ISBN: 978-981-18-5852-9
WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications
doi: 10.18178/wcse.2022.04.171

Intelligent Reading for Multi-Scale Engineering Drawings based on Adaptive Object Detection and Optical Character Recognition

Rui Yao+, Xin Cheng, Zhilei Hui, Zexin Li, Zhaoe Min

School of Computer science, Nanjing University of Posts and Telecommunications, Nanjing, China

Abstract. Aiming at the problem that a single OCR model cannot extract signs from engineering drawings, and that the multi-size and super-size of input drawings seriously affect the performance of deep learning models, a new intelligent engineering drawing reading scheme is proposed in this paper. Our scheme contains several deep learning detectors, such as YOLO object detector, text angle predictor, DB text detector and CRNN line-level text recognizer. In addition, some important rules and methods are included in the pipeline to unify the input dimensions. We first set up the data set provided by the grid company. For the collected drawings, we carried out format conversion, angle adjustment and size adjustment, and constructed the annotated sample data set of secondary wiring drawings. Then, YOLO V5 model is used to detect the object in the drawing, DB+CRNN is used to detect and recognize the text in the drawing, and knowledge distillation strategy is used to train an ultra-lightweight small model for mobile inferring. The experimental results show that the YOLO V5 object detector achieves 0.98 score under map@0.5 index in this task, and the lightweight OCR model achieves 0.648 score on F1 index when its volume is compressed to 4.7% of the large model volume, which is only 0.031 lower than the large model. In addition, the size transformation rules made by us ensure that the model can keep relatively uniform performance for different sizes of drawings, and prevent the memory explosion caused by large size drawings in industrial applications.

Keywords: Digitization of Engineering Drawings, Optical Character Recognition, Object Detection, Secure Redundancy Segmentation Strategy.

1. Introduction

Today, the growing need to digitize and compute engineering drawings (EDs) is in conflict with the mounting piles of paper-based drawings and worried looking workers in factories. In the process of drawing digitization, some key technologies have been studied for a long time. For example, optical character recognition (OCR) is used to extract text from drawings. However, this technique is mainly suitable for dealing with prints with simple background, but not able to extract some signs or patterns that contain important information [1]. In addition, there are problems of multi-size and oversize in the recognition of EDs. In the power scenario, for example, some typical drawings are around 2000 by 3000 pixels in size, while others are 4000,5000 or even tens of thousands of pixels in length or width. It has a very negative effect on the deep learning model.

In this paper, we propose a method for Intelligent reading for multi-scale engineering drawings based on adaptive object detection and optical character recognition. The main contributions of this work are outlined as follows:

1) According to the latest research progress in the field of OCR, we introduce a two-stage OCR algorithm, i.e., Differentiable Binarization (DB) and Convolutional Recurrent Neural Network (CRNN), to locate and recognize text line.

2) We introduce the YOLO V5 model to detect symbols and patterns in EDs.

3) We propose methods to handle drawings with large variation in size i.e., secure redundancy segmentation strategy for oversized drawings, and scale-up and black trim filling strategy for small ones.

4) We apply the transfer learning method based on the pre-training model to train the OCR model, which greatly alleviates the problem of insufficient annotated data in industrial scenes.

+ Corresponding author.
E-mail address: b18011527@njupt.edu.cn.

5) We apply the knowledge distillation strategy to reduce the volume of the OCR model to improve the efficiency of the system on the mobile terminal.

2. Related Work

This section introduces relevant literature and techniques. Firstly, we discuss the application of OCR in the field of industrial drawing recognition. Common object detection techniques and how to apply them to industrial drawing symbol detection are then introduced.

2.1. Optical Character Recognition

OCR has a long history of research. Traditional methods rely on manual feature extraction, which has low recognition accuracy and poor generalization performance. After the emergence of deep learning methods, neural network models can automatically learn features and have strong nonlinear fitting ability.

OCR methods based on deep learning are generally divided into two stages: text detection and text recognition.

