

인공지능 데이터 구축 가이드라인

- 수직농장통합데이터 연계류 -

담당 역할	기관명
데이터 구축 총괄	CCmediaService
데이터 설계	CCmediaService
데이터수집 및 정제	 (주)유비엔  경상대학교 GYEONGSANG NATIONAL UNIVERSITY
	 경상남도농업기술원
가공(라벨링, 어노테이션)	 (주)유비엔
데이터 검수(자체 검수)	 (주)유비엔

구축 가이드라인 작성	(주)씨씨미디어서비스	강진석
	(주)유비엔	김지성
가이드라인 버전 (제작일자)	V0.5 2021.10.14	

목 차

1. 데이터 구축 개요	3
2. 문제 정의	3
2.1 임무 정의	3
2.2 데이터 구축 유의사항	3
3. 데이터 수집·정제	5
3.1 원시데이터 선정	5
3.2 수집·정제 절차 및 방법	10
3.3 데이터 적재방안	14
3.4 수집 및 정제 기준	15
3.5 수집·정제 도구	16
4. 데이터 가공	17
4.1 가공 절차 및 방법	17
4.2 가공 기준	19
4.3 가공 규격	21
4.4 가공 조직	23
4.5 가공 도구	24
5. 검수	26
5.1 검수 절차 및 방법	27
5.2 검수 기준	28
5.3 검수 조직	28
5.4 검수 절차	29
5.5 검수 도구	30
5.6 기타 품질관리 활동	30

1. 데이터 구축 개요

1.1. 데이터 구축 개요

수직농장통합 엽채류 데이터의 수집 목적은 수직농장, 스마트팜 환경에서 재배되는 엽채류의 생육 이미지와 온도, 습도, CO2 등 엽채류의 생육에 영향을 미치는 다양한 재배환경데이터를 센서 수집 기술을 활용하여 수집을 진행하고, 정제 가공 검수 과정과 부가 정보를 메타 데이터로 함께 구성하여 일반인, 농민등이 딸기의 재배 및 재배환경에 대한 응용서비스 개발 및 연구에 활용 할 수 있는 AI 학습용 자료를 만든 것을 목적으로 함.

수직농장 엽채류 AI 데이터 셋 구축 절차



2. 문제 정의

2.1. 임무 정의

본 가이드라인은 수직농장 및 스마트팜 온실에서 재배되는 엽채류 4개 작물에 대하여, 생육단계별 이미지를 스마트폰을 활용한 자동 수집 시스템을 이용하여 1시간단위로 수집 100,000만장을 수집하며, 재배환경을 해당 이미지와 함께 200,0000건을 수집 및 가공을 통하여 고품질의 인공지능 학습 데이터를 구축하기 위하여 작성되었음.

구축된 엽채류 데이터셋을 활용하여 영농인에게 현재 재배하고 있는 엽채류 작물에 대한 생장 및 생육상태 정보등을 제공하고, 다양한 AI 기반 기술을 활용할 수 있는 기반을 마련하고 농업기술에 발전에 기여함을 목적으로 함.

아울러 본 사업을 통하여 구축된 딸기 학습용 데이터는 AI-Hub 에 공개되어, 대국민 활용이 가능하도록 공공데이터로 활용가능 하도록 지원하며, 농업 AI 기술을 기반으로 하는 스마트 농업분야 국가 경쟁력 향상에 이바지 하고자 함.

또한, 데이터 가공 및 정제 과정에서 온라인 기반의 클라우드 소싱을 통해 일자리 창출을 지원하고, 농업 AI 기술 보유 기업을 지원함으로써 궁극적인 농업 AI 산업에 이바지 함을 그 임무로 함.

2.2. 데이터 구축 유의사항

- (데이터 수집) 업체류 데이터수집 시 여러 빈도의 정식 과정을 거치게 되므로 데이터의 겹침, 생육지표등의 지표가 겹치게 되는 현상을 사전에 예방 고려하여야 함.
- 생육단계의 변화와 재배 중 다양한 문제로 인하여 수집에 한계가 발생 할 수 있다. 따라서 사전 조사를 진행하여 전 생육단계에 대한 데이터를 확보하도록 한다.
- 자동 데이터 수집 환경을 유지하기위하여 지속적인 모니터링 과 네트워크 환경을 체크 점검하여 데이터의 결측이 최소화 될 수 있도록 관리체계를 구축 운영하여야 함.
- (개인정보)이미지 데이터 중 재배 중인 농민이 촬영된 경우 해당 데이터를 정제 과정에서 삭제 처리 진행함. 본 데이터수집의 경우 개인정보라 할 수 있는 수집 장소의 지번, 또는 위치정보를 포함하지 않음.
- (저작권 및 활용범위) 수집 농장과 임대 계약시 사전 저작권 및 공공활용이 가능한 내용이 포함된 촬영동의서 기명 후 서명날인 청구.
- (데이터 가공 및 배포) 가공 및 배포에 대하여 NIA 와 저작권 공유 기간 협의를 진행하고, 모든 데이터는 2차 가공 및 배포를 금지함.

2.3. 원시데이터 선정

2.3.1 업체류개체 클래스 정의

- 업체류를 대상개체로 정의 하며, 클래스의 구분 은 품종과 생육단계를 기준으로 데이터의 다양성을 확보하도록 구성 함.

대상 작물		산출 (작물의 생육특성 및 작황 현장 수집환경에 따라 변동 가능)									
연번	품목	※ 현장 수집환경 및 작물생육 상황에 따라 품종별 생육단계별 수집을 및 개체 수량 변동 가능									
업체류 작물		개체수	품종	대상	정식기	생육기	수확기	소계			
		25	상추	이미지	2,500	15,000	7,500	25,000		이미지	
				재배환경 데이터	5,000	30,000	15,000	50,000			재배환경 데이터
				생육 지표 데이터	7(개월)*4(주)*25(개체수)=700 SET 1주 단위 측정(측정 샘플 수량은 작물 재배 특성 및 정식일자에 따라 수량 변동 가능)						
		25	케일	이미지	2,500	15,000	7,500	25,000		재배환경 데이터	
				재배환경 데이터	5,000	30,000	15,000	50,000			생육지표 데이터
				생육 지표 데이터	7(개월)*4(주)*25(개체수)=700 SET 1주 단위 측정(측정 샘플 수량은 작물 재배 특성 및 정식일자에 따라 수량 변동 가능)						
		25	겨자채	이미지	2,500	15,000	7,500	25,000		생육지표 데이터	
				재배환경 데이터	5,000	30,000	15,000	50,000			2800건
				생육 지표 데이터	7(개월)*4(주)*25(개체수)=700 SET 1주 단위 측정(측정 샘플 수량은 작물 재배 특성 및 정식일자에 따라 수량 변동 가능)						
		25	근대	이미지	2,500	15,000	7,500	25,000		2800건	
				재배환경 데이터	5,000	30,000	15,000	50,000			2800건
				생육 지표 데이터	7(개월)*4(주)*25(개체수)=700 SET 1주 단위 측정(측정 샘플 수량은 작물 재배 특성 및 정식일자에 따라 수량 변동 가능)						

표 1) 업체류 개체 클래스 정의

- 이미지 데이터 각 (업채류) 100,000건, 재배환경데이터 (업채류) 200,000건
- 생육지표데이터는 주 단위 측정 요원에 의하여 수동 수집
- 획득한 이미지 데이터의 수량을 동일하게 정제,가공,검수 부분에 적용
- 작물의 생육특성, 재배환경 및 주변여건에 따라 수집 내역 및 수량 변동 가능

3. 데이터수집·정제

3.1 원시데이터선정

3.1.1 데이터 포맷 정의

- 학습용 데이터 포맷

구 분		항목		데이터 포맷
원본 데이터		데이터형식	엽채류 -이미지	JPG
			재배환경	CSV OR TEXT
			시계열데이터	
			생육지표-값	
학습용 데이터	이미지	작물 이미지		JPG
	어노테이션	특징 이미지		bbox[좌표값]
	라벨링 테이블	이미지메타정보		JSON (메타데이터)
		작물 메타정보		
		환경 메타정보		
생육지표 메타 정보				

표2) 학습용 데이터 포맷 정의

- 데이터 명세

구 분		데이터 명	데이터타입/단위
데이터 정보	일반정보	이미지 고유 번호	Number
		파일경로	String
		파일종류	String
		저작권	String
	작물정보	작물 코드	String
		농장 코드	String
		작물명	String
		품종	String
		생육단계	String
	이미지		이미지 이름
이미지 넓이			String
이미지 높이			String
이미지 촬영일			String
이미지 생성일			String
재배환경 정보		농장 코드	String

