Article

An Improved GrabCut Method Based on a Visual Attention Model for Rare-Earth Ore Mining Area Recognition with High-Resolution Remote Sensing Images

Yan Peng [1,2,3], Zhaoming Zhang [1,2,3],*, Guojin He [1,2,3],* and Mingyue Wei [4]

[1]   Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China; pengyan@radi.ac.cn

[2]   Key Laboratory of Earth Observation Hainan Province, Sanya 572029, China

[3]   Sanya Institute of Remote Sensing, Sanya 572029, China

[4]   College of Science, Central South University of Forestry and Technology, Changsha 410004, China; weimy@csuft.edu.cn

*   Correspondence: zhangzm@radi.ac.cn (Z.Z.); hegj@radi.ac.cn (G.H.); Tel.: +86-010-8217-8192 (Z.Z.); +86-010-8217-8190 (G.H.)

check for updates

Abstract: An improved GrabCut method based on a visual attention model is proposed to extract rare-earth ore mining area information using high-resolution remote sensing images. The proposed method makes use of advantages of both the visual attention model and GrabCut method, and the visual attention model was referenced to generate a saliency map as the initial of the GrabCut method instead of manual initialization. Normalized Difference Vegetation Index (NDVI) was designed as a bound term added into the Energy Function of GrabCut to further improve the accuracy of the segmentation result. The proposed approach was employed to extract rare-earth ore mining areas in Dingnan County and Xunwu County, China, using GF-1 (GaoFen No.1 satellite launched by China) and ALOS (Advanced Land Observation Satellite) high-resolution remotely-sensed satellite data, and experimental results showed that FPR (False Positive Rate) and FNR (False Negative Rate) were, respectively, lower than 12.5% and 6.5%, and PA (Pixel Accuracy), MPA (Mean Pixel Accuracy), MIoU (Mean Intersection over Union), and FWIoU (frequency weighted intersection over union) all reached up to 90% in four experiments. Comparison results with traditional classification methods (such as Object-oriented CART (Classification and Regression Tree) and Object-oriented SVM (Support Vector Machine)) indicated the proposed method performed better for object boundary identification. The proposed method could be useful for accurate and automatic information extraction for rare-earth ore mining areas.

Keywords: visual attention model; GrabCut; NDVI; mining area; high-resolution remote sensing image

---

1. Introduction

The Rare-earth Ore (REO) mining process, during which topsoil is stripped and large volumes of waste materials are removed from one place to another, leaving huge holes and piles on the Earth's surface [1], causes continuous change in topography and biodiversity, water pollution, soil erosion, and so on. These problems have disturbed human life and also restricted regional sustainable development, which requires an effective way to monitor and manage the surface mining activities. High-resolution remote sensing technologies have been recognized as promising tools for monitoring mining areas by several researchers [2–5].

Remote sensing classification methods can be generally divided into pixel-based approaches and object-oriented approaches. Pixel-based classification methods are generally applied to classify medium or coarse spatial resolution satellite images [6], and are not suitable for high spatial resolution mining region mapping. Object-oriented classification methods, which can use spectral, spatial, textural, and contextual information, were adopted to monitor mining activities with high-resolution satellite images by several researchers [2,7,8]. This method can obtain accurate mining area extraction results, however, it can be time-consuming, and the process usually depends on manual intervention. With the development of artificial intelligence, Song introduced a visual attention model to extract the mining areas with higher precision, speed, and automatic degree from high-resolution satellite images [3]. Inspired by human behavior, where humans usually make decisions using small local Regions of Interests (ROIs) of desired targets, the visual attention mechanism can focus attention on small regions of images [9]. However, the visual attention model itself has limited information processing capability [10,11]; the object boundary can hardly be detected accurately with only the visual attention model, therefore, the visual attention model should be combined with the image segmentation method. Traditional image segmentation methods include supervised methods, unsupervised methods, and interactive methods [12–14], among which interactive methods can achieve better segmentation results than other methods. As an interactive method, the GrabCut algorithm has been widely used because of its simple interactivity and satisfactory image segmentation results [15–19]. It has been applied to resolve different segmentation problems, such as medical computerized tomography (CT) and Positron Emission Computed Tomography (PET) image segmentation [16,17], human face segmentation [18], vehicle plate number recognition [19], and building extraction [20]. Until now, few studies have been performed using GrabCut for mining area segmentation with high-resolution remote sensing images. It should be noted the GrabCut method also has its drawbacks, e.g., it requires manual initialization [20]. Liu et al. used a salient region generated by the ITTI model as the initial inputs of the GrabCut method, instead of manual initialization, to segment the PET image, and good results were achieved [17]. This study was based on Liu et al's work. However, compared with traditional images (e.g., PET image), high-resolution satellite remote sensing images are multi-dimensional and highly complex, therefore, Liu et al.'s method must be improved and adapted in order to be applied to high-resolution satellite images.

Concerning the above issues, in order to make use of advantages of both the visual attention model and the GrabCut method, in this study the visual attention model was employed to generate a saliency map as the initial inputs of the GrabCut method instead of manual initialization, and NDVI (Normalized Difference Vegetation Index), a frequently used vegetation index in vegetative the remote sensing community, was designed as a bound term added into the Energy Function of GrabCut to reduce influences of vegetation and further improve the accuracy of the segmentation result. In this way, an improved GrabCut method based on the visual attention model is proposed in this paper to extract REO mining area information from high-resolution remote sensing images.

2. Materials and Methods

2.1. Research Area, Data, and Preprocessing

The southern part of Jiangxi province is rich in mineral resources, especially ion-absorbed REO mines. The terrain of the southern part of Jiangxi is dominated by hills and mountains. In order to test the universality of the proposed method, Lingbei REO mining region and Shipai REO mining region, two of the most prominent REO mining areas in the south of Jiangxi, and with over 20 years of REO exploitation history, were chosen as the study areas of this research. The locations of the study areas are shown in Figure 1. Furthermore, different spatial resolution data for each study area, including GF-1 (GaoFen No.1 satellite launched by China) and ALOS (Advanced Land Observation Satellite) satellite remote sensing images, were mainly used to extract REO mining area information, and the details of these images are listed in Table 1. GF-1 multispectral data have four spectral bands (band 1: 450–520 nm,

blue; band 2: 520–590 nm, green; band 3: 630–690 nm, red; band 4: 770–890 nm, near-infrared). ALOS multispectral data also have four spectral bands (band 1: 420–500 nm, blue; band 2: 520–560 nm, green; band 3: 610–690 nm, red; band 4: 760–890 nm, near-infrared). Geometric correction and image fusion were conducted for the satellite images before information extraction. The GF-1 data and ALOS were geometrically corrected using RPC (Rational Polynomial Coefficient) model, and the geometric errors of the corrected images were within 1 pixel. Subsequently, The PANSHARP method was introduced to fuse multispectral and panchromatic images. The PANSHARP fusion model is known as Pan-sharpening and tends to produce superior sharpening results, while preserving the spectral characteristics of the original images. And the final fused images are showed as Figure 2.
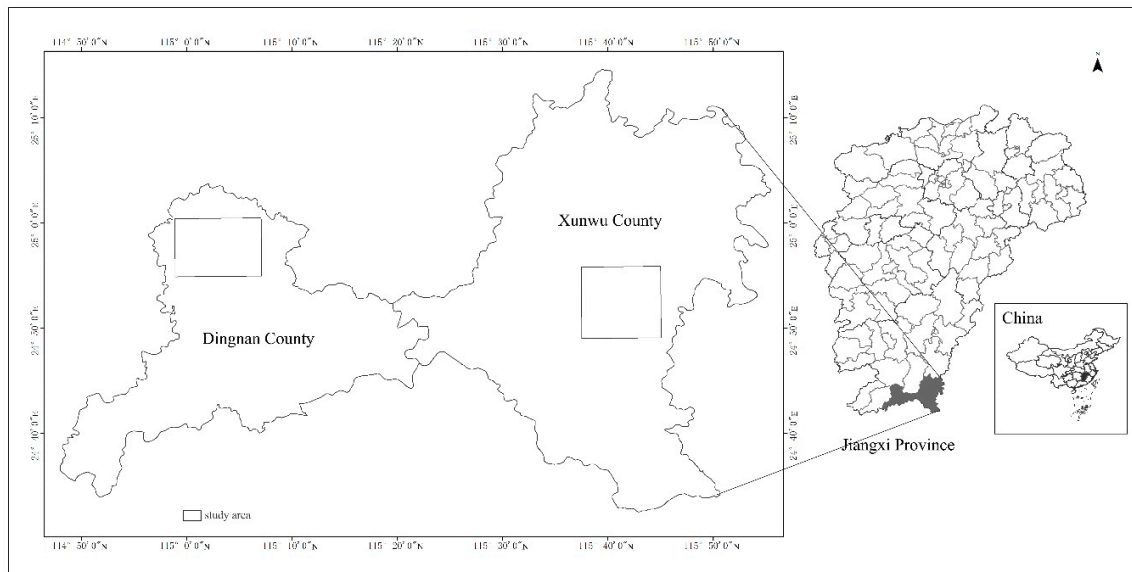

