Multi-Resolution Airport Detection via Hierarchical Reinforcement Learning Saliency Model

Danpei Zhao, *Member, IEEE,* Yuanyuan Ma, Zhiguo Jiang, and Zhenwei Shi, *Member, IEEE*

***Abstract*—Traditional airport detection methods usually utilize geometric characteristics to locate targets, but they are not suitable for low-resolution remote sensing images. Taking both low and high resolution into account, we present a novel hierarchical reinforcement learning (HRL) saliency model to detect airport target. Different from conventional saliency models focusing on nature images, our HRL model is more effective for multi-resolution remote sensing images. According to airport characteristic, we design a reinforcement learning structure to suppress background and highlight interesting airport regions level by level. To generate a final saliency map, we fuse bottom-up region features with top-down line feature based on target attribute, which can restrain other salient regions except for airports. Moreover, a learning stop criterion based on Latent Dirichlet Allocation (LDA) topic model is proposed at each level to judge the state of saliency detection, thus learning process can be adaptively controlled. Besides, a back-level propagation mechanism is employed to reinforce airport target between levels. HRL saliency model can take the advantage of hierarchical structure to quickly locate interest regions in remote sensing images with large cover area. Furthermore, HRL is robust for illumination and resolution variety. Extensive experimental results on a remote sensing dataset containing 730 images of 40 different airports demonstrate that the proposed HRL model outperforms 18 state-of-the-art saliency models in terms of two popular evaluation measures. Besides, it has significantly higher detection rate than other 6 airport detection methods.**

***Index Terms*—airport detection, multi-resolution , hierarchical reinforcement learning, LDA topic model, back-level propagation mechanism.**

I. Introduction

WITH the increasingly development of sensor technology, the application area of remote sensing images is more and more abroad, such as scene classification ([1], [2]), semantic annotation [3], target detection [4]. And automatic airport detection technology plays an important role in target detection, attracting more and more attention in military and civil application, such as precision guidance, aerial reconnaissance, security monitoring. Because airports are usually located on cluttered ground surroundings including buildings, mountains, rivers or vegetation, accurate airport detection becomes a challenging problem under the influence of various disturbance factors.

Most of existing airport detection methods usually adopt geometrical characteristics of airport such as straight or parallel line feature aiming at high-resolution images. Liu et al. [5] searched for elongated rectangles in the image, then considered these detected rectangles as runways. Zhu et al. [6] applied long straight line as airport top-down feature and the method in [7] used parallel information to determine the interest regions. However, these methods could be hardly used to locate airport regions in low-resolution images because linear feature of airport has some problems such as zigzag broken or parallel lines overlap by the resolution restriction. Tao et al. [8] proposed an improved SIFT matching strategy to detect regions of interest (ROIs), being followed by a SVM classifier to refine detection result. But searching and matching in the whole image, can undoubtedly cause extensive computation.

In recent years, various saliency models were put forward to detect targets, which are most designed for nature images. Traditional saliency models ([9], [10]) are based on the biological vision mechanism, such as Itti model [9], which calculates global or local center-surround feature differences to determine saliency maps. But it is very hard to suppress complicated backgrounds. Perazzi [11] (SF model) and Cheng [12] (GC model) choose element uniqueness and its distribution as features to generate saliency map. SF model [11] treats superpixel as basic element to perform the operation, while GC model in [12] uses Gaussian Mixture Models (GMM) to decompose an image to get basic elements. Besides that, there are some saliency models using frequency characteristics. The prominent superiority of these methods is their easy operability. Hou and Zhang [13] (SR model) extract the spectral residual of an image in the spectral domain, and obtain saliency map by inverse transform of spectral residual. DSR model [14] computes dense and sparse reconstruction errors putting the boundary pixels as background according to the pre-knowledge hypothesis that most of targets usually appear in the center of an image while surrounding outer boundary are the background regions. While Yang et al. [15] (GBMR model) combines background and foreground queries to generate saliency map by using the background prior information. GBMR can effectively resolve the problem that targets appear on the edge of an image by computing four independent saliency maps in four side areas. The above-mentioned saliency models have good performance for natural images, which can quickly locate target regions. So, saliency models are introduced to airport detection in RSIs with mass data, which can greatly reduce searching time. But these models cannot get the same accuracy for remote sensing images (RSIs), due to the significant differences between RSIs and natural images in the aspect of resolution, texture, structure, and illumination intensity. The natural image datasets applied

Danpei Zhao (Corresponding Author, e-mail: zhaodanpei@buaa.edu.cn), Yuanyuan Ma, Zhiguo Jiang and Zhenwei Shi are with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, and with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China.

to saliency detection are most close shot photography. These images usually have high contrast, bright color, clear outline and simple texture. Therefore, edge detection, differences about color and texture can be used to detect salient regions, which makes most saliency detection models work well. But for RSIs, especially for low-resolution RSIs, there are no obvious boundary between airport target and background, and texture contrast is not obvious either. Moreover, different airport targets have different texture features and structure information. Due to the differences, existing saliency models designed for natural images cannot be used directly in RSIs. This is because edge and texture features are ineffective to detect airport in low-resolution RSIs, it is difficult to get accurate detection result.

Therefore the general structure that airport detection methods based on saliency model is designed to be a saliency detector followed by a classifier aiming at high-resolution RSIs. The saliency detectors either employ the existing saliency model or make a slight change on it. Just as those methods proposed in [16], [17] and [6], they apply saliency model to extract ROIs containing an airport, and then use a classifier to refine these regions until airport was detected. [16] and [6] use the GBVS saliency model [18] to detect ROIs , which views maps as a graph model, using Markovian algorithm to search key locations. And Yao et al.[17] employs the FT saliency model proposed in [19], which chooses a band pass filter to eliminate noises and backgrounds and retain salient regions. Thus, unsupervised model becomes supervised through refinement process, which increases the training samples and complexity of algorithm undoubtedly. But it is difficult to acquire high-resolution images in many aviation reconnaissance missions. Therefore, the problem of how to fast accurately detect airports in low-resolution RSIs becomes urgent, which we should settle in practical engineering application.

To solve above problems, this paper proposes a new hierarchical reinforcement learning saliency model to detect airport target in terms of airport imaging characteristic. HRL model not only can accurately detect airport region in low-resolution RSIs which includes many similar structures and textures, but also can fast locate airport target in high-resolution RSIs with massive data and wide field of view. So, the adaptive ability for multi-resolution images is conspicuous superiority of our model. Furthermore, Latent Dirichlet Allocation (LDA) [20] model is embedded into our saliency model rather than following behind saliency detection. Thus, our saliency model can efficiently detect certain type of salient target depending on the particular task demands.

The main contributions of our approach are summarized in three aspects below:

(1). We propose a novel hierarchical reinforcement learning structure which can selectively approximate airports area level by level by a back-level propagation mechanism of saliency map. It is helpful to better suppress complex background and other salient targets and highlight airport region at the same time.

(2). We build a saliency detection model which fuses bottom-up lower-features map with top-down object-based feature map. Thus, airports region can be highlighted well; meanwhile, other salient targets and background are all effectively suppressed. That is to say, special advantage of our model is that it can only detect airport target rather than all salient targets in RSIs compared with other saliency models.

(3). We design a judgment strategy to autonomously determine the number of learning level and the time of learning stop applying LDA topic model. It can intelligently control learning processing by identifying the similarity between saliency regions and training airport target features in each level.

