Deep Learning for Table Detection and Structure Recognition: A Survey

MAHMOUD SALAHELDIN KASEM, Faculty of Computer and Information, Assiut University, Assiut,
Egypt and Chungbuk National University, Cheongju, Republic of Korea

ABDELRAHMAN ABDALLAH, Faculty of Computer and Information, Assiut University, Assiut, Egypt and Ca' Foscari University of Venice, Venezia, Italy

ALEXANDER BERENDEYEV, Satbayev University, Almaty, Kazakhstan

EBRAHEM ELKADY, Faculty of Computer and Information, Assiut University, Assiut, Egypt

MOHAMED MAHMOUD, Faculty of Computer and Information, Assiut University, Assiut, Egypt and Chungbuk National University, Cheongju, Korea (the Republic of)

MAHMOUD ABDALLA, Information Technology Institute, Alexandria, Egypt

MOHAMED HAMADA, Department of Information System, International IT University, Almaty, Kazakhstan

SEBASTIANO VASCON, Ca' Foscari University of Venice, Venezia, Italy

DANIYAR NURSEITOV, JSC NC KazMunayGas, Astana, Kazakhstan

ISLAM TAJ-EDDIN, Faculty of Computer and Information, Assiut University, Assiut, Egypt

Tables are everywhere, from scientific journals, articles, websites, and newspapers all the way to items we buy at the supermarket. Detecting them is thus of utmost importance to automatically understanding the content of a document. The performance of table detection has substantially increased thanks to the rapid development of deep learning networks. The goals of this survey are to provide a profound comprehension of the major developments in the field of Table Detection, offer insight into the different methodologies, and provide a systematic taxonomy of the different approaches. Furthermore, we provide an analysis of both classic and new applications in the field. Lastly, the datasets and source code of the existing models are organized to provide the reader with a compass on this vast literature. Finally, we go over the architecture of utilizing various object detection and table structure recognition methods to create an effective and efficient

Authors' Contact Information: Mahmoud SalahEldin Kasem, Faculty of Computers and Information, Assiut University, Assiut, Egypt and Chungbuk National University, Cheongju, Republic of Korea; e-mail: mahmoud.salah@aun.edu.eg; Abdelrahman Abdallah, Faculty of Computers and Information, Assiut University, Assiut, Egypt and Ca' Foscari University of Venice, Venezia, Veneto, Italy; e-mail: abdelrahmanelsayed@aun.edu.eg; Alexander Berendeyev, Satbayev University, Almaty, Kazakhstan; e-mail: aberendeyev@gmail.com; Ebrahem Elkady, Faculty of Computers and Information, Assiut University, Assiut, Egypt; e-mail: ebrahemelkady@aun.edu.eg; Mohamed Mahmoud, Faculty of Computers and Information, Assiut University, Assiut, Egypt and College of Electrical and Computer Engineering, Chungbuk National University, Cheongju, Chungcheongbuk-do, Korea (the Republic of); e-mail: mohamedabokhalil@aun.edu.eg; Mahmoud Abdalla, Information Technology Institute, Alexandria, Cairo, Egypt; e-mail: mahmoudelsayed201999@gmail.com; Mohamed Hamada, Department of Information System, International IT University, Almaty, Almaty, Kazakhstan; email: M.hamada@iitu.edu.kz; Sebastiano Vascon, Ca' Foscari University of Venice, Venezia, Veneto, Italy; e-mail: sebastiano.vascon@unive.it; Daniyar Nurseitov, JSC NC KazMunayGas, Astana, Kazakhstan; e-mail: D.Nurseitov@niikmg.kz; Islam Taj-Eddin, Faculty of Computer and Information, Assiut University, Assiut, Egypt; e-mail: itajeddin@aun.edu.eg.

ACM 0360-0300/2024/10-ART305 https://doi.org/10.1145/3657281

ACM Comput. Surv., Vol. 56, No. 12, Article 305. Publication date: October 2024.
system, as well as a set of development trends to keep up with state-of-the-art algorithms and future research. We have also set up a public GitHub repository where we will be updating the most recent publications, open data, and source code. The GitHub repository is available at https://github.com/abdoelsayed2016/tabledetection-structure-recognition.

CCS Concepts: • Computing methodologies → Object detection; Object recognition;

Additional Key Words and Phrases: Convolutional neural networks, deep learning, table detection, table structure recognition

ACM Reference Format:
Mahmoud SalahEldin Kasem, Abdelrahman Abdallah, Alexander Berendeyev, Ebrahem Elkady, Mohamed Mahmoud, Mahmoud Abdalla, Mohamed Hamada, Sebastiano Vascon, Daniyar Nurseitov, and Islam TajEddin. 2024. Deep Learning for Table Detection and Structure Recognition: A Survey. *ACM Comput. Surv.* 56, 12, Article 305 (October 2024), 41 pages. https://doi.org/10.1145/3657281

1 INTRODUCTION

Textbooks, lists, formulae, graphs, tables, and other elements are common in documents. Most articles, in particular, contain several sorts of tables. Tables, as a significant part of articles, may convey more information in fewer words and allow readers to quickly explore, compare, and comprehend the content. Table detection (TD) and structure identification are crucial tasks in image analysis because they allow retrieving vital information from tables in a digital format. Because of the document's type and the variety of document layouts, detecting and extracting images or document tables is tough. Researchers have previously used heuristic techniques to recognize tables or to break pages into many parts for table extraction [51]. Few studies have focused on table structure recognition (TSR) in documents following TD [45, 141].

The layout and content analysis of documents are used to detect tables. Tables come in a number of layouts and formats. As a result, creating a general method for TD and TSR is quite difficult. TD is regarded as a difficult subject in the scientific community. A large number of studies have been conducted in this sector, although the majority of them have limitations, such as Table areas cannot be fully detected from document images (DIs) using current commercial and open-source document analysis algorithms, such as Tesseract [44].

Detecting and Structure recognition of tables in documents is challenging due to their varied layouts and formats, making the development of a universal detection and recognition method difficult. Despite extensive research, current algorithms like Tesseract struggle to accurately identify table areas, underscoring the complexity of this issue in document analysis [44].

In recent years, a variety of remarkable and creative strategies have been used to improve deep learning model detection accuracy and solve complex challenges encountered during the training and testing process of deep learning object recognition models. Modification of the activation function of deep convolutional neural networks (CNNs) [136], Transfer learning [71, 83], Image Inpainting [79, 138], cancer diagnosis, detection [1, 46], and classification [20], and medical question answers [2–4, 84], as well as software engineering applications such as optimizing the time

and schedule of software projects[8, 34], Customer Segmentation [6, 54], Intrusion Detection in IoT [80, 133] and handwritten recognition for various languages [55, 81, 91, 125], and inventive ways in the combined selection of the activation function and the optimization system for the proposed deep learning model are among these unique strategies. Among the various variables and initiatives that have contributed to the rapid advancement of TD algorithms, the development of deep convolutional neural networks (DCNNs) and GPU computational capacity should be credited. Deep learning models are now widely used in many aspects of computer vision, including general TD [28, 109]. Table structures, on the other hand, receive far less attention, and the table structure is typically characterized by the rows and columns of a table [52, 82].

Figure 1 shows a basic pipeline comparison of deep learning techniques and conventional approaches for the task of understanding tables. Traditional table recognition (TR) techniques either can't handle varied datasets well enough or need extra metadata from PDF files. Extensive pre- and post-processing were also used in the majority of early approaches to improve the effectiveness of conventional TR systems. However, deep learning algorithms retrieve features using neural networks, primarily CNNs [126], instead of manually created features. Object detection or segmentation networks then try to differentiate the tabular portion that is further broken down and recognized in a DI.

This survey examines deep learning-based TD, recognition, and classification architectures in depth. While current evaluations are comprehensive [19, 139], the majority of them do not address recent advancements in the field.

TD [78, 94, 105] is a foundational task in the domain of DI analysis. This process seeks to identify the presence and precise location of tables within a document or image. The primary goal of TD is not to interpret or understand the data within the table but rather to demarcate its boundaries within the broader document space. Tables are structured data representations that carry substantial informational weight in documents, making their accurate detection crucial. This is especially significant in scanned documents or PDFs where tables cannot be programmatically accessed but need to be extracted for further data analysis or transformation. While TD is about finding where a table is, TSR [48, 109] delves deeper. It involves understanding the internal layout, organization, and relationships of components within a detected table. Specifically, this means identifying individual rows, columns, headers, footers, and cells. Recognizing the structure is pivotal for any subsequent data extraction or transformation tasks. Without a clear understanding of the table's structure, the data within it can be misinterpreted. For instance, mistaking a header for a data row could lead to incorrect data parsing. Table classification is the process of categorizing tables based on various criteria, such as layout, content type, purpose, or complexity. For instance, tables could be classified as full-line tables, partial-line tables, and more. Not all tables serve the same purpose, and understanding the type or category of a table can aid in subsequent processing steps. The primary contributions of this article include:

(1) A comprehensive overview of historical and contemporary Table Datasets, emphasizing their distinct characteristics.
(2) An in-depth review of pivotal TD methodologies, tracing their development and evolution.
(3) An exhaustive exploration of TSR techniques, providing a deep dive into their intricacies.
(4) A comparative study of various Table Classification methods, filling a noticeable gap in the existing literature where such a broad summary was previously absent.
(5) Presentation of experimental results based on several datasets related to TD.

There are several challenges associated with TD and structure recognition. Some of these challenges include:

(1) Tables can have a wide range of shapes, sizes, and styles, which can make it difficult for algorithms to accurately detect and recognize them.
(2) Tables can be located in a variety of different contexts, such as in documents, web pages, or natural images, which can make it difficult for algorithms to generalize to different settings.
(3) Tables can contain a wide range of different types of information, such as text, numbers, and images, which can make it difficult for algorithms to extract and interpret this information.
(4) Tables can be distorted or occluded by other objects in the scene, which can make it difficult for algorithms to accurately detect and recognize them.

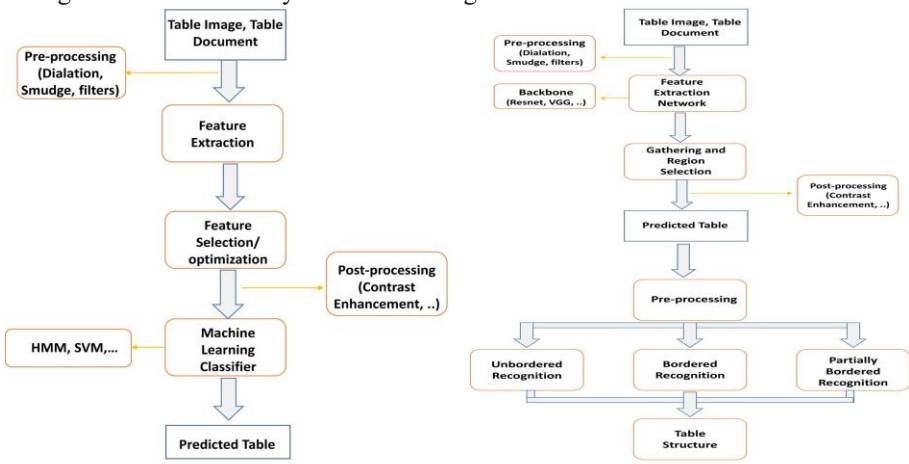


Fig. 1. Table analysis pipeline comparison of conventional and deep learning methods. While convolutional networks are used in deep learning techniques, classical approaches primarily perform feature extraction through image processing techniques. Deep learning methods for interpreting tables are more generalizable and independent of data than conventional approaches.

(5) Tables can be presented in a variety of different formats, such as HTML tables, PDF tables, or scanned images, which can make it difficult for algorithms to handle different input formats.

Overall, these challenges can make it difficult for algorithms to accurately and reliably detect and recognize tables in a wide range of different settings.

1.1 Comparison with Previous Reviews

For many years, the issue with table analysis has been widely acknowledged. Figure 2 shows the upward trend in publications during the previous eight years; these analysis values were derived from Scopus. There have been notable TD and table classification surveys published. There are outstanding studies on the subject of TD in these surveys [19, 139]. There have been few recent surveys that specifically address the subject of TD and classification. B. Coüasnon [15] released another review on TR and forms. The review gives a quick rundown of the most recent techniques at the time, S. Khusro [58] stands as the latest comprehensive survey on PDF table extraction, to our knowledge. Despite deep learning's breakthroughs in fields like visual recognition and medical image analysis, there's a gap in exhaustive surveys on deep learning approaches for TD. A detailed review is essential for progress in this area, particularly benefiting researchers new to the field.

1.2 Scope

The vast number of studies on deep learning for TD precludes a full review within a single article, necessitating selective focus on top-tier journal and conference publications. This article aims at offering a detailed survey of deep learning techniques for detecting, recognizing, and classifying tables, including a taxonomy for understanding these approaches based on datasets, evaluation metrics, and methods. The taxonomy is designed to clarify the similarities and differences between various strategies, aiding readers and guiding future research directions.

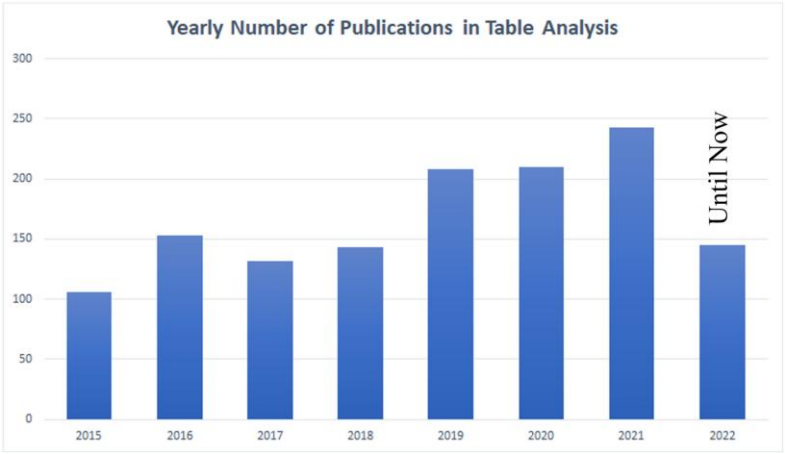


Fig. 2. shows an illustration of an expanding trend in the area of table analysis. This information was gathered by looking through the annual reports on TD and table identification from the years 2015 to 2022, this analysis values were derived from Scopus.

2 MAJOR CHALLENGES

2.1 Object Detection Challenges

Developing a general-purpose algorithm that fulfills two competing criteria of high quality/accuracy and great efficiency is ideal for object detection. High-quality detection must accurately localize and recognize objects in images or video frames, allowing for the distinction of a
wide range of object categories in the real world and localization and recognition of object instances from the same category, despite intra-class appearance variations, for high robustness. High efficiency necessitates that the full detection process is completed in real time while maintaining reasonable memory and storage requirements.

2.2 TD Challenges

Although a trained segmentation model can accurately locate tables, conventional machine learning techniques have flaws in the structural identification of tables. A major issue is the large number of things in such a little space. As a result, the network misses out on critical visual cues that may aid in the detection and recognition of tables [109]. As physical rules are available, intersections of horizontal and vertical lines are computed to recognize table formations. The Hough transform is a prominent approach in computer vision that aids in the detection of lines in document scans [123]. Length, rotation, and average darkness of a line are utilized to filter out false positives and determine if the line is, in fact, a table line [67]. The intersections of the remaining horizontal and vertical lines are computed after the Hough lines have been filtered. Table cells are created based on the crossings.

2.3 TSR Challenges

TR in document analysis is a multifaceted task that involves comprehending the intricate structures of tables within textual content. In the realm of TSR, scholars and researchers have identified two fundamental aspects: logical structure recognition and physical structure recognition. Logical structure recognition delves into the semantic meaning of the table, focusing on understanding relationships and hierarchies among different elements within the table, such as headers, rows, and columns. On the other hand, physical structure recognition centers on the spatial arrangement of table elements on a document page, concentrating on precise localization, boundary delineation, and positional information of cells. In this comprehensive exploration, we delve into these two pivotal aspects separately, discussing the diverse methodologies and techniques employed to tackle each facet [65, 100].

3 A QUICK OVERVIEW OF DEEP LEARNING

From image classification and video processing to speech recognition and natural language understanding, deep learning has transformed a wide range of machine learning activities. Given the incredible rate of change [74], there is a plethora of current survey studies on deep learning [31, 73, 137, 142], medical image analysis applications [73], natural language processing [137], and speech recognition systems [142].

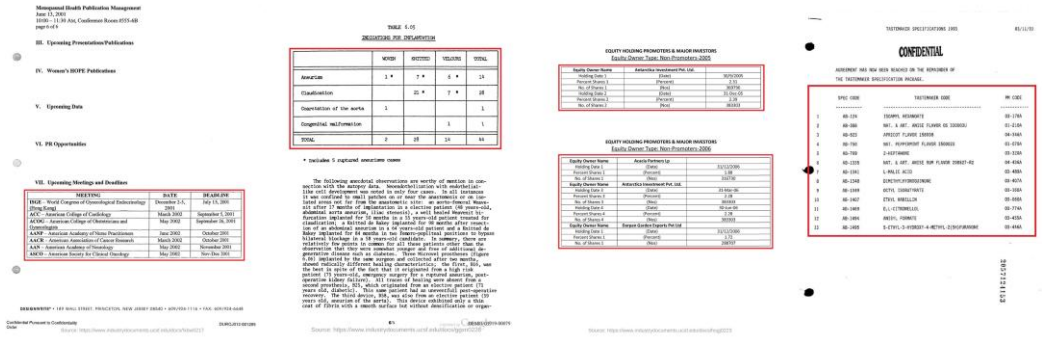
CNNs, the most common deep learning model, can use the fundamental properties of actual signals: translation invariance, local connection, and compositional hierarchies. A typical CNN comprises a hierarchical structure and numerous layers for learning data representations at different levels of abstraction [66]. We start with a convolution.

$$x^{l-1} * w^{l}, \qquad x_{jl} = \sigma \sum_{i=1}^{N_{l-1}} x_{il-1} * w_{il,j} + b_{lj}, \qquad \sigma(x) = max\{x, 0\} \tag{1}$$

between a feature map from the previous layer l-1 and an input feature map x^{l-1}, convolved using a 2D convolutional kernel (or filter or weights) w^{l}. This convolution is seen as a series of layers that have been subjected to a nonlinear process σ, such that with a bias term b^{l}_{j} and a convolution between the N^{l-1} input feature maps x_i^{l-1} and the matching kernel $w_i^{l}{}_{,j}$. For each element, the element-wise nonlinear function $\sigma(.)$ is commonly a rectified linear unit (ReLU) for each element, Finally, pooling is the process of downsampling and upsampling feature maps. DCNNs are CNNs with a large number of layers, often known as “deep” networks. A CNN’s most basic layers consist of a series of feature maps, each of which operates as a neuron. A set of weights $w_{i,j}$ connects each neuron in a convolutional layer to feature maps from the preceding layer (essentially a set of 2D filters). Whereas convolutional and pooling layers make up the early CNN layers, the subsequent layers are usually completely connected. The input picture is repeatedly convolved from earlier to later layers, and the receptive field or region of support grows with each layer. In general, the first CNN layers extract low-level characteristics (such as edges), whereas subsequent layers extract more generic features of increasing complexity [9, 66].

DCNNs have a hierarchical structure that allows them to learn data representations at numerous levels of abstraction, the ability to learn highly complicated functions, and the ability to learn feature representations directly and automatically from data with minimum domain expertise. The availability of huge-size labeled datasets and GPUs with extremely high computational capabilities is what has made DCNNs so successful.

Despite the enormous achievements, there are still acknowledged flaws. There is a critical need for labeled training data as well as expensive computational resources, and selecting proper learning parameters and network designs still requires substantial expertise and experience. Trained networks are difficult to comprehend, and lack resistance to degradations, and many DCNNs have been proven to be vulnerable to assaults [31], all of which restrict their applicability in real-world applications.



