Research on abnormal object detection in specific region based on Mask R-CNN

Haitao Xiong1,2 @, Jiaqing Wu1, Qing Liu1 and Yuanyuan Cai1

Abstract
As an information carrier with rich semantics, image plays an increasingly important role in real-time monitoring of logistics management. Abnormal objects are typically closely related to the specific region. Detecting abnormal objects in the specific region is conducive to improving the accuracy of detection and analysis, thereby improving the level of logistics management. Motivated by these observations, we design the method called abnormal object detection in a specific region based on Mask R-convolutional neural network: Abnormal Object Detection in Specific Region. In this method, the initial instance segmentation model is obtained by the traditional Mask R-convolutional neural network method, then the region overlap of the specific region is calculated and the overlapping ratio of each instance is determined, and these two parts of information are fused to predict the exceptional object. Finally, the abnormal object is restored and detected in the original image. Experimental results demonstrate that our proposed Abnormal Object Detection in Specific Region can effectively identify abnormal objects in a specific region and significantly outperforms the state-of-the-art methods.

Keywords
Logistics management, abnormal object, object detection, instance segmentation, Mask R-CNN

Date received: 17 January 2020; accepted: 6 April 2020

Introduction
In the background of reform and opening-up policy in China, fully developed, modern information technology has gained explosive development and application. The new management methods, represented by multidimensional, multi-angle, and real-time monitoring, have gradually penetrated into every level of logistics management, which has a profound impact on the management mode of logistics. In particular, these new management methods generate and store unstructured content such as text and massive video every day, carrying very important information, which can strengthen the supervision of logistics, detect the abnormal situation in logistics in time, and reduce the occurrence of accidents.1 As the main monitoring methods, video monitoring is widely used in logistics management, and video is mainly composed of images. Therefore, automatic analysis and processing of

1 School of Computer and Information Engineering, Beijing Technology and Business University, Beijing, China
2 National Engineering Laboratory for Agri-product Quality Traceability, Beijing Technology and Business University, Beijing, China

Corresponding author:
Haitao Xiong, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China; National Engineering Laboratory for Agri-product Quality Traceability, Beijing Technology and Business University, Beijing 100048, China.
Email: xionghaitao@th.btbu.edu.cn
Object detection is an important step in the automatic detection of abnormal objects. The task of object detection mainly focuses on the location information of all kinds of objects in the picture, such as marking the position of the detected object in the picture with a rectangular box. Deep learning makes a breakthrough in object detection research using its powerful feature learning ability. In view of the problems of low detection accuracy in the visual recognition technology in logistics transportation, difficulty in accurately extracting the object for recognition from the video image, difficulty in identifying and classifying the object subclass, difficulty in considering the recognition accuracy and detection efficiency, and so on, Zhu et al. propose a convolutional neural network (CNN) model to learn high-level features for saliency detection. Compared to other methods, their method presents two merits. First, when performing features extraction, apart from the convolution and pooling step in our method, they add restricted Boltzmann machine into the CNN framework to obtain more accurate features in the intermediate step. Second, to avoid manual annotation data, they add depth belief network classifier at the end of this model to classify salient and nonsalient regions. The experimental results show that the average accuracy of the proposed method is improved, and the detection and recognition time is shortened; compared with the traditional classification method and the deep convolution neural network before the improvement, the recognition accuracy and efficiency are significantly improved, and the detection robustness is significantly improved.

Li et al. proposed a new pallet detection algorithm to solve the problem of low detection accuracy of traditional object detection algorithm in the complex environment. They collect a large number of pictures of people and pallets in the real warehouse for labeling, build the pallet database of the logistics warehouse, and improve the basic network of the single multi-box detector detection algorithm into DenseNet network, using the labeled pallet database for training and testing. In the test phase, the multi-scale feature map with a different resolution is combined to enhance the adaptability of the network to the detected object, and a single network is used to achieve the detection task. The experimental results show that the detection accuracy of this algorithm is higher than that of the Yolo algorithm. At the same time, it provides experience for the selection of object detection methods in this article.