II. HRL SALIENCY MODEL

HRL saliency model consists of hierarchical learning process and learning stop judgment. As for the learning part, we fuse the bottom-up latent feature with top-down object-based feature to highlight airport target. And we reinforce these conspicuous regions and suppress background by hierarchical learning structure. For estimating learning processing, we design an adaptive learning stop criteria using LDA topic model which are trained in advance. Fig. 1 shows the flowchart of our HRL saliency model.

A. Reinforcement learning saliency model

For each level of HRL saliency model, as shown in Fig. 1(a), by bottom-up and top-down saliency detection and feature fusion, we get a saliency map of corresponding level. Through back-level propagation mechanism among levels, as shown in Fig. 1(b), the saliency difference between target and background is reinforced. Specific implementation details are as follows.

1) Bottom-up saliency map: a) Superpixel clustering

For capturing the structural information of an image and reducing computational complexity, we first cluster all pixels into different regions to form superpixels using the simple linear iterative clustering (SLIC) algorithm [21]. According to color features and space distance constraints, similar pixels are partitioned into the same superpixel region which has obvious consistency in structure. So, superpixel segmentation is more conducive to highlight the structural information of the image than in pixels.

Given the input image *I*, we use SLIC to segment it into *m* superpixels (regions), getting the global feature set $\mathbf{P}$ as

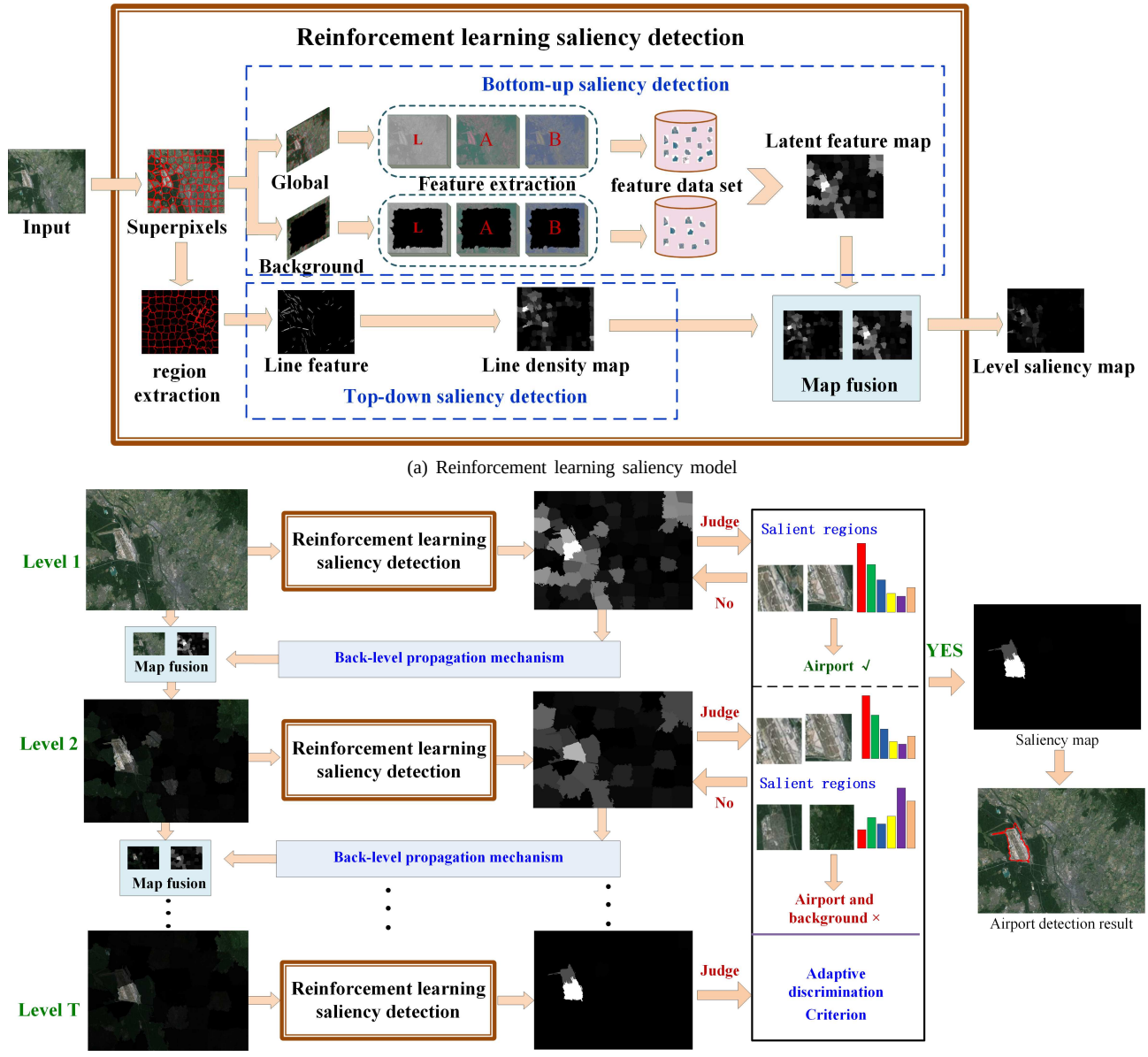
$$\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_m\}. \tag{1}$$

In Eq. (1), $\mathbf{p}_i$ represents feature vector of region *i*. In this paper, we select the color features in CIELab color space as the feature vector because of the physical characteristics based on CIELab which is provided with the wide color gamut and rich chroma. The feature vector of each region is defined as average color feature of pixels it consists in each color channel in CIELab space. That is, for the region *i*, feature vector $\mathbf{p}_i$ can be described as

$$\mathbf{p}_i = (\bar{l}_i, \bar{a}_i, \bar{b}_i), \tag{2}$$

where $\bar{l}_i$, $\bar{a}_i$, $\bar{b}_i$ represent the average color of all pixels in region *i* of *L*, *A*, *B* color channel respectively.

b) Background pre-knowledge



(a) Reinforcement learning saliency model

(b) Hierarchical reinforcement learning saliency model

Fig. 1. Flowchart of HRL saliency model.

As we known, RSIs usually contain complicated background such as sea, forest and land in one image. In these situations, it is difficult to guarantee airport target as salient regions according to human visual attention mechanism. So, background pre-knowledge is introduced to our model, which has been used to saliency detection in natural images. According to this principle, we suppose that interesting targets always tend to appear at central regions of an image.

As for RSI, especially low-resolution RSI often contain large cover area in which most of the area is the background, only a small fraction is the target. Moreover, the possibility of target appear on image boundary is very small, so it is reasonable to treat image edges as the background in RSIs. Even if the target appears in one boundary, it cannot appear in four boundaries simultaneously. Consequently based on this assumption that targets will appear in the image center while background regions usually are located in the image boundaries, we can construct a background data set using feature vectors of boundary regions. Thus, background template $\mathbf{B}$ is represented as

$$\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_j, \cdots, \mathbf{b}_n\}, 0 < n < m, \tag{3}$$

$$\mathbf{b}_j = (\bar{l}_j, \bar{a}_j, \bar{b}_j), 1 \leq j \leq n, \tag{4}$$

where n is the number of image boundary regions, and $\mathbf{b}_j$ represents feature vector of the background region j. Because those superpixels locating in the boundaries are considered as background which can almost cover all kinds of background features, our model can better suppress complex background to ensure highlighting airport.

c) Similarity measure

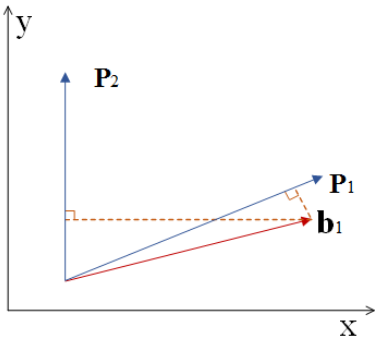


Fig. 2. Similarity measure map.

