Scientific
Research
Publishing

Detection of Water Meter Digits Based on Improved Faster R-CNN

Liyun Sun, Yuying Yuan*, Shichao Qiao, Ruijie Qi

School of Computer Science and Technology, Shandong University of Technology, Zibo, China
Email: *88617060@qq.com

Abstract

In order to more accurately detect the accuracy of word-wheel water meter digits, 2000 water meter pictures were produced, and an improved Faster-RCNN algorithm for detecting water meter digits was proposed. The improved Faster-RCNN algorithm uses ResNet50 combined with FPN (Feature Pyramid Network) structure instead of the original ResNet50 as the feature extraction network, which can enhance the accuracy of the model for small-sized digit recognition; the use of ROI Align instead of ROI Pooling can eliminate the error caused by the quantization process of the ROI Pooling twice, so that the candidate region is more accurately mapped to the feature map, and the accuracy of the model is further enhanced. The experiment proves that the improved Faster-RCNN algorithm can reach 91.8% recognition accuracy on the test set of homemade dataset, which meets the accuracy requirements of automatic meter reading technology for water meter digital recognition, which is of great significance for solving the problem of automatic meter reading of mechanical water meters and promoting the intelligent development of water meters.

Keywords

Faster RCNN, Identification of Water Meter Numbers, ResNet50, FPN, ROI Align

1. Introduction

Mechanical water meters are widely used as mainstream water meters in people's daily life because of their simple structure and low price. Mechanical water meters require manual meter reading, which wastes a lot of human resources and has the disadvantage of untimely meter reading. With the development of deep learning technology, the use of target detection algorithms to automatically identify the numbers in the water meter picture can replace manual meter read-

ing, greatly improving the efficiency of the meter reading company and saving human resources.

Most of the water meters used in people's daily life are character wheel type water meters with rectangular reading area frame. The recognition of characters in the reading area is similar to OCR (Optical Character Recognition) [1]. OCR technology is widely used in the field of digital and English symbol recognition, where the algorithms based on OCR technology in word-wheel water meter digital detection are mainly divided into traditional algorithms and deep learning algorithms. The literature [2] [3] uses traditional algorithms such as image binarization, image segmentation, and pattern matching for character recognition. Although the traditional reading recognition method used can also identify water meter digits, but there are shortcomings of low detection accuracy and recognition speed, and the traditional method requires the use of different methods for different reading scenarios, the detection process is complex. With the continuous development of computer vision and deep learning, deep learning can automatically extract features in images and train the corresponding algorithms, and finally predict the results of the images and output them. In paper [4], Liu *et al.* proposed to use BP-bagging algorithm to recognize images, output classification information with BP network and generate multiple classifiers by bagging. The algorithm can extract 25-dimensional features of the image compressed to 5-dimensions and input the results to the BP-bagging classifier, and finally recognize digital characters by multiple voting. Yang *et al.* proposed a full convolutional sequence recognition network to read water meter numbers and designed an incremental loss function to manage the intermediate states of the numbers, although it improved the accuracy of water meter reading, it only studied for the dataset after cropping the reading area, and it was more difficult to recognize the numbers for the complete water meter pictures taken in natural scenes [5].

In order to further improve the digit recognition accuracy of word-wheel water meters, this paper collects and produces a dataset of 2000 word-wheel water meter images and uses the improved Faster-RCN algorithm to detect and recognize water meter digits [6]. In order to improve the detection accuracy of small numbers, ResNet50 combined with FPN (Feature Pyramid Network) is used as the feature extraction network. In order to reduce the error of candidate region location caused by the two rounding operations during the ROI pooling operation in Faster-RCNN, ROI Align is used instead of ROI Pooling to further improve the accuracy of model training. The experimental results show that the improved Faster-RCNN model has high detection accuracy in the water meter digit recognition task and can meet the accuracy requirements of the automatic water meter digit recognition task.