(a) ICDAR 2013 (b) ICDAR 2017 (c) ICDAR 2019 (d) RVL-CDIP

Fig. 3. Examples of images in ICDAR 2013, ICDAR 2017, ICDAR 2019, and RVL-CDIP.

4 DATASETS AND EVALUATION METRICS

4.1 Datasets

This section will describe datasets that are available and have been most commonly used for TD, TSR, and classification tasks.

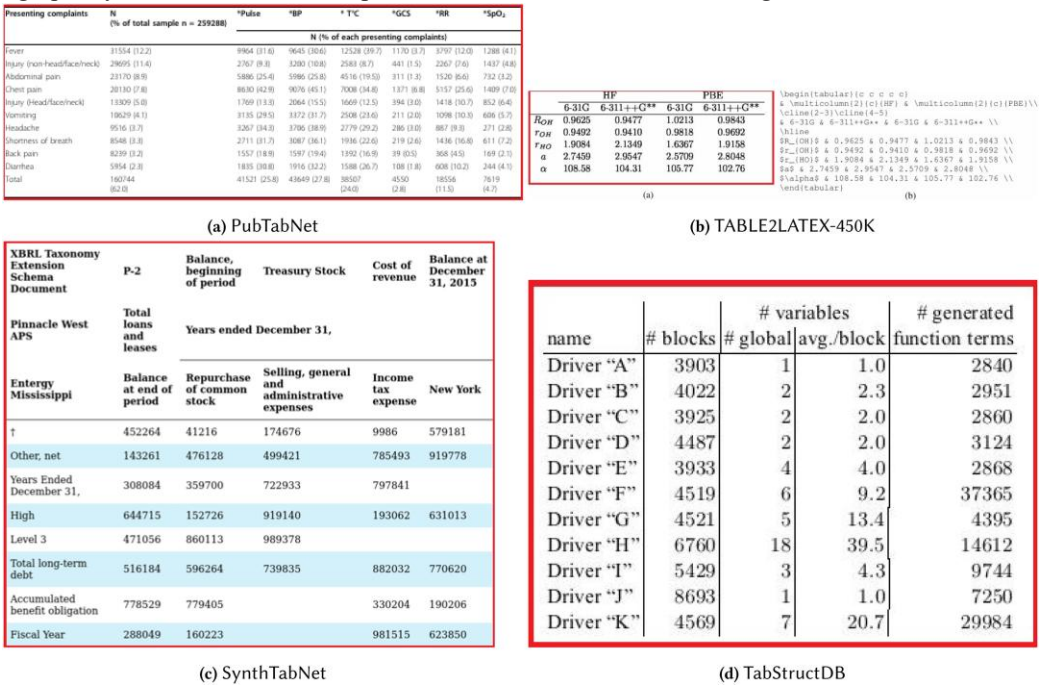
4.1.1 ICDAR 2013. The ICDAR2013 dataset, used as the official practice dataset for the ICDAR2013 competition, was created by collecting PDFs from Google searches, limited to europa.eu and *.gov sites for public domain documents [30]. It includes 150 tables from 27 EU and 40 US Government documents, focusing on TD and structure recognition tasks. The dataset challenges methods in identifying table cell structures and spans multiple pages, as shown in Figure 3.

4.1.2 ICDAR 2017 Page Object Detection *(*POD*).* The ICDAR2017 POD dataset [26], published for testing TD methods, contains 2,417 images from 1,500 CiteSeer scientific articles, including figures, tables, and formulae. It's larger than the ICDAR2013 table dataset, with 1,600 images for training (731 tabular areas) and 817 for testing (350 tabular regions). Examples are shown in Figure 3.

4.1.3 ICDAR2019. ICDAR2019 [25] introduced a dataset for TD (TRACK A) and recognition (TRACK B), divided into historical and modern types. The modern dataset includes diverse formats from scientific articles, forms, and financial documents, while the historical dataset features images from sources like handwritten ledgers and old books. It consists of 1,600 training images and 839 testing images, with TRACK A providing images containing tables and TRACK B divided into two sub-tracks for TSR with or without prior knowledge. Annotations follow a format similar to ICDAR 2013 [30], using XML files to detail table and cell positions. Examples are shown in Figure 3.

4.1.4 TabStructDB. TabStructDB is a different publicly available image-based TSR dataset that was promoted by SA Siddiqui [115]. The authors enhanced the well-known POD dataset from ICDAR-17 by incorporating detailed structural information for all the tabular regions within the dataset. Figure 4 illustrates two examples of this dataset.

4.1.5 TABLE2LATEX-450K. TABLE2LATEX-450K [16], a large dataset released at the latest ICDAR conference, comprises 450,000 annotated tables and associated images. It was created by crawling LaTeX source documents and ArXiv publications from 1991 to 2016, leading to a highquality, refined dataset. Examples from this dataset are shown in Figure 4.



Presenting complaints	N (% of total sample n = 259288)	*Pulse	*BP	* T°C	*GCS	*RR	*SpO₂
		N (% of each presenting complaints)					
Fever	31554 (12.2)	9964 (31.6)	9645 (30.6)	12528 (39.7)	1170 (3.7)	3797 (12.0)	1288 (4.1)
Injury (non-head/face/neck)	29695 (11.4)	2767 (9.3)	3200 (10.8)	2583 (8.7)	441 (1.5)	2267 (7.6)	1437 (4.8)
Abdominal pain	23170 (8.9)	5886 (25.4)	5986 (25.8)	4516 (19.5)	311 (1.3)	1520 (6.6)	732 (3.2)
Chest pain	20130 (7.8)	8630 (42.9)	9076 (45.1)	7008 (34.8)	1371 (6.8)	5157 (25.6)	1409 (7.0)
Injury (Head/face/neck)	13309 (5.0)	1769 (13.3)	2064 (15.5)	1669 (12.5)	394 (3.0)	1418 (10.7)	852 (6.4)
Vomiting	10629 (4.1)	3135 (29.5)	3372 (31.7)	2508 (23.6)	211 (2.0)	1098 (10.3)	606 (5.7)
Headache	9516 (3.7)	3267 (34.3)	3706 (38.9)	2779 (29.2)	286 (3.0)	867 (9.1)	271 (2.8)
Shortness of breath	8548 (3.3)	2711 (31.7)	3087 (36.1)	1936 (22.6)	219 (2.6)	1436 (16.8)	611 (7.2)
Back pain	8239 (3.2)	1557 (18.9)	1597 (19.4)	1392 (16.9)	39 (0.5)	368 (4.5)	169 (2.1)
Diarrhea	5954 (2.3)	1835 (30.8)	1916 (32.2)	1588 (26.7)	108 (1.8)	608 (10.2)	244 (4.1)
Total	160744 (62.0)	41521 (25.8)	43649 (27.8)	38507 (24.0)	4550 (2.8)	18556 (11.5)	7619 (4.7)

(a) PubTabNet

	HF		PBE	
	6-31G	6-311++G**	6-31G	6-311++G**
R_{OH}	0.9625	0.9477	1.0213	0.9843
r_{OH}	0.9492	0.9410	0.9818	0.9692
r_{HO}	1.9084	2.1349	1.6367	1.9158
α	2.7459	2.9547	2.5709	2.8048
α	108.58	104.31	105.77	102.76

(a) (b)

```
\begin{tabular}{c c c c c}
& \multicolumn{2}{c}{HF} & \multicolumn{2}{c}{PBE}\\
\cline{2-3}\cline{4-5}
& 6-31G & 6-311++G** & 6-31G & 6-311++G** \\
\hline
$R_{OH}$ & 0.9625 & 0.9477 & 1.0213 & 0.9843 \\
$r_{OH}$ & 0.9492 & 0.9410 & 0.9818 & 0.9692 \\
$r_{HO}$ & 1.9084 & 2.1349 & 1.6367 & 1.9158 \\
$\alpha$ & 2.7459 & 2.9547 & 2.5709 & 2.8048 \\
$\alpha$ & 108.58 & 104.31 & 105.77 & 102.76 \\
\end{tabular}
```

(b) TABLE2LATEX-450K

XBRL Taxonomy Extension Schema Document	P-2	Balance, beginning of period	Treasury Stock	Cost of revenue	Balance at December 31, 2015
Pinnacle West APS	Total loans and leases	Years ended December 31,			
Entergy Mississippi	Balance at end of period	Repurchase of common stock	Selling, general and administrative expenses	Income tax expense	New York
†	452264	41216	174676	9986	579181
Other, net	143261	476128	499421	785493	919778
Years Ended December 31,	308084	359700	722933	797841	
High	644715	152726	919140	193062	631013
Level 3	471056	860113	989378		
Total long-term debt	516184	596264	739835	882032	770620
Accumulated benefit obligation	778529	779405		330204	190206
Fiscal Year	288049	160223		981515	623850

(c) SynthTabNet

name	# blocks	# variables # global	# variables avg./block	# generated function terms
Driver "A"	3903	1	1.0	2840
Driver "B"	4022	2	2.3	2951
Driver "C"	3925	2	2.0	2860
Driver "D"	4487	2	2.0	3124
Driver "E"	3933	4	4.0	2868
Driver "F"	4519	6	9.2	37365
Driver "G"	4521	5	13.4	4395
Driver "H"	6760	18	39.5	14612
Driver "I"	5429	3	4.3	9744
Driver "J"	8693	1	1.0	7250
Driver "K"	4569	7	20.7	29984

(d) TabStructDB

Fig. 4. Examples of images in PubTabNet, TABLE2LATEX-450K, SynthTabNet, and TabStructDB.

4.1.6 RVL-CDIP (SUBSET). The RVL-CDIP dataset, a prominent collection in document analysis, contains 400,000 images across 16 categories [37]. P. Riba [106] created a subset of this dataset by annotating 518 invoices specifically for TD research. This subset, publicly available, is vital for testing table identification methods in invoice DIs. Examples from this dataset are illustrated in Figure 3.

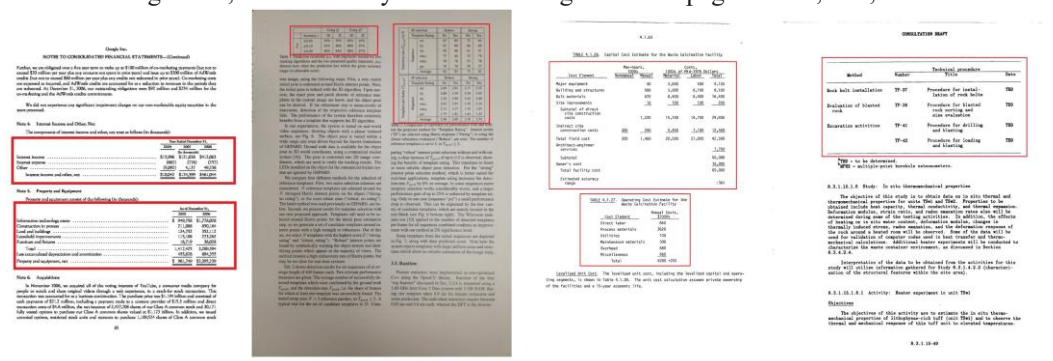
4.1.7 IIIT-AR-13K. IIT-AR-13K, introduced by A Mondal [85], is a new dataset formed from publicly available annual reports in multiple languages, and is the largest manually annotated dataset for graphical page object recognition. It includes diverse annotations like figures, natural images, logos, signatures, and tables, with 11,000 training samples, and 2,000 and 3,000 samples for validation and testing, respectively. Examples from this dataset are in Figure 5.

4.1.8 CamCap. CamCap, proposed by W. Seo [110], is a dataset of camera-captured photos comprising only 85 images, including 38 tables on curved surfaces (1,295 cells) and 47 tables on planar surfaces (1,162 cells). It is publicly available for detecting and identifying table structures

and is crucial for assessing the accuracy of table identification techniques in camera-captured DIs. Two examples from this dataset are shown in Figure 5.

4.1.9 UNLV Table. The UNLV Table dataset [112] consists of 2,889 pages of scanned DIs from diverse sources such as magazines, newspapers, and business letters, available in bitonal, grayscale, and fax formats with resolutions between 200 to 300 DPI. It includes ground truth data with manually marked zones, detailed in text format. Examples from this dataset are displayed in Figure 5.

4.1.10 UW-3 Table. The UW-3 Table dataset [96] contains 1,600 skew-corrected English DIs from books and magazines, with manually edited bounding boxes for page frames, text, non-text



(a) IIIT-AR-13K (b) CamCap (c) UNLV (d) UW3

Fig. 5. Examples of images in IIIT-AR-13K, CamCap, UNLV, and UW3.

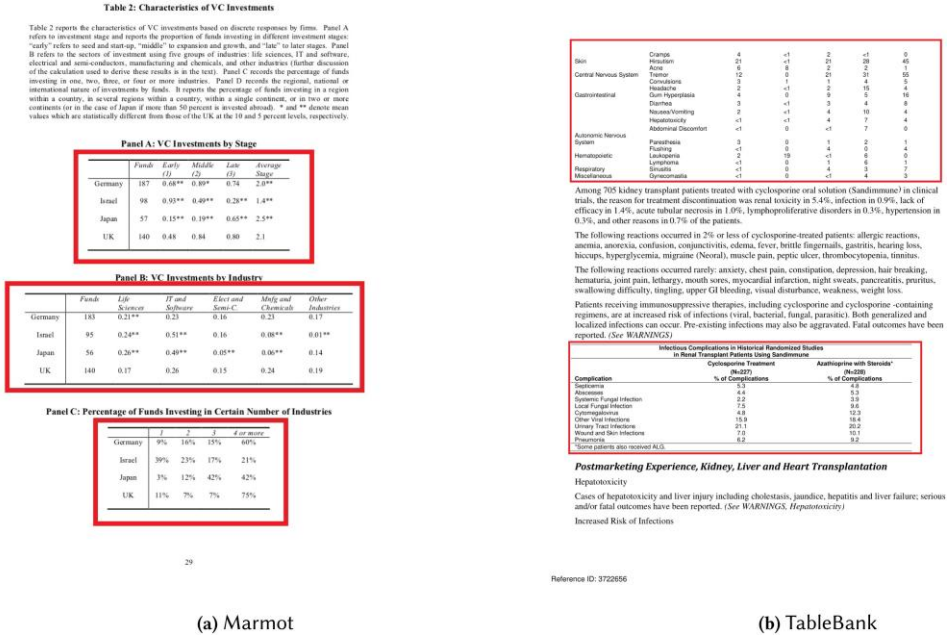
zones, lines, and words. Approximately 120 images include at least one marked table zone. Ground truth, stored in XML, was prepared using the T-Truth tool, with manual validation and corrections for accuracy. Challenges in labeling, especially for column-spanning cells and varying table structures, are noted. Examples from this dataset are in Figure 5.

4.1.11 Marmot. The Marmot dataset [23], a pioneer in TD, comprises 2,000 PDF pages from conference papers in both English and Chinese, ranging from 1970 to 2011, and includes ground truth data. Labeling was standardized and double-checked by 15 people to ensure consistency. The dataset, still expanding, features a balance of Chinese and English pages, with the Chinese pages sourced from over 120 e-Books in Founder Apabi's digital library, and the English pages in both one and two columns. It covers a variety of table types, including ruled, partially and nonruled, horizontal, vertical, inside-column, and span-column tables. Samples from this dataset are displayed in Figure 6.

4.1.12 TableBank. The TableBank dataset [70] introduced a novel weak supervision method for automatically creating a dataset that is significantly larger and of higher quality than existing human-labeled datasets for table analysis. It was compiled by systematically gathering .docx documents from online sources and LaTeX documents from the arXiv database. This approach involves modifying Office XML code for Word documents and LaTeX code to identify table boundaries, resulting in high-quality labeled data across various domains like business documents, official filings, and research articles. The TableBank dataset comprises 417,234 high-quality labeled tables and their original documents. Samples from this dataset are illustrated in Figure 6.

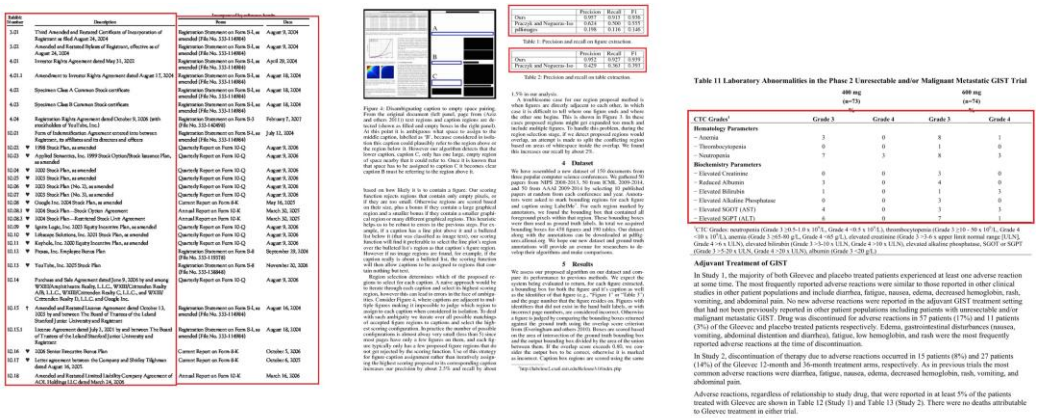
4.1.13 DeepFigures. DeepFigures [119], a dataset for figure extraction, was created without human assistance using scientific articles from databases like arXiv and PubMed. It comprises around 5.5 million tables and figures-induced labels, making it 4,000 times larger than its predecessor and achieving an average precision of 96.8%. This substantial dataset supports the development of modern, data-driven approaches for figure extraction, with samples shown in Figure 7.

4.1.14 PubTables-1M. PubTables-1M [121] is a dataset comprising nearly one million tables from scientific articles. It supports multiple input modalities and offers detailed header and location information for table structures, suitable for various modeling approaches. The dataset introduces a novel canonicalization procedure to address over-segmentation, a common issue in previous datasets, enhancing training performance and providing a more accurate assessment of model performance for TSR. Additionally, transformer-based object detection models trained on



(a) Marmot (b) TableBank

Fig. 6. Examples of images in Marmot and TableBank.

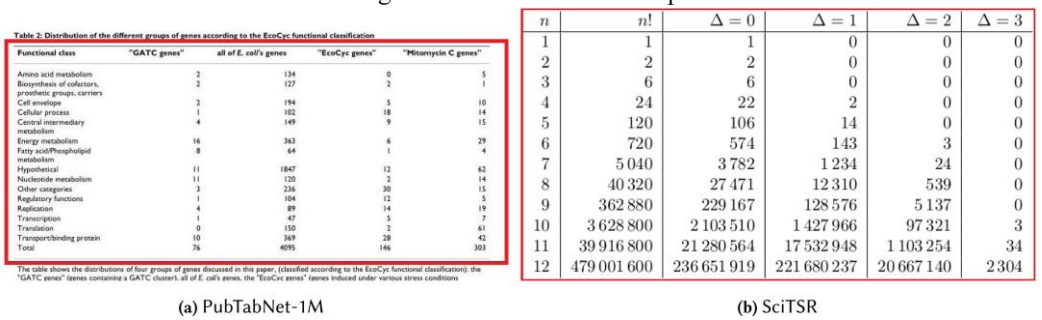


(a) FinTabNet (b) DeepFigures (c) TNCR

Fig. 7. Examples of images in FinTabNet, DeepFigures, and TNCR.

PubTables-1M have shown excellent results in detection, structure recognition, and functional analysis without task-specific customizations. Two examples from this dataset are displayed in Figure 8.

4.1.15SciTSR. SciTSR [14] presents a large-scale TSR dataset compiled by systematically collecting LaTeX source files from the arXiv repository. that comprise 15,000 tables from PDF files and their related structural labels. Figure 8 illustrates two examples of this dataset.



(a) PubTabNet-1M (b) SciTSR

Fig. 8. Examples of images in PubTabNet-1M and SciTSR.

