IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 4, 2019, accepted April 30, 2019, date of publication May 7, 2019, date of current version May 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2915368

Vehicle Detection in Aerial Images Using Rotation-Invariant Cascaded Forest

BODI MA[1], ZHENBAO LIU[1], (Senior Member, IEEE), FEIHONG JIANG[1], YUEHAO YAN[2], JINBIAO YUAN[1], AND SHUHUI BU[1], (Member, IEEE)

[1]School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China
[2]UAV Industrial Technology Research Institute, Chengdu Technological University, Chengdu 611730, China

Corresponding author: Zhenbao Liu (liuzhenbao@nwpu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 61672430, in part by the Equipment Advance Research under Grant JZX7Y20190241011501, in part by the Shaanxi Key Research and Development Program under Grant S2019-YF-ZDCXL-ZDLGY-0227, in part by the Aeronautical Science Fund under Grant BK1829-02-3009, and in part by the NWPU Basic Research Fund under Grant 3102018jcc001.

ABSTRACT Vehicle detection in aerial images has been taking great interest to researchers in recent years. It plays a crucial part in multidirectional applications, such as traffic surveillance, urban planning, and so on. However, the vehicle detection field faces many difficulties owing to the small size of the vehicles, different orientations, and the complex background. To solve this problem, this paper introduces a novel rotation-invariant vehicle detection method which is accurate, stable and has a simple structure compared with region-based convolutional network method. First, the data-driven method has been employed to generate the proposal region which will be applied for data augmentation. Second, this paper designs a method to obtain the rotation invariant descriptors by using radial gradient transform descriptors. Then, the rotation invariant descriptors are fed into the cascaded forest based on auto-context for feature learning and classification. The comprehensive experiments are conducted on the Munich vehicle dataset and UAVDT dataset. The results of experiment illustrate the satisfactory performance of the proposed method.

INDEX TERMS Rotation invariant, vehicle detection, cascaded forest, aerial images.

I. INTRODUCTION

In high resolution aerial images, vehicle detection plays an essential part for many practical applications such as safety assistant driving [1], traffic monitoring [2], [3], and city planning, etc. Consequently, vehicle detection task has aroused the wide attention of researchers. However, there are still a lot of challenges in this task, such as the varying orientation of vehicles, relatively small size (a car might be only 40*20 pixels), and intricate backgrounds. Additionally, some objects (e.g., billboard, and road marks) which are similar to vehicles in appearance, can cause wrong detection result. Furthermore, different from the some large public datasets (such as: ImageNet and CIARF) with huge amounts of data sets, the vehicle detection in aerial images lacks sufficient number of labeled data, which also increases the challenges in this field. Fig. 1. demonstrates some difficulties of vehicle detection. In early studies, the common used methods of vehicle detection in aerial images are based on shallow learning or sliding window method [4]–[9]. Liu *et al.* [10] proposed an approach which could detect the position of the objects without using geographical reference information. This method has some drawbacks, because the sliding window search method for a large scale aerial image will increase calculation burden and it is hard to distinct the vehicle from intricate backgrounds by shallow-learning-based method. In the late years, region-based convolutional neural network (CNN) [11] is widely utilized in the object detection methods. In particularly, Faster RCNN [12] generate the proposal region by region proposal network (RPN), which has better performance than sliding window search method.

However, the RCNN model has some drawbacks which limit its application in vehicle detection.

1) Different from traditional object detection with constant orientation, vehicles in aerial images usually have different orientation. That means the detector needs to recognize objects with various rotation. Furthermore, small size of vehicle and intricate backgrounds will also increase the difficulty of vehicle detection.

The associate editor coordinating the review of this manuscript and approving it for publication was Krishna Kant Singh.

FIGURE 1. Example of some difficulties in vehicle detection (arbitrary orientation, small size of vehicles, and complex background in aerial images).

2) RCNN method usually requires large dataset for training. The small dataset often cause an over-fitting problem for RCNN model.
3) In RCNN model, there are too many hyper-parameters which will directly affect the performance. It takes a lot of time to do the fine tune job.

To address these problems, many researches have focused on deep forest model [13] which was an ensemble method of the forest. It has been successfully applied for object detection, classification, face recognition, etc. Inspired by this manner, we propose a novel and accurate vehicle detection framework called rotation-invariant cascade forest (RICF) which can effectively detect vehicles in aerial images. Different from RCNN based method, we apply the forest based structure with less hyper-parameters to detect the vehicle in aerial images. Radial gradient transform(RGT) has been applied to deal with arbitrary orientation of vehicle in aerial images. Briefly, RICF consists of following two major stages:

In the data augmentation extraction stage, a proposal region generating approach has been employed to generate the object-like regions automatically. To obtain more dataset, we apply rotation transformations to all object-like regions with various rotation direction. In the feature extraction stage, the rotation- invariant descriptors has been utilized in our method to address the problem of orientations. The radial gradient transform is the crucial case on extracting the rotation-invariant descriptors. The rotation-invariant descriptors has been obtained by the RGT method.

In the classification stage, we introduce the cascade forest model based on auto-context. The feature is fed into the first cascade forest layer. And the output of each cascade forest concatenated with the original feature will be fed into next cascade forest layer iteratively. Because both the origin feature and context feature have been taken into account by auto-context model, the last layer will produce the final class vector with the high precision.

The main contributions of our work are as follows:

1) A novel detection model with satisfactory performance is proposed for vehicle detection in aerial images. The multi-layer structure of cascade forest improves the detection performance obviously. Furthermore, compared with the RCNN method, it is more convenient to train the RICF, because the cascade forest model has much fewer hyper-parameters. The multi-layer structure of cascade forest improve the performance obviously. When RICF is applied to different dataset, good performance can even be achieved by almost same settings of hyper-parameters.
2) The rotation-invariant feature has been extracted from the given object-like region. Furthermore, we introduce the approximate radial gradient transform using lookup table, which has much less computation load than traditional radial gradient transform method.
3) Due to expensive data annotations cost, the vehicle detection in aerial images lacks sufficient amounts of training data. To address this problem, the data augmentation manner is applied to generate more training samples.

