PARCS: A Deployment-Oriented AI System for Robust Parcel-Level Cropland Segmentation of Satellite Images

Chen Du,[1,*] **Yiwei Wang,**[2,3,*,†] **Zhicheng Yang,**[1] **Hang Zhou,**[1] **Mei Han,**[1] **Jui-Hsin Lai**[1]
[1]PAII Inc., Palo Alto, CA 94306, USA
[2]University of Science and Technology of China, Hefei, Anhui 230026, China
[3]Ping An Technology, Shenzhen, Guangdong 518048, China
chendu.future@gmail.com, zcyangpingan@gmail.com, juihsin.lai@gmail.com

Abstract

Cropland segmentation of satellite images is an essential basis for crop area and yield estimation tasks in the remote sensing and computer vision interdisciplinary community. Instead of common pixel-level segmentation results with salt-and-pepper effects, a parcel-level output conforming to human recognition is required according to the clients' needs during the model deployment. However, leveraging CNN-based models requires fine-grained parcel-level labels, which is an unacceptable annotation burden. To cure these practical pain points, in this paper, we present PARCS, a holistic deployment-oriented AI system for **PAR**cel-level **C**ropland **S**egmentation. By consolidating multi-disciplinary knowledge, PARCS has two algorithm branches. The first branch performs pixel-level crop segmentation by learning from limited labeled pixel samples with an active learning strategy to avoid parcel-level annotation costs. The second branch aims at generating the parcel regions without a learning procedure. The final parcel-level segmentation result is achieved by integrating the outputs of these two branches in tandem. The robust effectiveness of PARCS is demonstrated by its outstanding performance on public and in-house datasets (an overall accuracy of 85.3% and an mIoU of 61.7% on the public PASTIS dataset, and an mIoU of 65.16% on the in-house dataset) . We also include subjective feedback from clients and discuss the lessons learned from deployment.

1 Introduction

Cropland segmentation is an integral part of utilizing cropland satellite image data to efficiently achieve crop types and areas without cumbersome on-site measurement. Compared to pixel-level cropland segmentation, a more favorable segmentation output format is *parcel*-level, partitioning cropland into individual reasonable pieces. Numerous previous studies focused on this segmentation task in a supervised manner (M Rustowicz et al. 2019; Garcia-Pedrero et al. 2019; Sun, Di, and Fang 2019; Garnot and Landrieu 2021; Martinez et al. 2021). However, the cost of parcel-level annotation for training cannot be ignored. Many other research studies explored unsupervised approaches (Yang et al. 2021; Cheng et al. 2020), but along with the research work using supervised learning, their work used limited study sites to demonstrate the effectiveness of the proposed methods. Therefore, insufficient generalization capacity is still a challenge for confidently carrying out their solutions in practice.

One of the barriers to an AI system's effective deployment is the gap in domain knowledge among different disciplines. To confront this challenge, we aim to propose a deployment-oriented AI system to bridge this gap. As the clients provide a large area of interest (AoI) for cropland segmentation prediction, they also need to annotate some sample regions of this AoI for our model training. In this process, the clients mainly suffer from two pain points. First, the parcel-level annotation cost is not acceptable in terms of time and labor. Second, instead of a pixel-level segmentation output with severe salt-and-pepper effects, a meaningful parcel-level segmentation output is required, which should be consonant with human recognition.

To address the pain points above, we divide the parcel-level cropland segmentation into two sub-tasks: 1) *pixel-level crop segmentation* and 2) *parcel region extraction*. In the first sub-task, we still leverage the powerful deep learning-based training and inference paradigm, however, the parcel-level annotation is not needed. Instead, we design and develop a user-friendly annotation tool to obtain an annotator's limited effort on a few pixel samples, and leverage the active learning strategy (Settles, Craven, and Friedland 2008) with human-in-the-loop advantages to continuously improve the pixel-level segmentation accuracy. Compared with the fine-grained annotation in computer vision, both our pixel-level annotation and active learning-based training procedures are lightweight. In the second sub-task, we design our image processing-based algorithm dedicated to effectively extracting parcel regions. This design not only avoids annotating parcel-level labels, but also enables our system to provide a parcel-level output without any further training process. Admittedly, a parcel-level result might be generated by post-processing on a pixel-level segmentation output, such as using morphological methods to eliminate the salt-and-pepper effects. Nevertheless, it is difficult to set the kernel size or keep the parcel shape from distorting. Obtaining a parcel-level result by post-processing is thus still

[*]These authors contributed equally.

[†]This work was done during Yiwei Wang's internship at Ping An Technology, Shenzhen, China.

