Radiant MLHub
EARTH IMAGERY FOR IMPACT

Dataset Documentation

Dataset Name: East Africa Agricultural Field Centers Dataset

Description

Georeferenced crop yield prediction is a valuable tool for agronomists and policymakers. One challenge with many existing datasets is that of location accuracy. GPS locations for fields can end up offset from the true location due to sensor inaccuracies or from locations being collected at the edges of fields rather than the field centers. This makes it harder to connect remote-sensed data to the yield values. The goal of this project was to produce a method that can help correct these location offsets by finding the most probable field center given an input location.

Citation

Amer, Karim; Eissa, Kareem (2021): GPS Coordinates of 18,482 Crop Fields in East Africa with Improved Accuracy using Planet Imagery and Yolo v5 Object Detection Model. figshare. Dataset. https://doi.org/10.6084/m9.figshare.15157263.v1

License

Creator

CGIAR

Contact

D.GUERENA@cgiar.org

Radiant MLHub
EARTH IMAGERY FOR IMPACT

Location and boundaries

Overall Location Method

- [x] 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
- [x] 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
- [x] Other N/A
- [ ] Unknown

Imagery used (Skip if no imagery used)

Sensor: PlanetScope

Date(s): June 2017, December 2017, June 2018, December 2018

List scenes used in Appendix B

Radiant MLHub
EARTH IMAGERY FOR IMPACT

Data Properties

Property name	Property Description	Parameters/Allowed responses (optional)
ImgID	*Corresponding ID to the images in the data folder. For each season in June'17, December'17, June'18, and December'18.*	
Quality	*Human annotated confidence score of the correct field centers.*	*0: Unavailable* *1: Low Quality, Best Guess Candidate Field* *2: Medium Quality, High Confidence Candidate Field* *3: High Quality, Almost Certain Candidate Field* *Only available for manually annotated data.*
human_lat	Human annotated latitude of the correct field centers. Only available for manually annotated data.	
human_lon	Human annotated longitude of the correct field centers. Only available for manually annotated data.	
model_lat	Model predicted latitude for the correct field centers.	
model_lon	Model predicted longitude for the correct field centers.	

Appendix A: Describe the method of geographic ground data collection

We have a set of GPS coordinates corresponding to the centers of maize fields. There were some issues in the data collection process which sometimes resulted in GPS coordinates that do not precisely coincide with their correct maize field centers. The errors in the GPS coordinates are due to recording on the edges of the field rather than the center, or in the house which owns the farm, or under the shade of a nearby tree, or even on the main road leading to the farm. Only unique datapoints of (longitude, latitude, plot-size) triples are kept, while duplicate ones are dropped. The problem we are trying to solve is to correctly identify the original field centers given the GPS coordinates.

Manual Annotation

The GPS coordinates are projected on the map as shown in the figures below. Two target icons are drawn on the map; black indicates original position and red indicates corrected position. The area of these icons is calculated from the "plot

size" variable but may not exactly match its size. Note that the icons are always circular but the field itself may be an elongated rectangular stretch or even triangular. Also, the corrected (red) field location may be somewhat far from the original (black) Location.