To evaluate the proposed RICF, several experiments are conducted on the DLR 3K Munich Vehicle Aerial Image Dataset [14] and UAVDT Dataset [15]. Comprehensive comparisons and analysis indicate that the RICF achieve better performance on vehicle detection. The remainder of this paper is organized as follows: Section II discusses the related works. Section III describes the proposed method. Section IV reports the experimental results. Finally, Section V concludes the paper.

II. RELATED WORKS

In this section we briefly introduce the recent target detection methods. Moreover, some methodologies based on random forest related to target detection are reviewed as well.

A. VEHICLE DETECTION IN AERIAL IMAGES

Vehicle detection in aerial imagery data is an interesting study nowadays. Many researchers have proposed suitable methods with good performance to detect the vehicle. Manual features are applied extensively in vehicle detection, for instance, in [4] and [5] hand-crafted features [e.g., histogram of oriented gradients (HOG), local binary pattern, Haar-like feature], have been utilized for vehicle detection which have achieved promising results. Cheng *et al.* [16] designed a dynamic Bayesian network to detect the vehicle utilizing color features. Moranduzzo and Melgani [17], presented a vehicle detection method using SIFT features. However, these methods do not consider the arbitrary orientations of objects in aerial images, which will cause inconsistent performance on different aerial image datasets. To address this problem, Liu and Shi [18] used rotation invariant sparse coding to detect the objects in aerial images. Zhou *et al.* [19] introduced a proper search direction method using image local orientation, which achieved well performance in vehicle detection.

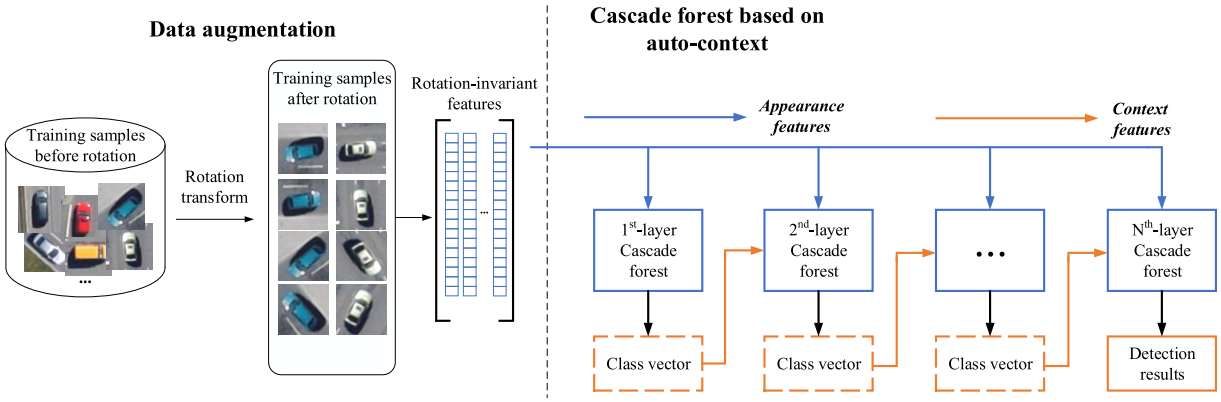

FIGURE 2. Architecture of the proposed rotation-invariant cascaded forest (RICF).

The dominant gradient orientations has been employed to achieve rotation invariance in [20].

B. CLASSIFICATION WITH RANDOM FOREST MODEL

By virtue of the sample architectures and excellent performance in practical applications, random forest based method has been widely used in target detection. Breiman [21] first used random forest to deal with the classification problem. Due to its performance on randomness which is introduced by the random sampling of training set provided to each tree, the forest model can work well on small-scale training data compared to convolution neural networks. In recent years, there have been many applications in object detection using the random forest in [22]–[24]. Bo [25] designed a framework using random forest model, which obtained good performance in vehicle detection. In [26] the structured random forest method was introduced to detect the target. Furthermore, Zhou *et al.* [13] explored the deep forest model based on random forest. The performance of deep forest approach on classification is verified by a series of experiments. Different from RCNN based method, the deep forest has much less hyper-parameters. And the cascade forest structure of deep forest model has powerful feature classification capability, which is suitable for vehicle detection task.

III. PROPOSED METHOD

Rotation-invariant cascaded forests model is proposed to detect the objects in aerial images. Particularly, different from CNN-based method, we use Auto-context cascaded forests to distinguish the positive targets and negative targets. As shown in Fig. 2, our RICF model firstly utilize a general-purpose proposal region generating method [27] and design some rotating operation to generate positive and negative training samples. The radial gradient transform is used to get the rotation-invariant features from image patch. Then, the proposed cascaded forest is trained by these rotation-invariant features. We introduce the auto-context model which was used to enhance the cascaded forests.