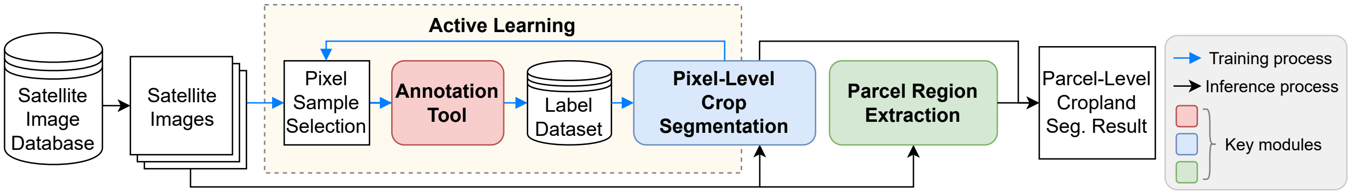


Figure 1: Framework of the proposed PARCS.

not reliable in clients' eyes. By consolidating these two subtasks, we remarkably relieve the heavy burden of the parcel-level annotation job originally assigned to an annotator, and guarantee the segmentation output in the parcel-level format.

In this paper, we propose PARCS, a deployment-oriented AI system for robust **PAR**cel-level **C**ropland **S**egmentation. Sentinel-2 satellite image data is utilized as our image source as commonly used in many previous studies (Garnot and Landrieu 2021; Garcia-Pedrero et al. 2019; Masoud, Persello, and Tolpekin 2019). Moreover, Sentinel-2 data has been included in our own satellite image database to efficiently support our system. Our key contributions are summarized as follows:

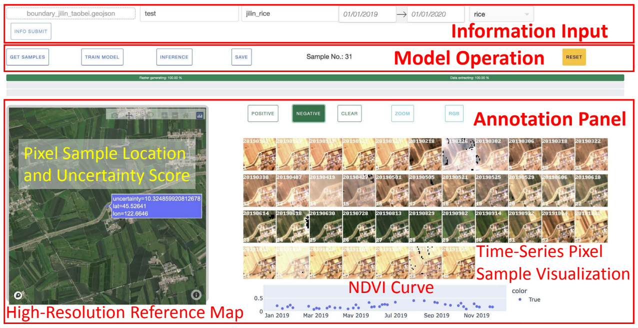
- PARCS is a holistic AI system for parcel-level cropland segmentation using satellite images. This system appropriately integrates multiple disciplinary knowledge from remote sensing, computer vision, image processing, and software engineering to precisely resolve clients' issues.
- With the designed annotation tool and the active learning strategy, PARCS needs pixel labels only and significantly improves the annotation paradigm by removing the parcel-level labeling costs.
- Our proposed method is robust to generate impressive parcel-level segmentation results, dramatically expediting the deployment of the entire system to clients. Our evaluation results demonstrate the outstanding effectiveness of our method on both public and in-house datasets.

2 System Design

Fig. 1 illustrates the framework of our proposed PARCS. We first acquire multi-temporal Sentinel-2 satellite images from our satellite image database relying on the given time period and area of interest (AoI). These images are used in both algorithm branches. In the pixel-level learning flow, the designed annotation tool initially collects pixel labels on a small number of samples. We then leverage the active learning strategy with the annotator's iterative input to constantly improve the prediction accuracy of pixel-level crop segmentation model. In the inference stage, a probability map generated by the trained pixel-level segmentation model is integrated with a corresponding parcel segmentation label map from the parcel region extraction flow. A parcel-level cropland segmentation result is finally achieved. In the following subsections, we elaborate on each module of our framework.

2.1 Image Data Source

As mentioned in Sec. 1, we use Sentinel-2 satellite images as our data source. First, they are free and easy to be obtained.


boundary_jilin_taobei.geojson
test
jilin_rice
01/01/2019 → 01/01/2020
rice

Sample No.: 31

Raster generating: 100.00 %
Data extracting: 100.00 %

uncertainty=10.324859920812678
lat=45.52641
lon=122.6646

Figure 2: User interface of the annotation tool.

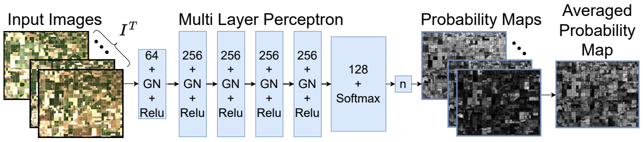


Figure 3: Model of pixel-level crop segmentation.

Second, Sentinel-2 satellites have a high revisit frequency of 5 days for most places in the world. Third, their highest resolution is better than that of other free satellites such as Landsat-8 (Chakhar et al. 2020). For example, Sentinel-2 satellites provide 10-meter (10-m) resolution images for red, green, blue, and near-infrared (RGBN) channels and 20-m or 60-m for other channels, while Landsat-8 only provides 30-m resolution images. Once the satellite images have been obtained, we resize all other channels to 10-m resolution with bilinear interpolation.

