# Electric Energy Meter Information Recognition System Based on Deep Learning

Shuai Gao, Bo Ning(✉), and Nan Jiang

Dalian Maritime University, Dalian 116000, LN, China
ningbo@dlmu.edu.cn

**Abstract.** A new type of electric energy meter information recognition system based on deep learning is proposed. The system is mainly divided into OCR character recognition system and electric meter information verification system. OCR character recognition system mainly includes two parts: character detection and character recognition. The text detection uses the CTPN model, and the text recognition uses the CRNN network in deep learning for recognition, and then uses the CTC loss function for sequence processing to improve the accuracy of text recognition. Through the RCTW-17 data set training, an OCR text recognition system with high accuracy, strong stability and fast speed is obtained. The identified results are automatically checked with the information in the background database, and finally an electric energy meter information identification system is obtained. The verification of the RCTW-17 data set and the actual photo identification of the electric energy meter prove the effectiveness of this method.

**Keywords:** OCR technology · CTPN network · CRNN network · Electric energy meter information recognition

## 1 Introduction

The identification of electric energy meter information has always been a very important issue. With the rapid development of my country's economy, more and more places need to use electric energy, which is particularly important for the management of electric energy meters. The electric energy meter plays a very important role as an instrument for measuring electric energy. The method of manually copying the electric energy meter is inefficient and prone to errors. Therefore, it is extremely important to automatically identify the electric energy meter data. This method can not only save labor costs, but also it can also improve the accuracy and save time. The methods of identifying the data of electric energy meters can be roughly divided into two categories. One is the method of positioning and identifying according to the characteristics of artificial design [1–4]. Shape features for digital recognition. Yang Juan [5] completed the identification of electric energy meter numbers through image preprocessing, target positioning, and identification, but the positioning accuracy in complex scenes is low and the robustness is insufficient. Wang Chen et al. [6] used the TM-MSER algorithm to complete the


Y. Chenggang et al. (Eds.): MobiMedia 2022, LNICST 451, pp. 340–349, 2022.
https://doi.org/10.1007/978-3-031-23902-1_26

positioning and segmentation of digital instruments. Another method is to use a deep learning framework, using models such as convolutional neural networks, to extract features from images, and then perform detection and recognition to finally obtain data. Wu Binbin [7] et al. combined template matching and deep neural network technology, made full use of template calibration information, turned the indication detection under complex conditions into simple and effective isometric segmentation, and then identified, with better robustness. Li Ming et al. [8] used connected domain segmentation and localization and BP neural network for digit recognition. Chen Ying [9] and others proposed the Enhanced Faster R-RCNN network by using the features in the Faster-RCNN network, which improved the accuracy in identifying the information of the electric energy meter. Gong An [10] et al. proposed a recognition method of electric energy representation based on YOLOv3 network, which showed that this method has higher localization accuracy and recognition accuracy. It can be seen from the above literature that most of the digital recognition algorithms for electric energy meters are based on traditional image processing methods. However, the recognition accuracy of electric energy meters in complex scenarios is low, and there is a problem of insufficient robustness. This paper is based on deep learning. The method proposes a combination of CTPN [11] text detection model and CRNN [12] text recognition model to realize the digital recognition algorithm of electric energy meter, which improves the robustness and accuracy of the algorithm. After identification, the identified data will be transferred to the background database for comparison. If there is an electric meter of this type, the data will be updated. If not, a new data will be inserted to obtain a complete electric energy information table. It is hoped that in the future, the content of the electric energy meter can be identified by uploading photos to the system through mobile phones, which will benefit the society. The flow chart of the whole system is shown in Fig. 1:

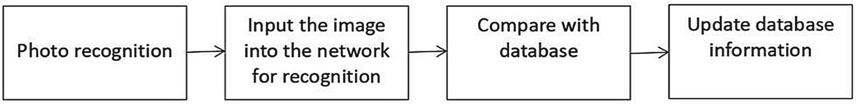


**Fig. 1.** System flow chart

## 2 OCR Text Detection Method

This paper mainly introduces the basic process of CTPN network detection and the basic structure of the network.

### 2.1 The Basic Process of CTPN Network Detection

The proposal of CTPN is based on the fact that text is usually written horizontally from left to right, and the width between words is roughly the same. Fixed width, to detect text height. It's essentially an RPN method that stitches together the detected boxes. Due to the above three characteristics, this model is particularly suitable for information identification of electric energy meters.