(시계열 센서데이터)	수집 일시	String
	온도 센서	String(℃)
	CO2 센서	String(ppm)
	습도 센서	String(%)
	(수직)광량 센서	String(W/m²)
	(수직양액) EC 센서	String(ms/cm)
	(수직양액) PH 센서	String(ph)
	(수직양액) 온도 센서	String(℃)
생육지표 (수기입력데이터)	초장	String(cm)
	엽수	String(cm)
	엽폭	String(cm)
	엽장	String(cm)
	엽형지수	String
	엽병장	String(cm)
	과중(생체중)	String(g)
	SPAD	String
기타 정보(ETC)	질병정보	String
	이상장애 정보	String
	이미지 경로	String
어노테이션	화방	String
	줄기	String
	과실	String
	엽	String
	주	String
	BBOX	Number

표 3)데이터 명세

3.1.2 데이터 속성 정의

○ 이미지 속성 정보

No.	항목		길이	타입	필수 여부	비고
	한글명	영문명				
1	이미지 고유번호	images_id	8	varchar	Y	
1-1	농장 코드	farm_id	8	varchar	Y	
1-2	작물 코드	Crops_id	8	varchar	Y	
1-3	작물명	Crops	8	varchar	Y	
1-4	품종명	Kind_Type	8	varchar	Y	
1-5	파일경로	file_path	100	varchar	Y	
1-6	이미지이름	fname	100	varchar	Y	
1-7	이미지종류	fext	100	varchar	Y	
1-8	이미지 넓이	width	1	varchar	Y	
1-9	이미지 높이	height	1	varchar	Y	
1-10	이미지 촬영일	captured_date	-	varchar	Y	
1-11	이미지 생성일	Create_date	-	varchar	Y	
1-12	이미지 넓이	width	1	varchar	Y	어노테이션 결과에 따라 결측 가능
1-13	이미지 높이	height	1	varchar	Y	
1-14	작물엽	leaf	1	varchar		
1-15	작물 주	plant_body	1	varchar		

표 4) 이미지정보 데이터셋 속성정의

○ 시설정보 속성

No.	항목		길이	타입	필수 여부	비고
	한글명	영문명				
1	농장코드	Farm_id	8	varchar	Y	
2	측정일시	Receive_date	20	varchar	Y	

표 5) 시설정보속성정의

○ 환경정보 데이터셋 속성

No.	항목				길이	타입	필수 여부	비고
	분류	한글명	영문명	단위				
1	내부환경	농장코드	Farm_id	-	8	varchar	Y	네트워크 환경에 따라 결측 가능
2								
3		수집일시	Receive_date	date	100	varchar	Y	
4		내부CO2	CI	ppm	20	varchar		
5		내부습도	HI	%	20	varchar		
7		일사량	SR	W/m-2·s	20	varchar		
8		광량	IR	μmol/m-2·s	20	varchar		
9		내부온도	TI	도	20	varchar		
11								
12	배지환경 양액환경	양액EC	EI	ds/m	20	varchar		
13		양액PH	PI	ph	20	varchar		
		양액온도	TL	도	20	varchar		

표 6)환경정보 데이터셋 속성 정의

○ 생육측정지표 정보 속성

No.	항목				길이	타입	필수 여부	비고
	분류	한글명	영문명	단위				
1	생육지표	농장코드	Farm_id	-	8	varchar	Y	
		작물코드	Crops_id	-	8	varchar	Y	
		초장	stem_length	길이(cm)	20	varchar		■
2		엽수	leaf_cnt	수량(장)	20	varchar		■
		엽폭	leaf_width	길이(cm)	20	varchar		■
3		엽장	leaf_length	길이(cm)	20	varchar		■
4		엽병장	petiole_length	길이(cm)	20	varchar		엽채류
5		엽형지수	leaf_index	-	20	varchar		엽채류
6		과중(생체중)	fr_weight	무게(g)	20	varchar		■
7		SPAD	SPAD		20	varchar		엽채류
8								
9		생육단계	growth_stage	-	20	varchar	Y	
10		ETC	inform	-	100	varchar		※발생시 병해충 이상장애

표 7)생육지표 속성정의

3.1.3 데이터 특성

구 분	데이터 포맷	주요 특성
작물이미지	이미지(JPG)	10만 화소 이상의 JPG 형식 이미지 - 시계열 특성 - 자동 수집 - 작물의 생육 특성 반영
재배 환경 데이터	JSON	- 시계열 특성 - 자동 수집 - 재배 온실, 시설 환경 반영
생육지표 데이터	JSON	- 딸기, 엽채류 각 작물마다 고유한 생육 특성 반영 - 직접 수집 - 작물 생육지표 반영

표 8) 데이터특성 정의

3.1.4 데이터 저장구조

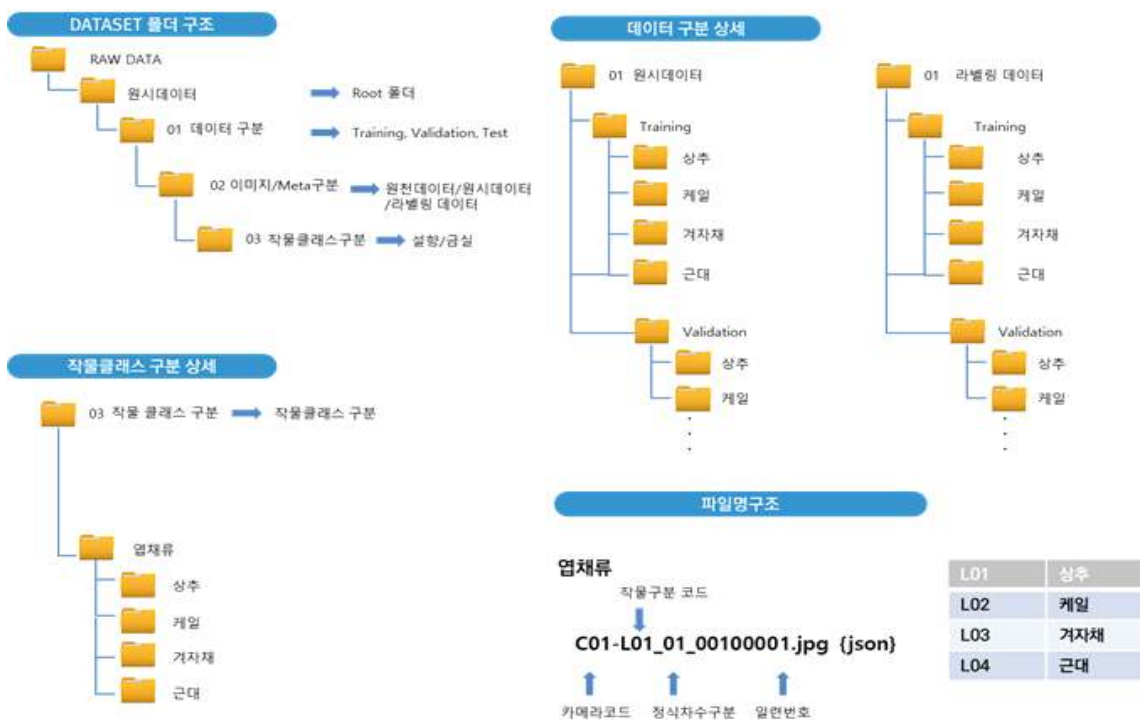


그림 1) 데이터저장구조

3.1.5 데이터의 출처

- 데이터의 수집은 경남 지역의 민간 농장, 기술원 등의 딸기재배 가능 장소를 섭외 하여 수집을 진행
- 데이터의 저작권은 기 수집 데이터의 활용이 아닌, 직접 사업자가 촬영 수집하여 저작권상의 문제는 일체 존재 하지 않음.

번호	품종	형태	장소	지역	시설유형	규모
1	엽채류	민간	민간농장 D	경상북도	수직농장	660㎡ 1실
2		공공	경상남도농업기술원	경상남도	수직농장	30㎡ 1실
3		공공	경상남도농업기술원	경상남도	수직농장	20㎡ 1실
4		공공	경상남도농업기술원	경상남도	수직농장	20㎡ 1실
5		민간	민간농장 E	경상남도	수직농장	165㎡ 1실

표9) 데이터 수집장소

3.2 수집·정제 절차 및 방법

3.2.1 데이터 획득 절차



그림 2) 데이터 수집 절차

3.2.2 수집 대상 정의

- 명확한 대상 정의를 통하여 AI 데이터셋 구축 목표를 설정 후 데이터 획득 진행
- 취득 데이터의 다양성 및 편향성 방지를 위한 지역, 장소, 품종 등 목표 수집 비율 목표를 사전에 정의하고 Baseline으로 삼아 품질 관리의 기준으로 삼음
- 고품질 딸기 이미지 영상 데이터 및 재배 환경 정보를 획득하기 위하여 설치 장소에 대한 사전 조사 진행을 통하여 최적의 설치 위치 선정
- 이미지 영상데이터 획득을 위한 고정 카메라, 환경 센서를 비롯한 재배환경 정보 수집을 위한 장비 설치와 현장 방문 시 생육정보 및 재배환경 예측 등 데이터 취득을 위한

