A Novel Query Strategy-Based Rank Batch-Mode Active Learning Method for High-Resolution Remote Sensing Image Classification

Xin Luo,1,2,3 Huaqiang Du,1,2,3 Guomo Zhou,1,2,3 Xuejian Li,1,2,3 Fangjie Mao,1,2,3 Di’en Zhu,4 Yanxin Xu,1,2,3 Meng Zhang,1,2,3 Shaobai He,1,2,3 and Zihao Huang,1,2,3

Abstract: An informative training set is necessary for ensuring the robust performance of the classification of very-high-resolution remote sensing (VHRRS) images, but labeling work is often difficult, expensive, and time-consuming. This makes active learning (AL) an important part of an image analysis framework. AL aims to efficiently build a representative and efficient library of training samples that are most informative for the underlying classification task, thereby minimizing the cost of obtaining labeled data. Based on ranked batch-mode active learning (RBMAL), this paper proposes a novel combined query strategy of spectral information divergence lowest confidence uncertainty sampling (SIDLC), called RBSIDLC. The base classifier of random forest (RF) is initialized by using a small initial training set, and each unlabeled sample is analyzed to obtain the classification uncertainty score. A spectral information divergence (SID) function is then used to calculate the similarity score, and according to the final score, the unlabeled samples are ranked in descending lists. The most “valuable” samples are selected according to ranked lists and then labeled by the analyst/expert (also called the oracle). Finally, these samples are added to the training set, and the RF is retrained for the next iteration. The whole procedure is iteratively implemented until a stopping criterion is met. The results indicate that RBSIDLC achieves high-precision extraction of urban land use information based on VHRRS; the accuracy of extraction for each land-use type is greater than 90%, and the overall accuracy (OA) is greater than 96%. After the SID replaces the Euclidean distance in the RBMAL algorithm, the RBSIDLC method greatly reduces the misclassification rate among different land types. Therefore, the similarity function based on SID performs better than that based on the Euclidean distance. In addition, the OA of RF classification is greater than 90%, suggesting that it is feasible to use RF to estimate the uncertainty score. Compared with the three single query strategies of other AL methods, sample labeling with the SIDLC combined query strategy yields a lower cost and higher quality, thus effectively reducing the misclassification rate of different land use types. For example, compared with the Batch_Based_Entropy (BBE) algorithm, RBSIDLC improves the precision of barren land extraction by 37% and that of vegetation by 14%. The 25 characteristics of different land use types screened by RF cross-validation (RFCV) combined with the permutation method exhibit an excellent separation degree, and the results provide the basis for VHRRS information extraction in urban land use settings based on RBSIDLC.

Keywords: spectral information divergence; query strategy; ranked batch-mode active learning; Worldview-3; urban land use
1. Introduction

Very-high-resolution remote sensing (VHRRS) images, which contain valuable features, particularly spatial information, have drawn much attention in urban land use monitoring in recent years [1–6]. Moreover, with the development of aerospace technology and sensor technology, VHRRS images are becoming increasingly easy and inexpensive to obtain, which raises the problem of enormous amounts of data being underutilized. Thus, unlabeled data are abundant, but labeling is time-consuming and expensive. Driven by the explosion of remotely sensed datasets, the establishment of accurate and effective methods for remotely sensed imagery information extraction is a prerequisite for applications and investigations of remote sensing technology.

The extraction methods of urban land use information mainly include support vector machines [7,8], decision trees [9], random forest (RF) models [10,11], and deep learning [12–14]. Most of these methods are based on supervised learning, which requires many labeled samples for model training [15]. Therefore, it is challenging to examine the training of classifiers with limited samples for the interpretation of remote sensing images [16]. Many scholars have explored methods to reduce the dependence of models on samples, such as transfer learning [17], few-shot learning [18], semi-supervised learning [19–22], unsupervised learning [23,24], and weakly supervised learning [25,26], and have achieved favorable results. However, the performance of these methods is not comparable to that of supervised learning methods.

Active learning (AL) finds the most “valuable” training samples through heuristic strategies and aims to achieve high accuracy using as few labeled samples as possible for the underlying classification task. It can achieve or even exceed the expected effect while minimizing the labeling cost [27–30]. This approach effectively solves the problem of classifier training with limited samples and has attracted attention from scholars worldwide [28,31,32]. To use AL on an unlabeled dataset, a very small sample of these data must first be labeled and a model must be trained. After the model is built, predictions should be made for all the unlabeled data. The labeling should be prioritized using a score. The selected sample should then be labeled and a model should be trained. The steps can be repeated iteratively to improve the model. AL includes two learning modes: stream-based selective sampling and pool-based sampling [33]. Stream-based selective sampling determines whether samples need to be labeled in sequence. The disadvantage is that it is impossible to obtain the structural distribution of samples. Pool-based sampling involves forming many unlabeled samples in the sample pool and then using a certain screening strategy to select the most “valuable” samples from the pool for priority labeling. However, there is a common problem between the two AL types; notably, only one sample can be selected for annotation in each iteration, resulting in exceedingly low efficiency. To improve the efficiency of AL sample labeling, researchers have proposed batch-mode active learning (BMAL) [34–37]. BMAL is a learning model based on a sample pool from which a batch of unlabeled samples is generated in each iteration. The samples are labeled through various methods, thus solving the problem of low sample labeling efficiency. However, BMAL only uses an active selection strategy based on a single uncertainty index or diversity index when screening samples, which leads to considerable information redundancy in the labeled samples and unnecessary labeling costs. Cardoso [38] proposed the ranked batch-mode active learning (RBMAL) framework, which overcomes the limitations of traditional BMAL methods and generates an optimized ranked list to determine the priority of samples being labeled. Therefore, the RBMAL method has higher flexibility than the classic methods.

