
데이터 설명서

[2-15] 고품질 과수작물 통합 데이터


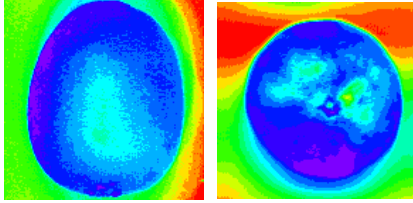

2022년 01월

지디에스컨설팅그룹 콘소시엄

(1) 데이터셋 설계

(가) 객체 클래스 정의

□ 구축데이터 명세

항목	내용	형식	예시
이미지 데이터	- 과수작물 과실 병충해 이미지	JPG	
	- 과수작물 초분광 이미지	HDR	
기상환경 데이터	- 장비로부터 수집된 환경 데이터 - 수치모델 데이터와 결합	CSV	
메타데이터	- 이미지 파일 정보 - 어노테이션 정보 - 환경 정보	JSON	<pre> { "interface" : { "id" : "000001", "filename" : "000001_Panorama.jpg", "path" : "00/panorama_image/", "resolution" : [4000, 2000], "location" : "서귀포", "datetime" : "2019-12-16 10:34:19.0", "annotations" : [{ "annotation_id" : 1593465, "annotation_type" : "bbox", "class_code" : "0203000100000", "class_name" : "감귤/잎", "coord_xy" : [[2191, 2222], [905, 925]], "meno" : "" }, { "annotation_id" : 1593470, "annotation_type" : "polygon", "class_code" : "0204001300000", "class_name" : "감귤/잎", "coord_xy" : [[107, 152, 225, 189, 105, 152, 225, 189, 105, 150, 225, 189, 103, 149, 224, 188, 103, 149, 238, 188], "meno" : "" }] } </pre>

□ 클래스 설계

- 최상위 대분류는 작물명(감귤, 키위)
- 중분류는 병충해명
예) 정상, 궤양병, 꺾임애, 진딧물, 점무늬병, 과실무름병, 총채벌레
- 세부 클래스 체계는 과수명/병충해명으로 구성함

대분류/중분류 (작물코드)	중분류 (병충해코드)	CLASSCODE
감귤(01)	정상	감귤_정상
	궤양병	감귤_궤양병
	꺾임애	감귤_꺾임애
	진딧물	감귤_진딧물
키위(02)	정상	키위_정상
	점무늬병	키위_점무늬병
	총채벌레	키위_총채벌레
	과실무름병	키위_과실무름병

□ 데이터 설계

구분	속성명	타입	필수여부	설명	범위
1	Info 기본 정보				
1-1	IMAGE_FILE_NM	string	Y	파일명	HF01_01FT_000001
1-2	IMAGE_FILE_TY	string	N	종류	jpg, hdr
1-3	STRE_COURS	string	Y	저장폴더	\1.Training\원천데이터 \01.감귤_TS \1.Training\라벨링데이 터\01.감귤_TL
1-4	RSOLTN	string	N	해상도	(1920,1080) or (1440,1080)
1-5	IMAGE_POTOGRF_DT	string	Y	시간정보	YYYY:MM:DD HH24:MI:SS
1-6	CMRA_INFO	string	N	카메라정보	“SAMSUNG”
1-7	F_STOP	string	N	조리개값	1.2 ~ 22
1-8	FILM_SPD	string	N	감광속도	25~19200
1-9	FLSLT_USE_AT	string	N	플래시사용여부	1~2 (1:사용,2:미사용)
1-10	LCINFO	string	N	위치정보	F01 ~ F50
1-11	IMAGE_OBTAIN_P	string	N	필드타입	“노지” or “온실”

구분	속성명	타입	필수여부	설명	범위
	LACE_TY				
1-12	GRWH_STEP_CODE	string	N	생육단계코드	1-9
1-13	BRX	number	N	*당도	0~20
1-14	DBYHS	string	N	병해충 코드	00~06
1-15	OCPRD	string	N	병해충 발생시기	01-01 ~ 12-31
1-16	SPCIES_NM	string	N	품종명	“온주”
2	Annotations 라벨링정보				
2-1	ANTN_ID	string	N		“1609”
2-2	ANTN_TY	string	N	어노테이션 타입	“polygon”
2-3	OBJECT_CLASS_CODE	string	N	객체 클래스 코드	감귤_정상, 감귤_궤양병, 감귤_굴응애, 감귤_진딧물, 키위_정상, 키위_점무늬병, 키위_총채벌레, 키위_과실무름병
2-4	ANTN_PT	string	N	폴리곤 정보	[818 763 726 726 770 837 928 1007 1071 1165 1238 1297 1316 1307 1284 1243 1158 1099 1010 952 889],...
3	Environment 환경정보				
3-1	ENVRN_INFO_FILE_NM	string	N	파일명	환경정보 파일명
3-2	MSRG_DATETM	string	N	시간정보	YYYY-MM-DD HH24:MI:SS
3-3	DATA_TY	string	N	데이터 형태	OBS/SIM
3-8	SOLRAD_QY	number	N	일사량	0~2000
3-9	AFR	number	N	강우	0~600
3-10	TP	number	N	온도	-30~70
3-11	HD	number	N	습도	0~100
3-13	MESURE_SNS_SPEC	string	N	측정센서 스펙	
4	Hyperspectral 이미지 정보				
4-1	HS_NON_BRX	number	N	비파괴검사 당도	0~100
4-2	HS_NON_DM	number	N	비파괴검사 건물률	0~100
4-3	HS_BRX	number	N	파괴검사 당도	0~100
4-4	HS_DM	number	N	파괴검사 건물률	0~100

구분	속성명	타입	필수여부	설명	범위
4-5	HS_CHROMA	number	N	파괴검사 색도	0~500
4-6	HS_HN	number	N	파괴검사 경도	0~100
4-7	HS_ACIDITY	number	N	파괴검사 산도	0~100

□ 데이터 표준화

- 메타데이터 표현 방식은 농진청 자문을 통해 표준화
- Microsoft COCO 기준을 바탕으로 필요속성을 반영하고, 영문 속성명은 행안부 공통표준용어와 농업진흥청의 표준용어를 준수하여 재설계하여 적용

(2) 원시데이터 수집

(가) 데이터 수집 요약

항 목	내 용
대상 데이터	- 감귤/키위 병충해 이미지 - 감귤/키위 초분광 이미지 - 환경/기상 데이터 (온도, 습도, 토양지표 등)
수집기간	- 2021. 08. 01.~ 2021. 12. 03.
장소	- 제주지역 감귤 및 키위 재배지 (온실 및 노지 재배지 대상)
담당자	- 사전 교육을 받은 클라우드 워커 - 현장 관리자에 의한 수집공정 관리
수집방법	- 감귤/키위 병충해 이미지와 초분광이미지는 관련장비를 활용하여 직접 촬영 - 과수작물 2종에 대한 병충해별, 부위별, 시기별, 농장별 소재 확보하여 촬영 진행 - 환경/기상 정보는 수집된 Observation 데이터를 이용하여 100m 해상도의 환경/기상 Simulation 데이터 생성 . 지역별 24시간의 예측 데이터 생성 . 200,000건 이상의 환경/기상 데이터 생성(4개월간 5~30분 간격으로 데이터 생성)

(나) 병충해 이미지 데이터 수집

□ 병충해 선정

- 어떤 과실이든 해마다 유행하는 병충해가 달라질 수가 있기 때문에 감귤과 키위 각각 병충해 전문가를 섭외하여 병충해를 선정
- 병충해의 특성상 여러 가지 병충해가 고르게 유행하기 보다는 한가지 병충해가 dominant할 경우 다른 병충해의 발생이 억제되는 현상이 있음
- 금년의 경우, 때 이른 장마와 9월에도 비가 많이 내렸던 날씨 영향으로 감귤의 경우 점무늬병의 발생이 많았음
- 키위 병충해 역시 점무늬병의 발생이 dominant하였으며, 총채벌레 데이터도 확보하기 용이하였으나, 세 번째로 유행하는 병충해를 찾기가 어려웠음

