

인공지능 학습용 데이터 구축 가이드라인

세부 데이터명	[데이터 107] 지능형 스마트측사 통합 데이터[한우]
---------	--------------------------------

인공지능 데이터 구축	사업 총괄	
	데이터 설계	 
	원천 데이터 수집/정제	  
	데이터 가공	
	데이터 검수	
	저작도구 개발	
	AI 모델 개발	
	응용 서비스 개발	
가이드라인 정보	v1.0	최초 작성, SK주식회사 C&C, 2021.12.30

목 차

1. 데이터 명세 정보	03
1.1 데이터 정보 요약	03
1.2 데이터 포맷 정의	03
1.3 라벨링 및 어노테이션 구조 정의	04
1.3.1 이미지 데이터	04
1.3.2 음성 데이터	04
1.4 데이터 분포	06
1.4.1 데이터 구축 규모	06
1.4.2 데이터 분포	06
 2. 데이터 구축 가이드	 07
2.1 데이터 구축 개요	07
2.2 데이터의 획득 방법	08
2.2.1 원시 데이터의 선정	08
2.2.2 획득 데이터의 특성 분석	11
2.3 데이터의 정제 방법	12
2.3.1 원시 데이터의 정제 방식	12
2.3.2 획득 데이터의 특성 분석	14
2.4 어노테이션/라벨링 방법	15
2.4.1 데이터 라벨링 방법 및 절차	15
2.3.2 데이터 어노테이션 포맷 및 형식 정의	17
2.5 검수 방법	21
2.5.1 검수 절차 및 기준	21
2.5.2 단계별 검수 방법	22

1. 데이터 명세 정보

1.1 데이터 정보 요약

데이터 명	지능형 스마트축사(한우,양돈) 데이터
활용 분야	[한우] 사육 상태를 진단하고 최적 상태로 유지하기 위한 사육환경, 농장 관리 등을 인공지능을 통해 분석할 수 있는 고품질 통합 데이터로 활용 [양돈] 출생부터 출하 및 유통관리 시스템 적용을 위한 통합 데이터로 돈방별 돼지 월령에 따른 사육정보 및 개체정보를 취득함으로써 양돈의 상태를 진단하고 개체의 실시간 관리를 위한 데이터로 활용
데이터 요약	[한우] 스마트 축사를 위한 월령별 한우의 상태를 진단하고 모체 정보를 연동시킴으로써 한우의 종자관리가 가능한 데이터 셋 [양돈] 돈방별 돼지의 상태를 진단하고, 개체의 실시간 관리 및 출하, 유통관리 시스템에 적용 가능한 통합 데이터 셋
데이터 출처	[한우] 강원도, 경기도, 충청도, 경상도, 전라도 등 지역별 특성을 고려하여 5개 농장 선정 [양돈] 충청도 2곳, 전라도, 경상도, 제주도 등 각 1개 농장씩 지역별 5개 농장 선정

[표 1] 데이터 정보의 요약

1.2 데이터 포맷 정의

데이터	포맷	데이터 취득 방법
영상/이미지 데이터	- MP4, JPG, XML format - 공간 해상도 : 512*512 - 시간 해상도 : 5fps 이상	데이터 수집 선정 농장에서 CCTV/Depth Camera를 이용하여 개체의 월령별 행동 상태 및 상태별 부위 영상 데이터 취득 후 단위 시간별로 슬라이싱된 이미지 데이터로 가공
음성 데이터	-WAV, JSON format	데이터 수집 선정 농장에서 특수 상황별 개체의 음성을 녹음기를 이용하여 1분 이상 취득
센서 데이터	- CSV, txt	사육환경 모니터링을 위한 축사 및 돈방 내/외부의 환경 데이터(온도/습도, 암모니아 등)를 IoT 장치 및 센서를 활용하여 5분 단위로 취득

[표 2] 데이터 포맷

1.3 라벨링 및 어노테이션 구조 정의

1.3.1 이미지 데이터

항목1	항목2	세부 항목	타입	설명
annotations	-	-	object	어노테이션 정보
	meta	-	object	메타 데이터
		task	object	메타 데이터 상세
		id	number	프레임 수
		name	string	파일명
		size	int	이미지의 픽셀
		mode	string	모드
		overlap	int	오버랩
		bugtraker	-	버그 추적
		created	date	파일 생성 일시
		updated	date	파일 업데이트 일시
		start_frame	int	시작 이미지
		stop_frame	int	종료 이미지
		frame_filter	string	추출 이미지 간격
		labels	object	객체 정의
		label	object	객체 정의 상세 (클래스별로 반복)
		name	enum	객체 이름
		color	string	색상 정보(Hex color)
		attributes	-	속성
		segments	object	세그먼트 단위
		id		
		start	int	시작 정보
		stop	int	종료 정보
		url	enum	URL 정보
		owner	object	작성자
		username	enum	계정 ID
		email	enum	이메일 주소
		assignee	enum	검수자
		original_size	object	Width, Height Size
		dumped	date	YYYY-MM-DD HH:MM:SS
		source	string	동영상 파일명
	image	-	object	프레임 정보
		id	int	프레임 정보
		name	string	파일명
		width	int	이미지 폭
		height	int	이미지 높이

