2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)

Deep Convolutional Neural Network Based Object Detector for X-Ray Baggage Security Imagery

Jinyi Liu
School of Computer Science and Technology, University of Chinese Academy of Sciences,
Beijing, China
liujinyi16@mails.ucas.ac.cn

Jiaxu Leng
School of Computer Science and Technology, University of Chinese Academy of Sciences,
Beijing, China
lengjiaxu17@mails.ucas.ac.cn

Ying Liu
School of Computer Science and Technology, University of Chinese Academy of Sciences, Key Lab of Big Data Mining and Knowledge Management,
Beijing, China
yingliu@ucas.ac.cn

***Abstract*—Detection and classification of X-ray baggage security imagery are of great significance not only in daily life. However, up to now, it is still difficult for the traditional image processing methods to detect and distinguish objects which we want to detect in X-Ray baggage security imagery. Moreover, it is even more challenging to classify different types of objects. Although the current state-of-the-art detectors have achieved impressive performance on various public datasets with visible images, these detectors fail to deal with objects in X-ray images. This paper proposes an object detection algorithm for X-Ray baggage security screening images. Firstly, to outline the detected object from X-Ray baggage security imagery, we propose a fore-ground-background segmentation method which based on color information. Then, to classify and outline different type of object in X-Ray im-age, a deep convolutional neural networks (DCNNs) based object detection framework Faster R-CNN are proposed. In this stage we also use transfer learning method to speed up Faster R-CNN. The proposed method is proved to achieve 77% mAP by in a real-word dataset which contains 32,253 sub-way X-Ray baggage security screening images.**

***Index Terms*—Convolutional neural network, deep learning, transfer learning, object detection, baggage X - ray security**

I. Introduction

X-ray baggage security screening is widely used to maintain security in public areas such as airports, subways, and railway stations. The screener are trained to identify the items in the baggage quickly and accurately. In public areas, the volume of the passenger flow rises significantly in peak hours, but the number of X-ray screening machines is limited. Therefore, how to identify the items in the baggage quickly is an urgent problem. Both the increasing passenger throughput in the transportation network and an increasing focus on wider aspects of extended public security pose an urgent need in automatic baggage detection task.

Different with optical images, the images under the X-ray screening machine lost the original illumination, color and texture information, only the internal structural information remained. On the other hand, the types and sizes of the objects in the baggage are diverse, and the shielding effects between the objects is serious, which lead to a number of adjustments to the object detection algorithm.

In this paper, we first propose a foreground-background segmentation method to out-line all class of detected object from X-Ray image, in this way, we can delete useless information in X-Ray image. Then we feed images which only contact detected objects into a deep convolutional neural networks based detector. By training the convolutional neural network on X-Ray baggage security imagery, the detector can learn the inner features automatically so that it can recognize the type of each objects show where each object is located in the image (Fig. 1).

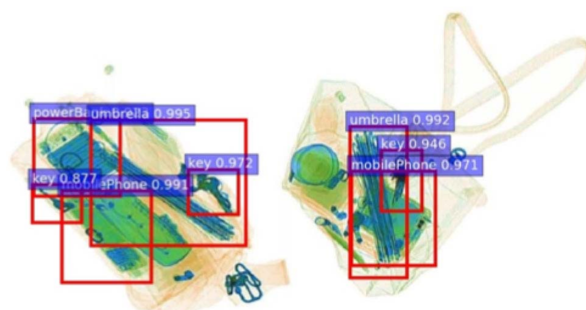

Fig. 1. Exemplar X-ray baggage imagery multiple objects.

The remainder of this paper is organized as follows. Section 2 introduces the related work. Section 3 presents our foreground-back-ground segmentation method in detail. Section 4 presents the object detector. Finally, we conclude our work in Section 5.

II. Related Work

X-Ray Baggage Imagery. Prior work on object detection in X-ray baggage imagery is limited. Most of the work within this context is primarily based on the bag of visual words model (BoVW) [3] and sparse representations [4]. Convolutional neural networks (CNN), is introduced into the field of X-ray baggage imagery by [5], comparing CNN to a BoVW approach with conventional hand-crafted features trained with a Support Vector Machine (SVM) classifier. Following the work of [5], [6] also studies X-ray baggage object classification with

2375-0197/19/$31.00 ©2019 IEEE
DOI 10.1109/ICTAI.2019.00262

CNN, and similarly compares it with the traditional classifiers. Transfer learning and image detector are proposed within the context of X-ray baggage security imagery by [7], exploring the applicability of multiple CNN driven detection paradigms, such as SW-CNN, Fast-RCNN [8], Faster RCNN [9], R-FCNN [10] and YOLOV2 [11].

