IEICE TRANSACTIONS

on Information and Systems

DOI:10.1587/transinf.2024EDL8029

Publicized:2024/05/30

This advance publication article will be replaced by the finalized version after proofreading.

A PUBLICATION OF THE INFORMATION AND SYSTEMS SOCIETY
The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3chome, Minato-ku, TOKYO, 105-0011 JAPAN

LETTER

TDEM: Table data extraction model based on cell segmentation

Zhe Wang†, ***Nonmember***, **Zhe-Ming Lu**†, ***Member,*** **Hao Luo** †, **and Yang-Ming Zheng** †, a), ***Nonmembers***

SUMMARY To accurately extract tabular data, we propose a novel cell-based tabular data extraction model (TDEM). The key of TDEM is to utilize grayscale projection of row separation lines, coupled with table masks and column masks generated by the VGG-19 neural network, to segment each individual cell from the input image of the table. In this way, the text content of the table is extracted from a specific single cell, which greatly improves the accuracy of table recognition.

key words: *Table Structure Recognition, Information Extraction, Text Recognition.*

1. Introduction

Image information has gradually replaced traditional paper document information and become an indispensable part of people's daily lives. Tables in electronic document images are an important form of presenting data and key information. Extracting the content of table areas from electronic document images correctly and completely is a challenging task.

Many previous works that extract table structure are heuristic methods-based systems that analyze PDF drawing commands [1, 2, 3, 4], they don't directly analyze the document image. Recently, deep learning has been proposed to learn table structure directly from images. The DeepDeSRT model [12] uses a neural network originally designed for semantic segmentation of natural scenes and thus primarily uses local information to classify pixels. However, the direct recognition without incorporating semantic information of table text leads to unsatisfactory results. Existing algorithms also fail to preserve the structured representation of the recognized results. To solve these issues, this letter proposes an accurate tabular data extraction model based on cell segmentation (TDEM).

For tables with complete frame lines, TDEM exploits the interdependence between the twin tasks of table detection and table structure recognition combined with pre-trained VGG-19 features to segment out the table and column regions. TDEM uses the grayscale projection of the table region to determine the position of the row separation lines, thereby cropping each cell in the table. In addition, we also produced the lined table dataset ICDAR-2019_line based on the ICDAR-2019 dataset [5]. Most of the existing table recognition models are trained and tested on the ICDAR-2013 dataset [6], and cannot achieve ideal results on the ICDAR-2019 dataset. Aiming at the extraction accuracy of the table content in the image, this letter proposes a new evaluation metric, which is different from the traditional F-measure, Recall value and Precision value.

In summary, the main contributions of this letter are as follows:

1) We propose TDEM: an accurate tabular data extraction model based on cell segmentation, achieving state-of-the-art performance on the ICDAR-2019_line dataset. Furthermore, our approach allows for structured storage of the results, e.g. in Excel format.

2) We also propose a novel evaluation criterion A_e for tabular text content extraction other than the traditional F-measure, Recall value and Precision value.

3) We select wired frame tables from the ICDAR-2019 dataset and the Marmot dataset and annotate the cells and table borders, resulting in the ICDAR-2019_line dataset for table data extraction.

2. Related Work

Before deep learning was applied to table structure recognition, traditional table structure recognition algorithms were mainly based on heuristic rules, that is, specifying a set of rules to make decisions in order to identify tables that meet specific conditions [7, 8, 9, 10]. However, the table recognition method based on heuristic rules is more complicated to design, it is difficult to obtain high accuracy in table recognition in various scenarios, and the robustness is relatively poor.

The table recognition task is often divided into two separate tasks to solve. First, the table detection is performed to locate the table area in the image, and then the structure recognition is performed on the segmented table, and finally the complete table structure information is obtained. A single model is difficult to solve practical problems, and an end-to-end table recognition system [11] is equally important. To overcome the shortcomings of traditional heuristic rule-based table recognition methods that are complex and have low generalization ability, a data-driven end-to-end table recognition system DeepDeSRT was proposed by Schreiber et al. [12] as early as 2017. The system consists of two independent table detection and structure recognition parts. Then in 2019, Tensmeyer et al. [13] proposed the deep learning model SPLERGE for table structure recognition, which consists of two models, split

