Object detection in remote sensing images based on deep transfer learning

Jinyong Chen[1] · **Jianguo Sun**[1] · **Yuqian Li**[2] · **Changbo Hou**[2]

Abstract

Object detection is a basic part in remote sensing image processing. At present, it is more common to conduct the topic based on deep learning, however the volume of remote sensing images has become a limitation. In order to solve the problem of small sample of remote sensing image, transfer learning is combined with deep learning in the research. First, the detection problem is caused by insufficient data, such as over-fitting, which is solved by model-based transfer learning. The structure of models and parameters obtained based on natural images are transferred to the detection task in remote sensing target domain. In addition, it is usually assumed that the distribution of training data and the testing data are the same in detection, but this is not the case. Therefore, how to improve the robustness of training models and widen the scope of application should be taken into consideration. In the research, Domain Adaptation Faster R-CNN (DA Faster R-CNN) algorithm is proposed for detecting aircraft in remote sensing images. Two domain adaptation structures are designed and selected as the criterion of similarity measurement between domains. Adversarial training is applied to alleviate the domain shift. Finally, the effectiveness of the algorithm is certified in the low brightness experiment. DA Faster R-CNN detection algorithm improves the accuracy of the original algorithm for low quality images. It is worth noting that the DA Faster R-CNN algorithm is a kind of unsupervised transfer learning method for remote sensing object detection.

Keywords Object detection · Remote sensing images · Transfer learning · Domain adaptation · Adversarial training

✉ Changbo Hou
  houchangbo@hrbeu.edu.cn

1   College of Computer Science and Technology, Harbin Engineering University, Harbin, China

2   College of Information and Communication Engineering, Harbin Engineering University, Harbin, China

1 Introduction

With the rapid development of remote sensing technology, remote sensing information is gradually introduced into more and more application scenes. Object detection is a key link in remote sensing analysis. For the input image, object detection requires not only the recognition results of a semantic category, but also the accurate location information, so it is more challenging. In the high-resolution remote sensing image, in addition to the forest, desert, tennis court and other large objects, aircraft, vehicles, oil barrels and other small targets are also clearly visible. The large volume of data promotes the development and update the remote sensing image processing technology.

Template matching method, feature analysis method and deep learning detection method are the three major directions of remote sensing object detection. Traditional detection methods mainly focus on feature-based research, such as feature extraction, feature selection, and so on [1, 9, 12, 20, 21, 24]. Adjustment and optimization algorithms can improve the detection accuracy and efficiency to some extent; however, these features are common image attributes. It is difficult to distinguish the target and background more effectively. And the extraction and representation of efficient features is not easy.

Since Krizhevsky et al. proposed the AlexNet [10] in 2012, the focus of computer vision research has focused on deep learning. With its full feature expression ability, it has achieved unprecedented excellent results in various fields of image analysis. The same is true for the object detection problem. In recent years, almost all detection methods with outstanding performance have been proposed based on convolutional neural networks [6, 7, 15, 18, 19]. Studies have shown that the end-to-end detection algorithm is the ahead of two-step detection method in speed. But the detection accuracy is sacrificed to some extent [15, 18]. In comparison, the two-step detection method is the process of coarse inspection and fine inspection firsts, so it has more advantages in detection accuracy [6, 7, 19].

In theory, deep learning can characterize the nuances of remote sensing images by extracting high-level features of remote sensing images. However, remote sensing images are not as easy to obtain as natural images. Currently, remote sensing fields do not have enough tagged remote sensing images to train a large number of structural parameters in network models. In addition, since the quality of remote sensing image is greatly affected by the acquisition conditions, in the real scene, many reasons can cause the distribution of data between the training datasets and the test datasets to be different. Therefore, this paper combines transfer learning to research the problem of remote sensing image object detection.

In this paper, the aircrafts are studied as representative examples of the object detection task in high resolution remote sensing image. The specific contributions include the followings: We combine transfer learning with the Faster R-CNN algorithm, which is applied to remote sensing image object detection task. First, models in the natural image source domain are used to solve the problem of limited data volume in the remote sensing target domain. Then, Domain Adaptive Faster R-CNN object detection algorithm is proposed, which is based on the domain adaptation in transfer learning. DA Faster R-CNN object detection algorithm adds two adaptive structures, image-level structure and instance-level structure for domain shift. The improved algorithm can be used in the situation that the training data and test data are in different distribution. What we research is an unsupervised method, which can perform target detection tasks without labeling in the target domain. Compared with the original Faster R-CNN object detection algorithm, the detection average precision of DA Faster R-CNN algorithm is greatly improved.

The remainder of this paper is as follows. In Section 2, a comprehensive survey of remote sensing technology based on transfer learning is given. In Section 3, the basic concepts of transfer's learning, domain adaptation and the Faster R-CNN algorithm are discussed. In Section 4, we describe the details of the proposed work. In Section 5, the experiments and results analysis are shown. Finally, the conclusion is drawn in Section 6.

2 Related work

Transfer learning is one of the most important research topics of machine learning. It can make full use of the knowledge in known fields to solve the related but different learning methods in another field [16]. In practice, transfer learning has unparalleled advantages: it can solve the training problem in the case of sample scarcity, and can also relax the condition that the training data and test data must satisfy the same distribution. Therefore, it is of practical significance to carry out research on this.

In the field of remote sensing image analysis, transfer learning is more useful to solve the problem of classification. In [13], a heterogeneous transfer learning algorithm is proposed to solve the supervised problem, which can make full use of remote sensing information of different dimensions of images, even remote sensing information between different sensors. Lang designed a transfer algorithm combining geometric features and classifier for ship classification, which trains the adaptive support vector machine (A-SVM) for SAR images with the auxiliary of the ship automatic identification system AIS as the source domain [11]. An automatic classification system is designed to solve the problem of land cover classification, combining active learning with transfer learning. This technique can effectively applied to land classification problems of very high-resolution images (VHR) images and hyperspectral images [17]. In 2017, J Xia et al. proposed Ensemble of Transfer Component Analysis (E-TCA) algorithm [22] based on Transfer Component Analysis (TCA), which combines ensemble strategy with transfer learning to solve the domain adaptive problem of hyperspectral remote sensing image classification process. Guo proposed a Temporal-Adaptive Support Vector Machine (TASVM) algorithm based on domain adaptation for multi-temporal remote sensing image classification [8]. In [14], zero-shot scene classification (ZSSC) algorithm is proposed for remote sensing scene recognition, which greatly reduces the demand for label.