1) Text Detection: There are two ways of text detection, based on regression and based on segmentation. The regression-based algorithm adds a textbox proposal layer on the basis of the convolutional network, then judges the probability that the proposal is the text and adjusts the position of the positive sample proposal. Based on this idea, Minghui Liao et al. proposed TextBoxes [2]. The main innovation is to adjust the convolution kernel to several strips with different aspect ratios shape, so that it is more robust to the detection of slender text lines. Jianqi Ma et al. proposed RRPN [3], Rotation Candidate Region Network, which can generate text Candidate regions with slanted angles. The main problem of text detection based on regression is that text boxes with present shapes cannot describe text with excessive aspect ratio or curved text well.

Different from the regression-based method, which needs to output rectangular object box, the segmentation-based method aims to find all the pixels corresponding to the same object. A typical sample of this kind of work is the fused text segmentation network [4] proposed by Yuchen Dai et al., which is a solution for curved text detection. The main problem with segmentation-based text recognition algorithms is that they often require complex post-processing, such as using clustering algorithms to combine pixel-level results into a text line, which is often expensive in prediction. Minghui Liao et al. [5] proposed a module called Differential Binarization (DB), which can perform binarization process in segmentation networks. This work greatly simplifies post-processing.

2) Text Recognition: In terms of text recognition, some common methods model character recognition as image classification, using CNN network; Other methods pay attention to the transcription of line-level text.

Methods using CNN to identify individual characters in EDs typically require training a robust character detector. This detector is responsible for accurately detecting individual characters and clipping each character from the original image.

Convolutional recursive neural network (CRNN) is a representative deep learning model for line-level text recognition [6]. It uses CNN features as input, bidirectional LSTM for sequence processing, and then translates the results through CTC to get the output results. This method greatly improves the accuracy of text recognition.

2.2. Object Detection

There are image segmentation techniques in the field of computer vision applied to the task of industrial drawing detection. For example, Carlos Francisco et al. [7] attempt to localise the actual pixels that constitute each shape and text in industrial engineering drawings. They propose a Convolutional Neural Network (CNN) capable of classifying each pixel, which is able to classify all pixels in the drawing as symbols, text, or connector pixels. However, it is difficult to obtain a reliable source of training samples to classify each pixel individually.

For symbols and other patterns in drawings, we can use the method of object detection. The mainstream algorithms of object detection are divided into two types: two-step method and one-step method, and both of them are applied to the task of industrial drawing detection. The famous R-CNN model [8] has proved that CNN can be used for region-based localization and segmentation of objects. Although R-CNN series of algorithms improved the accuracy of the object detection task, the two-step algorithm has also been criticized for the amount of calculation and the slow operation speed. In contrast, one-step method formulates the detection task as a unified, end-to-end regression problem [9]. In our work, we apply YOLO V5 network to detect industrial object in wiring drawing scenario. YOLO implicitly encodes contextual information about classes and their appearance, making the number of background errors significantly reduced, which especially makes sense for EDs are full of extraneous lines and it is challenging to distinguish the background from the object.

All of the aforementioned methods work with a varying degree of success for object detection task and optical character recognition task. Overall, in our work, we apply DB [5] and CRNN models [6] to OCR problems and YOLO V5 model to symbol detection problems. In different problem scenarios, different models can be replaced to perform the above tasks.

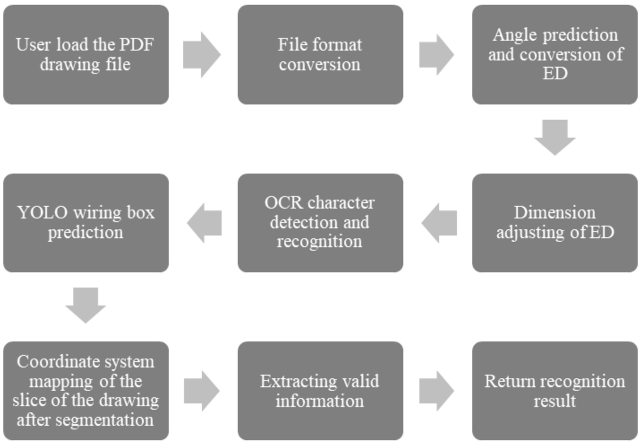

User load the PDF drawing file
File format conversion
Angle prediction and conversion of ED
Dimension adjusting of ED
OCR character detection and recognition
YOLO wiring box prediction
Coordinate system mapping of the slice of the drawing after segmentation
Extracting valid information
Return recognition result

Figure 1. Pipeline system overview.

