 배기량 1,000cc 이하	 <p>기아 올뉴모닝</p>
소형	전장 4.7m, 전폭 1.7m, 전고 2.0m 이내, 엔진 1,600cc 미만	 <p>QM3</p>
중형	전장, 전폭, 전고 중 소형차 기준을 하나라도 초과, 엔진 1,600cc 이상 2,000cc 미만	 <p>현대 아반떼MD(13)</p>
대형	전장, 전폭, 전고 중 소형차 기준을 하나라도 초과, 엔진 2,000cc 이상	 <p>기아 오피러스(06)</p>

■ 차량 부위 명칭

(각 차체의 구성 부위 별 class 번호 및 부위 명칭 매칭 기준)

□ 하체 38

● 주의사항
29, 30, 31, 32: 타이어
33, 34, 35, 36: 휠

	부위 이름		부위 이름		부위 이름		부위 이름		부위 이름
1	Front bumper	9	Rear lamp(R)	17	Rear door(L)	25	C pillar(L)	37	Rear windshield
2	Head Lights(L)	10	Rear bumper	18	Rear door(R)	26	C pillar(R)	38	Undercarriage
3	Head lights(R)	11	Front fender(L)	19	Rocker panel(L)	27	Rear fender(L)		
4	Bonnet	12	Front fender(R)	20	Rocker panel(R)	28	Rear fender(R)		
5	Windshield	13	Side mirror(L)	21	A pillar(L)	29,33	Front Wheel(L)		
6	Roof	14	Side mirror(R)	22	A pillar(R)	30,34	Front Wheel(R)		
7	Trunk lid	15	Front door(L)	23	B pillar(L)	31,35	Rear Wheel(L)		
8	Rear lamp(L)	16	Front door(R)	24	B pillar(R)	32,36	Rear Wheel(R)		

- 차량 부위 명칭 표준화 내역




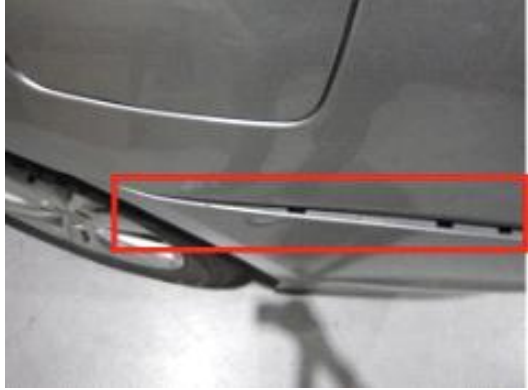
부위 표준명	견적서 내 부위명칭
a필러	앞윈도우필러, 프런트필러, 프론트필러, 후론트필러, 후런트필러
a필러(우)	(우)프론트필러, 우프론트필러, 우후론트필러, 프론트필러(우), 후론트A필러(우), 후론트윈도우필러(우), 후론트필러(외측우), 후론트필러(우), 후론트필러(우외측),후론트필러우
a필러(좌)	(좌)프론트필러, 윈도우필러(앞측좌), 윈도우필러(좌)후론트필러, 좌프론트필러, 좌후론트필러, 프론트윈도우필러(좌우), 프론트필러(좌), 프론트필러좌, 후론트윈도우필러(앞좌), 후론트필러(외측좌), 후론트필러(좌), 후론트필러(좌)상단, 후론트필러(좌)윈도우필러, 후론트필러(좌)윈도필러, 후론트필러(좌우), 후론트필러좌
b필러	B필러, 센터필러, 센타릴러

b필러(우)	(우)센터필러, 센타필러(우), 센타필러외측우, 센타필러우, 센타필러우외측, 센타필러패널우, 센터필러(실외우), 센터필러(아우터)(우), 센터필러(아우터)우, 센터필러(아우터우), 센터필러(외측우), 센터필러(우), 센터필러(우)전면, 센터필러외측중간(우), 센터필러우측, 우센타필러, 프론센타필러우
b필러(좌)	(좌)센터필러, 센타필러(아우터)(좌), 센타필러(외측좌), 센타필러(운), 센타필러좌, 센터필러(아우터)(좌), 센터필러(외측)(좌), 센터필러(외측)좌, 센터필러(외측좌), 센터필러좌, 센터필러좌측
c필러(우)	뒤필러패널(외측우), 뒤필러패널(우), 리어필러(우)
c필러(좌)	뒤필러인너패널(좌), 뒤필러패널좌, 리어C필러(좌), 리어필러(좌), 리어필러좌, 리어필러패널(좌), 리어필러패널좌, 좌측뒤필러
뒷도어	도어(상뒤), 도어(하뒤), 뒤도어, 뒤문, 뒷도어, 뒷문, 뒷문(Backoor), 리어도어, 백도어(리어)
뒷도어(우)	(우)뒤도어, 도어(뒤우), 도어(우뒤상), 도어(우뒤하), 도어뒤우, 뒤도어(뒤우), 뒤도어(우), 뒤도어(우상), 뒤도어(우전동), 뒤도어(우하), 뒤도어우, 뒤도어우측, 뒤문(우), 뒤우도어, 뒷도어(우), 뒷도어우, 뒷문(우), 리어도어(도어)(우), 리어도어(도어우), 리어도어(로워)(우), 리어도어(로워)우, 리어도어(상)우, 리어도어(어퍼)(우), 리어도어(어퍼)우, 리어도어(우), 리어도어(우뒤), 리어도어(우상), 리어도어(우전동), 리어도어(우하), 리어도어(측우), 리어도어(하)우, 리어도어어퍼(우), 리어도어우, 리어도어중간(우), 우뒷도어, 우리어도어, 우측뒤도어, 조수석뒷도어
뒷도어(좌)	(좌)뒤도어, 도어(뒤좌), 도어(뒤좌수동문), 도어뒤좌, 뒤도어(좌), 뒤도어좌, 뒤도어좌(상), 뒤도어좌(하), 뒤문(좌), 뒤좌도어, 뒷문(좌), 리어도어(도어좌), 리어도어(좌), 리어도어(좌)전면, 리어도어(좌)중간, 리어도어(좌리어), 리어도어(좌하), 리어도어뒤좌, 리어도어로워(좌), 리어도어외(좌_뒤), 리어도어좌후, 리어도어프런트(좌), 운전석뒷도어, 프런트도어리어(좌), 후도어뒤좌
뒷범퍼	뒤범퍼, 뒷범퍼, 리어(후범퍼), 리어로워범퍼, 리어범퍼, 리어범퍼(우), 리어범퍼(좌), 리어범퍼(좌)(우), 리어범퍼(좌)우, 범퍼(뒤), 후범퍼
뒷유리	뒤유리, 뒷유리
뒷타이어(우)	뒤타이어(우)휠, 타이어&휠(뒤우), 타이어(뒤우), 타이어(뒤우)뒤휠, 타이어(뒤우)타이어, 타이어(뒤우)휠, 타이어(리어)(우), 타이어(리어우), 타이어휠(뒤우), 타이어휠(리어우), 휠&타이어(뒤우), 휠&타이어(리어우), 휠(뒤우)타이어, 휠타이어(뒤우)
뒷타이어(좌)	타이어(뒤좌), 타이어(뒤좌)뒤휠, 타이어(뒤좌)휠, 타이어(리어)(좌), 타이어(리어)(좌)휠, 타이어(리어)좌)휠, 타이어(리어좌), 타이어뒤(좌), 타이어뒤좌, 타이어휠(리어)(좌), 타이어휠(리어)좌), 휠&타이어(뒤좌), 휠타이어(뒤좌), 휠타이어(리어)(좌)
뒷펜더	뒤휰다, 리어휰더, 리어휰다, 리어휰다(뒤), 리어휰더
뒷펜더(우)	(우)뒤휰다, (우)뒤휰더, 뒤휰다(우), 뒤휰다(우)휰, 뒤휰다(우측), 뒤휰다우, 뒤휰다우측, 뒤휰더(우), 뒤휰더), 뒤휰더우, 뒤휰(외측)(우), 리어휰더(우), 리어휰다(상단우), 리어휰다(우), 리어휰다(우상단), 리어휰다리어우측, 리어휰다우, 리어휰더(우), 우뒤휰다, 우리어휰다, 우측리어휰다, 조뒤휰더, 휰다(뒤우), 휰다(우뒤)