Research on object detection based on transfer learning theory is still in its infancy. ZP Dan et al. proposed a model for detecting unlabeled samples in remote sensing images based on LBP algorithm [4]. The model extracts the target domain feature vectors via the LBP algorithm and implements transferring through the hybrid regularization framework. Bin Pan proposed a target detection method based on geometric feature constraints in combination with transfer learning [2]. Two-stage convolutional neural network was designed to improve the efficiency of large-area remote sensing image detection. Chen designed a CNN for road information extraction based on transfer learning [3]. The detection results show that this scheme is superior to the traditional road information extraction method.

However, there are two problems in the current research on remote sensing object detection. One is the scarcity of remote sensing images. The other one is the situation that the training model does not match the target task, which caused by the difference in data distribution. How to improve the robustness of the detection model and solve the problem of mismatch between the training model and the task is also an important research content in the target detection.

In this paper, the transfer learning is used in the remote sensing image target detection task. In order to solve the problem of small sample of remote sensing image, model-based transfer learning is applied. Aiming at the detection problems caused by poor quality of remote sensing images, Domain Adaptive Faster R-CNN object detection algorithm is proposed, which is based on the domain adaptation in transfer learning.

3 Basic theory

3.1 Transfer learning

Transfer can be seen everywhere in real life, such as people who already play badminton are more likely to learn to play tennis because of the high similarities between the two sports. Transfer learning is a learning process based on the similarity between domains. It applies the acquired data, labels, models and other knowledge to different but related new fields. First there are two concepts: the domain $\mathcal{D} = \{\mathcal{X}, P(X)\}$ and the task $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$. The domain $\mathcal{D}$ consists of feature space $\mathcal{X}$ and probability distribution function $P(X)$. The task $\mathcal{T}$ is composed of labels $\mathcal{Y}$ and label prediction functions $f(\cdot)$, in which the label prediction function $f(\cdot)$ can predict the labels corresponding to the new variables.

Under conditions where the source domain $\mathcal{D}_s$ and source tasks $\mathcal{T}_s$ are known, transfer learning is based on the labeled source domain $\mathcal{D}_s = \{(x_1, y_1), \dots, (x_n, y_n)\}$ to learn the label prediction function $f_t(\cdot)$ in target domain [19]. There are two possible cases: one is that the domain is different $\mathcal{D}_s \neq \mathcal{D}_t$, the other one is that the task is different $\mathcal{T}_s \neq \mathcal{T}_t$. Therefore, when the domains are different, it can be expressed as $\mathcal{X}_s \neq \mathcal{X}_t$ or $P(X_s) \neq P(X_t)$, and when the tasks are different, it can be expressed as $\mathcal{Y}_s \neq \mathcal{Y}_t$ or $P(Y_s|X_s) \neq P(Y_t|X_t)$.

3.2 Domain adaptation

Domain adaptation is a branch of transfer learning. In general, transfer learning is the process of solving the prediction function $f_t(\cdot)$ of the target domain based on the source domain knowledge. Domain adaptation is a situation that the tasks are the same $\mathcal{T}_s = \mathcal{T}_t$ but the domains are different $\mathcal{D}_s \neq \mathcal{D}_t$.

Since the domain is consist of feature space $\mathcal{X}$ and distribution function $P(X)$, there are two possible scenarios: one is that the distribution is different $P_s(x_s) \neq P_t(x_t)$, and another is that the feature space is different $\mathcal{X}_s \neq \mathcal{X}_t$. The first domain adaptation scenario is the research content of this paper, that is, the task in target domain is predicted based on the labeled data in source domain when the feature space is the same $\mathcal{X}_s = \mathcal{X}_t$ as well as the probability distribution is different $P_s(x_s) \neq P_t(x_t)$.

3.3 Faster R-CNN

The Faster R-CNN algorithm [11] is what we research in the work. It is gradually developed based on the R-CNN algorithm and Fast R-CNN algorithm.

Convolutional Neural Network is used to solved the object detection problem for the first time in the R-CNN algorithm, which changes the limitation of traditional detection algorithms. Then, in order to reduce the computational complexity in the feature extraction process of the R-CNN algorithm, the Fast R-CNN algorithm is proposed. The above two methods still use

the Selective Search (SS) method to generate region proposal, which consumes a lot of time and restricts the detection speed. The Faster R-CNN algorithm came into being. The Faster R-CNN algorithm is divided into two parts as a whole: one is the region proposal network, the other is the Fast R-CNN for detection. The two parts can short the detection time by sharing the feature map.

The overall framework of the Faster R-CNN network is shown in Fig. 1. First, the CNN is used to extract the feature map of images, and then, the RPN network obtains the region proposal based on the convolutional feature map. The RoI pooling layer maps the output of RPN network back to the corresponding position, and outputs a fixed-size feature map. Finally, the above feature map is fully connected, and the final classification result and the higher-precision location information are determined.

In the training process of Faster R-CNN algorithm, it is necessary to consider the different tasks of two networks, and also consider the connection between networks. There are three implementations to share the convolutional layers between the networks: alternating training, approximate joint training, and non-approximate joint training. The alternating training and approximate joint training are compared in our research.

4 Proposed method

4.1 Model-transfer

At present, many efficient and cutting-edge deep learning algorithms are especially required a large number of images as the basis of research in natural images. However, the disclosed remote sensing image dataset cannot be compared with the natural image dataset in terms of quantity and scale, which restricts the development of deep learning algorithms in remote sensing image analysis to some extent.

