

테크니컬 리포트

2020년 1차
인공지능
학습용
데이터 구축

기타 영역

위성영상 객체판독

개요: 위성영상 객체판독 AI데이터셋이란?

위성영상에서 얻을 수 있는 정보는 현재 다양한 사회문제의 진단과 해결을 위해 사용되고 있다. 예를 들면, 위성영상은 항공 영상과는 달리 접근 불가지역을 아무런 외교적 마찰 없이 촬영할 수 있어, 타국의 핵시설이나 군사 활동 등을 감시 할 수 있고, 매우 넓은 지역을 짧은 시간 내에 촬영할 수 있어 지진과 같은 대형 재난 사고 시 신속히 상황을 파악하는데 사용될 수 있다. 또한, 위성영상은 전 지구를 주기적으로 촬영 가능해 전 세계에서 일어나는 변화를 관찰하는 데 유용하다.

위성 정보 서비스 분야 시장 규모는 매년 크게 증가하고 있으며 다양한 공공, 민간 서비스 수요에 대응하기 위해 위성 개발이 또한 확대되고 있다. 이처럼 기하급수적으로 증가하는 위성 정보를 효율적으로 분석, 활용하기 위해서는 AI 기술 개발이 꼭 필요한 실정이다.

한국항공우주연구원(이하 항우연)은 국내 위성영상을 활용한 AI 분석 기술 발전을 위해 한국정보화진흥원의 "인공지능 학습용 데이터 구축 사업"내에서 위성영상 AI 분석 전문기업인 (주)에스아이에이(SIA), 위성영상 처리 전문기업 (주)에스아이아이에스(SIIS), 그리고 데이터 가공 전문기업인 (주)슈퍼브에이아이(Superb AI)와 함께 아리랑 위성영상을 활용한 AI 학습 데이터셋 5종을 구축하였다.

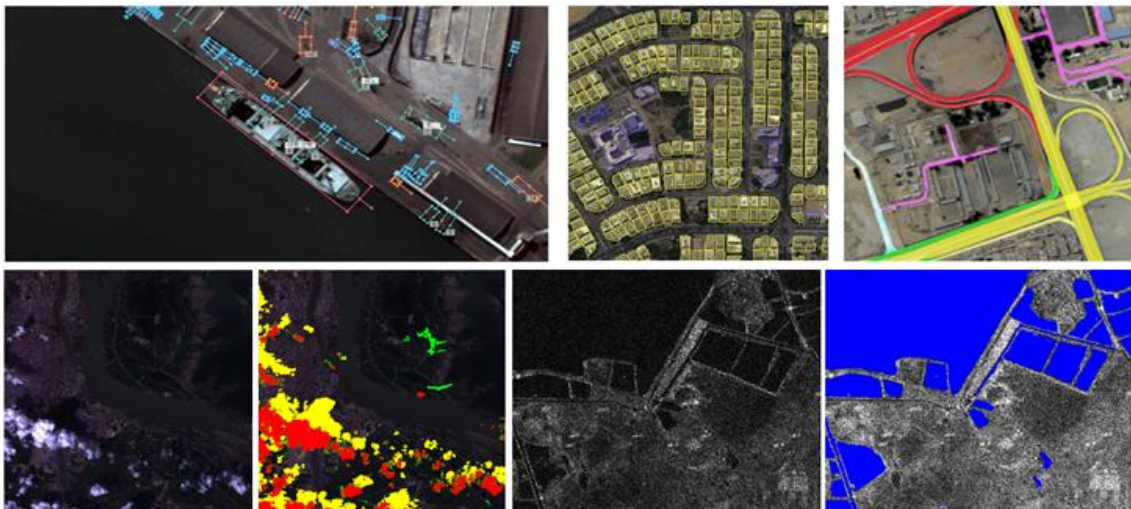


그림 1 위성영상 객체판독 AI 데이터 예시

데이터의 구성

본 데이터셋은 표 1과 같이 아리랑 3호/3A호에서 제공하는 고해상도 광학 영상과 아리랑 5호에서 제공하는 레이더 영상을 이용하여 범용의 위성정보 AI 데이터셋 5종으로 구성된다. 데이터별 구축 수량은 비슷한 종류의 해외 데이터셋을 참고하여 정하였으며, 단일 데이터셋으로는 세계 최대 규모를 자랑한다.

