Airbase Detection And Airship Recognition In High Spatial Resolution Remote Sensing Images

B.Bersi Beulah[1], T.Agnes Ramena[2]
[1]Dept of ECE
[2]Assistant Professor, Dept of ECE
[1, 2] PET Engineering College

Abstract- *Automatic target detection in satellite images is a challenging problem due to the varying size, orientation and background of the target object. Airship recognition in remote sensing images is a meaningful task. It remains challenging due to the difficulty of obtaining appropriate representation of airships for recognition. To solve this problem, we propose a novel approach that is based on spatial frequency visual saliency analysis and convolutional neural networks. First, airbase is identified using spatial-frequency visual saliency analysis algorithm that is based on a CIE Lab color space to reduce the interference of backgroundand efficiently detect well-defined airbase regions in broad-area remote-sensing images. Second,an airship segmentation network is designed to obtain refined airship segmentation results. Then, a keypoints' detection network is proposed to acquire airship' directions and bounding boxes. At last, apply a template matching method to identify airships. Experiments show that the proposed method outperforms the state-of-the-art methods and can achieve more than 98% accuracy on the challenging data set.*

I. INTRODUCTION

Due to the increasing ability to acquire remote-sensing images using various satellites and sensors, the detection of valuable targets from high spatial resolution remote-sensing images has become one of the most fundamental but challenging research tasks in recent years. However, human image analysts are unable to search targets by heavy manual examination due to the overwhelming number of remote-sensing images that are available on a daily basis. Thus, the need for automatic algorithms to preprocess the remote-sensing data and to extract meaningful information is critical.

An airbase, as an important traffic and military facility, has important practical value for the fields of aircraft navigation, military reconnaissance, and aircraft autopilot. The automatic detection of airbase will have a significant impact on the recognition of aircraft. Previous studies on airbase detection can be grouped into the following two categories: studies that are based on edge detection and studies that are based on image segmentation.

Han *et al.* in [2] proposed a method based on graph search strategy and improved Hough Transform for detection of oil tanks in satellite imagery. Yildiz*et al.* in [3] employed Gabor feature and used SVM classifier to detect different aircrafts. Gabor filter is also employed by authors in [4,5] for road crack detection in aerial images and settlement zone detection is satellite images respectively. Hsieh *et al.* in [6] employed Zernike moments, aircraft contour and wavelets and used SVM classifier for the detection of aircrafts in satellite images. Most of the methods discussed above use hand-crafted features and work effectively in their scenes only. Deep learning is a very effective method for learning optimum features directly from huge training dataset automatically. Now a day in numerous applications computer vision along with deep learning have outperformed humans. Furthermore, the use of Graphical Processing Units (GPUs) has decreased the training time of deep learning methods. Large databases of labelled data and pre-trained networks are now publicly available.

An automatic target detection method based on EdgeBoxes and Convolutional Neural Networks (CNN) is proposed in this paper. Fig. 1 illustrates the conceptual level block diagram of the proposed system. We use EdgeBoxes to produce object proposals in the initial stage. The candidate objects proposed by EdgeBoxes are filtered using some geometric checks while maintaining high recall rates. We then feed the potential object proposals to CNN for automatic feature extraction and classification. Finally, the performance of our method is evaluated on a large military target dataset which contains aircraft and non-aircraft patches for training and test satellite images from Google Earth.

The proposed algorithm can be used for detection of any type of targets. However, we have only detected aircrafts so far due to availability of the aircraft dataset. The literature related to detection of any military target is of interest here.

The proposed target detection system is explained in detail in Section II. The experimental analysis is presented in Section III followed by conclusion and future prospects in Section IV.

II. THE PROPOSED TARGET DETECTION SYSTEM

The proposed framework for target detection system is shown in Fig. 1. The FLS is approximately divided into four parts. First, low resolution images are generated by a Gaussian pyramid. Second, by converting an image to CIE Lab space, we obtain a saliency map by spatial-frequency visual saliency analysis

Third, saliency map is transformed to binary map after threshold segmentation, which produces the initial ROIs of the images. With binary map, we can compute the area for each connected region and extract the connected regions with the area of the topfive. Last, we adopt color and intensity features combined with the Hough transform to exclude images which do not include the airport region, and for images which contain airports, the well-defined airport region is detected by its salient value.