과실	병충해명	특징
감귤	궤양병	<ul style="list-style-type: none"> - 반점 형태로 외관을 해치고 심할 경우에는 잎이 뒤틀리며 낙엽이 되며 새순의 경우 순 전체가 죽고 과실은 낙과 될 수 있음 - 감염 7~10일 후에 첫 병징이 보이기 시작하며 초기 증상은 주위가 황화된 매우 작은 반점(직경 약 0.3mm ~ 0.5mm)으로 부터 시작하여 병이 진전되면서 점차 그 크기가 커지고 모양도 원형에서 불규칙한 모양으로 발전 - 잎의 양면, 특히 잎의 뒷면이 부풀어 오르고 나중에 이 부분이 코르크화 되며 분화구 모양이 됨 - 발생 초기에 연한 노란색의 기포 같은 것이 표면에 붙어 있는 모양이며 시간이 경과하면 점차 진한 갈색의 뚜렷한 병반을 형성
	꿀응애	<ul style="list-style-type: none"> - 알은 구형으로 크기는 0.13 ~ 0.15mm이며 홍색, 약충은 0.2 mm의 크기로 담홍색이고 3쌍의 다리, 전약충은 0.3mm전후로 유충과 비슷하지만 다리가 4쌍, 후약충은 체장이 0.3 ~ 0.45mm로 전약충과 거의 비슷 - 약충과 성충이 잎과 과실에 기생하여 조직 내의 세포액이나 엽록소를 흡수 - 피해 받은 잎은 잎 표면에 바늘로 찌른 듯한 하얀 반점이 나타남 - 피해받은 엽록소가 파괴되어 동화작용이 저하되고 흡즙한 상처부위부터 수분이 증산되어 생리기능이 현저히 떨어짐
	진딧물	<ul style="list-style-type: none"> - 조팝나무 진딧물과 마찬가지로 역시 주변의 기주식물에서 알 상태로 월동 - 감귤원 주변에서 월동한 알이 봄에 부화한 후 감귤원에 봄순이 발생하기 시작하면서 유시암컷성충이 감귤원으로 날아와 증식 - 일반적으로 5월 하순부터 발생하기 시작하여 봄순에서는 5

과실	병충해명	특징
		<p>월 하순 ~ 6월 상순에 그리고 여름순에서는 7월 하순 ~ 8월 상순에 밀도가 가장 높는데, 전반적으로 조팝나무 진딧물보다 발생량이 다소 적으며, 특히 여름순에서의 발생이 더 적음</p> <p>- 식물체에 달라붙어서 즙액을 빨아먹고 기생</p>
키위	점무늬병	<p>- 키위 잎에서 가장 많이 나타나는 병 중의 하나이나 피해는 크지 않음</p> <p>- 6월 이후부터 가을까지 계속해서 발생하며, 주로 강우가 많은 여름철에 많이 발생</p> <p>- 처음에는 작은 회갈색의 점무늬가 형성되고, 점차 진전되면서 원형 내지 불규칙한 병반으로 확대</p> <p>- 오래된 병반상에는 소립점(병자각)이 형성되며, 탄저병과 흡사하여 구분하기가 매우 어려움</p>
	총채벌레	<p>- 암컷성충의 크기는 약 1.4 ~ 1.7mm정도이며, 몸색은 황색에서 갈색으로 다양하고, 더듬이 첫마디는 황색, 둘째 마디는 갈색, 3 ~ 5번째 마디는 황갈색이고 나머지는 갈색</p> <p>- 수컷성충의 크기는 약 1.0 ~ 1.2mm정도이며 황색. 성충의 날개는 2쌍으로 되어 있으나 잘 발달하지 못하였고, 총채모양을 하고 있어서 잘 날지 못하므로 툭툭 튀듯이 이동</p> <p>- 피해는 수관의 하부보다는 중앙부 이상의 과실에서 피해가 많은 경향이며, 과실이 피해를 받으면 초기에는 가해 받은 부분이 구름 모양의 백색반점이 형성되고, 시간이 경과함에 따라 피해부위가 갈변되어 상품가치를 떨어뜨림</p>
	과실무름병	<p>- 병든 부위에서 병자각 형태로 월동하여 다음해 1차전염원이 됨</p> <p>- 병자각에서 월동된 병포자는 비바람에 의하여 전반, 과경을 침입하여 갈변시키고, 점차 진전되면 생육 중 또는 저장 중에 연부증상을 일으키는데, 특히 저장중의 피해가 큼</p> <p>- 담색의 병반으로 나타나고, 주위는 불분명한 농녹색을 띠며, 진전되면 무른 상태로 썩음</p>

※ 출처 : 국가농작물병충해관리시스템

□ 병충해 데이터 촬영

- 병충해에 걸린 개체를 채취하여 A4용지 등의 흰색 배경으로 실내 촬영을 진행함
- 일부 병충해는 채취하지 않고 매달린 상태로 촬영을 진행
- 채취하여 촬영하는 경우, 운반 및 보관 이슈를 고려하여 현장에서 직사광선을 차단시키는 촬영세트를 구성하여 실내촬영의 효과를 얻을 수 있었음



<과수작물 병충해 이미지 현장촬영 모습>

- 데이터 수집매뉴얼 작성 : 피사체 전체가 나오도록 정확한 화면 구도와 비율에 맞게 촬영하고, 일관된 이미지 품질 및 수량 범위를 위해 취득 품종별로 높이, 각도, 거리, 조명 등을 매뉴얼로 작성
- 스마트폰에 ‘오픈 카메라’ 라는 카메라 전용앱을 설치하여 촬영함으로써, 카메라 설정의 일관성을 확보함 (해상도 1920 X 1080, 중횡비율 4:3, 지오타킹 등)

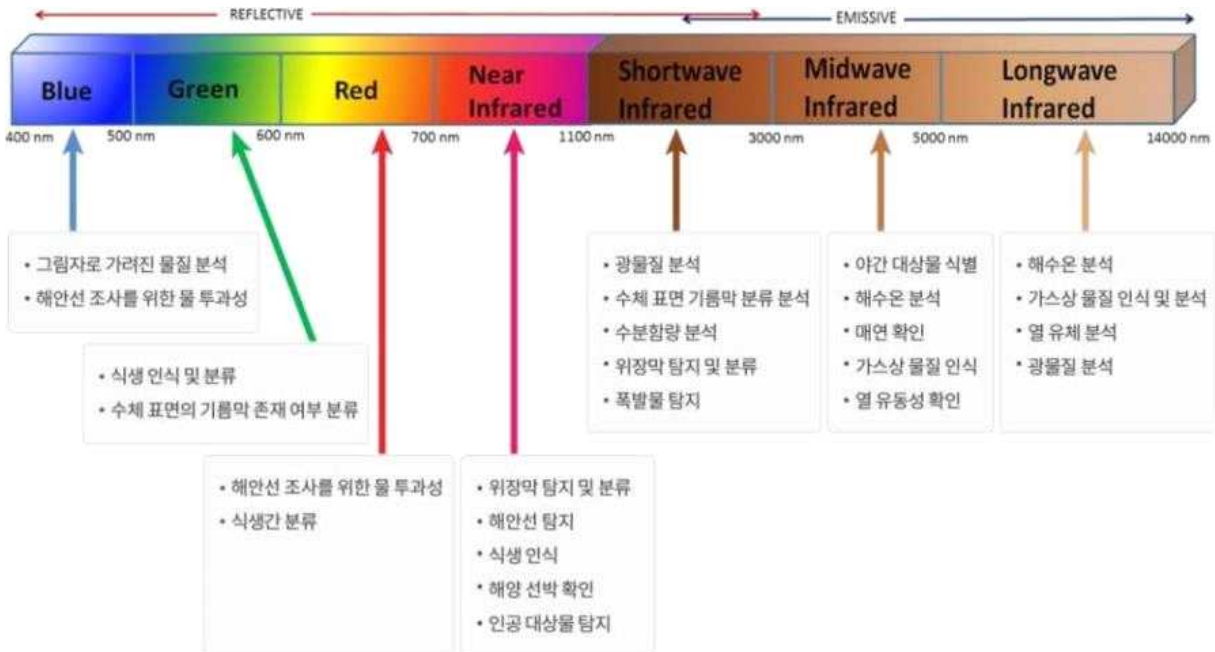


<오픈카메라 앱을 통해 촬영 설정 일관성 확보>

(라) 초분광 데이터 수집

□ 초분광 데이터 개요

- 초분광은 수백 개의 분광채널을 통해서 대상물의 특성을 나타내는 연속적인 스펙트럼(spectrum)을 수집하는 센서를 이용한 촬영기법을 의미하며, 각각의 화소는 대상체의 연속적인 분광스펙트럼을 가지고 있음



<초분광 데이터 대역별 활용범위>

- 과수작물에 대한 데이터의 활용목표는 크게 생산량 예측과 품질 예측 2가지로 나눌 수 있음
- 초분광 드론 촬영을 통해 과수원 전체에 대한 생산량 예측을 시도한 사례는 있으나, 이는 초분광 데이터 특정 밴드 데이터로부터 과실을 count하는 방법으로서 농민들이 농장을 운영하는 데 크게 도움이 되는 방법은 아님
- 본 과제에서는 개별 과실에 대한 초분광 데이터를 기반으로 품질을 예측하는 것으로 활용 목적을 설정함
- 이를 위해 감귤과 키위 과실에 대해 초분광이미지를 촬영하고 당도 등의 과실 품질을 측정하였음