순번	이름	활용위성	종류	구축수량	용도
1	관심 객체 검출	3호/3A호(광학)	객체검출	15종+, 50만건+	선박, 비행기, 자동차 등 다양한 위성정보 획득
2	건물 윤곽 추출 및 종류 구별	3호/3A호(광학)	영상분할	20만건+	변화탐지·재난재해 대응· 지도제작
3	도로 윤곽 추출 및 종류 구별	3호/3A호(광학)	영상분할	6000km+	변화탐지·재난재해 대응· 지도제작
4	위성영상 전처리를 위한 구름 검출	3호/3A호(광학)	영상분할	패치단위 4000장+	무효(구름) 화소 제거
5	레이더영상을 이용한 전천후 수계 검출	5호(레이더)	영상분할	패치단위 2400장+	전천후 홍수·가뭄 감시

표 2 아리랑 위성영상을 이용한 범용의 다양한 위성정보 시데이터 구축 (5종)

데이터셋의 설계 기준과 분포

관심 객체 검출

관심 객체 클래스는 위성영상 관련 학회 회원들을 대상으로 설문 조사 결과(그림 2)와 해외 데이터셋을 벤치마킹 하여 결정했다. 또한 사용자가 다양한 객체 정보를 활용할 수 있도록 15종의 목표 클래스 이외에 추가 클래스(6종)도 확보했다.

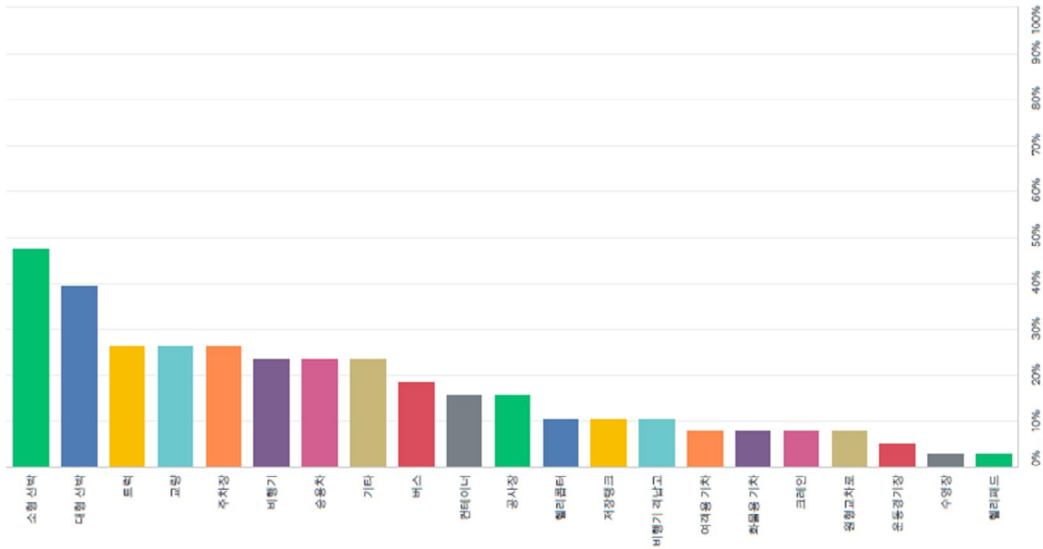


그림 2 클래스 결정을 위한 학회 설문 조사 결과

그림 3은 본 데이터셋에서 정의한 클래스를 나타낸다. 기존의 소형승용차, 트럭, 버스, 기차, 소형선박, 대형선박, 민간비행기, 군용비행기, 다리, 원형교차로, 정유탱크, 크레인, 댐, 헬리패드, 운동경기장(총 15종) 이외에 추가 클래스로 헬기, 수영장, 실외운동장, 개별 컨테이너, 그룹 컨테이너, 기타 (총 6종)을 가진다.

