TABLE STRUCTURE RECOGNITION

A Project report submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR'S OF TECHNOLOGY

in

Computer Science and Engineering/Electronics and Communication Engineering

Kaif Ahmed(112015083),

Rachit Jain(112016028),

Prabhat Todi(112016023),

Akash Malvathkar(112016017)

Under the Supervision of: Dr. Bhupendra Singh

Semester: V

Indian Institute of Information And Technology Pune

Ambegaon BK, Pune, Maharashtra 411041

Dec 2022

BONAFIDE CERTIFICATE

This is to certify that the project report entitled **“Table Structure Recognition”** submitted by **Kaif Ahmed** bearing the **MIS No:112015083**, **Rachit Jain** bearing the **MIS No:112016028**, **Prabhat Todi** bearing the **MIS No:112016023**, **Akash Malvathkar** bearing the **MIS No:112016017** in completion of his project work under the guidance of **Bhupendra Singh** is accepted for the project report submission in partial fulfillment of the requirements for the award of the degree of Bachelors of Technology in Computer Science and Engineering in the Department of Computer Science and Engineering , Indian Institute of Information Technology, Pune, during the academic year 2022-23.

Dr Bhupendra Singh
Project Guide
Assistant Professor
Department of CSE
IIIT Pune

Dr. Sanjeev Sharma
Head of the Department,CSE
Assistant Professor
Department of CSE
IIIT Pune

Project viva-voce held on 15/12/2022

Internal Examiner

External Examiner

ACKNOWLEDGEMENT

This project would not have been possible without the help and cooperation of many. I would like to thank the people who helped me directly and indirectly in the completion of this project work.

First and foremost, I would like to express my gratitude to our beloved director, **Prof. O.G. Kakde,**, for providing his kind support in various aspects.

I would like to express my gratitude to my project guides **Dr. Bhupendra Singh**, Assistant Professor, Department of CSE, for providing excellent guidance, encouragement, inspiration, constant and timely support throughout this M.Tech project.

I would like to express my gratitude to the head of department **Dr. Sanjeev Sharma**, Head of Department CSE and **Dr. Sandeep Mishra**, Head of Department ECE for providing his kind support in various aspects.

I would also like to thank all the faculty members in the Dept. of CSE and my classmates for their steadfast and strong support and engagement with this project.

Abstract

In document images, tables are organized objects that are information-rich. Only a few attempts have been made to recognise the structure of tables, despite the fact that extensive work has been done to localize tables as graphic objects in document pictures. The majority of the work currently in existence on structure recognition relies on the extraction of meta-features from PDF documents or on optical character recognition (OCR) models to extract low-level layout information from images. However, due to the lack of meta-features or mistakes caused by the OCR when there is a substantial variation in table layouts and text structure, these approaches do not generalize effectively. In our research, we concentrate on tables with intricate layouts, substantial material, and independent of meta-features and/or OCR. We make use of CasacdeTabNet, Multi-type-TD-DSR, and TGR-Net after making some changes and then test it on our dataset which we made from combining datasets from ICDAR 2013, ICDAR 2019, and Sci-TSR.

Keywords : Table Structure Recognition, Table detection, Cascade-Tab-Net , Multi-type-TD-DSR , TGR-Net.

Contents

List of Figures

Chapter 1

Introduction

An automatic table information extraction method involves two subtasks of table detection and table structure recognition. In table detection, the region of the image that contains the table is identified while table structure recognition involves identification of the rows and columns to identify individual table cells. To understand the structure of different tables, both the cell spatial and logical locations are of great importance for many applications.
For Evaluation We are considering the four models:-

1. CascadeTabNet
2. Multi-Type-TD-TSR
3. TGRNet

We used Several Datasets to train the model:-

1. ICDAR13
2. ICDAR19
3. PubTabNet
4. Table Bank
5. SciTSR

And to test the model we used the subset of PubTabNet-1M (where M stands for million).

Chapter 2
Motivation

2.1 Objective

To Overcome the challenge of retrieving tabular structures along with its data from images of official documents irrespective of its format.
To Improve the pre-existing model CascadeTabNet mainly at the post-processing module.

2.2 Motivation

Tabular data have been widely used to help people manage and extract important information in many real-world scenarios, including the analysis of financial documents, air pollution indices, and electronic medical records. The world is changing and going digital. The use of digitized documents instead of physical paper-based documents is growing rapidly. These documents contain a variety of table-based information with variations in appearance and layouts.
Although humans can easily understand tables with different layouts and styles, it remains a great challenge for machines to automatically recognize the structure of various tables. Considering the massive amount of tabular data presented in unstructured formats (e.g., image and PDF files). That's why we need an automatic table information extraction method.
The community will significantly benefit from an automatic table recognition system, facilitating large-scale tabular data analysis such as table parsing, patient treatment prediction and credit card fraud detection

Chapter 3

Literature Review

3.1 Data:

The analysis of tabular data in unstructured texts focuses on three issues: I table detection: locating the table table bounding boxes in documents, ii) table structure recognition: solely parsing the structural (row and column arrangement) information of tables, and iii) table recognition: parsing both the structural information and the content of table cells.