2.2 Annotation Tool

To facilitate the labeling experience and efficiency, we design an interactive annotation tool for technicians who annotate the sampled data. The user interface is shown in Fig. 2. A technician needs to provide a geospatial AoI description of the target area (e.g. a Geographic JavaScript Object Notation (GeoJSON) file) and a specific time period in the "information input" section. The system then extracts corresponding satellite images from our database. Initially, multiple samples are randomly selected to form a queue at the tool's backend and one sample is pushed to the frontend each time. To provide the technician with more surrounding visual information, we extract a small patch in which the sample point at the center. All the valid temporal patches of this sample point are displayed at the interface. A high-

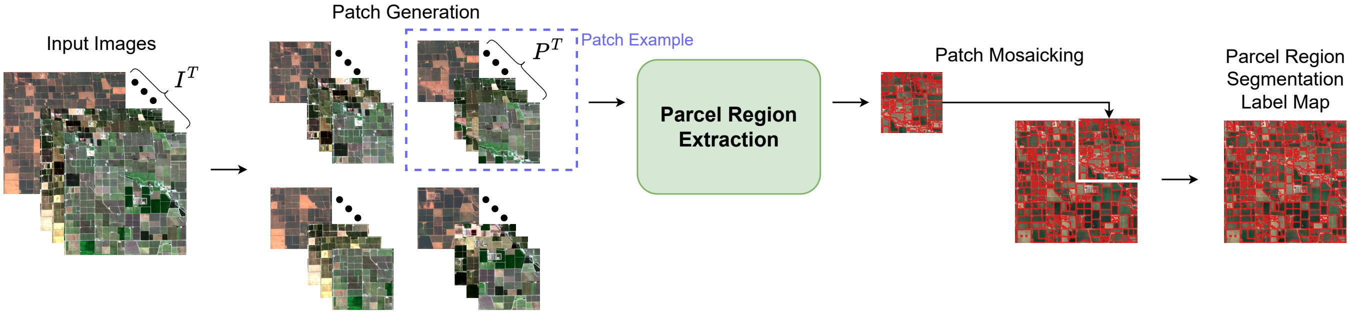


Figure 4: Workflow of the parcel region extraction module.

resolution image of the sample position is also provided via Mapbox (Mapbox 2003) on the left side of the interface for reference only. Meanwhile, a normalized difference vegetation index (NDVI) curve is also provided for each data captured time at the bottom, which is defined as NDVI=(NIR-Red)/(NIR+Red), where “NIR” and “Red” represent near-infrared and red bands, respectively. An NDVI value indicates the growth of plants on the grounds (Zheng et al. 2015) because each type of crops is supposed to have its own NDVI curve pattern. The technician can decide whether a sample belongs to a specific crop type based on the provided information. Once the label is selected, the annotation will be saved to the label dataset for model training.

After a few sample points are labeled, a pixel-level crop segmentation model can be trained by clicking the “train model” button. When the training is finished, the system next uses the trained model to re-evaluate all pixels in the AoI, and an uncertainty score from the active learning strategy (described in Sec. 2.4) is then calculated for each sample. New samples with ambiguous scores are chosen for the next round of annotation. The technician can iteratively label new samples and train new models until the scores for all samples show acceptable values. The “inference” button is designed for model inference on the AoI.

2.3 Lightweight Pixel-Level Crop Segmentation

Unlike other time-series data, due to the unforeseen conditions in space, it is common that satellite images have regions with cloud cover (unexpected white regions) or no data (abnormal black regions). Therefore, not all temporal data can be properly utilized. Moreover, the satellite trajectory causes different image capture time stamps for two adjacent image tiles. These uncertainties of time sequence hinder the utilization of time-series-based models such as RNN (Sun, Di, and Fang 2019) and LSTM (Shi et al. 2015; Rußwurm and Körner 2018), which assume that all input data are *valid* (i.e. without cloud cover or no data) and time stamps for all samples are the same. To mitigate this issue, we leverage the Sentinel-2 Scene Classification Layer (SCL) (Main-Knorn et al. 2017) to detect and replace the cloud-covered pixels with no data (0 values), so these two kinds of unavailable pixels can be filtered out by excluding 0 values.

