A Joint Network of Super-Resolution and Active Learning for Agricultural Land Classification

Jiachen Yang, *Senior Member, IEEE*, Zechen Wang, Desheng Chen, Shuai Xiao, *Member, IEEE*, and Ahmad Taher Azar

***Abstract*—Remote sensing image analysis plays a vital role in achieving intelligent agricultural monitoring. However, the acquisition of high-resolution agricultural remote sensing data can be resource-intensive, resulting in an imbalance between training samples and artificial intelligence model parameters. In order to achieve accurate agricultural land recognition of limited-resolution remote sensing images, this article proposes a joint network of super-resolution and active learning (AL). The network introduces a pretrained image super-resolution model and optimizes this for remote sensing classification tasks. It effectively detects detailed features and completes the reconstruction. Based on the reconstructed data, an AL algorithm is proposed with a DBSS. It balances the contributions between interclass and boundary samples. Furthermore, we propose a semisupervisory assistance strategy based on consistency, it fully utilizes the predictive power of deep learning models aiming to reduce labeling costs. This framework is proved effective by experiments on an agricultural remote sensing image dataset, it reduces the cost of agricultural data annotation and improves the efficiency of model learning for low-resolution agricultural remote sensing.**

***Index Terms*—Active learning (AL), land classification, remote sensing, super-resolution (SR).**

I. Introduction

With the global population on the rise and urbanization advancing, issues like environmental concerns, scarcity of land resources, and the availability of arable land are becoming increasingly prominent. This has put the development of agriculture constantly under threat. Achieving high-quality and sustainable agricultural development is therefore, a pressing research issue today. Remote sensing technology is constantly developing, thanks to the construction of various earth observation platforms and the popularity of automated equipment. Remote sensing image classification can quickly and accurately identify environmental changes based on spectral information for agricultural land planning assessment, providing a reliable method for intelligent agricultural monitoring.

The rapid development of artificial intelligence (AI) [1] technology over the past decade has brought about dramatic changes in various fields. Due to the great success of AlexNet [2], experts in various fields started to pay attention to the great significance of deep learning. The subsequent emergence of models such as VGG [3] and ResNet [4] has also made deep learning far superior to humans in image classification. In addition, deep learning has also achieved good results in many tasks such as agricultural detection [5], [6], marine survey [7], face-swap [8], and image segmentation [9], [10].

However, remote sensing images often suffer from low spatial resolution due to the limitations of imaging sensors and transmission capabilities, making high-quality agricultural data difficult to obtain and very costly to annotate. To overcome the challenge of acquiring high-resolution images with sufficient coverage and frequency, remote sensing image super-resolution (SR) technology has emerged as a promising solution to meet the growing demand for detailed and accurate information. Currently, deep learning-based image super-segmentation methods have made significant progress and play an important role in video processing [11], urban planning [12], land cover classification [13], and other fields. In this article, an SR network is introduced to detect detailed features and reconstruct high-quality datasets, thereby, recovering high-frequency details lost in low-resolution datasets and maintaining content consistency.

On the other hand, training SR models still require a large amount of labeled data, which will bring high training costs. In order to make more effective use of high-resolution data, this article introduces the idea of semisupervised active learning (AL) to select the most valuable samples for annotation. AL trains a neural network model with a small number of labeled samples. It uses the model to determine the value of unlabeled samples so that informative samples can be selected and labeled by an observer. The birth of this field has to some extent alleviated the problem of labeling agriculture datasets [14], [15], [16] and has been applied in other fields as well [17], [18], [19]. In addition, semisupervised learning learns the labeled samples so that the unlabeled samples can be labeled autonomously and participate in training. This has the advantage of effectively using the predictive power of the neural network model itself, rather than relying solely on expert labeling, which can further reduce

Manuscript received 8 April 2024; revised 5 June 2024 and 5 July 2024; accepted 21 July 2024. Date of publication 24 July 2024; date of current version 9 August 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62271345 and Grant 62301356, and in part by the Joint Fund of Ministry of Education for Equipment Pre-research under Grant 8091B032254. *(Corresponding author: Desheng Chen.)*

Jiachen Yang, Zechen Wang, Desheng Chen, and Shuai Xiao are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: chendesheng@tju.edu.cn).

Ahmad Taher Azar is with the College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia, also with the Automated Systems and Soft Computing Lab (ASSCL), Prince Sultan University, Riyadh 11586, Saudi Arabia, and also with the Faculty of Computers and Artificial Intelligence, Benha University, Benha 13518, Egypt (e-mail: aazar@psu.edu.sa).

Digital Object Identifier 10.1109/JSTARS.2024.3432976

the cost of labeling. Semisupervised learning has also achieved remarkable success in many fields [20], [21], [22].

Currently, there are numerous AL methods available for evaluating the information content of samples, each with different focuses on sample selection. The excessive focus on screening a single sample can lead to biased sample selection, reducing the efficiency of the dataset. It is important to consider integrating the evaluation scores of different AL algorithms. Therefore, a proposed dual-branch fusion selection method aims to adapt to different needs.

The main contribution points of this article are as follows.

1) A joint network of SR and AL is designed. The pretrained SR can effectively restore the informative details in LR remote sensing images; through AL, reconstructed valuable images are picked out for human annotation.
2) A dual-branch selection strategy (DBSS) is proposed for sample information assessment. It balances the contributions between the interclass selection and boundary selection, which helps the annotation of high-value samples.
3) A self-supervisory assistance strategy (SSAS) optimized for AL is designed to effectively utilize the predictive power of the neural network model itself for samples, thereby, reducing the annotation cost.