The PubTabNet dataset and the EDD model we develop in this paper aim at the image-based table recognition problem. Comparing to other existing datasets for table recognition (e.g. SciTSR3 , Table to Latex , and TIES), PubTabNet has three key advantages:

1. The tables are typeset by the publishers of over 6,000 journals in PMCOA, which offers considerably more diversity in table styles than other table datasets.
2. Cells are categorized into headers and body cells, which is important when retrieving information from tables.
3. The format of targeted output is HTML, which can be directly integrated into web applications. In addition, tables in HTML format are represented as a tree structure.

3.2 Model:

Pre-defined rules and statistical machine learning are used in traditional table detection and identification systems. Deep learning has recently demonstrated outstanding performance in image-based table identification and structure recognition. Hao et al. proposed possible table areas using a set of primitive rules and a convolutional neural network to identify whether the regions included a table. Table identification has also been done using fully convolutional neural networks followed by a conditional random field.

Furthermore, deep neural networks for object identification, such as Faster-RCNN, Mask-R CNN, and YOLO, have been used to recognise tables and segment rows and columns. Furthermore, by storing document pictures as graphs, graph neural networks are employed for table detection and identification.

There are various programmes available to transform text-based PDF tables into structured representations.

However, there has been little research towards image-based table recognition. Xu et al. introduced the first attention-based encoder-decoder for picture captioning. Deng et al. improved it by including a recurrent layer in the encoder to capture long horizontal spatial dependencies in order to translate pictures of mathematical formulae into LATEX format. To convert table pictures into LATEX form, the same model was trained on the Table Latex dataset. The performance of this approach on image-based table recognition is mediocre, as demonstrated in and in our experimental findings.

Chapter 4

Research Gap

We identified the following gaps in the existing approaches of Table Structure Recognition:

- **CascadeTabNet** requires extra pre-processing steps and hence this approach becomes computationally intensive
- **CDec-Net** has deformable convolutions along with a composite backbone which increases the complexity of the overall model and affects its accuracy.
- **TableNet** only works for column detection and not on row detection.
- **DeepDeSRT** is not as accurate as compared to the other models.
- **TGR-Net** has low F1 score tested on PubTabNet-1M

Chapter 5
Methodology

Re-evaluation of the scores reported in the study papers has been done for the four models namely -

- CascadeTabNet
- Multi-Type-TD-DSR
- TGRNet

Using the approach, a comparative study of all the models was conducted. The picture samples are divided as **Tables without Borders** ,**Tables with Partial Borders** and **Tables with Borders** in the datasets

5.1 CascadeTabNet

Cascade RCNN is a multi-stage model that solves the conundrum of high-quality detection in CNNs. A modified HRNet was also used to achieve accurate high-resolution representations and multi-level representations for semantic segmentation and item recognition.

CascadeTabNet is a Cascade mask R-CNN HRNet model with three stages. A backbone, such as a ResNet-50, lacking the last fully connected layer is a component of the image transformation model should include maps

By attaching a segmentation branch, the Cascade R-CNN architecture is extended to the instance segmentation task, as done in the Mask R-CNN.

The CNN HR NetV2p W32 backbone converts the image "I" to feature maps. The "RPN

Head" (Dense Head) forecasts first item ideas for these feature maps. The "Bbox Heads" accept RoI characteristics as input and forecast RoI-wise. Each head produces two predictions, which are represented as bounding box classification scores and box regression points. "B" denotes the bounding boxes predicted by the heads, and the classification scores are not shown in the figure for clarity. The "Mask Head" forecasts the object masks, and "S" signifies a segmentation output.

During the inference, "Bbox Heads" object detections are supplemented by "Mask Head" segmentation masks for all discovered objects.

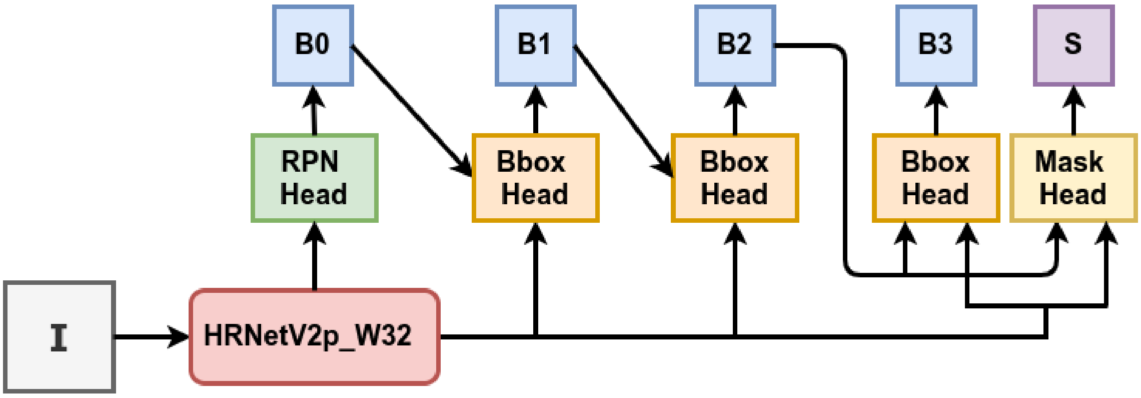


Figure 5.1: CascadeTabNet model architecture

Because text components in documents are quite tiny and the suggested model was used to detect real-world objects in photos, we aim to make the contents more intelligible to the object segmentation model by thickening the text areas and minimizing the blank space regions. We suggest picture alteration strategies to aid the model's learning process. The modified photos are incorporated into the original data-set, increasing the quantity of relevant training data for the model. In the dilation transform, we transform the original image to thicken the black pixel regions.

