



# Engineering Drawing Manager: A Smart Data Extractor and Management Software

Honglong Yang<sup>1</sup>, Yang Du<sup>1</sup>, Jingwei Guo<sup>1</sup>, Shiyi Wei<sup>1</sup>, and Hongbin Ma<sup>1,2</sup>(✉)

<sup>1</sup> School of Automation, Beijing Institute of Technology,  
Beijing 100081, People's Republic of China  
[mathmhb@139.com](mailto:mathmhb@139.com)

<sup>2</sup> State Key Laboratory of Intelligent Control and Decision of Complex Systems,  
Beijing 100081, People's Republic of China

**Abstract.** Engineering drawings play an important role in fields such as architecture, industrial engineering, and electric engineering, within which tables contain essential data and structures. However, most engineering drawings exist in the form of scanned PDFs or images, which is inconvenient for data management and storage, especially for table information. Also, many industries are in urgent need of data management software for engineering drawings to improve the degree of digital preservation and management. To this end, a software, ED Manager which is based on the fusion of deep learning and traditional image processing, is presented to detect the position and structure of the table, split and recognize characters, and reconstruct the table in a digital form. Further, we extract crucial information and develop a user interface and database to construct a comprehensive model that fits most engineering drawings. Our software can accurately locate tables for various complex drawings, extract structured information from tables, and build a better data management software for engineering drawings.

**Keywords:** Engineering drawings · Table reconstruction · Table structure recognition · Optimal character recognition

## 1 Introduction

In the context of “Made in China 2025” [16] and industrial upgrading, various fields have the demand for transformation to digitalization, intellectualization, and internationalization. In this process, engineering drawings, as information carriers for design and manufacturing, play an important role. For example, in the field of architecture, manufacturing, and energy, engineering drawings contain important information on design structure and component details.

The data presented in the title blocks of engineering drawings record important information, including the designer, design date, product parameters, etc. Such information can be used as a reference for new products or can be used to improve directly on the original drawings, thus shortening production cycles and reducing research and development costs.

Nowadays, with the development of the industrial manufacturing field, a huge number of engineering drawings have been accumulated. However, these drawings are difficult to manage. Although CAD software is widely used in the design process, most of the drawings in the fabrication plants exist in the form of scanned copies, prints, and other images. Since the title block, which contains unstructured information and is presented in various styles, cannot be directly recognized by computer system and can only be scanned and queried manually, it greatly reduces the efficiency of engineering drawings utilization and is not conducive to the digital preservation and management of engineering drawings. So far, there is no solution to this problem on the market. Therefore, the development of software that can provide a quick way to manage drawing information and a retrieval function is an urgent need. To achieve this solution, we combine methods including table detection, table line extraction, cell extraction, character segmentation, optical character recognition, and key information extraction.

For table recognition, many methods based on deep learning have been proposed. Gilani et al. [2] implemented and improved the Faster R-CNN [11] model and proposed an image transformation method that converts the table image to a natural image. Huang et al. [5] improved the precision of table recognition by introducing adaptive adjustments, an Anchor optimization strategy, and a post-processing method to the original YOLOv3 [10] model. Also, Qasim et al. [9] proposed an architecture based on graph networks for table recognition; and formulated the table recognition problem as a graph problem.

For table structure recognition, Suzuki et al. [1] proposed a border following algorithm to detect the borders of binary images, which can be used to detect the table cell well. Based on CNN, Siddiqui et al. [13] used semantic segmentation techniques based on FCN to recognize the rows and columns of tables well. Also, based on transformer architecture, Nassar et al. [8] presented a new end-to-end model by using a transformer-based decoder, which can solve complex column/row-header configurations, multiline rows, different variety of separation lines, missing entries, etc.

For character segmentation, many methods based on traditional image processing can segment images with merged characters, noise inference, and unknown size well. For instance, the Histogram-Based Character Segmentation method can both segment the horizontal and vertical characters with high accuracy. Many methods based on the connected components can also segment the characters well.

For optimal character recognition problems, many models based on recurrent neural networks and transformer architecture have been proposed to recognize characters of printed and natural images. SHI et al. [12] presented the CRNN model which can handle arbitrary lengths sequence without requiring character segmentation or horizontal scale normalization. Li et al. [7] proposed an end-to-end model named TrOCR, which is based on the transformer architecture without using CNN as the backbone.

