

메타데이터 정보 (다중기입가능)	분야	데이터 유형 ¹⁾	구축 데이터량	원천데이터 형식 ²⁾	라벨링 형식 ³⁾	라벨링 유형 ⁴⁾
	교통물류	이미지	632,694	jpg	json	세그먼테이션
	데이터 출처 ⁵⁾	데이터 구축년도	구축기관(총괄)	가공기관	검수기관	
	쏘카, 자동차정비업 체	2021년	(사)한국전기차 리빌딩협회	(주)슈퍼브에이 아이	(사)한국자동차 검사정비사업 조합연합회 (사)경기도북부 자동차정비사 업조합	
	데이터 문의처	기관명	문의담당자명	전화번호 (유선전화번호기 입)	메일주소	
		(사)한국전기차 리빌딩협회	김현식	02-2219-5777	hskim@kevra.o r.kr	
	데이터 소개	자동차 정비소 및 쏘카에 축적되어 있는 사고 자동차 이미지를 수집하여 손상부 위, 차량부위, 손상종류, 손상정도에 대해 라벨링한 데이터 셋				
	주요키워드	차량 파손, 차량 손상종류, 차량 손상부위				
카테고리 정의서		(2-08)_차량파손이미지_카테고리정의서.xlsx 첨부의 카테고리 정의서 엑셀파일에 데이터카테고리 작성하여 제출(예시참고)				

1) 텍스트, 오디오, 이미지, 비디오,

2) txt, jpg,.....

3) json, csv,.....

4) 내용요약(텍스트), 번역(자연어), 질의응답(자연어), 바운딩박스(이미지/동영상), 키포인트(이미지/동영상), 세그먼테이션(이미지/동영상), 전자(음성)

5) 4대 언론기사, 자체 수집,.....

데이터셋명	국문영문	차량 파손 이미지 데이터																							
		Vehicle Damage Image Dataset																							
구축목적	자동차 사고시 발생하는 다양한 유형의 손상 이미지로부터 손상유형, 손상부위, 손상심각도를 도출해낼 수 있도록 인공지능을 훈련하기 위한 데이터 셋																								
활용서비스	보험사, 자동차 정비업소, 렌터카 업체, 부품업체 등에서 사고자동차 이미지를 이용하여 수리비를 산출하는 서비스																								
소개	자동차 정비소 및 렌터카 업체에 축적되어 있는 차량사고 이미지에 대해 정제, 가공을 통하여 AI 데이터 셋 구축하고 사고차량 견적을 위한 AI 데이터 셋을 기반으로 사고 견적을 도출할 수 있도록 하기 위하여 다양한 유형의 차량, 손상유형, 손상부위를 촬영한 원천데이터를 확보																								
	<div><p>가공 데이터 50만건 15만+15만+20만건 Delivery</p><p>사고차량 이미지 → 사고영역 폴리곤 → 검수 작업 → Annotation 생성 → 손상유형, 차량부위 탐지모델 → AI 학습을 통한 데이터 유효성 검증 및 피드백 → (loop back to 검수 작업)</p></div>																								
데이터셋 통계 (구축 규모 및 분포)	1. 데이터 구축 규모																								
	<table><tr><th colspan="2">데이터 종류</th><th>데이터 형태</th><th>어노테이션 규모</th><th>결과물 규모</th></tr><tr><td rowspan="2">사고 자동차 이미지</td><td>손상종류</td><td rowspan="2">jpg</td><td>2,013,758 건</td><td>504,450 장</td></tr><tr><td>차량부위</td><td>179,357 건</td><td>128,244 장</td></tr><tr><td colspan="2">사고 자동차 견적 정보</td><td>xls</td><td>-</td><td>125,007 건</td></tr></table>		데이터 종류		데이터 형태	어노테이션 규모	결과물 규모	사고 자동차 이미지	손상종류	jpg	2,013,758 건	504,450 장	차량부위	179,357 건	128,244 장	사고 자동차 견적 정보		xls	-	125,007 건					
데이터 종류		데이터 형태	어노테이션 규모	결과물 규모																					
사고 자동차 이미지	손상종류	jpg	2,013,758 건	504,450 장																					
	차량부위		179,357 건	128,244 장																					
사고 자동차 견적 정보		xls	-	125,007 건																					
데이터셋 통계 (구축 규모 및 분포)	2. 데이터 분포																								
	<div>- 차량 크기 분포: 경형, 소형, 중형, 대형 차량 분포</div> <table><tr><th>차량 크기</th><th>수량</th><th>비율(%)</th></tr><tr><td>경형</td><td>109,132</td><td>17.2</td></tr><tr><td>소형</td><td>62,689</td><td>9.9</td></tr><tr><td>중형</td><td>319,207</td><td>50.5</td></tr><tr><td>대형</td><td>141,666</td><td>22.4</td></tr><tr><td>계</td><td>632,694</td><td>100.0</td></tr></table> <div>※ 라벨링 데이터 내 영어 표기법: 경형=CityCar, 소형=Compact car, 중형=Mid-size car, 대형=Full-size car</div> <div>- 차량색상 분포: 흰색, 회색, 검정, 빨강, 파랑, 기타 색상 분포</div> <table><tr><th>색상 유형</th><th>수량</th><th>비율(%)</th></tr><tr><td>흰색, 회색</td><td>421,562</td><td>66.6</td></tr></table>		차량 크기	수량	비율(%)	경형	109,132	17.2	소형	62,689	9.9	중형	319,207	50.5	대형	141,666	22.4	계	632,694	100.0	색상 유형	수량	비율(%)	흰색, 회색	421,562
차량 크기	수량	비율(%)																							
경형	109,132	17.2																							
소형	62,689	9.9																							
중형	319,207	50.5																							
대형	141,666	22.4																							
계	632,694	100.0																							
색상 유형	수량	비율(%)																							
흰색, 회색	421,562	66.6																							

