Research on Water Meter Digital Recognition Method based on DBNET-CRNN Framework

Weiye Li

Anhui University, Hefei, China

2132862056@qq.com

Abstract. With the prominent drawbacks of traditional manual meter reading mode, the automatic recognition of water meter digits with high efficiency and high accuracy using image processing technology has become one of the issues that more and more researchers pay attention to. In order to more accurately and effectively count the groundwater representation numbers in natural scenes, this paper presents a detection and recognition algorithm for water meter digits based on DBNET-CRNN framework. First, we design text detection model recognition text box based on DBNET network. Next, we use a text recognition algorithm based on CRNN network to predict the text box area and identify the final result. In addition, we have independently analyzed the half-word to calibrate the results, which further improves the accuracy of the model. The experimental results show that the model achieves good results in the water meter number recognition task. It can be applied to water meter number detection and recognition in natural scenes, effectively improve the half-word recognition effect, and is superior to other recognition algorithms.

Keywords: Water Meter; Image Recognition; Deep Learning.

1. Introduction

1.1 Background and Significance of the Study

Established in 1941, Changzhi City Urban Water Supply Group Limited Company is responsible for the water supply tasks for the production and use of domestic water by residents in Luzhou District, Lucheng District and along pipelines of Changzhi City. Its business scope is mainly urban water supply, which integrates water supply design, municipal engineering construction and water quality analysis and testing. After more than 80 years of progress, the company's tap water supply capacity is currently 1728,000 cubic meters per day, with an average daily water supply of 1616,000 cubic meters and a service population of about 1 million people. However, the difference between production and sales is low (i.e., the water supply efficiency is low). The company arranges meter readers to download the meter reading task for the day through the palm-in-one machine to the water supply department, and then rush to the site with the meter reader to complete the meter reading task, and then upload the results to the meter reading system after all the meter reading tasks for the day are completed. There are many drawbacks in such a working mode: First, meter readers have a large workload and low reading efficiency; Second, there are no errors or errors in the results of meter reading by meter readers. Thirdly, this method is very costly, in addition to the salaries of many meter readers, meter reading fees and outsourcing fees, etc. In the 21st century, with the continuous improvement of computer technology, all fields are striving to change to the direction of intelligence. Therefore, for the current manual meter reading problem, it is very necessary to have an effective method to extract and identify the water meter image more accurately, which can not only improve the efficiency but also reduce the expense of personnel.

Most of the traditional solutions use a global perspective to identify and split the reading area of the water meter. Then use the split model to further divide the reading area into individual characters. Finally, individual characters are recognized respectively. Although the traditional methods of recognition and segmentation are easy to understand and operate, there are some problems such as image noise interference and uneven distribution of recognition difficulty. In recent years, many scholars have further improved and deepened it. Compared with traditional methods, our water meter

digital recognition algorithm combined with DBNET-CRNN framework is more accurate and can handle the errors caused by more complex underwater environment.

1.2 Research Status at Home and Abroad

Researchers have developed a wheel meter reading detection and recognition algorithm based on Deep Learning. The Yolov4 target detection algorithm is used to detect and locate the image, which solves the problem of insufficient generalization ability of traditional methods and large recognition model, but the solution to the problem of missing or obscured numerical reading is still not in place. [1]

To solve the recognition problems caused by the serious contamination of the dial, camera overexposure, multiple shooting angles, and multiple dial directions, some researchers proposed a target detection algorithm based on Deep Learning: R-FCN algorithm, full name: Region-based fully convolutional network, the algorithm first designs specific labeling rules, and then used R-FCN algorithm to recognize and detect the text in the labeled frame. By using this algorithm, the problem of reading occlusion is solved, but the recognition accuracy can be further improved. [2]

In the reading recognition of wheel water meter based on the optimized convolution neural network, the author puts forward an improved method based on classical CNN (Convolutional Neural Network) model, uses SVM for character recognition and dropout function to speed up network convergence. This method solves the problems of poor recognition performance and slow convergence of existing traditional shallow learning algorithms. [3]

In the water meter reading recognition based on deep learning and its application, researchers not only solved some of the reading disturbances, but also proposed “an enhanced loss function” to solve the half-word reading problem by identifying the number in the “middle state” as the corresponding number in the “lower state”. [4]

A convolution neural network is built under the framework of Tensor Flow. Some researchers selected 3x3 convolution kernels to form a three-layer convolution neural network, which has achieved higher recognition accuracy. [5]

Some researchers have proposed a water meter digital recognition algorithm based on Lenet-5 network combination feature fusion. A convolution neural network is built under Tensor Flow deep learning framework, and it selects the convolution layer whose convolution kernel is 3×3, 1×3, 3×1 and increase the number of layers of the convolution layer. By this means, the recognition rate is greatly improved compared with the original Lenet-5 network. [6]

