A Unified Framework for Cropland Field Boundary Detection and Segmentation

Rodrigo Fill Rangel
Seedz
rodrigo.rangel@seedz.ag

Vítor Nascimento Lourenço
Gaivota

Lucas Volochen Oldoni
Seedz

Ana Flavia Carrara Bonamigo
Seedz

Wallas Santos
Gaivota

Bruno Silva Oliveira
Seedz

Mateus Neves Barreto
Seedz
874, Gonçalves Dias st. 30140-091
Belo Horizonte, MG, Brazil

Abstract

Agricultural monitoring is essential to ensure food security while minimizing the environmental impacts generated by the activity. Crop fields are the basic units of management in farmland, and the delimitation of their boundaries is useful for farmers and field-level analysis. In this work, we address the cropland field boundaries segmentation challenge by proposing an end-to-end novel segmentation framework. Our framework comprises three main pipelines: data preparation, in which Sentinel-2 MSI sensor images are handled; segmentation, where we propose the use of three different methods to obtain a cropland field map, the Felzenswalb's segmentation algorithm, and the neural networks U-Net and R2AttU-Net; and post-processing, where we propose a novel temporal aggregation and filtering methodology to enhance crop field boundary delineation. Our results show that our end-to-end framework is able to outline cropland field boundaries from Sentinel-2 data. The U-Net segmentation achieved overall good results, although some small fields may not be correctly identified. On the other hand, the post-processing was able to mitigate most incorrectly segmented cropland field polygons, significantly improving results in most metrics, removing isolated pixels, and better delimiting fields.

1. Introduction

In the coming decades, the world has the challenge of providing an increase in agricultural production to meet the growing demands of food for human and animal consumption and for the production of biofuels [10, 34], while minimizing the environmental impacts generated by the activity [24, 36]. Agricultural monitoring becomes essential to achieve these goals [33]. Remote sensing is one of the most promising data sources for agricultural monitoring as it allows obtaining objective information from large areas periodically [5, 21, 40].

Crop fields are the basic units of management in farmland [39]. Delimiting crop field boundaries are helpful inputs to precision agriculture and digital agriculture services [8], as well as, enable field-level analysis of crop type mapping [3], crop phenology and biomass [31], and crop yield mapping [30]. Manual digitization with visual interpretation is one of the methods used for delimiting crop fields [4]. However, this method requires experienced interpreters and is time-consuming. Traditional automatic segmentation methods including edge detection [9], region-based [23], a combination of the two [28], or graph-based [15] are, also, adopted. While edge-based detection does not guarantee closed polygons; region-based detection boundaries are not always located at the natural edges. The combination of the two approaches can overcome some of the weaknesses of each one individually, but usually require a fine-tune parameterization that may depend on the characteristics of the analyzed region and the image used. The graph-based is also usually dependent on fine-tune parameterization.

Current advances in computer vision and machine learning applied to remote sensing data have brought breakthroughs in cropland field boundary delimitation. Deep learning methods offer a vessel for an all-inclusive process including feature extraction and classification, which

makes the field boundary delimitation process much more polished [6]. However, field delimitation offers some challenges not commonly encountered in other deep learning applications, requiring learning complex and specific temporal, spatial, and spectral patterns from the differences in plant phonological profiles, sub-pixel border information, swift human interventions such as harvests, and the high agricultural dynamics [14].

A strategic interpretation of the cropland field boundaries segmentation challenge is approximate to a semantic segmentation problem. This interpretation enables leveraging deep convolution neural networks (CNN) to analyze remote sensing imagery, exploiting different hierarchical (local to global) features in images [38], such as used in some methodologies [3, 38]. Nonetheless, CNN-based approaches still lack represent the spectro-temporal variation that exists due to the change in phenology during crop growing.

In this manuscript, we address the cropland field boundaries segmentation challenge by proposing a novel end-to-end segmentation framework. Our framework comprises three main pipelines: data preparation pipeline, in which Sentinel-2 MSI sensor images are handled; segmentation pipeline, where we propose the use of three different methods to obtain a cropland field map; and post-processing pipeline, where we propose a novel temporal aggregation methodology to enhance crop field boundary delineation. We highlight as main contributions of this manuscript:

(i) A novel cropland boundaries segmentation framework;

(ii) A temporal aggregation post-processing method;

(iii) Set of object-based metrics to evaluate resulting identified cropland polygons.

2. Related Work

Field boundaries can be defined as the boundaries where a change in crop type, crop mixture, or farm management practice takes place, or where two similar cultivation are separated by a disruption in the landscape, *e.g*., a road [25]. Conceptually, extracting crop field boundaries consists of labeling each pixel of a multi-spectral image with one of two classes "boundary" or "not boundary" [38]. While crop field boundaries can be manually delimited by visual interpretation of images, it is time-consuming, and as farmers can change the distribution of their fields from one crop season to another, manually delimiting them can be costly as well. Field boundaries delimitation automatic methods can be roughly divided into traditional algorithms and deep learning algorithms [19].