[표 3] 이미지 라벨링 및 어노테이션 구조

[이미지 정보에 따른 객체 구성]

Polygon 객체 <polygon>	
<label>	해당 Polygon의 레이블 명
<occluded>	가려짐 여부
<z_order>	Polygon의 z축 정렬 값
<points>	점의 위치(';'으로 구분)
<group_id>	Polygon에서 그룹화 작업의 정보
Bounding Box 객체 <box>	
<label>	해당 Box의 레이블 명
<occluded>	가려짐 여부
<xtl>	좌상단 x 좌표 (X Top Left)
<ytl>	좌상단 y 좌표 (Y Top Left)
<xbr>	우하단 x 좌표 (X Bottom Left)
<ybr>	우하단 y 좌표 (Y Bottom Left)
<z_order>	Box의 z축 정렬 값
Keypoint 객체 <points>	
<label>	해당 KeyPoint의 레이블 명
<occluded>	가려짐 여부
<truncated>	잘려짐 여부
<points>	점의 위치
<z_order>	Keypoint의 z축 정렬 값

1.3.2 음성 데이터

구분	항목	타입	설명
한우	fileName	string	파일 이름 (A~E)
	farm	string	수집 농장
	voiceType	string	음성 종류(배고픔음, 분만음, 발정음)
	recordDate	string	수정 날짜(YYYY-MM-DD)
	cowType	string	종류(번식우, 어린송아지, 육성우)
	audioLength	string	시간(10/60초)

[표 4] 음성 라벨링 및 어노테이션 구조

1.4 데이터 통계

1.4.1 데이터 구축 규모

데이터 종류		주요 내용	구축량(TB)
원시 데이터	한우	512*512, 5pts 이상, MP4 형식, 416시간 이상	5TB 이상
원천 데이터	한우	이미지 33만장 이상	5TB 이상
라벨링 이미지	한우	512*512 화질의 JPG 이미지 파일 생성 - 총 약 33만장 구축	200GB 이상
어노테이션 파일	한우	 XML 파일 33만개 (이미지당 1개)	300MB 이상

[표 5] 데이터 종류별 구축 규모

1.4.2 데이터 분포

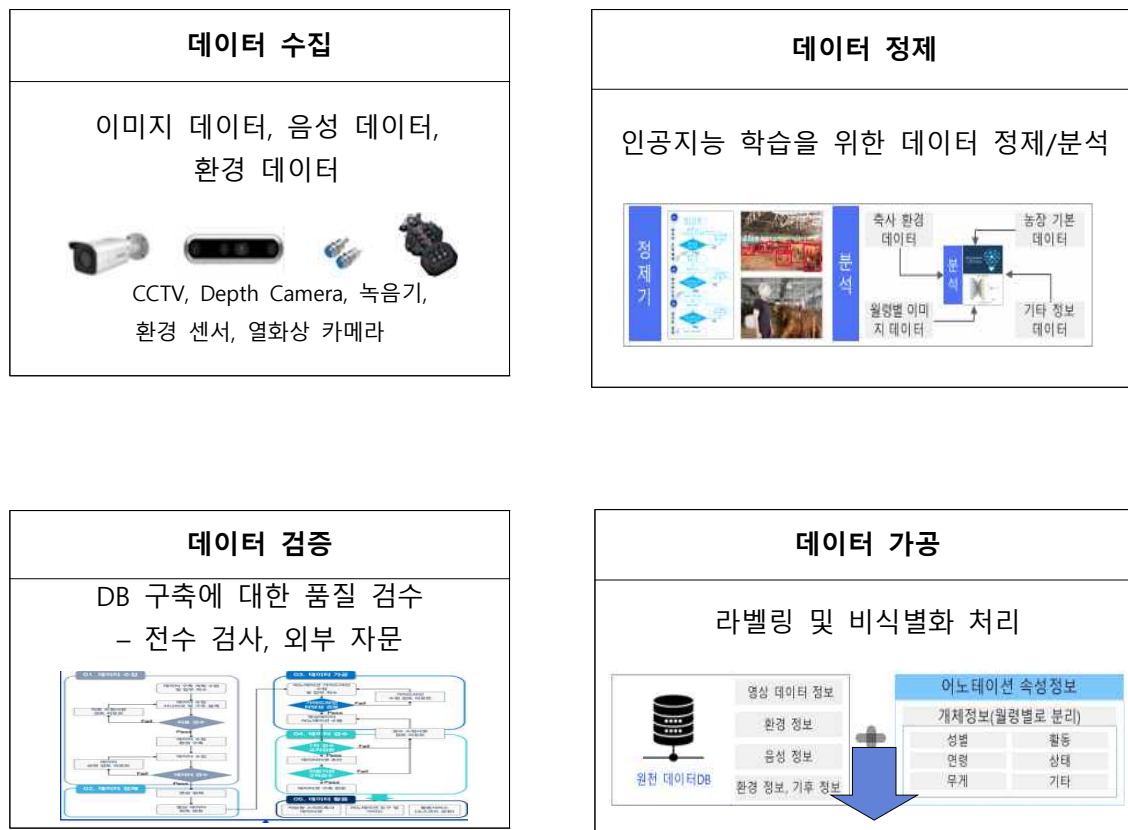
구분		산출물량	
		시간	수량
한우	일반 상황별 이미지	391	31만장
	특수 상황별 이미지	25	2만장
	음성 데이터	20	1,100건

