FIELDS OF THE WORLD: A FIELD GUIDE FOR EXTRACTING AGRICULTURAL FIELD BOUNDARIES

Isaac Corley*
Wherobots

Hannah Kerner
Arizona State University

Caleb Robinson
Microsoft AI for Good Research Lab

Jennifer Marcus
Taylor Geospatial

ABSTRACT

Field boundary maps are a building block for agricultural data products and support crop monitoring, yield estimation, and disease estimation. This tutorial presents the Fields of The World (FTW) ecosystem: a benchmark of 1.6M field polygons across 24 countries, pre-trained segmentation models, and command-line inference tools. We provide two notebooks that cover (1) local-scale field boundary extraction with crop classification and forest loss attribution, and (2) country-scale inference using cloud-optimized data. We use MOSAIKS random convolutional features and FTW derived field boundaries to map crop type at the field level and report macro F1 scores of 0.65–0.75 for crop type classification with limited labels. Finally, we show how to explore pre-computed predictions over five countries (4.76M km^2), with median predicted field areas from 0.06 ha (Rwanda) to 0.28 ha (Switzerland). Code and notebooks are available here.

1 INTRODUCTION

Agricultural monitoring at continental scales informs food security, land-use policy, and compliance frameworks such as the EU Deforestation Regulation (EUDR) (European Parliament and Council, 2023). Field boundaries define parcels that are the unit of analysis for many agricultural task: crop mapping, yield estimation, subsidy allocation, and deforestation screening (see Figure 1 for and example of field polygon predictions). Further, automated approaches for generating field boundaries from satellite imagery are necessary, as manual digitization of fields is not feasible at scale.

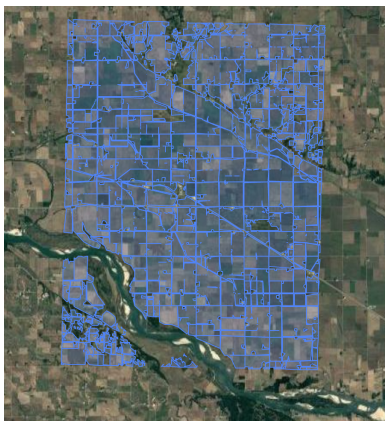

Figure 1: **FTW field boundary predictions for MGRS tile 14TPN in Iowa, USA.** Blue polygons show extracted agricultural parcels overlaid on Sentinel-2 true-color imagery.

The Fields of The World (FTW) project (Kerner et al., 2024) provides a benchmark dataset with 1.6M field polygons across 24 countries, pre-trained segmentation models that use bi-temporal Sentinel-2 imagery as input, and open-source tools for inference and evaluation. The models are distributed through the `ftw-tools` package which provides a command line interface for performing inference on Sentinel-2 satellite imagery and postprocessing the result.

This tutorial provides two Python notebooks that implement end-to-end workflows using the `ftw-tools` package, specifically covering: (1) local-scale field boundary inference used for crop classification and forest loss attribution, and (2) exploration of pre-computed country-scale field boundary outputs and agricultural change detection. Through these workflows, we contribute a reproducible pipeline to download Sentinel-2 imagery and polygonize

*Corresponding author: isaac@wherobots.com

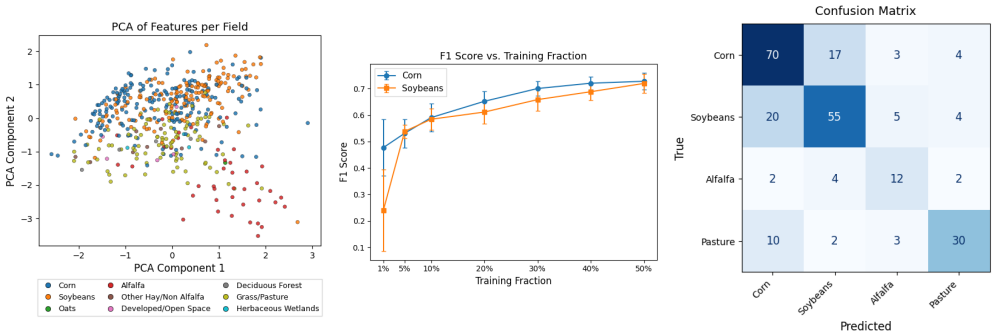

Figure 2: **Crop classification results on Iowa fields**: (**left**) PCA of 256-dim MOSAIKS embeddings colored by CDL crop type showing separable clusters; (**center**) macro F1 score versus training data fraction demonstrating few-shot capability; (**right**) confusion matrix revealing high accuracy for dominant crops (corn, soybeans) with more errors on minority classes.