검은색	96,004	15.2
파란색	45,115	7.1
빨간색	37,511	5.9
기타	32,502	5.1
계	632,694	100.0

※ 라벨링 데이터 내 영어 표기법: 흰색=White, 회색=Gray tone, 검은색=Black, 파란색=Blue tone, 빨간색=Red tone, 기타=Others

- 손상 종류 분포: 스크래치, 찌그러짐, 이격, 파손에 대한 분포

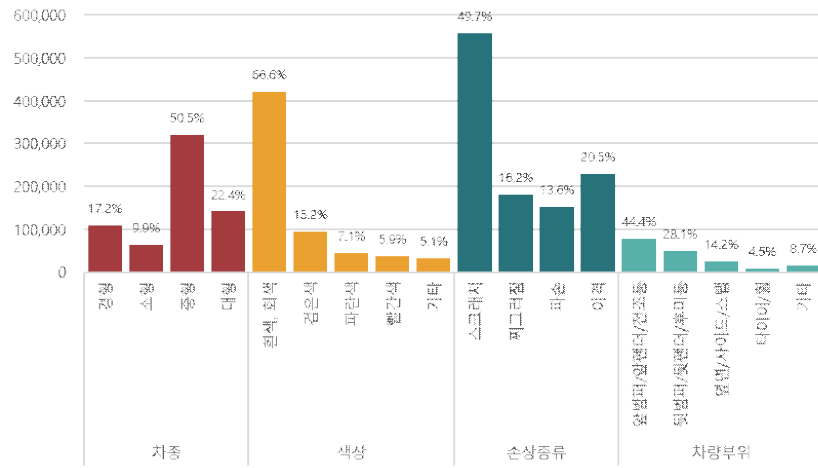
손상 종류	수량	비율(%)
스크래치	558,284	49.7
찌그러짐	182,196	16.2
파손	229,648	20.5
이격	152,531	13.6
계	1,122,659	100.0

※ 라벨링 데이터 내 영어 표기법: 흰색=White, 회색=Gray tone, 검은색=Black, 파란색=Blue tone, 빨간색=Red tone, 기타=Others

- 차량 부위 분포: 정면, 후면, 측면, 타이어, 기타 차량 부위에 대한 분포

차량 부위	수량	비율(%)
앞범퍼/앞펜더/전조등	77,830	44.4
뒷범퍼/뒷펜더/후미등	49,275	28.1
옆면/사이드/스텝	24,917	14.2
타이어/휠	7,960	4.5
기타	15,274	8.7
계	175,256	100.0