3. Proposed Method

3.1. Image preprocessing module

After collecting drawings, we first preprocess the files properly to make detection and recognition effective.

1) Format conversion: EDs on PDF format offered by design firms must be converted into JPG format and the appropriate DPI should be set so that calculation will not occupy too much memory space under the condition of acceptable resolution of images.

2) Direction adjustment: Directions of EDs have a significant impact on detecting text and symbols, and hence we first run Angle Classifier to calculate the orientation of input drawings [10]. For non-forward drawings, the direction of the drawing is rotated to forward at this stage.

3) Adaptive scale adjustment: To provide a uniform size of image input for YOLO model and avoid memory explosion, we have to enlarge small drawings and segment large ones to ensure that the input size of the final image is 3488*2560 [11].

First, small drawings need to be enlarged to the expected size, but this may make the text and symbols on it to be out of shape and hard to recognize, hindering the follow-up work. Therefore, we develop the strategy of scale-up and black trim filling. With scale-up of small drawings, the rest blank of the expected dimension needs to be filled with black trim, guaranteeing the expected input size and symbol features. An example of scale-up and black trim filling is demonstrated in Figure 2.

When it comes to oversized drawings, downscale their size is not feasible, for the accuracy of the recognition model will decrease with the reduction of the target in the later stage. We choose to split the picture into several small pieces and recognize them in turn, and then aggregate the final results. For fear of cutting apart wiring boxes and affecting recognition results in pieces, we propose the secure redundancy segmentation strategy. There contains a public area of m*n case between neighbouring pieces, namely redundant block. Even though a target to be recognized can be incomplete in one piece, another piece will retain the entire object.

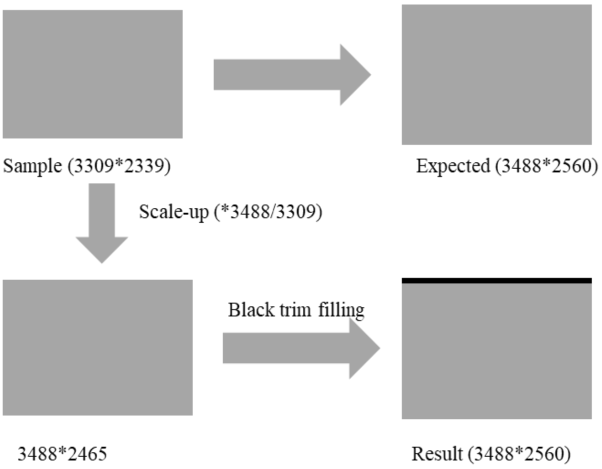


Figure 2. An example of scale-up and black trim filling.

Specifically, we adopt the following implementation method of secure redundancy segmentation strategy. Firstly, the anchor sizes of all targets are clustered and it is found that redundant block of 300*300 can cover targets to be recognized. Therefore, when the drawing is segmented horizontally, m=3488 and n ⩾ 300 are required for public redundant regions; while it is segmented vertically, n=2560 and m ⩾ 300 are required. An example of secure redundancy segmentation strategy is demonstrated in Figure 3. Block A is a public redundant block of two transverse pieces. When the subsequent prediction model completes the prediction and obtains the coordinates of the target, all coordinates are mapped to the coordinate system of the original drawing. In this way, we gain exactly the same recognition result as that of direct recognition of original drawings. In addition, input size of YOLO model is standardized, preventing system from crash, which is resulted from recognizing oversized EDs directly.

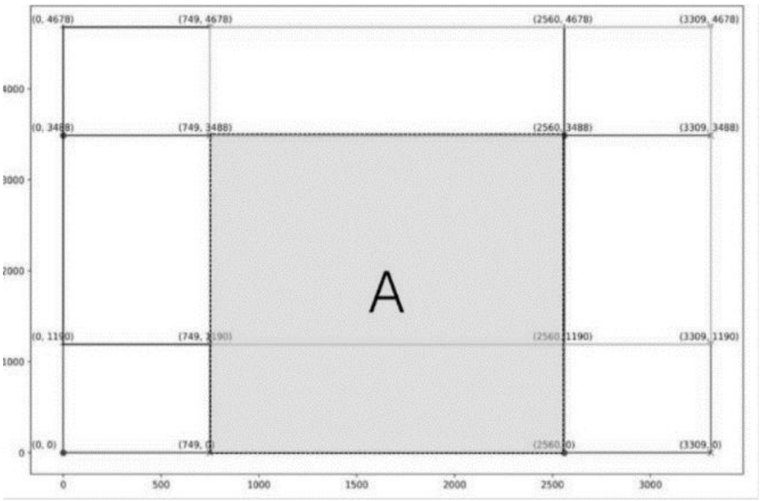


Figure 3. An example of secure redundancy segmentation strategy.

