LiSegAgr:Labeled Instance Segmentation for Agricultural Remote Sensing Images through Iterative SAM

Yunkai Wang[1] and Yanfeng Lu[2⋆]

[1] School of Artificial Intelligence, Nanjing University, Nanjing, Jiangsu 210093, China
211300067@smail.nju.edu.cn
[2] Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Science (CASIA), Beijing 100190, China
yanfeng.lv@ia.ac.cn

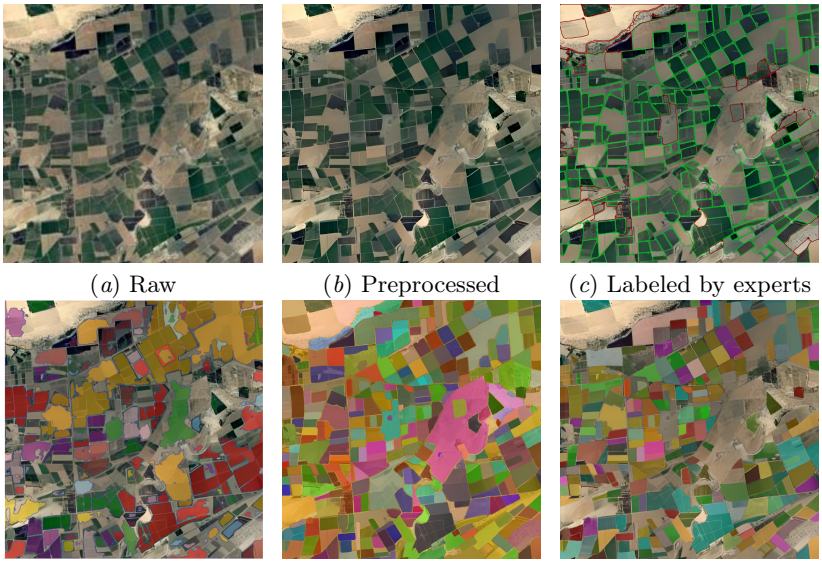
Abstract. The precise segmentation of agricultural remote sensing images is pivotal for the effective monitoring and management of cultivated land resources. Traditional approaches often fall short in accurately delineating agricultural areas, hampered by complex surface features and the presence of non-arable elements such as buildings, roads, and wastelands. Addressing these issues, our study introduces a novel method for the unsupervised segmentation of agricultural remote sensing images through iterative application of the Segment Anything Model*(SAM)*. Additionally, we significantly enhanced a classification model improved from ResNet50 by integrating two attention mechanisms. By integrating the SAM with classification model, we innovatively tackle the fundamental challenge of SAM—its segmentation results lack labels—thereby providing an inspiring solution for the entire field of image segmentation, including remote sensing imagery. This enhancement also enables precise instance segmentation for specific category, effectively retaining instances of cultivated land, which are more valuable for further agricultural analysis. Experiments demonstrate the high-efficiency of our proposed LiSegAgr, which has achieved state-of-the-art(SOTA) performance in the domain of agricultural remote sensing image segmentation. Compared to the classic instance segmentation algorithm Mask R-CNN, our method enhances the IoU metric by 17.2%. The code of our method is available on GitHub: https://github.com/WangYunKa/ISAgrSC2.

Keywords: SAM · Instance Segmentation · Remote Sensing Images.

1 Introduction

In the contemporary era, characterized by rapid population growth and accelerated urbanization, the efficient management and utilization of agricultural land resources have emerged as critical concerns[1]. Agricultural remote sensing analysis technology, as an advanced method of monitoring, offers innovative perspectives and tools for the assessment, classification, and management of arable land

⋆ Corresponding author



(*a*) Raw (*b*) Preprocessed (*c*) Labeled by experts

(*d*) By HQ-ISNet (*e*) By AgriSeIS (*f*) By LiSegAgr

Fig. 1. This is a complex agricultural remote sensing image with high richness. Here we present the raw image alongside its preprocessed effect and the results of expert annotation. Compared with previous best performing method HQ-ISNet[23](d), our segmentation method AgriSeIS(e) can tackle the challenges posed by such complex images and successfully segment all plots. (f) is the instance segmentation result of only cultivated-labeled lands after the classification phase.

resources[2,3]. This technology captures comprehensive surface information, enabling swift and precise analysis of agricultural resources on a macro scale. Such analyses are crucial for precision agriculture management, crop monitoring, yield estimation, and related applications.