[표 6] 데이터 분포

2. 데이터 구축 가이드

2.1 데이터 구축 개요

▣ 데이터 구축 절차



[표 6] 데이터 구축 절차

- **(수집)** 한우 5개 농장, 양돈 5개 농장에 이미지 데이터(CCTV, Depth Camera), 음성 및 환경 센서 데이터 수집
- **(정제)** 클라우드 소싱 및 직접 고용 인력을 통해 편집 도구를 활용한 정제 작업, 개인정보 포함 영상 및 문구 제거 작업
- **(가공)** 오픈 소스 기반 자체 커스터마이징 어노테이션 도구 활용, 라벨링 및 비식별화 처리를 위한 1차 가공 후 클라우드 소싱 및 자체 인력은 보정 및 수정 2차 가공 작업 수행
- **(검증)** 직접 고용 인력을 통해 모든 데이터를 1차 전수 검사 진행하고 외부 전문기관 의뢰를 통한 2차 검수

2.2 데이터의 획득 방법

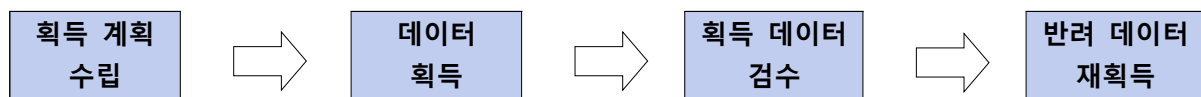
2.2.1 원시 데이터의 선정

▣ 원시 데이터의 구성

	영상	음성	환경 및 추가 데이터
데이터 구성	한우/양돈 CCTV 촬영 영상	한우/양돈 소리	온도/습도 Depth, 열화상(한우), 암모니아/메탄(양돈)
포맷	MP4	WAV	CSV, txt
예시			

[표 8] 원시 데이터의 구성

▣ 데이터의 획득 과정



- 획득 계획 수립
 - 농장 내외부 데이터 획득 장소 선정
 - 데이터 획득시 중복을 회피하기 위한 사전 협의
- 데이터 획득 : 설치된 장비(CCTV, 센서 등)의 특성 및 사용법 및 데이터 획득시 고려사항 등을 준수하여 데이터 획득
- 획득 데이터 검수 : 데이터 정제 1차 편집 작업 시 원천데이터 품질기준에 준수하여 검수
- 반려 데이터 재획득 : 검수 후 반려된 데이터 재획득

▣ 일반/특수 상황별 데이터 수집 시나리오 (한우)