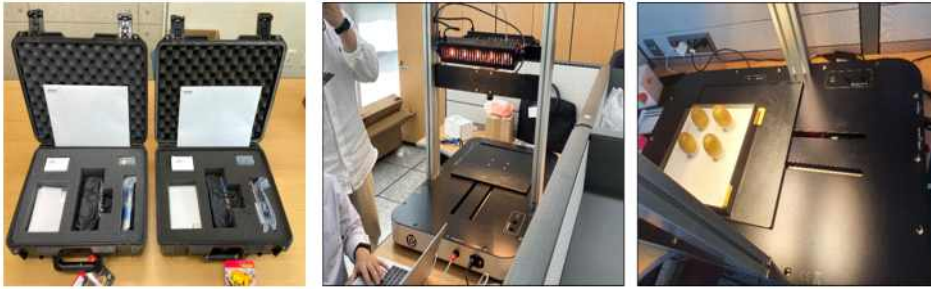
□ 초분광 카메라 Specification

- 본 과제에서는 총 3대의 초분광 카메라를 사용하였으며, 이중 2대는 본 과제 예산으로 구매하였고, 일정만회를 위해 추가로 1대를 임대하여 활용
- 본 과제 활용목적인 과실 품질 예측에 활용될 수 있도록 400~1,000nm 대역의 카

메라를 사용하였음

항 목	사 양
Wavelength band	400 - 1000 nm
Spectral resolution FWHM	7nm
Spatial Sampling	512 pix
Spectral bands	204
F/number	1.7
Peak SNR	> 400:1
FOV	31 x 31 deg
Viewfinder camera	5 Mpix

<초분광카메라 Specification>



<본 과제에서 사용된 초분광카메라>

□ 초분광데이터 촬영



<초분광 데이터 수집 프로세스>

- 초분광카메라는 빛에 매우 민감하기 때문에 현장촬영을 통해서는 품질예측에 활

용할 수 있는 고품질의 데이터를 확보하기는 어려워 전량 실내 촬영을 진행

- 농장주에 의해 과실을 획득한 후 클라우드워크 수거팀이 농장을 방문하여 수거하며, 수거 시점에 열매별로 농장, 품종, 획득일시 등을 라벨링
- 클라우드워커 촬영팀에 의해 초분광 촬영을 진행한 후, 비파괴검사와 파괴검사의 방법을 병행하여 과실의 품질을 측정
- 측정된 품질은 각 이미지에 속성으로 라벨링



<초분광데이터 촬영 현장사진>

(마) 기상환경 데이터 수집

□ 기상데이터 수집

- 제주대학교에서 보유한 AWS(자동기상관측장비) 20곳으로부터 기상정보를 수집하고, 기상청 데이터는 공개된 데이터를 활용



<제주지역 AWS(자동기상관측장비) - 초록색: 제주대 보유, 파란색: 기상청>



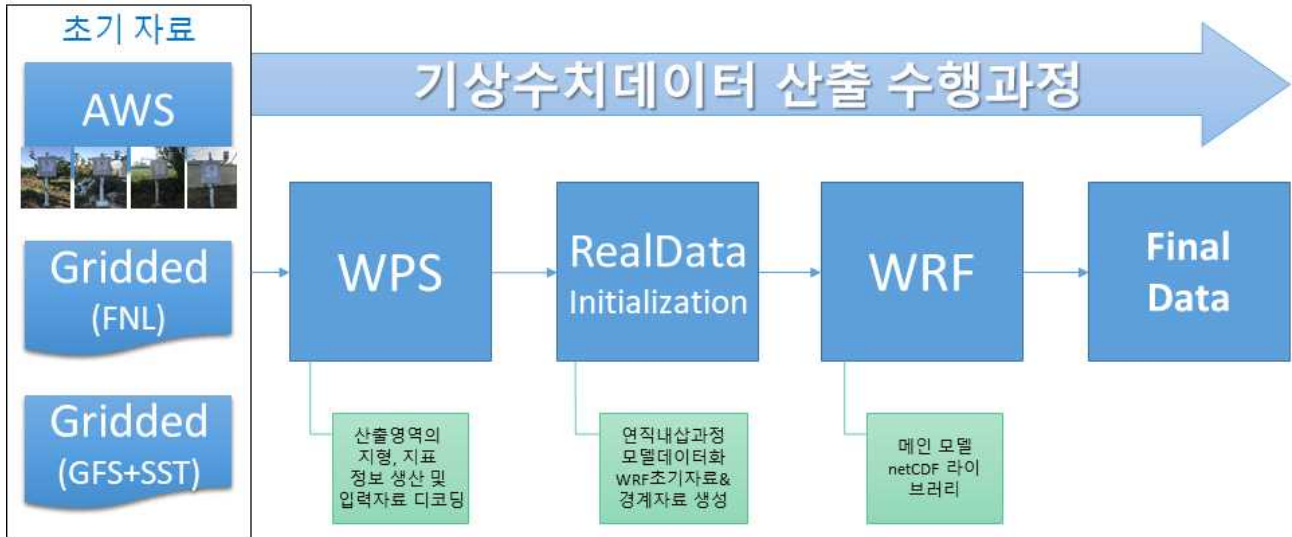
<제주지역에 설치된 AWS(자동기상관측장비) 현장사진>

□ 기상데이터 수치모델링

- 수집된 AWS 데이터는 FNL/GFS/SST 등의 글로벌 기상자료와 병합하여 WRF 수치

모델링 프로세싱을 거쳐 최종 기상환경 데이터를 구축

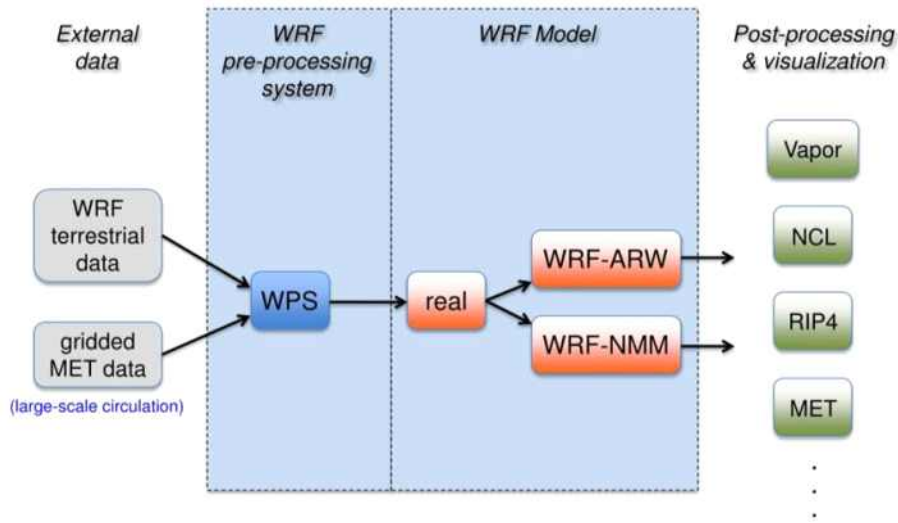
※ WRF 모델(Weather Research and Forecasting Model, 중규모 기상 수치 모델)은 미국 국립대기환경연구소를 중심으로 개발된 수치모형으로서 자료동화, 예보 연구, 모델간 결합, 실시간 예보 등에 활용



<기상데이터 WRF 수치모델링 프로세스>

- 기상데이터 수치모델링은 단기간에 강력한 서버 리소스를 필요로 하는 작업이며, 본 과제에서는 NBP 클라우드 기반으로 진행됨
- WRF 수치모델링에 사용된 데이터는 다음과 같음
 - . AWS : 제주 지역 AWS로부터 수집된 데이터
 - . FNL : 미국 국립대기환경연구소의 Final Global Data Assimilation System 자료
 - . GFS : Global Forecast System 데이터
 - . SST : Sea Surface Temperature (해수면 온도) 데이터
- 1단계 WPS Processing
 - . 초기자료 검토 (AWS데이터, FNL, GFS, SST)
 - . 지형자료를 이용하여 해당 실험 영역의 지형, 지표 정보 생산
 - . 중간단계 파일 생성 : 입력자료 구축을 위한 자료 디코딩
- 2단계 Real Data Initialization
 - . 연직내삽 : 중간단계파일(결과물) 활용
 - . 모델레벨로 변환 -> WRF 초기자료& 경계자료 생산
- 3단계 WRF Processing
 - . 메인 모델링 과정
 - . nc파일형태 -> netCDF 라이브러리 활용

