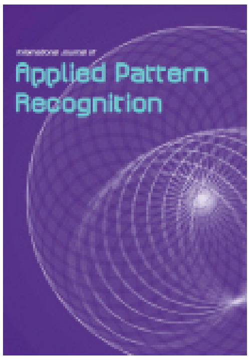
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Applied Pattern Recognition

ISSN online: 2049-8888 - ISSN print: 2049-887X
https://www.inderscience.com/ijapr

High-precision recognition method for sparsely arranged financial tables

Xiaofeng Shi, Yahui Ji, Menghui Liu

DOI: 10.1504/IJAPR.2026.10078484

Article History:

Received:	15 August 2025
Last revised:	24 March 2026
Accepted:	16 April 2026
Published online:	10 August 2026

Int. J. Applied Pattern Recognition, Vol. 8, No. 4, 2026

High-precision recognition method for sparsely arranged financial tables

Xiaofeng Shi*

School of Computer and Information Engineering,
Fuyang Normal University,
Fuyang Anhui, 236037, China
and
School of Computer and Information Engineering,
Hefei University of Technology,
Hefei Anhui, 230601, China
Email: shixiaofeng@fynu.edu.cn
*Corresponding author

Yahui Ji

School of Computer and Information Engineering,
Fuyang Normal University,
Fuyang Anhui, 236037, China
Email: yahui_ji@126.com

Menghui Liu

School of Computer and Information Engineering,
Hefei University of Technology,
Hefei Anhui, 230601, China
Email: 3287960253@qq.com

Abstract: Table recognition and parsing is an important task in the field of information extraction. Especially in financial printing sparse layout forms, the cell spacing within the form is obvious, and the arrangement structure is relatively loose, the form frame line is complex, and the multi-line text within the cell brings a big challenge to form recognition. In order to solve this problem, an end-to-end efficient recognition algorithm for financial sparse layout forms is proposed, which realises form structure recognition and content filling by processing text semantics and structure, using dynamic difference detection algorithm, overlap-driven algorithm, and differential method. A large number of experiments have been done on 13 bank flow forms, and the results show that the mean overall accuracy is as high as 99.85%, which meets the requirements of financial-grade accuracy, and especially demonstrates strong robustness in complex borderline scenarios. The research aims to provide new theoretical support and technical reference for the field of form recognition.

Keywords: table analysis; sparse layout; end-to-end analysis; high precision.

Reference to this paper should be made as follows: Shi, X., Ji, Y. and Liu, M. (2026) 'High-precision recognition method for sparsely arranged financial tables', *Int. J. Applied Pattern Recognition*, Vol. 8, No. 4, pp.1–23.

Biographical notes: Xiaofeng Shi received her PhD from Xiamen University in 2014. She joined Fuyang Normal University in 2014 and currenlty an Associate Professor. Her research interests include optoelectronic integrated chips, silicon-based OEIC, and mixed-signal IC design for optical communications and detection. She has led Anhui provincial projects, published over 20 papers, and filed multiple patents.

Yahui Ji obtained his Bachelor of Science degree from the College of Information Engineering, Fuyang Normal University in 2023. Since then, he has been pursuing a Master's degree in Artificial Intelligence at Fuyang Normal University, Anhui Province, China.

Menghui Liu obtained his Bachelor of Engineering from Fuyang Normal University in 2019. Since then, he has been pursuing a Master's in Computer Technology at Fuyang Normal University, Anhui Province, China.

1 Introduction

Financial forms are widely used to convey structured, high-accuracy information. Table recognition aims to extract structured information from images/documents into editable spreadsheets for efficient data management, but it faces key challenges: diverse table styles (wired, unwired, lineless, multi-line cell text), high annotation costs (especially for complex semantically rich tables), limited multi-domain datasets, and demands for model efficiency/memory on resource-constrained devices.

To address these challenges, methods primarily rely on vision-based approaches using convolutional neural networks (CNNs) or vision transformers (ViTs) to process table images and extract features. Models like CascadeTabNet (Prasad et al., 2020) and TableNet (Paliwal et al., 2019) segment table grids to determine structure, performing well on bordered tables but struggling with borderless or ambiguous ones. Graph neural networks model tables as graphs (nodes = cells, edges = semantic relationships), capturing row-column relationships and handling gridless tables (Zhou et al., 2020; Liu et al., 2021; Li et al., 2021; Dong et al., 2019). However, they face challenges with complex graph construction for multi-line cells and poor generalisation. These methods fail to meet financial industry accuracy standards, highlighting the need for more efficient, accurate algorithms tailored to financial tables.

A promising alternative leverages textual semantics and structural information to avoid interference from table borders and background lines, thereby enhancing robustness. For quantitative definition, sparse layout is evaluated by measuring the horizontal distance between adjacent text instances in the image. Specifically, a layout is regarded as sparse if the horizontal spacing between adjacent texts is greater than the width of one character; it is considered dense if the distance is less than the width of one character.

This paper proposes an end-to-end sparse layout-based form recognition algorithm, whose core reconstructs the logical structure of forms by parsing text block organisation. Furthermore, a chunking strategy is employed for content parsing, whereby the image is segmented into individual cell patches along row and column boundaries inferred from the parsed table structure, and text recognition is performed on each patch independently. Experimental results on a large-scale dataset comprising bank statement forms from 13 institutions demonstrate that the proposed method achieves an accuracy of 99.85%. In addition, the approach is lightweight, requiring no high-performance computing resources, and satisfies the stringent requirements of the financial industry.

2 Related research

Traditional methods for table recognition often employ edge detection algorithms, such as Canny edge detection or the Hough transform (Zhao et al., 2022), to extract table lines and generate grid structures based on intersections and lines. However, these approaches perform poorly when dealing with nonstandard tables that lack or have minimal borders. With the rapid advancement of deep learning technologies (Hashmi et al., 2021), numerous innovative methods have emerged. Image segmentation methods: the TableNet algorithm treats table detection as an image segmentation task (Paliwal et al., 2019), utilising deep CNN to segment rows, columns, and cell regions. This approach is simple and efficient for regular tables with clear borders, but performs suboptimally with complex backgrounds and spanning cells. Object detection algorithms: algorithms such as CascadeTabNet (Prasad et al., 2020), DeepDeSRT (Schreiber et al., 2017), and YOLO-based models (Zhang et al., 2023; Huang et al., 2019) approach the detection of table cells, rows, or columns as object detection problems, employing modified CNN models to extract bounding boxes for each region. These methods support the handling of spanning and borderless tables; however, they are less effective when processing complex cell content, such as multiline text, and entail high annotation costs and significant hardware resource requirements. End-to-end parsing algorithms: the TabStruct-Net (Kashinath et al., 2022; Wang et al., 2023) algorithm directly generates structured table data from images, eliminating intermediate steps and simplifying the table parsing process. However, the generated table structures are susceptible to noise from the table content, resulting in reduced flexibility. Multimodal learning methods: the LayoutLM algorithm combines visual and textual features (Xu et al., 2020), utilising multimodal models to jointly process table images and content. This approach is suitable for complex borderless tables and handwritten tables. However, it requires simultaneous annotation of text and layout information, relies heavily on visual features, and exhibits suboptimal real-time performance. Both traditional methods and deep learning methods are triggered by the features of the image itself to analyse the overall structure of the form.

In recent years, methodological advancements in financial table recognition have focused on complex document structures. Khandokar and Deshpande (2025) proposed an end-to-end computer vision framework that decomposes the task into detection, structure recognition, and content extraction modules. By leveraging synthetic data and augmentation techniques, their approach mitigates data scarcity issues and achieves 85%

extraction accuracy on invoice datasets. Peng and Li (2025) developed an ante hoc enhancement method targeting tables with multi-line text and missing demarcation lines. Their approach integrates OCR-based text localisation, heuristic candidate generation, and semantic matching for line refinement, improving existing models' F1 scores by 5.10–14.36 percentage points.

Recent advances in document layout analysis focus on complex sparse layouts. Wu et al. (2025) proposed Light-HookNet, a lightweight cross-attention framework that integrates global and local features to distinguish visually similar layout elements in historical documents points. Singh et al. (2025) developed a Devanagari character encoded mix-merge vision transformer (MViT), combining character-level encoding with spatial relationship modelling to achieve 97.4% segmentation accuracypoints. These studies highlight the trend toward semantic and structural cues for sparse document analysis.