2. Water Meter Digit Recognition Dataset

2.1. Dataset Collection and Production

The dataset uploaded in github in literature [5] is the only publicly available water meter number recognition dataset, but this dataset is for recognizing images

after cropping out the water meter reading area, while the purpose of this experiment is to recognize the numbers in a complete water meter image, so this dataset is not suitable for the research of this paper. Since there is no publicly available and suitable dataset for water meter number recognition, a dataset is produced in this paper. A total of 1000 pictures of wheel-type water meters were collected from the field shooting and online, in order to improve the accuracy and generalization ability of the model, the training set was enhanced by panning, rotating, increasing the brightness and random cropping, etc. The training set after data enhancement was 2000 pictures, among which 1800 pictures were randomly selected for training and 200 pictures were used for testing. With reference to the PascalVOC2007 dataset format, the 2000 images in the training and test sets were labeled by LabelImg software. Figure 1 shows a few sample pictures of the dataset. There are many other interfering numbers in the water meter in addition to the reading area numbers that need to be recognized. When labeling the reading area numbers, the black box on the four sides of the numbers will also be boxed in, so as to distinguish the valid numbers from the interfering numbers. Such as Figure 2 is part of the water meter picture labeling diagram.

2.2. Character Annotation and Analysis

In view of the characteristics of the water meter dataset, special attention should be paid to the digital labeling of the "intermediate state" when labeling the dataset. The red box in Figure 3 is an example of digital labeling of "intermediate state" of water meter. Figure 3, the first picture as an example, the water meter from the state of "00742" has not fully reached the state of "00743", at this time the reading should be read as "00742.5". So separate labeling labels are set up for figures in this transition state. Figure 4 shows an example of labeling for a single digit that is labeled. For "complete numbers", the labeling is $t \in [0 - 9]$. For "intermediate" digits, the labeling label $s \in [10 - 19]$ is used, which corresponds to the transition state from digit "s-10" to "s-9" when the label s is set. Take the fifth character of the first picture in Figure 3 as an example, set the character label to "12", which represents the transition state from the number "12 – 10 = 2" to "12 – 9 = 3".

Figure 1. Some images of the water meter dataset.

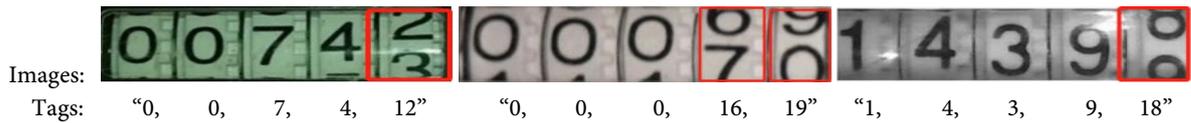**Figure 2.** Lartially labeled map of the water meter dataset.

Images:

Tags: "0,   0,   7,   4,   12"   "0,   0,   0,   16,   19"   "1,   4,   3,   9,   18"

Figure 3. Example of a label for a water meter reading area.

Images:

Tags: 0   1   2   3   4   5   6   7   8   9   10   11   12   13   14   15   16   17   18   19
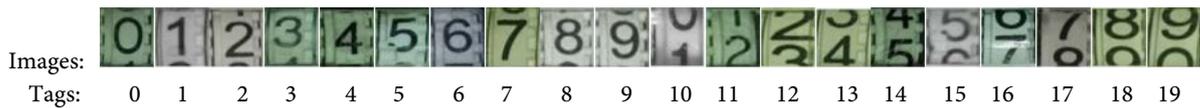
Figure 4. Example of a single number and label in a water meter image dataset.

After analysis, it is found that the model has more difficulty in recognizing "intermediate" numbers than "complete" numbers, mainly for two reasons. The first reason is that there are fewer "intermediate" numbers in the data set. Table 1 shows the number of various characters in the data set. Table 1 shows that the number of characters in the "intermediate state" is much smaller than the number of "complete numbers". The second reason is that there is a big difference in the number of "intermediate state" characters corresponding to the same labels. Figure 5 shows three sets of pictures with the same label corresponding to different transition states of numbers. Each group of images has the same label for the "intermediate state" of the digits, but the difference in the degree of digit transition leads to a large difference in the image features.

3. Faster-RCNN Algorithm and Improvement of the Network

3.1. Faster-RCNN Algorithm

Faster-RCNN algorithm is a target detection algorithm proposed in 2015. Compared with Fast-RCNN, Faster-RCNN has higher detection accuracy and less detection time. Currently, target detection algorithms are mainly categorized into two main groups: one is the single-stage detection algorithms represented by YOLO (You Only Look Once) and SSD (Single Shot Multibox Detector) algorithms, and the two-stage detection algorithms represented by RCNN, Fast-RCNN, and Faster-RCNN [7] [8] [9] [10]. The two-stage detection algorithms mainly divide the detection task into two stages, the first stage is to find the location of the object and generate the region suggestion, and the second stage is to classify and regress the region suggestion to obtain a more accurate

Table 1. Number of characters corresponding to each character in the water meter image dataset.

characters	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
quantities	2681	803	997	711	702	491	486	506	410	427	198	160	178	249	211	124	131	128	197	210

Images:
Tags: 11   15   10

Figure 5. Images with the same label corresponding to different numeric transition states.

location of the object. Compared with the single-stage detection algorithm, Faster-RCNN includes an additional RPN (Region Proposal Network) structure, which is used to generate high-quality region proposal boxes on the feature map, so usually Faster-RCNN has higher detection accuracy than the single-stage algorithm.