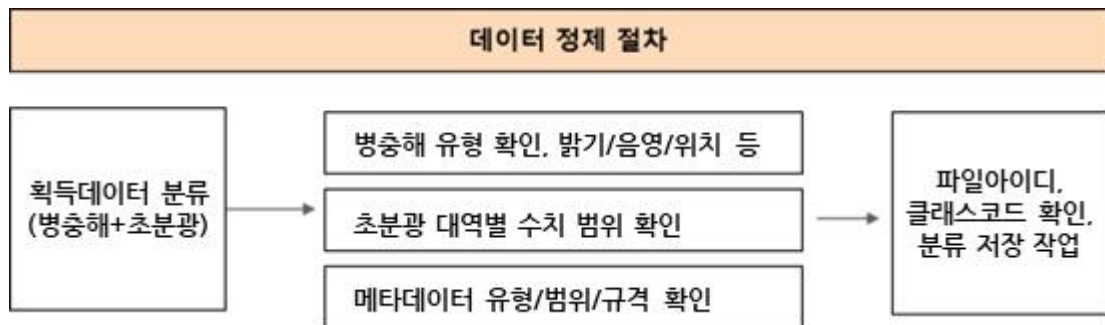
. 분석 및 Final Data 산출



<WRF Flow Chart>







(3) 데이터 정제

(가) 데이터 정제 절차



- 어노테이션 단계에 들어가기 전에 학습용 데이터로 적합한 데이터를 선별하고 처리하는 정제 프로세스를 획득데이터 유형별로 수행
- 과수작물 데이터는 병충해이미지와 초분광데이터로 구성되어 있어서 데이터 정제 작업도 각 유형에 맞게 진행
- 병충해 데이터는 병충해 유형이 맞는지 확인하고 밝기, 음영, 위치, 해상도 등이 규격에 맞게 수집되었는지 확인
- 또한 과실, 잎의 이미지를 분류하여 정제하고 해당 객체의 차지하는 비율과 이미지 내 객체 외 여백이 과다하게 많은 경우 등을 점검
- 초분광 데이터는 RGB상태에서 이상이 없는지 기본 점검 후, 각 대역별로 수치의 이상값 등을 확인
- 기타 AI 학습용 데이터의 정확도의 대부분은 획득데이터의 품질에 좌우되는 만큼 고품질의 원천데이터를 확보할 수 있도록 필요한 정제 작업을 수행

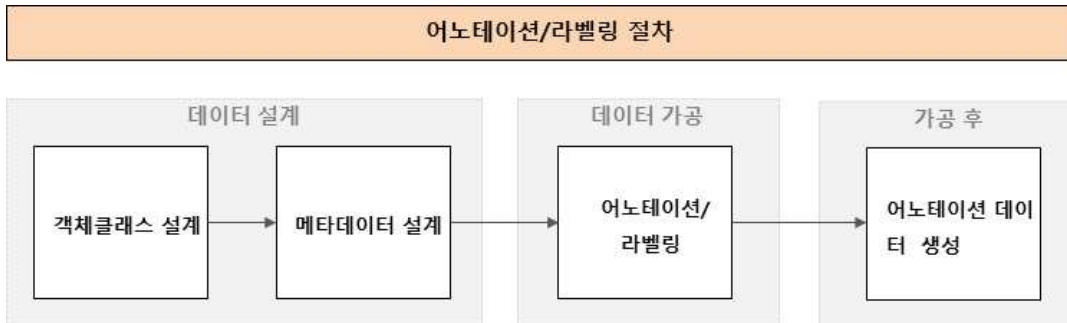
(나) 이미지 데이터 정제기준

구분	적합	부적합
밝기		
	<ul style="list-style-type: none"> - 객체 전체의 밝기가 균등함 - 객체를 식별하기에 적절한 밝기(색농도, 색상, 질감이 실물과 유사한 정도) 	<ul style="list-style-type: none"> - 지나치게 어두움(노출부족) - 지나치게 밝음(노출과다) - 객체의 부분의 밝기 차이가 심함
음영		
	<ul style="list-style-type: none"> - 객체 주변 배경의 그림자가 거의 없음 - 객체와 주변 배경을 구별할 수 있는 윤곽선(색상, 색 농도 등)이 명확 	<ul style="list-style-type: none"> - 객체와 주변 배경의 그림자를 구분 또는 구별할 수 없음 - 객체와 그림자의 색상 및 색 농도가 비슷
위치		
	<ul style="list-style-type: none"> - 객체가 이미지의 중앙에 위치함 - 검수도구(SW)의 가이드 박스의 중앙에 있는 경우 	<ul style="list-style-type: none"> - 객체가 이미지의 중앙을 벗어난 경우 - 검수도구(SW)의 가이드 박스의 중앙을 벗어난 경우

구분	적합	부적합
초점		
	<ul style="list-style-type: none"> - 객체에 초점이 맞음 - 객체 표면의 질감이 식별됨 	<ul style="list-style-type: none"> - 객체에 초점이 어긋남 - 객체 표면의 질감을 식별할 수 없음

(4) 데이터 가공

(가) 데이터 가공 절차



- 본 과제의 데이터 가공은 병충해 이미지에 대한 어노테이션과 라벨링으로 분류됨
- 어노테이션은 병충해 객체를 포함하는 도형을 그리는 작업이며, 라벨링은 해당 이미지 또는 도형에 대해 속성을 입력하는 작업을 말함
- 데이터 가공을 통해 이미지와 메타데이터를 생성하여 정해진 규격에 맞게 저장

(나) 병충해 데이터 어노테이션

- 어노테이션은 AI모델 학습목표(Classification)에 맞게 학습용 데이터셋 활용 전 이미지의 대상 객체와 배경의 분리가 필요한 경우(현장 촬영)에 한하여 구획함
- 채집하여 하얀 배경으로 촬영한 이미지의 경우는 AI학습에 그대로 활용할 수 있기 때문에 별도의 어노테이션을 하지 않음

- 현장사진의 경우 배경과 대상 객체의 구분이 필요하기 때문에 해당 객체를 폴리곤으로 구획하고 해당 어노테이션 정보를 라벨링
- 어노테이션 도형 타입은 Polygon으로 하며, 추후 데이터의 활용도를 높이기 위하여 Fine Segmentation 방식으로 진행
- 어노테이션은 병진이 아니라 과실/잎 객체를 대상으로 하며, 해당 객체를 완전하게 포함하도록 구획

분류	이미지 예시	어노테이션 방법
채집 촬영	 <p data-bbox="453 837 592 869"><채집촬영></p>	<ul style="list-style-type: none"> ▪ 채집 대상은 어노테이션 구획하지 않음 ▪ 라벨 데이터(작물명, 정상, 병충해명, 부위명 등) 메타데이터에 정의한 값 기록함
현장 촬영	 <p data-bbox="400 1198 651 1229"><현장촬영 - 과실 ></p>	<p data-bbox="791 920 1038 952">【 Annotation 기준 】</p> <ul style="list-style-type: none"> ▪ 병충해에 걸린 객체를 대상으로 Polygon 도형으로 구획 ▪ 대상 객체 전체를 포함하도록 하며, Fine Segmentation방식으로 구획
	 <p data-bbox="416 1579 639 1610"><현장촬영 - 잎 ></p>	<p data-bbox="791 1162 1102 1193">【 Annotation 작업 방법 】</p> <ul style="list-style-type: none"> ▪ 병진이 아닌 대상 객체 이미지 영역 전체가 포함되도록 작업 ▪ 폴리곤 어노테이션으로 작업 ▪ 라벨 클래스는 작물명, 부위명, 정상, 병충해명(3종)으로 구별함
	<p data-bbox="791 1460 1187 1491">【 Annotation 작업 시 주의사항 】</p> <ul style="list-style-type: none"> ▪ 객체 이미지 영역 내부를 침범하여 일부가 어노테이션 되는 경우 오류로 처리 	

(다) 데이터 라벨링

- 데이터 라벨링 작업에서는 병충해 유형의 클래스를 이미지 데이터에 태깅하고 메타데이터 속성 정보로 저장
- 클래스 코드 이외의 정보는 엑셀 등으로 입력한 후 설계내역에 맞게 메타데이터로 구축

분류	라벨링 정보	예시
기본 정보	파일명	HF01_01FT_000001
	종류	jpg, hdr
	저장폴더	\1.Training\원천데이터\01.감굴_TS
	해상도	1920,1080
	시간정보	2021:10:15 15:32:25
	카메라정보	SAMSUNG
	조리개값	1.2
	감광속도	19200
	플래시사용여부	1 (1:사용,2:미사용)
	위치정보	F01 ~ F50
	필드타입	노지 or 온실
	생육단계코드	1-9
	당도	0~20
	병해충 코드	00~06
	병해충 발생시기	01-01 ~ 12-31
	품종명	온주
어노테이션 도형 정보	어노테이션 ID	1609
	어노테이션 타입	polygon
	객체 클래스 코드	감굴_궤양병
	폴리곤 정보	[818 763 726 726 770 837 928...
기상환경 정보	파일명	환경정보 파일명
	시간정보	2021:10:15 15:32:25
	데이터 형태	OBS/SIM
	일사량	0~2000
	강우	0~600
	온도	-30~70
	습도	0~100
과실 품질 정보	비파괴검사 당도	0~100
	비파괴검사 건물률	0~100
	파괴검사 당도	0~100
	파괴검사 건물률	0~100
	파괴검사 색도	0~500
	파괴검사 경도	0~100
	파괴검사 산도	0~100