4.1.16FinTabNet. FinTabNet [144] introduces GTE, a vision-guided framework for TD and cellstructured identification, adaptable to any object detection model. GTE-Table uses a penalty based on cell containment constraints for training, while GTE-Cell detects cells using table layouts. The authors developed a method for automatically labeling table and cell structures in texts, creating a large training and testing corpus cost-effectively. FinTabNet comprises real-world scientific and financial datasets with detailed structure annotations. Collaboration with PubTabNet creators enriched FinTabNet with cell labels from PubMed scientific articles. Examples from this dataset are shown in Figure 7.

4.1.17PubTabNet. PubTabNet [146] is a large open-access TR collection with 568k table images and corresponding HTML representations, automatically constructed by comparing XML and PDF

formats of scientific publications from the PubMed CentralTM Open Access Subset (PMCOA). The authors introduced an attention-based encoder-dual-decoder (EDD) architecture for converting table images to HTML code, featuring a structure decoder for table reconstruction and a cell decoder for cell content recognition. They also proposed a new Tree-Edit-Distance-based Similarity (TEDS) metric for TR, effectively addressing multi-hop cell misalignments and OCR errors. Examples from this dataset are displayed in Figure 4.

4.1.18 TNCR. TNCR [1], a new table collection, features images of varied quality sourced from free access websites, and is designed for recognizing and classifying tables in scanned DIs into five categories. The dataset includes approximately 6,621 images with 9,428 captioned tables. Using state-of-the-art deep learning approaches for TD, the study established robust baselines. Notably, Deformable DERT with a Resnet-50 Backbone Network achieved the best performance on the TNCR dataset, with an accuracy of 86.7%, recall of 89.6%, and an F1 score of 88.1%. Samples from this dataset are presented in Figure 7.

4.1.19 SynthTabNet. SynthTabNet, proposed by A Nassar [88], is a synthetic dataset of 600 k samples, developed to diversify appearance styles and complexity in table datasets. It synthesizes elements from Tablebank, PubTabNet, and FinTabNet into four distinct styles, ranging from realistic appearances to colorful, high-contrast, and minimal-content tables. This dataset aims at correcting imbalances in existing datasets. Samples are illustrated in Figure 4.

4.1.20 WiredTableintheWild *(*WTW*).* R. Long [76] introduces a solution for parsing table structures from diverse images, including those with deformations and occlusions, focusing on real-world scenarios with a novel method called Cycle-CenterNet. Built on the CenterNet architecture, Cycle-CenterNet features a cycle-pairing module for detecting and grouping tabular cells into structured tables. Additionally, the article presents the WTW dataset, a comprehensive collection of well-annotated tables from photos, scanned files, and web pages, emphasizing various table styles and scenes.

4.1.21 WikiTableSet. NT Ly [77] introduces WSTabNet, a weakly supervised model for TR in images using HTML or LaTeX annotations instead of detailed cell annotations. This end-to-end system, comprising an encoder, structure decoder, and cell decoder, is trained using images and their HTML/LaTeX codes. The WikiTableSet dataset, sourced from Wikipedia, supports this approach with millions of table images in English, Japanese, and French, including their HTML representations.

4.1.22 STDW. M. Haloi [33] introduces a comprehensive dataset for TD to overcome the limitations of current benchmarks. This dataset, consisting of over seven thousand diverse table samples, was collected from scanned documents, Word files, and searchable PDFs, providing a varied resource for analysis and research. The article showcases baseline results using a CNN-based approach, demonstrating its superiority over traditional computer vision methods in detecting table structures in documents.

4.1.23 TabRecSet. F. Yang [135] delves into TR in pattern recognition, encompassing TD, TSR, and table content recognition (TCR). The study introduces the Table Recognition Set (TabRecSet), a comprehensive dataset and the first to include both English and Chinese languages, tailored for end-to-end TR research. TabRecSet features 38.1 K tables (20.4 K English, 17.7 K Chinese) in various formats, including complete and incomplete borders, regular and irregular shapes, and sourced from diverse scenarios like scanned and camera-taken images, documents,

Excel tables, educational papers, and financial invoices. Additionally, the article presents TableMe, an annotation tool designed for improved efficiency and quality in annotation through visualization and interactivity.

4.1.24ICT-TD. B. Xiao [132] improves TD datasets by enhancing annotations in the "OpenTables" dataset and introducing the "ICT-TD" dataset, which contains 175,682 PDF documents across 370 ICT commodities. These datasets, manually annotated for quality, offer a reliable resource for cross-domain research, with experiments showing their effectiveness for cross-domain evaluations and their ability to improve model performance in such settings.

4.1.25DECO. E Koci [63] introduces the DECO dataset, a collection of 1,165 Enron corpus spreadsheet files annotated for both layout and contents, with assigned roles like Header and Data. The dataset includes marked table borders and categorization for files without tables. The article extensively analyzes the dataset and annotations, offering insights for future research. The detailed annotation methodology, along with the DECO dataset and tools, is openly accessible to the research community.

Table 1 presents a comparison between some of the popular datasets of TD and structure recognition.

4.2 Dataset Challenges

The spectrum of table data analysis is broad and fraught with intricacies. While the presented datasets offer a treasure trove of data for researchers, they also embody an array of challenges, each distinct and demanding.

Starting with foundational datasets like ICDAR2013 and ICDAR2017-POD, one can discern the intricacies tied to source variety. These datasets, which feature data from diverse sources like books, journals, and magazines, present challenges linked to varied layouts and structures. Furthermore, the latter's inclusion of diverse objects elevates the domain of multi-object detection tasks.

Table 1. The Table Illustrates a Quantitative Comparison between Some Famous Datasets in TD

Dataset	Total pages	Total Tables	Table detection	Table Structure	Classification	Document Type
ICDAR2013	462	150	✓	✓	✗	Scanned
ICDAR2017-POD	2,417	-	✓	✗	✗	Scanned
TabStructDB	2.4 k	-	✗	✓	✗	Scanned
TABLE2LATEX-450K	-	450,000	✗	✓	✗	Scanned
RVL-CDIP (SUBSET)	518	-	✓	✗	✗	Scanned
IIIT-AR-13K	13 K	-	✓	✗	✗	Scanned
CamCap	85	-	✓	✓	✗	Camera Capture
UNLV	2,889	-	✓	✓	✗	Scanned
UW-3 dataset	1,600	-	✓	✓	✗	Scanned
Marmot	2,000	-	✓	✗	✗	Scanned
TableBank	-	417,234	✓	✓	✗	Scanned
ICDAR2019	-	2,000	✓	✓	✗	Scanned
DeepFigures	-	5.5 million	✓	✗	✗	Scanned
PubTables-1M	460,589	1 million	✓	✓	✗	Scanned
SciTSR	-	15,000	✗	✓	✗	PDF
FinTabNet	89,646	112,887	✓	✓	✗	PDF and HTML
PubTabNet	-	568 k	✗	✓	✗	Scanned
TNCR	6621	9,428	✓	✗	✓	Scanned
SynthTabNet	600 k	-	✓	✓	✓	Scanned

WTW	14,581	-	✗	✓	✗	Scanned
WikiTableSet	-	5 M	✗	✓	✗	HTML or LaTeX
STDW	7 K	-	✓	✗	✗	Scanned
TabRecSet	32.07 K	38.17 K	✓	✓	✗	Scanned
ICT-TD	5,000	3,581,805	✓	✗	✗	PDF
DECO	1, 165	-	✗	✓	✗	Spreadsheets

However, as we move to Marmot and UNLV, the complexity deepens. Chinese and English language intricacies in Marmot, coupled with the vast array of scanned document challenges in UNLV, like skewing, low-resolution, and diverse layout arrangements, highlight the need for robust preprocessing and detection mechanisms. Meanwhile, DeepFigures and PubTables-1M, due to their volume and figure diversity, require refined segmentation techniques to ensure accurate data extraction. Over-segmentation, particularly in PubTables-1M, emerges as a primary concern, necessitating intelligent interpretation of table structures.

SciTSR and FinTabNet, being domain-specific, carry their unique set of hurdles. SciTSR, centered around scientific articles, grapples with elements like footnotes, subscripts, and superscripts, making data extraction an intricate task. On the other hand, FinTabNet, rooted in the financial domain, presents challenges like intricate layouts, merged cells, and domain-specific jargon and structures. Such nuances can easily lead to misinterpretations if not handled adeptly.

WikiTableSet and TableBank confront linguistic and format diversity. The former's multilingual array and the latter's duality of Word and Excel data sources mandate a versatile extraction and interpretation strategy. ToTTo and WikiSQL, being centered around natural language interfaces, challenge researchers with ensuring context retention and semantic understanding.

TabFact and SQA, while seemingly traditional in format, introduce complexities in reasoning and question answering, requiring models not just to extract but also to infer and deduce. TABMCQ and TURL, being tailored for educational and URL-centered tasks, respectively, present challenges of context sensitivity and accurate source linking.

Datasets like TabbyQA, WikiTables, and OpenTable emphasize scale and structural diversity. The vastness of the data combined with variations in table presentations calls for robust and adaptable analysis techniques. The likes of SemTab, TaPas, and Table-Pretrain introduce semantical, context-driven, and pretraining challenges, urging researchers to not just perceive tables as data structures but as entities with inherent meanings.

Finally, datasets like ExTab, TableNet, DocBank, and TableSet further widen the challenge spectrum. From extending to non-tabular elements in ExTab to grappling with annotations in TableNet and diverse OCR challenges in DocBank, these datasets push the boundaries of table analysis. TableSet, with its focus on adversarial examples, introduces the need for resilient models capable of
withstanding intentionally misleading data.

In essence, the expansive list of datasets, while providing rich opportunities for research, also underscores the multifaceted challenges in table data analysis. As the field progresses, it becomes imperative to develop techniques that are not only accurate but also versatile across varied datasets.

4.3 Metrics

Evaluation in TD, and more critically, in TSR, requires a careful selection of metrics to ensure robustness and accuracy. While table detectors utilize metrics such as frames per second (FPS) for speed evaluation, precision, recall, and mean Average Precision (mAP) are common for performance accuracy.

4.3.1 TD. Precision is derived from Intersection over Union (IoU), which is the ratio of the area of overlap and the area of union between the ground truth and the predicted bounding box. A threshold is set to determine if the detection is correct. If the IoU is more than the threshold, it is classified as True Positive, while an IoU below it is classified as False Positive. If the model fails to detect an object present in the ground truth, it is termed a False Negative. Precision measures the percentage of correct predictions, while recall measures the correct predictions with respect to the ground truth.

$$\text{Average Precision (AP)} = \frac{\text{True Positive (TP)}}{\text{AllObservations}} = \frac{\text{TruePositive}}{\text{(True Positive (TP) + False Positive (FP))}} \quad (2)$$

$$\text{Average Recall (AR)} = \frac{\text{True Positive (TP)}}{\text{(True Positive (TP) + False Negative (FN))}} = \frac{\text{TruePositive}}{\text{AllGroundTruth}}. \quad (3)$$

$$\text{F1-score} = \frac{2 * (\text{AP} * \text{AR})}{(\text{AP} + \text{AR})}, \qquad \text{IOU} = \frac{\text{Area of intersection}}{\text{area of union}}. \quad (4)$$

Based on the above equation, average precision is computed separately for each class. To compare performance between the detectors, the mean of average precision of all classes, called mAP is used, which acts as a single metric for final evaluation.

IOU is a metric that finds the difference between ground truth annotations and predicted bounding boxes. This metric is used in most state-of-the-art object detection algorithms. In object detection, the model predicts multiple bounding boxes for each object, and based on the confidence scores of each bounding box, it removes unnecessary boxes based on their threshold value. We need to declare the threshold value based on our requirements.

4.3.2 TSR. Unlike TD, TSR delves deeper into understanding the components of the table, such as rows, columns, headers, cells, and their inter-relationships.

Directed adjacency relations (DAR) [29, 60]: This metric considers the connectivity of cells in a table, represented as a directed graph. The DAR score is calculated as the fraction of correctly predicted edges in the graph.

$$\text{DAR} = \frac{\text{correctly predicted edges}}{\text{total number of edges}}. \quad (5)$$

Tree edit distance similarity (TEDS) [118]: This metric considers the logical structure of a table represented as a tree. The TEDS score is calculated as the minimum number of edits required to transform the predicted tree into the ground truth tree.

$$\text{TEDS} = \tau \min_{\in T} \text{dist}(T, T), \quad (6)$$

where

T represents the set of all possible trees, T' is the ground truth tree, T (without the prime) is the predicted tree.

dist(T, T') denotes the distance (or the number of edit operations) between the predicted tree T and the ground truth tree T'. 4-gram BLEU score (BLEU-4) [95, 120]: This metric considers the text

content of cells in a table, represented as a sequence of words. The BLEU-4 score is calculated as the similarity between the predicted and ground truth sequences.

$$\text{BLEU-4} = BP \cdot \exp\sum_{n=1}^{4} w_n \log p_n. \tag{7}$$

Here, BP is the brevity penalty.w_n is the weight assigned to the n-gram precision.p_n is the modified n-gram precision.

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-\frac{r}{c})} & \text{if } c \leq r \end{cases}. \tag{8}$$

c is the length of the candidate sequence. r is the length of the reference sequence.

$$p_n = \frac{\sum_{\text{ngram} \in C} \min(\text{count}_C(\text{ngram}), \text{count}_R(\text{ngram}))}{\sum_{\text{ngram} \in C} \text{count}_C(\text{ngram})}. \tag{9}$$

C is the set of n-grams in the candidate sequence. R is the set of n-grams in the reference sequence. $count_C$(ngram) and $count_R$(ngram) are the counts of ngram in the candidate and reference sequences, respectively. TEDS-based IOU similarity (TEDS (IOU)) [65, 100]: This metric combines aspects of TEDS and DAR, considering both the logical and physical structure of a table. The TEDS (IOU) score is calculated as the weighted average of the TEDS score and the IOU score between the predicted and ground truth bounding boxes of the cells.

$$\text{TEDS (IOU)} = \alpha \cdot \text{TEDS} + (1 - \alpha) \cdot \text{IOU}. \tag{10}$$

TEDS (IOU) represents the combined TEDS and IOU similarity metric. α is the weight assigned to the TEDS score. TEDS is the Tree Edit Distance Similarity score. IOU is the Intersection over Union score. $(1 - \alpha)$ is the weight assigned to the IOU score. Grid table similarity metric (GriTS) [120]: This metric evaluates the correctness of a predicted table directly in its natural form as a matrix. To create a similarity measure between matrices, the authors generalize the two-dimensional largest common substructure (2D-LCS) problem to the 2D most similar substructures (2DMSS) problem and propose a polynomial-time heuristic for solving it.

$$GriTS_f(A,B) = \frac{2\, i^{\sim}f(A^{\sim},B^{\sim})}{|A| + |B|}. \tag{11}$$

In order to provide a comprehensive understanding of the various metrics utilized in TSR, a comparison of the most prevalent evaluation metrics is presented. Table 2 shows these metrics, breaking down their components by cell attributes they target, the data structures they represent, their criteria for matching, and their respective scoring methods. As illustrated, different metrics prioritize different aspects of table structure, from content to topology, and their corresponding scoring methods vary accordingly.

Researchers are still actively developing new evaluation metrics for TSR. This is because the task is challenging, and there is no single metric that can perfectly capture all aspects of table structure.

Table 2. Evaluation Metrics for TSR

Evaluation Metric	Cell Attributes	Structure Representation	Matching Criteria	Scoring
DAR [60]	Content	Adjacency Relations Set	Exact match	F-score
DAR [29]	Location	Adjacency Relations Set	Average (Multiple IoU)	F-score
BLEU-4	Topology and Role	Token Sequence	Exact match	BLEU-4
GriTS$_{Top}$	Topology	Cell Matrix	IoU	F-score
GriTS$_{Con}$	Content	Cell Matrix	Normalized LCS	F-score
GriTS$_{Loc}$	Location	Cell Matrix	IoU	F-score

5 TD AND STRUCTURE RECOGNITION MODELS

TD has been studied for an extended period of time. Researchers used different methods that can be categorized as follows: heuristic-based methods, machine learning-based methods, and deep learning-based methods. Primarily heuristic-based methods were mainly used in the 1990s, 2000s, and early 2010. They employed different visual cues like lines, keywords, space features, and so on to detect tables.

P Pyreddy [98] proposed an approach to detecting tables using character alignment, holes, and gaps. Y Wangt [129]. used a statistical approach to detect table lines depending on the distance between consecutive words. Grouped horizontal consecutive words together with vertical adjacent lines were employed to propose table entity candidates. MACA Jahan [49] presented a method that uses local thresholds for word spacing and line height for detecting table regions.

K Itonori [48] proposed a rule-based approach that led to the text-block arrangement and ruled line position to localize the table in the documents. S Chandran [13] developed another TD approach based on vertical and horizontal lines. W Seo [110] used junctions (intersection of the horizontal and vertical line) detection with further processing.

T Hassan [39] locates and segments tables by analyzing spatial features of text blocks. E Oro [93] introduced PDF-TREX, a heuristic bottom-up approach for TR in single-column PDF documents. It uses the spatial features of page elements to align and group them into paragraphs and tables. A Nurminen [90] proposed a set of heuristics to locate subsequent text boxes with common alignments and assign them the probability of being a table.

J Fang [22] used the table header as a starting point to detect the table region and decompose its elements. G Harit [36] proposed a technique for TD based on the identification of unique table start and trailer patterns. S Tupaj [127] proposed an OCR-based TD technique. The system searches for sequences of table-like lines based on the keywords

The above methods work relatively well on documents with uniform layouts. However, heuristic rules need to be tweaked to a wider variety of tables and are not really suited for generic solutions. Therefore, machine learning approaches started to be employed to solve the TD problem.

T Kieninger [59] applied an unsupervised learning approach by clustering word segments. F Cesarini [12] used a modified XY tree supervised learning approach. M Fan [21] uses both supervised and unsupervised approaches to TD in PDF documents. Y Wang [128] applied Decision tree and SVM classifiers to layout, content type, and word group features. T Kasar [53] used the junction detection and then passed the information to the SVM classifier. AC e Silva [18] applied joint probability distribution over sequential observations of visual page elements (Hidden Markov Models) to merge potential table lines into tables. S Klampfl [61] compares two unsupervised TR

methods from digital scientific articles. L O'Gorman's Docstrum algorithm [92] applies KNN to aggregate structures into lines and then uses perpendicular distance and angle between lines to combine them into text blocks. It must be noted that this algorithm was devised in 1993, earlier than other methods mentioned in this section.

F Shafait [111] proposes a useful method for TR that performs well on documents with a range of layouts, including business reports, news stories, and magazine pages. The Tesseract OCR engine offers an open-source implementation of the algorithm.

As neural networks gained interest, researchers started to apply them to document layout analysis tasks. Initially, they were used for simpler tasks like TD. Later on, as more complex architectures were developed, more work was put into table columns and overall structure recognition.

L Hao[35] employed CNN to detect whether a certain region proposal is a table or not. A Gilani [28] proposed a Faster R-CNN-based model to make up for the limitations of L Hao [35] and other prior methodologies.

S Schreiber [109] was the first to perform TD and structure recognition using Faster RCNN. D He [40], used FCN for semantic page segmentation. S Arif [7] attempted to improve the accuracy of Faster R-CNN by using semantic color-coding of text. MM Reza [105] used a combination of GAN-based architecture for TD. M Agarwal [5] used a multistage extension of Mask R-CNN with a dual backbone for detecting tables.

Recently transformer-based models were applied to document layout analysis, B Smock[121] applied N Carion[10] DEtection TRansformer framework, a transformer encoder-decoder architecture, to their table dataset for both TD and structure recognition tasks. J Li [69] proposed a self-supervised pre-trained DI Transformer model using large-scale unlabeled text images for document analysis, including TD

5.1 TD Models

In this section, we examine the deep learning methods used for document image TD. We have divided the methods into several deep-learning ideas for the benefit of our readers' convenience. Tables 3 and 4 list all the object identification-based TD strategies. It also discusses various deep learning-based methods that have been used in these methods.