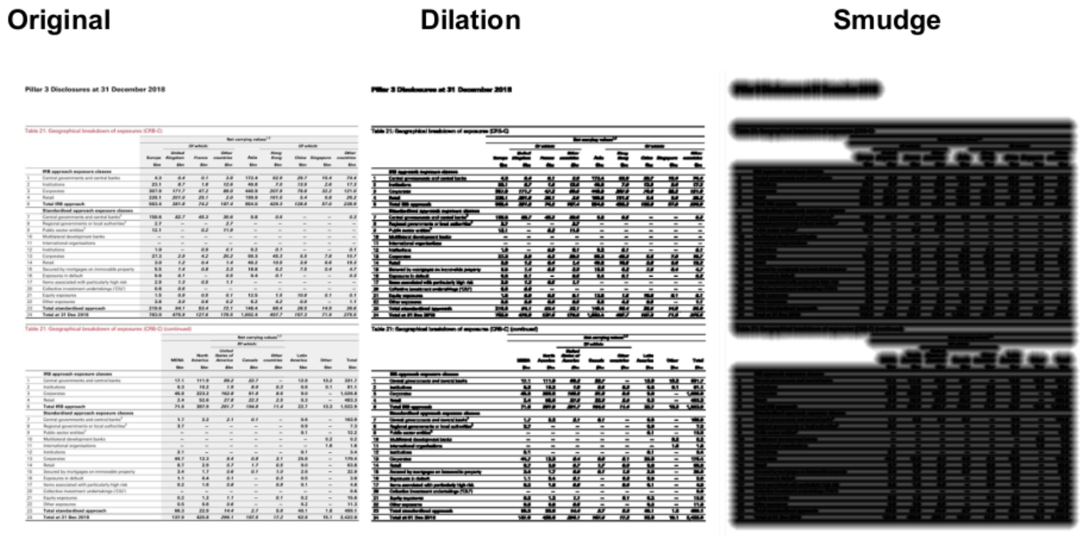


Figure 5.2: Dilation and Smudge Transformation

The Pipeline for the CascadeTabNet Model is as follows -

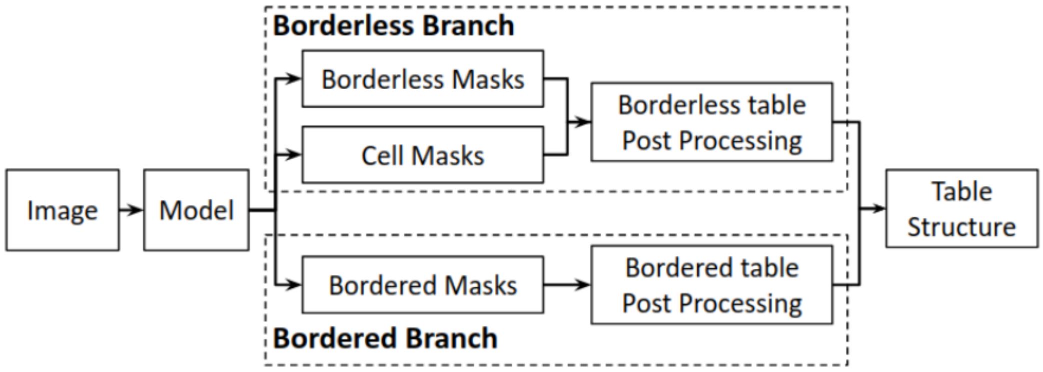
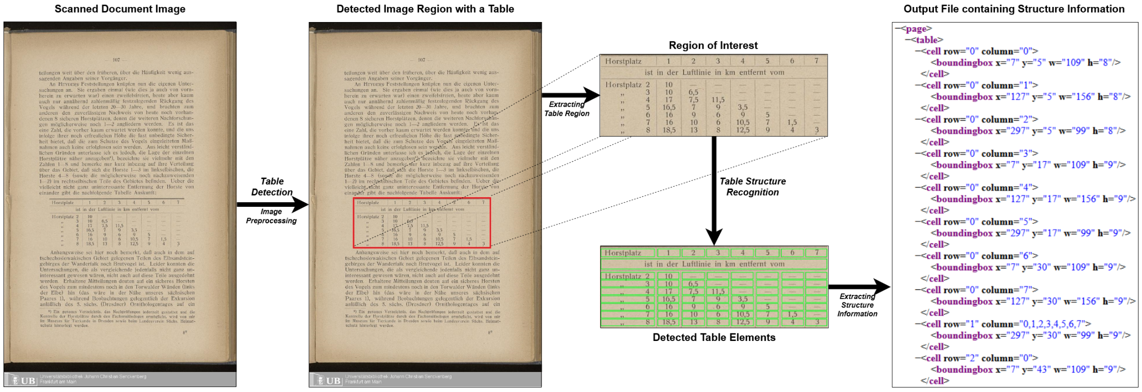