The analysis and interpretation of agricultural remote sensing images necessitate highly precise image segmentation technologies to accurately extract information on cultivated land amidst complex surface features[4]. However, the segmentation of agricultural remote sensing images is fraught with challenges. The complex and variable nature of land surfaces complicates the delineation of boundaries between cultivated and non-cultivated lands. Environmental factors such as vegetation cover, soil types, and water distribution can influence remote sensing images and impact segmentation accuracy. Additionally, the presence of non-cultivated elements introduces further complexity. Man-made structures like buildings and roads, interspersed within farmlands, along with natural features such as mountains and rivers, may display characteristics similar to those of cultivated areas in remote sensing images, complicating accurate segmentation. Moreover, common issues in agricultural remote sensing datasets include low resolution, which directly impairs the performance of segmentation models, as highlighted by[15]. Furthermore, remote sensing images captured at higher angles often cover an extensive number of sub-regions, making comprehensive area identification exceedingly challenging.

Consequently, the development of advanced image segmentation algorithms that can effectively address these challenges, improving the accuracy and completeness of agricultural resource monitoring and management has become a crucial area of research. The new strategy should ensure precise instance segmentation of agricultural remote sensing images and facilitate the identification of cultivated land for comprehensive agricultural science analysis based on the segmentation results.

In the field of agriculture remote sensing segmentation, previous methods predominantly employ supervised learning paradigms, necessitating substantial volumes of annotated data. They can be roughly categorized into three distinct groups: models that solely conduct semantic segmentation of cultivated land areas without delineating the boundaries of each land instance[5,6,7]; approaches that utilize object detection techniques to identify remote sensing images and demarcate the approximate boundaries of plots[8,9,17]; and strategies that integrate with traditional supervised segmentation models, such as Mask R-CNN, for supervised instance segmentation[10,11,12,23], as illustrated in Fig.1(d).

The majority of methods mentioned above rely on supervised model training, leading to a significant dependency on labeled data. It presents a challenge in agricultural remote sensing domain, where images often comprise numerous plots with diverse shapes, complicating the task of precise plot labeling, as shown in Fig.1(c). Therefore, unsupervised segmentation approaches will be our primary focus. The Segment Anything Model (SAM)[14] marks a significant breakthrough in image segmentation technology, offering the ability to segment relevant regions without relying on pre-annotated data. Despite SAM's potential in various applications, findings from[15] and our own empirical research indicate its limitations in discerning dense plot structures and achieving complete image recognition.

To maximize the efficacy of the SAM in this tasks, we develop a segmentation method named AgriSeIS *(AGRicultural remote sensing image SEgmentation based on Iterative Sam)*. This method segments images using a stepwise iterative application of SAM, enhanced with auxiliary techniques such as the canny edge detection algorithm and color entropy value analysis to improve the model's segmentation capability. AgriSeIS facilitates comprehensive instance segmentation of significant agricultural plots within an unsupervised framework. The iterative application of SAM addresses its limitations in achieving complete segmentation of entire images. This enhancement significantly improves the model's generalization performance, enabling its application beyond agricultural remote sensing to other fields. The segmentation results are illustrated in the Fig.1(e), with additional outcomes detailed in the supplementary material.

Our complete framework LiSegAgr initiates with the application of Real-ESRGAN [16] super-resolution technology to increase resolution of typically low-resolution agricultural remote sensing images. Subsequent steps involve iterative segmentation by SAM, with some auxiliary methods including Gaussian blur and impurity filtration. Then further refined segmentation of partially segmented areas ensures maximal segmentation performance. Following segmentation, the extracted instance plots are utilized for improving and train classification models

to label field instance into cultivated and non-cultivated lands. We enhance the ResNet50 model by incorporating two attention mechanisms that focus on color richness and hue, achieving a classification accuracy of 96.21% on the validation set. By integrating segmentation and classification models, we successfully achieve labeled instance segmentation in remote sensing images, characterized by high levels of completeness and accuracy. Considering the practical value and labeling accuracy, we propose categorizing the label information into two categories: cultivated land and non-cultivated land. The main goal is to identify cultivated land that are more meaningful for subsequent agricultural research. Experiment verifies the superior segmentation capabilities of our approach, particularly notable in scenarios featuring a high number of plots and extensive interference areas, where it demonstrably outperforms existing methods.

Our contributions are delineated as follows:

- **SOTA Performance of SAM in Instance Segmentation of Agricultural Remote Sensing Images**: We introduce an unsupervised segmentation method for agricultural remote sensing images utilizing SAM, achieving SOTA performance in both completeness and accuracy.
- **Creative implementation of iterative SAM**: We innovatively apply iterative use of the SAM in AgriSeIS, which substantially enhances the segmentation integrity of SAM. This algorithm will have broad applicability across various fields.
- **Integrating SAM with Classification Models for Labeled Instance Segmentation**: To address the limitation of label-free instance segmentation results from SAM, we innovatively combined it with optimized classification model, enabling SAM to produce labeled segmentation outputs.