A. DATA AUGMENTATION

Due to the labeling cost, vehicle detection task lacks large amount of training data. To increase the number of training samples, we implement dataset augmentation which is significant to avert over-fitting. Instead of only using ground truth objects, we generate training samples by similar method in [11] and [28]. Specifically, we produce multiple object proposals by an effective object proposal region generation method [27], which could generate independent proposal regions. If the object proposal regions has the intersection over union (IoU) larger than 0.6 with the ground truth bounding box. We label it as a positive samples. However, we assign a negative label to the proposal regions, if the IoU ratio is lower than 0.3. The following is the definition of IoU ratio:

$$IoU = \frac{area(B_p \cap B_g)}{area(B_p \cup B_g)} \tag{1}$$

where $area(B_p \cap B_g)$ stands for the intersection of the proposal region and the ground truth bounding box, and $area(B_p \cup B_g)$ denotes their union. In this way, the number of training samples could increase significantly. Furthermore, we explicate N rotation angles $\varphi = \{\varphi_1, \varphi_2, \ldots, \varphi_n\}$ and their rotation transformations $T\varphi = \{T_{\varphi 1}, T_{\varphi 2}, \ldots, T_{\varphi n}\}$ with $T_{\varphi n}$ stands for the rotation of the proposal region with the angle of φ_n. The amounts of training data have been expanded by rotating all the training data $X = \{x_1, x_2, \ldots, x_n\}$ with $T\varphi$. Finally, we construct the new data set i.e,$\mathcal{X} = \{X, T_\varphi X\}$, which has been utilized jointly to increase the training data.

B. ROTATION-INVARIANT GLOBAL GRADIENT DESCRIPTOR

Because the shape of the object is symmetrical, the histogram of oriented gradients (HOG) descriptor [29] is widely utilized in objects detection. In the course of aerial photography, aircraft adopts multi-angle and multi-direction shooting mode making objects have multiple rotation angles. Traditional HOG descriptors do not have rotation invariance, thus it is difficult to employ the HOG descriptor dealing with vehicle

detection task in aerial images. To address this shortfall, the PCA method was used to compute the direction of dominant axis, then the object candidates would be rotated to the setting direction [30]. Analogously, Xu *et al.* [31] applied the segmentation method to estimate the direction of the targets. However, the estimated dominate orientation was not very efficient and accurate. Taking this into consideration, this paper applies the radial gradient transform (RGT) [32] which could compute radial and tangential gradients to achieve rotation invariant. Furthermore, we improved approximate RGT (ARGT) with lookup table to reduce computation of the algorithm.

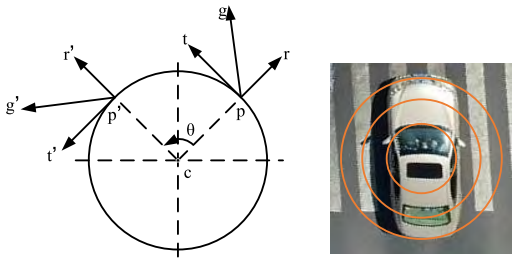

FIGURE 3. (a) Illustration of the radial gradient transform. (b) Extract RGT features from three annular spatial bins.

1) RADIAL GRADIENT TRANSFORM

As the Fig. 3 shows, r and t are two orthogonal basis vectors which are associated with the point p, and c is the center point of the proposal region. The direction of vector r is from the point c towards the point$p$. Meanwhile, unit vector t is the tangential directions at a point p. By projecting onto r and t, we reformulate the gradient g as $\left(g^T\right) r+\left(g^T\right) t$. R_θ represents the rotation matrix for angle θ. If the region has been rotated about its center by the angle θ, a new local coordinate system and gradient will be expressed as:

$$R_\theta = p', \quad R_\theta r = r', \; R_\theta t = t', \; R_\theta g = g' \tag{2}$$

We can obtain a new radial gradient vector $\left(g'^T\right) r' + \left(g'^T\right) t'$ It can be easily proved that these two vectors are invariant to the rotation:

$$\begin{aligned}\left(g'^T r', g'^T t'\right) &= \left((R_\theta g)^T R_\theta r, (R_\theta g)^T R_\theta t\right) \\ &= \left(g^T {R_\theta}^T R_\theta r, g^T {R_\theta}^T R_\theta t\right) \\ &= \left(g^T r, g^T t\right)\end{aligned} \tag{3}$$

Thus, the gradient vector of each point on the image is invariant when the object was rotated some angle around the center of the patch. To improve distinctiveness while maintaining rotation invariance, we subdivide the patch into annular spatial bins. In order to obtain the rotation invariant of the remote sensing object candidates, we compute RGT features in the annuli which was shown in Fig. 3. Using the radial gradient vector, the rotation-invariant descriptors have been obtained.

2) ARGT BASED ON LOOKUP TABLE

However, it is very complicated to directly compute the HOG in RGT method. Takacs *et al.* [32] proposed approximate RGT to reduce computation load. We improved ARGT method with lookup table which could save the gradient direction information for each point and eliminate the projected calculation. The improved algorithms are as follows: assume that the local-coordinate of the point w is (u, v), then we compute the ordinal number of radial coordinate axis:

$$R_\theta = p', R_\theta r = r', R_\theta t = t', R_\theta g = g' \tag{4}$$

$$\begin{cases} i_r = rem\left(floor\left(\frac{N}{2\pi}\phi\left(u, v\right) + \frac{1}{2}\right), N\right), \\ i_t = rem\left(floor\left(\frac{N}{2\pi}\phi\left(u, v\right) + \frac{\pi}{2}\right), N\right), \end{cases} \tag{5}$$

where N represents the number of approximate transform directions, we set $N = 8$. And *floor* (.) means obtaining the maximum integer which is no larger than the input value, *rem* $(., N)$ outputs the remainder that divide inputs by N. Then we can determine the offset (δ_r, δ_t) by the results of formula 4.

$$(\delta_r, \delta_t) \begin{cases} (1, 0), & if \; i = 0 \\ (1, 1), & if \; i = 1 \\ (0, 1), & if \; i = 2 \\ (-1, 1), & if \; i = 3 \\ (-1, 0), & if \; i = 4 \\ (-1, -1), & if \; i = 5 \\ (0, -1), & if \; i = 6 \\ (1, -1), & if \; i = 7 \end{cases} \tag{6}$$

The radial gradient vector g will be computed by formula (6):

$$\begin{bmatrix} g^r \\ \overline{g^t} \end{bmatrix} = \begin{bmatrix} I(u + \delta_u^r, v + \delta_v^r) - I(u, v) \\ I(u + \delta_u^t, v + \delta_v^t) - I(u, v) \end{bmatrix} \tag{7}$$

We use the local-coordinate(u, v) and the calculation results from formula (5). Generating the LUT as follows: $\left\{u_1, v_1, i_1^r, i_1^t, \ldots, u_k, v_k, i_k^r, i_k^t\right\}$. Vector quantizer method was utilized to divide the radial gradient into seventeen specific bins. We will get seventeen-dimensional histogram from the gradient image by performing radial gradient transform. We divide each proposal region into three concentric circles. Then, the feature vector $f = [f_1, f_2, \ldots, f_{51}]$ denote the radial gradient in 51 bins. The Fig.4 shows that the radial gradient features extracted from the same example with arbitrary rotation is obviously invariant.