The related studies mentioned above provide partial solutions to the problem of identifying and managing information in the table of engineering drawings and provide algorithmic processes for drawing information extraction. Although

the software available in the market can achieve information extraction, content recognition, and text reconstruction of tables, they have some shortcomings. Firstly, they do not provide solutions in an engineering context. Secondly, they do not have an automatic positioning function to detect the table. Thirdly, they can only recognize tables with certain formats, such as title bars with less dense lines. In addition, they do not provide functions for data management and key information extraction.

The rest of the paper is organized as follows: Sect. 2 describes software structure. Section 3 explains the key algorithms used in ED Manager. Section 4 explicitly illustrates the improvement of the histogram projection method and an innovative algorithm called dots-connection method. Section 5 contains the software recognition results on different drawings and comparative analysis with other software. Section 6 concludes the software and provides some directions for future research.

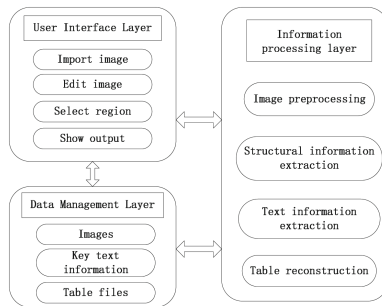
The contributions of this paper are summarized as follows:

- (1) we improved the character segmentation algorithm to extract words from a sentence with a traditional image processing method.
- (2) we improved the table line extraction algorithm to avoid the problem of text misidentification due to text sticking.
- (3) Based on deep learning and traditional image processing techniques, the software extracts both the structure information and content information, reconstructs the tables, and builds a data management supporting keywords search, which provides a solution to the problem of digital preservation and information management of engineering drawings.

## 2 Overview of ED Manager

### 2.1 Software Framework

Figure 1 shows the framework of ED Manager, which can be divided into three main parts: user interface layer, information processing layer and data management layer.



**Fig. 1.** Software framework

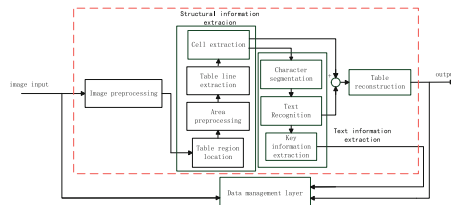
The user interface layer provides basic operating functions and interfaces of other layers' functions. After reading the instructions, the user can run the software through the interface layer such as importing and editing images, selecting processing regions, and viewing the final result.

The information processing layer reconstructs the table and extracts the key information through four main steps. The input of this layer is an selected image. The output is a html and csv file transmitted to the user interface layer for instant check and to the data management layer for long-term storage.

The data management layer defines several operations for users to add, delete, modify and require the stored data. These functions will be shown graphically on the user interface layer. The data management layer contains three kinds of data: image, text and table file. Image is from user input and used for long-term storage and checking results; text comes from the key information extraction in the information process layer; table file from the table reconstruction step in the information process layer includes both the text content and location information.

## 2.2 Workflow of Information Processing Layer

The workflow of the key algorithms is shown in the red box area of Fig. 2, and the four main steps are image preprocessing, structural information extraction, text information extraction and table reconstruction.



**Fig. 2.** Workflow of information processing layer

- (1) Image preprocessing: At this stage, operations include image binarization, resizing, and enhancement of the whole image.
- (2) Structural information extraction: The Structural information extraction includes four steps: table region location, area preprocessing, table line extraction, cell extraction. Table region location aims to automatically locate the tables on the image; also, the final table region can be chosen and modified manually. Table line extraction extracts the table line and reconstructs the table. The cell extraction extracts each cell by the findContours algorithm.
- (3) Text information extraction: Character segmentation and text recognition are applied to each cell to convert the optical character into a digital format; after that, key information such as title, date and designers are identified, extracted and stored in the database.

- (4) Table reconstruction: Table reconstruction combines the location information and text information to reconstruct the comprehensive information of the table.