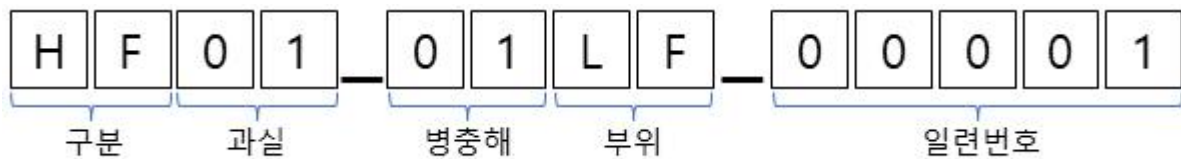
(다) 데이터 저장

□ 파일 포맷

- 어노테이션 포맷은 JSON 파일 형태의 메타데이터로 구성하여 특정 소프트웨어에 독립적으로 구축 데이터셋을 활용할 수 있도록 함
- 과수작물의 원천 이미지 데이터는 다양한 연구에 활용도를 높이기 위하여 JPG 포맷으로 저장하여 제공

□ 파일명명 규칙

- 파일아이디는 총 13자리이며, 타이틀, 과실분류, 병충해분류, 부위분류, 일련번호 등의 5개의 카테고리로 구분하고 4_4_5 구조로 “_”를 삽입



구분	과실분류	병충해		부위	일련번호
HF (고품질 과수작물)	01 : 감귤 02 : 키위	00 : 정상 01 : 감귤_궤양병 02 : 감귤_굴응애 03 : 감귤_진딧물	04 : 키위_점무늬병 06 : 키위_총채벌레 07 : 키위_과실무름병	FT : 과실 LF : 잎	000001, 000002, ...

<파일아이디 체계>

[파일명명규칙]

1) 병충해 데이터 명명 규칙

□ 과수/병충해/부위별 구분 : 파일명으로 구별

- 자리수 : 1 2 3 4 5 6 7 8 9 A B C D E F 0
- 파일명 : H F 0 0 _ 0 0 X X _ 0 0 0 0 0 0

□ 과수별 구분 : 자리수 위치 - 3, 4(데이터 형태 : 숫자)

- 자리수 : 1 2 3 4 5 6 7 8 9 A B C D E F 0
- 감 귤 : H F 0 1 _ 0 0 X X _ 0 0 0 0 0 0
- 키 위 : H F 0 2 _ 0 0 X X _ 0 0 0 0 0 0

□ 병충해별 구분 : 자리수 위치 - 6, 7(데이터 형태 : 숫자)

- 자리수 : 1 2 3 4 5 6 7 8 9 A B C D E F 0
- 정 상 : H F 0 1 _ 0 0 X X _ 0 0 0 0 0 0
- 궤양병 : H F 0 2 _ 0 1 X X _ 0 0 0 0 0 0

정상(00), 궤양병(01), 굴응애(02), 진딧물(03), 점무늬병(04), 총채벌레(06), 과실무름병(07)

□ 부위별 구분 : 자리수 위치 - 8, 9(데이터 형태 : 알파벳)

- 자리수 : 1 2 3 4 5 6 7 8 9 A B C D E F 0
- 과 실 : H F 0 1 _ 0 0 F T _ 0 0 0 0 0 0
- 앞 : H F 0 2 _ 0 0 L F _ 0 0 0 0 0 0

2) 초분광 데이터 명명 규칙

□ 초분광 영상 구분 : 자릿수 위치 - B, C (데이터 형태 : 알파벳)

- 자리수 : 1 2 3 4 5 6 7 8 9 A B C D E F 0 1 2 3
- 초분광(감굴) : H F 0 1 _ 0 0 F T _ H S _ 0 0 0 0 0 0
- 초분광(키위) : H F 0 2 _ 0 0 F T _ H S _ 0 0 0 0 0 0

3) 기상환경 데이터 명명 규칙

□ 기상환경 농장코드 구분 : 자릿수 위치 - 6, 7, 8 (데이터 형태 : 알파벳&숫자)

날짜 구분 : 자릿수 위치 - A~1A (데이터 형태 : 날짜)

- 자리수 : 1 2 3 4 5 6 7 8 9 A B C D E F 0 1 2 3 4 5 6 7 8 9 A
- 기상환경(감굴농장) : H F 0 1 _ F 0 1 _ Y Y Y Y M M D D - Y Y Y Y M M D D
- 기상환경(키위농장) : H F 0 2 _ F 0 3 _ Y Y Y Y M M D D - Y Y Y Y M M D D

□ 데이터 저장 구조

- 어노테이션 정보는 JSON 파일 형태의 구조체 형식으로 각 대상 객체의 클래스별로 어노테이션 정보를 저장
- 각 이미지 데이터와 일대일 매핑이 될수 있도록 파일명 형식을 갖추어 저장
- 저장 디렉토리는 각 종별로 분류하고 종별로 분류된 저장공간에 이미지, 어노테이션 메타데이터를 구조체 형태로 저장

1단계	2단계	3단계	파일명
1.Data	/라벨링데이터	/01.감굴	HF01_00FT_000001.json ~ HF01_03FT_999999.json
		/02.키위	HF02_00FT_000001.json ~ HF02_07FT_999999.json
		/03.초분광_감굴	HF01_00FT_HS_000001.json ~ HF01_00FT_HS_999999.json
		/04.초분광_키위	HF02_00FT_HS_000001.json ~ HF02_00FT_HS_999999.json

1단계	2단계	3단계	파일명
	/원천데이터	/01.감굴	HF01_01FT_000001.jpg ~ HF01_03FT_999999.jpg
		/02.키위	HF02_01FT_000001.jpg ~ HF02_07FT_999999.jpg
		/03.초분광_감굴	HF01_00FT_HS_000001.dat HF01_00FT_HS_000001.hdr ~ HF01_00FT_HS_999999.dat HF01_00FT_HS_999999.hdr
		/04.초분광_키위	HF02_00FT_HS_000001.dat HF02_00FT_HS_000001.hdr ~ HF02_00FT_HS_999999.dat HF02_00FT_HS_999999.hdr
		/05.환경기상	HF01_F01_20210801-20211130.csv ~ HF02_F22_20210801-20211130.csv
2.Validation	위와 동일	/01.감굴 /02.키위	위와 동일
3.Test	위와 동일	/01.감굴 /02.키위	위와 동일

<데이터 저장 구조>

(라) 어노테이션/라벨링 도구

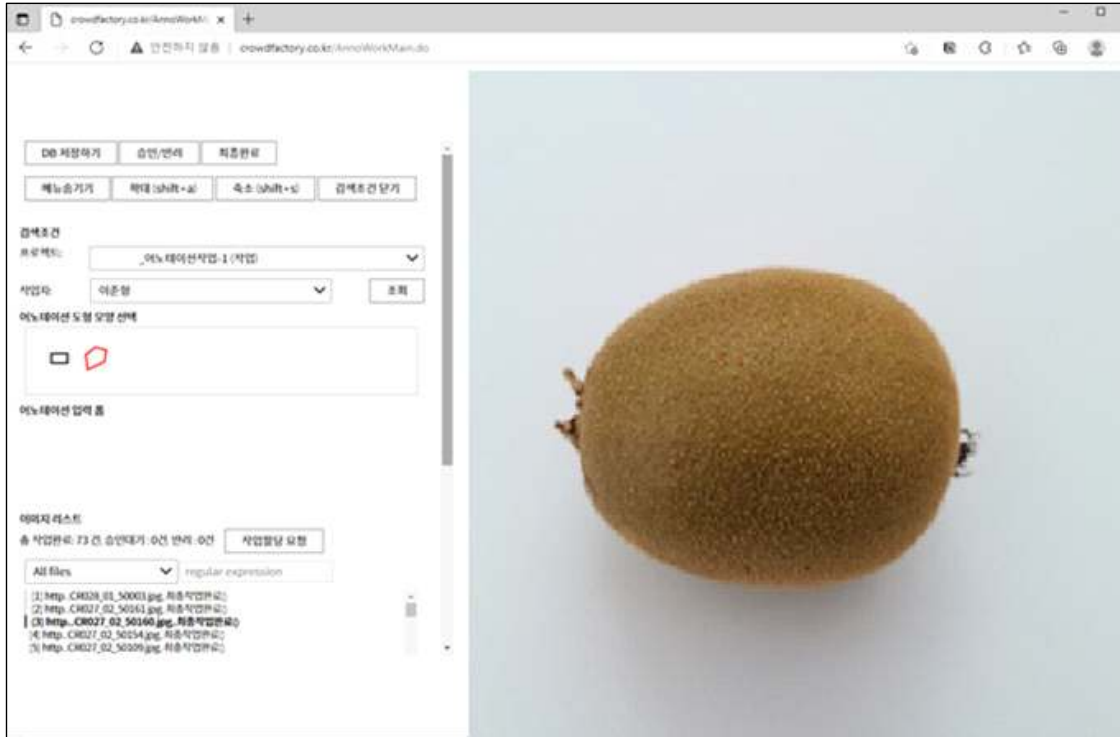
□ 클라우드팩토리 소개

- 본 과제에서는 주관기관인 지디에스컨설팅그룹이 보유한 어노테이션 통합솔루션을 사용하여 어노테이션 및 검수 작업까지 수행
- 클라우드팩토리는 클라우드워커 인력들이 재택근무로 활용할 수 있도록 웹기반의 서비스 제공
- VGG 오픈소스SW를 기반으로 개발된 클라우드팩토리는 기존 인공지능 데이터셋 구축 사업에서 활용하여 검증된 바 있음