대분류	중분류	데이터명	데이터 취득 방법
성장 발달	이미지 데이터	소 외부 촬영 이미지	CCTV+ 열화상 카메라+Depth Camera

영양 사양	섭취와 관련된 행동	사료조 방문 횟수 음수조 방문 횟수	CCTV
	이미지 데이터	소 내부 촬영 이미지 (육질, 마블링 섬세도)	Depth Camera
발정기 예측	음성 데이터	Vocalization	음성기록 마이크 장치
	이미지 데이터	mounting	CCTV
		안절부절하는 행동	
		walking	
		Eating	
분만 일시 예측	음성 데이터	Vocalization	음성기록 마이크 장치
	이미지 데이터	mounting	CCTV
		안절부절하는 행동	
		walking	
		Eating	
대사성 질병	이미지 데이터	분변 배출빈도, 분변 양	CCTV
		분변 상태 점수	CCTV
	이미지 데이터	Lameness	CCTV
		발굽부분 관찰	CCTV
고온 스트레스	이미지 데이터	부위별 이미지	CCTV+열화상 카메라
저온 스트레스	이미지 데이터	축사내부 바닥 축사내 소들의 위치	CCTV+열화상 카메라
동물 복지	거세 스트레스 행동	Stand	CCTV
		Lie	
		Eat	
		Drink	
		Walk	
		Stiff gait	
		Head shake	
		Head turn	
		Head paw	
		Head rubbing	
		Kick	
		Stamp	
		Ear flick	
		Tali flick	
		Lick	
	이미지 데이터	거세 상처	CCTV
	음성 데이터	축사 바닥	음성기록 마이크장치
		Vocalization	

[표 9] 한우 데이터 수집 시나리오

2.2.2 획득 데이터 특성 분석

▣ 개체 활동 영상 데이터의 특성

구분	분류	수집 방법	수집 항목 객체 정보
일반 상황 개체 활동 영상 데이터	일반적인 활동 상태 영상 데이터	CCTV 열화상 카메라 Depth Camera	동적 객체 특수 객체
특수 상황 개체 활동 영상 데이터	개체의 특수 상황 대응 영상 데이터	CCTV Depth Camera	동적 객체 특수 객체

[표 11] 개체 활동 영상 데이터의 특성

▣ 획득 데이터의 요구사항

○ 적합성

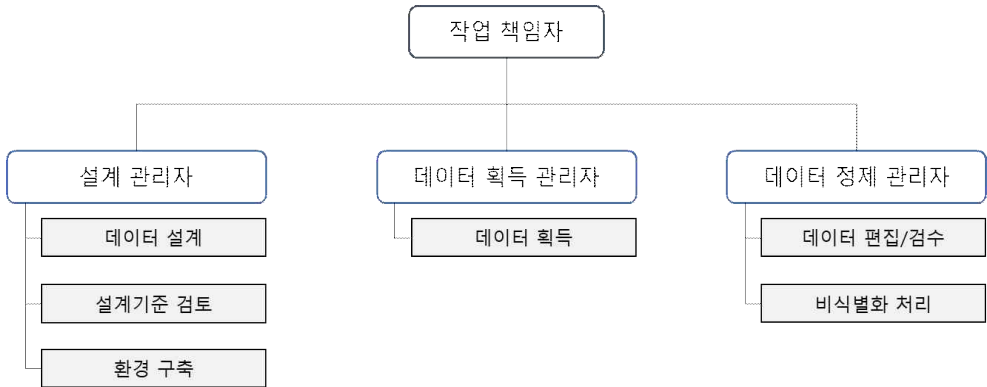
개체의 행동 상태 분석이 가능한 물리적 공간 기반의 다양한 환경에서 영상 데이터 획득

- 개체의 활동이 이루어지는 다양한 장소를 구분하여 편중 없이 영상 데이터 취득
- 일상생활과 밀접한 장소를 선정하여 실제 사육환경과 흡사한 영상 데이터 획득
- 일반상황에서 월령별 개체의 행동 영상 획득
- 돌발상황에서 상황별 대처/대응의 활동을 통한 개체의 활동 영상 획득

2.3 데이터의 정제 방법

2.3.1 원시 데이터 정제 방식

▣ 획득·정제 조직 구성



조직		역할 및 책임	교육 및 훈련
작업 책임자		- 획득 및 정제 작업 진행 관리 - 데이터 검수 및 편집 인력 할당	- 프로젝트 및 인력관리 교육
설계 관리자	데이터 설계	- 획득 데이터 분류 기준 수립 - 데이터 획득 및 정제 기준 수립 - 데이터 획득 시나리오 설계	- 데이터 관리 교육 - 데이터 속성 분석 교육
	설계 및 기준 검토	- 데이터 분류 기준 검토 - 획득 및 정제 기준 검토 - 데이터 획득 시나리오 검토	- 일반상황, 특수상황 내용 교육 - 영상 관련 특성 교육
	환경 구축	- 데이터 수집/관리 환경 구축 - 촬영 장비 구축 - 장소 및 클라우드워커 섭외	- 장비 사용법 및 주의사항 교육 - 데이터 업로드 방식 교육
데이터 획득 관리자	데이터 획득	- 데이터 수집 관리 - 클라우드워커를 통한 영상 수집	- 클라우드워커 교육
데이터 정제 관리자	데이터 검수	- 데이터 포맷, 길이, 화질 검수 - 데이터 중복, 무객체, 품질 검수 - 정제 완료 데이터 작업 내용 검수 - 폐기 및 제외 데이터 확인	- 동영상 정보 확인, 원천 데이터 품질 지표 관련 검수 방식 교육
	데이터 편집	- 25분 단위 영상편집 - 동영상 편평화 및 수평화 작업 - 개인정보 비식별화 작업	- 동영상 편집 프로그램사용법

