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

Abstract. There are five parts in this supplementary:
- Hyperparameter settings of LiSegAgr and detailed dataset introduction.
- Supplementary ablation study that investigates the influence of iteration rounds.
- Details of the implementation involve employing Canny edge detection and color entropy to identify areas of inadequate segmentation.
- Experimental details of attention mechanism selection and setting.
- Additional results demonstrate the effectiveness of our proposed method LiSegAgr.

1 Hyperparameter settings and dataset of LiSegAgr

Table 1. Hyperparameter settings of LiSegAgr.

Hyperparameter	Value
ratio *(Contrast)*	1.2
ratio *(Saturation)*	1.1
ratio *(Sharpen)*	1.1
ratio *(Super-resolution)*	4
model-type *(SAM)*	vit-h
momentum *(SGD)*	0.9
learning-rate	0.0001
weight-decay *(L_2 regularization)*	0.1
num-epochs	50

1.1 Hyperparameter settings

Within LiSegAgr, procedures such as preprocessing, SAM segmentation, and binary classification are incorporated. In the image preprocessing phase, We enhance the image's contrast, saturation, and sharpness by factors of 1.2, 1.1, and 1.1, respectively. Furthermore, we apply Real-ESRGAN super-resolution technology to achieve a 4 times increase in the image's resolution. This enhancement

not only improves the image resolution but also preserves the operational efficiency of the model without negative impacts. SAM employs various pre-training weight parameters (model-types). For performance optimization, we choose the pre-training parameter vit-h, which possesses the highest number of parameters, to guarantee the comprehensiveness and precision of SAM's recognition capabilities. Throughout the training phase, we employed an enhanced ResNet50 model, utilizing the cross-entropy loss function and the Stochastic Gradient Descent (SGD) optimizer, which includes a momentum of 0.9. The learning rate was established at 0.0001, augmented by L_2 regularization featuring a weight decay of 0.1, over a duration of 50 epochs. Detailed settings for these hyperparameters are presented in Table 1.

1.2 Dataset introduction

We collect over 3,000 unlabeled agricultural remote sensing images from the NWPU-RESISC45, DeepGlobe, BigEarthNet, AID and USGS datasets, of which 1500 contain annotations on categories and locations. Each annotated image was reviewed and validated by at least three senior experts in agricultural sciences to ensure reliability for model performance testing.

To optimize the effectiveness of the segmentation model, we preprocess the images before using LiSegAgr and all comparative methods, standardizing the image dimensions to 1000×1000 pixels. For models requiring labeled samples for training, we allocate 1200 of the 1500 labeled images for training purposes and reserved 300 for performance evaluation, maintaining a 4:1 training to testing ratio.

Each plot segmented by AgriSeIS is converted into a regular image framed by its circumscribed rectangle and uniformly resized to 224×224 pixels for classification model training. This phase of annotation, unlike the initial manual expert process, capitalized on the segmentation outcomes. It involved solely selecting and labeling effective segmentation areas, thus greatly enhancing the efficiency of the annotation process. By merging these annotations with previously annotated plots, we assembled a dataset of 10,000 images featuring both cultivated and non-cultivated land. Of these, 9,000 were designated for training and testing, while the remaining 1,000 formed the validation set.

2 Supplementary ablation study

We perform additional ablation study to investigate the impact of iteration rounds on model segmentation performance, as illustrated in Fig.1. The data reveal that, initially, both IoU and mIoU experience significant improvements with an increase in iteration rounds, stabilizing after approximately four rounds. This indicates that under normal circumstances, the segmentation results of most areas can be obtained after about four rounds of iteration.