뒷펜더(좌)	뒤웬다(좌), 뒤웬다좌, 뒤웬다좌측, 뒤웬더(좌), 뒤웬더좌, 리어좌웬다, 리어웬더(좌), 리어웬더(좌상단), 리어웬더사이드(좌), 리어웬다(좌), 리어웬다간이실(좌), 리어웬다간이실(좌)(우), 리어웬다사이드(좌), 리어웬다좌, 리어웬다좌우, 리어웬다휠(좌), 리어웬더(좌), 리어휠(리어)(우), 리어휠(프런트)(우), 운전석뒤웬다, 운전석뒷웬다, 웬다(뒤좌), 웬다(좌뒤)
루프	루프, 루프(뒤), 루프(우), 루프(전체), 루프(좌), 루프(좌우), 루프(지붕), 루프랙, 루프랙우, 루프랙좌, 루프랙(우), 루프랙(좌), 루프랙좌
범퍼	범퍼
본넷	본넷, 본닛, 본네트
사이드미러	사이드미러
사이드미러(우)	사이드미러(R), 사이드미러(우), 우사이드미러, 사이드미러우
사이드미러(좌)	사이드미러(L), 사이드미러(우), 좌사이드미러, 사이드미러좌
스텝	보조발판, 사이드발판, 사이드스텝발판, 사이드스테프발판, 사이드스텝발판, 사이드스텝, 사이드스텝(발판형), 스텝, 스텝(발판), 스텝(발판)좌우
스텝(우)	발판스텝(우), 사이드(우)스텝, 사이드발판(우), 사이드발판스텝(우), 사이드스텝발판(우), 사이드스텝(뒤우), 사이드스텝(발판)(우), 사이드스텝(발판)우, 사이드스텝(발판우), 사이드스텝(외측우), 사이드스텝(우, 사이드스텝(우), 사이드스텝(우)발판, 사이드스텝(패널)우, 사이드스텝우
스텝(좌)	발판스텝(좌), 사이드발판(좌), 사이드발판스텝(좌), 사이드스텝발판(좌), 사이드스텝(리어좌), 사이드스텝(발판좌), 사이드스텝(앞좌), 사이드스텝(외측)(좌), 사이드스텝(외측)좌, 사이드스텝(외측좌), 사이드스텝(좌), 사이드스텝(좌)발판, 사이드스텝(좌발판), 사이드스텝(좌스텝분리형), 사이드스텝(패널)좌
앞도어	도어(앞), 앞문, 프론트도어, 후론트도어
앞도어(우)	도어(앞도어우), 도어(앞우), 도어(우앞뒤), 도어(우앞상), 도어(우앞하), 도어앞우, 도어외(앞우), 도어외(우)앞뒤, 앞도어(앞우), 앞도어(우), 앞도어(우상), 앞도어(우수동), 앞도어(우측), 앞도어(우하), 앞도어앞우, 앞도어우, 앞도어우측, 앞도어필러(우), 앞문(우), 앞우뒤도어, 앞우전도어, 앞전도어(우), 조수석앞도어, 프론트도어(우), 프론트도어(우앞), 프론트도어우, 후론트도어우)
앞도어(좌)	(좌)앞도어, 도어(앞)좌, 도어(앞뒤)좌, 도어(앞좌), 도어앞좌, 앞도어(상)좌, 앞도어(좌), 앞도어좌(하), 앞도어좌측, 앞뒤도어(좌), 앞문좌, 운전석후론트도어, 프론트도어(좌), 프론트도어좌, 후론트도어(좌), 후론트도어(좌)), 후론트도어좌, 후론트도어중간(앞좌), 후론트도어중간(좌)
앞범퍼	범퍼(앞우), 범퍼(앞좌), 앞범퍼, 앞범퍼(우), 프런트범퍼, 프런트범퍼, 프런트범퍼(상단), 프런트범퍼(좌), 프런트범퍼(좌우), 프런트범퍼(하), 프런트범퍼(하부), 후론트범퍼
앞유리	앞유리
앞타이어(우)	(타이어)앞우, 앞타이어(우), 타이어&휠(스틸앞우), 타이어&휠(앞우), 타이어(앞우), 타이어휠(앞우), 휠(앞우)타이어, 휠타이어(앞우)