We let T refer to the set of the total temporal sequence of Sentinel-2 data queries. $I^T \in \mathbb{R}^{|T|\times C\times H\times W}$ denotes the multi-temporal images acquired from our satellite image database, where I refers to a Sentinel-2 image tile. H, W, and C are the height, width, and number of channels of I, respectively. $s_k^{T_k^+}$ presents all the *valid* visual data at the location of k-th pixel sample. We use 70% of all K labeled pixel samples in $\mathcal{S} = \{s_1^{T_1^+}, ..., s_K^{T_K^+}\}$ as training set and the remaining 30% as validation set and apply the pixelwise Focal Loss(Lin et al. 2017) as our loss function. During the training process, the pixel sample data at every valid time stamp $s_k^{t_k}, t_k \in T_k^+$ is *individually* fed into the model. In the inference process, our model generates the prediction result for every pixel captured at each individual time stamp for $I^{T^+_{H,W}} = (I_{y,x}^{T^+_{y,x}})_{H\times W}$, and calculate the average probability map output.

Fig. 3 shows our model of pixel-level crop segmentation. The model adopts multilayer perceptron (MLP) to learn the patterns from the input of valid images along with the calculated NDVI channel with a limited number of annotated pixel samples. It predicts the probabilities of the specific crops for each pixel at every image captured time. Note that we use group normalization (GN) instead of batch normalization (BN) due to the small amount of annotated samples, so the batch size is not large enough for batch normalization.

2.4 Active Learning Strategy

Even though inference results can be obtained by labeling a moderate amount of sample points, we leverage the active learning strategy (Settles 2009) to improve model performance with the annotator’s effort on demand. Specifically, once obtaining the inference results on the AoI, we calculate a score to indicate the *uncertainty* of samples using entropy of the probability $u = \Sigma_t^{T^+} -p_t \log(p_t)$, where p_t is the prediction probability of the pixel sample at the time stamp t. Once the uncertainty value is calculated based on entropy, we use it to sort all the samples from large to small, and push them back to the front-end for the annotation by a user.

2.5 Parcel Region Extraction

Patch Generation. Fig. 4 depicts our designed parcel region extraction module. Due to the huge size of one image tile (e.g. 10,980×10,980), we crop it into smaller patches. Let $P^T \in \mathbb{R}^{|T|\times C\times H_P\times W_P}$ represent one multi-temporal patch from I^T, where H_P and W_P are the height and width of a patch P, respectively. Every patch is processed separately and has a one-pixel height/width overlap with the adjacent patches, so that all patch-level segmentation outputs can be correctly mosaicked back to the image-level result.

Parcel Region Extraction. In natural images, objects always have the distinct visual appearance, leading to satisfactory semantic contour extraction using traditional or deep learning-based edge detectors (Sobel 1982; Canny 1986; Xie and Tu 2015). Nonetheless, this observation is not easily applicable to satellite images. For example, some recent studies explored outlining parcels using deep learning methods (Masoud, Persello, and Tolpekin 2019; Qiao et al. 2019; Garnot and Landrieu 2021; Martinez et al. 2021; Huang et al. 2022), but their shown performance is not acceptable by the clients in our case. We find that the CNN-based models suffer from the highly frequent failure of complete boundary closure of parcels, even though a boundary is obviously identified by human eyes (e.g. Fig. 5c in Garcia-Pedrero et al. (2019)). This observation echos the distinction between a segmentation task of natural and satellite images. Meanwhile, other studies exploited unsupervised methods to delineate parcels, such as *superpixel*-based approaches, but over-segmentation and under-segmentation are always observed due to the difficult control of a superpixel's compactness (Garcia-Pedrero et al. 2018). The reasons for their moderate performance are mainly two-fold: 1) adjacent parcels might have very similar color and texture; 2) for 10-m resolution Sentinel-2 images, a parcel's boundary has two or even one-pixel width only. Mechanically applying the successful models and algorithms from computer vision without considering domain knowledge gap is not able to meet the actual clients' requirement. Moreover, these methods are not capable of deployment or generalization, using either a small study site or supervised approaches.

In order to address the issues above, we formulate the problem of parcel region extraction (i.e. parcel delineation) as a *search-and-compare* task on pixels within an image. Inspired by various search algorithms (Cormen et al. 2022) and similarity measures (Moreira, Carvalho, and Horvath 2018), we design the algorithm dedicated to extracting parcel regions. To better describe our algorithm, we let $P^T_{i,j} \in \mathbb{R}^{|T| \times C \times 1 \times 1}$ denote a pixel in a multi-temporal patch P^T, where $i \in \{1, ..., H_P\}$ and $j \in \{1, ..., W_P\}$, respectively. Let $\Omega(i, j)$ define the set of the neighboring pixels of each $P^T_{i,j}$. We calculate the similarity of $P^T_{i,j}$ and its neighbor $P^T_{m,n}$ for $\forall(m, n) \in \Omega(i, j)$, and adopt a search algorithm to partition the maximum regions which have the similar pixels. We then define $P^{T'}_{i,j}$ and $P^{T'}_{m,n}$, respectively as the adjacent pixels which share valid acquisition time existing in both T sequences (i.e. $T' = T^+_{i,j} \cap T^+_{m,n}$). The detailed algorithm flow is described in Alg. 1.