### 3 Key Algorithms

#### 3.1 Image Preprocessing

Image preprocessing aims to improve the effectiveness of information extraction and computational efficiency and is usually a necessary process to realize the key image processing algorithms. Specifically, grayscale and binarization are used to reduce the amount of calculation; erosion and dilation are used for filtering out the text and extracting the straight line; the median filter is used for removing noise points; image pyramid is used for expressing the image at multiple resolutions so the segment algorithm can fit images in various sizes. These methods mainly refer to the book *Digital Image Processing* [3]

#### 3.2 Structure Information

**Table Detection.** In ED Manager, the detection of title blocks in the drawings is treated as a task of table detection, which means locating tables in an image using bounding boxes. We directly apply the X101 model trained on the dataset TableBank [6] to do this task. TableBank is an image-based table detection and recognition dataset, containing 417K high-quality labeled tables. X101 model is trained on this dataset with the object detection framework Detectron2 [14]. Faster R-CNN algorithm with the ResNeXt [15] is used as the backbone network architecture in training. Other training parameters are specified in [6]. Figure 3 shows the performance of the model on table detection.

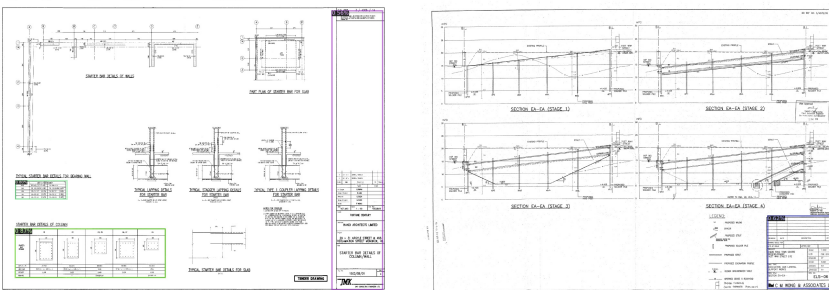


Fig. 3. Results of table detection

**Table Line Extraction.** In ED Manager, three methods based on traditional image processing, including Hough function method, dilating and eroding method, and an innovative dots-connection method, are used to implement table line extraction.

- (1) The Hough transformation method uses the Hough function to transfer a line in the image coordinate  $(x - y)$  into a point in the parametric coordinate  $(\rho - \theta)$ . The process of transformation is

$$y = kx + b \implies \rho = x \cos \theta + y \sin \theta \quad (1)$$

Specifically, the method can be divided into three steps: using the Hough function to transfer the discretized points; discretizing the parametric coordinate and generating a Hough matrix; finding the peak value of the Hough matrix.

The advantages of the Hough transformation method include strong anti-interference ability and high tolerance of gaps. The disadvantages include high time and space complexity and missing length information.

- (2) The dilating and eroding method uses the OpenCV to set a specific kernel to extract the horizontal and vertical lines.

In order to extract the horizontal and vertical lines of images in different sizes, an adaptive algorithms is applied to choose appropriate kernel size. The algorithm can be described as

---

**Algorithm 1.** Adaptive Dilating and Eroding Algorithm for Line Extraction

---

**Input:**  $w$ : image width;  $h$ : image height;  $r$ :  $\frac{w}{h}$

**Output:**  $k_w$ : kernal width;  $K_h$ : kernal height

```

1: initial  $k_w = w/15$ ;  $k_h = h/15$ 
2: if  $r \leq 0.5$  then
3:    $k_w = w/15$ 
4:    $k_h = h \times r/15$ 
5: else if  $r \geq 2$  then
6:    $k_w = w/(15 \times r)$ 
7:    $k_h = h/15$ 
8: else
9:    $k_w = w/15$ 
10:   $k_h = h/15$ 
11: end if

```

---

- (3) The innovative dots-connection method uses dilating and eroding method to extract the dots of the table, and uses the connected component method to delete the misidentified table lines. The method provides a way to eliminate the influence of the characters' size, table's aspect ratio, and complex structure on the table line extraction.

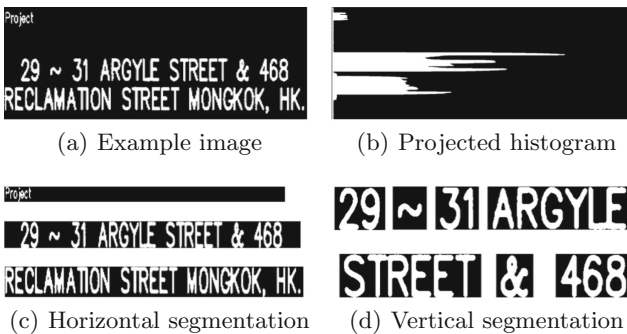
**Cell Extraction.** In ED Manager, the findContours function in OpenCV is used to extract the topological structure of a binary picture, and convert the picture into the border representation. FindContours is a boundary tracking algorithm that scans a binary image by the raster scan method and could distinguish between the outer border and the hole border. With this algorithm, we can

extract the surrounding relationship between the boundaries, which corresponds to the surrounding relationship between the connected components [1].