Figure 1. Location of Study Area.

Table 1. Remote sensing images detail list.

ID	Sensor	Resolution	Acquired Time	Study Area
1	GF-1 MSS2 [1]	8 m	2015-10-16	Lingbei
	GF-1 PMS2 [2]	2 m	2015-10-16	
2	ALOS AVNIR-2 [3]	10 m	2010-11-01	
	ALOS PRISM [4]	2.5 m	2010-11-01	
3	GF-1 MSS1 [5]	8 m	2014-12-12	Shipai
	GF-1 PMS1 [6]	2 m	2014-12-12	
4	ALOS AVNIR-2	10 m	2008-11-24	
	ALOS PRISM	2.5 m	2008-11-24	

[1] Multispectral scanning system type 2, [2] Panchromatic multispectral scanning system type 2, [3] Advanced visible and near infrared radiometer type 2, [4] Panchromatic remote-sensing instrument for stereo mapping, [5] Multispectral scanning system type 1, [6] Panchromatic multispectral scanning system type 1.

2.2. Methods

Figure 3 illustrates the overall flow chart of the proposed method. Aiming to detect the REO mining area automatically and accurately, the ITTI visual attention model was applied to produce a saliency map as the initial inputs of the GrabCut method, and in the improved GrabCut method NDVI was considered as a bound term of the Energy Function, mainly to restrict the vegetation information or other non-REO mining area features.
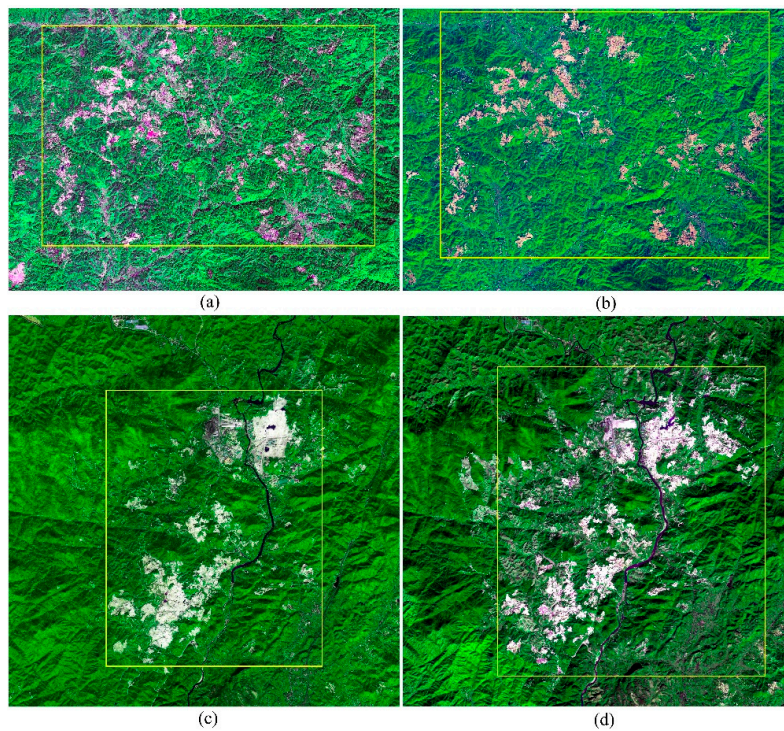
Figure 2. Four fused images used in this paper. (**a**) GF-1 image composed of R(3) G(4) B(1) in Lingbei, Dingnan County; (**b**) ALOS image composed of R(3) G(4) B(1) in Lingbei; (**c**) GF-1 image composed of R(3) G(4) B(2) in Shipai, Xunwu County; (**d**) ALOS image composed of R(3) G(4) B(2) in Shipai.
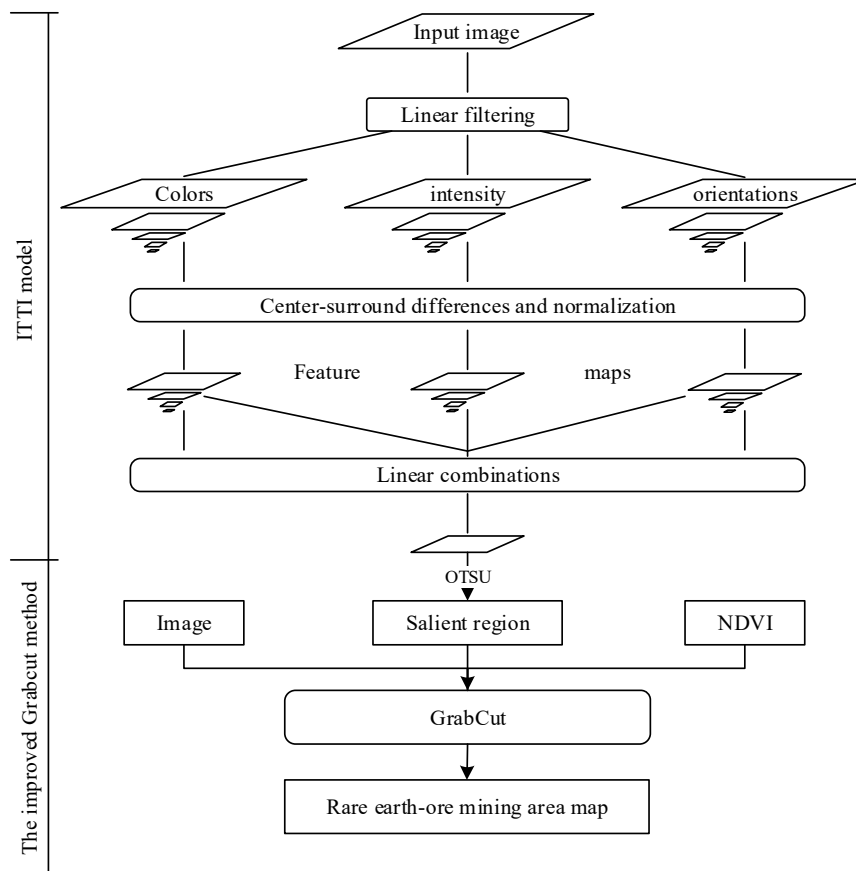

Figure 3. Flowchart of REO (Rare-earth Ore) mining area detection methods.

2.2.1. ITTI Visual Attention Model

The ITTI visual attention model is based on the visual attention mechanism of the human visual system, which solves the complex problem in scene understanding by quickly selecting salient regions for detailed analysis. It is a typical bottom-up significant viewpoint prediction model, which tries to quantitatively calculate the appearance of each point in the scene through the stimulation driven by the most basic image features of color, brightness, and orientation, and thus predicts the gaze point of the human eye [11]. The ITTI model was used to generate a REO mining area saliency map, and then form areas of interest in the scene. In general, there are three steps to generate the saliency map. Firstly, color, brightness, and orientation feature channels were extracted with different parts of the Gaussian pyramid according to the center-surround difference mechanism. Then, each feature map was integrated as a saliency map using the ITTI model normalization operator, which simulates the lateral cortical inhibition mechanism of human beings, and can enhance the significant feature regions and restrain the salient background peak value regions; this is the key procedure in the ITTI model. Finally, a saliency map was generated by calculating the average of the three saliency maps of each feature channel.