Traditional methods include edge-based, region-based, graph-based, and hybrid approaches. Edge-based relies on a spatial gradient from discontinuities in images where pixel values change rapidly by applying various filters with specific kernels such as Scharr, Sobel, and Canny operators [17,35]. The region-based method groups adjacent pixels into the same field according to a predefined uniformity criterion [42]. In graph-based, weight is associated with each edge based on some property of the pixels that it connects, this weight represents a dissimilarity between the two pixels connected by that edge, such as their image intensities [15]. Sahadevan [29] applied the Felzenszwalb's algorithm [15], a graph-based method, on airborne hyperspectral images to extract homogeneous crop fields, obtaining an Error of 0.276, Precision of 0.568 and Recall of 0.999. Wagner and Oppelt [37] applied image enhancement techniques and merged edge-based and graph-based techniques to delimit the field boundaries in Germany, identifying the number of fields with a difference between -8.9% and 8.3% for the actual number of fields.

Recently, deep learning networks have gained attention in the cropland field boundary detection task. Among the deep network typologies, Convolutional Neural Networks (CNNs) became very popular in image analysis and semantic segmentation because of their capability to learn a hierarchy of spatial features at different levels of abstraction [25]. The U-Net architecture [26], made major breakthroughs in image semantic segmentation task when introducing the encoder-decoder paradigm. The U-Net has also been used for crop field boundary delimitation. Ajadi *et al*. [3] combined U-Net with the Fully Convolutional Neural Network (FCN), what they called the Boundary Net model, to extract crop field boundary and improve the crop type and crop area mapping across Brazil combining synthetic aperture radar and optical imagery. Their field boundary delimitation approach achieved an F1 score of 93%. Unlike traditional CNNs, which predict one class label per input image, FCNs are designed to infer pixel-wise predictions directly, independently from the size of the input image [25], being essential for semantic segmentation. Several works also used U-Net enhancements. Waldner and Diakogiannis [38] used ResUNet-a, a fully connected U-Net architecture that features dilated convolution, and conditioned inference to identify the extent of fields, the field boundaries, and the distance to the closest boundary. To achieve instance segmentation of individual fields they used post-processing after the ResUNet-a. However, these approaches do not take advantage of the unique spectro-temporal variation that exists due to the change in phenology during crop growing.

3. End-to-end cropland field boundaries segmentation framework

In this section, we propose our framework for cropland field boundaries segmentation. As seen in Fig. 1, the framework comprises three main pipelines: *(i)* data preparation; *(ii)* segmentation; and *(iii)* post-processing. Briefly, the data

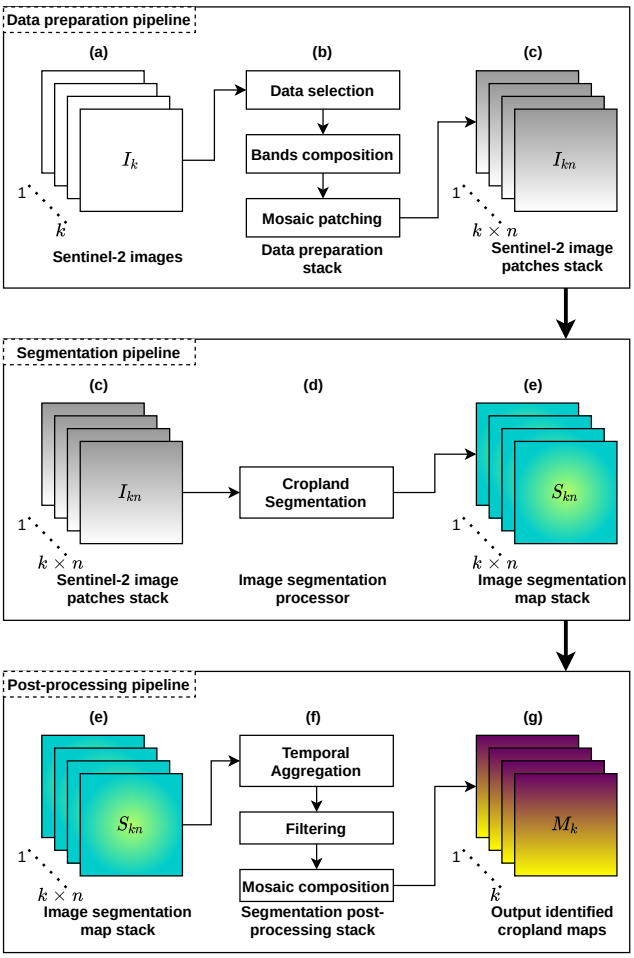


Figure 1. Overview of our proposed cropland field boundaries segmentation framework.

preparation pipeline manipulates Sentinel-2 raw data to filter cloudy dates, compute proper bands, and patch data into segmentation-ready images. The segmentation pipeline employs the segmentation processor to obtain a segmentation map for each image. Finally, our post-processing pipeline temporally aggregates the segmentation maps, and, then, composes patches to form a unique mosaic of identified cropland fields.