In this paper, we propose a high-precision end-to-end recognition method for sparse layout tables, which determines the structure and content of the table through the text block and its distribution relationship, the advantage of the algorithm is that it is triggered from the text content of the image, not affected by the complex table line and table background, and it is very applicable to blank cells, complex cell content, the algorithm is low in complexity and high in versatility, and it has been effectively verified in the financial field.

3 Recognition algorithms for sparsely arranged financial tables

The algorithms in this paper achieve the following goals: stable recognition of sparse layout of various types of box and line forms, including wireframe, less wireframe, and no wireframe; effective handling of complex text styles within the cell, such as blank text, single-line text, multi-line text, etc.; financial level accuracy guarantee to meet the stringent requirements of financial business on data accuracy.

3.1 Algorithm details

Inspired by the layout fusion paradigm of LayoutLM (Xu et al., 2020), the proposed algorithm adopts a pure text-block parsing strategy and discards visual features to reduce complexity. Furthermore, to address the poor performance of CascadeTabNet (Prasad et al., 2020) in handling multi-row cells, a block-based content parsing strategy is introduced, which eliminates the requirement for bounding box annotations. The core idea of the algorithm is to analyse the text layout, construct the logical table structure, and populate the content. The main steps are as follows:

Step 1 adaptive scaling

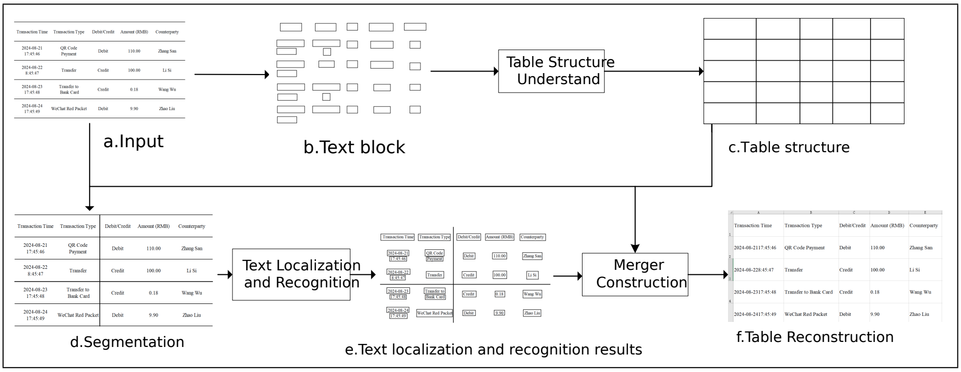
Step 2 text block detection

Step 3 table structure parsing via local peak detection

Step 4 block detection

Step 5 structure merging and table reconstruction (as illustrated in Figure 1).

Figure 1 Algorithm flowchart



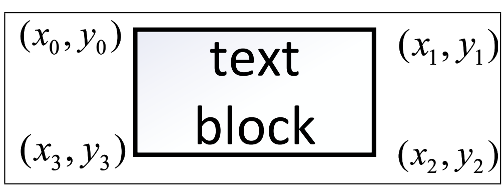
3.1.1 Adaptive scale scaling

The input table image is scaled to match the optimal input size required by subsequent text block detection. The scaling factor is determined according to the original image resolution and average text block size, ensuring scaled text blocks fit within the receptive field of the detection network. For low-resolution images, bilinear interpolation is used for upscaling to reduce pixel distortion, whereas high-resolution images are processed by downsampling.

3.1.2 Text block detection

All text blocks in the image are detected, and their four-corner coordinates are extracted to obtain the bounding box (four-point coordinates). Text blocks are scanned in left-to-right and top-to-bottom order. The coordinates are illustrated in Figure 2.

Figure 2 Text block coordinate definition diagram



3.1.3 Local peak detection algorithm

As shown in Figure 2, the column left coordinate of each text block is defined as the mean of its left boundary coordinates x_0 and x_3. These values form a list, which is denoised using the dynamic local difference method. We then compute the first-order difference of the list, traverse the sequence, and identify the positions of maximum positive change within local regions. The corresponding coordinates constitute the column left coordinate list.

Similarly, the column right coordinate of each text block is defined as the mean of its right boundary coordinates x_1 and x_2. These values form another list, which is also denoised via the dynamic local difference method. After computing the first-order difference, we traverse the sequence to locate the positions of maximum negative change in local regions, yielding the column right coordinate list.

Finally, we iterate over the two lists. The column coordinate of the $(n+1)^{\text{th}}$ column is calculated as the mean of the n-th column right coordinate and the $(n+1)^{\text{th}}$ column left coordinate.The first column coordinate is set to 0, and the last column coordinate equals the width of the reconstructed table.

3.1.4 Parsing the table structure

Once the column coordinates are determined, the row coordinates can be calculated directly. Financial transaction documents typically contain fixed fields such as transaction time and amount. Considering the fields common to all 13 banks, this paper selects the columns containing transaction time and amount as reference columns (used to establish a fixed reference for table rows). Reference columns can be efficiently identified through simple matching strategies, including time format matching, amount format matching, or text matching. Next, we calculate the y_0 coordinates of the text blocks in the reference column. The y_0 coordinate of the nth text block is defined as the coordinate of the nth row, and the coordinate of the last row is set to the height of the input table. Finally, by combining the row and column coordinates, we obtain the complete table structure, as shown in module c in Figure 1.

3.1.5 Image chunking detection

Text block detection can obtain coordinates and content, but cannot guarantee a high recognition accuracy. Therefore, segmentation (dividing the image into appropriate regions) is required to ensure that the text fits within the receptive field of the text recognition network. The segmentation process follows the rule of ‘no cell division’. Based on the table structure obtained in the previous section, the image is segmented along row and column lines, as shown in module d in Figure 1.

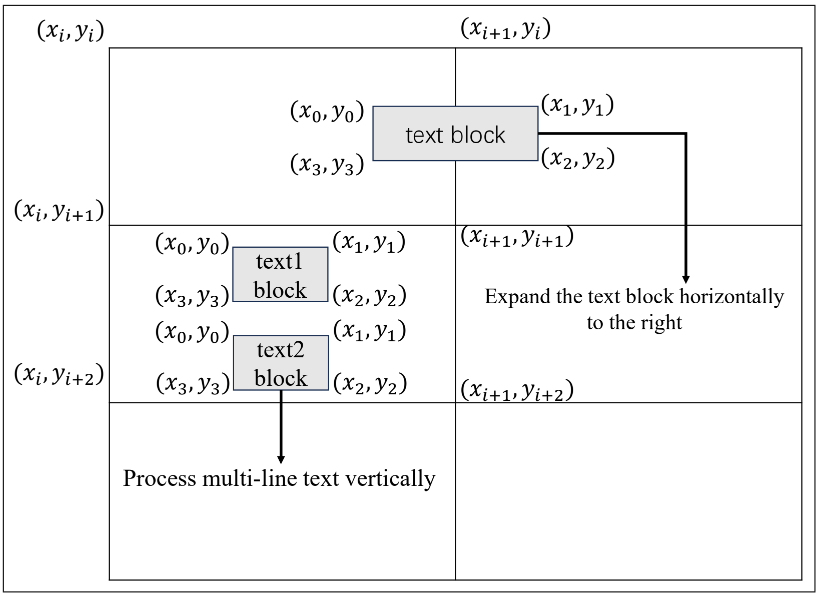
3.1.6 Reconstruction of the recognised text content into the table structure

Using the table structure from the previous section, we determine each cell’s bounding box. An overlap-driven cell attribution algorithm verifies cell validity: for each text block, we compute its overlap with cell boundaries to assign it to the corresponding cell. Blocks overlapping multiple cells are treated as merged cells. After processing all text blocks, we reconstruct the table and output results in JSON or Excel format.