†The authors are with the School of Aeronautics and Astronautics, Zhejiang University, No.38, Zheda Road, Hangzhou 310027, P. R. China.
a) E-mail: zymsun2002@zju.edu.cn (Corresponding Author)

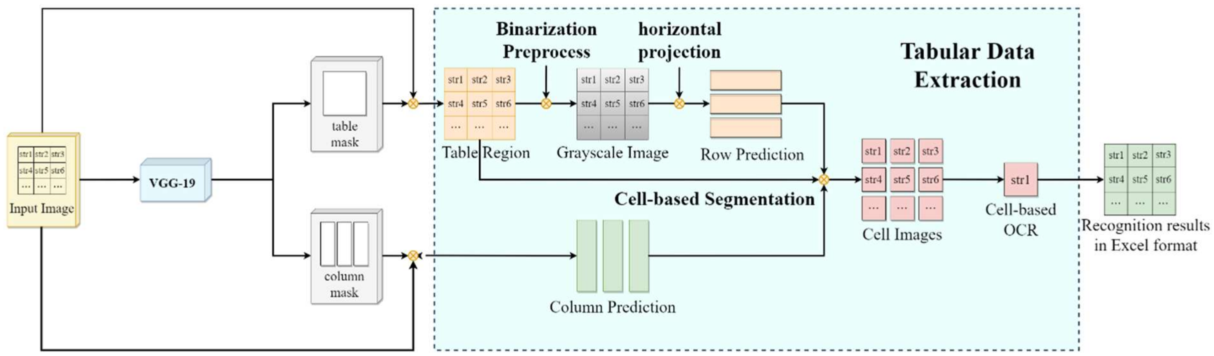

Fig. 1: TDEM takes in a document image and two decoder branches generate separate table predictions and column predictions. The cell images are segmented by grayscale projection.

and merge. In the same year, Paliwal et al. [14] proposed an end-to-end image semantic segmentation model TableNet which uses the interdependence between the two tasks of table detection and table structure recognition to segment table and column areas, further improving the recognition accuracy. Recently, Prasad et al. [15] proposed a deep learning end-to-end convolutional neural network model CascadeTabNet that uses instance segmentation technology to complete table recognition tasks, and obtain state-of-the-art results on the corresponding dataset. Currently, existing table recognition algorithms can only provide the recognition accuracy of the table structure, but they cannot combine the textual content of the table to extract the complete table for users from the input image. Our TDEM method effectively solves this problem by combining accurately recognized table structures with table text content, outputting the complete table contained in the image.

3. Methodology

3.1 Table and column detection module based on VGG-19

As shown in Figure 1, the input image to the model is first converted to RGB image and then resized to 1024×1024 resolution. The fully connected layers of VGG-19 are replaced by two (1×1) convolutional layers, which become two different branches of the decoder network. In each branch, additional layers are appended to filter out the respective active regions. In the table branch of the decoder network, an additional (1×1) convolutional layer is used, followed by a series of fractional stride convolutional layers to upscale the image. Finally, the final feature map is up-scaled to meet the original image size. In the other branch of the detection column, there is an additional convolutional layer with ReLU activation function and a dropout layer with the same dropout probability among the additional convolutional layers. After this layer, the feature maps are upscaled to the original image. Thus, the outputs of the two branches of the computation graph yield masks of table and column regions as shown in Figure 2 (b), (c).

3.2 Cell Data Extraction Based on Grayscale Projection

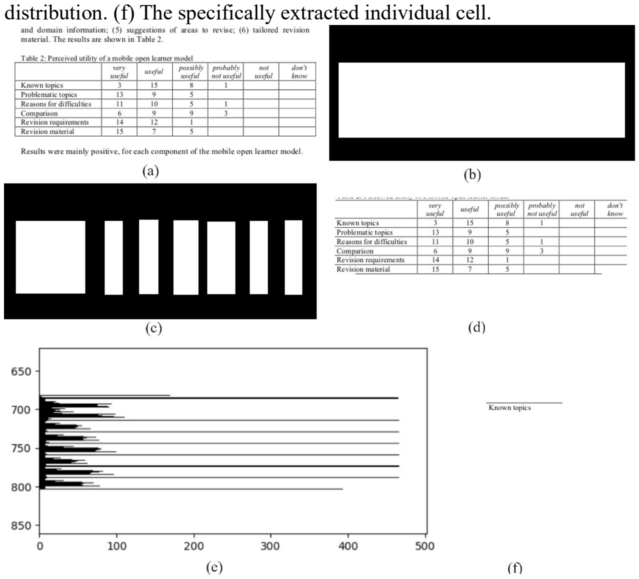

Fig. 2: (a) The raw document image. (b) Generated table mask after VGG-19. (c) Generated column mask after VGG-19. (d) The grayscale image of the extracted table region. (e) The histogram of horizontal projection distribution. (f) The specifically extracted individual cell.

