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]

Abstract

Ground-truth labels on crop type and other variables are critically needed to develop machine learning methods that use satellite observations to combat climate change and food insecurity. These labels are difficult and costly to obtain over large areas, particularly in Sub-Saharan Africa where they are most scarce. We propose Street2Sat, a new framework for obtaining large data sets of geo-referenced crop type labels obtained from vehicle-mounted cameras that can be extended to other applications. Using preliminary data from Kenya, we present promising results from this approach and identify future improvements to the method before operational use in 5 countries.

1. Introduction

Climate change is already impacting global food production (Wolfram & Lobell, 2010). Increasing temperatures, changing precipitation patterns, increased weather variability, and more frequent extreme events are projected to further intensify in many developing countries where food production is already challenged (Nakalembe, 2018; Parry et al., 2004). Decision makers need more accurate and timely information on what, where, and how crops are performing to assess food security situations and ground data are required to derived those insights from Earth observations (EO) data (Becker-Reshef et al., 2020; Nakalembe et al., 2021a). Farmers are also desperate for actionable information to improve their practices and break the cycle of poverty. EO derived projects on crop conditions, yield and their changes as climate changes can inform programs and policies to guide climate resilient planning, implementation, and program management that directly lead to better outcomes for farmers (Becker-Reshef et al., 2019; Nakalembe, 2018).

[1]University of Maryland, College Park, MD, USA [2]NASA Harvest [3]University of Guelph, Canada. Correspondence to: Catherine Nakalembe <cnakalem@umd.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

Recent advances in machine learning (ML), cloud computing, and EO technologies have spurred promising solutions toward addressing the UN's Sustainable Development Goals and Targets and provide critical data to inform climate resilient programs and adaptation planning (Whitcraft et al., 2019; Nakalembe, 2018). The limiting factor today is a lack of ground-truth data for training and validating ML methods that use EO data as inputs. Ground-truth data are particularly lacking in rural and underdeveloped regions that are home to most of the world's 500 million smallholder farmers (Nwanze, 2017). Smallholders produce about 80% of the food consumed in Asia and sub-Saharan Africa. They face the greatest losses due to climate change but can also make the biggest contributions toward addressing food insecurity in these regions (Nwanze, 2017; McCarthy et al., 2001; Nakalembe et al., 2021a; Nakalembe, 2020; nat, 2020).

Ground-truth data required to develop these critical information products are scarce, often collected opportunistically and on a project-to-project basis (Kerner et al., 2020; Tseng et al., 2020; 2021). There is an urgent need to collect high-quality data sets to advance ML methods that can inform programs to benefit people and ecosystems under-served regions. Given the large geographical areas, complexity of smallholder systems, and need for continuous data collection due to the changing nature of land cover/land use, transformative data collection approaches must be scalable, affordable, and equitable (Nakalembe et al., 2021b).

We propose a new method for data collection that performs automated windshield surveys by collecting images of roadside objects at regular intervals using vehicle-mounted cameras and predicting the locations and labels of objects of interest in the images. Our proposed system, called Street2Sat, transforms a set of geo-tagged images collected from a vehicle on the road to a set of labeled geo-referenced points with locations corresponding to the object(s) of interest detected in the images. These points can then be used as labels for satellite images. We focus on the use case of mapping smallholder agriculture given urgency for ensuring food security in a changing climate (Nakalembe et al., 2021b; Nakalembe, 2020; nat, 2020). However, this approach could be applied for any use case where objects of interest can be seen from a road, including monitoring housing quality, construction, wildlife, damage assessment and more. In this paper, we describe the details of our Street2Sat system and present initial results from a pilot study in western Kenya in 2020.

Our key contributions include:

- A low-cost approach to creating ground-truth geo-referenced labeled datasets from street-level images
- Open code for replicating the process for other data sets or applications[1]
- Open data set that can be used as a benchmark for crop type mapping methods