- Center-surround difference.

  The center-surround difference means the differences between "center" fine scale c yield and "surround" coarser scale s yield of the feature maps [11]. Both types of sensitivities are simultaneously computed in a set of six maps $CSD(c, s)$ [11]. Setting $P(n), n = 1, 2, \cdots, N$ as the pyramid image, the center-surround difference $CSD(c, s)$ of a feature can be obtained by Equation (1), and the feature map $\overline{F}$ can be calculated by Equation (2).

$$CSD(c, s) = \left| P(c) \ominus P(s) \right|, \tag{1}$$

$$\overline{F} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+\delta, \delta=3}^{\delta=4} CSD(c, s), \tag{2}$$

  where $\ominus$ is the difference between two different level images, which are resampled to the same resolution, $|\cdots|$ means absolute value, $\oplus$ is across-scale addition consist of reduction of each map to scale four and point-by-point addition, $c \in \{2, 3, 4\}, s = c + \delta, \delta \in \{3, 4\}$.

- Normalization Operator.

  The normalization operator is a key process in the ITTI model, and it mainly includes three steps. The first step is to unify the dimension among these feature maps, i.e., these maps are normalized to a fixed value range $[0, M]$. Secondly, the location of the maximum feature value M is calculated and the mean of the maximum values for all other local regions ($\overline{m}$) is also calculated. Finally, the feature maps are multiplied by $(M - \overline{m})^2$ pixel by pixel.

- Saliency Map Generation.

  In order to widen the gap among different center-surround differences of the same feature map in the saliency, and to ensure that effects of different features on the overall saliency map are independent, it is necessary to independently generate a conspicuity map for each channel's features before generating the overall saliency map, and the detailed process is expressed as Equations (3)–(5) [11]. The feature conspicuity maps include intensity, color, and orientation conspicuity maps. Then, the three conspicuity maps are normalized and weighted into the final saliency map, expressed as Equation (6).

$$\overline{I} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+\delta, \delta=3}^{\delta=4} N(I(c, s)), \tag{3}$$

$$\overline{C} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+\delta, \delta=3}^{\delta=4} \left[N(RG(c, s) + N(BY(c, s)) \right], \tag{4}$$

$$\overline{O} = \sum_{\theta \in \{0°,45°,90°,135°\}} N(\overset{4}{\underset{c=2}{\oplus}} \overset{\delta=4}{\underset{s=c+\delta,\delta=3}{\oplus}} N(O(c,s,\theta))), \tag{5}$$

where $\overline{I}, \overline{C}, \overline{O}$ indicate intensity, color, and orientation, respectively; $\oplus$ has been defined previously. For orientation, four θ (0°, 45°, 90°, 135°) values were given.

$$S = \frac{1}{3}[N(\overline{I}) + N(\overline{C}) + N(\overline{O})], \tag{6}$$

where N is the normalization, and S is the final saliency map.

2.2.2. Rare-earth Ore Mining Area Extraction Based on GrabCut

(1) GrabCut method

The GrabCut technique proposed by Rother et al. in 2004 is known as one of the state-of-the-art unsupervised semi-automatic methodologies for image segmentation, and it is developed to segment color images based on the graph cut algorithm [15]. It can obtain a minimum energy segmentation by building an energy model based on the Min-Cut Max-Flow algorithm [21]. GrabCut adopts Gaussian mixture models (GMMs) to build color distribution models of the foreground and background based on the probability of each given pixel and the foreground and background, which is given by interactively drawing a rectangle around the desired foreground object to assign only the background pixels [18].

The image is an array $z = (z_1, \ldots, z_n, \ldots, z_N)$ of RGB values $z_i = (R_i, G_i, B_i), i \in [1, \ldots, N]$. Segmentation of the image is expressed as an array $\alpha = (a_1, \ldots, a_N), \alpha_i \in \{0, 1\}$, with 0 for background and 1 for foreground. A trimap T is provided by the user with a semi-automatic interactive model, which includes initial background T_B, foreground T_F, and uncertain pixels T_U. Then, GMMs (Gaussian Mixture Models) are used to construct distribution histograms θ for the background and foreground, respectively, and each GMM is taken to be a full covariance with K components (typically $K = 5$); θ is expressed as Equation (7) [15]:

$$\theta = \left\{ \pi(\alpha,k), \mu(\alpha,k), \sum(\alpha,k), a = 0, 1, k = 1 \ldots K \right\}, \tag{7}$$

where π is the weights, μ is the means of the GMMs and $\sum$ the covariance matrices of the model. The Gibbs energy function for segmentation is then expressed as Equation (8) [15]:

$$E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z), \tag{8}$$

where U represents a data term to calculate the probability of a pixel to belong to some label and V represents a smoothness term, which is a regularizing prior term supposing that segmented objects should be consistent in light the of colour, taking the neighbourhood C around each pixel into account. The data term is composed of the Gaussian probability distributions of the GMM $p(z_i|\alpha_i k_i, \theta)$ and mixture weighting coefficients $\pi(\alpha_i, k_i)$, therefore, it is expressed as Equation (9) [15]:

$$U(\alpha, k, \theta, z) = \sum_i -\log p(z_i|\alpha_i k_i, \theta) - \log \pi(\alpha_i, k_i), \tag{9}$$

(2) Improved GrabCut model

The improvement is mainly reflected in two aspects:

- Energy Function.

  NDVI, a commonly used vegetation index in the quantitative remote sensing community, was added to the original energy function as a bound term, therefore, the improved energy function is expressed as Equation (10):

  $$E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z) + N(\alpha), \tag{10}$$

  where N is a bound term of NDVI to assist in extracting the REO mining area. It signifies the weight of a pixel belonging to the corresponding category identified by the NDVI data, and can be expressed as Equation (11):

  $$N(\alpha) = \omega \sum [N_i \neq \alpha_i], \tag{11}$$

  where ω is the weight of the added bound term, and it can be adjusted according to the actual situation. N_i represent the category tag of pixel i.

- Initial setting.

  For the original GrabCut method, user interaction is generally needed to fulfil satisfactory segmentation work. The initial and incomplete user-labeling, which is drawn as a rectangle by users, may finish the entire segmentation, but further user editing is required sometimes. Moreover, a remote sensing image is usually larger, more fragmented, and more complex than natural pictures; user interaction with labeled seed points will result in an inefficient segmentation process when GrabCut is applied for remote sensing image segmentation. Therefore, in this study the binarized map generated from the saliency map with the ITTI model was employed as an initial of the improved GrabCut method in order to accomplish the entire segmentation process efficiently and automatically.

2.2.3. Accuracy Evaluation Metrics

In order to judge whether a segmentation method is useful and effective, performance of the proposed method must be evaluated thoroughly by comparison with existing methods, such as SVM (Support Vector Machine) and CART (Classification and Regression Tree), using standard and well-known metric in many aspects including execution time and accuracy [22]. It's hard to evaluate execution time of SVM and CART method, because SVM and CART methods refer to select samples manually. However, the proposed method does not need to select samples, and the whole process is automatic without manual intervention. Thus, only the accuracy of the proposed method is evaluated by comparison with SVM and CART methods. There are many evaluation measures for assessing the accuracy of any segmentation method; these measures are usually variants of pixel accuracy (PA) and Intersection over Union (IoU). In this paper, False Positive Rate (FPR), False Negative Rate (FNR), PA, mean pixel accuracy (MPA), mean Intersection over union (MIoU), and frequency weighted intersection over union (FWIoU) were chosen to assess the accuracy of the proposed method. In all the metrics described below, it is assumed that there are a total of $k + 1$ classes (including background), then p_{ij} is the amount of pixels of class i inferred to belong to class j. Namely, p_{ii} means the number of true positives, p_{ij} and p_{ji} usually represent false positives and false negatives, respectively. In this paper, only one target is classified, namely, $k = 1$. Thus, p_{11}, p_{00}, p_{01}, and p_{10} represent true positive (TP), true negative (TN), false negative (FN), and false positive (FP), respectively.