Up to now, we have got the global dataset $\mathbf{P}$ and background dataset $\mathbf{B}$. Next, we mine the latent feature of each region in global data set by compare it with background information.

Given data pairs $(\mathbf{p}_i, \mathbf{b}_j)$, where $\mathbf{p}_i \in \mathbf{P}$ and $\mathbf{b}_j \in \mathbf{B}$, we use Eq. (5) to calculate the similarity coefficient α_{ij} between each global region (that is, $\mathbf{p}_i$) and each background one (that is, $\mathbf{b}_j$),

$$\alpha_{ij} = \arg\min_{\alpha_{ij}} \|\mathbf{b_j} - \alpha_{ij}\mathbf{p_i}\|_2. \tag{5}$$

We define the model in Eq. (5) as Least Distance Similarity Measure (LDSM) operator. And α_{ij} is the learning coefficient of $\mathbf{p}_i$ corresponding to $\mathbf{b}_j$. As shown in Fig. 2, taking 2D feature space for example, it describes three feature vectors including one background region ($\mathbf{b}_1$) and two global regions ($\mathbf{p}_1$ and $\mathbf{p}_2$). If a global vector is similar to background ones, just as $\mathbf{p}_1$ and $\mathbf{b}_1$ in Fig. 2 (By projecting between two vectors we can have an intuitive understanding of two vectors similarity.), the similarity coefficient α_{11} is approximate to 1. When $\mathbf{p}_i$ equals to $\mathbf{b}_j$, α_{ij} equals to 1 exactly. While there is a great difference (difference on angle or length of vectors) between global and background regions, just as $\mathbf{p}_2$ and $\mathbf{b}_1$ in Fig. 2, α_{21} tends to be bigger or smaller than 1. Therefore, we can define the similarity measure of each region to background features with the difference between α_{ij} and 1. To standardize and simplify it, we normalize the learning coefficient α_{ij} to β_{ij} as follows.

$$\beta_{ij} = \mathcal{N}\left(|\alpha_{ij} - 1|\right), \tag{6}$$

where $\mathcal{N}(x)$ denotes normalizing *x* into [0,1]. After this normalization, the closer β_{ij} is to 0, the corresponding superpixel region is closer to background.

By solving this optimization problem, we can get a similarity coefficient matrix:

$$\begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mn} \end{pmatrix}, \tag{7}$$

where the *i*th row is a learning vector representing the differences of all background regions to region *i* in global data set. Each element shows the relevance between a background region and a global region. It will get a low value for element β_{ij} when there is a strong correlation between $\mathbf{p}_i$ and $\mathbf{b}_j$ because $\mathbf{p}_i$ tends to preferably represent $\mathbf{b}_j$ to obtain a small error. Otherwise, if there is large difference between these two regions, that is, $\mathbf{p}_i$ has a fair chance to be a target region, $\mathbf{p}_i$ cannot represent $\mathbf{b}_j$ perfectly and β_{ij} tends to take a high value to meet with $\mathbf{b}_j$'s projection. For this reason, salient targets can be distinguished from background areas due to its higher β_{ij}-value than other regions'.

For each row of the similarity coefficients above, we define latent feature β_i as the average of the similarity coefficients on all global superpixels, implying distributing equal weight to each coefficient

$$\beta_i = \sum_{j=1}^{n} \beta_{ij}/n. \tag{8}$$

Then, the latent feature map is $F = (\beta_1, \beta_2, \cdots, \beta_m)$. Eq. (8) denotes that the latent feature is calculated by averaging the similarity coefficients on all global superpixels, implying distributing equal weight to each coefficient. When background consists of more than one texture structure in RSIs, (for example, background is composed of land and sea.) this strategy uniformly treating all background regions can learn approximately equal latent features for different background. Meanwhile, these latent features background holding are lower than target regions'. Therefore, we calculate latent feature as Eq. (8), and its value shows the possibility that region *i* is a target. The bigger β_i is, the more likely region *i* to be a target.

2) top-down saliency map: Considering that the airport is a special kind of remote target with many particular characteristics, among which straight line represents most significant feature of airport target. Compared with mussy land, forest cover and calm sea, airport regions consist of lots of densely straight lines, which are long or short, parallel or intersecting. So, we think that these regions with higher linear density distribution are most likely the airport area. For distinguishing airport target from many salient regions, we introduce a top-down strategy driven by certain task to build our saliency model which extract linear feature to construct airport-based feature map.

The line detection operator (LSD) [22] is a linear-time line segment detector that gives accurate results, a controlled number of false detections, and requires no parameter tuning. We use LSD to detect line segment in input images. Due to the differences between high and low resolution in RSIs, there are no obvious long straight lines after LSD detecting on low-resolution image. Instead, it will produce some problems such as zigzag broken or parallel lines overlap. It is difficult to form accurate saliency map using high-density broken short lines. Thus, we adopt line density map as top-down feature to reflect linear characteristics of airport area. Then we calculate the line density of each superpixel. In theory, line distribution should be dense in airport region. Though there are no clear rules of these line distributions, the line density of these regions is usually higher than other regions, that is to say, dense area can be considered as proposal of airport target in line density map. Therefore, we use line density information as the particular characteristic of airport to distinguish it from other salient targets.

For the initial input image *I*, we use LSD model obtain the line information of the whole image. Then, corresponding superpixel segment above, we get line density in each region by a statistic about line length in one superpixel. For the region *i*, the line density d_i can be calculate as

$$d_i = \frac{N_L(region(i))}{N(region(i))}, i = 1, 2, \cdots, m, \tag{9}$$

where $N_L(region(i))$ denotes the number of pixels on the lines in region *i* and $N(region(i))$ is total number of pixels in region *i*. Then, the target feature map (that is, line density map) is $D = (d_1, d_2, \cdots, d_m)$. Airport regions consist of numerous line structures, which is very important characteristic. Although it is difficult to extract the geometric characteristics of these lines (such as parallel, vertical) in low-resolution image, the line density can imply more useful information in locating airport targets. Therefore, we believe that those regions with higher line density are the airport targets with greater probability.

3) Top-down and Bottom-up feature fusion: Next, in order to bring in airport information to feature maps for highlighting airport regions meanwhile suppressing other targets, we fuse the latent feature map *F* with line density map *D* as Eq. (10),

$$S = D \cdot F, \tag{10}$$

where '$\cdot$' means element-wise product, and *S* means the result map after feature fusion. This step can be seen as putting mask over latent feature map. For the salient regions in *F*, if the corresponding value in *D* is large, its salient characteristic can be retained. That is, we keep those regions which are salient in both saliency map and airport-based feature map. Otherwise, if it is inconspicuous in line feature map, the corresponding region would be suppressed. As a result, this feature fusion procedure can be seen as a filtering airport step from a series of salient targets. For achieving our aim and simplicity, we choose multiplex operation for fusion here. And careful adjustment for the features is implemented during constructing reinforcement matrix process following.