FTW inference outputs, the integration of these polygons with MOSAIKS embeddings for few-shot crop classification (Rolf et al., 2021), parcel-level forest loss attribution using Hansen Global Forest Change (Hansen et al., 2013), and standardized access patterns for pre-computed country-scale predictions stored as cloud-native Zarr archives.

2 Related Work

Field boundary extraction has trended from classical image processing and edge detection to deep learning-based modeling (Lavreniuk et al., 2025; Rufin et al., 2024; Waldner & Diakogiannis, 2020). Edge-based and region-growing approaches perform poorly under spectral variability and irregular parcel geometry. Encoder-decoder architectures (e.g., U-Net) now dominate agricultural segmentation Estes et al. (2024), and FTW (Kerner et al., 2024) provides a globally diverse training set that supports generalization across different geographies and climate zones.

3 Methods

Data Pipeline. Our workflow begins with bi-temporal Sentinel-2 imagery for planting and harvest windows. We query the Microsoft Planetary Computer STAC catalog (Microsoft, 2022) for cloud-free Level-2A surface reflectance scenes. For each MGRS tile (100×100 km), we define seasonal windows from local crop calendars and compute median composites to reduce atmospheric artifacts. The MGRS grid provides a standardized tiling scheme for reproducible analysis.

FTW models take two 4-channel inputs (RGB+NIR at 10 m) from early and late season observations. The bi-temporal setup captures phenological differences between crops and non-agricultural land cover. The acquisition gap is typically 3–6 months, depending on the growing season.

Field Boundary Extraction. The FTW segmentation models use a U-Net model with an ImageNet pretrained EfficientNet-B3 encoder. Training data span 24 countries and a wide range of field sizes ($<$1 ha to $>$100 ha). The model outputs per-pixel probabilities for field interior, field boundary, and background. The `ftw-tools` package provides a command-line interface for inference:

```
# Install FTW package
pip install ftw-tools

# Download bi-temporal Sentinel-2 imagery and run inference
ftw inference download --win_a <id1> --win_b <id2> --out input.tif --bbox <bbox>
ftw inference run input.tif --out preds.tif --model FTW_PRUE_EFNET_B3_CCBY

# Filter by land cover and convert raster predictions to polygons
ftw inference filter-by-lulc preds.tif --out filtered.tif
ftw inference polygonize filtered.tif --out boundaries.gpkg
```

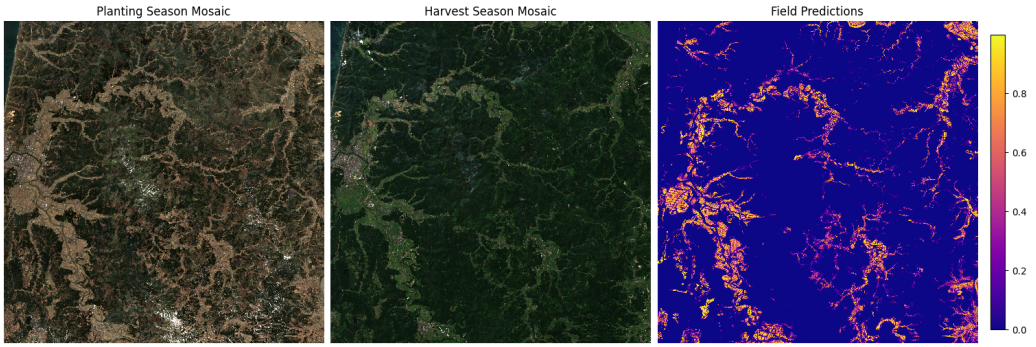

Figure 3: **Country-scale FTW inference for Japan.** Left: Sentinel-2 RGB mosaic. Right: per-pixel field probability (softmax output). Data accessed via cloud-optimized Zarr archives.

The post-processing steps threshold softmax outputs at 0.5, apply morphological opening to remove small artifacts, and convert rasters to polygons via connected components. We optionally filter predictions with land-cover masks (e.g., ESA WorldCover) to remove forests, water, and urban areas. Figure 1 shows example predictions over Sentinel-2 imagery.