3.2 Overlap-driven cell attribution algorithm

Overlap degree-driven cell attribution schematic shown in Figure 3, the cell boundaries expressed as $(x_i, y_i, x_{i+1}, y_{i+1})$, then the overlap region of the coordinates of the boundary for $x_{\text{left}} = \max(x_0, x_i)$, $y_{\text{top}} = \max(y_0, y_i)$, $x_{\text{right}} = \min(x_2, x_{i+1})$, $y_{\text{bottom}} = \min(y_2, y_{i+1})$, if the overlap region in the cell bounding box, then determine the text block in the current cell. If x_2 is greater than x_{i+1} or y_2 is greater than y_{i+1}, according to the overlap of the boundary value, to the right to expand horizontally as shown in Figure 3 in the text block, downward longitudinal processing of multiple lines of text as shown in Figure 3 in the text1 block and text2 block, to continue the above overlap degree detection to determine the cell attribution of all text blocks. The algorithm has a time complexity of $O(n \cdot m)$ and a space complexity of $O(m + n)$, where n is the number of text blocks and m is the number of cells.

Figure 3 Schematic illustration of the overlap-driven cell attribution algorithm



Notes: Text block ownership is determined via boundary coordinate intersection, supporting both horizontal expansion and vertical multi-line processing.

3.3 Dynamic local difference method

Use dynamic local difference method to remove noise points for the list, the algorithm idea: sort the input list, calculate the absolute difference of the median of the list, multiply it by a scale factor scale, get a dynamic threshold, then traverse all the data points, carry out the calculation of the local difference, when the forward difference diff_p and backward difference diff_n of the data points at the same time exceed the dynamic threshold, then it will be judged to be an anomalous data point. The dynamic threshold calculation formula is shown in equation (1) and the local difference is shown in equation (2).

$$\text{MAD} = \text{median}\left(|x_i - \text{median}(x)|\right) \times \text{scale} \tag{1}$$

where x_i represents an individual data point; median denotes the median function; scale is a scaling factor, set to 0.1.

$$\begin{aligned} \text{diff}_p(i) &= |x_i - x_{i-1}| \\ \text{diff}_n(i) &= |x_i - x_{i+1}| \end{aligned} \tag{2}$$

where

- $\text{diff}_p(i)$ denotes the absolute difference between the current point x_i and the preceding point x_{i-1} (forward difference).
- $\text{diff}_n(i)$ denotes the absolute difference between the current point x_i and the succeeding point x_{i+1} (backward difference).
- the subscript p stands for *previous* (forward), and the subscript q stands for *next/queue* (backward), serving to distinguish the direction of the difference.
- $|\cdot|$ represents the absolute value operation.

Note: For notation consistency throughout the manuscript, the i^{th} data point is uniformly denoted as x_i in both equations and algorithm descriptions, where x_i represents the i^{th} data value in the dataset.

Algorithm 1 Dynamic local difference algorithm

Require: data (list of values), scale (optional control factor, default = 0.1)
Ensure: cleaned_data (processed data list with outliers replaced by 0)
sorted_data $\leftarrow$ **sorted**(data)
threshold $\leftarrow$ **calculate_threshold**(sort_data, scale)
for $i \leftarrow 1$ to $\text{len}(data) - 1$ **do**
  $\text{diff}_p(i) \leftarrow$ |data[i] – data[$i-1$]|
  $\text{diff}_n(i) \leftarrow$ |data[i] – data[$i+1$]|
  if $\text{diff}_p(i) >$ threshold **and** $\text{diff}_n(i) >$ threshold **then**
    cleaned_data[i] $\leftarrow 0$
  end if
end for
return cleaned_data

The proposed method leverages the median absolute deviation (MAD; see Table 3) to adaptively determine the threshold, eliminating the need for manual tuning. This enhances robustness against outliers and ensures higher computational efficiency. The algorithm operates with a time complexity of $\text{O}(n \log n)$ due to the sorting step and requires $\text{O}(n)$ space complexity, making it well-suited for large-scale data processing. By dynamically adjusting the threshold based on MAD, the method effectively suppresses noise while preserving essential data structures, ensuring reliable performance across varying datasets.

3.4 Improved text block detection

This study focuses on table structure analysis, with an emphasis on enhancing the performance of text block detection. To better demonstrate the advantages of the proposed algorithm, the open-source PaddlePaddle optical character recognition (PaddleOCR; see Table 3) model is employed. Through targeted improvements, the model's accuracy in detecting text blocks is refined, ensuring the feasibility of subsequent processing steps.

3.4.1 Adjusting detection model parameters

To optimise text block detection, several key model parameters are adjusted. The unclip ratio ($unclip\_ratio$; see Table 3) is reduced, as this parameter controls the expansion ratio of the detected bounding boxes. A high $unclip\_ratio$ may cause adjacent text blocks to merge incorrectly. Additionally, the bounding box confidence threshold (box_thresh; see Table 3) parameter, which defines the confidence threshold for bounding boxes, is slightly increased. A low box_thresh value may retain excessive unnecessary bounding boxes, leading to incorrect merges and reduced detection accuracy.

3.4.2 Improved post-processing strategy

Enhancements to the post-processing strategy to refine text block merging. The non-maximum suppression (NMS; see Table 3) threshold is lowered to better handle overlapping bounding boxes. If the intersection over union threshold (iou_thresh; see Table 3) is set too high, an excessive number of bounding boxes may be merged, negatively impacting detection performance. Furthermore, connectivity analysis is improved to ensure more precise bounding box merging. Before merging, the distance and angle between text blocks are evaluated to determine whether merging is necessary. Geometric filtering is applied to prevent unreasonable merges, and a distance threshold is set to regulate merging decisions. For instance, if the distance between the centre points of two text blocks exceeds a predefined threshold, they are not merged. These refinements contribute to more accurate text block detection, ultimately improving the reliability of table structure analysis.

4 Application research and experimental evaluation

4.1 Application research experiments

In order to advance the practical application of the algorithm, application research experiments were done on the algorithm, and the experimental procedure is as follows:

- Step 1 The PDF document through the pre-model or algorithm to get the form image.
- Step 2 The algorithm in this paper to deal with the form image.
- Step 3 The output of generalised form data, such as JSON, excel tables and other documents.

The main source of data from 13 banks (WeChat Pay, Alipay, Industrial and Commercial Bank of China, Agricultural Bank of China, Bank of China, China Construction Bank, Bank of Communications, Postal Shanghai Pudong Development Bank, Ping An Bank, China Everbright Bank, Hua Xia Bank), including electronic documents in PDF format or scanned documents, form types include wireframe, wireless frame, less wireframe, cell types include blank text, single line text, multi-line text, the table layout is mostly sparse layout, there is a clear spacing between the cell text, and the actual form screenshots are shown in Figure 4.

Figure 4 Screenshot of actual table, (a) WeChat Pay (b) Bank of China (c) Agricultural Bank of China

Serial No.	Transaction Date	Currency	Transaction Amount	Account Balance
1	2023-10-25	CNY	5000.00	14210.97
2	2023-10-22	CNY	5000.00	9210.97
3	2023-10-22	CNY	-4022.85	4210.97
4	2023-10-22	CNY	-300.00	8233.82
5	2023-10-22	CNY	6100.00	8533.82
6	2023-10-22	CNY	-7.50	2433.82

(a)

Transaction Date	Transaction Time	Transaction Description	Amount	Balance After Transaction
20230601	175607	UnionPay credit-in	+600.00	601.53
20230601	175622	Alipay debit	–584.29	17.24
20230601	175857	Third-party agency payment	+69.33	86.57
20230602	153815	Douyin Pay debit	–20.75	65.82
20230602	154047	Third-party agency payment	+85.02	150.84

(b)

Transaction Date	Transaction Counterparty	Transaction Type	Debit (Dr) / Credit (Cr)	Amount (CNY)	Balance After Transaction (CNY)
2023-12-03	Tenpay Payment Technology Co., Ltd.	Online payment	Dr	700.00	4008.08
2023-12-03	Tenpay Payment Technology Co., Ltd.	Refund	Cr	700.00	4708.08
2023-12-04	Meituan Designated Merchant	Other transaction	Dr	2.50	4705.58
2023-12-07	Meituan Designated Merchant	Other transaction	Dr	9.90	4695.68
2023-12-07	Meituan Designated Merchant	Other transaction	Cr	9.90	4705.58
2023-12-09	Other (credit card repayment)	Credit card transfer repayment	Dr	4453.98	251.60
2023-12-10	Payment platform	Transfer from demand-deposit wealth management product	Cr	302.40	554.00
2023-12-10	Meituan Designated Merchant	Other transaction	Dr	554.00	0.00

(c)

4.2 *Objective evaluation*

4.2.1 *Data sources*

To ensure statistical reliability, we employed a large-scale dataset consisting of 1,000 pages. Such a sample size is sufficient to minimise random errors and guarantee that our results are robust and representative. To verify the practical application performance of the proposed algorithm, we note that among the 13 banks investigated, only the WeChat platform supports the generation of both sheet forms and electronic PDF bills. Therefore, WeChat electronic bills are adopted for objective evaluation, while bills from other banks are used for professional assessment. The experimental dataset comprises 1,000 pages of PDF bill images, including 800 pages of wireframe forms and 200 pages of manually annotated non-wireframe forms. The electronic form bills directly exported from the platform with complete structural information serve as the benchmark for algorithm performance comparison.

