An Algorithm of Object Detection Based on Regression Learning for Remote Sensing Images

Jiajia Wang1,2, Ran Wu3* and Xiaqiong Yu4

1 Beijing Jinghang Research Institute of Computing and Communication, Beijing, 100076, China
2 Applied Mathematics Research Center of CASIC, Beijing, 100076, China
3 Beijing Air Traffic Control Engineering Technology Co., Ltd, Beijing, 100621, China
4 32021 Troops Beijing, 100094, China

*Corresponding author’s e-mail: appliedmath@163.com

Abstract. Saliency detection method is widely used in the field of object detection. It can greatly improve the efficiency of detection by extracting the region that may contain the objects and reduce the processing of background area. It is very difficult to detect the airports in remote sensing image for its changeable structure and complex background. Considering this situation, this paper proposes a object detection method of saliency model based on regression learning, this method extracts salient regions that may contain airports, and then detects salient regions through feature bag model. Our method uses super-pixel segmentation, minimization of absolute shrinkage and selection operator and hierarchical learning theory based on back-propagation. The salient regions that contain potential objects are extracted and sent to the feature bag model for classification, and it can location the final objects. This paper constructs the airport training dataset and test dataset, and the experimental results fully prove the effectiveness and superiority of this method.

1. Introduction

Object detection for Remote sensing images is a key research technology in the field of pattern recognition and computer vision. It detects the objects from the complex and diverse background in the Remote sensing image. If the object exists, it can be extracted by using relevant technology, and then it can be accurately located and analyzed. In recent years, visual saliency detection has been applied to the remote sensing objects detection and recognition. Regions of interest in remote sensing images are extracted to obtain regions containing potential targets as the next detection or recognition objects [1-2]. Through saliency detection, the background region is avoided to be detected, which greatly reduces the amount of calculation. The existing saliency detection methods are not ideal for remote sensing images with complex background. In addition, the airport structure in the remote sensing image is complex, and there are buildings similar to the airport in the background, which is relatively difficult for the airport detection, and the existing saliency model is difficult to accurately detect the airport. In view of this situation, this paper proposes an airport detection method based on regression learning salience model (RL). Based on the existing bottom-up data-driven saliency detection methods, this paper focuses on the research and implementation of saliency detection model by generating high-level features from bottom-up features. On this basis, a regression learning saliency model based on adaptive control and adaptive learning is proposed, which can accurately extract the salient region of the airport...
while maintaining the airport structure, so as to further judge the region. When the significant region is extracted by regression learning saliency model, SIFT feature is used to describe the region, and feature bag model is used to classify and detect the region. The airport detection results based on regression learning saliency model are obtained.

2. Regression learning saliency algorithm

This paper presents a regression learning saliency model of adaptive control and adaptive learning. The model uses the theory based on Reinforcement Learning [10], super-pixel segmentation algorithm, minimum absolute shrink and selection operator (Lasso) and hierarchical learning theory based on back-propagation to realize the extraction of salient regions. The model can decide the number of learning layers according to the given termination conditions.

2.1. Region feature extraction based on super pixel segmentation

In the regression learning saliency model, SLIC algorithm is used for image segmentation preprocessing, and super pixel is used to replace image sub region operation to reduce the computational complexity. For an input image \( I \) whose size is \( W \times H \), if the image is divided into \( k \) sub regions, the original image contains approximately \( k \) super pixels, and the set of super pixels is \( P = \{ p_1, p_2, \ldots, p_n \} \). Using the theory of background priori [6], the \( n \) super pixels close to the boundary in the image are extracted to form the background set \( B = \{ b_1, b_2, \ldots, b_n \}, 0 < n < k \). Next, we use the knowledge of graph theory to extract the boundary superpixels. Assuming that \( G = (V, E) \), \( V \) is the set of all the superpixels \( V = P \), and \( E \) is the set of boundary superpixels, we extract the boundary by an affine matrix \( W = (w_y)_{n \times k} \). For the boundary superpixels, a connected set can be formed, as shown in Figure 1. If the i-th superpixel is connected to the j-th superpixel \( w_{ij} = 1 \), and vice versa \( w_{ij} = 0 \), then most of the elements in the affine matrix are 0. Through the affine matrix, the set of super pixels \( B \) [at the boundary can be extracted [6].

![Figure 1. Affine Matrix.](image)

2.2. Region feature extraction based on super pixel segmentation

For the segmented image, the background set \( B = \{ b_1, b_2, \ldots, b_n \}, 0 < n < k \) is used to optimize all the superpixels, and the coefficients of the superpixels similar to the background set are shrunk, and the coefficients of the superpixels that are different from the background set are set to 0. The input image \( I \) after super-pixel segmentation contains \( k \) super pixels, and the set \( P = \{ p_1, p_2, \ldots, p_n \} \) is the background set \( B = \{ b_1, b_2, \ldots, b_n \} \) composed of \( n \) super pixels on the boundary

\[
\hat{\beta} = \arg \min \left\{ \sum_{i=1}^{n} \left( p_i - \beta b_i \right)^2 \right\}, \quad \text{s.t.} \ |\beta| \leq t
\]

(1)

Through the logical regression of all the superpixels sets \( P \) to the background set \( B \), the representation coefficient is reduced to obtain \( \hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k) \). At the same time, this is also the learning feature of image sub region corresponding to K superpixels calculated by lasso.