Researchers applied batch normalization to the Tensor Flow deep learning framework and then optimized the objective function, resulting in a new network whose recognition accuracy and speed significantly improved compared with the traditional VGG-16 network. [7]

Other researchers have retrained the output layer of Inception based on Tensor Flow, and proposed a water meter character recognition method with high recognition rate and high practicability. [8]

In the research of real-time water meter automatic recognition system based on BP network, the author obtains the text area by image pretreatment, then extracts and uses BP network to recognize a single character. This system can not only display the water meter data in real-time, but also has a high recognition accuracy. [9]

1.3 Research Content and Technical Framework

Text recognition is a common problem in the field of image processing. However, for images of natural scenes, the text position in the image must be positioned before it can be recognized. So in general, text recognition from natural scene pictures involves two steps:

(1) Text detection: The problem solved by text detection is where there is text and what is the scope of text.

(2) Text recognition: Text recognition is used to recognize the located text area. The main problem to be solved is what each text is, and convert the text area in the image into character information.

The water table digital recognition model proposed in this paper is also based on the above two steps, and the specific algorithm is shown in Figure 1:



Figure 1. Flow-process diagram

2. Digital Text Detection Model for Water Meter based on DB Network Module

2.1 Data Set Introduction

The training set contains 1000 water meter pictures and corresponding labeled answer results. The test set contains 500 pictures of water meters to be identified. The label of the training set is shown in Figure 2:



Figure 2. Training set example

The first eight values in the label are the four corners of the water meter dial (x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4). The origin of the coordinate values is the upper left corner of the picture, and the order is clockwise, starting from the upper left corner. The label is the value of the dial, and since the last digit has a half character, there are two values.

2.2 Traditional Text Detection Model

There are three traditional text detection models: The first is based on threshold, such as bimodal method, OSTU method, and so on. The second is based on region, such as regional growth, Watershed Algorithm, and so on. The third is based on clustering, such as K-means, FCM, and so on. Threshold-based methods are fast and efficient, but do not take into account differences in spatial characteristics. Region-based methods have high robustness but also bring high algorithm complexity. Cluster-based methods are good at solving fuzzy problems, but they are sensitive to initial parameters.

2.2.1 Threshold-based Text Detection Model

The threshold-based text detection model is divided into two steps: first, an optimal threshold is determined, then the image gray values are compared with the threshold values and classified. This method is simple and easy to understand, but how to determine an appropriate threshold becomes a challenge. In addition, this method is more sensitive to noise.

The histogram bimodal method is the most common one. It uses the corresponding gray level at the bottom of the bimodal valley of the image gray histogram as the threshold of the algorithm. The disadvantage of this method is that there can be many different images corresponding to the same histogram. Therefore, this algorithm not only has poor accuracy, but also investigates the knowledge base of the experimenter's image verification and to some extent improves the difficulty of the algorithm.

The OTSU algorithm takes the gray histogram of the whole image into account, and its gray level range is derived from the overall variance between the target and background of the image, by which

the appropriate gray level is selected. Because maximizing the segmentation by this method means minimizing the probability of misclassification. When the threshold falls in or near the histogram valley, the variance between the two classes is maximized, and the separation between the background and the target of the image is optimal, and the corresponding threshold is the optimal segmentation threshold for the image. OTSU algorithm is considered to be the best algorithm for threshold selection in image segmentation because it is simple to calculate and is not affected by image brightness and contrast.

2.2.2 Cluster-based Text Detection Model

Cluster-based text detection models use image similarity to divide an image into several homogeneous superpixel subregion. In a clustering-based text detection model, clusters generated by clustering are a set of data objects. Objects in the same cluster are similar, but objects in different clusters are different. Cluster analysis can maximize the similarity of objects in the same class and minimize the similarity of objects in different classes. Clustering is more suitable for images with low resolution, blurred edges and low contrast.

The basic idea of the classical K-means clustering algorithm is to cluster k points in space, classify the objects closest to them, and the number of categories is k. Keep iterating and update the values of each cluster center one by one until you get the best clustering results.

2.2.3 Region-based Text Detection Model

The area-based text detection model uses the similarity between the object and the background gray distribution to find the area directly. It can be divided into two categories: regional growth and regional division and merge. Specific examples are region growing method, region splitting, watershed, and so on.

Region Growing aggregates pre-defined pixels or subareas into larger regions. The basic principle is to merge adjacent pixels or regions with similar characteristics from a set of seed pixels (a single pixel or a specific small area) and gradually expand outward until no more similar feature pixels or regions exist. The key of the region growing algorithm is: first, the selection of seed pixels that correctly represent the desired region; Second, criteria to include adjacent pixels; Third, the conditions or rules for stopping the growth process. However, since the algorithm of region growing method relies on manual selection of seed points, it will require a greater amount of manual work for more complex or multi-area images that interfere with each other. At the same time, the anti-noise capability of the region growing method is weak, and separate regions will be connected.