Parcel Mosaicking. Once the patch-level parcels are extracted, all patches are mosaicked together to generate the image-level result with the original image height and width ($H \times W$). The one-pixel overlap at the margin of each patch is updated depending on the length of the shared border.

Algorithm 1: Parcel Region Extraction

```
Input: one multi-temporal patch $P^T \in \mathbb{R}^{|T| \times C \times H_P \times W_P}$; a
    similarity threshold $\theta$
Output: one segmentation label map $L \in \mathbb{N}^{H_P \times W_P}$; a la-
    bel index $l$
Initialize: $L \leftarrow (L_{i,j} \leftarrow 0)_{H_P \times W_P}$; $l \leftarrow 1$; a queue
    $q \leftarrow \varnothing$ to save pixel position to be accessed.
for each pixel $P^T_{i,j}$ do
  if $L_{i,j} = 0$ then
    push pixel $(i, j)$ to the back of $q$
    $L_{i,j} \leftarrow l$
    while q do
      pop out the visited pixel from the front of $q$
      for each pixel $P^T_{m,n}$,$\forall(m, n) \in \Omega(i, j)$ do
        update $P^{T'}_{i,j}$, $P^{T'}_{m,n}$
        calculate similarity $g(P^{T'}_{i,j}, P^{T'}_{m,n})$
        if $L_{m,n} = 0$ and $g(P^{T'}_{i,j}, P^{T'}_{m,n}) > \theta$ then
          $L_{m,n} \leftarrow l$
          push pixel $(m, n)$ to the back of $q$
        end if
      end for
    end while
    $l \leftarrow l + 1$
  end if
end for
```

2.6 Parcel-Level Cropland Segmentation Integration

After achieving the averaged pixel-level crop probability map and the parcel region segmentation label map of I^T, we use a simple voting scheme to fill a parcel region relying on the most dominant crop class inside it, eliminating the salt-and-pepper effects within the parcel and outlining the parcel boundary. The final segmentation result can be exported as the GeoJSON and GeoTIFF formats for clients' use, which have prediction masks with geospatial information.

3 Performance Evaluation

In this section, we describe our implementation details and conduct experiments on a multi-temporal public dataset and our in-house dataset collected during deployment. We also present the subject feedback from clients on PARCS.

3.1 Implementation Details

Datasets. The public dataset we use is PASTIS (Garnot and Landrieu 2021), which provides multi-temporal agricultural parcel-level annotation using Sentinel-2 satellite data. It collects 2,433 image patches in France and annotates 18 crop types. The dataset divides images into 5 folds, and we perform 5-fold cross-validation to accomplish our results.

Our in-house dataset is captured in 2019, comprising 10 labeled areas for rice, corn, and wheat across 6 provinces in China. Each area has 1 crop type, and contains about 50 Sentinel-2 satellite images and the average image size is 6,000×5,000 pixels. Each area has fine-grained parcel-level

Model	Param. #	OA	mIoU
Train w/ parcel anno.			
ConvLSTM (2015; 2018)	1.010M	77.9	49.1
FPN-ConvLSTM (2021)	1.261M	81.6	57.1
U-ConvLSTM (2019)	1.508M	82.1	57.8
3D U-Net (2019)	1.554M	81.3	58.4
U-TAE (2021)	1.087M	83.2	**63.1**
Train w/ pixel anno.			
Ours	0.335M	**85.3**	61.7

Table 1: Performance comparison of our proposed method and completing methods on the public PASTIS dataset. The bold number refers to the best result and the underlined number denotes the second best result.

annotations for clients to verify our model performance. The original Sentinel-2 data has `uint16` data type, ranging from 0 to 65,535. We clip this value range to 0∼4,000 to keep the vast majority of meaningful data and then normalize all channel values to 0∼1. The NDVI channel with the values from -1 to 1 is concatenated with the normalized Sentinel-2 channels. Regarding the time stamp, we calculate the day of a year divided by 366 as a normalized input for the model.

Pixel-Level Segmentation Model Training. As aforementioned in Sec. 2.3, for training the pixel-based MLP model, we use 70% of all labeled samples as the training set and the rest of the labeled ones as validation data. The learning rate is set to 0.001 and the cosine annealing learning rate scheduler is used. The training epoch is set to 3000, taking around 10 to 20 minutes for the entire training process. We pick the best model on the validation set as our final model.