(1) FPR

FPR simply computes a ratio between the amount of false positive classified pixels and the number of actual negative pixels, and it can be expressed as Equation (12).

$$FPR = \frac{FP}{FP + TN}, \tag{12}$$

(2) FNR

FNR simply computes a ratio between the amount of false negative classified pixels and the number of actual positive pixels, and it can be expressed as Equation (13).

$$FPR = \frac{FN}{TP + FN}, \tag{13}$$

(3) PA

PA simply calculates a ratio between the amount of properly classified pixels and the total number of them, and it can be expressed as Equation (14) [22].

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j}^{k} p_{ij}}, \tag{14}$$

(4) MPA

MPA is a slightly improved PA, which computes a ration of correct pixels based on class and then averages these over the total number of classes, and it can be expressed as Equation (15) [22].

$$MPA = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}}, \tag{15}$$

(5) MIoU

MIoU calculates a ratio between the intersection (the number of true positives) and the union (the sum of true positives, false negatives, and false positives) of two sets (the ground truth and the predicted segmentation), and it can be expressed as Equation (16) [22].

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}, \tag{16}$$

(6) FWIoU

FWIoU is an improved MIoU, and it can be expressed as Equation (17) [22].

$$FWIoU = \frac{1}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \sum_{i=0}^{k} \frac{\sum_{j=0}^{k} p_{ij} p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}, \tag{17}$$

3. Results

3.1. REO Mining Information Extraction Result from High-Resolution Remote Sensing Images

Figure 4 shows the results generated by the ITTI model described in Section 2.1. Figure 4(a1–a4) represent overall saliency maps of the study areas and Figure 4(b1–b4) are the salient regions in the form of binary maps. The overall saliency map was an average of the three saliency maps of intensity,

color, and orientation feature channel. Otsu's method [23,24] was used to automatically perform clustering-based image thresholding and then to reduce the saliency map to an initial binary map. Figure 4(c1–c4) are the NDVI data added as bound terms of energy function in the improved GrabCut model, Figure 4(a1–c1) are the results of the GF-1 image in Lingbei, Figure 4(a2–c2) are the results of the ALOS image in Lingbei, Figure 4(a3–c3) are the results of the GF-1 image in Shipai, and Figure 4(a4–c4) are the results of ALOS image in Shipai. Consequently, the extracted REO mining areas of the study areas can be achieved, as demonstrated in Figure 5, when the salient regions and NDVI data are, respectively, input into relative GrabCut models. Figure 5a is the extracted result of the GF-1 image in Lingbei. Figure 5b is the result of ALOS image in Lingbei, Figure 5c is the result of GF-1 image in Shipai, and Figure 5d is the result of ALOS image in Shipai.
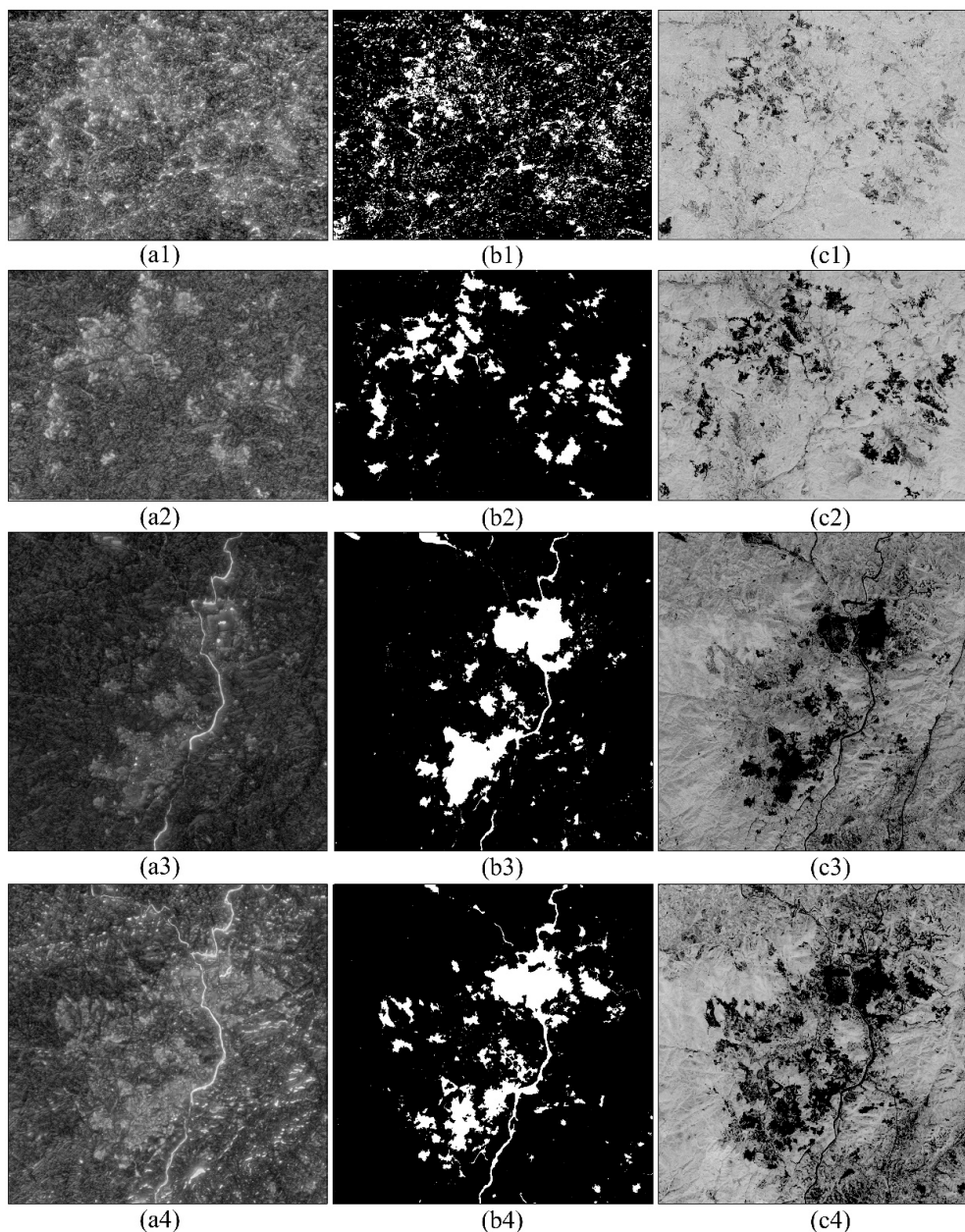

(a1)   (b1)   (c1)

(a2)   (b2)   (c2)

(a3)   (b3)   (c3)

(a4)   (b4)   (c4)