Figure 6 shows the network structure of the Faster-RCNN algorithm, which has four main components: feature extraction network, RPN layer, ROI pooling layer, classification and regression network [11] [12]. The detection process of Faster RCNN algorithm: 1. Adjust the image uniformly to M × N size and input it into the feature extraction network to generate a feature map. 2. The feature map generates high-quality target suggestion boxes through the RPN structure, and preliminarily obtains the approximate position of the target. RPN consists of two parts. One part uses the Softmax function to determine the probability of a candidate region being a target, while the other part uses the bounding box regression function to correct the position of the candidate boxes. Finally, in the Proposal layer, the probability of positive and negative samples and the position of the bounding boxes are combined to filter and obtain the region's candidate boxes. 3. Based on the candidate boxes generated by RPN, they are projected onto the feature map to obtain the corresponding feature matrix. Each feature matrix is scaled to a 7 × 7 size feature map through the ROI Pooling layer. 4. The feature map further classifies and adjusts the position of bounding boxes through fully connected layers.

3.2. Improvement of Feature Extraction Network

When the water meter image is subjected to digital detection using the Faster RCNN algorithm, the first step is to extract features through a feature extraction network. The original Faster RCNN often used convolutional neural networks such as VGG16, ResNet50, and MobileNetV2 as feature extraction networks. Because residual networks (ResNet) can solve the problem of network degradation while increasing the depth of convolutional layers, and have good detection performance, this paper chooses ResNet50 as the original feature extraction network. This article uses ResNet50 combined with FPN as an improved feature extraction network. FPN can fully utilize the information of all feature layers in the feature extraction network, effectively improving the model's detection accuracy for small number targets. Figure 7 shows the structure of ResNet50 combined with FPN as a feature extraction network. The FPN can fuse the feature
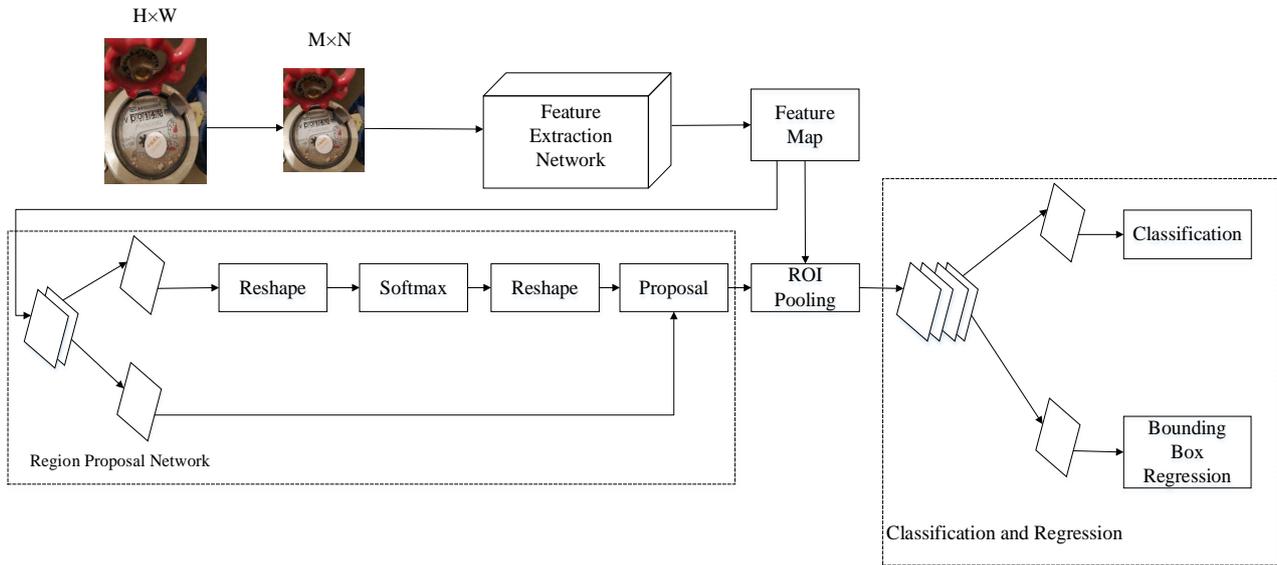
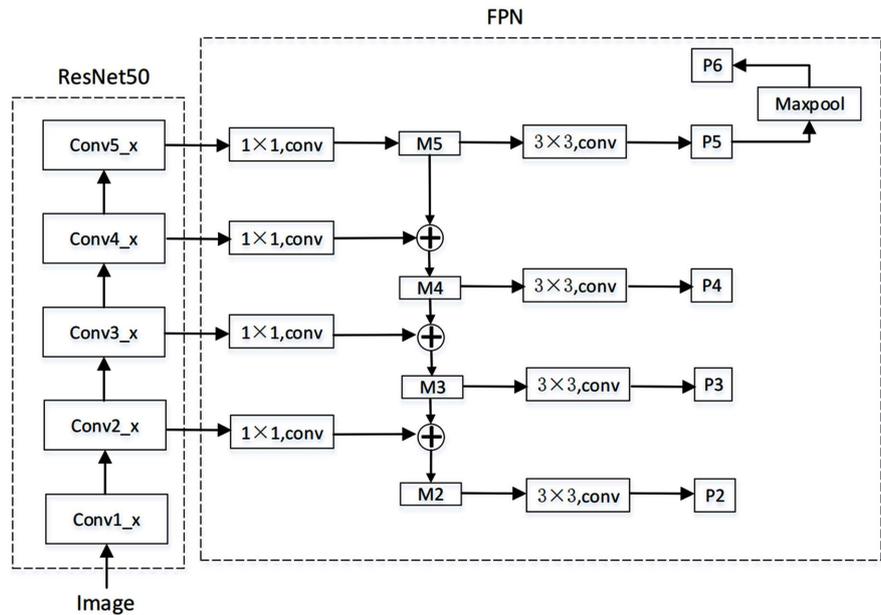**Figure 6.** Faster-RCNN network structure.