앞타이어(좌)	타이어&휠(앞좌), 타이어(앞좌), 타이어앞좌, 타이어휠(앞좌), 프론트(타이어형)좌, 휠&타이어(앞좌), 휠타이어(앞좌)
앞펜더	앞펜더, 앞펜다, 앞휠더, 앞휠다, 프론트펜더, 프론트휠다, 후론트휠다, 후론트휠더
앞펜더(우)	(우)앞휠다, 앞뒤휠더(우), 앞우전휠다, 앞우휠다, 앞휠다(우), 앞휠다우, 앞휠다우), 앞휠다우측, 앞휠더(앞우), 앞휠더(우), 우앞휠다, 우측앞휠다, 프론트펜다(우), 프론트펜더(우), 프론트휠다(우), 프론트휠다우, 프론트휠더(우), 후론트우휠다, 휠다(앞우), 휠다(우앞), 휠다패널앞우
앞펜더(좌)	(좌)앞휠다, 리어휠다(좌)앞, 앞뒤휠더(좌), 앞유리(좌우), 앞좌우휠다, 앞휠다(좌, 앞휠다(좌), 앞휠다(좌)), 앞휠다(좌)휠더, 앞휠다(좌우), 앞휠다(휠좌우), 앞휠다램프(좌), 앞휠다좌, 앞휠다좌(휠다), 앞휠다좌우, 앞휠다좌측, 앞휠다형(좌), 앞휠더(좌), 앞휠더(좌우), 좌앞휠다, 프론트펜더좌, 프론트휠다(좌), 프론트휠다좌, 프론트휠더(좌), 후론트전휠다(좌), 후론트좌휠다, 후론트휠다(좌), 휠다(앞좌)
전방휠	후론트휠, 휠(운전석앞), 휠앞
전방휠(우)	앞우전휠, 앞우측휠, 앞우휠, (우)앞휠, 우전프론트휠, 우측(앞뒤), 우프론트휠, 프론트휠우, 후론트휠(앞우), 후론트휠(우, 후론트휠(우), 휠(앞우), 휠앞우, 휠우앞
전방휠(좌)	앞(좌)휠, 앞좌휠, 앞휠(좌), 앞휠(좌우), 앞휠좌, 좌프론트휠, 좌프론트휠좌, 좌휠프론트, 프론트휠좌, 후론트휠(앞좌), 후론트휠(어퍼)좌, 후론트휠(좌), 후론트휠(좌앞), 휠(앞좌), 휠(좌앞), 휠앞(좌), 휠앞좌
전조등(우)	(우)헤드램프, 우헤드램프, 전조등(우), 헤드램프(우), 헤드램프우
전조등(좌)	(좌)헤드램프, 전조등(좌), 좌헤드램프, 헤드램프(좌), 헤드램프좌
트렁크	트렁크
하체	하체
후방휠	리어휠, 뒤휠, 휠뒤
후방휠(우)	(뒤우)휠, (우)뒤휠, 뒤우휠, 뒤휠우, 리어(우), 리어우휠, 리어휠(리어)우, 리어휠(외측우), 리어휠(우), 리어휠우, 우뒤휠, 휠(뒤우), 휠(리어)우, 휠(리어우), 휠뒤(우), 휠뒤우
후방휠(좌)	(뒤좌)휠, (좌)뒤휠, (좌뒤)휠, 뒤(좌), 뒤휠(좌), 뒤휠좌, 리어(좌), 리어좌휠, 리어휠(뒤좌), 리어휠(리어)좌, 리어휠(외측좌), 리어휠(인너)좌, 리어휠(좌), 리어휠(프런트)좌, 리어휠가아드(좌), 리어휠좌, 좌뒤휠, 휠(리어)좌, 휠(리어좌), 휠(좌뒤), 휠뒤(좌), 휠뒤좌, 휠리어(뒤좌)

■ 손상종류

손상 종류	설명	예시
Scratched (스크래치)	차량의 표피 성분의 변화 없이 일부의 굽 힘 상태 -> 사포로 문지른 것 과 같은 형태, 굽힌 형태	
Crushed (찌그러짐)	차량 외관의 철제상 결함이 발생한 수준 -> 움푹 들어간 부분	
Breakage (파손)	외부 충격 등으로 인 해 차량의 형태 일부, 혹은 전체가 기존의 형태와 완전 달라진 경우 -> 깨지거나 부서진 경우, 기존 형태를 가 지고 있지 않음	
Separated (이격)	차량의 각 부위가 외 부 충격으로 인해 어 긋남이 발생한 경우 -> 이음부만 떨어진 경우, 차량의 서로 다 른 부위 사이로 떨어 진 경우만 이격으로 처리	