Classic Object Detectors: The progress of the past decades on various visual recognition tasks are considerably based on SIFT [12] and HOG [**?**]. CNNs have been used for at least two decades, typically on constrained object categories. At present, there are two types of popular detectors, one-stage detector and two-stage detector.

Two-stage Detectors: The dominant paradigm in modern object detection is based on a two-stage approach. As pioneered in Selective Search [14], the two-stage detection algorithm divides the detection problem into two stages. Firstly, region proposals are generated and candidate regions are classified. The typical representation of such algorithm is R-CNN [13], which combines region proposals with CNN for the first time. Over these years, R-CNN was improved in speed and the learned object proposals were proposed to use. As the state-of-the-art detector, Faster-RCNN [1] integrates the proposal generation RPN with the second-stage classifier into a single convolutional network. Numerous extensions to this framework have been proposed, e.g. [15] [16] [17].

One-stage Detectors: Comparing with the two-stage detectors, one-stage detectors do not require the region proposal stage. As it generates the category probability and the position coordinates of the object in one pass, it is a faster detector producing lower accuracy. OverFeat [18] was one of the first modern one-stage object detectors based on deep networks. More recently, SSD [19] and YOLO [20] have renewed interest in one-stage methods. These detectors have been tuned for speed but their accuracy trails that of two-stage methods. A new one-stage detector, RetinaNet, is proposed in [2], which is able to match the speed of previous one-stage detectors while surpassing the accuracy of all the existing state-of-the-art two-stage detectors.

III. Foreground-background Segmentation

As shown in Fig. 2, although the input to the detector is the entire baggage, only part of the input contains objects to be detected, where the remaining background is use-less, such as the schoolbags as well as its strap. However, such useless information occupies a large area of the picture.

Fig. 2. Baggage under X-ray screening machine.

In the detector, all the input images will be resized to 600×600, and then the feature values in the image will be extracted by VGG16. After the convolution layer and the pooling layer of VGG16, the size of the object in the baggage will be reduced, especially the small-sized object, such as the key, may only retain 1-2 pixels or dis-appear completely. In this way, the detector's detection accuracy for small-sized objects will be greatly reduced.

It can be seen from Fig. **??** that different colors are used in the images under the X-ray screening machine to indicate different materials. As shown in Fig. 2, object with large density is darker than that with less density. In contrast with bags, power banks, laptops and mobile phones are composed of circuit boards and batteries, umbrellas, keys and some bottles are made of metal. Therefore, the foreground (the targeting objects) is much darker than the background (bags). Thus, we propose a novel color-based foreground-background segmentation algorithm.

As shown in Fig.3, The color distribution of the six types of objects to be detected is significantly different from that of the background. R(224-255) G(216-255) B(201-255) is selected as the threshold, where the part of the image falls in the threshold will be regarded as the background, otherwise, the foreground.

The output of our proposed segmentation is presented in Fig. 4. Evidently, the pro-portion of the objects to be detected in the original image is greatly increased. If the input image contains multiple bags or the area of the object to be detected in a bag is not connected, our algorithm can automatically split an image into two or more separate images, and then feed them into the detector.

To verify the reliability of our foreground-background segmentation method, we compare the outputs of our method with 32,253 labeled images. It is considered to be correctly if all labeled objects in one image are contained in the area of foreground we segmented. Through calculating, 31,834 images are correctly segmented, the rate of correct segmentation is up to 98.7%

IV. Object detection

The detection of the objects in X-Ray images is to further classify and location of objects in each image. In this work, we need to classify six types of objects in our datasets. To solve this problem, we propose to introduce a color-based foreground-background segmentation method in the data preprocessing stage. In addition, we propose combine a state-of-the-art classifier (VGG16) with a fast detection framework (Faster-RCNN). Faster RCNN (F-RCNN) is a two-stage detector composed of a back-bone network and two subnetworks, backbone network extracted feature map from input images. VGG16 and ResNet are always used as backbone networks. Two sub-networks containing a unique region proposal network (RPN) and a Fast RCNN net-work. The RPN is a full convolutional network that estimate whether the region is an object or background.(shown in Fig.5).