3.1. Input data and preparation

Our framework uses as input images acquired by the Multi-Spectral Instrument (MSI), onboard of the twin satellites Sentinel-2A and Sentinel-2B, launched in 2015 and 2017, respectively, within the European Copernicus program [11]. MSI sensor acquires images of the Earth's surface in 13 spectral bands from visible, Near InfraRed (NIR) and Short Wave InfraRed (SWIR) regions of the electromagnetic spectrum at 10, 20, and 60 m [16]. Sentinel-2/MSI Level-1C (L1C) images from 01-Sep-2020 to 30-

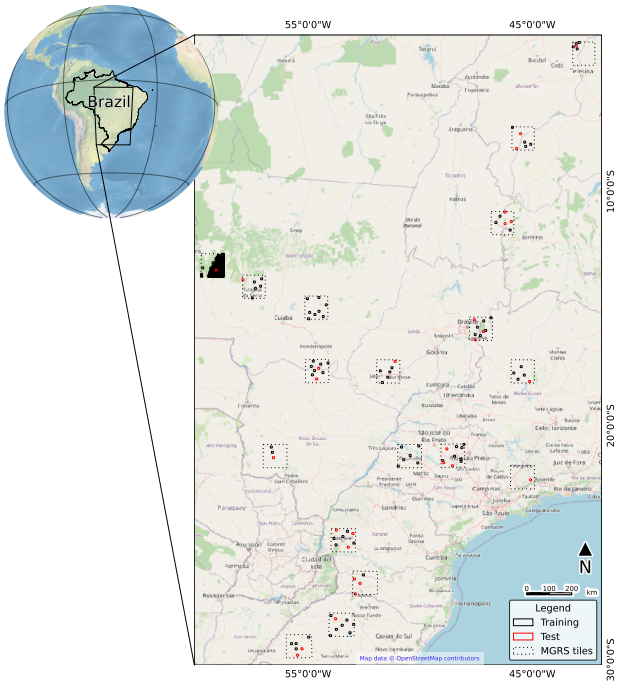


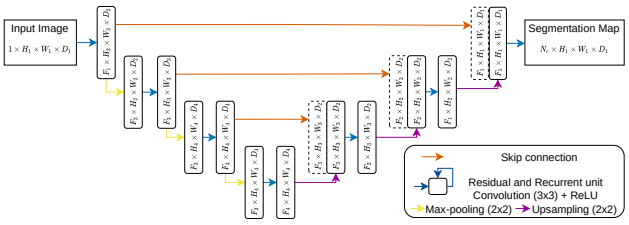
Figure 2. Locations of the training and test patches.

May-2021 with less than 15% cloud cover were accessed from Amazon Web Services [1] and stored in Amazon Simple Storage Service (Amazon S3) bucket for further processing. The range of date of the images was selected to include the period of growth of the major agricultural crops in the first crop season in the principal farmland regions in Brazil. Sentinel-2/MSI L1C is a product radiometrically and geometrically corrected with ortho-rectification, and provided in Top Of Atmosphere (TOA) reflectances [13]. We used all the spectral bands provided at 10 m and 20 m, *i.e.*, bands Blue (B2), Green (B3), Red (B4), Red-edge 1 (B5), Red-edge 2 (B6), Red-edge 3 (B7), NIR (B8), NIR Narrow (B8A), SWIR 1(B10), and SWIR 2 (B11) [2]. The 20 m spectral bands were resampled to 10 m using the nearest neighbor resampling method. In addition to the 10 spectral bands, the Normalized Difference Vegetation Index (NDVI) [27] (calculated using the B8) and the Roughness [41] of the B8 were used as additional input.

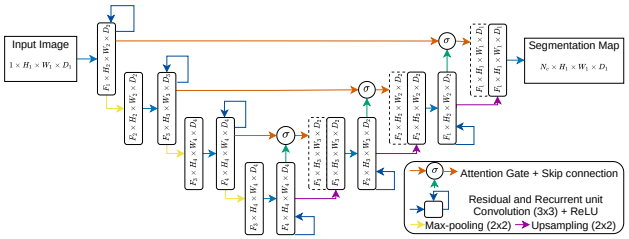
The selected images are illustrated in Fig. 2 and cover 18 Military Grid Reference System (MGRS) tiles distributed over Brazil, covering the 314 patches used in our analyses. MSI images were clipped according to each patch, which has dimensions of 500x500 10m pixels. The patches were divided into training and test, with 285 for the further and 29 for the latter (black and red dots in Fig. 2). Each patch was labeled by visual interpretation of Sentinel-2/MSI

[1] https://registry.opendata.aws/sentinel-2

[2] https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument



(a) U-Net architecture.



(b) Residual and Recurrent Attention U-Net architecture.

Figure 3. Convolution neural networks segmentation methodologies architecture.

false-color composition (NIR-SWIR-red) with cloud cover less than 10%. Cropland and consolidated pasture fields were delimited and labeled as one class, and the other targets were labeled as another class. Three validation patches were randomly selected in each Brazilian state where there were patches, to ensure that they are representative of the different characteristics of size and shape of fields, agricultural crops, cloud cover, relief, and soil, among others, that exist in Brazil.