- 차량 색상
 - 흰색/회색
 - 블랙
 - 블루계열
 - 레드계열
 - 기타

○ 라벨링 대상

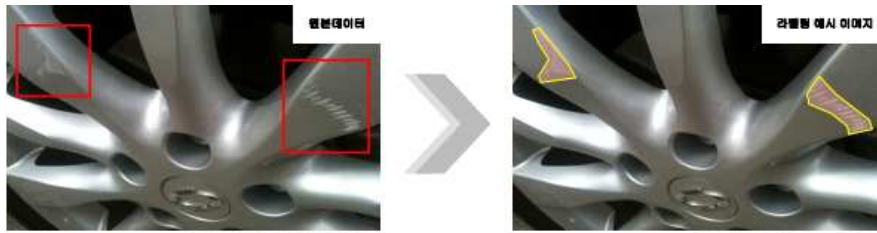
- 객체 클래스(손상부위, 손상된 차량부위)에 대해서 모두 어노테이션함
- 라벨링 제외 대상
 - 손상으로 인해 걸어서 드러나는 차량의 내부 영역(엔진부, 외판의 내부 영역 등)은 라벨링 하지 않음.
 - 거울에 비친 차량 부위는 라벨링 제외
 - 안에서 바깥으로 촬영된 유리는 제외
 - 차량에 글씨, 숫자가 적혀 있는 경우
 - 손가락이나 사람이 등장하는 이미지
 - 차량에 페인팅 작업을 해 놓은 이미지
 - 이미지에 그림을 그려 놓은 경우
 - 차량을 비닐로 덮어놓았거나, 테이프 등으로 붙여 놓은 이미지



(그림) 라벨링 제외 이미지 예시

○ 폴리곤 작업하기

- 손상부위를 포함하는 최소의 영역으로 다각형을 그림
- 단, 다음과 같이 동일 손상 종류가 완전히 분리되어 있는 경우와 주변에 묻쳐있는 경우 다음과 같이 작업한다.
 - 완전히 분리되어 있는 경우: 각각의 영역으로 작업



- 동일손상이 그룹지어 분포되어 있는 경우, 하나의 영역으로 작업, 그 외는 분리 작업



- 유사 영역에 2가지 이상의 손상이 발생한 경우 각각 작업함: 스크래치(주황), 찌그러짐(노랑)
- 아래의 좌측 이미지와 같이 손상이 발생했을 경우 우측 이미지와 같이 작업함



- 손상부위 라벨링 및 ground truth 예시

사고데이터	라벨링 데이터	AI 모델링 결과
		

○ 후처리를 통한 차량정보 적용(대상 항목: 차량크기, 차량연식, 손상심각도)

- 차량크기, 차량연식에 대해 수리내역서에서 수집한 정보를 이용하여 annotation의 해당 란에 삽입
- 다음과 같은 기준을 적용하여 손상심각도 계산 및 annotation의 해당 란에 삽입
 - 계산식: $\text{손상 면적(폴리곤크기)} / \text{차량부위 면적(폴리곤크기)} * 100(\%) * \text{손상 종류 가중치}(1\sim3) / 3$
 - 손상 종류별 가중치: 스크래치(1.0), 이격(1.2), 찌그러짐(2.0), 파손(3.0)
 - 손상심각도 (4)등급 범위: 4등급[0,10], 3등급[11,40], 2등급[41,80], 1등급[80,100]

○ 비식별화 작업

- 후처리를 통한 차량 번호판 비식별화
- 가공단계에서 차량번호판에 대해 폴리곤 작업한 부분에 대해 후처리 단계에서 번호판에 Blurring 처리를 통하여 비식별화 실현

○ 손상이미지 라벨링 시의 난제

- 손상 판단이 어려운 경우 발생
 - 차량에 나타난 손상이 얼룩인지, 비치는 모습인지, 먼지 등의 불순물이 묻은 것인지 구분이 어려운 경우가 많음
- 손상 유형의 모호성
 - 차량에 발생한 손상의 종류를 판정하기 어려운 경우 존재
 - 예: 스크래치인지 파손인지 구분이 어려운 경우
- 손상 영역 지정의 모호성
 - 차량이 파손되어, 부위가 절단되어 떨어져 나간 경우 잘려진 부분 또는 내부가 함께 보이는 경우 범위 지정의 어려움(사례가 다양하여 획일적 지정이 어려움)

4.3. 라벨링 규격

○ 라벨링 요소 정의

No	항목		길이	타입	필수여부
	한글명	영문명			
1	데이터셋정보	info		JsonObject	Y
1~1	데이터셋명	name	숫자,문자,특수문자 100 자리 이내	String	Y

1~2	데이터셋생성일자	date_created	MM/DD/YYYY	String	Y
2	이미지정보	images		JsonObject	Y
2~1	이미지식별자	id	1	Integer	Y
2~2	이미지너비	width	800~10000	Integer	Y
2~3	이미지높이	height	600~10000	Integer	Y
2~4	이미지파일명	file_name	숫자,문자,특수문자 100 자리 이내	String	Y
3	라벨링정보	annotations		JsonObject	Y
3~1	라벨링식별자	id	숫자 100 자리 이내	Integer	Y
3~2	연관이미지식별자	image_id	1	Integer	Y
3~3	사고 ID	category_id	숫자,문자,특수문자 100 자리 이내	String	Y
3~4	라벨링폴리곤	segmentation	[X,Y]	List	Y
3~5	라벨링폴리곤크기	area	0.00 ~	float	Y
3~6	라벨링바운딩박스	bbox	[X,Y]	List	Y
3~7	손상종류	damage	[스크래치, 찌그러짐, 파손, 이격]	String	
3~8	차량부위	part	숫자, 문자 100 자리 이내	String	
3~9	차량색상	color	숫자, 문자 100 자리 이내	String	Y
3~10	손상 심각도	level	[1, 2, 3, 4]	Integer	
3~11	차량 연식	year	[2012 ~ 2021]	Integer	Y
3~12	수리방법	repair	[부위명:수리방법,..]	List	N
4	카테고리 정보	categories		Object	Y
4~1	사고 ID	id	숫자,문자,특수문자 100 자리 이내	String	Y
4~2	차량 크기	supercategory_name	숫자,문자,특수문자 100 자리 이내	String	Y

■ 수리방법 적용을 위한 구조

- repair: [부위명:수리방법1, 수리방법2..]