2.3. Layered learning theory based on back propagation

For the regression learning saliency model, according to the idea of back-propagation, the learning features calculated by lasso are used as the weight coefficients \( \hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k) \) of the secondary input
image to realize multiple learning and obtain the ideal detection results. Regression learning model for more than one learning, so that more intelligent judgment of the effectiveness of learning.

For the input image \( I \), the first time after SLIC super-pixel segmentation, the image \( I \) contains the number of \( k \) super pixels is, the total set is \( P \), the number of super pixels in the background set is \( n \), and the background set is \( B \). Use lasso to learn the set \( P \) and the background set \( B \) to get the learning coefficient \( \hat{\beta} \), and calculate the mean square error of the learning coefficient \( \overline{\beta} \), if it is less than the set threshold value \( T \). For the super pixels in the corresponding sub region, \( T_{\beta} \), the elements in the learning coefficient are used as weights to regenerate the secondary input image, as shown in formula (2)

\[
I_2 = I \ast (\hat{\beta})_{w=\hat{\beta}}
\]

At the same time, SLIC super-pixel segmentation is used again to make the number \( I_2 \) of super pixels is \( k_2 \) (\( k_2 < k \)), the set of all super pixels is \( P_2 \), and the corresponding background set is updated to \( B_2 \). Again, Lasso is used to learn the set \( P_2 \) and background set \( B_2 \) to get the learning coefficient \( \hat{\beta}_2 \). The mean square error of the learning coefficient \( \overline{\beta}_2 \) is calculated. If it is less than the set threshold value \( T_{\beta2} \), the learning coefficient \( \overline{\beta}_2 \) is output, otherwise the learning coefficient \( \hat{\beta}_1 \) is changed. Back propagation is applied to the input image \( I \). For the super pixels in the \( k \) corresponding sub region, the sub input image \( I_2 \) is reconstructed by taking the elements in the learning coefficient \( \hat{\beta} \) as weights

\[
I_3 = I_2 \ast (\hat{\beta})_{w=\hat{\beta}}
\]

In this way, the learning is stopped until the mean square error of the learning coefficient is less than the set threshold, and the learning coefficient learned by the current layer is output. In hierarchical learning, the threshold value of each layer to judge the end condition of learning is different, which can avoid the large error caused by using the unified threshold judgment. For the threshold to judge the learning condition, its size depends on the learned coefficient, where, the threshold is defined as its mean square error, as shown in the following formula:

\[
T_{\beta} = \frac{1}{k} \sum_{k} (\hat{\beta} - \bar{\beta})^2, \quad \bar{\beta} = \frac{1}{k} \sum_{k} \beta
\]

In the process of current layer learning, when the number of learning coefficients greater than the threshold value is greater than \( m \), the learning will terminate and all output coefficients will be output to generate saliency map. Assuming that the learning has been conducted for a total of \( t \) times, the final output learning coefficient is \( \hat{\beta}_t \).

2.4. salient region extraction of visual constraints

After the back-propagation hierarchical learning, the learning coefficient \( \hat{\beta}_t \) of the last layer reaches the learning termination condition, and the output after stopping learning is the first learning coefficient. At this time, in the learning coefficient, the characteristics of a total of \( m \) super pixels are greater than the mean square error of the whole learning coefficient, and the elements in the \( \hat{\beta}_t \) corresponding \( k \) super pixels, namely \( k \) sub regions. Each element value of \( \hat{\beta}_t \) is taken as the saliency score of the region, and the final saliency map \( S_{\text{output}} \) is obtained by normalizing the image \( I_t \), as shown in the following formula:
\begin{equation}
I_t = \left( \hat{\beta}_t \right)_{w_{\text{eff}}}
I_t = 1 - I_t
S_{\text{output}} = (I_t - \min(I_t)) / (\max(I_t) - \min(I_t))
\end{equation}

The final output $S_{\text{output}}$ is the final saliency map detected by regression learning saliency model.

3. **Airport detection driven by regression learning saliency algorithm**

After the salient region of the image is extracted by the regression learning saliency model, the salient region is extracted and matched to further determine whether the salient region is the target. The regression learning saliency model is used to predict the airport target, and the results are good, and most of the airport target positions can be predicted accurately.

3.1. **Method flow**

The method flow: Firstly, the codebook is generated by using the samples in the training sample library, which includes two categories: target and background; (2) the saliency of the input image is detected by using the regression learning saliency model, and the saliency region containing potential airport target is extracted; (3) the extracted saliency region is processed to generate the feature bag model for classification; (4) the airport detected after classification is classified The target is marked in the original image and the detection is completed. The specific flow chart is shown in Figure 2.