□ 주요 기능

- Annotation 속성 파일을 저장 또는 로딩
- 주석을 만들 이미지를 불러오는 기능

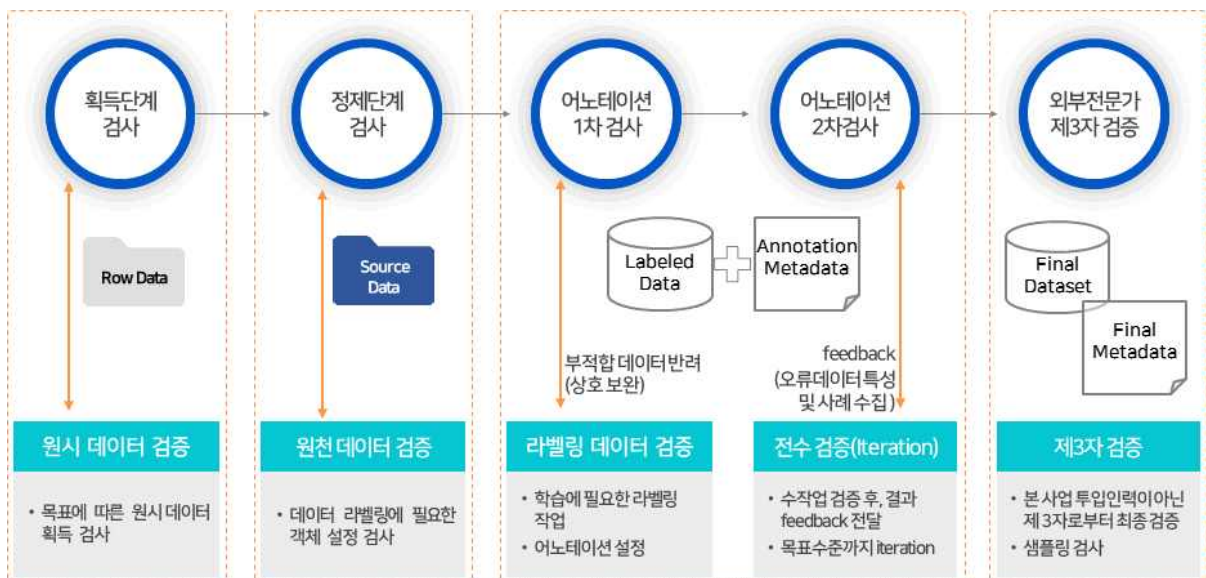
- 객체 종류 등 속성 확인 기능
- Annotation 좌표는 DB에 저장되며, 최종적으로 JSON, CSV, XML 형태로 Export
- 영상이미지 전환 기능



<클라우드팩토리 솔루션 작업화면>

(5) 데이터 검수

(가) 데이터 검수 절차



(나) 데이터 검수

구분	세부 내용
테스트 결과 검수	- 관리자의 어노테이션 테스트 결과 확인 및 합격/불합격 통보 * 불합격 이유(미분류, 과분류, 오분류 등) 통보
균질성 교육 및 재교육	- (합격) 어노테이션 결과의 균질성 확보를 위한 분류결과 교육 - (불합격) 분류항목, 분류기준 및 방법 등 재교육
교차검수 실시	- 작업자 A와 B의 어노테이션 결과 교차 검수
어노테이션 수정	- 작업자 교차검수에 따른 수정

- 작업자의 숙련도 부족에 따른 오류 발생을 최소화하고자 어노테이션 교육 및 테스트를 진행하며, 테스트 결과 검수를 통해 합격된 인원이 어노테이션을 실시
- 합격된 인원은 작업자간의 균질성 확보를 위해 테스트 분류결과에 대한 교육을 실시
- 테스트 결과에 따라 불합격된 인원은 분류항목, 분류기준 및 방법에 대해 재교육을 실시하며, 어노테이션 테스트 및 결과 검수 과정을 반복
- 어노테이션 결과에 대해 작업자 2인 이상을 1개 조로 편성하여, 작업자 간의 교차검수를 통해 검수 실시
- 수정사항 확인 시, 해당 작업자에게 오류사항에 대해 FeedBack 하며 어노테이션 수정 실시
- 필요 시, 어노테이션 실시 인원도 분류항목, 분류기준 및 방법에 대한 재교육 실시

검사 절차	검사 항목	검사 내역
획득단계 검사	법·제도 준수	개인정보, 법·제도 관련 검사
	사실적인 획득 환경 구성	원시데이터를 목적에 맞는 환경 구성 여부 검사
	데이터 동기화	이미지 데이터와 메타정보 간의 동기화
	편향성 방지	병충해 특성을 대표할 수 있는 분포 구성 여부 검사

정제단계 검사	정제 기준의 명확성	데이터 사용 목적에 적합한 정제 기준 수립 여부 검사
	중복성 방지	데이터 정제 시 중복 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
어노테이션 1차 검사	라벨링 가이드	목적에 맞게 작성된 라벨링 가이드에 대한 타당성 여부를 검사 후 라벨링 작업자들에게 내용 전달
	어노테이션 항목	목적에 맞는 어노테이션 구성인지 여부를 검사 후 확인된 내용을 포함하도록 작업자들에게 전달
	라벨링 검사 도구	자동화 도구를 통해 검사(적합성, 구문정확성 등) 후 검사자가 육안으로 부적합 데이터 여부 2차 확인(의미 정확성)과 촬영된 영상 이미지의 누락, 번짐 및 조건 오류를 점검하기 위한 도구를 활용하여 전수 검사를 통해 중복 없이 빠른 검사를 진행
어노테이션 2차 검사	부적합 판정 데이터 분포 확인	데이터의 오류율, 특성 분포 확인을 통한 데이터 수집, 정제, 라벨링, 부문 최적화
	교차/전수 검수	1차 검사자와 별도의 인력이 2차 검사 진행하며, 전수 검사 시행
제3자 검사	외부 전문가에 의한 제3자 검사	본 과제의 참여기관 소속이 아닌 제3자로부터 최종 검증

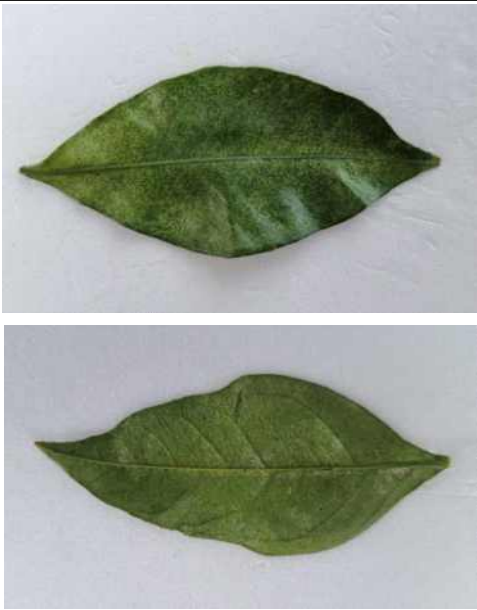
(다) 품질검증 지표 및 목표




품질특성	항목명	측정 지표	정량 목표	지표 및 목표 설정 근거
다양성	클래스 (감굴 부위) 별 라벨 분포	비율	분포확인	일부 병충해가 우점종이 되면 경쟁 병충해가 감소하는 현상 발생으로 병충해가 균등하게 발생하지 않음
	클래스 (키위 부위) 별 라벨 분포	비율	분포확인	일부 병충해가 우점종이 되면 경쟁 병충해가 감소하는 현상 발생으로 병충해가 균등하게 발생하지 않음
	클래스 (병충해) 별 라벨 분포	비율	분포확인	일부 병충해가 우점종이 되면 경쟁 병충해가 감소하는 현상 발생으로 병충해가 균등하게 발생하지 않음
	감굴 초분광 생육 단계 분포	비율	분포확인	데이터구축방법 및 검증계획서
	키위 초분광 품종 분포	비율	분포확인	데이터구축방법 및 검증계획서
구문 정확성	구문 정확도	정확도(%)	99.9% 이상	TTA 권고사항