CNN-based Models. D Prasad [97] presents an automatic TD approach for interpreting tabular data in document pictures, which primarily entails addressing two issues: TD and TSR. Using a single CNN model, provide an enhanced deep learning-based end-to-end solution for handling both TD and structure recognition challenges. CascadeTabNet is a Cascade mask Region-based CNN High-Resolution Network (Cascade mask R-CNN HRNet)-based model that simultaneously identifies table areas and recognizes structural body cells from those tables.

L Hao [35] offers a new method for detecting tables in PDF documents that are based on CNNs, one of the most widely used deep learning models. The suggested method begins by selecting some table-like areas using some vague constraints, then building and refining convolutional networks to identify whether the selected areas are tables or not. Furthermore, the convolutional networks immediately extract and use the visual aspects of table sections, while the non-visual information contained in original PDF documents is also taken into account to aid in better detection outcomes.

DD Nguyen [89] introduces TableSegNet, a fully convolutional network (FCN) with a compact design that concurrently separates and detects tables. TableSegNet uses a shallower path to discover table locations in high resolution and a deeper path to detect table areas in low resolution, splitting the found regions into separate tables. TableSegNet employs convolution blocks with broad kernel

sizes throughout the feature extraction process and an additional table-border class in the main output to increase the detection and separation capabilities.

AA Gurav [32] devised an innovative approach to automate data extraction from diverse digital documents (DDs), including images, scanned files, e-mails, and books. Focusing on DIs, like office documents and scans, they employed CNNs for superior performance. Their unique method, based on weakly supervised learning, detects and recognizes table locations in DI without the need for bounding box annotations. This groundbreaking approach promises efficient and accessible automation of tabular data extraction from varied DDs.

M Haloi [33] addressed limitations in existing TD benchmarks by introducing a large-scale, diverse dataset comprising over seven thousand samples with varied table structures from multiple sources. They employed CNN-based methods, demonstrating their superiority over classical computer vision techniques in detecting table structures within documents. This dataset offers a valuable resource for developing efficient deep learning methods for document layout understanding and tabular data processing.

H Dong[17] developed TableSense, an innovative framework for spreadsheet TD, which is crucial for spreadsheet data intelligence. They used a CNN model tailored for precise table boundary detection, leveraging an active learning approach to create a diverse training dataset. TableSense achieved remarkable performance with 91.3% recall and 86.5% precision, surpassing both existing detection algorithms in common spreadsheet tools and state-of-the-art CNNs in computer vision.

RPN Models. A Gilani [28] has shown how to recognize tables using deep learning. Document pictures are pre-processed initially in the suggested technique. These photos are then sent into a Region Proposal Network for TD, which is followed by a fully connected neural network. The suggested approach works with great precision on a variety of document pictures, including documents, research articles, and periodicals, with various layouts.

Á Casado-García [11] Uses object detection techniques, The authors have shown that finetuning from a closer domain improves the performance of TD after conducting a thorough examination. The authors have utilized Mask R-CNN, YOLO, SSD, and Retina Net in conjunction with object detection algorithms. Two basic datasets are chosen to be used in this investigation, TableBank, and PascalVOC.

N Sun [122] presents a corner-finding approach for faster R-CNN-based TD. The Faster R-CNN network is first used to achieve coarse table identification and corner location. then, coordinate matching is used to group those corners that belong to the same table. Untrustworthy edges are filtered at the same time. Finally, the matching corner group fine-tunes and adjusts the table borders. At the pixel level, the suggested technique enhances table boundary finding precision.

A Samari [108] developed an innovative approach for detecting tables in digitized historical print, addressing challenges in varied table characteristics and their visual similarity to other elements. They introduced the NAS dataset, enhancing evaluation diversity. Their method utilized the Gabor filter for dataset preparation and Faster-RCNN for detection, overcoming labeled data limitations with weakly supervised bounding box extraction and pseudo-labeling, improving model generalization.

GenerativeAdversarialNetwork (GAN) Models. Y Li [72] provides a new network to produce the layout elements for table text and to enhance the performance of less ruled table identification. The GANs and this feature generator model are comparable. The authors mandate that the feature generator model extract comparable features for both heavily governed and loosely ruled tables.

Adaptive and Hybrid Models. Y Huang [47] describes a table-detecting algorithm based on the YOLO principle. The authors offer various adaptive improvements to YOLOv3, including an anchor optimization technique and two post-processing methods, to account for the significant differences between document objects and real objects. also employ k-means clustering for anchor optimization to create anchors that are more suited for tables than natural objects, making it easier for our model to find the exact placements of tables. The additional whitespaces and noisy page objects are deleted from the projected results during the post-processing procedure.

D Zhang [140] suggests a YOLO-table-based TD methodology. To enhance the network's capacity to learn the spatial arrangement aspects of tables, the authors incorporate involution into the network's core, and the authors create a simple Feature Pyramid Network to increase model efficacy. This research also suggests a table-based enhancement technique.

X Zheng [145] provides Global Table Extractor (GTE), a method for jointly detecting tables and recognizing cell structures that can be implemented on top of any object detection model. To train their table network with the help of cell placement predictions, the authors developed GTE-Table, which introduces a new penalty based on the inherent cell confinement limitation of tables. A novel hierarchical cell identification network called GTE-Cell makes use of table styles. Additionally, in order to quickly and inexpensively build a sizable corpus of training and test data, authors develop a method to automatically classify table and cell structures in preexisting texts.

I Kavasidis [56] proposes a method for detecting tables and charts using a combination of deep CNNs, graphical models, and saliency ideas. M Holeček [43] presented the concept of table understanding utilizing graph convolutions in structured documents like bills, extending the applicability of graph neural networks. A PDF document is used in the planned research as well. The job of line item TD and information extraction are combined in this study to tackle the problem of TD. Any word may be quickly identified as a line item or not using the line item technique. Following word classification, the tabular region may be easily identified since, in contrast to other text sections on bills, table lines are able to distinguish themselves rather effectively.

R Liu [75] introduced FewTUD, a benchmark dataset focusing on few-shot table understanding, a challenging task due to limited annotations. They addressed the scarcity of public Chinese tables by creating a large-scale corpus. Additionally, they developed FewTPT, a novel pre-trained language model, and extensively evaluated its performance on the FewTUD benchmark.

P Fischer [24] developed Multi-Type-TD-TSR, an end-to-end solution for TR in scanned documents. This multistage pipeline employs deep learning models and differentiates between three types of tables based on their borders. The system addresses challenges such as rotated images and noise artifacts. Their approach also includes specific algorithms for non-bordered and bordered tables, achieving comprehensive TSR.

T Shehzadi [113] proposes an innovative semi-supervised TD method utilizing the deformable transformer, a deep learning technique. Traditional deep learning methods for TD demand extensive labeled data, but this approach significantly reduces the need for labeled samples. By leveraging the deformable transformer, this method achieves outstanding results on various datasets including PubLayNet, DocBank, ICADR-19, and TableBank. It surpasses both fully supervised methods and previous semi-supervised approaches, demonstrating superior performance with limited labeled data.

5.2 Discussion on TD

The intricate landscape of TD in DIs has witnessed a seismic shift with the proliferation of deep learning methodologies. Within this sphere, several researchers have designed innovative strategies to navigate the nuances and challenges inherent to detecting tables in varied document formats.

The core tenet of many methodologies, as portrayed by A Gilani [28], revolves around preprocessing DIs followed by leveraging neural architectures like the Region Proposal Network. D Prasad [97]'s CascadeTabNet further encapsulates the essence of simultaneous TD and structure recognition, illustrating the benefit of end-to-end solutions. These methodologies showcase the power of employing CNN models, underscoring their adeptness at handling the intricacies of documents ranging from periodicals to research articles.

Table 3. A Comparison of the Benefits and Drawbacks of Several Deep Learning-based TD Methods

Literature	Method	Benefits	Drawbacks
A Gilani [28]	Faster R-CNN	(1) On scanned document pictures, this is the first deep learning-based table detection method. (2) The object detection technique is made easier by converting RGB pixels to distance measures.	There are additional phases in the pre-processing process.
S Schreiber [109]	transfer learning methods + Faster R-CNN	end-to-end strategy for detecting tables and table structures that is straightforward and efficient	When compared to other state-of-the-art techniques, it is less accurate.
SA Siddiqui [117]	Deformable CNN + Faster R-CNN	Deformable convolutional neural networks' dynamic receptive field aids in the reconfiguration of multiple tabular boundaries.	When compared to standard convolutions, deformable convolutions are computationally demanding.
P Riba [106]	OCR-based Graph NN that makes use of textual characteristics	The suggested technique makes use of more data than only spatial attributes.	(1) No comparisons to other state-ofthe-art strategies. (2) Additional annotations are needed using this strategy in addition to the tabular data.
N Sun [122]	Faster R-CNN + Locate corners	(1) Better outcomes are obtained using a novel technique. (2) Faster R-CNN is used to identify not just tables, but also the corners of tabular borders.	(1) It is necessary to do postprocessing operations such as corner refining. (2) Because of the additional detections, the computation is more involved.
I Kavasidis [56]	combination of deep CNNs, graphical models, and saliency	(1) Dilated convolutions rather than conventional convolutions are used. (2) Using this technique, saliency detection is performed in place of table detection.	To provide equivalent results, many processing stages are necessary.
M Holeček [43]	Graph NN + line item identification Method	This approach yields encouraging outcomes when used to layout-intensive documents like invoices and PDFs.	(1) Limited baseline approach without comparisons to other state-of-the-art techniques . (2) No publicly accessible table datasets are used for the evaluation of the approach.
Y Huang [47]	YOLO	In comparison, a quicker and more effective strategy	The suggested methodology relies on data-driven post-processing methods.
Y Li [72]	GAN	For ruling and less ruled tables, the GAN-based strategy drives the network to extract comparable characteristics.	In document images with different tabular layouts, the generator-based model is susceptible.
M Li [70]	Faster R-CNN	This method demonstrates how a basic Faster R-CNN can yield excellent results when used with a huge dataset like TableBank.	Just a simple Faster-RCNN implementation
D Prasad [97]	Cascade mask Region-based CNN High-Resolution Network-based model	The study shows how iterative transfer learning may be used to transform pictures, which can lessen the need for huge datasets.	The same as [28], There are additional phases in the pre-processing process.
Á Casado-García [11]	Liken fine-tuning + Mask R-CNN, RetinaNet, SSD, and YOLO	Describe the advantages of using object detection networks in conjunction with domain-specific fine-tuning techniques for table detection.	Closed domain fine-tuning is still insufficient to get state-of-the-art solutions.
M Agarwal [5]	multistage extension of Mask R-CNN with a dual backbone	(1) A comprehensive object detection-based framework utilizing a composite backbone to deliver state-of-the-art outcomes (2) Extensive tests on benchmark datasets for table detection that are openly accessible.	The technique is computationally expensive since it uses a composite backbone in addition to deformable convolutions.
X Zheng [145]	GTE which is general method for object detection	(1) The problem of table detection is benefited by the extra piece-wise constraint loss introduced. (2) A complete method that is compatible with all object detection frameworks.	Annotations for cellular borders are necessary since the process of table detection depends on cell detection.
AA Gurav [32]	CamNet (ResNet 50 + CAM map prediction)	(1) It does not require detailed bounding box annotations. (2) Enables efficient extraction of structured data	Document layout, fonts, and languages' variability require extra preprocessing for accuracy

Several methodologies have extended the foundational principles of popular object detection strategies to suit the TD landscape. For instance, Y Huang [47]'s YOLO-based approach accentuates the essential modifications needed, such as anchor optimization, to tailor YOLOv3 for document

structures. The emphasis on pre-processing and post-processing to eliminate noise and refine detections offers a holistic view of the entire TD pipeline.

The realm of TD isn't just confined to structured documents. L Hao [35]'s methodology, focusing on PDF documents, epitomizes the importance of preliminary selection of table-like areas, refining detection through convolutional networks. This approach underscores the essence of intertwining visual with non-visual information for enhanced detection outcomes.

Innovative strategies like SA Siddiqui [117]'s usage of deformable CNN paired with Faster R-CNN/FPN further delineate the adaptability of deep learning models. By accommodating variable

Table 4. A Comparison of the Benefits and Drawbacks of Several Deep Learning-based TD Methods (Continue Table 3)

Literature	Method	Benefits	Drawbacks
A Samari [108]	Faster R-CNN	(1) The article addresses the scarcity of comprehensive datasets for table detection, introducing two new datasets with diverse table structures and classes. (2) Innovative table detection approach.	The unavailability of public access to the two datasets prevents the evaluation of state-of-the-art detection results.
R Liu [75]	FewTPT (Table PreTraining)	(1) Provides a comprehensive benchmark dataset for few-shot table understanding. (2) Introduces a comprehensive benchmark dataset for few-shot table understanding, covering five downstream tasks.	Require substantial time, and computational resources.
M Haloi [33]	CNN (RetinaNet)	(1) Diverse dataset reflecting real-world scenarios. (2) CNN methods outperform classical techniques.	They do not compare various stateof-the-art approaches on the STDW dataset.
H Dong [17]	TableSense (CNN and employs active learning)	(1) Effectiveness table detection approach. (2) The introduction of a Precise Bounding Box Regression (PBR) module contributes to more accurate predictions of table boundaries.	Need More Pre-processing Efforts.
P Fischer [24]	CNN Multi-Type-TD (ResNeXt-152)	(1) Utilizes advanced deep learning models, leveraging recent trends in transfer learning, to enhance accuracy and adaptability. (2) The combination of two conventional algorithms into a third, unified algorithm demonstrates an insightful strategy.	(1) The algorithms are designed for tables with basic cell structures, lacking a comprehensive solution for more complex, recursive structures often found in tables. (2) The proposed algorithm's F1-score diminishes at higher IoU thresholds due to the inability to detect sharp borders.
T Shehzadi [113]	Semi-supervised Deformable DETR	(1) Reduces the dependency on large-scale annotated datasets, making the method more practical and cost-effective. (2) effective for handling spatial deformations in document images.	(1) Require significant computational resources. (2) They does not provide insights into the potential limitations or challenges associated with varying levels of annotated data.

table sizes and orientations, it tailors its receptive field, emphasizing the customization and flexibility deep learning offers in detection methodologies.

It's also noteworthy to highlight the dedicated efforts towards refining the precision of TD, such as N Sun [122]'s corner-finding approach. By integrating coordinate matching and filtering untrustworthy edges, this strategy emphasizes the importance of pixel-level precision in delineating table boundaries.

Beyond traditional TD, approaches like I Kavasidis [56]'s combination of deep CNNs, graphical models, and saliency ideas, or M Holeček [43]'s exploration of graph convolutions, extend the boundaries of what's achievable. These methods indicate the continued blurring of lines between classical computer vision techniques and deep learning methodologies.

However, the landscape is further enriched by the inclusion of methods that cater to specialized scenarios. AA Gurav [32]'s approach focuses on automating data extraction from diverse DDs, leveraging CNNs and emphasizing the significance of weakly supervised learning. This methodology exemplifies the potential of deep learning in handling varied DD formats without extensive annotations. Similarly, A Samari [108]'s strategy for detecting tables in historical prints,

R Liu [75]'s emphasis on few-shot table understanding, and M Haloi's large-scale dataset introduction echo the sentiment of embracing diversity in data and challenges.

Innovative frameworks like H Dong [17]'s TableSense accentuate the need for precision in unique scenarios such as spreadsheet TD, exemplifying the adaptability of CNN models. Meanwhile, P Fischer [24]'s Multi-Type-TD-TSR underscores the importance of end-to-end solutions tailored for varied table types.

T Shehzadi [113]'s semi-supervised approach, capitalizing on the deformable transformer, captures the overarching theme of the current research landscape—the quest for optimizing performance while minimizing the need for extensive labeled data.

5.3 Case Study Analysis: Evaluating Methodologies in TR

This section delves into the practical application of TR methodologies through detailed case studies. By examining specific implementations and their outcomes, we aim at highlighting the real-world challenges and benefits associated with these methods. The analysis not only sheds light on the efficacy of various approaches but also underscores the adaptability and limitations of TR technologies in addressing diverse data extraction needs.

5.3.1 Introduction to Case Study Selection. The case studies were carefully selected to cover a wide range of applications, from academic research articles to business financial reports and medical records. This diversity ensures a comprehensive understanding of how TR methodologies perform across different domains. The selection criteria focused on the complexity of the table structures, the document formats, and the specific challenges each application presented.

5.3.2 Case Study 1: Academic Research Article Data Extraction. The goal was to automate data extraction from tables in environmental science academic articles using a CNN-based model, overcoming challenges like diverse table formats and mixed content types. Implementing multi-step preprocessing for format standardization and symbol accuracy, along with semantic analysis in post-processing, enhanced data extraction and organization. This method significantly cut down on manual data compilation time despite requiring substantial computational effort. Automating this process boosted meta-analysis efficiency, enabling the analysis of larger datasets more quickly. This advancement not only streamlines research workflows but also sets a precedent for applying similar technologies in other scientific domains.

5.3.3 Case Study 2: Financial Report TR for Business Intelligence. This case involved extracting financial data from tables in quarterly and annual reports of publicly traded companies to enhance business intelligence analyses. An ensemble approach combining OCR technologies with machine learning-based TR algorithms was utilized to cater to both scanned and digitally generated financial reports. The primary challenge was dealing with the high variability in report formats and the accuracy of financial data extraction critical for analysis. Custom OCR correction algorithms were developed to address common errors in financial data recognition. Additionally, a domain-specific adaptation of the machine learning model was trained on a dataset of financial tables to improve accuracy. This approach enabled highly accurate extraction of financial data across a wide range of report formats, significantly enhancing the business intelligence process. However, the system required ongoing training and adaptation to new report formats, presenting scalability challenges. The implementation led to a more efficient and accurate business intelligence process, enabling deeper and faster financial analyses of competitor and market trends.

5.3.4 Comparative Analysis and Lessons Learned. The case studies underscore the potential of TR methodologies to streamline data extraction across diverse domains. While each approach has its strengths, common challenges include the need for domain-specific adaptations and the balance between accuracy and computational efficiency. These insights pave the way for future innovations in TR technology, emphasizing the importance of flexible, adaptable solutions capable of handling the complexities of real-world applications.

5.4 TSR Models

In order to recognize table structures in DIs, deep learning approaches are reviewed in this part. We divided the methods into discrete deep-learning principles for the benefit of our readers. Tables 5 and 6 list all methods for recognizing table structures based on object detection, as well as their benefits and drawbacks. It also discusses various deep learning-based methods that have been used in these methods.

CNN Based Models. SS Paliwal [94] presents TableNet which is a new end-to-end deep learning model for both TD and structure recognition. To divide the table and column areas, the model uses the dependency between the twin objectives of TD and TSR. Then, from the discovered tabular sub-regions, semantic rule-based row extraction is performed. SA Siddiqui [116] described the structure recognition issue as the semantic segmentation issue. To segment the rows and columns, the authors employed FCNs. The approach of prediction tiling is introduced, which lessens the complexity of table structural identification, assuming consistency in a tabular structure. The author imported pre-trained models from ImageNet and used the structural models of FCN's encoder and decoder. The model creates features of the same size as the original input picture when given an image.

SA Khan [57] presents a robust deep learning-based solution for extracting rows and columns from a recognized table in document pictures in this work. The table pictures are pre-processed before being sent into a bi-directional Recurrent Neural Network (RNN) using Gated Recurrent Units (GRUs) and a fully-connected layer with softmax activation in the suggested solution.

A Nassar [88] provides a fresh identification model for table structures. The latter enhances the most recent EDD from PubTabNet end-to-end deep learning model in two important aspects. First, the authors provide a brand-new table-cell object detection decoder. This allows them to easily access the content of the table cells in programmatic PDFs without having to train any proprietary OCR decoders. The authors claim that this architectural improvement makes table-content extraction more precise and enables them to work with non-English tables. Second, transformer-based decoders take the place of LSTM decoders.

C Tensmeyer [124] has presented SPLERGE (Split and Merge), another method using dilated convolutions. Their strategy entails the use of two distinct deep learning models, the first of which establishes the grid-like layout of the table and the second of which determines if further cell spans over many rows or columns are possible.

