

Street2Sat: A Machine Learning Pipeline for Generating Ground-truth Geo-referenced Labeled Datasets from Street-Level Images

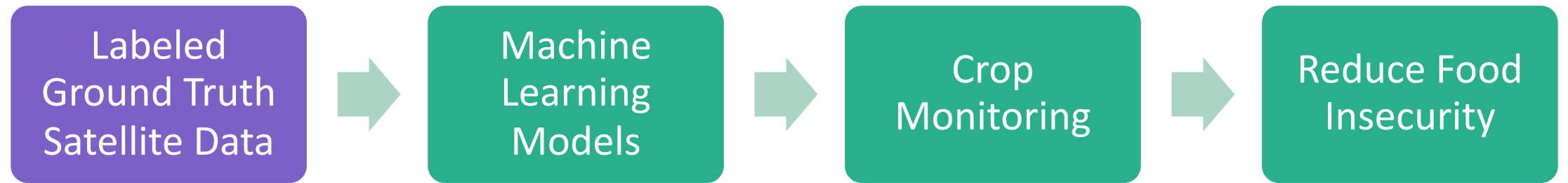
Madhava Paliyam^{1 2}, Catherine Nakalembe^{1 2}, Kevin Liu^{1 2}, Richard Nyiawung^{3 2},
Hannah Kerner^{1 2}

¹ University of Maryland, College Park, MD, USA

² NASA Harvest

³ University of Guelph, Canada.

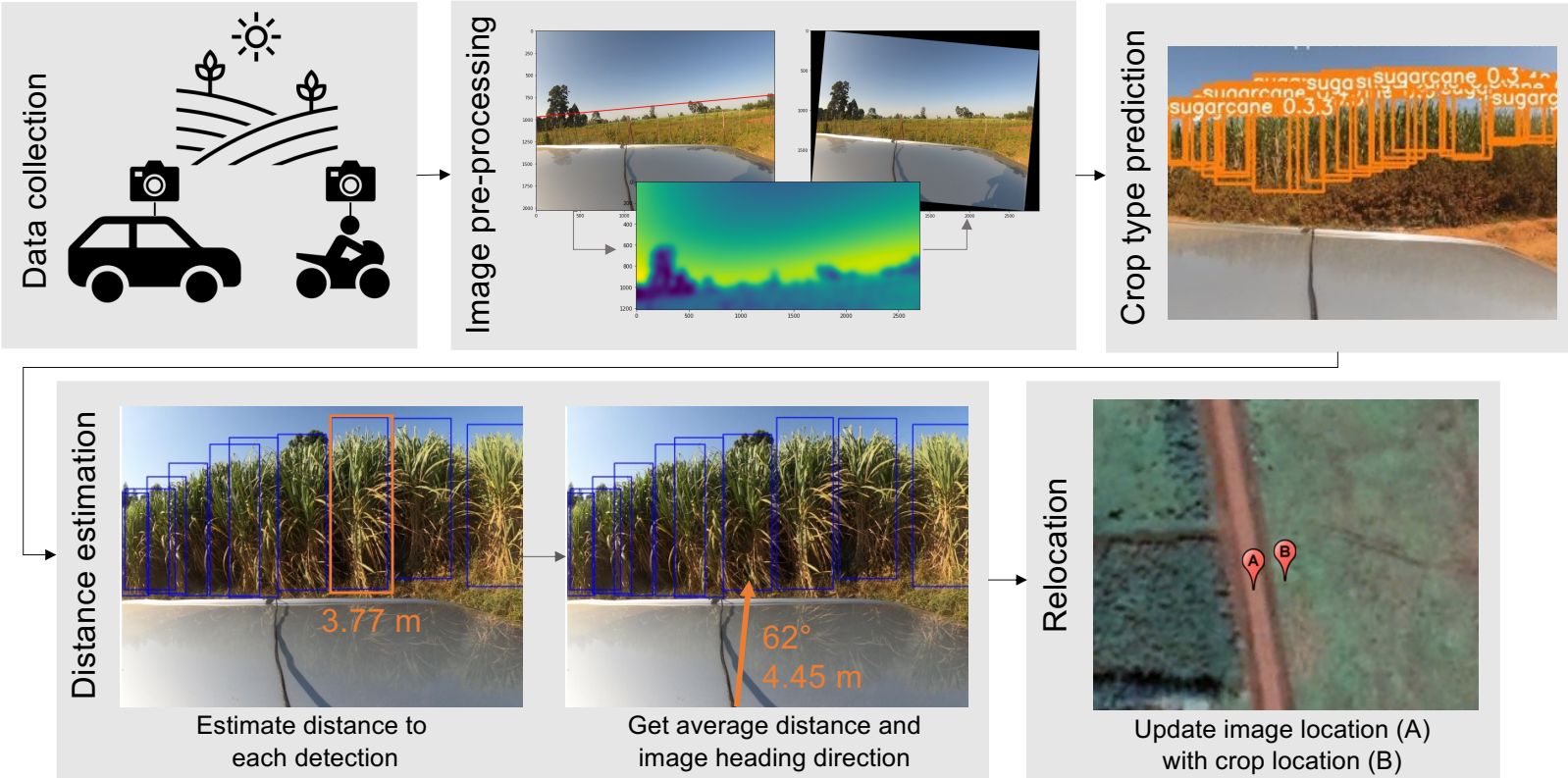
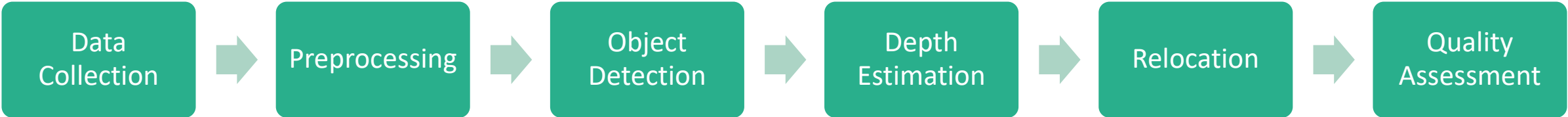
Motivation