3.2. Cropland field boundaries segmentation

Aiming to achieve crop field boundaries segmentation, *i.e.*, identify and outline croplands, we define the challenge as semantic segmentation, or image segmentation, which is a form of pixel-level prediction, defined by clustering parts of an image together that belong to the same object class. As such, we propose and evaluate the employment of the three different segmentation methodologies. One using a graph-based image segmentation algorithm (Felzenszwalb's algorithm [15]) and two others using convolution neural networks (U-Net [26] and Residual and Recurrent Attention U-Net (R2AttU-Net) [2, 22]).

Felzenszwalb's algorithm Felzenszwalb and Huttenlocher [15] developed an efficient graph-based image segmentation approach, commonly referred to here as Felzenszwalb's algorithm. Their approach is formulated as given an image I in the form of an undirected graph $G = (V, E)$, where $v_i \in V$ is a vertex within G (a pixel in I) and $(v_i, v_j) \in E$ corresponding to pairs of neighboring vertices (adjacent pixels in I), the goal is to produce a new graph $G_{seg} = (V, E')$, such that G_{seg} is a segmentation of G, containing distinct and likely-homogeneous regions of G, and $E' \subset E$. In our framework, we use Felzenszwalb's algorithm as a segmentation technique, and I being each patch and G_{seg} an identified cropland field.

U-Net The U-Net architecture, proposed by Ronneberger *et al.* [26] and illustrated in Fig. 3a, was first developed aiming to perform biomedical image segmentation, but soon became the state-of-the-art solution for image segmentation and the baseline for newer segmentation neural network. The main idea of the architecture is to compile an encoder, known as contraction path, with a decoder, known as expansion path. The contraction path is composed of stacked convolutional and max pooling operations, which learn the image contexts' representation. The expansion path is responsible to unfold the learned representation and spatially align the learned information. To do so, this path consists of stacked upsampling operations that enable precise localization by transposing convolutions through a skip connection [18] mechanism.

Residual and Recurrent Attention U-Net The Residual and Recurrent Attention U-Net (R2AttU-Net) model, illustrated in Fig. 3b, is a variation of the U-Net segmentation model. The model combines residual and recurrent units [2] in the encoding path to enhance the feature learning processing with attention gates that spatially filter the features propagated through the skip connections in the decoding path [22].

We also introduce two modifications to the CNN-based segmentation methodologies. We replace the traditionally used Rectified Linear Unit (ReLU) activation function [2, 22, 26] to using the Sigmoid-Weighted Linear Unit (SiLU) activation function [12]. Also, we modified the traditionally used cross-entropy loss function to using Lovász-Softmax loss [7]. Both modifications improve significantly our training efficiency by fasting model training convergence while keeping results stable.

3.3. Post-processing

To enhance our results and better outline the detected croplands, we propose the use of a series of filters and post-processing methods. Thus, in the following, we discuss the temporal aggregation post-processing method as well as area-based and MapBiomas-based filters.

Temporal aggregation The proposed post-processing pipeline uses temporal information by combining multiple predictions to better outline the target croplands, which can increase overall accuracy, especially when more than four observations are available [38]. Thus, as described in Section 3.1, all images with less than 15% clouds from between 01-Sep-2020 to 30-May-2021 in the 314 patches were selected in our experiments. Thus, we combine the multiple patches' segmentation map of different dates within a single stack. Then, for each pixel, we propose computing the element corresponding to a specific percentile and retrieving this value as the single value that represents the pixel.

Filtering To remove pixels that may have been misclassified, we use the MapBiomas land use and land cover (LULC) map as a mask. MapBiomas [32] is the most recent LULC mapping initiative in Brazil, covering the entire country annually between 1985 and 2020 in Collection 6. All non-crop and non-pasture-related pixels in MapBiomas were filtered out from our results. In addition to MabBiomas filter, areas with less than 1ha were filtered out, to remove isolated pixels and noise from small areas, because in Brazil, usually, the fields are larger than 1ha.

4. Experiments

We conduct experiments to evaluate the performance of the proposed cropland field boundaries segmentation framework. In the following, we present the experimental settings, our results, and our quantitative and qualitative analyses.

4.1. Experimental setting

Datasets In our experiments we aim to evaluate Felzenswalb's segmentation algorithm, and the neural networks U-Net and R2AttU-Net, described in Section 3.2. Both neural networks are trained using the training and results are gathered from the test dataset, both described in Section 3.1.

Implementation details We implemented the neural networks using PyTorch [3] and trained using a single Nvidia k80 GPU [4]. In our best experimentation settings, U-Net uses a learning rate of 10^{-5} with Adam optimizer [20], batch size of 32, and 20 training epochs with early stopping of 0.01 delta and 5 patience. R2AttU-Net uses the same settings as U-Net, with the addition of recurrent units of 2. The Felzenswalb's algorithm uses the scikit-image [5] publicly available implementation with the scale factor of 20, sigma of 0.5, and 100 being the minimum number of pixels as settings.