B. Hierarchical reinforcement learning structure

The fused feature map above reflects the feature difference between airport target and background, which we can use to reinforce input image and enlarge the contrast of foreground and background. And the implementation details are as follows.

Firstly, we stretch the fused feature using a quadratic function,

$$R = f(S) = \left[{x_i}^2\right]. \tag{11}$$

In Eq. (11), x_i means any element in matrix *S*. Selecting polynomial transformation for element stretching in matrix *F* can suppress background meanwhile remain object features unchanged approximately. We define the feature *R* in above equation as reinforcement matrix, and use it to reinforce input image,

$$I_2 = I \cdot R. \tag{12}$$

We treat the image I_2 getting from *I* as a new input of next layer, and use the similar procedures above to learning features in the 2^{nd} level. In this way, we can construct hierarchical learning structure, that is,

$$I_{t+1} = I_t \cdot R_t, \tag{13}$$

where I_t and I_{t+1} mean the input images in *t*th and *t+1*th level respectively, and R_t represents the reinforcement matrix in *t*th level. By above updating, the new input image I_{t+1} has larger salient pixel values in target regions and further inhibits background area.

C. Learning stop criteria based on LDA

By hierarchical feature learning, we can distinguish salient regions from background according to the feature value. Then, a criterion is needed to judge whether the target regions are salient enough to segment them from background regions.

At the beginning of hierarchical learning process, some background regions cannot be suppressed totally while target regions are highlighted, which means that salient regions contain target as well as part of background at this moment. As the learning going on, background regions are suppressed gradually level by level until totally, when the learning process will complete.

To realize these, we design a stop criteria based on LDA topic model to judge whether it is enough to distinguish target and background at the current level, that is to say, when learning should stop. Because LDA topic model is not trained as a supplement for airport features, we just only measure the similarity degree between the detected salient regions and airport targets using LDA model. We construct LDA model by the same features as ones in the reinforcement learning stage. First, we train LDA model using color features of training images in CIELab space and get the topic model of background $p\left(z|b\right)$ and airport (foreground) $p\left(z|f\right)$. Then, for the saliency map in each level, we calculate the topic model $p(z|s_i)$ of salient superixel *i*. We will end the learning process when Eq. (14) is satisfied.

$$sim(p\left(z|s_i\right), p(z|f)) < sim(p(z|s_i), p(z|b)), \forall s_i \in S, \tag{14}$$

where $sim(\mathbf{A}, \mathbf{B})$ means cosine distance of vector $\mathbf{A}$ and $\mathbf{B}$. And s_i is a salient region in saliency map *S*. We can see from Eq. (14), if all of the salient regions are more like to target samples rather than background ones, then there is strong reason to believe that the entire background region has been suppressed totally and the reinforcement learning process has finished.

By the end of this hierarchical learning process, we can get the final saliency map S_{final}:

$$S_{final} = S_T, \tag{15}$$

where S_T is the saliency map obtained from *T*th learning layer, and *T* is the total layer number.

III. Experiments

To validate our proposed method, we construct a database of RSIs which include real images taken by Gaofen II satellite with resolution of 3.2m/pixel and a dataset with multi-resolution taken by Google Earth. Our database of RSIs

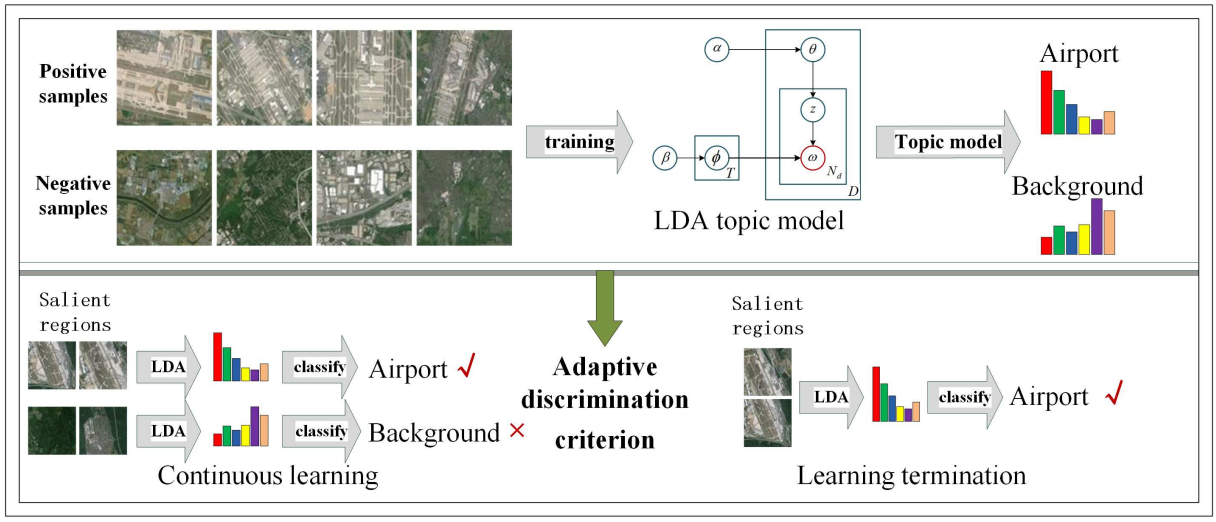


Fig. 3. LDA-based learning stop criterion.

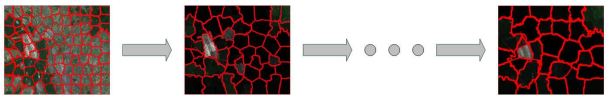


Fig. 4. Multi-scale superpixel segment results.

include 730 images containing 40 different airports , which have an image size of 500×600 pixels and multi-resolution varying from 30m to 50m. Furthermore, these images contain large cover regions with complicated background such as forest, sea, land and buildings, existing various viewpoints and illumination intensity. We carry out all experiments on this database. Ground-truth dataset was manually formed by marking tarmac and main runways. During the experiments, we compare our method with other 18 saliency detection models (wCtr [23], AIM [24], CA [25], DSR [14], FT [19], GBMR [15], GC [12], HC [26], RC [26], HS [27], Itti [9], GB [18], LC [28], LR [29], MSS [30], RA [31], SF [11] and SR [13]) and 6 other airport detection approaches ([8], [17], [6], [32], [7], [33]).

A. Parameter setting

For the hierarchical learning structure model, one key problem is how to determine the superpixel number in each level. Suppose that we segment input image I_t into m_t superpixels in the tth level, then the superpixel number in each level is subjected to:

$$m_1 \geq m_2 \geq \cdots \geq m_t \geq \cdots \geq m_T. \quad (16)$$

Here, as shown in Fig. 4, a fine-to-coarse framework is used in the scale selection of SLIC method considering accuracy and rapidity. In the process of feature learning, a prior fine segment can capture the tiny differences of boundary precisely. By features reinforcing, these differences become obvious. Therefore, the coarse segment later can also get these region differences with a lower computational cost at the same time.

For training learning stop criteria, we construct training samples set by randomly selecting 100 images from RSI dataset, training LDA topic model with one positive sample patch (20×20) and two negative patches in each training image.