C. CASCADE FOREST MODEL

Random forest has been successfully used to object detection application. The cascade forest model is the multi-layer structure based on the Auto-context method. Inspired by ensemble learning [13], each layer of cascade combined two random forests and two complete-random forest, and each forest is trained independently. Fig. 2 demonstrates the framework of the proposed cascade forest model.

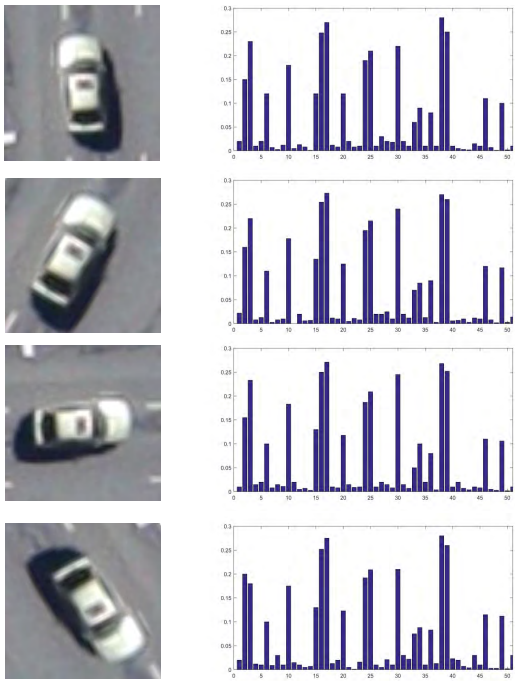

FIGURE 4. Gradient statistics describe the radial gradient histogram based on annulus between three circles. The x-coordinate is the 51 signed orientation bins; the y-coordinate is the gradient statistic information.

1) CASCADE FOREST

In this section, we directly introduce the training/testing process of random forest and complete-random tree forest.

In the training stage, the rotation-invariant vector $X_i = [f_1, f_2, \ldots, f_{51}]$ represents radial HOG of the proposal region i-th, and let y_i be the label of the region i-th. Then the training set D can be denoted as:

$$D = (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n), \quad y_i = 0 \; or \; 1 \tag{8}$$

Random forest is composed of numerous decision trees. It is worth denoting that the decision tree will be trained independently. Each decision tree is composed of leaf nodes and splitting nodes. The spilt function was applied to distinguish positive training set between negative training set:

$$S_L = X \in D | f_k < \tau \tag{9}$$

$$S_R = X \in D | f_k > \tau \tag{10}$$

where, f_k denotes the rotation-invariant features extracted from object candidates, and τ is the random or pre-defined threshold value which separates the samples into left and right sub-nodes (corresponding to S_L and S_R). Decision trees select the features from the feature subset, according to the best *gini* value:

$$Gini(T) = 1 - \sum_{i}^{c} p_i^2 \tag{11}$$

$$Gini(T) = \sum_{j=1}^{k} \frac{|T_j|}{|T|} Gini(T_j) \tag{12}$$

In the testing phase, the rotation-invariant features extracted from positive/negative samples are sent into the forest model, and it can be driven to the leaf node of decision tree. The split node will then divide the samples into different leaf nodes (i.e., arrive at left leaf node if $f_k < \tau$, and go right otherwise). N_l denotes the number of positive samples in leaf node, and N_r denotes the number of negative samples in leaf node. Then the proportion of samples going to left are used as the probability of the positive sample:

$$p(c = 1|x_i) = \frac{N_L}{N_L + N_R} \tag{13}$$

The classification results of each forest are obtained by averaging the results of all trees in random forest.

$$p(c = y|x_i) = \frac{1}{k} \sum_{j=1}^{k} p_j(c = y|x_i) \tag{14}$$

where k is the number of trees in forest, p_j represents the output of j-th tree, y (y= 0,1) is the label of testing set.

Similar to the random forest, complete-random tree forest has been utilized to improve the robust performance of proposed method. Each complete-random tree forest contains multiple complete-random trees [33]. Different from the decision tree of random forests, the training process of complete-random tree is by splitting the samples into different nodes with random features. The complete-random tree will stop expanding when there are only the same class of instances at the leaf nodes. Hence, the output of each forest is a two dimensional class vector which includes the probability of vehicle target and negative target.

2) AUTO-CONTEXT

In the combined random forest, the classification results are estimated from rotation-invariant features extracted by radial gradient transform. To obtain better detection performance, we further incorporate context features to enhance the training process. Particularly, we apply an auto-context model [34] to refine the prediction results iteratively. Fig. 5 illustrates the structure of multi-layer random forest model. It must be noted that, the first layer only use the original feature as input. The amounts of cascade layers is five. Specifically, for detection vehicle in aerial images, after the 1st-layer cascade forest, a class vector can be obtained which denotes the context feature of the proposal regions. The new representation features which concatenate the context feature and the original appearance features will be sent into the 2nd-layer cascade forest. The rotation-invariant features extracted from positive/negative samples has been considered as the original appearance features. The proposed approach has perform the same process iteratively, resulting in a string of cascaded forests. The original appearance features are incorporated with the updated context features which will improve the performance of the cascaded forests obviously with this layer-by-layer manner.

