electronics

MDPI

Article

Convolution Feature Inference-Based Semantic Understanding Method for Remote Sensing Images of Mangrove Forests

Shulei Wu [1,2,*], Yuchen Zhao [1], Yaoru Wang [1], Jinbiao Chen [3], Tao Zang [1] and Huandong Chen [1,2]

[1] School of Information Science and Technology, Hainan Normal University, Haikou 571158, China
[2] Hainan Provincial Key Laboratory of Ecological Civilization and Integrated Land-Sea Development, Haikou 571158, China
[3] Smart Police College, People's Police University of China, Langfang 065000, China
* Correspondence: wsl@hainnu.edu.cn

Abstract: The semantic segmentation and understanding of remote sensing images applying computer technology has become an important component of monitoring mangrove forests' ecological changes due to the rapid advancement of remote sensing technology. To improve the semantic segmentation capability of various surface features, this paper proposes a semantic understanding method for mangrove remote sensing images based on convolution feature inference. Firstly, the sample data is randomly selected, and next a model of convolution feature extraction is used to obtain the features of the selected sample data and build an initial feature set. Then, the convolution feature space and rule base are generated by establishing the three-dimensional color space distribution map for each class and domain similarity is introduced to construct the feature set and rules for reasoning. Next, a confidence reasoning method based on the convolution feature region growth, which introduces an improved similarity calculation, is put forward to obtain the first-time reasoning results. Finally, this approach adds a correction module, which removes the boundary information and reduces the noise from the results of the first-time reasoning as a new sample to correct the original feature set and rules, and uses the corrected feature set and rules for reasoning and understanding to obtain the final image segmentation results. It uses the corrected feature set and rules for reasoning and understanding to obtain the final image segmentation results. Experiments show that this algorithm has the benefits of a simple process, a short training time, and easy feature acquisition. The effect has been obviously improved compared to a single threshold segmentation method, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and other image segmentation methods.

Keywords: convolution features; semantic understanding; remote sensing image segmentation; semantic inference; feature rule base

check for updates

Citation: Wu, S.; Zhao, Y.; Wang, Y.; Chen, J.; Zang, T.; Chen, H. Convolution Feature Inference-Based Semantic Understanding Method for Remote Sensing Images of Mangrove Forests. *Electronics* **2023**, *12*, 881. https://doi.org/10.3390/electronics12040881

Academic Editors: Xin Ning, Achyut Shankar and Praveen Kumar Donta

Received: 30 December 2022
Revised: 31 January 2023
Accepted: 4 February 2023
Published: 9 February 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The intertidal zones of tropical and subtropical regions are home to mangrove trees, which are evergreen woody plants that can withstand high salt levels [1,2]. Mangroves play an important role in wind prevention, coastal stability, carbon sequestration, and other applications [3]. Mangrove forests in China have shrunk from 420 km^2 in 1950 to 220 km^2 in 2000 due to land reclamation for agriculture, urbanization, industrialization, and aquaculture [4–7]. In view of various laws and regulations issued by the Chinese government on wetland protection, it is crucial to track the ecological changes of mangroves. However, mangrove distribution data for extensive field measurement and sampling is complex due to the density of mangroves in intertidal zones and their periodic submergence by periodic seawater [8]. The widespread adoption of remote sensing technology has resulted in the use of satellite imagery for various environmental protection purposes, including the ecological change for mangroves [9–12].

In recent years, scholars have further explored deep learning research, which provides a highly positive effect for the semantic segmentation of remote sensing images and meets the accuracy requirements of computer vision applications [13–21]. The fully convolutional neural network was proposed in 2015 [22]. This method is modified based on a convolutional neural network to classify images at the pixel level. Bittner et al. [23] proposed a Fused-FCN4s model consisting of three parallel FCN4s networks to improve the convolution method. Three-band (R, G, B), panchromatic (PAN), and Normalized Digital Surface Model (NDSM) images are used as inputs to parallel networks to extract features from high-resolution remote sensing images. Chen et al. [24] proposed a symmetric FCN model, including symmetric normal fast FCN (SNFCN) and symmetric dense quick FCN (SDFCN) with fast connections. In terms of improving the effect of semantic segmentation, the asymmetric convolution net (ACNet) proposed by Hu et al. [25] in 2019 uses ResNet for feature extraction and sets the attention mechanism through the amount of information in the features at different levels, to achieve the balance between characteristics and the segmentation of effective regions. In 2019, Jun Fu et al. [26] proposed a new type of scene segmentation network, which creatively presented a dual attention module after the network consisted of the fully convolutional network and the hollow convolution. To cope with these factors, the proposed method by Guillaume et al. [1] combines deep learning-based enhancement of ITCs with a marker-controlled watershed segmentation algorithm.

Deep convolutional neural networks have achieved great success in many fields and have proven their superior performance in many applications in recent years [27]. McGlinchy explored the use of fully convolutional neural networks (FCNNs), specifically UNet [28], to map these complex features at the pixel level of high-resolution satellite images. This trend has also attracted many researchers to apply deep convolutional neural networks to the field of semantic segmentation of remote sensing images [29,30]. Based on the popular UNet model, by modification and adjustment Ge et al. [31] improved the prediction performance of forest parameters in the improved model.

Although the various FCN-based methods mentioned above have achieved remarkable performances in remote sensing image segmentation, their recognition relies heavily on large-scale datasets because millions of parameters in the network need to be trained [32]. In addition, recent studies have shown that the more profound the network, the better the performance of deep convolutional networks [33]. Unfortunately, as the number of neural network layers increases, gradient vanishing and gradient explosion problems may occur. The disappearance of gradients causes the model to fail to get updates from the training data. Gradient explosion can cause model instability, significant changes in loss during an update, or null model loss during training.

To solve the problem of insufficient labeled data, transfer learning, as a deep learning strategy [34], provides an effective way to train large networks with limited data without overfitting. Transfer learning can reduce the pre-training time and resource overhead of new models and can solve the problem of insufficient samples in new prediction problems. By fusing CNNs with models such as conditional random fields (CRF) and support vector machines (SVM) to create a new fusion model [35], Dong et al. tested an approach based on the fusion of RF classifier and CNN with ultra-high resolution remote sensing (VHRRS) [36]. To solve the problem of vanishing gradients and gradient explosions, Szegedy et al. [37] proposed ResNet with typical residual connections, which allows gradients to propagate flexibly by bypassing paths.

Many challenges remain when using a convolutional neural network for image semantic segmentation. For instance, pooling operation and long convolutional steps cause the detailed features of the image to be lost; the spatial position information in the image to not be used efficiently; and the high algorithm complexity and number of datasets large [38,39].

Various scholars are currently using the heuristic learning method of artificial intelligence to solve the uncertainty problem. Based on the expert system of evidence inference and generative rule base, confidence rule inference is a heuristic method based on artificial intelligence, which has been widely used and studied [40,41]. Loannou et al. [42] introduced

the confidence framework into the traditional IF-THEN rule expression, covering theories such as evidence theory, fuzzy set theory, and decision theory, and proposed a new inference method based on evidential inference (ER) of the belief rule base. Lin et al. [43] proposed an inference method which introduced the attenuation factor to correct the incomplete rule weights for incomplete data sets for the extended confidence rule base expert system. Liu et al. put forward a classification method based on the inference of the confidence rule base by introducing the linear combination of the confidence rule base [44]. Niu et al. proposed an algorithm based on importance and visual attention fusion confidence [45]. Charels et al. [46] proposed a target criterion for model confidence.

The above research focuses on acquiring feature sets and dividing similar classes in the inference process. However, it does not strictly consider the validity and credibility of feature sets.

For the segmentation of remote sensing images of mangrove forests, there are the following difficulties:

- Mangroves have less distribution, and remote sensing images are challenging to obtain, resulting in insufficient data set for classification;
- Mangrove reserves belong to the natural environment, and their distribution and growth area are irregular. Some parts are mixed with other shrubs, making distinguishing difficult;
- The remote sensing images of the interested areas show that the houses in residential areas are relatively small, the tidal flats and bare soil appear irregular, and the interspecific characteristics are very similar, such as oceans and rivers, mangroves, and other shrubs.

Hence, it is difficult for neural networks and traditional image inference methods to deal with the above problems. Although the use of neural networks can usually achieve better segmentation effects, a large enough data set to support the network model for training and verification is needed. Using methods such as knowledge graphs also requires a large amount of data to build a knowledge base. To solve the above problems, a semantic understanding of remote sensing images is proposed to construct the feature space of different ground objects in remote sensing images, and to build the mapping relationship between these features and different ground objects. Therefore, this paper put forward an approach based on convolutional feature inference-based semantic understanding, combining convolution transformation with confidence inference. The feature inference is interpreted as adopting the rule base for inference and prediction, by constructing the appropriate convolution kernel for the feature extraction to build the rule base, for inference and prediction.

The main contributions of this study are:

- This study proposes a new approach to solve the segmentation issue of mangroves in different regions using an improved convolution feature extraction model;
- A spatial confidence inference method is proposed for the regional growth of convolutional features. This method introduces an improved similarity calculation method to divide similar classes, to obtain the first-time reasoning results;
- To improve the semantic segmentation ability of various features, the proposed algorithm takes the first-time result as a new sample for boundary exclusion and noise reduction and builds the final feature space and rule base for inference by establishing a three-dimensional color space distribution map of each category and the similarity of the introduced domain.

2. Proposed Method

Convolution can extract image features with specific texture information and compresses images. Thus, convolution is introduced when processing the original sample images in this study. This paper proposes a semantic segmentation method for remote-sensing images based on convolutional feature confidence inferences to extract convo-

lutional features with certain texture information. The research method flow is shown in Figure 1.

