

Dataset Documentation

Dataset Name: Semantic Segmentation of Crop Type in South Sudan

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

- 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	Land cover or 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

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

Pixel Value	Class Type
0	unknown
1	sorghum
2	maize
3	rice
4	groundnut
5	sesame
6	bush
7	bare soil
8	forest
9	shrubland
10	grasslands and meadows
11	water bodies
12	built-up surface
13	wetland
14	fallow
15	intercrop
16	bulrush millet
17	beans
18	cassave

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

Land cover types were determined by visual inspection of direct ground observation. Polygons containing more than one crop were labeled as intercrop.

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]

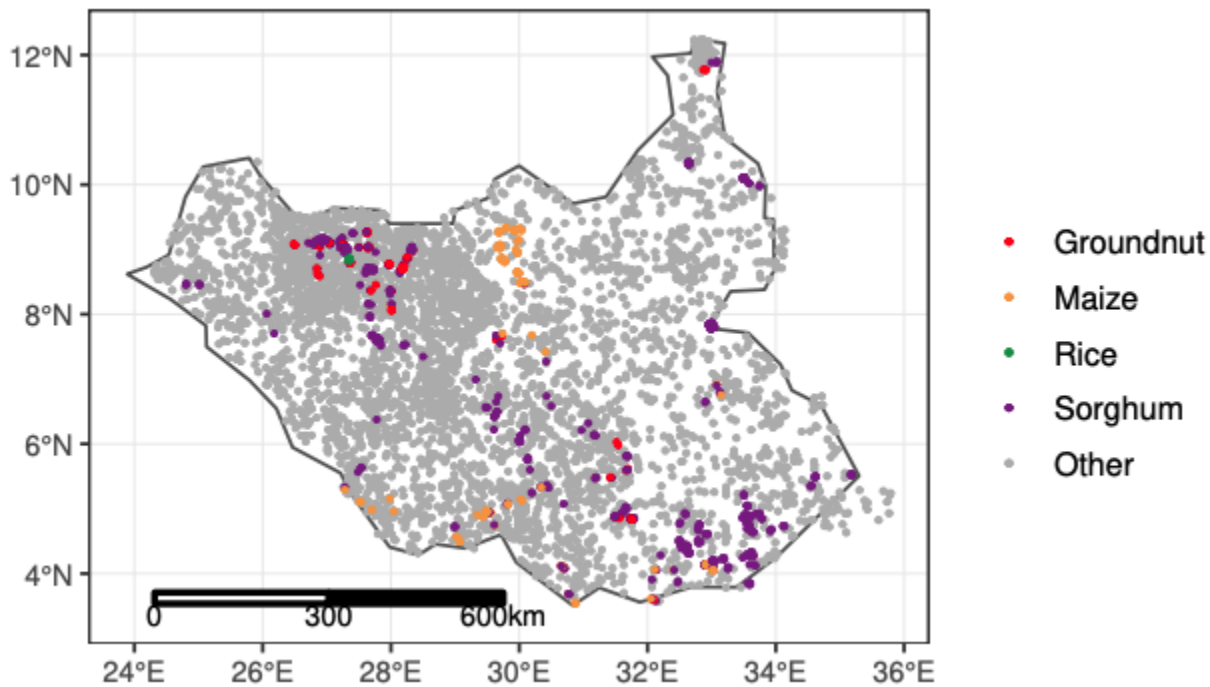
PlanetScope

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

All satellite imagery time series span an entire year, from January 1st - December 31st, 2017. Planet imagery was filtered to 10% or less cloud cover, while all Sentinel-1 and Sentinel-2 scenes were kept.

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 South Sudan. Up to 3km of random jitter have been added to locations for privacy.