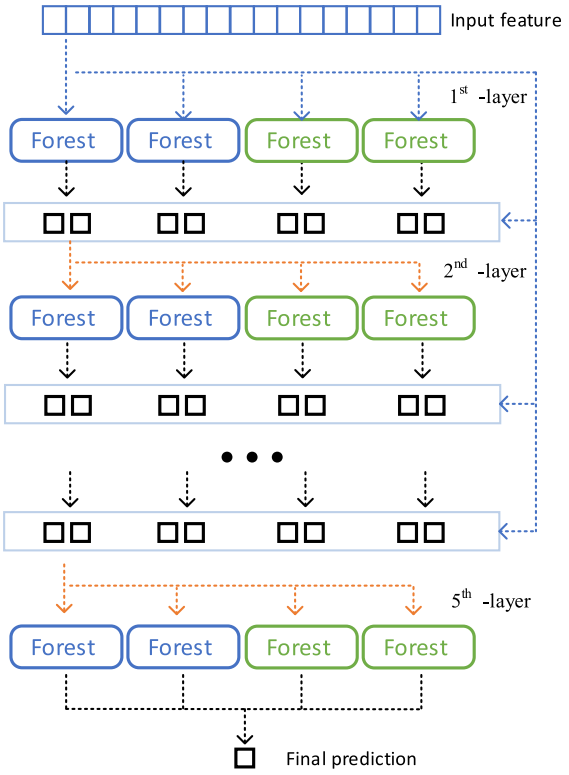

FIGURE 5. Illustration of cascade forest. There are 5 layers in proposed model. Each layer of the Cascade Forest includes 2 random forests (blue) and 2 complete-random forests (green).

As Fig. 5 shows, the number of layers in proposed model is 5. There are 2 complete-random tree forests and 2 random forests in cascade layer. Each complete-random tree forest contains 500 complete-random trees, and similarly each random forest include 500 random trees. In the detection stage, the results of each forest is defined as a two dimensional class vector. And then we concatenate class vector from same layer and send it into the next layer, iteratively. In this way, the last layer of cascade forest will average the class vector from four forests in this layer and out put the final classification results with high accuracy. Furthermore, the influence of hyper-parameters to the detection results will be studied in Section 4.

IV. EXPERIMENTS

In this section, we test the proposed approach experimentally and evaluated the performance of RICF method on the Munich vehicle dataset. We also compare our approach with several advanced methods under the same experiment setting. The experiments are applied on a computer with Intel core i7-7700 CPU, a NVIDIA GTX-1060 GPU, and 16 GB RAM. The operating system was Ubuntu 16.04.

A. DATASETS DESCRIPTION

We test the performance of the proposed method on two benchmark datasets: the DLR Munich vehicle dataset [14] and the UAVDT dataset [15]. The dataset is annotated with rectangular bounding box surrounding the vehicles.

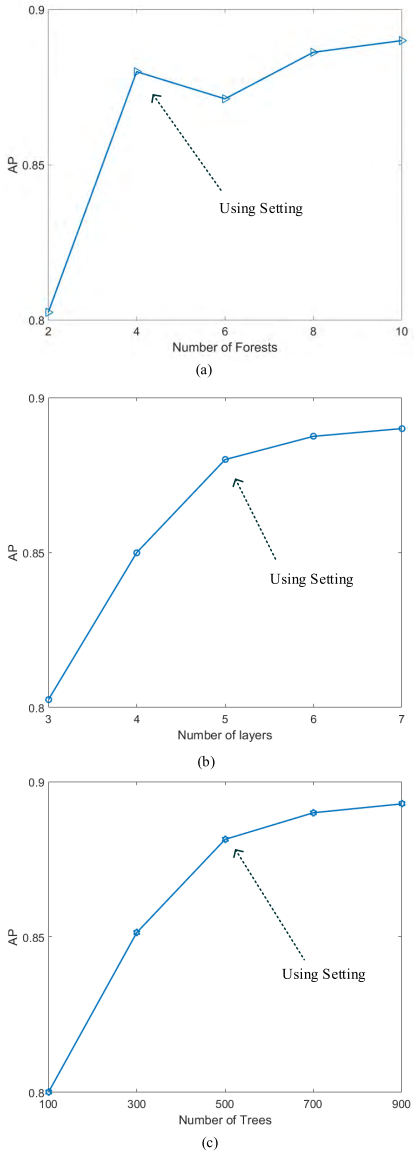

FIGURE 6. Influence of hyper-parameters.The larger models, the better performances will be obtained. But it will cost more computational resource. (a) Performance with increasing number of forests. (b) Performance with increasing number of layers. (c) Performance with increasing number of trees.

1) DLR MUNICH DATASET

The data set is a publicly available aerial image dataset provided by online by the Germany Aerospace Center. It contains 20 images, covering the city of Munich, with the size of 5616 pixels × 3744 pixels. Munich dataset was taken at a height of 1000m by DLR 3K camera, and the ground sampling distance is approximately 0.13 m. The total amounts of vehicles annotated in the training data is 3418, and there are 5799 vehicles in the test set. Each car contains approximately 40×20 pixels.

2) UAVDT

UAVDT is the benchmark for visual tasks, such as detection, multiple object tracking. For vehicle detection, there

TABLE 1. Comparison results of various detection methods on Munich vehicle dataset.