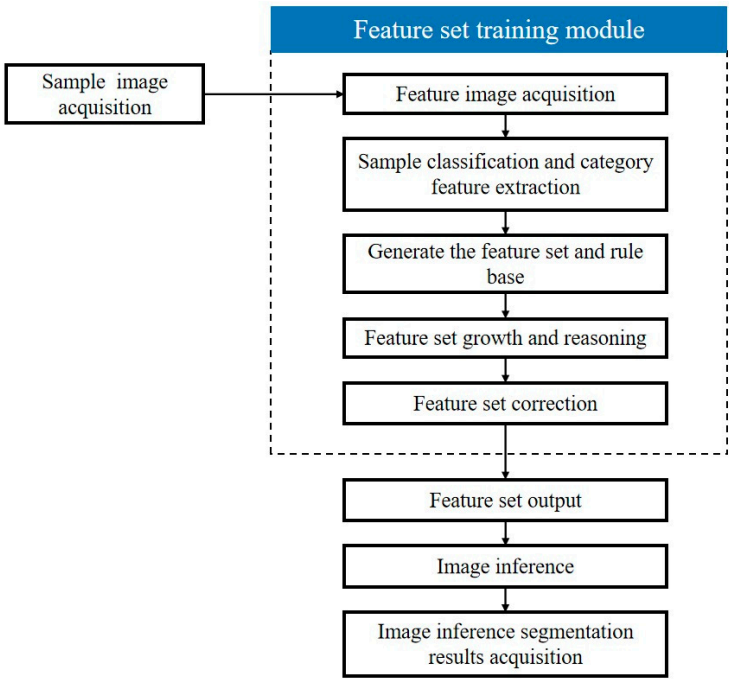


Figure 1. Flowchart of the proposed method.

Based on the small segmentation subgraph of high-resolution remote sensing images, this method extracts colors and textures in a certain way to construct an inference feature set. It then makes the feature set have different types of semantic comprehension and generalization abilities to obtain the segmentation result of the whole image through the semantic reasoning model. The detailed method of image semantic understanding, as described above, is shown in Figure 2.

This section will introduce the process, construction method, and theoretical analysis of the semantic inference model based on convolution features.

2.1. Convolution Feature Extraction

The collection of inference samples has a crucial impact on the inference results, and a reasonably prosperous and correct sample set determines the predictive ability of the inference model. Image inference requires a certain amount of data samples as a reference, and the corresponding feature sets and rules are set according to the sample characteristics, so that the reasoning model can grasp the relevant feature rules of the sample such that the model has the reasoning ability. This paper randomly extracts local remote sensing images as samples for feature extraction. A rule base of samples is generated, and the rule base is used to divide the entire remote sensing image inferentially. The process of the feature set training module is shown in Figure 3.

As is known, objects appear macroscopic in a remote sense image. Many entities are often concentrated in a particular area, and under most conditions present random and nonuniform distribution characteristics. Some large-scale, densely distributed objects occupy only a very small spot in remote sensing images. Therefore, the local image can contain abundant feature information. The color characteristics of the "ocean" class in the remote sensing image are taken as an example for analysis. The pixel color in the image is divided into four ranges from high to bottom by red, yellow, blue, and green, respectively. The distribution characteristics of its color can be observed, as shown in Figure 4.

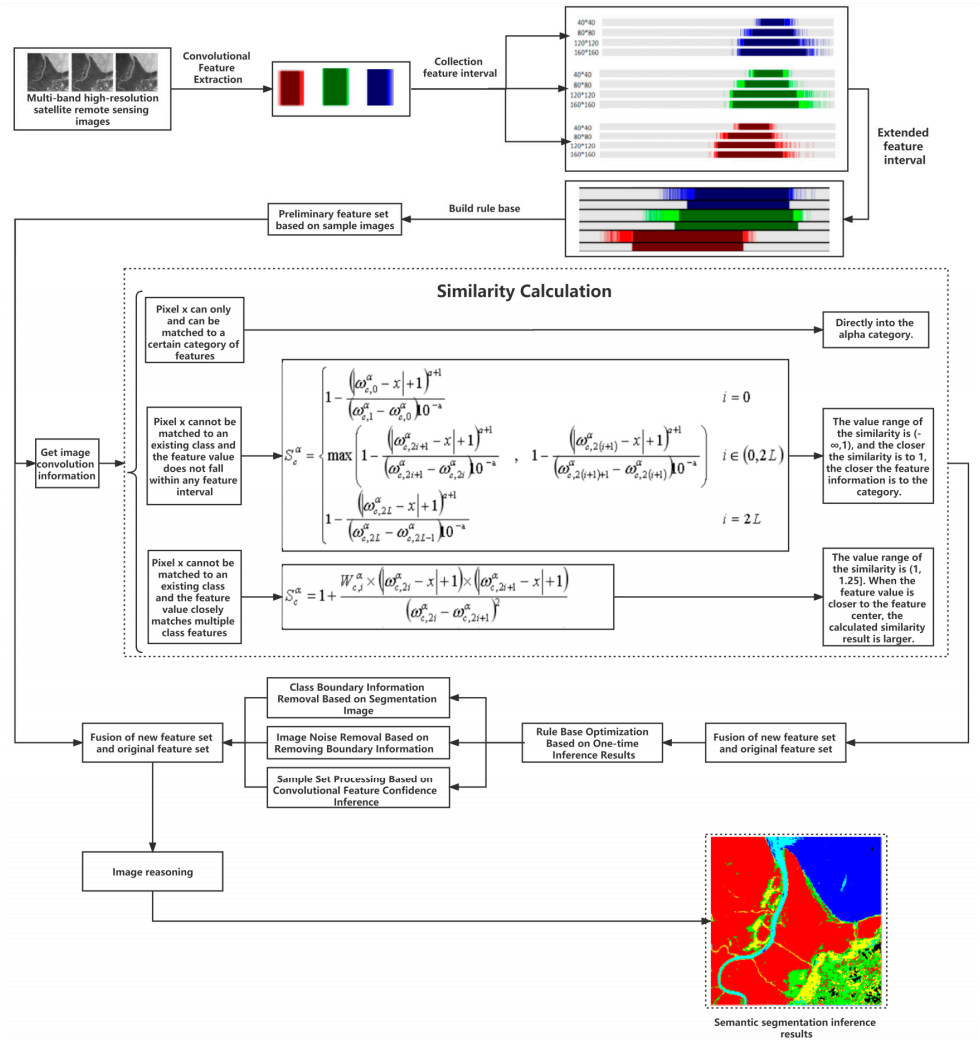


Figure 2. Flow of rule construction and semantic inference method.

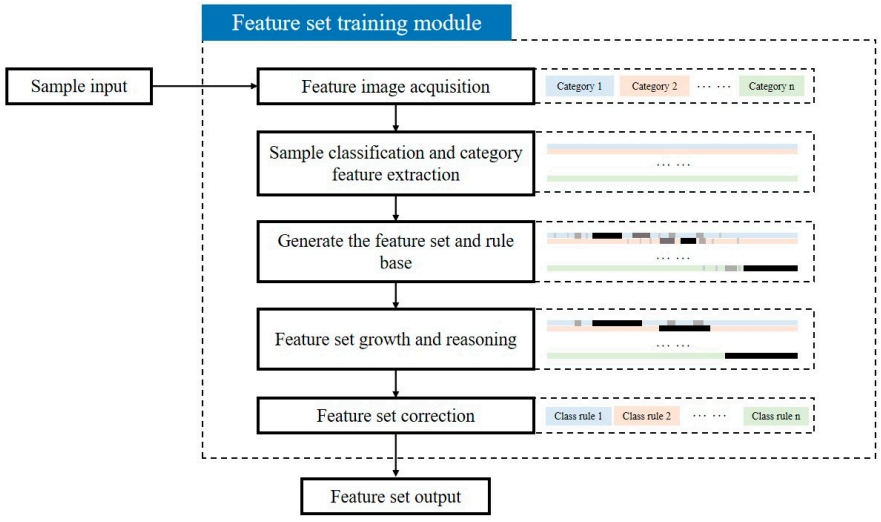


Figure 3. The flow of the feature set training module.

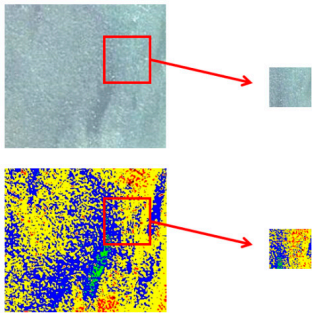


Figure 4. Sample of characteristics' distribution.

Notably, if the inference model masters the feature laws in the local feature image, it can master the critical part of the feature rules in a single category, and this model has preliminary feature inference ability. However, it is not easy to generate accurate segmentation results solely based on the color feature of the image itself to build an inference model. In object detection and image segmentation research in recent years, many scholars have used the idea of convolution in image processing. The convolution idea is introduced when processing the original sample image. When extracting convolutional features, it must be assumed that the input image is defined as $x(n)$ and thar the convolution kernel is defined as $h(n)$. The convolution operation can be expressed according to the formula below:

$$y(n) = \sum_{i=-\infty}^{\infty} x(i)h(n-i) = x(n)h(n) \tag{1}$$

In the above Formula (1), $y(n)$ is the result of a convolution transformation. It can be seen that in places with dense single-category pixels, the features tend to be relatively stable, while for areas between different categories, the features will show obvious changes. Therefore, using the convolution kernel should extract the image edge information better. The setting of the convolution kernel is variable and can be adjusted according to the different images. The segmentation results obtained by comparing different convolution kernels are shown in Figure 5.

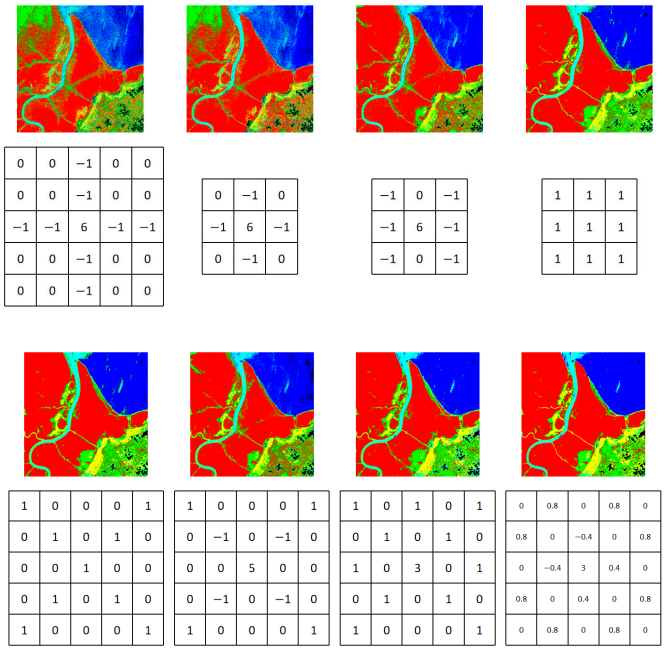


Figure 5. Different effects of different convolution kernels.

