

●○ 피복지도 및 산림수종

## 토지 피복지도 항공위성 이미지 데이터(강원 및 충청)



●○ 개요: 강원충청권 토지피복 AI 학습데이터셋이란?

토지피복은 인공위성이나 항공기로 촬영한 이미지를 이용하여 지표면의 상태를 표현한 지도로, 국토의 현 상태를 수집하고, 변화를 파악하는데에 활용된다.

Semantic Segmentation은 이미지에서 객체를 분류하는 기술로, 이 중 Deeplab Network-Semantic Segmentation은 최고의 성능을 나타낸다.

(주)네이버시스템은 세종청사이전, 택지 개발 등으로 인한 국토 변화를 신속히 탐지하고 분석하기 위해, 토지피복 Deeplab Network-Semantic Segmentation 기술 개발에 활용할 수 있는 강원 충청권 토지피복 AI 학습데이터셋을 구축하였다. 강원충청권 토지피복 AI 학습데이터셋은 총 10만 건의 이미지데이터(픽셀크기 512×512)로 구축하였으며, 건물, 주차장, 도로, 가로수, 논, 밭, 산림, 나지 8개 클래스로 구성되어 있다.

강원충청권 토지피복 Semantic Segmentation 사례는 그림 1과 같다. 왼쪽 그림과 같이 사용자가 항공이미지를 입력하면, 논, 밭, 도로 등 토지피복 객체를 segmentation 하여 오른쪽 그림과 같은 결과를 출력한다.



그림1 | 강원충청권 토지피복 Semantic Segmentation 사례

### ●○ 데이터셋의 구성

강원충청권 토지피복 AI학습데이터셋은 항공이미지로 작성한 해상도 0.51m의 학습데이터셋 99400건과, 위성이미지로 작성한 해상도 10m의 학습데이터셋 600건으로 구성되어 있으며, 총 8종의 클래스를 포함한다. 단, 10m 토지피복 AI 학습데이터셋은 해상도의 제약에 따라 건물, 도로, 논, 밭, 산림 5종의 클래스를 포함한다.

데이터 명칭	데이터 종류 (이미지종류)	포함내용	제공방식
강원충청권 토지피복 AI학습 데이터셋	0.51m 토지피복 AI데이터셋 (항공)	건물, 주차장, 도로, 가로수, 논, 밭, 산림, 나지 (99400건)	Tiff, JSON 포맷 파일
	10m 토지피복 AI데이터셋 (위성)	건물, 도로, 논, 밭, 산림 (600건)	Tiff, JSON 포맷 파일

### ●○ 데이터셋의 설계 기준과 분포

강원충청권 토지피복 AI 학습데이터셋은 데이터 활용도를 고려하여 설계하였다. 다양한 환경변화를 효율적으로 탐지하기 위해 0.51m해상도의 항공이미지와 10m 해상도 위성이미지를 원시이미지로 선택하여 multi resolution 피복지도 학습용 data sets을 구성하였다. 강원충청권 토지피복 특징을 잘 반영할 수 있도록 지역적으로 골고루 분포하여 학습지역을 선정하였고, Fine annotation과 Coarse annotation방법을 병행하여 많은 양의 학습용 데이터를 확보할 수 있도록 하였다. 또한, 512\*512 크기 학습데이터 외, 1024\*1024크기 학습데이터를 추가 제공하여 AI 데이터셋 활용도를 제고하였다.