To further improve the rigor and reproducibility of the experimental results, statistical analyses including repeated trials and mean value computation were performed on the 1,000-page dataset. Concretely, the recognition experiment was implemented five independent repetitions to mitigate random errors, based on which the mean overall accuracy of the proposed algorithm was derived. The experimental results demonstrate that the model yields a mean overall accuracy of 99.12%, with a 95% confidence interval of [99%, 99.5%], thereby validating the statistical reliability and stability of the proposed method.

4.2.2 Assessment methodology and indicators

Step 1 Use the PDF bill as input and parse it through the algorithm to generate a spreadsheet.

Step 2 Export the spreadsheet file through the platform.

Step 3 Traverse the cell structure and content of the algorithm-generated spreadsheet and the platform-exported spreadsheet to calculate the accuracy of form recognition.

The evaluation indexes mainly include the structure accuracy rate, which is used to evaluate the recognition accuracy of the algorithm on the cell structure, and the content accuracy rate, which is used to evaluate the accuracy of text recognition. The average value of the structure accuracy rate the content accuracy rate is taken to comprehensively evaluate the overall recognition effect of the algorithm.

4.2.3 Experimental results

The general form recognition model introduced by Baidu (Li et al., 2022) and the Huawei (Wang et al., 2023) form model are selected as comparison algorithms, and the experimental results of the comparison algorithms come from the online experience platform. Because the online experience platform cannot directly convert PDF documents, the PDF to image for the experiment, the output results of the integration, and benchmark data from the structure and content of the comparison. The experimental results are shown in Table 1.

Table 1 Comparison of experimental results

Algorithms	*Structural acc (%)*	*Content acc (%)*	*Overall acc (%)*
Baidu	61.20	99.00	80.10
Huawei	89.00	99.30	94.15
Ours (proposed)	99.90	99.80	99.85

Notes: Overall accuracy is defined as the weighted average of structural accuracy and content accuracy. Specifically, it is computed as the arithmetic mean of the two metrics, i.e., (structural accuracy + content accuracy) / 2.

The results show that all selected algorithms demonstrate high accuracy in text content recognition, and their performance is largely comparable when processing tables with borders and fully bordered tables. In such standard scenarios, the structural cues provided by complete borders enable traditional models to achieve satisfactory results, as shown in Figures 6 and 7. However, as shown in Figure 5, significant performance differences emerge among the algorithms when dealing with unbordered tables. The primary discrepancies lie in the handling of tables without structural lines, the erroneous segmentation of multi-line text cells into multiple independent cells, and inaccuracies in parsing multi-line text. Although all models perform comparably in bordered environments, their limitations become glaringly apparent in less structured layouts such as unbordered tables.

In contrast, the algorithm proposed in this paper reconstructs table structures by analysing the spatial and semantic relationships between text blocks, without relying on visual cues such as borders or lines. This method performs exceptionally well on borderless tables, while also achieving performance comparable to other methods on tables with borders and fully bordered tables, it maintains a high structural recognition accuracy even in scenarios where other models begin to fail. Regarding the processing of multi-line text, the algorithm effectively utilises reference columns to guide cell segmentation, achieving optimal results particularly in blank or text-sparse areas. Furthermore, the adoption of a chunking strategy mitigates interference caused by variations in input scale during content recognition, thereby significantly improving the overall accuracy of text recognition.

Figure 5 Comparison of recognition results for borderless tables, (a) original image (b) Baidu method (c) Huawei method (d) this paper method (see online version for colours)

Transaction Time	Transaction Type	Income/Expenditure	Amount (CNY)	Counterparty
2124-08-21 17:45:46	QR code payment	Expenditure	110.00	Zhang San
2124-08-22 8:45:47	Transfer	Income	100.00	
2124-08-23 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
2124-08-24 17:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu
2124-08-25 17:45:46	QR code payment	Expenditure	110.00	Zhang San
2124-08-26 8:45:47	Transfer	Income	100.00	
2124-08-27 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
2124-08-28 17:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu

(a)

transaction type income expenditure transaction time 2124-08-21 scan QR code 17:45

transaction counterparty amount yuan expenditure Zhang San 110.00

transfer income 100.00 8:45:47 2124-08-22

transfer to bank income Wang Wu 0.18 2124-08-23 17:45:48 bank card

expenditure Zhao Liu WeChat red packet 17:45:49 scan QR code Zhang San

expenditure 110.00 2124-08-25 17:45:46 payment

transfer income 100.00 2124-08-26

8:45:47

transfer to bank income 0.18 Wang Wu 2124-08-27

bank card

17:45:48

WeChat red packet expenditure 9.90 Zhao Liu 17:45:49 2124-08-28

(b)

Transaction Time	Transaction Type QR code scan	Income/Expenditu re Expenditure	Amount (CNY) 110.00	Counterparty Zhang San
2124-08-21 17:45:46 2124-08-22	payment Transfer	Income	100.00	
8:45:47 2124-08-23 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
2124-08-24 17:45:49	WeChat red packet	Expenditure	9.90	Zhao Liu
2124-08-25 17:45:46	QR code scan payment	Expenditure	110.00	Zhang San
2124-08-26 8:45:47	Transfer	Income	100.00	
2124-08-27 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
2124-08-28 17:45:49	WeChat red packet	Expenditure	9.90	Zhao Liu

(c)

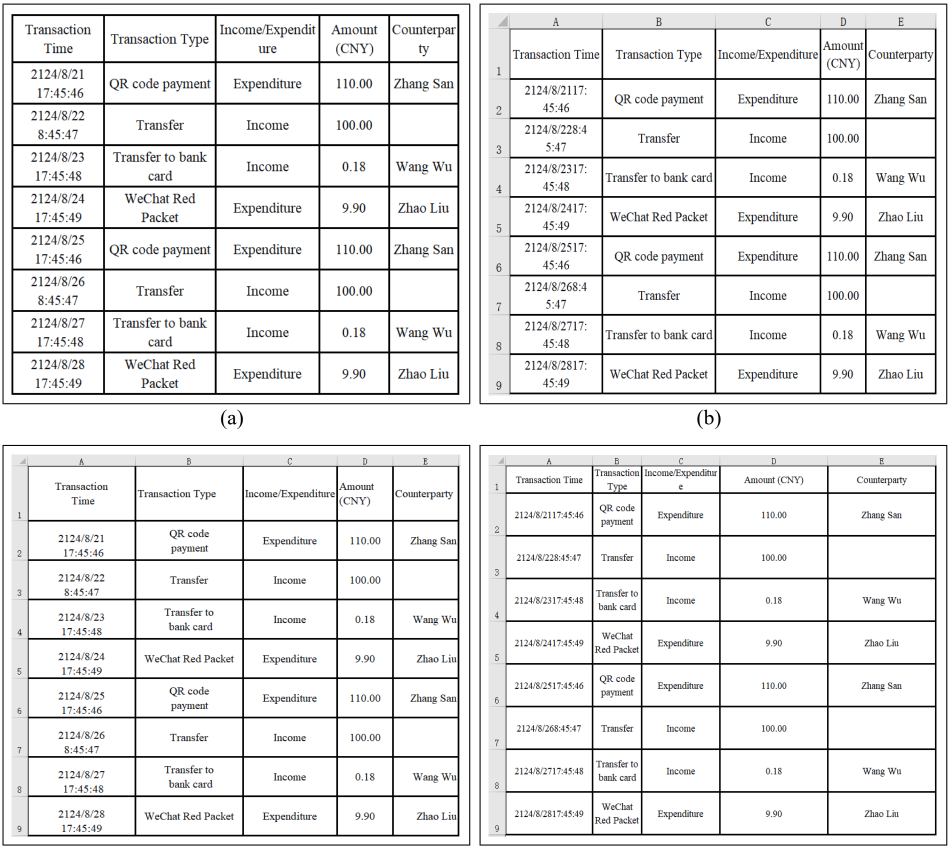
A	B	C	D	E
Transaction Time	Transaction Type	Income/Exp enditure	Amount (CNY)	Counterparty
2124-08-2117:45:46	QR code payment	Expenditure	110.00	Zhang San
2124-08-228:45:47	Transfer	Income	100.00	
2124-08-2317:45:48	Transfer to bank card	Income	0.18	Wang Wu
2124-08-2417:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu
2124-08-2517:45:46	QR code payment	Expenditure	110.00	Zhang San
2124-08-268:45:47	Transfer	Income	100.00	
2124-08-2717:45:48	Transfer to bank card	Income	0.18	Wang Wu
2124-08-2817:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu

(d)

4.2.4 *Efficiency analysis*

To rigorously validate the lightweight design of our proposed algorithm, we conducted a quantitative comparison of inference time and memory consumption against models from Baidu and Huawei. The results are summarised in Table 2.

Figure 6 Comparison of recognition results for full-border tables, (a) original image (b) Baidu method (c) Huawei method (d) this paper method



Transaction Time	Transaction Type	Income/Expenditure	Amount (CNY)	Counterparty
2124/8/21 17:45:46	QR code payment	Expenditure	110.00	Zhang San
2124/8/22 8:45:47	Transfer	Income	100.00	
2124/8/23 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
2124/8/24 17:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu
2124/8/25 17:45:46	QR code payment	Expenditure	110.00	Zhang San
2124/8/26 8:45:47	Transfer	Income	100.00	
2124/8/27 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
2124/8/28 17:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu

(a)

	A	B	C	D	E
1	Transaction Time	Transaction Type	Income/Expenditure	Amount (CNY)	Counterparty
2	2124/8/2117:45:46	QR code payment	Expenditure	110.00	Zhang San
3	2124/8/228:45:47	Transfer	Income	100.00	
4	2124/8/2317:45:48	Transfer to bank card	Income	0.18	Wang Wu
5	2124/8/2417:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu
6	2124/8/2517:45:46	QR code payment	Expenditure	110.00	Zhang San
7	2124/8/268:45:47	Transfer	Income	100.00	
8	2124/8/2717:45:48	Transfer to bank card	Income	0.18	Wang Wu
9	2124/8/2817:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu

(b)

	A	B	C	D	E
1	Transaction Time	Transaction Type	Income/Expenditure	Amount (CNY)	Counterparty
2	2124/8/21 17:45:46	QR code payment	Expenditure	110.00	Zhang San
3	2124/8/22 8:45:47	Transfer	Income	100.00	
4	2124/8/23 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
5	2124/8/24 17:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu
6	2124/8/25 17:45:46	QR code payment	Expenditure	110.00	Zhang San
7	2124/8/26 8:45:47	Transfer	Income	100.00	
8	2124/8/27 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
9	2124/8/28 17:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu

(c)

	A	B	C	D	E
1	Transaction Time	Transaction Type	Income/Expenditure	Amount (CNY)	Counterparty
2	2124/8/2117:45:46	QR code payment	Expenditure	110.00	Zhang San
3	2124/8/228:45:47	Transfer	Income	100.00	
4	2124/8/2317:45:48	Transfer to bank card	Income	0.18	Wang Wu
5	2124/8/2417:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu
6	2124/8/2517:45:46	QR code payment	Expenditure	110.00	Zhang San
7	2124/8/268:45:47	Transfer	Income	100.00	
8	2124/8/2717:45:48	Transfer to bank card	Income	0.18	Wang Wu
9	2124/8/2817:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu

(d)

All experiments were conducted on a laptop equipped with an Intel Core i7-9750H processor, 8 GB of RAM, and an NVIDIA GeForce GTX 1650 graphics card with 4 GB of dedicated video memory, running the Windows 10 Professional operating system. The input image resolution was fixed at 1920 × 1080 pixels. Furthermore, an inverse correlation was observed between inference speed and batch size, where a larger batch size led to degraded processing speed, whereas a smaller batch size contributed to higher throughput.

Table 2 Comparison of computational efficiency

Algorithms	*Reasoning time (s)*	*Memory usage (MB)*
Baidu	9.3	223.6
Huawei	7.8	336.4
Ours (proposed)	*3.5*	*254.6*

Notes: The proposed algorithm achieves the best trade-off between reasoning speed and memory consumption.

Figure 7 Comparison of recognition results for half-bordered tables, (a) original image (b) Baidu method (c) Huawei method (d) this paper method

Transaction Time	Transaction Type	Income/Expenditure	Amount (CNY)	Counterparty
2124/8/21 17:45:46	QR code payment	Expenditure	110.00	Zhang San
2124/8/22 8:45:47	Transfer	Income	100.00	
2124/8/23 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
2124/8/24 17:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu
2124/8/25 17:45:46	QR code payment	Expenditure	110.00	Zhang San
2124/8/26 8:45:47	Transfer	Income	100.00	
2124/8/27 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
2124/8/28 17:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu

(a)

	A	B	C	D	E
1	Transaction Time	Transaction Type	Income/Expenditure	Amount (CNY)	Counterparty
2	2124/8/2117: 45:46	QR code payment	Expenditure	110.00	Zhang San
3	2124/8/228:4 5:47	Transfer	Income	100.00	
4	2124/8/2317: 45:48	Transfer to bank card	Income	0.18	Wang Wu
5	2124/8/2417: 45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu
6	2124/8/2517: 45:46	QR code payment	Expenditure	110.00	Zhang San
7	2124/8/268:4 5:47	Transfer	Income	100.00	
8	2124/8/2717: 45:48	Transfer to bank card	Income	0.18	Wang Wu
9	2124/8/2817: 45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu

(b)

	A	B	C	D	E
1	Transaction Time	Transaction Type	Income/Expenditure	Amount (CNY)	Counterparty
2	2124/8/21 17:45:46	QR code payment	Expenditure	110.00	Zhang San
3	2124/8/22 8:45:47	Transfer	Income	100.00	
4	2124/8/23 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
5	2124/8/24 17:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu
6	2124/8/25 17:45:46	QR code payment	Expenditure	110.00	Zhang San
7	2124/8/26 8:45:47	Transfer	Income	100.00	
8	2124/8/27 17:45:48	Transfer to bank card	Income	0.18	Wang Wu
9	2124/8/28 17:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu

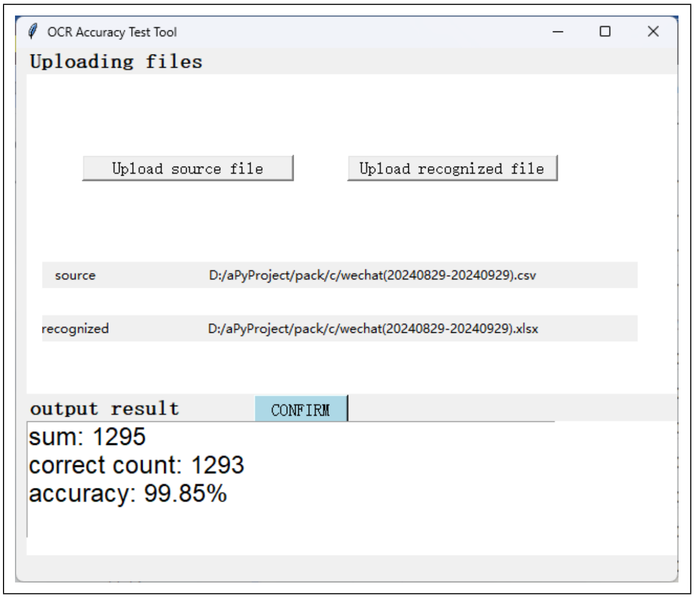
(c)