Region splitting is a cyclic algorithm that divides an entire image into any disjoint regions and then merges them. The advantage of this algorithm is that it can reduce noise very well and further improve the accuracy of the image. However, the operation of this algorithm will destroy the boundary of the area.

The watershed method, which is based on the mathematical morphology of topology, can not solve the weak edge problem very well although it has a high efficiency of segmentation.

2.3 Text Detection Model based on Deep Learning Technology

The text detection model based on in-depth learning technology does not select a fixed threshold artificially like the traditional method, but uses the continuous learning of the neural network to predict the threshold at each location, which can better separate the background from the foreground. At the same time, such operations bring gradient non-differentiable situation to the network update, so we introduce DBnet, which can not only solve the gradient non-differentiable situation but also have a better partitioning effect.

2.3.1 Text Detection Model based on DBNET

As shown in Figure 3, the initial text detection model artificially sets a threshold for segmentation map to be converted to a binarization map. Then, slowly expand through the red area in the binarization map to find the text area(detection results). It's a bottom-up process. This method of

finding pixel points before determining the text area is artificially limited by the segmentation map area selection threshold.

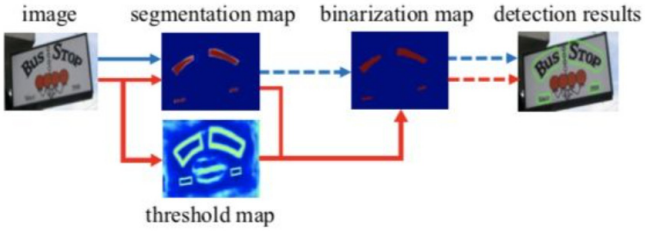


Figure 3. Evolution of Text Detection Model

Later, some scholars put forward improvements to this, the specific difference is in the selection of threshold value. That is, instead of using a fixed value, the network can predict the threshold value at each location. This adaptively separates the background from the foreground. However, gradients may not be differentiable during training. DBNET networks emerged as the times require. This Split-based scene text detection method is often more accurate, especially for curly text. Therefore, by introducing the DB module, we can not only simplify the binarization method better, but also solve absolutely indistinguishable problems.

The key to optimizing DBNET over traditional segmentation methods is to obtain approximate binary map from probability map and threshold map.

The inner area of the probability map text block is determined by the formula $D = \frac{A(1-r^2)}{L}$. The block edge area of threshold map is the middle area determined by the formula$D = \frac{A(1-r^2)}{L}$ and $D = \frac{A(1+r^2)}{L}$. Then P and T are used to denote the probability that each point in probability map and threshold map may be a text area, and the closer the point is to the border, the greater the probability. Last followed by formula $\hat{B}_{i,j} = \frac{1}{1+e^{-k(P_{i,j}-T_{i,j})}}$ determines the B value of each point in the approximate binary map. The specific process is shown in the Figure 4.:

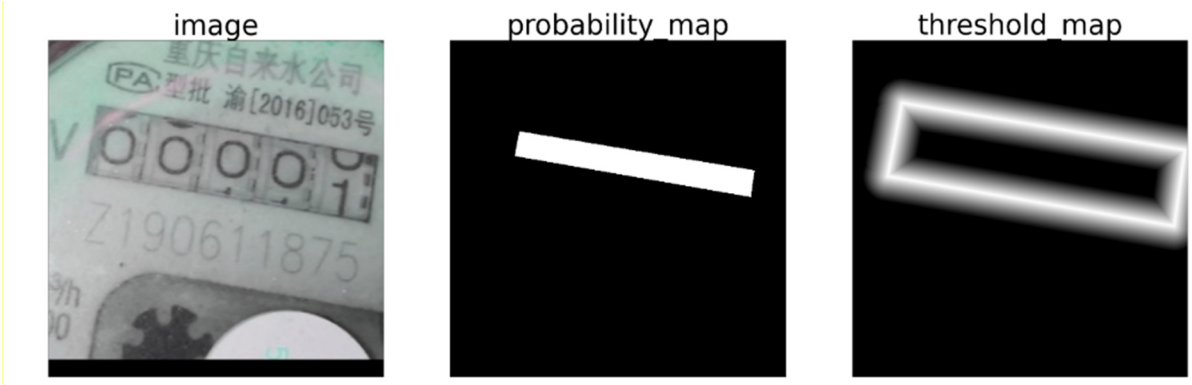


Figure 4. Example of image processing

3. Water Meter Digital Text Recognition Model based on Deep Learning Technology

After locating the position of the digital text, identify the number in the text area. Because the number in the text box is multi word, this paper uses a CRNN model that can generate serialized tags with high accuracy.