In foreign object detection applications, the method based on computer vision has been widely used, and the shape of each object and other information has been obtained. Therefore, it is the premise to segment each object image accurately from the background, which is an important step to achieve the real-time acquisition of the object individual information. Qiao et al. proposed an instance segmentation method based on Mask-R-CNN deep learning framework, which is used to solve the problem of object instance segmentation and contour extraction in the actual environment. This method mainly includes the following steps: key frame extraction (detecting the huge moving frame of the object), image enhancement (reducing the influence of light and shadow), object segmentation, and abnormal object contour extraction. They are trained and tested the proposed method on a challenging image data set. The experimental results show that this method can achieve better segmentation results, the average pixel accuracy is improved, and the average distance error of contour extraction is reduced, which is better than the most advanced SharpMask and DeepMask instance segmentation methods. It also serves as a reference for the selection of instance segmentation methods.

In the process of automatic object detection, instance segmentation is also necessary. Liu et al. propose an improved fully CNN which fuses the feature maps of the deeper layers and shallower layers to improve the performance of image segmentation. In the process of feature fusion, adaptive parameters are introduced to enable different layers to participate in feature fusion as different proportion. The deep layers of the neural network mainly extract the abstract information of the object, and the shallow layers of the neural network extract the refined features of objects, such as edge information and precise shape. Adaptive parameters can speed up the training speed and improve prediction accuracy. In the early stages of training, the feature maps of shallow layers have a larger fusion coefficient that allows the neural network to learn the feature of object’s location and shape quickly. As the training progresses, gradually weakening the fusion coefficient of shallow layers and increasing the fusion coefficient of deep layers can enhance the network’s ability to predict the details of the objects. Experiments show that the method proposed in this article speeds up the training and improves the pixel prediction accuracy. It also provides experience for the selection of neural network in this article.

Recently, neural network has been widely used. Especially, CNN-based object detection methods have shown superior performance of object detection against traditional object detection methods in images. But there are few applications of abnormal object detection in logistics management, such as in the field of logistics and transportation, due to the large number of express goods, it is not uncommon to lose a product. In the process of loading and unloading transportation, the soft and uneven goods at the bottom of the package will often drop when they are transported on the conveyor belt. However, at present, the object detection method is mainly used in logistics management to identify the goods in logistics transportation. It is unable to further distinguish the goods and detect abnormal objects in the
goods, abnormal objects are goods dropped in logistics transportation. For example, the traditional object detection method can only identify all express goods in the belt region and cannot directly check abnormal objects, which express goods are under the belt and find out whether the express goods have dropped. With the development of deep learning research, Mask R-CNN improves on Faster R-CNN by adding a branch of Mask prediction, which puts forward a flexible framework for object recognition and positioning. Mask R-CNN improves the speed and accuracy and makes the instance segmentation more accurate. However, in the prediction stage, the effect of using Mask R-CNN to identify the conveyor belt is not very good, so it is not possible to use Mask R-CNN directly to segment instances of the conveyor belt.

For these reasons, this article designs an Abnormal Object Detection in Specific Region (AODinSR) based on Mask R-CNN. This article studies the problem of abnormal object detection in a specific region, such as a conveyor belt. AODinSR realizes instance segmentation through Mask R-CNN algorithm and overlaps with the specific region to detect abnormal objects, such as belt region and out of the belt region. The application focuses on detecting normal, potential abnormal, and abnormal handling goods according to different regions, focusing on abnormal handling goods, and automatically identifying the falling phenomenon of goods.

**Related work**

Through the development of computer vision, object detection and instance segmentation has been an interesting and meaningful research topic recently. Object detection and instance segmentation are the main methods of image recognition in the process of loading and unloading. It is also one of the main research contents in the field of computer vision. Through all-round and real-time monitoring, we can timely find and pay attention to the real-time situation in the process of loading and unloading transportation of logistics warehouse. In view of the images obtained by real-time monitoring, how to accurately distinguish normal, potential abnormal, and abnormal goods handling and provide data support for the supervision of logistics warehouse is the research focus in the field of logistics management.