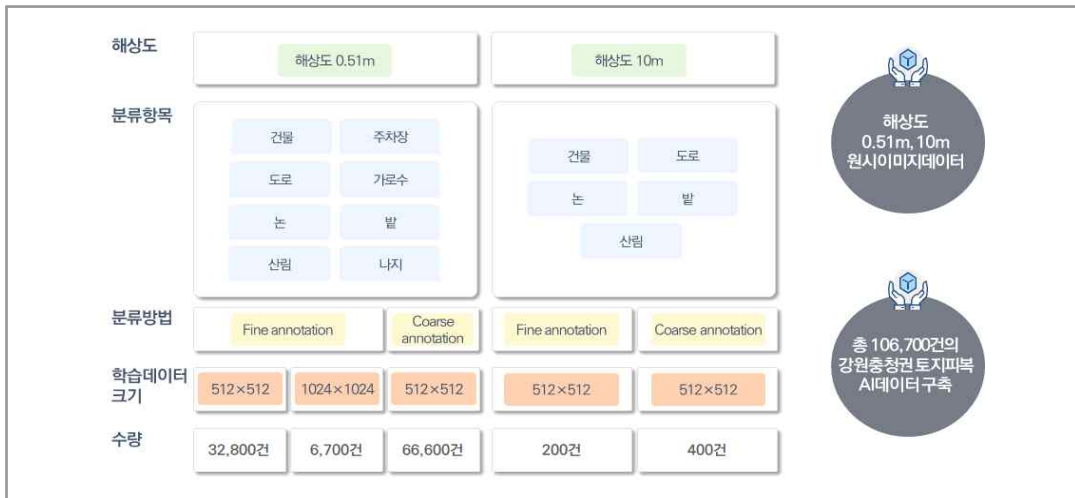


그림2 | 데이터셋 구성 개요

강원충청권 토지피복 AI 학습데이터셋 학습지역 분포는 그림 3과 같다. 0.51m 학습데이터셋의 가로수 항목의 경우 절대적인 분포가 적어, 범위 내 가로수 지역을 최대한 많이 포함하도록 선정하였다.

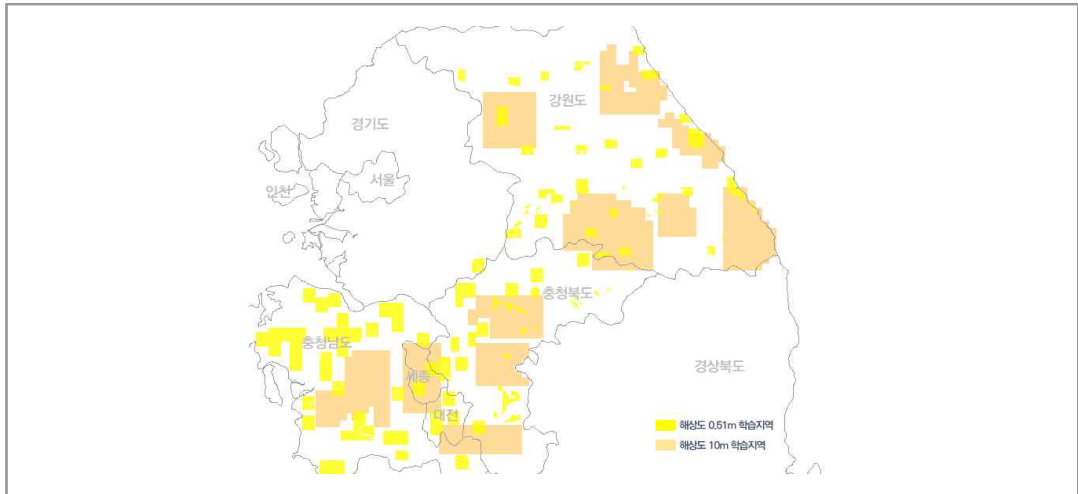


그림3 | 강원충청권 토지피복 학습지역 분포

### ●○ 데이터 구조

데이터셋의 어노테이션 정보는 gray scale의 Tiff 이미지 형태로 제공하고, 원시이미지, 데이터 너비, 높이 등 데이터에 대한 정보를 Json 포맷으로 제공한다.

