
데이터 설명서

세부 데이터명	[데이터 108] 지능형 스마트측사 통합 데이터[양돈]
---------	--------------------------------

2021.12

SK컨소시엄

메타데이터 정보 (다중기입가능)	분야	데이터 유형 ¹⁾	구축 데이터량	원천데이터 형식 ²⁾	라벨링 형식 ³⁾	라벨링 유형 ⁴⁾
	농축수산	비디오(MP4) 이미지(JPG) 음성(WAV) 주석(txt)	[양돈] 이미지:100만장 음성: 1,100개	비디오(MP4) 이미지(JPG) 음성(WAV) 주석(txt)	XML JSON	바운딩 박스 (이미지/영상), 폴리곤 (이미지/영상), 키포인트 (이미지/영상), 개체 소리(음성), 주석(텍스트)
	데이터 출처 ⁵⁾	데이터 구축년도	구축기관(총괄)	가공기관	검수기관	
	양돈 전국 각 5개 농장	2021년	SK컨소시엄	디타스	디타스	
	데이터 문의처	기관명	문의담당자명	전화번호	메일주소	
		SK컨소시엄	박영익	02-6400-5881	12blue@sk.com	
	데이터 소개	[양돈] 돈방별 돼지의 상태를 진단하고, 개체의 실시간 관리 및 출하, 유통관리 시스템에 적용 가능한 통합 데이터 셋				
	주요키워드	인공지능, 지능화 축사, 행동 정보, 상태 예측, 동물 이상 행동, 데이터, 서비스, 개체 인식, DB				
카테고리 정의서		*첨부 파일 : 스마트 축사_준비 단계_데이터 설명서_카테고리 정의서				

데이터셋명	국문영문	지능형 스마트축사(양돈) 데이터																				
		Intelligent Smart livestock farm(Korean cattle/pig) data																				
구축목적	(양돈) 축산 분야의 각 단계에 인공지능을 적용하여 축사 내·외부 생육, 상태 진단 및 최적의 상태로 유지하기 위한 사육환경, 농장관리 등을 인공지능(AI)을 통해 분석할 수 있는 통합(영상 이미지+센서 등) 데이터 구축을 목표로 함																					
활용서비스	[양돈] 태생부터 출하 및 유통관리 시스템에 적용을 위한 통합 데이터로, 돈방별 돼지의 월령에 따른 사육 정보 및 개체 정보, 행동 및 음성 정보를 제공하여 상태를 진단하고 개체의 실시간 관리가 가능한 서비스로 활용 함.																					
소개	[양돈] 양돈의 출생부터 출하 및 유통관리 시스템 적용을 위한 통합 데이터로 돈방별 돼지 월령에 따른 사육 정보 및 개체정보를 취득함으로써 양돈의 상태를 진단하고 개체의 실시간 관리를 위한 데이터 구축 스마트 축산 질병 관리 · 행동 관찰 CCTV · 생체 삽입 장비 (반추 활동, 반추위 pH, 활동량, 체온 등) 환경 관리 · 온도도 센서 · 기후 데이터 · 악취 및 가스 센서 사양 관리 · 사료 자동 급이기 · 음수 급이기 · 체중 측정 장치 데이터 활용 AI서비스 Big Data Machine Learning 조기 진단 & 자동 알람 시스템 질병 맞춤형 관리 & 예방 사료 자동 급여 수집할 데이터 행동 분석 - 질병 특이 행동 지표 · 절뚝거림 · 반추 활동 저하 등 체액 분석 - 대사 질병 생리적 지표 · 혈액 내 대사 물질 · 반추위 pH, 미생물 · 유량, 유성분 사양 실험 - 질병제어 첨가제 연구 · 효모, 완충제 동물 질병 예방 동물 복지 구현 생산성 극대화 노동력 감축 친환경 축산 정밀 사양																					
데이터셋 통계 (구축 규모 및 분포)	1. 데이터 구축 규모																					
	양돈 - 이미지 110만 장, 음성 1,100분, 환경 11만 건의 일반 및 특수 데이터																					
	<table><tr><th>데이터명</th><th>유형</th><th>구축량</th><th>데이터 설명</th></tr><tr><td>스마트축사</td><td>이미지</td><td>이미지 100만 장</td><td>- 돼지 행동별 이미지 데이터</td></tr><tr><td>통합</td><td>음성</td><td>음성 1,100분</td><td>- 임신돈 분만 일시 예측을 위한 음성/이미지 데이터</td></tr><tr><td>데이터(양돈)</td><td>환경</td><td>환경 22만 건</td><td>- 분만돈 포유 여부 확인을 위한 음성 및 이미지 데이터</td></tr><tr><td></td><td></td><td></td><td>- 비육돈 출하 예측을 위한 이미지 및 체중 데이터</td></tr></table>		데이터명	유형	구축량	데이터 설명	스마트축사	이미지	이미지 100만 장	- 돼지 행동별 이미지 데이터	통합	음성	음성 1,100분	- 임신돈 분만 일시 예측을 위한 음성/이미지 데이터	데이터(양돈)	환경	환경 22만 건	- 분만돈 포유 여부 확인을 위한 음성 및 이미지 데이터				- 비육돈 출하 예측을 위한 이미지 및 체중 데이터
	데이터명	유형	구축량	데이터 설명																		
스마트축사	이미지	이미지 100만 장	- 돼지 행동별 이미지 데이터																			
통합	음성	음성 1,100분	- 임신돈 분만 일시 예측을 위한 음성/이미지 데이터																			
데이터(양돈)	환경	환경 22만 건	- 분만돈 포유 여부 확인을 위한 음성 및 이미지 데이터																			
			- 비육돈 출하 예측을 위한 이미지 및 체중 데이터																			
2. 데이터 분포																						
<u>지역별 분포</u>																						
양돈 지역별 5개 농장 선정 - 충청도 2곳, 전라도, 경상도, 제주도 등 각 1개 농장씩 선정																						
<u>데이터 시나리오별 분포</u>																						
양돈																						