3.2. Model prediction, coordinate transformation and result returned

In order to illustrate this process vividly, we take a real application as an example. In the substation secondary wiring check scenario, the task is to get wire boxes and look for the characters on the wiring signs corresponding to them. Based on the coordinates of the wiring boxes, the wiring signs are horizontally matched to wire information. As shown in Figure 4, YOLO v5 detect wiring boxes on the right of Sign 4 and

Sign 5, which means we should return the wiring information of Sign 4 and Sign 5. So, in this example, the system needs to return (4, YMB803) and (5, YMB804).

The model works as follows:

1) YOLO wiring box prediction module is run and the coordinates of each wiring box is got.

2) OCR module is run and the coordinates of each character line and their recognition results are obtained. It can be known that the labels corresponding to the two wiring boxes are 4 and 5 respectively.

3) For each piece in the oversized drawing, Transformation program of coordinates is run to map all the recognition locations in pieces to the coordinate system of the original drawing. The recognition effect obtained by this mapping is exactly the same as that obtained by direct recognition of large drawings.

4) Based on the coordinates of the wiring boxes, the wiring signs are horizontally matched to wiring information. Label 4 and 5 are horizontally matched to the wiring information YMB803 and YMB804 respectively.

5) Information of each wired pair of sign and wiring is returned.

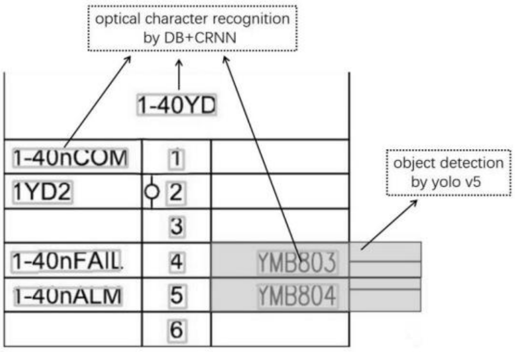

optical character recognition by DB+CRNN
1-40YD
1-40nCOM
1
1YD2
2
3
1-40nFAIL
4
YMB803
1-40nALM
5
YMB804
6
object detection by yolo v5

Figure 4. Task specification

4. Experiments and Results

4.1. Experimental Environment and Data Set

The experimental environment used in this paper is Linux operating system, Intel(R) Xeon(R) Gold 6148 4-core CPU @ 2.40GHz, 32GB RAM, Tesla V100 GPU with 32GB video memory. The deep learning framework is PaddlePaddle.

Our original data set contains 84 secondary-wiring drawings of variable size from the grid corporation. After the pre-processing operations described above, the dataset was expanded to 211 drawings with a fixed size of 2560*3488 pixels. A sample of the dataset is shown in Figure 5.

4.2. Evaluation Indicators

1) Object Detection: We mainly use map@0.5 score to evaluate the model. In this example, map@0.5 reflects the mean average precision of the predicted results in all drawings when IOU is set to 0.5.

2) Text Detection: In this phase, we observed the performance of the fine-tuned model under precision, recall and hmean (F-Score) to objectively evaluate our model.

3) Text Recognition: Since the pre-trained CRNN model can be directly used for the prediction of this project with excellent accuracy, we did not conduct the text recognition experiment.