In ED Manager, the function returns the inflection point information of the profile and saves the information in the contours vector. By finding the upper-left and lower-right corners of these returned contour inflection information, we can find the corresponding areas in the original map and intercept and save these areas, thus achieving cell extraction.

### 3.3 Content Information

**Character Segmentation.** The segmentation process consists of two parts: horizontal segmentation and vertical segmentation. The former divides the cell into lines of words; the latter segments a line into words. Both two parts apply the histogram projection method. This method counts the number of white pixels in each row (or column). If successive lines (or columns) contain a large number of white pixels in the histogram, it means there are characters in this area and those lines (or columns) should be cut out as a whole. Figure 4 shows the results of projected histogram, images after horizontal segmentation, and vertical segmentation.



**Fig. 4.** Results of character segmentation

**Character Recognition.** In ED Manager, to recognize a word instead of a single character, the CRNN with LSTM [4] (long short-term memory) model is used to recognize arbitrary-length words with high accuracy. The principle of the CRNN-LSTM model is based on a feature extraction network and bidirectional LSTM, which can combine long-term memory and short-term memory to avoid the error back-flow problems. Therefore, the CRNN-LSTM model can efficiently recognize arbitrary words by considering the previous and subsequent input.

For ED Manager, we adopt three datasets to train the model. Firstly, an OCR dataset containing English words from Kaggle is implemented. Secondly, we use TRDG (TextRecognitionDataGenerator), a synthetic data generator for text recognition, and an English word list to generate more data. Thirdly, we

use pygame to generate a dataset containing out-of-order characters. By combining the three different datasets, the CRNN-LSTM model can recognize various English words accurately. Figure 5 shows the characters that ED Manager is able to recognize.

```

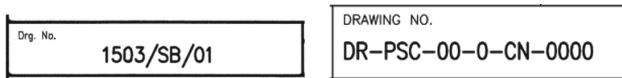
0 1 2 3 4 5 6 7 8 9 a b c
d e f g h i j k l m n o p
q r s t u v w x y z A B C
D E F G H I J K L M N O P
Q R S T U V W X Y Z . ! /
? £ , % - \ € : ( ) + = $
ç ® @ # ^ * & ~ < > [ ]
\ ; ' ° { } “ ” | - © é
J æ ð Æ Ç ' \

```

**Fig. 5.** Supported recognized characters

**Key Information Extraction.** To construct an effective database, we extract five types of key information from the drawing: designer name, company name, project title, drawing date, and drawing number.

The key information is extracted by searching with proper keywords. For example, by searching for the characters ‘No.’, we can directly find the cell with the drawing number, as shown in Fig. 6.



**Fig. 6.** Examples of drawing number

## 4 Improvement

### 4.1 Improvement of the Histogram Projection Method

To achieve a better performance of character segmentation, we improve the histogram projection method in the following aspects:

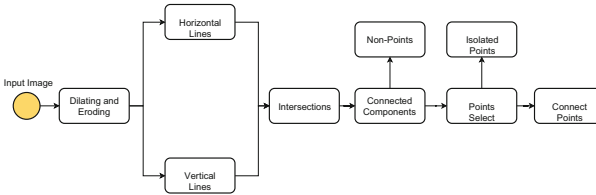
- (1) The top/bottom 10 rows (left/right 10 columns) are not considered in the histogram. This is because the extracted cell often has white table lines around the characters, which should be ignored for the purpose of character recognition. Otherwise, the white lines will be taken as characters and they may yield errors in the recognition process.
- (2) We do horizontal dilation to the characters before vertical segmentation. The dilation operation makes the characters in a word closer to avoid cutting a word wrongly in the middle.



## 4.2 Dots-Connection Method

The dots-connection method combines the dilating and eroding method, connected components algorithm, and the prior knowledge of the table to extract the table line more robust and avoid extracting misidentified table lines, which are caused by some large size characters.

Figure 7 shows the workflow of the dots-connection method.