2 Related Work

Agricultural remote sensing image segmentation is a key technology for achieving precision agriculture management and enhancing crop yield. Current segmentation algorithms, as detailed in [5,6,7], predominantly concentrate on semantic segmentation. This method lacks precise location and shape information for each plot, complicating the execution of targeted agricultural analysis. Secondly, numerous algorithms, as outlined in [8,9,17], are primarily suited for delineating the general boundaries of plots with the help of target detection algorithm, which also offers limited utility for in-depth agricultural analysis. Additionally, the studies in[10,11,12,23] attempt to guide agricultural production using instance segmentation of agricultural plots. These methods represent the most precise approach for analyzing agricultural images, but the drawback is that most of the current agricultural remote sensing image instance segmentation methods rely on supervised methods. Labeling samples for instance segmentation incurs the highest labor cost, consequently resulting in operational inefficiencies. Compared with these methods, our proposed unsupervised algorithm AgriSeIS can segment all kinds of lands including arable land from complex remote sensing images and does not require labeled data for training.

Segment anything model (SAM), developed by Meta AI, is an advanced large-scale image segmentation model that represents a significant advancement in the field. SAM's architecture includes an encoder, prompt encoder, and mask decoder, allowing for high-quality segmentation from diverse prompts without specific adjustments. SAM's zero-shot capability is particularly remarkable, which surpasses traditional supervised methods in segmenting a wide range of objects. It is noteworthy that extant literature[15] has verified the feasibility of SAM for agricultural remote sensing image segmentation, although they only tested the "Anything" mode of SAM through web page demonstrations. We observe similar experimental phenomena on the agricultural remote sensing image data set, but experiments shows that the effect of directly using SAM for segmentation is not ideal.

3 Method

The LiSegAgr we proposed is specifically designed for the task of land labeled instance segmentation within agricultural remote sensing imagery. This algorithm operates in two distinct phases:

Firstly, it undertakes the comprehensive segmentation of individual plot instances across the entirety of the image. The core purpose is dedicated to the thorough segmentation of all areas within the image deemed significant for subsequent analysis, which we refer to as "Valid Segmentation Instance"*(VSI)*. In order to achieve the goal mentioned above, we are supposed to achieve complete segmentation to facilitate subsequent classification efforts, ensuring the extraction of fully classifiable information. Specifically, the segmentation algorithm is designed to support a wide range of tasks, including buildings, urban green spaces, water bodies etc. The methodology unfolds through three defined steps to address challenges presented by agricultural remote sensing images, thereby improving the segmentation and subsequent classification outcomes.

1. Preprocess images through super-resolution and other technologies, obtaining images with same size and enhanced color information such as contrast and saturation (Sec. 3.1);
2. Utilize SAM to segment the preprocessed image iteratively. Auxiliary techniques, such as Gaussian blur, are integrated to facilitate comprehensive segmentation, ensuring that all original images are ultimately identified as masks (Sec. 3.2);
3. Canny edge detection and color entropy value are utilized to assess the completeness of segmentation for each mask layer. Areas recognized as incompletely segmented are subjected to refined segmentation procedures to achieve the highest degree of segmentation completeness (Sec. 3.3).

The second phase of our study utilizes the segmentation results from the initial phase to conduct classification, effectively getting label information of cultivated lands, which are of greater value for subsequent research. We evaluate the efficacy of many commonly used classification networks in addressing the

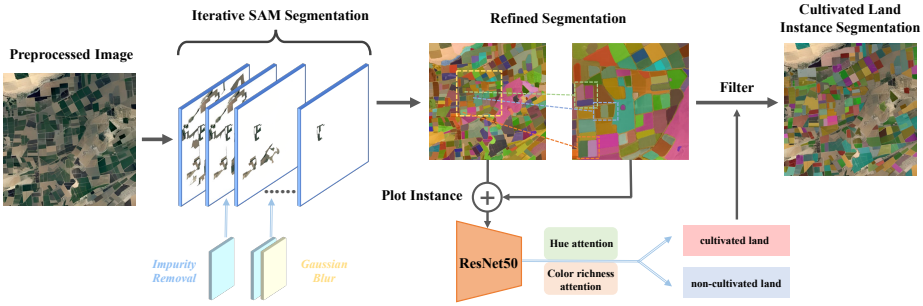


Fig. 2. The pipeline of our LiSegAgr. Through iterative application of SAM and refined segmentation on incompletely segmented regions, we obtain segmentation results for remote sensing images. We then enhance the ResNet50 network through the incorporation of attention mechanisms for classification. By integrating two models, we achieved precise labeled instance segmentation results for agriculture images.

binary classification challenge pertaining to agricultural plots. Drawing upon the outcomes of these experiments, we introduce two attention mechanisms based on hue and color richness into ResNet50 network(ResNet50 demonstrates the best performance without further enhancement and under the same limited training data conditions, so it is selected as the starting point for model improvement), which significantly enhance the classification model's performance(Sec. 3.4). The pipeline of our LiSegAgr is shown in Fig.2.