The 3 × 3 convolution kernel has fewer parameters and an insufficient sensing field, and the inference results are greatly affected by the category boundary information. Convolution kernels larger than 5 × 5 take much time in actual operations, affecting the inference speed. In the algorithm, a 5 × 5 convolution kernel is used, as shown in Figure 6, in which the parameters of the middle 1 × 1 part of the convolution kernel are set as large as possible to expand the feature range. The edge convolution kernel is introduced in the middle 3 × 3 part of the convolution kernel, and eight parameters are added to the outermost layer for collecting texture information around the pixel.

0	0.8	0	0.8	0
0.8	0	−0.4	0	0.8
0	−0.4	3	0.4	0
0.8	0	0.4	0	0.8
0	0.8	0	0.8	0

Figure 6. 5 × 5 convolution kernels.

Taking the original sample map as the input layer, the original sample map is convolved to obtain a set of feature values for the corresponding category. If conduct feature extraction for the image information of the channel c in the class α is conducted, the category feature value $\xi_{c,i}^{\alpha}$ is obtained. All categorical feature values can form a one-dimensional feature value table, including the size of each feature value and the weight of the feature value $w_{c,i}^{\alpha}$, which is the proportion of the feature value to the percentage of the total pixel value. Taking the mangrove class as an example, the extraction method of feature values is shown in Figure 7.

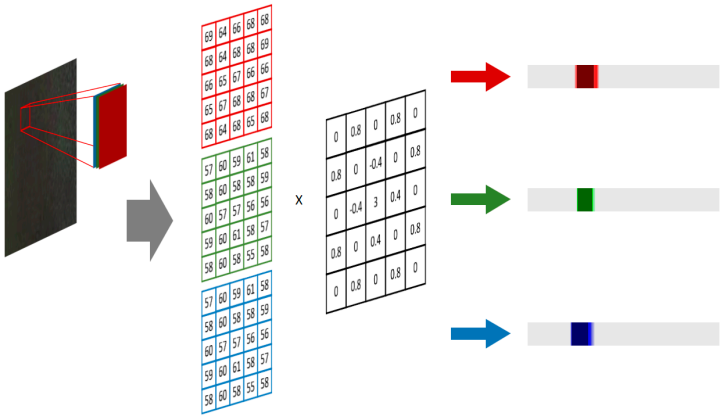


Figure 7. The extraction method of feature values.

The original sample is the category sample image as it is manually collected. Its three channels (RGB) are multiplied by the convolution kernel for feature extraction to obtain feature information, respectively. Statistics on the feature value distribution interval for each channel are employed to obtain a set of categorical features for that category, such as q_R^1, q_G^1, and q_B^1.

2.2. Semantic Rule Base Construction

The feature distribution interval of each category $\left[\omega_{c,2i}^{\alpha}, \omega_{c,2i+1}^{\alpha}\right]$ and distribution weight of each feature in each feature interval $W_{c,i}^{\alpha}$ can be obtained through statistics. All feature intervals make up a subset of category features q_c^{α}. The feature set Q^{α} corre-

sponding to a category contains subsets of its three color channels (R, G, and B), namely $q_R^1 \cup q_G^1 \cup q_B^1 = Q^\alpha$. The final feature set in the output is the collection of all feature sets of categories. The relationships among feature set, category feature set, category sub-feature set, and feature table are shown in Figure 8.

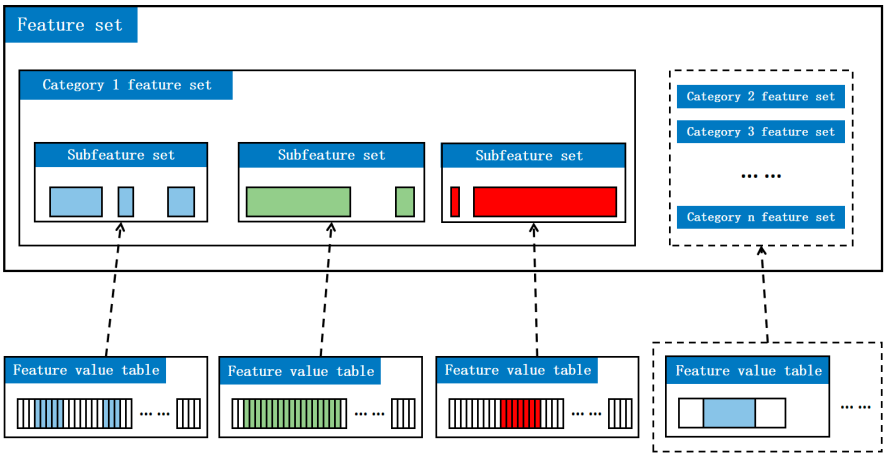


Figure 8. The relationship diagram.

The manually collected samples have important drawbacks wherein the lack of sample data results in the lack of features. If the sample is limited, then the part with a dense but discontinuous feature distribution is also very likely to be a feature of the actual model.

To solve this problem, under the condition that the feature information provided by the sample is certain, it is necessary to adjust the initially obtained feature set, that is, to expand the feature interval of each sub-feature set reasonably. The expansion limit M, which is a number between $(1, +\infty]$, is introduced to limit the feature interval's expansion range, indicating that the maximum number of the expanded interval can be transformed to a multiple of the original interval. Let the original interval length be l_a and the extended interval length is l_b, then:

$$M \geq \frac{l_b}{l_a} \tag{2}$$

The expansion limit is usually set within the range $(1, 1.2]$. If the expansion limit is too large and results in too significant changes to the feature set, the segmentation result will be unstable. Supposing there are n feature intervals for a subset of features, the weight of each feature interval $\left[\omega_{c,2i}^{\alpha}, \omega_{c,2i+1}^{\alpha}\right]$ is $W_{c,i}^{\alpha}$, and the extension limit is M. Then, the extension length L of the feature intervals is calculable by Equation (3):

$$L = W_{c,i}^{\alpha}\left(\omega_{c,2i}^{\alpha} - \omega_{c,2i+1}^{\alpha}\right)(M-1) \tag{3}$$

If the minimum value of the feature table is 0 and the maximum value of the feature table is H, then the result of the original feature interval after interval expansion can be expressed as Formula (4):

$$\left[\min\left(0, a - \frac{L}{2}\right), \max\left(H, b + \frac{L}{2}\right)\right] \tag{4}$$

After the initially extracted feature intervals are expanded in the above way, some feature intervals in the feature set will intersect. The intersection interval is prone to errors in the inference logic in the process of confidence inference, so it is necessary to process the feature interval that produces the coincident region.

After the feature interval is expanded, the information is extended to a certain extent. The feature interval that still accounts for a small proportion of the new sub-feature set is overwhelmingly likely to be an outlier interval and may not belong to its classified

features. If this part of the feature value is still retained for subsequent inference, then it will also have an impact on the part of the inference information. Therefore, a smaller feature interval in each subset must be removed to attenuate this effect.

2.3. Semantic Feature Inference

Feature interval expansion enriches the feature information to a certain extent, which can improve the problem of the insufficient data volume of samples. Although the feature set obtained based on the sample image can grasp part of the critical feature information of each classification, it will also miss more classification features. During semantic inference, this part of the uncollected features can easily lead to misjudgment. To improve the segmentation ability of the inference model, this paper uses the method of confidence inference to solve this problem.

Confidence rule inferences are the process of inference through data statistics, processing, and analysis in a given data or data set to obtain confidence rules. The conclusion of inferences is generated based on probability. Three parts are needed to represent a particular inference method: one is the feature sets, the second is the rules of reasoning, and the third is the result of reason. The problem that confidence inference needs to solve is that when inferential segmentation is performed on images, pixels that are not within the scope of feature set segmentation ability are inferred into the corresponding classification by inference.

To complete the inference task on unknown classification information, it is necessary to calculate the similarity of each image pixel's feature values to each category's features. There are roughly three cases in this inference process. In the ideal case, the feature value can be matched to a particular class of feature information. Supposing a pixel x contains the feature values x_R, x_G, and x_B of the three color channels R, G, and B, the pixel can be perfectly matched to the feature information of the category α to complete the inference, as shown in Equation (5):

$$\mathrm{x_B} \in \left[\omega^{\alpha}_{B,2i}, \omega^{\alpha}_{B,2i+1}\right] \wedge \mathrm{x_G} \in \left[\omega^{\alpha}_{G,2i}, \omega^{\alpha}_{G,2i+1}\right] \wedge \mathrm{x_R} \in \left[\omega^{\alpha}_{R,2i}, \omega^{\alpha}_{R,2i+1}\right] \tag{5}$$

If this happens, the feature value will be directly assigned to the class α. However, in addition to this situation, the feature values may not be ideally matched into any category of feature ranges, and other discrimination methods should be set for such pixels.

The feature value of a pixel should not be allowed to fall in any of the feature ranges of the feature set $q^{\alpha}(x \notin q^{\alpha})$ corresponding to channel c of the category α. Then, the similarity calculation method should be introduced to calculate the similarity of the feature value and the category to represent the degree of a match between the feature value and the category x_c. Assuming that the feature set has L standard feature intervals, and the feature value x is between the feature intervals $\left[\omega^{\alpha}_{c,2i}\ ,\ \omega^{\alpha}_{c,2i+1}\right]$ and $\left[\omega^{\alpha}_{c,2(i+1)}\ ,\ \omega^{\alpha}_{c,2(i+1)+1}\right]$, its class similarity S^{α}_{c} is calculable as follows:

$$\mathrm{S_c^{\alpha}}\begin{cases} 1-\frac{\left(\left|\omega^{\alpha}_{c,0}-\mathrm{x}\right|+1\right)^{\mathrm{a}+1}}{10^{-\mathrm{a}}\left(\omega^{\alpha}_{c,1}-\omega^{\alpha}_{c,0}\right)} & \mathrm{i}=0 \\ \max\left(1-\frac{\left(\left|\omega^{\alpha}_{c,2i+1}-\mathrm{x}\right|+1\right)^{\mathrm{a}+1}}{10^{-\mathrm{a}}\left(\omega^{\alpha}_{c,2i+1}-\omega^{\alpha}_{c,2i}\right)},\ 1-\frac{\left(\left|\omega^{\alpha}_{c,2(i+1)}-\mathrm{x}\right|+1\right)^{\mathrm{a}+1}}{10^{-\mathrm{a}}\left(\omega^{\alpha}_{c,2(i+1)+1}-\omega^{\alpha}_{c,2(i+1)}\right)}\right) & \mathrm{i}\in(0,2\mathrm{L}) \\ 1-\frac{\left(\left|\omega^{\alpha}_{c,2L}-\mathrm{x}\right|+1\right)^{\mathrm{a}+1}}{10^{-\mathrm{a}}\left(\omega^{\alpha}_{c,2L}-\omega^{\alpha}_{c,2L-1}\right)} & \mathrm{i}=2\mathrm{L} \end{cases} \tag{6}$$

where a is an adjustable parameter. The values of a can be used to magnify the difference between feature values and feature sets. The larger the value of a, the lower the similarity if the distance between the feature value and the feature interval is fixed. When processing different segmented images, the number of categories of image segmentation and the feature relationship between different types can be adjusted according to the image scale.