Project Configuration



Data Type	Image	
Annotation Type	Tiltedbox	
Object Class	<ul style="list-style-type: none"> ● 소형 선박 Tiltedbox ● 대형 선박 Tiltedbox ● 민간 항공기 Tiltedbox ● 군용 항공기 Tiltedbox ● 소형 승용차 Tiltedbox ● 버스 Tiltedbox ● 트럭 Tiltedbox ● 기차 Tiltedbox ● 크레인 Tiltedbox ● 다리 Tiltedbox 	<ul style="list-style-type: none"> ● 정유탱크 Tiltedbox ● 댐 Tiltedbox ● 운동경기장 Tiltedbox ● 헬리패드 Tiltedbox ● 원형 교차로 Tiltedbox ● 기타 Tiltedbox ● 헬기 Tiltedbox ● 실외운동장 Tiltedbox ● 수영장 Tiltedbox ● 개별컨테이너 Tiltedbox ● 그룹컨테이너 Tiltedbox

그림 3 객체의 종류들 (15종 + 추가 라벨)

건물 윤곽 추출 및 종류 구별

현재 공개 데이터셋은 건물 윤곽 데이터만 있지만 이번 과제에서는 활용성을 높이기 위해서 건물의 크기와 분포를 고려하여 소형건물, 아파트, 공장, 중형건물, 대형건물(5종)로 나누었다. 예를 들어 소형 건물의 경우 30m 미만의 단독 주택으로 정의되며 위성영상은 GSD가 존재하여 크기를 알 수 있으므로 크기로 분류가 가능했다. 또한 4개 도시 (미국 LA, 중국 상해, 독일 볼프스부르크, 이집트 뉴카이로)에서 취득한 영상으로 국가별/대륙별 다양한 건물 형태를 데이터셋에 담았다.

도로 윤곽 추출 및 종류 구별

도로 데이터셋은 Open Street Map(OSM)을 기반으로 데이터셋을 구축하였다. 국제적으로 널리 활용되는 OSM의 클래스를 기준으로 다른 데이터셋의 클래스와 호환성을 고려하여 제작하였으며 취득 국가는 건물과 동일하게 4개국에서 취득하였다.

위성영상 전처리를 위한 구름 검출

구름 데이터셋은 아리랑위성 3/3A호(광학)의 다중 분광 (Multi-Spectral) 채널 영상으로부터 제작되었다. 아리랑위성 광학영상은 Red, Green, Blue, NIR의 4개 채널을 촬영하며, 해상도는 각각 2.8 m와 2.2 m의 고해상도를 갖는다. 구름은 영상 전반에 불규칙적으로 존재하여 지표 정보 습득을 방해하는 요소로 지표 정보 습득 여부에 따라, 두꺼운 구름, 얇은 구름, 구름 그림자, 청천으로 각 화소를 정의하였다. 영상에서 구름 비율이 50 % 이상으로 존재하는 경우, 영상 정합 품질에 영향을 미치므로, 20 ~ 50 % 수준으로 구름이 존재하는 영상을 데이터셋 구축 대상으로 삼았다. 또한, 구름은 지역에 따라 그 특징이 다양하게 나타나므로, 다양한 구름 데이터셋 구축을 위해 지면 분류가 농경지, 도심지, 초지, 습지, 눈/얼음 등 서로 다른 영상을 선정하였다.

레이더영상을 이용한 전천후 수계 검출

수계 검출 데이터셋은 아리랑위성 5호(레이더) 영상으로부터 제작되도록 하였다. 아리랑위성 5호 영상은 촬영모드, 편파, 처리레벨에 따라 상이한 특징을 갖는데, 수계 검출 데이터셋을 위한 촬영모드, 편파, 처리레벨은 각각 ES 모드, HH 편파, L1D 레벨로 선정되었다. 또한, 촬영시기와 촬영지역에 대해 다양한 분포를 갖도록 아리랑위성 5호 영상이 수집되었고, 딥러닝 모델의 입력 크기를 적절히 조절하기 위해 수집된 아리랑위성 5호 영상으로부터 수계가 포함된 1,024 x 1,024 크기의 패치들이 추출되었다. 각 패치는 도심지, 농경지, 초지, 산지, 습지 등의 다양한 지역에서 추출되었으며, 철도, 도로, 건물 그림자 등의 다양한 객체를 포함 가능하도록 설계되었다.