We detect objects in input image using EdgeBoxes and apply geometric checks to select military targets among the object proposals. A well-trained CNN is used to extract features of the proposed objects and classify them as aircraft or non-aircraft objects.

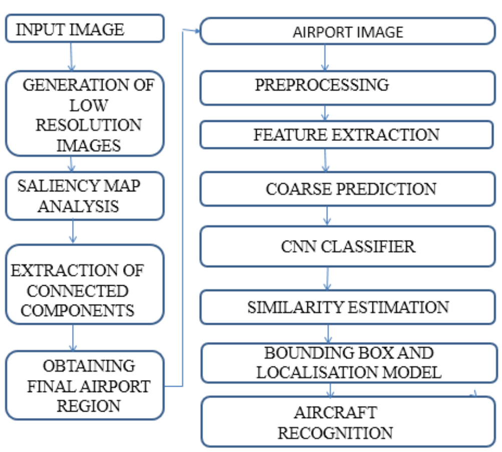

Fig 1 Block diagram

A.AIRBASE DETECTION MODEL:

In this paper, we adopt a spatial-frequency visual saliency analysis algorithm, which introduces a frequency enhancement filter to sharpen edges. First, the CIE Lab color space [22], rather than the RGB color space, has been chosen for saliency detection, as the fact that the Lab color space is uniform and similar to the human perception, with a luminance channel and two-chromatic channels (RG and BY). Given an image I, as previously mentioned, the grayscale of an airport region is usually lighter than the grayscales of its surroundings; thus, the role of the luminance channel in saliency detection is more important than the role of the other two channels. In the two-chromatic channels, we use a Gaussian blurred version to eliminate fine texture details in the spatial domain to achieve a fast speed. For the luminance channel, we adopt a high-frequency enhancement Butterworth high-pass filter to compute feature maps. An image can be described as an amplitude spectrum and a phase spectrum in frequency domain as shown in (1), whereas a phase spectrum retains some important features of an image, such as the edge information.

$$F(u,v) = |F(u,v)| e^{\mu \Phi(u,v)} \qquad (1)$$

where$|F(u, v)|$ is the amplitude spectrum and $\Phi(u, v)$ is the phase spectrum. If we only use the phase spectrum and disregard the amplitude spectrum, the received feature map will contain part of the background information, which is unfavorable for thefollowing segmentation. High-pass filter can sharpen the edge of targets and maintain the edge information by attenuating and suppressing the low-frequency component. If we use a high pass filter, the image quality may not be satisfactory due to the substantial noise and suppressed salient features. To enhance image details and reduce noise, the high-frequency enhancement Butterworth high-pass filter is adopted. The n-order Butterworth high-pass filter with a cutoff frequency of $D0$ is defined as

$$H(u,v) = \frac{1}{1 + [D_0 / D(u,v)]^{2n}} \qquad (2)$$

Where

$$D(u,v) = [(u - \frac{M}{2})^2 + (v - \frac{N}{2})^2]^{1/2}$$

represents the distance between the midpoint of frequency domain (u, v) and the center of frequency rectangle $(M2, N2)$, n is the order of filter (n = 1is adopted in this paper), and $D0$ is the cutoff frequency ($D0$ is 0.04% of the image size in this paper). Given an image with width W and height H pixels, the final saliency map can be formulated as follows:

$$SM(x,y) = | \, |I_L - I_L(x,y)|| + ||I_a - I_a(x,y)| + ||I_b - I_b(x,y)|| \qquad (3)$$

where *IL* (*x*, *y*) is the image pixel vector value of the luminance channel that is filtered by the high-frequency enhancement Butterworth high-pass filter, *Ia*(*x*, *y*), *Ib*(*x*, *y*) is the corresponding image pixel vector value of two-chromatic channels, *IL* , *Ia*, *Ib* is the mean feature map vector, and _ · _ is the *L*2 norm. Equation (3) summarizes our approach, and the received saliency map, which has highlighted salient region with well defined boundaries, is a complete resolution saliency map with the same size of the original image. After threshold segmentation with the Otsu operator, the saliency map is transformed to binary map for target extraction. Considering the larger size of an airport target, we choose to extract the connected regions with the area of the top five among all connected regions in an image as the target candidates. Finally, we adopt color and intensity features combined with Hough transform to exclude images that do not include the airport region, and for images which contain airports, target candidates are identified based on salient value. As the gray value of an airport target is lighter than its surroundings, we treat the target that involves the highest salient value among the five connected regions as the airport target and label the airbase region with a red rectangle.