Faster-RCNN includes a feature extraction backbone network and two subnets re-sponsible for detection and classifi-

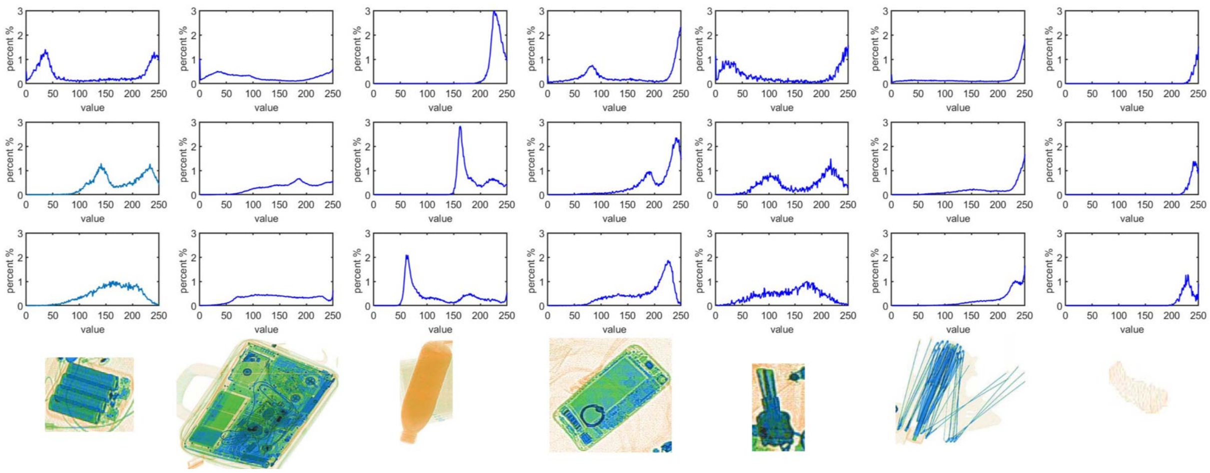

Fig. 3. RGB channel value(top to bottom) in power bank, laptop, bottle, mobile phone, keys, umbrella and background, the abscissa is the pixel value. They-axis indicates the proportion of the pixel in the original image .

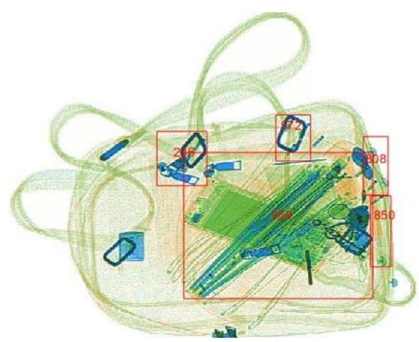

Fig. 4. Segmentation outputs of the proposed color-based foreground-background segmentation algorithm.

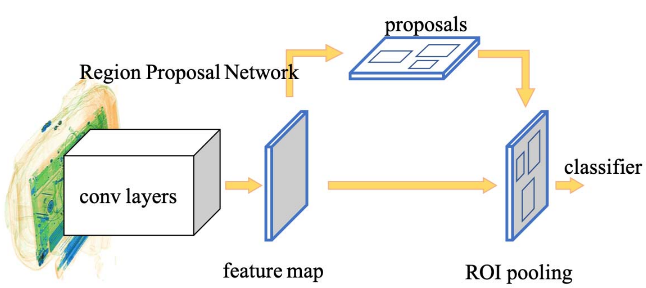

Fig. 5. Structure of Faster R-CNN

cation. If no pre-training model is used, the speed of convergence will decrease. So far, VGG16-based transfer learning model is built on the ImageNet which is an optical image dataset. However, compared with optical image, x-ray baggage imagery lacks illumination, texture, and color information. The picture under the X-ray screening machine only contains its internal structure, and is dyed into different colors according to different materials. As shown in Figure 1, the same object under the X-ray screening machine is completely different with those in optical images, so the ImageNet-based transfer learning model is useless in our data set and completely unable to achieve the expected goals of transfer learning.