Evaluation protocol As described in Section 1, our key challenge is to best outline cropland areas from a given image. like so, we evaluate the solutions based on their capacity to correctly outline these areas. We report results from a set of metrics, properly divided into pixel-based and object-based. The pixel-based metrics set includes Precision (Pr), Recall (Re), F_1-score (F1), and Intersection over Union (IoU) and are performed at a pixel-level over the raster images. The object-based metrics set are performed at an object-level over the derived polygons from the original raster images, and includes Under-Segmentation (US), Over-Segmentation (OS), General Segmentation Error (GSE), and Polygons and Line Segments (PoLiS) [1].

[3] https://pytorch.org/

[4] https://www.nvidia.com/en-gb/data-center/tesla-k80/

[5] https://scikit-image.org/

The pixel-based metrics Precision, Recall, and F_1-score shown in Equations 1, 2, and 3, respectively, measure the pixel classification provided by each model on the raster images. The Precision evaluates the ratio of cropland pixels correctly classified given all classifications. The Recall measures the fraction of correctly classified pixels overall. The F_1-score takes the harmonic mean between Precision and Recall to combine both into a single meaningful metric. The fourth pixel-based metric is the Intersection over Union (IoU), Equation 4. This metric assesses the overlapping pixels of predicted cropland pixels (P) given their ground truth pixels (GT).

$$\text{Pr} = \frac{TP}{TP + FP} \quad (1) \qquad \text{Re} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1} = 2 * \frac{Pr * Re}{Pr + Re} \quad (3) \qquad \text{IoU} = \frac{|P \cap GT|}{|P \cup GT|} \quad (4)$$

On the object-based evaluation side, we evaluate the derived polygons. To do so, we propose the use of three metrics to understand and assess the matching areas of predicted cropland objects versus their ground truth objects. Like so, the Under-Segmentation metric (Equation 5) measures how much area of the prediction object (P_o) is missing from the ground truth object (GT_o). The Over-Segmentation metric (Equation 6) measures how much area of the prediction object is outside the borders of its correspondent ground truth. Following, we employ the root mean squared error over both Under- and Over- Segmentation metrics to evaluate a single polygon area error, called General Segmentation Error (Equation 7).

$$\text{US} = 1 - \frac{|P_o \cap GT_o|}{|P_o|} \quad (5)$$

$$\text{OS} = 1 - \frac{|P_o \cap GT_o|}{|GT_o|} \quad (6)$$

$$\text{GSE} = \sqrt{\frac{US^2 + OS^2}{2}} \quad (7)$$

Finally, we also employ the Polygons and Line Segments metric [1], shown in Equation 8. The PoLiS is a metric originally designed to compare building polygons, however, we employ it over the cropland polygons. The metric considers polygons as a sequence of connected edges to compute differences, approximately linearly, in translation, rotation,

Table 1. Results obtained from pixel- and object- based metrics from the cropland field boundaries segmentation methodologies with the post-processing pipeline and without post-processing. **Bold** results indicates the best results from each pixel-based metric, and **bold with underline** indicates the best results from each object-based metric.

2* Approach	*Pixel-based*				*Object-based*			
	Pr	Re	F1	IoU	US	OS	GSE	PoLiS
Felzenswalb	0.8083	0.8628	0.8347	0.7163	0.6338	0.6841	0.6594	647.15
U-Net	0.8728	0.7923	0.8306	0.7103	0.6735	0.8862	0.7870	709.62
R2AttU-Net	0.7825	**0.8914**	0.8334	0.7172	0.6476	0.5747	0.6122	731.49
Felzenswalb *+post-processing*	0.8375	0.8379	**0.8377**	**0.7208**	0.3802	0.4032	0.3918	177.30
U-Net *+post-processing*	**0.9380**	0.7568	**0.8377**	**0.7208**	**0.3506**	**0.3711**	**0.4521**	**175.46**
R2AttU-Net *+post-processing*	0.9262	0.7640	0.8373	0.7201	0.5386	0.4056	0.4767	289.91

and scale between predicted and ground truth polygons. p_j and g_k represent each vertex of the polygons P_o and GT_o, respectively. g and p represent the points of the polygons GT_o and P_o closest to the vertices p_j and g_k, respectively. q and r represent the number of vertices in polygons P_o and GT_o, respectively.

$$\text{PoLiS} = \frac{1}{2q}\sum_{p_j \in P_o} \min_{g \in \partial GT_o} ||p_j - g|| + \frac{1}{2r}\sum_{g_k \in GT_o} \min_{p \in \partial P_o} ||g_k - p|| \quad (8)$$