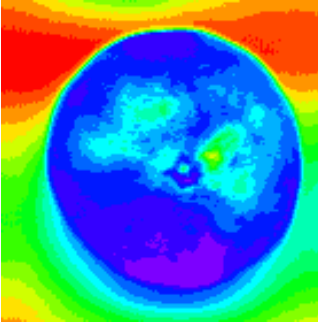
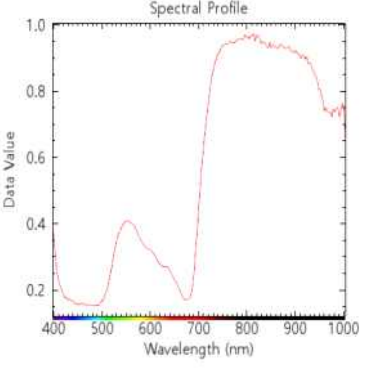
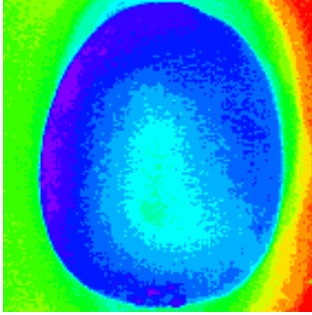
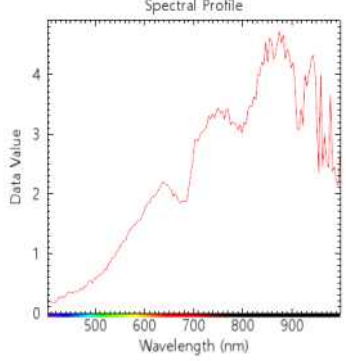
품질특성	항목명	측정 지표	정량 목표	지표 및 목표 설정 근거
의미 정확성	폴리곤 라벨 정확도	정밀도 재현율 F1-score	95% 이상	데이터구축방법 및 검증계획서
	분류 라벨 정확도	정밀도 재현율 F1-score	95% 이상	초기데이터 검수 결과 Precision: 92%, Recall: 100%, F-1 Score : 95.83%, Accuracy: 91.16%
유효성	객체 분류	F1-score	85% 이상	본 과제는 농업현장에서 구축된 데이터셋 으로 학습하므로 실질적으로 실현 가능한 85%를 목표로 설정함

(6) 최종데이터셋

분류	과실	클래스	이미지
병충해 이미지	감귤	감귤_정상	 <p>The image shows two photographs of a healthy citrus fruit and leaf. The top photograph is a whole, round, green citrus fruit with a small stem at the top, set against a plain white background. The bottom photograph is a single, dark green, ovate leaf with prominent veins, also set against a plain white background.</p>
		감귤_궤양병	 <p>The image shows two photographs of citrus fruit and leaf affected by scab disease. The top photograph is a whole citrus fruit with a light blue background, showing numerous small, brown, scab-like lesions on its green surface. The bottom photograph is a single leaf with a light blue background, showing several yellowish-green, irregular spots on its surface, characteristic of scab infection.</p>

분류	과실	클래스	이미지
		감귤-귤응애	
		감귤-진딧물	
	키위	키위-정상	

분류	과실	클래스	이미지
		키위_점무늬 병	
		키위_총채벌 레	
		키위_과실무 름병	

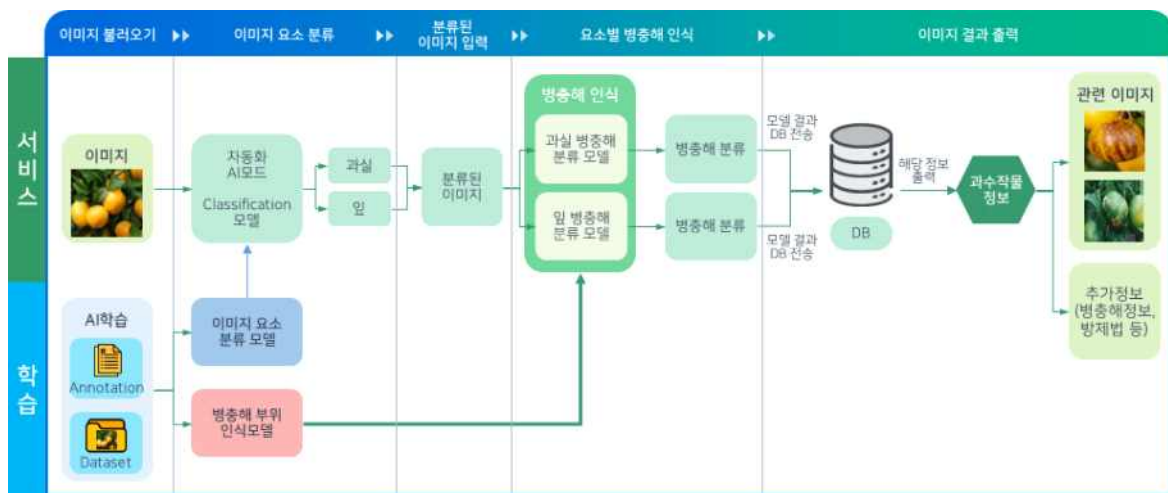
분류	과실	클래스	이미지
초분광 이미지	감귤	감귤_초분광	 
	키위	키위_초분광	 

(7) AI학습모델 개발

(가)모델 적합성 검토

- Classification 모델에는 기존 연구결과¹⁾에 따라 아래 표와 같이 3개의 모델에 대하여 선정 기준과 절차에 따라 학습모델 선정을 진행함

Model	VGGNet	ResNet	DenseNet
Accuracy	74.5%	83.2%	78.44%



<AI모델 개발 프로세스>

- 과수작물 병충해 데이터 비교 사례 논문을 통해 실현 가능한 목표 설정
 - 논문²⁾에서는 ResNet50을 바탕으로 구성된 PD2SE-Net 모델을 사용하여 사과, 배, 후추 등 9종의 서로 다른 식물의 질병 심각도 추정, 종 인식 및 질병 분류 학습 모델
- 과적합 방지를 위한 모델 검증 및 대상 병충해 변경
 - 과적합 (Overfitting) 방지를 위해 모델 학습시 후보 학습셋 데이터를 모델 구축용 (training dataset)과 모델 검증용 데이터셋(validation dataset)으로 분리
 - 두 데이터에 대해 동일한 모형을 적용하고 두 모델의 설명력(결정계수 : R^2)기준 10% 이상 차이나는 경우 병충해 대상에 대한 변경 검토
 - 병충해 대상 변경 및 선정은 NIA와 협의 후 변경

1) <https://paperswithcode.com/sota/image-classification-on-imagenet>

2) Liang, Qiaokang, et al. "PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network." Computers and electronics in agriculture 157 (2019): 518-529.

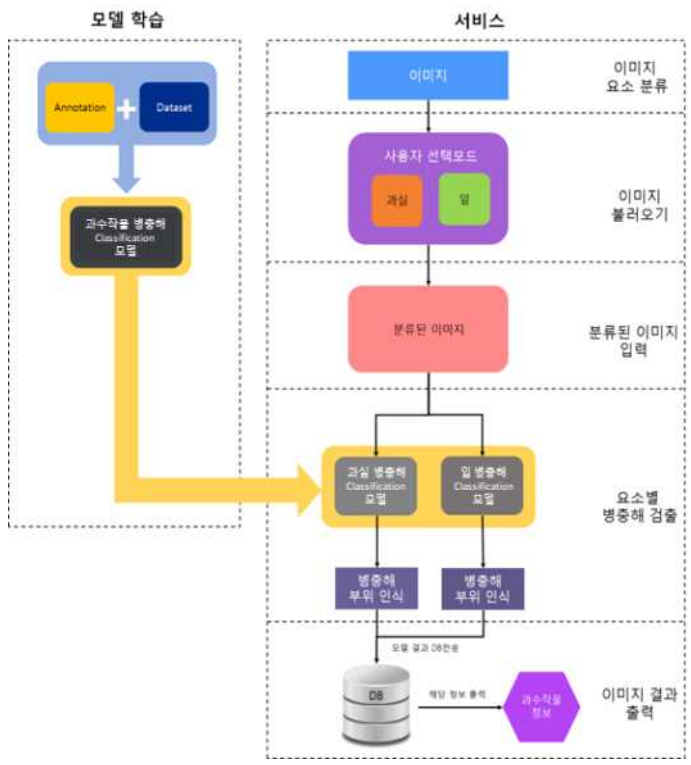
(나)인공지능 학습 모델 선정

- 대규모 ImageNet 데이터셋에서 우수한 성능을 보인 VGGNet, ResNet, Densenet 모델을 Classification 인공지능 학습 모델로 선정함
- VGGNet, ResNet 모델은 ImageNet 데이터셋에 대해 높은 정확도를 보이며, 국제 ILSVRC 대회에서 이미지 분류 오류율이 각각 7.3%, 3.57% 로 낮은 결과로 좋은 성능을 나타냄
- Inception V3 모델의 경우 높은 정확도를 보이나 소모하는 컴퓨팅 파워가 크므로 Densenet 모델이 고품질 과수작물 병충해 분류 작업에 더 적합할 것으로 예상함