[표 12] 데이터 획득·정제 조직

데이터 정제 과정

2단계의 정제 과정을 진행하며, 1차에서는 기본적인 영상 품질과 데이터 활용목적에 부합하는 여부, 25분 1개의 영상으로 편집을 시행하고, 2차에서는 개인정보 등의 민감정보 비식별화 처리를 목표로 함

데이터 정제 기준

○ 1차 편집 (원천 데이터 편집)

- 원천데이터 품질기준에 미달되는 영상 제거 -> 1개 동영상 기준 25분 영상으로 편집
- 1차 편집 예시

항목	라벨링 정확도 오류	오류코드	AD(Accuracy Detection)
기준	객체의 크기보다 박스 또는 폴리라인이 크거나 작게 라벨링 됨 * 허용범위 : 객체 픽셀의 $\pm 10\text{pixel}$ 오차 범위내에 바운더리 박스 생성		
승인예시		반려예시	반려사유
			바운더리 박스가 객체보다 큼

항목	라벨링 정확도 오류	오류코드	AD(Accuracy Detection)
기준	객체의 크기보다 박스 또는 폴리라인이 크거나 작게 라벨링 됨 * 허용범위 : 객체 픽셀의 $\pm 10\text{pixel}$ 오차 범위내에 바운더리 박스 생성		
승인예시		반려예시	반려사유
			바운더리 박스가 객체보다 큼

[표 13] 1차 편집 예시

○ 2차 편집 (개인정보, 민감정보 비식별화)

- 원천 데이터의 정제(비식별화) 절차는 25분 단위로 편집된 원천데이터(MP4 영상)를 비식별화 도구(도구명 : DTAAS imagemasking Tool)에 import – frame을 10FPS로 설정 – Face Detection from File 선택하여 자동 비식별화 처리 실시
- 비식별화 완료 후 1번 이미지 프레임으로 돌아와 작업자가 비식별화 처리가 잘되었는지 모든 이미지 샷을 검수 및 수정 진행함
- 정면이 아닌 특정 각도에서는 얼굴 인식이 어려울 수 있으며, 문자나 숫자형식의 개인 정보는 반드시 작업자가 직접 영역을 지정하여 비식별화 처리

2.3.2 데이터 정제시 고려사항

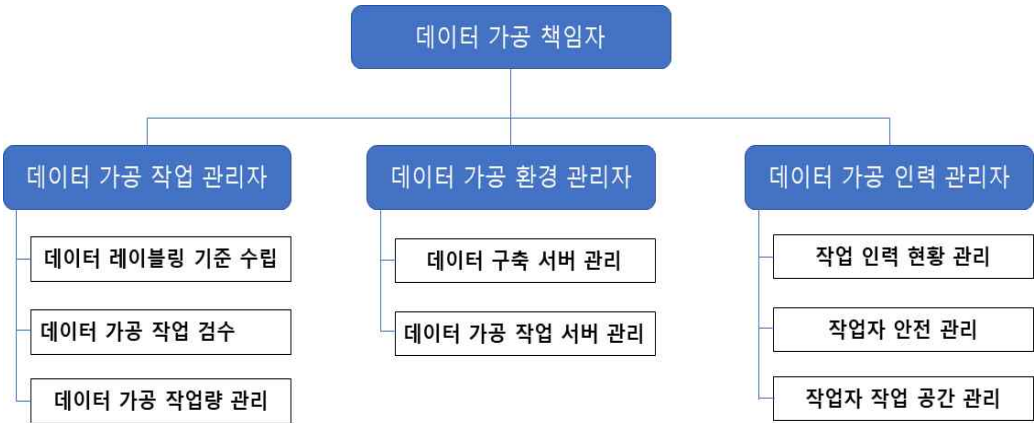
고려사항	내용	예방 방안
중복성	- 획득 단계에서 같은 장소, 같은 구도로 여러번 촬영 - 같은 영상을 이용하여 반복 정제/편집하는 경우	- 데이터 획득 시 건당 모니터링 후 작업자 피드백 - 정제 작업자의 영상 배정은 관리자가 직접 수행(권한 배정을 통해 영상 편집 및 복사에 제한)
표준화	정제 작업자의 개개인의 편차로 기준이 모호한 정제	품질관리자의 모니터링으로 신속한 피드백과 명확한 정제 기준 확립
개인 정보 및 저작권	개인정보나 민감정보의 유출 가능성	데이터 획득 단계에서 협약서, 동의서 등 관련 서류 구비, 또는 철저한 비식별화 수행
데이터의 균형	촬영의 편의성에 의한 유사 또는 동일한 장소 촬영	관리자의 명확한 임무 전달 및 확인