Crop Classification with MOSAIKS. For crop mapping we use MOSAIKS (Rolf et al., 2021), which produces fixed-dimensional embeddings from random convolutional filters, that are shown to have high performance under linear probing in a variety of downstream tasks.

For each FTW polygon, we extract the corresponding Sentinel-2 pixels, apply 256 random 3×3 filters (Gaussian initialization), apply ReLU, and compute a masked global average pooling to obtain a 256-dim embedding. We train a logistic regression classifier using Cropland Data Layer (CDL) from the USDA (USDA NASS, 2024) as labels.

Forest Loss Monitoring. We integrate Hansen Global Forest Change (GFC) (Hansen et al., 2013) to attribute forest loss to field polygons. GFC provides annual tree cover loss predictions at 30 m/px spatial resolution from 2000 to 2023. We spatially join FTW derived field boundary polygons with loss-year rasters from GFC and compute zonal statistics (mean, min, max loss year) per field, which identifies parcels with recent deforestation, which is relevant to measuring EUDR compliance after 2020.

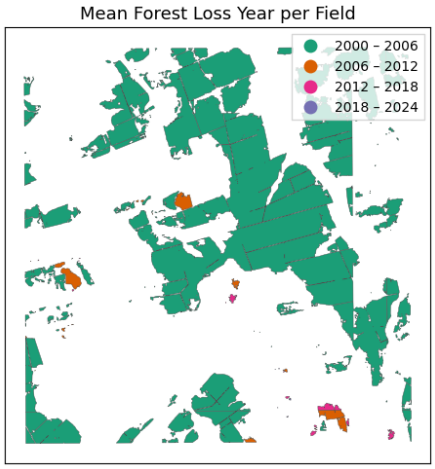

Figure 4: **Forest loss attribution for MGRS tile 21LXF in Mato Grosso, Brazil.** Colors indicate mean deforestation year per field polygon from Hansen GFC data. Pink and Purple colors denote recent clearing (2012–2024), relevant to EUDR compliance thresholds.

Country-Scale Inference. We provide pre-computed predictions for five countries (Japan, Mexico, Rwanda, South Africa, Switzerland) totaling 4.76M km^2 across two seasons and two years (2023–2024). Seasonally-aligned mosaics are constructed using latitude-dependent heuristics to identify optimal planting and harvest windows; a greedy algorithm selects cloud-free Sentinel-2 scenes to maximize coverage. Inference uses a sliding window of 256×256 pixels with 25% overlap. Gaussian-weighted averaging (apodization) merges overlapping patches, suppressing lower-quality edge predictions and reducing tiling artifacts. Outputs are re-projected to Web Mercator (EPSG:3857) on-the-fly and written in parallel to cloud-optimized Zarr archives, enabling lazy loading and spatial queries without full downloads. Median predicted field areas range from 0.06 ha (Rwanda) to 0.28 ha (Switzerland), reflecting variation between smallholder and mechanized systems.

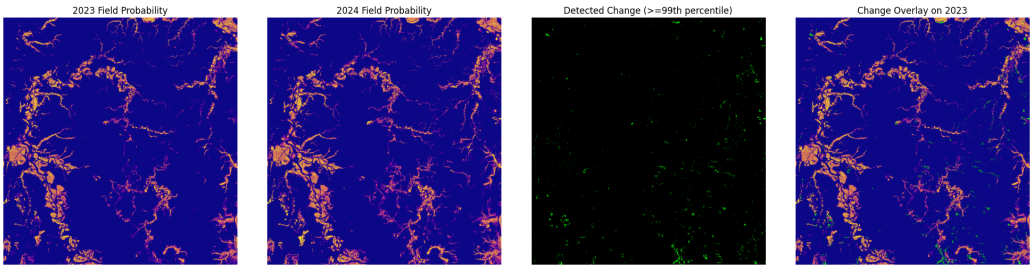

Figure 5: **Year-over-year change detection from FTW field probability maps.** Panels show 2023 and 2024 field probability, detected changes ($\geq$99th percentile), and a change overlay on the 2023 map.

4 Results

Crop Classification. We evaluate crop classification on MGRS tile 14TPN (Iowa, USA) using FTW polygons and the following CDL labels: corn, soybeans, grass/pasture, and alfalfa. Logistic regression on MOSAIKS embeddings yields macro F1 scores of 0.65–0.75 depending on training size (Figure 2). PCA of embeddings shows separable clusters by crop type.