수리부위 명칭	수리방법
a필러(좌), a필러(우), b필러(좌), b필러(우), c필러(좌), c필러(우), 전조등(좌), 전조등(우), 뒷도어(좌), 뒷도어(우), 뒷범퍼, 뒷유리, 후방휠(좌), 후방휠(우), 뒷펜더(좌), 뒷펜더(우), 루프, 본넷, 사이드미러(좌), 사이드미러(우), 스텝(좌), 스텝(우), 앞도어(좌), 앞도어(우), 앞범퍼, 앞유리, 전방휠(좌), 전방휠(우), 앞펜더(좌), 앞펜더(우), 트렁크, 후미등(좌), 후미등(우), 하체	도장, 판금, 교환, 수리
'A pillar(R)', 'A pillar(L)', 'B pillar(R)', 'B pillar(L)', 'C pillar(R)', 'C pillar(L)', 'Head lights(R)', 'Head lights(L)', 'Rear door', 'Rear door(R)', 'Rear door(L)', 'Rear bumper', 'Rear windshield', 'Rear Wheel(R)', 'Rear Wheel(L)', 'Rear fender(R)', 'Rear fender(L)', 'Roof', 'Bonnet', 'Side mirror',	coating, sheet_metal, exchange, repair

'Side mirror(R)', 'Side mirror(L)', 'Rocker panel(R)', 'Rocker panel(L)', 'Front door(R)', 'Front door(L)', 'Front bumper', 'Windshield', 'Front Wheel', 'Front Wheel(R)', 'Front Wheel(L)', 'Front fender(R)', 'Front fender(L)', 'Trunk lid', 'Rear lamp(R)', 'Rear lamp(L)', 'Undercarriage'	
---	--

* 견적서의 수리부위 명칭에 (앞, 뒤, 좌, 우) 가 없는 경우 (앞, 뒤, 좌, 우) 없이 기록함

○ 이미지 파일명 규칙

- 숫자, 문자, 특수문자 100자리 이내.jpg (이미지ID_사고ID.jpg),
- 숫자, 문자, 특수문자 100자리 이내.json (이미지ID_사고ID.json)

데이터종류	손상부위 유형	파일명
1.원천데이터	damage	20180329135555_1267487_sc-112145.jpg
	damage_part	20180329135555_1267488_sc-112145.jpg
2.라벨링데이터	damage	20180329135555_1267487_sc-112145.json
	damage_part	20180329135555_1267488_sc-112145.json

4.4. 라벨링 조직

○ 라벨러 교육 현황

구분	주요 내용	비고
클라우드워커 (신입)	<ul style="list-style-type: none"> - AI와 데이터 산업에 대한 이해 - 데이터 라벨링에 대한 이해 - 데이터 라벨링 방법론 - 데이터 라벨링 소프트웨어 학습 	교육 수강 후, Test 시행 (교육/Test 급여 제공)
클라우드워커 (1~3개월)	<ul style="list-style-type: none"> - 차량 스크래치 데이터 중 초기 데이터 작업 - 틀린 부분에 대한 수정 작업 진행 	클라우드워커가 원하는 형태로 협의 후, 급여 제공
검수 (3~5개월 Leader)	<ul style="list-style-type: none"> - 데이터 품질 관리 수행 1. 반려 데이터에 대한 피드백 2. 클라우드워커 재교육 - 라벨링 가이드 업데이트 	클라우드워커 대비 10% 인상된 급여 제공
PM(Project Manager, 6개월 이후)	<ul style="list-style-type: none"> - 데이터 프로젝트 부분 담당 1. 성과 기반의 추가 급여 제공 2. 과제 이후 채용 및 전문 프리랜서 전향 지원 - 슈퍼브에이아이의 주기적인 육성 및 프로젝트 담당 	클라우드워커의 희망에 따라 계약형태/업무 형태 논의 예정

4.5. 라벨링 도구

○ 데이터 라벨링 도구 선정

- ‘데이터 프로젝트 관리/ 품질 확보/ 라벨링 생산성/ 클라우드소싱/ 안정성-보안’을 고려하여 도구 선정

대분류	소분류	주요 내용	본 프로젝트 중요도
프로젝트 관리	데이터 관리기능	<ul style="list-style-type: none"> • 실시간 데이터 구축 현황 확인 기능 • 데이터 자동 분배 기능 	상
	데이터 분석기능	<ul style="list-style-type: none"> • 실시간 데이터 통계 기능 • 실시간 label 통계 기능 	중
품질 관리	데이터 피드백 기능	<ul style="list-style-type: none"> • 데이터 품질 향상을 위한 관리 기능 • 데이터 품질 검사/피드백 기능 	상(품질 관리)
	데이터 검수 기능	<ul style="list-style-type: none"> • 데이터 검수 결과 기입 기능 • 데이터 검수 필터 기능 	상(품질 관리)
라벨링 생산성	라벨링 사용성(Work app)	<ul style="list-style-type: none"> • 사용자 친화적인 데이터 가공 기능 • 추가 프로퍼티 기록 가능 여부 	상
	AI 활용 가능성	<ul style="list-style-type: none"> • AI를 활용한 자동 라벨링 기능 보유 여부 	상(라벨링/검수 효율성)
클라우드소싱	클라우드소싱 활용가능성	<ul style="list-style-type: none"> • 클라우드소싱 라벨링 기능 보유 	상(일자리 창출)
	클라우드소싱 인력 생산성 관리	<ul style="list-style-type: none"> • 클라우드소싱 인력의 퍼포먼스 트레이킹 • 생산성 관리 평가 기능 보유 여부 	상(생산성/품질 관리)
안정성 /보안	데이터 보안	<ul style="list-style-type: none"> • SaaS/ Hybrid On-prem 가능성 	중
	서버 안정성	<ul style="list-style-type: none"> • 10,000이상의 라벨러 동원 가능성 • 1,000,000개 이상 데이터 업로드/운영 가능성 	상(운영 중단시, 작업 일정 차질 발생)

- 관련 기준에 따라 자체 개발한 스위트(Suite) 플랫폼을 데이터 구축 도구로 선정

○ 도구 특징(Superb AI Suite)

