Weak labeling for cropland mapping in Africa

Gilles Hacheme
AI for Good Lab, Nairobi, Kenya
ghacheme@microsoft.com

Girmaw Abebe Tadesse
AI for Good Lab, Nairobi, Kenya
gtadesse@microsoft.com

Caleb Robinson
AI for Good Lab, Redmond, USA
caleb.robinson@microsoft.com

Rahul Dodhia
AI for Good Lab, Redmond, USA
rahul.dodhia@microsoft.com

Juan Lavista Ferres
AI for Good Lab, Redmond, USA
jlavista@microsoft.com

Abstract

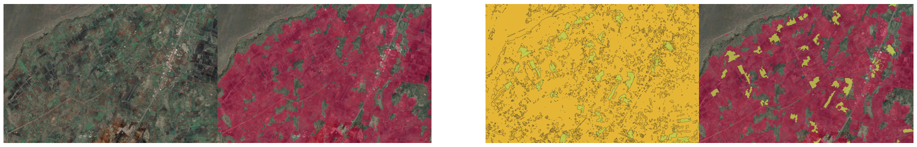
Cropland mapping is an essential task in addressing environmental, agricultural, and food security challenges. Unfortunately, most research works and products only offer low to medium-sized resolution cropland mapping based on satellite imagery, and their practical usage in Africa is often limited. Creating high-resolution cropland maps requires extensive human labeling, which is a bottleneck for scaling. This paper suggests a new method that leverages K-means clustering to improve existing weak labels (e.g. from noisy global cropland maps) that can be used to train higher-resolution cropland mapping models. The human and improved weak labels can then be used in a deep semantic segmentation neural network to detect the croplands. We perform simulations that showcase the added value of the improved weak labels we generated.

1 Introduction

Up-to-date and high-resolution data on the spatial distribution of crop fields is critical for environmental, agricultural, and food security policies, especially in Africa, as most of the countries' economies heavily depend on agriculture. Cropland mapping from satellite imagery has been an essential topic for the research community [1, 2, 3, 4, 5, 6]. Unfortunately, most research works and products only offer low to medium-sized resolution cropland mapping based on satellite imagery datasets such as Landsat [7]. Additionally, as they heavily rely on machine learning (ML) methods trained on data from Western fields, their practical usage in Africa is often limited.

Semantic segmentation has been used extensively for cropland mapping and, more broadly, land cover mapping [8, 9, 10, 11, 12]. In this work, we use the Torchgeo package [13] to implement a deep semantic segmentation training workflow on satellite imagery, specifically a U-Net [14] with a Resnet-50 [15] backbone model. To increase the size and coverage of our training dataset, we developed a method that improves existing weak labels using the K-means unsupervised clustering method [16, 17]. Our simulations reveal that the improved weak labels, mostly the negative samples, can improve the cropland mapping system in a scenario where human labels are limited.

5th Deep Learning Indaba Conference (DLI 2023).



(a) An area from the CHEF and TNC weak labels (b) K-means polygons before and after filtering
Note: The red areas are the potential cropland areas

2 Problem statement

Consider an area of interest (AOI) represented by a $k \times k$ dimensional matrix A where a_{ij} is the pixel from A located at (i, j). We assume that we have a corresponding mask M with the same dimensions, where each pixel from M, $m_{ij} \in \{0, 1, 2\}$, and where $0 = unknown$, $1 = non - cropland$, $2 = cropland$. In a cropland mapping semantic segmentation problem, let's consider the distribution of M's pixels in different scenarios:

1. Complete ground truth (CGT): we would like the distribution of m_{ij} to be $\{0 : 0, 1 : 1 - p, 2 : p\}$, where $p \in [0, 1]$ is the ground truth proportion of pixels covering croplands. This scenario matches with the case where we have ground truth labeling for every single pixel in M. This is the idealistic situation but the least realistic in real-life scenarios, given the large AOIs used for semantic segmentation and the limited budget and time organizations and researchers have for their projects.
2. Complete labeling (CL): in this case, we may want to label through a labeling tool every single pixel a_{ij}. In this case, we can reasonably assume that the distribution of m_{ij} is $\{0 : 0, 1 : 1 - (p - \epsilon), 2 : p - \epsilon\}$ where $(p - \epsilon) \in [0, 1]$ and $\epsilon \in [0, 1]$ is the level of noise (mislabels) introduced by the labeling. Lower ϵ, the better the labeling. In a perfect labeling scenario, though most of the time unrealistic, $\epsilon = 0$.
3. Partial labeling (PL): in this case, labeling is not performed on the whole AOI, which is often the case in real-life settings. So, the distribution of m_{ij} is $\{0 : q, 1 : 1 - (p - \epsilon + \alpha q), 2 : p - (\epsilon + (1 - \alpha)q)\}$ where $(p - (\epsilon + (1 - \alpha)q) \in [0, 1])$, and $q \in [0, 1]$ is the share of unknown/unlabeled pixels.