Figure 5.3: CascadeTabNet Pipeline

5.2 Multi-Type-TD-TSR

Without proper alignment of the entire table image, accurate bounding boxes cannot be generated, reducing the overall performance of table representation. As a result, the correct alignment of table images should be regarded as a necessary step in the computer vision task at hand.

Multi-Type-TD-TSR is a multistage pipeline that solves the task of extracting tables from table images and representing their structure end-to-end.


Scanned Document Image
Table Detection
Image Preprocessing
Detected Image Region with a Table
Extracting Table Region
Region of Interest
Table Structure Recognition
Detected Table Elements
Extracting Structure Information
Output File containing Structure Information

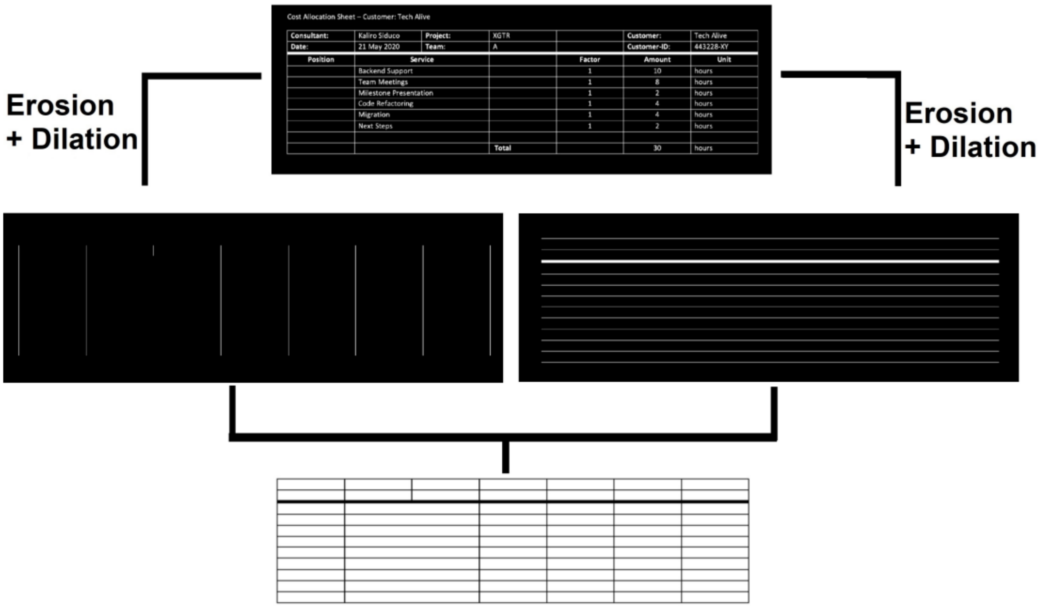
Horstplatz	1	2	3	4	5	6	7
ist in der Luftlinie in km entfernt vom							
Horstplatz 2	10	—	—	—	—	—	—
„ 3	10	6,5	—	—	—	—	—
„ 4	17	7,5	11,5	—	—	—	—
„ 5	16,5	7	9	3,5	—	—	—
„ 6	16	9	6	9	5	—	—
„ 7	16	10	6	10,5	7	1,5	—
„ 8	18,5	13	8	12,5	9	4	3

```
−<page>
  −<table>
    −<cell row="0" column="0">
        <boundingbox x="7" y="5" w="109" h="8"/>
     </cell>
    −<cell row="0" column="1">
        <boundingbox x="127" y="5" w="156" h="8"/>
     </cell>
    −<cell row="0" column="2">
        <boundingbox x="297" y="5" w="99" h="8"/>
     </cell>
    −<cell row="0" column="3">
        <boundingbox x="7" y="17" w="109" h="9"/>
     </cell>
    −<cell row="0" column="4">
        <boundingbox x="127" y="17" w="156" h="9"/>
     </cell>
    −<cell row="0" column="5">
        <boundingbox x="297" y="17" w="99" h="9"/>
     </cell>
    −<cell row="0" column="6">
        <boundingbox x="7" y="30" w="109" h="9"/>
     </cell>
    −<cell row="0" column="7">
        <boundingbox x="127" y="30" w="156" h="9"/>
     </cell>
    −<cell row="1" column="0,1,2,3,4,5,6,7">
        <boundingbox x="297" y="30" w="99" h="9"/>
     </cell>
    −<cell row="2" column="0">
        <boundingbox x="7" y="43" w="109" h="9"/>
     </cell>
```

Figure 5.4: Schematic representation of Multi-Type-TD-TSR

The pipeline is made up of four main modules that were developed independently and can thus be developed in a modular fashion in future work. Unlike similar approaches, the pipeline begins by addressing rotation and noise issues that are common when scanning documents. To localize tables inside images and forward them to TSR, uses a fully data-driven approach based on a Convolutional Neural Network (CNN).