The high-resolution remote sensing images studied in this paper refers to the high spatial resolution remote sensing images, of which measurement method is the area corresponding to the unit pixel. At present, the spatial resolution of satellites is generally less than 1 m. Under this condition, the geometric structure of the object is more obvious, the position layout is clearer, and the information is more accurate, such as texture and size. With the increase of resolution, the original difference between natural images and remote sensing images are also gradually reduced, which provides a good feasibility for the introduction of deep learning methods into high resolution remote sensing image analysis.
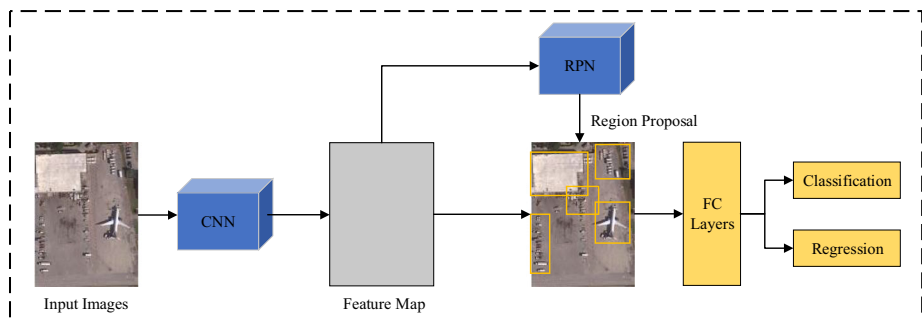

Fig. 1 The block diagram of faster R-CNN

The process of model-based transfer learning is shown in Fig. 2. The source domain is natural images, and the source task is to perform multi-class classification tasks on the natural image dataset. The source models are the ZF network, the VGG_M network and the VGG16 network, which are 1000-classes classification models trained by the ImageNet dataset. The target domain is remote sensing images, and the target task is aircraft object detection in remote sensing images. The specific implementation process is transferred the parameters and structures to initialize the weights of the RPN network and the Fast R-CNN network in the Faster R-CNN. The parameters in the models also need to be modified based on the target task, and then the network is trained in the combination with the aircraft dataset.

4.2 Domain adaptation faster R-CNN algorithm

The research of object detection usually assumpted that the training data and the test data belong to the same distribution, but this is not the case in real-world applications. Many factors may cause the training models won't match the task. In our work, according to the initial setting of domain adaptation, the training image is regarded as the source domain, and the test image is regarded as the target domain. The final task is to obtain a detection model suitable for the target domain data based on the labeled data in source domain, where the labeled contents include both the real bounding box of the object and the category.

Since the domain adaptation does not require the assumption that the data in source domain and the data in target domain data should satisfy the same distribution, which brings a new idea to the performance improvement of the object detection model. The difference between the training datasets and the test datasets in the remote sensing images can be divided into two types, namely the domain shift in the domain adaptation:

(1)   Image-level shift: macroscopic differences caused by factors such as acquisition scales, acquisition angles and lighting;
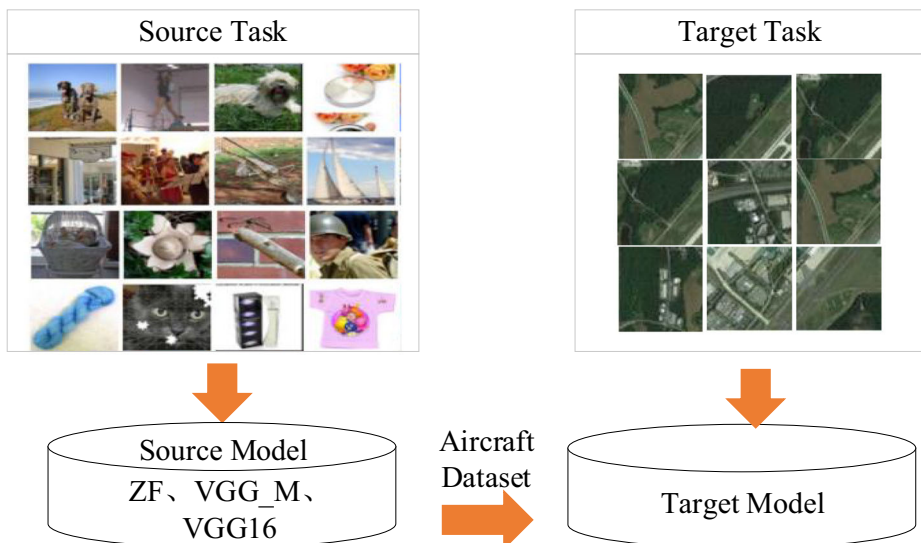

Fig. 2 Model-transfer based on faster R-CNN

(2) Instance-level shift: individual differences caused by the appearance and size of aircrafts.

Therefore, two corresponding adaptation structures are designed for transferring between the two domains. The improved domain adaptation Faster R-CNN algorithm is proposed. K-distance is selected as a measurement of similarity between domains. K is defined as a hypothesis class to represent the domain classifiers, which are used to distinguish whether the sample belongs to the source domain or the target domain.

$$d_{\mathrm{H}}(\mathcal{D}_s, \mathcal{D}_t) = 2(1 - err_s - err_t) \tag{1}$$

where err_s and err_t represent the error probabilities of the domain classifiers.

In the work, the feature vector is represented by x, and the sample from the source domain is recorded as x_s, the sample from the target domain as x_t. At the same time, $h : x \rightarrow \{0, 1\}$ is used to represent the classifier between the two domains, where {0} represents the sample from the source domain and {1} represents the sample from the target domain.