Figure 4. Saliency map and NDVI (Normalized Difference Vegetation Index). (**a1–c1**) Saliency map and NDVI of the GF-1 image in Lingbei; (**a1**) saliency map; (**b1**) salient region; (**c1**) NDVI; (**a2–c2**) saliency map and NDVI of ALOS image in Lingbei; (**a3–c3**) saliency map and NDVI of the GF-1 image in Shipai; (**a4–c4**) saliency map and NDVI of ALOS image in Shipai.
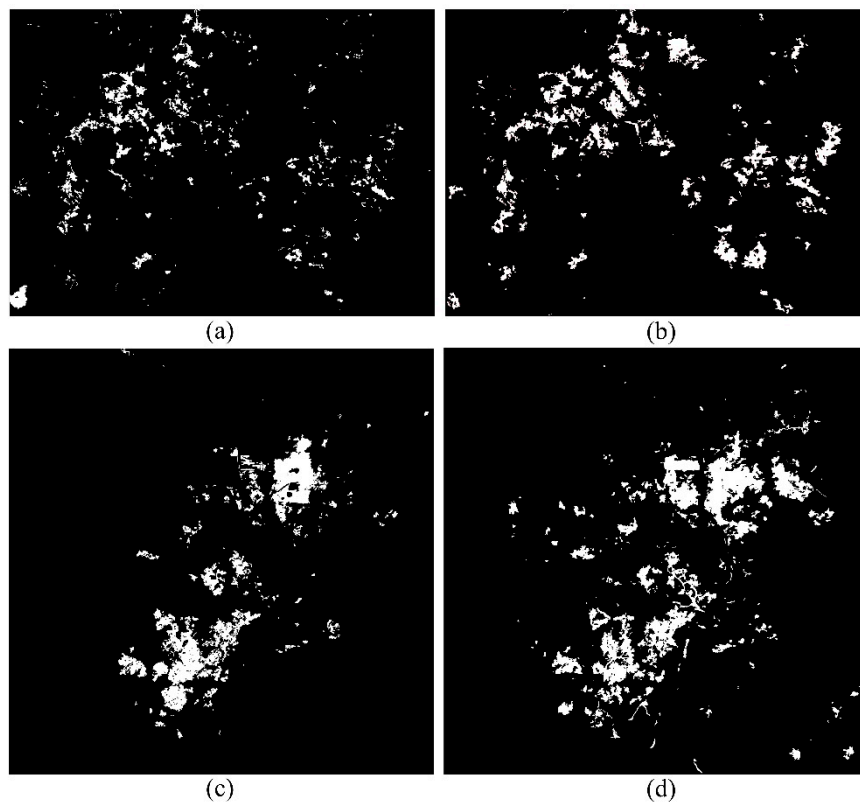
Figure 5. REO (Rare-earth ore) mining area extraction results. (**a**) REO mining area extraction result of the GF-1 image in Lingbei; (**b**) result of ALOS image in Lingbei; (**c**) result of the GF-1 image in Shipai; (**d**) result of the ALOS image in Shipai.

3.2. Precision Verification

In order to quantitatively test the precision of the experimental results, visual interpretations with Google Earth map and GF-1 images or ALOS images were conducted to extract the REO mining area. A field campaign was carried out in October 2018 to improve the visual interpretation result. In the field, the suspected REO mining areas were determined, and photos were taken with a digital camera for future reference (demonstrated as Figure 6). Considering the occurrence of the discrepancy between the field campaign and the acquisition of the remote sensing data, we further consulted regional experts for the land-cover changes in recent years to avoid possible errors. After the field campaign, the reference maps were finally produced for precision verification. For Lingbei, there are 1,044,681 pixels and 1,260,770 pixels for the mining area, and 30,129,969 pixels and 18,700,059 pixels for other land cover types, for the GF-1 and ALOS reference maps, respectively. For Shipai, there are 1,591,366 pixels and 1,474,408 pixels for the mining area, and 37,483,635 pixels and 23,525,592 pixels for other land cover types, for the GF-1 and ALOS reference maps, respectively.
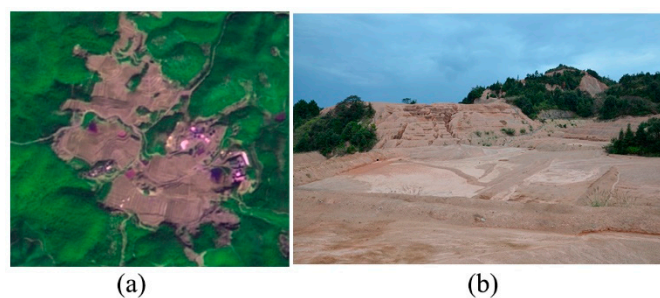

Figure 6. REO mining area field investigation photo. (**a**) GF-1 image; (**b**) REO mining area field photo.

3.2.1. Effectiveness Evaluation

To demonstrate the effectiveness of the improved GrabCut method, firstly, as usual, self-drawing rectangles (shown in Figure 2 as yellow rectangles) are set as the initial inputs of the original GrabCut method to extract the REO mining areas, and the extraction results are demonstrated in Figure 7(a1–a4). Then, salient regions (Figure 4(b1–b4)) generated by the ITTI model are regarded as the initial inputs of the original GrabCut method without adding NDVI data to extract the REO mining areas, and the results are shown in Figure 7(b1–b4). Finally, the two experimental results were compared with the results of the proposed improved GrabCut method (as shown in Figure 7(c1–c4)). It can be seen from Figure 7 that the normal GrabCut method could not be suitable for remote sensing image segmentation. Table 2 quantitatively lists six accuracy metrics of the three extraction methods for all experiments, and it shows that the six accuracy metrics of the normal GrabCut method were greatly worse than the other two methods. The MIoU reached up to 60% when salient regions were used as initial inputs of the original GrabCut method, the MIoU reached up to 90% in all segmentation results using the improved GrabCut method, and the FPR and FNR, respectively, were lower than 12.5% and 6.5%. In other words, accuracies of the GrabCut method with salient regions as initial inputs were greatly improved compared with the GrabCut method with self-drawing rectangles as initial inputs in four experiments, but they were still not satisfactory. The accuracies met our demands when NDVI was introduced as a bound term of the GrabCut method based on the salient region as the initial input. Therefore, it can be inferred that combining the visual attention mechanism with the image segmentation method greatly improves the segmentation result and the improved GrabCut method can greatly improve the accuracy of the REO mining area extraction result.

Table 2. Accuracy of REO mining area extraction with various methods in different study areas.

Areas	Methods	FPR	FNR	PA	MPA	MIoU	FWIoU
Lingbei GF-1	Normal GrabCut	93.4	4.0	54.3	74.4	29.7	51.2
	Salient region as initial	69.1	1.6	92.6	95.3	61.5	90.2
	The improved GrabCut	**9.1**	**4.9**	**99.5**	**97.4**	**93.2**	**99.1**
Lingbei ALOS	Normal GrabCut	91.6	1.3	31.8	63.0	17.9	26.2
	Salient region as initial	36.0	1.2	96.4	97.5	79.9	94.1
	The improved GrabCut	**4.6**	**6.5**	**99.3**	**96.6**	**94.4**	**98.6**
Shipai GF-1	Normal GrabCut	88.5	0.1	68.6	83.6	39.4	65.0
	Salient region as initial	61.9	0.1	93.4	96.5	65.6	90.9
	The improved GrabCut	**9.9**	**5.7**	**99.3**	**96.9**	**92.4**	**98.8**
Shipai ALOS	Normal GrabCut	85.9	2.0	64.7	80.3	38.3	59.7
	Salient region as initial	50.2	1.1	94.1	96.3	71.6	91.1
	The improved GrabCut	**12.5**	**5.1**	**98.9**	**97.0**	**91.2**	**97.9**