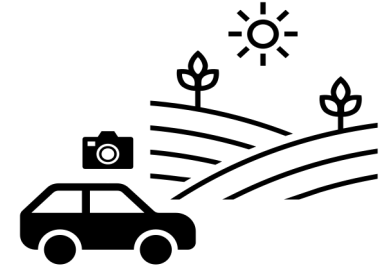
Difficult to obtain and scarce
in many areas of the world

Street2sat provides a solution to create large datasets of geo-referenced labels to be used in machine learning and other applications.

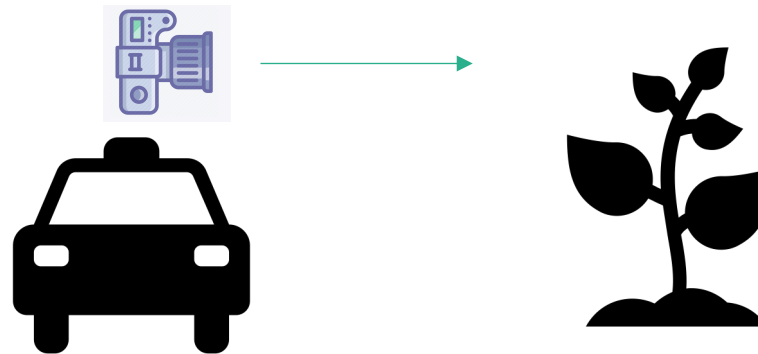
Approach



Data Collection

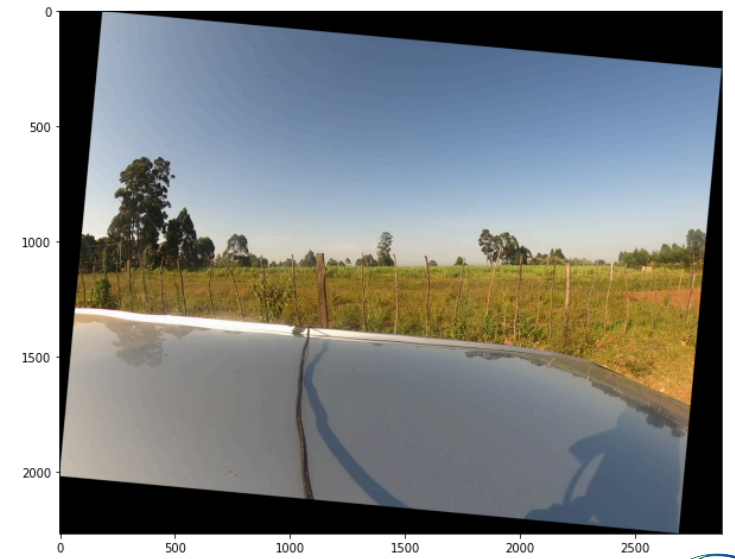
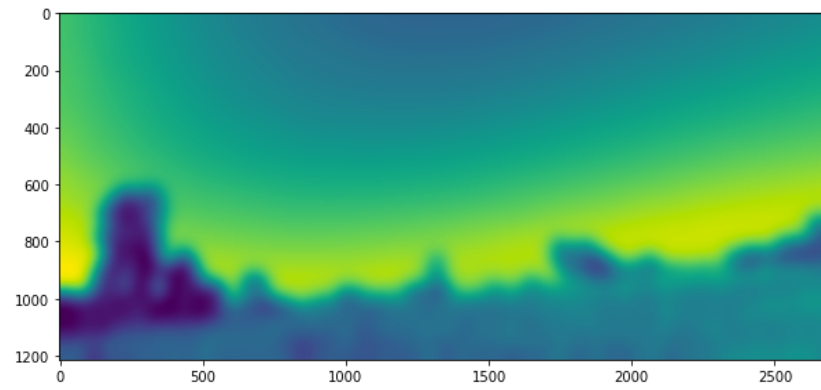
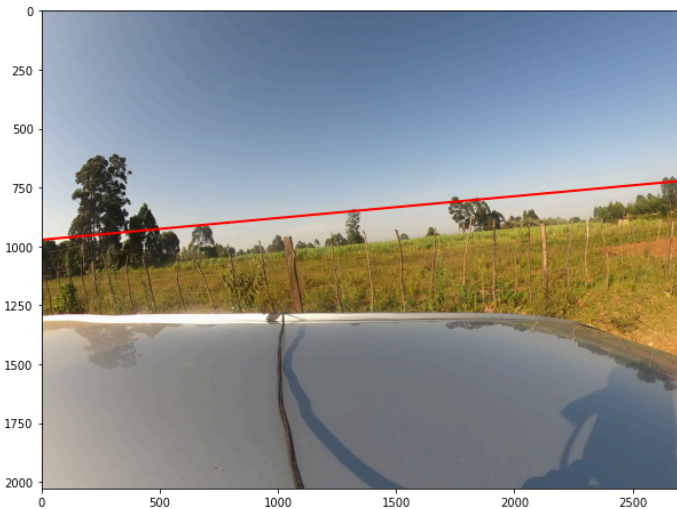