The range of similarity is $(-\infty, 1]$; the closer the similarity is to 1, the closer the feature information is to the category.

In addition to the above two cases, when segmenting an image, if the segmentation categories are too much or the features between the segmentation categories are too similar, the feature values of a pixel may closely match the characteristics of multiple categories. At this point, further reasoning and division of the input information of this class are required. The feature value x_c should fall in the feature interval $\left[\omega_{c,2i}^{\alpha}\ ,\ \omega_{c,2i+1}^{\alpha}\right]$ of the feature set q_c^{α} corresponding to the channel c in the category α. According to the previous definition of similarity, the similarity match degree of the feature value and the feature of the class is 1. Thus, according to the previous definition of similarity, the similarity match degree of the feature value and the quality of the type is 1. To determine which of the categories with similarity 1 is more closely related to this information, the similarity must be calculated according to the method of Formula (7), as follows:

$$\mathrm{S}_{\mathrm{c}}^{\alpha} = 1 + \frac{W_{c,i}^{\alpha}\left(\left|\omega_{c,2i}^{\alpha} - \mathrm{x}\right| + 1\right)\left(\left|\omega_{c,2i+1}^{\alpha} - \mathrm{x}\right| + 1\right)}{\left(\omega_{c,2i}^{\alpha} - \omega_{c,2i+1}^{\alpha}\right)} \tag{7}$$

The range of similarity $\mathrm{S}_{\mathrm{c}}^{\alpha}$ in this equation is $(1, 1.25]$. The closer the feature values are to the center of the feature, the greater the calculated similarity result. $W_{c,i}^{\alpha}$ is the weight of the feature interval in the feature set of its class. With the effect of $W_{c,i}^{\alpha}$, the similarity calculation formula will infer a more accurate inference result for the pixel corresponding to the feature value and each similar category based on the input feature to determine the classification of the pixel.

The similarity calculation result of each pixel's feature value in each channel in the category can be obtained by calculating the above method, and the similarity of the input information in the category can be obtained by adding the calculation results, namely:

$$\mathrm{S}_{\mathrm{x}}^{\alpha} = \sum_{\mathrm{c}=1}^{\mathrm{n}} \mathrm{S}_{\mathrm{c}}^{\alpha} \tag{8}$$

where n is the total number of channels. According to the above method, the similarity of the input information to each category can be obtained. The corresponding pixel can be classified into the category with the most remarkable similarity by comparing the similarity with each type to complete the inference.

To summarize the semantic feature reasoning in the above different cases, the semantic feature reasoning can be summarized in the following three points:

- If all feature values $\mathrm{x_c}$ of the pixel x belong to one category n only, then the reasoning result is category n;
- A feature value of the pixel x belongs to multiple categories $\mathrm{n_1}$, $\mathrm{n_2}$, $\mathrm{n_3}$ $\cdots$, and the similarity $\mathrm{S_c^1}, \mathrm{S_c^2}, \mathrm{S_c^3}$ of each category is calculated according to Formula (7);
- A feature value of the pixel x does not belong to any category, and the similarity $\mathrm{S_c^1}, \mathrm{S_c^2}, \mathrm{S_c^3}$ of each category is calculated according to Formula (6);
- x should be inferred as the largest similar class $\max\left\{\mathrm{S_x^1}, \mathrm{S_x^2}, \mathrm{S_x^3} \cdots\right\}$.

The above reasoning process is articulated with the "if … Then" sentences in the pseudocode as follows:

Input any pixel x:
if $\mathrm{x_B}$ in $\left[\omega_{B,2\mathrm{i}}^{\alpha}, \omega_{B,2i+1}^{\alpha}\right]$ and $\mathrm{x_G}$ in $\left[\omega_{G,2i}^{\alpha}, \omega_{G,2i+1}^{\alpha}\right]$ and $\mathrm{x_R}$ in $\left[\omega_{R,2i}^{\alpha}, \omega_{R,2i+1}^{\alpha}\right]$
then
    calculate $\mathrm{S_c^{\alpha}}$ in case one
else if $\mathrm{x_c^{\alpha}}$ in $\left[\omega_{\mathrm{c,2i+1}}^{\alpha}, \omega_{\mathrm{R,2(i+1)}}^{\alpha}\right]$
then

calculate S_c^{α} in case two
else if x_c^{α} in $\left[\omega_{c,2i}^{\alpha}, \omega_{R,2i+1}^{\alpha}\right]$
then
calculate S_c^{α} in case three

$$S_x^{\alpha} = \sum_{c=1}^{n} S_c^{\alpha}$$

$$x \in \max\left\{S_x^1, S_x^2, S_x^3 \cdots\right\}$$

2.4. Correction Module Based on the Results of First-Time Inference

The preliminary division of the segmented image is obtained through the above operations, but problems still exist in the feature set. On the one hand, the feature set is not rich enough. On the other hand, the segmentation ability of the image is poor. Therefore, further processing of the feature set is required to improve the segmentation ability of the model. Since the image segmented by the first inference also has a considerable number of incorrectly segmented pixels which will interfere with the construction of the feature set, it is therefore necessary that a reasonable training method is set to adjust the feature set. It is found that roughly three types of misjudgment areas affect the construction of feature sets after experiments. The first type is the boundary area between categories, and the feature performance in this region is unstable. The second type is a small area of noise information, which cannot be well reasoned into the corresponding correct classification due to the limitations of the original feature set. The third type, where the data being wrongfully segmented is fused with the category being wrongfully segmented, is complicated and challenging to eliminate. The main task of the correction module is the rule base optimization based on the results of first-time inference, which needs to be completed for feature set expansion with minimal error information, and its flow is shown in Figure 9.

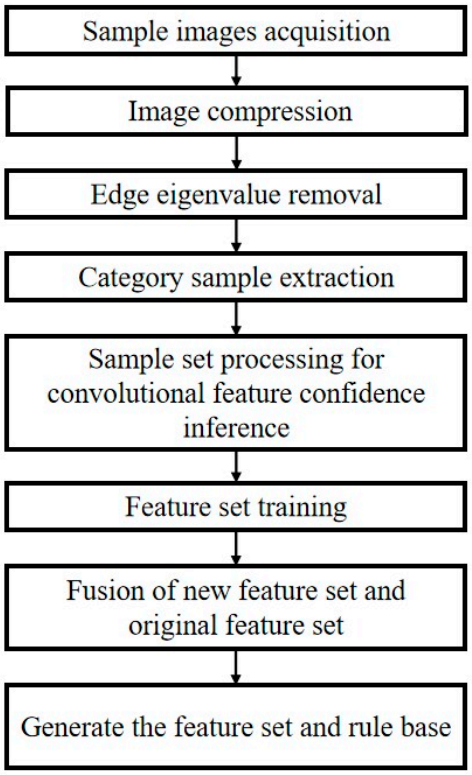


Figure 9. Optimization of the feature set.

2.4.1. Category Boundary Information Removal Based on Segmented Images

If the first segmented image is input as a new sample set and the first type of misjudged area, the boundary area between categories is processed first. The edge of each segmented class will intersect with other styles, affecting the feature performance. Hence, this part of the information needs to be removed when generating a feature set.

Therefore, this paper uses the mean convolution kernel for convolution, which compresses and then samples the image while ensuring that the features are not lost, meanwhile removing the edge information according to the convolution result. The mean convolution is equivalent to a filter for smoothing the image, if the convolution kernel size is KW, the convolution kernel is calculated as follows:

$$\mathrm{K} = \frac{1}{\mathrm{K_W^2}} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \tag{9}$$

2.4.2. Noise Information Removal

After removing the boundary information, the sample image cannot be used as an input sample for feature extraction because it still contains noise information interference. If the noise information is added to the feature construction process, it will impact the category features. Therefore, the image is first split according to the spatial position. Each independent region in the space is divided into a separate sample image $\mathrm{E_i^\alpha}$. The number of pixels in the sample image is I_i^α, and all the sample image sets of each category shape the sample set of the category $\mathrm{R}^\alpha = \left\{E_1^\alpha, E_2^\alpha, \cdots E_n^\alpha\right\}$. The noise information in the sample set is then processed, and the dropout parameters are set as D. Supposing the sample set of class α is processed at this time, when a particular class in the sample image satisfies Formula (10), the sample is considered a noise sample and discarded.

$$\mathrm{D} = \frac{\mathrm{E_i^\alpha}}{\sum_{\mathrm{j}=0}^{\mathrm{n}} \mathrm{E_j^\alpha}} \tag{10}$$

2.4.3. Sample Set Processing Based on Convolutional Feature Confidence Inference

Following the above method, the noise information in the small and boundary areas between categories was processed to obtain a sample set. However, the sample still exists the problem of the third type. Therefore, the sample set still cannot be used for feature extraction, and further processing is required.

Assuming that class α contains n sample images, the RGB channel values of the pixels in each sample image are concentrated in a particular area U_i^α in the color space. If most of the pixels in the sample image are correctly divided, then the area U_i^α will also be distributed stably. For the pixels of each class sample, their distribution in the color feature space is drawn as shown in Figure 10.