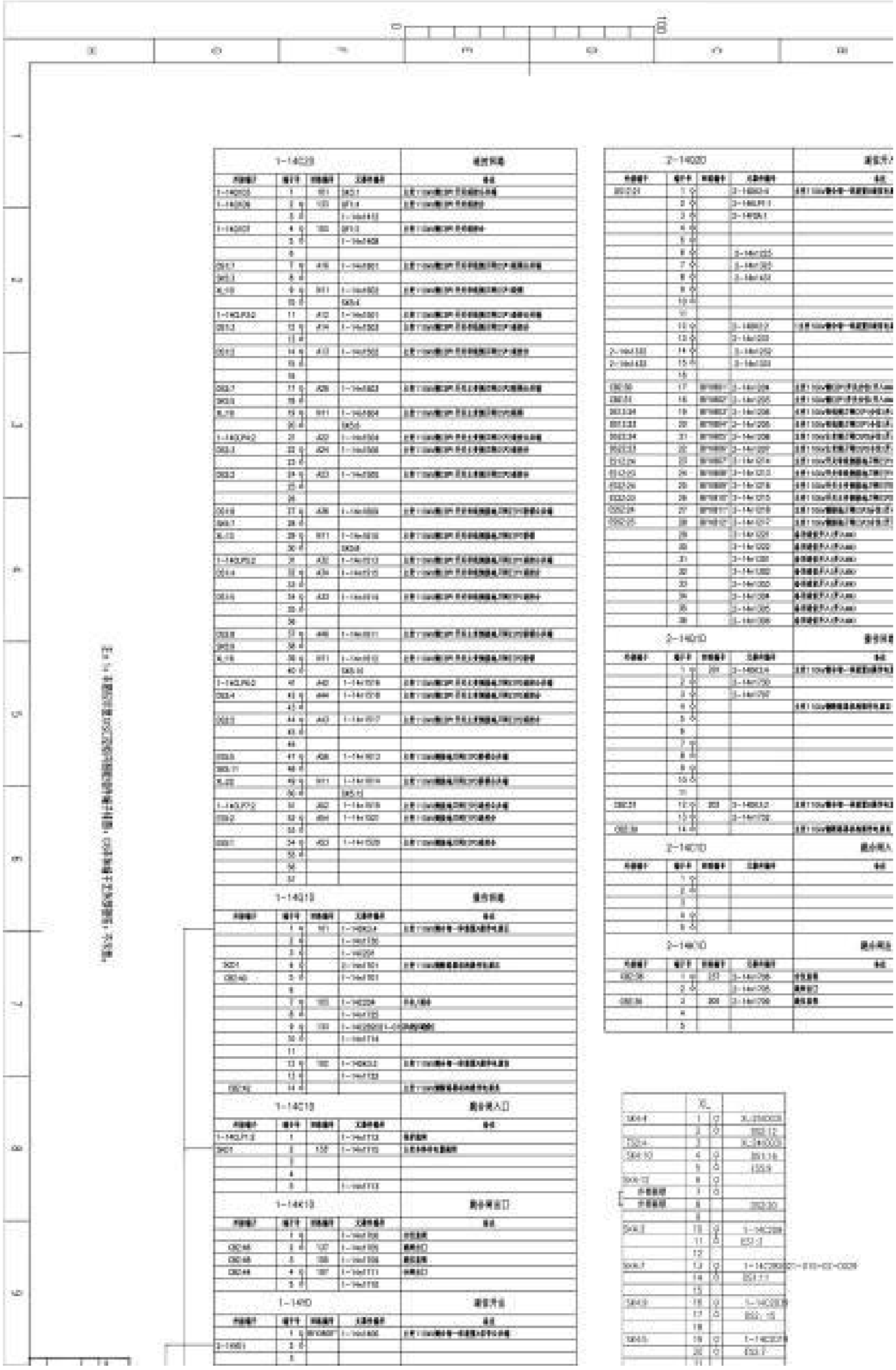


Figure 5. sample dataset

4.3. Results

1) Object Detection: After training for 731 epochs, the model triggered the early-stop strategy (the last 100 epochs did not result in a decrease in loss for the evaluation data set) to avoid over-fitting. Finally, our best model under the indicator map@0.5 reached more than 0.98. Figure 6 shows the training curve along the way.

2) Text detection: When it comes to the training of OCR model, it is difficult to achieve a good training result completely based on this data set, which has at least two problems: slow loss decline and poor prediction accuracy. Therefore, we used the idea of transfer learning for reference. Based on the pre-trained model provided by Baidu Inc., we combined our project data for fine-tuning. Then, we used the knowledge

distillation strategy to train the smaller model based on the finetuned large model to facilitate end-side deployment and inference in actual application.