3.2.2. Comparison with Traditional Methods

In order to further test the performance of the proposed method, Object-oriented CART (Classification and Regression Tree) and Object-oriented SVM (Support Vector Machine), two commonly used information extraction methods in the high-resolution remote sensing community, were employed for a comparative study. The two methods were both carried out using the eCognition Developer 9.4 software. Features, such as spectral information, brightness, maximum difference (Max.diff), GLDV (Gray Level Difference Vector) texture, NDVI, and NDWI (Normalized Difference water Index), were used for SVM and CART classifiers to extract REO-mining areas by comprehensively analyzing the characteristics of REO-mining areas in remote sensing images. The detail parameters of the SVM and CART methods are listed in detail in Table 3, while the sample numbers of REO mining areas and non-REO mining areas are listed in Table 4. Finally, the classification results are shown in Figure 8. Extraction accuracy of the improved GrabCut was compared with the two different classification algorithms. Table 5 shows that differences of PA, MPA, and FWIoU between the improved GrabCut method and the other two methods were not significant, and the reason seems to be that the number of non-REO mining area pixels was about 30 times that of the REO mining area. MIoU values of the

two traditional methods were lower than 85%, obviously worse than the improved GrabCut method, and FPR and FNR values of the two traditional methods were apparently higher than the improved GrabCut method. In other words, all metrics of the improved GrabCut method outperformed that of SVM and CART classifiers in the four experiments. It can be convincingly stated that the accuracy of the proposed method is apparently better than the two traditional methods.
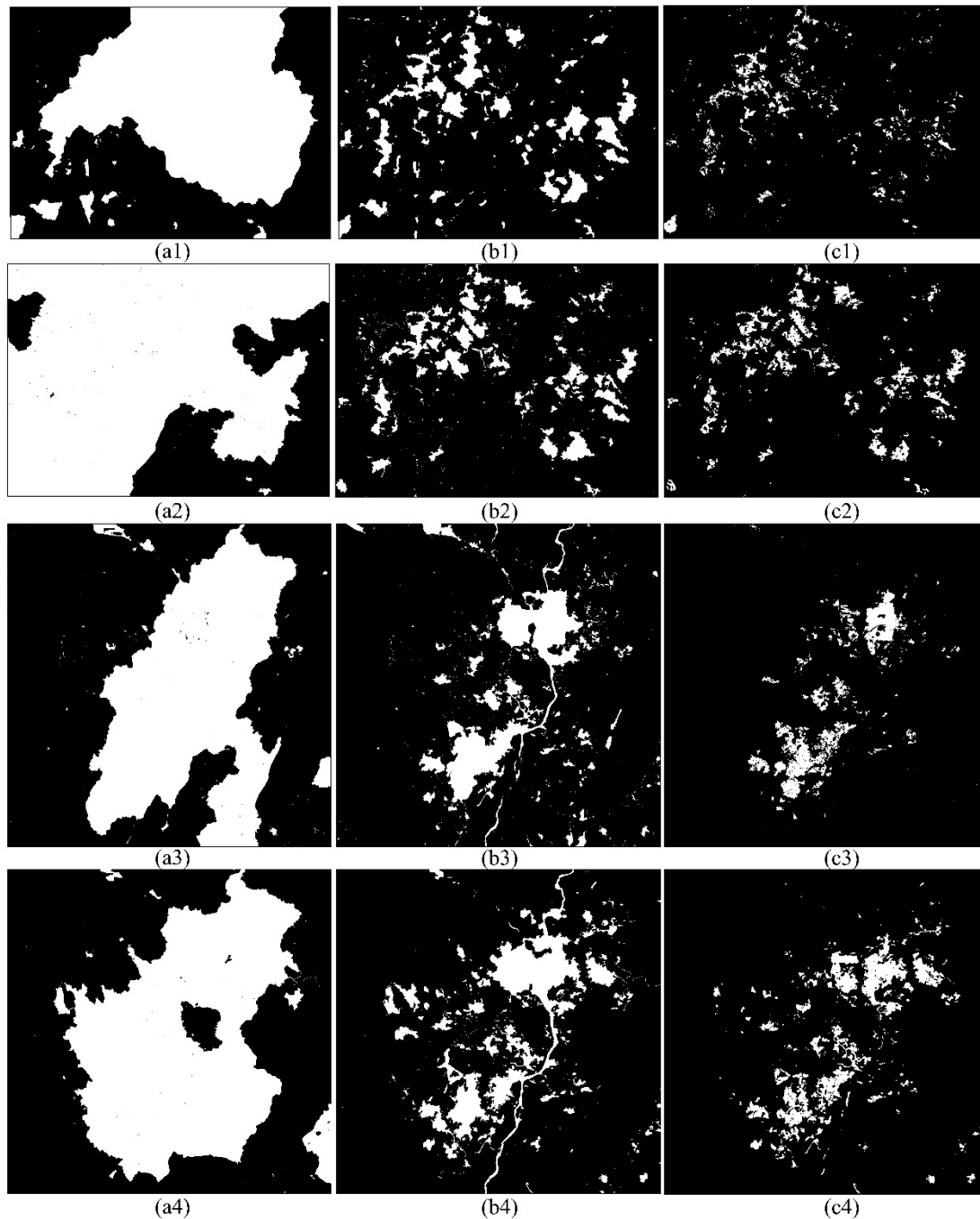

Figure 7. REO mining area extraction results with different methods. (**a1**–**c1**) Results of GF-1 images in Lingbei; (**a2**–**c2**) results of ALOS images in Lingbei; (**a3**–**c3**) results of GF-1 images in Shipai; (**a4**–**c4**) results of ALOS images in Shipai; (**a1**–**a4**) results of the normal GrabCut method with self-drawing rectangles as initial inputs; (**b1**–**b4**) results of the original GrabCut method with the salient region as the initial input; (**c1**–**c4**) results of the proposed method.

Table 3. Parameters of the SVM and CART methods.

	SVM		CART	
kernel type	linear		depth	0
c	2		max categories	16
gamma	0		cross validation folds	3
features	NDVI and (NDWI); Mean Blue, Mean Red, Mean NIR, Brightness, Max. diff; GLDV Entropy (all directions).		features	NDVI and (NDWI); Mean Blue, Mean Red, Mean NIR, Brightness, Max.diff.

Table 4. Sample numbers of the study areas for SVM and CART methods (unit: objects).

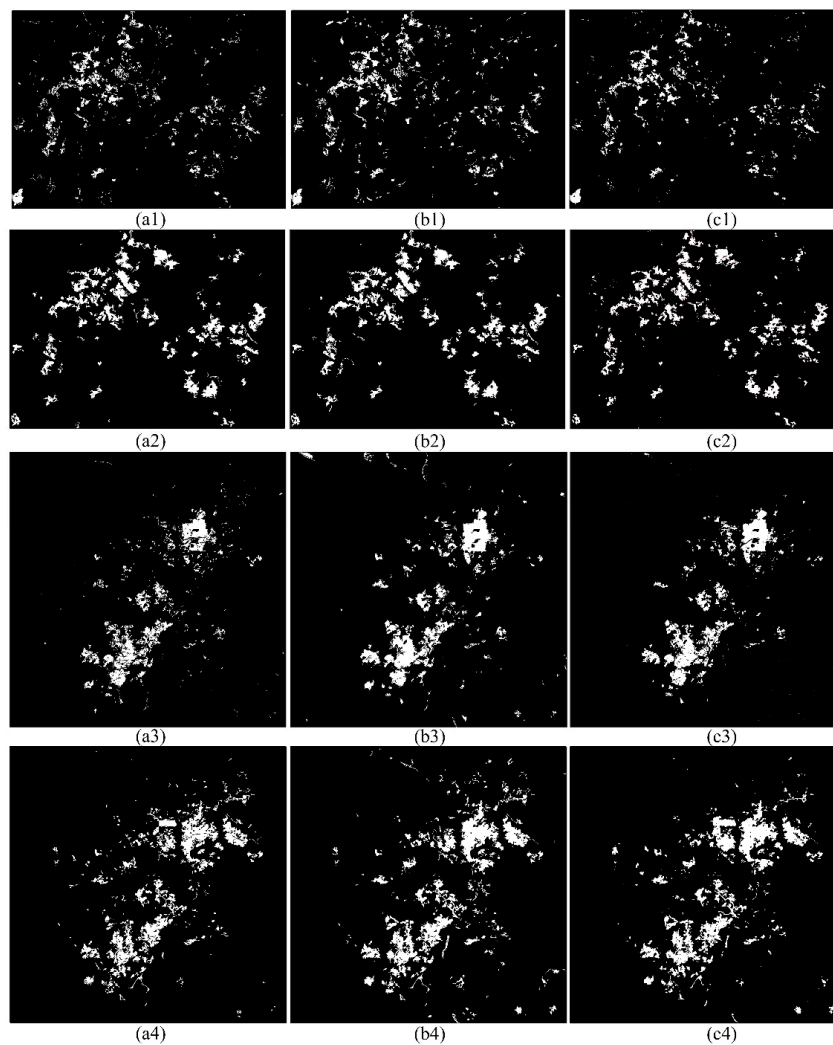| Study Areas	SVM		CART	
	REO	Non-REO	REO	Non-REO
Lingbei GF-1	76	138	76	138
Lingbei ALOS	76	132	76	132
Shipai GF-1	23	48	77	131
Shipai ALOS	40	109	40	109

Figure 8. REO mining area extraction results with different classification algorithms. (**a1–c1**) Results of GF-1 images in Lingbei; (**a2–c2**) results of ALOS images in Lingbei; (**a3–c3**) results of GF-1 images in Shipai; (**a4–c4**) results of ALOS images in Shipai; (**a1–a4**) object-oriented CART; (**b1–b4**) object-oriented SVM; (**c1–c4**) the improved GrabCut.