The specific process is as follows: First, the feature extraction is carried out through the backbone network VGG16, and the Conv5 layer outputs the feature map of N × C × H × W. Since the convolutional network of VGG16 has a cumulative stride of 16 after 4 pooling layers. That is, one pixel in the feature map output by the Conv5 layer corresponds to 16 pixels of the original image. Then do a 3 × 3 sliding window on Conv5, that is, each point is combined with the surrounding 3 × 3 area features to obtain a feature vector of length 3 × 3 × C. The final output is a feature map of N × 9C × H × W, which is still the spatial feature learned by CNN. Then continue to reshape the feature map output in the previous step:

$$\text{Reshape:}\quad \mathrm{N} \times 9\mathrm{C} \times \mathrm{H} \times \mathrm{W}(\mathrm{NH}) \times \mathrm{W} \times 9\mathrm{C} \tag{1}$$

Then feed the Bi-LSTM with a data stream of Batch = NH and maximum time length Tmax = W, and learn the sequence features of each row. The Bi-LSTM output is (NH) × W × 256, which is then reshaped to restore the shape:

$$\text{Reshape:}\quad (\mathrm{NH}) \times \mathrm{W} \times 256\ \mathrm{N} \times 256 \times \mathrm{H} \times \mathrm{W} \tag{2}$$

This feature includes both spatial features and sequence features learned by Bi-LSTM. Then through the fully connected layer, it becomes a feature of N × 512 × H × W. Finally, through an RPN network similar to Faster RCNN, the text detection frame is obtained.

Since CTPN is aimed at the detection of horizontally arranged text, it adopts a set of (10) fixed reference frames of equal width to locate the text position. The fixed reference frame width and height are:

$$\text{Widths} = [16] \tag{3}$$

$$\text{Heights} = [11, 16, 23, 33, 48, 68, 97, 139, 198, 283] \tag{4}$$

Since CTPN uses the VGG16 model to extract features, the width and height of the Conv5 feature map are 1/16 of the width and height of the input original image. At the same time, the fully connected layer is equal to the width and height of Conv5. CTPN is equipped with 10 above-mentioned fixed reference frames for each point of the feature map of the fully connected layer. After obtaining the fixed reference frame, similar to Faster R-CNN, CTPN will do the following processing: use Softmax to determine whether the fixed reference frame contains text, That is, a positive fixed reference frame with a large Softmax score is selected; then the center y-coordinate and height of the fixed reference frame containing the text are corrected using bounding box regression.

$$\mathrm{Vc} = \left(C_y - C_y^a\right)/h^a \tag{5}$$

$$V_h = \log\left(\frac{h}{h^a}\right) \tag{6}$$

$$V_c^* = \left(C_y^* - C_y^a\right)/h^a \tag{7}$$

$$V_h^* = \log\left(\frac{h^*}{h^a}\right) \tag{8}$$

Equation (5) and (6) are the coordinates of the regression prediction; $C_y^a, h^a$ is the center y coordinate and height of the fixed reference frame, and $V_c^*$, $V_h^*$ is the Ground Truth. After the fixed reference frame is processed by the above Softmax and bounding box regression, a set of vertical strip text detection boxes will be obtained. Subsequent text detection boxes only need to be connected together with the text line construction algorithm to obtain the text position.

The Loss function is:

$$\mathrm{Loss}(S_i, V_j, O_k) = \frac{1}{N_s}\sum\nolimits_i L_s^{cls}(S_i, S_I^*) + \frac{\lambda_1}{N_v}\sum\nolimits_j L_v^{reg}\left(v_j, v_j^*\right) + \frac{\lambda_2}{N_o}\sum\nolimits_k L_o^{reg}(O_k, O_k^*) \tag{9}$$

The main three errors of the above loss function come from the classification error of the fixed reference frame and the background, the vertical coordinate offset regression error, and the correction error of the fixed reference frame x at the boundary.