Method	Ground turth	True Positive	False Positive	Recall Rate	Precision Rate	F1-score
ACF 2015 [10]	5892	3078	4602	52.24%	43.31%	0.47
Faster R-CNN 2017 [12]	5892	4050	503	68.74%	88.95%	0.78
H-Fast 2017 [35]	5892	4363	696	74%	86.2%	0.80
AVPN_large 2017 [36]	5892	4538	630	77.02%	87.81%	0.82
Our method	5892	5367	329	**91.1%**	**94.23%**	**0.93**

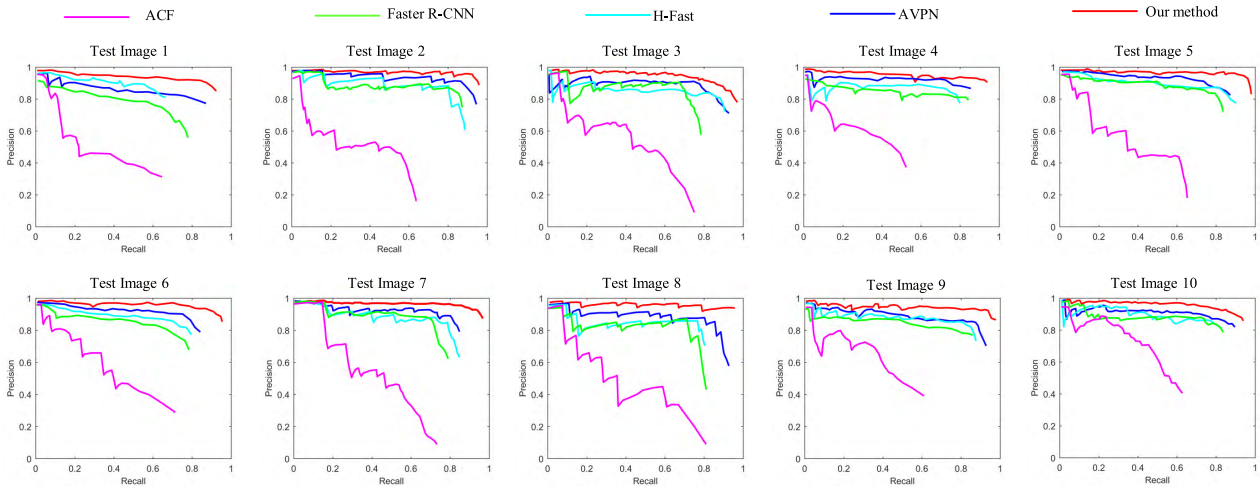

FIGURE 7. PRCs of the proposed RICF method and other state-of-the-art approaches for vehicle detection in ten test images.

are 80000 frames are manually labeled with 0.84 million rectangular bounding boxes. Furthermore, this dataset contains multiple aspects (e.g., camera view, and flying altitude), which is suitable to verify the robustness of the proposed method.

B. EVALUATION CRITERIA

To quantitatively assess our method for vehicle detection, we introduce several widely utilized evaluation criteria: the precision-recall curve (PRC), average precision (AP), recall rate and F1-score. The precision reflects the fraction of true positive detections, and the recall represents the fraction of correctly identified positive detections. The formulation is given as follows:

$$precision = \frac{TP}{TP + FP} \tag{15}$$

$$recall = \frac{TP}{TP + FN} \tag{16}$$

where TP, FP, and FN denote the number of correctly detected vehicles, the number of false detected vehicles, and the number of undetected vehicles. The precision-recall curve shows the recall and precision of the detector on the whole dataset. The average precision denotes the surrounding area of curve, which has positive correlation with the performance of vehicle detection. Furthermore, the formula of F1-score is shown as follows:

$$F1 = \frac{2 * recall * precision}{recall + precision} \tag{17}$$

In brief, AP and F1-score are two crucial parameters which could evaluate proposed method.

C. IMPLEMENTATION DETAILS AND RESULTS ANALYSIS

The RICF method has been tested on the DLR Munich Dataset. In the first stage, we utilized a data-driven method [27] to generate the object-like region with the IoU exceeding 0.6. The training data was augmented by rotating proposals at intervals of 15° (where $K = 23$). In the next stage, RIFF descriptors was calculated by lookup table. Since the size of each vehicle in aerial images is approximate 40×20 pixels, we set the center of the proposal region as the center of the circle, and generate three concentric annulus with five pixels as the radius. A 51-dimensional feature $f = [f1, f2, \cdots, f51]$ will be obtained by vector quantizer method [32]. Then send the rotate-invariant descriptors into the RICF and train the model. In the cascade forest model, the value of hyper-parameters is critical, which directly effects the performance of the proposed method. Fig. 6 indicates the relation between the performance (measured with evaluation metrics) and the value of the hyper-parameters (the layers of cascade forests, the number of the forests, and the amounts of trees) on using test data set. Therefore, we draw the following conclusions: The average precision and recall

FIGURE 8. Detection and annotation results from the Munich test aerial images. A red box denotes correct localization, a yellow box denotes false alarm and a blue box denotes missing detection.

rate raise rapidly and then tend to stabilize with the increment of the value of hyper-parameters. Specifically, by taking both the computation load and detection accuracy into account, a good performance/calculation tradeoff was obtained (Recall is greater than 91% and AP is about 94%) while the number of layers is 5, each layer contains 4 forests, and the amounts of trees in each forest is 500. That shows that the cascade forest model is effective for distinguishing a positive sample and a negative sample using the limited calculation.

D. RESULTS ON MUNICH IMAGES

Using the trained RICF model, we performed vehicle detection on the Munich Vehicle dataset. Four competitive approaches were utilized for the comparison:

1) ACF

The aggregated channel features detectors is the classic method designed by the work [10].