Another effort to segment tabular structures is the ReS2TIM article by W Xue [134] which describes the reconstruction of syntactic structures from the table. Regressing the coordinates for each cell is this model's main objective. A network that can identify the neighbors of each cell in a table is initially built using the new technique. In the study, a distance-based weighting system is given that will assist the network in overcoming the training-related class imbalance problem.

To identify rows and columns in tables, KA Hashmi [38] suggested a guided technique for table structure identification. The localization of rows and columns may be made better, according to this study, by using an anchor optimization approach. The boundaries of rows and columns are detected in their proposed work using Mask R-CNN and optimized anchors.

Another study by Y Zou [147] called for the development of an image-based table structure identification technique using FCNs. the shown work divides a table's rows, columns, and cells. All of the table components' estimated bounds are enhanced using connected component analysis. Based on the placement of the row and column separators, row and column numbers are then allocated for each cell. In addition, special algorithms are used to optimize cellular borders. X Shen [114] suggested two modules, referred to as Rows Aggregated (RA) and Columns Aggregated (CA). First, to produce a rough forecast for the rows and columns and address the issue of high

Table 5. A Comparison of the Benefits and Drawbacks of Several Deep Learning-based Table Structure Recognition Methods

Literature	Method	Benefits	Drawbacks
SF Rashid [104]	Uses the geometric position of words + A neural network model (autoMLP)	No reliance on complex layout analysis Mechanism. Can be used on the diverse set of documents with different layouts.	limitation is in marking columns boundaries due to variations in the number of words in each column.
E Koci [62]	Encoding of spatial interrelations between these regions using a graph representation, as well as rules and heuristics	(1) Recognition for single-table and multitable spreadsheets. (2) No reliance on any assumptions with what regards the arrangement of tables.	Tables with few columns and empty cells are not handled well.
SA Siddiqui [115]	deformable CNN + Faster R-CNN	(1) The use of deformable convolution can handle various tabular structures. (2) released a new dataset that contained table structure data.	The tables in the proposed approach won't operate co-rrectly if they have a row and column span.
SA Siddiqui [116]	Fully CNNs	The complexity of the task of identifying table structures is reduced by the proposed prediction tiling approach.	(1) Additional post-processing processes are necessary when rows or columns are excessively fragmented. (2) The technique is based on the tabular structures' consistency assumption.
SR Qasim [99]	Graph NN + CNN	(1) This article also presents a unique, memoryefficient training strategy based on Monte Carlo. (2) The suggested approach makes use of both textual and spatial characteristics.	The publicly accessible table datasets are not used to test the system.
W Xue [134]	Graph NN + weights depending on distance	For the cell relationship network, the class imbalance issue is solved using the distancebased weighting method.	When dealing with sparse tables, the approach is insecure.
C Tensmeyer [124]	Dilated Convolutions + Fully CNN	The technique is effective with both scanned and PDF document images.	The post-processing heuristics determine how the merging portion of the method works.
SA Khan [57]	RNN	The reduced receptive field of CNNs is solved by the bi-directional GRU.	Pre-processing procedures including binarization, noise reduction, and morphological modification are necessary.
P Riba [106]	Graph Neural Networks approach	(1) It is not constrained to rigid tabular layouts in terms of single rows, columns or presence of rule lines. (2) The model is language independent.	(1) The method may have problems when dealing with border cond-itions. (2) There is a small amount of training data in the RVL-CDIP dataset and F1, Precision and Recall metrics are lower than other methods.
Y Deng [16]	Encoder decoder net	(1) In the work that is given, issues with endto-end table recognition are examined. (2) Mad -e a contribution with yet another sizable data -set in the area of table comprehension.	The other publicly accessible table recognition datasets are not used to assess the suggested baseline technique.
E Koci [64]	Graph model + Appl-ication of genetic-based approaches	Requires little to no involvement of domain experts .	The accuracy of GE depends on the number of edges. Specifica-lly, we determined that GE achieves an accuracy of only 19% for multi-table graphs.
SS Paliwal [94]	Networks with fully convolutions	(1) First attempt at combining a single solution to handle both the problem of table detection and structure recognition. (2) A comprehensive method for structure recognition and detection in document pictures.	This approach only functions on column detection when used for table structure extraction.
D Prasad [97]	Cascade mask Regionbased CNN High-Resolution Networkbased model	Direct regression occurs at cellular boundaries using an end-to-end method.	Tables with/out ruling lines must undergo further post-processing.

S Raja [101]	Mask R-CNN + ResNet-101 based Net	(1) An additional alignment loss is suggested for precise cell detection. (2) A trainable top-down for cell identification and bottom-up for structure recognition collection is proposed.	When cells are empty, the strategy is weak.

error tolerance, feature slicing, and tiling are applied. Second, the attention maps of the channels are computed to further obtain the row and column information. In order to complete the rows segmentation and columns segmentation, the authors employ RA and CA to construct a semantic segmentation network termed the Rows and Columns Aggregated Network (RCANet).

C Ma [78] presents RobusTabNet, a novel method for recognizing the structure of tables and detecting their borders from a variety of document pictures. The authors suggest using CornerNet

Table 6. A Comparison of the Benefits and Drawbacks of Several Deep Learning-based TSR Methods (Continue Table 5)

Literature	Method	Benefits	Drawbacks
B Xiao [131]	cells' bounding boxes + conditional attention network	Only utilizes visual features without any metadata .	(1) Assumes that the coordinates of cells in the table are known. (2) Difficulties with tables without borders.
Y Zou [147]	Fully CNNs	(1) Using linked component analysis enhances the outcomes. (2) In a table, cells are segmented in addition to the rows and columns.	To provide comparison findings, a small number of post-processing procedures utilizing specific algorithms are necessary.
X Zhong [146]	Dual decoder with attention-based enc-oding	(1) To assess table recognition techniques, the methodology offers a unique evaluation metric called TEDS. (2) released a huge table dataset.	The technique cannot be readily compared to other state-of-theart techniques.
KA Hashmi [38]	Utilizing an optimization technique for anchors+ Mask RCNN	Networks of region proposals converge more quickly and effectively thanks to optimized anchoring.	This study relies on the pre liminary pre-processing phase of clustering the ground truth to find appropriate anchors.
A Zucker [148]	Character Region Awareness for Text Detection (CRAFT) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN)	A bottom-up method, which emphasizes that the table structure is formed by relative positions of text cells, and not by inherent boundaries .	Cannot handle spreading rows or columns well.
X Zheng [145]	Method for object detecting generally	An additional innovative cluster-based tech nique combined with a hierarchical network to detect tabular forms.	Accurately classifying a table is a prerequisite for final cell structure identification.
Z Zhang [143]	A combination of FCN+ RoI-Align + the pretrained BERT model + GRU decoder	Directly operates on table images with no dependency on meta-information, can process simple and complex tables.	Oversegments tables when space between cells is large, doesn't handle merged cells well.
M Namysl [86]	Rule-based algorithms + graph-based table interpretation method	(1) Approach allows processing images and digital documents. (2) Processing steps can be adapted separately.	(1) Support the most frequent table formats only. Reliance on the presence of predefined keywords. (2) Prone to the errors propagated from the upstream components of system. (3) Focus on the tables with rulings.
A Nassar [88]	End-to-end neural network + CNN Backbone + transformer based layers	(1) Handles different languages without being trained on them. (2) Predicts tables structure and bounding boxes for the table content.	Work with PDF documents.
A Jain [50]	spatial associations + dynamic programming techniques	Recognizing complex table structures having multi-span rows/columns and missing cells.	Uses OCR to read words from images Not language agnostic.
S Raja [102]	object detection	Better detection of empty cells.	Fails for very sparse tables where most of the cells are empty.
J Herzig [42]	Tabular Pre-trained Language Model	Simplifies question-answering by directly predicting denotations from tables, outperforming traditional methods in accuracy, and showcasing efficient transfer learning capabilities.	Limited scope beyond table-related tasks, requires substantial computational resources, and depends heavily on the quality of pre-training data.
SX Rao12 [103]	Weak Supervision + Mask R-CNN	(1) Handles both native PDFs and scanned images. (2) Provides TableAnnotator and ExcelAnnotator, fostering collaborative research.	(1) Computationally demanding. (2) Accuracy hinges on the quality of training data, impacting performance if data is noisy or limited.

M Namysł [87]	heuristic-based structure recognition, and graph-based semantic interpretation	(1) Flexible and adaptable to various document layouts. (2) Handles both image and PDF formats (3) Effective when extracting specific data from chosen table columns.	May require adjustments for unconventional layouts or formats .
NT Ly [77]	WSTabNet (a weakly supervised table recognition model)	(1) Achieves top-tier accuracy on benchmark datasets (2) Simplifies training, enhancing model efficiency.	Relies heavily on specific HTML annotations, limiting applicability to datasets without such annotations.
A Ghosh Chowdhury [27]	self-supervised image classifier + pix2pix GAN	(1)Accurately detects tables and recognizes structures, demonstrated through evaluations on multiple datasets. (2) Reduces dependency on manual annotations.	Requires significant computational resources.

as a new region proposal network to produce higher-quality table proposals for Faster R-CNN, which has greatly increased the localization accuracy of Faster R-CNN for table identification. by utilizing only the minimal ResNet-18 backbone network. Additionally, the authors suggest a brandnew split-and-merge approach for recognizing table structures. In this method, each detected table is divided into a grid of cells using a novel spatial CNN separation line prediction module, and then a Grid CNN cell merging module is used to recover the spanning cells. Their table structure recognizer can accurately identify tables with significant blank areas and geometrically deformed (even curved) tables because the spatial CNN module can efficiently transmit contextual information throughout the whole table picture.

A Jain [50] suggests training a deep network to recognize the spatial relationships between various word pairs included in the table picture in order to decipher the table structure. The authors offer an end-to-end pipeline called TSR-DSAW: TSR through Deep Spatial Association of Words, which generates a digital representation of a table picture in a structured format like HTML. The suggested technique starts by utilizing a text-detection network, such as CRAFT, to identify every word in the input table picture. Next, using dynamic programming, word pairings are created. These word pairings are underlined in each individual image and then given to a DenseNet-121 classifier that has been trained to recognize spatial correlations like same-row, same-column, samecell, or none. Finally, The authors apply post-processing to the classifier output in order to produce the HTML table structure.

SX Rao12 [103] developed TableParser, a system adept at parsing tables in native PDFs and scanned images with high precision. They emphasized the significance of parsing table structures and extracting bounding content from various formats such as PDFs, images, spreadsheets, and CSVs. The study highlighted the efficacy of domain adaptation techniques in developing TableParser. Additionally, they introduced TableAnnotator and ExcelAnnotator, enabling weak supervision and facilitating table parsing. These resources were shared with the research community to encourage further exploration in this area.

NT Ly [77] introduced WSTabNet, a novel weakly supervised model for TR, reducing dependency on detailed and costly annotations. Their approach utilizes only HTML (or LaTeX) codelevel annotations of table images. WSTabNet includes components for feature extraction, table structure generation, and cell content prediction. The model trained end-to-end using stochastic gradient descent, demonstrated superior or comparable accuracy to state-of-the-art methods. To support deep learning in TR, the authors curated WikiTableSet, a vast dataset from Wikipedia, containing millions of table images in multiple languages, enabling extensive experiments and validations.

GAN Models. A Ghosh Chowdhury [27] explores self-supervised learning in document TD, addressing the challenges of extracting tabular information from complex documents. They use a self-supervised image classifier as a primary backbone for supervised object detection and employ a pix2pix GANs approach for TSR. Their proposed methods form a robust machine learning pipeline for TD and structure recognition. Evaluation across various datasets, including domainspecific ones,

demonstrates the effectiveness of these approaches in extracting tabular information from intricately structured documents.

Adaptive and Hybrid Models. A Zucker [148] presents CluSTi, a Clustering approach for recognizing the Structure of Tables in invoice scanned images, as an effective way. CluSTi makes three contributions. To begin, it uses a clustering approach to eliminate high noise from the table pictures. Second, it uses state-of-the-art text recognition to extract all text boxes. Finally, CluSTi organizes the text boxes into the correct rows and columns using a horizontal and vertical clustering technique with optimum parameters. Z Zhang [143] presents Split, Embed, and Merge (SEM) as a table structure recognizer that is accurate. M Namysl [86] presents a versatile and modular table extraction approach in this research.

E Koci [62] offers a new method for identifying tables in spreadsheets and constructing layout areas after determining the layout role of each cell. Using a graph model, they express the spatial interrelationships between these areas. On this foundation, they present Remove and Conquer (RAC), a TR algorithm based on a set of carefully selected criteria.

Using the potential of deformable convolutional networks, SA Siddiqui [115] proposes a unique approach for analyzing tabular patterns in document pictures. P Riba [106] presents a graph-based technique for recognizing tables in document pictures in this article. also employ the location, context, and content type instead of the raw content (recognized text), thus it's just a structural perception technique that's not reliant on the language or the quality of the text reading. E Koci [64] uses genetic-based techniques for graph partitioning, to recognize the sections of the graph matching to tables in the sheet. SR Qasim [99] presents a graph network-based architecture for table recognition as a superior alternative to typical neural networks. S Raja [101] describes a method for recognizing table structure that combines cell detection and interaction modules to locate the cells and forecast their relationships with other detected cells in terms of row and column. Also, structural limitations to the loss function for cell identification as extra differential components. The existing issues with end-to-end table identification were examined by Y Deng [16], who also highlighted the need for a larger dataset in this area. S Raja [102] suggests a novel object-detection-based deep model that is tailored for quick optimization and captures the natural alignments of cells inside tables. Dense TR may still be problematic even with precise cell detection because multi-row/column spanning cells make it difficult to capture long-range row/column relationships. Therefore, the authors also seek to enhance structure recognition by determining a unique rectilinear graph-based formulation. The author emphasizes the relevance of empty cells in a table from a semantics standpoint by introducing a novel loss function designed to capture the natural alignment of cells within a cell detection network. Additionally, they proposed a graphbased approach to establish connections between the identified cells, enabling a more comprehensive understanding of their relationships. The authors recommend a modification to a well-liked assessment criterion to take these cells into consideration. To stimulate fresh perspectives on the issue, then provide a moderately large assessment dataset with annotations that are modeled after human cognition.

B Xiao [131] postulates that a complex table structure may be represented by a graph, where the vertices and edges stand in for individual cells and the connections between them. Then, the authors design a conditional attention network and characterize the table structure identification issue as a cell association classification problem (CATT-Net).

H Li [68] formulates the issue as a cell relation extraction challenge and provides T2, a cuttingedge two-phase method that successfully extracts table structures from digitally preserved texts. T2 offers a broad idea known as a prime connection that accurately represents the direct relationships between cells. To find complicated table structures, it also builds an alignment graph and uses a message-passing network.

Z Chi [14] introduced GraphTSR, a novel graph neural network designed for recognizing intricate table structures within PDF files. Their approach, GraphTSR, utilizes table cells as input and predicts relationships among these cells to understand the table layout accurately, even in complex scenarios involving spanning cells that occupy multiple columns or rows.

M Namysł [87] developed an advanced table extraction system to extract quantitative data from documents with diverse layouts. Their hybrid approach integrates a deep learning-based TD module, heuristic-based structure recognition, and graph-based semantic interpretation. This modular system handles both image format and PDF files, outperforming baseline methods and achieving results comparable to state-of-the-art techniques. Additionally, the system demonstrates high performance, especially when extracting targeted information from specific table columns.

NLP models. J Herzig[42] introduced TAPAS, a novel method for answering natural language questions over tables. Unlike traditional semantic parsing approaches, TAPAS avoids generating complex logical forms, instead predicting answers directly from weak supervision in the form of denotations. The model operates by selecting relevant table cells and applying aggregation operators. TAPAS extends BERT's architecture to encode tables and is trained from a joint pre-training of text segments and Wikipedia tables. In evaluations across three semantic parsing datasets, TAPAS outperformed or matched the accuracy of traditional semantic parsing models, achieving significant improvements in question-answering accuracy, particularly on the SQA dataset. Importantly, it achieved this while utilizing a simpler model architecture.

5.5 Discussion on TSR

TSR in DIs is pivotal for information retrieval and data digitization, particularly in documents that are dense with tabular data. Recent advancements in deep learning have paved the way for a multitude of models and algorithms designed to tackle this challenge. This discussion offers an overview and insight into the key methods and their respective merits and drawbacks.

At the heart of TR lies the problem of understanding spatial relationships between various elements in a document, be they textual or graphical. Most contemporary approaches, such as CluSTi [148] and SEM [143], focus on effectively segmenting the table, recognizing its structure, and then extracting data from it. The use of clustering and embedding techniques showcases the shift towards unsupervised and semi-supervised methodologies, reducing the need for exhaustive manual annotations.

Models like TableNet [94] and ReS2TIM [134] highlight the interconnected nature of TD and structure recognition, arguing that a holistic view of both processes can improve accuracy. Such an integrated approach also allows these models to be more flexible and adaptable to varied table structures.

A trend noticeable in the recent literature is the drift towards more context-aware models. These models, such as the ones proposed by SA Siddiqui [116] and SA Khan [57], emphasize understanding the underlying context and content, moving away from purely structural analysis. This shift provides two significant advantages: language independence and robustness against varying text quality, as highlighted by P Riba [106].

Transformers, originally designed for NLP tasks, have made a notable entrance into the TR domain as well. Nassar's TableFormer [88] exemplifies the adaptability of transformer-based architectures for spatial tasks. Given their capability to capture long-range dependencies, transformers are particularly suited for TR, especially when dealing with complex structures.

The aspect of granularity in TR cannot be overlooked. While some models strive for a macrolevel understanding, identifying tables' boundaries and general layout, others delve into microlevel

details. These models, such as the one proposed by Raja [102], emphasize detecting individual cells and their inter-relationships, which is especially crucial for tables with multi-row/column spanning cells.

Datasets play an undeniable role in the advancement of any machine learning task. The need for extensive and diverse datasets for TR has been accentuated by Y Deng [16]. Recent efforts, such as the WikiTableSet introduced by NT Ly [77], cater to this demand, providing rich training material in multiple languages.

A noteworthy approach to the challenge of TR is self-supervised learning, as advocated by A Ghosh Chowdhury [27]. This method's elegance lies in reducing the dependency on labeled data, which is often a significant bottleneck for deep learning projects.

In summary, TSR has witnessed a paradigm shift in the past few years. From heuristic-based methods to advanced deep learning architectures, the field has evolved rapidly. Each method has its unique strengths, catering to different challenges within TR. Future advancements may well see a fusion of these techniques, aiming for a universal model adept at handling any table structure in DIs.

5.6 Case Study Analysis: Evaluating Methodologies in TSR

5.6.1 Introduction to Case Study Analysis. Concrete case studies provide invaluable insights into the practical application, challenges, and benefits of TSR methodologies. This analysis aims at bridging the gap between theoretical research and real-world application, offering a deeper understanding of how these methodologies perform under various conditions.

5.6.2 Selection Criteria for Case Studies. The case studies were selected based on several criteria: the complexity of the table structures, the diversity of the document formats (including scientific articles, financial reports, and medical records), and the unique challenges each case presented. These criteria ensure a broad perspective on the applicability and performance of TSR methods.

5.6.3 Case Study 1: Financial Report Analysis. The first case study focused on automating data extraction from financial tables in multinational corporation reports to improve the efficiency of quarterly financial analyses. Challenges included variable table formats and the precision required for fine-grained numerical data. To overcome these, the study used a custom version of the TableNet deep learning model, enhanced with specialized OCR for better numerical recognition and fine-tuned on financial tables. Despite the high accuracy achieved in data extraction, the need for detailed fine-tuning and preprocessing underscored the model's limitations in handling diverse tables without specific adjustments. This adaptation of TableNet significantly streamlined the data extraction portion of financial analysis, marking a substantial step toward automating and enhancing financial report processing. The success of this approach opens avenues for applying similar methodologies across different sectors requiring detailed data extraction. Furthermore, it underscores the potential for AI to transform traditional business processes, making them more efficient and less reliant on manual labor.