The rest of this article is organized as follows. Section II describes the work on remote sensing image classification, image SR and AL. Section III describes the proposed AL method, DBSS and SSAS. Section IV describes the results of the proposed methods. Finally, Section V concludes this article.

II. Related Work

A. Remote Sensing Image Classification

As neural networks continue to evolve [23], [24], remote sensing image classification using a neural network is of huge value. In 2010, Lienou et al. [25] introduced a theme model of the machine learning field to remote sensing image scene classification to realize the marking and classification tasks of high-resolution remote sensing scenes. At the same time, Yang collected and arranged datasets containing 21 remote sensing scenes [26]. This is a high-resolution remote sensing scene dataset for the first time available, widely used by domestic and foreign researchers, and is the most common dataset in this field. Li et al. [27] introduced a pretrained convolutional neural network (CNN) model which is used as a feature extractor. It constructs a feature representation of remote sensing images for classification. However, when classifying remote sensing images based on CNN networks, the accuracy of the model is constrained by insufficient training samples. To address the problem, Marco et al. [28] compared the performance between the traditional training and pre-trained way. It provides a constructive solution for insufficient samples. Wang et al. [29] proposed an end-to-end attention recurrent convolutional network (ARCNet) for scene classification and fine-tunes on the target dataset. The features learned by these models are not exactly suited to the features of the target dataset, and these methods suffer from high labeling costs and unbalanced processing samples.

B. AL

After the rise of deep learning, there has been some research on AL based on deep learning. Gal et al. [30] proposed a method called deep Bayesian active learning (DBAL), which can be used in high dimensional data, an extremely challenging task for conventional AL methods. Zhu et al. [31] proposed an AL method for query synthesis using generative adversarial networks (GAN), a method that adaptively generates instances for querying, thus, speeding up training. Sener et al. [32] defined the AL problem as core-set selection and proposed a core set approach, according to which the most selective (a.k.a. highest accuracy) samples are selected as the core-set. Yoo et al. [18] proposed a learning-loss method to score samples by predicting loss using a loss prediction module.

Since remote sensing datasets inherently have the problem of high annotation costs, many AL algorithms are based on remote sensing applications. Wang st al. [33] addressed semisupervised object detection (SSOD) in remote sensing images characterized by a long-tailed distribution. The experimental results show that the introduction of AL performs well in the remote sensing dataset DOTA-v1.0. Lars et al. [34] validated the effectiveness of the joint use of self-supervised pretraining with AL.

Semisupervised learning refers to pattern recognition using a small amount of labeled data and a large amount of unlabeled data. Commonly used methods include self-training, graph-based semisupervised learning, co-training [35], and label propagation [36]. With the development of deep learning, deep semisupervised learning for deep learning also appeared. Li et al. [37] proposed a semisupervised learning method using pseudolabels by simply using the predicted maximum class to train on unlabeled data, and to prevent network degradation, a growth function was used to control the amount of unlabeled data used by the network while making the iteration speed as fast as possible, and making the obtained pseudolabels as accurate as possible. This method is very simple to implement, but yields good results. Laine et al. [38] introduced a self-integration method that uses network predictions of unlabeled data from different epochs to synthesize consensus predictions to generate pseudolabels. Self-ensemble prediction uses the network to predict unknown labels under different regularization and input enhancement conditions, so the resulting labels will be more consistent and semisupervised. Tarvainen et al. [39] proposed mean teacher, a semisupervised learning method using unlabeled data using average model weights instead of label predictions. This method can effectively improve test accuracy and can be trained with fewer labels than temporal ensembling.

C. Image SR

High-resolution images often contain more detailed visual information. Image SR can help enhance image analysis by addressing the issues of blurriness and missing details in low-resolution images. This technique has wide-ranging applications in target monitoring, video reconstruction, and earth observation. Image SR methods can be divided into interpolation-based methods, reconstruction-based methods and deep learning-based methods. Traditional interpolation-based methods include

nearest neighbor interpolation, bicubic interpolation [40], bilinear interpolation and other methods, which are computationally simple and easy to implement, but are unable to more accurately restore the lost information of an image in complex application scenarios. Reconstruction-based methods aim to construct a priori constraints for high-resolution images using regularisation methods [41], including a priori knowledge of the image's edge texture, local smoothing features, nonnegativity of pixel values, etc., and complete the image SR with the idea of solving an optimization problem. Currently, deep learning-based methods are mainstream. Super-Resolution convolutional neural network (SRCNN) [42] is the first neural network approach to achieve the single image SR by end-to-end mapping. But the quality of reconstruction is insufficient for early methods like this, and the computational complexity is high because the convolutional layer is in HR space. MEN [43] used multiscale features of remote sensing images to enhance the network's reconstruction capability. Besides, VDSR [44], DRRN [45], ESRGAN [46], RFB-ESRGAN [47], and other methods emerged one after another, and these methods can provide effective solution ideas for the reconstruction of low-resolution remote sensing images.

III. Methodology

In this section, we introduce the following three main parts: 1) a joint network of SR and AL that can be excellent on agricultural remote sensing datasets; 2) a DBSS is proposed for sample information assessment, and can balance the contributions between the interclass selection and boundary selection; and 3) an SSAS that optimizes AL and can effectively exploit the predictive power of the neural network model itself for the samples, thus, further reducing the labeling cost.

A. Remote Sensing Images Reconstruction

Since texture details vary greatly from one remote sensing image to another, this article uses a pretrained SR model, RFB-ESRGAN [47], to implement the reconstruction of the low-resolution dataset. In order to make this pretrained model better applicable to our agricultural remote sensing dataset, we introduce an auxiliary training module based on visual geometry group (VGG) [3] for feature extraction and further fine-tuning. The training of the SR model is divided into two stages.