Through the masks generated by the VGG-19 network, we filter out the table regions from the images. We process the images of the table regions by binary preprocessing to obtain grayscale images of the tables as shown in Figure 2 (d). Projection processing is then applied to the grayscale images in the horizontal direction, and the cumulative number of black pixels in each row is calculated to obtain the histogram of horizontal projection distribution of the tables as shown in Figure 2 (e). Since there are blank character gaps in the text content of the cells and the row separator lines are composed of continuous black pixels, it can be observed that the pixel accumulation values of the row separator lines after projection are much larger than those in other areas of the tables. Then the vertical coordinates of the row separator lines can be selected based on a certain threshold. The procedure for determining the vertical coordinates of the row separator lines is as follows:

1) Let N denote the number of rows of the table region image, for $1 \leq i \leq N$, select all the i's that satisfy $A(i) > minHor$, and store them in array $H[y]$. The threshold $minHor$ can be determined by $\max(A(i)) \times p$, where p is a hyperparameter and we set it to 0.7 in this work.

2) Further filter $H[y]$ based on the threshold $lineHor$. If the difference between several coordinates is less than

lineHor, then select the median value as the final vertical coordinate of the row separator line, and store the final vertical coordinate in the array *finlH*[*y*]. The threshold *lineHor* can be interpreted as the maximum thickness of the row separator lines, and we set it to 4. The details of the table row separator locator are summarized in Algorithm 1.

Algorithm 1: Table Row Separator Locator.

Data: input image p, the binarized image of the table area p_B

Result: the array of vertical coordinates for row separator lines $finlH[y]$

1 $p \rightarrow p_B = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1M} \\ u_{21} & u_{22} & \cdots & u_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ u_{N1} & u_{N2} & \cdots & u_{NM} \end{bmatrix}, u_{ij} = \begin{cases} 1 & \text{black} \\ 0 & \text{white} \end{cases}, 1 \le i \le N, 1 \le j \le M;$

2 Horizontal projection pixel accumulation value:

$$A(i) = \sum_{j=1}^{M} u_{ij}, 1 \le i \le N;$$

3 For $1 \le i \le N$,

if $A(i) > minHor \ (= 0.7\max(A(i)))$,

$i \rightarrow H[y]$;

4 if $|y_1 - y_2| < lineHor(lineHor = 4).y_1, y_2 \in H[y]$,

$$y = \frac{y_1 + y_2}{2},$$

$y \rightarrow finlH[y]$.

After obtaining the accurate vertical coordinate values of the row separator lines, combined with the table column regions filtered out by the VGG-19 column region masks, we can accurately crop out each cell in the table as shown in Figure 2 (f). By performing OCR text recognition on specific cells, the accuracy of table content recognition in TDEM has been significantly improved.

4. Experiments

4.1 Dataset and Criterion

To effectively test the model's ability to extract complex lined table data, we created an ICDAR-2019_line dataset containing 3000 images and 4739 tables by merging the two datasets of ICDAR-2019 (TRACT B) and Marmot. Annotations to table columns and rows are missing in the Marmot dataset. By adding labels to the bounding boxes around each column and row of the tabular region, we manually annotate the dataset for table structure identification. The ICDAR-2019_line dataset contains not only English tables but also Chinese tables. Most of these Chinese tables are in the form of financial statements and ownership structures. We randomly split the ICDAR-2019_line dataset into 2.4k/0.3k/0.3k for training, validation, and testing. Tables in the ICDAR-2019_line dataset exhibit greater complexity, visual appearance, and structure compared to the ICDAR-2013 dataset.