As shown in the figure above, each category of pixels are distributed in three-dimensional feature space and the front view (RB), side view (GB) and top view (RG) of this three-dimensional space are displayed. In the RG diagram, the abscissa is R, and the ordinate is G, and so forth. In the 3D diagram, the x-axis corresponds to feature channel R, the y-axis corresponds to channel G, and the z-axis corresponds to channel B. It can be seen that the pixel values of each category have a relatively concentrated distribution area and show more stable characteristics. Furthermore, a sample image in class α will have a higher coincidence and similarity with the area composed of other sample maps when there are fewer wrongly segmented pixels.

Conversely, supposing the region contains a portion of pixels that are not properly segmented, the area will be less coincident with the area formed by other sample maps in the class α, and this area will be discarded. Thus, a method is needed to determine the point distribution of each sample in the feature space.

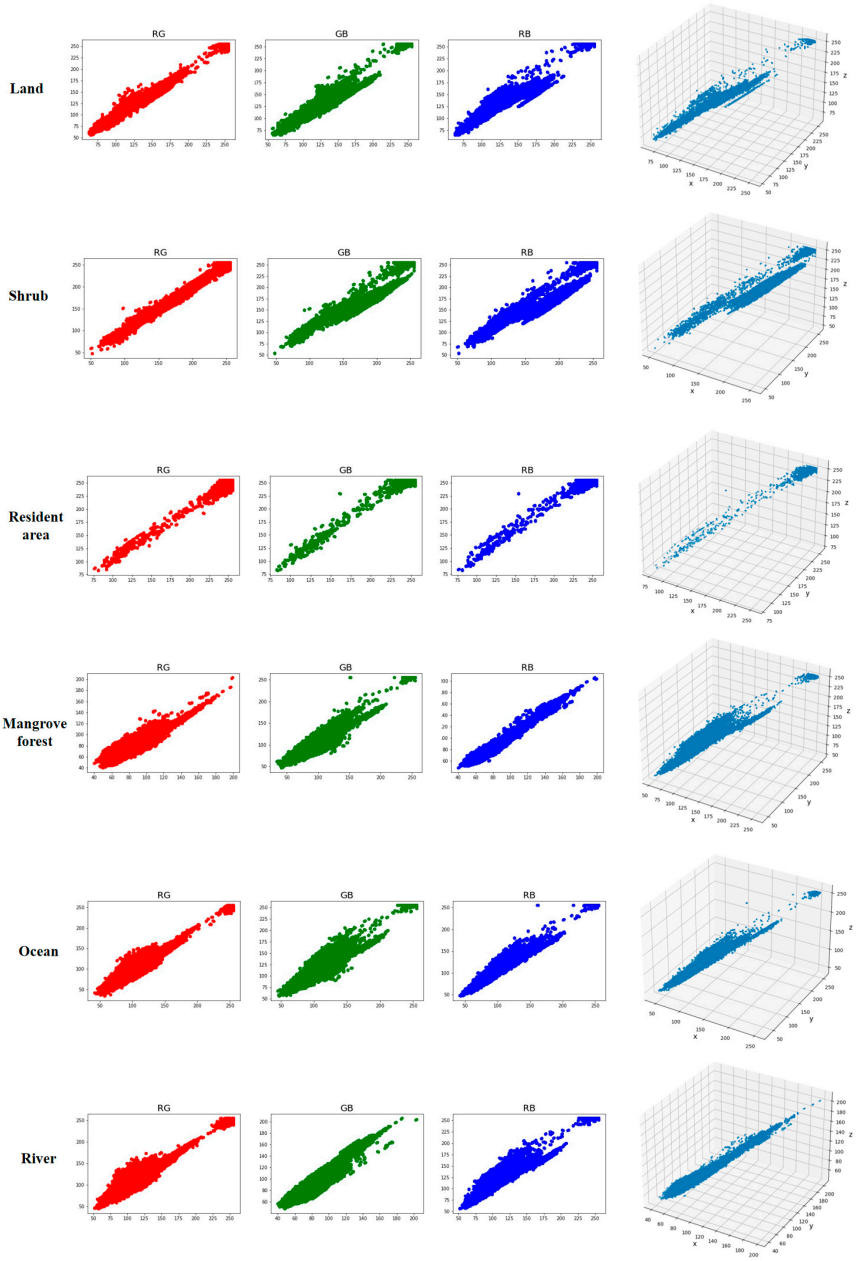


Figure 10. Distribution of color features.

Taking a pixel x of the ith sample in the category α, its values corresponding to the three channels are x_R, x_G, x_B. Then, a set of characteristics for this pixel x can be calculated according to the Equation (11):

$$\begin{aligned} \mathrm{K} &= \{\mathrm{k}_1, \mathrm{k}_2, \mathrm{k}_3, \mathrm{k}_4, \mathrm{k}_5, \mathrm{k}_6, \mathrm{k}_7, \mathrm{k}_8, \mathrm{k}_9, \mathrm{k}_{10}, \mathrm{k}_{11}, \mathrm{k}_{12}\} \\ &= \left\{ \frac{x_R}{x_G}, \frac{x_G}{x_B}, \frac{x_R}{x_B}, \frac{(\mathrm{R}-\mathrm{H})}{\mathrm{G}}, \frac{(\mathrm{G}-\mathrm{H})}{\mathrm{B}}, \frac{(\mathrm{R}-\mathrm{H})}{\mathrm{B}}, \frac{(\mathrm{H}-\mathrm{R})}{(\mathrm{H}-\mathrm{G})}, \frac{(\mathrm{G}-\mathrm{H})}{\mathrm{B}}, \frac{(\mathrm{R}-\mathrm{H})}{\mathrm{B}}, \frac{\mathrm{R}}{(\mathrm{G}-\mathrm{H})}, \frac{\mathrm{G}}{(\mathrm{B}-\mathrm{H})}, \frac{\mathrm{R}}{(\mathrm{B}-\mathrm{H})} \right\} \end{aligned} \tag{11}$$

where k_1 to k_{12} is the position of input information in RG, GB, RB, which are the slope of the line where the four vertices $(0,\ 0)$, $(0,\ 256)$, $(256,\ 256)$, and $(256,\ 0)$ are located (taking 256 instead of the maximum value of 255 prevents the denominator from being zero). In order to show the meaning of these features more clearly, $x_R = 50$, $x_G = 100$, and $x_B = 150$ can be assumed. The relationship between the feature information is shown in Figure 11.

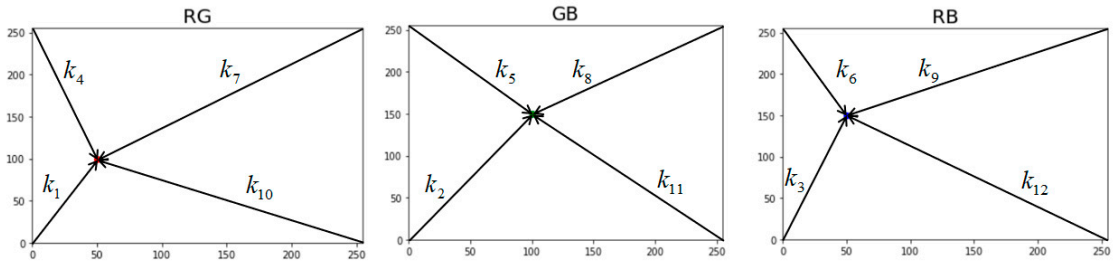


Figure 11. k1–k12 feature schematic.

The domain U_i^{α} of the ith sample in the class α eventually constitutes a space limited by the above conditions. The points in the sample are distributed within this spatial range. It is necessary to determine the supremum and infimum of each condition for space determination. The state of each threshold in U_i^{α} is as Equation (12) outlines:

$$U_i^{\alpha} = [(\mathrm{R_L}, \mathrm{R_R}), (\mathrm{G_L}, \mathrm{G_R}), (\mathrm{B_L}, \mathrm{B_R}), (\mathrm{k_{1L}}, \mathrm{k_{1R}}), (\mathrm{k_{2L}}, \mathrm{k_{2R}}), \cdots, (\mathrm{k_{12L}}, \mathrm{k_{12R}})] \tag{12}$$

where R_L is the lower limit of the channel R and R_R is the upper limit of the channel R. By analogy, the threshold with subscript L is the infimum of the condition, and the threshold with subscript R is the infimum of the condition. Then, after calculating the value of each condition for x, the upper and lower bounds of each condition for the domain U_i^{α} should be updated. After processing all of the pixels of the sample, the distribution domain U_i^{α} of the sample in the feature space is finally obtained.

Taking the sample image n and $n+1$ of the category α as an example, the features of two sample images are extracted to obtain the feature values $\zeta_{\mathrm{c,i}}^{\alpha,\mathrm{n}}$ and the feature value weights $\mathrm{W}_{\mathrm{c,i}}^{\alpha,\mathrm{n}}$ of each channel for each sample. Supposing that the characteristic intervals of two samples n and $n+1$ intersect, and each intersection interval is expressed with $[w_a, w_b]$, then the similarity of each two sample images is calculated according to Formula (13):

$$S_{(n,n+1)}^{\alpha} = \frac{\sum\sum_{j=\omega_a}^{\omega_b}\left(w_{c,j}^{\alpha,n} + w_{c,j}^{\alpha,n+1}\right)}{\sum_{k=0}^{L} w_{c,k}^{\alpha,n} + \sum_{m=0}^{L} w_{c,m}^{\alpha,n+1}} \tag{13}$$

where L is the length of the feature table. A decision threshold v is set to determine the minimum similarity between two samples. When the similarity between a sample domain and all other sample domains falls below this threshold, the anomaly segmentation value of the sample area is considered too high and needs to be discarded. On the contrary, if a sample is very similar to any other sample, then the information of the two samples is similar. The domains corresponding to the two samples need to be fused. Taking the minimum value of the infimum of features and the maximum value of the infimum of features in the two domains, that is, taking the union of the two domains to get a new domain that can contain the features of the two original domains, the result of processing according to the above description is the final sample set for training.

The sample set processed according to the above method has the first type of outliers removed, the second type of outliers, and part of the third type of outliers. The sample data obtained is larger and the category rules contained are richer. The newly obtained sample set when input into the model of rule extraction to get a new feature set, obtain a feature set which can used for reasoning segmentation. The feature set nevertheless contains a considerable amount of error information, which will have a great impact on the reasoning results. At this time, the newly obtained feature set must be fused with the initial feature set obtained from the first segmentation by optimizing and adjusting this new feature set based on the feature set obtained from the first training.