품질검사영역	구문 정확성 검사	의미 정확성 검사	학습모델 유효성 검사
도구명	Superb AI Suite 자체개발 RuleMaker	Superb AI Suite	자체개발 (Python & Tensorflow)
설명	구축된 데이터에 대한 통계 및 산출물에 대한 양식의 오류를 검사하는 소프트웨어	Computer Vision 분야에서 정의하고 있는 대표적인 품질 측정 지표 활용 - Classification Error - Low Recall - Low Precision - Low IOU	학습된 AI가 원시/원천데이터를 대상으로 Segmentation 결과값을 생성할 수 있는지 확인

도구유형	자체개발 / JAVA	SaaS	자체개발 / python
주요기능	<ul style="list-style-type: none"> 어노테이션 파일의 구조적 문제를 검사 어노테이션 파일 내 항목에 대한 구조적 문제를 검사 어노테이션 파일 내 항목별 형식 검사 검사대상 데이터 등록관리 검사기준관리 품질검사관리 품질검사이력관리 보고서 	<ul style="list-style-type: none"> 참값 오류 육안 검사 데이터 종류별에 해당하는 어노테이션과 참값 간의 중첩율을 계산하여 정확도 확인 데이터 종류별로 생성된 어노테이션 중 정답 어노테이션 비율을 계산하여 정밀도 확인 데이터 종류별로 가능한 모든 어노테이션 중 정답 어노테이션 비율을 계산하여 재현율 확인 	<ul style="list-style-type: none"> 2개의 Json 파일의 결과를 비교하여 객체인식률을 측정 (AI Inference output과 Ground Truth의 IOU 비교)
사용환경	-OS:Linux server -H/W:Linux server -S/W: Web 기반 소프트웨어 (Chrome 최적화 솔루션)	Web 기반 소프트웨어 (Chrome 최적화 솔루션)	python

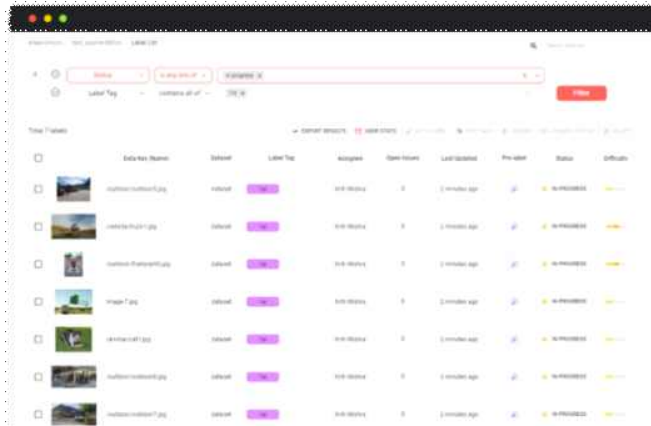
○ 도구 강점

- 본 과제 참여사인 (주)슈퍼브에이아이에서는 데이터 가공 및 검증, 관리에 최적화된 SaaS기반 머신러닝 데이터 플랫폼을 제공하여 고품질 데이터셋 구축작업을 수행함



(그림) 데이터 구축 플랫폼 SUITE 특징

■ Data 통합 관리/분석



Data Storage

- 고객사 Cloud Storage와의 Secure Link (AWS, MS, GCP, NCP 등)
- 고객사 IDC와의 Secure Link (On-Premise 형태)

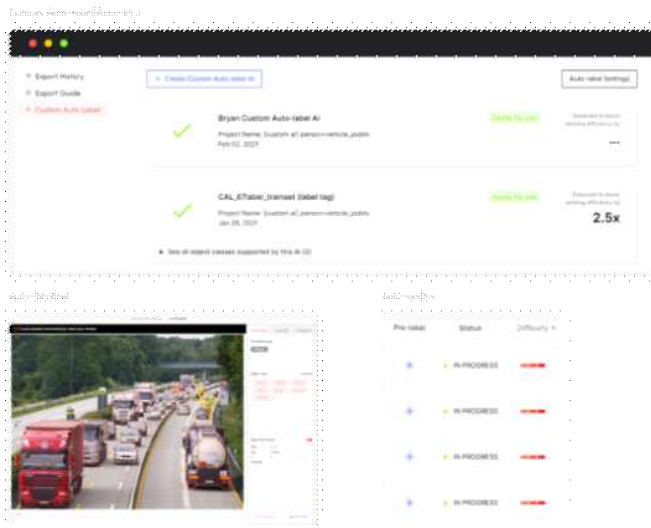
Label Upload

- 기존 구축된 json 파일 업로드 (AI Inference 결과 및 공개 데이터셋 업로드)

Filter/Search

- 10여종의 다중 필터 기능을 활용한 빠른 Label 검색 및 분류

■ AI기반 자동 라벨링



Custom Auto-Label (Auto-ML)

- 기존 라벨링된 데이터를 활용한 실시간 AI 생성 기능 제공
- Custom Auto Label AI의 효과성 예측 빠른 프로토타입을 세 개발 가능

Auto Labeling

- Common Object 110종에 대한 AI 자동 라벨링
COCO dataset 기법 라벨링 정확도 향상
mAP52, LowIOU 1%

Auto Review

- 정확도 보정(Accuracy AI) 기술을 활용한 고품질 데이터 생성 지원
- 불확실성 평가(Uncertainty Estimation) 기술을 활용한 데이터 검증 기준 제공
중요도 평가 5점 평가