3.1 Preprocessing of agricultural remote sensing images

Remote sensing images in agriculture typically exhibit lower resolutions, as illustrated in Fig.1(a), which substantially impairs the model's segmentation capability. We employ the Real-ESRGAN super-resolution technology to address this issue, enhancing the resolution of these images by a factor of four. Real-ESRGAN, a deep learning-based image super-resolution method, utilizes generative adversarial networks (GAN)[22] to enhance both the resolution and quality of images. The application of super-resolution notably promotes the completeness segmentation, and won't introduce error or distortion because the boundaries of fields have not changed. Because SAM is highly sensitive to the color and texture characteristics of images, we utilize the Pillow library in Python to increase the image's saturation, contrast, and sharpness for the segmentation integrity.

3.2 Iteratively segment the image through SAM

The AgriSeIS we proposed uses a progressive and iterative approach to segment remote sensing images. In each round of iteration, SAM is first used for segmentation to obtain the masks of the identified multiple plots. Our goal is to obtain *VSI* of the highest possible quality. Therefore, after each round of

(*a*) Original area (*b*) Preliminary Seg. (*c*) Refined Seg.

Fig. 3. Complex areas in the original image, comparison of initial recognition results and refined segmentation results.

recognition, masks with too small areas and high overlap rates are removed to avoid obstructing subsequent segmentation and classification task.

After each round of iteration, we whiten the areas recognized in this round to avoid repeated recognition. As the iteration progresses, we gradually add operations such as Gaussian blur and removal of tiny impurity color patches to promote the integrity of model segmentation, thereby efficiently obtaining complete recognition results. It has been verified through experiments that even for relatively complex remote sensing images, after processing the above operations, complete segmentation results can be obtained in up to four rounds.

3.3 Refined segmentation of incomplete areas

After obtaining the complete segmentation results, we found that the areas corresponding to many masks are collections of multiple lands, which are not our ideal single plot of *VSI*, so further segmentation are needed. We introduce the canny edge detection algorithm and average color entropy to determine whether each area needs further segmentation. We use the canny algorithm to determine whether there are clear edges inside each area, so as to determine whether there may be multiple plots; by judging the average color entropy of each area, we determine the color change range of the corresponding plot, the higher the degree of color confusion, the more likely it is to contain multiple plots. Detailed descriptions of the screening method are provided in the supplementary material. Based on the above two criteria, we screen out the plots that need to be refined and segmented, and use small parameterized lightweight SAM to refine the segmentation, thus obtaining more refined segmentation results. Fig.3 shows our preliminary segmentation effect and the comparison effect of refined segmentation.

3.4 Optimization of classification model

After the segmentation task is completed, we obtain the segmentation results of a large number of plots, including agricultural lands, houses, wastelands, shrubs, etc. Our goal is to get label information of plot instances, so the segmented plots need to be classified. We tested simple convolutional neural networks designed by ourselves, VGG16[19], ResNet50, ResNet101[20], EfficientNet v0-v1[21] and

other networks, but did not get the results we satisfied. The accuracy of the classification model can usually only reach 86% to 91% on the verification set, among which the ResNet50 network achieved the highest accuracy of 91.22%. Such accuracy cannot meet our expectation, so we consider introducing attention mechanism into the currently best-performing ResNet50 network to improve the classification performance.

The main feature that distinguishes cultivated land from others is its color or texture characteristics. Agricultural land is usually closer to a solid color than non-cultivated land, and the hue is mainly green and wheat. We compare different color related values of cultivated lands and non-cultivated lands, including color entropy, color richness, hue, color variance, color uniformity and other indicators. The change curves will be shown in the supplementary materials. We found that in the two graphs of color richness and hue, the curves of cultivated land and non-cultivated land have a tendency to separate, which means that they can be used as the main features to distinguish the two types of lands.

Inspired by the above experimental results, we integrated two attention mechanisms, hue and color richness, into the ResNet50 network, establishing thresholds to regulate the degree of attention. Specifically, the hue attention mechanism converts the input image to the hue component within the HSV color space, applying convolution operations and activation functions to generate attention weights. The color richness attention mechanism quantifies color richness by calculating the number of colors in the image and utilizes a sigmoid function to produce corresponding attention weights. The color richness is defined as follows:

$$R = |\{\text{unique colors}\}| \tag{1}$$

where R is the color richness, $|\cdot|$ denotes the cardinality (size) of the set, and {unique colors} is the set of unique colors in the image.