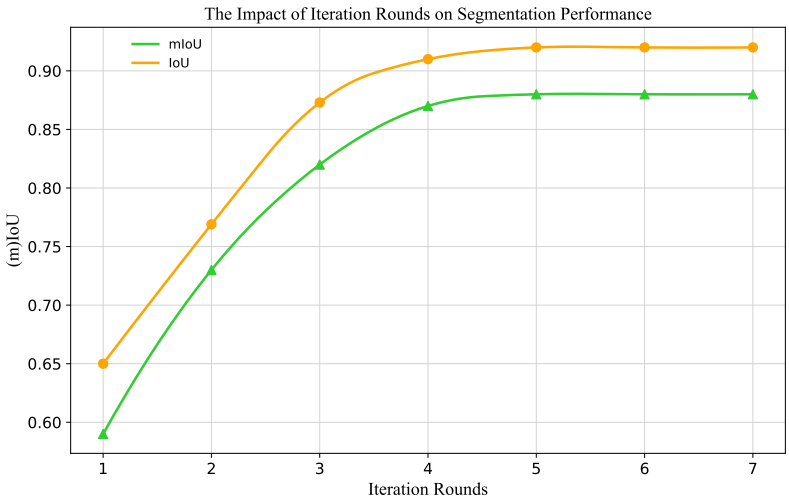


Fig. 1. The impact of angle threshold and distance threshold on LiSegAgr.

3 Canny edge detection and color entropy

As described in the main paper, upon completing the iterative segmentation process, we employ Canny edge detection and color entropy to identify areas requiring refined segmentation. Here we show the effect and implementing of these two discrimination methods. Canny edge detection is a multi-stage algorithm introduced by John F. Canny in 1986 for accurate and efficient edge detection in images. Edge detection on each previously segmented area is performed using the Canny method, with results depicted in Fig.2(b). Following this, we delineate the edge for the entire area as shown in Fig.2(c), eliminate the peripheral edges, and assess the quantity and ratio of edge pigment points within the area to ascertain the necessity for refined segmentation. Typically, an area requiring no further division contains few or no internal edge lines. Consequently, the presence of numerous internal edge lines suggests undivided plots, indicating the need for additional segmentation.

In addition to the Canny method, we employ color entropy as an auxiliary criterion. Fig.3 illustrates that, unlike fully segmented areas which often exhibit pure colors, regions with unsegmented plots typically encompass a wider array of colors, resulting in higher color entropy. We quantified the color entropy across the three RGB channels and computed their average. Utilizing this metric in conjunction with the Canny criterion, we identified areas necessitating refined segmentation.

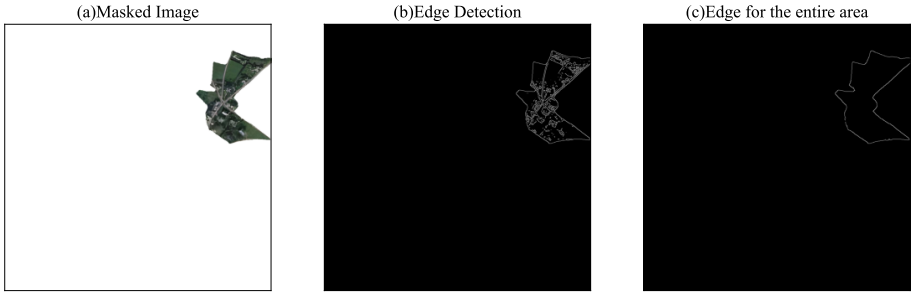
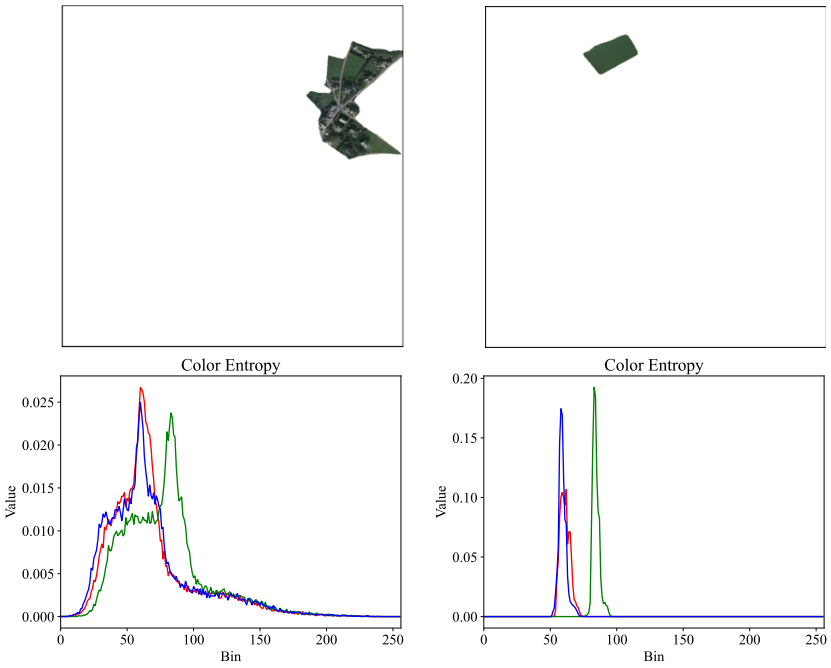