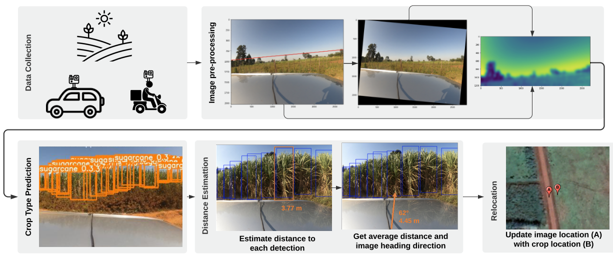

Figure 1. Street2Sat pipeline

2. Related Work

Matching ground and satellite images Several solutions for finding the satellite location or image corresponding to a ground image have been proposed in prior work. Some approaches use the image GPS location. Yan & Ryu (2021) used Google Street View images to create a data set of geo-referenced crop type points. They assigned crop type labels by classifying the crop type in the image using a Convolutional Neural Network (CNN), then used the image heading and location to move the point location from the road to the field using constant offsets. They verified their predictions by comparing to the USDA Cropland Data Layer. This is similar to our goal except they assign one crop type to the entire image and use a constant relocation offset, which is problematic for smallholder fields with irregular shapes and mixed crops. In lieu of GPS location, Viswanathan et al. (2014) warped the ground image to obtain an estimation of the satellite image. They found a whole image descriptor for the warped ground image and matched it with the most similar descriptors from many satellite images. Other approaches like Hu & Lee (2019) used multiple frames from a video and Cai et al. (2019) used triplet loss and channel attention to improve the geo-location accuracy of images by matching street view images with satellite images.

Crowdsourced labels Crowdsourcing is another approach to scalable ground-truth data collecting and is particularly relevant to Street2Sat because the low-cost image collection setup is amenable to crowdsourced data collection in the future. For example, Wang et al. (2020) used geo-located images from Plantix, a smartphone app to identify crop diseases. User-identified labels were paired with satellite imagery to train ML classifiers. To filter out points with noisy location data, they trained a separate CNN to identify points with locations not on fields.

Monocular depth estimation Traditional approaches use photogrammetry to estimate the distance to an object in an image if certain parameters like focal length or object height are known. There have been many recent deep learning approaches to monocular depth estimation (Khan et al., 2020), such as AdaBins (Bhat et al., 2020), SGDepth (Klingner et al., 2020), Monodepth2 (Godard et al., 2018). Since many approaches do not generalize well to images outside the training domain, MiDaS (Lasinger et al., 2019) and DiverseDepth (Yin et al., 2020) used many data sets to build a more generalizable model that could be used off-the-shelf.

The goal of this work is to locate crops in street-level images and translate the image location to the crop location. While some prior studies tackle similar problems, they do not provide a complete solution and may have limited performance for regions like Sub-Saharan Africa where smallholder fields contain multiple crops with similar foliage.

3. Street2Sat Pipeline

The goal of Street2Sat is to turn geo-referenced images acquired from roads into geo-referenced labeled point samples with locations corresponding to objects of interest in the images. The pipeline consists of data collection, pre-processing, object detection, depth estimation, relocation, and quality assessment/control (QA/QC) (Fig. 1).

Data collection The images for our initial study were collected in Western Kenya (Bungoma, Kakamega, and Busia counties) in November 2020. Two teams drove cars with GoPro Max 360 cameras pointing orthogonal to the drive direction toward the nearest roadside (left/driver side). Driving routes were determined using existing roads data and cropland maps to ensure full coverage of the environmental and crop production gradients (Waldner et al., 2019).

Pre-processing Since crops may be tilted in some images due to inclined roads or mounting, we applied an automatic straightening procedure to images in pre-processing. This is important because we use the height of the bounding box to estimate the crop height and distance from the sensor. We straightened images using Otsu's method in OpenCV for automatic image thresholding to separate the background from foreground. The background and foreground locations were used to find a horizon line to straighten the image.

[1] https://github.com/nasaharvest/street2sat_website/tree/ICML_paper_code

Object detection We used YOLO-v5 (Jocher et al., 2021)[2] to predict bounding boxes around crops since it has shown good performance for a variety of domains. We initialized the network using pre-trained weights from COCO (Lin et al., 2014) and fine-tuned for crop type classes using manually labeled images from the data collection stage.

Depth estimation We used a simple method based on ratios for estimating the distance to each predicted bounding box. We assume the true height of each bounding box is known based on the crop type and its growth stage, which we store as a lookup table. For example, sugarcane crops in the Kenyan Sugar Belt range 3-4 m at maturity while maize averages 3 m in Kenya (Juma & George, 2018; Melik et al., 2013). While most crops are expected to be at similar growth stages during data collection, there is still substantial variability in growth stages due to the heterogeneity of smallholder farming practices which could result in errors. We will explore approaches to resolve this in future work. We used the following equation to estimate the distance:

$$d = \frac{(l_{focal} * h_{crop} * h_{image})}{(h_{bbox} * h_{sensor})} \tag{1}$$

The focal length, l_{focal}, was obtained from the EXIF image metadata; the crop height, h_{crop}, is from the height lookup table; h_{bbox} is the height of the bounding box in pixels; h_{image} is the height of the image in pixels; and the GoPro sensor height h_{sensor} is 4.55 mm. We calculated the distance to each bounding box and then calculated the average depth for all of the boxes of the same crop type class to get a single depth for the field.