3.2. **Construction of remote sensing image training sample set and test data set**

This section creates a training sample set and a test data set. The training sample set contains 60 positive and 60 negative samples, named TDA (training dataset of airports), The test data set ad (Airport dataset) contains 140 images. The images in the data set are obtained from Google Earth. The visual altitude is 5-20km. The images are obtained under various visual angles, visibility and light changes. The location of the airport is random, and the size is different, and the background is diverse. The sample images are shown in Fig. 4 and Fig. 5 respectively.
3.3. experimental parameter setting

The image size of the test data set is $w = 1024$, $H = 590$. In the process of hierarchical learning, when SLIC is used to segment the image, the number of super pixels is determined by the nature of the image when learning $t$ times, which only needs to satisfy $K_1 > K_2 > \cdots K_t$. The number of super pixels $[K_1, K_2, K_t] = [200, 150, 100, 80, 50, \ldots]$. In section 4.3.3, if $M = 3$ is set, it is considered that there may be three significant regions containing potential targets in an image. For TDA in training sample set, the selection of codebook size generated by feature extraction of positive and negative samples is determined by the experimental results, and the highest detection value is 150. The specific experimental results are shown in Figure 6.

3.4. airport test results

For the test data set driven by regression learning saliency model, the accuracy and false detection rate of airport target detection results in AD are taken as the standard of target detection efficiency. Compared with GBMR model [7], GC model [8] and LC model [9], the method of feature bag detection is used, and compared with the method of SVM classifier using RL method. The saliency model of regression learning is an algorithm based on the structural characteristics of the airport, so the effect of airport detection is far better than the existing saliency model, as shown in the table. As shown in Figure 1, compared with the LC model with the best airport detection effect, the detection accuracy of the significance model of regression learning is increased by 27.76%, and the false detection rate is reduced by 27.51% compared with the GBMR model. Due to the limited number of images in the ad database and the high similarity between the buildings in the image and the target structure, the false detection rate is generally high. Similarly, the effect of the feature bag model as a classifier is much better than that of the SVM classifier, with the detection accuracy increased by 20.79% and the false detection rate reduced by 68.94%. Because the number of airports in the airport training sample set is small, the SVM classifier is not suitable, and the SIFT features used in the feature bag model can still produce a large number of SIFT feature descriptors [10]. For classification, less samples has little effect, so in the experiment, feature bag model is used to classify the significant regions extracted by significance model. The final experimental contrast effect of airport detection is shown in Figure 6.

| Significance driven detection model | RL-BoF | GBMR-BoF | GC-BoF | LC-BoF | RL-SVM |
|-----------------------------------|--------|----------|--------|--------|--------|
| The correct detection rate        | 82.14% | 43.57%   | 45.71% | 64.29% | 68.00% |
| The false detection rate          | 20.71% | 28.57%   | 84.29% | 52.14% | 66.67% |
4. Conclusion
This paper mainly introduces the object detection method for low-resolution remote sensing images, which is driven by regression learning saliency model, and focuses on the principle, algorithm and steps of regression learning saliency model. The model has the ability of adaptive control and adaptive learning in saliency detection, determines the number of learning layers independently, extracts the saliency region more accurately. This method is a more intelligent saliency detection method. Compared with the existing saliency model, the regression learning saliency model can accurately locate the location of the airport and keep the original structure of the airport to the greatest extent. Therefore the regression learning saliency model is proved to be more effective in airport detection.

References
[1] Huaguang Zhang, Zhanshan Wang, Derong Liu. A Comprehensive Review of Stability Analysis of Continuous-Time Recurrent Neural Networks [J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(7), 1229-1262.
[2] Minghui Tian, Shouhong Wan and Lihua Yue. A Novel Approach for Change Detection in Remote Sensing Image Based on Saliency Map [J]. Computer Graphics, Imaging and Visualisation, 2010, pp. 397-402.
[3] Radhakrishna Achanta, Appu Shaji, Kevin Smith, et al. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11), pp. 2274-2281.
[4] Sunil L. Kukreja, Johan Lofberg and Martin J. Brenner. A Least Absolute Shrinkage and Selection Operator (LASSO) for Nonlinear System Identification [C]. 14th IFAC Symposium on System Identification, 2006: 814-819.
[5] Qiong Yan, Li Xu, Jianping Shi, et al. Hierarchical Saliency Detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 1155 – 1162.
[6] Xiaohui Li, Huchuan Lu, Lihe Zhang, et al. Saliency Detection via Dense and Sparse Reconstruction [C]. IEEE International Conference on Computer Vision, 2013: 2976-2983.
[7] Chuan Yang, Lihe Zhang and Huchuan Lu, et al. Saliency Detection via Graph-Based Manifold Ranking [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 3166-3173.
[8] Mingming Cheng, Jonathan Warrell and Wenyuan Lin et al. Efficient Salient Region Detection with Soft Image Abstraction [C]. IEEE International Conference on Computer Vision, 2013: 1529-1536.
[9] Yun Zhai, Mubarak Shah. Visual Attention Detection in Video Sequences Using Spatiotemporal Cues [C]. The 14th annual ACM international conference on Multimedia, 2006: 815.
[10] Yang Cao, Changhu Wang, Zhiwei Li, et al. Spatial-bag-of-features [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2010: 3352-3359.