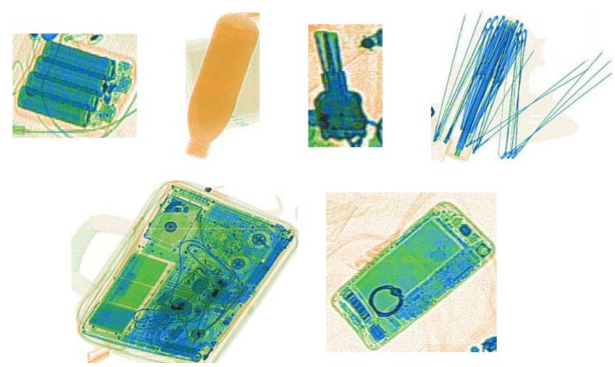

Fig. 6. Six type of detected objects in datasets (power bank, bottle, key, umbrella, laptop and mobile phone)

In order to solve this problem, As show in Fig.6 we cut the object marked in training data and design a VGG16 network with input image of 224×224. As shown in Fig. 7, this paper directly transfer the pretrained VGG16 parameters to VGG16 in Faster-RCNN. In the network, the convergence speed of the model is greatly accelerated and the training time is reduced.

The dataset contains 32,253 real images collected from subway stations in China cities such as Hangzhou, Wuhan and Nanchang. Since dangerous items such as guns and knives are extremely few in our dataset, we marked six types of common objects, they are computers, mobile phones, power banks, umbrellas, bottles and keys, which appear frequently. Finally, we generated 19,570 positive samples and 12,683 negative samples, including 9,540 mobile phones, 2,384 power banks,

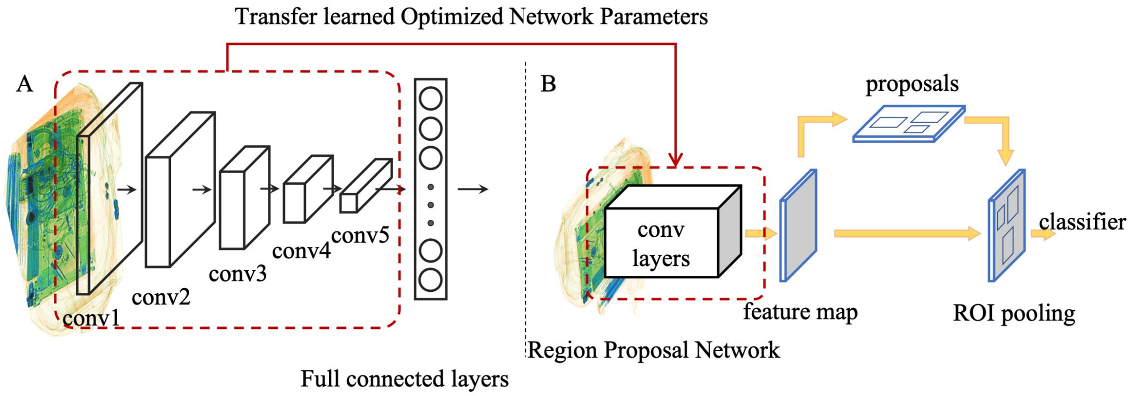

Fig. 7. Transfer learning pipeline. (A) shows classification pipeline for a source task, while (B) is a target task, initialized by the parameters learned in the source task.

7,574 umbrellas, 1,395 laptops, 7,171 bottles and 11,765 keys.

We split the datasets into training (25%), validation (25%) and test sets (50%) such that each split has similar class distribution but unseen test set contains some-what challenging samples never trained before. Besides, we also weight the data when sampling to cope with class imbalances. We also perform random flipping, random cropping, and rotation to each sample to augment the datasets. For learning, we use stochastic gradient descent (SGD) with momentum and weight decay of 0.9 and 0.0005, respectively. The initial learning rate of 0.001 is divided by 10 with step down method in every 10, 000 iteration.

In this paper, we used the open-source toolkit Tensorflow, which uses the NVIDIA GPU and CUDA library (e.g., cuDNN [21]) to speed up the computation. We used the NVIDIA GeForce GTX Titan X edition GPU (with a 12-GB memory) and In-tel(R) Xeon(R) CPU(4 cores) with a 8-GB memory in our experiments.