Parcel Region Extraction. In our search-and-compare scheme, we adopt Breadth-first search (BFS) as our search algorithm. The time complexity of parcel region extraction is $O(H_P W_P)$ and extra space complexity is also $O(H_P W_P)$ since we need a queue to save neighbors. Note that Depth-first search (DFS) is also applicable in our case. We use 8-connected neighbors as $\Omega(i, j)$, and leverage temporal average cosine similarity defined as in Eq. 1 to compare the neighboring pixels. In other words, we compute cosine similarity for each time stamp, and then calculate the mean of similarity over all time stamps. The default similarity threshold θ is set to 0.98.

$$g(P_{i,j}^{T'}, P_{m,n}^{T'}) = \left(\sum \frac{P_{i,j}^{t_{i,j}} P_{m,n}^{t_{m,n}}}{|P_{i,j}^{t_{i,j}}||P_{m,n}^{t_{m,n}}|} \right) / |T'|, t_{i,j}, t_{m,n} \in T' \tag{1}$$

Evaluation Metric. We use overall accuracy (OA) and mean intersection over union (mIoU) to align the other competing methods on the public dataset. We average the IoU values across all 10 areas for the in-house dataset.

Development. Our system is mainly built on Python3, using Plotly Dash for the user interface and PostgreGIS as the database. We use Cython, Numba as well as parallel processes to accelerate the speed of the system.

Model	Train #	mIoU
Ours w/ Initial Round	100/100	37.81
Ours w/ AL Round 1	50/150	53.32
Ours w/ AL Round 2	50/200	61.80
Ours w/ AL Round 3	50/250	63.93
Ours w/ AL Round 4	50/300	**65.16**
Pixel-Level w/ AL R4+FH (2004)	300/300	45.32
Pixel-Level w/ AL R4+SNIC (2017)	300/300	50.29

Table 2: Performance comparison of our proposed method and ablation studies on our in-house dataset. (AL: Active Learning; R4: Round 4; (#/#) in Train #: new/total pixels involved in each training round for each area)

3.2 Experiment Results

Results of Public Dataset. In this experiment, we use the active learning strategy and let the model select the samples that need to be labeled. We stop iteration when the uncertainty scores of new selected samples are less than the default threshold. Table 1 lists the performance comparison of our approach and other competing models. Previous studies mainly leverage CNN-based or time-series LSTM-based methods with parcel-level annotations. To the best of our knowledge, U-TAE (U-Net with Temporal Attention Encoder) is a state-of-the-art model for the PASTIS dataset (Garnot and Landrieu 2021). As we can see, our model is able to achieve competitive results compared with U-TAE, however, we train it with fewer labeled pixels rather than parcel-level labels, and our model size is ∼1/3 of U-TAE's.

Results of In-House Dataset. First, we evaluate the efficacy of active learning with 4 rounds. We label 50 pixel samples for each round except the initial round in which we label 100 samples. The first part of Table 2 unveils that our model accomplishes the acceptable mIoU values (>50.0) even has only one round of active learning. As the number of iterative rounds increases, the performance gradually rises up and tends to be converged. Second, the effectiveness of our parcel region extraction module is evaluated. We keep the best pixel-level crop segmentation model from Round 4 of active learning in the experiment above, and replace our parcel extraction module with some representative superpixel algorithms Felzenszwalb and Huttenlocher (FH) (Felzenszwalb and Huttenlocher 2004) and Simple Non-Iterative Clustering (SNIC) (Achanta and Susstrunk 2017), respectively. We here choose superpixel-based approaches because these algorithms are able to aggregate satellite image pixels into regions of varying sizes in an unsupervised way (Yang et al. 2021), and provide complete boundary closure results desired for parcel shapes. Note that the over-segmentation effect of superpixels is not an issue because our pixel-level output can visually aggregate two adjacent superpixels if the dominant crop class in them is the same.

The second part of Table 2 reveals the performance of our pixel-level models with these superpixel methods. Even though we have tuned the hyper-parameters of superpixel methods to achieve the relative best results, our model with only one round still greatly outperforms them. It is difficult to improve the local segmentation without impact on global performance via hyper-parameter tuning for superpixel algorithms. Specifically, FH is not stable for aggregating pixels near a parcel's boundaries. The cluster-based superpixel algorithm SNIC tends to generate superpixels with similar compactness, downgrading the parcel boundary delineation.

Fig. 5 provides three visualized results of our final parcel-level cropland segmentation in different locations across three provinces in our in-house dataset. The results show that our approach is greatly capable of segmenting the croplands in accurate parcel shapes regardless of various crop landscape. It also clearly separates all the city areas, most of the lanes, and even some areas hard to distinguish in the visual images, such as the holes on the top left quarter of the third position. Note that most of the defective predictions occur on the boundaries of the croplands, because the Sentinel-2 satellite images have the inherent spatial error of less or approximately 1 pixel. Hence, the pixels at the boundaries of the croplands possibly have 1-pixel misalignment with the cropland ground truth at different time stamps, but this irreparable error is acceptable.