Object detection identifies and locates multiple objects in the image. Object detection mainly resolves the following two problems: firstly, what object is available in the image (object recognition); secondly, where the object exists (object location). Therefore, the task of abnormal object detection is not only to confirm the category of the object in the image but also to determine the pixel range of the object. Instance segmentation can mark different individuals of the same object in the image by the boundary box of object detection accurate to the edge of the object. Instance segmentation uses object detection method to frame different instances in the image and makes pixel by pixel prediction in different instance regions. Using the mask as the label of pixel prediction, the mask is a binary image composed of 0 and 1. Generally, the image mask is defined by the region of interest. The function of masking is to extract the region of interest and mask the region of interest on the image. Image segmentation is the basis of image understanding computer vision and is one of the most important steps in the process of image analysis. The segmented region can be used as the object for subsequent feature extraction.

There are two kinds of common abnormal object detection and instance segmentation algorithms: One is CNN series object detection algorithm based on region which is the main research direction due to its high precision, such as Fast R-CNN, Faster R-CNN, Mask R-CNN; the other is to convert the classification problem of abnormal object detection into regression, for example, YOLO, SSD, and so on, but there are some problems, such as low precision and poor detection effect for small objects. Deep learning method is a multilevel feature learning method, which can transform the features of each layer (starting from the original data) into higher level, more abstract-level features. In the field of object recognition, CNN can effectively capture the deep semantic features of images, get a large number of representative feature information, and finally classify and predict the samples with higher accuracy. With the continuous breakthrough of deep learning in the field of computer vision, R-CNN algorithm is a widely used object detection algorithm. However, due to the problem of repeated calculation of feature links, Fast R-CNN algorithm is proposed on the basis of R-CNN. Fast R-CNN performs a feature extraction on the image to be detected in the convolution layer; secondly, the region of interest (ROI) pooling layer is introduced to unify the feature scale; then, the normalized exponential function softmax is used to replace support vector machine (SVM), combining classification and border regression, reducing repeated calculation, and improving the detection speed. Faster R-CNN is an improved algorithm based on Fast R-CNN. It introduces the region proposal network model to carry out two-stage object detection. By generating candidate regions, extracting features, distinguishing feature categories, and correcting the position of candidate frames, the speed and accuracy of detection are greatly improved.

Mask R-CNN is a simple, flexible, and general object instance segmentation framework. Mask R-CNN adds a segmentation branch to Fast R-CNN to segment the instance at the same time of detection. Mask R-CNN improves the loss function of segmentation, from the polynomial cross-entropy based on single-pixel softmax to the binary cross-entropy based on single-pixel sigmoid. If the candidate frame is detected as a category, the binary cross-entropy will make the cross-entropy of this category to be calculated as the error value, only contribute to the ground
truth (GT) of the specific K class, while the other classes do not contribute to the loss, and in the back propagation, the loss function $L$ only calculates and back propagates the GT classes, which decouples the segmentation and classification and effectively avoids the competition between classes. Another improvement is that Mask R-CNN adds the RoI align layer, Mask R-CNN uses bilinear interpolation to make the pooling result of the region of interest closer to the features before the non-pooling, thus reducing the error. Although the structure is simple, with the help of a series of practical technologies such as feature pyramid networks (FPN), Mask R-CNN has achieved good results.

Since the mask branch is added to Mask R-CNN, the loss function $L$ is calculated as follows

$$L = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}}$$

$L_{\text{cls}}$ represents the classification loss function, $L_{\text{box}}$ represents the bounding box location loss function, $L_{\text{mask}}$ represents the instance segmentation loss function.

To improve the accuracy of oriented FAST and rotated BRIEF (ORB) matching of multi-object images, Bo et al. proposed a method of image ORB mismatch removal based on Mask R-CNN. Firstly, the image is recognized by Faster R-CNN method, and the region of interest and category label marked by a rectangle frame is obtained by regional recommendation network. In this step, the predicted category and coordinate information of the region of interest can be obtained, and the pixel-level correction can be carried out by convolution layer of full convolution network to get the category of the pixel-level object, and then the object segmentation is carried out. Finally, on the basis of the original ORB feature point matching, the mismatches outside the same object segmentation region in the two images are eliminated. To verify the effectiveness of this method, the traditional ORB matching and the ORB matching based on this method are simulated. The results show that the accuracy of the algorithm is higher than that of the traditional ORB matching algorithm.