Table data extraction is based on table detection and structure recognition, with an additional step of cell text recognition. To achieve optimal extraction results, higher demands should be placed on the accuracy of cell text recognition. Traditional evaluation metrics such as F-measure, Recall, and Precision are not comprehensive as they only reflect the accuracy of models in table detection and structural recognition. In order to more intuitively demonstrate the effectiveness of data extraction, we additionally incorporate a comparison of the recognition results of table text content and the ground truth, introducing a new evaluation metric: Accuracy of Table Data Extraction (A_e), as shown in formula (1). In the formula, C represents the number of cells whose text content and structural position are correctly identified in the table, T represents the total number of cells in the table, and A_e is the final accuracy of tabular data extraction.

$$A_e = \frac{C}{T} \times 100\% \quad (1)$$

4.2 Results of table structure recognition algorithm

This section presents the experimental results of all models on the ICDAR-2019_line dataset. Model performance is evaluated based on traditional metrics including F-measure, recall and precision values. These measures are computed for each image and averaged over all document images.

As TDEM is a tabular data extraction algorithm composed of a combination of deep learning preprocessing methods and traditional projection post-processing steps, we not only conduct experiments to compare it with table structure recognition algorithms based on deep learning methods [12-15], but also include comparisons with traditional methods [3, 4]. As shown in Table 1, our approach achieves an F-measure value of 96.64% on the ICDAR-2019_line dataset, surpassing not only the F-measure value of 79.45% obtained by the traditional method TEXUS [4] but also outperforming the state-of-the-art deep learning method, the CascadeTabNet [15] model, with an F-measure value of 94.92%. Through comparative analysis, it is evident that deep learning methods generally outperform traditional methods in terms of table structure recognition accuracy. Hence, TDEM utilizes the VGG-19 neural network as a preprocessing network to accomplish table region detection and column region recognition tasks.

Table 1: F-measure, Recall value and Precision value of models on ICDAR-2019_line table dataset.

Model	F-measure	Recall	Precision
J. Li et al	0.7748	0.7842	0.7953
TEXUS	0.7945	0.7823	0.8071
TableNet	0.8864	0.8995	0.9235
DeepDeSRT	0.8911	0.8628	0.9193
SPLERGE	0.9353	0.9434	0.9527
CascadeTabNet	0.9492	0.9549	0.9435
TDEM	**0.9664**	**0.9728**	**0.9835**

4.3 A_e of table structure recognition algorithm

Table 2 further reveals the regular patterns of the A_e values assessed for all models on the ICDAR-2019_line dataset. As depicted, the TDEM model, based on cell instance segmentation, distinctly outshines the comparative methods. Through the fusion of semantic information from table text and accurate cell segmentation leveraging row separator lines projection, TDEM achieves the highest accuracy in tabular data extraction among all models, achieving an A_e value of 89.26%. Throughout our experimentation, we observed that all table recognition algorithms perform well in processing tables like the one depicted in Figure 3 (b), which features clear boundary lines, uniform inter-row spacing, and straightforward textual content. However, when confronted with tables like that shown in Figure 3 (a), which features lack of column separator lines, irregular inter-row spacing, and intricate textual content, traditional table recognition methods, along with most deep learning methods, are ineffective. This accentuates the importance of semantic information within table text and validates that recognition methods based on cell segmentation can significantly outperform alternative recognition approaches.

Table 2: Experimental results with the metric A_e for extraction accuracy analysis.

Model	A_e
J. Li et al	54.69%
TEXUS	60.86%
DeepDeSRT	44.69%
TableNet	52.48%
CascadeTabNet	67.40%
SPLERGE	75.57%
TDEM	89.26%

(a) (b)

Table 4: Comparisons of the enhanced SPLERGE and TSR-Former on different datasets.

Methods	Dataset	Prec. (%)	Rec. (%)	F1. (%)	TEDS-Struct (%)
SPLERGE	SciTSR	99.3	98.9	99.1	-
TSRFormer	SciTSR	**99.5**	**99.4**	**99.4**	-
SPLERGE	SciTSR-COMP	98.8	98.0	98.4	-
TSRFormer	SciTSR-COMP	**99.1**	**98.7**	**98.9**	-
SPLERGE	PubTabNet	-	-	-	97.1
TSRFormer	PubTabNet	-	-	-	**97.5**
SPLERGE	In-house	85.4	82.3	83.8	-
TSRFormer	In-house	**95.1**	**95.3**	**95.2**	-

(a)

which can only be accomplished in 25μs if the processor has a clock rate of 1.7 GHz. Based on our actual measurement of 100.21μs, the controller is only able to attain an average IPC rate of 1.4.