- Images from initial study collected in Western Kenya
- GoPro Max 360 cameras pointing orthogonal to the side of the car closest to the crops
- Routes mapped to ensure maximal coverage of crops



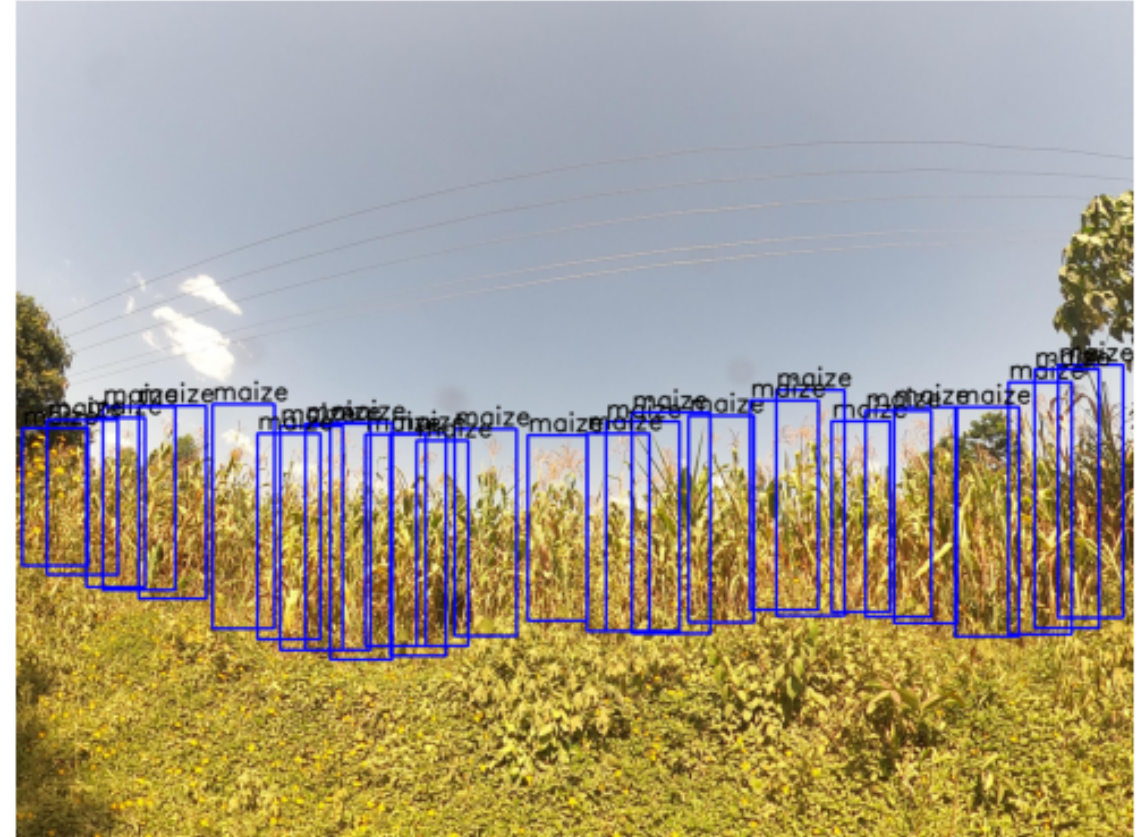
Preprocessing

- Images were straightened using Otsu's method to create training and validation sets
- Straightening was needed so that the predicted bounding boxes would have accurate height representation