5. 검수

5.1. 검수 절차 및 방법

○ 검수 프로세스



<4단계에 걸친 데이터 품질 검수>

- 1단계: 데이터 수집, 구축 기업인 쏘카는 자체 전수 검수를 거쳐 라벨링 기업인 슈퍼브에이아이에 데이터 전달
- 2단계: 슈퍼브에이아이 및 풀비는 수집데이터에 대한 전수 검수 및 라벨링 단계를 통하여 피드백 전달, 목표 성능 달성을 위해 반복적으로 검수 실시 (외부 공개 가능한 수준의 품질까지 도달 목표), 목표 구축량을 초과하였더라도 성능 개선에 필요시 추가 데이터 요청.
- 3단계: 자동차 수리 전문가와 함께 구성된 주관기관(한국전기차리빌딩협회) 검수단이 내부 최종 검수. 데이터 규모를 감안해 품질관리 전담기관인 풀비의 인력에 전문가 교육을 통하여 품질 검수 진행. (풀비는 품질관리에 대한 모든 단계를 모니터링 및 검수 수행)
- 4단계: 외부 검수 기관(TTA)이 중간, 최종 산출물에 대해 최종 검수

○ 수행 단계별 검사 기본항목

- 데이터 수집 단계

검사 항목	검사 내용
데이터 정확성	제시된 segment에 손상 부위가 제대로 촬영되었는가
이미지 품질	이미지의 크기는 800x600 이상인가
데이터 구성	수리내역서와 이미지가 pair를 이루고 있는가
데이터 균등성	수집된 데이터가 한쪽으로 편향되지 않았는가
데이터 라이선스	수집된 데이터의 라이선스는 해결되었는가 (초상권/재산권/저작권, 2차 저작권)

- 데이터 정제 단계

검사 항목	검사 내용
정제 기준의 명확성	수집된 데이터의 이슈는 없는가?(흔들림, 노이즈, 초점불량 등) - 데이터 사용 목적에 적합한 정제 기준 수립 여부 검수
중복성 방지	동일 데이터가 2장 존재하지 않는가? - 데이터 정제 후 정보 비교 등을 통한 중복도 여부 검수
정제 작업 매뉴얼	정제 작업의 프로세스를 절차대로 진행하였는가? - 정제 작업을 위한 매뉴얼 작성 및 관리 여부 검수
정제 도구	정제도구를 올바르게 사용하였고, 정확하게 정제를 수행하였는가? - 정제 작업에 사용될 소프트웨어(S/W) 도구를 확보 및 사용 방법을 숙지
정제 작업 방법	정제방법에서 프로세스(HOW)와 정제항목(What)을 정확히 수행하였는가? - 데이터 특성 및 활용 목적에 맞는 적절한 정제 방식 선정 여부 및 선정 기준 타당성 여부 검수

■ 데이터 라벨링 단계

검사 항목	검사 내용
라벨링 가이드라인	라벨링은 손상부위에 정확히 Polygon segmentation 처리 되었는가? 해당 라벨링의 오차범위는 15픽셀 이하인가? - 목적에 맞게 작성된 라벨링 가이드라인에 대한 타당성 여부 검사 후 라벨링 작업자들에게 해당 내용의 가이드라인 전달
어노테이션 항목	라벨링에 필요한 4단계 정보가 차례대로 기입되었는가? - 목적에 맞는 어노테이션 구성 여부 검수 후 확인된 내용을 포함하도록 작업자들에게 전달
가공 검수 도구	가공도구는 올바르게 사용되었고, 정확한 값을 입력하였는가? - 자동화 도구를 통해 검수 후 검수자가 육안으로 부적합 데이터 여부에 대해 2차 확인 및 조건 오류 전수 검수

■ 전수 검사 단계

검사 항목	검사 내용
부적합 판정 데이터 분포 확인	데이터 획득/정제/라벨링 단계의 항목을 모두 준수하였는가? - 목적에 맞게 작성된 라벨링 가이드라인에 대한 타당성 여부 검사 후 라벨링 작업자들에게 해당 내용의 가이드라인 전달
외부 검수자	데이터 획득/정제/라벨링 단계의 항목을 전부 준수하였는가? - 목적에 맞는 어노테이션 구성 여부 검수 후 확인된 내용을 포함하도록 작업자들에게 전달

○ 정제,가공에 대한 검수 방법

- 정제, 가공 검수를 위하여 Superb-AI Suite를 활용, 검수방법 동일
- 모든 검수자에 대해 각각 Label Tag를 부여하여 1, 2, 3차 검수를 진행함

<input type="checkbox"/>	Data Key (Name)	Dataset	Label Tag	Assignee	Last Updated ↓	Auto-Label	Reviews
<input type="checkbox"/>	 EDIT LABEL	EX2015	CS_물... move... divisio... Conve...	01_설화민	13 days ago		2 
<input type="checkbox"/>	 external/2015_re...	EX2015	CS_물... move... divisio... Conve...	01_설화민	13 days ago		2 
<input type="checkbox"/>	 external/2017_res...	EX2017	CS_물... move... divisio... Conve...	01_설화민	13 days ago		2 
<input type="checkbox"/>	 external/2015_re...	EX2015	CS_물... move... divisio... Conve...	EY_홍진경	13 days ago		2 
<input type="checkbox"/>	 external/2017_res...	EX2017	CS_물... move... divisio... Conve...	DL(B)_Woojin K...	14 days ago		2 






- 정제검수 -> 손상부위 검수 -> 차량부위 검수 순으로 검수 진행
- 검수자(작업자)는 데이터 가공 도구를 이용하여 라벨러의 작업물을 보며 도구의 'issue' 에 검수 내용을 기록함
- 가공 데이터에 대해 미태깅, 과태깅, 오타깅 등에 대해 의견 및 가공 제외 사유 등을 기술함
- 아래 그림과 같이 검수자가 검수에 대한 의견을 실제 이미지 화면의 해당 위치에 기록하여 라벨러가 쉽게 인지 할 수 있도록 함