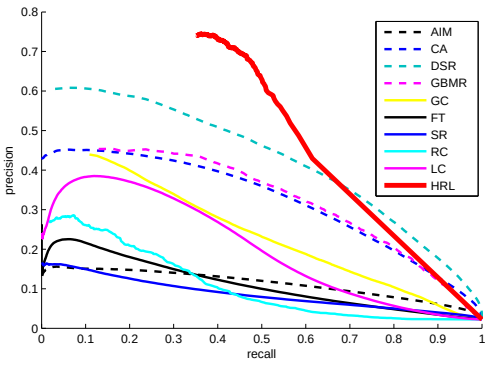
For each saliency map we need to select salient regions to judge learning progress. And salient regions are determined by doing thresholding to the saliency map with threshold *th* set as 0.25. If the salient value of one superpixel is larger than *th*, we consider it as a salient region and put it into LDA model to learn its topic features.

B. Comparison on saliency detection performance

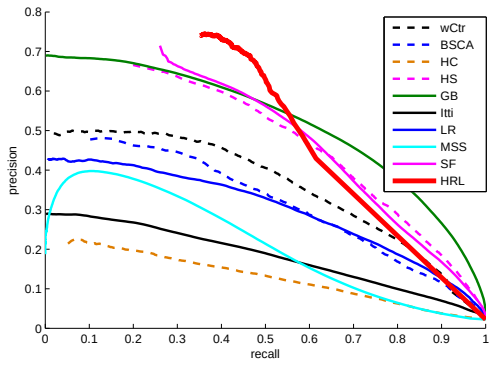
Similar to [23], we use precision-recall curve (*PR*-curve) to evaluate these saliency models, and adopt mean absolute error (*MAE*) as a supplement to *PR*-curve. Just like definitions in [15] and [23], precision shows the proportion of actual salient targets in the salient regions, while recall represents the detected proportion in actual salient targets. The *PR*-curves are plotted with the binary saliency map whose threshold varies from 0 to 255. We define *MAE* to measure average pixel difference between saliency map and ground truth as:

$$MAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S_{final}(x,y) - GT(x,y)|, \quad (17)$$

where *W* and *H* represent image width and height respectively. And $S_{final}(x,y)$ is the saliency value in pixel (x,y) and *GT* means ground truth. The *PR*-curve results are shown in Fig. 5. We find out that *PR*-curve of our method begins with a high recall, which is because our saliency maps are based on superpixel level and maximum of gray image can always hit the target regions. At the same time, this phenomenon also reflects high detection rate of our model. However, those recent popular saliency models that have high detection accuracy for natural images hold a low precision value generally, which means that they are not quite suitable for RSIs. Because there are significant differences between natural images and RSIs. The natural image datasets applied to saliency detection are most close shot photography with big and distinct target.



(a)



(b)

Fig. 5. Comparison on *PR*-curves of various methods.

TABLE I
COMPARISON OF *MAE*S.

Method	AIM	CA	DSR	GBMR	LC
MAE	0.135	0.205	0.104	0.224	0.081
Method	GC	FT	SR	RC	SF
MAE	0.175	0.088	0.096	0.213	0.118
Method	HC	wCtr	HS	Itti	GB
MAE	0.230	0.073	0.259	0.289	0.178
Method	BSCA	LR	MSS	**HRL**	
MAE	0.232	0.106	0.050	**0.019**	

These images usually have high contrast, bright color, clear outline and simple texture. Therefore, edge detection, feature difference about color and texture can be used to detect salient regions, which makes most saliency detection models work well. But for RSIs, especially for low-resolution RSIs, there are no obvious boundary between airport target and background, and texture contrast is not obvious either. Moreover, different airport targets have different texture features and structure information. As a result, edge contrast and texture information are ineffective for airport detection in RSIs. That is the reason why the existing saliency models cannot obtain good detection performance for low-resolution RSIs. Table I shows the *MAE* value of 19 saliency detection models. It shows that our model has a lower *MAE* apparently than others. Therefore, our proposed saliency model, specifically designed for RSIs, has a better performance than other state-of-the-art models on this RSIs dataset. And Fig. 7 presents saliency detection results of several representative models, which can give us intuitive understanding for the performance of these models. As shown in Fig. 7, it is difficult to precisely highlight airport target as salient regions for other saliency models in RSIs which usually contain large cover regions with complicated background such as forest, sea, and land. But our saliency model can better suppress complex background to ensure highlighting airport.

C. Comparison of detection performance at different resolutions

HRL saliency model is designed for low-resolution RSIs, furthermore, taking into account multi-resolution RSIs, which

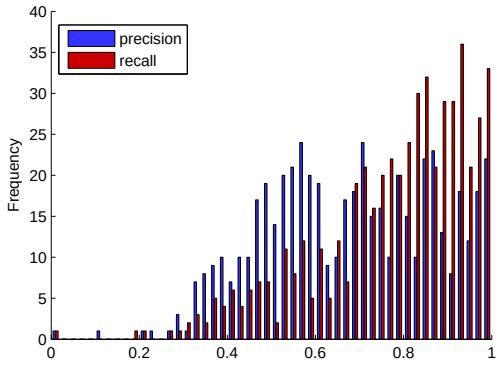
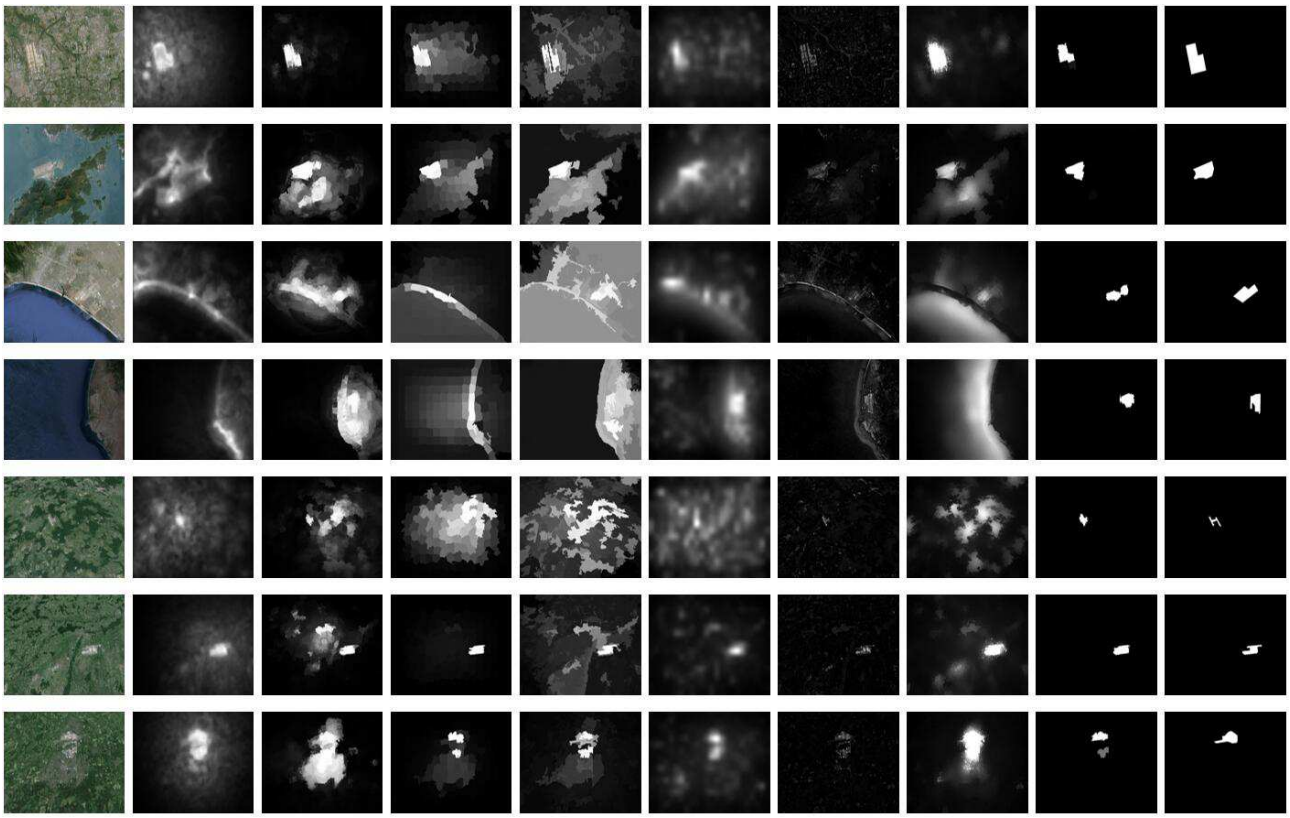