Considering the PL scenario as the most realistic one, many techniques have been developed by the community to leverage weak labels for semantic segmentation tasks [18, 19, 20]. The goal is to decrease q as much as possible without affecting ϵ too much. It is straightforward to virtually decrease the missing labels proportion q, but it is hard to decrease it without substantially increasing ϵ. For instance, we can reach $q = 0$ by simply classifying any pixel a_{ij} randomly, but this would increase ϵ by a large margin and likely hurt our segmentation model training more. In this work, we propose an approach to decrease q, by leveraging existing weak labels, while keeping the noise ϵ as low as possible. In such a case, we hypothesize the semantic segmentation model should be improved by using the proposed data augmentation approach.

3 Using K-means to strengthen cropland weak labels

Semantic segmentation is an essential task in computer vision that involves classifying each pixel in an image into one of multiple categories. By utilizing unsupervised learning, we aim to generate enhanced weak labels to augment the existing strong/human labels available to train the segmentation model.

For the experiments in this paper, we use cropland weak labels obtained from The Nature Conservancy (TNC). These labels cover the Central Highlands Ecoregion Foodscape (CHEF) in Kenya. However, these labels seem ambiguous[1], making them challenging to be used to train a segmentation model (Figure 1a). For the input imagery, we use Planetscope Basemap imagery provided by theNorwegian International Climate and Forests Initiative (NICFI).

The K-means algorithm is used to cluster the pixels in the PlanetScope imagery into 10 clusters based on their spectral characteristics. The goal is to generate weak labels with better boundaries.

[1]The cropland boundaries are not well delineated.

After clustering, we convert each cluster of pixels produced into a polygon and filtered out tiny and too-large polygons to reduce noise. The filtering of tiny and very large polygons is done sequentially using a quantile-based approach, where polygons covering areas smaller than a threshold are filtered out. Specifically, the 99th area quantile threshold is first used to filter out tiny polygons. Then the top 25% polygons (from the first filter) in terms of area are filtered out. This approach has been validated visually as the vast majority of the clusters are tiny and overlay other clusters of pixels while a tiny fraction of the clusters represents very large areas. The input of the K-means is made of a concatenation of pixels randomly sampled from the 88 imagery quads covering the CHEF region. We randomly sample 1 million pixels out of the $4096 \times 4096 = 16,777,216$ pixels per quad, resulting in a sample size of 88 million pixels, each with five features representing each band. Finally, we predict the cluster class for each pixel in the original quad (4096x4096), save the prediction as a GeoTIFF, and extract polygons from the pixel clusters (e.g. see figure 1b). Following this, we estimate the proportion of crop cover in each polygon by measuring the area of the polygon that intersects with the TNC label polygons. The determination of cropland vs. non-cropland is then based on a threshold value of the intersection. From visual inspection, we classify as cropland any polygon with an intersection strictly greater than 0.8 and as non-cropland any polygon with an intersection strictly lower than 20%. These enhanced weak labels are then used to augment the original training dataset for the semantic segmentation task.

4 Simulation experiments

To validate our method, we run simulation experiments where we consider an ideal case and more realistic scenarios. The experiments help to identify the best way to leverage the weak labels and quantify their potential benefit for cropland semantic segmentation. Our experiments are as follows:

1. Human labels: we train the model on the AOI with the complete set of human labels, and we evaluate on the exact same AOI. This experiment is conducted for the sole reason of having the best performance level our system can potentially achieve given a more limited or noisier set of labels. In this experiment, we have 67 human labels (polygons) covering 4.056% of the AOI.
2. Human mined labels: we train the model on the AOI with the complete set of human labels and improved weak (mined) labels.
3. Human mined negative labels: we train the model on the AOI with the complete set of human labels and improved weak (mined) negative labels (i.e., Non-cropland labels only).
4. Human mined positive labels: we train the model on the AOI with the complete set of human labels and improved weak (mined) positive labels.
5. Half human labels [mined [negative/positive] labels]: we conduct the same experiment as previously but with only half human labels. This case is for simulating more realistic real-world scenarios where we only have a fraction of the whole data labeled by humans.
6. Human TNC labels: we train the model on the AOI with the complete set of human labels and TNC's raw positive weak labels.
7. Human TNC mined negative labels: we train the model on the AOI with the complete set of human labels, TNC's raw positive weak labels, and the improved weak (mined) negative labels.