5.6.4 Case Study 2: Medical Records Extraction. The case study aimed at enhancing digitization accuracy of patient data from scanned medical records into a hospital's electronic system, utilizing Faster R-CNN for TD and an LSTM-based model for recognizing structures despite poor scan quality and varied layouts. Key challenges were low-quality scans, handwritten notes, and maintaining data privacy and security. Solutions included advanced denoising, handwriting recognition, and training on a secure, anonymized medical dataset. This approach improved digitization accuracy and reduced manual errors, though its scalability was limited by the need for extensive preprocessing and a secure training setup. The hybrid deep learning technique

significantly enhanced the efficiency and accuracy of converting medical records into digital form, aiding better patient data management and care.

5.6.5 Comparative Analysis and Lessons Learned. The case studies illustrate the potential of deep learning methodologies to transform TSR across different domains. However, they also underscore the importance of domain-specific adaptations, the challenges posed by diverse document formats, and the critical role of preprocessing steps. Lessons learned include the need for targeted dataset preparation, the potential for hybrid models to address complex recognition tasks, and the importance of privacy considerations in medical applications. These insights contribute to advancing the field of TSR, offering guidance for future research and application.

6 EXPERIMENTS RESULTS

6.1 TD Results

TD is crucial for analyzing the structure of documents by identifying tables and their boundaries within images. We conduct a comparative study on different TD techniques using benchmarks like ICDAR and UNLV, assessing them with the IOU metric detailed in Tables 7 and 8. The evolution

Table 7. TD

Approach	Dataset	Method		IoU											Year
				50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	50%:95%	
Tesseract [111]	UNLV	Tab-stop Detection	Precision Recall F1-Score	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	86.00 79.00 82.35	- - -	- - -	2010
A Gilani [28]	UNLV	Faster R-CNN	Precision Recall F1-Score	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	82.30 90.67 86.29	- - -	- - -	2017
SA Siddiqui [117]	UNLV	Deformable CNN + Faster R-CNN	Precision Recall F1-Score	78.6 74.9 76.7	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2018
Á Casado-García [11]	UNLV	YOLO	Precision Recall F1-Score	- - -	- - -	93.0 95.0 94.0	- - -	92.0 94.0 93.0	- - -	83.0 85.0 84.0	- - -	48.0 49.0 49.0	- - -	- - -	2020
M Agarwal [5]	UNLV	Cascade mask R-CNN	Precision Recall F1-Score	96.0 77.0 86.5	- - -	94.4 75.8 85.1	- - -	91.5 73.4 82.5	- - -	82.6 66.3 74.4	- - -	61.8 49.6 55.7	- - -	- - -	2018
S Schreiber [109]	ICDAR2013	Mask R-CNN	Precision Recall F1-Score	97.40 96.15 96.77	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2017
SA Siddiqui [115]	ICDAR2013	Deformable CNN	Precision Recall F1-Score	99.6 99.6 99.6	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2018
I Kavasidis [56]	ICDAR2013	Semantic Image Segmentation	Precision Recall F1-Score	97.5 98.1 97.8	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2019
Y Huang [47]	ICDAR2013	YOLO	Precision Recall F1-Score	100 94.9 97.3	- - -	98.6 93.6 96.1	- - -	- - -	- - -	89.2 84.6 86.8	- - -	- - -	- - -	- - -	2019
SS Paliwal [94]	ICDAR2013	fully convolutions	Precision Recall F1-Score	96.97 96.28 96.62	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2019
Á Casado-García [11]	ICDAR2013	Mask R-CNN	Precision Recall	- -	- -	70.0 97.0	- -	70.0 97.0	- -	70.0 97.0	- -	47.0 65.0	- -	- -	2020

			F1-Score	-	-	81.0	-	81.0	-	81.0	-	54.0	-	-	
D Prasad [97]	ICDAR2013	Cascade mask R-CNN HRNet	Precision Recall F1-Score	100 100 100	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2020
M Li [70]	ICDAR2013	Faster R-CNN	Precision Recall F1-Score	96.58 95.94 96.25	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2020
M Agarwal [5]	ICDAR2013	Cascade mask R-CNN	Precision Recall F1-Score	100.0 100.0 100.0	- - -	100.0 100.0 100.0	- - -	98.7 98.7 98.7	- - -	94.2 94.2 94.2	- - -	66.0 66.0 66.0	- - -	- - -	2021
X Zheng [145]	ICDAR2013	object detection networks	Precision Recall F1-Score	98.97 99.77 99.31	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2021
SA Siddiqui [115]	ICDAR2017	Deformable CNN	Precision Recall F1-Score	- - -	- - -	96.5 97.1 96.8	- - -	- - -	- - -	96.7 93.7 95.2	- - -	- - -	- - -	- - -	2018
Y Huang [47]	ICDAR2017	YOLO	Precision Recall F1-Score	- - -	- - -	97.8 97.2 97.5	- - -	- - -	- - -	97.5 96.8 97.1	- - -	- - -	- - -	- - -	2019
Abdallah [1]	TNCR	HRNets Cascade Mask R-CNN	Precision Recall F1-Score	88.8 97.0 92.7	88.7 97.0 92.6	88.7 97.0 92.6	88.6 96.7 92.4	88.5 96.7 92.4	88.4 96.5 92.2	87.2 95.5 91.1	85.8 94.2 89.8	82.8 91.8 87.0	73.2 83.6 78.0	81.0 90.3 90.3	2022
Abdallah [1]	TNCR	HRNets - Mask R-CNN	Precision Recall F1-Score	85.9 97.1 91.1	85.7 96.9 90.9	85.7 96.9 90.9	85.7 96.9 90.9	85.2 96.5 90.4	84.8 96.0 90.0	83.3 94.7 88.6	81.6 93.4 87.1	76.4 88.9 82.1	58.5 74.4 65.4	81.6 93.4 87.1	2022
Abdallah [1]	TNCR	HRNets - HTC	Precision Recall F1-Score	88.5 98.7 93.3	88.5 98.7 93.3	88.3 98.4 93.0	88.2 98.4 93.0	88.1 98.2 92.8	87.5 97.6 92.2	86.2 96.6 91.1	84.9 95.4 89.8	80.8 91.5 85.8	69.1 81.6 74.8	78.8 90.1 84.0	2022
Abdallah [1]	TNCR	HRNets - Faster R-CNN	Precision Recall F1-Score	86.7 97.2 91.6	86.5 97.0 91.4	86.3 96.8 91.2	85.9 96.4 90.8	85.3 95.9 90.2	84.5 95.2 89.5*	82.7 94.0 87.9	80.6 91.5 85.7*	75.0 86.9 80.5*	55.6 71.1 62.4	71.1 84.2 77.0	2022
Abdallah [1]	TNCR	HRNets - Cascade R-CNN	Precision Recall F1-Score	89.3 96.7 92.8	89.1 96.5 92.6	89.1 96.5 92.6	89.1 96.4 92.6	88.8 96.1 92.3	88.0 95.6 91.6	87.1 94.8 90.7	85.4 93.5 89.2	83.1 91.4 87.0	70.5 81.1 75.4	79.9 88.9 84.1	2022
Abdallah [1]	TNCR	Mask R-CNN - ResNeXt-101	Precision Recall F1-Score	77.8 97.5 86.5	77.7 97.4 86.4	77.4 96.8 86.0	76.9 96.4 85.5	75.9 95.2 84.4	74.9 94.1 83.4	71.3 91.3 80.0	65.1 85.6 73.9*	47.7 72.5 57.5	40.7 69.5 51.3	43.4 62.6 51.2	2022
Abdallah [1]	TNCR	Faster R-CNN - ResNeXt-101	Precision Recall F1-Score	88.4 97.2 92.5	88.4 97.0 92.5	88.0 96.9 92.2	87.9 96.7 92.0	87.6 96.5 91.8	87.1 96.1 91.3	85.6 95.0 90.0	83.3 93.1 87.9	78.0 88.4 82.8	58.1 72.4 64.4	73.3 84.8 78.6	2022

from basic strategies like Tesseract's tab-stop detection to advanced CNNs like the Faster R-CNN by A Gilani[28] shows significant improvements in accuracy. Recent methods have improved precision and recall across various IOU thresholds, though challenges remain at higher thresholds indicating the need for further refinement. The comparison suggests that newer methods, particularly those leveraging CNNs, offer promising advancements in detecting complex table structures across diverse datasets.

Table 8. TD (Continue Table 7)

Approach	Dataset	Method		IoU											Year
				50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	50%:95%	

Y Li [72]	ICDAR2017	GANs	Precision Recall F1-Score	- - -	- - -	94.4 94.4 94.4	- - -	- - -	- - -	90.3 90.3 90.3	- - -	- - -	- - -	- - -	2019
N Sun [122]	ICDAR2017	Faster R-CNN	Precision Recall F1-Score	- - -	- - -	- - -	- - -	- - -	- - -	94.3 95.6 94.9	- - -	- - -	- - -	- - -	2019
Á Casado-García [11]	ICDAR2017	RetinaNet	Precision Recall F1-Score	- - -	- - -	92.0 87.0 89.0	- - -	92.0 87.0 89.0	- - -	89.0 84.0 86.0	- - -	79.0 75.0 77.0	- - -	- - -	2020
M Agarwal [5]	ICDAR2017	Cascade mask R-CNN	Precision Recall F1-Score	- - -	- - -	96.9 89.9 93.4	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2021
D Prasad [97]	ICDAR2019	Cascade mask R-CNN HRNet	Precision Recall F1-Score	- - -	- - -	- 94.3 -	- - -	- 93.4 -	- - -	- 92.5 -	- - -	- 90.1 -	- - -	- - -	2020
M Agarwal [5]	ICDAR2019	Cascade mask R-CNN	Precision Recall F1-Score	98.7 94.6 96.6	- - -	98.0 93.9 95.9	- - -	97.7 93.6 95.6	- - -	97.1 93.0 95.0	- - -	93.4 89.5 91.5	- - -	- - -	2021
X Zheng [145]	ICDAR2019	object detection networks	Precision Recall F1-Score	- - -	- - -	- - -	- - -	- - -	- - -	96.0 95.0 94.0	- - -	90.0 89.0 94.0	- - -	- - -	2021
DD Nguyen [89]	ICDAR2019	fully convolutional network	Precision Recall F1-Score	- - -	- - -	- 92.8 -	- - -	- 91.7 -	- -	- 91.0 -	- - -	- 87.4 -	- - -	- - -	2022
J Li [69]	ICDAR2019	Vanilla Transformer architecture	Precision Recall F1-Score	- - -	- - -	- 97.89 -	- - -	- 97.22 -	- -	- 97.00 -	- - -	- 93.88 -	- - -	- - -	2022
SA Siddiqui [117]	Mormot	Deformable CNN	Precision Recall F1-Score	84.9 94.6 89.5	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2018
M Agarwal [5]	TableBank	Cascade mask R-CNN	Precision Recall F1-Score	93.4 92.4 92.9	- - -	99.5 97.8 98.6	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2021
P Riba [106]	RVL-CDIP	Graph NN	Precision Recall F1-Score	15.2 36.5 21.5	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2019
P Riba [107]	RVL-CDIP	Graph NN	Precision Recall F1-Score	30.80 25.20 39.60	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2022
P Riba [107]	RVL-CDIP	GAT	Precision Recall F1-Score	30.80 25.20 39.60	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2022
P Riba [107]	RVL-CDIP	GAT	Precision Recall F1-Score	30.80 25.20 39.60	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2022
C Ma [78]	ICDAR2019	Faster R-CNN	Precision Recall F1-Score	- - -	- - -	98.4 94.0 96.1	- - -	98.2 93.9 96.0	- - -	97.7 93.3 95.4	- - -	95.0 90.8 92.9	- - -	- - -	2022
C Ma [78]	IIIT-AR-13K	Faster R-CNN	Precision Recall	- -	- -	- -	- -	- -	- -	- -	- -	99.0 97.8	- -	- -	2022

			F1-Score	-	-	-	-	-	-	-	-	98.4	-	-	
Abdallah [1]	TNCR	Dynamic R-CNN	Precision Recall F1-Score	85.5 97.8 91.2	85.4 97.7 91.1	85.3 97.5 90.9	84.9 97.1 90.5	83.9 96.3 89.6	82.3 94.3 87.8	80.2 92.5 85.9	76.4 88.8 82.1	64.6 79.3 71.1	26.7 45.1 33.5	56.1 71.4 62.8	2022
Abdallah [1]	TNCR	Faster R-CNN	Precision Recall F1-Score	89.3 98.1 93.4	89.3 97.9 93.4	89.0 97.7 93.1	88.8 97.5 92.9	87.9 96.7 92.0	87.6 96.3 91.7	86.2 95.0 90.3	82.3 92.1 86.9	74.7 86.1 79.9	49.5 64.5 56.0	69.4 81.3 74.8	2022
Abdallah [1]	TNCR	Cascade R-CNN	Precision Recall F1-Score	90.5 98.5 94.3	90.3 98.4 94.1	90.2 98.3 94.0	89.9 97.9 93.7	89.3 97.6 93.2	89.1 97.2 92.9	88.4 96.5 92.2	87.6 95.8 91.5	82.6 91.7 86.9	69.3 81.1 74.7	79.9 89.8 84.5	2022
Abdallah [1]	TNCR	HRNets - FCOS	Precision Recall F1-Score	79.0 98.3 87.5	78.8 97.8 87.2	78.2 97.2 86.6	77.9 96.9 86.3	77.0 95.9 85.4	75.9 94.7 84.2	72.9 91.7 81.2	69.1 87.8 77.3	59.6 78.6 67.7	33.5 54.5 41.4	56.3 76.4 64.8	2022

Tables 7 and 8 delve into the specifics of various TD methodologies across different datasets. A notable observation is the employment of GANs by Y Li [72] and the impressive performance of Faster R-CNN by N Sun [122] on the ICDAR2017 dataset. On the ICDAR2017 dataset, Y Li [72] used GAN and reported an F1-Score of 90.3% at an IoU of 80% in 2019. On the same dataset, N Sun [122] employed the Faster R-CNN method, achieving an F1-Score of 94.9% at an IoU of 80% in 2019. Á Casado-García [11] utilized RetinaNet and attained an F1-Score of 86.0% at an IoU of 80% in 2020. M Agarwal [5], using the Cascade mask R-CNN approach on the ICDAR2017 dataset, reported a 93.4% F1-Score at 60% IoU in 2021. On the ICDAR2019 dataset, D Prasad [97] employed the Cascade mask R-CNN HRNet and achieved a 94.3% F1-Score at 60% IoU in 2020. Again, M Agarwal [5] on the ICDAR2019 dataset with the Cascade mask R-CNN reported an F1-Score of 95.0% at 80% IoU in 2021. X Zheng[145] proposed the use of object detection networks for the ICDAR2019 dataset and reached a 94.0% F1-Score at 80% and 90% IoU in 2021. DD Nguyen [89] adopted a fully convolutional network for the ICDAR2019 dataset and reported an F1-Score of 91.0% at 80% IoU in 2022. Meanwhile, J Li [69] implemented the Vanilla Transformer architecture on the same dataset and achieved a remarkable F1-Score of 97.00% at 80% IoU in 2022. SA Siddiqui [117] proposed the use of a Deformable CNN on the Mormot dataset, achieving an F1-Score of
89.5% at 50% IoU in 2018. On the TableBank dataset, M Agarwal [5] employed the Cascade mask RCNN and reported a 98.6% F1-Score at 55% IoU in 2021. On the RVL-CDIP dataset, P Riba [106, 107] utilized a Graph NN in 2019 and 2022, achieving F1-Scores of 21.5% and 39.60%, respectively. He also implemented the Graph Attention Neural Networks (GAT) in 2022 for the same dataset, reporting a consistent F1-Score of 39.60%.

On the TNCR dataset, The Faster R-CNN model has achieved good performance in TD compared with Cascade-RCNN and Cascade Mask-RCNN in most of the backbones. We have trained the Faster R-CNN model with L1 Loss [130] with Resnet-50 for bounding box regression. As shown in Tables 7 and 8, it achieves an f1-score of 0.921. Resnet-101 backbone achieves the highest F1 score over 50% to 65%, ResNeXt-101-64x4d achieves the highest F1 score over 70% to 95%, and ResNeXt101-64x4d achieves the highest F1 score over 50%:95% of 0.786. Resnet-50 backbone with 1× Lr schedule achieves the lowest performance over 50% to 60% IoUs. Also, the Resnet-50 backbone
with L1 Los achieves the lowest performance from 65% to 95% IoUs and also achieves the lowest performance over 50%:95%. HRNets Faster R-CNN detector with various backbone structures with combinations of Lr Schedule. The HRNetV2p-W18 with 1× Lr Schedule backbone shows a low

performance compared with other backbones. it achieves an f1 score of 0.770. It achieves 3.2% less than HRNetV2p-W18 with 2× Lr Schedule. HRNetV2p-W40 with 1× Lr Schedule backbone achieves better performance over 50% to 85% IoUs and HRNetV2p-W40 with 2× Lr Schedule backbone achieves better performance over 90% and 95% IoUs. HRNetV2p-W18 with 2× Lr Schedule backbone achieves an f1 score of 0.802 over 50%:95%. HRNetV2p-W32 with 1× Lr Schedule backbone share the same performance over 50% to 60%.

Also, on the TNCR dataset, We implemented Mask R-CNN [41] to use R-CNN for table objects in an image and also for performing object segmentation for each ROI. As seen in Table 7, Mask R-CNN shows good performance in our dataset in precision, recall, and F1 score for all backbones. Resnet-101 backbone has achieved the highest F1 score of 0.774 over 50%:95% and maintains the highest F1 score at various IoUs. ResNeXt-101-32x4d achieves the lowest performance over 50% to 95% IoUs and also achieves an f1 score of 0.512 over 50%:95%. ResNeXt-101-64x4d also achieves the lowest performance at various IoUs except for 95% IoU.

This comparative analysis underscores the dynamic nature of TD research. From basic methods to sophisticated CNN frameworks, the trajectory has been marked by innovation and integration. With continual advancements, the quest for the ideal TD algorithm, one that marries precision with robustness across diverse challenges, continues.