Figure 7. ResNet50 combined with FPN for feature extraction network.

maps of different layers output from ResNet50. Specific implementation steps: firstly, adjust the Conv2_x-Conv5_x layer to have the same channel through 1×1 convolutional layer, and then start from the Conv5_x layer sequentially with 2 times of up-sampling, and then fused with the previous layer of the feature map, got the feature map M2-M5, and then respectively carried out 3×3 convolutional operation to get the feature map P2-P5, and then carried out the maximum pooling down-sampling of the M2 in the step length of 2 to get the feature map P6, and finally got the feature map P2-P6 which fused semantic information of the feature map of the multi-layer. Fifteen anchor frames with five areas (32^2, 64^2, 128^2, 256^2, 512^2) and three aspect ratios (1:1, 1:2, 2:1) are set up for dif-

ferent predicted feature layers P2-P6, which is compared to the original Faster-RCNN setup of three anchor frames with three areas (128^2, 256^2, 512^2) and three aspect ratios (1:1, 1:2, 2:1), a total of 6 more different anchors, further improving the model's prediction accuracy for different shaped figures.

Input the predicted feature maps P2-P6 into the RPN layer to get propose, then the proposal box needs to be mapped again to the feature map to obtain the feature matrix, the formula for determining which layer of the feature map the propose needs to be mapped to is (1) [6].

$$k = \left\lfloor k_0 + \log_2\left(\sqrt{wh}/224\right) \right\rfloor \tag{1}$$

where 224 is the ImageNet pretrained image size; w and h are the width and height of the propose obtained from the RPN layer prediction corresponding to the original image; and k_0 is the baseline value, which is set to 4.

3.3. Introducing ROI Align

In the original Faster-RCNN, the role of ROI Pooling is to generate a 7×7 sized feature map of region proposals. The specific steps are firstly to project the RPN-generated candidate box onto the feature map, and to correspond the four vertices of the candidate box to the places on the feature map where there are pixel-value positions, and this process will round the coordinates of the candidate box. Then the feature matrix obtained earlier is adjusted to a feature map of 7×7 size, this step needs to divide the feature matrix into pooling regions. If the feature area is not divisible by the pooling area, another rounding operation will be performed to obtain a different pooling area. The two rounding operations of ROI Pooling will cause some deviation in the positions of the candidate areas corresponding to the feature maps, which will affect the detection accuracy of the water meter digits.