#### ● 메타데이터 항목

No	항목		타입	필수 여부	예시
	영문명	한글명			
1	Image	이미지 정보			
1-1	img_id	이미지 식별자	String	Y	LC_AP_00000000_001
1-2	img_width	이미지 너비	Number	Y	512
1-3	img_height	이미지 높이	Number	Y	512
1-4	img_type	이미지 종류	String	Y	aerophoto
1-5	img_coordinate	이미지 좌표계	String	Y	EPSG:5186
1-6	coordinates	좌상단 좌표	String	Y	000, 000
1-7	img_resolution	이미지 해상도	Number	Y	0.51
2	annotations	어노테이션 정보			
2-1	ann_id	어노테이션 식별자	String	Y	LC_AP_00000000_001_FGT
2-2	ann_type	어노테이션 타입	String	Y	polygon
2-3	ann_file_type	어노테이션 파일 유형	String	Y	tif

- 이미지 식별자는 이미지 파일명을 나타냄
- 이미지 종류, 좌표계, 해상도는 어노테이션에 사용한 원시이미지의 종류, 좌표계, 해상도를 나타냄
- 좌상단 좌표는 학습용 이미지 데이터의 좌상단 X, Y 좌표를 나타냄
- 이미지 너비 및 높이 정보는 학습을 위한 데이터의 크기를 나타냄 (기본 512x512)
- 어노테이션 식별자는 어노테이션 파일명을 나타냄
- 어노테이션 타입은 어노테이션 툴에서 지원하는 모양 중 사용한 모양을 나타냄
- 데이터셋 폴더 구조 및 파일명 규칙
  - 강원충청권 파복지도 106,700장의 이미지 데이터에 대한 폴더구조 및 파일 명 규칙은 다음과 같다.

| File ID체계(13자리) |

구분(2)	원시이미지종류(2)	도엽번호(8)(5)	일련번호(3)	GT구분(3)
LC : 토지피복	항공 : AP	00000000	001,002,...	Fine : FGT Coarse :CGT
	위성 : SN	00000		

| 적용형식 및 예시 |

파일종류	형식	예시
원시이미지	FileID.tif	LC_AP_00000000_001.tif LC_AP_00000000_001_1024.tif LC_SN_00000_001.tif
Annotation	FileID_FGT.tif FileID_CGT.tif	LC_AP_00000000_001_FGT.tif LC_AP_00000000_001_FGT_1024.tif LC_AP_00000000_001_CGT.tif LC_SN_00000_001_FGT.tif LC_SN_00000_001_CGT.tif
JSON	FileID_FGT.json FileID_CGT.json	LC_AP_00000000_001_FGT.json LC_AP_00000000_001_FGT_1024.json LC_AP_00000000_001_CGT.json LC_SN_00000_001_FGT.json LC_SN_00000_001_CGT.json
SHP	FileID_FGT.shp	LC_AP_00000000_001_FGT.shp LC_AP_00000000_001_FGT_1024.shp LC_SN_00000_001_FGT.shp