The query strategy is a critical part of the AL method. The objective is to select valuable samples for model training, an approach which is directly related to reducing the cost of annotation. At present, the commonly used query strategies can be divided into uncertainty sampling methods and query-by-committee (QBC) methods. The uncertainty-based methods include least confidence [39], margin sampling [40], and entropy-based sampling [41], and the QBC methods include the vote entropy and Kullback–Leibler maximization methods. A single query strategy usually leads to sampling bias [42]; that
is, the selected samples may not effectively reflect the distribution characteristics of the sample dataset. Therefore, combined query strategies have become popular [43,44].

In this study, we design a novel combined query strategy of spectral information divergence lowest confidence uncertainty sampling (SIDLC) based on RBMAL (RBSIDLC), to achieve high accuracy in extracting urban land use information from Worldview-3 VHRRS data by using as few labeled samples as possible. The combined query strategy, combining the RF as the base classifier for uncertainty estimation and a SID similarity measure function to form a novel query strategy, aimed to select the most informative and “valuable” samples for labeling. First, the RF is initialized with a few initial training sets, and the classification uncertainty score is obtained through the analyses of unlabeled samples. Then, the SID function is used to calculate the similarity score, and according to the final score, the unlabeled samples are ranked in a descending list. Third, the most uncertain samples are selected according to the ranked list and labeled by the oracle. Finally, after labeling, such samples are incorporated into the training set, which is used to retrain the RF in the next iteration until a satisfactory result is obtained. The research results would potentially introduce a new AL method and the AL-based VHRRS information extraction.

2. Materials and Methods
2.1. Study Area

The research area is the West Lake District of Hangzhou City, located in Zhejiang Province in Southeast China (Figure 1). The topography of the study area is plains, and rivers run vertically and horizontally, with a high degree of urbanization and highly complex ground features. There are mainly buildings with relatively regular shapes, grassland, barren land, large and regular street trees, regular roads, prominent small squares and playgrounds, etc. The main feature types basically cover the types of urban land. Although a standard Worldview-3 image has 8 MS bands with 1.24 m spatial resolution, and a panchromatic with 0.30 m spatial resolution [45], in this study, the collected Worldview-3 image over the study area was obtained on 28 October 2018, including four multispectral bands (R, G, B, and NIR) and one panchromatic band. The spatial resolution of the four multispectral bands is 2 m, and that of the panchromatic band is 0.5 m. ENVI 5.3 Gram–Schmidt pan sharpening [46] was used to fuse multispectral images with the panchromatic band to obtain a multispectral remote sensing image with a resolution of 0.5 m.

![Figure 1](https://via.placeholder.com/150)

**Figure 1.** Location of the study area: (a) Zhejiang Province. The blue polygon represents Hangzhou; (b) the subregion of the West Lake District of Hangzhou.
In this study, according to a field investigation, the USGS land cover classification system [47] and the FROM-GLC10 [48,49] classification system include five land cover types: barren land, built-up land, water, grassland, and forest. In addition, Figure 1b shows an image acquired on 28 October 2018. The solar elevation angle was approximately 46°, and consequently, the effects of shadowing were compounded in regions where there were dramatic changes in surface elevation; that is, in urban areas. The tall buildings and trees depicted cast shadows that obscured many other surface features. Even the smaller buildings cast shadows that obscured details of the surrounding streets [50,51]. Therefore, the shadows of trees, grasslands, and buildings are regarded as a sixth type of land use. Detailed descriptions of the land use classes and the corresponding subclasses are listed in Table 1.

Table 1. Urban land use classes and corresponding subclass components.

<table>
<thead>
<tr>
<th>Land Use Classes</th>
<th>Subclass Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barren land</td>
<td>Dry salt flats, bare exposed rock, and sandy areas other than beaches</td>
</tr>
<tr>
<td>Built-up land</td>
<td>Residential, industrial, transportation, commercial, and service lands</td>
</tr>
<tr>
<td>Shadows</td>
<td>Shadows of trees, grasslands, and buildings</td>
</tr>
<tr>
<td>Water</td>
<td>Streams, rivers, and ponds</td>
</tr>
<tr>
<td>Grassland</td>
<td>Natural grassland and planted grassland</td>
</tr>
<tr>
<td>Forest</td>
<td>Deciduous forest and evergreen forest</td>
</tr>
</tbody>
</table>

2.2. Training Sample Set

The training sample set used in this study combined (1) the initial training samples obtained from a field investigation integrated with a visual interpretation method and (2) highly homogeneous image segmentation objects generated by multi-scale segmentation.

The multi-scale object segmentation is employed to increase the size of training sample sets. The valuable detailed information in VHRRS images can interfere with the extraction of object boundaries because the boundaries of multi-scale segmentation may not completely match those of objects. Only the center pixel of each object is selected and added to the training set to avoid mislabeling samples due to the incorrect extraction of segmentation object boundaries [21]. In the following sections, we describe in detail how to assign the label to the center pixel. First, multi-scale segmentation is performed based on the VHRRS images [13], as shown in Figure 2a. Second, the initial training sample set is used to train the temporary RF₀, and the initial classification map can be obtained by RF₀, as shown in Figure 2b. Finally, the segmentation map is projected onto the initial classification map, as shown in Figure 2c. When the pixels within the segmented object have the same predicted label, the center pixel of the corresponding object is used and the predicted label is added to the training set to increase the training samples, as shown in Figure 2d.

The training sample sets obtained through the above steps are shown in Table 2, where the initial sets are the initial training sample sets, expansion sets are the expanded training sample sets, and the datasets are the combined sample sets. A total of 273 objects are selected, and 31,636 pixels are labeled.

Table 2. Reference samples collected.