※ 라벨링 데이터 내 영어 표기법: 앞범퍼=Front bumper, 앞펜더(좌)=Front fender(L), 앞펜더(우)=Front fender(R), 전조등(좌)=Head lights(L), 전조등(우)=Head lights(R) / 뒷범퍼=Rear bumper, 뒷펜더(좌)=Rear fender(L), 뒷펜더(우)=Rear fender(R), 후미등(좌)=Rear lamp(L), 후미등(우)=Rear lamp(R) / 앞도어(좌)=Front door(L), 앞도어(우)=Front door(R), 뒷도어(좌)=Rear door(L), 뒷도어(우)=Rear door(R), 사이드미러(좌)=Side mirror(L), 사이드미러(우)=Side mirror(R), 스텝(좌)=Rocker panel(L), 스텝(우)=Rocker panel(R) / 전방바퀴(좌)=Front Wheel(L), 전방바퀴(우)=Front Wheel(R), 후방바퀴(좌)=Rear Wheel(L), 후방바퀴(우)=Rear Wheel(R) / 본넷=Bonnet, 앞유리=Windshield, 루프=Roof, 트렁크=Trunk lid, A필러(좌)=A pillar(L), A필러(우)=A pillar(R), C필러(좌)=C pillar(L), C필러(우)=C pillar(R), 뒷유리=Rear windshield, 하체=Undercarriage