※ 1024 : 학습용데이터 크기 (기본크기 512), SHP : fine annotation 중간산출물

| SHP 데이터명세서 |

- 환경부 토지피복지도 작성 지침 상 분류항목명을 준용하여 정의

컬럼명	데이터타입(길이)	컬럼명 설명
ann_code	숫자(3)	어노테이션 코드
ann_name	텍스트(10)	어노테이션 항목명

ann_code	ann_name
10	건물
20	주차장
30	도로
40	가로수
50	논
60	밭
70	산림
80	나지
0	비대상지

| 파일 저장 디렉토리 구조 |

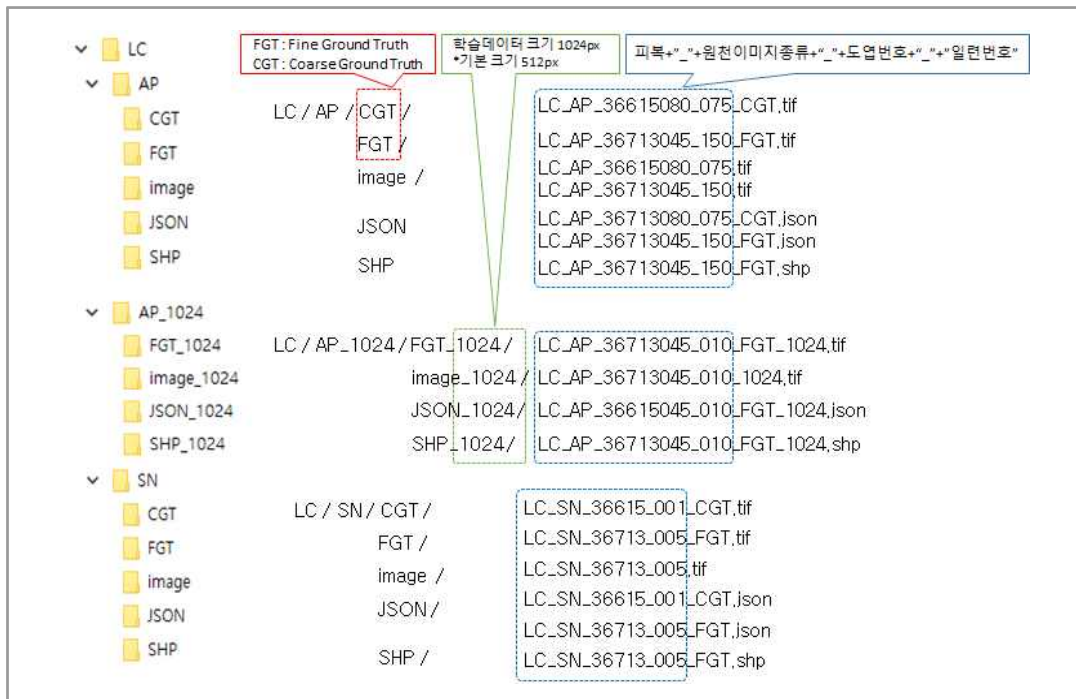


그림4 | 토지피복 AI 데이터셋 파일 저장 디렉토리구조

| 클래스항목별 코드 및 Gray Scale 값 |



ann_code	ann_name	Gray Scale	비고
10	건물	10	항공/위성
20	주차장	20	항공
30	도로	30	항공/위성
40	가로수	40	항공
50	논	50	항공/위성
60	밭	60	항공/위성
70	산림	70	항공/위성
80	나지	80	항공
0	비대상지	100	항공/위성

※ 클래스 8종에 해당하지 않는 지역은 '비대상지'로 분류

●○ 데이터 예시


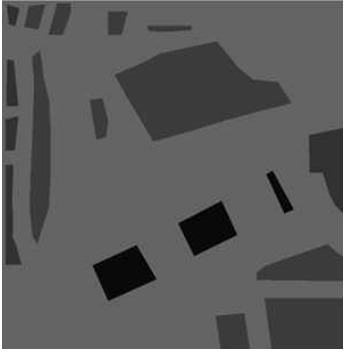
토지피복 AI 학습용 데이터의 Fine 및 Coarse annotation 결과물은 각 클래스정보의 코드값에 따라 정의된 색상으로 변환하여 Gray Scale의 8bit Tiff로 구축하며, 데이터 포맷 및 예시는 다음과 같다.

- Fine annotation 데이터 포맷 및 예시

원시이미지	학습용데이터	
Tiff	Tiff	Json
		<pre>[   {     "img_id": "LC_AP_37710090_095",     "img_width": 512,     "img_height": 512,     "img_type": "aerophoto",     "img_coordinate": "EPSG:5186",     "coordinates": "243275.99, 520906.070136",     "img_resolution": 0.51,     "img_winter": 0,     "ann_id": "LC_AP_37710090_095_FGT",     "ann_type": "polygon",     "ann_file_type": "tif"   } ]</pre>

\* fine annotaon은 중간산출물인 shp포맷파일을 함께 제공



• Coarse annotation 데이터 포맷 및 예시

원시이미지	학습용데이터	
Tiff	Tiff	Json
		<pre> {   "img_id": "LC_AP_37711081_155",   "img_width": 512,   "img_height": 512,   "img_type": "aerophoto",   "img_coordinate": "EPSG:5106",   "coordinates": "246213.59, 520122.710136",   "img_resolution": 0.51,   "img_winter": 0,   "ann_id": "LC_AP_37711081_155_C6T",   "ann_type": "polygon",   "ann_file_type": "tif" }                 </pre>

※ 위 그림과 같이 Fine annotation은 객체의 경계를 분류하고, Coarse annotation은 객체의 내부를 분류한 결과물이다.