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Barren Land (Objects)</th>
<th>Built-Up Land (Objects)</th>
<th>Shadows (Objects)</th>
<th>Water (Objects)</th>
<th>Grassland (Objects)</th>
<th>Forest (Objects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial sets</td>
<td>10 (723)</td>
<td>10 (1226)</td>
<td>10 (854)</td>
<td>10 (826)</td>
<td>10 (873)</td>
<td>10 (926)</td>
</tr>
<tr>
<td>Expansion sets</td>
<td>10 (2513)</td>
<td>90 (9034)</td>
<td>52 (9435)</td>
<td>5 (2747)</td>
<td>16 (2400)</td>
<td>40 (4079)</td>
</tr>
<tr>
<td>Datasets</td>
<td>20 (3236)</td>
<td>100 (10,260)</td>
<td>62 (6289)</td>
<td>15 (3573)</td>
<td>26 (3273)</td>
<td>50 (5005)</td>
</tr>
</tbody>
</table>

2.3. Feature Setting and Optimization

In this study, the classification features included 4 spectral features, 14 vegetation index features, and 192 texture features calculated from 4 spectral bands (Table 3). The texture features of the four spectral bands include eight variables: the mean, variance, homogeneity,
contrast, dissimilarity, entropy, angular second moment, and correlation [52–54]. There are also different windows: 3 × 3, 5 × 5, 7 × 7, 9 × 9, 11 × 11, and 13 × 13 [55]. The number of features selected was directly related to the computational efficiency of classification. This study used an RF cross-validation (RFCV) method for the above 210 out-of-bag (OOB) error score features to determine the number of optimal features to reduce the dimensionality of features and improve the speed of operation. Then, the permutation method was used to rank all variables and select the optimal features [10,56]. Finally, the optimal features were used as input variables.

**Table 3.** Used variables and formulations of variables.

<table>
<thead>
<tr>
<th>Feature Types</th>
<th>Feature Names</th>
<th>Details</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral</td>
<td>Blue band (B)</td>
<td>450–510 nm</td>
<td>Worldview-3 data</td>
</tr>
<tr>
<td></td>
<td>Green band (G)</td>
<td>510–580 nm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Red band (R)</td>
<td>630–690 nm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Near-infrared band (NIR)</td>
<td>770–1040 nm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ratio vegetation index (RVI) [57]</td>
<td>NIR/R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Difference vegetation index (DVI) [59]</td>
<td>NIR − R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normalized difference vegetation index (NDVI) [59]</td>
<td>(NIR − R)/(NIR + R)</td>
<td></td>
</tr>
</tbody>
</table>
|                      | Green normalized difference vegetation index (GN
DVI) [60]                                        | (NIR − G)/(NIR − G)                          |                             |
|                      | Soil adjusted vegetation index (SAVI) [61]        | (NIR − R)/(NIR + R + L), L = 0.5             |                             |
|                      | Triangular vegetation index (TVI) [59]            | (G − R)/(G + R)                              |                             |
|                      | Green vegetation index (Vg) [62]                  | (NIR/R − 1)/((NIR/R)1/2 + 1)                 |                             |
|                      | Modified simple ratio index (MSRI) [63]           | (NIR − R − 0.2(NIR − G) + (NIR/R)            | X take value for 0.16       |
|                      | Modified chlorophyll absorption in reflection inde
x (MCARI) [64]                                  | (NIR − R)/(NIR − G)                          | L take value for 0.5 [58]   |
|                      | Transformed chlorophyll absorption in reflectance
dex (TCARI) [65]                                | 3((NIR − R) − 0.2(NIR − G) + (NIR/R)         |                             |
|                      | Renormalized difference vegetation index (RDVI) [6
6]                                                | (NIR − R)/√((NIR − R)                      |                             |
|                      | Greenness index (GI) [67]                         | G/R                                          |                             |
|                      | Green leaf index (GLI) [68]                       | ((G − R + (G + R))/2(G + R + B)              |                             |
|                      | Enhanced vegetation index (EVI) [69]              | 2.5(NIR − R)/(NIR + 6R − 7.5B + 1))         |                             |

<table>
<thead>
<tr>
<th>Texture features based on the gray-level co-occurrence matrix (GLCM) [70]</th>
<th>Mean (ME)</th>
<th>$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} iP(i,j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance (VA)</td>
<td>$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - \text{mean})^2 P(i,j)$</td>
<td></td>
</tr>
<tr>
<td>Entropy (EN)</td>
<td>$-\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j) \log(P(i,j))$</td>
<td></td>
</tr>
<tr>
<td>Angular second moment (SE)</td>
<td>$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)^2$</td>
<td></td>
</tr>
<tr>
<td>Homogeneity (HO)</td>
<td>$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)/(1+P(i,j))$</td>
<td></td>
</tr>
<tr>
<td>Contrast (CON)</td>
<td>$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1}</td>
<td>i-j</td>
</tr>
<tr>
<td>Dissimilarity (DI)</td>
<td>$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1}</td>
<td>i-j</td>
</tr>
<tr>
<td>Correlation (COR)</td>
<td>$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)(i-j)^2/(1+P(i,j))$</td>
<td></td>
</tr>
</tbody>
</table>

$\mu_y = \sum_{i=0}^{N-1} P(i,j)$

$\sigma_y = \sum_{i=0}^{N-1} (i - \mu_y)^2 P(i,j)$

$\sigma_y^2 = \sum_{i=0}^{N-1} (i - \mu_y)^2 P(i,j)$

$P(i,j) = V(i,j)/\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} V(i,j)$

$V(i,j)$ is the $i$th row of the $j$th column in the $N \times N$ moving window.

Figure 2. Multi-scale segmentation-based sample selection. Segment the VHRRS image with multi-scale segmentation and select the center pixel of each object with a “pure” classification to enlarge the training set. (a) Multi-scale image segmentation; (b) initial classification diagram of the RF0; (c) selection of the center pixel of an object; (d) expanded training set.
2.4. Ranked Batch-Mode Active Learning

RBMAL was proposed by Cardoso in 2017 [38], and the approach has three key steps. (1) Uncertainty estimation. A classifier is trained with the initial labeled samples. After the model is built, make predictions for all samples in unlabeled sample pool U. The probability of each classification is then used to associate a score with the corresponding samples. (2) Ranked batch construction. Using a score, the samples in the unlabeled sample pool U are repeatedly selected to generate a descending ranking Q. (3) Labeling. The oracle labels one or more samples in the ranked list and requests another iteration. This whole process is detailed in Algorithm 1.