스마트폰 구비

3.2.3 장비설치사항

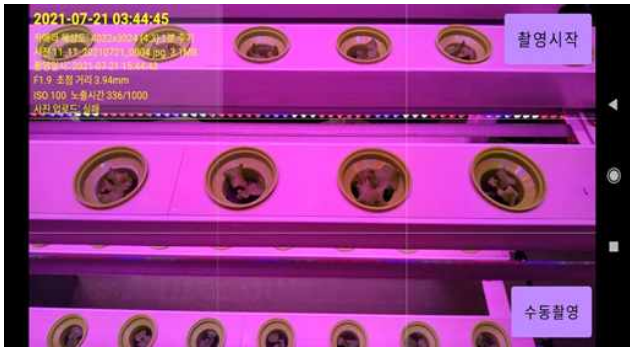
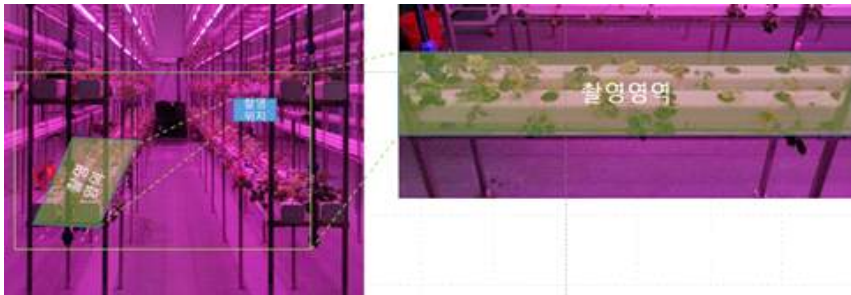
고정형 카메라 설치 사항			
구분	품 명	관측(설치)위치	
내부	이미지 획득	스마트폰 카메라 활용	· 장비는 (주)유비엔 담당자가 현장 방문 및 시찰 후 최적 위치에 설치(현장 상황에 따라 유동적 변경가능)
	규격	· 해상도 10만 화소 이상 · 데이터 통신이 가능한 기기 선정 - 병해 발생시 생육지표데이터 수집용 스마트폰을 이용하여 촬영 · 일정 시간 단위 규칙적인 촬영 기능을 하는 자동촬영 앱 개발  · 수직농장의 촬영 환경을 고려하여 카메라 위치 선정 	

표 10) 고정형 카메라 설치 방안

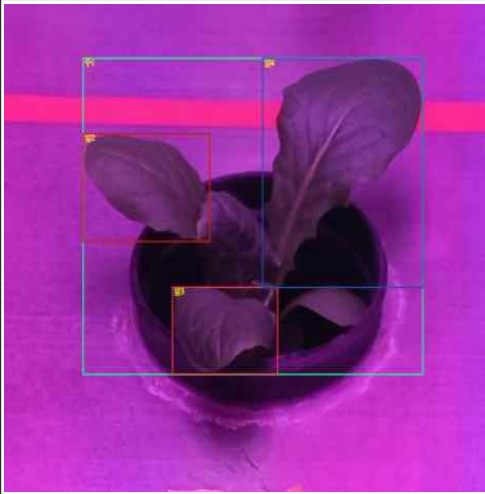
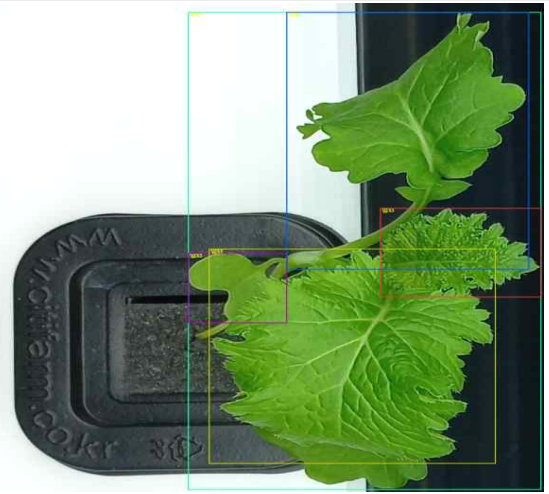
고려사항		
항목	딸기	
고려 사항	- 작물의 생육 특성을 반영하여 설치 -엽채류 : 재배 진행시 좌우로 퍼져 성장하는 특성 반영 - 수집장소(수직농장)등의 현장 상황에 유연하게 설치 운영 - 객체의 상이 틀어지거나, 변형되지 않도록 수집/플래쉬 사용금지 - 장비별, 객체별, 시간별, 지역별, 검토 (부가적으로 병해 발생시 병해충 정보입력) - 촬영에 적합한 거리, 각도, 해상도, 조도를 확보 후 촬영(농장환경에 따라 변경가능) - 병해충 및 이상장애 발생시 : 고정카메라 외 생육지표 수집 앱을 활용하여 병해충 및 생육장애 이미지 근접 이미지 취득	
일반 수집 (고정형 카메라) 예시	대각선 상부 위치 운영 예시	수직 상부 설치 운영 예시
		
병해충발생시 예시 (근접촬영) 스마트폰 활용		
병해충 발생시	- 근접 촬영 시행을 통해 병증이 선명하게 보이도록 수집 - 병징이 양면으로 발생시 각면의 병징을 별도로 수집 - 부위별, 감염정도별, 시기별 다양한 증상 데이터 확보 - 여러개의 병증이 복합될 경우 제외 ※ 병해충 및 이상장애 발생 시 농가의 사정으로 중단되어 질수 있으며, 발생시에 한하여 데이터를 수집하고자 함.	

표 11)엽채류 데이터 수집시 고려사항

센서 설치 공통사항			
구분	품 명	관측(설치)위치	
내부	미기상 센서	온도센서	· 생장환경 차이는 주로 입구 쪽에서 발생하며 온실내부와 온도는 3-5도, 습도는 20-30% 정도의 편차가 발생하므로 이를 감안하여 설치한다. · 설치수량은 3,333㎡ 기준으로 온습도 2개, CO₂는 1개를 권장한다. · 생장점에 근접하여 설치하여야 하며, 생장점 이동이 필요한 작물에 대한 센서의 설치 위치 이동이 가능하도록 안전성을 감안하여 설치한다. · 작물의 특성에 따라 근권부, 줄기부, 상층부 중 선택하여 설치할 수 있다. · 기본수량(1개)을 설치하는 경우 센서 데이터 값이 온실 환경의 대표성을 가질 수 있도록 하여야 한다.
		습도센서	
		CO ₂ 센서	
		광량센서	
	양액/ 토양 센서	토양수분, EC, 지온	· 근권부의 특성을 적절히 파악할 수 있는 위치에 설치하여야 한다. · 유지관리를 위하여 안전과 센서 위치의 파악이 가능한 지역을 선택하여 설치한다. · 기본수량(1개)을 설치하는 경우 센서 데이터 값이 온실 환경의 대표성을 가질 수 있도록 하여야 한다.

표 12) 센서 설치 공통사항

센서규격				
구분	측정범위	해상도 (정밀도) (분해능)	오차범위	운용환경
온도	-20℃ ~ +80℃	0.1℃	상온에서±0.3℃	-50℃ ~ +60℃
습도	0 ~ 100%	±2% RH	±2.0 %(0 ~ 90 %), ±3.0 %(90 ~ 100%)	-40℃ ~ +60℃
광량	400~700nm		95% 신뢰수준	-40℃~+80℃
CO ₂	0 ~ 3,000ppm		±50ppm이내	온도 : 0 ~ 60℃ 습도 : 0 ~ 95% RH
양액온도	-20℃ ~ +80℃	0.1℃	상온에서±0.3℃	-50℃ ~ +60℃
양액EC	0~100000us			
양액PH	0~14ph			

표 13) 센서규격

- ※ 센서의 입력은 위 표에 준한다.
 ※ 독립센서는 출력이 4~20mA를 갖는 센서를 사용한다.(감우감지센서는 제외)
 ※ 센서는 임차 기준으로 기간 운영 하며 사업 완료 후 철거

- 획득한 원시데이터를 체계적으로 분류 및 중복 방지 및 원자성을 유지하기 위하여 데이터 종류 및 분류에 따른 라벨링 데이터 파일을 파일 명명법과 파일 저장구조로 정의하고, 정의된 내용에 맞게 파일을 저장 및 관리

3.4 수집 및 정제 기준

3.4.1 원시데이터 수집 기준

구 분		세부 항목	세부 사양
원천 데이터	엽채류 생육 이미지	데이터형식	JPG
		이미지 해상도	10만 화소 이상
		이미지 메타 정보	CSV → JSON 변환
	재배 환경데이터	데이터형식	CSV → JSON 변환
	수동 수집 생육데이터	데이터형식	CSV → JSON 변환