	
해당 위치에 검수 의견이 기록되어 있는 사례	검수 의견 기록 사례

5.2. 검수 기준

○ 검수 기준

검수 기준	예시1
수리중 또는 수리완료된 프레임은 수집 제외(정제시 필터링 함)	
손상이 보이는 각도 이외의 다른 각도에 결함이 없는 프레임은 라벨링 대상 제외	
사람 얼굴이 나오거나 차량부위가 분리되어 있는 경우는 라벨링대상 제외	
결함부위가 있는 이미지 내에서 개인정보 노출 시 해당 영역은 번호판 클래스를 지정(비식별화 처리)	
차량부위 라벨링시, 결함을 포함한 차량부위만 라벨링하고, 결함이 없는 부위는 라벨링대상 아님	
차량의 부위 라벨링시, 해당 차량부위 전체가 보일 때에만 라벨링 하며, 조금이라도 잘려진 부분이 있다면 라벨링 하지 않음	

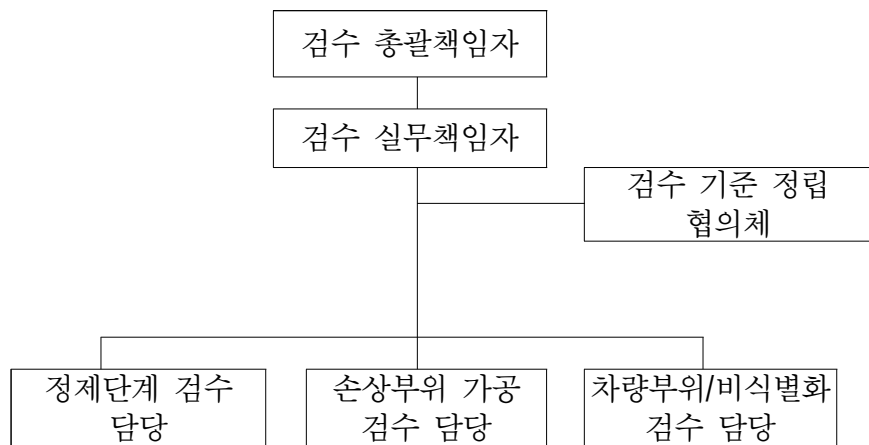
	
차량부위 및 손상종류를 정확하게 선택 지정해야 함	
손상영역을 너무 넓게 지정하면 안됨(15 pixel이내)	
결함 종류가 여러 종류일 경우, 모두에 대해 각각 라벨링 함 (예시 이미지는 찌그러짐과 이격이 동시에 발생한 경우임)	
스크래치에 대해 근접 부위는 그루핑하여 라벨링 함	

○ 오테깅, 미테깅, 과테깅 사례

오테깅	미테깅	과테깅
결함종류 또는 차량부위에 대한 라벨링 분류가 잘못된 경우	결함종류나 결함종류가 있는 차량부위를 라벨링하지 않은 경우	수리중이거나 해당 부위가 없는 경우에 라벨링하는 경우
		
결함종류가 ‘이격’ 인데 ‘스크래치’로 잘못 지정함	손상부위를 라벨링하지 않음	수리를 위하여 탈거한 부위를 라벨링함

5.3. 검수 조직

○ 검수 조직



조직 구분	역할과 책임
검수관리 총괄책임자	<ul style="list-style-type: none"> ◆ 학습용 구축 데이터 검수 관리 총괄 ◆ 구축과 검수 연계 업무에 대한 일정 조율 ◆ 검수 문제 발생시 개선 및 처리에 대한 의사 결정
검수관리 실무책임자	<ul style="list-style-type: none"> ◆ 인공지능 학습용 구축 데이터 검수 실무 총괄 ◆ 구축 단계에 따른 각 파트 별 일정 관리 ◆ 검수 관련 회의 주재 및 인력 관리
검수기준 정립 협의체	<ul style="list-style-type: none"> ◆ 다양한 유형의 레벨링 이슈 발생에 따른 결정 협의체 ◆ 라벨링 기준 정립 ◆ 정기적 검수 회의 진행

	<ul style="list-style-type: none"> ◆ 결정사항에 대해 상위 조직에 보고
정제단계 검수 담당	<ul style="list-style-type: none"> ◆ 정제를 위한 가이드라인 작성 ◆ 수집된 데이터에 대해 대상/비대상 분류 ◆ 정제 수행자에 대한 교육 시행 및 검수 ◆ 정제 검수도구 관리 및 검수 인력들에 대한 교육 진행
손상부위 가공 검수 담당	<ul style="list-style-type: none"> ◆ 손상부위 라벨링 가이드 기준 수립 ◆ 손상부위 라벨링 가이드 교육 및 검수 시행 ◆ 손상부위 검수 결과에 대해 해당 라벨러와의 신속한 피드백을 통한 보완/수정 유도 ◆ 손상부위 검수도구 관리 및 검수 인력들에 대한 교육 진행 ◆ 손상부위 가공/검수 관련 이슈 발생시 협의체에 상정
차량부위/비식별화 가공 검수 담당	<ul style="list-style-type: none"> ◆ 차량부위/차량번호 라벨링 가이드 기준 수립 ◆ 차량부위 라벨링 가이드 교육 및 검수 시행 ◆ 차량부위 검수 결과에 대해 해당 라벨러와의 신속한 피드백을 통한 보완/수정 유도 ◆ 차량부위 검수도구 관리 및 검수 인력들에 대한 교육 진행 ◆ 차량부위 가공/검수 관련 이슈 발생시 협의체에 상정

○ 실 데이터를 활용한 검수 교육 및 피드백을 통한 실시간 교육

- 각 이미지에 작업된 내용에 대해 정확한 ‘위치’에 ‘검수의견’을 기록하고, 이 의견은 실시간으로 라벨러에게 통지되어 라벨러로 하여금 쉽게 수정할 수 있도록 함으로써 동일 실수 발생을 줄임
- 검수자의 모니터링을 통하여 특정 라벨러의 작업 성향을 파악하고 통지함으로써 잘못된 작업 습관을 수정함
- 정기적/비정기적 라벨러 교육 이외에 교육 동영상을 제작하여 수시로 조회할 수 있도록 제공
- 라벨링 도구의 커뮤니티에 라벨링 가이드를 업로드 하되, 새로운 이슈사례 발생시 추가 결정된 가이드라인에 대해 지속적인 업데이트 시행 및 공지