In the first stage, given the raw set S_{Raw} as ground truth, low-resolution images are generated from the raw set as LR dataset S_{LR}.

$$S_{SR^1} = \mathcal{F}_\theta\left(S_{LR}\right). \tag{1}$$

$\mathcal{F}_\theta$ presents the pretrained SR network function. Then, LR dataset is input into a SR framework to restore lost high-frequency detail while maintaining content consistency. We compare the effect between the newly generated dataset S_{SR^1} and the raw set S_{Raw}, and calculate the auxiliary loss function L_{aux} to better fine-tune

$$L_{\text{aux}} = \left\|VGG\left(S_{SR^1}\right), VGG\left(S_{\text{Raw}}\right)\right\|_2 \tag{2}$$

where L_{aux} denotes the manhattan distance between the VGG feature representations of S_{SR^1} and S_{Raw}, and VGG represents the feature map. In the second stage, the raw set is directly input into the fine-tuned model, and the SR model reconstructs an SR dataset S_{SR^2} as the unlabeled samples with higher information. Then, S_{SR^2} will be sent into AL network for selection. Both stages share the same parameters θ of model

$$S_{SR^2} = \mathcal{F}_\theta\left(S_{\text{Raw}}\right). \tag{3}$$

As shown in Fig. 1, the main architecture includes the following five parts: 1) the first convolution module, 2) the Trunk-a module, 3) the Trunk-RFB module, 4) the upsampling module, and 5) the final convolution module. In order to avoid overfitting the model and to reduce the difficulty of fine-tuning, this article adapts part of the structure of the SR network. Trunk-a adopts eight residual of receptive field dense blocks and each block consists of five layers of convolution; Trunk-RFB adopts four residual of receptive field dense blocks and each block consists of five layers of receptive field block (RFB). Through RFB, multiple small kernels replace the large convolutional kernels in the sensory field block. This process extracts detailed multiscale features from remote sensing images while reducing computational complexity.

After one RFB layer used for feature fusion, upsampling operations of nearest neighbor interpolation and sub-pixel convolution are used alternately, and one RFB layer and activation PReLU layer are picked up after each upsampling. Nearest neighbor interpolation scales up the feature map spatially, while sub-pixel convolution further enhances the details by rearranging the feature channels. Finally, one convolution layer is attached. Through SR reconstruction, the original low-quality remote sensing images are transformed into high-resolution datasets with richer detail information and higher sampling values.

B. DBSS

The uncertainty-based AL algorithm utilizes the feature distribution extracted by the network to create a probability distribution. The algorithm then calculates the uncertainty (i.e., entropy) of this probability distribution and uses it as a score for the sample. A higher score indicates a higher value for the sample. In this study, the upper layer of the network's output layer is considered as the feature distribution for the uncertainty calculation. This layer is chosen over the output layer or a shallower layer because it retains more information about the image and is less abstract than the output layer. In addition, using a shallower layer can result in a chaotic distribution, making it difficult to design an effective probability distribution. Therefore, the upper layer of the network's output layer is chosen as the embedding layer for these reasons.

As illustrated in Fig. 1, the primary structure of the AL network utilizes a DBSS. This prediction strategy is chosen because we develop AL algorithms with distinct selection strategies for each branch, as detailed in Sections III-B1 and III-B2. We applied t-SNE dimensionality reduction to the embedded features of the training set samples extracted by the network, and the outcomes are presented in Fig. 2. The network-extracted features are predominantly distributed in various clusters. Consequently,

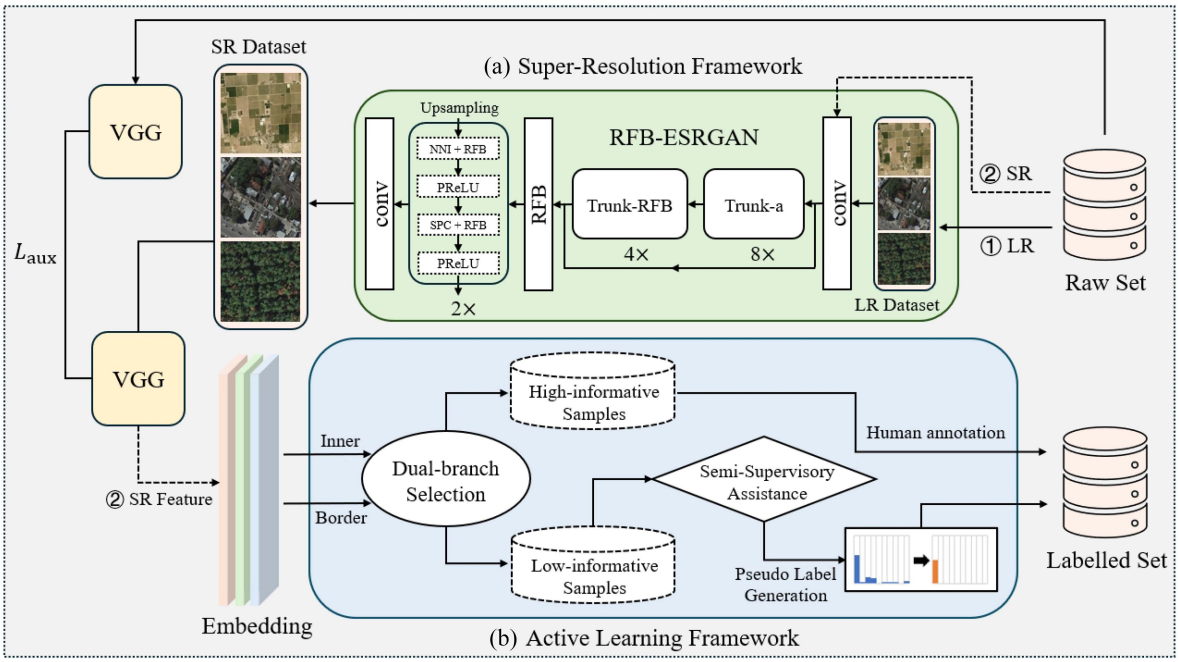

Fig. 1. Architecture of the proposed joint network of SR and AL. The whole architecture consists of two parts. (a) Pretrained SR framework, which is the main architecture of RFB-ESRGAN. (b) AL framework, which introduces a dual-branch AL selection strategy and an SSAS.