Algorithm 1: Ranked batch-mode active learning algorithm.

Input: A set with labeled samples \( L_0 \)
Input: A set with unlabeled samples \( U \)
1: Train the classifier with \( L_0 \); then, perform uncertainty estimation for the samples in \( U \): \( U_{\text{uncertainty}} \)
2: The sample sets are labeled or ranked: \( L_0_{\text{estimated}} \)
3: EmptyList of the sample ranking: \( Q \)
4: for \( u < |U| \) do
   \( u_0 = \text{SelectSample}(L_0_{\text{estimated}}, U_{\text{uncertainty}}) \)
   \( L_0_{\text{estimated}} = L_0_{\text{estimated}} \cup u_0, U_{\text{uncertainty}} = U_{\text{uncertainty}} - u_0 \)
   Insert into List \( Q, u_0 \)
   \( u_0 = u_0 + 1 \)
end for
5: \( l = \text{Oracle label}(Q) \)
6: \( L_0 = L_0 \cup l, U = U - l \)
7: return \((L, U)\)

2.5. Combined Query Strategy—SIDLC

This section proposes a combined query strategy of SIDLC, selecting the most informative sample at each step. This approach consists of two parts. One is uncertainty estimation, in which the probability that each unlabeled sample belongs to a known category is determined through an uncertainty estimator. Then, the uncertainty score is calculated. The other part is spectral similarity measurement, in which the similarity score between unlabeled and labeled samples is calculated as being equal to 1.0 minus the highest similarity between a sample of \( U_{\text{uncertainty}} \) and \( L_{\text{estimated}} \). High scores can be assigned to samples with low similarity to identify samples with high uncertainty among the unlabeled samples.

SIDLC prioritizes unlabeled samples with low scores. The function for calculating the score of unlabeled samples is as follows:

\[
\text{score} = \alpha(1 - \text{SID}) + (1 - \alpha)U(x)
\]

where \( U(x) \) is the uncertainty of the forecast, \( \text{SID} \) is a similarity measure based on spectral information divergence, and \( \alpha = \frac{|X_{\text{labeled}}|}{|X_{\text{labeled}}| + |X_{\text{unlabeled}}|} \) is the \( \alpha \) parameter responsible for weighting the impact of the two scores on the final score. \( \alpha \) is dynamically set based on the size of the labeled and unlabeled sample sets. Thus, it is feasible to shift the sample prioritization scheme from diversity to uncertainty. The \( \alpha \) parameter is updated when new labeled content is provided. The entire process is detailed in Algorithm 2.

2.5.1. Uncertainty Score Estimation Based on an RF

The classifier is the core of an uncertainty estimator. The higher the classifier’s uncertainty is based for an unlabeled sample, the more informative the sample is to the classification process. If informative, the sample can be used to determine the probabilities of samples in the unlabeled sample pool \( U \) belonging to a known class. These probabilities are then further processed to obtain an uncertainty score.
Algorithm 2: Combined query strategy.

Input: A set with labeled samples $L_0$
Input: A set with unlabeled samples $U$
Input: Current $\alpha$ parameter

1: best score = $-1$
2: InformationSample = nil
3: For $u_0 \in U$ do
4: UncertaintyScore = UncertaintyScore ($U$)
5: similarity SID = SID–Function ($L_0$, $U$)
6: score = $\alpha \times (1 - \text{similarity}_{\text{SID}}) + (1 - \alpha) \times \text{UncertaintyScore}$
7: If score > best Score then
8: best score = score
9: InformationSample = $u_0$
10: end if
11: end for
12: return InformationSample

The RF method is widely used in the classification of remote sensing images [11,71]. This study uses RF to calculate the predicted class probabilities of the input samples. The predicted class probabilities of input samples are computed as the mean predicted class probabilities of the trees in the forest, which reflect the probability that a given unlabeled sample belongs to a known class.

The probabilities of unlabeled samples produced by the RF model must be converted to uncertainty scores. In this study, the least-confident uncertainty score is used as follows:

$$X^*_{LC} = \arg \max (1 - \hat{p}^*(k|x))$$  \hspace{1cm} (2)

where $\hat{k} = \arg \max (\hat{p}^*(k|x))$; $\arg \max$ involves selecting the least-confident sample, which is the largest value in brackets; and $k$ denotes taking the subscript with the highest probability.

2.5.2. Similarity Function Based on the SID

Generally, the similarity function in RBMAL adopts the Euclidean distance and does not consider the spectral characteristics of samples [38]. The purpose of spectral similarity measurement in the combined query strategy in this study is to determine the similarity between the unknown spectrum and a known spectrum according to a spectral similarity measurement function and then divide the attributes of unknown categories according to the similarity results [72–74]. This approach is consistent with the description of the similarity function in RBMAL. Therefore, in this study, the SID proposed by Chang [75,76] is used to replace the Euclidean distance [77,78].

The SIDs between $x_{\text{labeled}} = (x_{\text{labeled}1}, \ldots, x_{\text{labeled}M})^T$ and $x_{\text{unlabeled}} = (x_{\text{unlabeled}1}, \ldots, x_{\text{unlabeled}M})^T$ are the labeled sample set and unlabeled sample set, respectively, as follows:

$$\text{SID}(x_{\text{labeled}}, x_{\text{unlabeled}}) = \left( \sum_{m=1}^{M} r_m \log (r_m/s_m) + \sum_{m=1}^{M} s_m \log (s_m/r_m) \right)$$  \hspace{1cm} (3)

where $r_m = x_{\text{labeled}M} / \sum_{t=1}^{M} x_{\text{labeled}t}$, $s_m = x_{\text{unlabeled}M} / \sum_{t=1}^{M} x_{\text{unlabeled}t}$. $M$ represents spectral dimensionality, and $x_{\text{labeled}t}$ and $x_{\text{unlabeled}t}$ represent the $t^{th}$ elements of vectors $x_{\text{labeled}}$ and $x_{\text{unlabeled}}$. The larger the value of $\text{SID}(x_{\text{labeled}}, x_{\text{unlabeled}})$ is, the less similar $x_{\text{labeled}}$ and $x_{\text{unlabeled}}$ are. The implementation of the similarity function is detailed in Algorithm 3.
Algorithm 3: Set the SID function.