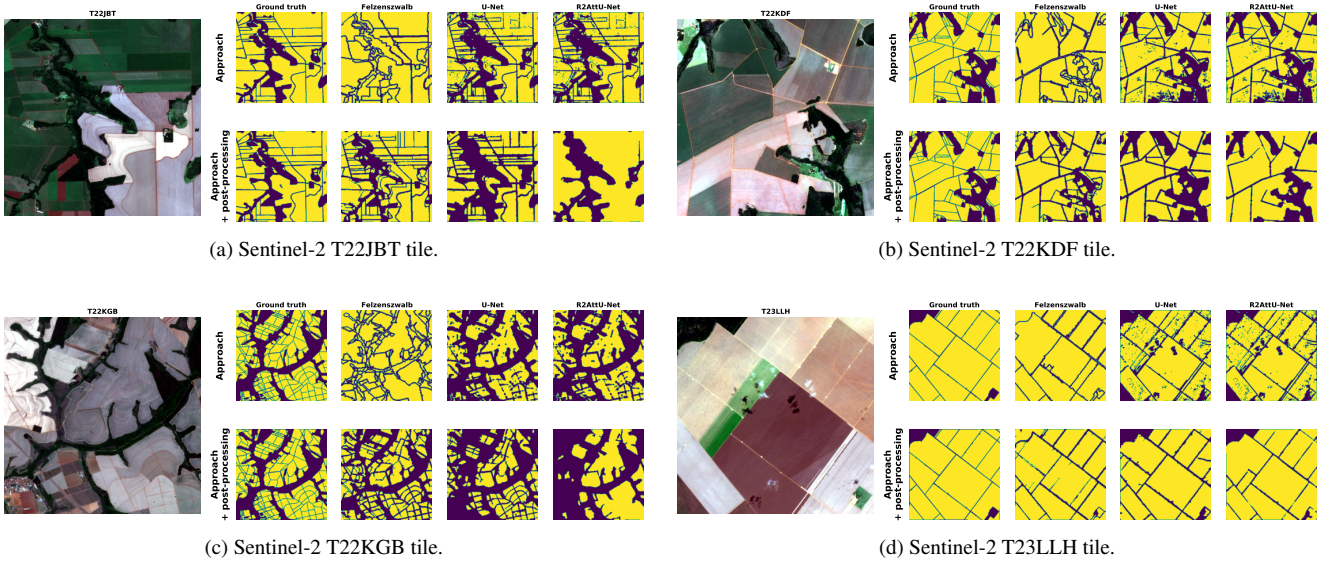
All accuracy metrics range from 0 to 1, with the exception of PoLiS. Values of Precision, Recall, F1-score, and Intersection over Union equal to 1 represent the ideal cropland field segmentation. While for Under-, Over- Segmentation, and General Segmentation Error, values equal to 0 represent the ideal cropland field segmentation. Lower values of PoLiS indicate the edges of polygons closer to the edges of ground truth polygons.

4.2. Quantitative evaluation

Results displayed in Table 1 show that Felzenswalb's segmentation algorithm has overall better sensitivity than other methods while having the worst precision. This result aligns with our initial expectations that this approach is able to cover most of the existing cropland, however, it suffers from the identification of true cropland areas. R2attU-Net has overall mixed results but lacks quality with the derived cropland areas. Finally, U-Net has the most balanced results. It combines high precision with good sensitivity. Besides, it features the smaller GSE and PoLiS, showing that the identified croplands better fit in scale, rotation, and translation of the ground truth polygons. Regarding the use of the post-processing pipeline, with the exception of recall, it improved all other pixel-based metrics. Besides, the post-processing pipeline efficiently improves results from object-based metrics, reducing up to 43% of the GSE metric and 76% of the PoLiS metrics on U-Net. These results prove the effectiveness of the proposed percentile-based temporal aggregation method as a post-processing solution to improve the cropland field boundary delimitation.

4.3. Qualitative evaluation

Fig. 4 illustrates the results of crop field segmentation using U-Net, R2AttU-Net, and Felzenszwalb's algorithm before and after post-processing for four patches from test dataset. Visually, in general, the results from U-Net and Felzenszwalb's method are more similar to the ground truth, although in some cases U-Net fails to segment part of the cropland fields (*e.g.*, patch from T22KGB) and the Felzenszwalb's algorithm lacks precision (*e.g.*, patch from T22JBT). On the other hand, R2AttU-Net has more difficulty in identifying roads and field boundaries leaving connected fields, as can be seen in the T22JBT and T22KGB patches. The post-processing is an important step, in which it removes non-cropland in the Felzenszwalb's approach, and helps to better delineate the fields and remove isolated pixels in the U-Net approach, as can be seen in the T22JBT and T22KGB patches. The results of the cropland segmentation by the three methods are dependent on the characteristics of the cropland fields (*e.g.*, size and shape) and on the characteristics of the images (*e.g.*, reflectance according to the phenological stage of the crops). Overall, the R2AttU-Net is able to identify croplands well but is not able to delimit the fields correctly. In a few cases, the post-processing method ends up dissolving the smaller identified polygons (*e.g.*, patch from T22JBT), however, it significantly improves overall results. Felzenszwalb's algorithm was able to segment the fields well but generates over-segmentation in some cases. It is dependent on the quality of the input data for post-processing, often segmenting areas that are not cropland. U-Net showed good results overall,



(a) Sentinel-2 T22JBT tile.