Because of the low quality of the image, the lack of annotation data, and the complex shape of the object, the existing detection methods only predict the center of the object, and the size of the object is a very important detection standard, but it is ignored. Liu et al. used the powerful object detection neural network Mask R-CNN to segment the object and provide the contour information. Because of the imbalance between positive and negative samples, the block-based classification network is used. The classification network with the highest accuracy is selected. As the backbone of the image segmentation network Mask R-CNN, the selected classification network has a good segmentation effect on natural images.

Detection methods of abnormal objects in a specific region based on Mask R-CNN

In the prediction stage, the instance segmentation of the conveyor belt with Mask R-CNN is not very good. In this article, an AODinSR method based on Mask R-CNN is designed, which aims to use Mask R-CNN to segment the instance and calculate the overlap with the specific region. Finally, it improves with abnormal object detection in the specific region.

The overall framework of the AODinSR model is shown in Figure 1. First of all, we use instance segmentation method to learn the instance segmentation model on the training images and obtain the instance segmentation model for specific object. As the most popular method, Mask R-CNN is widely used for this purpose. In AODinSR, we also use Mask R-CNN to generate the region of the object that we want to detect. Then, using the given specific region, such as conveyor belt in logistics transportation, the overlapping ratio of each instance segmentation and the specific region can be calculated in AODinSR. After that, the outputs of Mask R-CNN and overlapping regions with overlapping ratios lower than specify threshold are fed into fully connected (FC) layer. We took FC output separately to emphasize that the number of FC output is changed to the number of status of object for classifying the object into regular object and abnormal object and then restore the abnormal object to determine its original location and other
information. AODinSR believes that the object of the image should be detected according to different regions. And the traditional object detection method Faster R-CNN has a strong ability in detecting objects in the whole image. But Faster R-CNN has poor performance in detecting the same type of object in a more detailed subtype, such as normal box and abnormal box. Moreover, the traditional instance segmentation method Mask R-CNN has a strong ability in generating instance segmentation of objects in the whole image. Similar to Faster R-CNN, Mask R-CNN has poor performance in detecting the same type of object in a more detailed subtype. For these reasons, we consider using Mask R-CNN to get the mask of detected objects. Combined with a specific region, we calculate the overlapping ratio of the mask of the detected object and a specific region. Therefore, based on this idea, we use specific region and traditional Mask R-CNN method to detect abnormal objects.

Based on the objects detected by Mask R-CNN, AODinSR classifies the abnormal object. The abnormal object classification aims to deal with the problem of two classification prediction, which is the original object divided into regular object and abnormal object. It is assumed that there are two object types in the final prediction, including abnormal object $c_a$ and normal object $c_n$, including $N$ instance segmentations $i = \{i_1, i_2, \ldots, i_N\}$ and the region overlapping ratio $r = \{r_1, r_2, \ldots, r_N\}$, the corresponding object class label set is $l = \{l_1, l_2, \ldots, l_N\}$.

We use $f(x_n; w)$ to represent the number of full connection layer output activation value, which includes two parts: activation value $f_a(x_n; w)$ for abnormal objects and activation value $f_n(x_n; w)$ for normal objects. $w$ is the weight parameter of the network. At the same time, softmax function is used to standardize the output activation value and convert it into probability values belonging to object with a different status: abnormal and normal. As the previous definition, we use $f_a$ and $f_n$ to represent the final output. So, according to the function of softmax, we divide $f_a$ and $f_n$ by the sum of $f_a$ and $f_n$ to get the final probabilities which are $\frac{f_a(x_n; w)}{f_a(x_n; w) + f_n(x_n; w)}$ and $\frac{f_n(x_n; w)}{f_a(x_n; w) + f_n(x_n; w)}$. The final object with a different status is predicted as follows:

$$\hat{l}_n = \begin{cases} 
  c_a & \text{IF } f_a(x_n; w) \geq f_n(x_n; w) \\
  c_n & \text{IF } f_a(x_n; w) < f_n(x_n; w)
\end{cases} \quad (2)$$

The network is optimized by minimizing the following objective functions

$$w = \arg \min_w \frac{1}{N} \sum_{n=1}^{N} L(I_n, \hat{l}_n) + R(w) \quad (3)$$