In order to address all three table types, use a deterministic algorithm. The model perform a pre-processing step between TD and TSR to create font and background color invariance, so that the tables only contain black as the font color and white as the background color. The model has been trained on TableBank and ICDAR 2019 Datasets.



Cost Allocation Sheet – Customer: Tech Alive

Consultant:	Kaliro Siduco	Project:	XGTR		Customer:	Tech Alive
Date:	21 May 2020	Team:	A		Customer-ID:	443228-XY
Position	Service			Factor	Amount	Unit
	Backend Support			1	10	hours
	Team Meetings			1	8	hours
	Milestone Presentation			1	2	hours
	Code Refactoring			1	4	hours
	Migration			1	4	hours
	Next Steps			1	2	hours
			Total		30	hours

Figure 5.5: Erosion and Dilation Operations on the Input Images

Dilation - This operation entails filtering an image I with a kernel K, which can be any shape or size but is in this case a rectangle. K has a defined anchor point, in our case the kernel's center. The maximum pixel value overlapped by K is determined as K slides over the image, and the image pixel value at the anchor point position is overwritten by this maximum value. Bright areas of an image are magnified when it is maximized.

Erosion - It works similarly to dilation, but it finds a local minimum over the entire kernel area. As k moves over I, it determines the smallest pixel value overlapped by K and replaces it with the pixel value under the anchor point. In contrast to dilation, erosion causes the image's bright areas to become thinner while the dark areas become larger

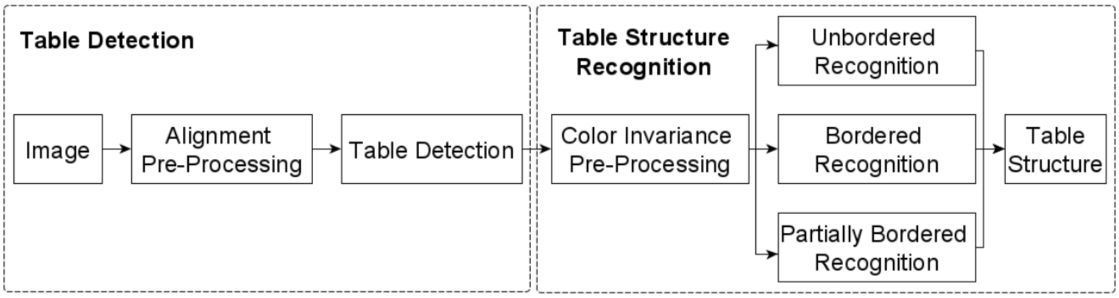


Figure 5.6: Two-Staged Process of Multi-Type-TD-TSR

It is divided into two parts: Table Detection (TD), which processes the entire image, and Table Structure Recognition (TSR), which only processes the recognised sections from TD. Before passing the scan to the next step, a pre-processing function is applied to the scanned document image in the first step to correct the image alignment. In order to perform TD, the aligned image is fed into a ResNext152 model of the type proposed by [9]. The recognised tables are cropped from the image and successively passed to TSR based on the predicted bounding boxes.

To create a color-invariant image, our algorithm performs a second preprocessing step in which the foreground (lines and fonts) is converted to black and the background is converted to white. Three branching options are available in the following step. The first employs an algorithm designed specifically for unbordered tables. The second employs a standard algorithm based on [10] that is tailored to bordered tables.

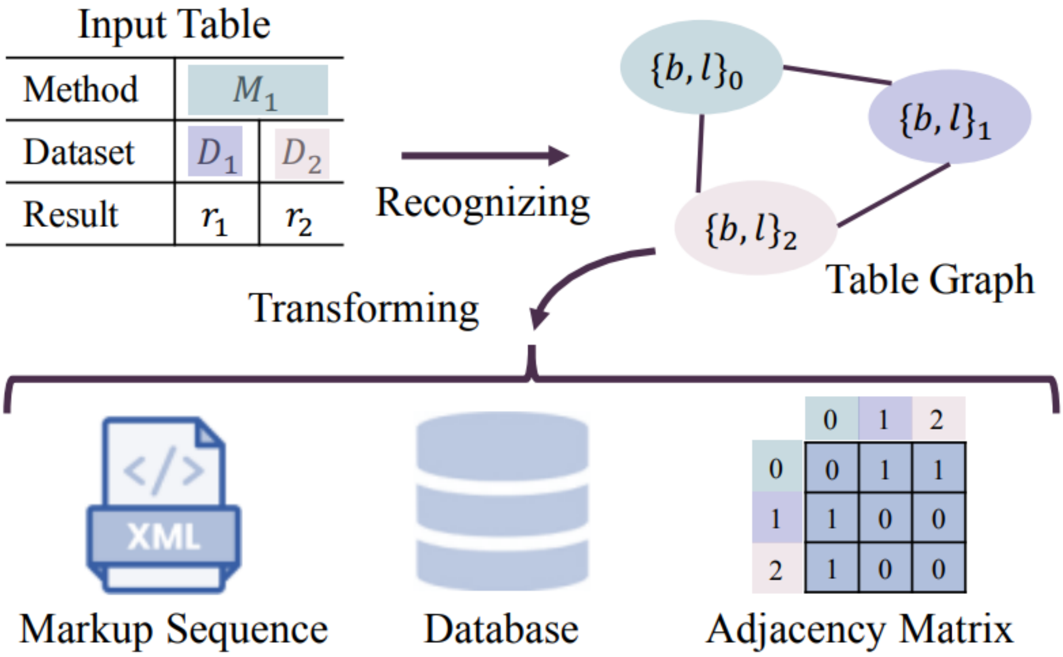
The third option is a hybrid of the first two; it works on partially bordered tables, including fully bordered and fully unbordered tables.