The adjustment of the feature set must expand each feature interval moderately, and the extension length needs to be adjusted in a certain way according to the situation of the interval. The interval weight, interval width, and the factors that segment the image need

to be comprehensively considered. Assuming that the characteristic interval of the channel for a category α is $\left[\omega_{c,2i}^{\alpha}, \omega_{c,2i+1}^{\alpha}\right]$, and the weight of the interval is $W_{c,i}^{\alpha}$, then the interval can be expanded according to Equations (14) and (15). Equation (14) is the length to extend.

$$L' = \frac{P \times W_{c,j}^{\alpha}}{\left(\omega_{c,2i+1}^{\alpha} - \omega_{c,2i}^{\alpha}\right)} \times 10^4 \tag{14}$$

The result of the expansion of the interval can be expressed by Equation (15):

$$\left[\min\left(0, \omega_{c,2i}^{\alpha} - L'\right), \max\left(\omega_{c,2i+1}^{\alpha} + L',\ L\right)\right] \tag{15}$$

where L is the length of the feature value. Thus at this stage, the operation of the correction module is complete, and the feature value table has been expanded.

3. Experiment Results and Analysis

3.1. Experiment Data

For the monitoring of mangroves in Hainan, this paper takes Guilinyang Mangrove Nature Reserve in Haikou City, Hainan Province, China, as an example to segment the ground features of the remote sensing image of the mangrove reserve. The experimental data are taken from three remote sensing images on 14 September 2010 (Data 1, worldview2), 4 December 2010 (Data 2, worldview2), and October 2020 (Data 3, spot6) respectively, as shown in Figure 12.



(**a**) Data 1 (**b**) Data 2 (**c**) Data 3

Figure 12. Remote sensing image used in the experiment.

3.2. Validation Parameters

Precision, also known as accuracy, represents the probability that a certain category in the prediction result is correct. Recall, also known as recall rate, represents the probability that a class in the real value is predicted correctly. IoU is usually used to evaluate the effect of the segmentation model, which is to calculate the intersection and union ratio between the prediction mask and the true mask. The values of image labeling results and prediction results can be divided into four categories: positive classes predicted as positive classes (TP), positive classes predicted as negative classes (FP), negative classes predicted as positive classes (FN), and negative classes predicted as negative classes (TN). Then, the calculation method of precision is in Formula (16). The higher the accuracy, the better the model has a segmentation effect on the category:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{16}$$

The recall rate is calculated as Formula (17), and the higher the recall rate, the better the model segments the class:

$$P = \frac{TP}{FN + TP} \tag{17}$$

The calculation of the intersection and union ratio can be expressed as the ratio of the intersection and union of a single category labeled image to a segmented image. Taking the mangrove category as an example, in the original image, we label the collection of areas where the mangrove forest is located in U_1. In the segmentation result obtained after the original image is imported into the segmentation model, the area where the mangrove forest is located is set as U_2. The intersection and union ratio of the mangrove category is calculated as follows:

$$IoU = \frac{U_1 \cap U_2}{U_1 \cup U_2} \tag{18}$$

3.3. Comparison and Analysis

3.3.1. Data 1

According to the idea of our proposed algorithm in this paper, the local data image is extracted from the original image as the sample used for training. Since some categories tend to not present a single feature form, the extracted feature image needs to be comprehensive and representative of the category. Otherwise, the segmentation model cannot be grasped. Taking the mangrove category as an example, four areas were extracted from the mangrove feature image, which were the areas of degradation due to geographical changes in the coastal part. Namely, degraded areas at the junction of mangroves and tidal flats or streams; areas where mangroves grow luxuriantly and densely; and areas where mangroves grow sparsely. The feature images of the remaining categories are extracted, as shown in Figure 13, for the rule generation of the segmentation model.

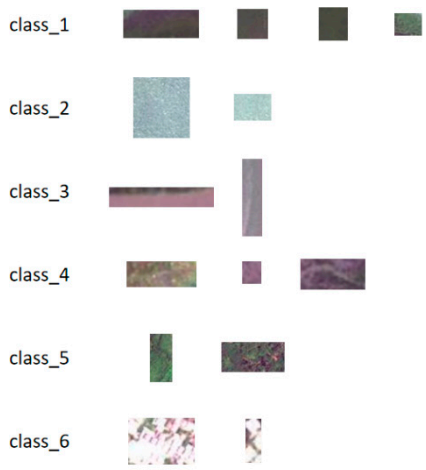


Figure 13. Different categories with classes labeled.

Among them, the object of category 1 is the mangrove forest, which is the target category of this paper. Category 2 is an ocean. Category 3 is a river. Due to river degradation resulting in a part of the riverbed being exposed, the middle and lower reaches of the river are more similar to the tidal flats, so the river and the sea are separated as distinct categories. The objects corresponding to category 4 are bare soil or tidal flats, including cultivated land mined by artificial mining, bare ground exposed by forest land degradation, or bare soil formed by river changes. The object corresponding to category 5 is a shrub, similar to the mangrove category, but its texture characteristics are different from mangroves. Category 6 is housing, containing central areas of human activity. For each category image, the correspondence between the category representation name and the category object is shown in Table 1:

Table 1. Semantically Segmented Categories and Category Names.

Category	Category Names
class_1	Mangrove forest
class_2	Ocean
class_3	River
class_4	Land, tidal flat
class_5	Shrub
class_6	House, residential area

The original sample set is obtained by collecting training samples from the remote sensing image. The preliminary segmentation model and feature set are obtained through the feature acquisition module, and the preliminary segmentation results are input into the correction model to expand the feature set to obtain the final feature set. The original image is reasoned based on the final feature set to receive the reasoning result. K-means clustering algorithm, KNN algorithm, GA genetic algorithm, and SVM algorithm were compared with the reasoning methods in this paper and the results are as follows.

When verifying the inference model, it is difficult to label the image because there are many shrubs and residential settlements in the lower right corner of the remote sensing image. So, this paper crops the original image and uses the part of a single image that can determine the category attribution as the verification image, which is used as a reference for evaluating the results of the model reasoning. The image inference results and evaluation parameters of the verification samples are shown in Tables 2 and 3:

Table 2. Test Results for IoU Parameter (Data 1).

Classification Name	GA	K-Means	KNN	SVM	This Paper
Mangrove forest	0.867	0.847	0.759	0.561	0.947
Ocean	0.779	0.957	0.62	0.729	0.929
River	0.909	0	0.141	0.079	0.905
Tidal flat, bare soil	0.672	0.635	0.03	0.513	0.838
Shrub	0.009	0.016	0	0	0.814
House, residential area	0.751	0.703	0	0.001	0.831

Table 3. Test Results for Accuracy (P) Parameter (Data 1).

Classification Name	GA	K-Means	KNN	SVM	This Paper
Mangrove forest	0.877	0.861	0.769	0.719	0.978
Ocean	0.804	0.957	0.689	0.729	0.929
River	0.921	0	0.154	0.258	0.941
Tidal flat, bare soil	0.693	0.692	0.098	0.529	0.91
Shrub	0.964	0.112	0	0	0.962
House, residential area	0.872	0.715	0	0.25	0.853

The output results of Data 1 are shown in Figure 14 and Tables 2 and 3. Intuitively, from the visual point of view, the experimental results of the segmentation in Figure 14 show that the algorithm proposed in this paper is relatively clear for the segmentation of various ground objects, namely mangroves, oceans, rivers, lands, and tidal flats. The mangrove part of the segmentation effect is especially good, and the river, tidal flat, and bare soil are also well segmented. There are many mixed shrubs and land residential areas in the lower right corner in Figure 14c, and it is difficult to label the image during preprocessing. It also affects the segmentation effect, but the segmentation effect is still better than other methods. In Figure 14a,b,d,e, for shrub and mangrove objects, and stream and ocean objects, which have very similar characteristics, it is difficult to separate them, thus resulting in unsatisfactory segmentation results. GA also mistakenly identified shrubs as mangroves. The SVM algorithm confuses and misjudges a small part of the mangroves,

bare soil area, and shrubs, while the KNN algorithm misjudges a large area of mangroves and shrubs as a tidal flat and bare soil.

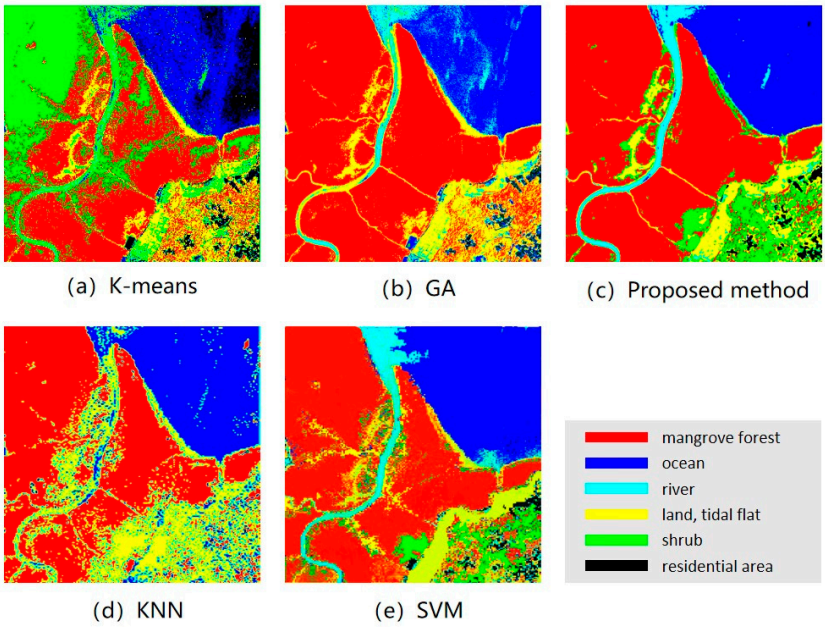


Figure 14. Output results for Data 1.

From the experimental index in Tables 2 and 3, the IoU index and accuracy (P) index of the algorithm in this paper reached 94.7% and 97.8%, respectively, which are superior indexes as compared to other algorithms. In other categories of data, this algorithm also achieves better results. However, the K-means algorithm cannot segment the river class, and its segmentation effect on shrubs is also very poor. Its accuracy rate is only 11.2%. KNN algorithm cannot recognize shrubs and houses, and its indicators are all zero. The SVM algorithm cannot recognize shrubs, the recognition effect of residential areas is poor, and the accuracy rate is only 25%.