[표 14] 데이터 정제시 고려사항

2.4 어노테이션/라벨링 방법

2.4.1 데이터 라벨링 방법 및 절차

▣ 어노테이션/라벨링 조직 구성



조직		역할 및 책임	교육 및 훈련
데이터 가공 책임자		- 데이터 가공 진척도 관리 - 데이터 중간 관리자 업무 관리	- 인공지능 데이터 셋 구축 사업 및 데이터 가공 작업 관리자 교육
데이터 가공 작업 관리자	데이터 레이블링 기준수립	- 레이블링 작업 기준 수립 - 라벨링 객체 추가 및 업데이트	- 어노테이션 작업 교육 - 객체 정의 분류 안내
	데이터 가공작업 검수	- 레블링 작업 작업자 검수 안내 - 레이블링 작업 샘플링 검수	- 가공 작업자 검수 교육
	데이터 가공작업량 관리	- 작업자 일일 작업 할당 관리 - 작업자 일일 프레임작업량 관리	- 작업 업무 일지 작성 안내
데이터 획득 관리자	데이터 구축 서버관리	- NAS 산출물 구축량 관리 - NAS 서버 네트워크 관리	- NAS 계정 안내 - 영상 다운로드 산출물 게재 안내
데이터 가공 서버관리	어노테이션 가공 서버 에러 관리, 서버 네트워크 관리	- 가공 서버 계정 안내 - 데이터셋 추출 작업 방법 교육	- 동영상 정보 확인, 원천 데이터 품질 지표 관련 검수 방식 교육
	데이터 가공 인력 관리자	작업 인력 현황 관리	- 출근 및 재택근무 현황 관리 - 근무 시간 관리

[표 15] 어노테이션/라벨링 수행 조직

■ 어노테이션 도구의 구조

○ 어노테이션 도구 구조는 TensorFlow 2.3 + Keras + YOLO 4.0 모델 + 커스터마이징으로 구성 됨.

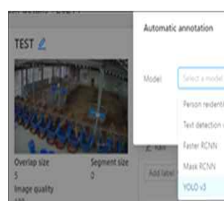
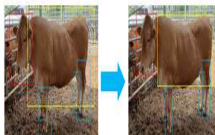
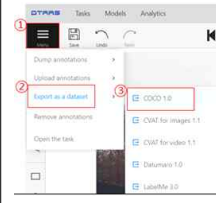
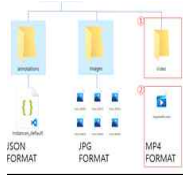
- TensorFlow : 구글에서 오픈 소스로 개방한 머신러닝을 위한 End to End 오픈 소스 플랫폼(보편적 라이브러리를 불러와 활용 됨)
- Keras : TensorFlow 2.0 이상 버전에서 지원하는 High-Level API 라이브러
- YOLO 4.0 모델 : 2020년 4월 출시된 CNN 알고리즘 기반 실시간 Object Detection 신경망 모델로 고속자율주행에서 현존 모델 중 최고의 포퍼먼스 구현

○ 데이터셋 구축을 위한 저작도구는 Content Input Layer, Repository Layer, Content Management Layer, Presentation Layer의 4가지 주요 기능 모듈로 구성 됨

기능 모듈	설명
Content Input Layer	컨텐츠를 수집 및 생성하는 부분으로 DB to DB 또는 운영자에 의한 직접 입력 등을 통해 컨텐츠의 수집 및 생성 가능
Repository Layer	수집 및 입력된 컨텐츠와 관련 메타 데이터를 저장
Content Management Layer	컨텐츠의 수정, 삭제, 복사, 이동, 버전 및 히스토리 관리 및 승인에 따른 해당 컨텐츠 승인 체계 관리
Presentation Layer	축척된 컨텐츠를 해당 매체에 맞게 실제 서비스 서버로 배포

[표 16] 저작도구의 주요 기능 모듈

■ 어노테이션/라벨링 절차

도구 준비	1차 라벨링	1차 검수	산출물 추출	완료
·저작 도구 세팅 ·계정 권한 부여	·자동 어노테이션 세팅을 통해 1차 라벨링 작업 수행	·1차 자동 어노테이션 완료한 데이터의 모든 프레임을 작업자가 직접 검수 및 수정 작업 실시	·라벨링 작업 완료 및 산출물 추출	·산출물 파일 및 이미지 파일을 해당 폴더에 저장
				