In order to overcome the error brought by ROI Pooling, the idea of Mask-RCNN model is referred to, and ROI Align is used to replace ROI Pooling. ROI Align has no quantization operation, and retains the floating-point numbers generated by pooling, and obtains the surrounding adjacent eigenvalues by bilinear interpolation, the specific principle is shown in Figure 8.

The black dashed line box with the largest area in the figure is the feature map, and the black solid line box is the ROI area, which is divided into a × a (2 × 2 in the figure as an example) sized areas. Take the first copy of the region in the upper-left corner of the ROI in the figure as an example, divide this region into four parts evenly, find the center point of the four parts of the region, and use the idea of bilinear interpolation to get the pixel value of the center point through the four pixel points of 1, 2, 3 and 4 of the feature map near the center point. Then use the same method to get the pixel values of the other center points, and output the pixel values of the region in the red box in the ROI region after the maximum pooling operation. According to the same method can get the other three copies of the region's pixel values, is the final ROI Align output results.
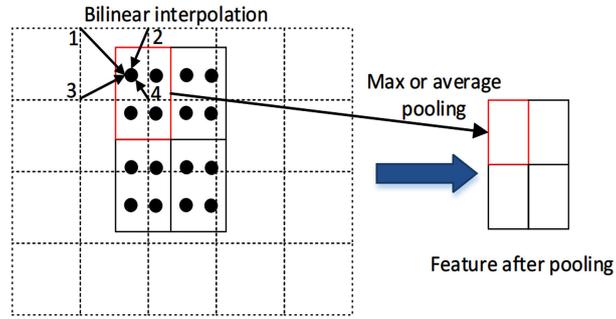
Figure 8. ROI align principle.

4. Experimental Results and Analysis

4.1. Experimental Environment and Hyperparameter Settings

This paper is a model built based on the deep learning framework pytorch1.7, water meter digit detection and recognition using python programming language. The main configuration of the computer is: Ubuntu 20.04 operating system, 12 core Intel (R) Xeon (R) Platinum 8255C CPU @ 2.50GHz, 43GB of memory, RTX 3090 graphics card, and 24GB of graphics memory. The training parameters of the experiment: each batch contains 8 images, the momentum value is 0.9, the weight decay is set to 0.001, the initial learning rate is set to 0.01, with the number of iteration increase gradually decreases the learning rate, the model is pre-trained on the ImageNet dataset and then trained on the water meter reading dataset for 50 iterations.

4.2. Evaluation Indicators

The metrics used to evaluate the model in this paper are detection time and mAP (mean average precision). ap numerically represents the area under the precision versus recall curve for each category, while mAP is the average of the AP values for all categories. mAP is calculated using the formula (2)

$$\text{mAP} = \frac{\sum_{i=1}^{N_c} \text{AP}_i}{N_c} \tag{2}$$

where N_c is the number of detection categories; the formula for AP is (3)

$$\text{AP} = \int_0^1 P(R)\,\mathrm{d}R \tag{3}$$

where P is the precision and R is the recall, calculated as (4) and (5)

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

In this equation TP is the number of detected frames with IOU (intersection over union) > 0.5, FP is the number of detected frames with IOU < 0.5, and FN is the number of correct targets not detected.