Relocation Moving the location from the image/vehicle location to the crop location d m away requires knowing the compass direction to move the point along. Since the image heading is not recorded in GoPro metadata, we obtain the direction of the vehicle by using the closest point in time available in the drive data set to compute the velocity vector. We assume that the camera was pointed towards the left of the car orthogonal to the drive direction and move the point d m in the direction 90 degrees west of the heading.

QA/QC Errors could occur in this pipeline due to a variety of factors such as poor lighting, mixed crop fields, occlusions, and object detection errors. In future work, we plan to apply techniques for quality assessment and control (QA/QC) to identify points with possible errors and correct them, e.g., out-of-distribution detection to find outliers located on roads or other objects.

4. Preliminary results

We conducted a preliminary experiment using the data collected in Kenya to test the pipeline. We labeled 296 training

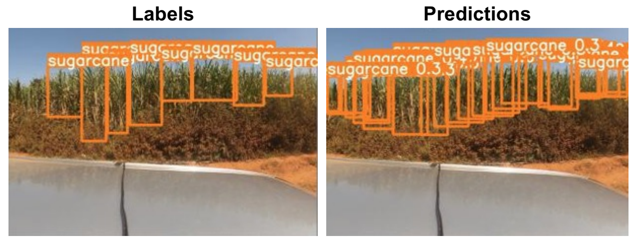

Figure 2. Example test image with labeled (left) and predicted (right) bounding boxes. Due to coarse and non-exhaustive labeling, there are often more correct bounding boxes predicted in the image than are included in the labels.

and 53 test images containing maize and sugarcane from 3 drives. The resulting data set included 755 instances (bounding boxes) of maize and 1,795 of sugarcane in the training set and 253 maize and 229 sugarcane in the test set. Labeling was guided by recommendations from agricultural experts. The test set performance metrics are as follows: Precision: 0.41, Recall: 0.59, mAP @ 0.5: 0.45, and mAP @.5-.95: 0.13. Due to the difficulty of drawing a bounding box around individual plants in a crop field, bounding box labels often included multiple plants and did not cover every crop instance in the image. As a result, there are often more correct bounding boxes predicted in the test image than are included in the labels for that image (e.g., Fig. 2), which can result in low precision. In addition, finding an accurate height of the crops is more important than the number of detected crops for accurate distance prediction.

Since we did not have ground-truth estimates of the distance from the cameras to the crops in the images, we evaluated the relocation results in a similar manner to Yan & Ryu (2021) by comparing the new locations to a 10 m/pixel cropland map of Kenya (Tseng et al., 2020). Out of 85 test points from one drive, 73 (86%) coincided with pixels also classified as crop in the map.

5. Conclusion and future work

We present a pipeline for obtaining geo-referenced points of objects of interest in images taken from vehicles on the road. We show that our pipeline has promising results using pilot data and are working to resolve challenges with image heading accuracy, setting up/securing cameras, and QA/QC. We plan to use Street2Sat to process images from upcoming data collection in 5 African countries. Though our focus is smallholder agriculture, our approach can be applied to other applications (e.g., wildlife tracking and damage assessment) and images from Google Street View. In future work we will implement field-based validation of height and distance and explore a Street2Sat approach to yield estimation, which is highly correlated with crop height (Melik et al., 2013; Juma & George, 2018). This could provide critical information for

[2] https://github.com/ultralytics/yolov5

food security policies, supply chain optimization to reduce food waste, a leading cause of food insecurity globally.

Acknowledgements

This work was funded by the "Helmets Labeling Crops" grant from the Lacuna Fund, supported by the NASA Harvest program and the SwissRe Foundation funded project EO-Farm. The street-level images used in this study were collected by LocateIT in Kenya.

References

Ending hunger: science must stop neglecting smallholder farmers. *Nature*, 586(336), 2020. ISSN 1476-4687. doi: https://doi.org/10.1038/d41586-020-02849-6.

Becker-Reshef, I., Barker, B., Humber, M., Puricelli, E., Sanchez, A., Sahajpal, R., McGaughey, K., Justice, C., Baruth, B., Wu, B., Prakash, A., Abdolreza, A., and Jarvis, I. The GEOGLAM crop monitor for AMIS: Assessing crop conditions in the context of global markets. *Glob. Food Sec.*, 23:173–181, 2019. ISSN 22119124. doi: 10.1016/j.gfs.2019.04.010. URL `https://doi.org/10.1016/j.gfs.2019.04.010`.

