Small Object Detection in High-Resolution Images Based on Multiscale Detection and Re-training

Guang-min SUN¹, Jia-yang CHEN¹, Bing Li², Dong YAN¹, Yu Li¹,* and Gang XIE²

¹Faculty of Information Technology, Beijing University of Technology, Beijing, China, 100124
²China National Tobacco Corp. Beijing Corp, No. 2, Nanxinyuan Road, Chaoyang District, Beijing, China, 100020

*Corresponding author

Keywords: High resolution image, Small object detection, Multiscale, Negative feedback mechanism, Defect detection of building wall surface.

Abstract. Most of the current small object detection algorithms are designed for low-resolution images. They can neither directly process high-resolution images nor make full use of the information contained. In this paper, an algorithm is proposed to detect small objects in high-resolution images directly. The process of the algorithm is as follows: Firstly, an original detection task is split logically into several relevant detection sub-tasks at different scales. Secondly, a corresponding low-resolution object detector is trained for each sub-task. Thirdly, the detectors are deployed to get detection results at different scales. Finally, the multi-scale detection results are logically combined to derive the final detection results of the small objects. Besides, the detector is re-trained at small scale by introducing negative samples. The algorithm proposed in this paper was tested in the task of defect detection on building wall surface. Experimental results show the reliability and efficiency of our approach.

Introduction

With the continuous development of deep learning techniques[1], many object detection networks have been proposed, which achieved superior performance compared with traditional algorithms. At present, widely used object detectors based on deep learning can be mainly divided into two categories. The first type is two-stage object detectors, such as Fast R-CNN [2], Faster R-CNN [3], and Mask R-CNN [4]. The process of these algorithms is composed of two parts: region proposals and classification. The second type is one-stage object detection algorithms, such as SSD [5], YOLO [6], YOLO 9000 [7], and YOLO V3 [8]. Such end-to-end object detection algorithms don not need region proposal in advance but directly complete the regression of position and the judgment of category through the prior boxes in the network. In the scene where the objects to be detected is large and sparse, both the two-stage and one-stage object detectors could have high detection precision. Moreover, the latter has a faster detection speed than the former.

However, due to factors such as the actual size of the object, imaging equipment, imaging distance and observation scale, pixels of objects may vary in different images. Compared to large objects, small ones have fewer pixels and more ambiguous features. So it is challenging to obtain excellent detection results of small objects by both the two-stage and the one-stage detectors. Therefore, small object detection is still a challenging task and has become a hot research topic in recent years.

Nowadays, a variety of optimization methods have been put forward for small object detection. However, these algorithms mainly focus on the improvement of the framework itself. One of the most famous structure is the Feature Pyramid Network [9]. The network fuses low-level feature map and high-level feature map generated by the backbone network in a specific manner to finally acquire a reconstruction feature pyramid. It makes semantic information enhanced and significantly improve the accuracy of the small object detection.
With the application of high-resolution (HR) cameras, small objects in the image can be portrayed by more pixels. This characteristic of HR images provides adequate data support for the small object detection task. Although the number of pixels of small objects in a high image increases, they are still relatively small compared to the whole picture. So we still refer to them as small objects. However, current detection algorithms are not suitable for images whose resolution is up to tens of millions of pixels. If the HR image is down-sampled to fit the input of current detection models, it is difficult to detect the small object in a low-quality image. To solve the problem, Adam Van Ette proposed a high-resolution satellite image-based object detection process named Satellite Imagery Multiscale Rapid Detection with Windowed Networks (SIMRDWN) [10]. The algorithm uses a fast detector to detect candidate regions acquired through the sliding window and can complete the fast detection task for high-resolution images of any size. However, it takes a long time to process a whole image, and false-alarms will come up in a complex environment. In response to these problems, we propose a simple and effective method for the small object detection task in HR images. The algorithm converts a primary detection task into a multi-scale detection task group with logical correlation. Then, low-resolution object detectors are trained for the detection tasks at different scales. Eventually, the detection results at each scale are logically combined and screened to obtain the final small object results. The proposed algorithm share a similar essence with human observation and make full use of the existing low-resolution detection models.