●○ 데이터 구축 과정

데이터 구축은 촬영계절, 운량을 고려하여 최신의 항공이미지(2018-2019년 촬영) 및 위성이미지(Sentinel-2 2019년-2020년 촬영)를 획득·정제하고, 강원충청권을 대상으로 지역적으로 골고루 분포하여 학습지역을 선정 한 후, 총 106,700건의 학습용데이터를 작성하였다.

구분	항공이미지	위성이미지
의 세 도	<p><b>국토지리정보원 항공영상</b></p>  <p>항공영상 다운로드</p>  <p>수도권, 강원, 충청 평중영상</p> <ul style="list-style-type: none"> <li>• 통합검색-영역을 클릭하고, 지도에서 원하는 범위를 설정한 다음, 항공사진 선택 후 정비구니에 답이 다운로드</li> </ul>	<p><b>Copernicus Open Access Hub Sentinel-2 영상</b></p>  <p>Sentinel-2 영상 다운로드</p>  <p>수도권, 강원, 충청 Sentinel-2 영상</p> <ul style="list-style-type: none"> <li>• 검색설정에서 Sensing period에서 촬영기간을 선택하고, Mission : Sentinel-2의 Product Type(S2MSI1C) 선택</li> <li>• 버튼을 선택하고, 지도화면에 와 같이 다운로드 받을 영역을 선택하고 검색 및 다운로드 실시</li> <li>• 다운로드 받은 압축파일을 풀면 jp2포맷으로 된 13개 밴드의 영상파일을 확인할 수 있음</li> </ul>

구분	항공이미지	위성이미지
정제	<p>국토지리정보원 제공 항공이미지는 Tiff 포맷의 RGB 이미지 자료로 별도의 데이터 정제를 실시하지 않음</p>	 <p>Sentinel-2 위성영상 전처리</p> <ul style="list-style-type: none"> <li>· QGIS에서 영상 전처리를 실시하여 각 밴드별 Geotiff로 변환 (SCP - Preprocessing - Sentinel2)</li> <li>· QGIS에서 layer stack으로 Red, Green, Blue, Nir 4밴드 영상으로 정제 (SCP - Bandset)</li> </ul> <p>Sentinel-2 Red, Green, Blue, NIR밴드 layer stack</p> <p>Band 2 (Blue) Band 3 (Green) Band 4 (Red) Band 8 (NIR)</p> <p>Sentinel-2 정제영상</p>

그림5 | 데이터 획득 및 정제

토지피복 AI 학습용 데이터의 Fine annotation은 객체의 경계를 분류하는 정밀한 고품질 어노테이션으로, 공간정보 편집을 위한 오픈소스 기반의 편집도구인 QGIS를 활용하여 구축하였다.