Becker-Reshef, I., Justice, C., Barker, B., Humber, M., Rembold, F., Bonifacio, R., Zappacosta, M., Budde, M., Magadzire, T., Shitote, C., Pound, J., Constantino, A., Nakalembe, C., Mwangi, K., Sobue, S., Newby, T., Whitcraft, A., Jarvis, I., and Verdin, J. Strengthening agricultural decisions in countries at risk of food insecurity: The GEOGLAM Crop Monitor for Early Warning. *Remote Sens. Environ.*, 237, feb 2020. ISSN 00344257. doi: 10.1016/j.rse.2019.111553.

Bhat, S. F., Alhashim, I., and Wonka, P. Adabins: Depth estimation using adaptive bins. *CoRR*, abs/2011.14141, 2020. URL `https://arxiv.org/abs/2011.14141`.

Cai, S., Guo, Y., Khan, S., Hu, J., and Wen, G. Ground-to-aerial image geo-localization with a hard exemplar reweighting triplet loss. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8390–8399, 2019. doi: 10.1109/ICCV.2019.00848.

Godard, C., Aodha, O. M., and Brostow, G. J. Digging into self-supervised monocular depth estimation. *CoRR*, abs/1806.01260, 2018. URL `http://arxiv.org/abs/1806.01260`.

Hu, S. and Lee, G. H. Image-based geo-localization using satellite imagery. *CoRR*, abs/1903.00159, 2019. URL `http://arxiv.org/abs/1903.00159`.

Jocher, G., Stoken, A., Borovec, J., NanoCode012, Chaurasia, A., TaoXie, Changyu, L., V, A., Laughing, tkianai, yxNONG, Hogan, A., lorenzomammana, AlexWang1900, Hajek, J., Diaconu, L., Marc, Kwon, Y., oleg, wanghaoyang0106, Defretin, Y., Lohia, A., ml5ah, Milanko, B., Fineran, B., Khromov, D., Yiwei, D., Doug, Durgesh, and Ingham, F. ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations, April 2021. URL `https://doi.org/10.5281/zenodo.4679653`.

Juma, E. O. A. and George, M. D. M. O. *Antagonistic Potential Of Selected Fungal And Bacterial Isolates From Rhizosphere Of Sugarcane Variety Co 421 Against Sporisorium Scitamineum In Kibos , Kisumu*. PhD thesis, Maseno University, 2018.

Kerner, H., Tseng, G., Becker-Reshef, I., Nakalembe, C., Barker, B., Munshell, B., Paliyam, M., and Hosseini, M. Rapid Response Crop Maps in Data Sparse Regions. *KDD '20 Humanit. Mapp. Work.*, 7, jun 2020. ISSN 23318422. URL `http://arxiv.org/abs/2006.16866`.

Khan, F., Salahuddin, S., and Javidnia, H. Deep learning-based monocular depth estimation methods—a state-of-the-art review. *Sensors*, 20(8):2272, 2020.

Klingner, M., Termöhlen, J., Mikolajczyk, J., and Fingscheidt, T. Self-supervised monocular depth estimation: Solving the dynamic object problem by semantic guidance. *CoRR*, abs/2007.06936, 2020. URL `https://arxiv.org/abs/2007.06936`.

Lasinger, K., Ranftl, R., Schindler, K., and Koltun, V. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *CoRR*, abs/1907.01341, 2019. URL `http://arxiv.org/abs/1907.01341`.

Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014. URL `http://arxiv.org/abs/1405.0312`.

McCarthy, J. J., Canziani, O. F., Leary, D. J., Dokken, D. J., and White, K. S. Climate change 2001: impacts, adaptation, and vulnerability. In McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J., and White, K. S. (eds.), *Contrib. Work. Gr. II to Third Assess. Rep. Intergov. Panel Clim. Chang.*, pp. 1–1042. Cambridge University Press., 2001. ISBN 0 521 01500 6.

Melik, C., Njoka, F., Ombakho, G., and Omari, O. Genetic variability analysis for growth and yield parameters in double cross maize (Zea mays L.) genotypes in Kitale county of Kenya. *J. Plant Breed. Genet.*, 1:7–11, 2013. URL `http://www.escijournals.net/JPBG`.

Nakalembe, C. Characterizing agricultural drought in the Karamoja subregion of Uganda with meteorological and satellite-based indices. *Nat. Hazards*, 91(3):837–862, feb 2018. ISSN 15730840. doi: 10.1007/s11069-017-3106-x. URL http://link.springer.com/10.1007/s11069-017-3106-x.