Input: A sample set \( x_{labeled} \)
Input: A sample \( x_{unlabeled} \) to be evaluated
1: SID-MAX = 0.0
2: for all \( u_0 \in x_{unlabeled} \) do
3: \( \text{Sid} = \text{SID-Function}(u_0, x_{labeled}) \)
4: If \( \text{Sid} > \text{SID-MAX} \) then
5: \( \text{SID-MAX} = \text{Sid} \)
6: End if
7: End for
8: return \( \text{SID-MAX} \)

2.6. The RBSIDLC Framework

The RBSIDLC framework for urban land use VHRRS information extraction mainly includes four parts and eight steps (Figure 3).

Figure 3. Overall workflow of the proposed AL algorithm.

Initial sample part (Step 1):
(1) According to a field investigation and visual interpretation, the initial training set \( l_1 \) is obtained. The RF has two parameters, and the number of trees is 800. This parameter is determined by 10 cross-validations based on OOB scores, and the other parameters are defaults used to set temporary RF\(_0\) values.

Sample expansion part (Steps 2–3):
(2) Train the temporary RF\(_0\) with the initial training set \( l_1 \), obtain the initial classification map, and segment the VHRRS image with a multi-scale segmentation method.
(3) The segmentation image is projected onto the initial classification image, and multiple pixels in the centers of segmentation objects are selected as the expanded sample set \( l_2 \), where \( L = l_1 \cup l_2 \).

Divide the dataset and set the parameters (Step 4):
(4) In this section, dataset \( L \) is divided into two parts: 70% for training and 30% for testing. Then, from the training data, 20 samples are randomly selected for label training \( L_0 \), which includes two parts: class labels and the feature vector. Then, the remaining samples are treated as an unlabeled pool (U), and other samples form a feature vector. In the AL selection iteration, the parameter BATCH-SIZE is used to control the number of samples that need to be labeled by the oracle in each iteration. Referring to related studies, 10 was selected as the BATCH-SIZE [79,80].
**AL part (Steps 5–8):**

(5) Train an RF based on training label set $L_0$, use the RF to classify the samples in $U$, and obtain the probability that each sample belongs to each class. Then, compute the least-confident uncertainty scores by using Equation (3). The similarity scores are calculated according to Algorithm 3 (SID function).

(6) Use the combined query strategy SIDLC to calculate the final score and rank. Algorithm 2 is used to select 10 samples from set $U$ for oracle labeling. The labeled samples form a set $l$.

(7) Let $U = U - l$ and $L_0 = L_0 \cup l$, and use $L_0$ to retrain each tree in the RF.

(8) Repeat steps 5 to 8 multiple times; when $U = \emptyset$ or the model reaches the predetermined accuracy criterion, the algorithm stops, and the final RF model is obtained.

3. Experimental Results and Analysis

3.1. Feature Importance Screening Results

The variation in the OOB error with the number of variables calculated by the RFCV method is shown in Figure 4. The most prominent finding is that with the decreasing variables, the OOB score gradually decreases. When the number of variables is greater than 25, changes in the curve tend to be stable. Therefore, 25 features are selected from 210 features to extract the VHRRS urban land use information.

![Figure 4. Effect of the number of variables on OOB error obtained by RFCV for feature selection with all 210 features.](image)

For different land use types, permutation algorithms can be employed to evaluate and rank all 210 features in variable importance. A total of 25 variables are selected for further investigation based on the decrease in accuracy for every urban land use type and mean decrease accuracy (MDA) over all classes (Figure 5). The texture is an important feature for extracting forests and grasslands. For forests, SEC_NIR_13 has the highest importance score, followed by COR_R_13, and for grassland, the red band has the highest MEA_R_13 importance score, followed by MEA_R_5. The NIR-band and GLI are important features that affect information extraction for built-up land. The vegetation index is an important feature for water and barren land. For water, the importance scores of the GNDVI and SAVI are high, and the importance scores of VIgreen and TVI for barren land are high. MEA_G_3 and MEA_G_5 are the most important features for shadow extraction.

<table>
<thead>
<tr>
<th>Spectral</th>
<th>G</th>
<th>R</th>
<th>NIR</th>
<th>Vegetation Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based on the G</td>
<td>Based on the R</td>
<td>Based on the NIR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIR</td>
<td>MEA_G_3</td>
<td>COR_R_11</td>
<td>MEA_NIR_9</td>
<td>RDVI</td>
</tr>
<tr>
<td></td>
<td>MEA_G_5</td>
<td>COR_R_13</td>
<td>MEA_NIR_11</td>
<td>RVI</td>
</tr>
<tr>
<td></td>
<td>VAR_G_13</td>
<td>COR_R_13</td>
<td>MEA_NIR_13</td>
<td>SAVI</td>
</tr>
<tr>
<td></td>
<td>HOM_G_13</td>
<td>MEA_R_11</td>
<td>COR_NIR_11</td>
<td>GLI</td>
</tr>
<tr>
<td></td>
<td>MEA_R_13</td>
<td>COR_NIR_13</td>
<td>DVI</td>
<td></td>
</tr>
<tr>
<td>SE_C_NIR_13</td>
<td>EVI</td>
<td>TVI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GNDVI</td>
<td>VIgreen</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 Note: Variable names are given as (feature name)-(base spectral band)-(window size). For example, MEA_NIR_13 indicates an MEA based on an NIR band with a window size of 13 × 13.
Figure 5. Variable importance scores of top 15 features. Variable names are given as (feature name)-(base spectral band)-(window size). For example, MEA_NIR_13 indicates an MEA based on an NIR band with a window size of 13 × 13. Bar color indicates the spectral bands, where green band is in green, red band in red, blue band in blue, NIR band in orange, and vegetation index variables in dark blue.