Fig. 6. The statistical result of precision and recall of HRL model in RSI dataset.

TABLE II
PRECISION AND RECALL IN DIFFERENT RESOLUTIONS.

Resolution	5.3m	6.4m	8m
Image size	4380×4146	3650×3455	2920×2764
Precision	0.7345	0.7255	0.7694
Recall	0.9428	0.9338	0.9446
Resolution	10.7m	16m	32m
Image size	2190×2073	1460×1382	730×691
Precision	0.7879	0.7031	0.6878
Recall	0.9207	0.8773	0.9575

is aimed at quickly and precisely locating airport regions in large cover area. Detection results can get from saliency maps by thresholding with 0.25. Fig. 6 describes the statistic results of our HRL model in term of precision and recall in RSI dataset, from which the high recall value reflects accuracy hitting in airport location using this method. With reference to [6], we defined it as a successful detection if precision is higher than 0.4 and recall is higher than 0.3. That is to say, if the detected regions contain more than 30% of the precisely labeled ground truth, and the ground truth contains more than 40% of the detection regions, it is defined a successful detection in our model. Thus, our HRL model has obtained detection rate of 91.22%, with the 67.68% precision and 77.44% recall in average on the basis of the statistic results of



(a)Input (b)CA[25] (c)DSR[14] (d)GBMR[15] (e)HS[27] (f)GB[18] (g)MSS[30] (h)SF[11] (i)ours (j)Ground truth

Fig. 7. Comparison results of several saliency models for airport detection in RSIs.

Fig. 6. It indicates that HRL model can get a good robustness no matter how illumination and resolution change. Fig. 8 shows several experiment results of HRL model under the condition of different resolution and illumination intuitively, which indicates that the adaptive ability for multi-resolution images is a superiority of our model.

Similarly, our approach has the good adaptability for high-resolution RSIs. As shown in Fig. 9(a), high-resolution RSI is acquired from Gaofen II satellite with the resolution of 3.2m/pixel and the size of 7300$\times$ 6910 pixels. First, this image is reduced to different size by down-sampling, then HRL model is applied to do target detecting. Fig. 9(b) shows the down-sampling images and their detection results in different resolutions, and Table II describes the precisions and recalls of the corresponding detection results. From Table II, we find out that HRL model is robust to resolution variation, which can get similar high precisions and recalls no matter high-resolution or low-resolution. Based on the analysis of experiment results, HRL model can exactly detect airport targets for various resolutions images. Therefore, we provide a good solution saving more calculating time for RSI with massive data, which can quickly locate airport region in small size image with low-resolution by down-sampling.

D. Comparison with other airport detection methods

The following measure is used for performance evaluation of airport detection: detection rate (*DR*). We define *DR* similar to [6]:

$$DR = x/N \times 100\%. \tag{18}$$

In Eq. (18), *N* means the total image number, and *x* is the number of images that have successful detection. As defined above, if a detection result is satisfied that precision is higher than 0.4 and recall is higher than 0.3, it is defined as a successful detection. We compare our airport detection method with other four methods ([8], [17], [6], [32]) in terms of *DR*. The codes of these four methods are realized by us or received from the authors, and parameters in them are adjusted according to the corresponding dataset.

Table III shows the comparison results with other four airport detection methods. On the basis of the definition of successful detection that precision is higher than 0.4 and recall is higher than 0.3, we can see that our method has a high airport *DR* by statistical analysis of experimental results. For

TABLE III
COMPARISON OF AIRPORT DETECTION RESULTS.

Models	Ref [6]	Ref [8]	Ref [17]
Average precision	0.9340	0.4048	0.6175
Average recall	0.2994	0.3242	0.3865
DR	43.27%	40.46%	57.06%
Models	Ref [32]	HRL	
Average precision	0.5798	0.6768	
Average recall	0.4958	0.7744	
DR	67.75%	91.22%	

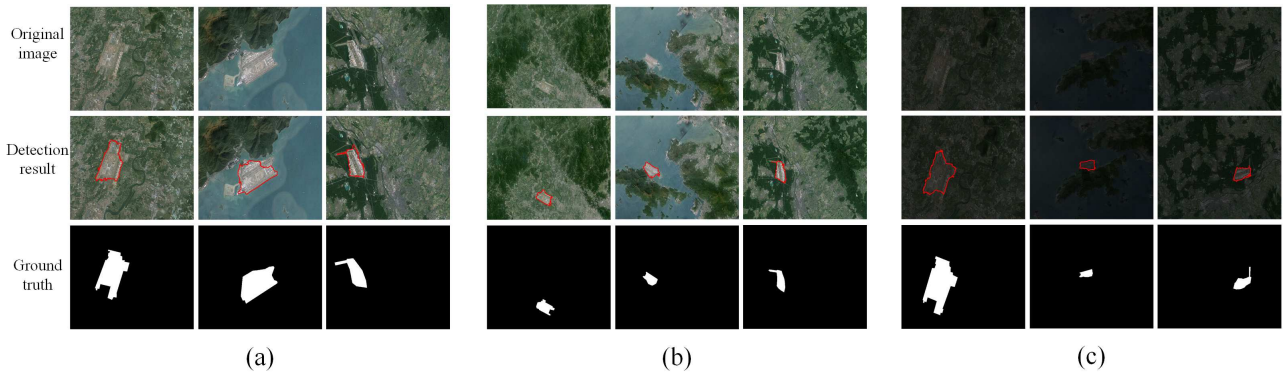


Fig. 8. Experimental results of the HRL model. (a) Images of relatively high space resolution, (b) Images of relatively low space resolution, (c) Images of condition on light deficiency.