[표 16] 어노테이션/라벨링 절차

2.4.2 데이터 어노테이션 포맷과 형식 정의

어노테이션이 완료된 데이터는 어노테이션 도구를 통해 산출하게 되면 .jpg 이미지 파일과 라벨링 정보를 담고 있는.xml 파일이 생성

■ 한우 XML 데이터 구조

	XML 구조	예시
Bounding Box	<pre><?xml version="1.0" encoding="UTF-8"?> -<annotations> <version>1.1</version> -<meta> -<task> <id>219</id> <name>한우_덕풍_일반_이승엽_0818_수정 김예지</name> <size>300</size> <mode>interpolation</mode> <overlap>5</overlap> <bugtracker/> <created>2021-08-18 09:57:20.892013+00:00</created> <updated>2021-09-06 02:20:19.186262+00:00</updated> <start_frame>0</start_frame> <stop_frame>5980</stop_frame> <frame_filter>step=20</frame_filter> -<labels> -<label> <name>Right_behind_leg</name> <color>#df32db</color> <attributes> </attributes> </label> . . <dumped>2021-09-16 00:39:42.930962+00:00</dumped> <source>01_D05_F4_99_Z_210802_0659_0701.mp4</source> </meta> -<image height="1656" width="2944" name="frame_000000" id="0"> <box z_order="0" ybr="786.15" xbr="1009.94" ytl="611.90" xtl="943.33" source="manual" occluded="1" label="Right_behind_leg"> </box> <box z_order="0" ybr="475.81" xbr="1033.21" ytl="272.89" xtl="875.00" source="manual" occluded="0" label="Hip"> </box> <box z_order="0" ybr="718.51" xbr="631.61" ytl="569.50" xtl="572.00" source="manual" occluded="1" label="Right_front_leg"> </box> <box z_order="0" ybr="421.29" xbr="461.10" ytl="249.20" xtl="304.07" source="manual" occluded="1" label="Hip"> </box> <box z_order="0" ybr="827.42" xbr="1103.95" ytl="533.93" xtl="887.27" source="manual" occluded="1" label="Left_behind_leg"> </box> <box z_order="0" ybr="770.10" xbr="608.68" ytl="538.52" xtl="489.45" source="manual" occluded="0" label="Left_front_leg"> . . .</pre>	원천 데이터  라벨링 데이터 

	XML 구조	예시
Polygon	<pre> <?xml version="1.0" encoding="utf-8"?> <annotations> <version>1.1</version> <meta> <task> <id>691</id> <name>한우_우영_일반_문지원_0827</name> <size>600</size> <mode>interpolation</mode> <overlap>5</overlap> <bugtracker></bugtracker> <created>2021-08-27 11:41:14.199398+00:00</created> <updated>2021-09-07 19:02:47.387379+00:00</updated> <start_frame>0</start_frame> <stop_frame>5990</stop_frame> <frame_filter>step=10</frame_filter> <labels> <label> <name>Mating Acceptance</name> <color>#f0f00b</color> <attributes> </attributes> </label> </labels> </task> </meta> <source>04_D04_M1_F5_99_A_F_J_210821_1840_1845.mp4</source> <image id="0" name="frame_000000" width="2944" height="1656"> <polygon label="Stand" occluded="1" source="manual" points="1051.70,913.00;1047.10,902.90;1045.50,890.80;1047.10,879. 70;1048.30,870.60;1046.97,860.15;1048.93,850.35;1058.73,851.01;10 68.20,857.50;1075.40,864.60;1080.10,872.50;1083.70,881.30;1086.60 ,890.20;1089.43,899.34;1097.93,905.22;1105.76,911.76;1113.60,918. 29;1122.09,924.17;1130.59,929.39;1137.77,936.58;1144.30,944.42;11 50.18,952.91;1152.50,962.10; </polygon> </image> </meta> </annotations> </pre>	원천 데이터  라벨링 데이터 
Key point	<pre> <?xml version="1.0" encoding="utf-8"?> <annotations> <version>1.1</version> <meta> <task> <id>7</id> <name>test_0810_1</name> <size>11</size> <mode>interpolation</mode> <overlap>5</overlap> <bugtracker></bugtracker> <created>2021-08-10 07:25:11.253381+00:00</created> <updated>2021-08-12 02:51:58.303549+00:00</updated> <start_frame>0</start_frame> <stop_frame>200</stop_frame> <frame_filter>step=20</frame_filter> </task> </meta> <image id="160" name="frame_000160" width="2944" height="1656"> <points label="sitting" occluded="0" source="manual" points="169.33,1374.08;200.66,1225.25;106.66,1193.92;130.16,1084.25;247. 66,1136.47;289.44,1034.64;485.27,932.81;824.71,932.81;1057.10,971.97;12 73.82,1094.70;1145.87,1261.80;1096.26,1389.75;952.65,1468.08;897.82,154 9.02;803.82,1499.41;842.99,1439.36;989.21,1368.86;576.66,1415.86;422.60, 1426.30;534.88,1319.25;370.38,1285.30" z_order="0"> </points> </image> <image id="180" name="frame_000180" width="2944" height="1656"> </image> <image id="200" name="frame_000200" width="2944" height="1656"> </image> </annotations> </pre>	원천 데이터  라벨링 데이터 