B.AIRCRAFT RECOGNITION MODEL

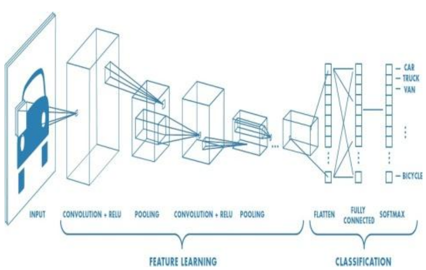

Fig 2 CNN Framework used for feature Extraction and classification in the proposed method

A. Candidate Objects Proposal using EdgeBoxes

The edge information of objects is very useful in remote sensing because it contains very prominent and concise attributes as shown. The EdgeBoxes technique presented in [14] leverages the edge information to detect objects. In EdgeBoxes, a single score from contours confined in a bounding box of the candidate object is calculated and edges with high affinity are grouped together using a greedy approach. The affinity between two groups is given by:

$$a(s_i, s_j) = |\cos(\theta_i - \theta_{ij})\cos(\theta_j - \theta_{ij})|^{\gamma} \quad \textbf{(4)}$$

where s_i and s_j represents the pairs of groups, θ_i the mean orientation of a group s_i and the angle between mean positions of groups s_i and s_j is represented by $\theta_{i,j}$. The sensitivity of affinity to orientation variations is controlled by γ. The score of a bounding box of a candidate object is given by:

$$h_b = \frac{\sum_i (w_b(s_i) m_i)}{2(b_w + b_h)^k} \quad \textbf{(5)}$$

where width and height of a box are represented by b_w and b_h respectively, the sum of magnitudes of all edges in a group s_i is represented by m_i . The value of $w_b(s_i) \epsilon [0, 1]$ indicates whether b contains s_i or not. To normalize the score, the magnitude of edges from b^{in} box centered in b is subtracted to improve the accuracy of EdgeBoxes:

$$h_b^{in} = h_b - \frac{\sum_{p \epsilon b^{in}} (m_p)}{2(b_w + b_h)^k} \quad \textbf{(6)}$$

Where b^{in} has height and width equal to $\frac{b_h}{2}$ and $\frac{b_w}{2}$ respectively. For further details on EdgeBoxes, the reader is referred to [14].

B. Candidate Objects Selection

The candidate objects proposed by EdgeBoxes are very large in number for classification by CNN. We apply some geometric checks to discard the objects which are unlikely to be aircraft objects. The patches of aircraft in satellite images are generally small in size and square shaped, thus the objects with very large or very small area and high aspect ratio are discarded. The objects left behind in geometric filtering are passed to the CNN for automatic feature extraction and image classification.

C. Feature Extraction and Classification using CNN

The Convolutional Neural Network (CNN) is a modern deep learning method which is being widely used for image analysis tasks such as image classification and object detection and segmentation. Krizhevsky*et al.* [18] achieved excellent recognition rates on Large Scale Visual Recognition Challenge dataset using standard backpropagation for training a deep CNN.

A CNN consists of several layers: convolutional, activation and pooling layers in alternation followed by a fully connected layer that produces the output. Unlike typical neural networks, only a small region of input neurons known as Local Receptive Field (LRF) is connected to the hidden neurons. LRF is translated across the image using convolution to map the input to hidden neurons. The hidden layers in CNN learn to detect different features in an image. The weights and biases for all neurons is a hidden layer are the same. Thus, all hidden neurons detect the same features such as edges and blobs in different regions of an image, making the CNN tolerant to translation of objects in an image. Activation transforms the output of each neuron by using activation functions such as Rectified Linear Unit (ReLU) which maps the output of a neuron to the highest positive value, or if the output is negative, ReLU maps it to zero. Pooling reduces the dimensionality of the feature map by condensing the output of small regions of neurons into a single output, thus simplifying the following layers and reducing the number of parameters to learn. The final layer connects the neurons from the last hidden layer to the output neurons which produce the final output. The class probabilities are determined by the value of each node in the final layer.