3.1 Difficulties in Text Recognition

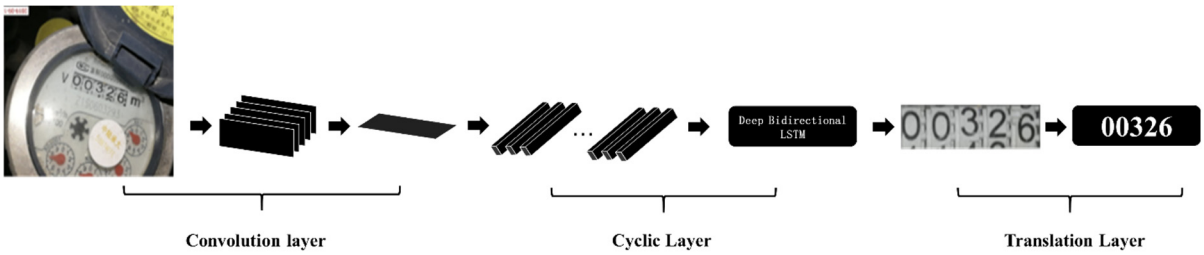


Figure 5. CRNN network structure diagram

3.2 Text Recognition Model based on CRNN Network

Input the text area of the water meter image selected by DBNET segmentation, and output its reading through the CRNN network. CRNN networks are divided into three layers: input the image from the first layer of the convolution layer, and extract its feature layer after convolution; In the second layer, the feature layer is divided into feature sequences, corresponding to a small part of the image, and then through the depth bidirectional LSTM to get the prediction results of each part of the feature sequence, so as to get a sequence prediction. In depth bidirectional LSTM, input/output or retention of information are judged by input sequence. Compared with the traditional RNN, the bidirectional LSTM increases the information of sequence inversion, so the trained accuracy is high and the stability is good. At the same time, due to the large number of bidirectional LSTM input parameters, its stable training times will increase relatively; Finally, through the translation layer, the output is integrated to get the final reading result.

3.3 Analysis of Experimental Results

Using the water meter digital recognition model combined with the DBNET-CRNN framework to analyze the image can effectively identify the text area, and the process is shown in:

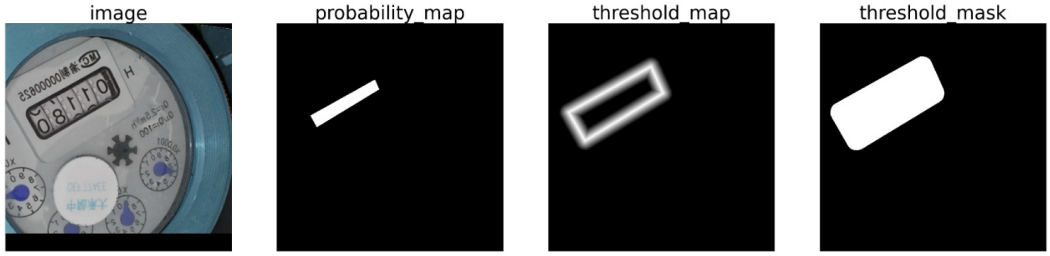


Figure 6. Text Detection Example

When we get the segmentation example as shown in Figure 6, CRNN network can be used for text recognition of the processed image, and the recognition accuracy can reach more than 98%. Analysis shows that only a few data sets that are difficult to distinguish by the naked eye are not recognized.



Figure 7. Text recognition example

4. Summarization and Prospect

This paper presents an algorithm of water meter digital recognition and detection based on deep learning. First, we design a text detection model based on DBNET network to recognize text area; Then, we use the text recognition algorithm based on CRNN network to predict and recognize the final result of the text area. The experimental results show that the model has achieved good results

in water meter number recognition task, and can be applied to water meter number detection and recognition in natural scenes.

References

[1] Research and Application of Reading Recognition of Word Wheel Water Meter Based on Deep Learning: (in Chinese).

[2] Water meter reading recognition based on depth convolution neural network: (in Chinese).

[3] Reading Recognition of Word Wheel Water Meter Based on Optimized Convolution Neural Network: (in Chinese).

[4] Water meter reading recognition based on depth learning and its application: (in Chinese).

[5] Water meter reading recognition based on convolutional neural network: (in Chinese).

[6] Water meter digital recognition based on Lenet-5 network combination feature fusion: (in Chinese).

[7] Water meter reading recognition method based on convolutional neural network: (in Chinese).

[8] Application of depth convolution neural network in water meter character image recognition: (in Chinese).

[9] Application of depth convolution neural network in water meter character image recognition: (in Chinese).