$L(\cdot, \cdot)$ is the loss function and $R(\cdot)$ is the regularization. In this article, KL diversity, which is commonly used in image classification and prediction, is chosen as the loss function in the proposed method. The main purpose of traditional $\ell_2$ regularizations is preventing overfitting problem in learning from training data. $\ell_2$ is the most common regularization in deep learning and used for weight decay. And Euclidean norm is chosen as the regularization to prevent model overfitting. Moreover, $\ell_2$ regularization is used to prevent overfitting. The definition of $R(w)$ is given below

$$R(w) = \|w\|_2^2 \quad (4)$$

The specific process of AODinSR algorithm is shown in Figure 2. In phase 1, AODinSR initiates instance segmentation and classification model learning. In the function of trainInstanceSegmentationModel, AODinSR firstly uses Mask R-CNN to generate the weights $w_{mask}$ of instance segmentation model and instance segmentation $i_D$ on training data set $D$. Then $i_D$, specific region $R$, and specify threshold $r$ are used to get overlapping regions $r_D$ accompany with in the function of generateRegionOverlap. Finally, we use function to get the weights $w$ of classification model. In phase 2, AODinSR calculates region overlap for test set. In detail, instance segmentation $i_T$ is calculated on testing data set $T$ in the function of instanceSegmentation. Region overlapping ratio $r_T$ can be generated based on instance segmentation $i_T$, specific region $R$, and specify threshold $r$. Especially, objects with overlapping ratio $r_T$ greater than $r$ are removed and no longer considered. In phase 3, AODinSR predicts that the probability $f(T; w)$ of object belongs to normal object and abnormal object using function ofpredictProbability. In the last phase, combined with the probability of an object, AODinSR fuses the total object detection of abnormal object to predict $o^*$ by the function of fuseObjectDetection.

**Experiments**

**Data set and experimental set**

To check the performance of the proposed method AODinSR for abnormal object detection in logistics warehouse management, we build the data set from real images coming from logistics warehouse management. In logistics warehouse management, the main concentrated abnormal object is a box not in the conveyor belt but in the ground or other place. Here, the conveyor belt is the specific region considered in AODinSR. These boxes may have accidentally fallen off the conveyor belt or they may have been left behind by workers when carrying them. These boxes are therefore considered abnormal boxes which are out target object. For the purpose of detecting abnormal boxes in logistics warehouse management, we collected 1500 pictures in the actual logistics warehouse. After collecting the images, with the help of software named “labelme,” which is an open-source data annotation tool, we labeled all boxes as normal boxes and abnormal boxes. The abnormal boxes are abnormal objects that we want to detect. The tool can directly generate the frame generated by the first and last connection of annotation into a lightweight data exchange
format JSON text file, which is similar to the annotation results of object detection in coco data set. The objects to be identified in each picture (especially, in this article, objects mainly refer to conveyor belt and box) are marked with the frame connected from the beginning to the end, and the corresponding JSON format file is generated. After all the logistics warehouse management images are marked, the data set used in this article for training, testing, and validating the abnormal object detection in the specific region which is the conveyor belt is formed.

Image examples for abnormal object detection in logistic transportation are shown in Figure 3. All three images show the target object which is box for detection. But there exist two different types of box, which are normal box and abnormal box. The left image shows several abnormal boxes in the top-right and left-bottom corner. In the center image, the abnormal box is covered by the conveyor belt and only better than 80% of the box is shown in the picture. The right image shows one abnormal box in the ground and not in the conveyor belt. For all three image examples, there exist lots of normal boxes in the conveyor belt. But these boxes are not what we considered. The main objects that we want to detect are the abnormal boxes not in the conveyor belt which is known as the specific region in this article.