Fig. 2. Distribution obtained by dimensionality reduction of all embedding features using t-SNE. Notice that the features are distributed in clusters; in-class features are close, and different classes' features are distant.

the underlying concept of the AL algorithm proposed in this article is to select samples that are situated as far as possible from the distribution of these training set embeddings.

1) Inner Branch: Given S_{SR^2} as base set B and $x_i \in B$ belonging to class i, the embedding layer is denoted as $e(x_i)$. Extract the embedding of all samples in the base set B, and the average for each category according to the category is calculated as

$$c(i) = \frac{1}{n(i)} \sum_{x_i \in B} e\left(x_i\right) \tag{4}$$

where $n(i)$ represents the number of samples in class i. The computational center calculates the average of all samples in each category and uses this average as a representative point for each category. The unlabeled set of samples contributes by being measured in terms of the distance to this representative point.

Once the center is found, the interclass branching can be measured for the unlabeled set samples. Since AL requires reselecting samples in the unlabeled set B that are as far away from the train set as possible, we design the following formula to measure the samples

$$D\left(x_i\right) = \sum_{l=1}^{C} \ln\left(\left\|e\left(x_i\right) - \mathrm{c}(l)\right\|_2\right) \tag{5}$$

where C represents the number of categories. This formula is actually the logarithm of the distance between the embedding $e(x_i)$ of the unlabeled samples and all centers, and then, summed. This design ensures that when $e(x_i)$ is sufficiently close to a specific average, its contribution will decrease infinitely. Conversely, the value will increase when $e(x_i)$ is far from the average of any class. This meets the criteria of AL for sample selection.

2) Border Branch: The interclass selecting method mentioned above effectively selects samples from the unlabeled set that are far from any class. However, this method does not select samples between any two classes because these samples are closer to both two classes. The contribution calculated according to the method in Section III-B1 will be smaller. Therefore, it is necessary to consider designing a new method to select the samples at the decision boundary between any two classes.

After calculating the center of each class, given class $m, n \in \{1, 2, \ldots, C\}$, the samples within the unlabeled set B are calculated as follows:

$$D\left(x_i\right) = \min_{m \neq n} \left\{\left\|e_{x_i} - \mathrm{c}(m)\right\|_2 - \left\|e_{x_i} - \mathrm{c}(n)\right\|_2\right\}. \tag{6}$$

The distance is calculated using the derived embedding with each type of center. Then, any two distances are subtracted. The

distance between a sample located at the decision boundary and the center will be of equal length. This formula can calculate the minimum distance of the unlabeled sample from the embedding of any two classes. When this distance is as low as possible, the sample is located at the decision boundary, and its contribution will be high.

3) Inner-Border Balance Selecting Strategy: In order to better integrate the two methods mentioned in Sections III-B1 and III-B2, we design a fusion contribution selection strategy to balance the contributions obtained from the above two. The proposed inner-boundary balanced selecting strategy is able to fuse the two in an arbitrary proportion.

Assuming that N samples need to be added in each round of additions, where the percentage of interclass methods is P_{In}, $N \cdot P_{\text{In}}$ samples are selected from the inner branch and $N \cdot (1 - P_{\text{In}})$ samples are selected from the boundary branch. The contributions obtained from predicting the two branches are ranked from highest to lowest to obtain two corresponding indices. Finally, the index with the highest contribution from the two branches is added to the base set. If this sample is already in the base set, it does not affect the number of samples already selected in this branch. The cycle continues until the corresponding branch reaches the number of added samples.

C. SSAS

The designed AL method can help in choosing informative samples, which reduces the need for labels and lowers the labeling cost. However, this approach may require discarding less informative samples, and it may not fully utilize the information from the entire unlabeled set. In addition, the model itself has some recognition ability during the learning process. Therefore, we have designed a SSAS to take advantage of the model's own prediction potential in this regard.

We believe that adding less informative samples to a certain base set and its corresponding unlabeled set does not provide a significant benefit. However, having less informative features can help semisupervised learning by enabling the model to better differentiate unlabeled samples. This means that the prediction result is closer to the actual truth. As a result, we process the prediction results and use them as labels for the unlabeled samples in subsequent training.

Given an unlabeled sample x_i of class i, the predicted distribution $\hat{y}(x_i)$ is obtained through AL model, after which a pseudolabel is generated using the following equation:

$$y_{\text{pseudo}}(i) = \begin{cases} 1, & \hat{y}_{\text{unlabeled}}(i) = \max \hat{y}\,(x_i) \\ 0, & \text{else} \end{cases}. \tag{7}$$

Among them, y_{pseudo} is the pseudolabel given by the semisupervised auxiliary strategy for unlabeled samples. This strategy enables the automatic labeling of samples with more accurate predictions, which can reduce labeling costs.