Nakalembe, C. Urgent and critical need for sub-Saharan African countries to invest in Earth observation-based agricultural early warning and monitoring systems. *Environ. Res. Lett.*, 15(12):1–3, 2020. ISSN 17489326. doi: 10.1088/1748-9326/abc0bb.

Nakalembe, C., Becker-Reshef, I., Bonifacio, R., Hu, G., Humber, M. L., Justice, C. J., Keniston, J., Mwangi, K., Rembold, F., Shukla, S., Urbano, F., Whitcraft, A. K., Li, Y., Zappacosta, M., Jarvis, I., and Sanchez, A. A review of satellite-based global agricultural monitoring systems available for Africa. *Glob. Food Sec.*, 29:100543, jun 2021a. ISSN 22119124. doi: 10.1016/j.gfs.2021.100543. URL https://linkinghub.elsevier.com/retrieve/pii/S2211912421000523.

Nakalembe, C., Justice, C., Kerner, H., Justice, C., and Becker-Reshef, I. Sowing Seeds of Food Security in Africa. *Eos (Washington. DC).*, 102, jan 2021b. ISSN 2324-9250. doi: 10.1029/2021EO153329. URL https://eos.org/science-updates/sowing-seeds-of-food-security-in-africa.

Nwanze, K. F. Smallholders can feed the world Smallholders can feed the world. *Int. Fund Agric. Dev.*, 2017. URL www.ifad.org.

Parry, M. L., Rosenzweig, C., Iglesias, A., Livermore, M., and Fischer, G. Effects of climate change on global food production under SRES emissions and socio-economic scenarios. *Glob. Environ. Chang.*, 14(1):53–67, 2004. ISSN 09593780. doi: 10.1016/j.gloenvcha.2003.10.008. URL https://www.sciencedirect.com/science/article/pii/S0959378003000827.

Tseng, G., Kerner, H., Nakalembe, C., and Becker-Reshef, I. Annual and in-season mapping of cropland at field scale with sparse labels. In *Proceedings of the Neural Information Processing Systems (NeurIPS) Workshops*, 2020.

Tseng, G., Kerner, H., Nakalembe, C., and Becker-Reshef, I. Learning to predict crop type from heterogeneous sparse labels using meta-learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2021.

Viswanathan, A., Pires, B. R., and Huber, D. Vision based robot localization by ground to satellite matching in gps-denied situations. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 192–198, 2014. doi: 10.1109/IROS.2014.6942560.

Waldner, F., Bellemans, N., Hochman, Z., Newby, T., de Abelleyra, D., Verón, S. R., Bartalev, S., Lavreniuk, M., Kussul, N., Le Maire, G., et al. Roadside collection of training data for cropland mapping is viable when environmental and management gradients are surveyed. *International Journal of Applied Earth Observation and Geoinformation*, 80:82–93, 2019.

Wang, S., Di Tommaso, S., Faulkner, J., Friedel, T., Kennepohl, A., Strey, R., and Lobell, D. B. Mapping crop types in southeast india with smartphone crowdsourcing and deep learning. *Remote Sensing*, 12(18), 2020. ISSN 2072-4292. doi: 10.3390/rs12182957. URL https://www.mdpi.com/2072-4292/12/18/2957.

Whitcraft, A. K., Becker-Reshef, I., Justice, C. O., Gifford, L., Kavvada, A., and Jarvis, I. No pixel left behind: Toward integrating Earth Observations for agriculture into the United Nations Sustainable Development Goals framework. *Remote Sensing of Environment*, 235:111470, dec 2019. ISSN 00344257. doi: 10.1016/j.rse.2019.111470.

Wolfram, S. and Lobell, D. B. Robust negative impacts of climate change on African agriculture. *Environ. Res. Lett.*, 5(1):1–8, jan 2010. ISSN 1748-9326. doi: 10.1088/1748-9326/5/1/014010. URL http://stacks.iop.org/1748-9326/5/i=1/a=014010?key=crossref.6f26401c5082faeb2f666118e1e02b98http://stacks.iop.org/1748-9326/5/i=1/a=014010.

Yan, Y. and Ryu, Y. Exploring google street view with deep learning for crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171:278–296, 2021. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2020.11.022. URL https://www.sciencedirect.com/science/article/pii/S0924271620303294.

Yin, W., Wang, X., Shen, C., Liu, Y., Tian, Z., Xu, S., Sun, C., and Renyin, D. Diversedepth: Affine-invariant depth prediction using diverse data. *CoRR*, abs/2002.00569, 2020. URL https://arxiv.org/abs/2002.00569.