(7) 응용(시범) 서비스 개발

(가) 과수작물 병충해 식별 사전 AI 서비스 개요

- 병충해를 자동으로 식별하는 시스템에서는 병해의 종류에 따라 과실, 잎에 나타나는 색깔, 무늬 등 시각적 특성이 다양하므로 판별에 어려움이 있음
- 병충해를 식별하는 시스템에 관해 본 서비스는 과수작물의 병충해 정보를 얻기 위하여 과실, 잎의 다양한 종류의 병충해 이미지들이 포함된 학습데이터를 활용한 분류 및 분류 모델을 개발함
- 본 서비스는 과수작물 이미지 내에서 AI 모델 개발을 통하여 병충해를 식별한 후, 해당 병충해의 종류 및 방제법 등의 다양한 정보를 출력해주는 서비스 제공



<서비스 모식도>

(8) 경진대회 및 홍보활동

(가) 해커톤 경진대회

□ 해커톤 개요

- 본 과제를 통해 수집한 고품질 과수작물 통합데이터를 활용하기 위하여 제주테크노파크의 주관하에 온라인 해커톤을 개최
- 해커톤 운영을 위해 나라장터 공개입찰을 통해 (주)디자인오아이스를 선정
- 온라인 플랫폼(<http://jejuhkt.kr/>)을 통하여 참가팀을 모집하였으며 COVID-19 및 원거리 참가자가 다수임을 감안하여 사이버 시상식(<http://jejuhkt.kr/vr/index.htm>)으로 운영

□ 해커톤 진행경과

일정	구분	주요내용
11월 1일 ~ 26일	준비/홍보	▶ 해커톤 데이터 검증 ▶ 온라인 해커톤 홍보물 제작 홍보 등
11월 23일~12월6일	접수	▶ 온라인 모집공고 및 접수
11월29~12월4일	설명회	▶ 온라인 사전 교육 1회 진행 ▶ 온라인 설명회 1회 진행 총 80팀 등록
12월 6일~13일	해커톤	▶ 플랫폼 기반으로 온라인 해커톤 진행 총 80팀 중 7팀 최종 결과물 제출
12월 14일~ 17일	평가	▶ 계량 평가 진행
12월 20일	시상식	▶ 우승자 PT발표 ▶ 시상식

□ 해커톤을 위한 데이터 확보 및 검증

- 해커톤을 운영하기 위하여 데이터 검증기관인 데이터웨이로부터 데이터를 제공받아 활용
- 제공된 데이터는 감귤과 키위 각각 12,000세트씩이었으며 제공된 데이터는 온라인 플랫폼 자료실에 업로드하여 참가자들이 손쉽게 활용할 수 있도록 함

항목	세부 내용	배점
문제 해결 방법론의 적절성	<ul style="list-style-type: none"> ● 해결하고자 하는 문제를 위한 접근 방법의 정확도 ● 모델 선정의 적합성 	20
결과 및 분석의 완성도	<ul style="list-style-type: none"> ● 테스트 결과의 정확도 ● 수행 결과에 대한 분석의 적절성 	20
소스 코드 구현 품질	<ul style="list-style-type: none"> ● 소스 코드 주요 내용 포함 여부 ● 소스 코드의 완결성 	20
보고서 완성도	<ul style="list-style-type: none"> ● 보고서 내용의 가독성 ● 보고서 콘텐츠 적절성 ● 보고서 주요 내용 포함 여부 	20

<해커톤 평가항목>

- 1차심사 결과 딥러닝하자팀이 총점 458점 최저/최고점 제외 279점으로 1위를 차지 (심사 결과는 아래 표 참고)

팀명	총점	평균점	순위	최고/최저 제외총점	최고/최저 제외평균	최고/최저 제외순위
딥러닝하자	458	91.60	1	279	93.00	1
eslab	447	89.40	2	272	90.67	2
DIPL	413	82.60	3	257	85.67	3
원서화이팅	413	82.60	3	255	85.00	4
리멤버	382	76.40	5	232	77.33	5
Badger & AI	380	76.00	6	228	76.00	6
누에고치	361	72.20	7	216	72.00	7

<1차심사 결과>

- 1차 심사결과에서 1~4위 팀인 딥러닝하자, eslab, DIPL, 원서화이팅 팀에 대하여 소스코드 검증을 한림성심대학교 디지털콘텐츠개발과에 의뢰하여 실시하였으며, 그 결과를 포함하여 최종 순위를 결정
- 최종심사결과는 딥러닝하자, 원서화이팅, DIPL, eslab 순으로 선정

팀명	1차 순위	1차 점수	2차 검증 순위	2차 검증 점수	최종합계	최종순위
딥러닝하자	1	93.00	1	20	113.00	1
eslab	2	90.67	4	10	100.67	4
DIPL	3	85.67	3	16	101.67	3
원서화이팅	4	85.00	2	18	103.00	2

<최종심사 결과>

- 1위를 한 딥러닝하자팀에게는 제주특별자치도지사상(대상)이 수여되었으며, 나머지 3개팀에게는 제주테크노파크 원장상이 수여

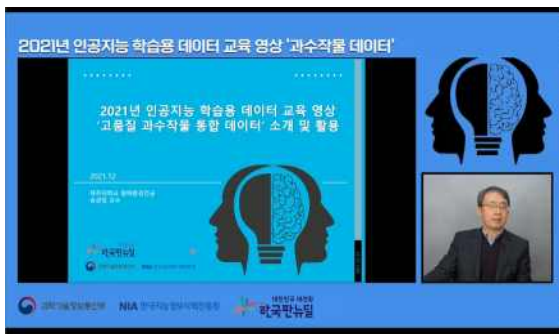
(나) 홍보동영상 제작

- 감귤/키위 과수작물에 대한 소개, AI알고리즘에 대한 소개, 과제에 대한 소개 등을 내용으로 홍보동영상을 제작

주제	내용	강사
Chapter 1. 과수작물의 특성	감귤/키위의 개요 및 특성	제주대학교 원예학과 송관정 교수
	감귤 재배현황 및 주요 병해충	
	키위 재배현황 및 주요 병해충	
Chapter 2. 인공지능 학습용 데이터 구축사업 소개	과기부 AI 학습용 데이터 구축사업	이화여자대학교 산학협력단 김규호 교수
	인공지능 학습 최신트렌드	
	데이터 분야 QnA 진행	
Chapter 3. 고품질 과수작물 통합 데이터 구축과제 소개	고품질 과수작물 통합 데이터 구축과제 소개	지디에스컨설팅그룹 박경채 이사

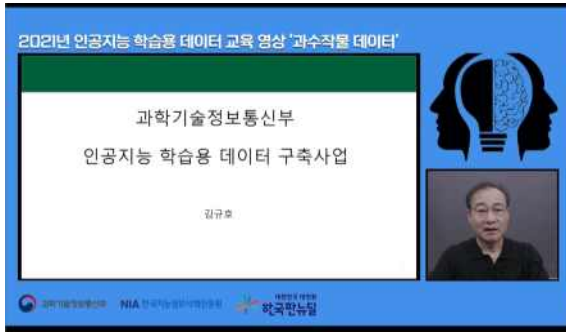
<홍보동영상 커리큘럼>

Chapter 1 : 과수작물의 특성			
강사	제주대학교 원예학과 송관정 교수	Running Time	30 - 45(m)
목 차	<ul style="list-style-type: none"> - 감귤 / 키위의 개요 및 특성 - 감귤 재배현황 / 키위 재배현황 - 감귤 주요 병해충(감귤 궤양병, 궤응애, 진딧물) - 키위 주요 병해충(키위 궤양병, 점무늬병, 총채벌레) 		



Chapter 2 : 인공지능 학습용 데이터 구축사업 소개			
강사	이화여대 김규호 교수	Running Time	20 - 30(m)
목 차	<ul style="list-style-type: none"> - 과학기술정보통신부 인공지능 학습용 데이터 구축사업 소개 - AI 허브 및 AI 허브 내 데이터 예시 소개 		

- 지정공모과제 발췌(예시)
- 인공지능 학습 최신 트렌드 소개
- 인공지능 9대 핵심 기술
- 최근 인공지능 모델 예시 사례
- 인공지능/AI/데이터 등 QnA 진행



Chapter 3. 고품질 과수작물 통합 데이터 구축과제 소개

강사	지디에스컨설팅그룹 박경채 이사	Running Time	10 - 20(m)
목 차	<ul style="list-style-type: none"> - 추진배경 - 사업목표 및 과제 범위 - 데이터 구축내역 - 데이터 소개(감귤/키위 병충해 데이터) - 데이터 수집/정제(병충해 데이터/초분광 데이터) - 데이터 구축(Annotation/Labeling/저작도구) - 데이터 검수(구조/검수도구 및 검수매뉴얼) - AI 학습모델 개발(알고리즘 선정) 및 기타(작업 가이드) 		