대분류	중분류	데이터명
성장발달	이미지 데이터	돼지 외부 촬영 이미지
영양 사양	섭취 관련 행동	사료조 방문 횟수, 음수조 방문 횟수
	이미지 데이터	돼지 내부 촬영 이미지(임신 진단 등)
발정 확인	음성 데이터	Vocalization
	이미지 데이터	귀 세우 기, 안절부절하는 행동, 앉았다/일어났다하는 행동, Eating, 배뇨, 배변
분만 예측	음성 데이터	Vocalization
	이미지 데이터	코로 바닥쓸기, 안절부절하는 행동, 눕기, Eating, 배뇨, 배변, 동물 내부촬영(임신 및 상태 진단)
질병 진단	이미지 데이터	분변 배출빈도, 분변 양, 분변 상태 점수
	개체 정보	활동, 움직임 빈도
	질병	홍맥폐렴, 어깨괴상, 발목부상
	이미지 데이터	지제 불량 발굽 부분 관찰
고온 스트레스	이미지 데이터	돈사 내부 바닥
		돈사내 돼지들의 위치
저온 스트레스	이미지 데이터	돈사 내부 바닥
		돈사내 돼지들의 위치
동물 복지	거세 스트레스	Stand, Lie, Eat, Drink, Walk
	이미지 데이터	거세 상처, 축사 바닥
	음성 데이터	Vocalization

1. 데이터 셋의 구조

데이터셋 구성

데이터 셋에 따른 항목과 해당 값을 아래 테이블과 같이 제시 함.

양돈

라벨링 방법	데이터 종류 및 형식	클래스 수량
·Bounding Box(객체의 부위) ·Polygon(객체의 정적 행동) ·Keypoint(객체의 동적 행동)	·원천 데이터 : 영상(MP4), 시나리오 주석(txt), 이미지(JPG), 음성(WAV) ·라벨 데이터 : XML(Bounding Box, Polygon, Keypoint), JSON (음성)	·대분류 돈방별 5개 ·상세 데이터 돈방별 15개