Then the distance between the two domains will be expressed as follows:

$$d_{\mathrm{H}}(s, t) = 2\left(1 - \min_{h \in \mathrm{H}}\left(err_s(h(x_s)) + err_t(h(x_t))\right)\right) \tag{2}$$

Therefore, the distance between the domains is inversely proportional to the prediction error of the classifier; that is, if the error is larger, the distance between the fields becomes closer to zero in this way, the two parts are difficult to distinguish. At this point, the transferring of the source domain to the target domain is achieved.

The networks that produce the feature vector x is denoted as f. In order to align the distribution, the distance $d_{\mathrm{H}}(s, t)$ should be minimized:

$$\min_f d_{\mathrm{H}}(s, t) \Leftrightarrow \max_f \left\{ \min_{h \in \mathrm{H}}\left\{err_s(h(x_s)) + err_t(h(x_t))\right\}\right\} \tag{3}$$

The block diagram of the Domain Adaptation Faster R-CNN algorithm is shown in Fig. 3. The network is divided into two parts. One is the original Faster R-CNN algorithm, and the other is the added domain adaptation structures. There are three modules: image-level domain classifier, instance-level domain classifier and consistency regularization. As shown in Fig. 3, the gradient inversion layer (GRL) is set in the network.
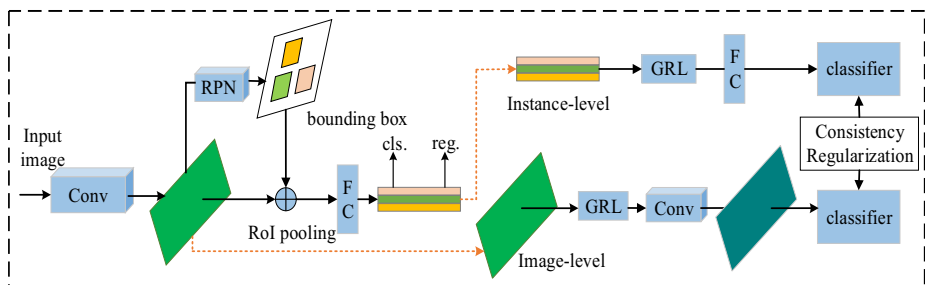

Fig. 3 The block diagram of domain adaptation faster R-CNN algorithm

Based on [5], the adversarial training can be achieved by GRL layer, which is able to maximize the classification error, and minimize the discrimination between source domain and target domain to finish the effective transferring.

(1) Image-Level Domain Classifier: In Faster R-CNN algorithm, the image-level representation of image refers to the feature map via the shared convolutional layer, so the image-level classifier is designed behind the shared convolutional layers in Faster R-CNN algorithm. In order to eliminate image-level distribution shift, the domain classifier will train based on sub-images. Specifically, the activations on the feature map are the training samples.

The label of the ith training image is marked as D_i. If the image comes from the source domain, it is represented as $D_i = 0$, if the image comes from the target domain, it is represented as $D_i = 1$. The ith image has a certain point (u, v) on the feature map F_i output through the convolutional layer, and the corresponding activation of the point (u, v) is expressed as $\phi_{u, v}(F_i)$, the loss function of the image-level adaptation structure is:

$$L_{img} = - \sum_{i,u,v} \left[D_i \log p_i^{(u,v)} + (1-D_i)\log\left(1-p_i^{(u,v)}\right) \right] \tag{4}$$

where $p_i^{(u,v)}$ denotes the output of the image-level domain classifier.

(2) Instance-Level Domain Classifier: In Faster R-CNN, the instance-level representation of an image refers to a feature vector that has not been classified behind the RoI pooling layer. Therefore, the instance-level domain classifier is located behind the RoI pooling layer. Reducing the difference in instance-level is useful for reducing detection errors caused by object factors such as the appearance, size, and the like.

Similar to the image-level domain adaptation described above, it is also necessary to train the classifier based on the feature vector to achieve alignment of the data distribution. The classifier output of the jth region of the ith image is recorded as $P_{i, j}$, then the loss function of the instance-level domain adaptation can be expressed as:

$$L_{ins} = - \sum_{i,j} \left[D_i \log p_{i,j} + (1-D_i)\log\left(1-p_{i,j}\right) \right] \tag{5}$$

(3) Consistency Regularization: Enhancing the consistency of the domain classifiers is benefit to enhance the cross-domain robustness of the RPN network, thus consistency regularization is proposed. Since the image-level domain classifier outputs corresponding score result for each activation on the feature map, the all-activated mean is selected to represent the image-level probability and is used as a regularization term. Then the consistency regularization can be expressed as:

$$L_{cst} = \sum_{i,j} \left\| \frac{1}{|F|} \sum_{u,v} p_i^{(u,v)} - p_{i,j} \right\|_2 \tag{6}$$

where the output of the image-level domain classifier is represented $p_i^{(u,v)}$, and the output of the jth region of the ith image is represented $P_{i, j}$. The total number of activations in the feature map is represented as $|F|$, and the 2-norm is represented as $\|\cdot\|_2$.

The loss function of the domain adaptive network is as follows:

$$L = L_{det} + 0.1 \times \left(L_{img} + L_{ins} + L_{cst} \right) \tag{7}$$

where the $L_{det} = L_{rpn} + L_{roi}$ indicates the loss of original Faster R-CNN algorithm.

5 Proposed method

In this work, the aircraft images in DOTA [23] are selected as the experimental data, a total of 244 images, containing more than 9000 objects. The images in DOTA are collected from many types of sensors and express the diversity of the sample. There is a large number of aircrafts in one image, and most of the images have dense aircraft sample areas. The original

Fig. 4 Aircrafts in DOTA dataset

images in the DOTA dataset are shown in Fig. 4. After the pre-processing of the segmentation by pixels, a total of 1429 sub-images containing the aircraft target were obtained.

The P-R curve and the average accuracy (AP) are selected as the experimental performance evaluation indicators. The P-R curve represents a two-dimensional curve with Recall as the abscissa and Precision as the ordinate. There is a constraint relationship between Precision and Recall. When different thresholds are set, different combinations of Precision value and Recall value can be obtained. As the Recall value grows and the Precision value remains at a very high level, the detection algorithm is excellent.