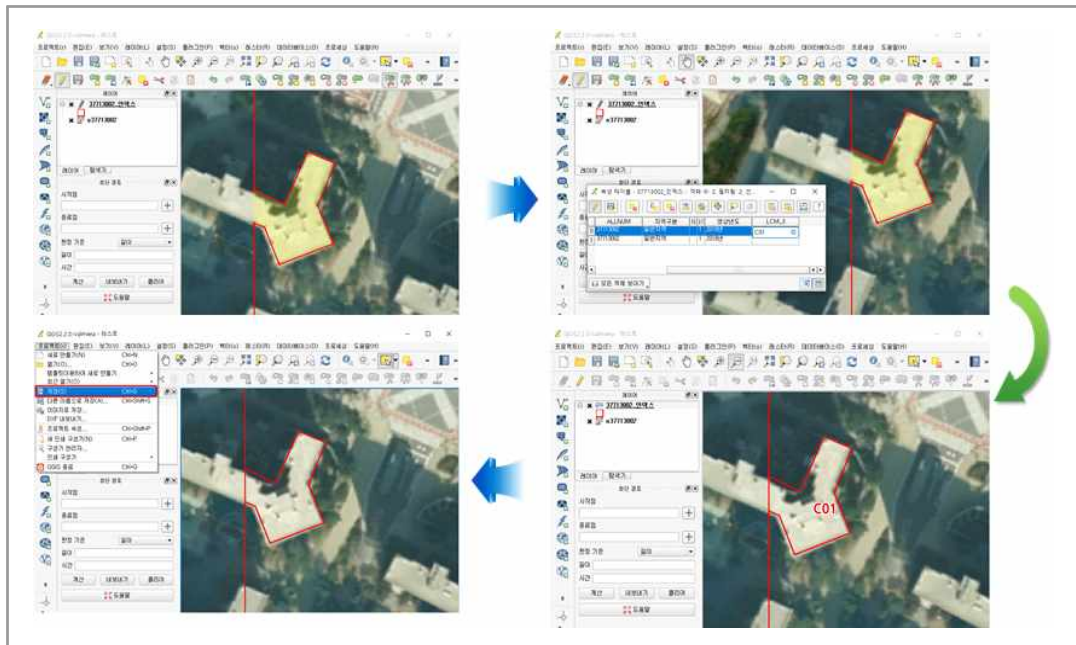


그림6 | 오픈소스 기반의 편집도구 QGIS를 이용한 Fine annotation



토지피복 AI 학습용 데이터의 Coarse annotation은 객체 내부를 분류하는 효율적인 어노테이션으로, 클라우드 기반 시스템을 통해 비대면 작업으로 구축하였다.

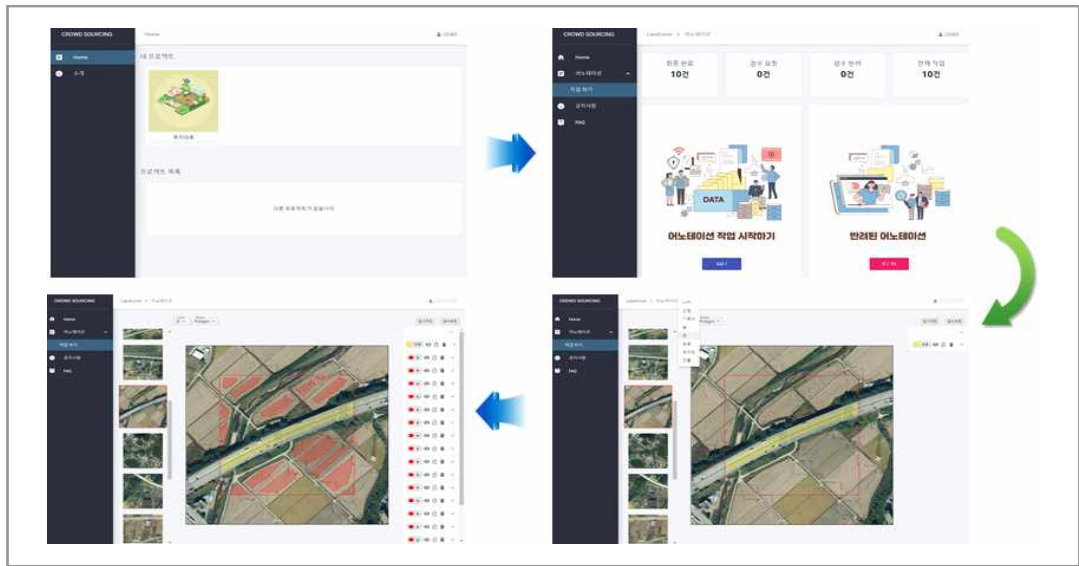
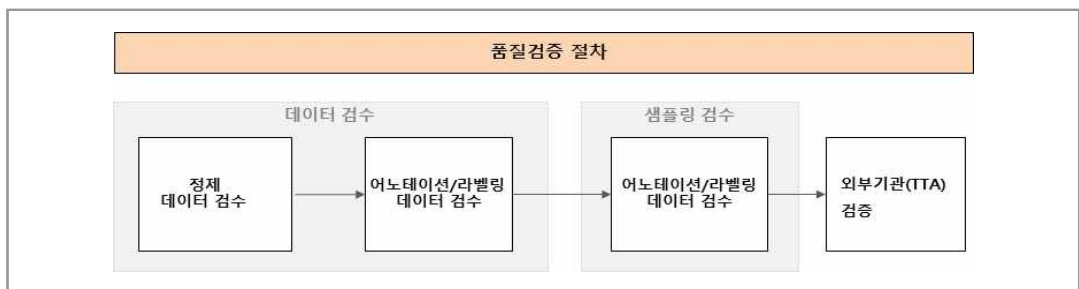