2. 라벨링 데이터 구성

구분	항목명	타입	필수여부	설명	범위
1	annotations	object	Y	어노테이션 정보	meta, image
1-1	meta	object	Y	메타 데이터	task, dumped, source
1-1-1	task	object	Y	메타 데이터 상세	id,name,size,mode, overlap,bugtracker, created,updated, start_frame,stop_frame, frame_filter,labels, segments, owner, assignee,original_size
1-1-1-1	id	number	Y	프레임 수	0~99999
1-1-1-2	name	string	Y	파일명	enum
1-1-1-3	size	int	Y	이미지픽셀	1920*1080
1-1-1-4	mode	string	Y	모드	-
1-1-1-5	overlap	int	Y	Overlap	-
1-1-1-6	bugtracker	-	-	-	-
1-1-1-7	created	date	Y	파일생성 일시	YYYY-MM-DD HH:MM:SS.
1-1-1-8	updated	date	Y	파일 업데이트 일시	YYYY-MM-DD HH:MM:SS.
1-1-1-9	start_frame	int	Y	시작 이미지	0
1-1-1-10	stop_frame	int	Y	종료 이미지	1~9999
1-1-1-11	frame_filter	string	Y	추출 이미지 간격	step=0~9999?
1-1-1-12	labels	object	Y	객체정의	label
1-1-1-12-1	label	object	Y	객체정의 상세(클래스별로 반복)	name,color,attributes
1-1-1-12-1-1	name	enum	Y	객체 이름	-
1-1-1-12-1-2	color	string	Y	색깔 정보 (Hex color)	#FFA000
1-1-1-12-1-3	attributes	-	-	-	-
1-1-1-13	segments	object	Y	-	segment
1-1-1-13	segment	object	Y	-	id,start,stop,url
1-1-1-13-1	id	int	Y	-	-
1-1-1-13-2	start	int	Y	-	-
1-1-1-13-3	stop	int	Y	-	-
1-1-1-13-4	url	enum	Y	-	-
1-1-1-14	owner	object	Y	작성자	username, email
1-1-1-14-1	username	enum	Y	계정 ID	-
1-1-1-14-2	email	enum	-	이메일 주소	-
1-1-1-15	assignee	enum	-	검수자	-
1-1-1-16	original_size	object	Y	-	wigth, height
1-1-1-16-1	wigth	int	Y	가로 사이즈	0~9999
1-1-1-16-2	height	int	Y	세로 사이즈	0~9999
1-1-2	dumped	date	Y	-	YYYY-MM-DD HH:MM:SS.

		1-1-3	source	string	Y	동영상 파일명	enum
		1-2	image	Object	Y	프레임 정보	id, name, Wigth, height
		1-3	id	int	Y	프레임 번호	0~999
		1-4	name	String	Y	파일명	frame_000000~999999
		1-5	wigth	int	Y	이미지 폭	2944
		1-6	height	int	Y	이미지 높이	1656
		1-6-1	box	Object	Y	어노테이션 형식	label,occluded,source, xtl,ytl,xbr, ybr, z_order
		1-6-2	label	String	Y	객체부위 정의	Right_behind_leg, Left_behind_leg, Right_front_leg, Left_front_leg, Head, Hip
		1-6-3	occluded	int	Y	가려진 객체	0,1
		1-6-4	truncated	int	Y	잘려진 객체	0,2
		1-6-5	xtl	floit	Y	좌측상단 X 위치	0~2944
		1-6-6	ytl	floit	Y	좌측상단 Y 위치	0~1656
		1-6-7	xbr	floit	Y	우측하단 X 위치	0~2944
		1-6-8	ybr	floit	Y	우측하단 Y 위치	0~1656
		1-6-9	z_order	int	-	중앙 가상점	0~3377
		1-6-1	polygon	Object	Y	어노테이션 형식	label,occluded,source, points,z_order
		1-6-2	label	String	Y	객체상태 정의	Standing, Walking, Running, Eating, Drinking, Searching, Scrubbing, Urinating,Defecting, Resting, Sitting,Coughing, Parturition, Lying, Suckling
		1-6-3	occluded	int	Y	가려진 객체	0,1
		1-6-4	truncated	int	Y	잘려진 객체	0,2
		1-6-5	points	floit	Y	포인트 위치	x=0~2944 y=0~1656
		1-6-6	z_order	int	-	중앙 가상점	0~3377
		1-6-1	keypoint	Object	Y	어노테이션 형식	labeloccluded,source, points,z_order
		1-6-2	label	String	Y	객체상태 정의	Standing, Walking, Running, Eating, Drinking, Searching, Scrubbing, Urinating, Defecting, Resting, Sitting ,Coughing, Parturition, Lying, Suckling
		1-6-3	occluded	int	Y	가려진객체	0,1
		1-6-4	truncated	int	Y	잘려진객체	0,2
		1-6-5	points	floit	Y	포인트 위치	x=0~2944 y=0~1656
		1-6-6	z_order	int	-	중앙 가상점	0~3377

3. 라벨링 데이터 실제 예시

XML 데이터 - Bounding Box 예

```
<?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">
.
.
```

JSON 데이터 (음성)

```
{
  "fileName": "A-5-1-31",
  "farmName": "새봄 ",
  "voiceType": "발정음",
  "recordDate": "2021-09-14",
  "cowType": "번식우",
  "recordTime": "10"
}
```

데이터셋 구축 수행기관 담당자	주관기관	기관명	책임자명	전화번호 (유선전화번호기 입)	메일주소	담당업무
		SK주식회사 C&C	박영역	02-6400-5881	yepark@sk.com	PM
	참여기관	기관명	담당업무	기관명	담당업무	
		디타스	데이터 가공	한경대학교	활용 서비스 제시	
		여덟끼니	데이터 정제	서울대학교	데이터 수집/분류	
		리얼팜	현장 시스템 구축			
		이지놈	AI 서비스 모델 개발			