5.1 Model-transfer experiment

In this paper, the remote sensing images are first rotated at 90°, 180°, and 270°, as shown in Fig. 5. The augmented remote sensing dataset containing 5716 images was divided into two parts randomly, 75% as training set and 25% for the testing set.
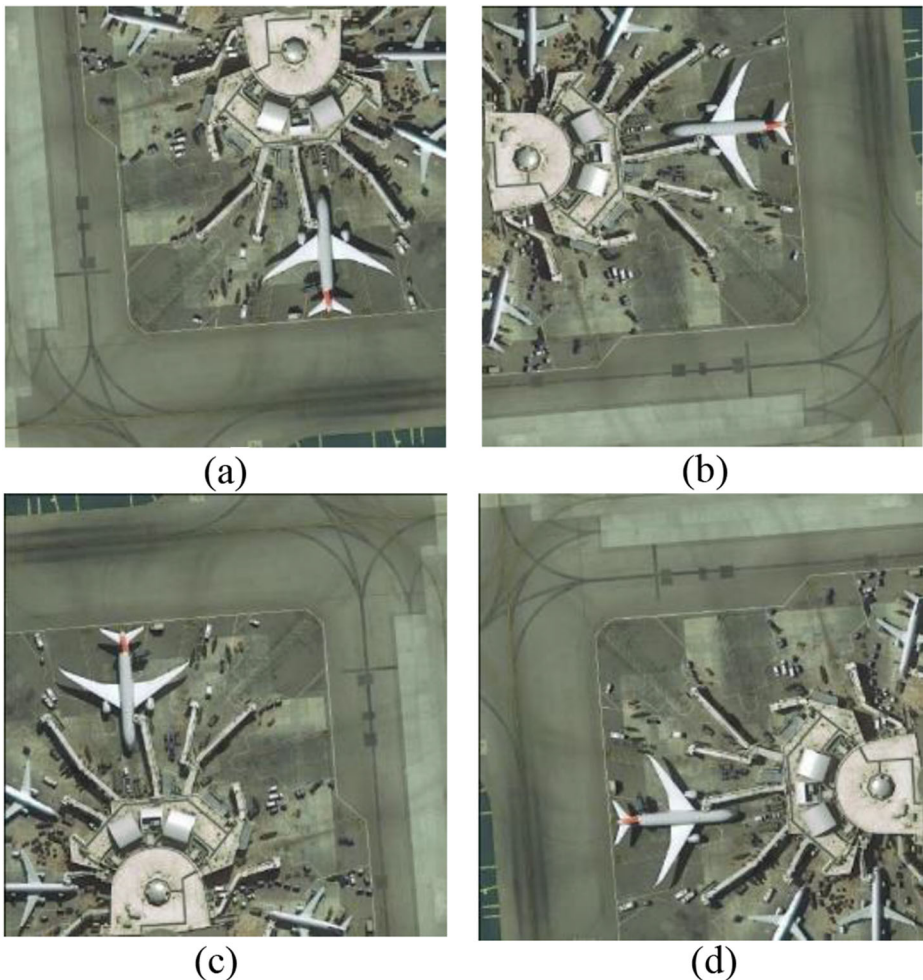

(a)   (b)

(c)   (d)
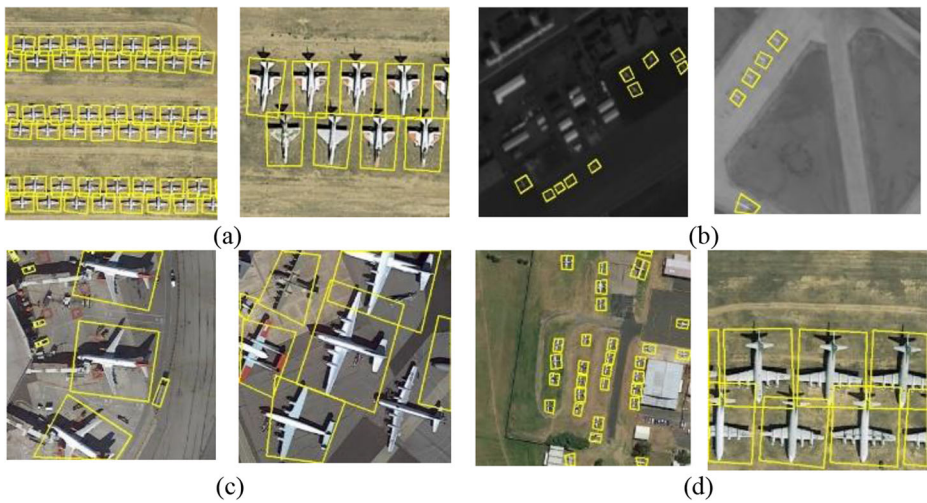
Fig. 5 The aircraft images vis rotation. **a** 0°, **b** 90°, **c** 180°, **d** 270°

Fig. 6 The images in testing set. (**a**) Aircraft graveyard, (**b**) Low brightness, (**c**) Different aircrafts, (**d**) Different heights of IMAQ

It is worth noting that there are various types of samples in testing set, including airliner, helicopters, fighters, etc. The images are collected in a variety of scenes, including civil airports, military airports, and aircraft graveyards. Light intensity is another factor influencing, and some images are collected in low brightness conditions. In addition, the size of the aircrafts is diverse. There are different heights of image acquisition (IMAQ), so there are large differences between the various instances, as shown in Fig. 6(d). The above information indicates that the testing set is complex enough to simulate a real object detection scenario to a certain extent. The more complex the testing set, the higher the requirements of the detection algorithm.