Fig. 2. Demonstration of the principle of Canny edge detection.

4 Selection and setting of attention mechanism

The primary distinctions between cultivated and non-cultivated land lie in their color, texture, and shape characteristics. However, in the design of attention mechanisms, shape and texture information are often not priority attributes because it is difficult to find general quantitative analysis values that can be prioritized by neural networks. Consequently, color features are given precedence. Nevertheless, it is premature to draw conclusions, as color features encompass a broad spectrum of quantitative analysis aspects, including color richness, color entropy, hue, and saturation. To identify which attributes are most appropriate for attention mechanisms, we randomly selected a sample of plots from both cultivated and non-cultivated land, calculated their distinctive characteristics, and illustrated the variation trends in the chart depicted in Fig.4. The results indicated clearer separation between the curves for color richness and hue, suggesting significant differences between the two types of plots in these aspects. Therefore, these characteristics are more apt for integration into attention mechanisms. Subsequent experiments of network training also confirmed our point of view.

The attention mechanisms were implemented within the ResNet50 architecture by introducing the Hue Attention Module and Color Richness Attention Module at strategic points in the network. The Hue Attention Module operates on the hue channel, which is extracted from the input images after converting them to the HSV color space. It uses a series of 1x1 convolution layers to generate attention weights, which are applied to the feature maps. Similarly, the Color Richness Module computes the richness of colors by analyzing the distribution of unique colors in the input. Both modules produce global feature vectors through pooling operations. These attention features are then concatenated with the deep visual features extracted from the ResNet50 network, right after the final convolutional layer and before the fully connected classification layer. This fusion is controlled through predefined scaling factors, which adjust the contribution of each attention mechanism to the overall model. By concatenating these features along the feature dimension, the model leverages both spatial and



(*a*) Area requires refined seg. (*b*) Area do not require refined seg.

Fig. 3. Color entropy analysis of different plots, (a)depicts an area necessitating refined segmentation, characterized by a histogram with extensive lateral spread and high color entropy(6.979). Conversely, (b)illustrates a plot that does not require further segmentation, where the color approximates a pure hue, the color entropy is minimal(2.347), and the histogram exhibits a narrow spread with pronounced peaks.

color-based information to enhance its classification performance, particularly for distinguishing between cultivated and non-cultivated land.

5 Additional segmentation results

In this section, we proide more experimental results of our LiSegAgr, as shown in Fig.5. We display the original image, followed by its preprocessing effect, the full segmentation effect of AgriSeIS, and the instance segmentation result for cultivated land. They all demonstrate high segmentation accuracy and completeness, maintaining comparable performance across complex images.