The architecture is shown in Fig. 3. The proposed SSAS seems to have the ability to annotate accurately. Still, this strategy is not so robust cause it is difficult to guarantee that the labels provided by semisupervised learning are 100% accurate. Incorrect labels will inevitably reduce the accuracy of deep learning models in

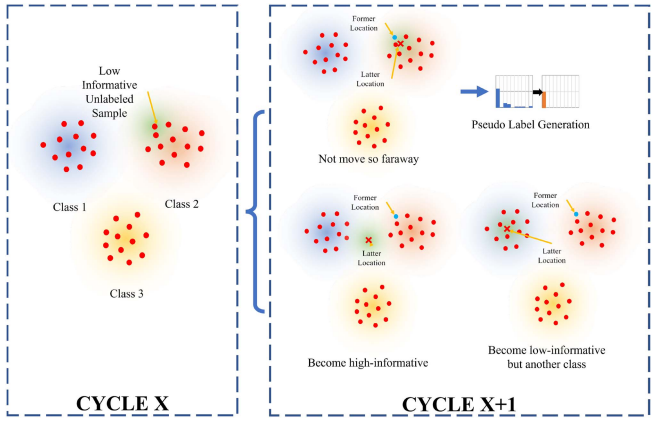

Fig. 3. Schematic representation of the proposed consistency-based SSAS.

a supervised learning task. Therefore, we propose a robust solution called consistency-based semisupervised learning. First, after training on the labeled set, save the model and select a certain amount of high information samples for annotation, then, the training of the next cycle continues. After the next cycle is trained, we use the model obtained from the two cycles to predict the unlabeled samples separately. Then, calculate the similarity of the two predictions using the following equation:

$$\text{Sim}(D(x^t), D(x^{t+1})) = \|D(x^t) - D(x^{t+1})\|_2 \tag{8}$$

where $D(x^t), D(x^{t+1})$ represents the same unlabeled sample score predicted by two cycles, at the time t and $t+1$. If the two predictions are consistent, the pseudolabel obtained by semisupervised learning is considered correct. Then, treat the hard labels predicted by (7) as the annotations, and move these samples to the labeled set. The consistency indicator $\mathcal{I}$ can be written as

$$\mathcal{I}(D(x^t), D(x^{t+1})) = D(x^t) \cdot \text{Sim}(D(x^t), D(x^{t+1})) \cdot D(x^{t+1}). \tag{9}$$

This concept is based on the following assumption: if two models, one before and one after a cycle, make different predictions on a sample, then, that sample is considered to be still highly informative, even if one of the models is highly confident in its prediction. However, if the two predictions are consistent, then, the sample is correctly predicted and can be annotated and moved to the labeled set. In subsequent experiments, it is observed that the impact of the consistent semisupervised learning strategy is greater than that of simple semisupervised learning.

IV. Results

A. Experimental Environment

1) Experimental Environment: To evaluate the effectiveness and robustness of the joint network, we conducted several experiments, including multiple rounds of parameter tuning and dual-branch ablation experiments. These experiments were carried out using our experimental platform, which is equipped with an Intel i7 12700 K CPU and two NVIDIA RTX 3090 Ti graphics cards. The platform runs on the Ubuntu 20.04

Fig. 4. NR09 dataset. The dataset contains 9 common agricultural land classes, including farmland, meadow, residential, forest, desert, hills, roads, lake and river. Each class consists of 200 images of 256 × 256 size.

TABLE I
COMPARATIVE EXPERIMENT TO VERIFY THE EFFECT OF VARIOUS AL ALGORITHMS ON THE NR09 DATASET

Method	10%	20%	30%	40%	50%	60%	70%	80%	90%
Core Set	80.22%	89.38%	91.54%	93.07%	93.27%	94.03%	94.44%	94.51%	94.70%
Learning loss	80.20%	88.45%	90.21%	92.55%	93.18%	93.87%	94.39%	94.48%	94.55%
Distance entropy	80.22%	83.04%	84.09%	84.75%	85.02%	87.49%	88.89%	90.03%	91.98%
Ours (without SR)	80.22%	89.75%	92.13%	93.15%	93.74%	94.21%	94.52%	94.52%	94.74%
Ours	**80.23%**	**89.97%**	**92.77%**	**93.84%**	**94.28%**	**94.43%**	**94.65%**	**94.65%**	**94.83%**

operating system, and we use PyTorch 1.9 as the deep learning framework.

2) Dataset: In order to construct a small dataset for agricultural land classification, this chapter extracted a shared database from the reference [48]. The original dataset contains 31 500 images belonging to 45 classes. However, this dataset suffers from intraclass variability and interclass data imbalance, which leads to the wastage of computational resources. To simplify the dataset and meet the demands of agricultural tasks, this chapter selected 9 common agricultural land classes (a total of 1800 samples) in this new balanced dataset, called NR09. The common agricultural land classes contain farmland, meadow, residential, forest, desert, hills, roads, lake, and river. Each class consists of 200 color images corresponding to the RGB channels of 256 × 256 size. In addition, 20 images (10%) from each class are uniformly selected as a test set for evaluating the performance of the selected dataset. Some samples are shown in Fig. 4.

B. Performance of the Joint Network

We co-train the joint network of SR and AL, using the sample selection effects of the active learning network to fine-tune the SR model. After SR modeling, the original low-resolution blurred remote sensing images are super-resolved by a factor of 8. The result of SR is shown in Fig. 5. After reconstruction, the lost texture features in the image are restored, which facilitates the subsequent embedding layer to extract the features of the agricultural images.

Fig. 5. Comparison of low-resolution remote sensing images before and after reconstruction.