In summary, the features related to the R-band, G-band, and NIR-band correspond to higher importance scores than other features, and the larger the texture calculation window is, the higher the proportion. The results demonstrate that texture features and vegetation indices are important features in extracting VHRRS urban land use information. Therefore, this study selects the top 25 most important feature variables based on MDA as input variables for further analysis (Table 4).

Table 4. Top 25 features.

<table>
<thead>
<tr>
<th>Spectral</th>
<th>G Based on the G</th>
<th>R Based on the R</th>
<th>NIR Based on the NIR</th>
<th>Vegetation Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIR</td>
<td>MEA_G_3</td>
<td>COR_R_11</td>
<td>MEA_NIR_9</td>
<td>RDVI</td>
</tr>
<tr>
<td>MEA_G_5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAR_G_13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOM_G_13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Variable names are given as (feature name)-(base spectral band)-(window size). For example, MEA_NIR_13 indicates an MEA based on an NIR band with a window size of 13 × 13.
3.2. Results of Urban Land Use Information Extraction Based on RBSIDLC

Figure 6a shows the urban land use classification result extracted with the RBSIDLC method, and Table 5 shows the classification accuracy evaluation. Notably, the accuracy of each extracted land use type is greater than 90%, and the overall accuracy (OA) is greater than 96%, indicating that the method based on the RBSIDLC can extract land use information with high accuracy.

As shown in Table 5, RBSIDLC urban land use information extraction yields the highest OA values (96.83%). Compared with the RBMAL and RF methods, the OA is improved by 2.99% and 2.48%, respectively. Compared with the RBMAL method, the RBSIDLC method improves the OAs of barren land, built-up land, water, grassland, and forest; thus, RBSIDLC yields land use extraction results with high accuracy. Compared with the RF algorithm, the RBSIDLC method yields higher OAs for land uses other than water, with improvements to varying degrees. For example, the extraction accuracy of forest increases by approximately 3%, that of grassland increases by approximately 5%, and that of barren land increases by 7%.
The analysis in Table 5 also shows that the OA values of the RF and RBMAL are above 90%. As an estimator of the uncertainty score, the results demonstrate that the RF lays an essential foundation for the extraction of VHRRS urban land use information. Moreover, the SID is used instead of the Euclidean distance in the RBMAL algorithm, which improves the accuracy of the RBSIDLC method and indicates that the similarity function based on the SID is superior to that based on the Euclidean distance.

Figure 6 compares the land use classification results of the RBSIDLC, RBMAL, and RF methods. Further analysis shows that the RBMAL and RF methods seriously confuse barren land and grassland areas (Figure 6e,f, red circle), while the RBSIDLC algorithm can properly distinguish them (Figure 6d, black circle). In addition, the RF algorithm mistakenly classifies barren land as built-up land (Figure 6i, red circle). Although the RBMAL method properly classifies barren land (Figure 6h,i, red circle), the classification result is still inferior to that of RBSIDLC (Figure 6g, black circle).

Overall, these results suggest that the proposed RBSIDLC approach has high accuracy and can effectively classify barren land, grassland, and built-up land compared to the RBMAL and RF algorithms, thus improving the identification accuracy of land use categories and the extraction of urban land use information from VHRRS images. Thus, this method has obvious advantages over traditional methods.

3.3. Comparison with Other AL Query Strategies

In this section, to fully analyze the proposed query strategies, three additional query strategies based on the batch-mode AL method are implemented and used for comparison. The details of BBLC, BBM, and BBE are listed in Table 6. The extraction results for VHRRS urban land use information are shown in Figure 7.

<table>
<thead>
<tr>
<th>Algorithm Scene</th>
<th>Query Strategy</th>
<th>Abbreviation</th>
<th>Extraction Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch-based</td>
<td>Uncertainty sampling least</td>
<td>Batch-based least-confident (BBLC) method</td>
<td>Figure 7a</td>
</tr>
<tr>
<td>Sampling</td>
<td>confident</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Margin sampling</td>
<td>Batch-based margin (BBM) method</td>
<td>Figure 7b</td>
</tr>
<tr>
<td></td>
<td>Entropy based</td>
<td>Batch-based entropy (BBE) method</td>
<td>Figure 7c</td>
</tr>
</tbody>
</table>

It is apparent that the BBLC, BBM, and BBE methods seriously confuse barren land and built-up land (Figure 7d–f, red circle). In the BBLC and BBE algorithms, an extreme situation occurs in which barren land is misclassified as built-up land and grassland, especially in the BBLC algorithm; moreover, many barren land areas are misclassified as built-up land (Figure 7d, red circle). Although the BBM algorithm corrects the error, it still incorrectly classifies grassland as barren land. The RBSIDLC algorithm can properly identify grassland, barren land, and built-up land areas, and the classification result is the best among the results of the considered methods (Figure 6d, black circle).

Thus, a single query with a high number of samples increases the risk of confusion when querying samples (Figure 7d,f, red circle). The RBMAL algorithm mitigates this phenomenon by ranking samples based on the amount of information (Figure 6e, red circle), and RBSIDLC inherits the advantages of the RBMAL algorithm and replaces the Euclidean distance in the similarity function with the SID to further improve the accuracy of recognition for each category (Figure 6d,g, black circle). This approach has obvious advantages in urban land use classification.