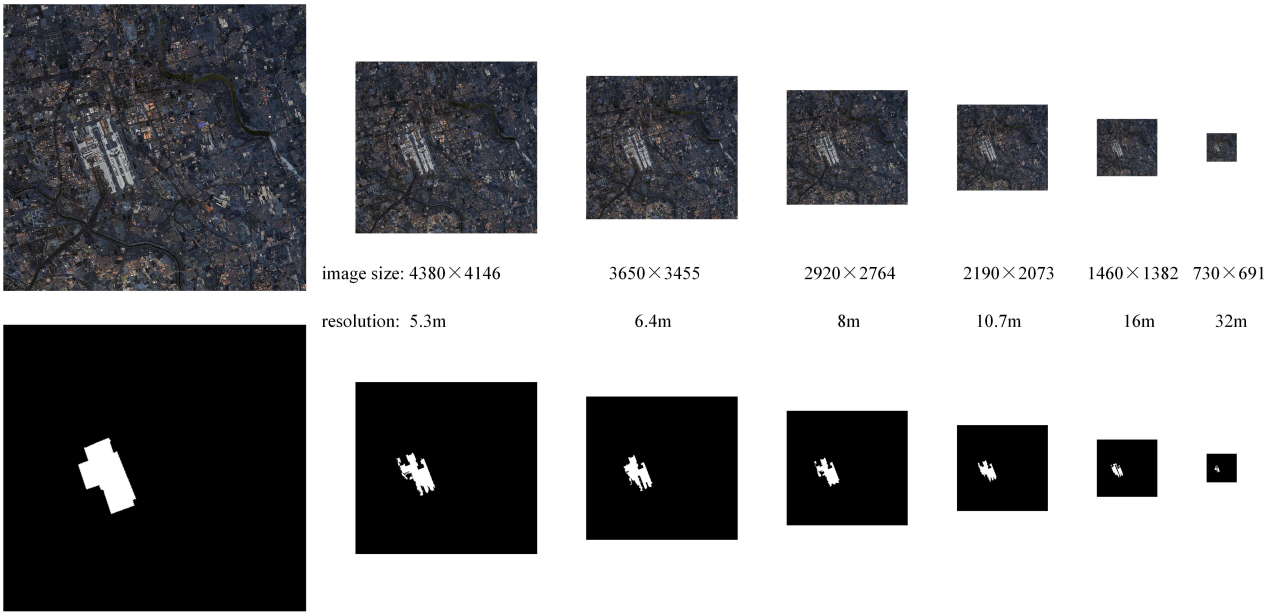
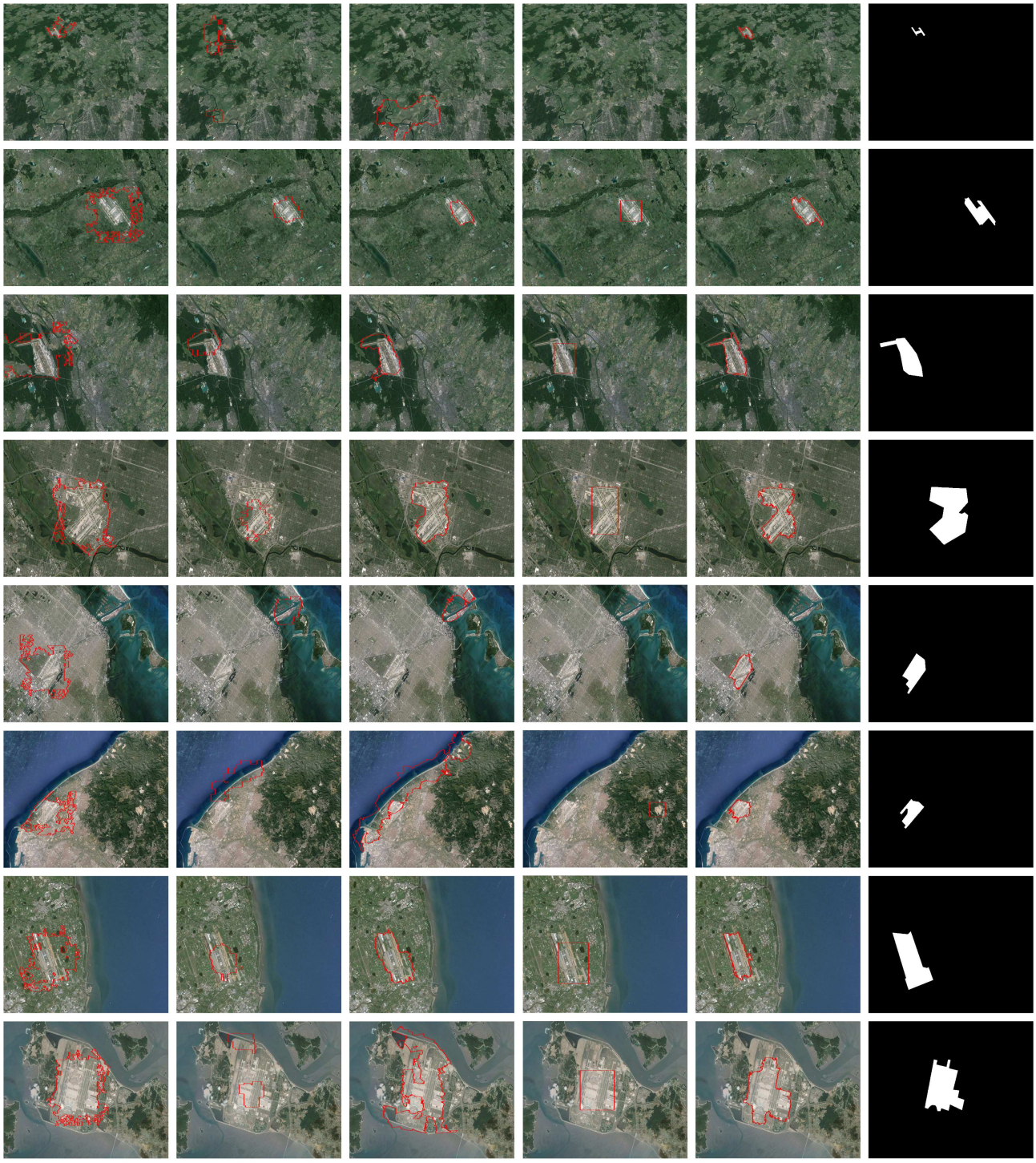


Fig. 9. Detection results of a high-resolution image.

the method in [8], which has a high average precision but low average recall, its results are unable to correctly distinguish most of airport and background. However, other three methods with both low average precision and average recall can hardly accurately locate airport region. All in all, as the methods in [6], [8], [17] and [32] are designed for high-resolution images in which line feature is significant, it is hard to highlight airport target from the complex background because of lacking texture features of airport regions in low-resolution images. Consequently, it is undisputed for them to hold a low *DR* value, which also is consistent with detection results in Fig. 10.

We also test the methods in [7] and [33] on our RSI dataset. And due to their excessive reliance on line feature, they could hardly detect any airport targets in low-resolution RSI dataset. Therefore, the results of [7] and [33] have not been filled into Table III.