표 15) 원시데이터 수집기준

- 고정형 스마트폰 카메라를 통하여 딸기 생육단계별 이미지를 1시간 단위로 획득하고 이와 함께 이미지 기본 정보에 대한 메타데이터를 획득
- 5분 간격으로 재배환경 데이터를 수집하여 수집 서버로 전송한 후 이미지 데이터와 어노테이션 작업을 통해 JSON 파일 형태로 변환한다
- 수동 수집데이터는 현장의 수집자의 스마트폰에 전용 촬영 APP을 설치하여 딸기 생육 정보가 획득되고, 데이터는 스마트폰 APP으로부터 수집 서버로 전송, 재배 환경데이터와 변환작업 프로세스를 거쳐 병합되며 이후 어노테이션 작업을 통해 JSON 파일 형태로 변환

3.4.2 원시데이터 정제 기준

- 원시데이터의 정제는 학습데이터로 적합한 데이터를 선별하고 학습데이터의 품질을 향상시키는 작업으로서 본 과제의 특성을 반영하여 정제 절차 및 기준을 수립하여 수행

단계	대상	정제기준
획득 시 정제	촬영 수집수단	• 스마트폰 카메라를 사용
	촬영 수집대상	• 수집대상을 제외한 타 작물 이미지 불허
	이미지 원시데이터 제약사항	• 촬영대상의 초점이 맞지 않은 데이터 불허
		• 심도가 낮은 사진 허용 불허
		• 이미지 흔들림 불허
		• 대상 잘림 혹은 가림 데이터 불허
		• 흐릿한 이미지, 과다 노이즈출이미지 불허
어노테이션	재배 환경 수집수단	• 이상 데이터(과다 측정 등) 불허
	개인정보 저작권	• 불특정한 시계열 특성 데이터 불허
		• 개인정보가 포함 시 비식별 처리
		• 현장 직접 수집으로 해당 없음
	어노테이션 작업시	• 중복/유사 이미지 제외

		• 작물의 데이터셋 구축 목적에 미달 시 제외 • (생육모델경우 지속적 반복 이미지 등)
--	--	--

표 16)원시데이터의 정제 기준

3.5 수집·정제 도구

○ 원시데이터 수집 획득 도구

생육이미지 수집	생육지표	내용
		• 생육이미지 수집용 카메라(고정형) • 생육지표정보 스마트폰 APP(전용APP) 또는 수기업로드 엑셀

표 17) 데이터 수집도구

○ 원시데이터 정제 도구

어노테이션 도구	내용
< 20200915_141941.jpg >  <자체 개발 라벨링 정제 도구>	• 어노테이션 도구를 활용하여 작업자가 직접 작업 진행 • 전수 정제 도구로 활용