First, we conducted experiments based on ZF network. The P-R curves of alternating training and approximate joint training are shown in Fig. 7. When alternating training is selected, the AP is 89.55%. When approximate joint training is conducted, the AP is 88.13%. The experimental P-R curves show that the Precision value of the algorithm can still maintain a
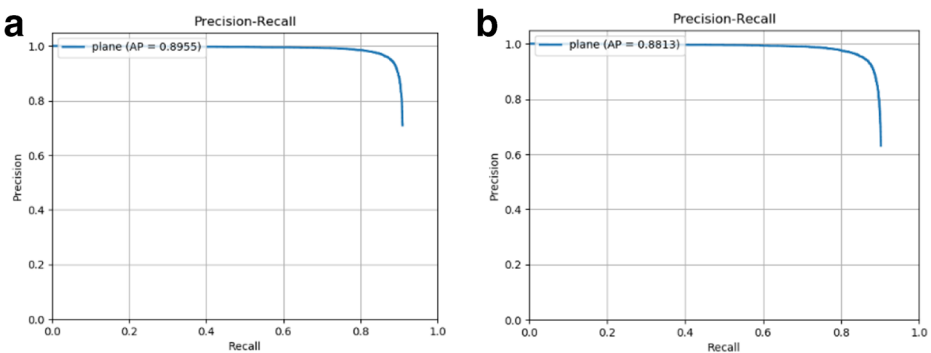

Fig. 7 The compasion of P-R curves. (**a**) The P-R curve of alternating training, (**b**) The P-R curve of approximate joint training

Table 1 Training result statistics for different network models

Model	Iteration time	Total time	Loss_cls	TLoss_bbox
ZF	0.237 s	276.5 min	0.087	0.134
VGG_M	0.326 s	380.33 min	0.078	0.092
VGG16	0.589 s	687.17 min	0.052	0.043

high level as the Recall value grows, which proves that the models have a good detection effect. Combining with the training time to analyze the experimental results, the time of the alternating training is 1.201 times that of the approximate joint training. Therefore, follow-up research also conducts the approximate joint training.

What's more, comparative experiments are set up to analyze the impact of the structure of transfer network on the results. The research work compared ZF network, VGG_M network and VGG16 network. Approximate joint training is selected, and the other experimental settings are the same. The training results are summarized in Table 1. The results show that the training time of VGG16 network is significantly higher than ZF and VGG_M, and the iteration time is also higher than them. The main reason is that ZF is a small network, and the structure of VGG16 is much more complicated than that of ZF. The consumption of training time is due to the increased complexity of the network.

The testing results based on different network models are shown in Table 2. The AP of trained model based on VGG16 network is the highest, reaching 90.17%. Compared with the ZF network, the accuracy is improved by 2.04%, which proves that the larger the network scale, the higher the accuracy. At the same time, the result shows the connection between the detecting time and the models that the larger the network, the more detection time is required.

5.2 Domain adaptation experiment

Due to the complex acquisition conditions of remote sensing images, the samples collected are not all clear images. As shown in Fig. 8, there are some samples under low brightness conditions. At the time, the difficulty of detection is much higher than that of normal brightness. The domain adaptive experiments in this part are used to solve the aircraft detection problem under low brightness conditions.

Since the number of low-brightness images is limited, this research performs brightness reduction processing on the images in DOTA and regards them as the target domain.

In addition, the normal images in the DOTA are selected as the source domain for training as an auxiliary. The source domain data volume in this experiment is 834, the target domain is 1112, and 25% of the target domain images are selected as the testing set. The images labeled

Table 2 Testing result statistics for different network models

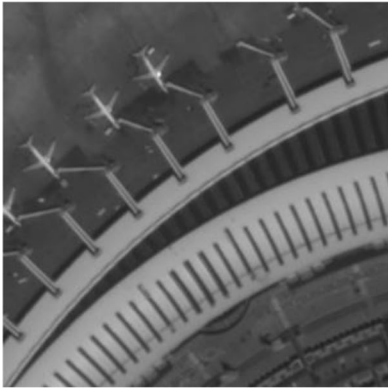| Model	AP	Testing time
ZF	88.13%	0.047 s
VGG_M	89.76%	0.093 s
VGG16	90.17%	0.130 s

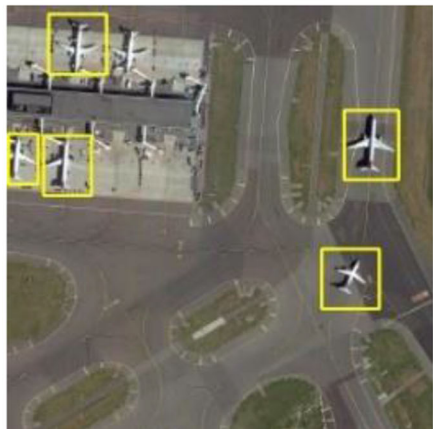Fig. 8 Original low brightness images in DOTA