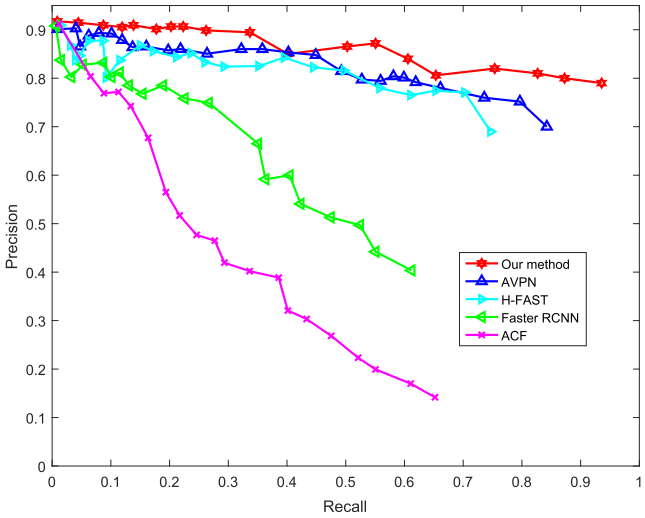

FIGURE 9. Precision-recall curves of test results by different methods in UAVDT.

2) FASTER RCNN

An advanced object detection method which is the combination of the fast RCNN and region proposal network [12].

3) H-FAST

It is the combination of HPRN and Fast RCNN which is used in [35].

4) AVPN_LARGE

Heretical feature maps is used in AVPN [36], which is conducive to detect small sized objects. AVPN_large utilized the bigger training data than AVPN method to get more accurate performance in vehicle detection.

As can be seen from Table. 1, the proposed method has achieved better results than other state-of-the-art methods according to recall, precision, and F1-score. To be precise, the recall rate outperforms the second best detector by 14.68%. The precision rate reaches a competitive level 94.23%. And the F1-score also achieves 13% improvements. It can be seen that our method has obtained the highest true positive rate and the lowest false positive rate. Fig.7 illustrates the PRCs of the four stat-of-art methods and our approach in the Munich dataset. The results demonstrate that our proposed method has better performance than other comparative methods in all test images. Fig. 8 shows several results of the detection on the test images. The results indicates that the proposed method is efficient to detect the vehicle with various rotation.

E. RESULTS OF THE UAVDT DATASET

To further verify the performance of our approach, we evaluated it on the UAVDT dataset as well. It is worth noting that we didn't change the hyper-parameters in the RICF model. Fig.9 shows the precision-recall curves of test results by our method and the above four methods in UAVDT data set. It can be seen that our method obtains better results compared with others. As shown in the Fig. 10, most of the vehicles in the aerial images have been detected by our proposed method. However, there are some vehicles have not been detected accurately due to the change of weather condition. In a future

FIGURE 10. Detection and annotation results from the UAVDT. A red box denotes correct localization, a yellow box denotes false alarm and a blue box denotes missing detection.

research, we will consider the influence of different weather conditions on the detection results.

V. CONCLUSION

In this study, we develop a novel vehicle detection method named Rotation-Invariant Cascaded Forest (RICF). The data augmentation manner is used to expand the training data of vehicle detection. Rotation-invariant descriptors are applied to address the problem of different orientations. Furthermore, cascade forest based on auto-context is utilized to improve the accuracy and robustness of the proposed method. The performance of RICF has been verified by the comprehensive experiments. In our future study, we will focus on improving the robustness of the proposed method under different weather conditions.