The data set used in this article is randomly divided into 75% training, 20% testing, and 5% validation sets. The validation set is used for choosing the best parameters of our method AODinSR. The whole experiment was carried out for 10 times, the mean average precision (mAP) was used as the evaluation index, and the average value of the prediction effect was calculated for 10 times. In the

*Input: image train set D, image test set T, Mask R-CNN algorithm, specific region R, specify threshold t*

*Output: abnormal object detection results  \( o^* \)************

**Begin**

// Phase 1: Initial instance segmentation and classification model learning

\[ w_{\text{mask}, D} = \text{trainInstanceSegmentationModel}(\text{Mask R-CNN}, D) \]

\[ r_D = \text{generateRegionOverlap}(i_D, R, t); \]

\[ w = \text{trainClassificationModel}(r_D, D); \]

// Phase 2: Region overlap calculation for test set

\[ i_T = \text{instanceSegmentation}(w_{\text{mask}}, T); \]

\[ r_T = \text{generateRegionOverlap}(i_T, R, t) \]

// Phase 3: Prediction of abnormal object probability

\[ f(T; w) = \text{predictProbability}(w, r_T); \]

// Phase 4: Abnormal object detection

\[ o^* = \text{fuseObjectDetection}(f(T; w), i_T); \]

**End**

Figure 2. The AODinSR algorithm. AODinSR: Abnormal Object Detection in Specific Region.

Figure 3. Image examples from data set.
Experiments, AODinSR is built on ResNet101 neural network and changes the number of final full connection layer output to the number of object detection status labels. That is 2, which concludes normal object and abnormal object. ResNet101 is a CNN that is trained on more than a million images from the ImageNet database. The learning rate is initialized to 0.0001. We fine-tune all layers by backpropagation through the overall neural network using mini-batches of 32, and the total number of epochs is 100 for abnormal object detection. Moreover, we had tried several different parameter configurations in the cross-validation fashion for parameter from 0.0001 to 10 using validation sets. For filtering measures to delete candidate objects in the abnormal object detection method, we exclude candidate objects whose overlapping ratios are greater than specify threshold \( t \) given by AODinSR. All experiments are carried out on NVIDIA GTX TITAN XP GPU with 12 GB memory.

### Experimental results and analysis

1. On the performance of abnormal object detection

To verify the superiority of the detection performance of AODinSR proposed in this article, we compare it with the traditional object detection method Faster R-CNN and the traditional instance segmentation method Mask R-CNN in the experiment and choose mAP as the evaluation metric to test the detection effect of abnormal objects. The AODinSR proposed in this article adds the specific region as a parameter to the abnormal object detection method, which can obtain the detection accuracy of the abnormal object, related to the specific region more accurately. Table 1 shows the detection results of AODinSR compared with the other two methods. Besides, results in bold indicate the best value of the mAP measure. It can be seen from the experimental results in Table 1 that compared with the traditional object detection method Faster R-CNN and traditional instance segmentation method Mask R-CNN, AODinSR pioneered the concept of specific region, combined with the region information in the process of object detection to judge abnormal objects, greatly improving the detection effect. The main reason is that both the Faster R-CNN and the Mask R-CNN can only identify a certain kind of object, such as box. So, no matter boxes in the conveyor belt or boxes in the ground are all detected as abnormal objects. This will result in bad detection performance. For this reason, both the Faster R-CNN and the Mask R-CNN cannot distinguish the different status of the certain kind of object that is regular object or normal object.

AODinSR shows superiority in the measure on real data set which demonstrates the effectiveness of AODinSR in abnormal object detection by considering the specific regions. Moreover, AODinSR can improve the recognition effect of the abnormal object by introducing the Mask R-CNN method in the specific region and finally improve the detection level of the abnormal object in the logistics transportation surveillance image.

Several images from the data set are shown in Figure 4, followed by the ground truth and detected abnormal object by AODinSR. From the results, we can see that AODinSR detects abnormal object similar to the ground truth and does not capture normal object. Detected abnormal objects of AODinSR on image examples are shown in bottom row. The ground-truth abnormal objects are shown in the top row. Especially, there are some wrong detected abnormal objects in the left image because there exist two abnormal objects near to each other.