The results of our sample addition experiments using coreset [32], learning loss [18], distance entropy [49], and our proposed method are shown in Table I. The comparative experiment starts with the same 10% samples and selects 10% samples every cycle, until the labeled set reaches 90% samples. The best results are highlighted in bold and the second best

TABLE II
COMPARATIVE EXPERIMENT TO VERIFY THE EFFECTIVENESS OF DOUBLE-BRANCH SELECTION STRATEGY

P_{In}	10%	20%	30%	40%	50%	60%	70%	80%	90%
AL	90.88%	91.17%	92.92%	**94.74%**	93.89%	92.55%	91.12%	90.33%	89.31%
AL + SR	91.34%	92.49%	93.76%	**94.83%**	94.17%	93.04%	91.91%	90.85%	90.08%

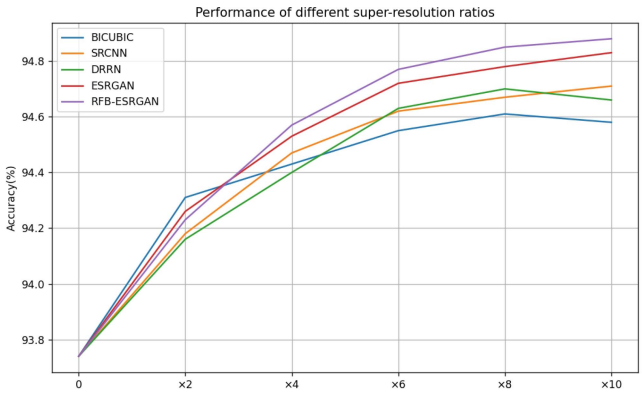

Fig. 6. Performance of different SR models in different SR ratios. It can be found that, the prediction accuracy of the joint networks embedded with different SR models all show an increasing trend as the SR ratio increases.

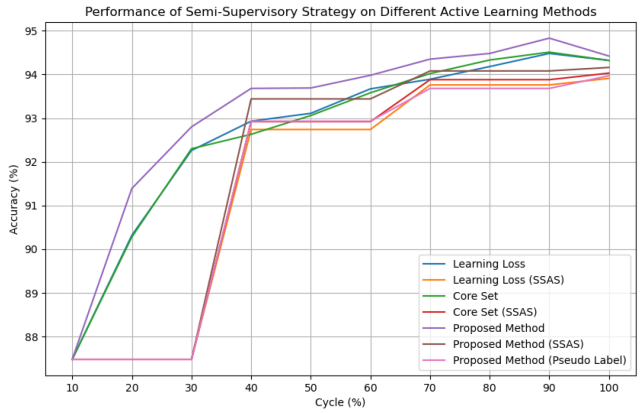

Fig. 7. Comparative experiment to verify the effect of SSAS. It can be found that the group using this strategy can achieve close to the test accuracy trained by full annotation but with less (40%) annotation cost.

are underlined. The experimental results demonstrate introduction of the SR network can bring about an improvement in the model performance, and our proposed joint network has a strong sample selection capability. It should be noted that we also compared the results of both branches for validation and found that the selecting results of both branches are more efficient than any of the branches, which effectively shows the effectiveness of our proposed dual-branch method. We added a set of hyperparametric experiments for this purpose. We took different hyperparameters P_{In} and compared these values with the two branches. We were able to find that the best selection was achieved when P_{In} was taken at 40%, indicating that both branches contributed more to the selection of the sample. The result is shown in Table II.

C. Experiments of Different SR Strategy

In order to further verify the role of SR models in improving the accuracy of remote sensing agricultural classification, we compare the performance of the joint network embedded in different SR models and at different SR ratios. The size of the original remote sensing images before reconstruction is 256 × 256. We set 2x, 4x, 6x, 8, and 10x magnification parameters, and select BICUBIC [40], SRCNN [42], DRRN [45], ESRGAN [46], and RFB-ESRGAN [47] as pretrained model for comparison. The results are shown in Fig. 6.

It is obvious that as the SR ratio increases, the prediction effect of the joint network has an overall upward trend, and the change in accuracy gradually slows down when the SR ratio is greater than 6. On the other hand, as the SR ratio increases, the SR model parameters and computations increase, and the model is difficult to converge, which requires a larger scale of sample data for optimization. Hence, we set the SR ratio as 8 to balance the performance.

D. Comparison Experiments of SSAS

In this study, we carry out a comparison experiment of a semisupervised learning strategy for the method proposed in this article. The control group uses an AL algorithm with 10% of the data as the base set. In each cycle, 10% of samples from the unlabeled set were selected and added to the base set. This process was repeated for a total of 10 cycles. On the other hand, the AL algorithm using the SSAS selected 10% of samples from the unlabeled set and moved them into the base set in each cycle, for a total of 4 cycles. The results of the experiment are presented in Fig. 7.

We can see that until the sample size approaches 100%, the performance of using the SSAS parallels that of the samples without the AL strategy, which is slightly lower at some stages. After comparison, we can find that the accuracy obtained using the consistent semisupervised learning strategy is higher than the ordinary semisupervised learning (pseudolabel) and also closer to the ordinary AL algorithm, yet the time consumed is only 40% of the ordinary AL, which is enough to prove the practicality of our method.

The results above show that with the assistance of the proposed SSAS, our proposed method can improve the performance of the labeled dataset as much as possible while further reducing human annotation. The semisupervised ratio in Fig. 7 is 1 (5% for human annotation and 5% for SSAS annotation). To explore the impact of using different semisupervised ratios in the proposed methods, we add experiments of 0.5 (2.5%), 1.5 (7.5%), 2 (10%), and 2.5 (12.5%).