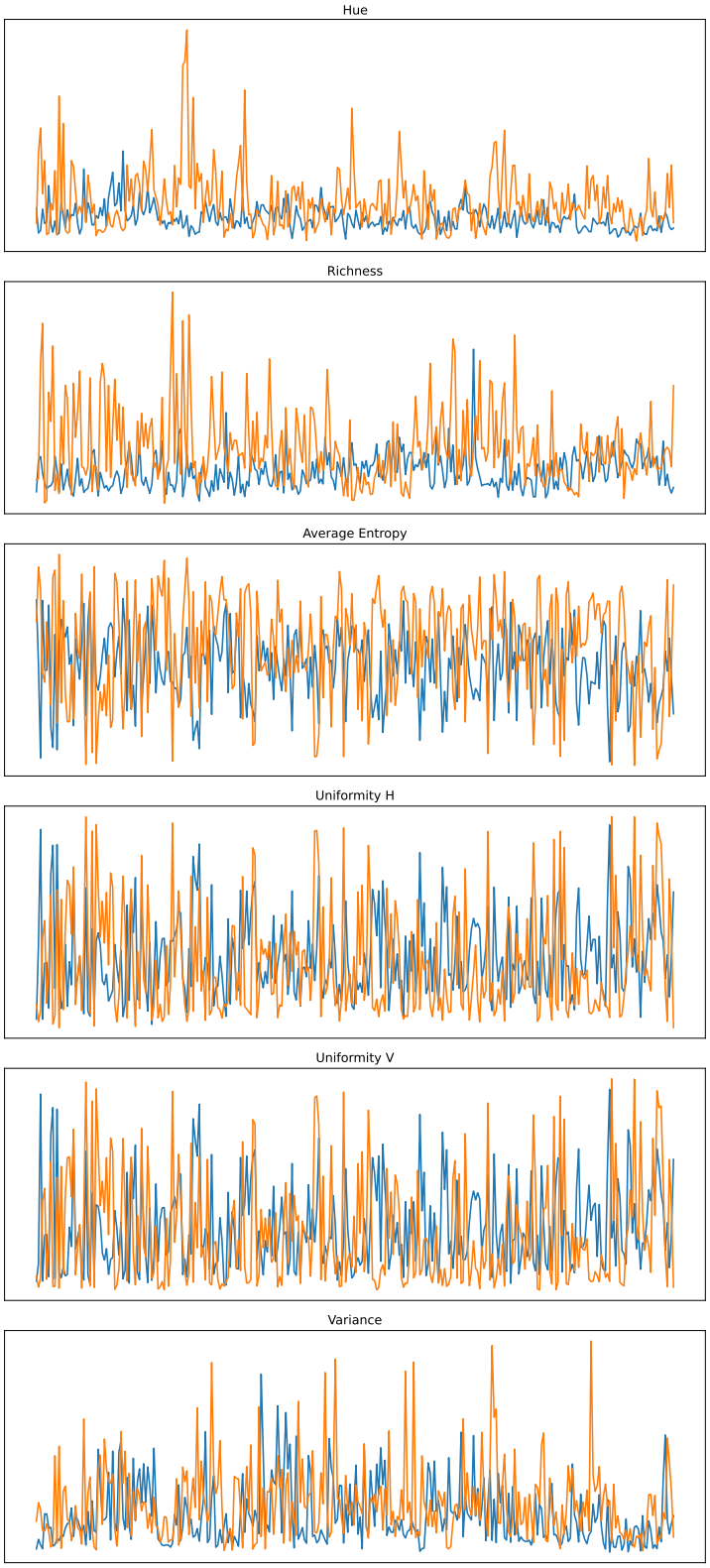


Fig. 4. Change curves of various color features including hue, color richness, average color entropy, hue uniformity, value(brightness) uniformity and color variance on cultivated land and non-cultivated land. We can see that in the first two graphs of color richness and hue, the two curves show a clearer trend of separation.

(*a*) Raw (*b*) Preprocessed (*c*) By AgriSeIS (*d*) By LiSegAgr

Fig. 5. Further experimental results demonstrate the effectiveness of our method. We display the raw images, preprocessed images, full scene instance segmentation results(c), and the instance segmentation results for cultivated land, following integration with the classification model(d).