Figure 1 Example of manual annotation

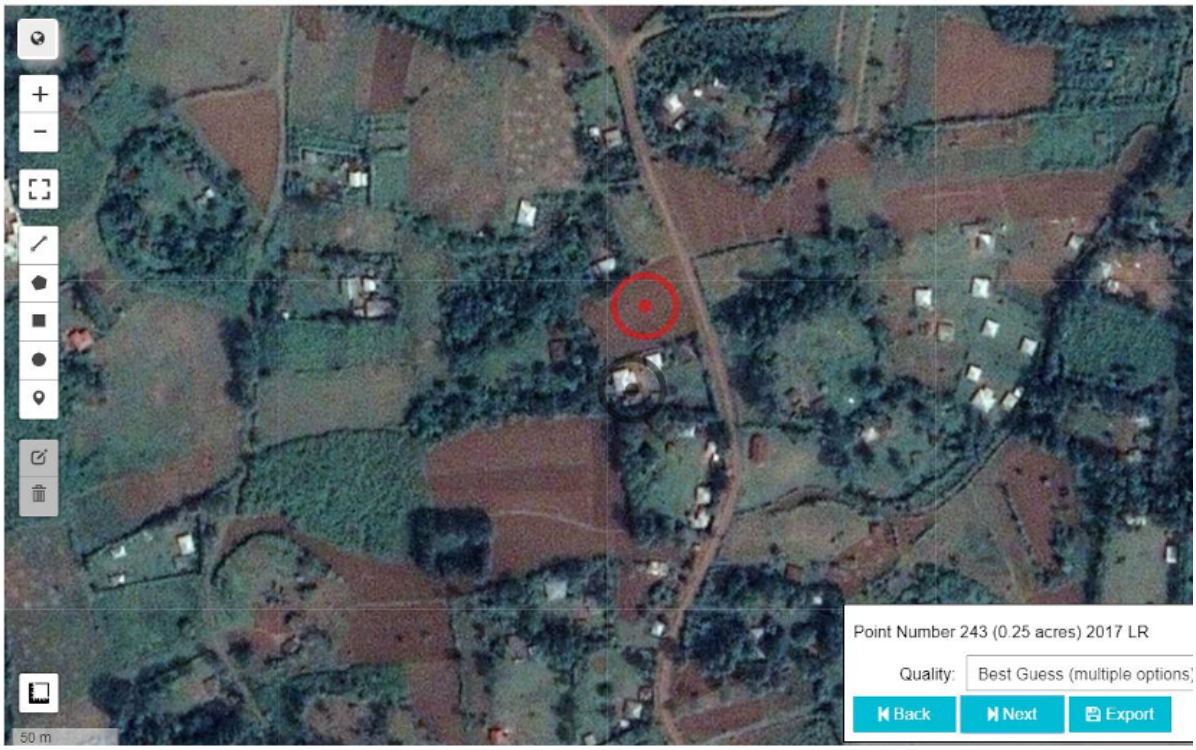

Figure 2 Example of manual annotation

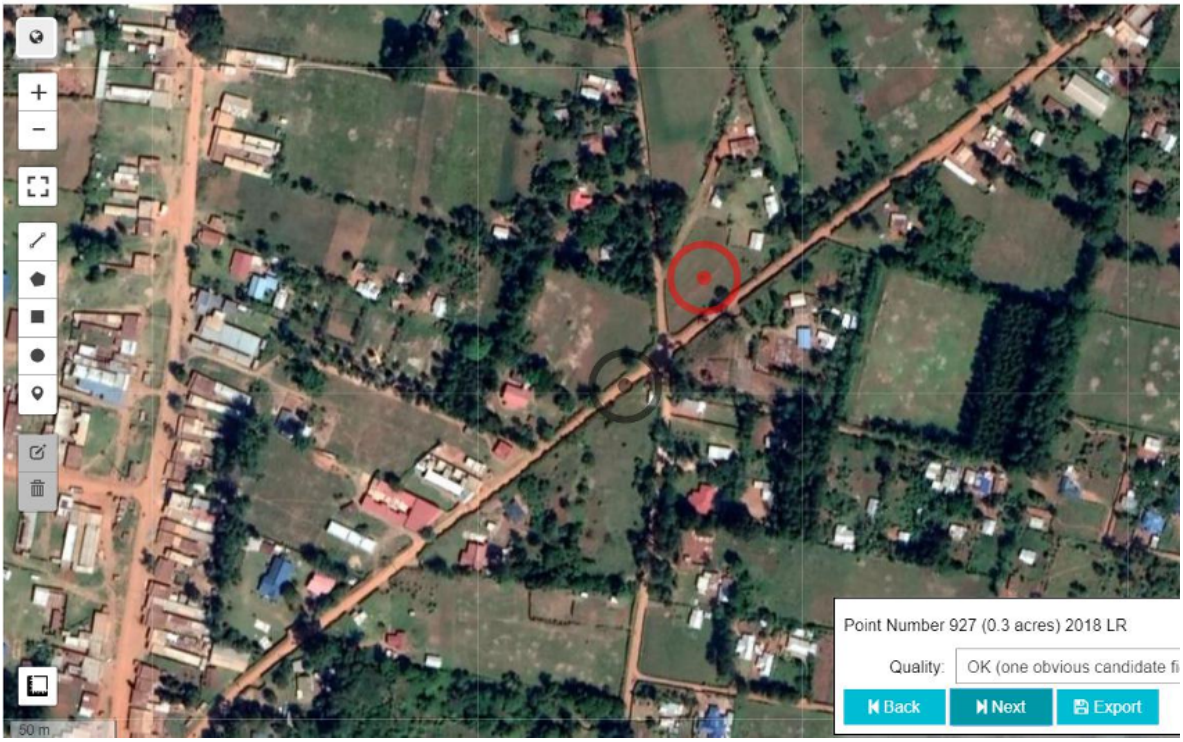

Figure 3 Example of manual annotation

Challenges

The figures 1, 2, and 3 demonstrate some of the challenges in this problem. The image resolution in the competition data is lower than these figures, making it harder to identify the fields. We couldn't use the same high-resolution images due to licensing issues. Furthermore, sometimes there are multiple possible candidate target fields and the annotator has to make a best guess based on size and proximity of the field. This is captured by the "Quality" variable in the data. Moreover, some field workers record multiple data points in the same spot making plot size the only distinguishing variable.

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

These are images captured using Planet Lab Satellite in ~4.7 m resolution in different timestamps: June'17, December'17, June'18, and December'18. Each image is 84*84 pixels and has RGB channels only. It should help identify the farm boundary and other objects inside beside Maize like houses and trees.

Appendix C: Dataset Structure

The dataset consists of 18,481 agricultural field center labels (GeoJSON) along with 73,924 PlanetScope PNG image chips (4 for each label). Each label has a unique identifier in the form of 8 alphanumeric characters. The matching PlanetScope chips have this unique identifier followed by the month and year that the PlanetScope imagery was captured in (jun17, dec17, jun18, or dec18). The geographic extent for the PlanetScope imagery matches the geographic extent of the label item and can also be found in the STAC metadata for the PlanetScope imagery item.