Using 10% of labeled fields for training yields F1 scores within 5% of full-data training. Figure 2c shows high accuracy for dominant classes (corn, soybeans; F1 > 0.8) and more confusion among minority classes, consistent with class imbalance.

Forest Loss Attribution. Applying the workflow to Brazil (MGRS tile 21LXF, Mato Grosso) links GFC loss years to individual fields (Figure 4). Recent deforestation (2015–2020) appears in warmer colors.

Country-Scale Predictions. The Zarr archives store both Sentinel-2 mosaics (RGB+NIR at 10 m) and raw softmax probabilities for background, field interior, and field boundary classes. Users can query arbitrary spatial extents without downloading complete datasets. Figure 3 shows country-scale predictions for Japan with input imagery and output probability maps. The raw probabilities support custom thresholding, watershedding, or vectorization beyond the defaults provided in the `ftw-tools` command line interface.

We also compute year-over-year change by differencing field probability maps from 2023 and 2024 (Figure 5). Pixels exceeding the 99th percentile of absolute difference are flagged as potential agricultural expansion or abandonment, demonstrating how the Zarr archives enable temporal analysis without rerunning inference.

5 Conclusion

This tutorial demonstrates how the Fields of The World ecosystem supports end-to-end agricultural monitoring workflows. We show that FTW segmentation models generalize across diverse regions and that the extracted polygons integrate with downstream tasks including crop classification via MOSAIKS embeddings and forest loss attribution via Global Forest Watch labels. Pre-computed country-scale predictions stored as cloud-optimized Zarr enable efficient spatial queries and temporal change detection without local reprocessing and can be openly used for a variety of downstream applications.

References

LD Estes, A Wussah, M Asipunu, M Gathigi, P Kovačič, J Muhando, BV Yeboah, FK Addai, ES Akakpo, MK Allotey, et al. A region-wide, multi-year set of crop field boundary labels for africa. *arXiv preprint arXiv:2412.18483*, 2024.

European Parliament and Council. Regulation (eu) 2023/1115 on deforestation-free products. https://eur-lex.europa.eu/eli/reg/2023/1115/oj, 2023. Official Journal of the European Union.

M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. High-resolution global maps of 21st-century forest cover change. *Science*, 342:850–853, 2013.

Hannah Kerner, Snehal Chaudhari, Aninda Ghosh, Caleb Robinson, Adeel Ahmad, Eddie Choi, Nathan Jacobs, Chris Holmes, Matthias Mohr, Rahul Dodhia, Juan M. Lavista Ferres, and Jennifer Marcus. Fields of the world: A machine learning benchmark dataset for global agricultural field boundary segmentation. In *AAAI Conference on Artificial Intelligence*, pp. 28151–28159, 2024.

Mykola Lavreniuk, Nataliia Kussul, Andrii Shelestov, Bohdan Yailymov, Yevhenii Salii, Volodymyr Kuzin, and Zoltan Szantoi. Delineate anything: Resolution-agnostic field boundary delineation on satellite imagery. *arXiv preprint arXiv:2504.02534*, 2025.

Microsoft. Microsoft planetary computer. https://planetarycomputer.microsoft.com, 2022. Accessed: 2024.

Esther Rolf, Jonathan Proctor, Tamma Carleton, Ian Bolliger, Vaishaal Shankar, Miyabi Ishihara, Benjamin Recht, and Solomon Hsiang. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Communications*, 12:4392, 2021.

Philippe Rufin, Sherrie Wang, Sá Nogueira Lisboa, Jan Hemmerling, Mirela G. Tulbure, and Patrick Meyfroidt. Taking it further: Leveraging pseudo-labels for field delineation across label-scarce smallholder regions. *International Journal of Applied Earth Observation and Geoinformation*, 134:104149, 2024. ISSN 1569-8432. doi: https://doi.org/10.1016/j.jag.2024.104149. URL https://www.sciencedirect.com/science/article/pii/S156984322400503X.

USDA NASS. Usda national agricultural statistics service cropland data layer. https://nassgeodata.gmu.edu/CropScape/, 2024. Accessed: 2024.

François Waldner and Foivos I Diakogiannis. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote sensing of environment*, 245: 111741, 2020.