4.3. Experimental Results and Analysis

In order to increase the training accuracy of the algorithm and reduce the number of training iterations, we use a migration learning approach to let the model be pre-trained on the ImageNet dataset first, and then trained on the water meter readings dataset for 50 iterations. Figure 9 shows the loss function curve and learning rate curve of the improved Faster R-CNN algorithm trained on the water meter readings dataset, which shows that the training loss is gradually reduced. Figure 10 shows the mAP curve of the improved Faster R-CNN algorithm verified on the validation set, and the algorithm recognition accuracy can reach more than 91% after about the tenth training iteration. From these two figures, it can be seen that the improved Faster R-CNN model learns better and has higher recognition accuracy.
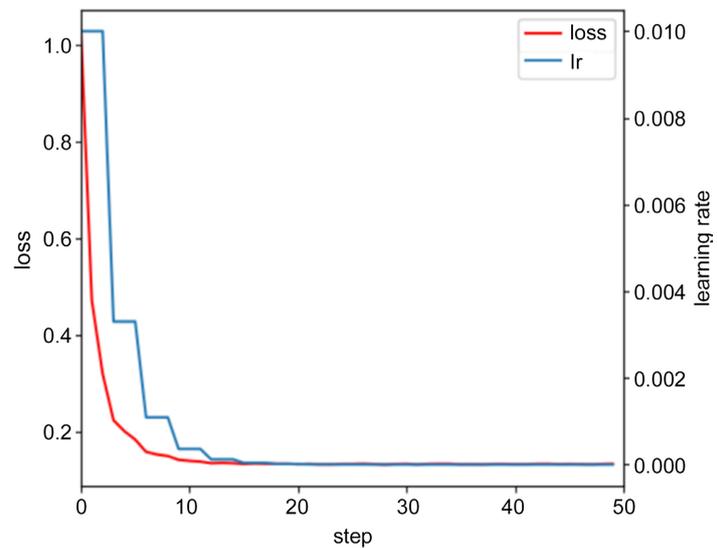

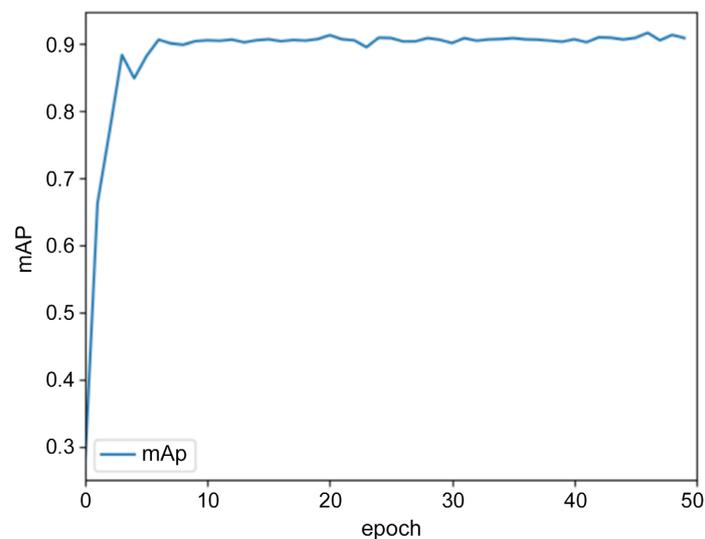**Figure 9.** Training loss and learning rate.

Figure 10. Validation set mAP.

In order to investigate the enhancement of each innovation point on the effect of recognizing water meter digits, three different experiments were set up and the results are shown in Table 2. Experiment 1 is an experiment using the original Faster R-CNN on the homemade water meter readings dataset, and the feature extraction network of the algorithm is ResNet50. Experiment 2 is to use ResNet50 plus FPN as the feature extraction network, and other things remain unchanged, and the mAP improves the accuracy of the mAP by 5.1% than that of the original Faster R-CNN algorithm. Experiment 3 is based on experiment 2, replacing ROI Pooling in Faster R-CNN with ROI Align, other things remain unchanged, and the mAP is improved by 3.5% over experiment 2. The experimental test results are shown in Table 2.

Experimental comparisons using the improved Faster R-CNN algorithm in this paper with the Faster R-CNN, SSD, and YOLOv5 algorithms were conducted, focusing on comparing the mAP and the detection time of a single photo of the four algorithms. From the experimental results shown in Table 3, it can be seen that the mAP of the improved Faster R-CNN algorithm is 8.6% higher than that of the Faster R-CNN in terms of accuracy, and the detection speed is not much different. The mAP of the improved Faster R-CNN algorithm is 31.5% and 6.4% higher than that of SSD and YOLOv5, respectively, and although the recognition speed is slower than that of SSD and YOLOv5, it can also satisfy the speed requirement for the task of automatically recognizing the numbers in the water meter. The experimental results prove that the improved Faster R-CNN algorithm can achieve higher detection accuracy while meeting the speed required for remote meter reading.

Three images were randomly selected in the test set to be detected using SSD, YOLOv5 and the improved Faster R-CNN algorithm in this paper, and the test results are shown in Figure 11. As can be seen from the three algorithms, the SSD algorithm, YOLOv5 algorithm for the "intermediate state" of the number of

Table 2. Comparison of detection results of Faster R-CNN algorithm with different schemes.