Similarly, the hue entropy is given by:

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{2}$$

where H is the hue entropy, p_i is the normalized histogram value for the i-th hue value, and n is the total number of histogram bins (usually 256).

These attention weights are incorporated into the feature extraction process of ResNet50, enhancing the model's sensitivity to color and hue information, thereby improving classification performance. After multiple rounds of experiments and adjustments, the finally trained improved ResNet50 network reaches the highest classification accuracy.

4 Experiment

We verify the effectiveness of our proposed method LiSegAgr on the agriculture remote sensing images data set. Both experimental data and visual results demonstrate that LiSegAgr achieves effective labeled instance segmentation.

4.1 Experimental Setups

Datasets. We collect over 2,000 unlabeled agricultural remote sensing images from NWPU-RESISC45, DeepGlobe and USGS datasets, 1000 of them have annotation information of categories and locations. And the annotation results of each image are checked and approved by at least three senior agricultural sciences experts for model performance testing. To enhance the segmentation model's effectiveness, we preprocessed the images prior to utilizing LiSegAgr and all comparative methods, standardizing the image size to 1000×1000 pixels. For those models that require labeled samples for training, we divide all 1000 existing labeled images according to 4:1, that is, 800 images are used for their training, and the remaining 200 images are used for performance evaluation.

Implementation details. Our method utilizes the SAM model and an improved ResNet50 model for segmenting and classifying remote sensing images, incorporating the large-parameter vit-h pre-trained weights for SAM. During the training process, we utilize the enhanced ResNet50 model, employing the cross-entropy loss function and the Stochastic Gradient Descent (SGD) optimizer with a momentum of 0.9. The learning rate is set to 0.0001, complemented by L_2 regularization with a weight decay of 0.1, across a training span of 50 epochs. The training is performed on NVIDIA Tesla V100 GPUs with 16GB of VRAM, with further details on hyperparameters available in the supplementary materials.

Evaluation Metrics.

Segmentation Performance Evaluation: We use IoU and mean IoU (mIoU) to evaluate the performance of different segmentation models, which are commonly used metrics in image segmentation. The IoU utilizes the entire image as the unit of measurement, concentrating on assessing the model's overall accuracy and the *completeness* of segmentation. In contrast, the mIoU emphasizes the accuracy of instance segmentation by evaluating the precision for each segmented instance. (m)IoU evaluates model performance by measuring the degree of overlap between the real areas (denoted as $A_i, i = 1, \cdots, N$) and the predicted areas (denoted as $B_i, i = 1, \cdots, N$). The higher the degree of overlap, the larger the index and the better the performance. The specific definition are as follows:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}, \quad \text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{|A_i \cap B_i|}{|A_i \cup B_i|} \right), \tag{3}$$

Before computing mIoU, we establish the optimal pairing relationship by evaluating the IoU ratios between each instance in the ground truth and the predicted outcomes. It is crucial to acknowledge that the segmentation results from the ground truth and certain methods might not align perfectly on a one-to-one basis, leading to a mismatch in quantities. In such instances, we apply a filtering criterion based on the principle that pairs with the highest IoU ratios

are selected for calculation. When there is a discrepancy in the number of instances between the ground truth and the segmentation results, we compute the maximum number of optimal matches following the aforementioned rules.

To complement our segmentation performance evaluation, we also employ the F1-Score, which is widely used in classification and segmentation tasks to measure a model's accuracy. The F1-Score provides a balance between precision(P) and recall(R), offering a single metric that considers both false positives and false negatives. The F1-Score is defined as follows:

$$P = \frac{|A \cap B|}{|B|}, \quad R = \frac{|A \cap B|}{|A|} \tag{4}$$

$$\text{F1-Score} = \frac{2 \cdot P \cdot R}{P + R} = \frac{2 \cdot |A \cap B|}{|A| + |B|} \tag{5}$$

Here, A denotes the set of actual positive instances, and B denotes the set of predicted positive instances. The intersection $|A \cap B|$ represents the number of correctly predicted positive instances.

Classification Performance Evaluation: To evaluate the performance of the classification model, we employ commonly used evaluation metric accuracy for a comprehensive assessment. The formal definition is given as follows.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{6}$$

Comparison Methods.