Specifically, we loaded the PP-OCRv2 pre-trained model and used a Tesla V100(with 32GB Video Memory) GPU for fine-tuning. Limited by computational resources, the large size of the drawings, and the time-consuming knowledge-distillation process, we conducted only 280 epoch experiments, with the best model occurring in the 201st epoch. The following figure shows the variation of the evaluation indicators on the evaluation data set during training.

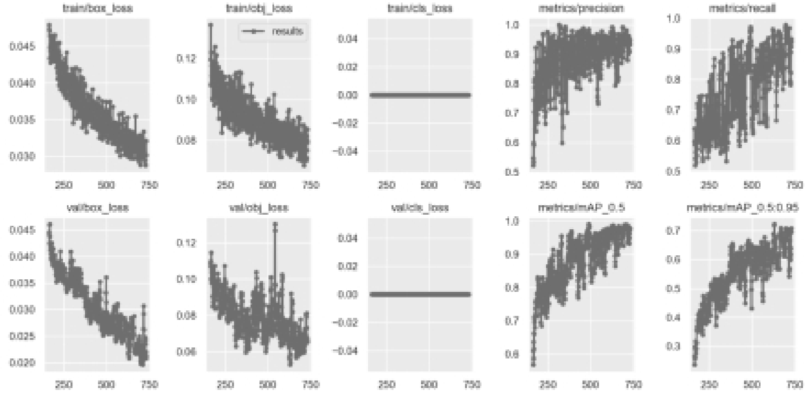


Figure 6. YOLO training curve

The control experiment on the testing data set using the teacher model and the student model is shown in Figure 10. According to the experimental results, the effectiveness of our training method can be proved: with limited computing resource and data, we have trained a model that is smaller but powerful enough. This is of great benefit to subsequent deployment and inference in mobile devices.

Further, we used the pre-trained model and the finetuned model for inferring on the testing data set, and representative results are shown in Figures 11 and 12.

Figure 11 shows the performance of the pre-trained model, which was not able to distinguish the circle marks from the text compared to the fine-tuned model in Figure 12. This distinction is not easily captured in Figure 10, but the ability to distinguish extraneous symbols has clear and important meanings for downstream tasks.

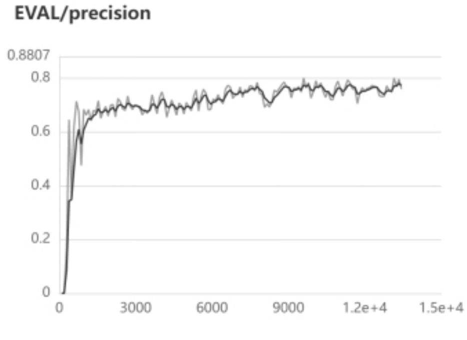


Figure 7. Precision curve of DB model during training process

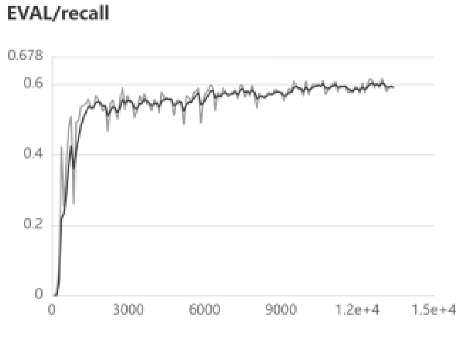


Figure 8. Recall curve of DB model during training process

5. Conclusions

In this paper, we proposed a solution to meet the actual demand which is mentioned in the introduction. The experiments done on real EDs from the power grid company demonstrate the effectiveness of our model in terms of excellent performance and the feasibility of the strategies such as secure redundancy segmentation for oversized drawings. Our method can be applied to other tasks of digitizing engineering drawings in many other fields and some practical problems which can be modelled as an object detection and OCR, not just limited to the substation secondary wiring check in this paper. In the future, we can replace pipeline components flexibly with better OCR and objective detection model to further enhance its performance. Additionally, we will collect richer datasets and use Generative Adversarial Network (GAN) and other tricks for enhancing image dataset.