Network basic structure (Fig. 2):

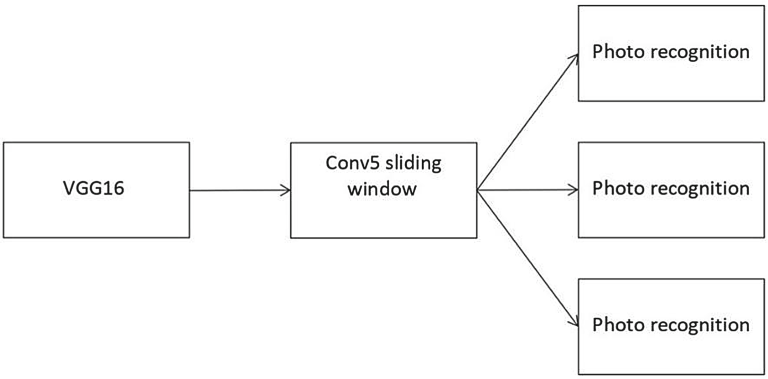


**Fig. 2.** Basic structure of CTPN network

Applied to the electric energy meter system proposed in this paper, the CTPN algorithm is used for segmentation detection, and the result is shown in Fig. 3:


**Fig. 3.** Text segmentation

# 3 OCR Text Recognition Method

This paper mainly introduces the basic process of CRNN network recognition and the basic structure of the network.

The CRNN model has many advantages. It can learn directly from sequence labels. Instead of labeling each character, you only need to label a sequence for a picture. For example, if the picture is "22.6 Kwh", the label is "22.6 Kwh" without having to label each character individually. Image features extracted by CNN. Using the input feature sequence of RNN training, the output is a sequence label. There is no length limit on the images to be trained, but normalize the height of the images. There are few parameters. Although CNN and RNN are combined, a loss function (CTC LOSS) is used for joint training in the end to achieve end-to-end training, which can directly get the final result we want and the recognition time is short.

## 3.1 The CRNN Network is Mainly Divided into Three Modules

1. Convolution module: The backbone network is an improved version based on the VGG model.
2. Recurrent network module: It consists of two steps of feature sequence extraction and two BLSTM training. B is bidirectional and can extract model information in two directions, not just one. The BLSTM of CRNN has two layers, and two layers can obtain higher-level sequence features.
3. Transcription module, the transcription module is mainly a CTC layer, and its main function is to train the loss function of the network.

### 3.2 The Specific Implementation Steps of CRNN

Feature sequence extraction is performed in the first step of the recurrent network module in the above figure. This is done by scaling the images to the same height. Then each feature vector of the feature sequence is generated from left to right in columns on the feature map. This means that the ith feature vector is the ith concatenation of all feature maps, and the width of each column is fixed to a single pixel in our setup. This approach can make the feature sequence in order, and can better perform the following cyclic network operations.

Transcription is the process of integrating all possible outcomes of the feature sequence predicted by the LSTM network into the final outcome. A CTC model is connected at the end of the bidirectional LSTM network to achieve end-to-end identification. The CTC model is connected to time classification. CTC can perform end-to-end training without requiring training data alignment and one-by-one labeling, and directly output sequence results of indeterminate lengths. CTC is generally connected in the last layer of the RNN network for sequence learning and training. For a sequence of length T, each sample point t (t is much larger than T) will output a softmax vector in the last layer of the RNN network, representing the predicted probability of the sample point, and these probabilities of all sample points are transmitted to After the CTC model, the most likely label is output, and after removing spaces and deduplication, the final sequence label can be obtained, that is, our final recognition result.

The structure of the CRNN model is shown in Fig. 4:

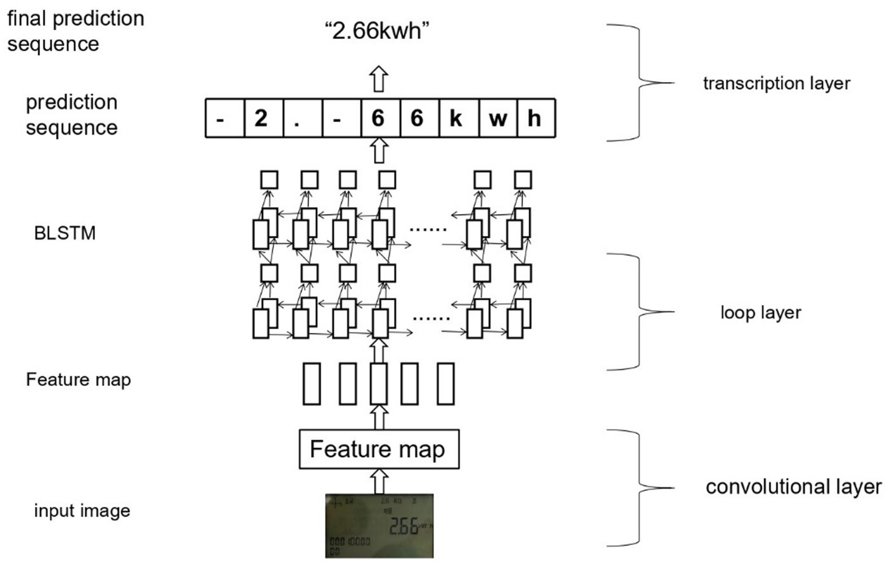


**Fig. 4.** CRNN model structure diagram

# 4 Experiment

## 4.1 Evaluation Indicators

The evaluation metrics for target detection are mAP (mean average precision) and accuracy. Before calculating mAP, it is necessary to understand the concepts of TP, TN, FP and FN. TP (true positives) refers to the predicted positive samples and matches the real results, TN (true negatives) refers to the predicted negative samples and matches the real results, FP (false positives) refers to predicting positive samples but does not match the real results, FN (false negatives) refers to predicting negative samples but does not match the real results. To judge whether it is consistent with the real results, it is necessary to obtain the intersection ratio (IoU) between the ground truthbox and the prediction box. When the IoU is greater than the set threshold (threshold_iou is generally set to 0.5), it means that the prediction box is the ground truth box, otherwise it means that the prediction box is the same as the predicted box. The actual results do not match.

The calculation of each category in object detection is shown in Eq. 10

$$precision_{class} = \frac{TP}{TP + FP} \tag{10}$$

The calculation formula of mAP is shown in 11, n represents the number of categories in target detection, and i represents the current category.

$$\mathrm{mAP} = \frac{1}{n} * \sum\nolimits_{t=0}^{n} precision_i \tag{11}$$

The calculation formula of accuracy is shown in Eq. 12

$$\mathrm{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

The evaluation standard of classification is the accuracy rate, and the calculation method is the ratio of the number of correctly classified samples to the total number of samples.

## 4.2 Electric Energy Meter Detection

**Table 1.** Electric energy meter test results

| Data set | Number of samples | CTPN | |
|---|---|---|---|
| | | mAP | Accuracy |
| Test set | 244 | 96.64% | 97.12% |
| Training set | 12034 | 93.34% | 91.33% |

Table 1 shows the experimental results of the electric energy meter detection. It can be seen from Table 1 that the mAP and accuracy detected by the electric energy meter are more than 91%, indicating that most of the images have detected the LCD screen area. The data set used in the detection process is 244 photos focusing on single-phase electric energy meters and three-phase electric energy meters taken from different angles. After experimental tests, as shown in Figs. 5 and 6, good results can be achieved.

**Fig. 5.** Electric energy meter pictures taken from different angles

**Fig. 6.** Processed grayscale image and detection frame

### 4.3 Electric Energy Meter Identification

**Table 2.** Electric energy meter identification results

| Data set | Number of samples | CTPN | |
|---|---|---|---|
| | | mAP | Accuracy |
| Test set | 244 | 92.42% | 93.52% |
| Training set | 12034 | 94.89% | 92.73% |

Table 2 shows the experimental results of digital recognition. It can be seen from Table 2 that the accuracy of digital recognition in the test set and training set is above 92%, and the number of samples is relatively large, which can represent the recognition results of electric energy meters taken from different angles, but A very small part of the recognition results have problems. The main reasons for the low accuracy in the test set are some blurred images, reflections and other problems. The recognition result is shown in Fig. 7.