Segmentation Comparison Methods: Segmentation comparison methods can be classified into two primary categories. The first category encompasses supervised segmentation methods, which are implemented using frameworks such as Mask R-CNN[13] and YOLACT[26]. Traditional agricultural remote sensing image segmentation techniques, including HQ-ISNet[23] and CATNet[12]. These supervised methods above utilize the model initially trained on the COCO dataset as a foundation, subsequently fine-tuning it with our annotated dataset. The second category comprises unsupervised instance segmentation methods, such as U2Seg[24] and CutLER[25]. Same as SAM, unsupervised methods are applied on the preprocessed images.

Classification Comparison Methods: The task of distinguishing between cultivated and non-cultivated land represents a typical binary image classification problem. We employ widely recognized supervised image classification architectures for comparing the effect of our improved ResNet50 model, including VGG16[19], the standard ResNet50, ResNet101[20], and EfficientNet networks ranging from v0 to v1[21].

4.2 Performance Evaluation

We evaluate the performance of LiSegAgr across three distinct stages: the full scene instance segmentation phase, the plot classification phase, and the class-specific (cultivated land) instance segmentation phase.

Table 1. IoU, mIoU and F1-Score of different methods on our collected dataset. All values are expressed as percentages, the best results are in bold.

Method	HQ-ISNet	CATNet	YOLACT	Mask R-CNN	U2Seg	CutLER	AgriSeIS
IoU	83.11	84.19	78.52	79.03	37.63	43.62	**90.82**
mIoU	81.43	81.89	71.22	71.39	24.52	29.33	**86.39**
F1-Score	90.78	91.41	87.96	88.30	54.70	60.74	**95.19**

Table 2. Accuracy of different methods. All values are expressed as percentages.

Method	VGG16	ResNet50	ResNet101	EfficientNet-v0	EfficientNet-v1	LiSegAgr
Accuracy	88.57	91.22	88.89	90.52	90.79	**96.21**

Full scene instance segmentation. In this phase, our primary focus is on evaluating the full-scene instance segmentation performance of AgriSeIS and comparison methods on agricultural remote sensing images, utilizing the IoU and mIoU metric for assessment as previously mentioned. The experimental outcomes are presented in Table 1. All the comparing methods and our AgriSeIS are performed on pre-processed images. Agricultural remote sensing images frequently encompass a vast number of plot instances, characterized by irregular shapes and indistinct background boundaries. These features present significant challenges for supervised segmentation methods like Mask R-CNN[13], YOLACT[26] and HQ-ISNet[23], particularly in achieving high segmentation accuracy and completeness with limited training data. Furthermore, these characteristics significantly complicate the process of data annotation, thereby hindering the enhancement of supervised models that depend on extensive labeled datasets for training. On the other hand, unsupervised instance segmentation methods, such as U2Seg[24] and CutLER[25], often rely on texture and structural patterns in images. Nevertheless, agricultural remote sensing images typically exhibit low contrast, weak texture features, and frequent small-block impurity interference, which poses a challenge for most unsupervised instance segmentation algorithms to achieve complete and high-quality masks in a single attempt. Our AgriSeIS attains high accuracy by rigorously testing and selecting the optimal unsupervised instance segmentation model, SAM, as a foundational starting point. Additionally, it addresses the challenge of incomplete segmentation and secures the highest level of segmentation completeness through the application of iterative approaches and the strategic subdivision of informative plots. This capability These enhancements and optimizations result in superior performance.

Plot classification. During the assessment of classification performance, we compare the efficacy of comparison classification models against the performance of the LiSegAgr's classification model. By choosing the ResNet50 network, known for its robust initial performance, as our base model and integrating a well-designed attention mechanism, we achieve a significant improvement in

Table 3. IoU, mIoU and F1-Score of fine-tuned comparison method and our LiSegAgr. All values are expressed as percentages, and the best results are in bold.

Method	U-Net	DeepLabV3$^+$	YOLACT	Mask RCNN	HQ-ISNet	CATNet	LiSegAgr
IoU	75.92	70.11	64.30	71.51	78.14	72.97	**88.73**
mIoU	-	-	54.72	62.13	73.43	68.78	**81.54**
F1-Score	86.32	82.39	78.31	83.41	87.71	84.39	**94.04**

classification performance over the comparison methods. The results of this evaluation are presented in Table 2.

Label segmentation results by introducing classification model. By integrating the full-scene instance segmentation model with the classification model, we develop a labeled instance segmentation model.