[표 17] XML 데이터 구조 - 한우

2.5 검수 방법

2.5.1 검수 절차 및 기준

▣ 검수 절차의 정의

절차	내용
1차 데이터 검수	검수 기준에 따른 작업된 메타 데이터 확인
2차 데이터 검수	오류가 검출 되었을 경우 검수 반려 파일로 작업자에게 전달
데이터 최종 검수	오류를 5% 미만을 목표로 진행
검수 완료	데이터 어노테이션 종료

[표 18] 검수 절차

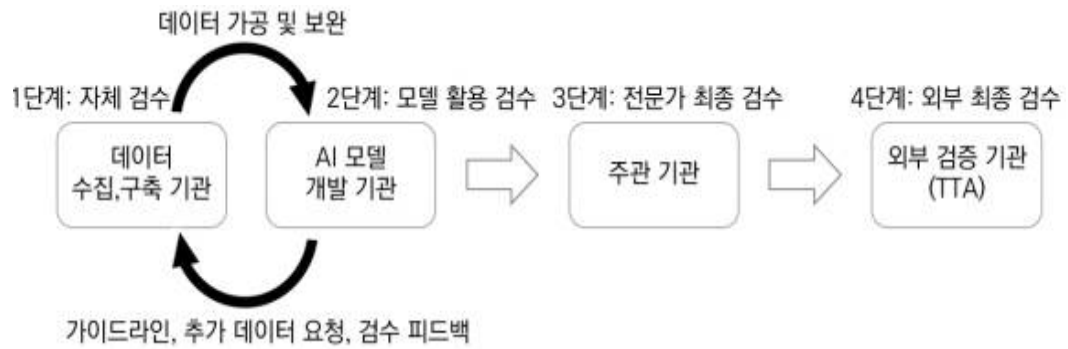
■ 검사 기준

- 작업자와 검수자로 구분하여 최종 검수 완료 파일을 생성함
 - 작업물은 미작업 파일, 작업 완료 파일, 검수 반려 파일, 최종 검수 완료 파일로 구성
 - 미작업 파일 : 작업물에 대해 작업자가 작업을 수행하지 않은 상태의 파일
 - 작업 완료 파일 : 작업자에 의해 작업 및 목시검수가 완료된 상태의 파일
 - 검수 반려 파일 : 검수 기준을 토대로 검수자에 의해 부적합 판정을 받은 작업물이 원작업자에게 배정되어 재작업을 필요로 하는 파일
 - 최종 검수 완료 파일 : 검수자에 의해 작업물이 검수 완료된 상태의 파일
- 작업자가 미작업 파일에서 목시 검수를 진행함
 - 고정형, 이동형 객체의 구분 확인
 - 과검출 및 미검출 데이터 확인
- 검수자가 비식별화 데이터 검수와 어노테이션 데이터 1차 및 2차 검수, 데이터 최종 검수를 진행 함
 - 비식별화 검수 기준 및 어노테이션 데이터의 검수 기준에 따라 오류율 5% 미만을 목표로 검수 진행
 - 과검출, 미검출 데이터 발생 시 검수 반려 파일로 작업자에게 전달 후 재작업 진행

2.5.2 단계별 검수 방법

■ 4단계 데이터 품질 검사 프로세스 적용

학계, 산업계에서 즉시 활용 가능한 수준의 품질확보를 위해 본 컨소시엄에서 설계한 4단계의 품질관리-검사 프로세스를 적용 함.



[4단계에 걸친 데이터 품질 검수]

- 1단계: 데이터 수집, 구축은 작업간에 작업자와 관리자에 의한 자체 전수 검사
- 2단계: AI 모델 개발 기관에서 AI 모델을 활용하여 데이터를 전수 검사 실시하고
검수 피드백 전달, 목표 성능 달성을 위해 반복적으로 검사 실시
- 3단계: 전문가로 구성된 검사단이 내부 최종 검수
- 4단계: 외부 검수 기관(TTA)이 중간, 최종 산출물에 대해 최종 검수