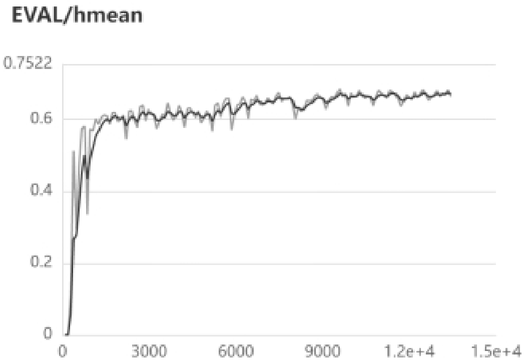


Figure 9. Hmean curve of DB model during training process

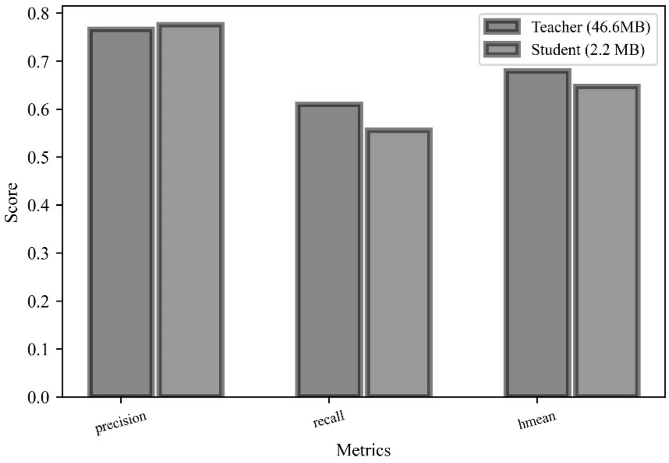
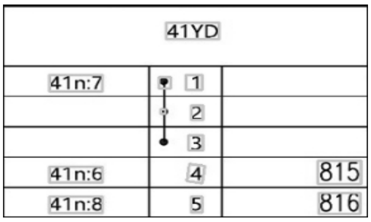
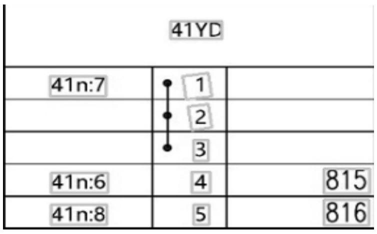


Figure 10. Evaluation indicators of the teacher model and the student model



41YD		
41n:7	1	
	2	
	3	
41n:6	4	815
41n:8	5	816

Figure 11. Performance of the pre-trained model



41YD		
41n:7	1	
	2	
	3	
41n:6	4	815
41n:8	5	816

Figure 12. Performance of the fine-tuned model

6. References

[1] M. T. Nguyen, V. L. Pham, C. C. Nguyen, and V. V. Nguyen, “Object detection and text recognition in largescale technical drawings,” 2021.

[2] M. Liao, B. Shi, X. Bai, X. Wang, and W. Liu, “Textboxes: A fast text detector with a single deep neural network,” in Thirty-first AAAI conference on artificial intelligence, 2017.

[3] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, and X. Xue, “Arbitrary-oriented scene text detection via rotation proposals,” IEEE Transactions on Multimedia, vol. 20, no. 11, pp. 3111–3122, 2018.

[4] Y. Dai, Z. Huang, Y. Gao, Y. Xu, K. Chen, J. Guo, and W. Qiu, “Fused text segmentation networks for multi-oriented scene text detection,” in 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018, pp. 3604–3609.

[5] M. Liao, Z. Wan, C. Yao, K. Chen, and X. Bai, “Real-time scene text detection with differentiable binarization,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, 2020, pp. 11 474– 11 481.

[6] B. Shi, X. Bai, and C. Yao, “An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition,” IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 11, pp. 2298–2304, 2016.

[7] C. F. Moreno-García, P. Johnston, and B. Garkuwa, “Pixel-based layer segmentation of complex engineering drawings using convolutional neural networks,” in 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020, pp. 1–7.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.

[10] Y. Du, C. Li, R. Guo, X. Yin, W. Liu, J. Zhou, Y. Bai, Z. Yu, Y. Yang, Q. Dang et al., “Pp-ocr: A practical ultra lightweight ocr system,” arXiv preprint arXiv:2009.09941, 2020.

[11] H. Allioui, M. Sadgal, and A. El Fazziki, “An improved image segmentation system: A cooperative multi-agent strategy for 2d/3d medical images,” Journal of Communications Software and Systems, vol. 16, no. 2, pp. 143–155, 2020.