We fine-tune the comparison methods using training data that retained only the labeling results for cultivated land and train two semantic segmentation models, U-Net[27] and DeepLab V3$^+$[28], for comparison with our LiSegAgr(Given that semantic segmentation models do not distinguish between multiple instances, we only employ IoU and F1-Score as the evaluation metric). The experimental outcomes detailed in Table 3. Because of indistinct feature distinction between cultivated and non-cultivated land, achieving precise segmentation of cultivated land instances through fine-tuning alone presents challenges. Comparative approaches typically suffer from inaccuracies such as the erroneous segmentation of non-cultivated land and omission of cultivated land instances, leading to diminished evaluation value. LiSegAgr, by partitioning the process into two distinct phases, though more complex, achieves higher accuracy and reliability. This method plays a more significant role in enhancing the support and advancement of subsequent agricultural research.

4.3 Ablation Study

We conduct ablation studies to evaluate the influence of critical processes, such as preprocessing, refined segmentation, and attention mechanisms, on the performance of LiSegAgr. During the segmentation phase, our focus is on examining the effects of preprocessing steps including super-resolution, refined segmentation, impurity removal, and Gaussian blur. The results of these studies are presented in Table 4. Preprocessing plays a crucial role in AgriSeIS, particularly super-resolution, which significantly enhances the accuracy of each SAM recognition cycle. Omitting this preprocessing step results in markedly poor outcomes. We observe that SAM's performance is highly sensitive to the dimensions of the input image and the proportion of the area targeted for recognition. Consequently, the refined segmentation step effectively identifies numerous regions that are challenging for the model to detect in a single iteration, thereby making a substantial contribution to the model's overall performance. The primary

Table 4. Ablation study on different parts of AgriSeIS, including preprocessing, refined segmentation, impurity removal, and Gaussian blur. Omitting Gaussian blur may lead to slight increases in the mIoU due to edge deformation in some plots. However, employing Gaussian blur significantly enhances the completeness of model recognition and substantially improves the IoU metric.

Super-resolution	Refined Seg.	Impurity Removal	Gaussian Blur	IoU	mIoU
✗	✓	✓	✓	84.55	76.73
✓	✗	✓	✓	**90.82**	84.06
✓	✓	✗	✓	89.31	84.15
✓	✓	✓	✗	88.27	**86.76**
✓	✓	✓	✓	**90.82**	86.39

Table 5. Ablation study on the influence of attention mechanism.

Method	Only Hue Attention	Only Color Richness Attention	LiSegAgr
Accuracy	94.89	93.49	**96.21**

purpose of impurity removal and Gaussian blur is to facilitate the complete segmentation through the iterative process. These steps are essential for accurately identifying and segmenting certain "stubborn plots", proving to be indispensable for the success of AgriSeIS.

Concentrating on the classification model, we individually assessed the enhancement effect of two attention mechanisms on the model's classification performance, with the results detailed in Table 5. Notably, the hue attention mechanism demonstrates a superior enhancement effect on classification due to its efficacy in delineating characteristic boundaries between cultivated and non-cultivated land. The synergistic application of both attention mechanisms markedly enhances the model's performance. Additional ablation study, focusing on the impact of iteration rounds, is detailed in the supplementary material.

5 Conclusion

We develop an unsupervised method, for the automated segmentation of agricultural remote sensing images. This method particularly tackles the challenge of comprehensively segmenting fields by iteratively utilizing the SAM model. Furthermore, we enhance classification model through attention mechanisms to classify between cultivated and non-cultivated lands. By integrating these two phases, LiSegAgr accomplishes detailed and precise labeled instance segmentation of agricultural remote sensing images. Our experimental findings highlight the superior segmentation capabilities of LiSegAgr over traditional classification and segmentation methods. The methodology delivers valuable perspectives on the unsupervised segmentation of complex images, encompassing a multitude of elements. This contribution has significant implications across various domains, including medical image segmentation and autonomous driving scene

segmentation, offering inspirational insights. Nevertheless, LiSegAgr still encounters difficulties in processing images with plots that are too small or exhibit irregular and complex shapes, such as variably planted plots with different crops at intervals. Consequently, future work must focus on targeted optimization for these scenarios to enhance the robustness of our method. And we will also aim to leverage the segmentation outcomes for the multi-classification of various field categories, thus getting more label information to aid agricultural scientists in the more effective utilization of remote sensing images for land management and agricultural analysis.

Acknowledgements. This work is supported by Beijing Natural Science Foundation (Grant L211023), the Strategic Priority Research Program of the Chinese Academy of Sciences under (Grants XDA0450200, XDA0450202) and National Natural Science Foundation of China (Grants 91948303, 61627808).