6.2 TR Results

Recognizing structured data from tables in images and documents involves accurately identifying components like rows and headers across diverse formats. Various methods have been developed to enhance this recognition, with evaluations often conducted on the widely-used ICDAR dataset, which includes table images and XML-based ground truth data. These methods are assessed based on precision, recall, F1-scores, and the IoU metric, which measures the accuracy of area predictions compared to the actual data. The research on TR has progressed from Fully Convolutional

Table 9. TSR

Approach	Dataset	Method		IoU											Year
				50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	50%:95%	
S Schreiber [109]	ICDAR2013	Fully CNN	Precision Recall F1-Score	95.93 87.36 91.44	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2017
SA Siddiqui [115]	ICDAR2013	Deformable CNN	Precision Recall F1-Score	93.19 93.08 92.98	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2019
W Xue [134]	ICDAR2013	Graph NN + weights depending on distance	Precision Recall F1-Score	92.6 44.7 60.3	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2019
SS Paliwal [94]	ICDAR2013	fully CNN	Precision Recall F1-Score	92.15 89.87 90.98	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2019
SA Khan [57]	ICDAR2013	Bi-directional RNN	Precision Recall F1-Score	96.92 90.12 93.39	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2019
C Tensmeyer [124]	ICDAR2013	Dilated Convolutions + Fully CNN	Precision Recall F1-Score	95.8 94.6 95.2	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2019
Z Chi [14]	ICDAR2013	Fully CNN	Precision Recall F1-Score	88.5 86.0 87.2	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2019

Á Casado-García [11]	ICDAR2013	Mask R-CNN	Precision Recall F1-Score	- - -	- - -	70.0 97.0 81.0	- - -	70.0 97.0 81.0	- - -	70.0 97.0 81.0	- - -	47.0 65.0 54.0	- - -	- - -	2020
S Raja [101]	ICDAR2013	Object Detection Methods	Precision Recall F1-Score	92.7 91.1 91.9	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2020
KA Hashmi [38]	ICDAR2013	Object Detection Methods	Precision Recall F1-Score	95.37 95.56 95.46	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2021
S Raja [102]	ICDAR2013	Object Detection Methods	Precision Recall F1-Score	93.3 91.5 92.4	- - -	93.0 90.8 91.9	- - -	80.0 79.1 79.5	- - -	63.8 62.4 63.1	- - -	29.1 28.4 28.7	- - -	- - -	2022
D Prasad [97]	ICDAR2019	Object Detection Methods	Precision Recall F1-Score	- - -	- - -	- - 43.8	- - -	- - 35.4	- - -	- - 19.0	- - -	- - 3.6	- - -	- - -	2020
Y Zou [147]	ICDAR2019	Fully CNN	Precision Recall F1-Score	- - -	- - -	18.79 10.07 13.11	- - -	- - -	- - -	1.71 0.92 1.19	- - -	- - -	- - -	- - -	2021
X Zheng [145]	ICDAR2019	Object Detection Methods	Precision Recall F1-Score	- - 54.8	- - -	- - 38.5	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	2021
S Raja [102]	ICDAR2019	Object Detection Methods	Precision Recall F1-Score	86.4 84.2 85.3	- - -	82.2 78.7 80.4	- - -	64.1 62.5 63.3	- - -	40.4 37.6 38.9	- - -	17.5 13.8 15.4	- - -	- - -	2022
S Raja [102]	UNLV	Object Detection Methods	Precision Recall F1-Score	86.4 84.2 85.3	- - -	84.9 82.8 83.9	- - -	73.5 71.1 72.3	- - -	55.8 53.2 54.5	- - -	17.3 14.8 16.0	- - -	- - -	2022
C Ma [78]	SciTSR	Spatial CNN	Precision Recall F1-Score	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	99.4 99.1 99.3	- - -	- - -	2022

Networks (CNN) to more advanced techniques involving Deformable CNNs, Graph Neural Networks, Bi-directional RNNs, and Object Detection Methods. As shown in Table 9 for instance, methods proposed by S Schreiber [109] (2017) and SS Paliwal [94] (2019) relied heavily on Fully CNN. In contrast, SA Siddiqui [115] (2019) introduced deformable structures into CNN, and W Xue[134] (2019) combined Graph Neural Networks with weight dependencies based on distances. Precision, recall, and F1-score are the primary metrics to evaluate performance. For instance, SA Khan [57] (2019) achieved an impressive precision of 96.92% on the ICDAR2013 dataset using Bidirectional RNNs. However, achieving high precision and recall simultaneously can be challenging. As seen by W Xue [134] (2019), while the precision was high at 92.6%, the recall was considerably low at 44.7%, reflecting the method's difficulty in detecting all relevant table regions.

IoU offers a multi-threshold evaluation. As seen in Table 9, while many studies reported metrics at the IoU of 50%, Á Casado-García [11] (2020) provided insights into performance across a wide range of IoU thresholds, from 60% to 90%. While most studies utilized the ICDAR2013 dataset, recent works like D Prasad [97] (2020) and Y Zou [147] (2021) have started using the ICDAR2019 dataset, potentially due to its updated and more challenging set of table images. It's intriguing to note the diversity in methods. For instance, Á Casado-García [11] (2020) used Mask R-CNN, a method predominantly known for its application in general object detection. On the other hand,

Table 10. Open Source Code for Most of the Studies Articles in TD and TSR

Article	Model	Year	Framework	Link
Z Chi [14]	SciTSR	2019	Pytorch	https://github.com/Academic-Hammer/SciTSR
D Prasad [97]	CascadeTabNet	2020	Pytorch	https://github.com/DevashishPrasad/CascadeTabNet
Á Casado-García [11]	-	2020	mxnet	https://github.com/holms-ur/fine-tuning
M Li [70]	TableBank	2020	Pytorch, Detectron2	https://github.com/doc-analysis/TableBank
S Raja [101]	TabStructNet	2020	tensorflow	https://github.com/sachinraja13/TabStructNet.git
X Zhong [146]	PubTabNet	2020	-	https://github.com/ibm-aur-nlp/PubTabNet
M Agarwal [5]	CDeC-Net	2021	PyTorch	https://github.com/mdv3101/CDeCNet

C Tensmeyer [124] (2019) introduced dilated convolutions into Fully CNN, indicating continuous innovations in network architectures for the task. TR is a dynamic field, facing challenges in achieving both high precision and recall, particularly at strict IoU thresholds. The diversity and complexity of tables in DDs highlight the need for models that can adapt to various structures. Despite these challenges, the progress shown in evaluations using the ICDAR dataset suggests promising directions for future research in this area.

6.3 Open Source Code

Several open-source frameworks for creating generic deep learning models, most of which are written in Python, are available online, including TensorFlow, Keras, PyTorch, and MXNet.The open-source projects for TD and structure recognition are summarized in Table 10. Many of the authors have also made open-source implementations of their proposed models available. TensorFlow and PyTorch are the most often utilized frameworks in these open-source projects.

7 CONCLUSION AND FUTURE WORKS

In the field of document analysis, table analysis is a significant and extensively researched problem. The challenge of interpreting tables has been dramatically transformed, and new standards have been set thanks to the use of deep learning ideas. As we said in the article's main contribution paragraph in the Introduction section, we have addressed several current processes that have advanced the process of information extraction from tables in document pictures by implementing deep learning concepts. We have discussed methods that use deep learning to detect, identify, and classify tables. We have also shown the most and least well-known techniques that have been used to detect and identify tables, respectively. all of the datasets that are publicly accessible and their access details have been compiled. On numerous datasets, we have presented a thorough performance comparison of the methodologies that have been addressed. On well-known datasets that are freely accessible to the public, state-of-the-art algorithms for TD have produced almost flawless results. Once the tabular region has been identified, the work of structurally segmenting tables and then recognizing them follows.

One potential area for future work in the field of TD using deep learning is the integration of additional document structure information into the models. Currently, many deep learning methods for TD primarily rely on the visual cues of tables within documents. However, incorporating supplementary details about the document's structure, such as identifying header rows and columns, could significantly enhance the model's performance.

Another promising direction for future research involves the exploration of more sophisticated deep learning architectures tailored for TD tasks. For instance, investigating the application of advanced techniques such as CNNs or RNNs hold promise in further enhancing the model's accuracy and robustness.

Furthermore, addressing the challenges posed by variations in table formatting and layout is a crucial area for future investigation. Tables exhibit diverse formats, making it essential to develop methods that can robustly detect tables in various layouts. Overcoming these challenges will undoubtedly lead to substantial improvements in the overall performance of TD models.

REFERENCES

[1] Abdelrahman Abdallah, Alexander Berendeyev, Islam Nuradin, and Daniyar Nurseitov. 2022. TNCR: Table net detection and classification dataset. *Neurocomputing* (2022), 79–97. DOI:https://doi.org/10.1016/j.neucom.2021.11.101

[2] Abdelrahman Abdallah, Daniel Eberharter, Zoe Pfister, and Adam Jatowt. 2024. Transformers and language models in form understanding: A comprehensive review of scanned document analysis. arXiv:2403.04080. Retrieved from https://arxiv.org/abs/2403.04080

[3] Abdelrahman Abdallah and Adam Jatowt. 2023. Generator-retriever-generator: A novel approach to open-domain question answering. arXiv:2307.11278. Retrieved from https://arxiv.org/abs/2307.11278

[4] Abdelrahman Abdallah, Mahmoud Kasem, Mahmoud Abdalla, Mohamed Mahmoud, Mohamed Elkasaby, Yasser Elbendary, and Adam Jatowt. 2024. ArabicaQA: A comprehensive dataset for arabic question answering. arXiv:2403.17848. Retrieved from https://arxiv.org/abs/2403.17848

[5] Madhav Agarwal, Ajoy Mondal, and CV Jawahar. 2021. Cdec-net: Composite deformable cascade network for table detection in document images. In *Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 9491–9498.

[6] Ahmed Alsayat. 2023. Customer decision-making analysis based on big social data using machine learning: A case study of hotels in mecca. *Neural Computing and Applications* 35, 6 (2023), 4701–4722.

[7] Saman Arif and Faisal Shafait. 2018. Table detection in document images using foreground and background features. In *Proceedings of the 2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 1–8.

[8] Anders Arpteg, Björn Brinne, Luka Crnkovic-Friis, and Jan Bosch. 2018. Software engineering challenges of deep learning. In *Proceedings of the 2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 50–59.

[9] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 8 (2013), 1798–1828.

[10] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. 2020. End-to-end object detection with transformers. In *Proceedings of the European Conference on Computer Vision*. Springer, 213–229.

[11] Ángela Casado-García, César Domínguez, Jónathan Heras, Eloy Mata, and Vico Pascual. 2020. The benefits of closedomain fine-tuning for table detection in document images. In *Proceedings of the International Workshop on Document Analysis Systems*. Springer, 199–215.

[12] Francesca Cesarini, Simone Marinai, L Sarti, and Giovanni Soda. 2002. Trainable table location in document images. In *Proceedings of the Object Recognition Supported by User Interaction for Service Robots*. IEEE, 236–240.

[13] Surekha Chandran and Rangachar Kasturi. 1993. Structural recognition of tabulated data. In *Proceedings of the 2nd International Conference on Document Analysis and Recognition (ICDAR'93)*. IEEE, 516–519.

[14] Zewen Chi, Heyan Huang, Heng-Da Xu, Houjin Yu, Wanxuan Yin, and Xian-Ling Mao. 2019. Complicated table structure recognition. arXiv:1908.04729. Retrieved from https://arxiv.org/abs/1908.04729

[15] Bertrand Coüasnon and Aurélie Lemaitre. 2014. Recognition of tables and forms.

[16] Yuntian Deng, David Rosenberg, and Gideon Mann. 2019. Challenges in end-to-end neural scientific table recognition. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 894–901.

[17] Haoyu Dong, Shijie Liu, Shi Han, Zhouyu Fu, and Dongmei Zhang. 2019. Tablesense: Spreadsheet table detection with convolutional neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 69–76.

[18] Ana Costa e Silva. 2009. Learning rich hidden Markov models in document analysis: Table location. In *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*. IEEE, 843–847.

[19] David W. Embley, Matthew Hurst, Daniel Lopresti, and George Nagy. 2006. Table-processing paradigms: A research survey. *International Journal of Document Analysis and Recognition* 8, 2 (2006), 66–86.

[20] Rasool Fakoor, Faisal Ladhak, Azade Nazi, and Manfred Huber. 2013. Using deep learning to enhance cancer diagnosis and classification. In *Proceedings of the International Conference on Machine Learning*. ACM, New York, USA, 3937–3949.

[21] Miao Fan and Doo Soon Kim. 2015. Table region detection on large-scale PDF files without labeled data. arXiv:1506.08891. Retrieved from https://arxiv.org/abs/1506.08891

[22] Jing Fang, Prasenjit Mitra, Zhi Tang, and C Lee Giles. 2012. Table header detection and classification. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*.

[23] Jing Fang, Xin Tao, Zhi Tang, Ruiheng Qiu, and Ying Liu. 2012. Dataset, ground-truth and performance metrics for table detection evaluation. In *Proceedings of the 2012 10th IAPR International Workshop on Document Analysis Systems*. IEEE, 445–449.

[24] Pascal Fischer, Alen Smajic, Giuseppe Abrami, and Alexander Mehler. 2021. Multi-type-td-tsr–extracting tables from document images using a multi-stage pipeline for table detection and table structure recognition: From ocr to structured table representations. In *Proceedings of the KI 2021: Advances in Artificial Intelligence: 44th German Conference on AI, Virtual Event, September 27–October 1, 2021*. Springer, 95–108.

[25] Liangcai Gao, Yilun Huang, Hervé Déjean, Jean-Luc Meunier, Qinqin Yan, Yu Fang, Florian Kleber, and Eva Lang. 2019. ICDAR 2019 competition on table detection and recognition (cTDaR). In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1510–1515.

[26] Liangcai Gao, Xiaohan Yi, Zhuoren Jiang, Leipeng Hao, and Zhi Tang. 2017. ICDAR2017 competition on page object detection. In *Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1417–1422.

[27] Arnab Ghosh Chowdhury, Martin ben Ahmed, and Martin Atzmueller. 2022. Towards tabular data extraction from richly-structured documents using supervised and weakly-supervised learning. In *Proceedings of the 2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, 1–4.

[28] Azka Gilani, Shah Rukh Qasim, Imran Malik, and Faisal Shafait. 2017. Table detection using deep learning. In *Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 771–776.

[29] Max Göbel, Tamir Hassan, Ermelinda Oro, and Giorgio Orsi. 2012. A methodology for evaluating algorithms for table understanding in PDF documents. In *Proceedings of the 2012 ACM Symposium on Document Engineering*. 45–48.

[30] Max Göbel, Tamir Hassan, Ermelinda Oro, and Giorgio Orsi. 2013. ICDAR 2013 table competition. In *Proceedings of the 2013 12th International Conference on Document Analysis and Recognition*. IEEE, 1449–1453.

[31] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.

[32] A. A. Gurav and Manisha J. Nene. 2020. Weakly supervised learning-based table detection. *SN Comput. Sci.* 1, 2 (2020), 90. DOI:https://doi.org/10.1007/S42979-020-0113-X

[33] Mrinal Haloi, Shashank Shekhar, Nikhil Fande, Siddhant Swaroop Dash, and Sanjay G. 2022. Table detection in the wild: A novel diverse table detection dataset and method. *arXiv preprint arXiv:2209.09207* (2022).

[34] Mohamed A Hamada, Abdelrahman Abdallah, Mahmoud Kasem, and Mohamed Abokhalil. 2021. Neural network estimation model to optimize timing and schedule of software projects. In *Proceedings of the 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*. IEEE, 1–7.

[35] Leipeng Hao, Liangcai Gao, Xiaohan Yi, and Zhi Tang. 2016. A table detection method for pdf documents based on convolutional neural networks. In *Proceedings of the 2016 12th IAPR Workshop on Document Analysis Systems (DAS)*. IEEE, 287–292.

[36] Gaurav Harit and Anukriti Bansal. 2012. Table detection in document images using header and trailer patterns. In *Proceedings of the 8th Indian Conference on Computer Vision, Graphics, and Image Processing*. 1–8.

[37] Adam W Harley, Alex Ufkes, and Konstantinos G Derpanis. 2015. Evaluation of deep convolutional nets for document image classification and retrieval. In *Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 991–995.

[38] Khurram Azeem Hashmi, Didier Stricker, Marcus Liwicki, Muhammad Noman Afzal, and Muhammad Zeshan Afzal. 2021. Guided table structure recognition through anchor optimization. *IEEE Access* 9 (2021), 113521–113534. DOI:https://doi.org/10.1109/ACCESS.2021.3103413

[39] Tamir Hassan and Robert Baumgartner. 2007. Table recognition and understanding from pdf files. In *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR 2007)*. IEEE, 1143–1147.

[40] Dafang He, Scott Cohen, Brian Price, Daniel Kifer, and C Lee Giles. 2017. Multi-scale multi-task fcn for semantic page segmentation and table detection. In *Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 254–261.

[41] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. 2017. Mask R-CNN. In *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*.

[42] Jonathan Herzig, Paweł Krzysztof Nowak, Thomas Müller, Francesco Piccinno, and Julian Martin Eisenschlos. 2020. TaPas: Weakly supervised table parsing via pre-training. arXiv:2004.02349. Retrieved from https://arxiv.org/abs/2004. 02349

[43] Martin Holeček, Antonín Hoskovec, Petr Baudiš, and Pavel Klinger. 2019. Table understanding in structured documents. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*. IEEE, 158–164.

[44] Jianying Hu, Ramanujan S Kashi, Daniel Lopresti, and Gordon T Wilfong. 2002. Evaluating the performance of table processing algorithms. *International Journal on Document Analysis and Recognition* 4, 3 (2002), 140–153.

[45] Yuan-Ting Hu, Jia-Bin Huang, and Alexander G. Schwing. 2017. MaskRNN: Instance level video object segmentation. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 325–334. Retrieved from https://proceedings.neurips.cc/paper/2017/ hash/6c9882bbac1c7093bd25041881277658-Abstract.html

[46] Zilong Hu, Jinshan Tang, Ziming Wang, Kai Zhang, Ling Zhang, and Qingling Sun. 2018. Deep learning for imagebased cancer detection and diagnosis- A survey. *Pattern Recognit*. 83 (2018), 134–149. DOI:https://doi.org/10.1016/J. PATCOG.2018.05.014

[47] Yilun Huang, Qinqin Yan, Yibo Li, Yifan Chen, Xiong Wang, Liangcai Gao, and Zhi Tang. 2019. A YOLO-based table detection method. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 813–818.

[48] Katsuhiko Itonori. 1993. Table structure recognition based on textblock arrangement and ruled line position. In *Proceedings of the 2nd International Conference on Document Analysis and Recognition (ICDAR'93)*. IEEE, 765–768.

[49] MAC Akmal Jahan and Roshan G Ragel. 2014. Locating tables in scanned documents for reconstructing and republishing. In *Proceedings of the 7th International Conference on Information and Automation for Sustainability*. IEEE, 1–6.

[50] Arushi Jain, Shubham Paliwal, Monika Sharma, and Lovekesh Vig. 2021. TSR-DSAW: Table structure recognition via deep spatial association of words. In *29th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2021, Online event (Bruges, Belgium), October 6-8, 2021*. DOI:https://doi. org/10.14428/ESANN/2021.ES2021-109

[51] K Jain, Anoop M Namboodiri, and Jayashree Subrahmonia. 2001. Structure in on-line documents. In *Proceedings of the 6th International Conference on Document Analysis and Recognition*. IEEE, 844–848.

[52] Ertugrul Kara, Mark Traquair, Murat Simsek, Burak Kantarci, and Shahzad Khan. 2020. Holistic design for deep learning-based discovery of tabular structures in datasheet images. *Eng. Appl. Artif. Intell.* 90 (2020), 103551. DOI:https://doi.org/10.1016/J.ENGAPPAI.2020.103551

[53] Thotreingam Kasar, Philippine Barlas, Sebastien Adam, Clément Chatelain, and Thierry Paquet. 2013. Learning to detect tables in scanned document images using line information. In *Proceedings of the 2013 12th International Conference on Document Analysis and Recognition*. IEEE, 1185–1189.

[54] Mahmoud SalahEldin Kasem, Mohamed Hamada, and Islam Taj-Eddin. 2024. Customer profiling, segmentation, and sales prediction using AI in direct marketing. *Neural Computing and Applications* 36, 9 (2024), 4995–5005.

[55] Mahmoud SalahEldin Kasem, Mohamed Mahmoud, and Hyun-Soo Kang. 2023. Advancements and challenges in Arabic optical character recognition: A comprehensive survey. arXiv:2312.11812. Retrieved from https://arxiv.org/abs/2312.11812

[56] Isaak Kavasidis, Carmelo Pino, Simone Palazzo, Francesco Rundo, Daniela Giordano, P. Messina, and Concetto Spampinato. 2019. A saliency-based convolutional neural network for table and chart detection in digitized documents. In *Proceedings of the International Conference on Image Analysis and Processing*. Springer, 292–302.

[57] Saqib Ali Khan, Syed Muhammad Daniyal Khalid, Muhammad Ali Shahzad, and Faisal Shafait. 2019. Table structure extraction with bi-directional gated recurrent unit networks. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1366–1371.

[58] Shah Khusro, Asima Latif, and Irfan Ullah. 2015. On methods and tools of table detection, extraction and annotation in PDF documents. *Journal of Information Science* 41, 1 (2015), 41–57.

[59] Thomas Kieninger and Andreas Dengel. 1998. The t-recs table recognition and analysis system. In *Proceedings of the International Workshop on Document Analysis Systems*. Springer, 255–270.

[60] Yeon-Seok Kim and Kyong-Ho Lee. 2008. Extracting logical structures from HTML tables. *Computer Standards and Interfaces* 30, 5 (2008), 296–308.