3.3.2. Data 2

The output results of Data 2 are shown in Figure 15 and Tables 4 and 5. From the perspective of vision, according to the experimental results in Figure 15, the segmentation effect of the algorithm proposed in this paper is still superior. The mangrove part is clearly segmented. The oceans, rivers, and tidal flats are relatively clear, and a small number of rivers are divided into oceans. The K-means algorithm can segment the mangrove part well, but it is difficult to separate the rivers, oceans, and tidal flats, resulting in unsatisfactory segmentation results. GA also confuses rivers, oceans, and tidal flats. SVM mistakenly identified oceans as rivers. The KNN algorithm seriously confuses and misjudges mangroves, rivers, and oceans.

According to the experimental index in Tables 4 and 5, all the algorithms except KNN have good segmentation results for mangroves. Among them, the algorithm in this paper has the highest accuracy rate in the validation data set. For mangrove segmentation, its IoU value is 97.5%, and the accuracy rate is 99.9%. GA, KNN, and SVM are not good at distinguishing ocean and river, and their intersection and combination ratio is less than 50%. Although the split intersection ratio of ocean and river reaches 99.5% and 98.9%, respectively, in the validation set for the K-means algorithm, from the perspective of the image segmentation effect, it is not accurate in distinguishing between oceans and rivers.

3.3.3. Data 3

The output results of Data 3 are shown in Figure 16 and Tables 6 and 7. From the overall view of Figure 16, the segmentation effect of the algorithm proposed in this paper

is still the best. Mangroves, oceans, rivers, shrubs, residential areas, tidal flats, and bare soil are relatively clear. The other four algorithms can also segment the mangrove part well, but it is difficult to separate rivers, oceans, and tidal flats, resulting in unsatisfactory segmentation results. The K-means and GA algorithms misjudge a large number of tidal flats as residential areas and oceans, and residential areas and shrubs as mangroves. KNN misjudges bare soil and tidal flats as oceans and rivers, and SVM misjudges residential areas as rivers and oceans.

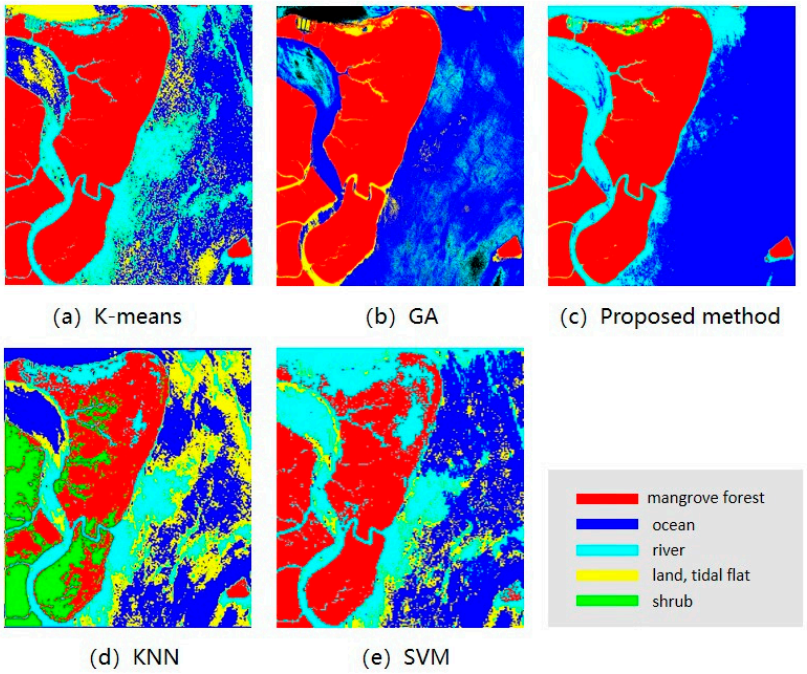


Figure 15. Output results for Data 2.

Table 4. Test Results for IoU Parameter (Data 2).

Classification Name	GA	K-Means	KNN	SVM	This Paper
Mangrove forest	0.975	0.971	0.682	0.772	0.975
Ocean	0.431	0.995	0.347	0.469	0.968
River	0.039	0.989	0.461	0.440	0.916

Table 5. Test Results for Accuracy (P) Parameter (Data 2).

Classification Name	GA	K-Means	KNN	SVM	This Paper
Mangrove forest	0.99	0.99	0.992	0.946	0.999
Ocean	0.434	0.907	0.622	0.972	0.987
River	0.85	0.718	0.644	0.546	0.926

According to indicators in Tables 6 and 7, the five algorithms have good effects on mangrove segmentation, but the highest accuracy of a proposed algorithm is 95.8%. The misclassification of this algorithm mainly focuses on the houses, shrubs, land and tidal flat categories. In addition to the algorithm in this paper, the accuracy of all kinds of algorithms for house segmentation is below 35%. Among them, the GA and K-means algorithm do not have the ability to segment land, tidal flat and shrubs. KNN algorithm can hardly distinguish between land and shrubs, while the SVM algorithm can distinguish land and tidal flats, but the accuracy of shrubs segmentation is 1.1%, and it also has little segmentation ability. In addition, the IoU value of the GA algorithm is 7.1%, and the effect is very poor.

In general, we can see from the results of this experiment that the algorithm in this paper also has a sound segmentation effect on three remote-sensing images of different devices in different periods.

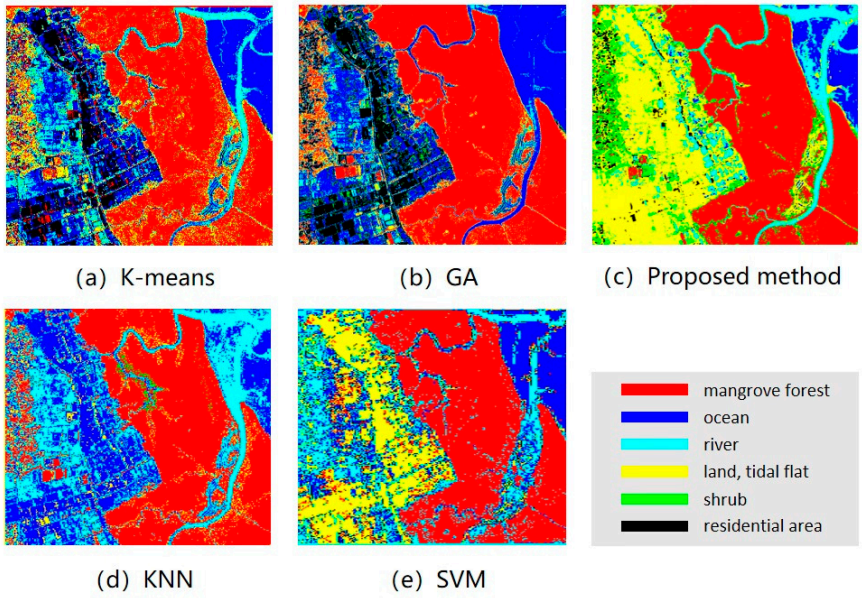


Figure 16. Output results for Data 3.

Table 6. Test Results for IoU Parameter (Data 3).

Classification Name	GA	K-Means	KNN	SVM	This Paper
Mangrove forest	0.943	0.954	0.57	0.952	0.958
Ocean	0.46	0.77	0	0.657	0.916
River	0.071	0.638	0.35	0.375	0.911
Tidal flat, bare soil	0	0	0.594	0.736	0.843
Shrub	0	0	0.135	0.011	0.829
House, residential area	0.34	0.009	0	0.093	0.781

Table 7. Test Results for Accuracy (P) Parameter (Data 3).

Classification Name	GA	K-Means	KNN	SVM	This Paper
Mangrove forest	0.956	0.962	0.966	0.98	0.98
Ocean	0.468	0.775	0	0.677	0.95
River	0.616	0.971	0.378	0.497	0.851
Tidal flat, bare soil	0	0	0.739	0.768	0.935
Shrub	0	0	0.143	0.36	0.944
House, residential area	0.34	0.011	0	0.35	0.972

4. Conclusions

Arming at mangrove ecological change monitoring in Hainan, China, semantic segmentation and semantic understanding of remote sensing images play an important role. In order to improve the semantic segmentation ability of various surface features, this paper proposes the semantic understanding method for mangrove remote sensing images based on convolutional feature inference. This method combines the spatial confidence inference method with the convolutional feature extraction model and adds the feature set and rule base correction process based on the primary inference results for the first time to solve the segmentation issue the remote sensing of mangroves. This method introduces an improved similarity calculation to divide similar classes, and it builds a unique feature set for inference by establishing a three-dimensional color space distribution map of each

category and the similarity of the introduced domain to improve the segmentation ability. Finally, this method adds a correction module, which removes the boundary information and reduces the noise from the result of the first segmentation as a new sample to correct the original feature set and rules and uses the corrected feature set and rules for reasoning and understanding to obtain the final image segmentation result. Experiments show that this approach has the advantages of the easy acquisition of a training feature set, short training time, and simple process. The effect improved compared with a single threshold segmentation method, KNN, SVM, and other image segmentation methods. Although the method proposed in this paper has obvious improvement, the segmentation effect of sea and river categories in the selected remote sensing images is poor. In the future, the combination of multi-band remote sensing images will be explored to improve the segmentation effect of category boundaries and small targets.

Author Contributions: Conceptualization, S.W.; Data curation, H.C.; Formal analysis, T.Z.; Funding acquisition, S.W.; Methodology, Y.Z.; Resources, H.C.; Validation, Y.W. and J.C.; Visualization, T.Z.; Writing—original draft, S.W. and Y.Z.; Writing—review & editing, Y.Z., Y.W. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No.61966013), the Hainan Natural Science Foundation of China (No.620RC602) and Hainan Provincial Key Laboratory of Ecological Civilization and Integrated Land-sea Development.