데이터의 구조

관심 객체 검출/건물 윤곽 추출/도로 윤곽 추출

객체/건물/도로는 표 2와 같은 구조를 지니며 객체 영상의 경우에는 객체의 바운딩 박스(bounding box)와 관련된 정보가 추가되고 건물/도로는 폴리곤(polygon) 정보가 추가된다.

	정보 종류	형태
영상(.tif) 및 보조자료(.kml)	위경도 좌표	숫자 - 패치 영상의 위경도 좌표 (wgs84)
	위성 이름	텍스트
정답지 (.json)	위경도 좌표	숫자 - 객체별 위경도 좌표 (wgs84)
	타입	텍스트 - 형태 (ex. Polygon)
	객체 위치	숫자 - Polygon 또는 Bounding Box 좌표값
	객체 각도	숫자 - Bounding Box 각도 (객체 파트 경우만)
	객체 ID	정수 - 형태의 Class 번호
	객체 이름	텍스트 - ID에 해당되는 이름
	취득일	숫자 - 영상 획득일

표 3 객체/건물/도로의 라벨 종류와 형태

위성영상 전처리를 위한 구름 검출

구름 데이터셋은 패치 단위가 아닌 전체 영상 단위로 제공되며, 원본영상(GeoTIFF), RGB 합성영상, NGR 합성영상 이미지 파일(PNG 형식), 구름 라벨 파일(PNG 형식)으로 이루어져 있다. 구름 라벨은 4가지 색으로 구성되어 있으며, 빨간색 라벨은 두꺼운 구름, 녹색 라벨은 얇은 구름, 노란색 라벨은 구름 그림자를 검정색은 청천 영역을 의미한다.

레이더영상을 이용한 전천후 수계 검출

수계 검출 데이터셋은 아리랑위성 5호 영상으로부터 추출한 패치와 각 패치로부터 제작된 레이블로 구성되어 있으며, 총 2,400 쌍 이상이 제공된다. 패치 및 레이블은 비트맵(GeoTIFF)

형식으로 제공된다.

데이터의 예시

관심 객체 검출/건물 윤곽 추출/도로 윤곽 추출

영상은 공통적으로 영상의 위경도 등의 정보를 나타내는 kml 파일을 가진다.

```
<?xmlversion="1.0"encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2" xmlns:gx="http://www.google.com/kml/ext/2.2" xmlns:kml="http://www.opengis.net/kml/2.2" xmlns:atom="http://www.w3.org/2005/Atom">
<Document>
  <name>KOMPSAT Product</name>
  <open>1</open>
  <Style id="strip_boundary">
    <LineStyle>
      <color>ffffff00</color>
    </LineStyle>
    <PolyStyle>
      <fill>0</fill>
    </PolyStyle>
  </Style>
  <Style id="scene_boundary">
    <LineStyle>
      <color>ffffff55</color>
    </LineStyle>
    <PolyStyle>
      <fill>0</fill>
    </PolyStyle>
  </Style>
  <Folder>
    <name>Patch - BLD00001_PS3_K3A_NIA0276</name>
    <open>1</open>
    <GroundOverlay>
      <name>Image</name>
      <Icon>
        <href>BLD00001_PS3_K3A_NIA0276.PNG</href>
        <scale>1.0</scale>
      </Icon>
      <gx:LatLonQuad>
        <coordinates>
31.4347031225,30.0413951468,0 31.4405428056,30.0414645232,0 31.4404632046
,30.0465452603,0 31.4346232236,30.0464758698,0
        </coordinates>
      </gx:LatLonQuad>
    </GroundOverlay>
  </Folder>
</Document>
</kml>
```