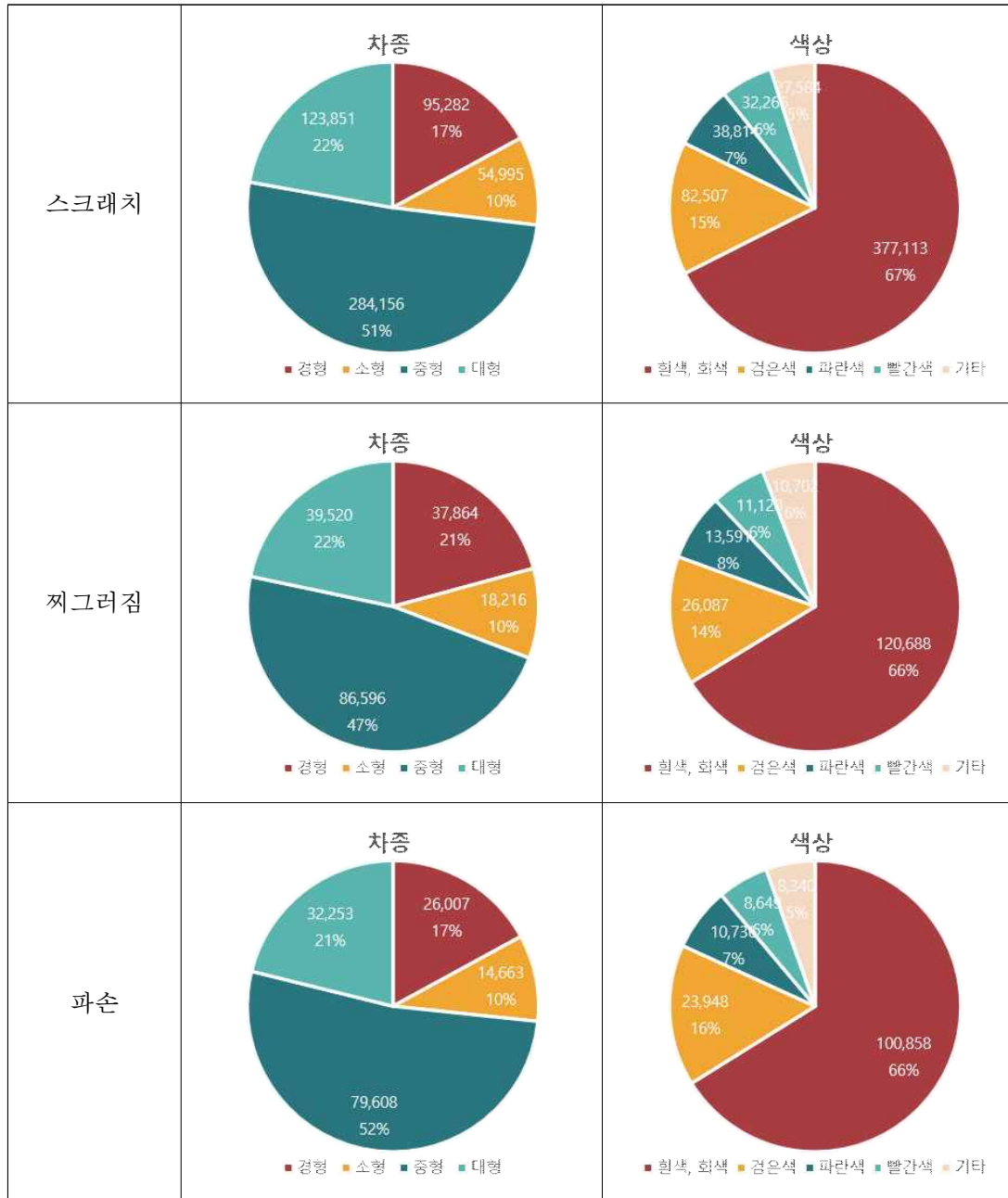
<그림> 데이터 분포

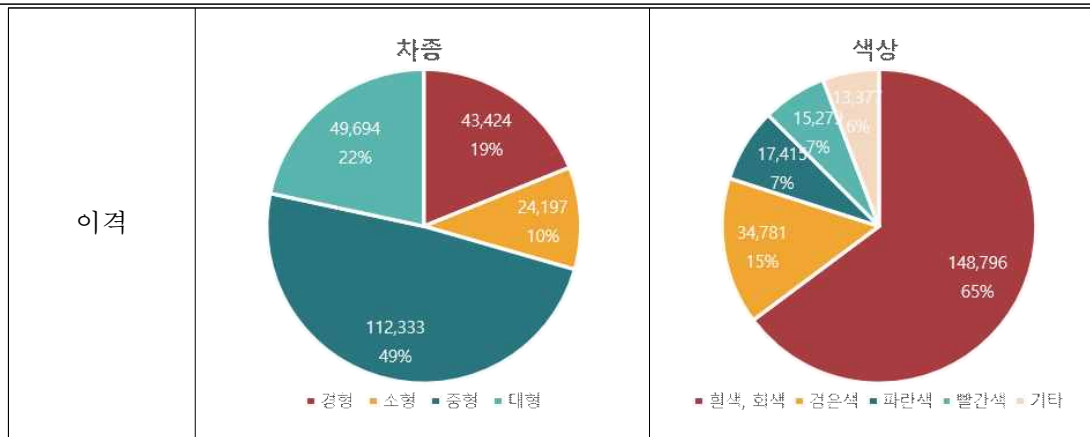
3. 상세 분포

- 손상 종류 상세 분포

구분			이미지 장수	비율
스크래치	차종	경형	95,282	17.1%
		소형	54,995	9.9%
		중형	284,156	50.9%
		대형	123,851	22.2%
	색상	흰색, 회색	377,113	67.5%
		검은색	82,507	14.8%
		파란색	38,814	7.0%
		빨간색	32,266	5.8%
		기타	27,584	4.9%
찌그러짐	차종	경형	37,864	20.8%
		소형	18,216	10.0%
		중형	86,596	47.5%
		대형	39,520	21.7%
	색상	흰색, 회색	120,688	66.2%
		검은색	26,087	14.3%
		파란색	13,591	7.5%
		빨간색	11,128	6.1%
		기타	10,702	5.9%
파손	차종	경형	26,007	17.1%
		소형	14,663	9.6%
		중형	79,608	52.2%
		대형	32,253	21.1%
	색상	흰색, 회색	100,858	66.1%
		검은색	23,948	15.7%
		파란색	10,736	7.0%
		빨간색	8,649	5.7%
		기타	8,340	5.5%

이격	차종	경형	43,424	18.9%
		소형	24,197	10.5%
		중형	112,333	48.9%
		대형	49,694	21.6%
	색상	흰색, 회색	148,796	64.8%
		검은색	34,781	15.1%
		파란색	17,415	7.6%
		빨간색	15,279	6.7%
		기타	13,377	5.8%