Data Availability Statement: The data reported in this study are included in the article.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Guillaume, L.; Matheus, P.F.; Laura, E.C.L.R.; Carlos, R.D.S.F. Deep learning-based individual tree crown delineation in mangrove forests using very-high-resolution satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *189*, 220–235.
2. Lee, W.K.; Tay, S.H.X.; Ooi, S.K.; Friess, D.A. Potential short wave attenuation function of disturbed mangroves. *Estuar. Coast. Shelf Sci.* **2021**, *248*, 106747. [CrossRef]
3. Han, H.; Huang, M.; Zhang, Y.; Bhatti, U.A. An extended-tag-induced matrix factorization technique for recommender systems. *Information* **2018**, *9*, 143. [CrossRef]
4. Bhatti, U.A.; Yu, Z.; Chanussot, J.; Zeeshan, Z.; Yuan, L.; Luo, W.; Mehmood, A. Local Similarity-Based Spatial–Spectral Fusion Hyperspectral Image Classification with Deep CNN and Gabor Filtering. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [CrossRef]
5. Wu, S.; Zhang, F.; Chen, H.; Zhang, Y. Semantic Understanding Based on Multi-feature Kernel Sparse Representation and Decision Rules for Mangrove Growth. *Inf. Process. Manag.* **2022**, *59*, 102813–102828.
6. Eddy, S.; Milantara, N.; Sasmito, S.D.; Kajita, T.; Basyuni, M. Anthropogenic drivers of mangrove loss and associated carbon emissions in South Sumatra, Indonesia. *Forests* **2021**, *12*, 187. [CrossRef]
7. Goldberg, L.; Lagomasino, D.; Thomas, N.; Fatoyinbo, T. Global declines in human-driven mangrove loss. *Glob. Change Biol.* **2020**, *26*, 5844–5855. [CrossRef]
8. Bhatti, U.A.; Yu, Z.; Yuan, L.; Zeeshan, Z.; Nawaz, S.A.; Bhatti, M.; Mehmood, A.; Qurat, U.A.; Zeehsan, Z.; Luo, W. Geometric algebra applications in geospatial artificial intelligence and remote sensing image processing. *IEEE Access* **2020**, *8*, 155783–155796. [CrossRef]
9. Wu, S.; Chen, H.; Bai, Y.; Zhu, G. A remote sensing image classification method based on sparse representation. *Multimed. Tools Appl.* **2016**, *75*, 12137–12154. [CrossRef]
10. Bhatti, U.A.; Huang, M.; Wu, D.; Zhang, Y.; Mehmood, A.; Han, H. Recommendation system using feature extraction and pattern recognition in clinical care systems. *Enterp. Inf. Syst.* **2019**, *13*, 329–351. [CrossRef]
11. Wu, S.; Bai, Y.; Chen, H. Change detection methods based on low-rank sparse representation for multi-temporal remote sensing imagery. *Clust. Comput. J. Netw. Softw. Tools Appl.* **2019**, *22*, 9951–9966. [CrossRef]
12. Shanmugapriya, P.; Rathika, S.; Ramesh, T.; Janaki, P. Applications of remote sensing in agriculture-A Review. *Int. J. Curr. Microbiol. Appl. Sci.* **2019**, *8*, 2270–2283. [CrossRef]
13. Hati, J.P.; Samanta, S.; Chaube, N.R.; Misra, A.; Giri, S.; Pramanick, N.; Gupta, K.; Majumdar, S.D.; Chanda, A.; Mukhopadhyay, A.; et al. Mangrove classifification using airborne hyperspectral AVIRIS-NG and comparing with other spaceborne hyperspectral and multispectral data. *Egypt. J. Remote Sens. Space Sci.* **2021**, *24*, 273–281.
14. Li, Q.; Wong, F.K.K.; Fung, T. Mapping multi-layered mangroves from multispectral, hyperspectral, and LiDAR data. *Remote Sens. Environ.* **2021**, *258*, 112403. [CrossRef]

15. Purwanto, A.D.; Asriningrum, W. Identifification of mangrove forests using multispectral satellite imageries. *Int. J. Remote Sens. Earth Sci.* **2019**, *16*, 63–86.
16. Abdel-Hamid, A.; Dubovyk, O.; El-Magd, A.; Menz, G. Mapping mangroves extents on the Red Sea coastline in Egypt using polarimetric SAR and high resolution optical remote sensing data. *Sustainability* **2018**, *10*, 646. [CrossRef]
17. Chen, N. Mapping mangrove in Dongzhaigang, China using Sentinel-2 imagery. *J. Appl. Remote Sens.* **2020**, *14*, 14508. [CrossRef]
18. Osei Darko, P.; Kalacska, M.; Arroyo-Mora, J.P.; Fagan, M.E. Spectral Complexity of Hyperspectral Images: A New Approach for Mangrove Classifification. *Remote Sens.* **2021**, *13*, 2604. [CrossRef]
19. Yassine, H.; Bhagawat, R.; Abhishek, T.; Abbes, A. Using artificial intelligence and data fusion for environmental monitoring: A review and future perspectives. *Inf. Fusion* **2022**, *86*, 44–75.
20. Xia, Q.; He, T.-T.; Qin, C.-Z.; Xing, X.-M.; Xiao, W. An Improved Submerged Mangrove Recognition Index-Based Method for Mapping Mangrove Forests by Removing the Disturbance of Tidal Dynamics and S. alterniflora. *Remote Sens.* **2022**, *14*, 3112. [CrossRef]
21. Liu, M.; Deng, H.; Dong, W. Identification of Mangrove Invasive Plant Derris Trifoliate Using UAV Images and Deep Learning Algorithms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 10017–10026. [CrossRef]
22. Wei, H.; Xu, X.; Ou, N.; Zhang, X.; Dai, Y. DEANet: Dual Encoder with Attention Network for Semantic Segmentation of Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 3900. [CrossRef]
23. Bittner, K.; Adam, F.; Cui, S.; Korner, M.; Reinartz, P. Building footprint extraction from VHR remote sensing images combined with normalized DSMs using fused fully convolutional networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2615–2629. [CrossRef]
24. Chen, G.; Zhang, X.; Wang, Q.; Dai, F.; Gong, Y.; Zhu, K. Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1633–1644. [CrossRef]
25. Hu, H.; Cui, J.; Wang, L. Region-aware contrastive learning for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021.
26. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual Attention Network for Scene Segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
27. Ghorbanian, A.; Ahmadi, S.A.; Amani, M.; Mohammadzadeh, A.; Jamali, S. Application of Artificial Neural Networks for Mangrove Mapping Using Multi-Temporal and Multi-Source Remote Sensing Imagery. *Water* **2022**, *14*, 244. [CrossRef]
28. McGlinchy, J.; Johnson, B.; Muller, B.; Joseph, M.; Diaz, J. Application of UNet Fully Convolutional Neural Network to Impervious Surface Segmentation in Urban Environment from High Resolution Satellite Imagery. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019.
29. Nawaz, S.A.; Li, J.; Bhatti, U.A.; Shoukat, M.U.; Ahmad, R.M. AI based Object detection latest trends in Remote sensing, Multimedia and Agriculture Applications. *Front. Plant Sci.* **2022**, *13*, 4273–4294. [CrossRef]
30. Chen, Q.; Huang, M.; Wang, H.; Zhang, Y.; Feng, W.; Wang, X.; Bhatti, U.A. A feature preprocessing framework of remote sensing image for marine targets recognition. In Proceedings of the 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO), Kobe, Japan, 28–31 May 2018.
31. Ge, S.; Su, W.; Gu, H.; Rauste, Y.; Praks, J.; Antropov, O. Improved LSTM Model for Boreal Forest Height Mapping Using Sentinel-1 Time Series. *Remote Sens.* **2022**, *14*, 5560. [CrossRef]
32. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Pohang University of Science and Technology, Seoul, Korea, 7–13 December 2015.
33. Zhou, X.; Takayama, R.; Wang, S.; Hara, T.; Fujita, H. Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. *Med. Phys.* **2017**, *44*, 5221–5233. [CrossRef]
34. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [CrossRef]
35. Moosaei, H.; Ganaie, M.A.; Hladík, M.; Tanveer, M. Inverse free reduced universum twin support vector machine for imbalanced data classification. *Neural Netw.* **2023**, *157*, 125–135. [CrossRef]
36. Dong, L.; Du, H.; Mao, F.; Han, N.; Li, X.; Zhou, G.; Liu, T. Very high-resolution remote sensing imagery classification using a fusion of random forest and deep learning technique—Subtropical area for example. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *13*, 113–128. [CrossRef]
37. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
38. Yi, Y.; Zhang, Z.; Zhang, W.; Zhang, C.; Li, W.; Zhao, T. Semantic segmentation of urban buildings from VHR remote sensing imagery using a deep convolutional neural network. *Remote Sens.* **2019**, *15*, 1774. [CrossRef]
39. Yuan, W.; Meng, C.; Tong, X.; Li, Z. Efficient local stereo matching algorithm based on fast gradient domain guided image filtering. *Signal Process. Image Commun.* **2021**, *95*, 116280–116284. [CrossRef]
40. Chen, X.; Zhao, Y.; Liu, C. Medical image segmentation using scalable functional variational Bayesian neural networks with Gaussian processes. *Neurocomputing* **2022**, *500*, 58–72. [CrossRef]
41. Mehrtash, A.; Wells, W.M.; Tempany, C.M.; Abolmaesumi, P.; Kapur, T. Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 3868–3878. [CrossRef]

42. Loannou, S.V.; Raouzaiou, A.T.; Tzouvaras, V.A.; Mailis, T.P.; Karpouzis, K.C.; Kollias, S.D. Emotion recognition through facial expression analysis based on a neurofuzzy network. *Neural Netw.* **2005**, *18*, 423–435.
43. Lin, Q.; Mao, R.; Liu, J.; Xu, F.; Cambria, E. Fusing topology contexts and logical rules in language models for knowledge graph completion. *Inf. Fusion* **2023**, *90*, 253–264. [CrossRef]
44. Liu, Y.; Palmedo, P.; Ye, Q.; Berger, B.; Peng, J. Enhancing evolutionary couplings with deep convolutional neural networks. *Cell Syst.* **2018**, *6*, 65–74. [CrossRef]
45. Niu, Y.; Zhang, S.; Wu, Z.; Zhao, T.; Chen, W. Image retargeting quality assessment based on registration confidence measure and noticeability-based pooling. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 972–985. [CrossRef]
46. Corbière, C.; Thome, N.; Saporta, A.; Vu, T.-H.; Cord, M.; Pérez, P. Confidence Estimation via Auxiliary Models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 6043–6055. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.