We compare our method with Faster R-CNN(without foreground-background seg-mentation and transfer learning) and state-of-the-art one stage detector RetinaNet. Faster R-CNN are trained with VGG16 and ReaNet101, RetinaNet are respectively trained with VGG16, VGG19, ResNet50 and ResNet101. We primarily evaluate detection mean Average Precision (mAP), because this is the actual metric for object detection (rather than focusing on object proposal proxy metrics).

TABLE I
DETECTION RESULT OF FASTER-RCNN AND RETINANET

Detector	Backbone Network	mAP
Faster-RCNN	VGG16	0.74
Faster R-CNN	ResNet101	0.74
RetinaNet	ResNet50	0.72
RetinaNet	ResNet101	0.71
RetinaNet	VGG16	0.70
RetinaNet	VGG19	0.71
ourmethod	VGG16	0.77

TableI show the detection result of our method and other state-of-the-art detectors, TableII show the detection outputs of Faster-RCNN and our method for different objects. Those two tables courtesy shows that mAP of our proposed algorithm exceeds the most advanced one-stage and two-stage detectors, and the detection of small objects such as keys is greatly improved, which means that the algorithm we pro-posed is better than state-of-the-art detectors.

TABLE II
DETECTION RESULT OF FASTER-RCNN AND RETINANET

	Faster R-CNN	Our Method
Laptop	0.86	0.87
Moblie Phone	0.84	0.86
Power Bank	0.69	0.71
Umbrella	0.86	0.88
Bottle	0.67	0.69
Key	0.53	0.58
mAP	0.74	0.77

As can be seen from Table. II, mAP of our proposed algorithm exceeds the most advanced one-stage and two-stage detectors, and the detection of small objects such as keys is greatly improved, which means that the algorithm we proposed is better than state-of-the-art detectors.

Fig. 9 illustrates qualitative examples extracted from the statistical performance analysis of Table. II. We see that our detection approaches can cope with cluttered datasets and small objects.

V. CONCLUSION

In this work, we exhaustively explore the use of CNN in the tasks of detection within X-ray baggage imagery. Firstly, we propose a foreground-background segmentation method based on color information. Then we propose an approach for adding context to Faster-RCNN. Finally, we established a transfer learning model based on real X-ray baggage image. This result illustrates that our method outperforming the performance of