표 18) 원시데이터 정제도구

4. 데이터 가공

4.1. 가공 절차 및 방법

4.1.1 가공 전체 프로세스

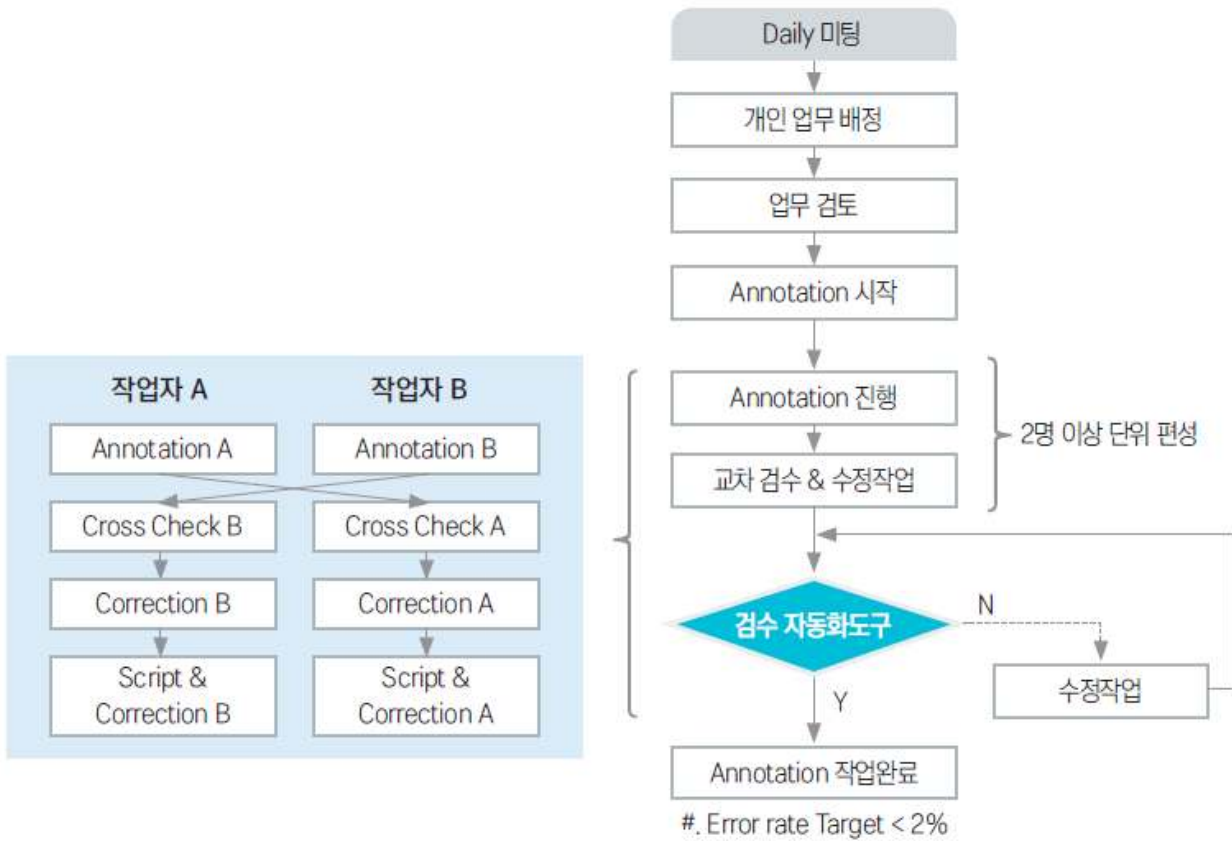


그림 4) 라벨링 전체 프로세스

4.1.2 가공 세부 프로세스

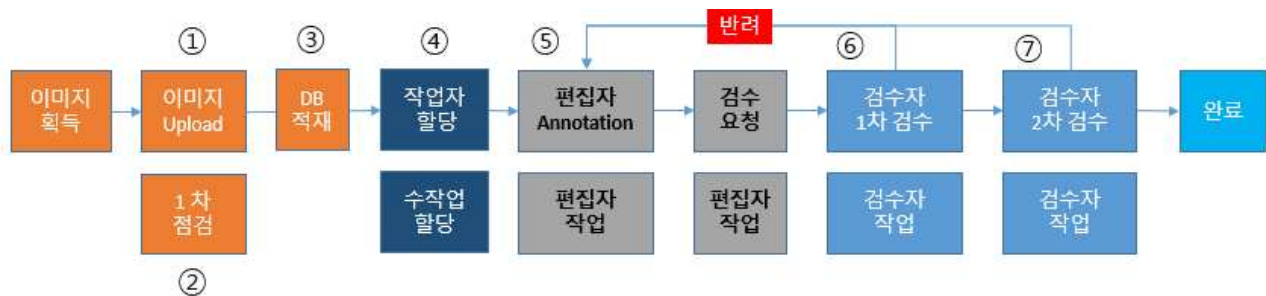


그림 5)라벨링 전체 프로세스

4.1.3 가공 업무 프로세스

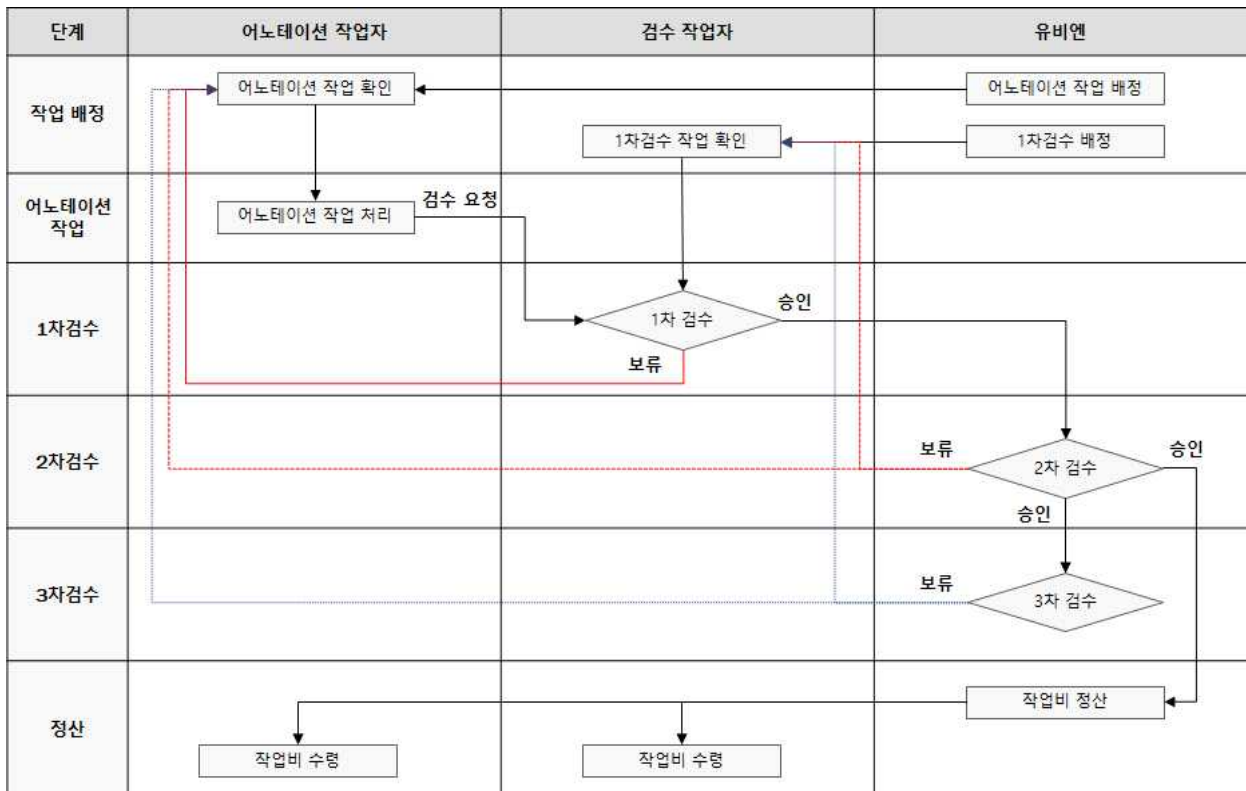
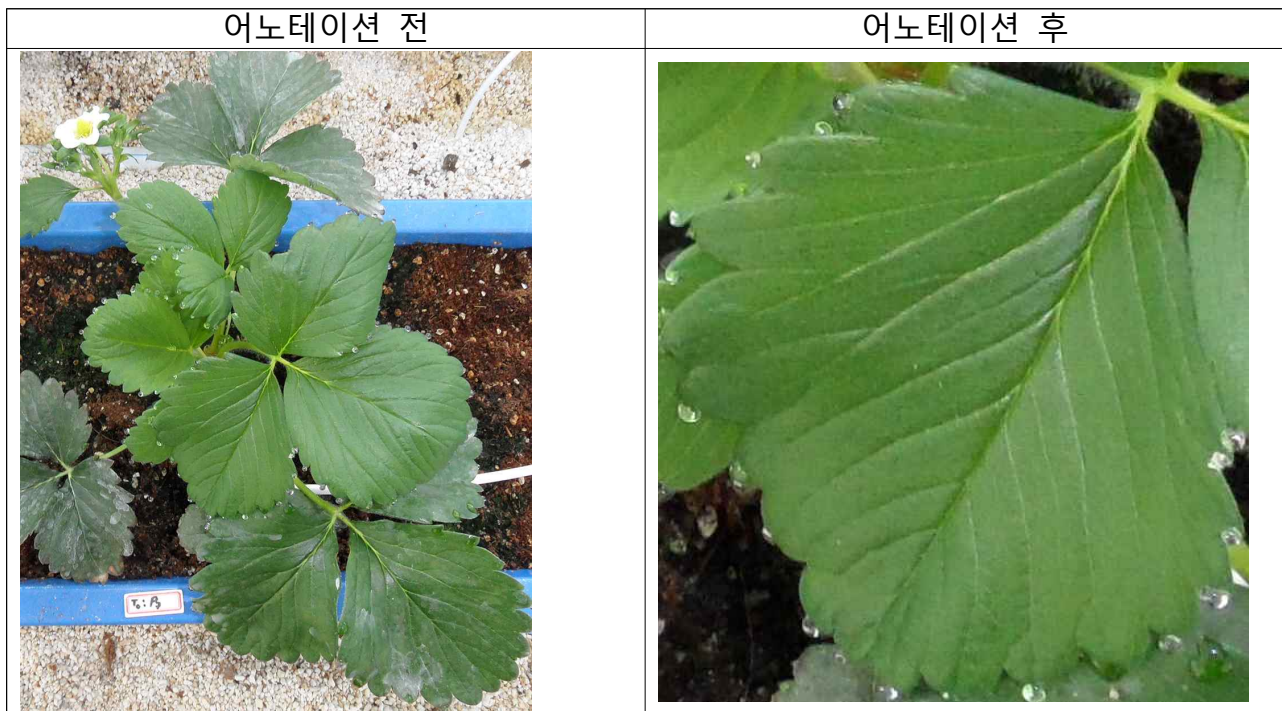


그림 6)라벨링 업무프로세스

○ 어노테이션/라벨링 예시



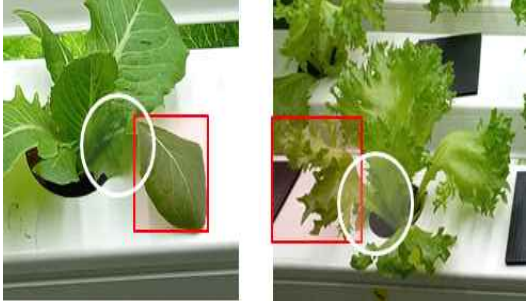
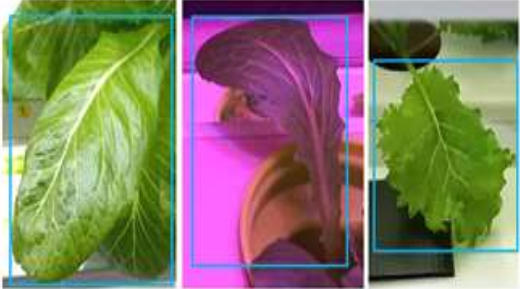
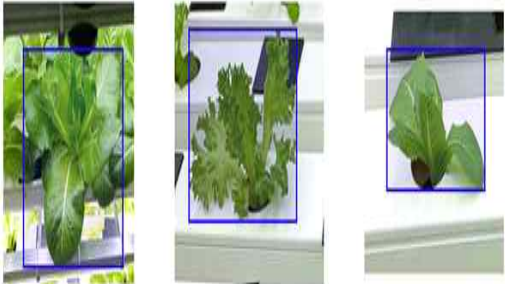
4.2. 가공 기준

○ 바운딩 박스/라벨링 기준

- Annotation의 목적이 객체에 대한 Detection 이므로 바운딩박스로 진행
- 주요 속성의 경우 객체의 분류체계 기준 폴더 구조를 통하여 자동입력
- 대상 객체의 외곽 기준 Box 처리(사람이 대상체를 인식하는 기준)
- 객체 외곽 기준 Box가 안쪽으로 들어가게 Box를 그리면 안 됨
- 객체 외곽 기준은 최대한 타이트하게 box처리 (약 5 Pixel)
- 객체 외의 사물이 함께 Box 처리 되면 안 됨
- 입력 필요 속성 처리, json 파일 내 속성 정보 확인 진행
- 농업전문가와 라벨링 전문가가 참여하여 정한 기준을 토대로 일반인들이 참여 가능

○ 바운딩 박스/라벨링 예시

- 샘플 예시

잘못된 예시	올바른 예시
 이미 20%이상 가려지면 안됩니다. 앞자루를 포함해 주세요 박스를 조금이라도 초과하면 안됩니다.  여백 흔들림 미태깅 호림  타 뿌리에서 나온 잎을 함께 지정하면 안됩니다. 박스를 조금이라도 초과하면 안됩니다.	  

- 어노테이션 규칙

주

*주



* 식재에 사용하는 뿌리가 붉은 식물개체를 일컫는 말

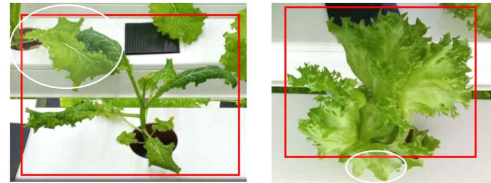
여백 없이 전체를 어노테이션 합니다.

통과 케이스



통과 케이스를 기준으로 지정하시고, 판단이 어려운 것은 하지 마세요.

반려 케이스



타 뿌리에서 나온
잎을 함께 지정하면 안됩니다.

박스를 조금이라도
초과하면 안됩니다.

통과 케이스를 기준으로 지정하시고, 판단이 어려운 것은 하지 마세요.

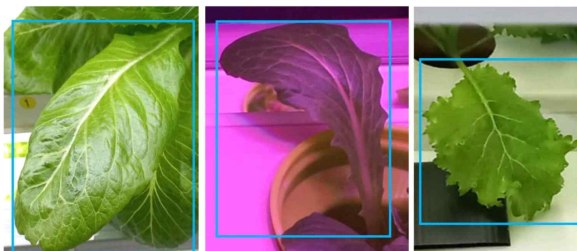
잎

잎



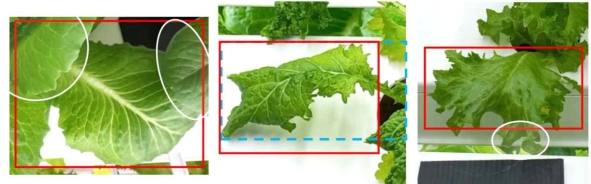
잎 전체를 지정하세요.