![Multiscale small object detection algorithm on high-resolution images.](image)

**Object Detection Algorithm**

Firstly, according to the theory of Gauss pyramid in image processing, the representation of an image under different scales can be obtained by repeated applications of Gauss filtering and down-sampling to the image. We establish a two-scale image pyramid for each image. One is a low-resolution image at a large scale generated by the operations above-mentioned, and the other is the original high-resolution image at a small scale without any processing.

Secondly, object segmentation is carried out at the large scale to obtain global information. The objects to be segmented should be set large rather than small. They should have an inclusion or exclusion relation with the small objects to be detected finally. That means the large objects must include or not include the small objects. Then, in low-resolution images, the selected large objects are segmented by the segmentation model, and a global semantic mask can be obtained. Finally, the mask is up-sampled to the resolution of the original image so that it can be applied to assist subsequent small object detection tasks.

Thirdly, object detection is executed at the small scale to get small objects. We used an overlapping sliding window to select sub-regions in the original image as the proposals to be detected and make another same window sliding in the mask image synchronously. The sub-region from the original image is called an image-proposal, while that from the mask is named a mask-proposal. The mask-proposal can decide whether the corresponding image-proposal will be sent to the detector. In a case of exclusion relation between the large and small objects, an image-proposal will not be detected subsequently if all the pixels in the mask-proposal belongs to the large objects. In this way, some unnecessary proposals can be avoided, and detection efficiency can be improved. Then, the detection results of all image-proposals are transformed back to the original image.
according to their location in the original image. In order to avoid the overlapping of detection boxes, it is necessary to apply Non-Maximum Suppression to these boxes as post-processing.

Finally, the multi-scale detection results are fused. Due to the factors such as rectangular sliding window, imprecise mask generation, loose region selection algorithm, false detection may occur inevitably. To solve this, we fuse the small objects detection results at small scale and the segmentation mask at large scale by rejecting the detection bounding boxes which conflict with the mask area. In order to avoid abandoning correct object bounding boxes near the boundaries in the mask, the mask is firstly processed by simple morphology before the final fusion. In this way, wrong detections can be screened, while correct detections can be retained. The overall framework of the algorithm is shown in Fig. 1.

**Training Procedure**

There are two models need to be trained: large-scale object segmentation and small-scale object detection. The former can be trained traditionally, while the training of the latter is more complex, which consists of initial training and re-training. The initial training process is as follows: Firstly, high-resolution images are divided into training, validation and test sets. Then slices containing small objects are cut from the high-resolution images of each set respectively, and the corresponding three slice-sets are obtained. These slices should conform to the input size of the detector to be trained, and each slice contains at least one object. Finally, the train slice-set is used for training the model, and the nodes which perform best on the validation slice-set are saved.

Then, we use the initial trained model to detect the original high-resolution image sets and introduce the slices near the false detections as negative samples to the original train and validation slice-sets. These negative sample slices are labeled as normal, and their detection boxes are uniformly set to the size of the slices. The number of negative sample slices needs to be guaranteed the ratio of 1:1 to the number of the original sample set. Finally, the secondary fine-tuning of the initial trained model can be applied using the new datasets. Because of the introduction of false detections, the detector pays more attention to these negative samples and better learn how to distinguish positive and negative samples. This negative feedback mechanism can correct the model training and significantly reduce the occurrence of false alarm. It should be noted that the normal class is not allowed to be output by the re-trained model, and just play a role during the training process.