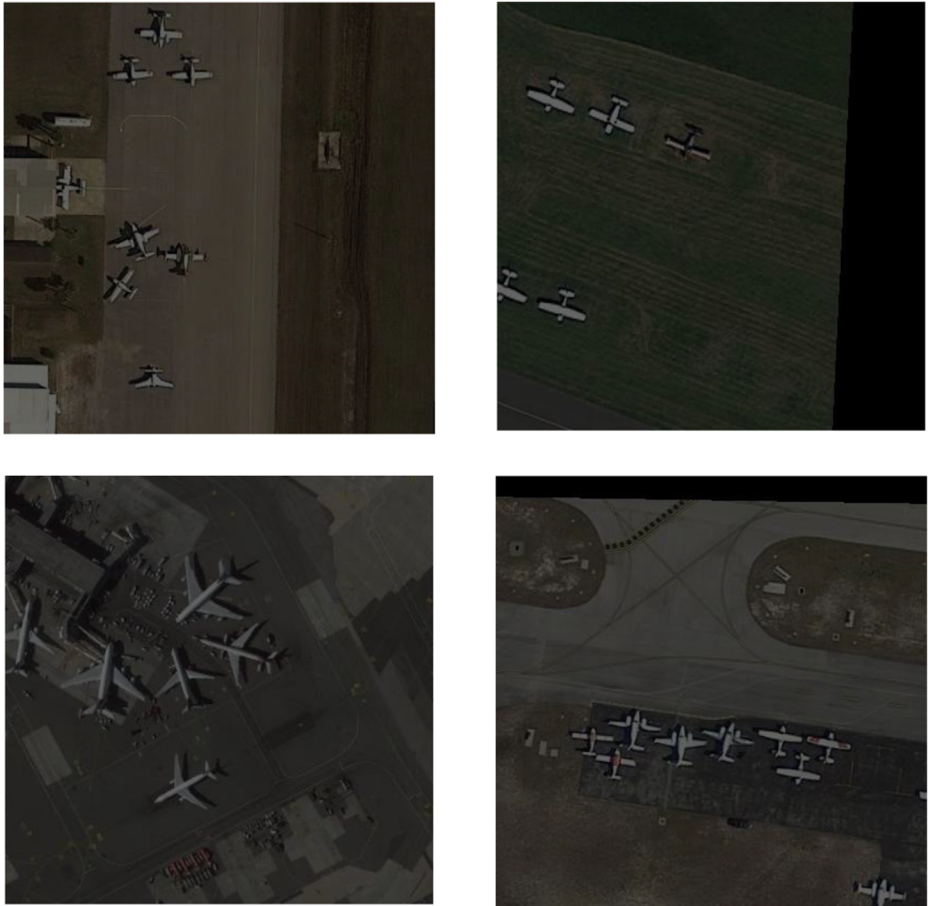
Fig. 9 Samples in the source domain

Fig. 10 Samples in the target domain

in the source domain are shown in Fig. 9. And Fig.10 shows the image samples in the target domain, which are unlabeled.

The results of the DA Faster R-CNN algorithm are as follows: the iteration time is 2.023 s, the total training time is 1333.17 min, the AP is 54.28%, and the testing time is 0.136 s. In order to analyze the influence of different adaptive structures on the detection accuracy, independent contrast experiments were performed. There are four cases of domain adaptation

Table 3 Result comparaison of different domain daptation structures

Method	Img-level	Ins-level	Cons	AP
Faster R-CNN				16.48%
DA Faster	√	0.093 s		44.23%
		√		42.48%
R-CNN	√	√		49.16%
	√	√	√	54.28%

Table 4 Result comparison on different methods

Method	Iteration time	Testing time	Cons
Faster R-CNN	0.589 s	392.67 min	16.48%
DA Faster R-CNN	2.023 s	1338.17 min	54.28%

to the model, as shown in Table 3. When the image-level domain adaptive structure is added, the detection accuracy is 44.23%, which is increased by 27.75%. When the instance-level domain adaptation structure is added, the detection accuracy is 42.48%, which is increased by 26%. When the two work together, the detection accuracy is 49.16%, which is an increase by 32.68%.
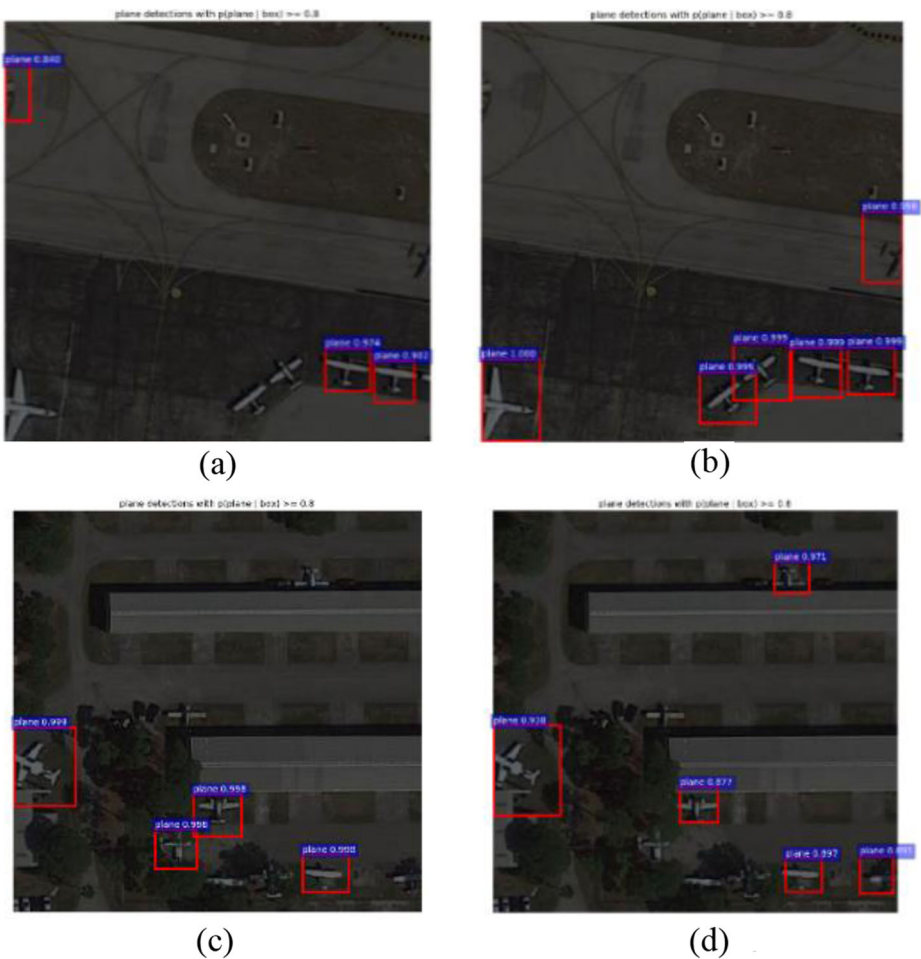

Fig. 11 Testing result of different algorithm. (**a**) Faster R-CNN, (**b**) DA Faster R-CNN, (**c**) Faster R-CNN, (**d**) DA Faster R-CNN

Therefore, the experimental results show that the data distribution of the source data and target data do differ at these two levels, and the image-level differences are larger. At the same time, it is shown that the structures can improve the performance of the detection to some extent. When image-level domain adaptation, instance-level domain adaptation, and consistency regularization are added to the training process, the highest detection accuracy can be achieved, reaching 54.28%. The AP values comparison between DA Faster R-CNN algorithm and the original Faster R-CNN algorithm are shown in Table 4. When testing set is the same, the AP of DA Faster R-CNN algorithm is 54.28%, and the AP of the original Faster R-CNN algorithm is only 16.48%, which is increased by 37.8%.