통과 케이스



통과 케이스를 기준으로 지정하시고, 판단이 어려운 것은 하지 마세요.

반려 케이스



앞이 20%이상
가려지면 안됩니다.

앞자루를
포함해 주세요

박스를 조금이라도
초과하면 안됩니다.

통과 케이스를 기준으로 지정하시고, 판단이 어려운 것은 하지 마세요.

4.3. 가공 규격

4.3.1 라벨링 형식 및 정의

NO	어노테이션 형태	항목 설명
1	annotations[].id	어노테이션 식별자
2	annotations[].image_id	연관 영상(동적/정적) 이미지 식별자
3	annotations[].classes	어노테이션 클래스
4	annotations[].segmentation	객체 영역 정보
5	annotations[].area	어노테이션 영역넓이
6	annotations[].bbox	어노테이션 바운딩박스 정보
7	annotaions[].isCrowd	어노테이션 툴 생성 정보

표 19) 라벨링 형식 과 정의

4.3.2 이미지 라벨링 데이터

NO	속성명		항목 설명	작성예시
	한글명	영문명		
1	이미지고유번호	images_id	데이터셋 업채류 이미지에 대한 순차적 ID 번호	"10003"
2	농장코드	Farm_id	딸기,업채류가 재배된 농장 ID	"AIF001"
3	작물개체코드	Crops_id	딸기/업채류에 배정된 작물 코드(개체) (S001~S200)	"C02_B01_001"
4	작물명	Crops	작물명	"딸기", "상추"
5	품종명	Kind_type	품종 또는 작물종	"설향""로메인"
6	파일경로	file_path	이미지파일의 경로정보	"../data/dataset/training"
7	파일종류	fext	해당 파일의 종류	"JPG"
8	이미지이름	fname	이미지의 이름	"DSC0001.JPG"
9	이미지 촬영일	Captured_date	이미지가 실제 촬영된 일시	2021.11.20. 17:08:15
10	이미지 생성일	Create_date	이미지가 변환 저장된 일시	2021.11.21. 18:09:12
11	이미지 가로	width	이미지 가로크기	"1920"
12	이미지 높이	height	이미지 세로크기	"1080"
13	작물엽	leaf	어노테이션 대상 잎	"1"or"0"
14	작물주	Plant_body	어노테이션 대상 주	"1"or"0"

표 20) 이미지라벨링데이터 정의

4.3.3 재배환경 텍스트 라벨링 데이터

NO	속성명		항목 설명	작성예시
	한글명	영문명		
1	농장코드	Farm_id	딸기/엽채류 재배 농장 코드	"f001"
2	작물코드	Corps_id	딸기/엽채류에 배정된작물코드(개체) (S001-S002)	"S001"
3	품종코드	Kind_type	품종코드	설향, 금실...
4	측정일시	Received_date	센서 측정일시	2021.09.20. 19:09:15
5	내부co2	CI	내부CO2 측정값	431 ppm
6	내부습도	Hi	내부습도 측정값	66%
7	내부일사량	SR	내부일사량 측정값	100 W/m-2·s
8	광량	IR	(수직농장) 광량측정값	300 μ mol/m-2·s
9	내부온도	TI	내부 온도 측정값	19.2도
10	양액 EC	EI	양액의 EC 측정값	0.9 ds/m
11	양액 PH	PL	양액의 PH 측정값	6.1 ph
12	양액 온도	TL	양액의 온도 측정값	22.4 도

표 21)재배환경데이터 정의

4.3.4 생육지표 텍스트 라벨링 데이터

NO	속성명		항목 설명	작성예시
	한글명	영문명		
1	작물코드	Corps_id	작물코드	"S001"
2	측정일시	measured_date	측정일시	2021.11.20.: 20:10:15
3	초장	stem_length	관부(지제부)에서 가장 긴 잎까지의 길이	"15cm"
4	엽수	leaf_cnt	한 개체 당 잎의 수 (완전 전개한 잎만 측정)	"3개"
5	엽장	leaf_length	완전히 전개한 잎을 기준으로 세 번째 잎을 선택후 중간 소엽의 가장 긴 부분을 측정	"7cm"
6	엽폭	leaf_width	완전히 전개한 잎을 기준으로 세 번째 잎을 선택한 후 중간 소엽의 가장 넓은 부분 측정	"5cm"
7	엽형지수	leaf_index	측정된 엽장의 길이/측정된 엽폭의 길이	"1.2"
8	엽병장	petiole_length	잎을 바치고 줄기에 붙어있는 자루 길이	"9cm"
9	SPAD	SPAD	완전전개한 입기준 세 번째잎 엽록소측정	
10	과중(생체중)	fr_weight	수확과의 무게/수확한 엽채류의 무게	"25g"
11	ETC	Crops_id	작물아이디	※발생시
		Create_date	수집일	
		inform	병해충 정보, 이상장애 정보	

표 22) 생육지표 라벨링데이터 정의

4.4. 가공 조직



4.5. 가공 도구

- 어노테이션/ 라벨링 도구는 오픈소스 도구를 기존 데이터 공유 플랫폼을 활용하여 통합 제공
 - Smartfarm 분석에 최적화된 빅 데이터 서비스 플랫폼인 “Smart DLSP”를 개발하여 활용
 - “Smart DLSP”은 인공지능 학습용 데이터 분석을 위한 분석환경을 포함하고, 각종 데이터에 대한 메타데이터를 관리하는 기능을 활용하여 제공.
 - “Smart DLSP”을 기반으로, 오픈소스 Annotation 관리 도구인 “Myvision.ai”를 적용하여 단기간에 관리 도구 구축 및 운영환경 전환 시행



다양한 산업의 흥미로운 데이터가 지금도 활용되고 있습니다.

그림6) 라벨링도구 접속 화면

- 라벨링 도구 웹 온라인 서비스를 활용하여 클라우드 워커를 활용한 가공 진행

○ 어노테이션 툴

- 오픈소스 어노테이션 툴인 "Myvision.ai"를 적용하여 클라우드 소싱 작업자에 제공
- 작업자의 수작업에 의한 바운딩박스 처리
- 바운딩박스 처리 결과는 바운딩박스 1개 단위로 레코드로 DB에 저장하여 관리
- DB 저장 세그먼테이션 예 :

```
{"segmentation": [767.79, 1809.66, 1261.28, 1809.66, 1261.28, 2354, 767.79, 2354], "area": 268626.34, "bbox": [767.79, 1809.66, 493.49, 544.34], "isCrowd": 0}
```