Fig. 2 represents the proposed architecture of CNN. There are a total of five layers, i.e. two convolutional and two pooling layers in alternation followed by a fully connected layer. There are 6 filters of mask size 5×5 in first convolutional and 12 filters in second convolutional layer. Pooling layer field has a size of 2×2. A sigmoid activation function processes the output of the last layer to generate class labels.

III. RESULT AND PARAMETER ANALYSIS

Fig 3 Remote sensed image

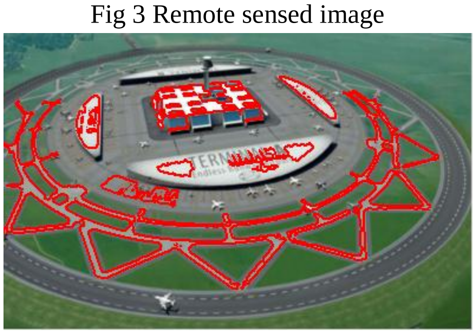
Fig 4 Airbase region detection

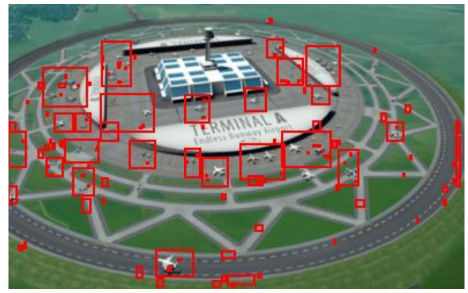

Fig 5 Airship recognition

The quantitative evaluation of the proposed method is based on computing accuracy, as shown in below equation. Here Acc represents accuracy; TP denotes the number of true positive samples; TN denotes the number of true negative samples; FP denotes the number of false positive samples; and FN denotes the number of false negative samples

$$\mathrm{Acc} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}. \quad (7)$$

true positive ratio (TPR) and the false positive ratio (FPR) are defined as

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \times 100\%$$
$$FPR = \frac{FP}{TN + FP} \times 100\%. \quad (8,9)$$

High TPR and low FPR usually contradict each other. Because the receiver operating characteristic (ROC) curve has the ability to show the comparison of TPR and FPR as classification decision criterion changes, it is widely adopted to compare the performance of two different classification systems. The area beneath the curve, which is referred to the ROC area, is also a reflection of the classification performance. Thus, a large ROC area indicates better performance Another quantitative measurement is based on the precision, recall, and *F*-measure, as denoted as *P*, *R*, and $F\beta$. Their definitions are given as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \times 100\%$$
$$R = \frac{TP}{TP + FN} \times 100\%$$
$$F_\beta = (1+\beta^2) \cdot \frac{P \cdot R}{\beta^2 \cdot P + R}. \quad (10\text{-}12)$$

Parameters	Value
Precision	0.9706
Recall	0.988
Time	4.857753
Accuracy	98%

In our experiment, we display classification results based on these measurements.

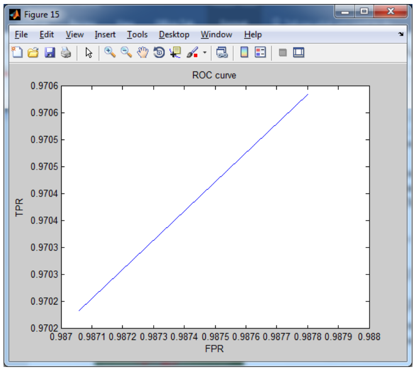

Fig 6 ROC curve

By using this curve we conclude that proposed method achieved 98% accuracy.

IV. CONCLUSION

Automatic target detection in satellite imagery has great significance in military applications. In our work, we propose spatial frequency visual saliency analysis algorithm and CNN for classification in satellite images. CNN effectively learns optimum features directly from huge amount of data automatically. Moreover, CNN is invariant to minor rotations and shifts in the target object. Encouraging experimental results have been obtained on a large dataset. The high precision and recall rates show the optimum performance and robustness of our system in complex scenes. In our future work, we will try to improve the performance of our system and lower the computational cost. We will also apply it in other areas where target detection is used.