[61] Stefan Klampfl, Kris Jack, and Roman Kern. 2014. A comparison of two unsupervised table recognition methods from digital scientific articles. *D-Lib Magazine* 20, 11 (2014), 7.

[62] Elvis Koci, Maik Thiele, Wolfgang Lehner, and Oscar Romero. 2018. Table recognition in spreadsheets via a graph representation. In *Proceedings of the 2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*. IEEE, 139–144.

[63] Elvis Koci, Maik Thiele, Josephine Rehak, Oscar Romero, and Wolfgang Lehner. 2019. DECO: A dataset of annotated spreadsheets for layout and table recognition. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1280–1285.

[64] Elvis Koci, Maik Thiele, Oscar Romero, and Wolfgang Lehner. 2019. A genetic-based search for adaptive table recognition in spreadsheets. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1274–1279.

[65] Tarun Kumar and Himanshu Sharad Bhatt. 2022. Evaluating table structure recognition: A new perspective. arXiv:2208.00385. Retrieved from https://arxiv.org/abs/2208.00385

[66] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.

[67] Benjamin Charles Germain Lee. 2017. Line detection in binary document scans: A case study with the international tracing service archives. In *Proceedings of the 2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2256–2261.

[68] Huichao Li, Lingze Zeng, Weiyu Zhang, Jianing Zhang, Ju Fan, and Meihui Zhang. 2022. A two-phase approach for recognizing tables with complex structures. In *Proceedings of the International Conference on Database Systems for Advanced Applications*. Springer, 587–595.

[69] Junlong Li, Yiheng Xu, Tengchao Lv, Lei Cui, Cha Zhang, and Furu Wei. 2022. DiT: Self-supervised pre-training for document image transformer. In *MM'22: The 30th ACM International Conference on Multimedia, Lisboa, Portugal, October 10 - 14, 2022*, ACM, 3530–3539. DOI:https://doi.org/10.1145/3503161.3547911

[70] Minghao Li, Lei Cui, Shaohan Huang, Furu Wei, Ming Zhou, and Zhoujun Li. 2020. Tablebank: Table benchmark for image-based table detection and recognition. In *Proceedings of the 12th Language Resources and Evaluation Conference*. 1918–1925.

[71] Shun Li, WeiDong Liu, and GongBing Xiao. 2019. Detection of srew nut images based on deep transfer learning network. In *Proceedings of the 2019 Chinese Automation Congress (CAC)*. IEEE, 951–955.

[72] Yibo Li, Liangcai Gao, Zhi Tang, Qinqin Yan, and Yilun Huang. 2019. A GAN-based feature generator for table detection. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 763–768.

[73] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. 2017. A survey on deep learning in medical image analysis. *Medical Image Anal.* 42 (2017), 60–88. DOI:https://doi.org/10.1016/J.MEDIA.2017.07.005

[74] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. 2020. Deep learning for generic object detection: A survey. *International Journal of Computer Vision* 128, 2 (2020), 261–318.

[75] Ruixue Liu, Shaozu Yuan, Aijun Dai, Lei Shen, Tiangang Zhu, Meng Chen, and Xiaodong He. 2022. Few-shot table understanding: A benchmark dataset and pre-training baseline. In *Proceedings of the 29th International Conference on Computational Linguistics*. 3741–3752.

[76] Rujiao Long, Wen Wang, Nan Xue, Feiyu Gao, Zhibo Yang, Yongpan Wang, and Gui-Song Xia. 2021. Parsing table structures in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 944–952.

[77] Nam Tuan Ly, Atsuhiro Takasu, Phuc Nguyen, and Hideaki Takeda. 2023. Rethinking image-based table recognition using weakly supervised methods. arXiv:2303.07641. Retrieved from https://arxiv.org/abs/2303.07641

[78] Chixiang Ma, Weihong Lin, Lei Sun, and Qiang Huo. 2023. Robust table detection and structure recognition from heterogeneous document images. *Pattern Recognit.* 133 (2023), 109006. DOI:https://doi.org/10.1016/J.PATCOG.2022. 109006

[79] Mohamed Mahmoud and Hyun-Soo Kang. 2023. GANMasker: A two-stage generative adversarial network for highquality face mask removal. *Sensors* 23, 16 (2023), 7094.

[80] Mohamed Mahmoud, Mahmoud Kasem, Abdelrahman Abdallah, and Hyun Soo Kang. 2022. AE-LSTM: Autoencoder with LSTM-based intrusion detection in IoT. In *Proceedings of the 2022 International Telecommunications Conference (ITC-Egypt)*. IEEE, 1–6.

[81] Sabri A Mahmoud, Irfan Ahmad, Wasfi G Al-Khatib, Mohammad Alshayeb, Mohammad Tanvir Parvez, Volker Märgner, and Gernot A Fink. 2014. KHATT: An open Arabic offline handwritten text database. *Pattern Recognition* 47, 3 (2014), 1096–1112.

[82] Song Mao, Azriel Rosenfeld, and Tapas Kanungo. 2003. Document structure analysis algorithms: A literature survey. In *Document Recognition and Retrieval X, Santa Clara, California, USA, January 22-23, 2003, Proceedings (SPIE Proceedings)*, SPIE, 197–207. DOI:https://doi.org/10.1117/12.476326

[83] Katleho L Masita, Ali N Hasan, and Satyakama Paul. 2018. Pedestrian detection using R-CNN object detector. In *Proceedings of the 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. IEEE, 1–6.

[84] Shervin Minaee and Zhu Liu. 2017. Automatic question-answering using a deep similarity neural network. In *Proceedings of the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 923–927.

[85] Ajoy Mondal, Peter Lipps, and CV Jawahar. 2020. IIIT-AR-13K: A new dataset for graphical object detection in documents. In *Proceedings of the International Workshop on Document Analysis Systems*. Springer, 216–230.

[86] Marcin Namysl, Alexander M Esser, Sven Behnke, and Joachim Köhler. 2022. Flexible table recognition and semantic interpretation system. In *Proceedings of the VISIGRAPP (4: VISAPP)*. 27–37.

[87] Marcin Namysł, Alexander M Esser, Sven Behnke, and Joachim Köhler. 2023. Flexible hybrid table recognition and semantic interpretation system. *SN Computer Science* 4, 3 (2023), 246.

[88] Ahmed Nassar, Nikolaos Livathinos, Maksym Lysak, and Peter Staar. 2022. TableFormer: Table structure understanding with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4614–4623.

[89] Duc-Dung Nguyen. 2022. TableSegNet: A fully convolutional network for table detection and segmentation in document images. *International Journal on Document Analysis and Recognition* 25, 1 (2022), 1–14.

[90] Anssi Nurminen. 2013. *Algorithmic Extraction of Data in Tables in PDF Documents*. Master's thesis.

[91] Daniyar Nurseitov, Kairat Bostanbekov, Daniyar Kurmankhojayev, Anel Alimova, Abdelrahman Abdallah, and Rassul Tolegenov. 2021. Handwritten Kazakh and Russian (HKR) database for text recognition. *Multimedia Tools and Applications* 80, 21 (2021), 33075–33097.

[92] Lawrence O'Gorman. 1993. The document spectrum for page layout analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 11 (1993), 1162–1173.

[93] Ermelinda Oro and Massimo Ruffolo. 2009. TREX: An approach for recognizing and extracting tables from PDF documents. In *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*. IEEE, 906–910.

[94] Shubham Singh Paliwal, D Vishwanath, Rohit Rahul, Monika Sharma, and Lovekesh Vig. 2019. Tablenet: Deep learning model for end-to-end table detection and tabular data extraction from scanned document images. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 128–133.

[95] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. 311–318.

[96] Ihsin Tsaiyun Phillips. 1996. User's reference manual for the UW english/technical document image database III. *UW-III English/Technical Document Image database Manual* (1996).

[97] Devashish Prasad, Ayan Gadpal, Kshitij Kapadni, Manish Visave, and Kavita Sultanpure. 2020. CascadeTabNet: An approach for end to end table detection and structure recognition from image-based documents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 572–573.

[98] P Pyreddy and WB Croft. 1997. Tinti: A system for retrieval in text tables title2.

[99] Shah Rukh Qasim, Hassan Mahmood, and Faisal Shafait. 2019. Rethinking table recognition using graph neural networks. In *2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, September 20-25, 2019*, IEEE, 142–147. DOI:https://doi.org/10.1109/ICDAR.2019.00031

[100] Liang Qiao, Zaisheng Li, Zhanzhan Cheng, Peng Zhang, Shiliang Pu, Yi Niu, Wenqi Ren, Wenming Tan, and Fei Wu. 2021. Lgpma: Complicated table structure recognition with local and global pyramid mask alignment. In *Proceedings of the International Conference on Document Analysis and Recognition*. Springer, 99–114.

[101] Sachin Raja, Ajoy Mondal, and CV Jawahar. 2020. Table structure recognition using top-down and bottom-up cues. In *Proceedings of the European Conference on Computer Vision*. Springer, 70–86.

[102] Sachin Raja, Ajoy Mondal, and CV Jawahar. 2022. Visual understanding of complex table structures from document images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2299–2308.

[103] Susie Xi Rao, Johannes Rausch, Peter H. Egger, and Ce Zhang. 2022. TableParser: Automatic Table Parsing with Weak Supervision from Spreadsheets. In *Proceedings of the Workshop on Scientific Document Understanding Co-Located with 36th AAAI Conference on Artificial Inteligence, SDU@AAAI 2022, Virtual Event, March 1, 2022 (CEUR Workshop Proceedings), CEUR-WS.org*. Retrieved from https://ceur-ws.org/Vol-3164/paper8.pdf

[104] Sheikh Faisal Rashid, Abdullah Akmal, Muhammad Adnan, Ali Adnan Aslam, and Andreas Dengel. 2017. Table recognition in heterogeneous documents using machine learning. In *Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 777–782.

[105] Mohammad Mohsin Reza, Syed Saqib Bukhari, Martin Jenckel, and Andreas Dengel. 2019. Table localization and segmentation using GAN and CNN. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*. IEEE, 152–157.

[106] Pau Riba, Anjan Dutta, Lutz Goldmann, Alicia Fornés, Oriol Ramos, and Josep Lladós. 2019. Table detection in invoice documents by graph neural networks. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 122–127.

[107] Pau Riba, Lutz Goldmann, Oriol Ramos Terrades, Diede Rusticus, Alicia Fornés, and Josep Lladós. 2022. Table detection in business document images by message passing networks. *Pattern Recognit.* 127 (2022), 108641. DOI:https: //doi.org/10.1016/J.PATCOG.2022.108641

[108] Arash Samari, Andrew Piper, Alison Hedley, and Mohamed Cheriet. 2021. Weakly supervised bounding box extraction for unlabeled data in table detection. In *Proceedings of the Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021*. Springer, 339–352.

[109] Sebastian Schreiber, Stefan Agne, Ivo Wolf, Andreas Dengel, and Sheraz Ahmed. 2017. Deepdesrt: Deep learning for detection and structure recognition of tables in document images. In *Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1162–1167.

[110] Wonkyo Seo, Hyung Il Koo, and Nam Ik Cho. 2015. Junction-based table detection in camera-captured document images. *International Journal on Document Analysis and Recognition* 18, 1 (2015), 47–57.

[111] Faisal Shafait and Ray Smith. 2010. Table detection in heterogeneous documents. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. 65–72.

[112] Asif Shahab, Faisal Shafait, Thomas Kieninger, and Andreas Dengel. 2010. An open approach towards the benchmarking of table structure recognition systems. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. 113–120.

[113] Tahira Shehzadi, Khurram Azeem Hashmi, Didier Stricker, Marcus Liwicki, and Muhammad Zeshan Afzal. 2023. Towards end-to-end semi-supervised table detection with deformable transformer. In *Proceedings of the International Conference on Document Analysis and Recognition*. Springer, 51–76.

[114] Xinyi Shen, Lingjun Kong, Yunchao Bao, Yaowei Zhou, and Weiguang Liu. 2022. RCANet: A rows and columns aggregated network for table structure recognition. In *Proceedings of the 2022 3rd Information Communication Technologies Conference (ICTC)*. IEEE, 112–116.

[115] Shoaib Ahmed Siddiqui, Imran Ali Fateh, Syed Tahseen Raza Rizvi, Andreas Dengel, and Sheraz Ahmed. 2019. DeepTabStR: Deep learning based table structure recognition. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1403–1409.

[116] Shoaib Ahmed Siddiqui, Pervaiz Iqbal Khan, Andreas Dengel, and Sheraz Ahmed. 2019. Rethinking semantic segmentation for table structure recognition in documents. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1397–1402.

[117] Shoaib Ahmed Siddiqui, Muhammad Imran Malik, Stefan Agne, Andreas Dengel, and Sheraz Ahmed. 2018. DeCNT: Deep deformable CNN for table detection. *IEEE Access* 6 (2018), 74151–74161. DOI:https://doi.org/10.1109/ACCESS. 2018.2880211

[118] Grigori Sidorov, Helena Gómez-Adorno, Ilia Markov, David Pinto, and Nahun Loya. 2015. Computing text similarity using tree edit distance. In *Proceedings of the 2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) Held Jointly with 2015 5th World Conference on Soft Computing (WConSC)*. 1–4. DOI:https://doi.org/10.1109/NAFIPS-WConSC.2015.7284129

[119] Noah Siegel, Nicholas Lourie, Russell Power, and Waleed Ammar. 2018. Extracting scientific figures with distantly supervised neural networks. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. 223–232.

[120] Brandon Smock, Rohith Pesala, and Robin Abraham. 2023. GriTS: Grid table similarity metric for table structure recognition. In *Proceedings of the International Conference on Document Analysis and Recognition*. Springer, 535–549.

[121] Brandon Smock, Rohith Pesala, and Robin Abraham. 2022. PubTables-1M: Towards comprehensive table extraction from unstructured documents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4634–4642.

[122] Ningning Sun, Yuanping Zhu, and Xiaoming Hu. 2019. Faster R-CNN based table detection combining corner locating.
In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1314–1319.

[123] Richard Szeliski. 2010. *Computer Vision: Algorithms and Applications*. Springer Science and Business Media.

[124] Chris Tensmeyer, Vlad I Morariu, Brian Price, Scott Cohen, and Tony Martinez. 2019. Deep splitting and merging for table structure decomposition. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 114–121.

[125] Nazgul Toiganbayeva, Mahmoud SalahEldin Kasem, Galymzhan Abdimanap, Kairat Bostanbekov, Abdelrahman Abdallah, Anel Alimova, and Daniyar B. Nurseitov. 2022. KOHTD: Kazakh offline handwritten text dataset. *Signal Process. Image Commun.* 108 (2022), 116827. DOI:https://doi.org/10.1016/J.IMAGE.2022.116827

[126] Mark Traquair, Ertugrul Kara, Burak Kantarci, and Shahzad Khan. 2019. Deep learning for the detection of tabular information from electronic component datasheets. In *Proceedings of the 2019 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 1–6.

[127] Scott Tupaj, Zhongwen Shi, C. Hwa Chang, and Hassan Alam. 1996. Extracting tabular information from text files. *EECS Department, Tufts University, Medford, USA* 1 (1996).

[128] Yalin Wang and Jianying Hu. 2002. A machine learning based approach for table detection on the web. In *Proceedings of the 11th International Conference on World Wide Web*. 242–250.

[129] Yalin Wangt, Ihsin T Phillipst, and Robert Haralick. 2001. Automatic table ground truth generation and a backgroundanalysis-based table structure extraction method. In *Proceedings of the 6th International Conference on Document Analysis and Recognition*. IEEE, 528–532.

[130] Shengkai Wu, Jinrong Yang, Xinggang Wang, and Xiaoping Li. 2022. IoU-Balanced loss functions for single-stage object detection. *Pattern Recognit. Lett.* 156 (2022), 96–103. DOI:https://doi.org/10.1016/J.PATREC.2022.01.021

[131] Bin Xiao, Murat Simsek, Burak Kantarci, and Ala Abu Alkheir. 2022. Table structure recognition with conditional attention. arXiv:2203.03819. Retrieved from https://arxiv.org/abs/2203.03819

[132] Bin Xiao, Murat Simsek, Burak Kantarci, and Ala Abu Alkheir. 2023. Revisiting table detection datasets for visually rich documents. arXiv:2305.04833. Retrieved from https://arxiv.org/abs/2305.04833

[133] Wen Xu, Julian Jang-Jaccard, Amardeep Singh, Yuanyuan Wei, and Fariza Sabrina. 2021. Improving performance of autoencoder-based network anomaly detection on NSL-KDD dataset. *IEEE Access* 9 (2021), 140136–140146. DOI:https: //doi.org/10.1109/ACCESS.2021.3116612

[134] Wenyuan Xue, Qingyong Li, and Dacheng Tao. 2019. ReS2TIM: Reconstruct syntactic structures from table images. In *Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 749–755.

[135] Fan Yang, Lei Hu, Xinwu Liu, Shuangping Huang, and Zhenghui Gu. 2023. A large-scale dataset for end-to-end table recognition in the wild. *Scientific Data* 10, 1 (2023), 110.

[136] Jing Yang and Guanci Yang. 2018. Modified convolutional neural network based on dropout and the stochastic gradient descent optimizer. *Algorithms* 11, 3 (2018), 28.

[137] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. 2018. Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine* 13, 3 (2018), 55–75.

[138] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. 2019. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4471–4480.

[139] Richard Zanibbi, Dorothea Blostein, and James R Cordy. 2004. A survey of table recognition. *Document Analysis and Recognition* 7, 1 (2004), 1–16.

[140] Daqian Zhang, Ruibin Mao, Runting Guo, Yang Jiang, and Jing Zhu. 2023. YOLO-table: Disclosure document table detection with involution. *Int. J. Document Anal. Recognit.* 26, 1 (2023), 1–14. DOI:https://doi.org/10.1007/S10032-02200400-Z

[141] Xi-wen Zhang, Michael R Lyu, and Guo-zhong Dai. 2007. Extraction and segmentation of tables from Chinese ink documents based on a matrix model. *Pattern Recognition* 40, 7 (2007), 1855–1867.

[142] Zixing Zhang, Jürgen Geiger, Jouni Pohjalainen, Amr El-Desoky Mousa, Wenyu Jin, and Björn Schuller. 2018. Deep learning for environmentally robust speech recognition: An overview of recent developments. *ACM Transactions on Intelligent Systems and Technology* 9, 5 (2018), 1–28.

[143] Zhenrong Zhang, Jianshu Zhang, Jun Du, and Fengren Wang. 2022. Split, Embed and Merge: An accurate table structure recognizer. *Pattern Recognit.* 126 (2022), 108565. DOI:https://doi.org/10.1016/J.PATCOG.2022.108565

[144] Xinyi Zheng, Doug Burdick, Lucian Popa, Peter Zhong, and Nancy Xin Ru Wang. 2021. Global table extractor (GTE): A framework for joint table identification and cell structure recognition using visual context. In *Proceedings of the IEEE/CVF Winter Conference for Applications in Computer Vision (WACV).*

[145] Xinyi Zheng, Douglas Burdick, Lucian Popa, Xu Zhong, and Nancy Xin Ru Wang. 2021. Global table extractor (GTE): A framework for joint table identification and cell structure recognition using visual context. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 697–706.

[146] Xu Zhong, Elaheh ShafieiBavani, and Antonio Jimeno Yepes. 2020. Image-based table recognition: Data, model, and evaluation. In *Proceedings of the European Conference on Computer Vision*. Springer, 564–580.

[147] Yajun Zou and Jinwen Ma. 2020. A deep semantic segmentation model for image-based table structure recognition. In *Proceedings of the 2020 15th IEEE International Conference on Signal Processing (ICSP)*. IEEE, 274–280.
[148] Arthur Zucker, Younes Belkada, Hanh Vu, and Van Nam Nguyen. 2021. ClusTi: Clustering method for table structure recognition in scanned images. *Mobile Networks and Applications* 26, 4 (2021), 1765–1776.

Received 13 December 2022; revised 11 February 2024; accepted 2 April 2024