Object Detection

- A custom YOLOv5 model was trained using hand labeled data
- Initialized using pretrained weights from COCO
- Fine-tuned using hand labeled data



Depth Estimation

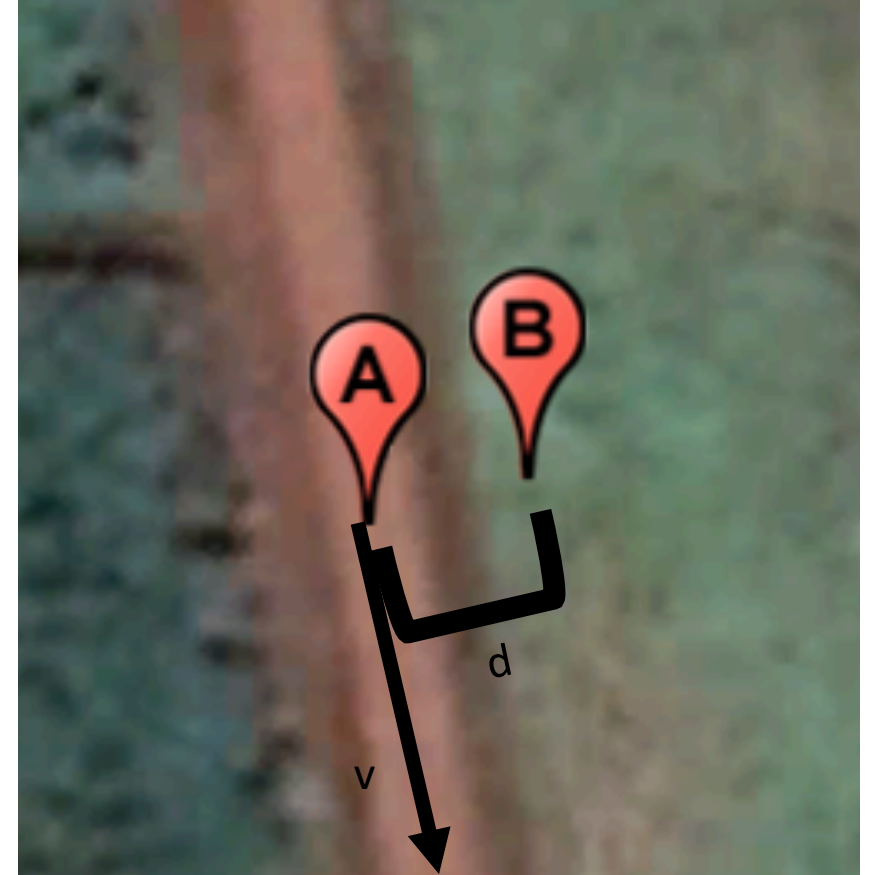
$$d = \frac{(l_{focal} * h_{crop} * h_{image})}{(h_{bbox} * h_{sensor})}$$

- l_{focal} : focal length of camera
- h_{crop} : height of crop based on lookup table
- h_{image} : image height in pixels
- h_{bbox} : detected bounding box height in pixels
- h_{sensor} : sensor height obtained from GoPro website
- d : distance to crop

In a given image, all predicted distances were averaged to get one distance per crop per image.

Relocation

- Once distance (d) is known, create a velocity vector based on the closest other image in time
- Since cameras are placed orthogonal to the drive direction, translate the image coordinates d meters 90 degrees west of the velocity vector



Experiment Setup

- Labeled 296 training images and 53 test images with expert guidance
- 755 instances of maize and 1795 instances of sugarcane in training set
- 253 maize and 229 sugarcane in test set
- Precision: 0.41, Recall: 0.59, mAP @ 0.5: 0.45, and mAP@.5-.95: 0.13
- Often more bounding boxes were predicted than labeled, which affected precision but is good for accurate crop distance.

Labels



Predictions



Quality Assessment/Quality Checks

- We did not have ground truth labels for the test set so we compared our predicted points with a 10 m/pixel crop/no crop map of Kenya
- 73/85 (86%) predicted crop points were also classified as crop on the map.
- In the future we hope to test on thousands of points in several locations
- Additional quality checks will be implemented in future work



Correspondence to:
mpaliyam@terpmail.umd.edu or
cnakalem@umd.edu.