REFERENCES

[1] L. Zhang, Z.W. Shi, X.R. Yu, "A hierarchical oil depot detector in high-resolution images with false detection control", 7th International. Congress on Image and Signal Processing (CISP), 2014, pp. 530-535.

[2] X.W. Han, Y.L. Fu, G. Li, "Oil Depots Recognition Based on Improved Hough Transform and Graph Search", Journal of Electronics & Information Technology, Vol. 33, No.1, pp. 66-72, 2011.

[3] C. Yildiz, E. Polat, "Detection of stationary aircrafts from satellite images", IEEE 19th Conference onSignal Processing and Communications Applications (SIU), 2011, pp. 518-521.

[4] H.A. Khan, M. Salman, S. Hussain, K. Khurshid, "Automation of Optimized Gabor Filter Parameter Selection for Road Cracks Detection", International Journal of Advanced Computer Science and Applications, Vol. 7. No. 3, pp. 269-275, 2016.

[5] H. Iftikhar, K. Khurshid, "Fusion of Gabor Filter and Morphological Operators for the Detection of Settlement Zones in Google Earth Satellite Images", IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Nov 2011.

[6] J.W. Hsieh, J.M. Chen, C.H. Chuang, K.C. Fan, "Aircraft type recognition in satellite images", IEE Proceedings - Vision, Image and Signal Processing, Vol. 152, No.3, pp. 307-315, 2005.

[7] G.E. Hinton, S. Osindero, Y.W. Teh, "A fast learning algorithm for deep belief networks", Neural Computation 2006, Vol. 18, No.7, pp.1527-1554, 2006.

[8] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov 1998.

[9] G.E. Hinton, "A practical guide to training restricted boltzmann machines", Neural Networks: Tricks of the Trade - Springer, 2012, pp.599-619.

[10] X.Y. Chen, S.M. Xiang, C.L. Liu, C.H. Pan, "Aircraft Detection by Deep Belief Networks", The 2nd IAPR Asian Conference on Pattern Recognition, 2013, pp. 54-58.

[11] W. Li, S.M. Xiang, H. B. Wang, C.H. Pan, "Robust airplane detection in satellite images", 18th IEEE International Conference on Image Processing (ICIP), pp. 2821-2824, 2011.

[12] F. Zhang, B. Du and L. Zhang, "Saliency-Guided Unsupervised Feature Learning for Scene Classification," in IEEE Transactions on Geoscience and Remote Sensing, Vol. 53, No. 4, pp. 2175-2184, April 2015.

[13] S. Sattar, H. A. Khan, K. Khurshid, "Optimized Class-Separability in HyperSpectral Images", 36th IEEE International Geo-science and Remote Sensing Symposium (IGARSS), July 2016

[14] C. L. Zitnick, P. Dollar, "EdgeBoxes: Locating object proposalsfrom edges", European Conference on Computer Vision Zurich, pp. 391-405, September 2014.

[15] M.M. Cheng, Z.M. Zhang, W.Y. Lin, P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.3286-3293, 2014.

[16] J.R. R. Uijlings, K.E. A. van de Sande, T. Gevers, A.W.M. Smeulders, "Selective search for object recognition", International Journal of Computer Vision - Springer, Vol. 104, No.2, pp.154-171, 2013.

[17] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), pp.580-587, 2014.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet Classification with Deep Convolutional Neural Networks”, The 26th Annual Conference on Neural Information Processing Systems (NIPS), pp. 1097-1105, 2012.

[19] A.J. Chen, J.Z. Li, “Automatic recognition method for quasicircular oil depots in satellite remote sensing images”, OptoElectronic Engineering, Vol. 33, No.9, pp. 96-100, 2006.

[20] P. Dollar, C.L. Zitnick, “Fast edge detection using structuredforests”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PP, No.99, pp.1, 2014.

[21] H. Wu, H. Zhang, J. Zhang and F. Xu, "Fast aircraft detection in satellite images based on convolutional neural networks," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, pp. 4210-4214,20