Table 5. Accuracy of REO mining area extraction with different algorithms.

Areas	Methods	FPR	FNR	PA	MPA	MIoU	FWIoU
	SVM	39.3	15.3	97.6	91.4	76.2	96.1
Lingbei GF-1	CART	28.2	15.4	98.4	91.7	80.9	97.2
	the improved GrabCut	**9.1**	**4.9**	**99.5**	**97.4**	**93.2**	**99.1**
	SVM	21.8	13.6	97.6	92.4	83.5	95.7
Lingbei ALOS	CART	21.1	13.9	97.7	92.3	83.8	95.8
	the improved GrabCut	**4.6**	**6.5**	**99.3**	**96.6**	**94.4**	**98.6**
	SVM	26.8	11.2	98.2	93.7	82.6	96.9
Shipai GF-1	CART	17.4	22.9	98.4	88.2	82.4	97.1
	the improved GrabCut	**9.9**	**5.7**	**99.3**	**96.9**	**92.4**	**98.8**
	SVM	20.5	11.3	97.9	93.6	85.0	96.4
Shipai ALOS	CART	15.9	20.8	97.9	89.1	83.3	96.1
	the improved GrabCut	**12.5**	**5.1**	**98.9**	**97.0**	**91.2**	**97.9**

4. Discussion

The original GrabCut model can fulfil the entire segmentation, generally using an initial and incomplete user-labelling manually drawn into a rectangle for a natural picture. However, it did not work for the high-resolution remote sensing image, which is multi-dimensional and highly complex. The segmentation result was significantly improved but not satisfactory, when the original GrabCut model used the salient region generated by ITTI visual attention model as initial. The experimental result was quite satisfactory when NDVI information was added to the GrabCut model as a bound term of energy function to reduce influences of vegetation. NDVI may be the most frequently used vegetation index in vegetative remote sensing analysis and applications. It has been proven to be a good indicator to distinguish vegetative surfaces from none vegetative surfaces, and also a highly sensitive parameter to represent vegetation growth status [25,26]. The experimental results indicate that the improved GrabCut model based on visual attention model can extract precise REO mining area information from high spatial resolution remote sensing image, and the whole process of REO mining area extraction was fully automatic, not relied on manual intervention.

Some facts can be discovered by comparing and analysing the extraction results. (1) As demonstrated in Figure 9, the object boundary with the improved GrabCut model more accurately coincided with the source satellite image than the two traditional methods, and reasons seem to be that some roads and reclamation areas were easily classified as REO mining areas (as yellow circles illustrated in Figure 9) using the two traditional methods and there were also some distinct missing error (as red circles shown in Figure 9) for the two traditional methods. (2) False extraction phenomena mainly lied in the partial impervious surface and partial reclaimed areas in the abandoned REO mining region, as exhibited in Figures 10 and 11. REO mining area is composed of digging area, leaching pools and higher-place ponds. Higher-place ponds are artificial buildings, and the spectral features are similar to the impervious surface, therefore some impervious surfaces are easily misinterpreted as the REO mining area. The partial reclaimed areas mistakenly identified as REO mining areas are usually places where the reclamation process has just begun, and economic forest (usually navel orange trees) has justly been planted in the abandoned REO mined areas. At the very beginning of reclamation process, the orange trees canopies are so small that the reclaimed areas are characterized by mined land in a remote sensing image, therefore it is difficult to distinguish these partial reclaimed areas from abandoned REO mined areas. Future advances in the high-resolution satellite remote sensing community, such as accurate spectral mixture analysis and machine learning technology, may be helpful to effectively distinguish REO mined areas from partial impervious and reclamation areas.
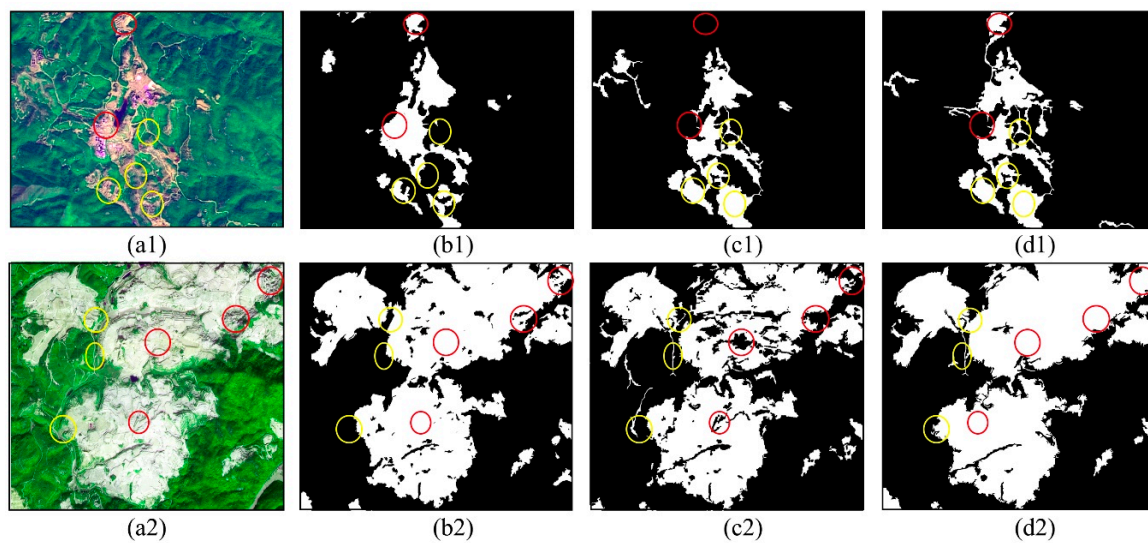
Figure 9. Extracted object contours with different extraction algorithms. (**a1**–**a2**) The source image; (**b1**–**b2**) the improved GrabCut; (**c1**–**c2**) CART; (**d1**–**d2**) SVM. Red circles represent missed regions in results of SVM or CART methods; yellow circles represent regions mistakenly classified as REO mining areas in results of SVM or CART methods.
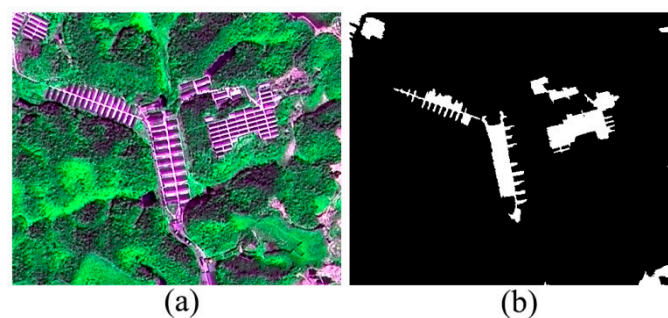

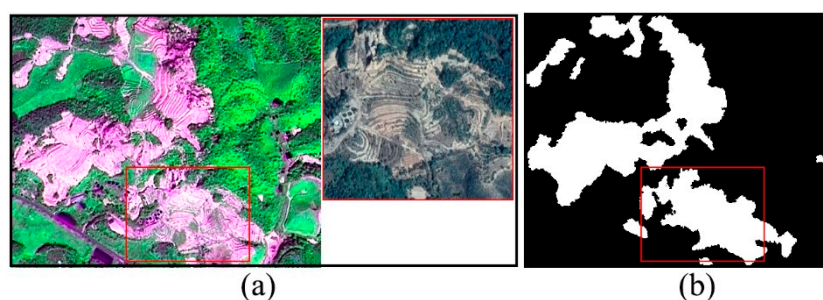**Figure 10.** Partial impervious surface mixed with REO mined area. (**a**) The source image; (**b**) experimental result.