객체/건물/도로의 json 파일의 구조이며 객체일 경우에는 “object_imcoords”와 “object_angle” 가 활성화 되며, building_imcoords와 road_imcoords는 EMPTY 상태가 된다. 다른 파트에서 활용될 때는 반대로 해당 파트가 활성화가 되며 polygon 값이 들어간다.

```

{"features": [{
  "geometry": {
    "coordinates": [[139.7704781771, 35.5525475742, 0.0], [139.778385
0987, 35.552627952, 0.0], [139.778287044, 35.5590895857, 0.0], [139.77037
94881, 35.5590091889, 0.0]],
    "type": "Polygon"},
  "properties": {
    "object_imcoords": "469.1820926230315,473.2955533775398,561.65439
96724264,473.2955533775399,561.6543996724264,563.5489576573916,469.182092
6230315,563.5489576573916",
    "object_angle": 6.283185307179586,
    "building_imcoords": "EMPTY",
    "road_imcoords": "EMPTY",
    "image_id": "OBJ00013_PS3_K3_NIA0078.png",
    "ingest_time": "2020-11-20T03:01:18.829265Z",
    "type_id": "3",
    "type_name": "civilian aircraft"},
    "type": "Feature"},
{"geometry": { ...
...
}]

```

위성영상 전처리를 위한 구름 검출

그림 4는 구축된 구름 검출 데이터의 예시이다. 각 쌍의 좌측 그림은 아리랑위성의 RGB 합성영상, 우측그림은 구름 레이블이다. 빨간색은 두꺼운 구름, 초록색은 얇은 구름, 노란색은 구름 그림자, 검정색은 청천을 의미한다.



그림 4 구름 검출 데이터 예시

레이더영상을 이용한 전천후 수계 검출

그림 5는 구축된 수계 검출 데이터의 예시이다. 예시는 총 4 쌍의 데이터가 도시되어 있으며, 각 쌍의 좌측그림은 아리랑위성 5호 영상 패치, 각 쌍의 우측그림은 레이블이다.

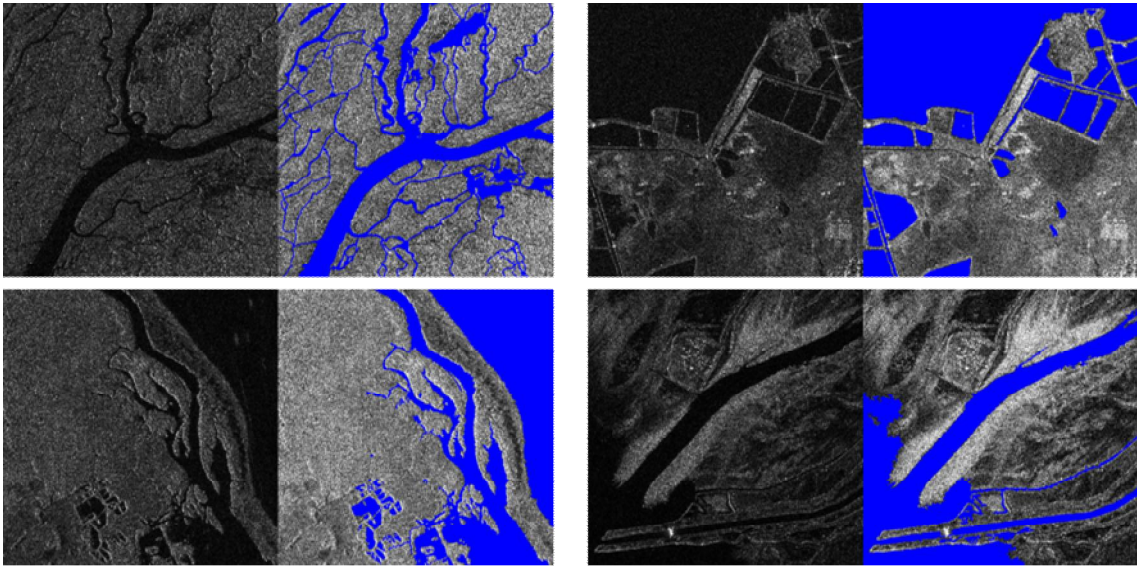


그림 5 수계 검출 데이터 예시

데이터 구축과정

원천 데이터는 항우연이 보유중인 영상을 위성영상 처리 전문 업체인 (주)에스아이아이에스(SIIS)가 그림 6과 같은 방법으로 수집·정제하였다.