The algorithm is type-agnostic. Finally, using a predefined data structure, the recognised table structure is exported per input table.

5.3 TGRNet

Table Graph is a more sophisticated graph-based table format that we introduce. Each table's structure, in particular, may be represented as a graph: Each node represents a table cell, and the edge connecting two nodes represents their logical relationship on the row and column dimensions, which are coupled to their respective row and column indices. A table cell in the image may be located using the suggested table graph by the position of its pixels, and its essential information can be accessed using the row and column indices.

In this research, we frame the issue of table structure identification as table graph reconstruction, which needs the model to forecast both the spatial and logical placement of the cells. The two primary components of TGRNet are then introduced: **Cell Spatial Location detection** and **Cell Logical Location prediction**.



$$\begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 \end{array}$$

Figure 5.7: Transformation of data from a table graph to other data formats. For clarity, three table cells are depicted. The adjacency matrix shows if two cells are in the same column.

First, we first use a backbone network, such as ResNet-50 with FPN, to extract multi-scale feature representations from the input table picture. We then execute cell spatial location identification and logical location prediction in parallel through two distinct head branches. For cell spatial localization, we first employ a widely established segmentation-based approach to produce a cell segmentation map, from which cells are recognised by calculating bounding boxes of related components. For cell logical location, we use a graph convolutional network to learn the table graph representation and solve it as an ordinal node classification issue. In addition, to solve the imbalance problem in cell logical location prediction, we combine a conventional ordinal regression loss with the focused loss as the objective function.

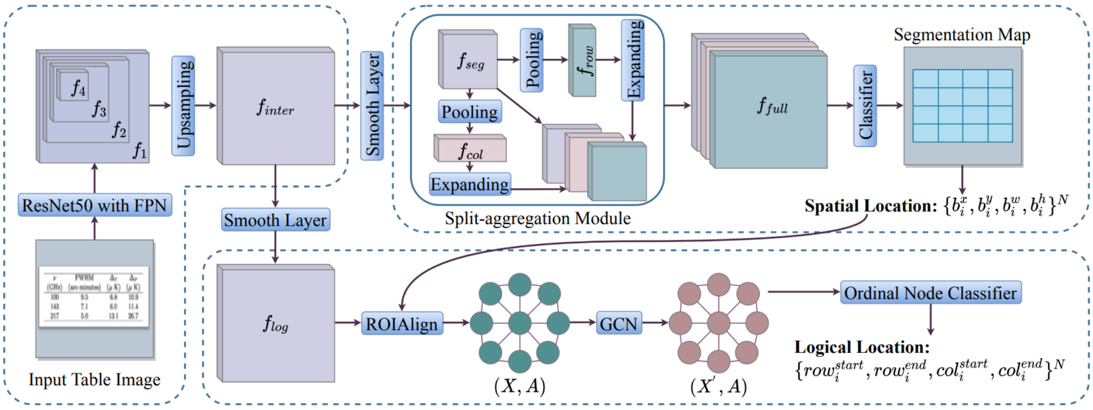


Figure 5.8: Model architecture for TGR-Net

TGRNet is trained independently for cell spatial and logical position prediction due to limited GPU memory. Meanwhile, the picture of the input table is scaled to 800 800 pixels.

- To separate compressed cells during cell spatial position recognition, TGRNet uses the morphological open operation with a 3 3 kernel on the segmentation map.
- Only 8 N edges with greater weights are reserved when the table graph is initialised for cell logical location prediction. The adjustment factor 3 has been adjusted to 10.
- The percentage of positive and negative neighbor relations in ReS2TIM is preserved at 1:4 by sampling cells around each target cell

Chapter 6

Results and Discussion

The Results for each models (As described by the Research Papers) are shown below on the **ICDAR-13 Dataset** and **ICDAR-19 Dataset**.