The results are shown in Fig. 8. By analyzing the results, We can draw the following two conclusions. 1) The model's strong prediction capability is sufficient for accurately labeling low-informative samples, and when combined with AL, it can efficiently annotate unlabeled samples. 2) The SSAS needs to be

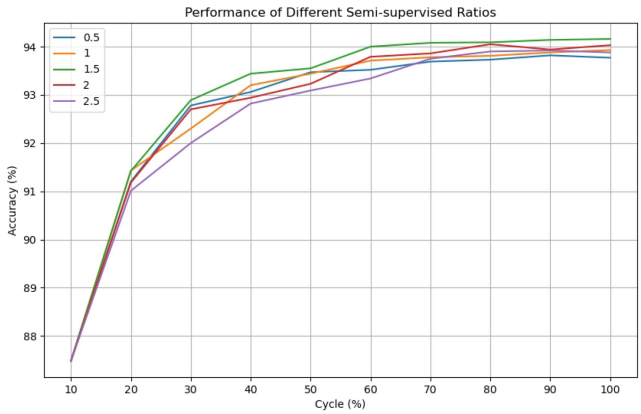

Fig. 8. Comparative experiment to verify the effect of different semisupervised ratios on the same AL backbone method.

at an optimal ratio to maximize its effectiveness. If the ratio is too low, the model's prediction cannot be effectively leveraged. Conversely, if the ratio is too high, the model's prediction capability may limit its accuracy on certain samples, leading to performance degradation.

V. Conclusion

Remote sensing datasets face the problem of low resolution and high labeling cost, which makes deep learning difficult to generalize in agriculture image processing under this limitation. This article provides a joint network of SR framework and semisupervised explanation-friendly AL framework. The introduction of pretrained SR models facilitates the reconstruction of important information such as texture details in low-resolution datasets. At the same time, the DBSS guarantees the selection and annotation of high-value samples. These enhance the highly accurate classification of agricultural remote sensing datasets.

Moreover, the traditional AL framework only relies on the representation of the unlabeled samples by the model, and selects the samples that are difficult to identify or have a great impact on the model for labeling. This ignores the gain of the samples with high confidence in the model, and the use of this part requires no extra manpower. SSAS can select samples with high confidence and participate in training with pseudolabels, which reduces the cost of manual labeling. The results also demonstrate the effectiveness of the semisupervised strategy. In the future, we will optimize the mechanism of the joint network, and explore the use of large models in remote sensing. At the same time, we will expand our work to image classification in multiple remote sensing scenarios such as urban land planning, maritime targets, etc.

Acknowledgment

The authors especially acknowledge the Artificial Intelligence and Marine Information Processing Lab (AIMIP) at Tianjin Univeristy, Tianjin, China and the Automated Systems and Soft Computing Lab (ASSCL) at Prince Sultan University, Riyadh, Saudi Arabia. The authors wish to acknowledge the editor and reviewers for their insightful comments, which have improved the quality of this publication.

References

[1] M. Xi et al., "A lightweight reinforcement learning-based real-time path planning method for unmanned aerial vehicles," *IEEE Internet Things J.*, vol. 11, no. 12, pp. 21061–21071, Jun. 2024.
[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[5] Z. Li, M. Dong, S. Wen, X. Hu, P. Zhou, and Z. Zeng, "CLU-CNNs: Object detection for medical images," *Neurocomputing*, vol. 350, pp. 53–59, 2019.
[6] P. F. Jaeger et al., "Retina U-net: Embarrassingly simple exploitation of segmentation supervision for medical object detection," in *Proc. Mach. Learn. Health Workshop*, 2020, pp. 171–183.
[7] J. He, J. Wen, S. Xiao, and J. Yang, "Multi-AUV inspection for process monitoring of underwater oil transportation," *IEEE/CAA J. Automatica Sinica*, vol. 10, no. 3, pp. 828–830, 2023.
[8] J. Yang, C. Cheng, S. Xiao, G. Lan, and J. Wen, "High fidelity face-swapping with style convtransformer and latent space selection," *IEEE Trans. Multimedia*, vol. 26, pp. 3604–3615, 2023.
[9] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet : A nested u-net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, Granada, Spain, 2018, pp. 3–11.
[10] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.
[11] S. Li et al., "Recent advances in intelligent processing of satellite video: Challenges, methods, and applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 6776–6798, 2023.
[12] M. Hao, G. Dou, X. Zhang, H. Lin, and W. Huo, "A subpixel mapping method for urban land use by reducing shadow effects," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2163–2177, 2023.
[13] Y. Zhang, Y. Du, F. Ling, S. Fang, and X. Li, "Example-based super-resolution land cover mapping using support vector regression," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1271–1283, Apr. 2014.
[14] Royal College of Radiologists, "Clinical radiology UK workforce census 2015 report," 2016, pp. 1–57.
[15] W. H. Beluch, T. Genewein, A. Nürnberger, and J. M. Köhler, "The power of ensembles for active learning in image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9368–9377.
[16] K. Wang, D. Zhang, Y. Li, R. Zhang, and L. Lin, "Cost-effective active learning for deep image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 12, pp. 2591–2600, Dec. 2017.
[17] Y. Li and J. Yang, "Meta-learning baselines and database for few-shot classification in agriculture," *Comput. Electron. Agriculture*, vol. 182, 2021, Art. no. 106055.
[18] D. Yoo and I. S. Kweon, "Learning loss for active learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 93–102.
[19] X. Chao and Y. Li, "Semisupervised few-shot remote sensing image classification based on KNN distance entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8798–8805, 2022.
[20] J. Nie, J. Jiang, Y. Li, H. Wang, S. Ercisli, and L. Lv, "Data and domain knowledge dual-driven artificial intelligence: Survey, applications, and challenges," *Expert Syst.*, 2023, Art. no. e13425.
[21] J. Tanha, M. Van Someren, and H. Afsarmanesh, "Semi-supervised self-training for decision tree classifiers," *Int. J. Mach. Learn. Cybern.*, vol. 8, pp. 355–370, 2017.
[22] J. Nie, Y. Wang, Y. Li, and X. Chao, "Sustainable computing in smart agriculture: Survey and challenges," *Turkish J. Agriculture Forestry*, vol. 46, no. 4, pp. 550–566, 2022.
[23] J. Wen, H. Dai, J. He, M. Xi, S. Xiao, and J. Yang, "Federated offline reinforcement learning with multimodal data," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 4266–4276, Feb. 2024.