Figure 11. Partial reclamation area mixed with REO mined area. (**a**) The source image, with the reclamation area framed in the red rectangle; (**b**) experimental result.

5. Conclusions

An improved GrabCut method based on a visual attention model is proposed in this paper to recognize REO mining areas from high-resolution remote sensing data, and the innovations mainly include two aspects. Firstly, the ITTI visual attention model was introduced to generate regions of interest quickly and automatically, and the salient region, instead of user interaction with labeled seed points, was employed as the initial input of the GrabCut model. Secondly, NDVI information was

added as a constraint term of the GrabCut Energy function, mainly to restrain vegetation and other non-REO information. Experimental results showed that:

1. Introducing the visual attention model to generate the salient region as the initial input of the GrabCut model made the extraction process fully automatic and improved extraction accuracy.
2. Adding NDVI information as the bound term of energy function achieved a higher precision than the original GrabCut model.
3. The proposed method outperformed the traditional CART and SVM methods.

Much work still remains to be done. For example, prior expert knowledge and time series NDVI data can be introduced to reduce false extraction phenomena, which mainly lie in the partial impervious surface and partial reclaimed areas in the abandoned REO mining region. Research in various mined areas and with more types of satellite images should be carried out to further test the performance of the approach proposed in this paper. Research in these directions should be conducted in the future.

Author Contributions: Conceptualization, Z.Z., Y.P., and G.H.; methodology, Y.P. and Z.Z. ; software, Y.P.; validation, Y.P. and M.W.; formal analysis, Y.P. and Z.Z.; investigation, Y.P. and Z.Z.; resources, Z.Z. and G.H.; data curation, Y.P. and M.W.; writing—original draft preparation, Y.P.; writing—review and editing, Z.Z., G.H., and Y.P.; supervision, Z.Z. and G.H.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Demirel, N.; Kemal Emil, M.; Sebnem Duzgun, H. Surface Coal Mine Area Monitoring Using Multi-temporal High-resolution Satellite Imagery. *Int. J. Coal Geol.* **2011**, *86*, 3–11. [CrossRef]
2. Zhang, Z.; He, G.; Wang, M.; Wang, Z.; Long, T.; Peng, Y. Detecting decadal land cover changes in mining regions based on satellite remotely sensed imagery: A case study of the stone mining area in Luoyuan county, SE China. *Photogramm. Eng. Remote Sens.* **2015**, *81*, 745–751. [CrossRef]
3. Song, X.; He, G.; Zhang, Z.; Long, T.; Peng, Y.; Wang, Z. Visual attention model based mining area recognition on massive high-resolution remote sensing images. *Clust. Comput.* **2015**. [CrossRef]
4. Karan, S.K.; Samadder, S.R.; Maiti, S.K. Assessment of the capability of remote sensing and GIS techniques for monitoring reclamation success in coal mine degraded lands. *J. Environ. Manag.* **2016**, *182*, 272–283. [CrossRef]
5. Yu, L.; Xu, Y.; Xue, Y.; Li, X.; Cheng, Y.; Liu, X.; Porwal, A.; Holden, E.; Yang, J.; Gong, P. Monitoring surface mining belts using multiple remote sensing datasets: A global perspective. *Ore Geol. Rev.* **2018**, *101*, 675–687. [CrossRef]
6. Prakash, A.; Gupta, R.P. Land-use mapping and change detection in a coal mining area—A case study in the Jharia coalfield, India. *Int. J. Remote Sens.* **1998**, *19*, 391–410. [CrossRef]
7. Kassouk, Z.; Thouret, J.; Gupta, A.; Solikhin, A.; Liew, S.C. Object-oriented classification of a high-spatial resolution SPOT5 image for mapping geology and landforms of active volcanoes: Semeru case study, Indonesia. *Geomorphology* **2014**, *221*, 18–33. [CrossRef]
8. Zeng, X.; Liu, Z.; He, C.; Ma, Q.; Wu, J. Detecting surface coal mining areas from remote sensing imagery: An approach based on object-oriented decision trees. *J. Appl. Remote Sens.* **2017**, *11*, 015025. [CrossRef]
9. Sun, W.; Zhao, H.; Jin, Z. A visual attention based ROI detection method for facial expression recognition. *Neurocomputing* **2018**, *296*, 12–22. [CrossRef]
10. Koch, C.; Ullman, S. Shifts in selective visual attention: Towards the underlying neural circuitry. *Hum. Neurobiol.* **1985**, *4*, 219–227. [CrossRef]
11. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]

12. Arbeláez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 898–916. [CrossRef]

13. Tosun, A.B.; Kandemir, M.; Sokmensuer, C.; Gunduz-Demir, C. Object-oriented texture analysis for the unsupervised segmentation of biopsy images for cancer detection. *Pattern Recognit.* **2009**, *42*, 1104–1112. [CrossRef]

14. Ning, J.; Zhang, L.; Zhang, D.; Wu, C. Interactive image segmentation by maximal similarity based region merging. *Pattern Recognit.* **2010**, *43*, 445–456. [CrossRef]

15. Rother, C.; Kolmogorov, V.; Blake, A. "GrabCut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **2004**, *23*, 309–314. [CrossRef]

16. Zhang, S.; Zhao, Y.; Bai, P. Object Localization improved GrabCut for Lung Parenchyma Segmentation. *Procedia Comput. Sci.* **2018**, *131*, 1311–1317. [CrossRef]

17. Liu, L.; Yu, X.; Ding, B. A Fast Segmentation Algorithm of PET Images Based on Visual Saliency Model. In Proceedings of the 2nd International Conference on Intelligent Computing, Communication & Convergence (ICCC-2016), Bhubaneswar, Odisha, India, 24–25 January 2016.

18. Khattab, D.; Theobalt, C.; Hussein, A.S.; Tolba, M.F. Modified GrabCut for human face segmentation. *Ain Shams Eng. J.* **2014**, *5*, 1083–1091. [CrossRef]

19. Salau, A.O.; Yesufu, T.K.; Ogundare, B.S. Vehicle plate number localization using a modified GrabCut algorithm. *J. King Saud Univ.-Comput. Inf. Sci.* **2019**. [CrossRef]

20. Zhang, C.; Hu, Y.; Cui, W. Semiautomatic right-angle building extraction from very graph cuts with star shape constraint and regularization. *J. Appl. Remote Sens.* **2018**, *12*, 026005. [CrossRef]

21. Boykov, Y.; Kolmogorov, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1124–1137. [CrossRef]

22. Alberto, G.G.; Sergio, O.E.; Sergiu, O.; Victor, V.M.; Jose, G.R. A Review on Deep Learning Techniques Applied to Semantic Segmentation. *Comput. Vis. Pattern Recognit.* **2017**, arXiv:1704.06857.

23. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Sys. Man. Cybern.* **1979**, *9*, 62–66. [CrossRef]

24. Sezgin, M.; Sankur, B. Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging* **2004**, *13*, 146–165. [CrossRef]

25. Zhang, Z.; He, G.; Wang, X.; Jiang, H. Leaf area index estimation of bamboo forest in Fujian province based on IRS P6 LISS 3 imagery. *Int. J. Remote Sens.* **2011**, *32*, 5365–5379. [CrossRef]

26. Meroni, M.; Fasbender, D.; Rembold, F.; Atzberger, C.; Klisch, A. Near real-time vegetation anomaly detection with MODIS NDVI: Timeliness vs. accuracy and effect of anomaly computation options. *Remote Sens. Environ.* **2019**, *221*, 508–521. [CrossRef]