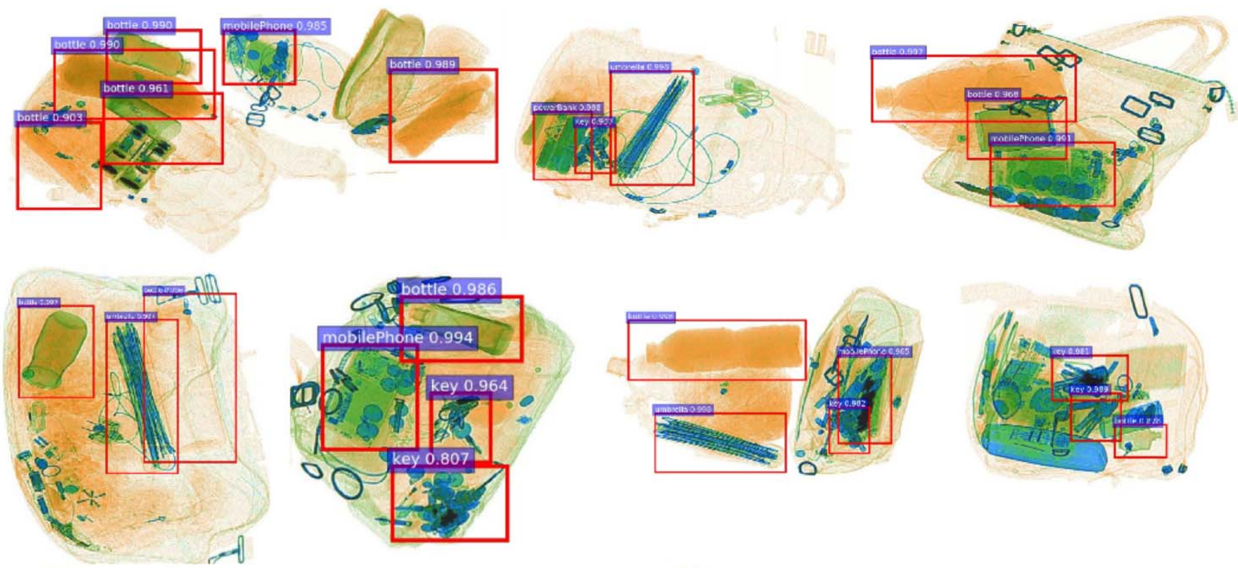

Fig. 8. Transfer learning pipeline. (A) shows classification pipeline for a source task, while (B) is a target task, initialized by the parameters learned in the source task.

the state-of-the-art Faster-RCNN and RetinaNet within this X-ray security imagery context, especially on small objects.

ACKNOWLEDGMENT

This project was partially supported by Grants from Natural Science Foundation of China ♯71671178 /♯ 91546201 /♯ 61202321, and the open project of the Key Lab of Big Data Mining and Knowledge Management. It was also supported by Hainan Pro-vincial Department of Science and Technology under Grant No. ZDKJ2016021, and by Guangdong Provincial Science and Technology Project 2016B010127004.

REFERENCES

[1] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
[2] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE transactions on pattern analysis and machine intelligence, 2018.
[3] M. Ba¸ stan, M. R. Yousefi, and T. M. Breuel, "Visual words on baggage X-ray images," in Computer Analysis of Images and Patterns.Berlin, Germany: Springer, 2011, pp. 360–368.
[4] D. Mery, E. Svec, and M. Arias, "Object recognition in baggage inspec tion using adap-tive sparse representations of X-ray images," in Image and Video Technology. Cham, Switzerland: Springer, 2016, pp. 709–720.
[5] S. Akçay, M. E. Kundegorski, M. Devereux, and T. P. Breckon, "Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery," in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2016, pp. 1057–1061.
[6] D. Mery, E. Svec, M. Arias, V. Riffo, J. M. Saavedra, and S. Banerjee, "Modern comput-er vision techniques for X-ray testing in baggage inspection," IEEE Trans. Syst., Man, Cybern., Syst., vol. 47, no. 4, pp. 682–692, Apr. 2017.
[7] S. Akcay, M. E. Kundegorski, C. G. Willcocks and T. P. Breckon, "Using Deep Convo-lutional Neural Network Architectures for Object Classification and Detection With-in X-Ray Baggage Security Imagery," in IEEE Transactions on Information Forensics and Security, vol. 13, no. 9, pp. 2203-2215, Sept. 2018.
[8] R. B. Girshick. (Sep. 2015). "Fast R-CNN." [Online]. Available: https://arxiv.org/abs/1504.08083
[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards realtime object detec-tion with region proposal networks," in Proc. Adv.Neural Inf. Process. Syst., 2015, pp. 91–99.
[10] J. Dai, Y. Li, K. He, and J. Sun. (Jun. 2016). "R-FCN: Object detection via region-based fully convolutional networks." [Online]. Available:https://arxiv.org/abs/1605.06409
[11] J. Redmon and A. Farhadi. (Dec. 2016). "YOLO9000: Better, faster,stronger." [Online]. Available: https://arxiv.org/abs/1612.08242
[12] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110. **b18** Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Confer-ence on. IEEE, 2005, 1: 886-893.
[13] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object de-tection and semantic segmentation[C]//Proceedings of the IEEE conference on com-puter vision and pattern recognition. 2014: 580-587. 23.
[14] recog-nition[J]. International journal of computer vision, 2013, 104(2): 154-171.
[15] Lin T Y, Dollár P, Girshick R B, et al. Feature Pyramid Networks for Object Detec-tion[C]//CVPR. 2017, 1(2): 4.
[16] Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 761-769.
[17] Shrivastava A, Sukthankar R, Malik J, et al. Beyond skip connections: Top-down modulation for object detection[J]. arXiv preprint arXiv:1612.06851, 2016.
[18] Sermanet P, Eigen D, Zhang X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks[J]. arXiv preprint arXiv:1312.6229, 2013.
[19] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.
[20] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
[21] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[J]. arXiv preprint, 2017.
[22] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[J]. arXiv preprint, 2017.