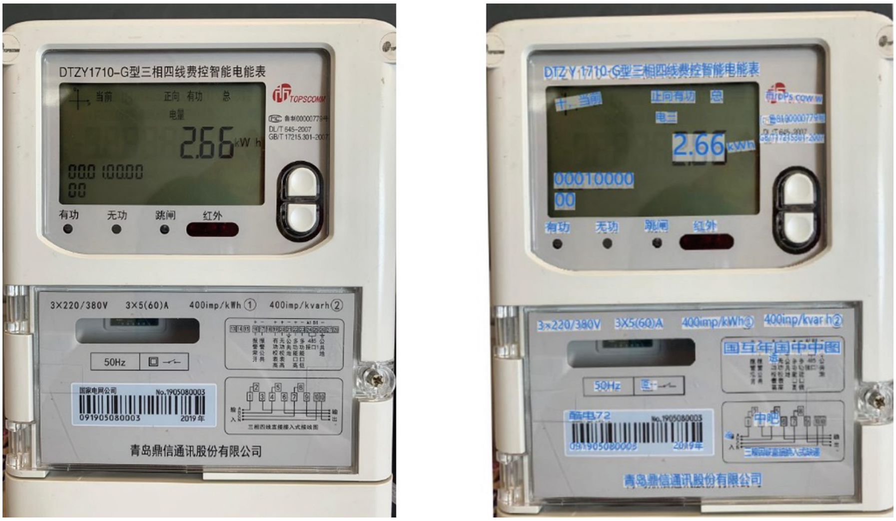


**Fig. 7.** Picture of a successfully detected electric energy meter

### 4.4 Model Comparison

**Table 3.** Model Comparison Results

| Model | mPA | Accuracy |
|---|---|---|
| EAST + CRNN | 89.24% | 88.45% |
| EAST + ATTENTION | 84.95% | 86.98% |
| SEGLINK + CRNN | 90.94% | 92.33% |
| SEGLINK + ATTENTION | 85.26% | 91.25% |
| FTSN + CRNN | 93.36% | 90.15% |
| FTSN + ATTENTION | 83.94% | 85.88% |
| **CTPN + CRNN** | **94.52%** | **94.26%** |
| CTPN + ATTENTION | 92.72% | 93.18% |

Table 3 presents the comparison of eight models. Through the comparison experiment, it can be seen that the CTPN + CRNN model has the best effect. The model in this paper can effectively identify the data in the electric energy meter, so as to achieve the function of collecting electricity consumption information.

## 5 Conclusion

In order to improve the accuracy of automatic identification of electric energy meters, a text detection and recognition model based on CTPN + CRNN network is constructed in this paper. Faster and more robust, it can detect and identify pictures of electric energy meters with high complexity taken from different angles, which has certain feasibility.

## References

1. Zhang, Z.: Research on the reading system of electric energy meter based on digital image recognition. Master's thesis. Beijing Jiaotong University, Beijing (2010)
2. Bei, C.: Research on image recognition technology of single-phase electricity meter. Master's Thesis. Nanjing University of Science and Technology, Nanjing (2014)
3. Zhang, X.: Research and implementation of digital image recognition technology of electric energy meter. Master's Thesis. Shanghai Jiaotong University, Shanghai (2007)
4. Ling, L.: Recognition technology of electric energy representation number based on image processing. In: Proceedings of the 2012 Academic Exchange Conference of the Electric Power System Automation Professional Committee of the Chinese Society of Electrical Engineering, pp. 1–5. Xiamen, China (2012)
5. Yang, J.: Research and Implementation of Electric Energy Meter Image Recognition Technology based on Digital Image Processing. Nanjing University of Science and Technology, Nanjing (2012)
6. Wang, C., et al.: Research on image-based digital instrument character wheel positioning and segmentation algorithm. Electron. Measur. Technol. **09**, 139–142 (2017)
7. Wu, B., Zhu, Y., Ge, Y., Lv, Y.: Information recognition of electric energy meter combined with template matching and deep neural network. Electr. Measur. Instrum. 1–9 (2021)
8. Li, M., et al.: Digital recognition method of electric energy meter based on image processing. Electron. Qual. **3**, 41–45 (2018)
9. Chen, Y., Jiang, W., Yang, F., Zhu, X.: Research on digital recognition algorithm of smart electric energy meter. Inform. Commun. (01), 17–21 (2020)
10. Gong, A., Zhang, Y., Tang, Y.: Recognition method of electric energy representation number based on YOLOv3 network. Comput. Syst. Appl. **29**(01), 196–202 (2020). https://doi.org/10.15888/j.cnki.csa.007215
11. Tian, Z., Huang, W., He, T., He, P., Qiao, Y.: Detecting text in natural image with connectionist text proposal network. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VIII, pp. 56–72. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_4
12. Baoguang, S., Xiang, B., Cong, Y.: An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Trans. Pattern Anal. Mach. Intell. **39**(11), 2298–2304 (2017). https://doi.org/10.1109/TPAMI.2016.2646371