〈그림〉 손상 종류 상세 분포

- 차량 부위 상세 분포

구분		이미지 장수	비율
앞범퍼/앞펜더/전조등	앞범퍼	51,585	28.8%
	앞펜더(좌)	9,963	5.6%
	앞펜더(우)	10,578	5.9%
	전조등(좌)	4,865	2.7%
	전조등(우)	4,939	2.8%
뒷범퍼/뒷펜더/후미등	뒷범퍼	32,241	18.0%
	뒷펜더(좌)	5,721	3.2%
	뒷펜더(우)	7,046	3.9%
	후미등(좌)	2,157	1.2%
	후미등(우)	2,110	1.2%
옆면/사이드/스텝	앞도어(좌)	2,725	1.5%
	앞도어(우)	3,518	2.0%
	뒷도어(좌)	2,965	1.7%
	뒷도어(우)	4,914	2.7%
	사이드미러(좌)	2,511	1.4%
	사이드미러(우)	3,193	1.8%
	스텝(좌)	1,715	1.0%
	스텝(우)	3,376	1.9%
타이어/휠	전방바퀴(좌)	2,065	1.2%
	전방바퀴(우)	3,430	1.9%
	후방바퀴(좌)	1,030	0.6%
	후방바퀴(우)	1,435	0.8%
기타	본넷	6,615	3.7%
	앞유리	170	0.1%
	루프	59	0.0%
	트렁크	7,831	4.4%
	A필러(좌)	48	0.0%
	A필러(우)	41	0.0%
	C필러(좌)	47	0.0%
	C필러(우)	78	0.0%
	뒷유리	315	0.2%
	하체	70	0.0%

1. 대표도면

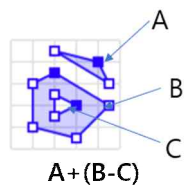


2. 라벨링데이터 구성

No	항목		길이	타입	필수여부
	한글명	영문명			
1	데이터셋정보	info		JsonObject	Y
1-1	데이터셋명	name	숫자,문자,특수문자 100 자리 이내	String	Y
	데이터셋생성일자	date_created	MM/DD/YYYY	String	Y
2	이미지정보	images		JsonObject	Y
2-1	이미지식별자	id	1	Integer	Y
	이미지너비	width	800~10000	Integer	Y
	이미지높이	height	600~10000	Integer	Y
	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,W,H]	List	Y
3-7	손상종류	damage	[스크래치, 찢그림, 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

※ 라벨링폴리곤(Segmentation) : [x1, y2, x2, y2, ...]

- GeoJSON 적용



<그림> GeoJSON 예시

※ 라벨링바운딩박스(bbox): [x, y, width, height]

데이터셋 구성

3. 라벨링데이터 실제예시

```
{
  "info":{
    "name":"external",
    "date_created":"02/17/2022"
  },
  "images":{
    "id":1,
    "width":800,
    "height":600,
    "file_name":"0000734_as-0000163.jpg"
  },
  "annotations":[
    1. {
      "id":2,
      "image_id":1,
      "category_id":"as-0000163",
      "segmentation":[..],
      "area":1252.5,
      "bbox":[
        627,
        342,
        86,
        57
      ],
      "damage":"Scratched",
      "part":null,
      "year":2015,
      "color":"White",
      "level":null,
      "repair":[
        "Front bumper:repair,coating",
        "Front fender(R):sheet_metal,coating",
        "Head lights:exchange"
      ]
    },
    {
      "id":5,
      "image_id":1,
      "category_id":"as-0000163",
      "segmentation":[..],
      "area":170492,
      "bbox":[
        42,
        199,
        694,
        286
      ],
      "damage":null,
      "part":"Front bumper",
      "year":2015,
      "color":"White",
      "level":4,
      "repair":"Front bumper:repair,coating"
    }
  ],
  "categories":{
    "id":"as-0000163",
    "supercategory_name":"Mid-size car"
  }
}
```


데이터셋 구축 수행기관 담당자	주관기관	기관명	책임자명	전화번호 (유선전화번호기입)	메일주소	담당업무
		한국전기차 리빌딩협회	김태훈	02-2219-5777	thkim@kevra.or.kr	총괄
	참여기관	기관명	담당업무	기관명	담당업무	
		(주)슈퍼브에이아이	정제,가공	(주)쏘카	데이터수집, AI 모델링	
		(주)꿀비	품질검증	(사)한국자동차검 사정비사업조합연 합회	검수	
		(사)경기도북부자동 차정비사업조합	검수			