References

1. Johnson, D.M.: A 2010 Map Estimate of Annually Tilled Cropland within the Conterminous United States. Agricultural Systems **114**, 95–105 (2013)
2. Carfagna, E., Gallego, F.J.: Using Remote Sensing for Agricultural Statistics. International Statistical Review **73**(3), 389–404 (2005)
3. Graesser, J., Ramankutty, N.: Detection of Cropland Field Parcels from Landsat Imagery. Remote Sensing of Environment **201**, 165–180 (2017)
4. Rudel, T.K., Schneider, L., Uriarte, M., Turner, B.L., DeFries, R., Lawrence, D., Geoghegan, J., Hecht, S., Ickowitz, A., Lambin, E.F., et al.: Agricultural Intensification and Changes in Cultivated Areas, 1970–2005. Proceedings of the National Academy of Sciences **106**(49), 20675–20680 (2009)
5. Zhang, J., Yang, X., Jiang, R., Shao, W., Zhang, L.: RSAM-Seg: A SAM-based Approach with Prior Knowledge Integration for Remote Sensing Image Semantic Segmentation. CoRR **abs/2402.19004** (2024).
6. Qi, L., Zuo, D., Wang, Y., Tao, Y., Tang, R., Shi, J., Gong, J., Li, B.: Convolutional Neural Network-Based Method for Agriculture Plot Segmentation in Remote Sensing Images. Remote Sensing **16**(2), 346 (2024)
7. Li, X., Li, Y., Ai, J., Shu, Z., Xia, J., Xia, Y.: Semantic Segmentation of UAV Remote Sensing Images Based on Edge Feature Fusing and Multi-level Upsampling Integrated with Deeplabv3+. PloS ONE **18**(1), e0279097 (2023)
8. Zheng, Z., Lei, L., Sun, H., Kuang, G.: A Review of Remote Sensing Image Object Detection Algorithms Based on Deep Learning. In: 2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC), pp. 34–43 (2020)
9. Gui, S., Song, S., Qin, R., Tang, Y.: Remote Sensing Object Detection in the Deep Learning Era - A Review. Remote. Sens. **16**(2), 327 (2024).
10. Zhong, B., Wei, T., Luo, X., Du, B., Hu, L., Ao, K., Yang, A., Wu, J.: Multi-Swin Mask Transformer for Instance Segmentation of Agricultural Field Extraction. Remote. Sens. **15**(3), 549 (2023).
11. Yang, F., Yuan, X., Ran, J., Shu, W., Zhao, Y., Qin, A., Gao, C.: Accurate Instance Segmentation for Remote Sensing Images via Adaptive and Dynamic Feature Learning. Remote. Sens. **13**(23), 4774 (2021).

12. Liu, Y., Li, H., Hu, C., Luo, S., Luo, Y., Chen, C.W.: Learning to Aggregate Multi-Scale Context for Instance Segmentation in Remote Sensing Images. IEEE Transactions on Neural Networks and Learning Systems (2024)
13. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)
14. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., et al.: Segment Anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026 (2023)
15. Gui, B., Bhardwaj, A., Sam, L.: Evaluating the Efficacy of Segment Anything Model for Delineating Agriculture and Urban Green Spaces in Multiresolution Aerial and Spaceborne Remote Sensing Images. Remote. Sens. **16**(2), 414 (2024).
16. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1905–1914 (2021)
17. Su, H., Wei, S., Yan, M., Wang, C., Shi, J., Zhang, X.: Object Detection and Instance Segmentation in Remote Sensing Imagery Based on Precise Mask R-CNN. In: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 1454-1457 (2019)
18. Teixeira, I., Morais, R., Sousa, J.J., Cunha, A.: Deep Learning Models for the Classification of Crops in Aerial Imagery: A Review. Agriculture **13**(5), 965 (2023)
19. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May, 2015, Conference Track Proceedings (2015).
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
21. Tan, M., Le, Q.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: International Conference on Machine Learning, pp. 6105–6114 (2019)
22. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Nets. Advances in Neural Information Processing Systems **27** (2014)
23. Su, H., Wei, S., Liu, S., Liang, J., Wang, C., Shi, J., Zhang, X.: HQ-ISNet: High-Quality Instance Segmentation for Remote Sensing Imagery. Remote Sensing **12**(6), 989 (2020)
24. Niu, D., Wang, X., Han, X., Lian, L., Herzig, R., Darrell, T.: Unsupervised Universal Image Segmentation. arXiv preprint arXiv:2312.17243 (2023)
25. Wang, X., Girdhar, R., Yu, S.X., Misra, I.: Cut and Learn for Unsupervised Object Detection and Instance Segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3124–3134 (2023)
26. Bolya, D., Zhou, C., Xiao, F., Lee, Y.J.: YOLACT: Real-Time Instance Segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9157–9166 (2019)
27. Ronneberger, O., Fischer, P., Brox, T.: 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, Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer (2015).
28. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)