[24] S. Xiao, J. Yang, and Z. Lv, "Protecting the trust and credibility of data by tracking forgery trace based on GANs," *Digit. Commun. Netw.*, vol. 8, no. 6, pp. 877–884, 2022.
[25] M. Lienou, H. Maitre, and M. Datcu, "Semantic annotation of satellite images using latent dirichlet allocation," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 28–32, Jan. 2010.
[26] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279.
[27] E. Li, J. Xia, P. Du, C. Lin, and A. Samat, "Integrating multilayer features of convolutional neural networks for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5653–5665, Oct. 2017.
[28] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," 2015, *arXiv:1508.00092*.
[29] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1155–1167, Feb. 2019.
[30] Y. Gal, R. Islam, and Z. Ghahramani, "Deep bayesian active learning with image data," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1183–1192.
[31] J.-J. Zhu and J. Bento, "Generative adversarial active learning," 2017, *arXiv:1702.07956*.
[32] O. Sener and S. Savarese, "Active learning for convolutional neural networks: A core-set approach," 2017, *arXiv:1708.00489*.
[33] Y. Wang, L. Yao, G. Meng, X. Zhang, J. Song, and H. Zhang, "Semi-supervised object detection in remote sensing images based on active learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 5571–5574.
[34] L. Möllenbrok and B. Demir, "Active learning guided fine-tuning for enhancing self-supervised based multi-label classification of remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 4986–4989.
[35] G. Lan et al., "Image aesthetics assessment based on hypernetwork of emotion fusion," *IEEE Trans. Multimedia*, vol. 26, pp. 3640–3650, 2023.
[36] S. Xiao, Z. Zhang, J. Yang, J. Wen, and Y. Li, "Manipulation detection of key populations under information measurement," *Inf. Sci.*, vol. 634, pp. 1–13, 2023.
[37] D.-H. Lee et al., "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Representation Learn.*, 2013.
[38] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," 2016, *arXiv:1610.02242*.
[39] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017.
[40] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.
[41] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Trans. Multimedia*, vol. 21, no. 12, pp. 3106–3121, Dec. 2019.
[42] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. 13th Eur. Conf. Comput. Vis.*, Zurich, Switzerland, 2014, pp. 184–199.
[43] Y. Wang, Z. Shao, T. Lu, C. Wu, and J. Wang, "Remote sensing image super-resolution via multiscale enhancement network," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5000905.
[44] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.
[45] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3147–3155.
[46] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 63–79.
[47] T. Shang, Q. Dai, S. Zhu, T. Yang, and Y. Guo, "Perceptual extreme super-resolution network with receptive field block," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 440–441.
[48] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017, doi 10.1109/JPROC.2017.2675998.
[49] Y. Li and X. Chao, "Distance-entropy: An effective indicator for selecting informative data," *Front. Plant Sci.*, vol. 12, 2022, Art. no. 818895.

Jiachen Yang (Senior Member, IEEE) received the M.S. and Ph.D. degrees in communication and information engineering from Tianjin University, China, in 2005 and 2009, respectively.

He is currently a Professor with Tianjin University. He was a Visiting Scholar with the department of computer science, School of Science, Loughborough University, Loughborough, U.K. His research interests include ocean information processing, Internet of Things, cloud computing, big data analytics, stereo vision research, pattern recognition, and image quality evaluation.

Zechen Wang received the B.S. degree in communication engineering in 2022 from Tianjin University, Tianjin, China, where he is currently working toward the M.S. degree in communication and information engineering with the School of Electrical and Information Engineering.

His research interests include artificial intelligence, information processing, and data mining.

Desheng Chen is currently working toward the Ph.D. degree in communication and information engineering with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China.

His research interests include marine exploration, marine intelligent equipment control, artificial intelligence, and Big data computing.

Shuai Xiao (Member, IEEE) received the Ph.D. degree in information and communication engineering from Tianjin University, Tianjin, China, in 2022.

He is currently an Associate Scientist with the School of Electrical and Information Engineering, Tianjin University. His research interests include artificial intelligence, generate adversarial networks, and cloud computing.

Ahmad Taher Azar received the M.Sc. and Ph.D. degrees from Faculty of Engineering, Cairo University, Giza, Egypt, in 2006 and 2009, respectively.

He is currently a Full Professor with the College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia. He is a Leader of Automated Systems and Soft Computing Lab (ASSCL), Prince Sultan University. He is also a Full Professor with the Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt. He has authored or co-authored more than 500 research papers in prestigious peer-reviewed journals, book chapters, and conference proceedings. His research interests include artificial intelligence, control theory and applications, robotics, machine learning, computational intelligence, and dynamical system modeling.

Dr. Azar is currently the Editor for IEEE SYSTEMS JOURNAL, IEEE TRANSACTIONS NEURAL NETWORKS AND LEARNING SYSTEMS, Springer's *Human-Centric Computing and Information Sciences*, and Elsevier's *Engineering Applications of Artificial Intelligence*.