We run experiments on the *L15-1237E-1025N* quad and evaluate on the same AOI but only on strong (human) labels. This means that the training and testing sets only overlap with the part of human labels contained in the training set. Our *baseline* segmentation model is based on the well-known U-Net architecture [14] with a ResNet50 backbone [15]. It is trained using a cross-entropy loss function and the Adam optimization algorithm [21]. The trained model is used to make predictions on the same imagery. The output produced by the model is a binary mask that shows the location of cropped regions in the input imagery.

Table 1 presents the F1 scores and additional metrics[2] of the cropland mapping segmentation task in different experiment settings, and across the two label classes *Cropland* and *Non-Cropland*.

[2]The number of mined labels, the area covered by these labels, the precision, and the recall for each class.

Experiment	Label	Mined Labels	Mined Area	F1 score	Precision	Recall
Human labels	Cropland	0	0.000	0.980	1.000	0.960
	Non-Cropland	0	0.000	0.991	1.000	0.982
Human mined labels	Cropland	606	11.016	0.979	0.999	0.960
	Non-Cropland	369	6.702	0.991	1.000	0.982
Human mined negative labels	Cropland	606	11.016	0.979	1.000	0.959
	Non-Cropland	369	6.702	0.991	1.000	0.982
Human mined positive labels	Cropland	606	11.016	0.979	0.999	0.960
	Non-Cropland	369	6.702	0.991	1.000	0.982
Human TNC labels	Cropland	0	0.000	0.976	0.995	0.957
	Non-Cropland	0	0.000	0.991	1.000	0.982
Human TNC mined negative labels	Cropland	606	11.016	0.978	0.997	0.959
	Non-Cropland	369	6.702	0.991	1.000	0.982
Half human labels	Cropland	0	0.000	0.533	0.408	0.767
	Non-Cropland	0	0.000	0.962	0.991	0.935
Half human mined labels	Cropland	606	11.016	0.694	0.553	0.931
	Non-Cropland	369	6.702	0.974	0.999	0.950
Half human mined negative labels	Cropland	606	11.016	0.841	0.916	0.777
	Non-Cropland	369	6.702	0.985	0.992	0.979
Half human mined positive labels	Cropland	606	11.016	0.324	0.196	0.929
	Non-Cropland	369	6.702	0.901	0.998	0.821
Half human TNC labels	Cropland	0	0.000	0.289	0.170	0.960
	Non-Cropland	0	0.000	0.880	1.000	0.785
Half human TNC mined negative labels	Cropland	606	11.016	0.581	0.417	0.959
	Non-Cropland	369	6.702	0.961	1.000	0.925

Table 1: Experiment results

Notes: In every experiment, we train and evaluate on the same area. Experiments only differ on the type and number of labels provided during training. A detailed description of each experiment can be found in the section4.

The first set of experiments leverages the complete human labels combined with eventually other labels, whether the mined labels or the raw weak labels. This first part simulates the ideal case where we have the complete set of labels from humans and, eventually, some additional weak labels. The *Human labels* experiment for *cropland* achieves, as expected, a very high F1 score of 0.980, indicating overfitting of the model. The F1 score for *non-cropland* is even higher (0.991). And therefore, we see no added value from the mined/weak labels in this scenario. These results are only helpful as they indicate results we could achieve if we had all the human labels at our disposal. But this scenario is usually less likely, and most of the time, we might get only a portion of the human labels.

The following set of experiments shows results where only half the human labels are used in the training sets. The results show that as the number of human labels decreases (by half in this case), the F1 scores globally decrease. The F1 score for *cropland* in the *Half-human labels* experiment is only 0.533, indicating a significant drop in performance. This drop is mainly due to a large decrease in the precision (only 0.408). However, the performance for *non-cropland* remains high, indicating that the segmentation task could still identify *non-cropland* areas relatively well, even with fewer human labels. Using all the mined labels in addition to half the human labels (*Half human mined labels*) improves the cropland F1 score from 0.533 to 0.694. But the highest F1 score is obtained when only the negative mined samples are used in addition to half the human labels (*Half human mined negative labels*). The cropland F1 score, in this case, reaches 0.841, with a precision of 0.916, while the recall is almost the same as the one obtained with the *Half human labels* experiment.

Using the raw (positive) weak labels from TNC in addition to half the human labels (*Half human TNC labels*), on the contrary, degrades the F1 score for *cropland* from 0.533 to 0.289. Even by combining the TNC raw (positive) weak labels, the mined negative labels, and half human labels (*Half human TNC mined negative labels*), the F1 score is only 0.581. This confirms our assumption that the raw weak labels should not be used directly to augment the training set, and implicitly show the added value of our mining approach.

These experiments indicate the potential of mining weak labels for large-scale cropland mapping.