Different programs	mAP (%)
Experiment 1	83.2
Experiment 2	88.3
Experiment 3	91.8

Table 3. Comparison of recognition results of different algorithms.

Different algorithms	mAP (%)	T (s)
Faster R-CNN	83.2	0.064
SSD	60.3	0.016
YOLOv5	85.4	0.024
The algorithms in this paper	91.8	0.071

(a) SSD   (b) YOLOv5   (c) Improved Faster R-CNN

(d) SSD   (e) YOLOv5   (f) Improved Faster R-CNN

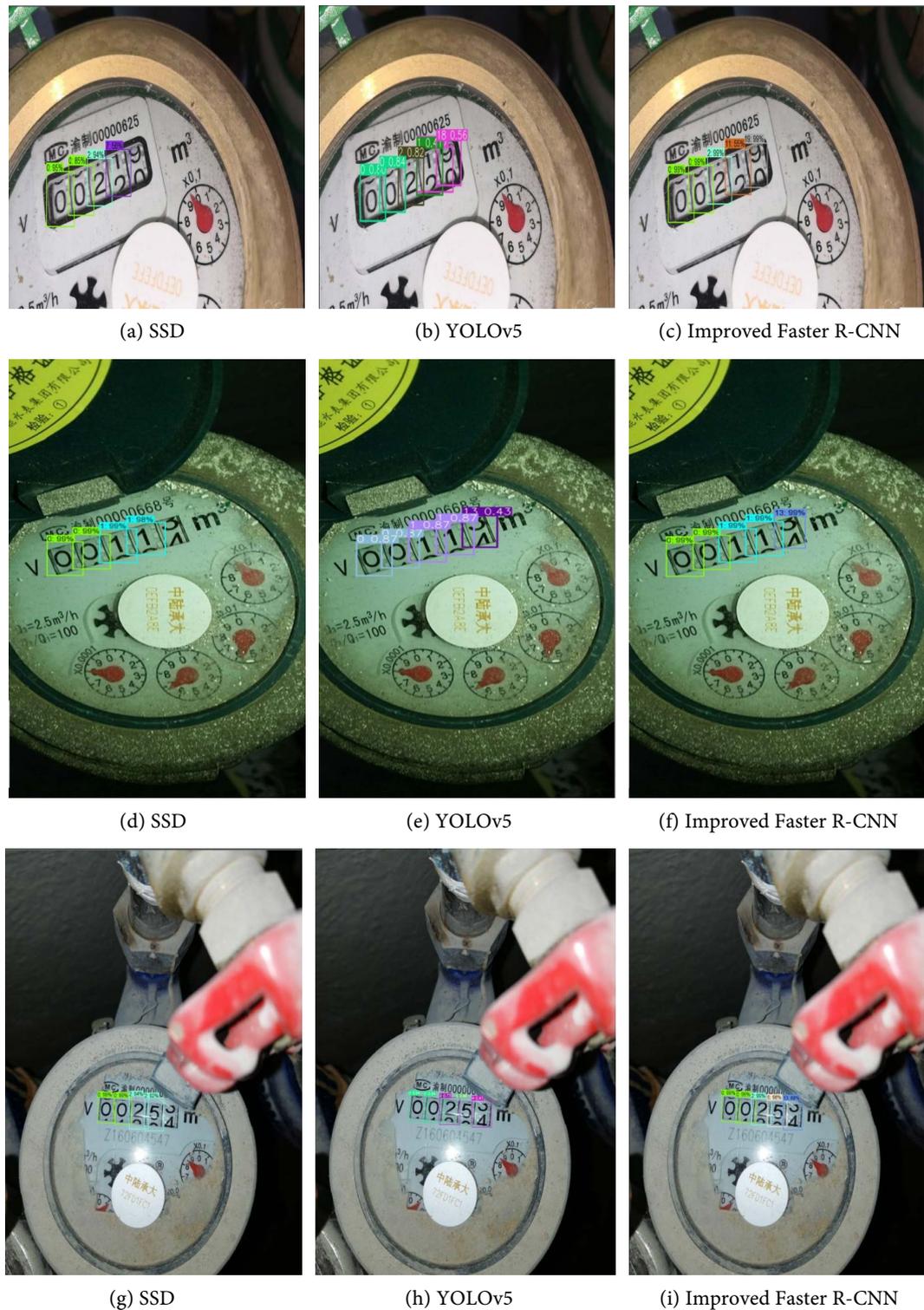(g) SSD   (h) YOLOv5   (i) Improved Faster R-CNN

Figure 11. Comparison of detection results of different algorithms.

leakage, wrong detection and multi-detection of the situation, the "complete number" also exists in the case of low detection accuracy. The improved Faster R-CNN algorithm in this paper has higher recognition accuracy for water meter numbers, and can better recognize the "intermediate state" and "complete state"

numbers, which can achieve the accuracy of remote meter reading technology for automatic identification of water meter numbers.