6.1 CascadeTabNet Model -

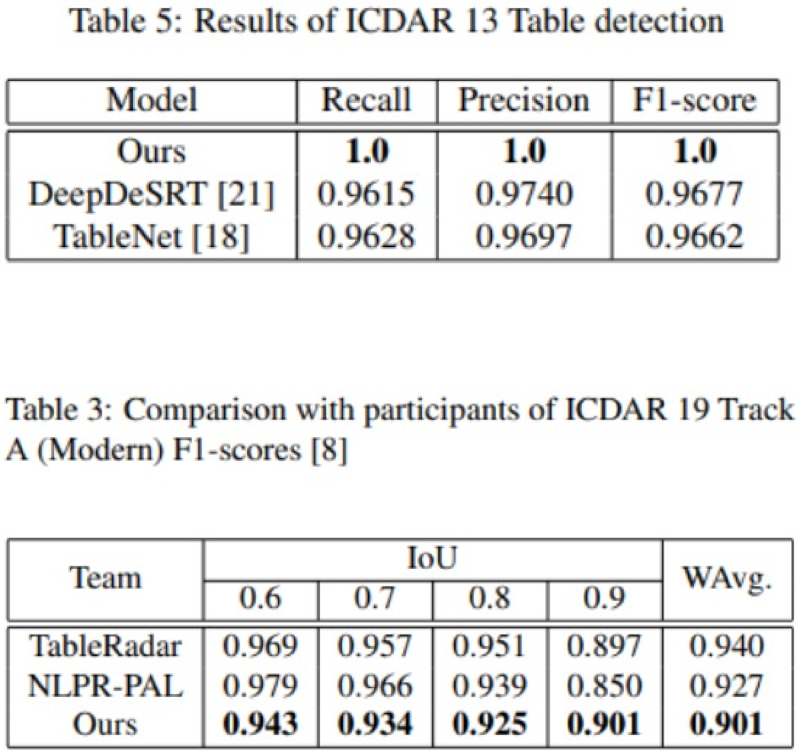


Table 5: Results of ICDAR 13 Table detection

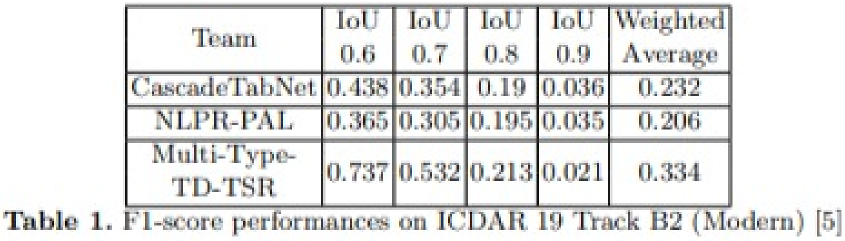
Model	Recall	Precision	F1-score
Ours	**1.0**	**1.0**	**1.0**
DeepDeSRT [21]	0.9615	0.9740	0.9677
TableNet [18]	0.9628	0.9697	0.9662

Table 3: Comparison with participants of ICDAR 19 Track A (Modern) F1-scores [8]

Team	IoU				WAvg.
	0.6	0.7	0.8	0.9	
TableRadar	0.969	0.957	0.951	0.897	0.940
NLPR-PAL	0.979	0.966	0.939	0.850	0.927
Ours	**0.943**	**0.934**	**0.925**	**0.901**	**0.901**

Figure 6.1: CascadeTabNet Results on ICDAR-13 and ICDAR-19 Track B2 Datasets

6.2 Multi-Type TD-TSR Model -



Team	IoU 0.6	IoU 0.7	IoU 0.8	IoU 0.9	Weighted Average
CascadeTabNet	0.438	0.354	0.19	0.036	0.232
NLPR-PAL	0.365	0.305	0.195	0.035	0.206
Multi-Type-TD-TSR	0.737	0.532	0.213	0.021	0.334

Table 1. F1-score performances on ICDAR 19 Track B2 (Modern) [5]

Figure 6.2: Multi-Type-TD-TSR Results on ICDAR-19 Track B2 Datasets

6.3 TGRNet Model -

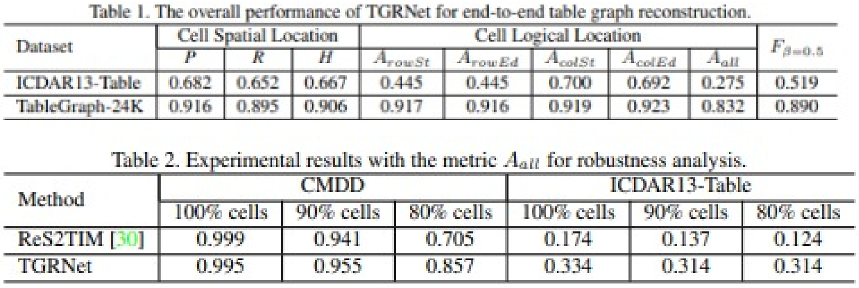


Table 1. The overall performance of TGRNet for end-to-end table graph reconstruction.