5 Conclusion

The accurate mapping of cropland fields through high-resolution satellite imagery is crucial for Africa's agricultural and food security policies. Unfortunately, labeling is the main bottleneck to building high-resolution cropland mapping systems. Our study presents a novel methodology to improve existing weak labels using K-means clustering, in order to augment existing training data, usually human labeled. The experimental results confirm that human labeling is vital for accurate results, while mining labels can significantly enhance large-scale cropland mapping. Therefore, the proposed system could be an essential tool for large-scale cropland mapping.

References

[1] Peter Potapov, Svetlana Turubanova, Matthew C Hansen, Alexandra Tyukavina, Viviana Zalles, Ahmad Khan, Xiao-Peng Song, Amy Pickens, Quan Shen, and Jocelyn Cortez. Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nature Food*, 3(1):19–28, 2022.

[2] Kwang-Hyung Kim, Yasuhiro Doi, Navin Ramankutty, and Toshichika Iizumi. A review of global gridded cropping system data products. *Environmental Research Letters*, 16(9):093005, sep 2021.

[3] Pradeep Adhikari and Kirsten M de Beurs. An evaluation of multiple land-cover data sets to estimate cropland area in west africa. *International Journal of Remote Sensing*, 37(22):5344–5364, 2016.

[4] Weston Anderson, Liangzhi You, Stanley Wood, Ulrike Wood-Sichra, and Wenbin Wu. A comparative analysis of global cropping systems models and maps. 2014.

[5] Claire Boryan, Zhengwei Yang, Rick Mueller, and Mike Craig. Monitoring us agriculture: the us department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, 26(5):341–358, 2011.

[6] M Santoro, G Kirches, J Wevers, M Boettcher, C Brockmann, C Lamarche, and P Defourny. Land cover cci: Product user guide version 2.0. *Climate Change Initiative Belgium*, 2017.

[7] Michael A Wulder, Joanne C White, Thomas R Loveland, Curtis E Woodcock, Alan S Belward, Warren B Cohen, Eugene A Fosnight, Jerad Shaw, Jeffrey G Masek, and David P Roy. The global landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, 185:271–283, 2016.

[8] Surabhi Lmgwal, Komal Kumar Bhatia, and Manjeet Singh. Semantic segmentation of landcover for cropland mapping and area estimation using machine learning techniques. *Data Intelligence*, pages 1–18, 2022.

[9] Michael Schmitt, Jonathan Prexl, Patrick Ebel, Lukas Liebel, and Xiao Xiang Zhu. Weakly supervised semantic segmentation of satellite images for land cover mapping–challenges and opportunities. *arXiv preprint arXiv:2002.08254*, 2020.

[10] Zhenrong Du, Jianyu Yang, Cong Ou, and Tingting Zhang. Smallholder crop area mapped with a semantic segmentation deep learning method. *Remote Sensing*, 11(7):888, 2019.

[11] Meiqi Du, Jingfeng Huang, Pengliang Wei, Lingbo Yang, Dengfeng Chai, Dailiang Peng, Jinming Sha, Weiwei Sun, and Ran Huang. Dynamic mapping of paddy rice using multi-temporal landsat data based on a deep semantic segmentation model. *Agronomy*, 12(7):1583, 2022.

[12] Zheng Shuangpeng, Fang Tao, and Huo Hong. Farmland recognition of high resolution multispectral remote sensing imagery using deep learning semantic segmentation method. In *Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence*, pages 33–40, 2019.

[13] Adam J Stewart, Caleb Robinson, Isaac A Corley, Anthony Ortiz, Juan M Lavista Ferres, and Arindam Banerjee. Torchgeo: deep learning with geospatial data. In *Proceedings of the 30th international conference on advances in geographic information systems*, pages 1–12, 2022.

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[16] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[17] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[18] Yujia Chen, Guo Zhang, Hao Cui, Xue Li, Shasha Hou, Jinhao Ma, Zhijiang Li, Haifeng Li, and Huabin Wang. A novel weakly supervised semantic segmentation framework to improve the resolution of land cover product. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196:73–92, 2023.

[19] Zhe Chen, Zhao Chen, Jingxin Liu, Qiang Zheng, Yuang Zhu, Yanfei Zuo, Zhaoyu Wang, Xiaosong Guan, Yue Wang, and Yuan Li. Weakly supervised histopathology image segmentation with sparse point annotations. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1673–1685, 2020.

[20] Seunghoon Hong, Suha Kwak, and Bohyung Han. Weakly supervised learning with deep convolutional neural networks for semantic segmentation: Understanding semantic layout of images with minimum human supervision. *IEEE Signal Processing Magazine*, 34(6):39–49, 2017.

[21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.