그림 6 원천 데이터 수집·정제 방법

데이터 가공 도구는 (주)슈퍼브에이아이(Superb AI)가 위성영상에 특화된 데이터별 가공 도구를 신규 개발하였으며, 클라우드 기반으로 대규모 인원이 웹상에서 동시에 가공/검수가 가능한 특징을 가진다.

관심 객체 검출/건물 윤곽 추출/도로 윤곽 추출

관심 객체 검출의 경우 희귀 객체(헬기와 같은 출현 빈도가 낮은 객체) 위주로 영상을 선정하여 그 영상 내 관심 객체를 모두 라벨링하는 방식으로 가공하였다. 건물과 도로 데이터는 참고 자료로 해당 위치의 Google Earth, Planet Explorer, OpenStreetMap 등의 정보를 이용해 구축하였다.

위성영상 전처리를 위한 구름 검출

구름 검출 데이터 제작에는 RGB 합성영상과 NGR 합성영상 두 개의 합성영상을 동시에 활용하였다. RGB 합성영상은 사람이 보는 영상과 동일한 형태를 가지며, NGR 합성영상의 경우 구름에 의해 발생한 그림자가 두드러지게 나타나 검출에 유용하다는 특징이 있다. 따라서,

두 영상을 동시에 활용하여 사람이 직접 포토샵 툴을 활용하여 구름 검출을 진행하였다. 일관성 있는 작업 진행을 위하여, 구름 아래의 지면 특성 판별 여부에 따라 구름을 두꺼운 구름과 얇은 구름으로 정의하였다. 또한 작업자간 상호검수 및 전문가 검수를 통해 작업물의 일관성 확보를 꾀하였다.

레이더영상을 이용한 전천후 수계 검출

수계 패치를 이용하여 레이블 데이터 제작 시, 각 패치는 수계 1종에 대하여 이진 분류가 수행되었다. 수계 지역에 대한 위치 정보를 획득하기 위해, 구글어스, 오픈스트리트맵 등 수계 표시가 가능한 오픈데이터를 활용하였다. 더불어, 오픈데이터와 수계패치의 획득 시기에 따른 수계 차이를 보정하기 위해, Planet Explorer를 활용하였다. 레이블링 작업은 전문 인력들이 포토샵 툴을 활용하여 직접 수행하였다.

검수와 품질 확보

위성 영상은 위성의 해상도 한계로 객체를 판독하기가 쉽지 않아 일반 영상보다 데이터를 가공하는 데 시간이 많이 소요되며 작업자의 숙련도와 피로도에 따라 오프기나 미표기와 같은 데이터의 품질에 문제가 생길 가능성 크다. 그래서 본 데이터 구축 사업에서는 높은 데이터 품질 확보를 위해 AI 모델 개발 기관인 항우연과 (주)에스아이에이(SIA)가 데이터별로 품질을 전적으로 담당하여 AI 모델과 육안을 통해 전수 검수를 하고 목표하는 품질이 나올 때까지 데이터를 수정·보완 하였다. 이러한 내부 반복 검수 프로세스에 합격한 데이터는 마지막으로 외부 검수 기관인 한국정보통신기술협회(TTA)의 검수를 한 번 더 다 받고 최종 공개된다.



그림 7 품질 확보를 위한 4단계 데이터 검수 프로세스

데이터 구축 담당자

수행기관(주관): 한국항공우주연구원

담당자명	전화번호	이메일	담당업무
오 한	042-870-3914	ohhan@kari.re.kr	과제 총괄, 구름/수계 데이터 설계 및 AI모델 개발

수행기관(참여)

기관명	담당업무
(주)에스아이에이(SIA)	객체/건물/도로 데이터 설계 및 AI 모델 개발, 레이더영상 수계 데이터 가공
(주)에스아이아이에스(SIIS)	원천 데이터 (위성영상) 수집 및 정제
(주)슈퍼브에이아이(Superb AI)	광학영상(객체/건물/도로/구름) 데이터 가공