(b) Sentinel-2 T22KDF tile.

(c) Sentinel-2 T22KGB tile.

(d) Sentinel-2 T23LLH tile.

Figure 4. Visual comparison between cropland field boundaries segmentation of Felzenswalb's algorithm, U-Net, and R2AttU-Net approaches with the post-processing pipeline and without post-processing on four patches from test dataset.

although some small fields may not be correctly identified (*e.g.*, patch from T22JBT). Post-processing improves the results in most cases, removing isolated pixels and delimiting the fields better (*e.g.*, patch from T22KDF and T23LLH).

5. Conclusion

In this paper, we propose a novel end-to-end framework for cropland field boundaries segmentation whose objective was to enhance the existing ML-based pipelines using image processing techniques. The study emphasizes the remarkable enhancements achieved through the application of the post-processing methodology, particularly for the U-Net segmentation approach. The application of post-processing not only greatly enhances the performance of U-Net but also outperforms other methods with a remarkable increase in precision (0.9380) and F1 score (0.8377).

Object-based metrics further underscore the effectiveness of the post-processing technique, with a notable 43% reduction in the GSE metric for U-Net+*post-processing*. These findings emphasize the critical role of the proposed percentile-based temporal aggregation method as a post-processing solution, making U-Net an even more powerful tool for cropland analysis and precision agriculture. In future work, we highlight the opportunity to combine both Felzewalb's identified fields and U-Net's identified cropland fields as a possible improvement.

References

[1] A metric for polygon comparison and building extraction evaluation. *IEEE Geoscience and Remote Sensing Letters*, 12:170–174, 2015. 5

[2] Recurrent residual u-net for medical image segmentation. *Journal of Medical Imaging*, 6:014006, 3 2019. 4

[3] Olaniyi A. Ajadi, Jeremiah Barr, Sang-Zi Liang, Rogerio Ferreira, Siva P. Kumpatla, Rinkal Patel, and Anu Swatantran. Large-scale crop type and crop area mapping across Brazil using synthetic aperture radar and optical imagery. *International Journal of Applied Earth Observation and Geoinformation*, 97:102294, may 2021. 1, 2

[4] Kwasi Appeaning Addo. Urban and Peri-Urban Agriculture in Developing Countries Studied using Remote Sensing and In Situ Methods. *Remote Sensing*, 2(2):497–513, feb 2010. 1

[5] Clement Atzberger. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing*, 5(2):949–981, 2013. 1

[6] Bahaa Awad and Isin Erer. One Stage Deep Learning Based Method for Agricultural Parcel Boundary Delineation in Satellite Images. In *2021 13th International Conference on Electrical and Electronics Engineering (ELECO)*, pages 1–4. IEEE, nov 2021. 2

[7] Maxim Berman, Amal Rannen Triki, and Matthew B. Blaschko. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 4413–4421, 5 2017. 4

[8] R. G.V. Bramley and J. Ouzman. Farmer attitudes to the use of sensors and automation in fertilizer decision-making: nitrogen fertilization in the Australian grains sector. *Precision Agriculture*, 20(1):157–175, 2019. 1

[9] John Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986. 1

[10] Emily S. Cassidy, Paul C. West, James S. Gerber, and Jonathan A. Foley. Redefining agricultural yields: From tonnes to people nourished per hectare. *Environmental Research Letters*, 8(3), 2013. 1

[11] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120:25–36, 2012. 3

[12] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, 107:3–11, 2 2017. 4

[13] ESA. SENTINEL-2 User Handbook. Technical Report 1, ESA, 2015. 3

[14] Vivien Sainte Fare Garnot and Loic Landrieu. Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4852–4861. IEEE, oct 2021. 2

[15] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59(2):167–181, sep 2004. 1, 2, 4

[16] F. Gascon, E. Cadau, O. Colin, B. Hoersch, C. Isola, B. López Fernández, and P. Martimort. Copernicus Sentinel-2 mission: products, algorithms and Cal/Val. *Earth Observing Systems XIX*, 9218(September 2014):92181E, 2014. 3

[17] Jordan Graesser and Navin Ramankutty. Semi-automatic detection of crop field parcels with Landsat imagery. *Remote Sensing of Environment*, 201(August):165–180, 2017. 2

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December:770–778, 12 2015. 4

[19] Hai Huan, Yuan Liu, Yaqin Xie, Chao Wang, Dongdong Xu, and Yi Zhang. MAENet: Multiple Attention Encoder-Decoder Network for Farmland Segmentation of Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*, 19, 2022. 2

[20] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. International Conference on Learning Representations, ICLR, 12 2015. 5

[21] Marcelo Nery, Rodrigo Santos, Wallas Santos, Vitor Lourenco, and Marcio Moreno. Facing digital agriculture challenges with knowledge engineering. In *2018 First International Conference on Artificial Intelligence for Industries (AI4I)*, pages 118–119, 2018. 1