Instructions	128,423
First-level instruction-cache misses	340
First-level data-cache misses	1750
Second-level (unified) cache misses	2
TLB misses	0
Branch mispredictions	3

(b)

Fig. 3: Various table images inducing alterations in A_e.

5. Conclusion

In this letter, we present an accurate tabular data extraction model based on cell segmentation (TDEM), aiming at the data extraction of lined table from document images. TDEM is the first model to extract tabular data from the cell level and save the result in Excel format. On the ICDAR-2019_line dataset, TDEM has achieved a far higher accuracy of tabular data extraction than other models. In the future, we plan to predict the appearance position of the row-column separation lines after obtaining the table area, and then divide the cell area, to further improve the tabular data extraction result of the complex table by the TDEM model.

Acknowledgements

This work was partially supported by Ningbo Science and Technology Innovation 2025 major project under grants 2020Z106 and 2023Z040.

References

[1] Liu J , Ding X and Wu Y, "Description and recognition of form and automated form data entry" 1995 Proceedings of the Third International Conference on Document Analysis and Recognition, doi: 10.1109 /ICDAR.1995.601963.

[2] Shigarov A, Mikhailov A and Altaev A, "Configurable Table Structure Recognition in Untagged PDF documents" 2016 Acm Symposium on Document Engineering, pp.119-122, 2016.

[3] Li J, Wang K, Hao S Q and Wang Q R, "Location and Recognition of Free Tables in Form" 2012 International conference on software engineering, knowledge engineering and information engineering, 2012.

[4] Rastan R, Paik H Y and Shepherd J, "TEXUS: A Task-based Approach for Table Extraction and Understanding" 2015 Proceedings of the 2015 ACM Symposium on Document Engineering, pp.25-34, 2015.

[5] L. Gao, Y. L. Huang, H. Déjean, J. L. Meunier, Q. Q. Yan, Y. Fang, F. Kleber, and E. Lang, "ICDAR 2019 competition on table detection and recognition (cTDaR)," 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, pp.1510-1515, 2019.

[6] M. Göbel, T. Hassan, E. Oro and G. Orsi, "ICDAR 2013 table competition," 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, pp.1449-1453, 2013.

[7] A. Laurentini and P. Viada, "Identifying and understanding tabular material in compound documents," Proc. 11th IAPR International Conference on Pattern Recognition. Vol.II. Conference B: Pattern Recognition Methodology and Systems, The Hague, Netherlands, pp.405-409, 1992.

[8] S. Chandran and R. Kasturi, "Structural recognition of tabulated data," Proc. 2nd International Conference on Document Analysis and Recognition (ICDAR '93), Tsukuba, Japan, pp.516-519, 1993.

[9] C. Akinlar and C. Topal, "Edlines: real-time line segment detection by Edge Drawing (ed)," Proc. 18th IEEE International Conference on Image Processing, Brussels, Belgium, pp.2837-2840, 2011.

[10] E. Koci, M. Thiele, O. Romero and W. Lehner, "A genetic-based search for adaptive table recognition in spreadsheets," 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, pp.1274-1279, 2019.

[11] E Silva, A. C., Jorge and Torgo, "Design of an end-to-end method to extract information from tables" 2006 International Journal of Document Analysis and Recognition, pp.144-171, 2006.

[12] S. Schreiber, S. Agne, I. Wolf, A. Dengel and S. Ahmed, "DeepDeSRT: deep learning for detection and structure recognition of tables in document images," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, pp.1162-1167, 2017.

[13] C. Tensmeyer, V. I. Morariu, B. Price, S. Cohen and T. Martinez, "Deep splitting and merging for table structure decomposition," 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, pp.114-121, 2019.

[14] S. S. Paliwal, V. D, R. Rahul, M. Sharma and L. Vig, "TableNet: deep learning model for end-to-end table detection and tabular data extraction from scanned document images," 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, pp.128-133, 2019.

[15] D. Prasad, A. Gadpal, K. Kapadni, M. Visave and K. Sultanpure, "CascadeTabNet: an approach for end to end table detection and structure recognition from image-based documents," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, pp.2439-2447, 2020.