	A	B	C	D	E
1	Transaction Time	Transaction Type	Income/Expenditure	Amount (CNY)	Counterparty
2	2124/8/2117:45:46	QR code payment	Expenditure	110.00	Zhang San
3	2124/8/228:45:47	Transfer	Income	100.00	
4	2124/8/2317:45:48	Transfer to bank card	Income	0.18	Wang Wu
5	2124/8/2417:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu
6	2124/8/2517:45:46	QR code payment	Expenditure	110.00	Zhang San
7	2124/8/268:45:47	Transfer	Income	100.00	
8	2124/8/2717:45:48	Transfer to bank card	Income	0.18	Wang Wu
9	2124/8/2817:45:49	WeChat Red Packet	Expenditure	9.90	Zhao Liu

(d)

As shown in the table, compared to the two baseline models, our algorithm is significantly faster in single-table inference. Specifically, its inference latency is reduced by more than 50% compared to the Baidu and Huawei models, indicating that our algorithm has a clear advantage in computational efficiency. This not only aligns with the lightweight design goal but also makes it highly suitable for real-time or near-real-time applications. In terms of memory usage, our algorithm maintains a moderate level of resource consumption: while slightly higher than the Baidu model, memory usage is significantly lower compared to the Huawei model. This result indicates that our method effectively controls memory overhead while maintaining high inference speed, achieving a more balanced trade-off between computational efficiency and resource utilisation. Overall, the proposed algorithm exhibits clear advantages in inference speed and competitive memory usage, which strongly corroborates the lightweight design claim stated in the introduction. It achieves a superior balance between speed and resource consumption, making it a promising candidate for deployment in resource-constrained environments.

Figure 8 User interface of the transaction statement conversion software (see online version for colours)



4.2.5 Real-world deployment

To validate the industrial applicability and deployability of the algorithm described in this paper, it was integrated into the batch statement conversion software independently developed by the authors. Deployment testing was conducted at a regional commercial bank, with the primary objective of addressing pain points such as low processing efficiency, insufficient accuracy, and non-standardised data in bank statement processing, thereby verifying the algorithm's practicality and stability in real-world financial scenarios. Figure 8 shows the core user interface of the transaction statement conversion software. It allows users to customise the upload path for statement files and the save path for results. The algorithm is triggered by clicking the 'confirm' button, and the progress bar at the bottom of the interface provides real-time feedback on processing progress, intuitively demonstrating the software's lightweight and user-friendly design. The software operates according to the algorithm logic described in this paper. It takes PDF statements as input and performs automatic parsing, recognition, and output through text block structure analysis and block-based recognition. Manual verification is initiated in special cases to ensure seamless integration with the bank's internal systems. Deployment results indicate that, after integration, the algorithm remains highly efficient and stable across multiple real-world financial scenarios, aligning with the design goals of lightweight and high-efficiency performance and meeting the core requirements of financial institutions.

4.3 *Analysis of non-financial sparse matrix data*

To verify the general applicability of the algorithm described in this paper beyond financial transaction reports, we selected non-financial sparse-layout tables – specifically, administrative and academic forms. The experimental results are shown in the Figure 9. Preliminary findings indicate that the algorithm is not only applicable to financial transaction reports but also demonstrates strong general applicability to sparse-layout tables in administrative, educational, and other scenarios. In the future, we plan to collect more types of non-financial tabular data (such as tables from research papers and medical forms) for systematic evaluation to further validate the algorithm's general applicability.

Figure 9 Visualisation results of non-financial forms: raw inputs and structured outputs, (a) administrative form entry (b) academic form entry (c) administrative form output (d) academic form output

Annual Work Items of Admin Dept	Status	Responsible Person	Remarks
Spring team building	Completed	Zhang San	
Office equipment procurement		Li Si	Pending approval
Employee physical examination	In progress		Expected completion in May
Factory landscaping			Outsourced to third party
Safety training	Completed	Wang Wu	Full staff coverage

(a)

Group	Temperature (°C)	Pressure (MPa)	Yield (%)
Group A	25	1	
Group B		1	
Group C	25		91
Group D		1.5	93.4

(b)

	A	B	C	D
1	Annual Work Items of Admin Dept	Status	Responsible Person	Remarks
2	Spring team building	Completed	Zhang San	
3	Office equipment procurement		Li Si	Pending approval
4	Employee physical examination	In progress		Expected completion in May
5	Factory landscaping			Outsourced to third party
6	Safety training	Completed	Wang Wu	Full staff coverage

(c)

	A	B	C	D
1	Group	Temperature (°C)	Pressure (MPa)	Yield (%)
2	Group A	25	1	
3	Group B		1	
4	Group C	25		91
5	Group D		1.5	93.4

(d)

4.4 *Professional evaluation*

In order to comprehensively verify the practical application effect of the proposed algorithm, we realistically evaluated the recognition effect and practical application value of the algorithm in the form recognition task by means of qualitative evaluation and feedback from experts.

Figure 10 Visualisation results of bank statement transaction records in different table formats, (a) table without borders (b) half-bordered table (c) half-bordered table (d) fully bordered table (e) output

Transaction Time	Transaction Type	Income/Expenditure	Amount (RMB)	Counterparty
2024-08-21 17:45:46	QR Code Payment	Expenditure	110.00	Zhang San
2024-08-22 8:45:47	Transfer	Income	100.00	
2024-08-23 17:45:48	Transfer to Bank Card	Income	0.18	Wang Wu
2024-08-24 17:45:49	WeChat Red Packet	Expenditure	9.90	

(a)

Transaction Time	Transaction Type	Income/Expenditure	Amount (RMB)	Counterparty
2024-08-21 17:45:46	QR Code Payment	Expenditure	110.00	Zhang San
2024-08-22 8:45:47	Transfer	Income	100.00	
2024-08-23 17:45:48	Transfer to Bank Card	Income	0.18	Wang Wu
2024-08-24 17:45:49	WeChat Red Packet	Expenditure	9.90	

(b)

Transaction Time	Transaction Type	Income/Expenditure	Amount (RMB)	Counterparty
2024-08-21 17:45:46	QR Code Payment	Expenditure	110.00	Zhang San
2024-08-22 8:45:47	Transfer	Income	100.00	
2024-08-23 17:45:48	Transfer to Bank Card	Income	0.18	Wang Wu
2024-08-24 17:45:49	WeChat Red Packet	Expenditure	9.90	

(c)

Transaction Time	Transaction Type	Income/Expenditure	Amount (RMB)	Counterparty
2024-08-21 17:45:46	QR Code Payment	Expenditure	110.00	Zhang San
2024-08-22 8:45:47	Transfer	Income	100.00	
2024-08-23 17:45:48	Transfer to Bank Card	Income	0.18	Wang Wu
2024-08-24 17:45:49	WeChat Red Packet	Expenditure	9.90	

(d)

	A	B	C	D	E
1	Transaction Time	Transaction Type	Income/Expenditure	Amount (RMB)	Counterparty
2	2024-08-2117:45:46	QR Code Payment	Expenditure	110.00	Zhang San
3	2024-08-228:45:47	Transfer	Income	100.00	
4	2024-08-2317:45:48	Transfer to Bank Card	Income	0.18	Wang Wu
5	2024-08-2417:45:49	WeChat Red Packet	Expenditure	9.90	

(e)

Since the method proposed in this paper adopts an end-to-end integrated pipeline, its core modules are highly coupled and mutually dependent. Particularly in the attribution module driven by dynamic local differences and overlapping cues, the removal or replacement of any key component may disrupt the overall inference logic and model structure, thereby leading to inaccurate table structure recognition. Consequently, the traditional ablation study based on module removal is difficult to be directly applied to the proposed method.