**Experiments**

To verify the proposed algorithms, we choose a task of detecting defects on building wall surface as background. In the task, the resolution of the wall surface image is as high as 7952×5304 pixels, while the size of the wall defect, such as missing bricks and broken bricks, is not more than 100×100 pixels. Their area covers the whole image less than 1/4000. The proposed algorithm is modified to be implemented on this defect detection task. At large scale, windows, air conditioners and other non-wall objects are detected with Mask-RCNN[4]. After up-sampling, non-wall masks can be obtained. At the small scale, we choose the SSD, whose backbone is Resnet50 model, combined with FPN.[9] to detect the wall defects. Fig.2 shows the overall procedure of the proposed small defect detection algorithm based on high-resolution wall surface image. All experiments were completed based on Google Object Detection API of Tensorflow platform [11].
Figure 2. Multiscale small defect detection algorithm on high-resolution images of building surface.

**Training Results**

31 samples are used to train the Mask RCNN, 25 for the train set, 6 for the validation set, whose resolution is generally 994×663 pixels. 105 samples are used to train the SSD, among them 88 for the train set, 9 for the validation set, and 10 for the test set, whose resolution is 7952×5304 pixels. According to the sample generation algorithm, 640×640 slices containing defects are cut from the original HR image to constitute initial datasets for training the SSD. The numbers of slices corresponding to the train set, validation set, and test set are 2660, 200, and 200, respectively. Due to the small size of the dataset, transfer learning is performed on the wall surface defect slice-sets using the pre-trained model on the COCO dataset[12]. Then based on the re-training skill, we create two new slice-sets. The numbers of new train and validation slice-set are 5320 and 400, respectively. The initial training model is fine-tuned again on the new dataset to obtain the final detection model. The detection results of the proposed algorithm are shown in the Fig.3.

(a) detection result of the Mask-RCNN  (b) detection results of the SSD  (c)detection result of the proposed algorithm

Figure 3. The detection results of Mask RCNN, SSD and the proposed algorithm.

**Comparison of Single-scale and Multiscale Frameworks**

To prove the validity of our method, we also apply the SIMRDWN [10] on the same dataset. The mAP@IOU0.5[12] value of each model calculated in this experiment on the high-resolution image is shown in Table 1. Compared with the multiscale fusion algorithm (the proposed algorithm) framework, the accuracy of the single-scale algorithm (SIMRDWN) is significantly reduced. At the same time, since the slices to be detected are screened, the time cost of our algorithm is smaller, as shown in Table 1.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>mAP@IOU0.5</th>
<th>Detection time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Train</td>
</tr>
<tr>
<td>SIMRDWN[10]</td>
<td>0.726</td>
<td>0.407</td>
</tr>
<tr>
<td>Proposed(Without Re-training)</td>
<td>0.754</td>
<td>0.419</td>
</tr>
<tr>
<td>Proposed(With Re-training)</td>
<td>0.760</td>
<td>0.551</td>
</tr>
</tbody>
</table>
Comparison of the Models before Re-training and after Re-training

To prove the validity of the re-training, we also set an experiment that compares the detection results of models before re-training and after re-training. It can be easily observed from the Table 1 that mAP at IOU=0.5 of the former is higher than that of the latter. We think it is because that some false detections are eliminated in the detection results of the secondary training model as shown in Fig.4. The result proves that the introduction of the negative samples is indeed practical. However, there may be some defects which resemble positive samples in the negative samples. It may inhibit the learning of the correct defect features, which limits the continued improvement of performance.

![Comparison of detections before and after re-training](image)

Figure 4. The comparison of detections (blue rectangles) before re-training and after re-training.

Conclusions

This paper proposes an algorithm that can directly deal with small object detection problems in high-resolution images based on current low-resolution detectors. The algorithm splits the original detection task into a multi-scale task group and fuses results from the task group to obtain the final detections of small objects. Moreover, a re-training scheme is proposed by introducing negative feedback to the small-scale detector, so as to reduce false alarms and improve detection accuracy in high-resolution images. The algorithm outperforms the single-scale detection framework in both accuracy and efficiency.

Acknowledgment

This research is jointly supported by the National Key Research and Development Program of China (2016YFB0501501), the Natural Scientific Foundation of China (41706201), and the 2018 International Research Cooperation Seed Fund of Beijing University of Technology (No. B1).

References