The above result shows that the improved DA Faster R-CNN algorithm reduces the difference between training data and testing data, and then, when the aircraft detection is performed under low brightness conditions, the higher AP can be obtained by the improved algorithm. In addition, the results in Table 4 also show that the total training time of the original Faster R-CNN target detection algorithm is 3.41 times that of the original Faster R-CNN algorithm, indicating that the detection accuracy is improved at the expense of training time.

The DA Faster R-CNN detection algorithm can detect the aircraft samples missed by the Faster R-CNN algorithm. The detection results are shown in Fig. 11.

6 Conclusion

The aircraft detection task in high-resolution remote sensing image is selected as the research example of our research, combined with transfer learning to improve the detection algorithm. Faster R-CNN algorithm for natural images is applied to the task of remote sensing image aircraft object detection. In order to solve the problem of small sample of remote sensing image, model-based transfer learning is applied. In this paper, two kinds of training methods, alternating training and approximate joint training, are compared. In addition, comparative experiments are set up to analyze the impact of the structure of transfer network on the results.

Aiming at the detection difficulties caused by the poor quality of remote sensing images, DA Faster R-CNN detection algorithm is proposed. Based on Faster R-CNN algorithm, H-distance is used to measure the similarity between domains in our research, and the distance between the domains is minimized by adversarial training. Low brightness domain adaptation experiments demonstrate the effectiveness of the proposed DA Faster R-CNN.

References

1. Acharya A (2014) Template matching based object detection using HOG feature pyramid[J]. Computer Science:689–694
2. Bin P, Jianhao T, Qi Z et al (2017) Cascade convolutional neural network based on transfer-learning for aircraft detection on high-resolution remote sensing images[J]. Journal of Sensors:1–14
3. Chen J, Liu X, Liu C, et al (2018) A Modified Convolutional Neural Network with Transfer Learning for Road Extraction from Remote Sensing Imagery[C]. Chinese Automation Congress (CAC)
4. Dan Z, Sang N, He Y, Sun S (2014) An improved LBP transfer learning for remote sensing object recognition[J]. Optik - International Journal for Light and Electron Optics 125(1):482–485
5. Ganin Y, Lempitsky V (2015) Unsupervised domain adaptation by Backpropagation[C]. 32nd international conference on machine learning. ICML 2:1180–1189
6. Girshick R (2015) Fast R-CNN[J]. IEEE International Conference on Computer Vision (ICCV):2380–7504

7. Girshick R, Donahue J, Darrelland T (2014) Rich feature hierarchies for accurate object detection and semantic segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition:1–21
8. Guo Y, Jia X, Paull D (2017) A domain-transfer support vector machine for multi-temporal remote sensing imagery classification[C]. Geoscience & Remote Sensing Symposium, IEEE
9. Huang F, Zhang X, Xu J, et al. (2019) Multimodal learning of social image representation by exploiting social relations[J]. IEEE transactions on cybernetics
10. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks[C]. International Conference on Neural Information Processing Systems:1097–1105
11. Lang H, Wu S, Xu Y (2018) Ship classification in SAR images improved by AIS knowledge transfer[J]. IEEE Geosci Remote Sens Lett 15:439–443
12. Li Z, Yang C, Xing X (2015) Object detection based on template matching by using enhanced global-best ABC[C]. Control & Decision Conference
13. Li X, Zhang L, Du B et al (2017) Iterative reweighting heterogeneous transfer learning framework for supervised remote sensing image classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10(5):2022–2035
14. Li A, Lu Z, Wang L et al (2017) Zero-shot scene classification for high spatial resolution remote sensing images[J]. IEEE Trans Geosci Remote Sens:1–11
15. Liu W, Anguelov D, Erhan D (2016) SSD: single shot MultiBox detector[J]. European Conference on Computer Vision:21–37
16. Pan SJ, Yang Q (2010) A survey on transfer learning[J]. IEEE Trans Knowl Data Eng 22(10):1345–1359
17. Persello C, Bruzzone L (2012) Active learning for domain adaptation in the supervised classification of remote sensing images[J]. IEEE Trans Geosci Remote Sens 50(11):4468–4483
18. Redmon J, Divvala S, Girshick R (2016) You only look once: unified, real-time object detection[J]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition:779–788
19. Ren S, He K, Girshick R (2017) Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149
20. Wen J, Weng J, Tong C, Ren C, Zhou Z (2019) Sparse Signal Recovery with Minimization of 1-Norm Minus 2-Norm. IEEE Trans. Vehicular Technology 68(7):6847–6854
21. Wen J, Li L, Tang X (2019) Wai ho mow: an efficient optimal algorithm for the successive minima problem. IEEE trans. Communications 67(2):1424–1436
22. Xia J, Yokoya N, Iwasaki A (2017) Ensemble of transfer component analysis for domain adaptation in hyperspectral remote sensing image classification[C]. IGARSS 2017–2017 IEEE International Geoscience and Remote Sensing Symposium. IEEE:4762–4765
23. Xia GS, Bai X, Ding J, et al. (2017) DOTA: A Large-scale Dataset for Object Detection in Aerial Images[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 3974–3983
24. Xiao Q, Hu X, Song G, et al. (2010) Object detection based on contour learning and template matching[C]. Intelligent Control & Automation