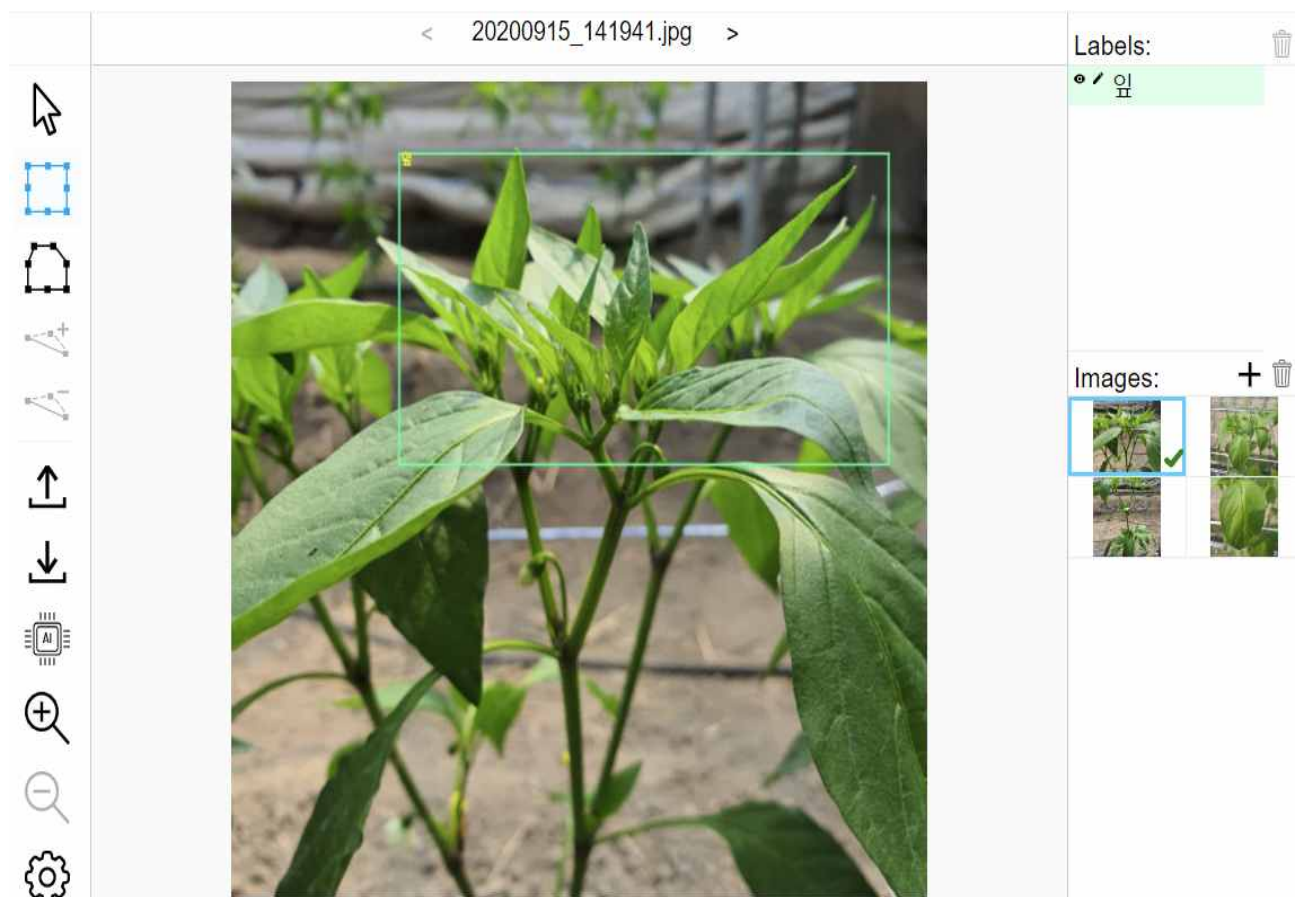


그림7) myvision 기반의 viewer 화면

5. 검수

5.1. 검수 절차 및 방법

5.1.1 검수절차

○ Annotation 데이터 품질 부문은 생육 단계별 객체 Annotation 다양성 검증과 Annotation 정보의 정확도에 대한 품질 검증으로 구성

- Annotation 정보의 다양성에 대한 품질검증은 생육 단계별 객체 Annotation (화방, 열매 등)을 검수 2단계에 걸쳐 검증
- Annotation 정보의 정확도에 대한 품질 검증은 Annotation 박스 기준으로 데이터 검수 3 단계 절차로 구성

○ Annotation 정보 정확도를 검증하기 위한 2단계 검수 절차

- 1단계. 준전문가 Review
 - 초기 어노테이션 작업자의 작업을 부분별 관리자가 평가하여, 검수 통과율이 높은 작업자를 1차 검수자(준전문가)로 지정
 - 1차 검수자는 타 작업자의 어노테이션 작업을 평가
- 2단계. 전문가 Review
 - 전문가(운영팀 검수 담당자)는 작업자가 입력한 데이터를 전수 Review를 통해 정확도 여부를 판단
- 3단계. 전문가 2차 Review
 - 전문가(운영팀 검수 담당자)는 2단계 검수를 마친 데이터 중 10%에 대해 정확도 여부 판단
- 데이터 품질 관리 구성
 - 데이터의 저장모델을 기준으로 최소값 / 최대값을 정의하고 최소값 또는 최대값의 범위를 벗어나는 경우 데이터를 재 확인할 수 있도록 조치



그림8) 품질관리 절차

5.2. 검수 기준

5.1.1 검수 기준

○검수기준은 해상도, 정확성, 완전성 및 일관성에 대해 객관적이고 정량적인 검수기준 목표를 제시

검수요소		검수 진단 내용
원시 이미지 데이터	해상도	·객체 식별이 가능한 해상도(번짐, 흔들림 등)인가
	적시성	·이미지 획득 시기는 언제인가?
학습용 이미지 데이터	정확성	·사용 목적에 부합한가?
	완전성	·필요한 데이터 식별 수준인가?
	일관성	·수집기준에 부합한가?

표23) 학습데이터 검수기준

○진단 대상 선정은 본 과제 목표 및 내용 부합성을 고려하여 아래와 같이 정의할 수 있음

데이터 구분	세부 데이터	진단 내용
기보유 데이터	영상 데이터 속성 데이터	저작권 판단 학습데이터 적용 가능 여부 검증
공개 자료	영상 데이터 속성 데이터	저작권 판단 학습데이터 적용 가능 여부 품질 검증
신규 취득 데이터	영상 데이터 학습 데이터	학습 적용

표24) 데이터선정시 고려사항

○상세 검수 기준

검사 절차	영상(동적/정적) 이미지 공통 항목	요구사항
초기 검사 (획득)	법·제도 준수	원시데이터 획득 시 관련 법·제도적 규정 등을 반드시 준수하여야 함
	사실적인 획득 환경 구성	원시데이터를 인위적인 환경과 조건 하에 획득해야 하는 경우 사실적인 획득 환경을 구성하여야 함
	데이터 동기화	다중 데이터 소스 간 정교한 동기화를 위한 절차를 마련하여야 함
	편향성 방지	데이터 편향을 방지하기 위한 절차를 마련하여야 함
(정제) 검사	정제 기준의 명확성	데이터 사용 목적에 적합한 정제 기준 수립 여부 검사
	중복성 방지	데이터 정제 후 정보 비교 후 중복도 여부 검사
	정제 작업 메뉴얼	정제 작업을 위한 메뉴얼 작성 및 관리 여부 검사
	정제 도구	정제 작업에 사용될 SW 도구를 확보 및 사용 방법을 숙지

	정제 작업 방식	데이터 특성 및 활용 목적에 맞는 적절한 정제 방식 선정 여부 및 선정 기준 타당성 여부 검사
(라벨링) 검사 2회 검수	라벨링 가이드	목적에 맞게 작성된 라벨링 가이드에 대한 타당성 여부를 검사 후 라벨링 작업자들에게 내용 가이드 전달
	어노테이션 항목	목적에 맞는 어노테이션 구성인지 여부를 검사 후 확인된 내용을 포함하도록 작업자들에게 전달
	라벨링 검사 도구	자동화 도구를 통해 검사 후 검사자가 육안으로 부적합 데이터 여부 2차 확인과 촬영된 영상(동적/정적) 이미지의 누락, 번짐 및 조건 오류를 전수 검사
최종 검사 (샘플)	부적합 판정 데이터 분포 확인	데이터의 오류율, 특성 분포 확인등 통계적 기법 활용을 통한 데이터 수집, 정제, 라벨링, 부문 최적화 및 정합률 확보 (라벨링 툴 내 통계 기능 활용)
		생육, 생육지표, 지표에 따른 이미지 정합성 최종 검수
	외부 검사자/전문가 검수	외부 검사자(TTA 등), 도메인 전문가, 데이터 요청자

표25)상세 검수기준

5.3. 검수 조직

○데이터 품질 검수조직 구성

구 분	역 할	책 임
최고의사 결정기구(주관)	·데이터에 대한 품질진단 및 개선과 관련한 최고 의사결정	·품질관리 우선순위 및 목표 설정 ·품질개선을 위한 의사결정
데이터 품질 전담 및 실무팀	·최고의사 결정권한자 결정사항 집행 ·데이터품질 의사결정 필요 사안 취합 ·품질진단 전반에 대한 실무적 책임과 의사결정 수행 ·품질진단을 위한 품질기준, 업무규칙 도출, 품질측정, 측정결과분석 및 개선활동 관리 등 수행	·구체적 품질목표 및 추진 방향 설정 ·품질진단계획 승인 및 확인 ·품질개선계획 승인 및 확인 ·품질측정 항목 관리 ·품질측정 대상 선정 ·업무규칙 도출 ·측정결과 및 개선결과 보고 ·개선활동관리, 담당자교육
데이터 품질 지원팀	·품질관리팀 지원 ·대상개체 및 시스템 인프라 관리	·품질측정 수행 지원 ·품질측정 결과분석 지원 ·품질개선활동 지원
업무팀	·1차적 품질관리 책임 수행 ·품질담당자와 협조체계 구축	·업무규칙 준수 및 점검 ·소관데이터 품질이슈 관리

표 26) 데이터 검수 조직 및 역할

5.4. 검수절차

○검수 절차 및 내용

- 작업자 중 정확도가 높은 인원을 선발 후 1차 검수자 참여 요청하여 30명 지정
- 사업 초기에는 작업자 : 검수자 비율 3:1을 유지
- 검수자의 숙련도가 올라가고 우수 작업자의 증가로 50명을 고정으로 지정
- 검수 작업을 진행하지 않기를 원하는 인원이 지속적으로 발생하여 추가 작업자 지정

○검수 작업기능

검수자에게 배정된 사진 목록

데이터 검수

날짜
 2020-10-02 ~ 2021-03-02

담당자
 라벨링 작업자

작업

품종

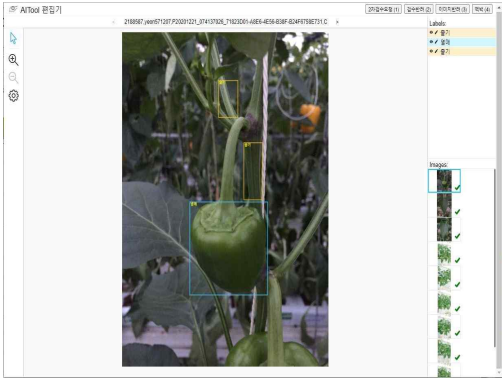
1차검수대기 : 9,544개 | 2차검수요청 : 11개 | 2차검수승인 : 168,355개 | 2차검수완료 : 390개