3.3 Deployment of PARCS

To collect subjective feedback from clients on our system deployment, we design a questionnaire to cover usability and reliability (Felderer and Ramler 2021), especially concentrating on the comparison of our previous pixel-level cropland segmentation system and the current PARCS. The overall clients' rating significantly rises from 3.65 to 4.87. The most impressive compliments include the robust parcel-level segmentation result on the in-house dataset, and lightweight annotation procedure to greatly reduce the labor costs. We believe PARCS confidently overcomes the salient pain points of clients.

4 Conclusions and Discussion

In this paper, we propose our deployment-oriented AI system PARCS for parcel-level cropland segmentation of satellite images. To meet the clients' challenging needs of parcel-level segmentation results without providing parcel annotations, we carefully design a two-branch method to address these challenges. To reduce the huge overhead of a parcel-level labeling task, we train an active learning-based crop segmentation model with limited pixel sample labels instead of expensive parcel-level annotation. To guarantee parcel-level output, we design an algorithm of parcel region extraction for outlining parcel boundaries and further removing salt-and-pepper effects of the pixel-level output from the crop segmentation model. The final parcel-level segmentation results of PARCS are evaluated using public and in-house datasets. Both the experiment results and the subjective feedback confirm the robust effectiveness of PARCS. We also discuss valuable aspects of our system as follows.

Deployment Experience. We observe that PARCS can remarkably reduce the burden of both clients and ours. On one hand, lightweight pixel-level annotation and limited re-engagement in the active learning loop substantially reduce the workload of an annotator. On the other hand, the robustness and the time consumption of PARCS are very competitive, enabling us to provide an agile response to more clients with diverse needs. Although a performance metric is easy to evaluate a model, it does not mean everything. There is no significant difference between the mIoU values of parcel-level segmentation and pixel-level with salt-and-pepper segmentation results, but the latter output format is not acceptable in practice. In our scenarios, a robust parcel-level output rather than a slightly improved OA/mIoU value is essential to boost the success of our deployment.

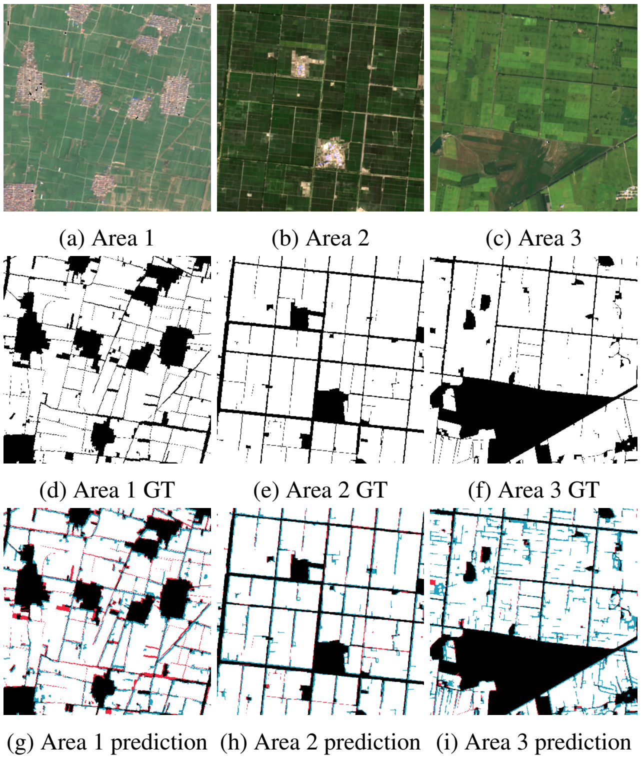


Figure 5: Visualization of our parcel-level segmentation results in three different areas in our in-house dataset. (In (d)-(f), GT: ground truth; In (g)-(i), white: true positive; black: true negative; red: false positive; blue: false negative)

Model Flexibility. It is natural that land cover in satellite images changes over time, such as from cropland to urban land. In this case, our system tends to be conservative to recognize and segment these changed areas out from the unchanged regions. Meanwhile, some clients need parcel boundary delineation without crop type identification. Under this circumstance, we deploy the standalone parcel region extraction module only to meet their demands.

Future work. One of our future tasks is to improve our capability to accept images with dense cloud cover or no data. We are also developing an online portal for PARCS with more capacity of human-computer interaction with potential customers. This way, users can gain real experience instead of inferring a system's feasibility from whitewashed demos.

References

Achanta, R.; and Susstrunk, S. 2017. Superpixels and polygons using simple non-iterative clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4651–4660.

Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6): 679–698.