[22] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. 4 2018. 4

[23] Sankar K. Pal and Pabitra Mitra. Multispectral image segmentation using the rough-set-initialized EM algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2495–2501, 2002. 1

[24] Nathan Pelletier and Peter Tyedmers. Forecasting potential global environmental costs of livestock production 2000-2050. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43):18371–18374, 2010. 1

[25] C. Persello, V. A. Tolpekin, J. R. Bergado, and R. A. de By. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sensing of Environment*, 231(June), 2019. 2

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer International Publishing, Cham, 2015. 2, 4

[27] J. W. Rouse, R. H. Haas, J. A. Schell, D. W. Deering, and J. C. Harlan. Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. Technical Report September 1972, NASA/GSFC, 1974. 3

[28] Anna Rydberg and Gunilla Borgefors. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11):2514–2520, 2001. 1

[29] Anand S Sahadevan. Extraction of spatial-spectral homogeneous patches and fractional abundances for field-scale agriculture monitoring using airborne hyperspectral images. *Computers and Electronics in Agriculture*, 188:106325, sep 2021. 2

[30] Raí A. Schwalbert, Telmo Amado, Geomar Corassa, Luan Pierre Pott, P. V.Vara Prasad, and Ignacio A. Ciampitti. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology*, 284(December 2019):107886, 2020. 1

[31] Hemant Servia, Sajid Pareeth, Claire I. Michailovsky, Charlotte de Fraiture, and Poolad Karimi. Operational framework to predict field level crop biomass using remote sensing and data driven models. *International Journal of Applied Earth Observation and Geoinformation*, 108(February), 2022. 1

[32] Carlos M. Souza, Julia Z. Shimbo, Marcos R. Rosa, Leandro L. Parente, Ane A. Alencar, Bernardo F.T. Rudorff, Heinrich Hasenack, Marcelo Matsumoto, Laerte G. Ferreira, Pedro W.M. Souza-Filho, Sergio W. de Oliveira, Washington F. Rocha, Antônio V. Fonseca, Camila B. Marques, Cesar G. Diniz, Diego Costa, Dyeden Monteiro, Eduardo R. Rosa, Eduardo Vélez-Martin, Eliseu J. Weber, Felipe E.B. Lenti, Fernando F. Paternost, Frans G.C. Pareyn, João V. Siqueira, José L. Viera, Luiz C.Ferreira Neto, Marciano M. Saraiva, Marcio H. Sales, Moises P.G. Salgado, Rodrigo Vasconcelos, Soltan Galano, Vinicius V. Mesquita, and Tasso Azevedo. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. *Remote Sensing*, 12(17), 2020. 5

[33] Prasad S. Thenkabail. Global croplands and their importance for water and food security in the twenty-first century: Towards an ever green revolution that combines a second

green revolution with a blue revolution. *Remote Sensing*, 2(9):2305–2312, 2010. 1

[34] Xiaoyu Tian, Bernie A. Engel, Haiyang Qian, En Hua, Shikun Sun, and Yubao Wang. Will reaching the maximum achievable yield potential meet future global food demand? *Journal of Cleaner Production*, 294:126285, 2021. 1

[35] Mustafa Turker and Emre Hamit Kok. Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 79:106–121, 2013. 2

[36] Hannah H.E. van Zanten, Herman Mollenhorst, Cindy W. Klootwijk, Corina E. van Middelaar, and Imke J.M. de Boer. Global food supply: land use efficiency of livestock systems. *International Journal of Life Cycle Assessment*, 21(5):747–758, 2016. 1

[37] Matthias P. Wagner and Natascha Oppelt. Extracting agricultural fields from remote sensing imagery using graph-based growing contours. *Remote Sensing*, 12(7), 2020. 2

[38] 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, aug 2020. 2, 4

[39] Sherrie Wang, Francois Waldner, and David B. Lobell. Unlocking large-scale crop field delineation in smallholder farming systems with transfer learning and weak supervision. jan 2022. 1

[40] Alyssa K. Whitcraft, Inbal Becker-Reshef, Christopher O. Justice, Lauren Gifford, Argyro Kavvada, and Ian Jarvis. No pixel left behind: Toward integrating Earth Observations for agriculture into the United Nations Sustainable Development Goals framework. *Remote Sensing of Environment*, 235(November 2018):111470, 2019. 1

[41] Margaret F. J. Wilson, Brian O'Connell, Colin Brown, Janine C. Guinan, and Anthony J. Grehan. Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope. *Marine Geodesy*, 30(1-2):3–35, 5 2007. 3

[42] Huanxue Zhang, Mingxu Liu, Yuji Wang, Jiali Shang, Xiangliang Liu, Bin Li, Aiqi Song, and Qiangzi Li. Automated delineation of agricultural field boundaries from Sentinel-2 images using recurrent residual U-Net. *International Journal of Applied Earth Observation and Geoinformation*, 105(September):102557, 2021. 2