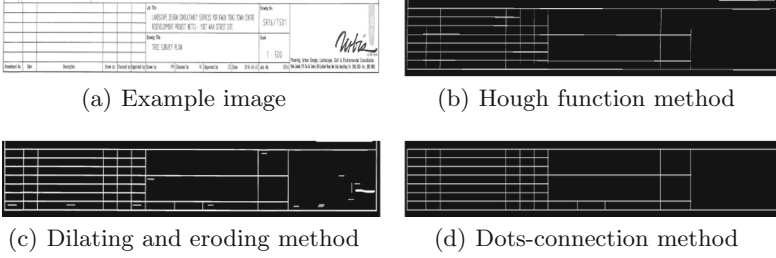
**Fig. 7.** Workflow of dots-connection method

In the workflow, the dots-connection algorithm can be divided into five steps. (1) Use dilating and eroding method to extract horizontal and vertical lines in advance. (2) With a bit-and logical operation obtain the intersections of vertical and horizontal lines. (3) Since the second step contains some misidentified areas which is the intersection of large size characters, we calculate the connected components areas and set a threshold value to delete the large areas. (4) To delete the misidentified corner points, we check whether each point is an isolated point. For a non-isolated point, there is at least one point in the same row and one point in the same column. To delete the isolated point, we set its coordinate equal to the previous point. (5) Connect those selected corner points to extract the table lines.

Step 3 and Step 4 are the major improvements of the dots-connection method. Since the intersections of large-size characters tend to be large size adhesion areas, and the corner points' areas are much smaller, the connected component algorithm can be used to calculate the area of intersections and delete the large parts. Also, with the prior knowledge that at least two points are needed to determine a line of tables, we can delete the misidentified corner points further.

Compared with the Hough function method and dilating and eroding method, our algorithm can get better performance, especially for tables with complex structures, extreme aspect ratios, and large size characters. Also, the time complexity and space complexity of our algorithm is smaller than the Hough function method.

Figure 8 shows different algorithms' performance for table with complex structures, extreme aspect ratios and large size characters. The experiment illustrates that the dots-connection method gets better performance since the former two methods cannot eliminate the misidentified lines caused by large size characters, which will affect the character segmentation and character recognition.



**Fig. 8.** Different algorithms' performance for tables with complex structures, extreme aspect ratio and large size characters

Furthermore, The dots-connection method not only improves the accuracy of extraction; but also guarantees the speed of extraction. Table 1 shows the comparison of the three algorithms on eight different tables (H represents Hough function method, D-E represents dilating dan eroding method, D-C represents dots-connection method). The experient shows that the accuracy of the dots-connection method is greater than the other two methods in nearly all situations. Since the method relies on the dilating and eroding method to obtain the intersections, the speed is slower than it. Yet the method is faster than the Hough function method. The error rate is defined as the ratio of misidentified cells to the number of total cells.

**Table 1.** Results of three methods

Image pixels	Algorithm time (s)			Error rate (%)		
	H	D-E	D-C	H	D-E	D-C
1992 × 2952	0.49	0.045	0.12	37.5	25.0	16.7
1763 × 3893	0.18	0.074	0.18	6.8	2.2	0.0
1904 × 2010	0.13	0.024	0.086	3.6	3.6	3.6
16294 × 742	0.39	0.07	0.25	30.8	15.4	0.0
5441 × 557	0.076	0.016	0.068	17.3	15.4	0.0
2446 × 11498	0.33	0.41	0.81	14.3	0.0	0.0
886 × 4402	0.11	0.027	0.09	31.0	10.3	17.2
1008 × 4500	0.13	0.028	0.097	51.9	18.5	7.4

**Table 2.** Comparison of different software

Image	Total cell	Errors				
		Tencent	iFLYTEK	Baimiao	Youdao AI	ED Manager
1	44	0	8	0	29	0
2	58	1	1	1	34	0
3	29	1	–	1	23	1
4	52	8	31	10	32	0
5	58	1	4	1	39	0
6	160	1	0	0	–	0
7	13	3	–	3	–	1
8	17	6	0	4	9	0
9	36	5	0	5	14	0
10	58	0	0	0	25	1
11	26	2	1	5	6	8
Sum	551	28	45	30	211	11
ER	–	5.1%	8.8%	5.4%	55.8%	1.9%

## 5 Experiements and Analysis

To evaluate the performance of ED Manager in table reconstruction, we compare it with some other OCR software, including Tencent OCR, Baimiao OCR, Youdao AI and iFLYTEK AI.