Figure 11 PDF image of bank transaction statements

Transaction Date	Transaction Direction	Transaction Amount	Counterparty Information	Transaction Memo
2023-06-01	UnionPay Credit	600.00	[redacted]Outbound Transition Account	
2023-06-01	Alipay	584.29	215500690	NA20[redacted]0525090211109 Zhang XX
2023-06-01	Agency Payment	69.33	Tenpay Payment Technology Co., Ltd.	NG202306[redacted]38S0200701 WeChat Balance Withdrawal
2023-06-02	Douyin Pay	20.75	Wuhan XX XX Technology Co., Ltd.	NA202[redacted]10000 Douyin Monthly Pay – Repayment
2023-06-02	Agency Payment	85.02		
2023-06-02	Zhejiang Vipshop	150.00	[redacted]E-commerce Co., Ltd.	UA060232385[redacted]Commerce Co., Ltd. - [redacted]Technology Co., Ltd.
2023-06-03	UnionPay Credit	15.00	[redacted]Outbound Transition Account	
2023-06-03	WeChat Pay	15.00	243300133	NA[redacted]308437274210310004 Beijing XX Trading Co., Ltd.
2023-06-10	UnionPay Credit	50.00		WY_TX[redacted]530301
2023-06-10	WeChat Pay	50.00	243300133	NA202306[redacted]0733310211407 WeChat Transfer
2023-06-11	Agency Payment	15.00	Wuhan[redacted]Co., Ltd.	NG20230[redacted]388684031000[redacted]
2023-06-14	Agency Payment	54.76	Tenpay Payment Technology Co., Ltd.	NG2023[redacted]64637S0300006 WeChat Balance Withdrawal
2023-06-14	Transfer Payment	70.00	XX Wealth (Shanghai) Fund Sales Co., Ltd.	E-commerce Payment
2023-06-15	UnionPay Credit	205.79		WeChat Balance Withdrawal
2023-06-15	Agency Payment	45.74	Alipay (China) Network Technology Co., Ltd.	NG20230[redacted]2509020130232 654
2023-06-15	Transfer Payment	47.07	XX Wealth (Shanghai) Fund Sales Co., Ltd.	E-commerce Payment

We invited researchers from agricultural and commercial banks as evaluation experts, who have been engaged in billing data management and analysis in the financial field for a long time and have rich practical experience and professional knowledge. Through their evaluation of the algorithm recognition results, they are able to provide objective and authoritative feedback from the practical needs of the application. The evaluation dataset consists of bill form data from the entire banking system, covering the current bills of 13 banks, and the form types are consistent with those in the application research experiments.

We evaluated the proposed algorithm on real-world billing data with the aid of a professional internal calibration system provided by the bank. The evaluation was performed from two key perspectives: table structure recovery and textual content accuracy. Combined with professional feedback from the expert team and validation on large-scale bank statement datasets, the proposed algorithm achieves an overall comprehensive accuracy of over 99%. This performance fully meets the stringent high-precision requirements for automated bill processing in practical banking scenarios. In addition, the document recognition and conversion software developed based on the proposed algorithm has been granted computer software copyright registration (Registration No. 2025SR0563817) by the National Copyright Administration of China.

4.5 Visualisation validation experiment

In order to present the research results more intuitively, a visual validation experiment was done with all possible table types, as shown in Figure 10.

Figures 10(a), 10(b), 10(c) and 10(d) are input form images, and Figure 10(e) is the output Excel. From the results, it is obvious that the algorithm in this paper performs well in handling various boxed tables and multi-line texts.

The visualisation of the actual bank statement image is shown in Figure 12.

Figure 12 Visualisation of structural reconstruction for bank transaction statements

Transaction Date	Transaction Direction	Transaction Amount	Counterparty Information	Transaction Memo
2023-06-01	UnionPay Credit	600.00	[redacted] Outbound Transition Account	
2023-06-01	Alipay	584.29	215500690	NA20[redacted]0525090211109 Zhang XX
2023-06-01	Agency Payment	69.33	Tenpay Payment Technology Co., Ltd.	NG202306[redacted]38S0200701 WeChat Balance Withdrawal
2023-06-02	Douyin Pay	20.75	Wuhan XX XX Technology Co., Ltd.	NA202[redacted]10000 Douyin Monthly Pay – Repayment
2023-06-02	Agency Payment	85.02		
2023-06-02	Zhejiang Vipshop	150.00	[redacted] E-commerce Co., Ltd.	UA060232385[redacted]Commerce Co., Ltd. - [redacted]Technology Co., Ltd.
2023-06-03	UnionPay Credit	15.00	[redacted] Outbound Transition Account	
2023-06-03	WeChat Pay	15.00	243300133	NA[redacted]308437274210310004 Beijing XX Trading Co., Ltd.
2023-06-10	UnionPay Credit	50.00		WY_TX[redacted]530301
2023-06-10	WeChat Pay	50.00	243300133	NA202306[redacted]0733310211407 WeChat Transfer
2023-06-11	Agency Payment	15.00	Wuhan [redacted] Co., Ltd.	NG20230[redacted]388684031000[redacted]
2023-06-14	Agency Payment	54.76	Tenpay Payment Technology Co., Ltd.	NG2023[redacted]64637S0300006 WeChat Balance Withdrawal
2023-06-14	Transfer Payment	70.00	XX Wealth (Shanghai) Fund Sales Co., Ltd.	E-commerce Payment
2023-06-15	UnionPay Credit	205.79		WeChat Balance Withdrawal
2023-06-15	Agency Payment	45.74	Alipay (China) Network Technology Co., Ltd.	NG20230[redacted]2509020130232 654
2023-06-15	Transfer Payment	47.07	XX Wealth (Shanghai) Fund Sales Co., Ltd.	E-commerce Payment

Figure 13 Reconstruction results of complex handwritten nested tables, (a) borderless handwritten-style table (b) output of a borderless handwritten-style table (c) half-bordered handwritten-style table (d) output of a table with half-bordered handwritten font

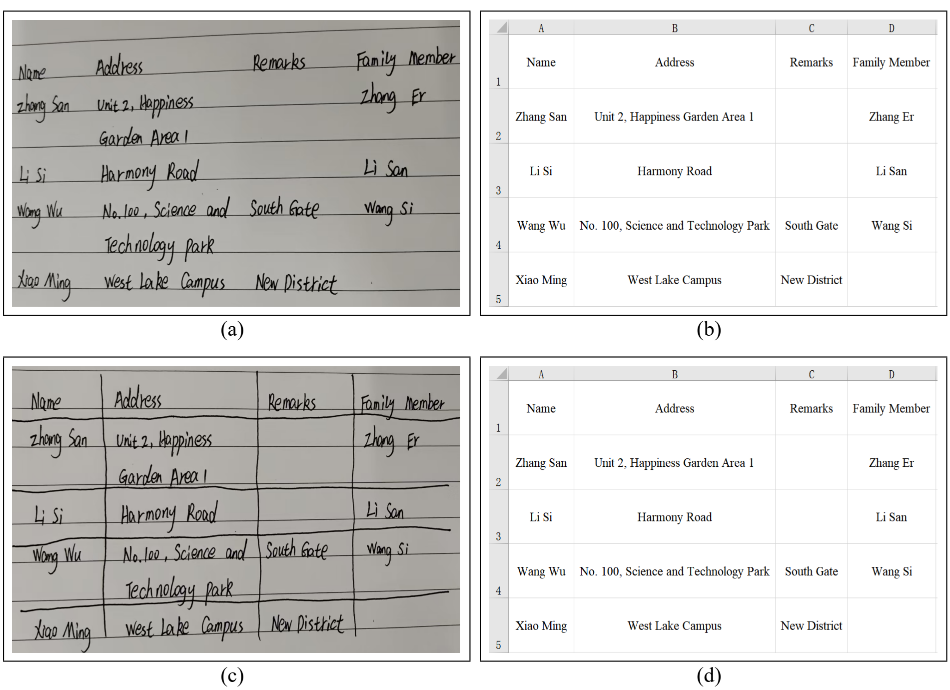


Figure 11 shows an image of a page from an actual bill, where sensitive information is processed in order to protect user privacy. As seen in the output result Figure 12, for this kind of high-resolution, sparsely-layouted forms, the algorithm in this paper shows excellent performance in terms of recognition accuracy, which effectively verifies its feasibility and practical value in real-world application scenarios.

Table 3 Definition of key terms and parameters

Term	*Full name*	*Definition*
PaddleOCR	PaddlePaddle optical character recognition	PaddleOCR is an open-source, free, high-performance text recognition toolkit built on Baidu's PaddlePaddle deep learning framework, dedicated to the accurate extraction of text from images.
NMS	Non-maximum suppression	Removes duplicate detection boxes to retain only the most accurate results.
MAD	Mean absolute deviation	Measures the magnitude of model prediction errors.
$unclip\_ratio$	Unclip ratio	The ratio used to expand the boundaries of text detection boxes.
box_thresh	Bounding box confidence threshold	A confidence threshold to filter out low-confidence detection boxes.
iou_thresh	Intersection over union threshold	A threshold for judging box overlap, applied for box deduplication.