그림 7 | 클라우드 소싱을 통한 Coarse annotation

## ●○ 검수와 품질 확보

### | 검수절차 |



- 작업자의 숙련도 부족에 따른 오류 발생을 최소화하고자 어노테이션 교육 및 테스트를 진행하며, 테스트 결과 검수를 통해 합격된 인원이 어노테이션을 실시
- 어노테이션 결과에 대해 작업자 2인 이상을 1개 조로 편성하여, 작업자 간의 교차검수를 통해 검수 실시
- 수정 사항 확인 시, 해당 작업자에게 오류사항에 대해 FeedBack 하며 어노테이션 수정 실시

## | 검수 기준 |

- 정제 데이터 검수
  - 항공이미지는 별도의 데이터 정제를 하지 않으므로 검수에서 제외하며, Sentinel-2 위성 이미지는 4개 밴드(Blue, Green, Red, Nir) 여부를 확인하고, 밴드 수량이 다를 경우 오류로 식별하여 데이터 획득·정제팀에 보완 요청
  - 오류 리포트를 통하여 검수팀에서 오류를 식별하고 획득·정제팀에게 오류 결과 보고하며 보고 결과를 전달받은 획득·정제 팀에서 재작업 후 검수팀에게 재검수 요청 실시
- 어노테이션/라벨링 검수
  - 미식별 오류
    - 정의 : 대상 객체이지만 어노테이션/라벨링을 미실시한 경우 미식별 오류로 체크
    - 기준 : 어노테이션/라벨링 클래스 대한 정의 및 기준에 따라 검수
  - 과식별 오류
    - 해당 객체 제외되어야 하지만 어노테이션을 실시한 경우 과식별 오류로 식별
  - 미대상 객체 식별 오류
    - 정의 : 대상 객체가 아니지만 어노테이션을 실시한 경우 미대상 객체 식별 오류로 체크
    - 기준 : 어노테이션/라벨링 클래스 대한 정의 및 기준에 따라 검수
  - 중복 식별 오류
    - 정의 : 동일 객체에 어노테이션을 실시한 경우 중복 식별 오류로 식별
    - 기준 : 어노테이션/라벨링 클래스 대한 정의 및 기준에 따라 검수
- 오식별 오류
  - 어노테이션 오류
    - 정의 : 대상 객체에 대한 어노테이션의 클래스 종류가 잘못 설정된 경우 어노테이션 오류로 식별
  - 라벨링 오류
    - 정의 : 정의된 클래스에 대한 라벨링 잘못이나 아이디 잘못된 경우 라벨링 오류로 식별

## ●○ 데이터 구축 담당자

참여기관 : (주)이테라 김대영대표

(전화: 010-\*\*\*\*-9333, 이메일: daeyoung\_72@naver.com)