5. Conclusion

Aiming at the shortcomings of low detection accuracy and poor generalization ability of the traditional water meter digit recognition algorithm, an improved Faster-RCNN algorithm is proposed in this paper. In order to improve the detection accuracy of the model for small numbers, ResNet50 combined with FPN instead of ResNet50 as the feature extraction network can extract richer semantic information. In order to solve the error generated by the two rounding operations of ROI Pooling in the arithmetic process, ROI Align is used instead of ROI Pooling, which can make the candidate regions more accurately mapped to the feature map and further improve the detection accuracy of the model. The improved Faster-RCNN algorithm can reach 91.8% detection accuracy on the homemade water meter digit recognition dataset, and it also has a high recognition rate for "intermediate state" characters. However, the improved Faster-RCNN algorithm still belongs to the two-stage target detection algorithm, and the detection speed is not as fast as the single-stage target detection algorithm, and the algorithm will be further researched in the future in terms of real-time detection, so as to make the model have better detection effect.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Divya, P., Varma, M., Mouli, U.R., *et al.* (2021) WITHDRAWN: Web Based Optical Character Recognition Application Using Flask and Tesseract. *Materials Today: Proceedings.* https://doi.org/10.1016/j.matpr.2020.10.850

[2] Lei, F., Xiong, Z. and Wang, X. (2019) A New Non-Smart Water Meter Digital Region Localization and Digital Character Segmentation Method. *International Conference on Cyberspace Data and Intelligence*, **1137**, 557-571. https://doi.org/10.1007/978-981-15-1922-2_39

[3] Jawas, N. and Indrianto (2018) Image Based Automatic Water Meter Reader. *Journal of Physics: Conference Series*, **953**, Article 012027. https://doi.org/10.1088/1742-6596/953/1/012027

[4] Liu, Z., Li, L. and Yu, M. (2015) An Algorithm of Handwritten Digital Recognition Based on BP-Bagging. *International Conference on Information Technology and Management Innovation*, Santiago, October 2015, 1164-1168. https://doi.org/10.2991/icitmi-15.2015.195

[5] Yang, F., Jin, L., Lai, S., *et al.* (2019) Fully Convolutional Sequence Recognition Network for Water Meter Number Reading. *IEEE Access*, **7**, 11679-11687. https://doi.org/10.1109/ACCESS.2019.2891767

[6] Ren, S., He, K., Girshick, R., *et al.* (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis*

& *Machine Intelligence*, **39**, 1137-1149.
https://doi.org/10.1109/TPAMI.2016.2577031

[7]   Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448.
https://doi.org/10.1109/ICCV.2015.169

[8]   Redmon, J., Divvala, S., Girshick, R., *et al.* (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788.
https://doi.org/10.1109/CVPR.2016.91

[9]   Liu, W., Anguelovd, Erhan, D., *et al.* (2016) SSD: Single Shot Multibox Detector. In: *European Conference on Computer Vision*, Springer, Heidelberg, 21-37.
https://doi.org/10.1007/978-3-319-46448-0_2

[10] Girshick, R., Donahue, J., Darrell, T., *et al.* (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587.
https://doi.org/10.1109/CVPR.2014.81

[11] Li, J., Zhu, Z., Liu, H., *et al.* (2023) Strawberry R-CNN: Recognition and Counting Model of Strawberry Based on Improved Faster R-CNN. *Ecological Informatics*, **77**, Article ID: 102210. https://doi.org/10.1016/j.ecoinf.2023.102210

[12] Hu, M., Wu, Y., Yang, Y., *et al.* (2023) DAGL-Faster: Domain Adaptive Faster R-CNN for Vehicle Object Detection in Rainy and Foggy Weather Conditions. *Displays*, **79**, Article ID: 102484. https://doi.org/10.1016/j.displa.2023.102484