Figure 8 shows the OA of urban land use information extraction, and the results of the RF, RBMAL, and RBSIDLC methods are compared with those of the methods in Table 6. The accuracies of BBLC and BBE are lower than that of the RF by 2.49% and 4.71%, respectively. Compared with that of the RF, the accuracy of the BBM algorithm is only 0.85% higher. This demonstrates that algorithms based on the batch-mode AL method may reduce the accuracy of RFs while increasing the number of labeled samples in a parallel environment. The accuracy of the RBMAL algorithm is 3% and 5.22% higher than the accuracies of the BBLC and BBE algorithms, respectively, which indicates that the
combined query strategy is more effective than single query strategies. The SID is used to replace the Euclidean distance in the similarity measurement function used in the RBMAL algorithm. The constructed RBSIDLC combined query strategy further improves urban land use classification accuracy, with the highest accuracy reaching 96.83%.

Figure 7. Optical classification results of the three AL methods when the “preferred feature” combination was used: (a) BBLC; (b) BBM; (c) BBE. And two subsets classification diagram (Subset (1) and Subset (2)). Subset (1): (d) BBLC; (e) BBM; (f) BBE. Subset (2): (g) BBLC; (h) BBM; (i) BBE.

Figure 8. OAs of the preferred feature classification results with different algorithms.
Figure 9 shows a confusion matrix that was used to further analyze the three AL methods in Table 6 and the results of urban land use information extraction with the RF, RBMAL, and RBSIDLC.

The RBMAL and RF misclassify 18% and 12% of barren land as built-up land, respectively, and 4% and 6% of grassland as built-up land, respectively (Figure 9). In contrast, the RBSIDLC algorithm reduces the misclassification to 4% between barren land and built-up land and 2% between grassland and built-up land. Therefore, the RBSIDLC method outperforms the other methods. Further analysis showed that the BBLC, BBM, and BBE algorithms mistakenly divide barren land into built-up land. The single most striking observation from the figure comparison is that the BBLC and BBE algorithms commonly misclassify barren land as built-up land and grassland. For example, the BBLC and BBE algorithms mistakenly classify 42% and 40% of barren land as built-up land, and 2% and 1% of barren land is misclassified as grassland, respectively. Compared with the BBLC, BBM, and BBE methods, the RBSIDLC algorithm improves the barren land, built-up land, and grassland classifications to varying degrees. For example, compared with that of the BBE algorithm, the barren land extraction accuracy of the RBSIDLC algorithm is increased by 37%, and the vegetation extraction accuracy is increased by 14%.

Together, the results confirm that each query strategy struggles to distinguish between barren land, grassland, and built-up land features. RBMAL can correctly identify most of the features by combining query strategies and RBSIDLC replaces Euclidean distance with the SID to obtain an improved classification result, indicating that a similarity function based on the SID is a better one based on Euclidean distance.

The sample labeling cost is an important index used to evaluate AL. In this study, referring to the relevant literature, we set the RF accuracy threshold and obtain the number of samples required by different AL algorithms. Then, all the methods in our experiments were implemented using Visual Studio Code platform python 3.7 with 3.7 GHz Intel i9-10900K CPU and 32 GB RAM. The SavedRate of different query strategies is then calculated [81] (Table 7).
### Table 7. Cost evaluation of five active learning methods with labeling.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Labeled Samples</th>
<th>Time(s)</th>
<th>SavedRate</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBLC</td>
<td>540</td>
<td>221.325</td>
<td>0.97562</td>
</tr>
<tr>
<td>BBM</td>
<td>440</td>
<td>221.371</td>
<td>0.98013</td>
</tr>
<tr>
<td>BBE</td>
<td>970</td>
<td>429.453</td>
<td>0.95619</td>
</tr>
<tr>
<td>RBMAL</td>
<td>420</td>
<td>212.562</td>
<td>0.98102</td>
</tr>
<tr>
<td>RBSIDLC</td>
<td>410</td>
<td>162.196</td>
<td>0.98149</td>
</tr>
</tbody>
</table>

The results indicate that the three query strategies based on the BBLC, BBM, and BBE methods require the most samples, followed by RBMAL, with RBSIDLC ranking last, which means that RBSIDLC is associated with the lowest sample labeling cost (Table 7). This result demonstrates that the proposed query strategy achieves the expected effect with as low an annotation cost as possible, accelerates the labeling of samples per unit time, and improves efficiency.

### 4. Discussion

The RBSIDLC method proposed in this study achieves high accuracy levels in the extraction of urban land use information from VHRRS images. Compared with other batch-mode AL query strategies, this method yields the highest accuracy, with an OA of 96.83%. These superior results can be explained from the following three perspectives.

First, an RF was adopted to estimate the uncertainty score in the SIDLC combined query strategy. Uncertainty estimation is the process of estimating the informativeness of an unlabeled sample. The labeled samples are added to the training set and the RF is used as a classifier to retrain the model for the next estimation. This cycle improves the estimator’s performance and the training accuracy of the RF, thus providing an important foundation for the extraction of VHRRS urban land use information. In addition, the SID is used to replace the Euclidean distance in the RBMAL algorithm, and the spectral and spatial characteristics of VHRRS images are fully considered, which effectively solves the misclassification issues for barren land, grassland, and built-up land observed for the RBMAL and RF algorithms, thereby improving the precision of land use category identification.

Moreover, the query strategy has a considerable impact on the accuracy of batch-mode AL information extraction. This study analyzes the OAs of different query strategies and RBSIDLC in extracting urban land use information (Figure 8). The BBLC, BBM, and BBE methods use a single query strategy to query many samples at a time, which increases the risk that confusion among query samples will affect the accuracy of the classifier (Figures 9 and 10). Although the RBMAL algorithm mitigates this phenomenon, the information extraction results are still not ideal (Figure 7e). By inheriting the advantages of the RBMAL algorithm, RBSIDLC adopts the SIDLC combined query strategy, which improves the recognition rate of each category and yields the highest accuracy in urban land use information extraction. These results support the hypothesis that the SIDLC combined query strategy is more effective than a single query strategy.