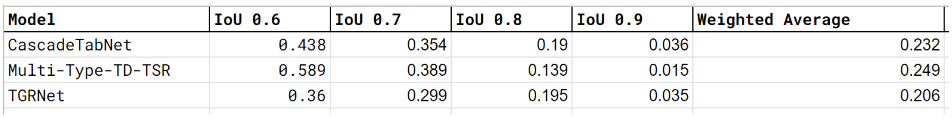
Dataset	Cell Spatial Location			Cell Logical Location					$F_{\beta=0.5}$
	P	R	H	A_{rowSt}	A_{rowEd}	A_{colSt}	A_{colEd}	A_{all}	
ICDAR13-Table	0.682	0.652	0.667	0.445	0.445	0.700	0.692	0.275	0.519
TableGraph-24K	0.916	0.895	0.906	0.917	0.916	0.919	0.923	0.832	0.890

Table 2. Experimental results with the metric A_{all} for robustness analysis.

Method	CMDD			ICDAR13-Table		
	100% cells	90% cells	80% cells	100% cells	90% cells	80% cells
ReS2TIM [30]	0.999	0.941	0.705	0.174	0.137	0.124
TGRNet	0.995	0.955	0.857	0.334	0.314	0.314

Figure 6.3: TGRNet Results on ICDAR-19 Track B2 Datasets

6.4 Final Outcome -



Model	IoU 0.6	IoU 0.7	IoU 0.8	IoU 0.9	Weighted Average
CascadeTabNet	0.438	0.354	0.19	0.036	0.232
Multi-Type-TD-TSR	0.589	0.389	0.139	0.015	0.249
TGRNet	0.36	0.299	0.195	0.035	0.206

Figure 6.4: Final Outcome for all the models on the Prepared Test Dataset.

Chapter 7

Conclusion and Future Scope

7.1 Conclusion

The current instance segmentation-based CNN architectures that were originally trained for objects in natural scene pictures can also be taught for objects in artificial scene images.

Table detection is quite effective. Iterative transfer learning and picture augmentation techniques may also be utilized to efficiently learn from limited amounts of data. The model begins by learning for a broad task and then iteratively learns to perform well on specialized tasks. CascadeTabNet, the suggested system, detects structures inside tables by predicting table cell masks while also utilizing line information. Improving the post-processing modules can improve the overall accuracy of the model. For both tasks, our system outperforms other available datasets.

We frame the issue of table structure identification as table graph reconstruction, which needs the model to forecast both the cell spatial position and the cell logical location. TGRNet is offered as a solution to this problem, which employs a segmentation-based module to identify cell spatial position and solves cell logical location prediction as an ordinal node classification problem.

Experiments on four datasets indicate the usefulness and resilience of TGRNeT.

We demonstrated a multistage pipeline for table identification and table structure recognition that included pre-processing for document alignment and color invariance. For this reason, we classified tables into three groups based on whether they have borders or not. Because big labeled datasets for table structure identification are unavailable, we opted to employ two standard algorithms: the first can handle tables without borders, while the second can handle tables with borders. Furthermore, we integrated these methods to create a third, more traditional table structure identification system that can handle all three types of tables.

7.2 Future Scope

1. None of the models use TableBank Dataset for TSR; images only contain table structure labels in the form of HTML tags. It lacks cell and column coordinates.
2. All of the approaches presented are based on the assumptions of the traditional tabular structure, which is made up of rows and columns intersecting. More complex cell arrangements, such as tables with multiple rows and columns, are not included.
3. The accuracy of these models is due to the presence of annotations for each image sample in the dataset; however, for table structure recognition, such a dataset is not yet available, which is why many data-driven algorithms use transfer learning to circumvent this issue.
4. The TGRNet and Multi-Type-TD-TSR models are still being evaluated and have not undergone extensive testing and experimentation. TGRNet concentrated on the author's locally prepared dataset, TABLE2LATEX-450K, rather than any other publicly available dataset.

Chapter 8

Reference

[1] An approach for end to end table detection and structure recognition from image-based documents - Devashish Prasad, Ayan Gadpal, Kshitij Kapadni, Manish Visave, Kavita Sultanpure Pune Institute of Computer Technology, India.

[2] T. Kasar, P. Barlas, S. Adam, C. Chatelain, and T. Paquet. Learning to detect tables in scanned document images using line information. In 2013 12th International Conference on Document Analysis and Recognition

Research Paper: https://arxiv.org/ftp/arxiv/papers/2004/2004.12629.pdf

[3] Complicated table structure recognition - Zewen Chi, Heyan Huang, Heng-Da Xu, Houjin Yu, Wanxuan Yin, and Xian-Ling Mao.

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

Research Paper: https://arxiv.org/pdf/2106.10598v3.pdf

[5] Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence.

[6] Seo, W., Koo, H.I., Cho, N.I.: Junction-based table detection in camera-captured document images. International Journal on Document Analysis and Recognition.

Research Paper: https://www.researchgate.net/publication/351840889$_{M}ulti - Type -$
$TD - TSR_{-} -_{E} xtracting_{T}ables_{f}rom_{D}ocument_{I}mages_{u}sing_{aM}ultistage_{P}ipeline$
$_{f}or_{T}able_{D}etection_{a}nd_{T}able_{S}tructure_{R}ecognition_{f}rom_{O}CR_{t}o_{S}tructured_{T}able_{R}epresentations$