보기

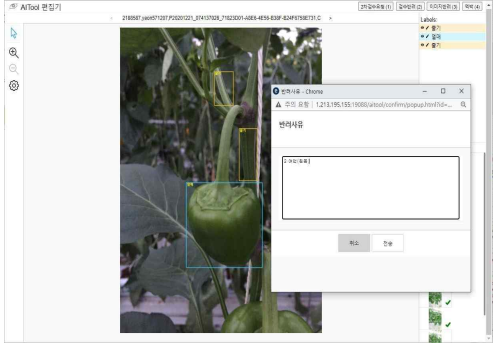
총 9,538 건 | 페이지 : 10

☐	인덱스	배정일	검수요청일	라벨링작업자	파일명	촬영거리	작업명	품종	가공수	반려율수	상태	반려사유
☐	2188587	2021-01-29	2021-01-29	조수현	P20201221_074137026_71823D01-ABE6-4E56-B38F-B24F6758E731	근거리	파프리카(불로키)	불로키	3	1	1차검수요청	1-알람-개체가 반이상 끈으로 가려져 있어서 라벨링 삭제 후 온전하게 채워주세요 아노테이션 필수
☐	2188585	2021-01-29	2021-01-29	조수현	P20201221_074115972_98449805-B919-41A4-A236-B71638ED0333	근거리	파프리카(불로키)	불로키	1	1	1차검수요청	알 아노테이션 가능
☐	2188566	2021-01-29	2021-01-29	조수현	P20201221_073917653_96585276-95F3-4686-BACC-8C9FE79FBF95	근거리	파프리카(불로키)	불로키	1	1	1차검수요청	아노테이션 가능

검수자 아노테이션 도구(AI-TOOL) 화면



수정이 필요한 사진에 반려(수정요청)사유 작성 화면



○아노테이션 작업 단계별 검사 화면

작업자에게 배정된 사진 목록

데이터 가공

배정일
 2020-10-01 ~ 2021-03-02

작업

품종

작업대기 : 481개 | 1차검수요청 : 922개 | 2차검수요청 : 89개 | 반려 : 3,125개 | 완료 : 4,319개 (번식 : 6,724개)

전망

총 481 건 | 페이지 : 10

☐	인덱스	배정일	생성자	파일명	촬영거리	작업명	품종	가공수	상태
☐	2137387	2021-01-31	김병현	20201130_135223	근거리	장미	치즈타르트	0	작업배정완료
☐	2137384	2021-01-31	김병현	20201130_135318	근거리	장미	치즈타르트	0	작업배정완료
☐	2137380	2021-01-31	김병현	20201130_135313	근거리	장미	치즈타르트	0	작업배정완료
☐	2137379	2021-01-31	김병현	20201130_135309	근거리	장미	치즈타르트	0	작업배정완료

- 29 -

1차 검수 요청 후 검수자가 수정 요청 보낸 사진 목록

검수 반려

반려일

2021-01-02

작물

품종

검색

작업대기: 481개 | 1차검수요청: 922개 | 2차검수요청: 89개 | 반려: 3,125개 | 완료: 4,319개 (번식: 6,724개)

보기

총 517건 | 페이지: 10

☐	원메스	작물명	작물명	품종	원거리	가공수	상태	배정일	1차검수요청일	반려일	반려원수	반려사유
☐	1430890	DSC02632	배추박	진한	근거리	3	2차검수반려	2020-12-10	2020-12-10	2021-01-08	0	열매 찢힘/ 열매받침 호랑에서 어노테이 신해주세요
☐	1304229	DSC09267	딸기	선허	원거리	1	1차검수반려	2020-11-30	2020-12-01	2021-01-04	2	화방 반대에서 포병침까지 채주세요
☐	1304215	DSC09252	딸기	선허	원거리	1	1차검수반려	2020-11-30	2020-12-01	2021-01-04	1	열매
☐	1304213	DSC09250	딸기	선허	원거리	1	1차검수반려	2020-11-30	2020-12-01	2021-01-06	1	3월 이월됨

어노테이션 도구(AI-TOOL)을 활용한 작업화면

어노테이션 라벨링 작업화면

작업 완료전 검수요청 화면

반려(수정요청)된 사진 재작업 화면

5.5. 검수 도구

- 수집단계에서 라벨링 항목을 입력함에 있어 설정된 유효값 유형만 입력될 수 있도록 도구 내 Error 방지 기능을 탑재
- 그룹별, 작물별 작업자에 있어서는 해당 작물에 대한 기본정보는 수동 입력이 아닌 자동입력을 통해 오류 라벨링을 방지
- 속성 값이 일괄 변경 또는 일부 작업량 범위 안에서 자동변경 될 수 있는 기능 탑재

5.6. 기타 품질관리 활동

- 데이터셋의 품질을 보장하기 위하여 이미지 데이터셋 수집 단계별 검증, 데이터 셋 유효성 확인, 품질보증 활동의 3가지 활동을 수행, 검수진단 과정에서 1차 육안검수, 교차검수를 통해 진행하며, 샘플검수를 통한 2차 검수 진행

- 30 -

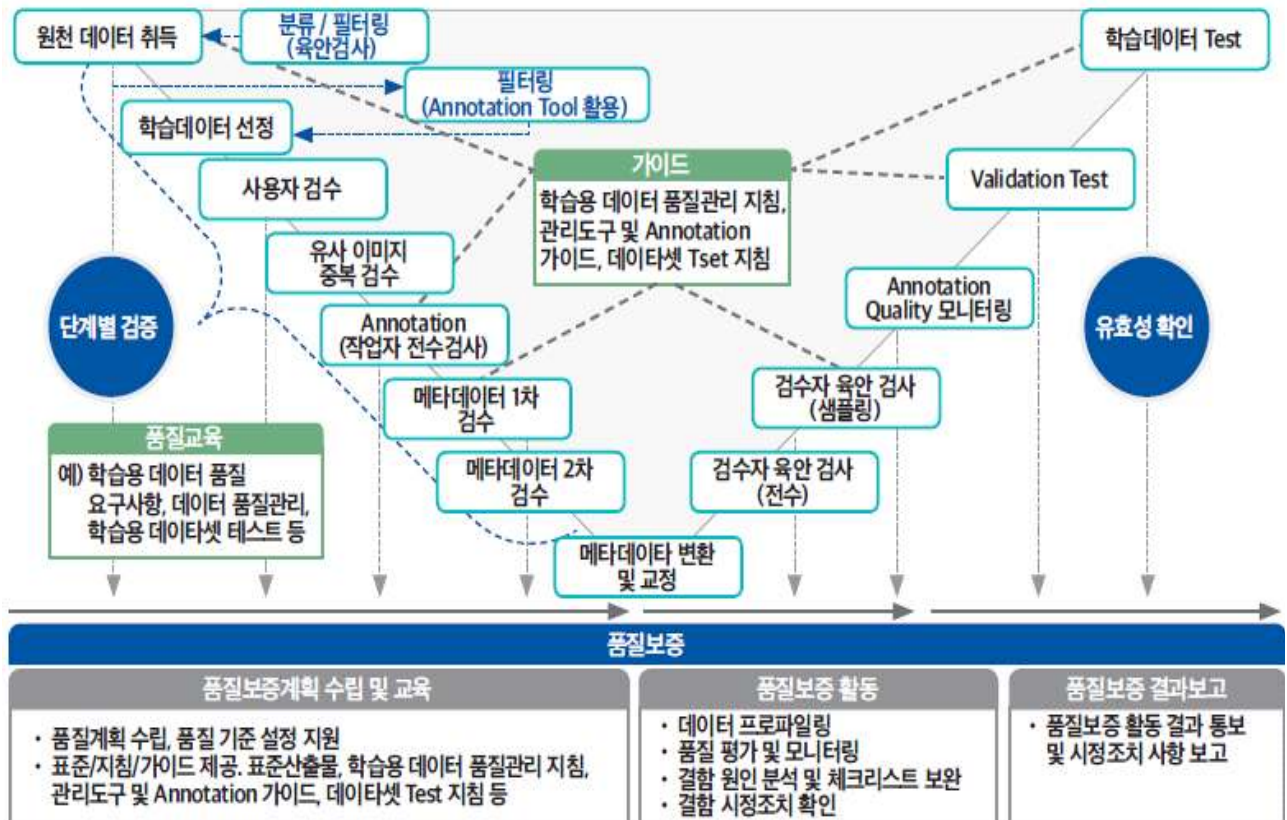


그림 9) AI 데이터셋 품질관리 활동 프로세스