We have successfully detected 671 images in the RSIs dataset with the total of 730. Comparison results of several airport detection algorithms are shown in Fig. 10. As it shows, our model has a better ability of edge preservation no matter for sample or complex background than other detection algorithms. Taking the sixth row in Fig. 10 for example, when airport target locates in complicated scene with sea and land, saliency models in [6] and [17] treat coast line as target regions and the method in [32] also gives wrong detection result. Although the method in [8] can always hit the target due to the SIFT feature matching strategy, but the ability of target edge segmentation is not good enough. Because these detection methods have been designed for high-resolution natural images, which have excessive reliance on line features and abundant textures, they could hardly fully detect airport targets in low-resolution RSI dataset. Our method combines bottom-up color features in LAB space and top-down line feature to generate saliency map, which take background pre-knowledge to suppress land and sea background simultaneous-



(a) Ref [8] (b) Ref [17] (c) Ref [6] (d) Ref [32] (e) HRL (f) Ground truth

Fig. 10. Comparison results of several airport detection algorithms.

ly. Furthermore, proposed hierarchical reinforcement learning structure is helpful to accurately distinguish airport target from the complex background and other salient regions, which can selectively approximate airports areas level by level by a back-level propagation mechanism of saliency map. Our strategy can ensure airport target as most salient regions. As discussed above, our model is suitable for low-resolution images with very large cover areas, it can be used to quickly locate airport targets in large size RSIs, but other methods just like [8], [17], [6], [32], [7] and [33] only can be used in high-resolution images with the small field of view.

IV. Conclusion

A hierarchical reinforcement learning saliency model for multi-resolution airport detection is presented in this paper. By combining bottom-up latent feature map driven by low-level cues with top-down line density map driven by task, the proposed HRL model can gradually reinforce feature difference degree between target and background using hierarchical learning structure, which make it more adaptive to low-resolution RSIs. Moreover, HRL model can accurately find airport target and quickly exclude other salient regions and background in comparison with other saliency model, which employs a novel learning stop criterion based on LDA topic model to control learning process. Comparisons of qualitative and quantitative analyses of the experimental results are implemented, which validate the effectiveness of our method for detecting airport target in multi-resolution RSI dataset. Not only our HRL model has more remarkable detection performance for low-resolution airports than the other latest saliency models, but also it can be applied for high-resolution RSIs with huge data to quickly locate airport by dimension reduction, which is one of our important contributions. In the future work, we will extend this model to automatically detect multiclass salient remote sensing targets at the same time.

Acknowledgment

This research was supported by the Fundamental Research Funds for the Central Universities and the National Natural Science Foundation of China (Nos. 60802043, 61071137, and 61271409), National Basic Research Program (also called the 973 Program, No. 2010CB327900), Aviation Science Foundation Project, and Space Support Foundation Project.

References

[1] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2175–2184, April 2015.

[2] ——, "Scene classification via a gradient boosting random convolutional network framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1793–1802, March 2016.

[3] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, "Semantic annotation of high-resolution satellite images via weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3660–3671, June 2016.

[4] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec 2016.

[5] D. Liu, L. He, and L. Carin, "Airport detection in large aerial optical imagery," in *Proc. ICASSP*, 2004, pp. V–761.

[6] D. Zhu, B. Wang, and L. Zhang, "Airport target detection in remote sensing images: A new method based on two-way saliency," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1096–1100, 2015.

[7] G. Tang, Z. Xiao, Q. Liu, and H. Liu, "A novel airport detection method via line segment classification and texture classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2408–2412, 2015.

[8] C. Tao, Y. Tan, H. Cai, and J. Tian, "Airport detection from large ikonos images using clustered sift keypoints and region information," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 128–132, 2011.

[9] L. Itti, C. Koch, E. Niebur *et al.*, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[10] A. Borji and L. Itti, "Exploiting local and global patch rarities for saliency detection," in *IEEE CVPR*, 2012, pp. 478–485.

[11] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *IEEE CVPR*, 2012, pp. 733–740.

[12] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *IEEE ICCV*, 2013, pp. 1529–1536.

[13] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *IEEE CVPR*, 2007, pp. 1–8.

[14] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *IEEE ICCV*, 2013, pp. 2976–2983.

[15] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *IEEE CVPR*, 2013, pp. 3166–3173.

[16] X. Wang, B. Wang, and L. Zhang, "Airport detection in remote sensing images based on visual attention," in *International Conference on Neural Information Processing*, 2011, pp. 475–484.

[17] X. Yao, J. Han, L. Guo, S. Bu, and Z. Liu, "A coarse-to-fine model for airport detection from remote sensing images using target-oriented visual saliency and crf," *Neurocomputing*, vol. 164, pp. 162–172, 2015.

[18] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2006, pp. 545–552.

[19] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *IEEE CVPR*, 2009, pp. 1597–1604.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[21] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.

[22] R. G. von Gioi, J. Jakubowicz, J. M. Morel, and G. Randall, "Lsd: A fast line segment detector with a false detection control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 722–732, April 2010.

[23] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *IEEE CVPR*, 2014, pp. 2814–2821.

[24] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *Advances in neural information processing systems*, 2005, pp. 155–162.

[25] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, 2012.

[26] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, 2015.

[27] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *IEEE CVPR*, 2013, pp. 1155–1162.

[28] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th ACM international conference on Multimedia*, 2006, pp. 815–824.

[29] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *IEEE CVPR*, 2012, pp. 853–860.

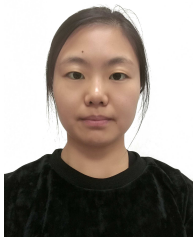
[30] R. Achanta and S. Süsstrunk, "Saliency detection using maximum symmetric surround," in *IEEE International Conference on Image Processing*, 2010, pp. 2653–2656.

[31] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *European Conference on Computer Vision*, 2010, pp. 366–379.

[32] X. Wang, Q. Lv, B. Wang, and L. Zhang, "Airport detection in remote sensing images: a method based on saliency map," *Cognitive neurodynamics*, vol. 7, no. 2, pp. 143–154, 2013.

[33] Z. Kou, Z. Shi, and L. Liu, "Airport detection based on line segment detector," in *IEEE CVRS*, 2012, pp. 72–77.

Danpei Zhao is an associate professor at Beihang University, and has been the Vice Director of the center of image processing at Beihang University. She currently serves as a standing member of the Executive Council of Beijing Society of Image and Graphics. She received her Ph.D. in Optical engineering from Changchun Institute of Optics, Fine Mechanics and Physics of Chinese Academy of Sciences in 2006. From 2006 to 2008, she was in Beihang University for postdoctoral research. She has been working at the Department of Computer Science, Rutgers, the State University of New Jersey, U.S.A. as a visiting Scholar from 2014 to 2015. Her research interests include saliency detection and its application in remote sensing target detection, remote sensing image understanding, target detection, tracking and recognition.



Yuanyuan Ma received her B.S. degree degree from the North-western Polytechnical University, China, in 2014. She is currently pursuing the M.S.degree at Beihang University. Her research interests include Saliency detection and object detection for remote sensing image.

Zhiguo Jiang is a professor at Beihang University, and has been the Vice Dean of the School of Astronautics at Beihang University since 2006. He currently serves as a standing member of the Executive Council of China Society of Image and Graphics and also serves as a member of the Executive Council of Chinese Society of Astronautics. He is an Editor for the Chinese Journal of Stereology and Image Analysis. His current research interests include remote sensing image analysis, target detection, tracking and recognition, and medical image processing.

Zhenwei Shi received his Ph.D. degree in mathematics from Dalian University of Technology, Dalian, China, in 2005. He was a Postdoctoral Researcher in the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was Visiting Scholar in the Department of Electrical Engineering and Computer Science, Northwestern University, U.S.A., from 2013 to 2014. He is currently a professor and the dean of the Image Processing Center, School of Astronautics, Beihang University. His current research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.