2. On sensitivity of specify threshold \( t \) for overlapping ratio

In AODinSR, \( t \) controls the relative overlapping ratio between detected objects and specific region for further classification. The bigger the value of \( t \), the more unimportant of overlapping ratio in deciding the abnormal object. On the other side, the smaller the value of \( t \), the more important of overlapping ratio in deciding the abnormal object. In the proposed AODinSR, \( t = 0 \) means all detected boxes are used for further classification of abnormal object, and \( t = 1 \) means none detected boxes are used for further classification of abnormal object. In this experiment, we also use mAP metric to demonstrate how \( t \) influences the performance of AODinSR on the real data set collected in logistics warehouse management. The results are shown in Figure 5. From the results in the figure, we find that (1) the performance of not considering and wholly considering of overlapping ratio is worse than using specify threshold \( t \) to filter object for next handling, which illustrates that to detect the abnormal object, if or if not considering specific region can handle only one aspect of the detection and using both can improve performance of abnormal object detection; (2) the performance of consideration of overlapping ratio is effective and stable when \( t \) increases from 0.3 to 0.6, which means addressing the usage of relative overlapping ratio between detected objects and specific region. Through these results, we can find out that our proposed specify threshold \( t \) used in image handling is robust for abnormal object detection in AODinSR.

| Table 1. Comparison of experimental results. |
|---------------------------------------------|
| Methods                | mAP   |
| Faster R-CNN          | 0.515 |
| Mask R-CNN            | 0.595 |
| AODinSR               | 0.768 |

AODinSR: Abnormal Object Detection in Specific Region; mAP: mean average precision; CNN: convolutional neural network.
Conclusion

At present, the method of object detection is mainly used in logistics management to identify the goods in logistics transportation. It cannot further distinguish the goods and detect abnormal objects in the goods. Therefore, this article discusses how to effectively use the information of a specific region to assist analysis in abnormal object detection and proposes an abnormal object detection method based on Mask R-CNN: AODinSR. This method first obtains the initial instance segmentation model through the traditional Mask R-CNN method and then considers the region overlapping ratio of the instance segmentation results and the specific region. And finally the two methods are combined to detect the abnormal object and verified it in the actual logistics monitoring image data set. In addition, this article selects the mean average accuracy (mAP) as the evaluation index and compares it with Faster R-CNN and Mask R-CNN to detect the effect of abnormal objects. The experimental results show that Faster R-CNN and Mask R-CNN can only recognize a certain kind of object and cannot distinguish different states of the object. For example, the object on the ground is an abnormal object, resulting in poor detection performance. AODinSR effectively uses the regional information in the image to judge the abnormal object and stably improves the effect of abnormal object detection.

Future work

There are many kinds of realization and basic network structure in the instance segmentation method, so future research will expand AODinSR and consider the influence of different network structure and instance segmentation methods on abnormal object detection effect.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Beijing Natural Science Foundation (No.4172014), Support Project of High-level Teachers in Beijing Municipal Universities in the Period of 13th Five–year Plan (No.CIT&TCD201804031), the R&D Program of Beijing Municipal Education Commission (No.KM202010011011), Humanity and Social Science Youth Foundation of Ministry of Education of China (No.17YJCZH007).

ORCID iD

Haitao Xiong  https://orcid.org/0000-0002-3500-1279

References

1. Cheng KF and Bell MGH. Attacker–defender model against quantal response adversaries for cyber security in logistics management: an introductory study. *Eur J Oper Res* 2019. DOI: 10.1016/j.ejor.2019.10.019.
2. Quiroz IA and Alférez GH. Image recognition of legacy blueberries in a Chilean smart farm through deep learning. *Comput Electron Agric* 2020; 168: 105044.

3. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 2004; 60(2): 91–110.

4. Oh J, Kim HI, and Park RH. Context-based abnormal object detection using the fully-connected conditional random fields. *Pattern Recogn Lett* 2017; 98: 16–25.

5. Kim J, Kang B, Wang H, et al. Abnormal object detection using feedforward model and sequential filters. In: *2012 IEEE ninth international conference on advanced video and signal-based surveillance*, China, 18–21 September 2012, pp. 70–75. IEEE.

6. Zhu D, Dai L, Shao X, et al. Image salient object detection with refined deep features via convolution neural network. *J Electron Imaging* 2017; 26(6): 1.

7. Li T, Huang B, Liu J, et al. Application of convolutional neural network object detection algorithm in logistics warehouse. *Comput Eng* 2018; 488(06): 182–187.

8. Qiao Truman M and Sukkarieh S. Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming. *Comput Electron Agric* 2019; 165: 104958.