Chakhar, A.; Ortega-Terol, D.; Hernández-López, D.; Ballesteros, R.; Ortega, J. F.; and Moreno, M. A. 2020. Assessing the accuracy of multiple classification algorithms for crop classification using Landsat-8 and Sentinel-2 data. *Remote sensing*, 12(11): 1735.

Cheng, T.; Ji, X.; Yang, G.; Zheng, H.; Ma, J.; Yao, X.; Zhu, Y.; and Cao, W. 2020. DESTIN: A new method for delineating the boundaries of crop fields by fusing spatial and temporal information from WorldView and Planet satellite imagery. *Computers and Electronics in Agriculture*, 178: 105787.

Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; and Stein, C. 2022. *Introduction to algorithms*. MIT press.

Felderer, M.; and Ramler, R. 2021. Quality Assurance for AI-Based Systems: Overview and Challenges (Introduction to Interactive Session). In *International Conference on Software Quality*, 33–42. Springer.

Felzenszwalb, P. F.; and Huttenlocher, D. P. 2004. Efficient graph-based image segmentation. *International journal of computer vision*, 59(2): 167–181.

Garcia-Pedrero, A.; Gonzalo-Martín, C.; Lillo-Saavedra, M.; and Rodríguez-Esparragón, D. 2018. The outlining of agricultural plots based on spatiotemporal consensus segmentation. *Remote Sensing*, 10(12): 1991.

Garcia-Pedrero, A.; Lillo-Saavedra, M.; Rodriguez-Esparragon, D.; and Gonzalo-Martin, C. 2019. Deep learning for automatic outlining agricultural parcels: Exploiting the land parcel identification system. *IEEE Access*, 7: 158223–158236.

Garnot, V. S. F.; and Landrieu, L. 2021. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4872–4881.

Huang, F.; Yang, Z.; Zhou, H.; Du, C.; Wong, A. J.; Gou, Y.; Han, M.; and Lai, J.-H. 2022. Unsupervised superpixel-driven parcel segmentation of remote sensing images using graph convolutional network. In *Companion Proceedings of the Web Conference 2022*, 1046–1052.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

M Rustowicz, R.; Cheong, R.; Wang, L.; Ermon, S.; Burke, M.; and Lobell, D. 2019. Semantic segmentation of crop type in africa: A novel dataset and analysis of deep learning methods. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 75–82.

Main-Knorn, M.; Pflug, B.; Louis, J.; Debaecker, V.; Müller-Wilm, U.; and Gascon, F. 2017. Sen2Cor for sentinel-2. In *Image and Signal Processing for Remote Sensing XXIII*, volume 10427, 37–48. SPIE.

Mapbox. 2003. Maps, geocoding, and navigation APIs & SDKs. https://www.mapbox.com/. Accessed: 2022-08-10.

Martinez, J. A. C.; La Rosa, L. E. C.; Feitosa, R. Q.; Sanches, I. D.; and Happ, P. N. 2021. Fully convolutional recurrent networks for multidate crop recognition from multitemporal image sequences. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171: 188–201.

Masoud, K. M.; Persello, C.; and Tolpekin, V. A. 2019. Delineation of agricultural field boundaries from Sentinel-2 images using a novel super-resolution contour detector based on fully convolutional networks. *Remote sensing*, 12(1): 59.

Moreira, J.; Carvalho, A.; and Horvath, T. 2018. *A general introduction to data analytics*. John Wiley & Sons.

Qiao, N.; Zhao, Y.; Lin, R.-S.; Gong, B.; Wu, Z.; Han, M.; and Liu, J. 2019. Generative-discriminative crop type identification using satellite images. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 1–5. IEEE.

Rußwurm, M.; and Körner, M. 2018. Convolutional LSTMs for cloud-robust segmentation of remote sensing imagery. *arXiv preprint arXiv:1811.02471*.

Settles, B. 2009. Active learning literature survey.

Settles, B.; Craven, M.; and Friedland, L. 2008. Active learning with real annotation costs. In *Proceedings of the NIPS workshop on cost-sensitive learning*, volume 1. Vancouver, CA:.

Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; and Woo, W.-c. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28.

Sobel, M. E. 1982. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological methodology*, 13: 290–312.

Sun, Z.; Di, L.; and Fang, H. 2019. Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *International journal of remote sensing*, 40(2): 593–614.

Xie, S.; and Tu, Z. 2015. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, 1395–1403.

Yang, L.; Wang, L.; Abubakar, G. A.; and Huang, J. 2021. High-resolution rice mapping based on SNIC segmentation and multi-source remote sensing images. *Remote Sensing*, 13(6): 1148.

Zheng, B.; Myint, S. W.; Thenkabail, P. S.; and Aggarwal, R. M. 2015. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *International Journal of Applied Earth Observation and Geoinformation*, 34: 103–112.