

Dataset Documentation

Dataset Name: Semantic Segmentation of Crop Type in Ghana

Location and boundaries

Overall Location Method

- Ground collection only
- Ground collection with boundary drawn using imagery
- Ground collection with spatial buffer added
- Boundary drawn from imagery
- Other _____
- Unknown

GeoLocation Device

- Industrial grade GPS (List model) _____
- Retail grade GPS
- Mobile Phone GPS
- N/A
- Unknown

Ground Boundary Method (Details explained in Appendix A)

- Live/Continuous point capture of walk-around
- Manual point capture of walk-around
- Manual point capture of polygon boundaries (not whole field)
- Manual point capture for later image annotation
- Manual point capture for spatial buffer within field
- Manual point capture while looking at but not in field, with heading recorded
- Other _____
- Unknown

Imagery used (Skip if no imagery used)

Sensor: _____

Date(s): _____

List scenes used in Appendix B

Imagery Annotation methods

- Boundaries drawn based on a single ground point captured
- Boundaries drawn/edited based on multiple ground points captured
- Buffer validated from ground point captured
- Boundary drawn without ground reference data (Include description of methods in Appendix C)
- Pixels annotated without ground reference data (Include description of methods in Appendix C)
- Unknown

Boundary inclusion

- Captured polygon includes the entire field/area
- Captured polygon includes only a sample of the field/area

Classification

Classification Type

- Land cover
- Crop type
- Other _____

Classes/fields used

Described in Appendix D

Ground Referenced Classification (TBD)

- Observation (Describe methods of determination in Appendix E)
- Survey/interview with land holder (Describe methods in Appendix E)
- Other (Describe methods in Appendix E)

Image Referenced Classification

Describe methods used in Appendix C

Image Description

Describe methods used in Appendix F

Data Properties

Property name	Property Description	Parameters/Allowed responses (optional)
field_id	A unique integer value for each identified field	
label	Crop type	See Appendix D

Appendix A: Describe the method of geographic ground data collection

Appendix B: List imagery scenes used for annotation (ideally also included in metadata)

N/A - Imagery was not used for annotation.

Appendix C: Describe how boundaries and classes were determined without ground reference data

Boundaries and classes were determined with ground truth data.

Appendix D: List all top-level classes or the classification guidance used

Pixel Value	Class Type
0	unknown
1	ground nut
2	maize
3	rice
4	soya bean
5	yam
6	intercrop
7	sorghum
8	okra
9	cassava
10	millet
11	tomato
12	cowpea
13	sweet potato
14	babala beans
15	salad vegetables
16	bra and ayoyo
17	watermelon
18	zabla
19	nili
20	kpalika
21	cotton
22	akata
23	nyenabe
24	pepper

Appendix E: Describe methods for determining classes based on direct/ground observation

To be added.

Appendix F: Describe imagery paired with the ground truth observations

The following files exist for each image tile:

Ground truth

- A .tif file with a ground truth array of integers that correspond to cover type, as described in Appendix D. The dimension of the array is: [rows, cols]

Sentinel-1

- A .tif file for each Sentinel-1 imagery in the time series. The dimension of the array is: [rows, cols, time]

Sentinel-2

- A .tif file for each Sentinel-2 imagery in the time series. The dimension of the array is: [rows, cols, time]
- A .tif file for cloud mask corresponding to each Sentinel-2 imagery. The dimension is: [rows, cols, time]

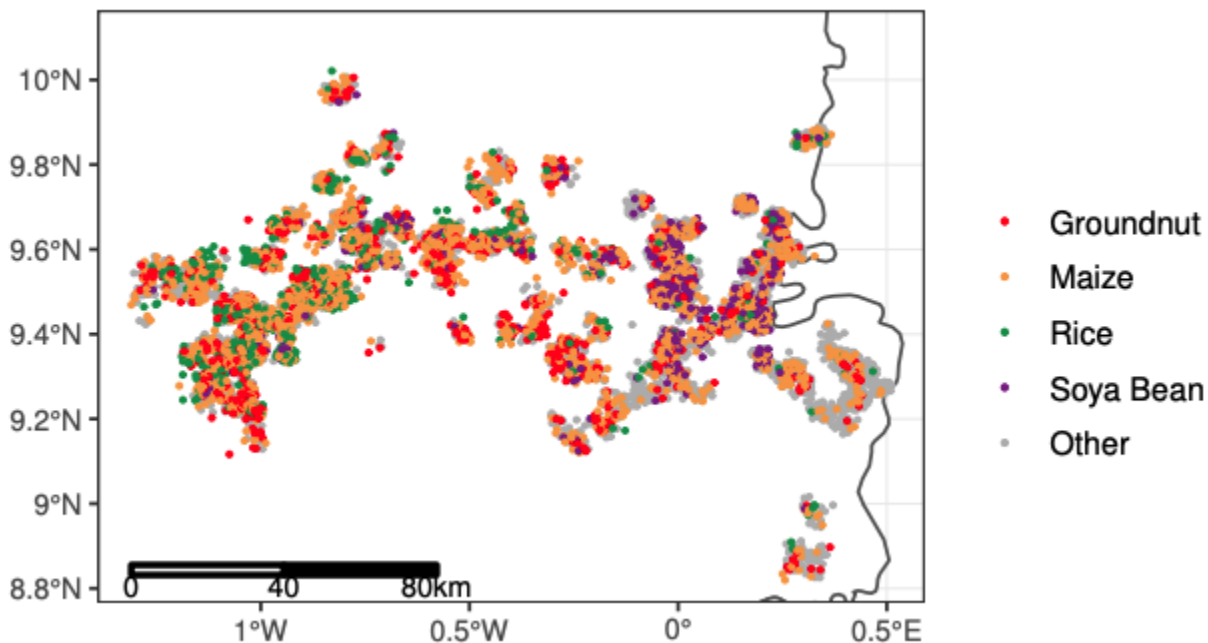
PlanetScope

- A .tif file for each Planet imagery in the time series. The dimension of the array is: [rows, cols, time]

All Sentinel-1 and Sentinel-2 time series span an entire year, from January 1st - December 31st, 2016. PlanetScope time series span an entire year, from January 1st - December 31st, 2017, due to a lack of data availability in 2016. This makes the assumption that cover type remains constant from 2016 to 2017, which may not always be true. Planet imagery was filtered to 10% or less cloud cover, while all Sentinel-1 and Sentinel-2 scenes were kept. A small amount of random noise was added to all satellite imagery.

Each tile is identified with a unique 6-digit integer identifier, which has been randomized in order to preserve location anonymity.

Appendix G: Spatial distribution of the dataset



Distribution of crop type in Ghana. Up to 3km of random jitter have been added to locations for privacy.