9. Liu A, Yang Y, Sun Q, et al. A deep fully convolutional neural network for semantic segmentation based on adaptive feature fusion. In: *2018 Fifth international conference on information science and control engineering (ICISCE)*, China, 20–22 July 2018, pp. 16–20. IEEE.

10. Roland SZ and Julien NS. Faster training of Mask R-CNN by focusing on instance boundaries. *Comput Vis Image Underst* 2019; 188: 9.

11. Deguchi M. Simple and low-cost object detection method based on observation of effective permittivity change. *Microelectron J* 2020, 95: 104678.

12. Boukhiris RR, Fendri E, and Hammami M. Moving object detection under different weather conditions using full-spectrum light sources. *Pattern Recogn Lett* 2020; 129: 205–212.

13. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceeding of IEEE conference on computer vision and pattern recognition*, USA, 24–27 June 2014, pp. 580–587. IEEE.

14. Ruiz-Santaquiteria J, Bueno G, Deniz O, et al. Semantic versus instance segmentation in microscopic algae detection. *Eng Appl Artif Intell* 2020; 87: 103271.

15. Cai Z and Vasconcelos N. Cascade R-CNN: high quality object detection and instance segmentation. *IEEE Trans Pattern Anal Mach Intell* 2019. DOI: 10.1109/TPAMI.2019.2956516.

16. Girshick R. Fast R-CNN. In: *Proceeding of IEEE international conference on computer vision*, Vol. 15, Chile, 7–13 December 2015, pp. 1440–1448. IEEE Press.

17. Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2015; 39(6): 1137–1149.

18. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. In: *Proceeding of IEEE conference on computer vision and pattern recognition*, USA, 26 June–1 July 2016, pp. 779–788. IEEE Computer Society.

19. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. In: *Proceeding of European conference on computer vision*, The Netherlands, 8–16 October 2016, pp. 21–37. Springer International Publishing.

20. LeCun Y, Bengio Y, and Hinton G. Deep learning. *Nature* 2015; 521(7553): 436.

21. Bae JW, Rykhlevskii A, Chee G, et al. Deep learning approach to nuclear fuel transmutation in a fuel cycle simulator. *Ann Nucl Energy* 2020; 139: 107230.

22. Chu M, Min B, Kwon S, et al. Determination of an infill well placement using a data-driven multi-modal convolutional neural network. *J Petrol Sci Eng* 2019; 12: 106805.

23. Ansari Y, Manti M, Falotico E, et al. Towards the development of a soft manipulator as an assistive robot for personal care of elderly people. *Int J Adv Robot Syst* 2017; 14: 1–17.

24. Wan S and Goudos S. Faster R-CNN for multi-class fruit detection using a robotic vision system. *Comput Netw* 2020; 168: 107036.

25. Zhu X, Chem C, Zheng B, et al. Automatic recognition of lactating sow postures by refined two-stream RGB-D faster R-CNN. *Biosyst Eng* 2020; 189: 116–132.

26. Lu X and Fei J. Velocity tracking control of wheeled mobile robots by iterative learning control. *Int J Adv Robot Syst* 2016; 13(3): 1.

27. He K, Gkioxari G, Dollar P, et al. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell* 2020; 42(2): 386–397 20181–1.

28. Hu Z, Fang W, Gou T, et al. A novel method based on a Mask R-CNN model for processing dPCR images. *Anal Methods* 2019; 11: 3410–3418.

29. Ganesh P, Volle K, Burks TF, et al. Deep orange: Mask R-CNN based orange detection and segmentation. *IFAC PapersOnLine* 2019; 52(30): 70–75.

30. Hu Z, Fang W, Gou T, et al. A novel method based on a Mask R-CNN model for processing dPCR images. *Anal Methods* 2019; 11(27): 9.

31. Zhang B and Han G. ORB mismatch elimination method based on Mask R-CNN. *LCD Display* 2018; 33(08): 690–696.

32. Liu M, Dong J, Dong X, et al. Segmentation of lung nodule in CT images based on Mask R-CNN. In: *2018 Ninth international conference on awareness science and technology (iCAST)*, Japan, 19–21 September 2018, pp. 1–6. IEEE.