Finally, the labeling quality of training samples is high for RBSIDLC to achieve VHRRS urban land use information. Figure 10a shows the training accuracy curves of all the compared algorithms based on preferred features. It is apparent from this figure that as the number of labeled samples increases, the changes in accuracy for RBMAL and RBSIDLC become more stable than those of the other methods; notably, RBSIDLC displays no apparent fluctuations. Further analysis showed that the RBSIDLC algorithm can query and obtain high-quality training samples. However, the results of the BBLC, BBM, and BBE algorithms fluctuate broadly, especially in the range of 0–100, which indicates that the training samples they query are chaotic. The results show clearly that RBSIDLC can select high-quality samples and reduce disturbances in the batch-mode AL method, thereby ensuring the accuracy of land use information extraction.
The parameter BATCH-SIZE in the iterative AL selection process is used to control the number of samples that need to be labeled by the oracle in each iteration. In this study, referring to relevant studies, we set BATCH-SIZE to 10 and achieved promising results. Additionally, the effect of BATCH-SIZE (set to 1, 3, 5, 7, and 9) on the accuracy is assessed, as shown in Figure 10b. The most prominent finding is that with increasing BATCH-SIZE, the model’s accuracy is improved, and the fluctuations in batch processing are reduced. The results of this study indicate that it is reasonable to adopt a value of 10 for BATCH-SIZE.

In addition, this paper uses the permutation method to rank the feature importance of all variables and selects 25 features from 210 feature variables to be included in the extraction of urban land use information. The results are improved with this approach. We further explored the importance of each feature for sample separation and selected nine important features for analysis. The results indicate that vegetation (forest and grassland) displays strong reflection in the NIR-band and for the vegetation indices, namely the TVI, DVI, and GNDVI (Figure 11). Forests and grasslands can have high reflection values based on different vegetation types. For VIgreen, GLI, MEA_G_5, and MEA_R_13, the six land use categories exhibit a favorable gradient distribution, and the average of each category is highly variable.

Further analysis shows that the selected features yield a good degree of separation for the six land use categories. This preliminary finding suggests that the proposed feature selection method outperforms most artificial feature combination methods.

Based on this research, there is a lot of work to be carried out in the future. First, we can further study the two parts of the query function: (1) other similarity functions can be considered, such as distance-based spectral similarity measurement method, spectral angle similarity measurement, etc.; (2) in the uncertainty estimation part, the combinations of different classifiers and uncertainty strategy will produce different results, and we can replace the base classifier, for example, KNN, SVM, etc. We can also choose different query strategies to find the best combination for remote sensing image information extraction. In addition, deep learning is an effective tool for dealing with complex classification problems. How to use deep learning models to improve the ability of active learning algorithms is also worth studying.
5. Conclusions

This paper proposes a ranked batch-mode AL classification framework with a new query strategy called SIDLC for improving the accuracy in extracting urban land use information from VHRRS images. For the optimal feature dataset, the OA of RF classification are both above 90% for the proposed method. As the estimator of uncertainty scores and the core classifier used in the framework, the RF lays the foundation for the high-precision extraction of VHRRS urban land use information. The RBSIDLC algorithm replaces the Euclidean distance in the RBMAL algorithm with the SID. Due to the advantages of RBMAL, the spectral and spatial information in VHRRS images is fully analyzed; notably, the classification OA reaches 96.83%. These experiments confirm that replacing the Euclidean distance with the SID in the similarity function can improve extraction performance. In addition, the SIDLC combined query strategy performs better than the batch-mode AL single query strategies, and the misclassification rates among different land types are reduced. Compared to the BBE algorithm, the extraction accuracy of barren land for the RBSIDLC algorithm increases by 37%, and the accuracy of vegetation extraction increases by 14%. This study finds that the SIDLC-based RBSIDLC algorithm has obvious advantages in extracting urban land use information from VHRRS images. Additionally, the Worldview-3 image is an important basis for feature extraction, and has great potential in urban land use information extraction. The RF feature selection method is used to select the optimal features for classification, thus avoiding subjective influence and dimensionality issues and improving model performance. The proposed approach provides a reference for selecting the optimal features and extracting VHRRS urban land use information. There is a lot of future work. Firstly, the two main components of combined query strategies could be further investigated. Secondly, how to use deep learning models to improve the ability of AL algorithms is also worth studying. Finally, studying the best components of the solution in remote sensing image information extraction will bring completely new challenges.
Author Contributions: Conceptualization, X.L. (Xin Luo) and H.D.; Data curation, X.L. (Xin Luo), X.L. (Xuejian Li), D.Z., Y.X., M.Z., S.H. and Z.H.; Formal analysis, X.L. (Xin Luo), X.L. (Xuejian Li), F.M., D.Z., Y.X., M.Z. and S.H.; Funding acquisition, H.D. and G.Z.; Investigation, X.L. (Xin Luo); Methodology, X.L. (Xin Luo); Project administration, H.D.; Software, X.L. (Xin Luo) and Z.H.; Validation, X.L. (Xin Luo) and S.H.; Visualization, X.L. (Xin Luo); Writing—original draft, X.L. (Xin Luo); Writing—review and editing, H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation (No. U1809208, No. 31901310, No. 31800538), the State Key Laboratory of Subtropical Silviculture (No. ZY20180201), the Key Research and Development Program of Zhejiang Province (2021C02005).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors gratefully acknowledge the supports of various foundations. The authors are grateful to the editor and anonymous reviewers whose comments have contributed to improving the quality.

Conflicts of Interest: The authors declare that they have no competing interest.

References


34. Haertel, R.; Felt, P.; Ringger, E.; Seppi, K. Parallel active learning: Eliminating wait time with minimal staleness. In Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing, Department of Computer ScienceBrigham Young University, Provo, UT, USA, 1 June 2010.


52. Fatih, B.; Abdelkader, A.; Latifa, H.; Mohamed, E. Spatio temporal analysis of vegetation by vegetation indices from multi-dates satellite images: Application to a semi arid area in ALGERIA. Energy Procedia 2013, 36, 667–675. [CrossRef]