4.6 Algorithm performance analysis

The limitation of this algorithm mainly lies in the text block localisation. The structural analysis function of the algorithm will fail when the spacing of the text blocks is too small resulting in the inability to accurately divide them. The algorithm in this paper has a high accuracy rate when processing 13 bank statements, the reason is that bank statements are all printed images with relatively sparse table layout, thus having a high accuracy rate in both text recognition and table structure. It should be noted that the recognition accuracy of the algorithm in this paper is not exactly equivalent to the recognition accuracy of the text recognition network model itself, as the datasets used are different and the evaluation criteria are also different.

In order to further verify the algorithm's ability to generalise the style of the form box lines, this paper tested the more complex box line structure. The results are shown in Figure 13, even under the interference of diverse box lines, the recognition effect is still stable, indicating that the algorithm has strong robustness and adaptability.

5 Results

We propose an end-to-end algorithm for financial sparse-layout tables, which reconstructs table structure by analysing both semantics and layout of textual content. The algorithm is robust to complex frame lines, blank cells, and diverse cell contents, achieving high accuracy in financial document processing. It features low complexity and easy integration with other systems. However, performance degrades on denser, more intricate table layouts, highlighting a need for further improvement. Future research will focus on two directions to enhance generalisation: replacing fixed morphological kernels with dynamic kernels that adjust to local text density to prevent over-merging in dense regions, modelling text clustering as a graph with spacing similarity as edge weights and applying graph cuts for semantic segmentation in dense areas.

Acknowledgements

We sincerely thank Professor Xiaofeng Shi and graduate student Menghui Liu for their constructive support and insightful discussions. We are also grateful to YingDong Rural Commercial Bank for providing valuable resources and assistance during the preparation of this paper.

Authors' contribution

Author contributions statement: conceptualisation, Xiaofeng Shi and Menghui Liu; methodology, Xiaofeng Shi; software, Yahui Ji and Menghui Liu; validation, Xiaofeng Shi and Menghui Liu; formal analysis, Menghui Liu; investigation, Xiaofeng Shi; resources, Xiaofeng Shi; data curation, Yahui Ji and Menghui Liu; writing – original draft preparation, Yahui Ji and Menghui Liu; writing – review and editing, Yahui Ji, Xiaofeng Shi and Menghui Liu; visualisation, Yahui Ji and Menghui Liu; supervision, Xiaofeng Shi; project administration, Xiaofeng Shi. All authors have read and agreed to the published version of the manuscript.

Data availability statement

The bank statement data used in this study contains sensitive information that is subject to privacy protection and relevant legal requirements and therefore cannot be publicly disclosed. The use of the data has been authorised by the relevant institutions and strictly adheres to confidentiality agreements. Readers are encouraged to conduct repeatability and replicability experiments based on their own bank statement data to verify the practicality and generalisation ability of the proposed method.

The source code of the proposed algorithm is publicly available in the *Repository Research-on-Table-Recognition-Algorithms-Based-on-Semantic-Interpretation-and-Structural-Parsing* hosted on the GitHub profile of markhard (https://github.com/markhard). This open-source project is implemented in Python.

Funding

This research was supported by the Anhui Province Electronic Information Peak Discipline Cultivation Program (Grant No. ICII202301), the Fuyang Normal University Graduate Academic Innovation Project (Grant No. FNU2023xscx012), and the Anhui

Engineering Research Center for Intelligent Computing and Information Innovation (Grant No. FYKFKT24054).

Declarations

This research does not involve human participants, animals, or sensitive data requiring ethical approval.

The authors declare no competing interests.

References

Dong, H., Liu, S., Han, S., Fu, Z. and Zhang, D. (2019) 'TableSense: spreadsheet table detection with convolutional neural networks', in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp.69–76.

Hashmi, K.A., Liwicki, M., Stricker, D., Afzal, M.A., Afzal, M.A. and Afzal, M.Z. (2021) 'Current status and performance analysis of table recognition in document images with deep neural networks', *IEEE Access*, Vol. 9, pp.87663–87685, DOI: 10.1109/ACCESS.2021.3087865.

Huang, Y., Yan, Q., Li, Y., Chen, Y., Wang, X., Gao, L. and Tang, Z. (2019) 'A YOLO-based table detection method', in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp.813–818.

Kashinath, T., Jain, T., Agrawal, Y., Anand, T. and Singh, S. (2022) 'End-to-end table structure recognition and extraction in heterogeneous documents', *Applied Soft Computing*, Vol. 123, p.108942, DOI: 10.1016/j.asoc.2022.108942.

Khandokar, I.A. and Deshpande, P. (2025) 'Computer vision-based framework for data extraction from heterogeneous financial tables: a comprehensive approach to unlocking financial insights', *IEEE Access*, Vol. 13, pp.17706–17723.

Li, Y., Huang, Z., Yan, J., Zhou, Y., Ye, F. and Liu, X. (2021) 'GFTE: graph-based financial table extraction', in Del Bimbo, A., Cucchiara, R., Sclaroff, S., Farinella, G.M., Mei, T., Bertini, M., Escalante, H.J. and Vezzani, R. (Eds.): *Pattern Recognition, ICPR International Workshops and Challenges*, pp.644–658, Springer International Publishing, Cham.

Li, C., Liu, W., Guo, R., Yin, X., Jiang, K., Du, Y., Du, Y., Zhu, L., Lai, B., Hu, X. et al. (2022) *PP-OCVR3: More Attempts for the Improvement of Ultra Lightweight OCR System*, arXiv preprint arXiv:2206.03001.

Liu, Y., Zhang, H., Yun, X-L., Ye, J-Y. and Liu, C-L. (2021) 'Table detection and cell segmentation in online handwritten documents with graph attention networks', in *Proceedings of the 2nd ACM International Conference on Multimedia in Asia, MMAsia '20*, Association for Computing Machinery, New York, NY, USA.

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

Peng, W. and Li, X. (2025) 'An ante hoc enhancement method for image-based complex financial table extraction', *Applied Science-Basel*, January, Vol. 15, No. 1, p.370, DOI: 10.3390/app15010370.

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

Schreiber, S., Agne, S., Wolf, I., Dengel, A. and Ahmed, S. (2017) 'DeepDeSRT: deep learning for detection and structure recognition of tables in document images', in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 1, pp.1162–1167.

Singh, S., Varshney, S. and Choudhary, A. (2025) 'Devanagari character encoded mix-merge vision transformer for robust document layout analysis', *International Journal on Document Analysis and Recognition*, 22 September, https://doi.org/10.1007/s10032-025-00556-4.

Wang, J., Lin, W., Ma, C., Li, M., Sun, Z., Sun, L. and Huo, Q. (2023) 'Robust table structure recognition with dynamic queries enhanced detection transformer', *Pattern Recognition*, Vol. 144, p.109817, DOI: 10.1016/j.patcog.2023.109817.

Wu, F., Seuret, M., Mayr, M., Kordon, F., Zoellner, J., Wind, S., Maier, A. and Christlein, V. (2025) 'Lightweight cross-attention-based hooknet for historical handwritten document layout analysis', *International Journal on Document Analysis and Recognition*, September, Vol. 28, pp.409–427.

Xu, Y., Li, M., Cui, L., Huang, S., Wei, F. and Zhou, M. (2020) 'LayoutLM: pre-training of text and layout for document image understanding', in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20*, Association for Computing Machinery, New York, NY, USA, pp.1192–1200.

Zhang, D., Mao, R., Guo, R., Jiang, Y. and Zhu, J. (2023) 'YOLO-table: disclosure document table detection with involution', *International Journal on Document Analysis and Recognition (IJDAR)*, Vol. 26, No. 1, pp.1–14.

Zhao, K., Han, Q., Zhang, C-B., Xu, J. and Cheng, M-M. (2022) 'Deep hough transform for semantic line detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 9, pp.4793–4806.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C. and Sun, M. (2020) 'Graph neural networks: a review of methods and applications', *AI Open*, Vol. 1, pp.57–81, DOI: 10.1016/j.aiopen.2021.01.001.