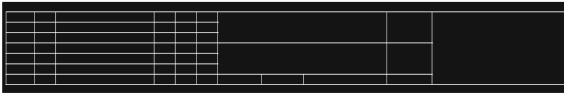
For each of the software mentioned above, we test it with 11 images and count the number of cells which are not reconstructed correctly. The total number of cells of each image is also counted for reference. The results are shown in Table 2. ('–' means the table cannot be reconstructed at all, ER represents the error

rate). As Table 2 illustrated, the error rate of ED Manager is the lowest among all software, which shows the accuracy of our algorithm.

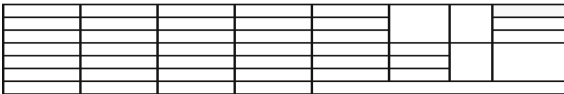
We notice that the aspect ratio of some images is extremely large and some OCR software has great difficulty in reconstructing such images. However, the dots-connection method in ED Manager enables it to extract this kind of table structure information accurately. Figure 9 shows the reconstruction results for one image with extreme aspect ratio. The result shows that compared with other software, our dots-connection method can extract tables with complex structures, extreme aspect ratios and large characters well.



(a) Example image (1754\*159)



(b) Reconstruction result of ED Manager



(c) Reconstruction result of Baimiao (missing 11 cells)

**Fig. 9.** Reconstruction results for images 4

## 6 Conclusion

In this work, we presented the ED Manager, a software based on the fusion of deep learning and traditional image processing, to detect, recognize, and manage the table of engineering drawings automatically with high speed and accuracy.

The ED Manager can finish various tasks including table detection, table structure extraction, table content recognition and table data management. For the table structure extraction, an improved method called dots-connection was presented to extract lines for tables with complex structures, extreme aspect ratios and large characters. And compared with both the existing methods and software, the dots-connection method achieves better performance.

In future work, we plan to enlarge the dataset of character recognition to improve the accuracy of CRNN-LSTM. Also, we intend to improve the dots-connection method to fit more situations such as tables without ruling lines. Furthermore, we hope to develop Android and ios application software, and a website to provide a variety of interfaces.

**Acknowledgements.** This work was partially funded by the National Key Research and Development Plan of China (No. 2018AAA0101000) and the National Natural Science Foundation of China under grant 62076028.

## References

1. Abe, S.K.: Topological structural analysis of digitized binary images by border following. *Graph. Image Process. Comput. Vis.* **30**, 32–46 (1985)
2. Gilani, A., Qasim, S.R., Malik, I., Shafait, F.: Table detection using deep learning. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 771–776. IEEE (2017)
3. Gonzalez, R.C., Woods, R.E.: Digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **28**(4), 484–486 (1980)
4. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
5. Huang, Y., et al.: A yolo-based table detection method. In: 2019 International Conference on Document Analysis and Recognition (ICDAR) (2020)
6. Li, M., Cui, L., Huang, S., Wei, F., Zhou, M., Li, Z.: TableBank: a benchmark dataset for table detection and recognition. *arXiv e-prints* (2019)
7. Li, M., et al.: TrOCR: transformer-based optical character recognition with pre-trained models. *arXiv e-prints* (2021)
8. Nassar, A., Livathinos, N., Lysak, M., Staar, P.: TableFormer: table structure understanding with transformers. *arXiv e-prints* (2022)
9. Qasim, S.R., Mahmood, H., Shafait, F.: Rethinking table recognition using graph neural networks. In: 2019 International Conference on Document Analysis and Recognition (ICDAR) (2020)
10. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. *arXiv e-prints* (2018)
11. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017)
12. Shi, B., Bai, X., Yao, C.: 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**, 2298–2304 (2017)
13. Siddiqui, S.A., Khan, P.I., Dengel, A., Ahmed, S.: Rethinking semantic segmentation for table structure recognition in documents. In: 2019 International Conference on Document Analysis and Recognition (ICDAR) (2019)
14. Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R.: Detectron2 (2019). <https://github.com/facebookresearch/detectron2>
15. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500 (2017)
16. Zhou, J.: Intelligent manufacturing—main direction of “Made in China 2025”. *China Mech. Eng.* **26**(17), 2273 (2015)