REFERENCES

[1] H. Xu, Z. Zhen, B. Sheng, and L. Ma, "Fast vehicle detection based on feature and real-time prediction," in *Proc. Int. Symp. Circuits Syst.*, May 2013, pp. 2860–2863.
[2] Y. Tang, C. Zhang, R. Gu, P. Li, and B. Yang, "Vehicle detection and recognition for intelligent traffic surveillance system," *Multimedia Tools Appl.*, vol. 76, no. 4, pp. 5817–5832, Feb. 2015.
[3] X. Wen, L. Shao, W. Fang, and Y. Xue, "Efficient feature selection and classification for vehicle detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 508–517, Mar. 2015.
[4] S. Wen, Y. Wen, L. Gang, and L. Jie, "Car detection from high-resolution aerial imagery using multiple features," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 4379–4382.
[5] S. Kluckner, G. Pacher, H. Grabner, H. Bischof, and J. Bauer, "A 3d teacher for car detection in aerial images," in *Proc. 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[6] K. Aniruddha, H. David, and L. S. Davis, "Vehicle detection using partial least squares," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1250–1265, Jun. 2011.
[7] Z. Chen *et al.*, "Vehicle detection in high-resolution aerial images via sparse representation and superpixels," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 103–116, Jan. 2015.
[8] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2017.
[9] C. Gong, 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.
[10] L. Kang and G. Mattyus, "Fast multiclass vehicle detection on aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1938–1942, Sep. 2015.
[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.
[12] S. Ren, K. He, R. Girshick, and S. Jian, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[13] Z. Zhou and J. Feng, "Deep forest: Towards an alternative to deep neural networks," in *Proc. 26th Int. Joint Conf. Artif. Intell. (IJCAI)*, Melbourne, Australia, 2017, pp. 3553–3559.
[14] K. Liu and G. Mattyus. *DLR 3k Munich Vehicle Aerial Image Dataset*. Accessed: Dec. 31, 2015. [Online]. Available: http://pba-freesoftware.eoc.dlr.de/3K_VehicleDetection_dataset.zip
[15] D. Du *et al.*, "The unmanned aerial vehicle benchmark: Object detection and tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 370–386.
[16] H.-Y. Cheng, C.-C Weng, and Y.-Y. Chen, "Vehicle detection in aerial surveillance using dynamic Bayesian networks," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2152–2159, Apr. 2012.
[17] T. Moranduzzo and F. Melgani, "Automatic car counting method for unmanned aerial vehicle images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1635–1647, Mar. 2014.
[18] L. Liu and Z. Shi, "Airplane detection based on rotation invariant and sparse coding in remote sensing images," *Optik*, vol. 125, no. 18, pp. 5327–5333, Sep. 2014.
[19] H. Zhou, L. Wei, D. Creighton, and S. Nahavandi, "Orientation aware vehicle detection in aerial images," *Electron. Lett.*, vol. 53, no. 21, pp. 1406–1408, Oct. 2017.
[20] Z. Lei, T. Fang, H. Huo, and D. Li, "Rotation-invariant object detection of remotely sensed images based on texton forest and hough voting," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1206–1217, Apr. 2012.
[21] L. Breiman, "Arcing classifier (with discussion and a rejoinder by the author)," *Ann. Statist.*, vol. 26, no. 3, pp. 801–824, 1998.
[22] M. Zhu, Y. Lang, S. Xia, and P. Hong, "Random Forests for Object Detection," in *Proc. Chin. Intell. Automat. Conf.*, Beijing, China, 2015, pp.267–274.
[23] A. González, D. Vázquez, A. M. López, and J. Amores, "On-board object detection: Multicue, multimodal, and multiview random forest of local experts," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. .3980–3990, Nov. 2017.
[24] R. Marée, P. Denis, L. Wehenkel, and P. Geurts, "Incremental indexing and distributed image search using shared randomized vocabularies," in *Proc. Int. Conf. Multimedia Inf. Retr.*, Mar. 2010, pp. 91–100.
[25] Y. Bo, W. Li, N. Zheng, M. Shakir, and X. Liu, "Constructions detection from unmanned aerial vehicle images using random forest classifier and histogram-based shape descriptor," *Proc. SPIE*, vol. 8, no. 1, Sep. 2014, Art. no. 083554.
[26] S. Wang, X. Liu, T. Yang, and W. Xing, "Panoramic crack detection for steel beam based on structured random forests," *IEEE Access*, vol. 6, pp. 16432–16444, 2018.
[27] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Sep. 2013.
[28] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *CORR*, vol. abs/1311.2524, Nov. 2013, pp. 1–21.
[29] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2005, pp. 886–893.
[30] S. Qi, M. Jie, L. Jin, Y. Li, and J. Tian, "Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1451–1455, Jul. 2015.
[31] F. Xu, J. Liu, M. Sun, D. Zeng, and X. Wang, "A hierarchical maritime target detection method for optical remote sensing imagery," *Remote Sens.*, vol. 9, p. 280, 03 Mar. 2017.
[32] G. Takacs, V. Chandrasekhar, S. Tsai, D. Chen, R. Grzeszczuk, and B. Girod, "Unified real-time tracking and recognition with rotation-invariant fast features," in *Proc. Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 934–941.
[33] F. T. Liu, M. T. Kai, Y. Yu, and Z. H. Zhou, "Spectrum of variable-random trees," *J. Artif. Intell. Res.*, vol. 32, no. 1, pp. 355–384, May 2008.
[34] Z. Tu, "Auto-context and its application to high-level vision tasks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[35] T. Tang, S. Zhou, Z. Deng, H. Zou, and L. Lei, "Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining," *Sensors*, vol. 17, no. 2, p. 336, Feb. 2017.
[36] Z. Deng, S. Hao, S. Zhou, J. Zhao, and H. Zou, "Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3652–3664, Aug. 2017.

BODI MA was born in Shaanxi, China, in 1993. He received the B.S. and M.S. degrees from Northwestern Polytechnical University, Xi'an, China, in 2016 and 2019, respectively, where he is currently pursuing the Ph.D. degree in means of transport applied engineering. His main research interests include image processing, machine learning methods, and UAV.

ZHENBAO LIU (M'11–SM'18) received the B.S. and M.S. degrees from Northwestern Polytechnical University, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree from the University of Tsukuba, Tsukuba, Japan, in 2009, all in electrical engineering and automation. He was a Visiting Scholar with Simon Fraser University, Canada, in 2012. He is currently a Professor with Northwestern Polytechnical University. His research interests include UAV, prognostics and health management, and computer vision. He is an Associate Editor of the IEEE Access.

FEIHONG JIANG was born in Jiangxi, China. He received the B.S. degree from the Xi'an Institute of Technology, in 2003, and the M.S. degree from the Xi'an Jiaotong University, in 2006. He is currently pursuing the Ph.D. degree in means of transport applied engineering with Northwestern Polytechnical University, Xi'an. His main research interests include UAV and flight control systems.

YUEHAO YAN received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 2001 and 2008, respectively. He is currently pursuing the Ph.D. degree in means of transport applied engineering with Northwestern Polytechnical University, Xi'an. He was a Visiting Scholar with Youngstown State University, Youngstown, OH, USA, in 2018. He is also an Associate Professor with the UAV Industrial Technology Research Institute, Chengdu Technological University. His main research interests include control theory and performance evaluation of UAV.

JINBIAO YUAN was born in Inner Mongolia, China. He received the B.S. degree from the Hefei University of Technology, in 2012, and the M.S. degree from the Shaanxi University of Science and Technology, in 2015. He is currently pursuing the Ph.D. degree in advanced manufacturing with Northwestern Polytechnical University, Xi'an. His main research interests include UAV and ground station applications.

SHUHUI BU (M'09) received the M.Sc. and Ph.D. degrees from the College of Systems and Information Engineering, University of Tsukuba, Tsukuba, Japan, in 2006 and 2009, respectively. He was an Assistant Professor with Kyoto University, Kyoto, Japan, from 2009 to 2011. He is currently an Associate Professor with Northwestern Polytechnical University, Xi'an, China. He has published approximately 30 papers in major international journals and conferences, including the IEEE Transactions on Biomedical Engineering, TVC, ACM MM, ICPR, CGI, and SMI. His research interests are concentrated on computer vision and robotics, including 3D shape analysis, image processing, pattern recognition, 3D reconstruction, and related fields.